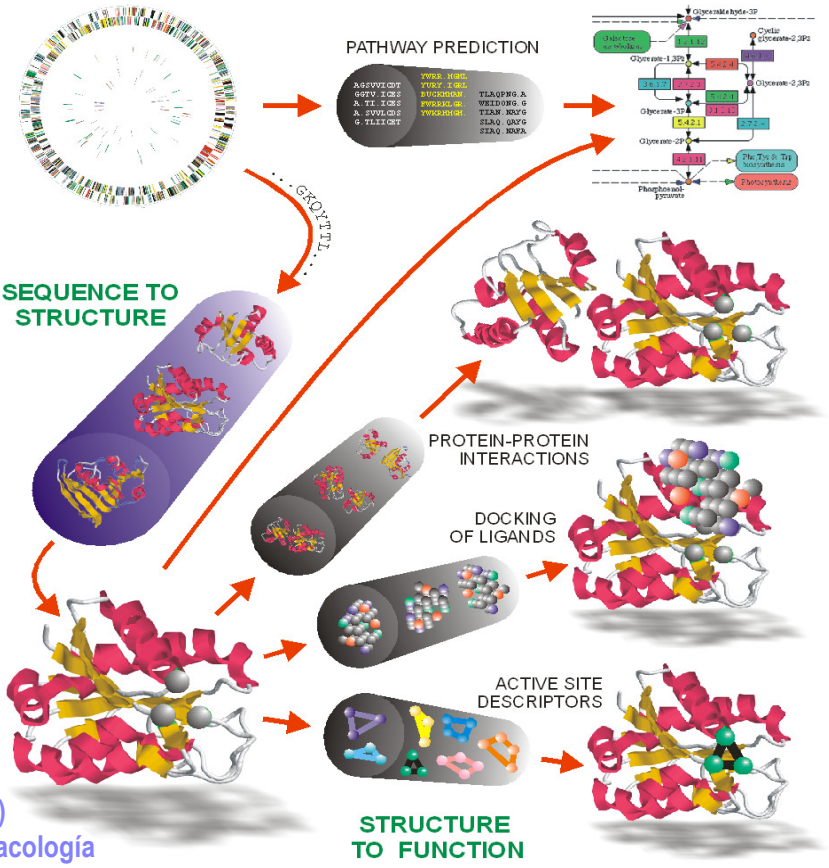
Master Dianas Terapéuticas en Señalización Celular: Investigación y Desarrollo

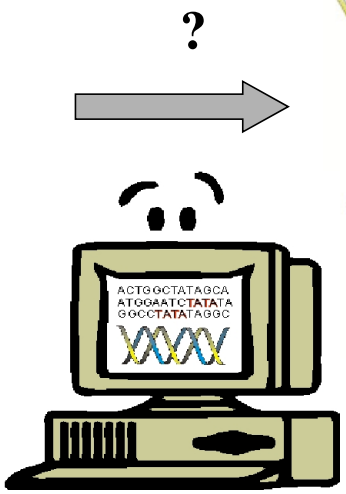
Modelado de proteínas por homología

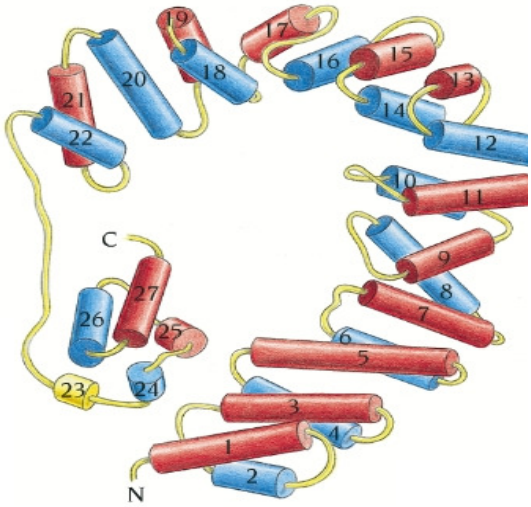


Federico Gago
(federico.gago@uah.es)
Departamento de Farmacología

Structure Prediction

GPSRYIV…

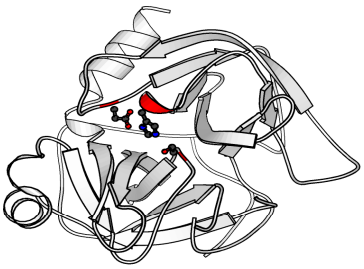




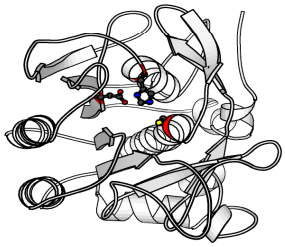
Problems with Structure-Based Function Predictions

