Comparative Protein Structure Prediction

Marc A. Marti-Renom
http://bioinfo.cipf.es/sgu/

Structural Genomics Unit
Bioinformatics Department
Prince Felipe Resarch Center (CIPF), Valencia, Spain

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

DISCLAIMER!

Name	Type[a]	World Wide Web address[b]
DATABASES		
CATH	S	http://www.biochem.ucl.ac.uk/bsm/cath/
DBAli	S	http://www.salilab.org/DBAli/
GenBank	S	http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html
GeneCensus	S	http://bioinfo.mbb.yale.edu/genome
MODBASE	S	http://salilab.org/modbase/
MSD	S	http://www.rcsb.org/databases.html
NCBI	S	http://www.ncbi.nlm.nih.gov/
PDB	S	http://www.rcsb.org/pdb/
PSI	S	http://www.nigms.nih.gov/psi/
Sacch3D	S	http://genome-www.stanford.edu/Sacch3D/
SCOP	S	http://scop.mrc-lmb.cam.ac.uk/scop/
TIGR	S	http://www.tigr.org/tdb/mdb/mdbcomplete.html
TrEMBL	S	http://srs.ebi.ac.uk/
FOLD ASSIGNMENT		
123D	S	http://123d.ncifcrf.gov/
3D-PSSM	S	http://www.sbg.bio.ic.ac.uk/~3dpssm/
BIOINBGU	S	http://www.cs.bgu.ac.il/~bioinbgu/
BLAST	S	http://www.ncbi.nlm.nih.gov/BLAST/
DALI	S	http://www2.ebi.ac.uk/dali/
FASS	S	http://bioinformatics.burnham-inst.org/FFAS/index.html
FastA	S	http://www.ebi.ac.uk/fasta3/
FRSVR	S	http://fold.doe-mbi.ucla.edu/
FUGUE	S	http://www-cryst.bioc.cam.ac.uk/~fugue/
LOOPP	S	http://ser-loopp.tc.cornell.edu/cbsu/loopp.htm
PDB-BLast/FASS	S	http://bioinformatics.ljcrf.edu/pdb_blast/
PHD, TOPITS	S	http://www.predictprotein.org/

http://sgu.bioinfo.cipf.es/home/?page=resources

Summary

- INTRO
- MODELLER
- MOULDER
- MODEL(S) --> FUNCTION
- MODELLER example

Nomenclature

Homology: Sharing a common ancestor, may have similar or dissimilar functions

Similarity: Score that quantifies the degree of relationship between two sequences.

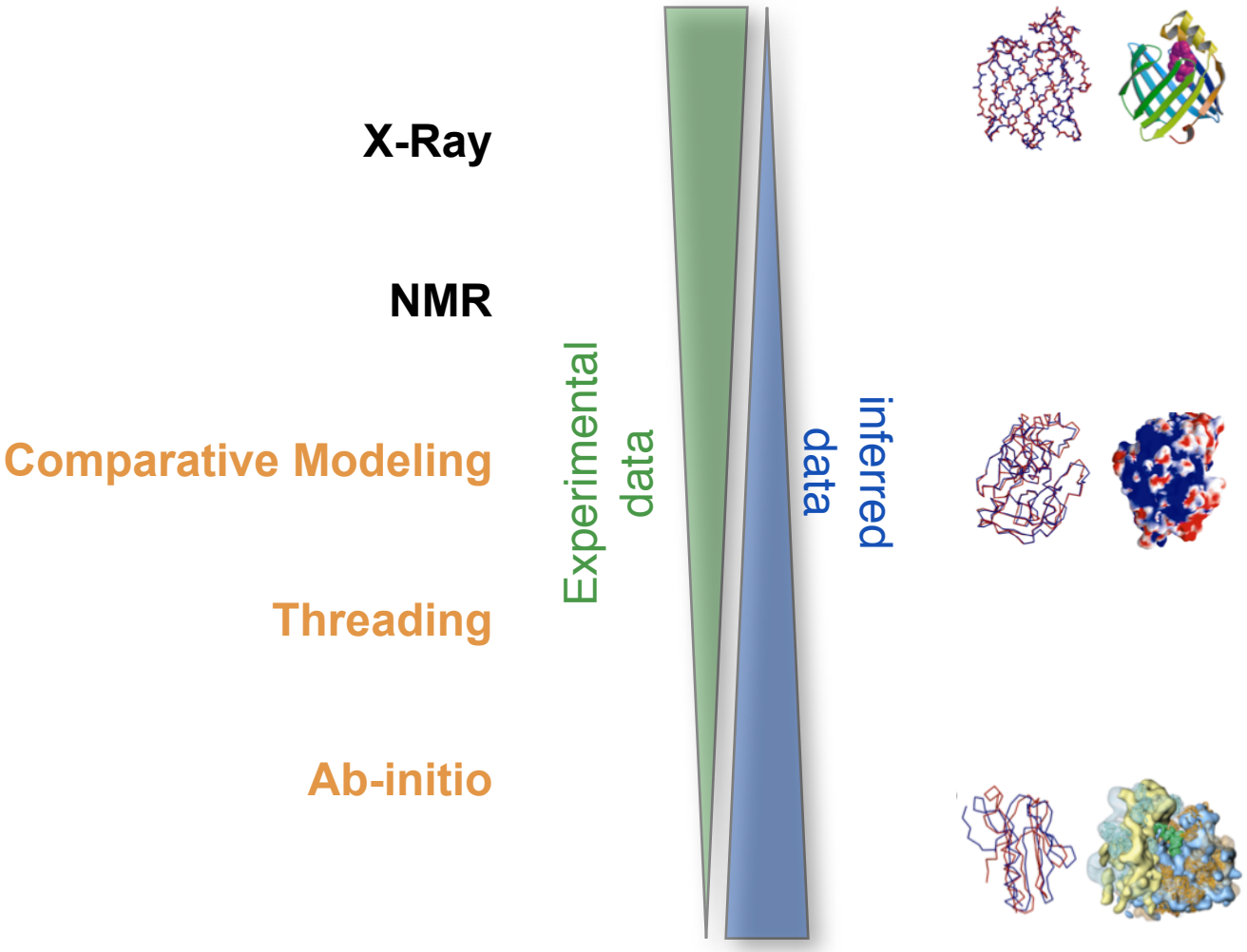
Identity: Fraction of identical aminoacids between two aligned sequences (case of similarity).

Target: Sequence corresponding to the protein to be modeled.

Template: 3D structure/s to be used during protein structure prediction.

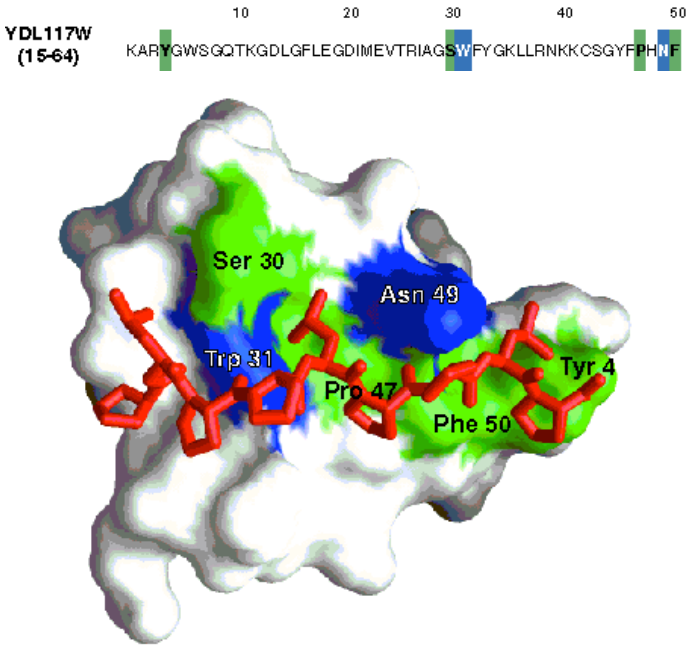
Model: Predicted 3D structure of the target sequence.

protein prediction .vs. protein determination



Why is it useful to know the structure of a protein, not only its sequence?

- The biochemical function (activity) of a protein is defined by its interactions with other molecules.
- The biological function is in large part a consequence of these interactions.
- The 3D structure is more informative than sequence because interactions are determined by residues that are close in space but are frequently distant in sequence.

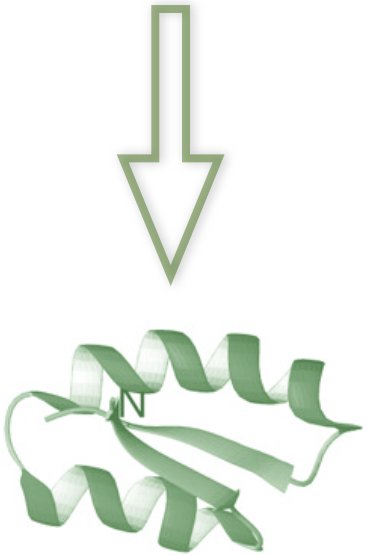


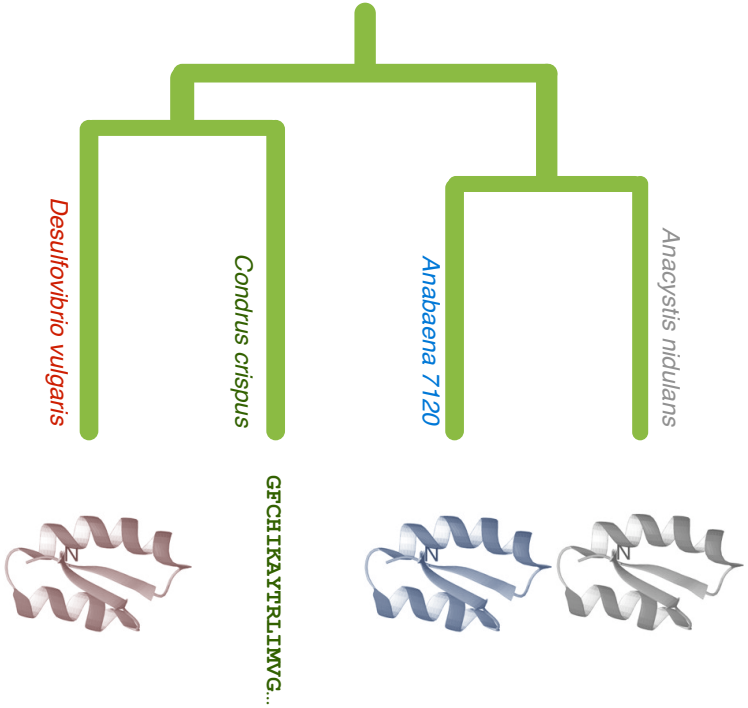
In addition, since evolution tends to conserve function and function depends more directly on structure than on sequence, **structure is more conserved in evolution than sequence**.

The net result is that **patterns in space are frequently more recognizable than patterns in sequence**.

Principles of protein structure

GFCHIKAYTRLIMVG...





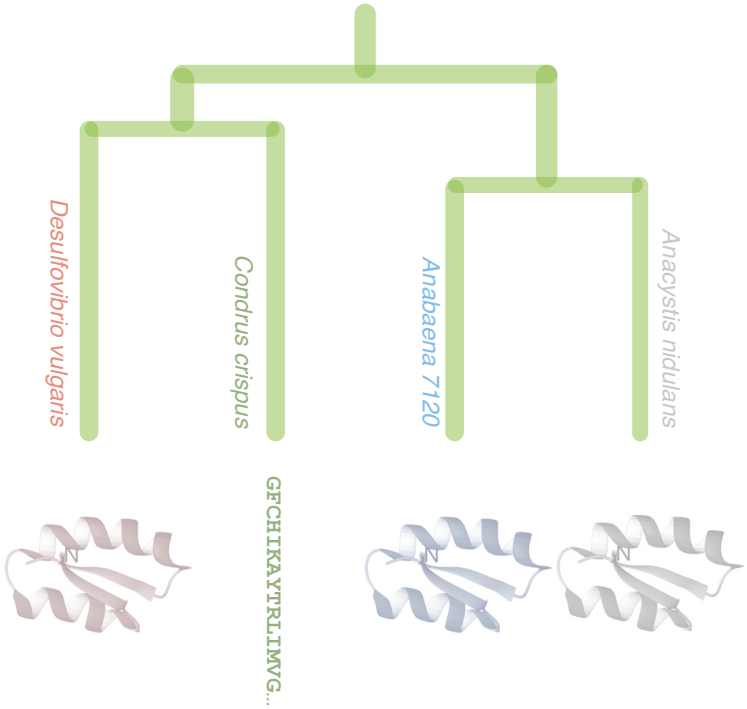
Folding (physics)

Ab initio prediction

Evolution (rules)

Threading
Comparative Modeling

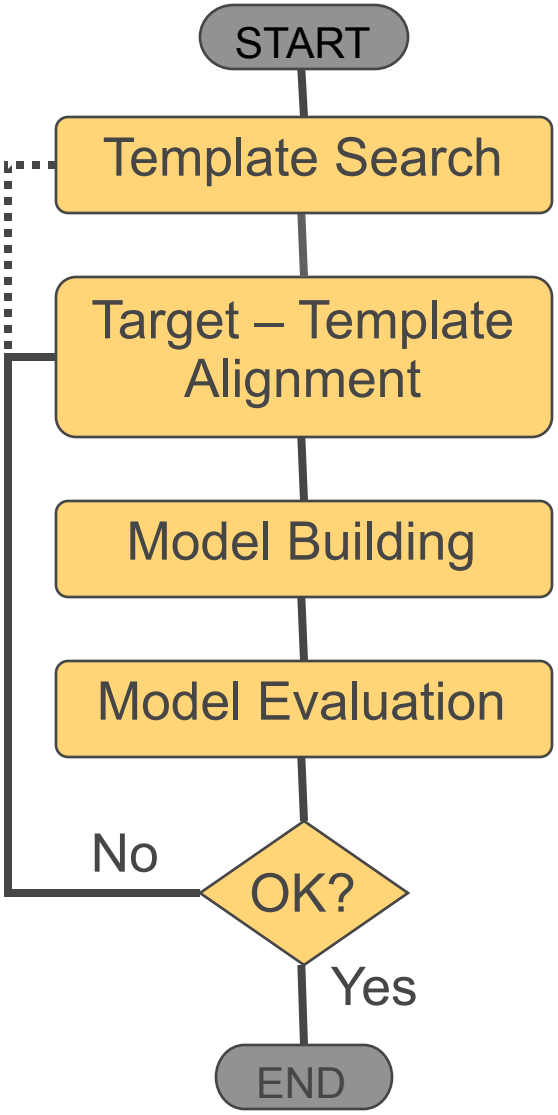
D. Baker & A. Sali. Science 294, 93, 2001.

