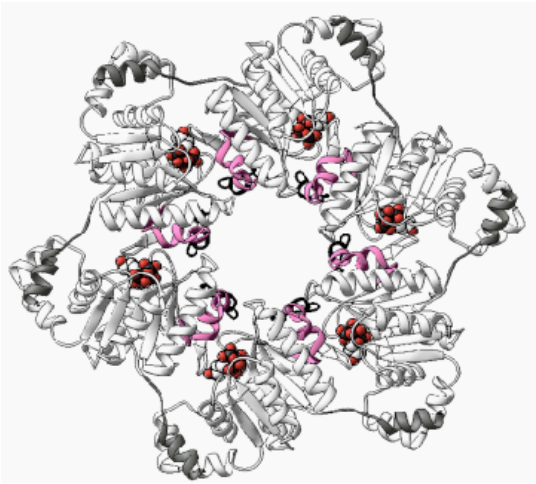
Tertiary Structure Prediction and Modelling

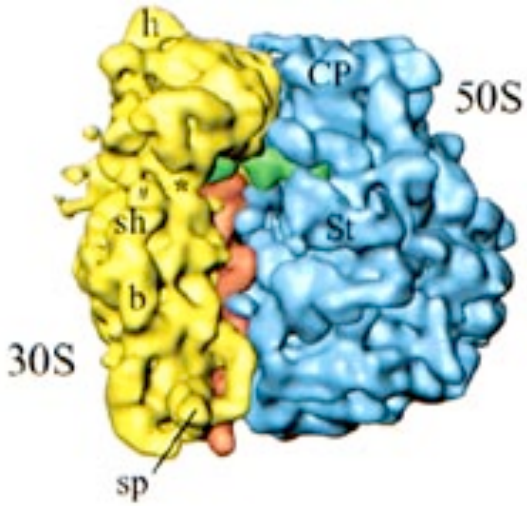
- Tertiary Structure and Domains
- Tertiary Structure Prediction
- Comparative Homology Modelling
- One Practical example: Retinoschisin
- Model evaluation
- Threading

Protein Structure Determination

- High-resolution structure determination
  - X-ray crystallography
  - Nuclear magnetic resonance (NMR)

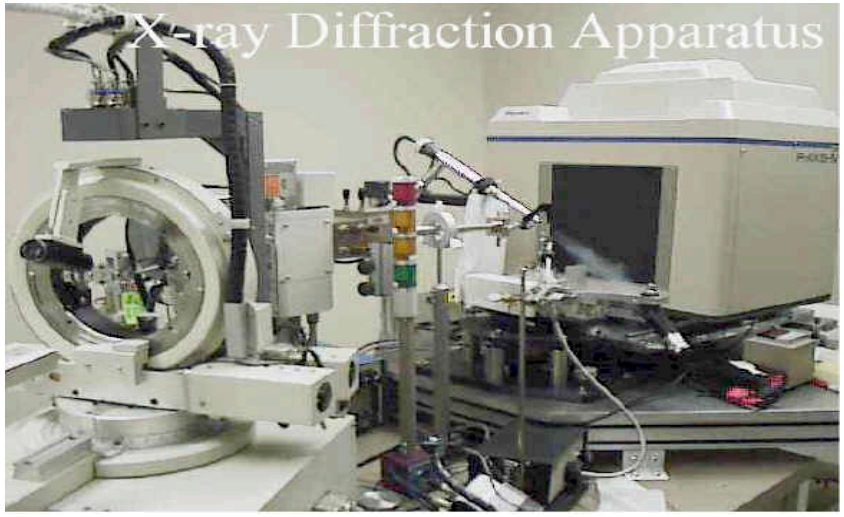


- Low-resolution structure determination
  - Cryo-EM (electron-microscropy) ~10-15Å



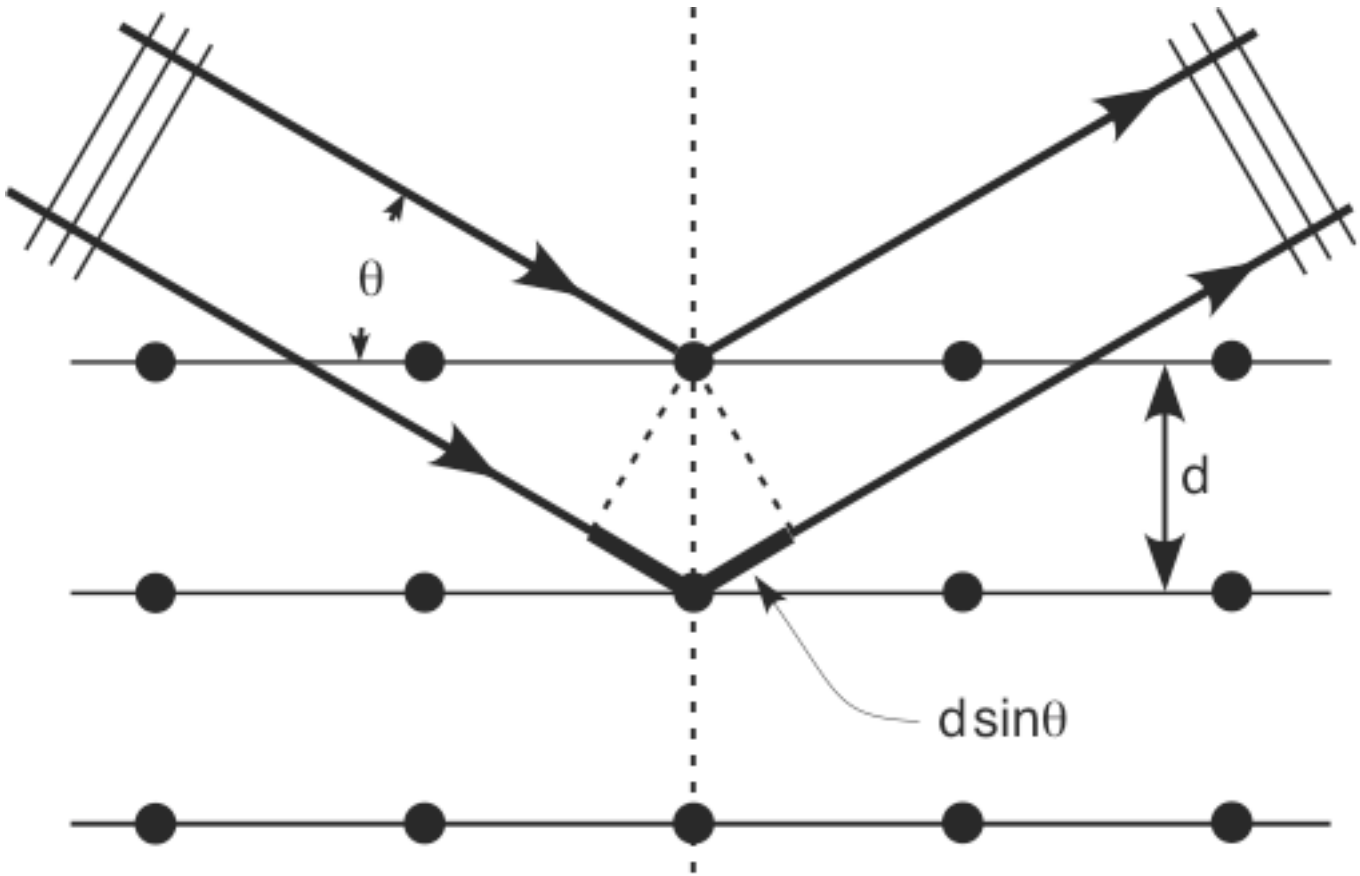
X-ray crystallography

- most accurate
- An extremely pure protein sample is needed.
- The protein sample must form crystals that are relatively large without flaws.
  Generally the biggest problem.
- Many proteins aren't amenable to crystallisation at all (i.e., proteins that do their work inside of a cell membrane).
- ~$100K per structure



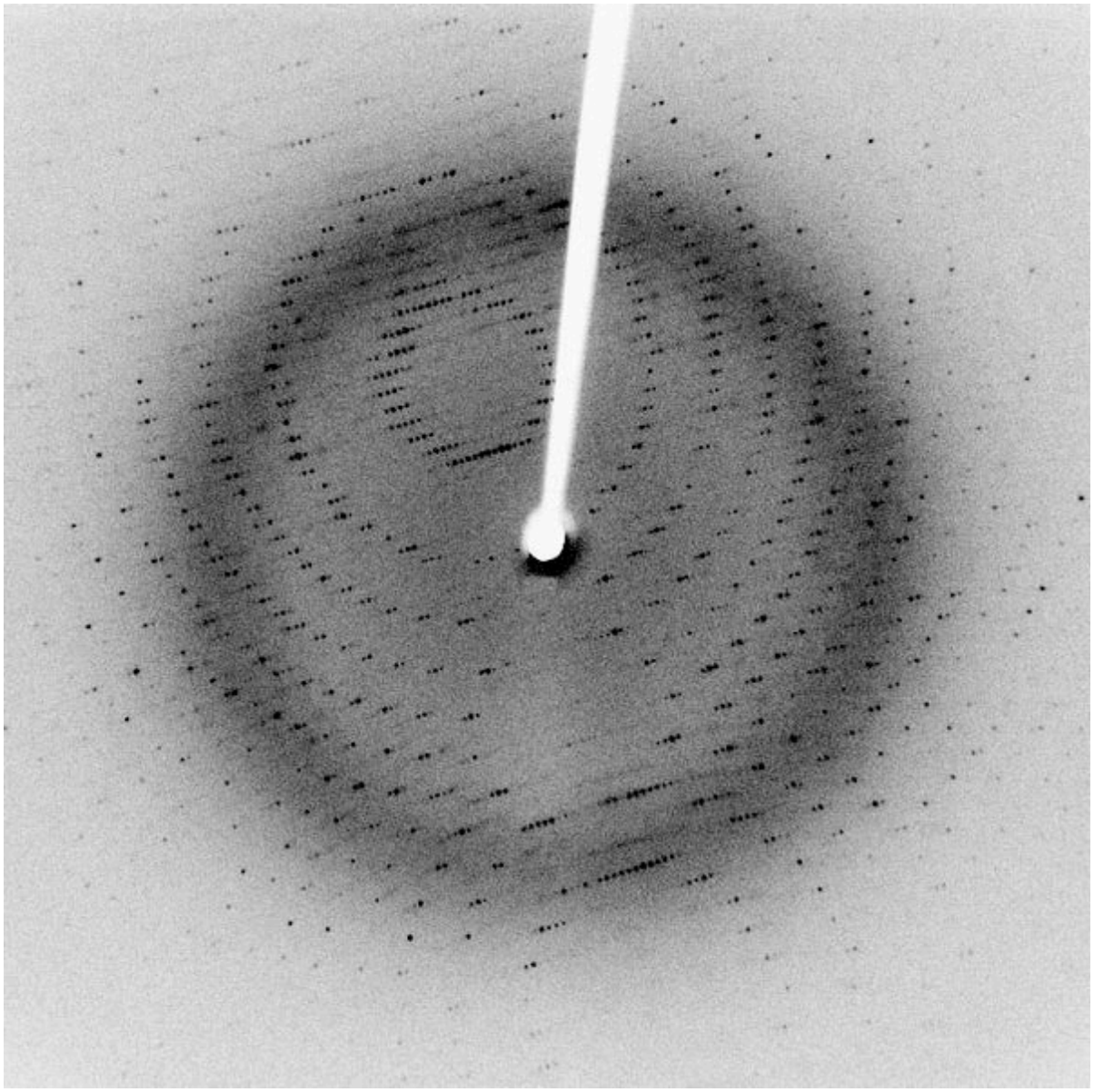
Bragg's Law

$$2d\sin\theta = n\lambda$$

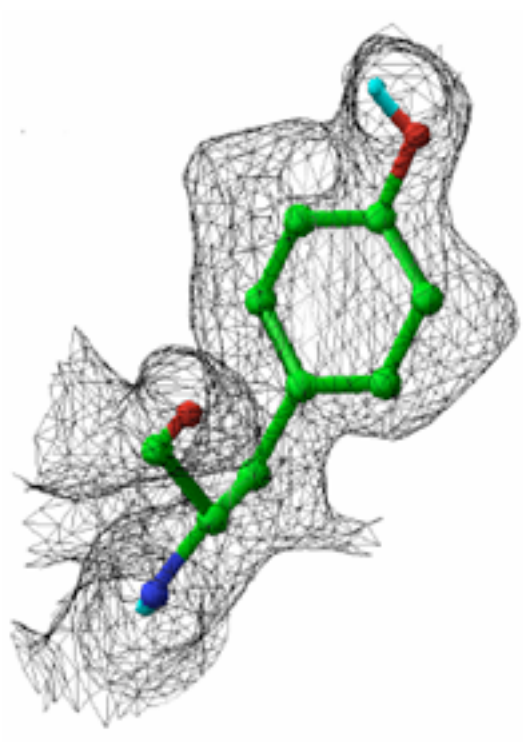


intensity is recorded, not phase information

Diffraction pattern (enzyme)



Electron Density Map and Molecular Replacement

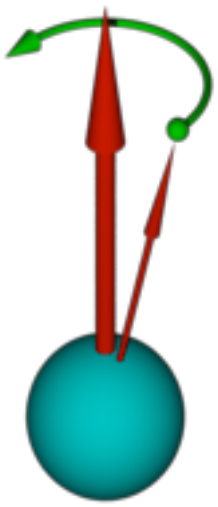


https://solve.lanl.gov/Resolve/html_resolve_manual/resolve_examples.htm

Nuclear Magnetic Resonance

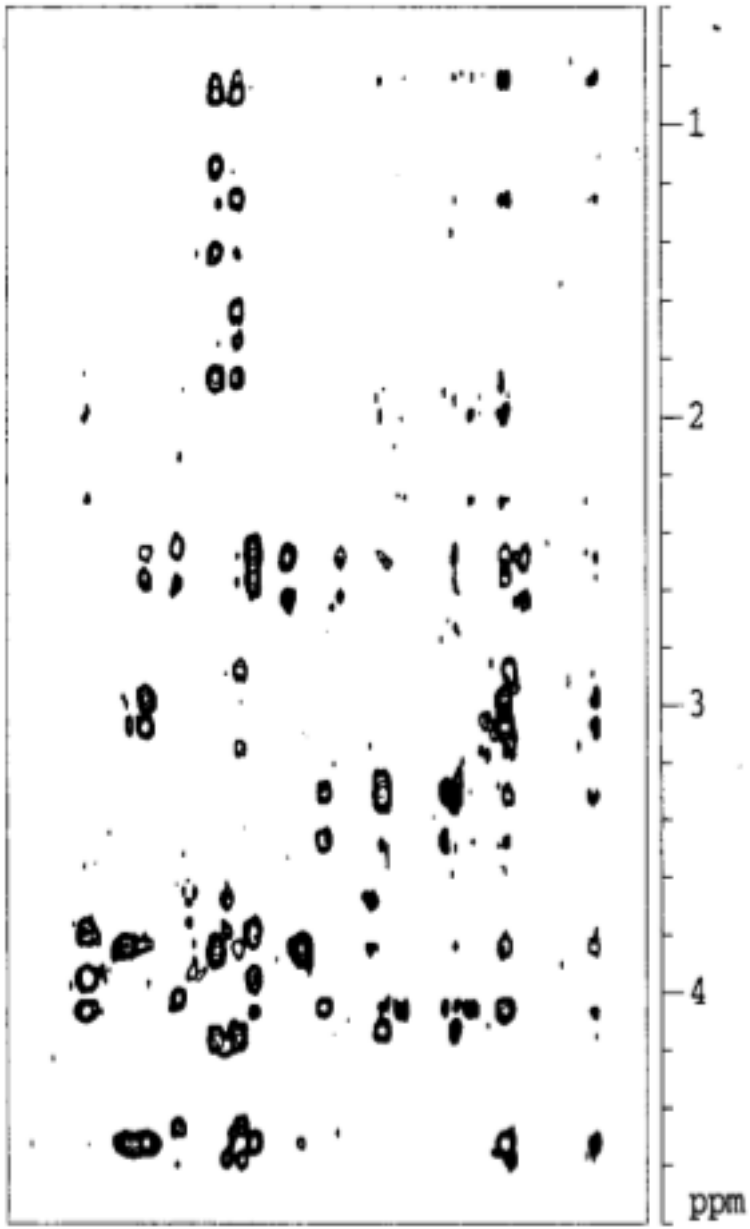
- Dynamics / kinetics
- No need for crystals
- Limited to small-medium size, soluble proteins only