Similar Function
Different Fold

Chymotrypsin

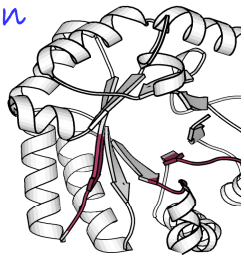


Subtilisin

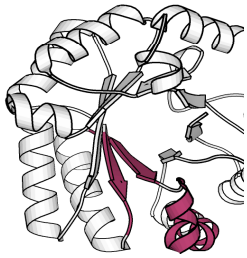


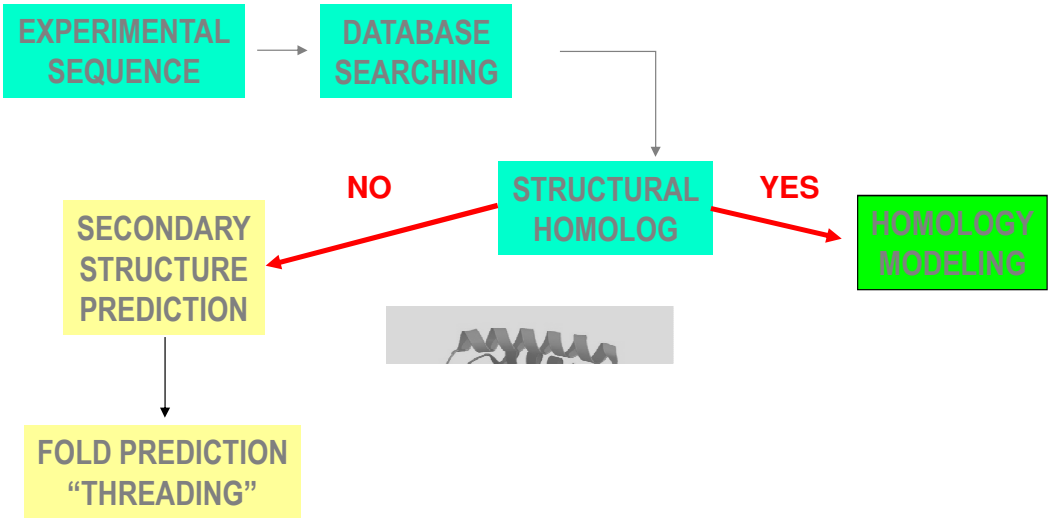
Similar Fold
Different Function

Dehydratase



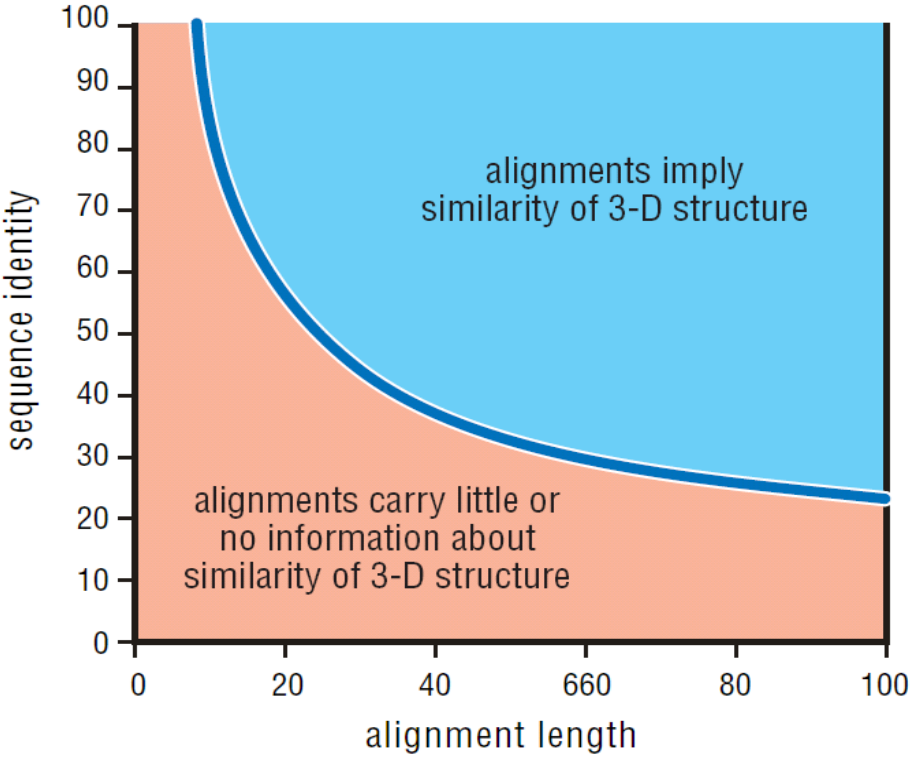
Hydrolase





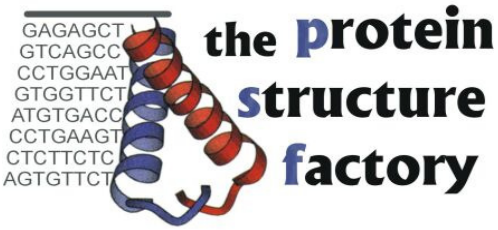
FINAL STRUCTURE ?

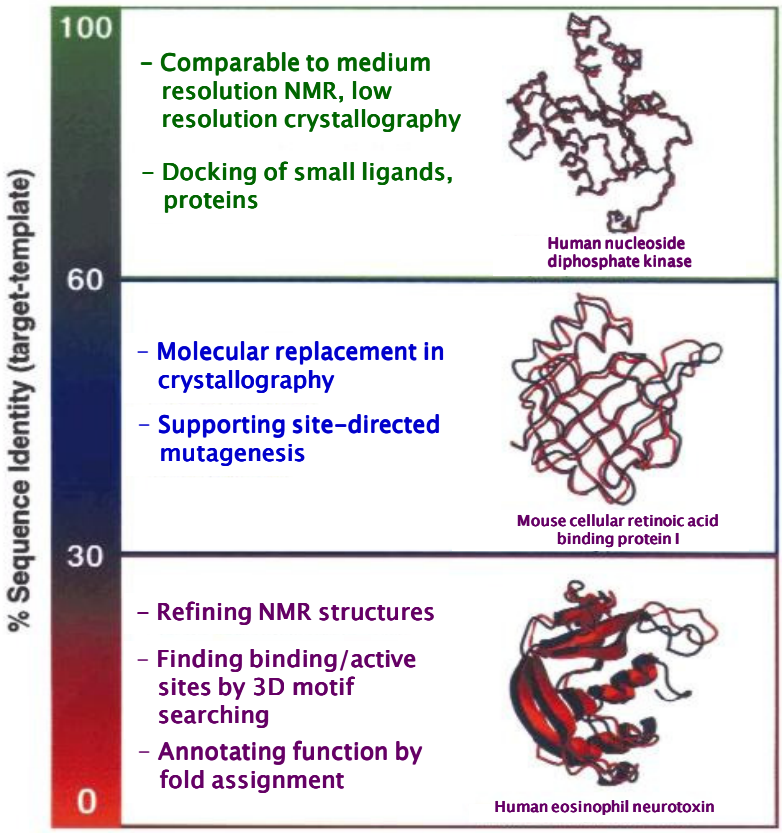
Homology Modelling: *a computational method for modelling the structure of a protein based on its sequence similarity to one or more other proteins of known structure.*



Comparative modelling

The potential use of a comparative model depends on its accuracy.





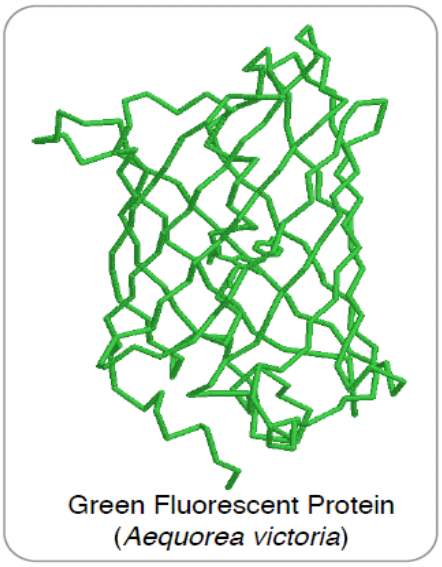
Sali, A. & Kuriyan, J.
Trends Biochem. Sci. 1999, *22*, M20–M24

Sample models and corresponding experimental structures

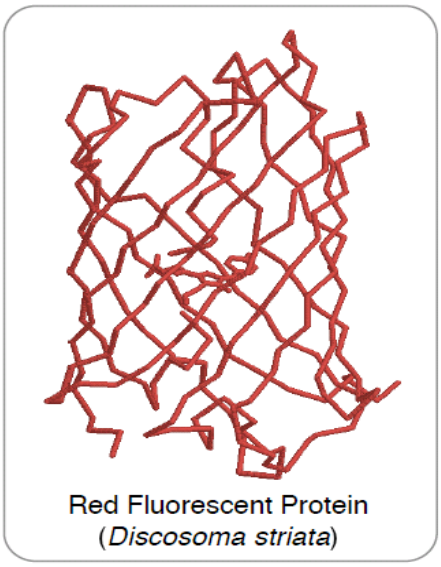
LOW SEQUENCE IDENTITY DOES NOT NECESSARILY IMPLY LOW ***STRUCTURAL HOMOLOGY***

GFP/RFP: 25% ID

```
GFP   MGKGEELFTGVVPILVELDGDVNGHKFSV
RFP   MRSSKNVIKEFMRFKVRMEGTVNGHEFEI

GFP   SGEGEGDATYGKLTLKFICTT.GKLPVPW
RFP   EGEGEGRPYEGHNTVKLKVTKGGPLPFAW

GFP   PTLVTTFSYGVQCFSRYPDHMKRHDFFKS
RFP   DILSPQFQYGSKVYVKHPADI..PDYKKL

GFP   AMPEGYVQERTIFFKDDGNYKTRAEVKFE
RFP   SFPEGFKWERVMNFEDGGVVTVTQDSSLQ

GFP   GDTLVNRIELKGIDFKEDGNILGHK.LEY
RFP   DGCFIYKVKFIGVNFPSDGPVMQKKTMGW

GFP   NYNSHNVYIMADKQKNGIKVNFKIRHNIE
RFP   EASTERLYPRDGVLKGEIHKALKLK....

GFP   DGSVQLADHYQQNTPIGDGPVLLPDNHYL
RFP   DGGHYLVEFKSIY..MAKKPVQLPGYYYV

GFP   STQSALSKDPNEKRDHMVLLEFVTAAGIT
RFP   DSKLDITSH....NEDYTIVEQYERTEGR

GFP   HGMDELY
RFP   HHLF
```