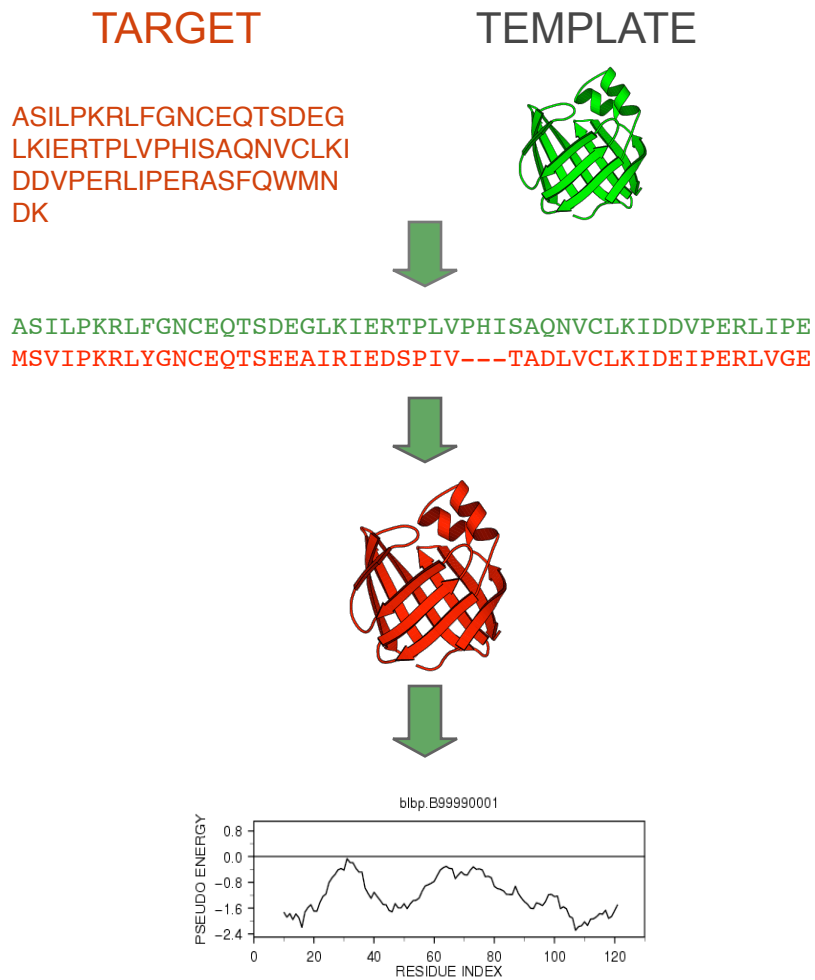


MODELLER

1. N. Eswar, et al. *Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 2008.*
2. M.A. Marti-Renom, et al.. *Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.*
3. A. Sali & T.L. Blundell. *Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.*
4. A. Fiser, R.K. Do, & A. Sali. *Modeling of loops in protein structures, Protein Science 9. 1753-1773, 2000.*

Steps in Comparative Protein Structure Modeling

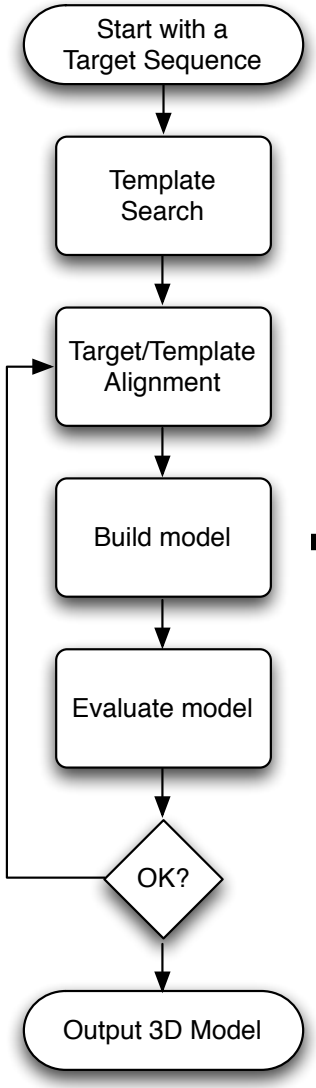




A. Šali, Curr. Opin. Biotech. 6, 437, 1995.
R. Sánchez & A. Šali, Curr. Opin. Str. Biol. 7, 206, 1997.
M. Marti et al. Ann. Rev. Biophys. Biomolec. Struct., 29, 291, 2000.

Comparative modeling by satisfaction of spatial restraints

MODELLER

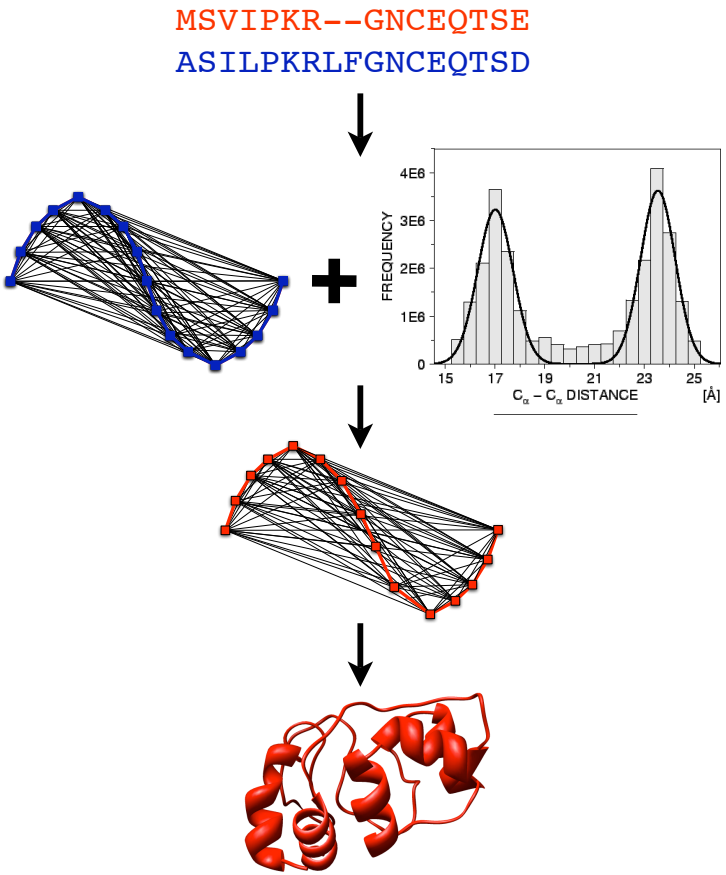


Given an alignment...

extract spatial features from the template(s) and statistics from known structures

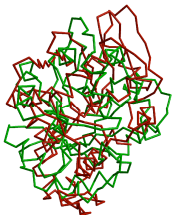
apply these features as restraints on your target sequence

optimize to find the best solution for the restraints to produce your 3D model

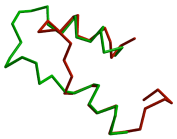


A. Šali & T. Blundell. J. Mol. Biol. 234, 779, 1993.
J.P. Overington & A. Šali. Prot. Sci. 3, 1582, 1994.
A. Fiser, R. Do & A. Šali, Prot. Sci., 9, 1753, 2000.

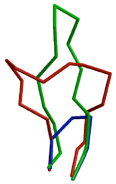
Comparative modeling by satisfaction of spatial restraints Types of errors and their impact



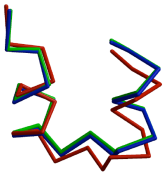
Wrong fold



Miss alignments

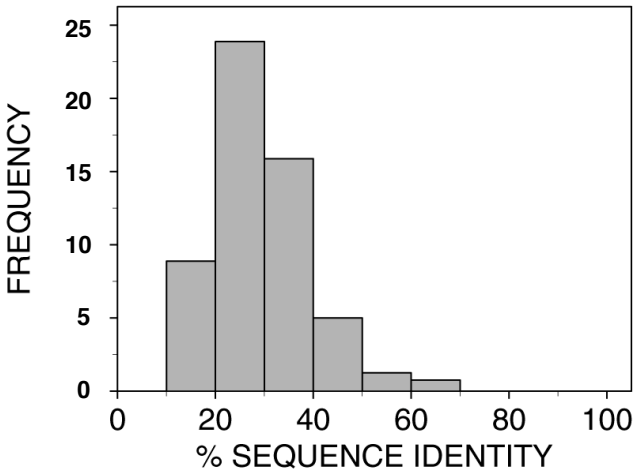


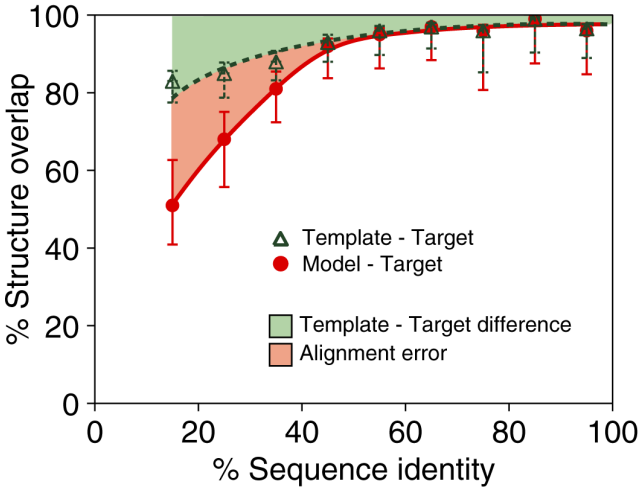
Loop regions



Rigid body distortions

Side-chain packing





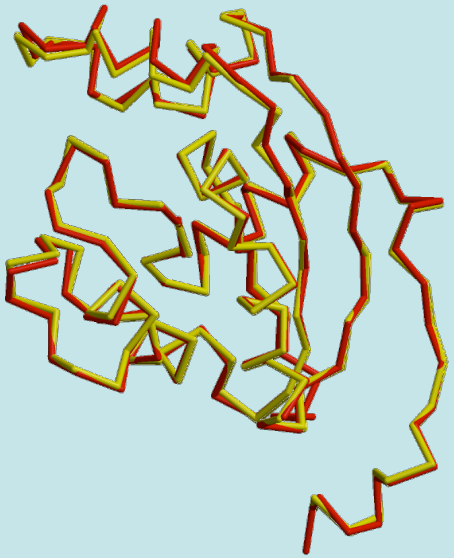
Marti-Renom etal. Ann Rev Biophys Biomol Struct (2000) 29, 291