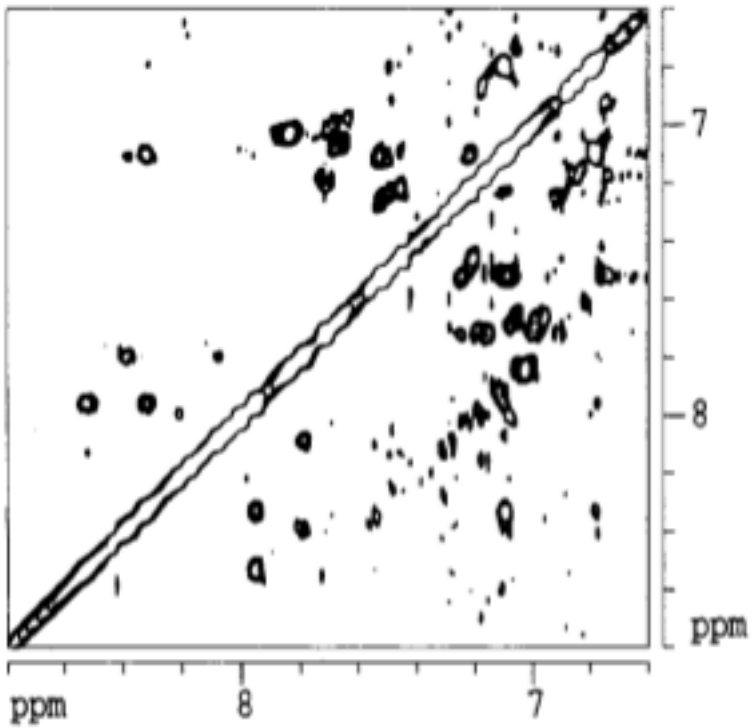
Macroscopic Magnetisation



Proton	Chemical Shift / ppm
H_3C–	0,9
$H_3CCR{=}C$	1,6
H_3C–Ar	2,3
H_3C–CO–R	2,2
H_3C–OR	3,3
R_2C–CH_2–CR_2	1,4
–C–CH_2–Cl	3,6
R–CO_2H	9 bis 13
ROH	0,5 bis 4,5
H–-Ar	7,2

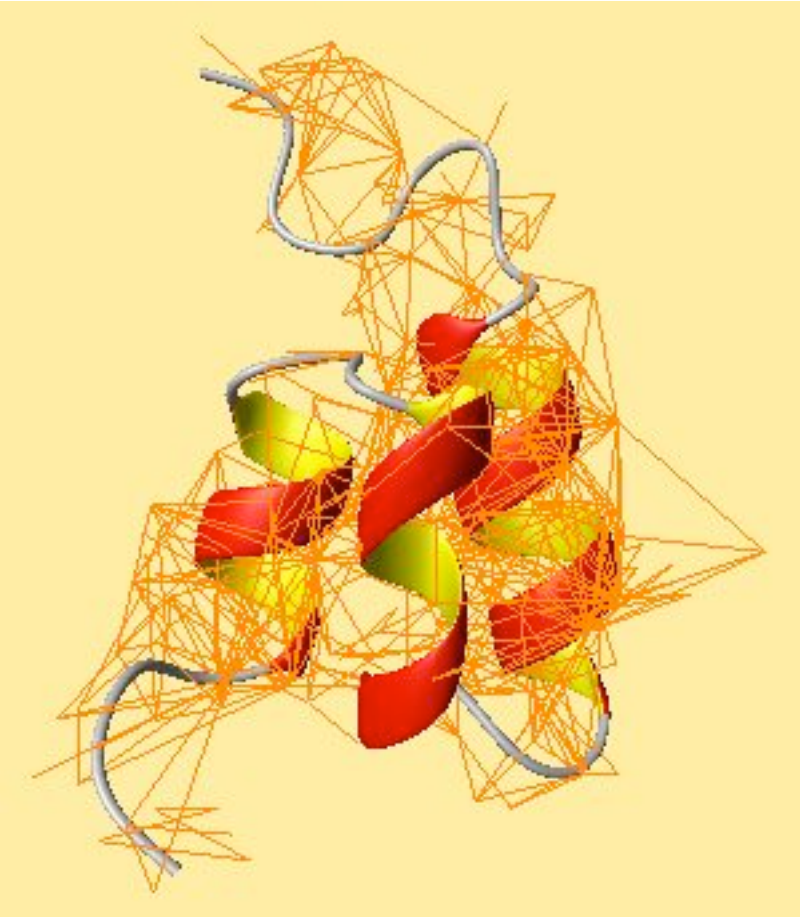
2D Spectra





Distance Geometry

Protons		r_{NMR}	r_{MD}	r_C
				Backbone
N^{68}—$C^{\alpha}H$	N^{68}—NH	2.7–3.3	3.0	3.0
P^{69}—$C^{\alpha}H$	G^{70}—NH	2.2–2.8	2.1	2.2
G^{70}—$C^{\alpha}H$	G^{70}—NH	2.4–3.0	2.6	2.8
G^{70}—NH	D^{71}—NH	2.7–3.3	3.2	3.3
G^{70}—$C^{\alpha}H$	D^{71}—NH	2.7–3.3	3.1	2.5
D^{71}—$C^{\alpha}H$	D^{71}—NH	2.3–2.9	3.0	2.9
D^{71}—NH	Y^{72}—NH	2.6–3.2	2.9	2.7
D^{71}—$C^{\alpha}H$	Y^{72}—NH	3.0–4.0[a]	3.2	3.6
Y^{72}—$C^{\alpha}H$	Y^{72}—NH	1.9–2.5	2.5	3.0
Y^{72}—$C^{\alpha}H$	G^{73}—NH	2.2–2.8	2.6	2.3
G^{73}—$C^{\alpha}H$	G^{73}—NH	2.3–2.9	2.6	2.4

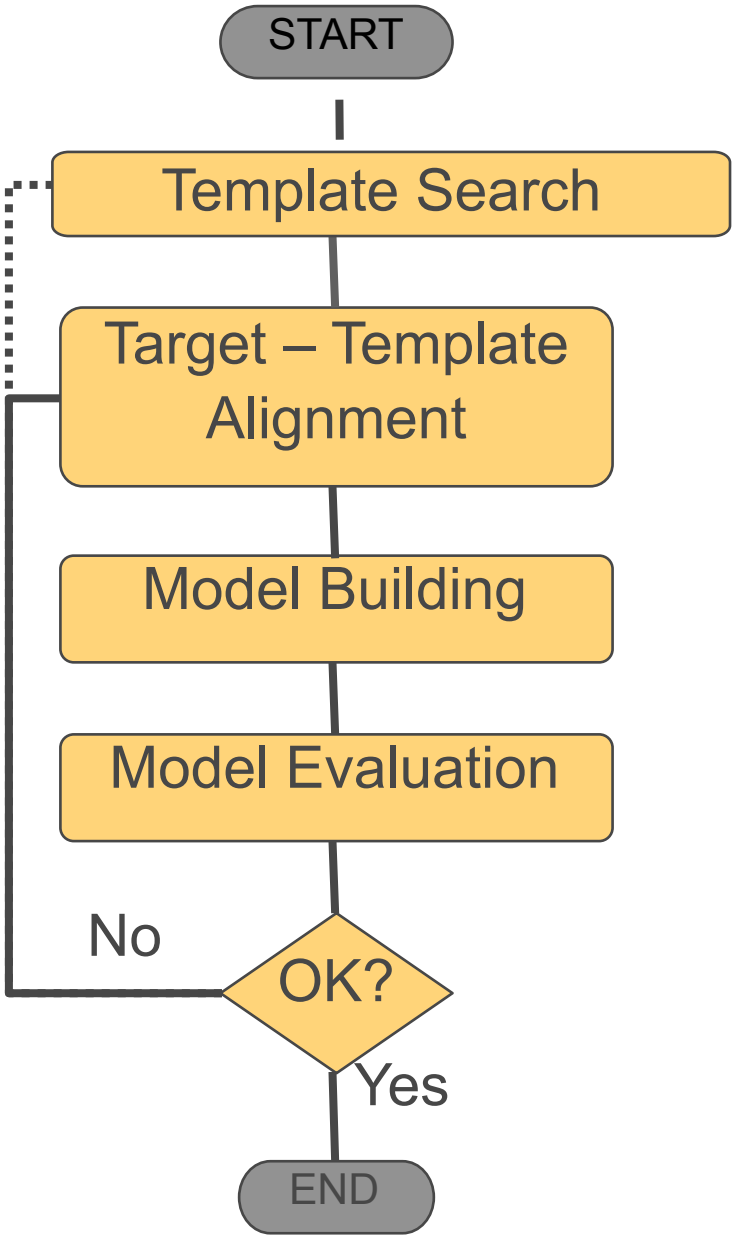


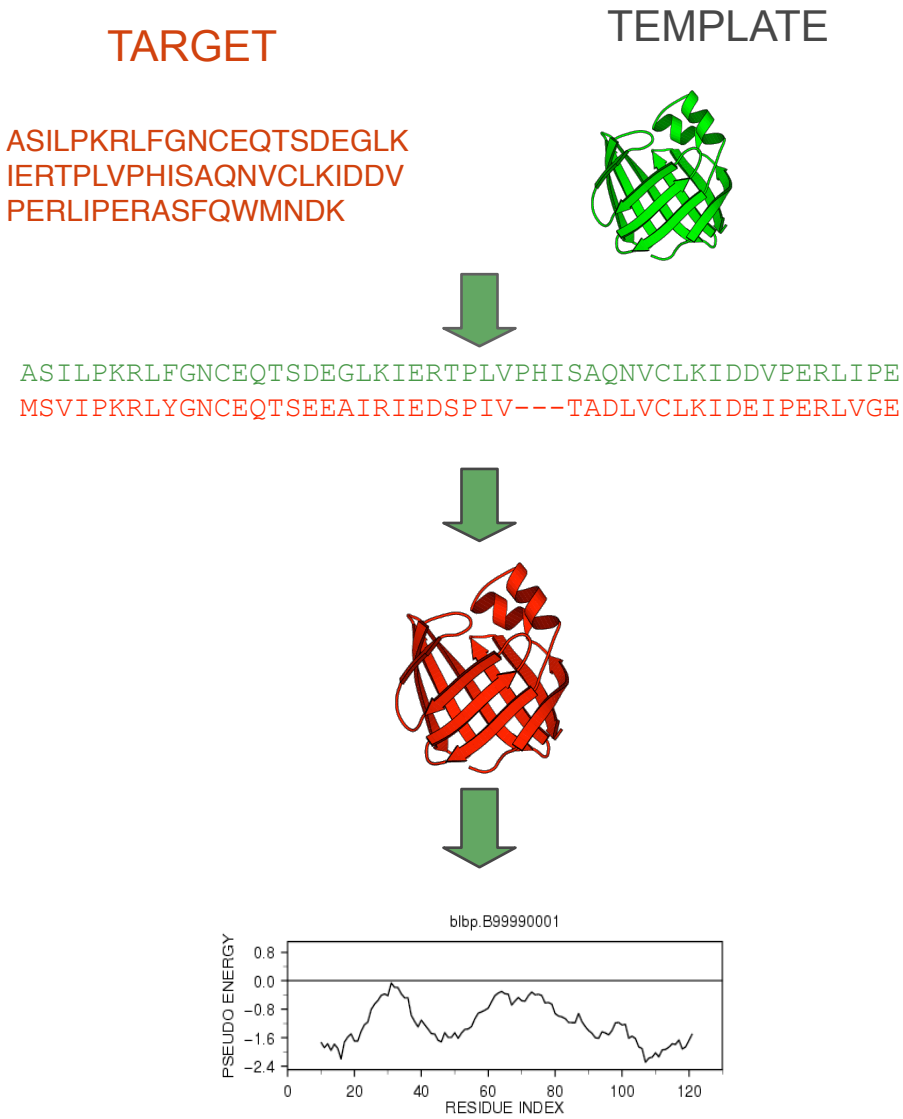
http://www.cs.duke.edu/brd/Teaching/Previous/Bio/dyana.jpg

Comparative Modelling

- Homology modelling
  - identification of homologous proteins through sequence alignment
  - structure prediction through placing residues into "corresponding" positions of homologous structure models
- Protein threading
  - make structure prediction through identification of "good" sequence-structure fit

Steps in Comparative Protein Structure Modeling





A. Šali, Curr. Opin. Biotech. 6, 437, 1995.
R. Sánchez & A. Šali, Curr. Opin. Str. Biol. 7, 206, 1997.
M. Marti et al. Ann. Rev. Biophys. Biomolec. Struct., 29, 291, 2000.