Green Fluorescent Protein
(*Aequorea victoria*)



Red Fluorescent Protein
(*Discosoma striata*)

Comparative modeling

✓ helps to bridge the gap between the available sequence and structure information

✓ is based on the general observation that evolutionarily related sequences have similar three-dimensional structures

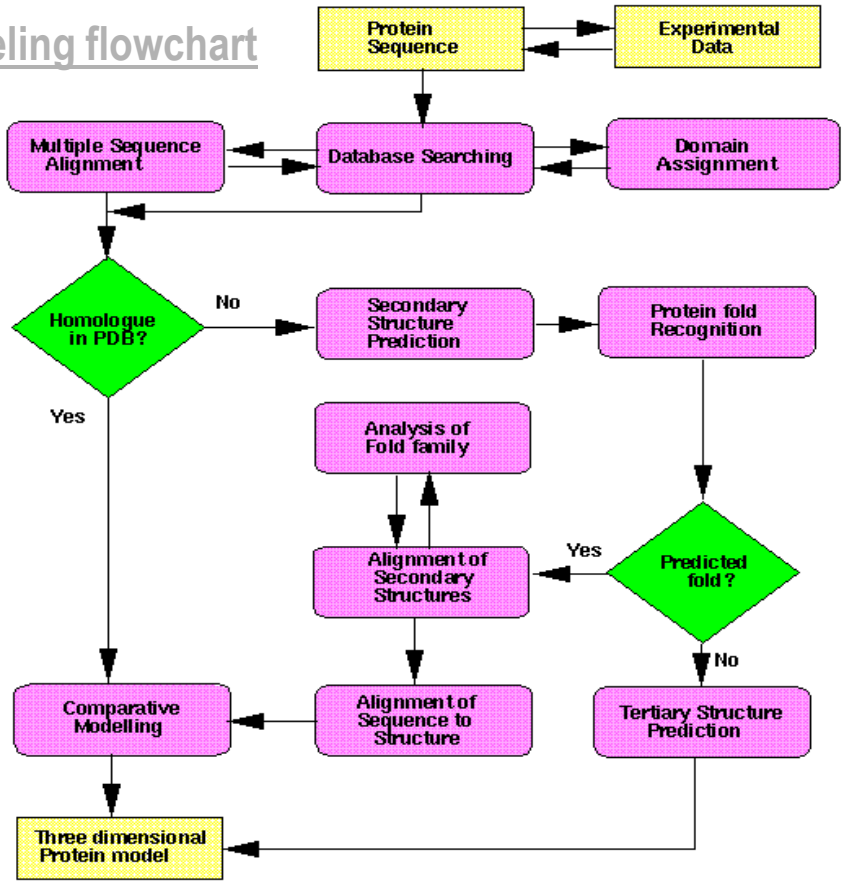
✓ allows building of a three-dimensional model of a protein of interest (target) from related protein(s) of known structure [template(s)] that share statistically significant sequence similarity.

Comparative modeling

Several consecutive steps are usually repeated iteratively until a satisfactory model is obtained:

- ✓ finding suitable template protein(s) related to the target
- ✓ aligning target and template(s) sequences
- ✓ identifying structurally conserved regions
- ✓ predicting structurally variable regions, including insertions and missing N and C termini
- ✓ modelling sidechains
- ✓ refining and evaluating the resulting model.

Comparative modeling flowchart



COMPARATIVE MODELLING

CPHmodels Tools (http://www.cbs.dtu.dk/services/CPHmodels/)

Sowhat: A neural network based method to predict contacts between C-alpha atoms from the amino acid sequence.

RedHom: A tool to find a subset with low sequence similarity in a database.

Databases: Subsets of the Brookhaven Protein Data Bank (PDB) database with low sequence similarity produced using the RedHom tool.

SDSC1 (http://cl.sdsc.edu/hm.html)

Sequence similarity search using intermediate sequence search concept.

SWISS-MODEL (http://www.expasy.ch/swissmod/SWISS-MODEL.html)

An Automated Comparative Protein Modelling Server

3D-JIGSAW (http://www.bmm.icnet.uk/servers/3djigsaw/)

Server that builds three-dimensional models for proteins based on homologues of known structure

SWISS-MODEL

An Automated Comparative Protein Modelling Server

http://swissmodel.expasy.org//SWISS-MODEL.html

MENU

Modeling requests:

- First Approach mode
- Alignment Interface
- Project (optimise) mode
- Oligomer modeling
- GPCR mode

Model Database

- SWISS-MODEL Repository, a database for theoretical protein models.

Interactive tools

- DeepView - Swiss-PdbViewer, a tool for viewing and manipulating protein structures and models.
- SWISS-MODEL Workspace, an interactive working environment for protein structure modelling and assessment.
- Lookup the ExPDB template codes accessible to SWISS-MODEL.
- Search the template sequences accessible to SWISS-MODEL.
- Examples using SWISS-MODEL and the Swiss-PdbViewer.
- ANOLEA Protein structure quality check (atomic non-local environment assessment)
- News from Swiss Model.

Other links

- 3DCrunch, a very large scale protein modeling project.
- Other Web-based Comparative Protein Modeling Servers.
- Course on protein structure and comparative modeling.
- PHYRE, fold recognition server at the ICRF.
- PredictProtein, Burkhard Rost's sequence analysis and structure prediction server.