Model Accuracy

HIGH ACCURACY

NM23
Seq id 77%

Cα equiv 147/148
RMSD 0.41Å



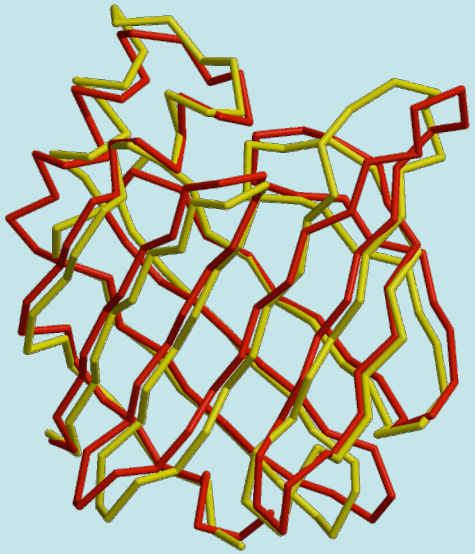
Sidechains
Core backbone
Loops

X-RAY / MODEL

MEDIUM ACCURACY

CRABP
Seq id 41%

Cα equiv 122/137
RMSD 1.34Å

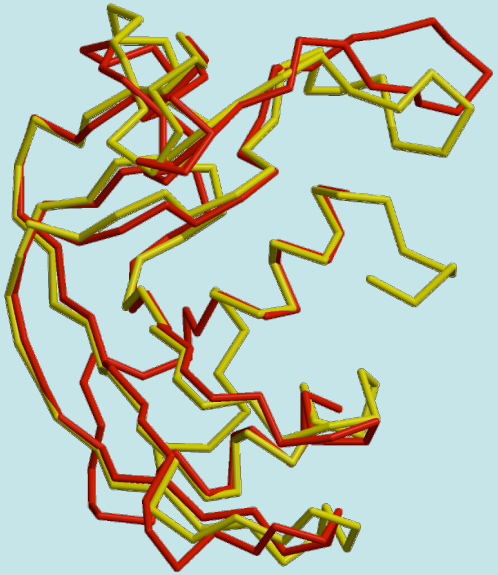


Sidechains
Core backbone
Loops
Alignment

LOW ACCURACY

EDN
Seq id 33%

Cα equiv 90/134
RMSD 1.17Å

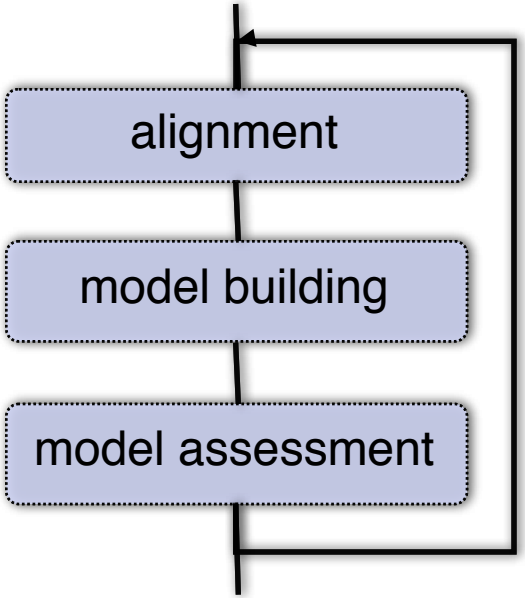


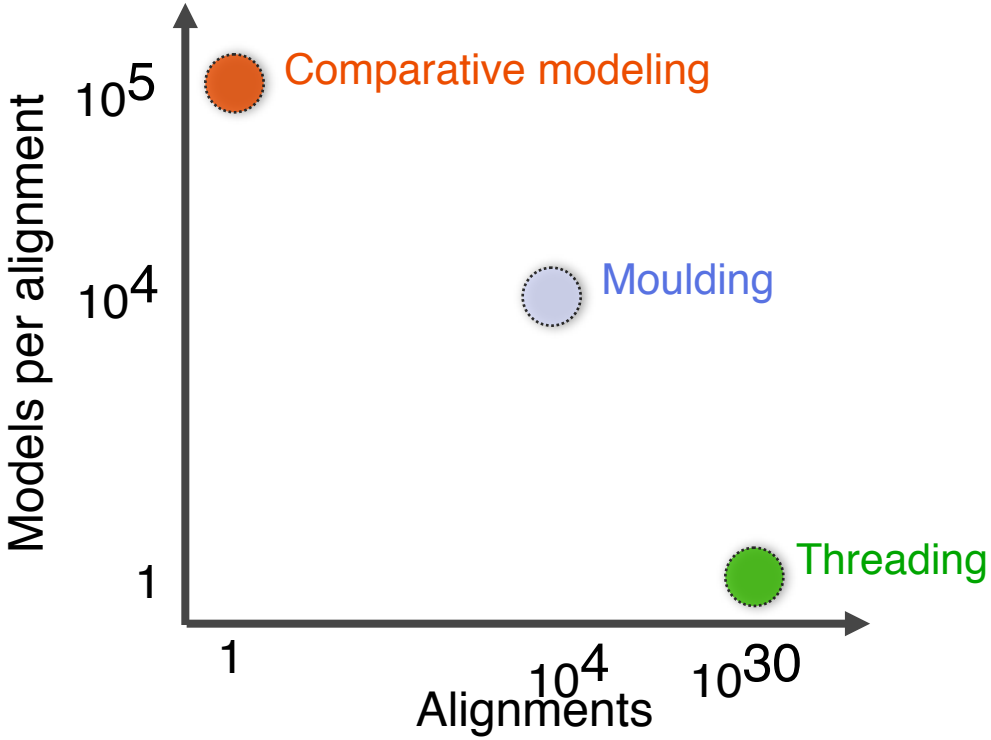
Sidechains
Core backbone
Loops
Alignment
Fold assignment

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

John, Sali (2003). NAR pp31 3982

Moulding: iterative alignment, model building, model assessment





Genetic algorithm operators

Single point cross-over

```
...TSSQ—NMKLGVFWGY———...
...V—SSCN———GDLHMKVGV...

...TSSQNMK———LGVFWGY...
...VSSCNGDLHMKV———GV...
```

→

```
...TSSQ—NMK———LGVFWGY...
...V—SSCNGDLHMKV———GV...

...TSSQNMKLGVFWGY———...
...VSSCN———GDLHMKVGV...
```

Gap insertion

```
...TSSQNMKLGVFWGY...
...VSSCNGDLHMKVGV...
```

→

```
...TSSQN——MKLGVFWGY...
...VSSCNGDLHMKVG——V...
```

Gap shift

```
...T——SSQNMKLGVFWGY...
...VSSCNGDLHMKVGV——...
```

→

```
...—T—SSQNMKLGVFWGY...
...VSSCNGDLHMKVGV——...

...T—S—SQNMKLGVFWGY...
...VSSCNGDLHMKVGV——...

...——TSSQNMKLGVFWGY...
...VSSCNGDLHMKVGV——...

...TS——SQNMKLGVFWGY...
...VSSCNGDLHMKVGV——...
```

Also, “two point crossover” and “gap deletion”.

Composite model assessment score

Weighted linear combination of several scores:

- Pair (P_p) and surface (P_s) statistical potentials;
- Structural compactness (S_c);
- Harmonic average distance score (H_a);
- Alignment score (A_s).

$$Z = 0.17\, Z(P_P) + 0.02\, Z(P_S) + 0.10\, Z(S_C) + 0.26\, Z(H_a) + 0.45\, (A_S)$$

$Z(\text{score}) = (\text{score} - \mu)/\sigma$

μ … average score of all models

σ … standard deviation of the scores

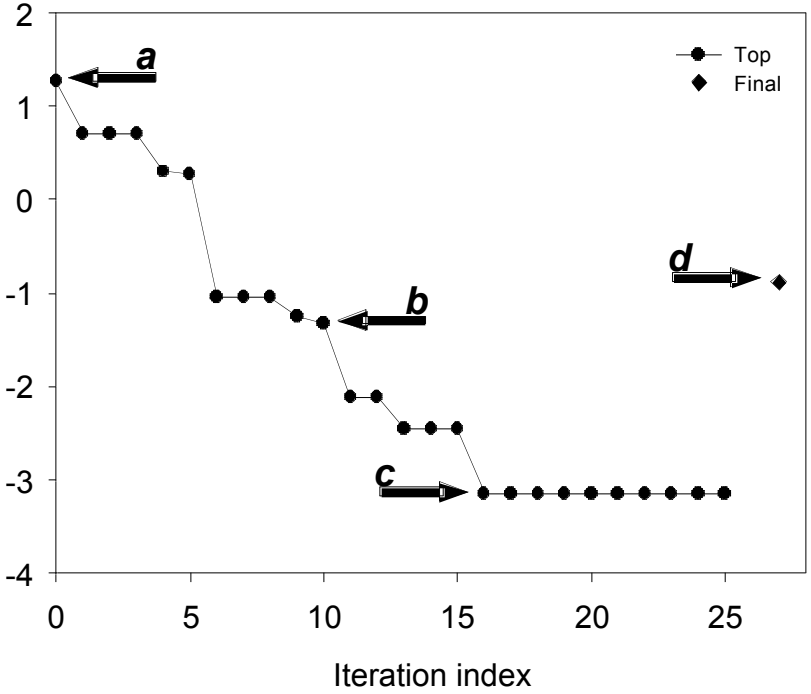
Benchmark with the "very difficult" test set

D. Fischer threading test set of 68 structural pairs (a subset of 19)

Target -template	Sequence identity [%]	Coverage [% aa]	Initial prediction		Final prediction		Best prediction	
			Cα RMSD [Å]	CE overlap [%]	Cα RMSD [Å]	CE overlap [%]	Cα RMSD [Å]	CE overlap [%]
1ATR-1ATN	13.8	94.3	19.2	20.2	18.8	20.2	17.1	24.6
1BOV-1LTS	4.4	83.5	10.1	29.4	3.6	79.4	3.1	92.6
1CAU-1CAU	18.8	96.7	11.7	15.6	10.0	27.4	7.6	47.4
1COL-1CPC	11.2	81.4	8.6	44.0	5.6	58.6	4.8	59.3
1LFB-1HOM	17.6	75.0	1.2	100.0	1.2	100.0	1.1	100.0
1NSB-2SIM	10.1	89.2	13.2	20.2	13.2	20.1	12.3	26.8
1RNH-1HRH	26.6	91.2	13.0	21.2	4.8	35.4	3.5	57.5
1YCC-2MTA	14.5	55.1	3.4	72.4	5.3	58.4	3.1	75.0
2AYH-1SAC	8.8	78.4	5.8	33.8	5.5	48.0	4.8	64.9
2CCY-1BBH	21.3	97.0	4.1	52.4	3.1	73.0	2.6	77.0
2PLV-1BBT	20.2	91.4	7.3	58.9	7.3	58.9	6.2	60.7
2POR-2OMF	13.2	97.3	18.3	11.3	11.4	14.7	10.5	25.9
2RHE-1CID	21.2	61.6	9.2	33.7	7.5	51.1	4.4	71.1
2RHE-3HLA	2.4	96.0	8.1	16.5	7.6	9.4	6.7	43.5
3ADK-1GKY	19.5	100.0	13.8	26.6	11.5	37.7	7.7	48.1
3HHR-1TEN	18.4	98.9	7.3	60.9	6.0	66.7	4.9	79.3
4FGF-81IB	14.1	98.6	11.3	24.0	9.3	30.6	5.4	41.2
6XIA-3RUB	8.7	44.1	10.5	14.5	10.1	11.0	9.0	34.3
9RNT-2SAR	13.1	88.5	5.8	41.7	5.1	51.2	4.8	69.0
AVERAGE	14.2	85.2	9.6	36.7	7.7	44.8	6.3	57.8

Application to a difficult modeling case

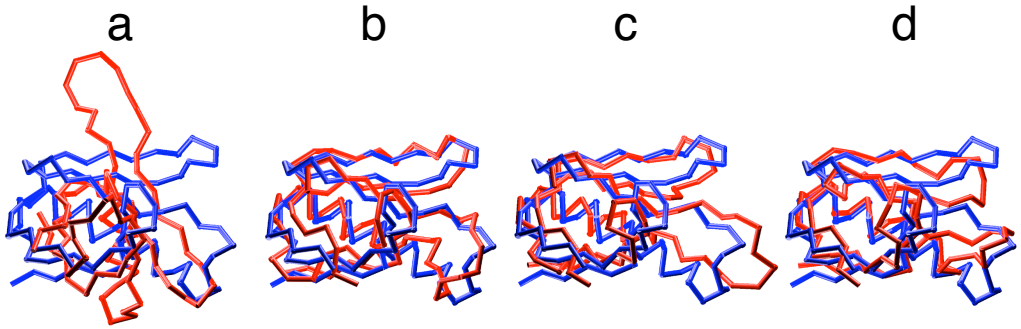
1BOV-1LTS



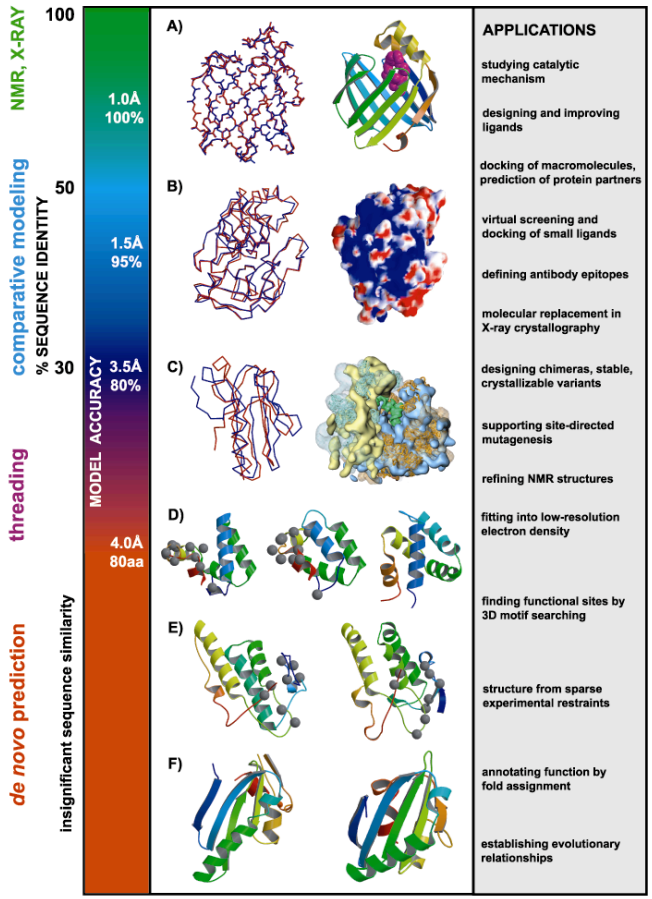
Sequence identity 4.4%

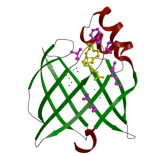
Initial model Cα RMSD 10.1Å

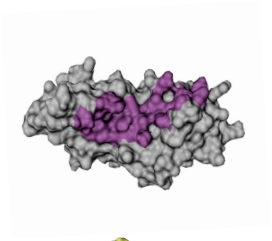
Final model Cα RMSD 3.6Å

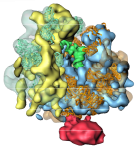


Can we use models to infer function?