Model Building

1. Identify cores and loops: conserved segments are cores, otherwise loops

2. Core modeling: copy backbone coordinates from the homologous one with know structure

3. Loop modelling

4. Side chain modelling

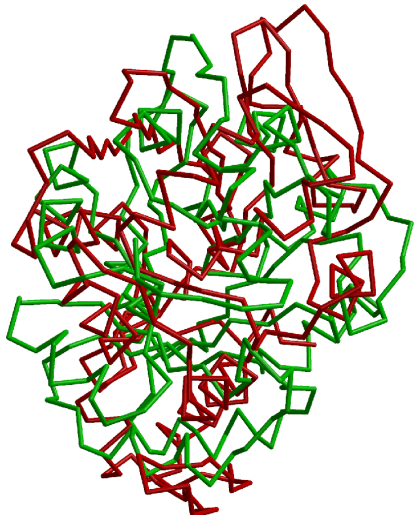
Typical errors in comparative models

MODEL

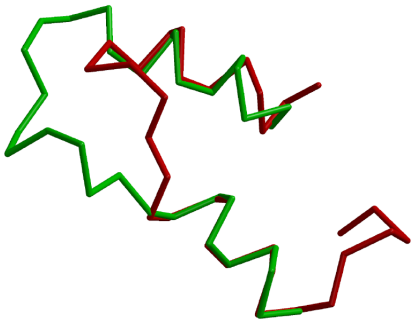
X-RAY

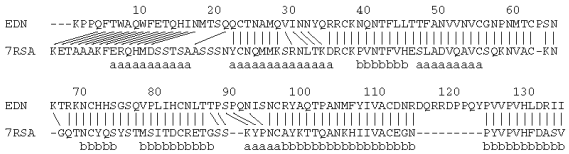
TEMPLATE

Incorrect template

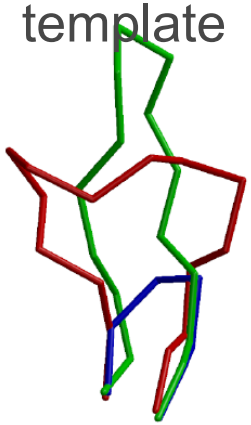


Misalignment

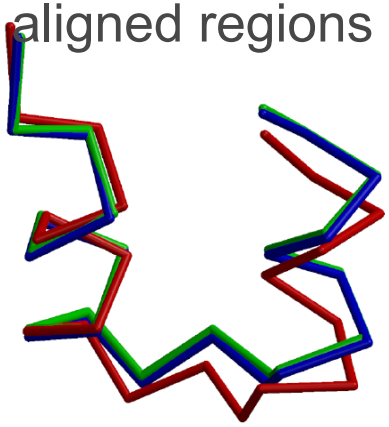




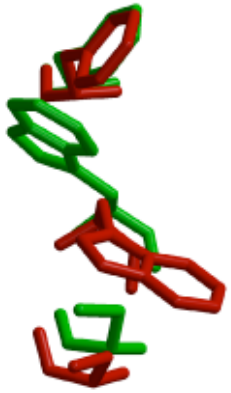
Region without a template



Distortion/shifts in aligned regions



Sidechain packing



Marti-Renom et al. Annu.Rev.Biophys.Biomol.Struct. 29, 291-325, 2000.

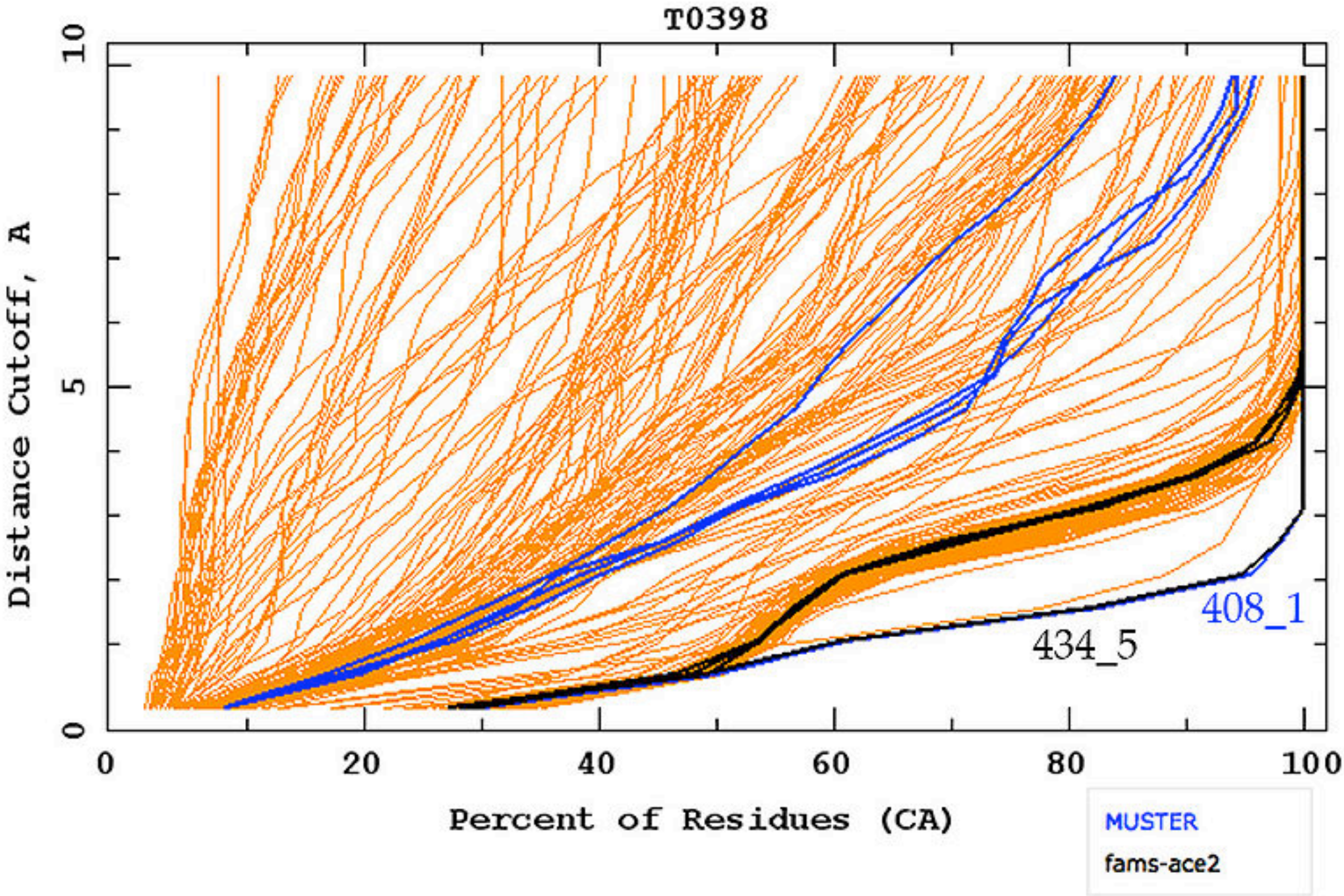
Evaluating Structural Alignments

to be considered:

1. Number of amino acid correspondences created.
2. RMSD of corresponding amino acids
3. Percent identity in aligned residues
4. Number of gaps introduced
5. Size of the two proteins
6. Conservation of known active site environments …

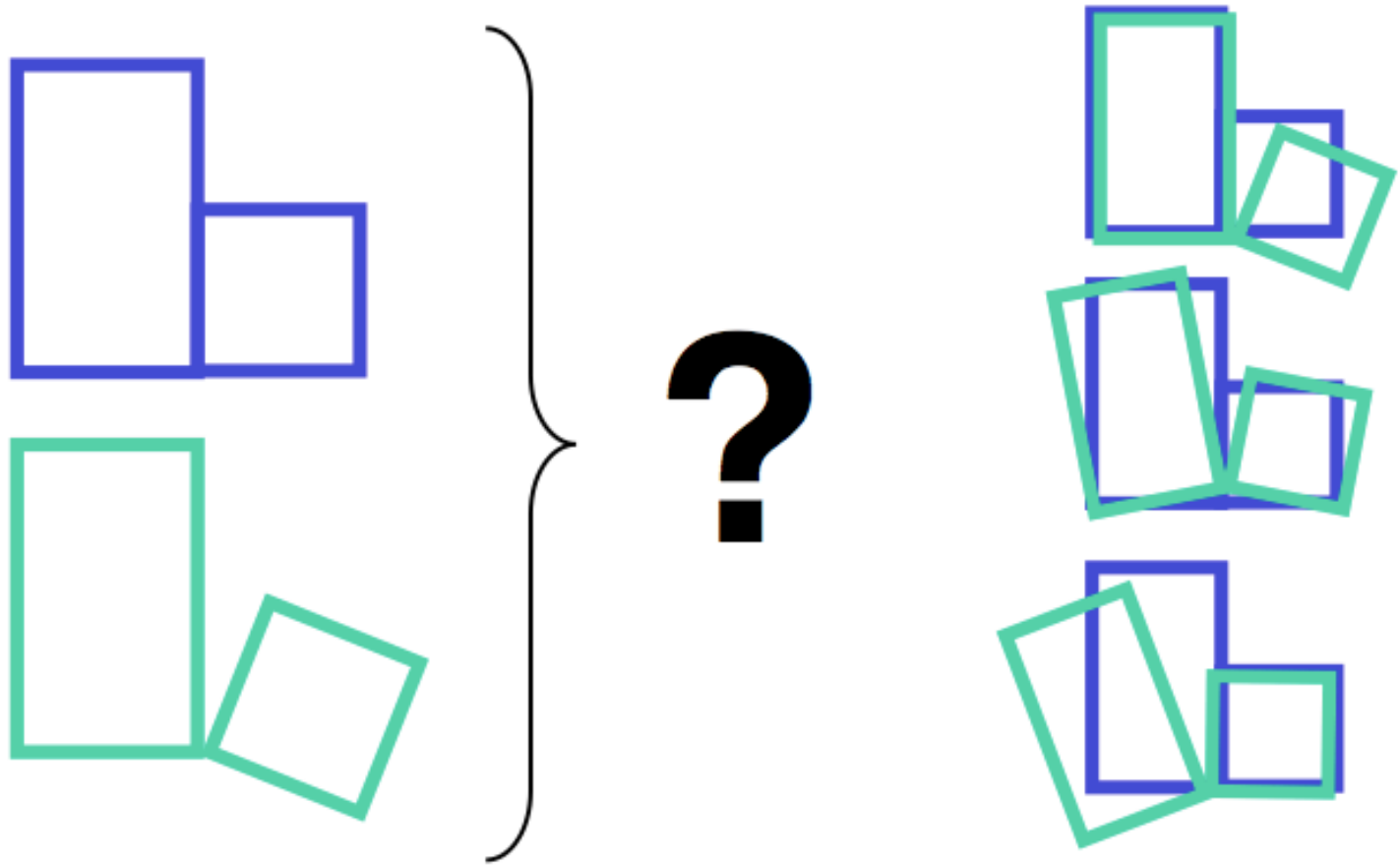
There are no universally agreed upon criteria. As usual, it depends on what you are using the alignment to do.

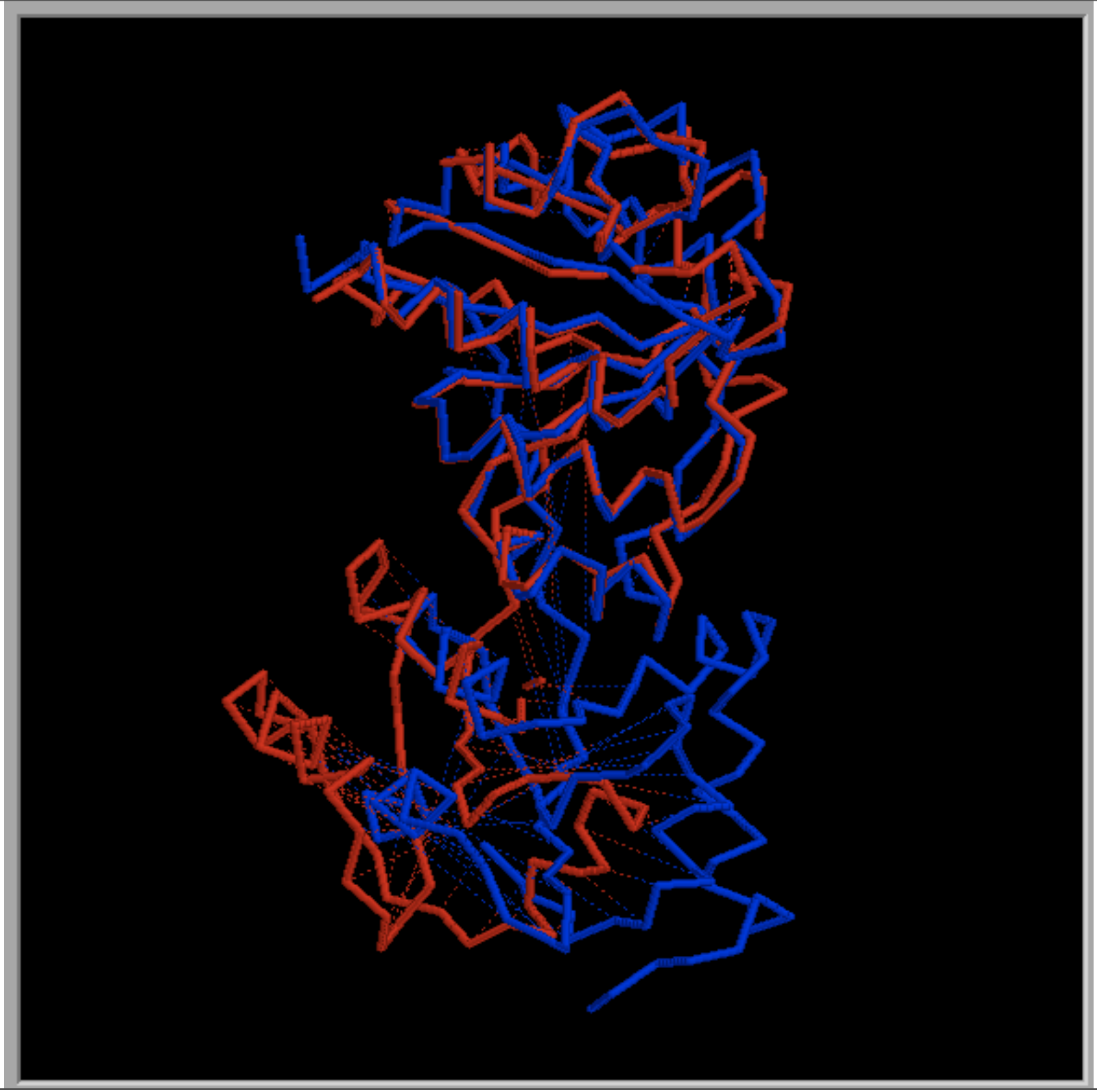
CASP Evaluation Measure



When is RMSD misleading ?

Rigid body movement of domains or subdomains ...