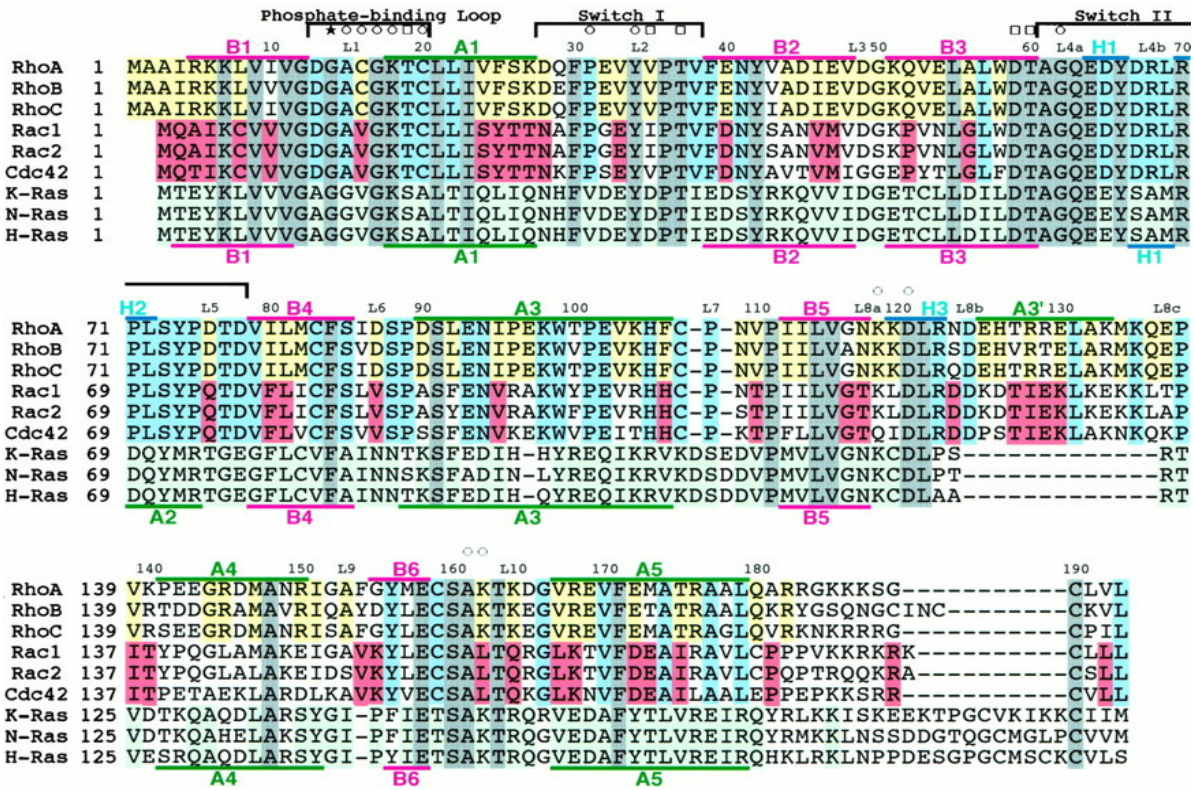
Comparative modeling

Most crucial determinants of final model quality:

- ✓ optimal use of structural information from available templates

- ✓ correctness of sequence-to-structure alignment

Multiple alignments: *how can similarity be quantified?*



Finding suitable template protein(s) related to the target

http://www.ncbi.nlm.nih.gov/BLAST

Simple pairwise BLAST alignment against PDB

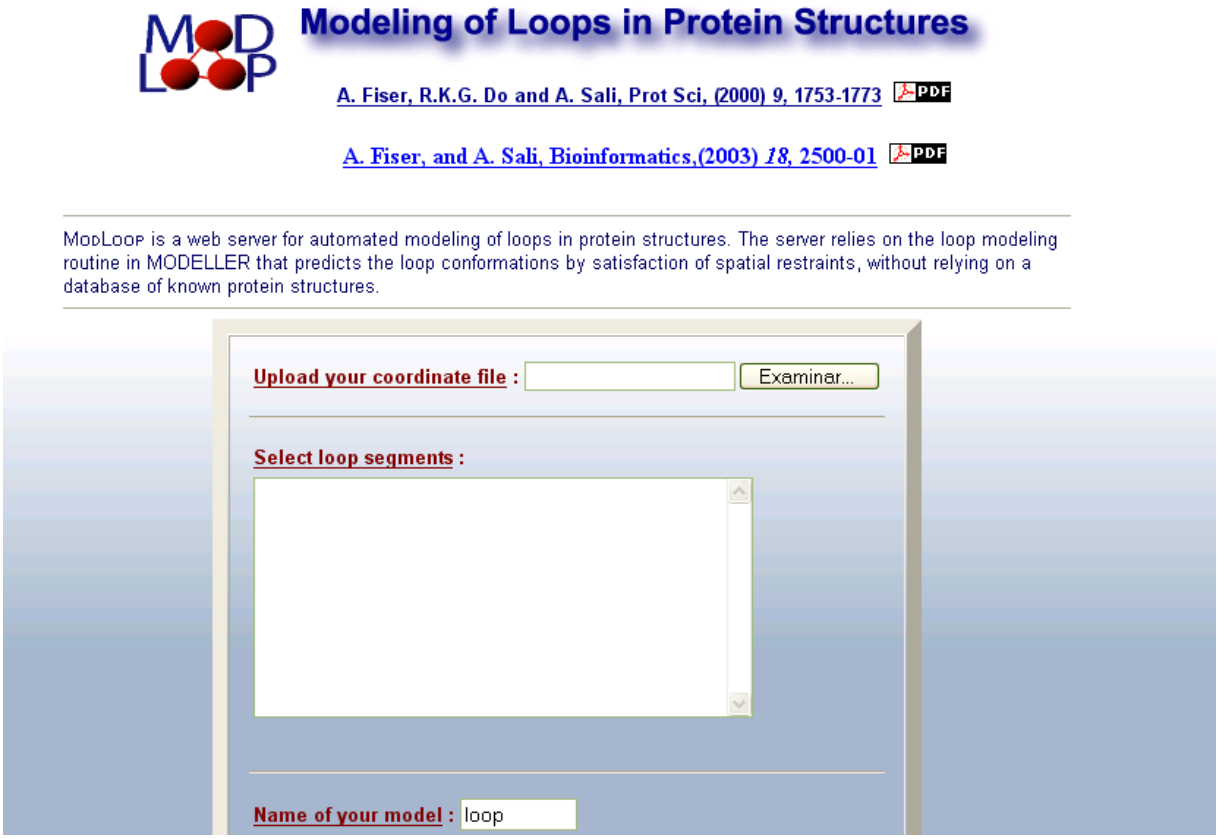
PSI-BLAST, etc

Profile-profile comparisons *NAR* (2005) 33:1874-1891

Aligning target and template(s) sequences

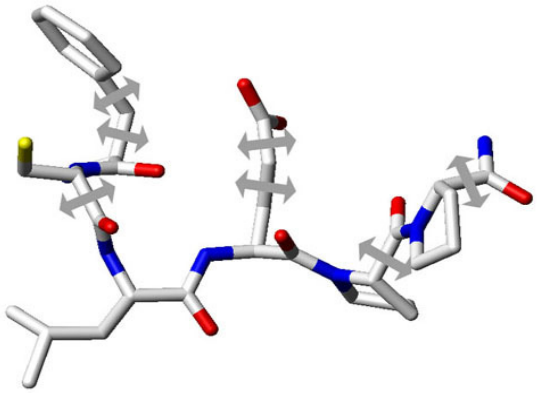
✓ Advantages of consensus strategies based on multiple templates or protein fragment recombination

✓ Benefits from extensive literature searches for any available biochemical information (mutations, catalytic residues, etc) that can lead to alignment anchors and improve the sequence-structure mapping in questionable regions



http://alto.compbio.ucsf.edu/modloop/

Modeling sidechains



✓ MaxSprout: a fast database algorithm for generating protein backbone and side chain co-ordinates from a Cα trace. The backbone is assembled from fragments taken from known structures. Side chain conformations are optimised in rotamer space using a rough potential energy function to avoid clashes

L. Holm, C. Sander (1991) *J. Mol. Biol.* 218:183-194

http://www.ebi.ac.uk/maxsprout/

Modeling sidechains

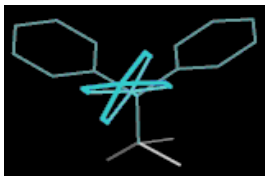
✓ Even in dihedral angle space, the conformational space accessible to all sidechains of a protein remains very large.

✓ In most existing methods for modelling sidechain conformation, sidechain conformation space is discretized, i.e. a sidechain is allowed to adopt only a discrete set of conformations.