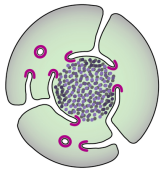








T. cruzi



What is the physiological ligand of Brain Lipid-Binding Protein?

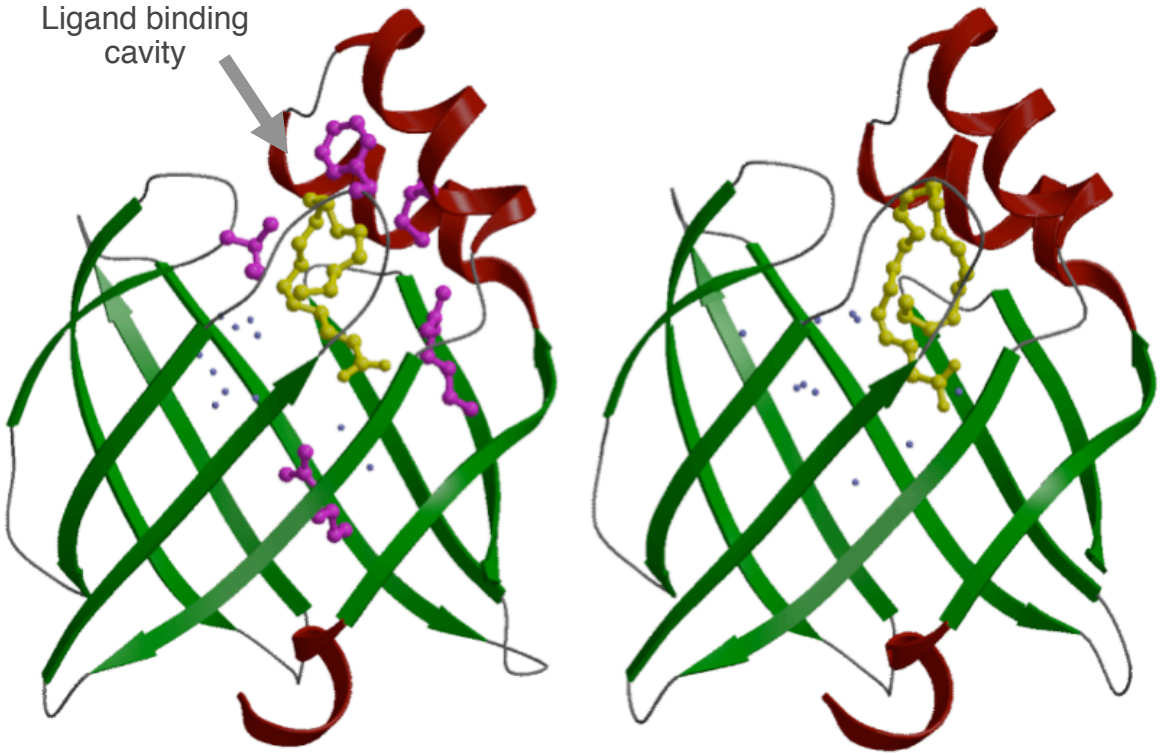
Predicting features of a model that are not present in the template

BLBP/oleic acid

Cavity is not filled

BLBP/docosahexaenoic acid

Cavity is filled



1. BLBP binds fatty acids.

2. Build a 3D model.

3. Find the fatty acid that fits most snuggly into the ligand binding cavity.

L. Xu, R. Sánchez, A. Šali, N. Heintz, J. Biol. Chem. 271, 24711, 1996.

Structural analysis of missense mutations in human BRCA1 BRCT domains

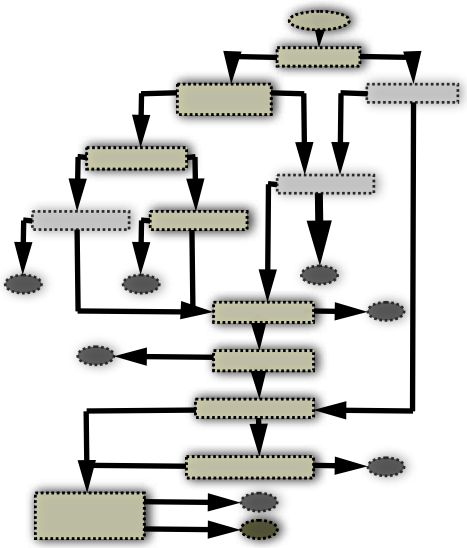
Nebojsa Mirkovic, Marc A. Marti-Renom, Barbara L. Weber, Andrej Sali and Alvaro N.A. Monteiro

Cancer Research (June 2004). 64:3790-97

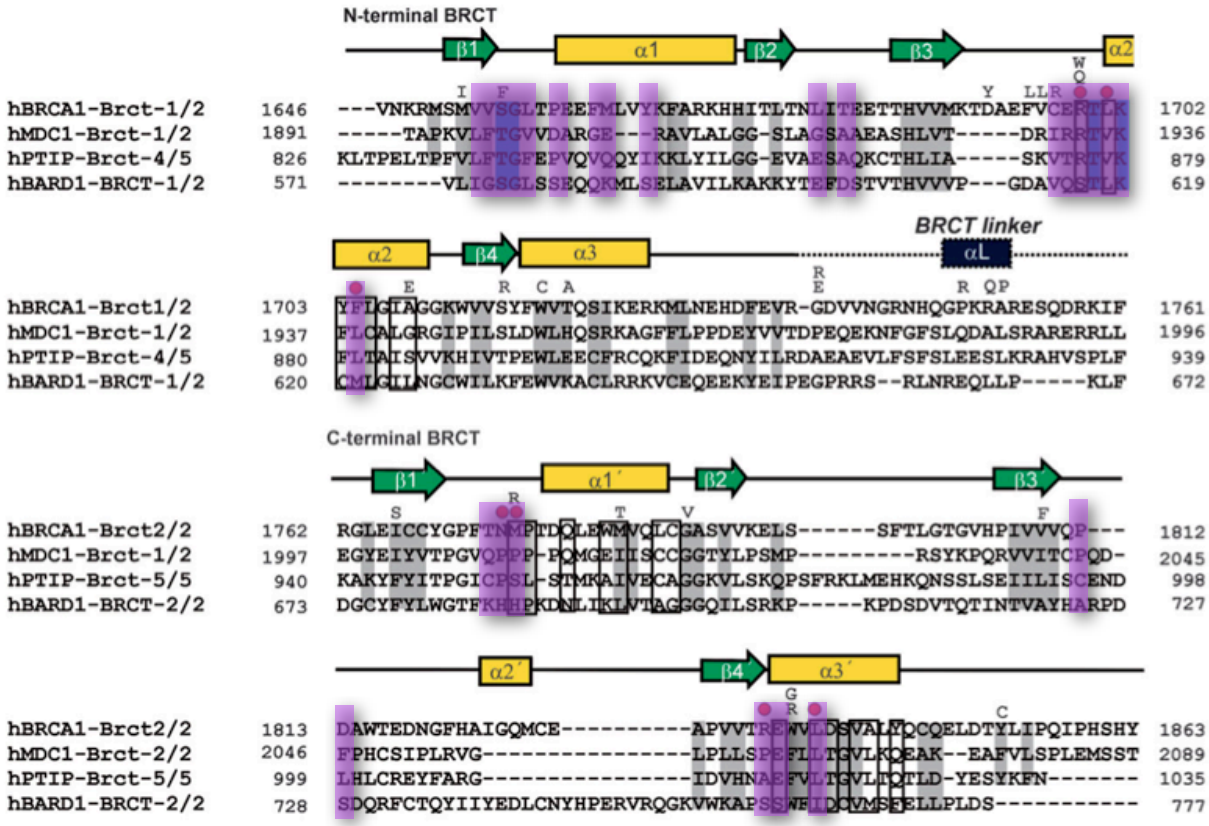
Cannot measure the functional impact of every possible SNP at all positions in each protein! Thus, prediction based on general principles of protein structure is needed.

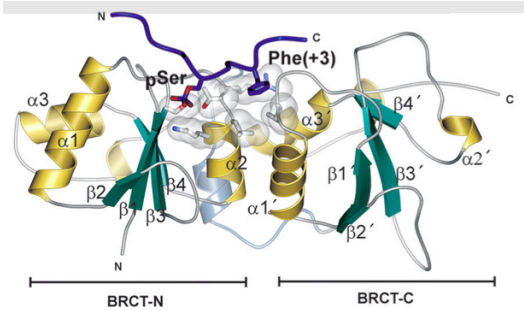
Missense mutations in BRCT domains by function

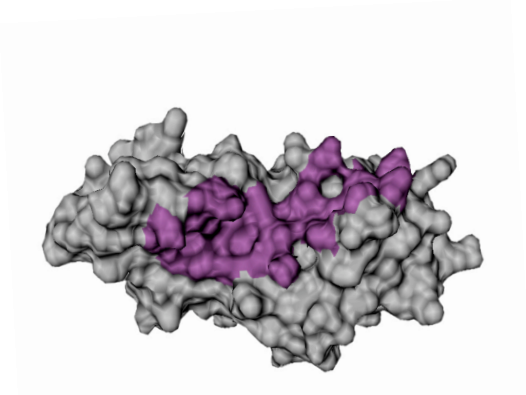
	cancer associated	not cancer associated	?
no transcription activation	C1697R R1699W A1708E S1715R P1749R M1775R		M1652K L1657P E1660G H1686Q R1699Q K1702E Y1703HF 1704S L1705PS 1715NS1 722FF17 34LG173 8EG1743 RA1752 PF1761I F1761S M1775E M1775K L1780P I1807S V1833E A1843T
transcription activation		M1652I A1669S	V1665M D1692N G1706A D1733G M1775V P1806A
?			M1652T V1653M L1664P T1685A T1685I M1689R D1692Y F1695L V1696L R1699L G1706E W1718C W1718S T1720A W1730S F1734S E1735K V1736A G1738R D1739E D1739G D1739Y V1741G H1746N R1751P R1751Q R1758G L1764P I1766S P1771L T1773S P1776S D1778N D1778G D1778H M1783T C1787S G1788D G1788V G1803A V1804D V1808A V1809A V1809F V1810G Q1811R P1812S N1819S A1823T V1833M W1837R W1837G S1841N A1843P T1852S P1856T P1859R



Putative binding site on BRCA1





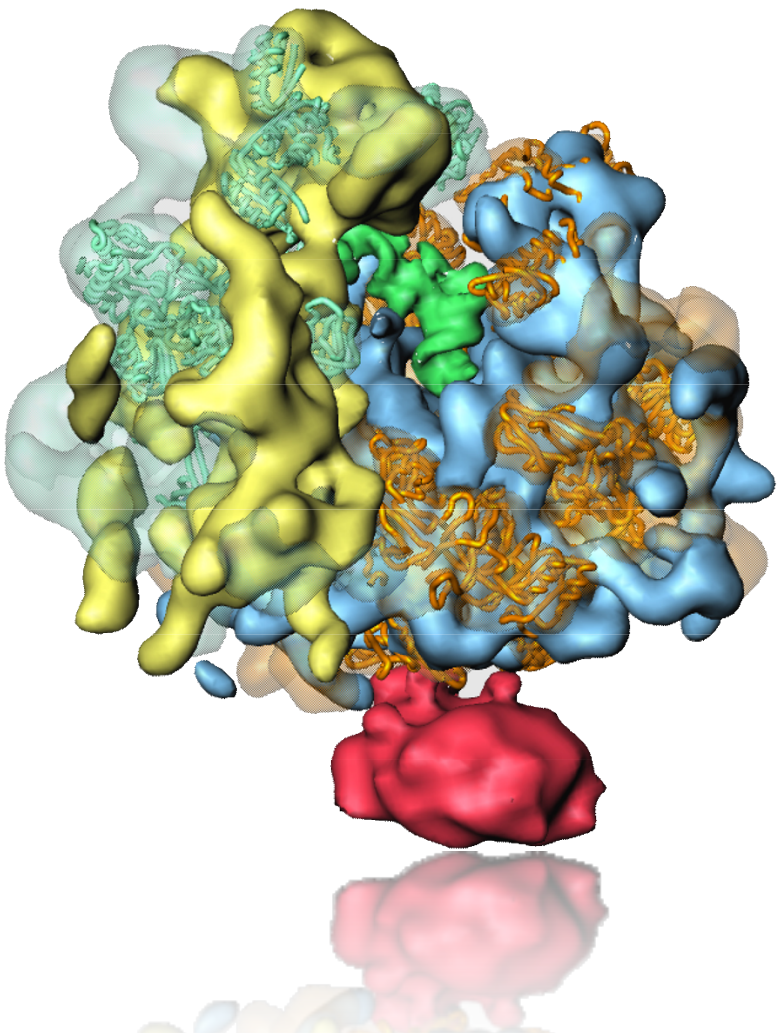


Putative binding site predicted in 2003 and accepted for publication on March 2004.