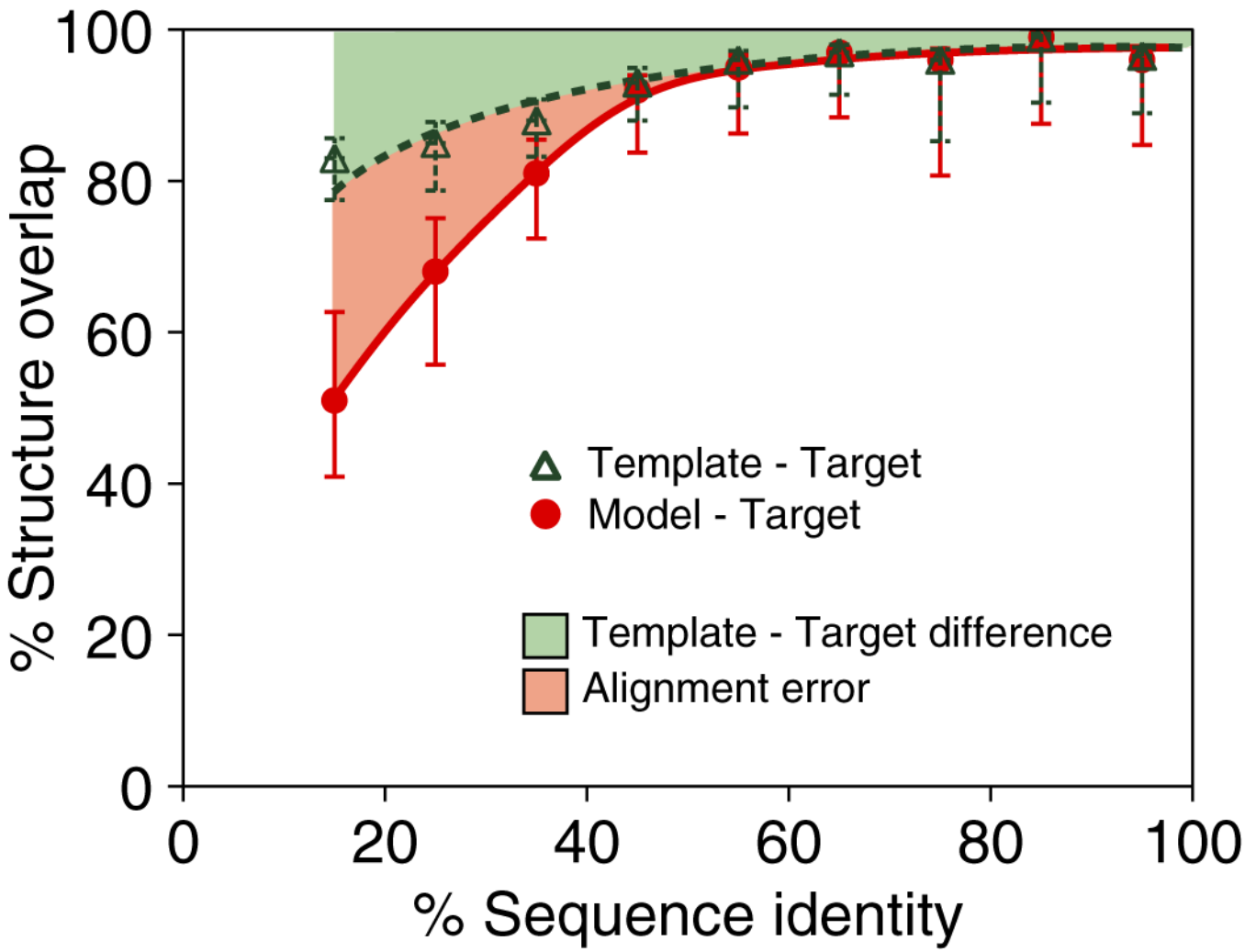
Sequence alignment

```
----SEIYGIHAVQALLERAPERFQEVFILKGREDKR-----LLPLIHALESQGVVIQL 50
MRITSTANPRIKELARLLERKHRDSQRRFLIEGAREIERALQAGIELEQALVWEGGLNPE 60

ANRQYLDE----KSDGAVHQGIIARVKPGRQYQENDLPDLIASLDQP----FLLILDGVT 102
EQQVYAALLALLEVSEAVLKKLSVRDNPAGLIALARMPERTLEEYRPSPDALILVAVGLE 120

DPHNLGACLRSADAAGVHAVIVPKDRSAQLNATAKKVACGAAESVPLIRVTNLARTRLQE 162
KPGNLGAVLRSADAAGAEAVLVAGGVDLYSPQVIRNSTGVVFSLRTLAASESEVLDWIKQ 180

ENIWIVGTAGEADHTLYQS--KTGRLALVGAEGEGRRLTR-EHCDELISIPAG-SVSSLN 218
HNLPLVATTPHAEALYWEANLRPPVAIAVGPEHEGLRAAWLEAAQTQVRIPMQGQADSLN 240

VSVATGICLFEAVRQRS- 235
VSVSAALLLYEALRQRLL 258
```

Structure alignment

```
----------------------SEIYGIHAVQALLERAPERFQEVFILKGREDKRLLPLI   38
MRITSTANPRIKELARLLERKHRDSQRRFLIEGAREIERALQAGIELEQALVWEGGLNPE   60

HALESQGVVIQLANRQYLDEKSDGAVHQGIIARVKPGRQYQENDLPDLIASLDQPFLLIL   98
EQQVYAALLALLEVSEAVLKKLSVRDNPAGLIALARMPERTLEEYR----PSPDALILVA  116

DGVTDPHNLGACLRSADAAGVHAVIVPKDRSAQL-NATAKKVACGAAESVPLIRVT--NL  155
VGLEKPGNLGAVLRSADAAGAEAVLVAG--GVDLYSPQVIRNSTGVVFSLRTLAASESEV  174

ARTRLQEENIWIVGTAGEADHTLYQSK-TGRLALV-GAEGEG-RRLTREHCDELISIP-A  211
LDW-IKQHNLPLVATTPHAEALYWEANLRPPVAIAVGPEHEGLRAAWLEAAQTQVRIPMQ  233

GSVSSLNVSVATGICLFEAVRQRS-  235
GQADSLNVSVSAALLLYEALRQRLL  258
```

Model Accuracy as a Function of Target-Template Sequence Identity



Sánchez, R., Šali, A. Proc Natl Acad Sci U S A. 95 pp13597-602. (1998).

A domain is a:

- Compact, semi-independent unit (Richardson, 1981).
- Stable unit of a protein structure that can fold autonomously (Wetlaufer, 1973).
- Recurring functional and evolutionary module (Bork, 1992).

"Nature is a tinkerer and not an inventor" (Jacob, 1977).

Domain connectivity

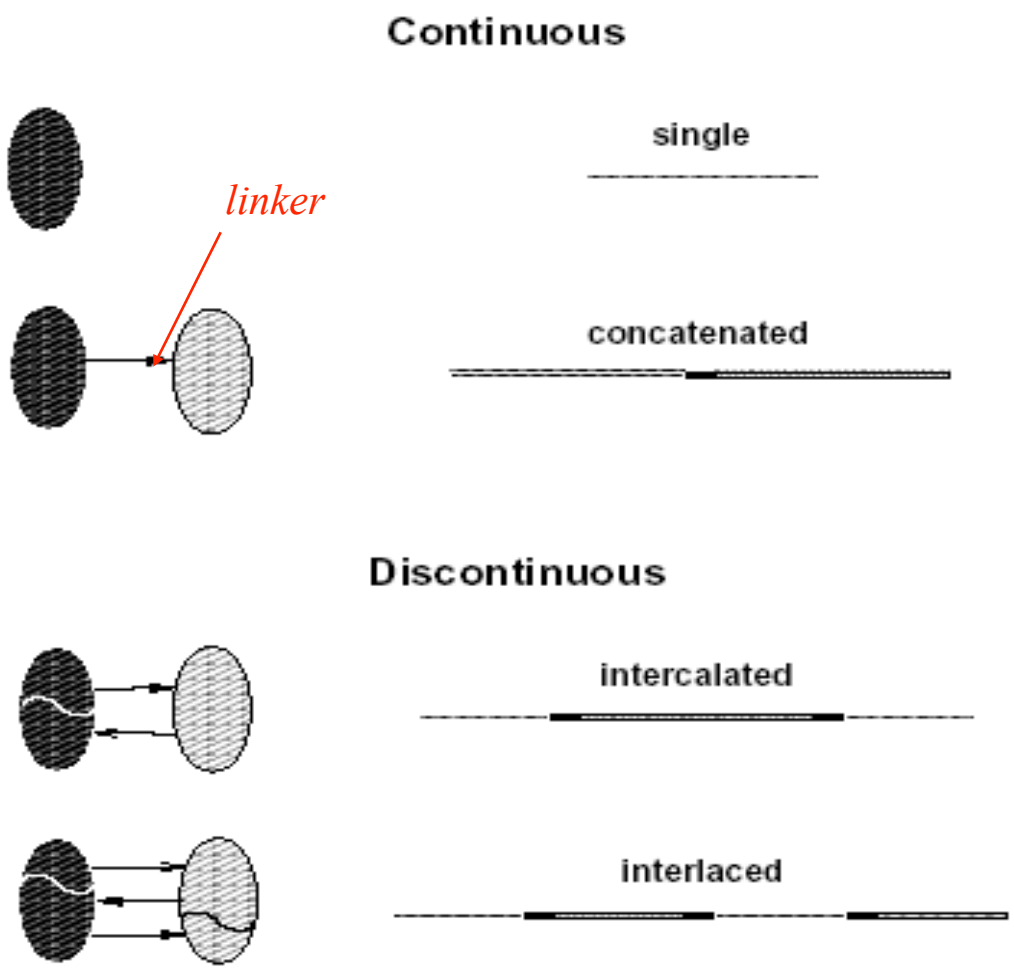


Figure 1: Scheme illustrating different types of connectivity in multidomain structures (left) and their sequences (right).

Domain databases

Table 1: Domain databases

Database	URL	Reference
Sequence		
BLOCKS	http://blocks.fhcrc.org	(Henikoff et al., 2000)
COGS	http://www.ncbi.nlm.nih.gov/COG	(Tatusov et al., 2001)
DOMO	http://www.infobiogen.fr/services/domo	(Gracy and Argos, 1998b)
InterPro	http://www.ebi.ac.uk/interpro	(Apweiler et al., 2000)
Pfam	http://www.sanger.ac.uk/Pfam	(Bateman et al., 2000)
PRINTS	http://www.bioinf.man.ac.uk/dbbrowser/PRINTS	(Attwood et al., 1999)
ProDom	http://www.toulouse.inra.fr/prodom.html	(Corpet et al., 2000)
PROSITE	http://www.expasy.ch/prosite	(Hofmann et al., 1999)
PROT-FAM	http://mips.gsf.de/proj/protfam	(Srinivasarao et al., 1999)
SBASE	http://www3.icgeb.trieste.it/ sbasesrv	(Murvai et al., 2000)
SMART	http://SMART.embl-heidelberg.de	(Schultz et al., 2000)
Structure		
3Dee	http://barton.ebi.ac.uk/servers/3Dee.html	(Siddiqui et al., 2001)
CATH	http://www.biochem.ucl.ac.uk/bsm/cath	(Orengo et al., 1997)
FSSP	http://www2.ebi.ac.uk/dali/	(Holm and Sander, 1997)
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop	(Murzin et al., 1995)

Model:Retinoschisin

http://www.expasy.org/cgi-bin/niceprot.pl?.O15537#comments

Retinoschisin is a 24-kDa discoidin domain-containing protein that is secreted from photoreceptor and bipolar cells as a large disulfide-linked multisubunit complex. It functions as a cell adhesion protein to maintain the cellular organization and synaptic structure of the retina. Over 125 different mutations in the RS1 gene are associated with X-linked juvenile retinoschisis, the most common form of early onset macular degeneration in males.

>O15537|XLRS1_HUMAN Retinoschisin - Homo sapiens

MSRKIEGFLLLLLFGYEATLGLSSTEDEGEDPWYQK
ACKCDCQGGPNALWSAGATSLDCIPECPYHKPLGFE
SGEVTPDQITCSNPEQYVGWYSSWTANKARLNS
QGFGCAWLSKFQDSSQWLQIDLKEIKVISGILTQ
GRCDIDEWMTKYSVQYRTDERLNWIYYKDQTGNN
RVFYGNSDRTSTVQNLLRPPIISRFIRLIPLGWHVRI

Model:Retinochisin

- 1) Blast search with the query sequence for homologous sequences with a solved structure (template)

blast search

- 2) align the query sequence and the template sequence

alignment

- 3) assign the coordinates of the template to the aligned query sequence

- 4) refine the model

Domain search: SMART

The domain within your query sequence starts at position 62 and ends at position 219; the E-value for the FA58C domain shown below is 2.47e-36.

```
ECPYHKPLGFESGEVTPDQITCSNPEQYVGWYSSWTANKARLNSQGFGCAWLSKFQDSSQ
WLQIDLKEIKVISGILTQGRCDIDEWMTKYSVQYRTDERLNWIYYKDQTGNNRVFYGNSD
RTSTVQNLLRPPIISRFIRLIPLGWHVRIAIRMELLEC
```

BLAST with Domain | Align your sequence against the SMART alignment

FA58C

Coagulation factor 5/8 C-terminal domain, discoidin domain

SMART accession number:	SM00231
Description:	Cell surface-attached carbohydrate-binding domain, present in eukaryotes and assumed to have horizontally transferred to eubacterial genomes.
Interpro abstract (IPR000421):	Blood coagulation factors V and VIII contain a C-terminal, twice repeated, domain of about 150 amino acids, which is called F5/8 type C, FA58C, or C1/C2- like domain. In the Dictyostelium discoideum (Slime mold) cell adhesion protein discoidin, a related domain, named discoidin I-like domain, DLD, or DS, has been found which shares a common C-terminal region of about 110 amino acids with the FA58C domain, but whose N-terminal 40 amino acids are much less conserved. Similar domains have been detected in other extracellular and membrane proteins (PUBMED:3092220), (PUBMED:8390675), (PUBMED:8639264) In coagulation factors V and VIII the repeated domains compose part of a larger functional domain which promotes binding to anionic phospholipids on the surface of platelets and endothelial cells (PUBMED:3125864). The C-terminal domain of the second FA58C repeat (C2) of coagulation factor VIII has been shown to be responsible for phosphatidylserine-binding and essential for activity (PUBMED:2110840), (PUBMED:7515064). It forms an amphipathic alpha-helix, which binds to the membrane (PUBMED:7893714). FA58C contains two conserved cysteines in most proteins, which link the extremities of the domain by a disulphide bond (PUBMED:8504111), (PUBMED:7613471), (PUBMED:8856064). A further disulphide bond is located near the C-terminal of the second FA58C domain in MFGM Q08431 (PUBMED:8856064).
GO process:	cell adhesion (GO:0007155)
Family alignment:	View Alignment consensus sequence or Family alignment in CHROMA format

```
  +-----------------------------------------------------------------------+
  |                                                                 +-+   |
  |                                                                 | |   |
  CxPLGxxQITASxxxxxRLxxxWxxxxWxxxxxxQGxxxxxxxxxxxxGNxxxxxxxxxxRxPxcxcLRxExGC

'C': conserved cysteine involved in a disulphide bond.
'c': cysteine involved in a disulphide bond in MFGM .
'x': any amino acid.
upper case letters: conserved residues.
```

NCBI PubMed www.pubmed.gov
A service of the U.S. National Library of Medicine and the National Institutes of Health
My NCBI [Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Go Clear Advanced Search
Limits Preview/Index History Clipboard Details
Display AbstractPlus Show 20 Sort By Send to
All: 1 Review: 0

1: J Biol Chem. 2005 Mar 18;280(11):10721-30. Epub 2005 Jan 11.

Final Version FREE J Biol Chem Links

RS1, a discoidin domain-containing retinal cell adhesion protein associated with X-linked retinoschisis, exists as a novel disulfide-linked octamer.

Wu WW, Wong JP, Kast J, Molday RS.

Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, V6T 1Z3, Canada.

RS1, also known as retinoschisin, is an extracellular protein that plays a crucial role in the cellular organization of the retina. Mutations in RS1 are responsible for X-linked retinoschisis, a common, early-onset macular degeneration in males that results in a splitting of the inner layers of the retina and severe loss in vision. RS1 is assembled and secreted from photoreceptors and bipolar cells as a homo-oligomeric protein complex. Each subunit consists of a 157-amino acid discoidin domain flanked by two small segments of 39 and 5 amino acids. To begin to understand how the structure of RS1 relates to its role in retinal cell adhesion and X-linked retinoschisis, we have determined the subunit organization and disulfide bonding pattern of RS1 by SDS gel electrophoresis, velocity sedimentation, and mass spectrometry. Our results indicate that RS1 exists as a novel octamer in which the eight subunits are joined together by Cys(59)-Cys(223) intermolecular disulfide bonds. Subunits within the octamer are further organized into dimers mediated by Cys(40)-Cys(40) bonds. These cysteines lie just outside the discoidin domain indicating that these flanking segments primarily function in the octamerization of RS1. Within the discoidin domain, two cysteine pairs (Cys(63)-Cys(219) and Cys(110)-Cys(142)) form intramolecular disulfide bonds that are important in protein folding, and one cysteine (Cys(83)) exists in its reduced state. Because mutations that disrupt subunit assembly cause X-linked retinoschisis, the assembly of RS1 into a disulfide-linked homo-octamer appears to be critical for its function as a retinal cell adhesion protein.