✓ This approximation is based on the observation that, in high-resolution experimental protein structures, side-chains tend to cluster around a discrete set of favored conformations, known as rotamers.

✓ In most cases, these rotamers correspond to local minima on the side-chain potential energy map.

For a review:	**Vasquez, M. Modeling sidechain conformation. *Curr. Opin. Struct. Biol.* 6, 217-221 (1996)**



Protein sidechain conformation - Rotamer libraries

- Ponder JW and Richards, FM. Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol*. 193, 775-791 (1987).

http://www.fccc.edu/research/labs/dunbrack/sidechain/ponder_richards.rot

- Dunbrack, RL and Karplus, M. Backbone-dependent rotamer library for proteins : application to side-chain prediction. *J. Mol. Biol*. 230, 543-574 (1993). Dunbrack, RL and Cohen, FE. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci*. 6, 1661-1681 (1997).

http://www.fccc.edu/research/labs/dunbrack/sidechain.html

- Tuffery, P, Etchebest, C, Hazout, S and Lavery, R. A new approach to the rapid-determination of protein side-chain conformations. *J. Biomol. Struct. Dyn*. 8, 1267-1289 (1991).

http://bioserv.rpbs.jussieu.fr/doc/Rotamers.html

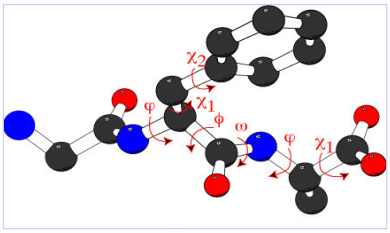
- DeMaeyer, M, Desmet, J and Lasters, I. All in one: A highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination. *Folding & Des*. 2, 53-66 (1997).

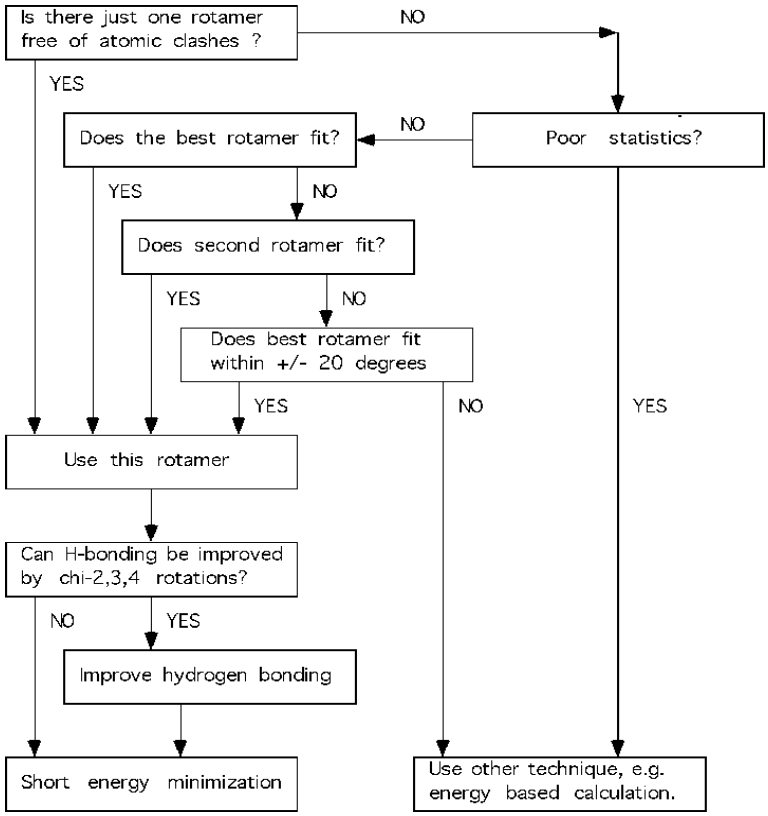
http://www.fccc.edu/research/labs/dunbrack/sidechain/demaeyer.rot

- SC Lovell, JM Word, JS Richardson and DC Richardson. The Penultimate Rotamer Library" *Proteins: Structure Function and Genetics* 40, 389-408 (2000).

http://kinemage.biochem.duke.edu/databases/rotamer.php

Decision scheme for the prediction of point mutant structures



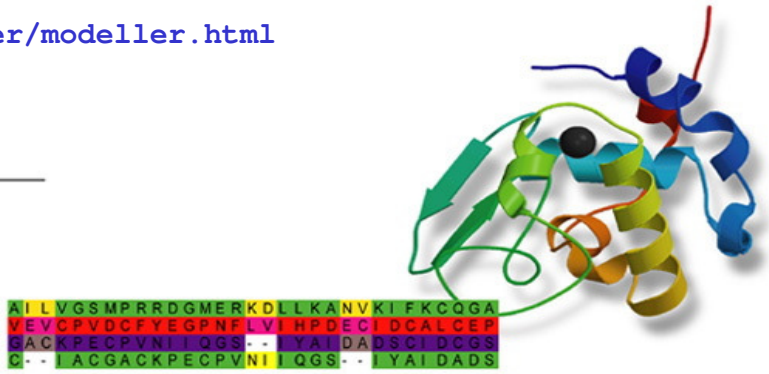


http://swift.cmbi.kun.nl/swift/whatif/courses.notes.html

http://salilab.org/modeller/modeller.html

Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



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. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints, and can perform many additional tasks, including *de novo* modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc.

FOLD RECOGNITION & THREADING METHODS

3D_PSSM (http://www.sbg.bio.ic.ac.uk/~3dpssm/)

A Fast, Web-based Method for Protein Fold Recognition using 1D and 3D Sequence Profiles coupled with Secondary Structure and Solvation Potential Information.

PHYRE (http://www.sbg.bio.ic.ac.uk/~phyre/)

Protein Homology/analogY Recognition Engine

FUGUE (http://www-cryst.bioc.cam.ac.uk/~fugue/)

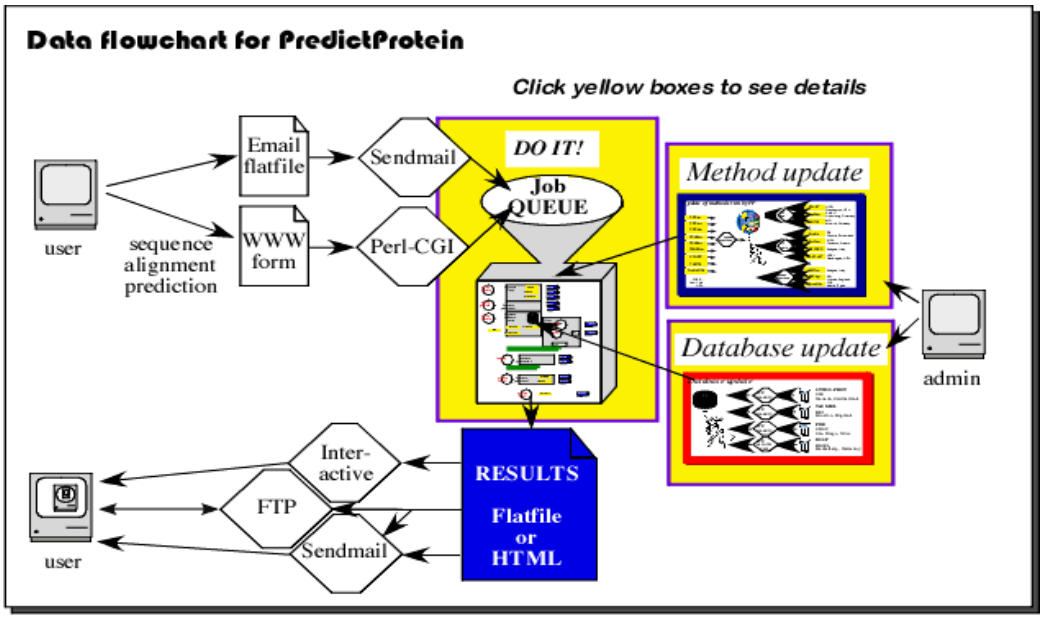
Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties

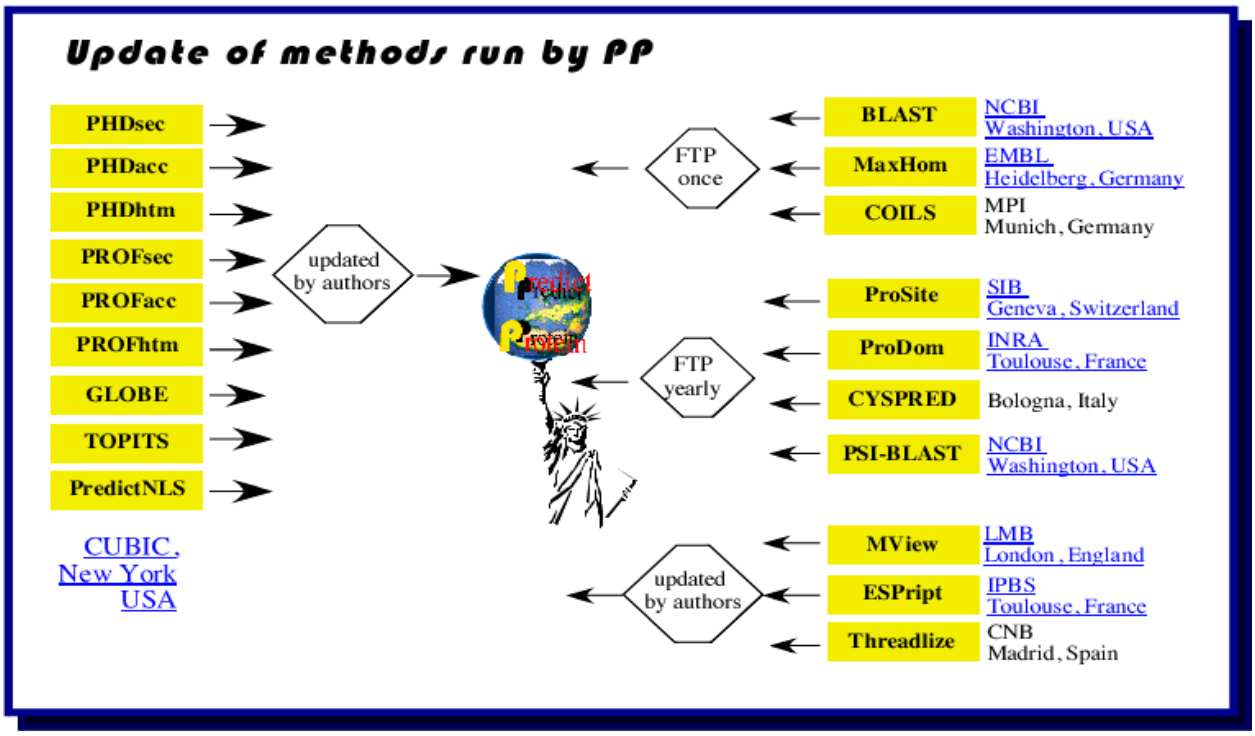
LOOPP (http://ser-loopp.tc.cornell.edu/loopp.html)

Learning, Observing and Outputting Protein Patterns (LOOPP)

Superfamily (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/)

Protein domain assignments to SCOP structural superfamilies using a hidden Markov model library.





Bioinformatics Unit UCL

McGuffin home>
Bryson home>
Jones home>

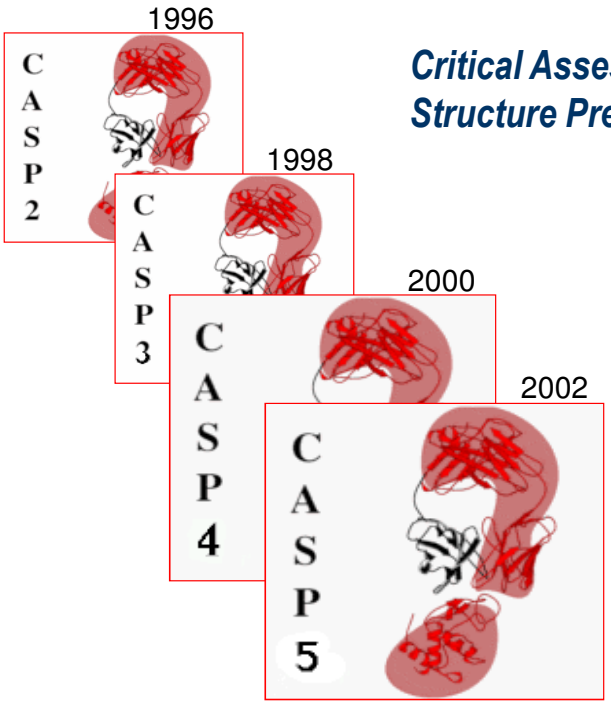
The PSIPRED Protein Structure Prediction Server

Liam J. McGuffin, Kevin Bryson & David T. Jones

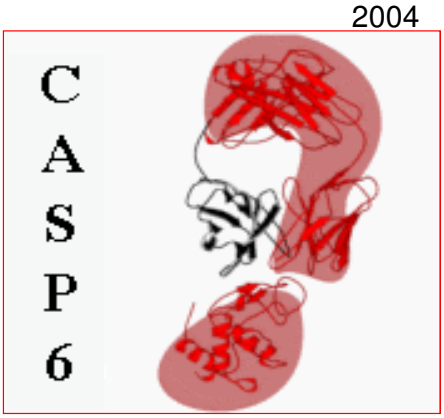
http://bioinf.cs.ucl.ac.uk/psipred/

The PSIPRED protein structure prediction server allows you to submit a protein sequence, perform a prediction of your choice and receive the results of the prediction via e-mail. You may select one of three prediction methods to apply to your sequence:

PSIPRED - a highly accurate method for protein secondary structure prediction,
MEMSAT3 - our widely used transmembrane topology prediction method and
GenTHREADER - a sequence profile based fold recognition method