Williams *et al.* 2004 Nature Structure Biology. **June 2004 11**:519

Mirkovic *et al.* 2004 Cancer Research. **June 2004 64**:3790

S. cerevisiae ribosome



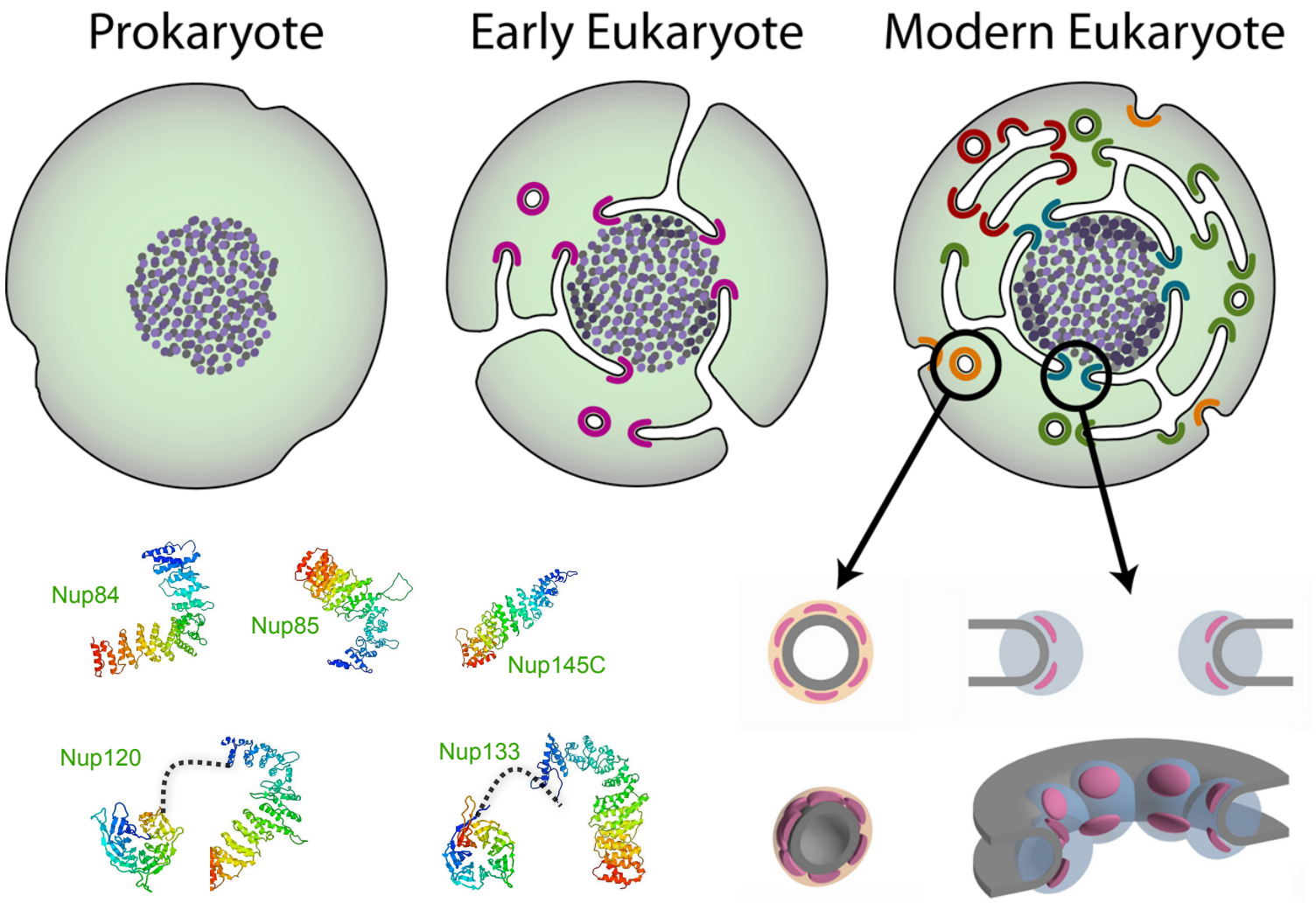
Fitting of comparative models into 15Å cryo-electron density map.

43 proteins could be modeled on 20-56% seq.id. to a known structure.

The modeled fraction of the proteins ranges from 34-99%.

C. Spahn, R. Beckmann, N. Eswar, P. Penczek, A. Sali, G. Blobel, J. Frank. Cell 107, 361-372, 2001.

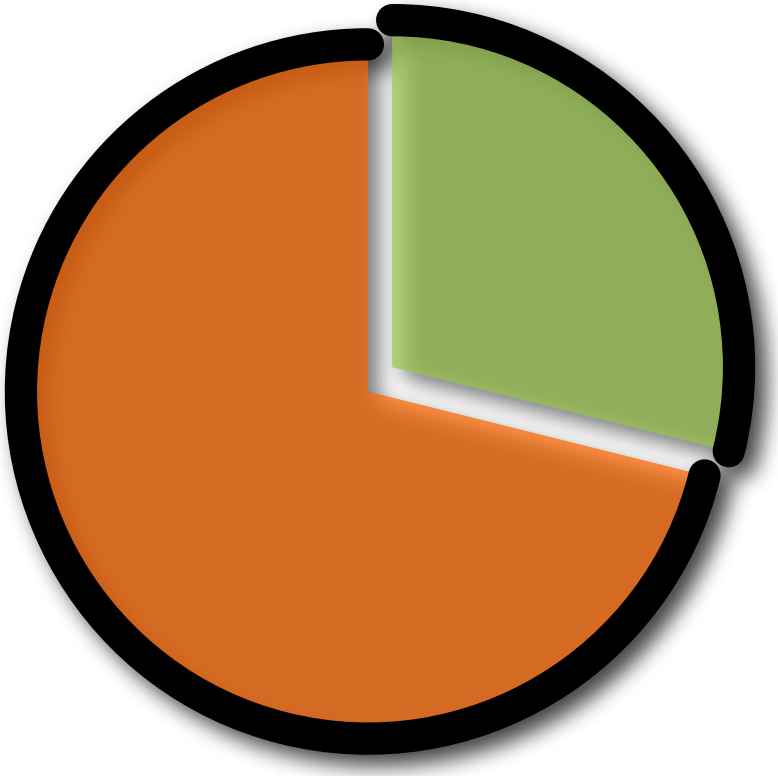
The Nucleopore complex
Cell evolution (?)



Devos et al. PLoS Biology 2, 1 (2004)

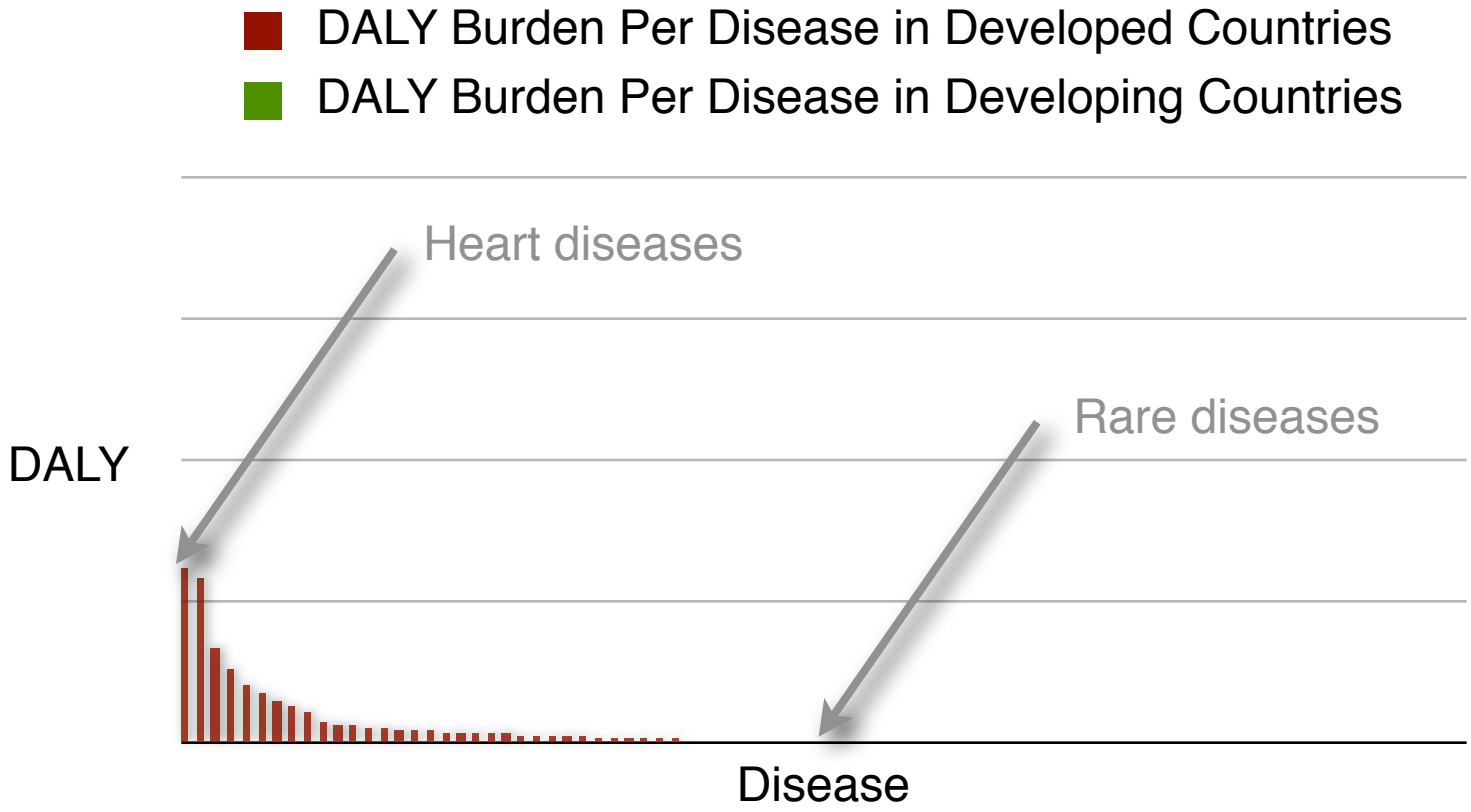
Tropical Disease Initiative (TDI)

Predicting binding sites in protein structure models.