PMID: 15644328 [PubMed - indexed for MEDLINE]

Related Articles

- Defective discoidin domain structure, subunit assembly, and endoplasmic reticulum processing of retinoschisin are primary mechanisms responsible for [J Biol Chem. 2003]
- Retinoschisin (RS1), the protein encoded by the X-linked retinoschisis gene, is anchored to the surface of retinal photoreceptor and bipolar cells through its in [J Biol Chem. 2007]
- Coexpression and interaction of wild-type and missense RS1 mutants associated with X-linked retinoschisis: its relevance to gene therapy. [Invest Ophthalmol Vis Sci. 2007]
- Review X-linked retinoschisis: an update. [J Med Genet. 2007]
- Review X-linked retinoschisis: a clinical and molecular genetic review. [Surv Ophthalmol. 2004]

» See Reviews... | » See All...

Cited by PubMed Central articles

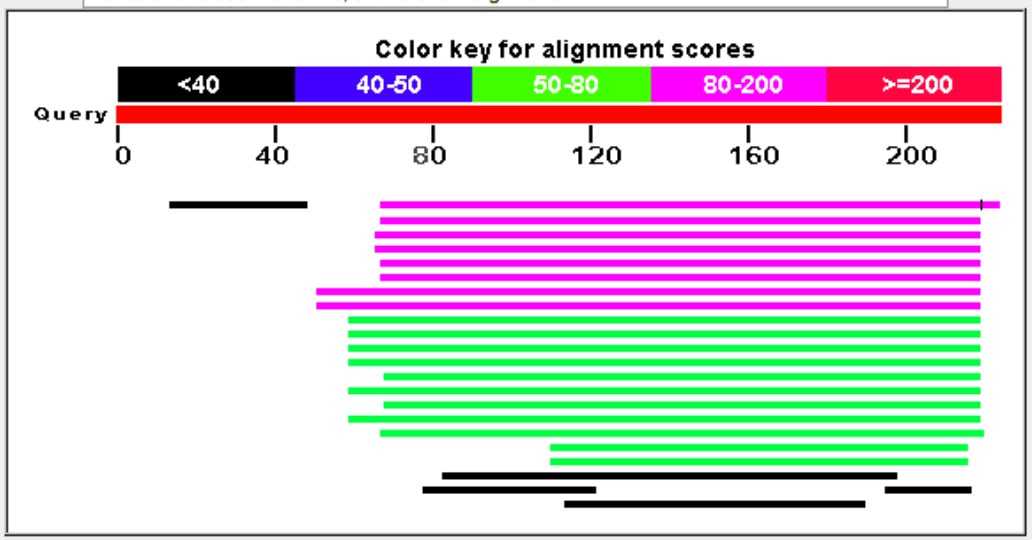
- Clinical and genetic findings in Hungarian patients with X-linked juvenile retinoschisis. [Mol Vis. 2008]
- CRX controls retinal expression of the X-linked juvenile retinoschisis (RS1) gene. [Nucleic Acids Res. 2008]
- Truncation of retinoschisin protein associated with a novel splice site mutation in the RS1 gene. [Mol Vis. 2008]

» See all (8)...

Blast search vs PDB

Distribution of 33 Blast Hits on the Query Sequence

Mouse over to see the defline, click to show alignments



Descriptions

Sequences producing significant alignments:		Score (Bits)	E Value	
pdb|1SDD|B	Chain B, Crystal Structure Of Bovine Factor Vai	115	1e-26	S
pdb|1CZS|A	Chain A, Crystal Structure Of The C2 Domain Of Hum...	109	8e-25	S
pdb|3BN6|A	Chain A, Crystal Structure Of The C2 Domain Of Bov...	104	3e-23	S
pdb|2PQS|A	Chain A, Crystal Structure Of The Bovine Lactadher...	104	3e-23	S
pdb|1D7P|M	Chain M, Crystal Structure Of The C2 Domain Of Hum...	97.4	4e-21	S
pdb|1IQD|C	Chain C, Human Factor Viii C2 Domain Complexed To ...	97.4	4e-21	S
pdb|2R7E|B	Chain B, Crystal Structure Analysis Of Coagulation...	97.1	4e-21	S
pdb|3CDZ|B	Chain B, Crystal Structure Of Human Factor Viii	97.1	5e-21	S
pdb|2QQK|A	Chain A, Neuropilin-2 A1a2b1b2 Domains In Complex ...	90.1	6e-19	S
pdb|2QQO|A	Chain A, Crystal Structure Of The A2b1b2 Domains F...	90.1	6e-19	S
pdb|2QQJ|A	Chain A, Crystal Structure Of The B1b2 Domains Fro...	89.7	8e-19	S
pdb|2ORX|A	Chain A, Structural Basis For Ligand Binding And H...	80.5	5e-16	S
pdb|2QQN|A	Chain A, Neuropilin-1 B1 Domain In Complex With A ...	76.3	8e-15	S

Align yourself query+template
Praline with homology extended alignment

IBIVU - www.ibi.vu.nl

Centre for Integrative Bioinformatics VU
vrije Universiteit *amsterdam*

Main page | Research | Teaching | Programs | Members | Publications | Vacancies | Related sites

PRALINE multiple sequence alignment

Advanced Interface

Options Help

PRALINE sample output

References and FAQs

PRALINE is a multiple sequence alignment program with many options to optimise the information for each of the input sequences; e.g. global or local preprocessing, predicted secondary structure information and iteration capabilities.

Paste in your sequences in FASTA format (MAX 500 sequences, length 2000):

```
>sp|O15537|XLRS1_HUMAN Retinoschisin OS=Homo sapiens GN=RS1 PE=1 SV=2
MSRKIEGFLLLLLFGYEATLGLSSTEDEGEDPWYQKACKCDCQGGPNALWSAGATSLDCI
PECPYHKPLGFESGEVTPDQITCSNPEQYVGWYSSWTANKARLNSQGFGCAWLSKFQDSS
QWLQIDLKEIKVISGILTQGRCDIDEWMTKYSVQYRTDERLNWIYYKDQTGNNRVFYGNS
DRTSTVQNLLRPPIISRFIRLIPLGWHVRIAIRMELLECVSKCA
>gi|50513523|pdb|1SDD|B Chain B, Crystal Structure Of Bovine Factor Vai
```

Or Upload a FASTA file (MAX 500 sequences, length 2000): Browse...

Enter a name for your job PRALINE Job

Options

Exchange weights matrix: BLOSUM62 Help **Associated gap penalties:** 12 Open 1 Extension Help

Global progressive alignment strategy: Help

- Standard progressive strategy
- Pre-profile global processing Iterations No Score Cutt-off 0
- Pre-profile local processing Iterations No Score Cutt-off 0
- PSI-BLAST pre-profile processing (Homology-extended alignment)

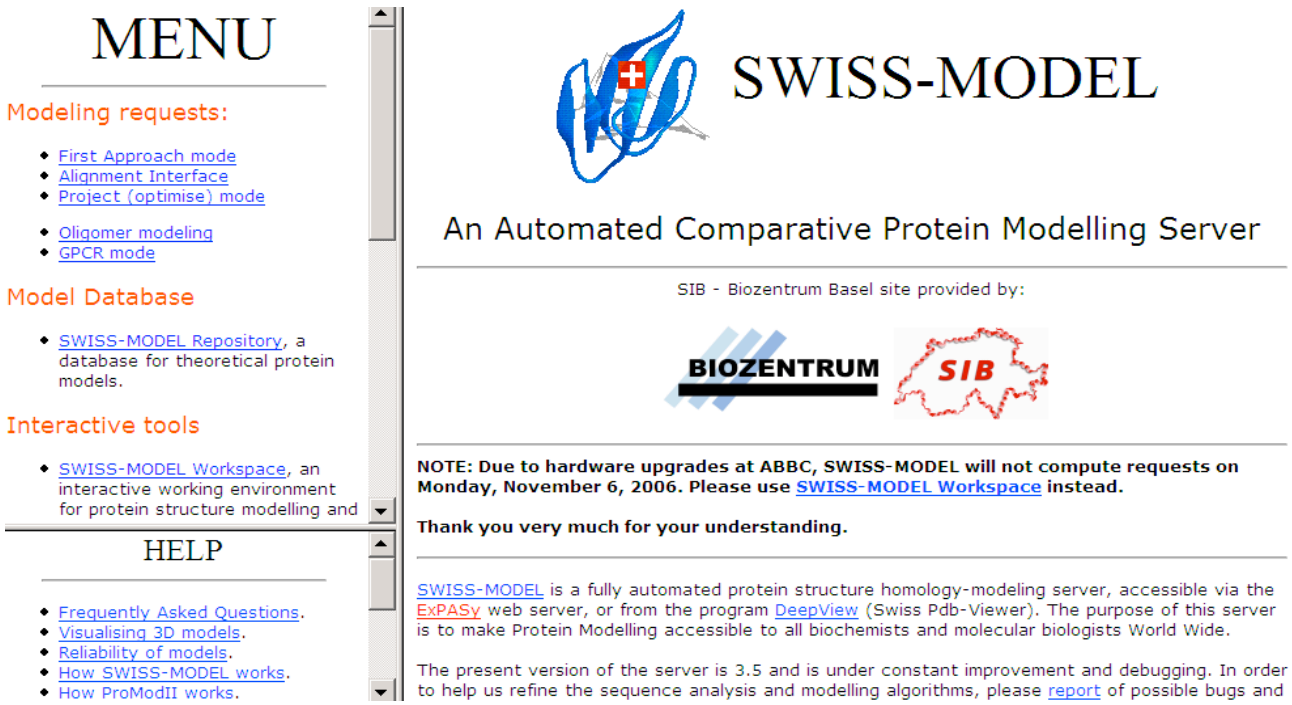
PSI-BLAST Iterations 3 Start e-value Cutt-off at 0.000001 DB NR

Secondary structure prediction No Help

DSSP-defined secondary structure search YES NO Help

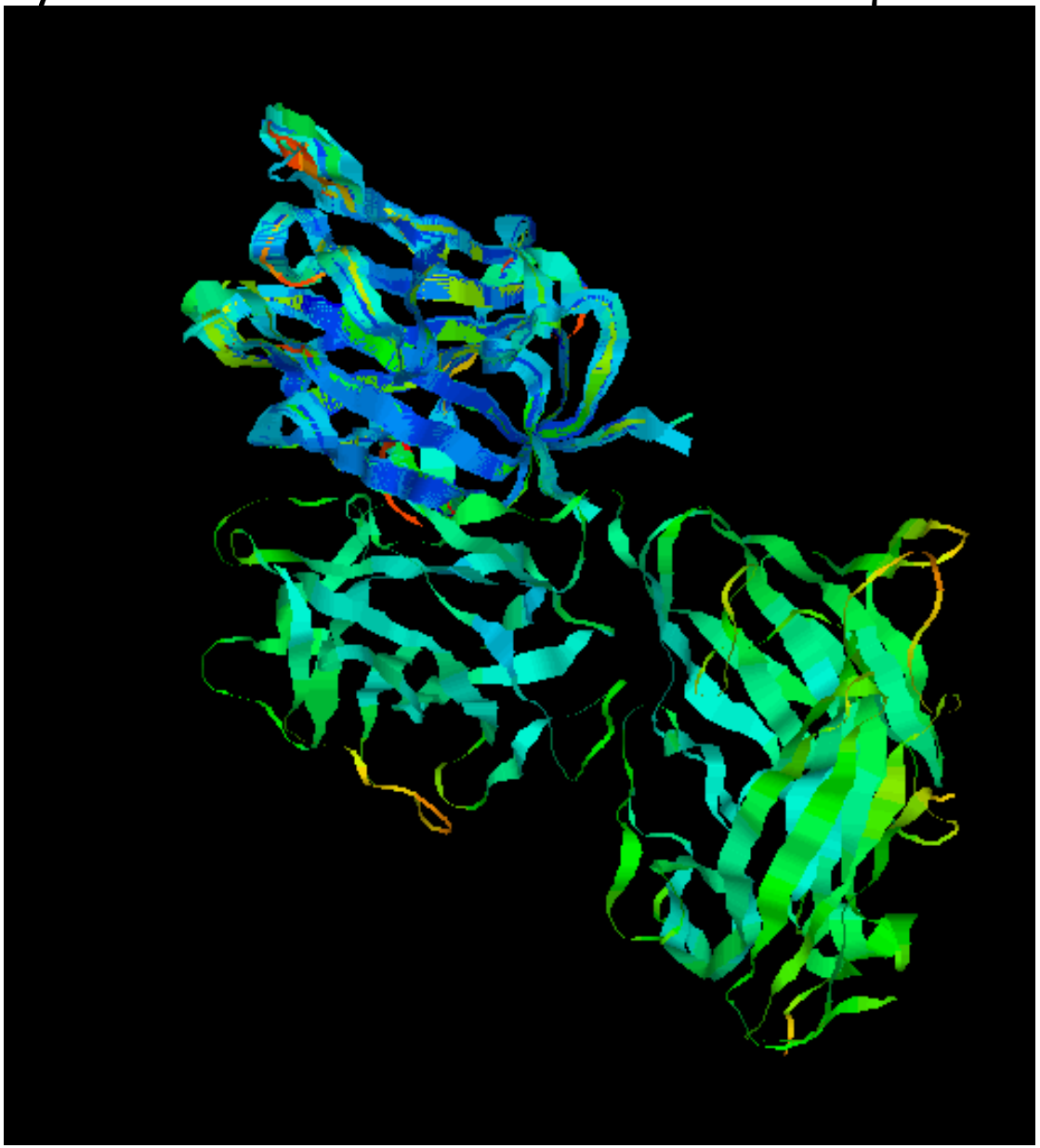
Modelling web server

- Swissmodel server: http://swissmodel.expasy.org/

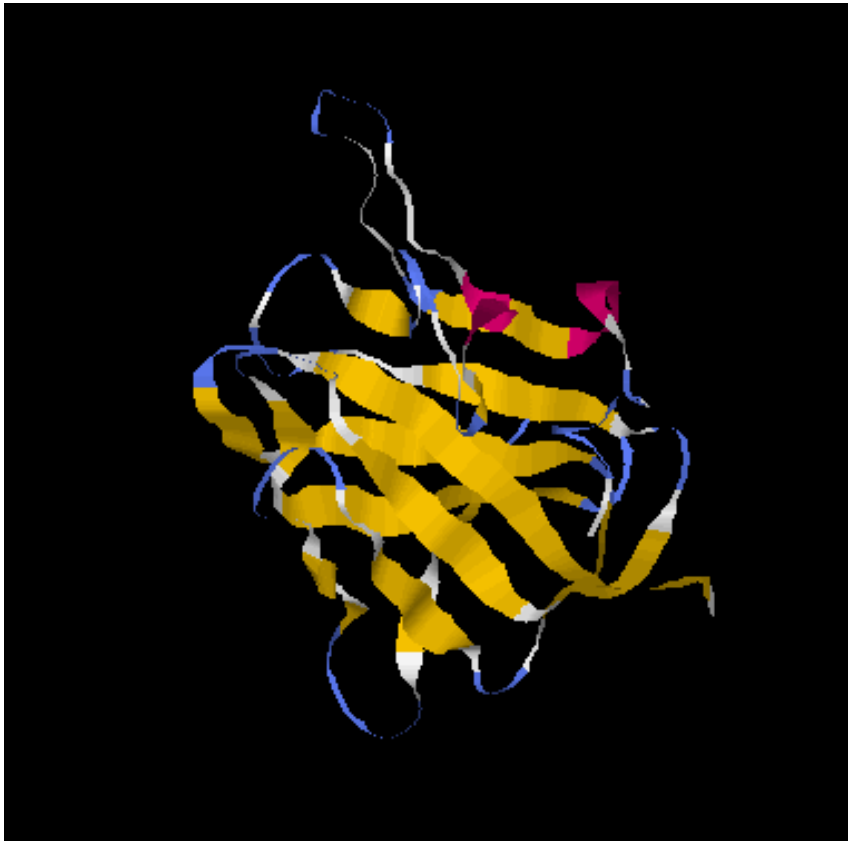


Model retrieved

If you take the entire retinoschisin sequence



If you take only the discoidin domain



Database of Comparative Protein Structure Models

Modbase http://modbase.compbio.ucsf.edu/modbase-cgi/search_form.cgi

UCSF University of California, San Francisco | About UCSF | UCSF Medical Center

ModBase | Home | User Login | ModBase Search Page | ModWeb Modelling Server | He

Current Lo

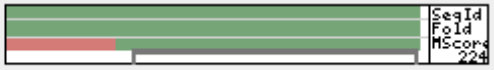
Sequence Information

Primary Database Link	O15537 (XLRS1_HUMAN)
Organism	*Homo sapiens*
Annotation	retinoschisin precursor (x-linked juvenile retinoschisis protein)
Sequence Length	224