Critical Assessment of Techniques for Protein Structure Prediction

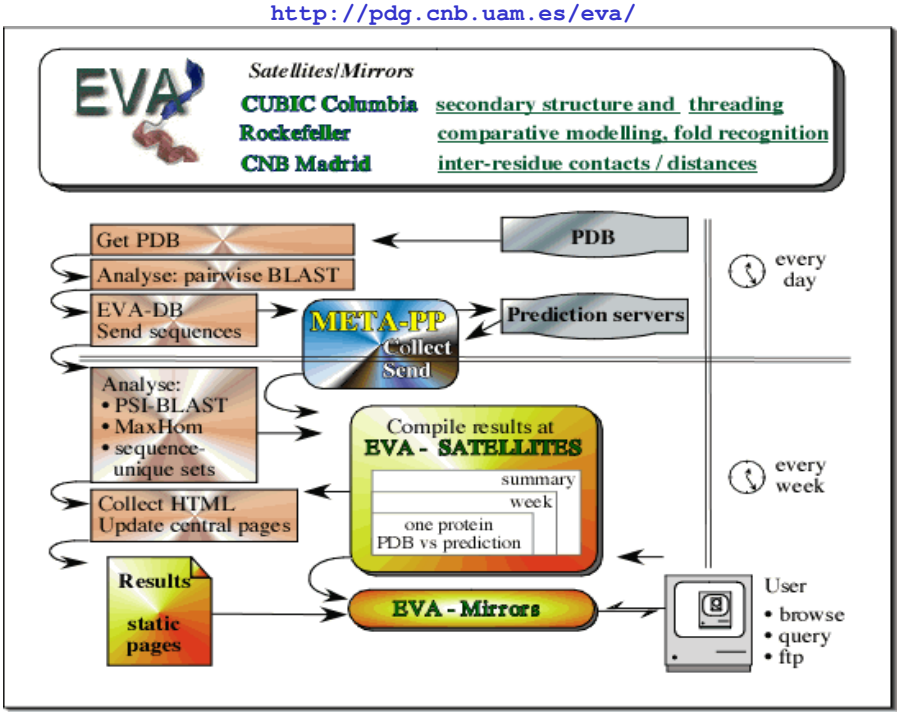


ASILOMAR, USA

GAETA, ITALY

EVA: continuous automatic evaluation of protein structure prediction servers

http://cubic.bioc.columbia.edu/eva/
http://pipe.rockefeller.edu/~eva/
http://pdg.cnb.uam.es/eva/



LiveBench:

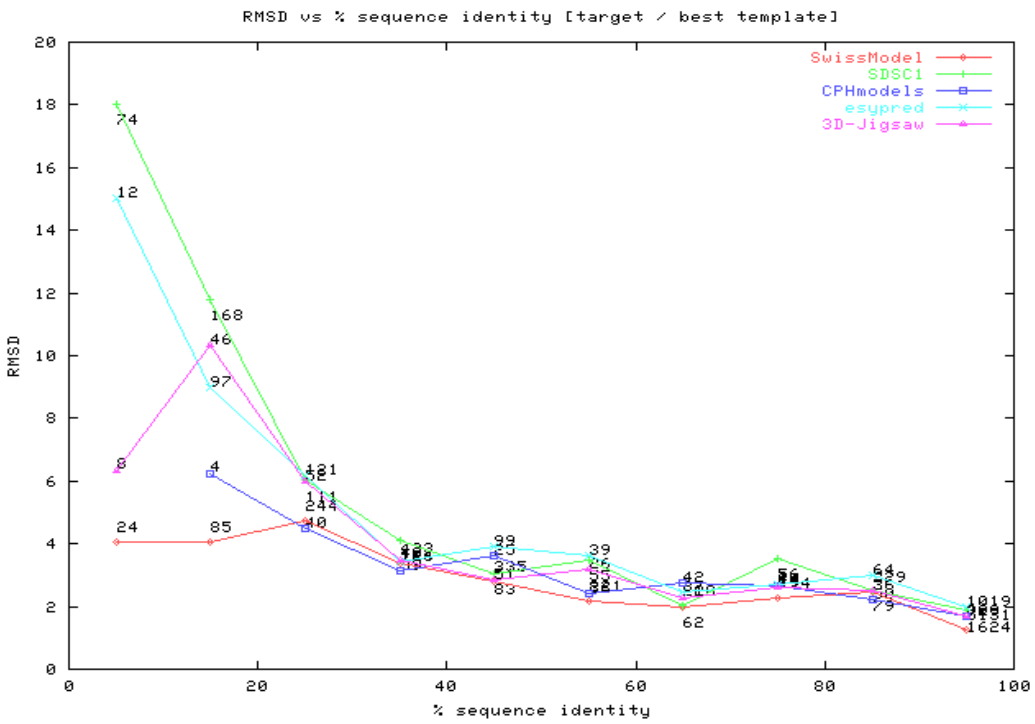
Continuous Benchmarking of Structure Prediction Servers

http://bioinfo.pl/meta/livebench.pl

Two main goals:

✓ The program provides simple evaluation of the structure prediction servers from the point of view of a potential user. The evaluation of sensitivity and specificity of the available servers can help the user to develop sequence analysis strategies and to assess the confidence of the obtained predictions.

✓ The program offers a simple weekly procedure for the prediction service providers, which can help to locate possible problems and tune the methods for best performance.



Meta-servers

✓ are servers that use the results of other autonomous servers to produce a consensus prediction

✓ outperform all the individual autonomous servers

✓ cannot run independently, explicitly requiring as input the predictions of at least one other participating server

✓ attempt to automate the process of selecting the top model

Meta-servers

- PCONS/PMOD series

  `http://www.sbc.su.se/~bjorn/Pcons5`

- 3D-SHOTGUN: INUB + SP3 + PROSPECTOR

  `http://inub.cse.buffalo.edu`

- 3D-JURY series

  `http://bioinfo.pl/meta/`

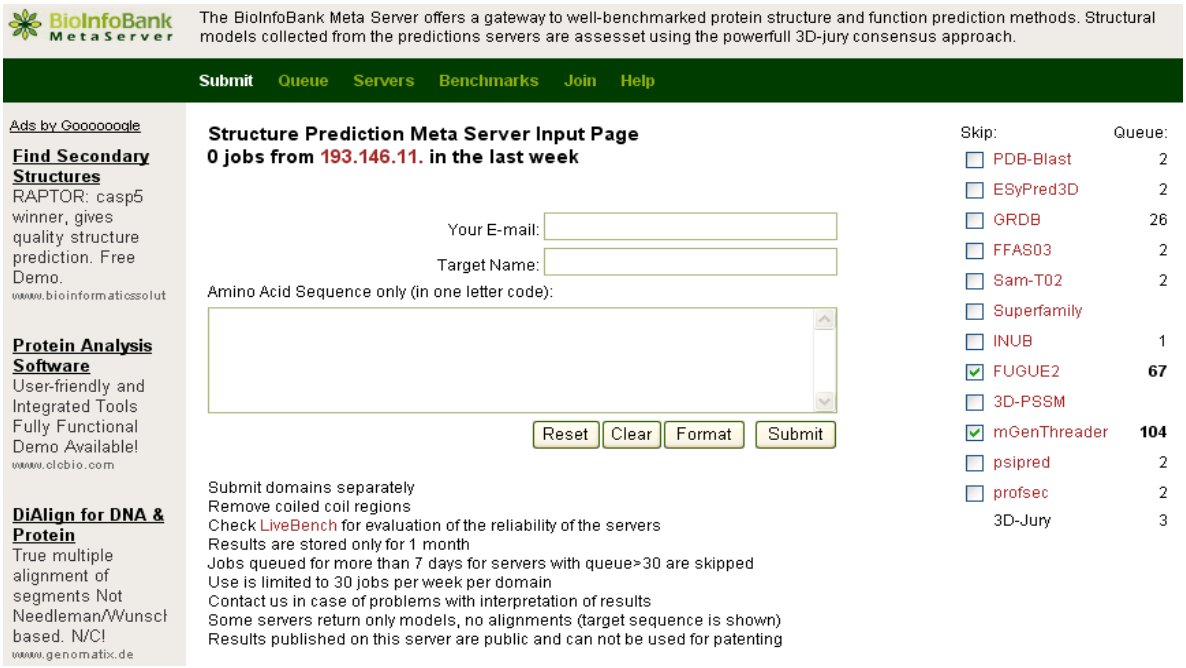
- PROTINFO

  `http://protinfo.compbio.washington.edu`

- Meta-BASIC, ORFeus, FFAS03, SP3, Robetta...