http://www.tropicaldisease.org

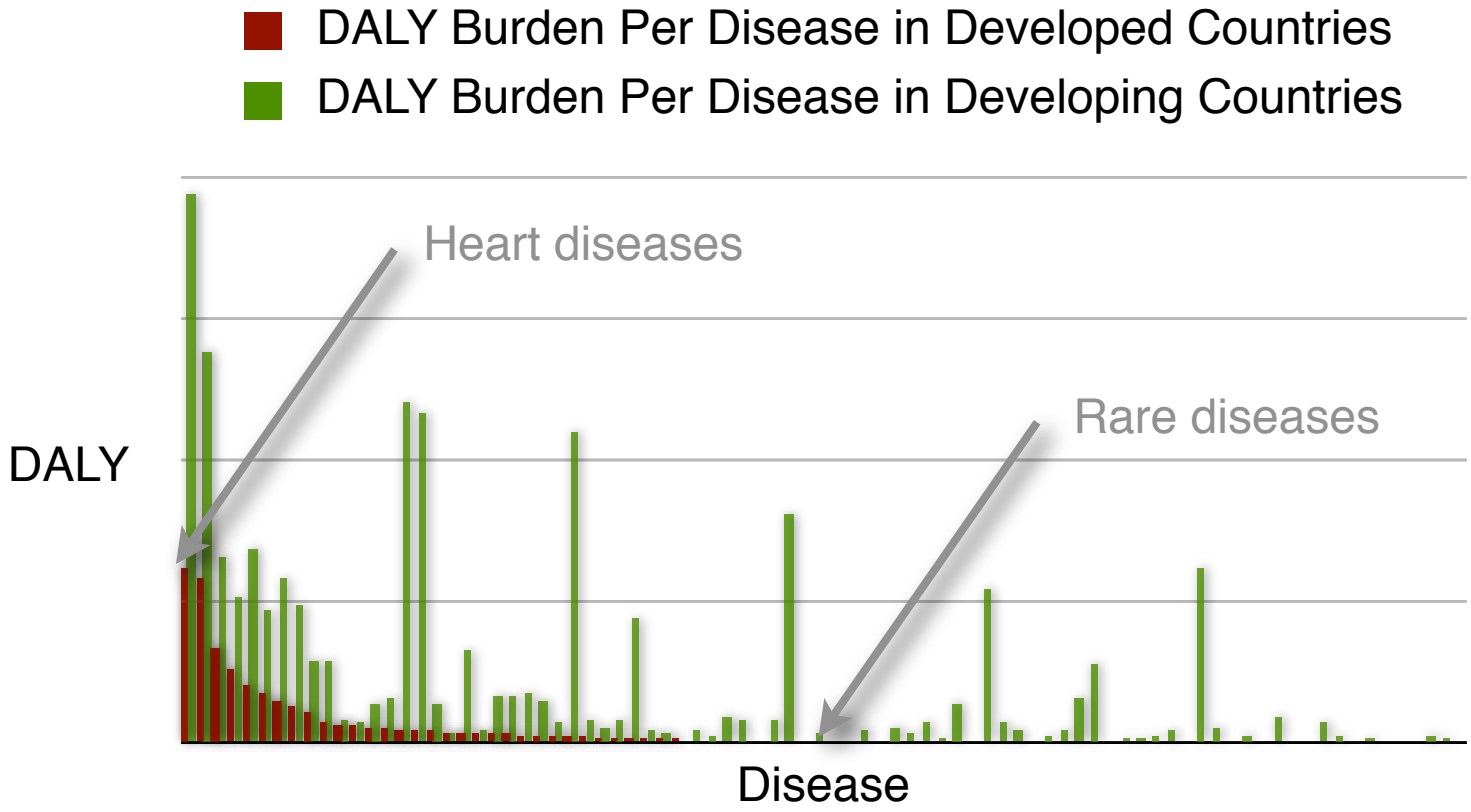
Need is High in the Tail



Disease data taken from WHO, *World Health Report 2004*
DALY - Disability adjusted life years
DALY is not a perfect measure of market size, but is certainly a good measure for importance.

DALYs for a disease are the sum of the years of life lost due to premature mortality (YLL) in the population and the years lost due to disability (YLD) for incident cases of the health condition. The DALY is a health gap measure that extends the concept of potential years of life lost due to premature death (PYLL) to include equivalent years of 'healthy' life lost in states of less than full health, broadly termed disability. One DALY represents the loss of one year of equivalent full health.

Need is High in the Tail



Disease data taken from WHO, *World Health Report 2004*
DALY - Disability adjusted life years
DALY is not a perfect measure of market size, but is certainly a good measure for importance.

DALYs for a disease are the sum of the years of life lost due to premature mortality (YLL) in the population and the years lost due to disability (YLD) for incident cases of the health condition. The DALY is a health gap measure that extends the concept of potential years of life lost due to premature death (PYLL) to include equivalent years of 'healthy' life lost in states of less than full health, broadly termed disability. One DALY represents the loss of one year of equivalent full health.

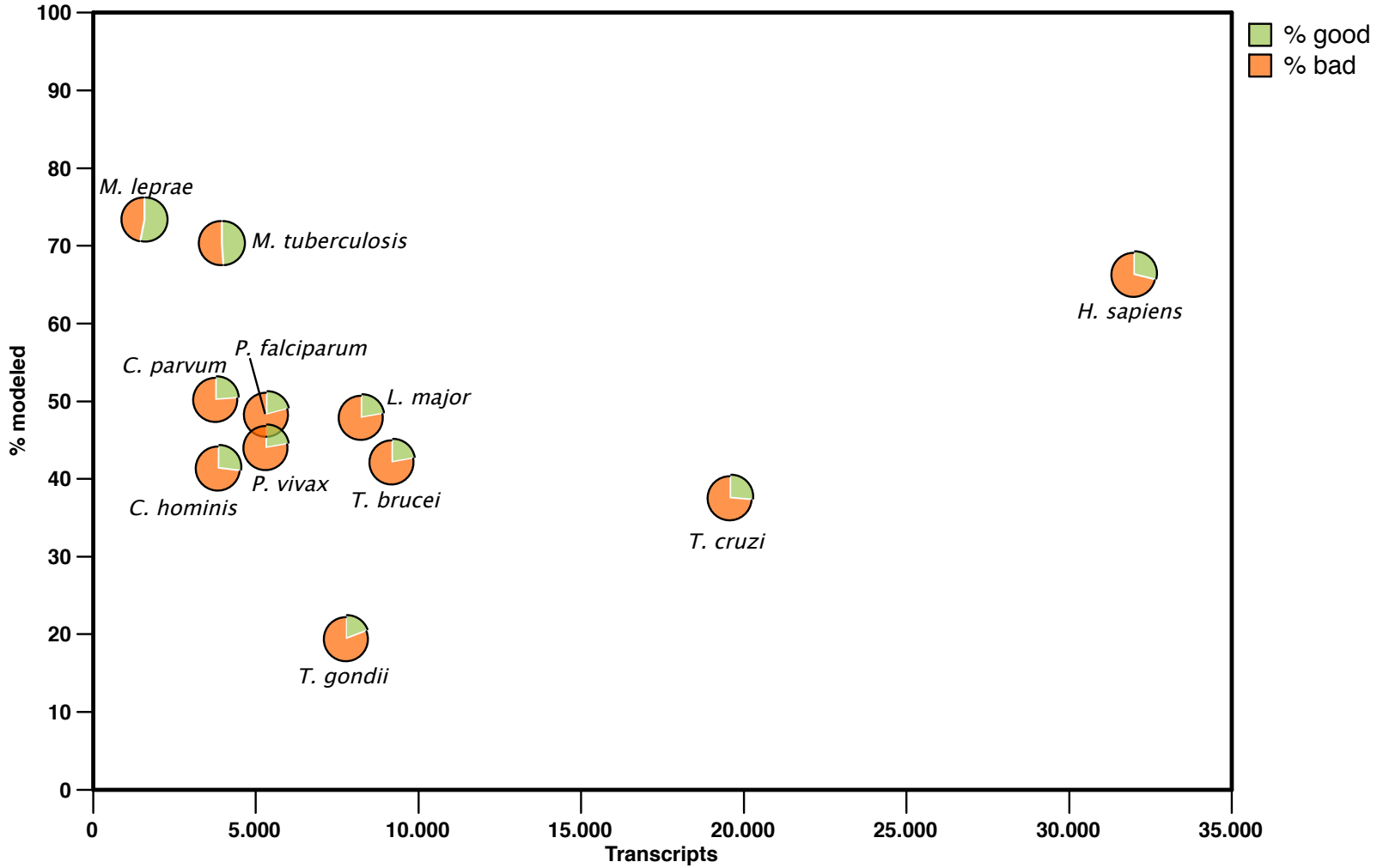
"Unprofitable" Diseases and Global DALY (in 1000's)

Malaria*	**46,486**	Trichuriasis	1,006
Tetanus	7,074	Japanese encephalitis	709
Lymphatic filariasis*	**5,777**	**Chagas Disease***	**667**
Syphilis	4,200	**Dengue***	**616**
Trachoma	2,329	**Onchocerciasis***	**484**
Leishmaniasis*	**2,090**	**Leprosy***	**199**
Ascariasis	1,817	Diphtheria	185
Schistosomiasis*	**1,702**	Poliomyelitise	151
Trypanosomiasis*	**1,525**	Hookworm disease	59

Disease data taken from WHO, *World Health Report 2004*
DALY - Disability adjusted life year in 1000's.
* Officially listed in the WHO Tropical Disease Research disease portfolio.

Modeling Genomes

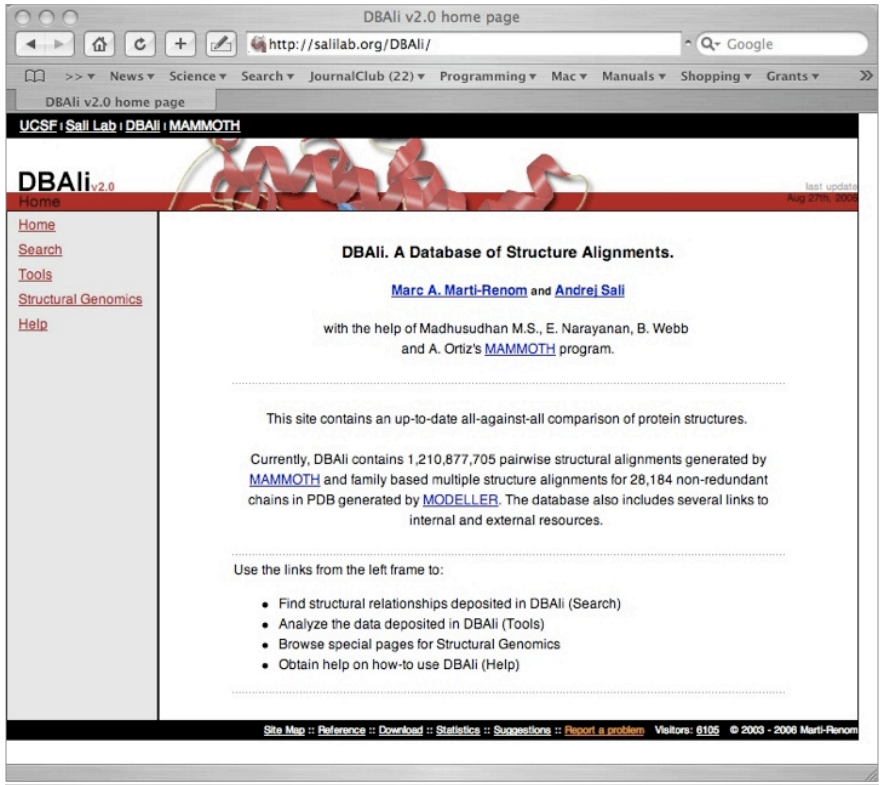
data from models generated by ModPipe (Eswar, Pieper & Sali)



A good model has MPQS of 1.1 or higher

DBAli$_{v2.0}$ database

http://www.dbali.org



- ✓ Fully-automatic
- ✓ Data is kept up-to-date with PDB releases
- ✓ Tools for "on the fly" classification of families.
- ✓ Easy to navigate
- ✓ Provides tools for structure analysis

Does not provide a stable classification similar to that of CATH or SCOP

Pairwise structure alignments	
Last update:	October 6th, 2007
Number of chains:	96,804
Number of structure-structure comparisons:*	1,748,371,897
Multiple structure alignments	
Last update:	August 1st, 2007
Number of representative chains:	34,637
Number of families:	12,732