Model Information

Perform action on this model : Select option

Sequence Model Coverage



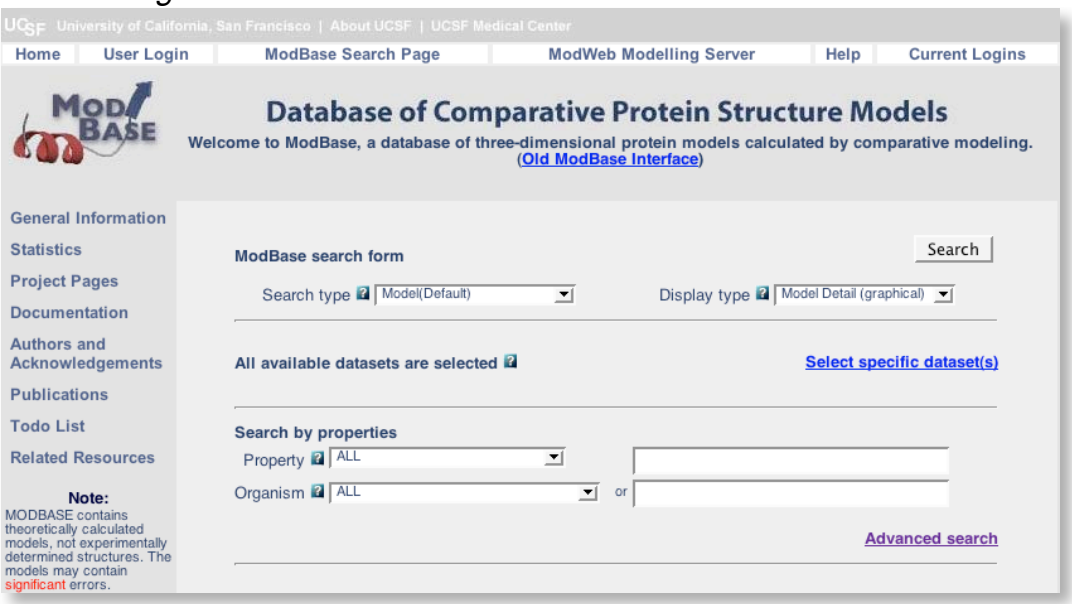
Sequence Identity	39.00%
E-Value	0
Model Score	1.00
Target Region	68-219
Protein Length	224
Template PDB Code	1cztA
Template Region	4-156
Dataset	H.sapiens

Filtered models for current sequence (Show all models)

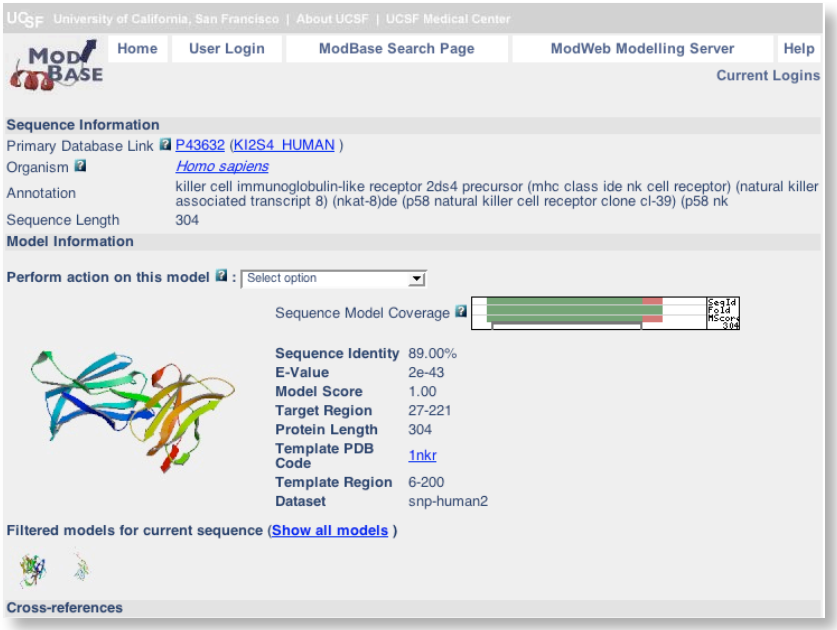
MODBASE

http://salilab.org/modbase

Search Page



Model Details



Sequence Overview

Q8G8A6	hypothetical protein	*Pseudomonas aeruginosa*	3738
Q8G9W1	hypothetical protein	*Escherichia coli*	1140
Q8CY62	hypothetical protein spr1965	*Streptococcus pneumoniae, Streptococcus pneumoniae R6*	1038

Model Overview

Q8G8C7	hypothetical protein	*Pseudomonas aeruginosa*	4996	2089-2158	70	37.00	7e-14	1.00	1dnyA	8-78
Q8G8C7	hypothetical protein	*Pseudomonas aeruginosa*	4996	492-1017	526	36.00	1e-82	1.00	1amuA	19-529
Q8G9W1	hypothetical protein	*Escherichia coli*	1140	349-1135	787	35.00	0	1.00	1r9dA	6-783

- *Pieper et al. (2004) Nucleic Acids Research 32, D217-D222*

http://www.salilab.org/modeller/tutorial/

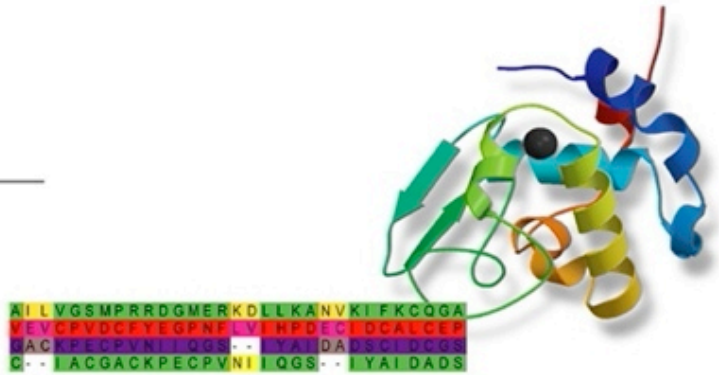
Tutorial

http://salilab.org/modeller/tutorial/

To main Sali lab pages

Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



- About MODELLER
- MODELLER News
- Download & Installation
  - Release Notes
- Registration
- Discussion Forum
  - Subscribe
  - Browse archives
  - Search archives
- Documentation
  - FAQ
  - Tutorial
  - Online manual
  - Wiki
- Developers' Pages
- Contact Us

Tutorial

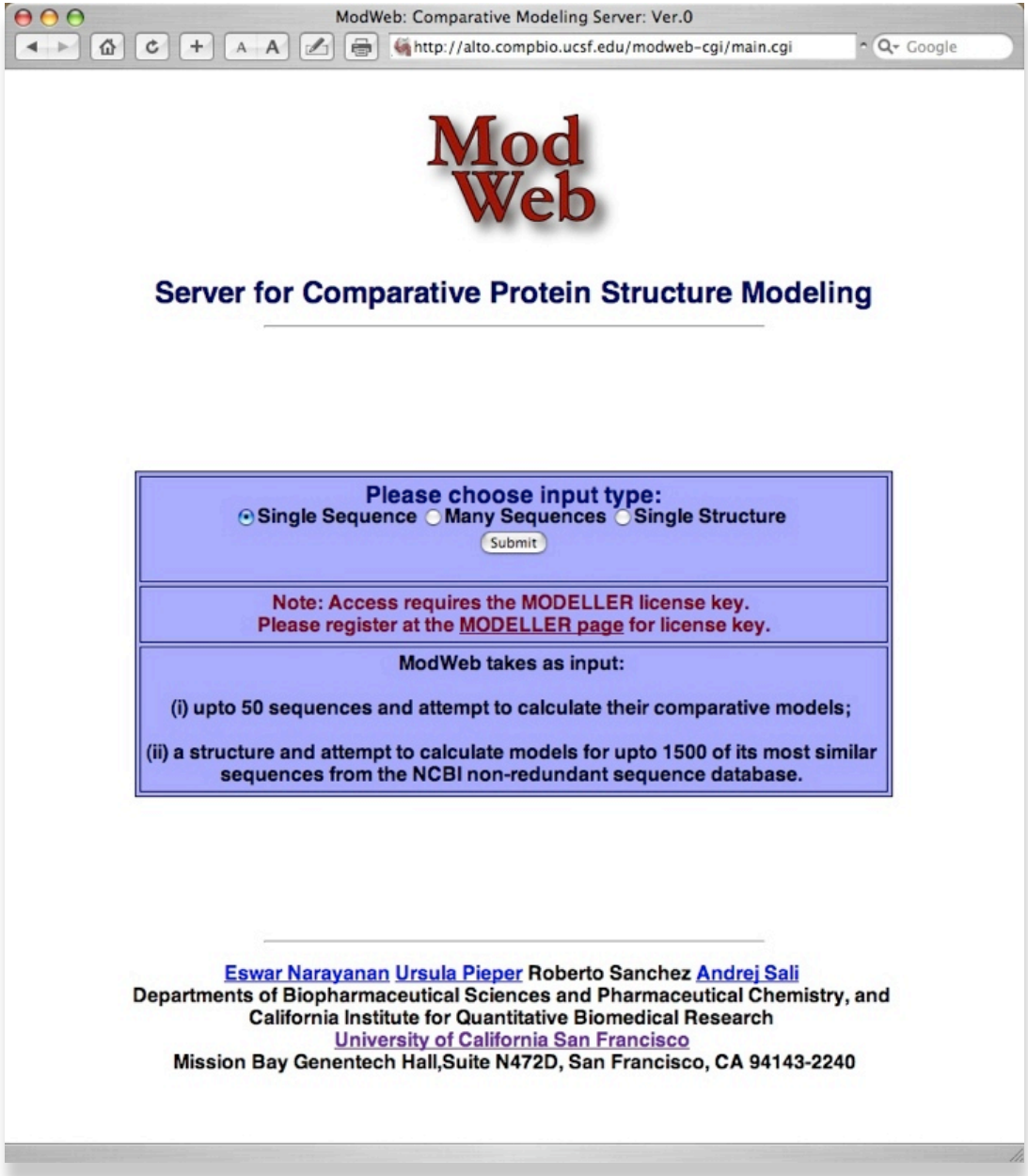
MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2,3). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms.

This web site presents a tutorial for the use of MODELLER 8v0 (for older versions of MODELLER, use the old MODELLER 7v7 tutorial). There are 4 modeling examples that the user can follow:

1. Basic Modeling. *Model a sequence with high identity to a template.*
   This exercise introduces the use of MODELLER in a simple case where the template selection and target-template alignments are not a problem.
2. Advanced Modeling. *Model a sequence based on multiple templates and bound to a ligand.*
   This exercise introduces the use of multiple templates and ligands in the process of model building with MODELLER.
3. Iterative Modeling. *Increase the accuracy of the modeling exercise by iterating the 4 step process.*
   This exercise introduces the concept of MOULDING to improve the accuracy of comparative models.
4. Difficult Modeling. *Model a sequence based on a low identity to a template.*
   This exercise uses resources external to MODELLER in order to select a template for a dificult case of protein structure prediction.