3D-jury consensus approach

http://bioinfo.pl/meta/



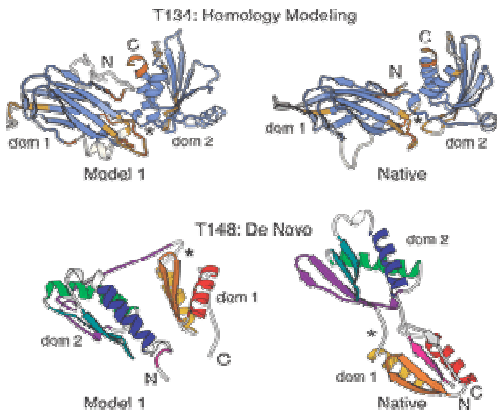
The Structure Prediction Meta Server provides access to various fold recognition, function prediction and local structure prediction methods.

www.bakerlab.org

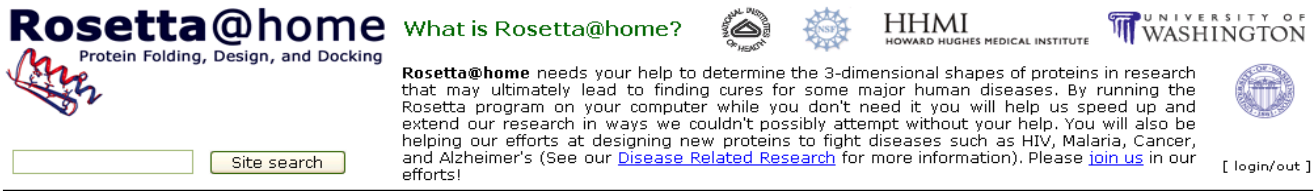
ROBETTA

Full-chain Protein Structure Prediction Server

http://robetta.bakerlab.org/index.html



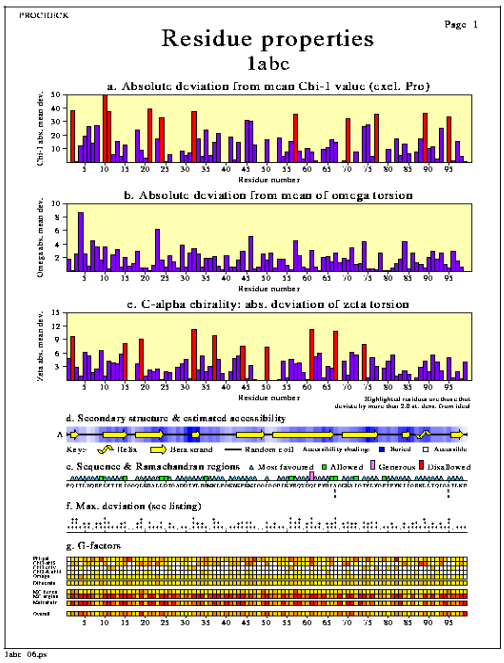
ROBETTA provides both *ab initio* and comparative models of protein domains. It uses the ROSETTA fragment insertion method [Simons et al. J Mol Biol 1997;268:209-225]. Comparative models are built from Parent PDBs detected by UW-PDB-BLAST, FFAS03, or 3DJury-A1 and aligned by the K*SYNC alignment method. Loop regions are assembled from fragments and optimized to fit the aligned template structure. The procedure is fully automated.

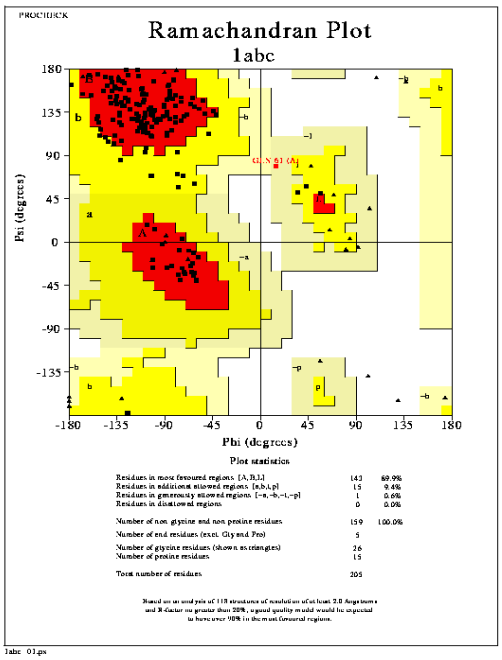


Structure Validation Servers

- PROCHECK
  - http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
- WHAT IF
  - http://swift.cmbi.kun.nl/WIWWWI/
- Verify3D
  - http://www.doe-mbi.ucla.edu/Services/Verify_3D/
- VADAR
  - http://redpoll.pharmacy.ualberta.ca

Procheck





The WHAT IF Web Interface

http://swift.cmbi.kun.nl/WIWWWI/

Name check: checks the nomenclature of torsion angles.

Coarse Packing Quality Control: checks the normality of the local environment of amino acids

Anomalous bond lengths: lists bond lengths that deviate more than 4 sigma from normal.

Planarity: checks if planar groups are planar enough.

Fine Packing Quality Control: checks the normality of the local environment of amino acids

Collisions with symmetry axes: lists atoms that are too close to symmetry axes.

Hand check: lists atoms with a chirality that deviates more than 4 sigma from normal.

Ramachandran plot evaluation: determines the quality of a Ramachandran plot.

Omega: checks if the distribution of omega angles is normal.

Proline puckering: checks if proline pucker falls in a normal range.

Anomalous bond angles: lists bond angles that deviate more than 4 sigma from normal.

VADAR Version 1.4

Please click here to do multiple chain analysis
Note: VADAR cannot process proteins > 2000 residues

VADAR (Volume, Area, Dihedral Angle Reporter) is a compilation of more than 15 different algorithms and programs for analyzing and assessing peptide and protein structures from their PDB coordinate data. The results have been validated through extensive comparison to published data and careful visual inspection. The VADAR web server supports the submission of either PDB formatted files or PDB accession numbers. VADAR produces extensive tables and high quality graphs for quantitatively and qualitatively assessing protein structures determined by X-ray crystallography, NMR spectroscopy, 3D-threading or homology modelling.

Please cite the following: Leigh Willard, Anuj Ranjan,Haiyan Zhang,Hassan Monzavi, Robert F. Boyko, Brian D. Sykes, and David S. Wishart "VADAR: a web server for quantitative evaluation of protein structure quality" Nucleic Acids Res. 2003 July 1; 31 (13): 3316.3319

For additional information on how to run VADAR or to process multiple chains via VADAR, click this button [HELP]

Select desired PDB file [Browse...]

Note: the uploaded file must be in PDB format in order for this form to work. Refer to the **HELP** button above.

OR Enter PDB accession number [1ad5]
(Please specify the chain e.g. 2TRXB (2TRX chain B), If not specified, the first chain will be processed. e.g. 2TRX)

[Submit] [Clear]