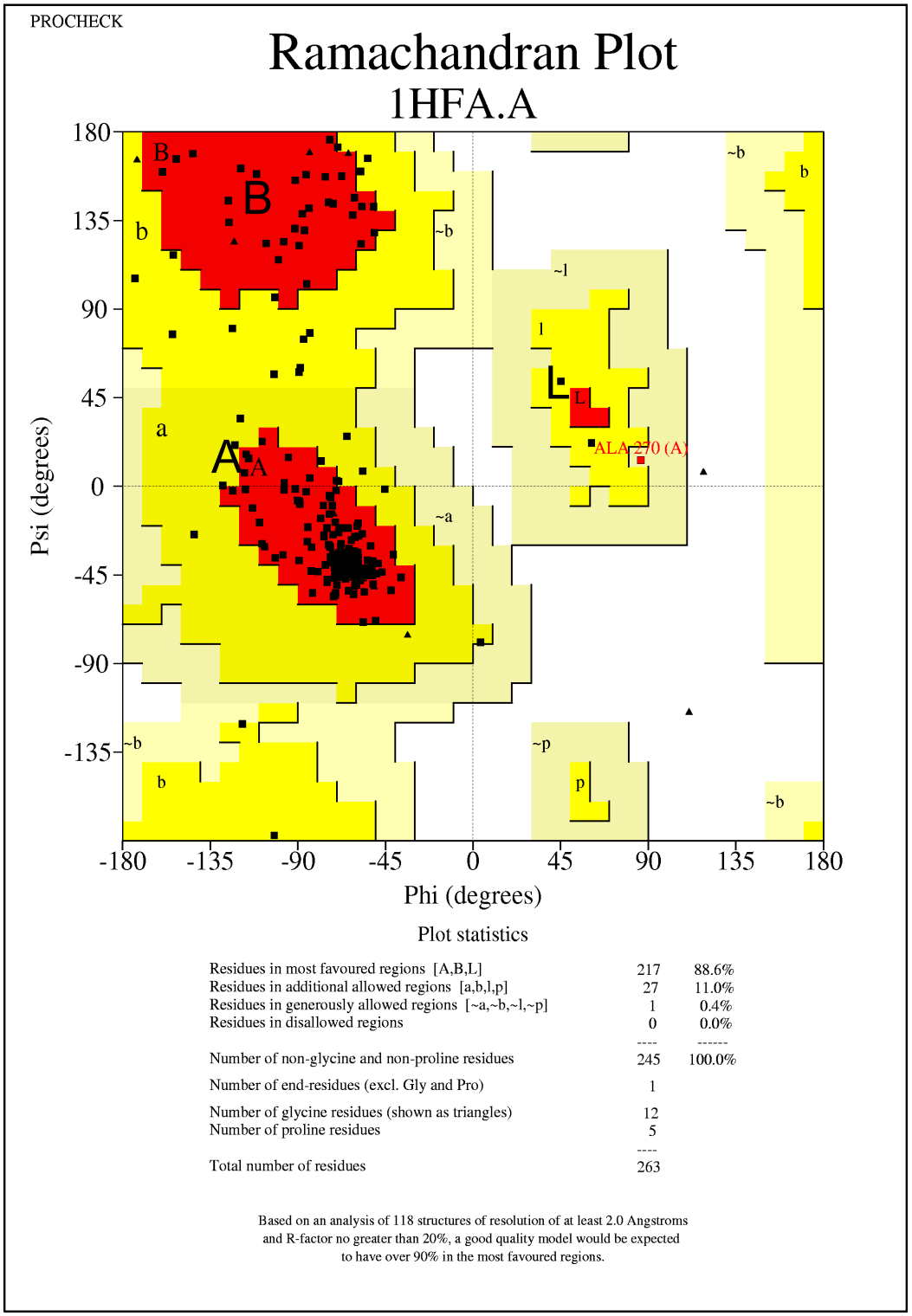
MODWEB

http://salilab.org/modweb



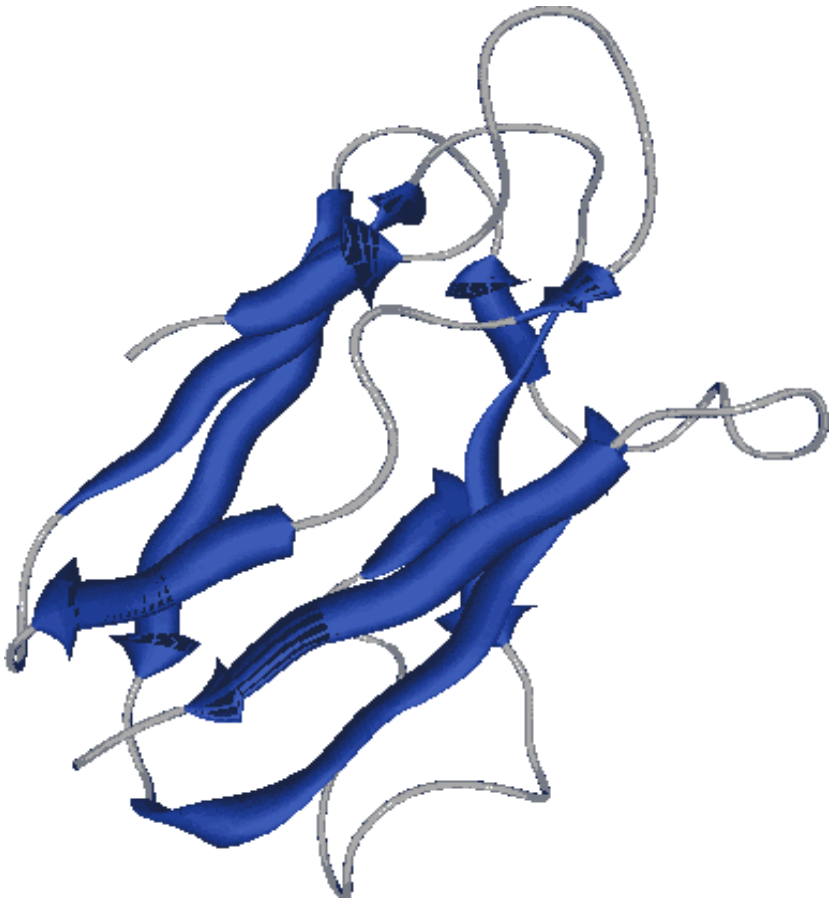
Ramachandran plot

- B. Beta strand
- A. Right handed helix
- L. Left handed helix
- Color coding
  - White. Disallowed
  - Red. Most favorable
  - Yellow. Allowed region
- Glycine triangles

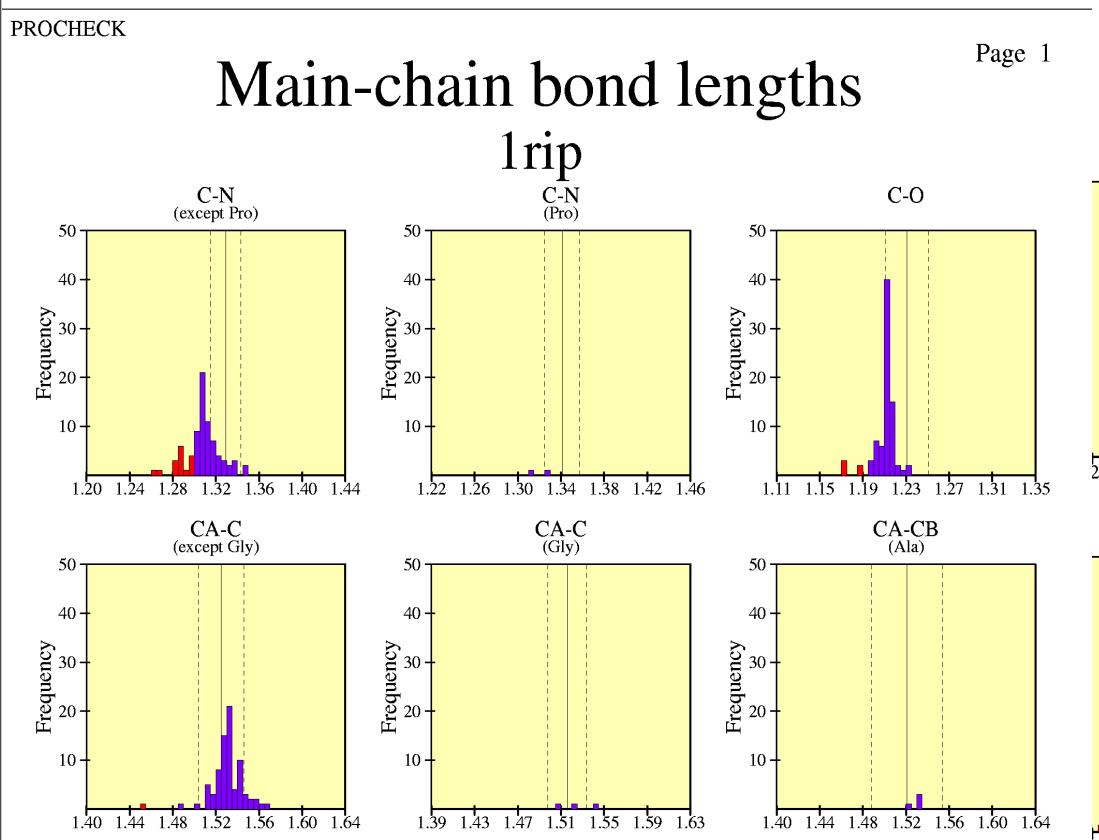


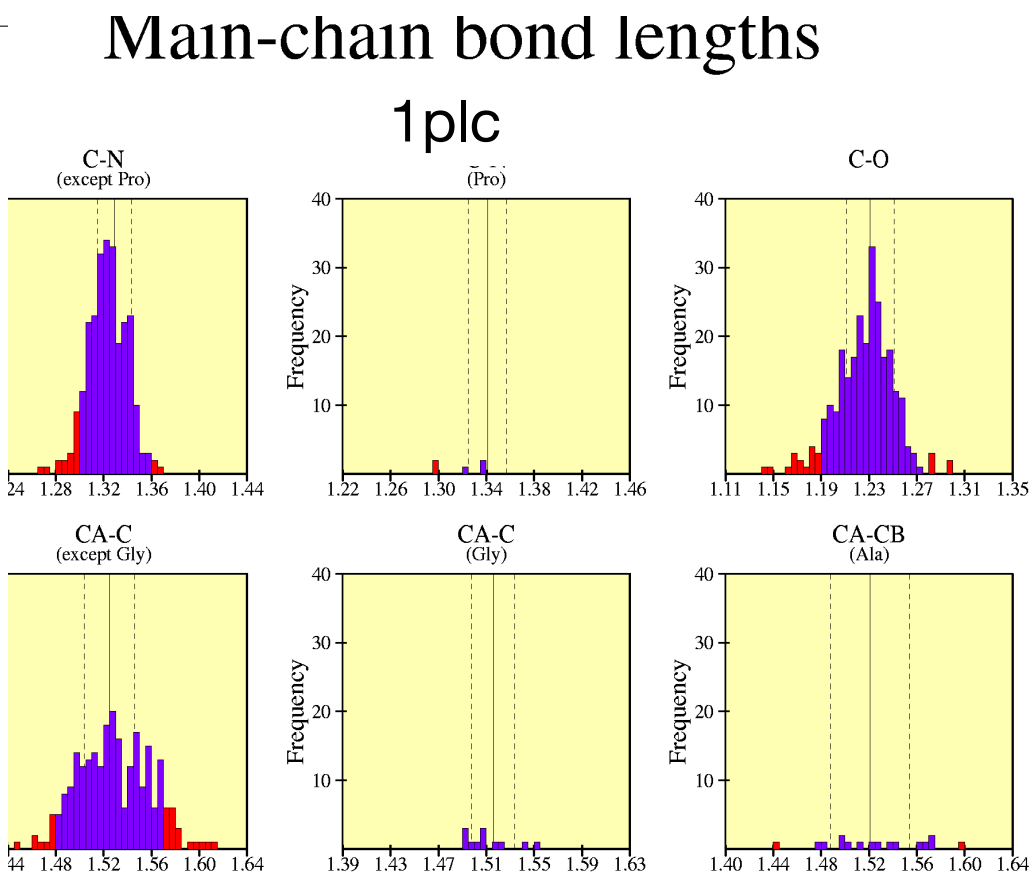
Let's find the wrong structure!

1PLC Electron transport protein



Procheck.Bond Lenghts

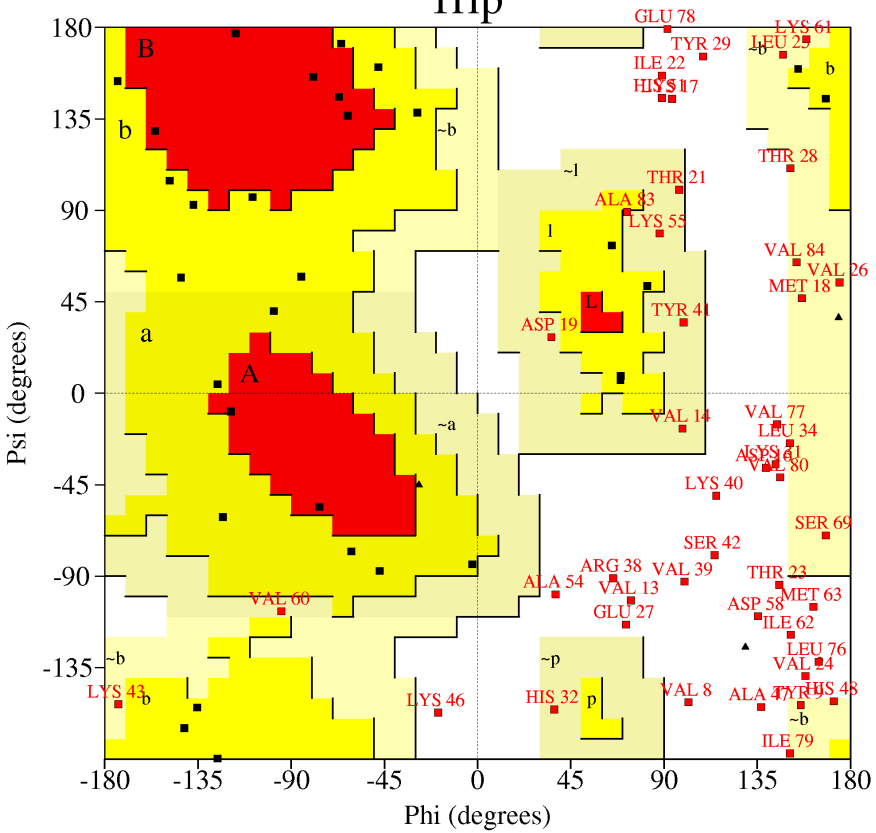




PROCHECK

Ramachandran Plot

1rip



Plot statistics

Residues in most favoured regions [A,B,L]	5	6.8%
Residues in additional allowed regions [a,b,l,p]	24	32.4%
Residues in generously allowed regions [~a,~b,~l,~p]	20	27.0%
Residues in disallowed regions	25	33.8%
	---	------
Number of non-glycine and non-proline residues	74	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	3	
Number of proline residues	2	

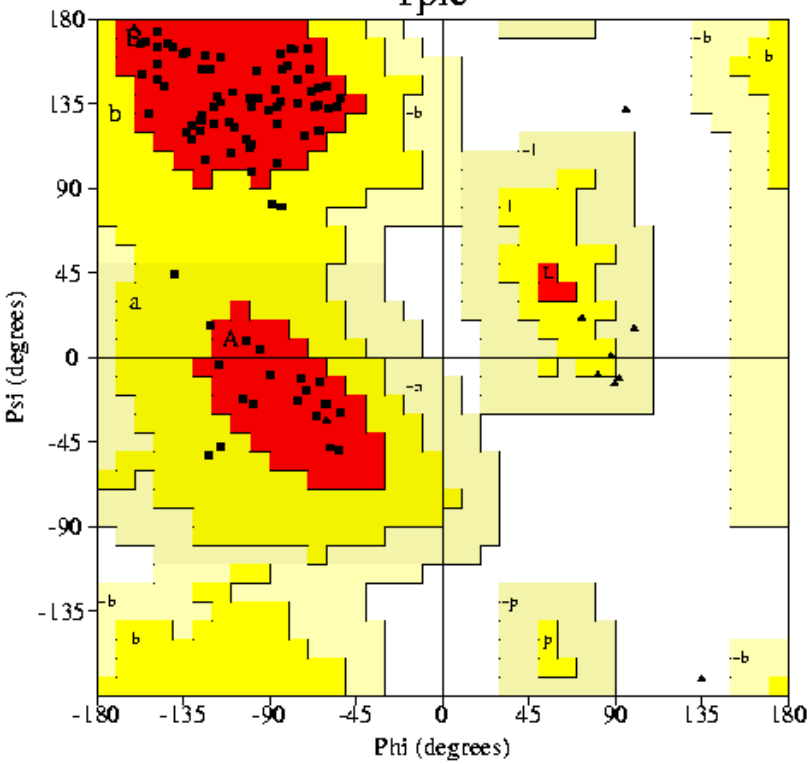
Total number of residues	81	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

PROCHECK

Ramachandran Plot

1plc



Plot statistics

Residues in most favoured regions [A,B,L]	76	92.7%
Residues in additional allowed regions [a,b,l,p]	6	7.3%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
	---	------
Number of non-glycine and non-proline residues	82	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	10	
Number of proline residues	5	

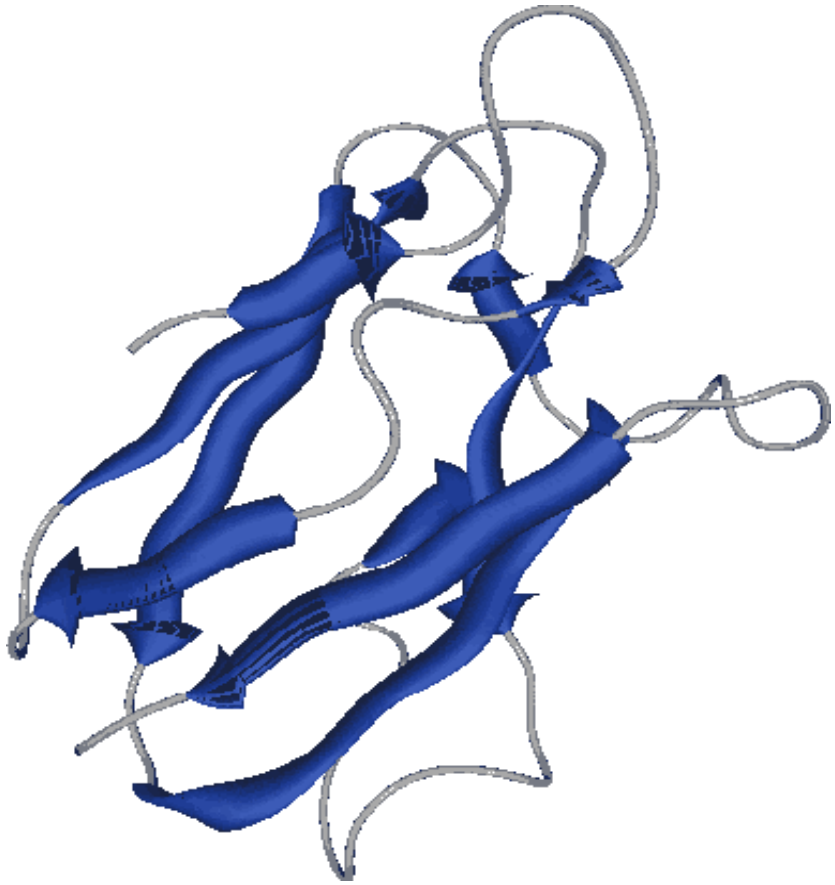
Total number of residues	99	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

1plc_01.ps

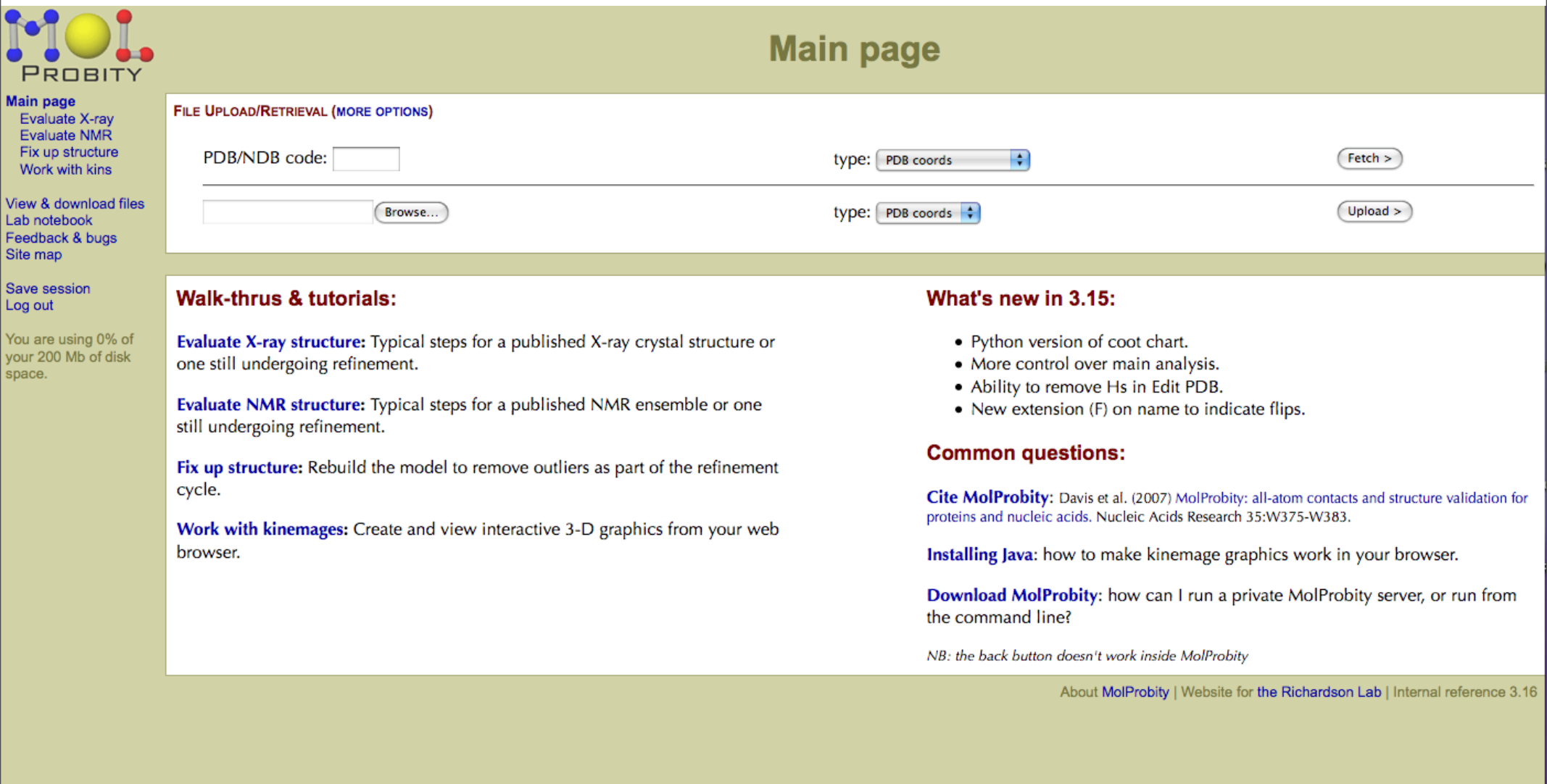
Let's find the wrong structure!

1PLC Electron transport protein



Check your structure/model

http://molprobity.biochem.duke.edu/



Threading

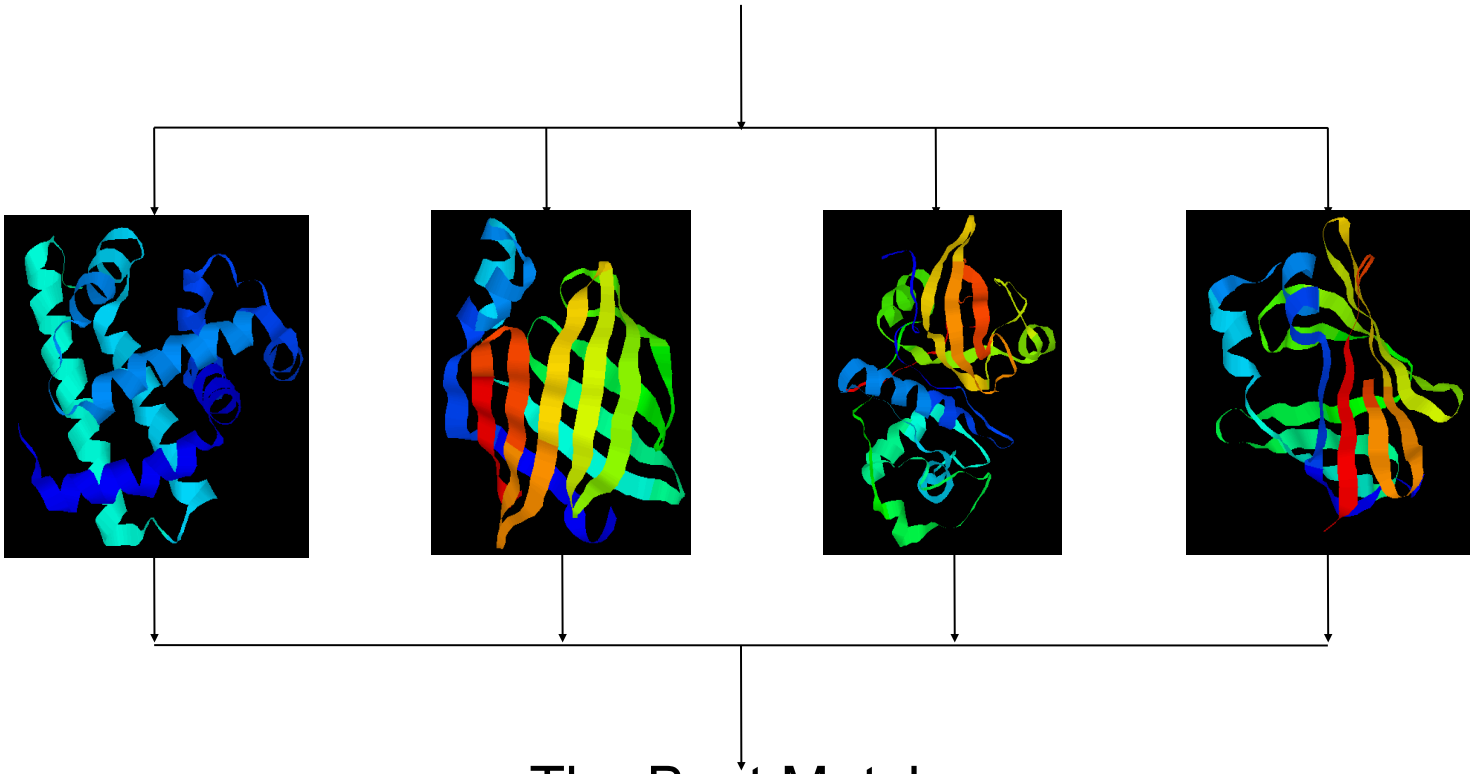
- Threading is a method for the computational prediction of protein structure from protein sequence.

- The query sequence is “threaded” through the backbone structure of a known protein structure. Then, the structure resulting from placing the amino acids of the query sequence in the positions of the model protein structure is evaluated according to the interactions between amino acid side chains and using a solvation potential.

Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) “A new approach to protein fold recognition”. Nature. 358, 86-89

Protein Threading

Query Sequence: **DRVYIHPF**A**DRVYIHPF**A



The Best Match

Threading – Energy Function

MTYKLILNGKTKGETTTEAVDAATAEKVFQYANDNGVDGEWTYTE

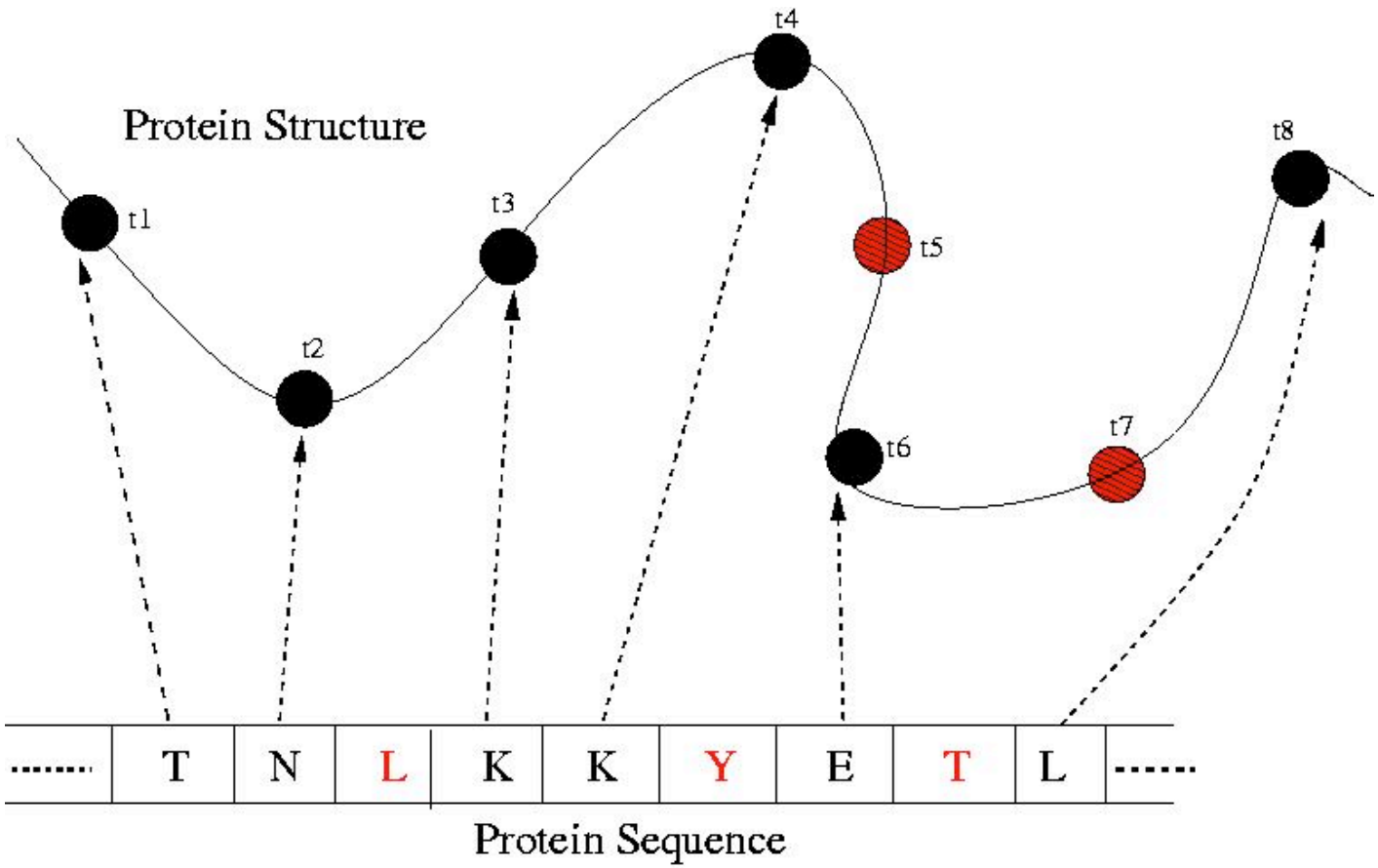
how preferable to put two particular residues nearby: E_p

how well a residue fits a structural environment: E_s

alignment gap penalty: E_g

total energy: E_p + E_s + E_g

Threading Example



Positions or residues in red are gaps

Scoring: Pairwise Potential

occurring probability of a and b with distance < cut-off

$$PairwisePotential(a,b) = -\log \frac{P(a,b)}{P(a)P(b)}$$

occurring probability of amino acid a

occurring probability of amino acid b

Learning Outcomes

- How can we predict the 3-D structure of a protein? When is it possible?
- Comparative homology modelling: When to use it and how it works
- Model and template structure evaluation
- Threading