Uses MAMMOTH for similarity detection

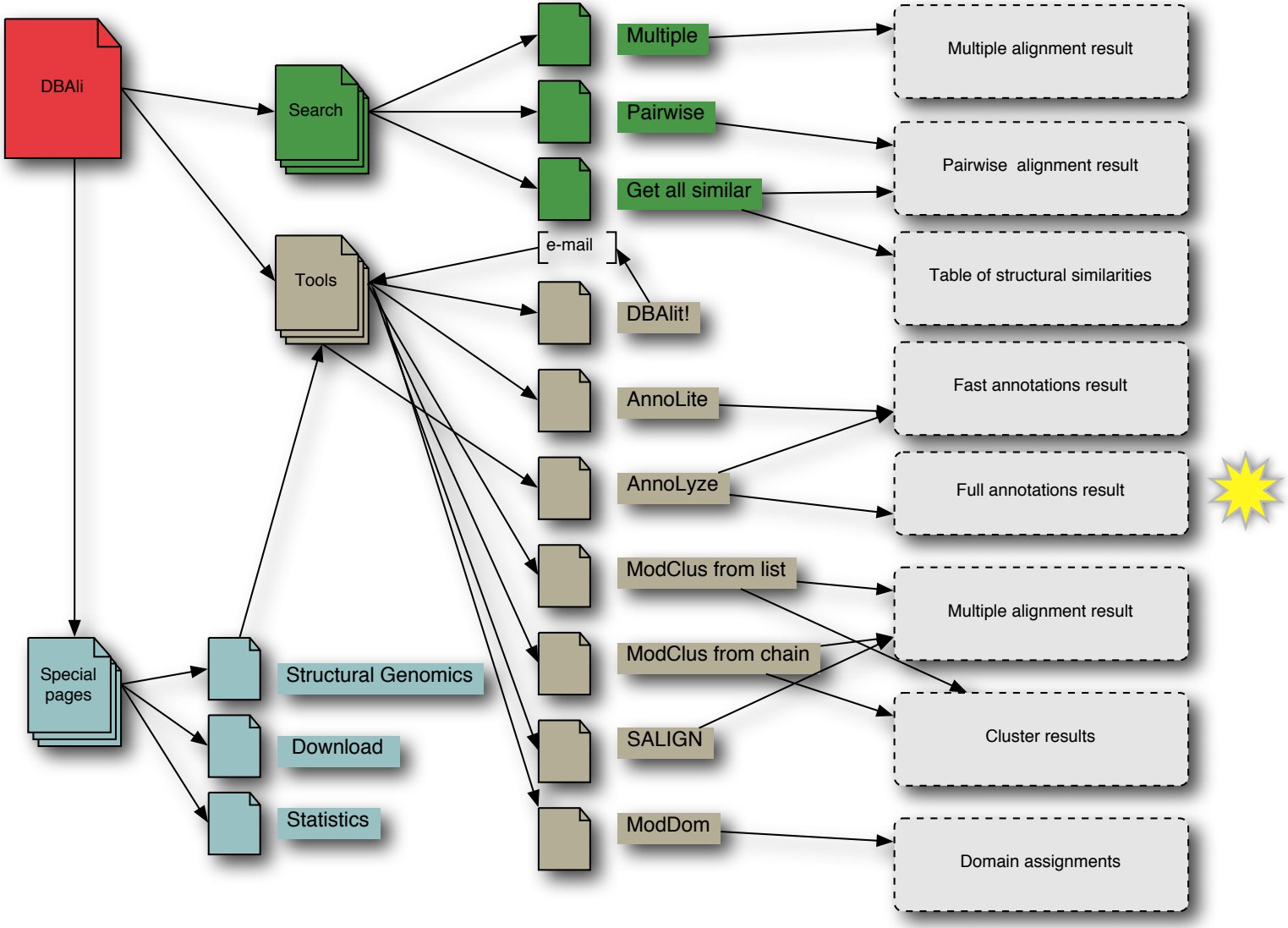
- ✓ **VERY FAST!!!**
- ✓ **Good scoring system with significance**

Ortiz AR, (2002) Protein Sci. 11 pp2606

Marti-Renom et al. 2001. Bioinformatics. 17, 746

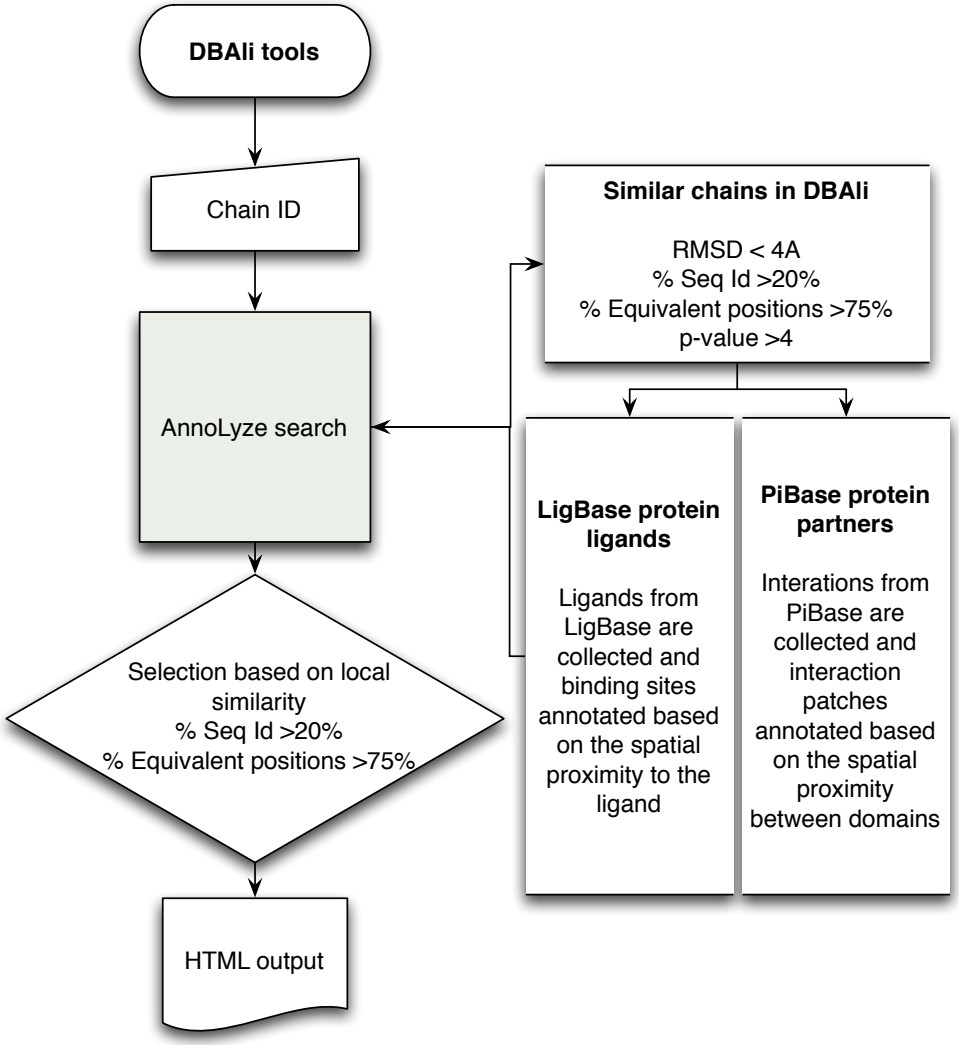
DBAli$_{v2.0}$ database

http://www.dbali.org



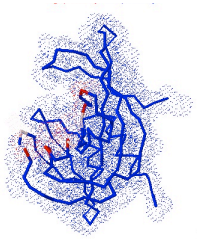
Marti-Renom et al. BMC Bioinformatics (2007) Volume 8. Suppl S4

Method



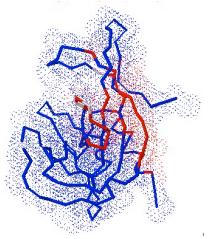
Inherited ligands: 4

Ligand	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
MO2	59.03	0.185	48 49 52 62 63 66 67 113 116
CRY	20.00	0.111	23 29 31 37 44 48 49 83 85 94 96 103 121
8OG	20.00	0.111	19 20 21 48 49 51 96 98 136
ACY	15.87	0.163	23 29 31 37 44 45 81 83 85 94 96 98 103 121 135



Inherited partners: 1

Partner	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
d.113.1.1	23.68	0.948	19 20 50 51 52 53 54 55 56 57 58 77 78 79 80 81 82 83 84 85 93 95 97 99 134 135 138 142 145



Sensitivity .vs. Precision

	Optimal cut-off	Sensitivity (%) Recall or TPR	Precision (%)
Ligands	30%	71.9	13.7

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad \text{Precision} = \frac{TP}{TP + FP}$$

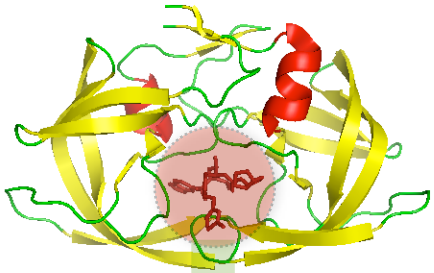
~90-95% of residues correctly predicted

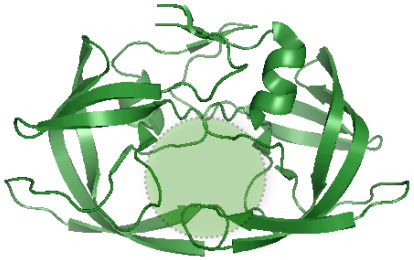
Marti-Renom et al. BMC Bioinformatics (2007) Volume 8. Suppl S4

Comparative docking

1. Expansion

co-crystalized protein/ligand

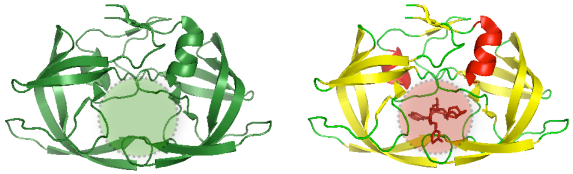




crystalized protein

2. Inheritance

model



template

Summary table

models with inherited ligands

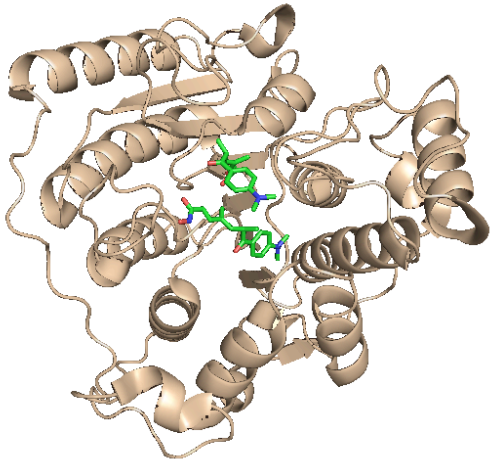
from 16,284 good models, 295 inherited a ligand/substance with at least one compound already approved by FDA and ready to be used from ZINC

	Transcripts	Good	Ligands	Lipinski	Lipinski+ZINC	FDA+ZINC
C. hominis	3,886	886	183	131	28	12 (10)
C. parvum	3,806	949	219	145	30	12 (10)
L. major	8,274	1,845	488	334	84	44 (34)
M. leprae	1,605	1,321	286	189	39	29 (25)
M. tuberculosis	3,991	2,887	404	285	71	44 (37)
P. falciparum	5,363	1,057	271	191	48	20 (16)
P. vivax	5,342	1,042	267	177	37	18 (15)
T. brucei	921	1,795	440	309	94	46 (36)
T. cruzi	19,607	3,915	730	493	127	62 (52)
T. gondii	7,793	587	174	124	28	8 (7)
TOTAL	**60,588**	**16,284**	**3,462**	**2,378**	**586**	**295 (242)**

Example of inheritance (expansion)

LmjF21.0680 from L. major *"Histone deacetylase 2" (model 1)*

Template 1t64A a human HDAC8 protein.



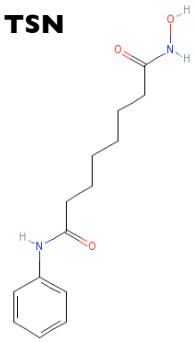
	Origen	Formula	Name	Cov.	Seq, Id. (%)
ZN	X-ray	Zn^{2+}	Zinc ion	--	--
NA	X-ray	Na^{+}	Sodium ion	--	--
CA	X-ray	Ca^{2+}	Calcium ion	--	--
TSN	X-ray	$C_{17}H_{22}N_2O_3$	Trichostatin A	--	--
SHH	Expanded	$C_{14}H_{20}N_2O_3$	Octadenioic acid hudroxyamide phenylamide	100.00	83.8

Example of inheritance (inheritance)

LmjF21.0680 from L. major *"Histone deacetylase 2" (model 1)*

	Formula	Name	Cov.	Seq, Id. (%)	Residues
TSN	$C_{17}H_{22}N_2O_3$	Trichostatin A	100.00	90.9	90 131 132 140 141 167 169 256 263 293 295
SHH	$C_{14}H_{20}N_2O_3$	Octadenioic acid hudroxyamide phenylamide	100.00	90.9	

TSN



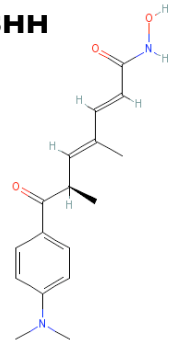
suberoylanilide hydroxamic acid

Pharmacological Action:

- Anti-Inflammatory Agents, Non-Steroidal
- Antineoplastic Agents
- Enzyme Inhibitors
- Anticarcinogenic Agents

Inhibits histone deacetylase I and 3

SHH



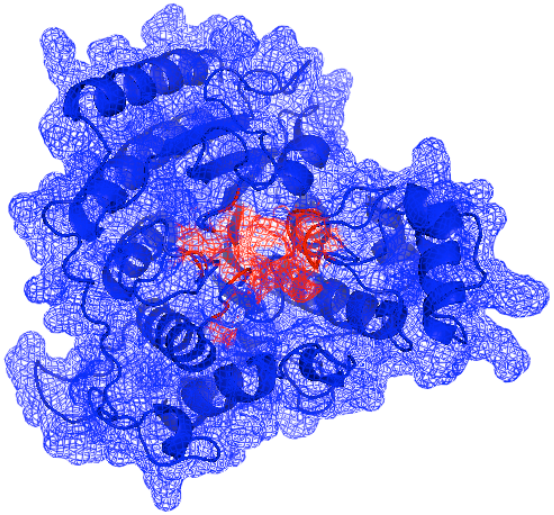
trichostatin A

Pharmacological Action:

- Antibiotics, Antifungal
- Enzyme Inhibitors
- Protein Synthesis Inhibitors

chelates zinc ion in the active site of histone deacetylases, resulting in preventing histone unpacking so DNA is less available for transcription

	LmjF21.0680.1.pdb
Template	1t64A
Seq. Id (%)	**38.00**
MPQS	1.47



Example of inheritance (CDD-Roos-literature)

LmjF21.0680 from L. major *"Histone deacetylase 2" (model 1)*

Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 13143–13147, November 1996
Medical Sciences

Apicidin: A novel antiprotozoal agent that inhibits parasite histone deacetylase

(cyclic tetrapeptide/Apicomplexa/antiparasitic/malaria/coccidiosis)

SANDRA J. DARKIN-RATTRAY*†, ANNE M. GURNETT*, ROBERT W. MYERS*, PAULA M. DULSKI*, TAMI M. CRUMLEY*, JOHN J. ALLOCCO*, CHRISTINE CANNOVA*, PETER T. MEINKE‡, STEVEN L. COLLETTI‡, MARIA A. BEDNAREK‡, SHEO B. SINGH§, MICHAEL A. GOETZ§, ANNE W. DOMBROWSKI§, JON D. POLISHOOK§, AND DENNIS M. SCHMATZ*

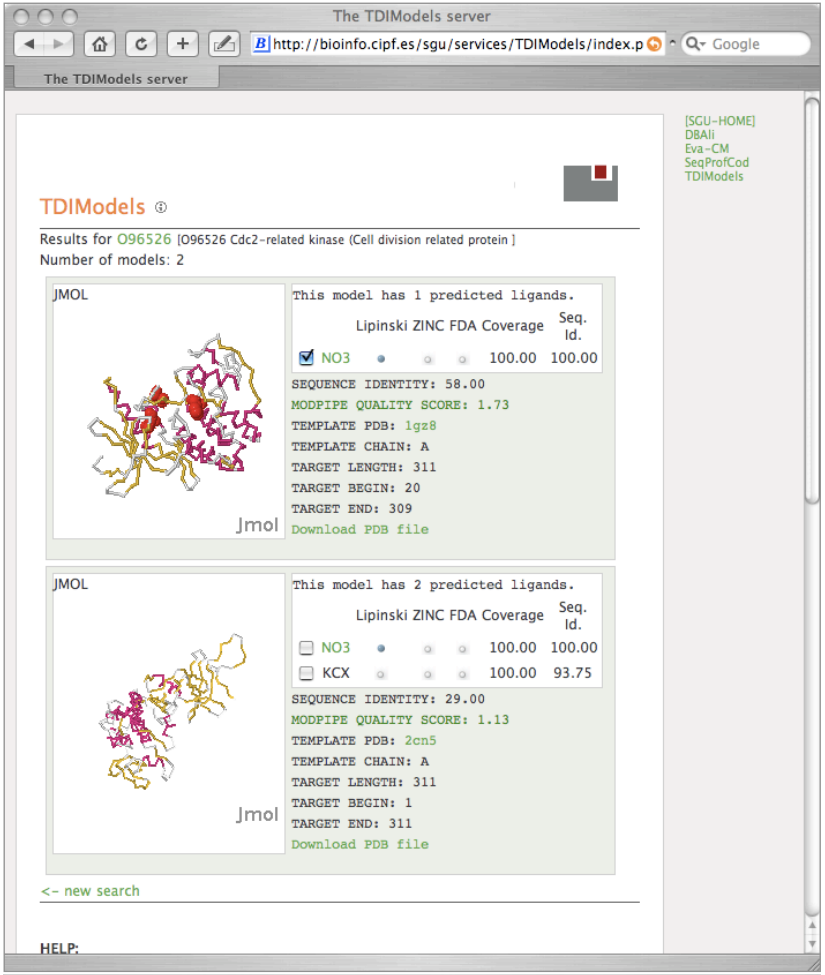
Departments of *Parasite Biochemistry and Cell Biology, ‡Medicinal Chemistry, and §Natural Products Drug Discovery, Merck Research Laboratories, P.O. Box 2000, Rahway, NJ 07065

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Apr. 2004, p. 1435–1436 Vol. 48, No. 4
0066-4804/04/$08.00+0 DOI: 10.1128/AAC.48.4.1435–1436.2004

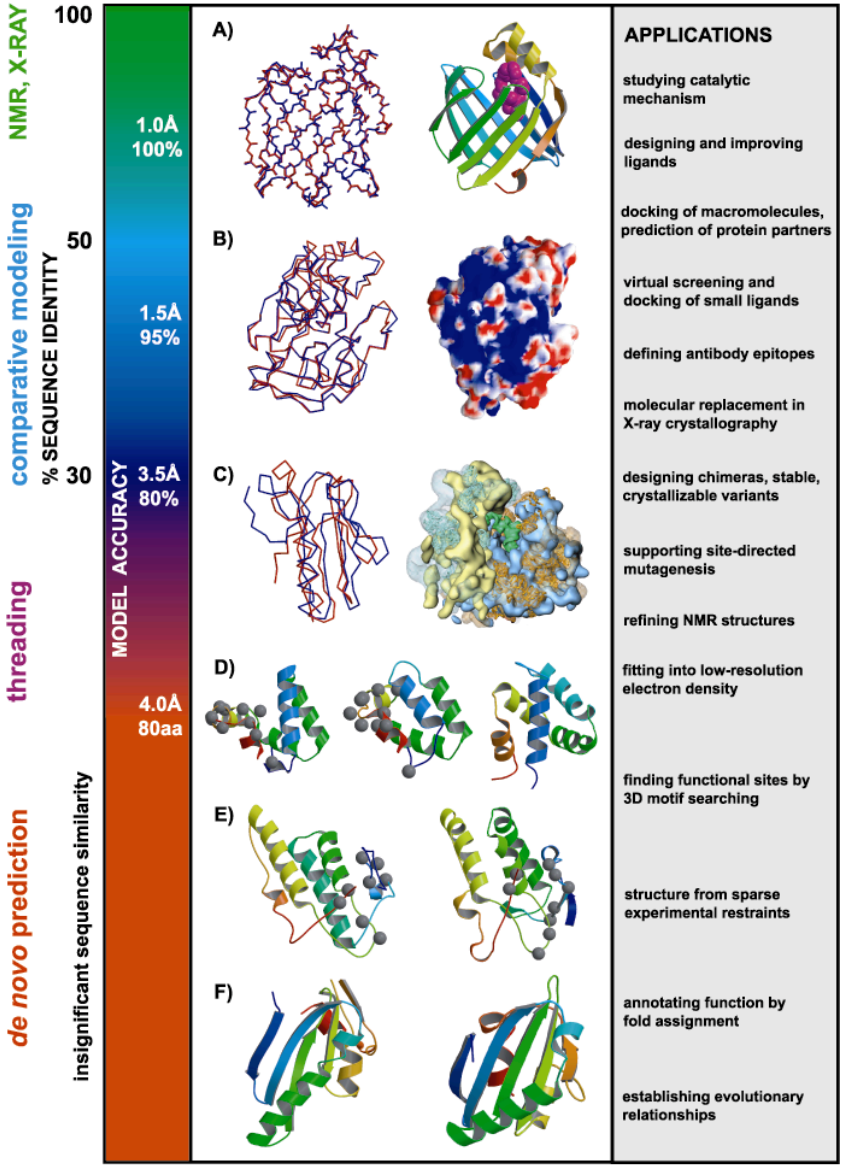
Antimalarial and Antileishmanial Activities of Aroyl-Pyrrolyl-Hydroxyamides, a New Class of Histone Deacetylase Inhibitors

Models database

http://bioinfo.cipf.es/sgu/services/TDIModels/



“take home” message



D. Baker & A. Sali. Science 294, 93, 2001.

Comparative Protein Structure Prediction

MODELLER tutorial

```
$>mod9v3 model.py
```

Marc A. Marti-Renom

http://bioinfo.cipf.es/sgu/

Structural Genomics Unit
Bioinformatics Department
Prince Felipe Resarch Center (CIPF), Valencia, Spain

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

Obtaining MODELLER and related information

- MODELLER (9v3) web page
- http://www.salilab.org/modeller/
  - Download Software (Linux/Windows/Mac/Solaris)
  - HTML Manual
  - **Join Mailing List**

Using MODELLER

- No GUI! ☹
- Controlled by command file ☹☹
- Script is written in PYTHON language ☺
- You may know Python language is simple ☺☺

MODELLER 9v3

Python interface

- Modeller Python interface uses classes, e.g.:
  - ‘alignment’ *holds and manipulates aligned sequences*
  - ‘model’ *holds and manipulates protein models*
  - ‘environ’ *keeps the configuration of the environment*
  - ‘profile’ *holds and manipulates sequence profiles*
  - ‘sequence_db’ *is for sequence databases*
- These behave just like ordinary Python classes, but Modeller Fortran code is linked to them
- The Modeller data is automatically freed when the Python object is deleted (explicitly or implicitly)

Using MODELLER

- INPUT:
  - Target Sequence (FASTA/PIR format)
  - Template Structure (PDB format)
  - Python file
- OUTPUT:
  - Target-Template Alignment
  - Model in PDB format
  - Other data

Modeling of BLBP
Input

- Target: Brain lipid-binding protein (BLBP)
- BLBP sequence in PIR (MODELLER) format:

```
>P1;blbp
sequence:blbp:::::::
VDAFCATWKLTDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGEEFEETSID
DRNCKSVVRLDGDKLIHVQKWDGKETNCTREIKDGKMVVTLTFGDIVAVRCYEKA*
```

Modeling of BLBP

STEP 1: Align blbp and 1hms sequences

Python script for target-template alignment

```
# Example for: alignment.align()

# This will read two sequences, align them, and write the alignment
# to a file:

log.verbose()
env = environ()

aln = alignment(env)
mdl = model(env, file='1hms')
aln.append_model(mdl, align_codes='1hms')
aln.append(file='blbp.seq', align_codes=('blbp'))

# The as1.sim.mat similarity matrix is used by default:
aln.align(gap_penalties_1d=(-600, -400))
aln.write(file='blbp-1hms.ali', alignment_format='PIR')
aln.write(file='blbp-1hms.pap', alignment_format='PAP')
```

Run by typing `mod9v3 align.py` in the directory where you have the python file.
MODELLER will produce a `align.log` file

Modeling of BLBP

STEP 1: Align blbp and 1hms sequences

Python script for target-template alignment

```
# Example for: alignment.align()

# This will read two sequences, align them, and write the alignment
# to a file:

log.verbose()
env = environ()

aln = alignment(env)
mdl = model(env, file='1hms')
aln.append_model(mdl, align_codes='1hms')
aln.append(file='blbp.seq', align_codes=('blbp'))

# The as1.sim.mat similarity matrix is used by default:
aln.align(gap_penalties_1d=(-600, -400))
aln.write(file='blbp-1hms.ali', alignment_format='PIR')
aln.write(file='blbp-1hms.pap', alignment_format='PAP')
```

Run by typing `mod9v3 align.py` in the directory where you have the python file.
MODELLER will produce a `align.log` file

Modeling of BLBP

STEP 1: Align blbp and 1hms sequences

Python script for target-template alignment

```
# Example for: alignment.align()

# This will read two sequences, align them, and write the alignment
# to a file:

log.verbose()
env = environ()

aln = alignment(env)
mdl = model(env, file='1hms')
aln.append_model(mdl, align_codes='1hms')
aln.append(file='blbp.seq', align_codes=('blbp'))

# The as1.sim.mat similarity matrix is used by default:
aln.align(gap_penalties_1d=(-600, -400))
aln.write(file='blbp-1hms.ali', alignment_format='PIR')
aln.write(file='blbp-1hms.pap', alignment_format='PAP')
```

Run by typing `mod9v3 align.py` in the directory where you have the python file.
MODELLER will produce a `align.log` file

Modeling of BLBP

STEP 1: Align blbp and 1hms sequences

Python script for target-template alignment

```
# Example for: alignment.align()

# This will read two sequences, align them, and write the alignment
# to a file:

log.verbose()
env = environ()

aln = alignment(env)
mdl = model(env, file='1hms')
aln.append_model(mdl, align_codes='1hms')
aln.append(file='blbp.seq', align_codes=('blbp'))

# The as1.sim.mat similarity matrix is used by default:
aln.align(gap_penalties_1d=(-600, -400))
aln.write(file='blbp-1hms.ali', alignment_format='PIR')
aln.write(file='blbp-1hms.pap', alignment_format='PAP')
```

Run by typing `mod9v3 align.py` in the directory where you have the python file.
MODELLER will produce a `align.log` file

Modeling of BLBP

STEP 1: Align blbp and 1hms sequences

Output

```
>P1;1hms
structureX:1hms:   1 : : 131 : :undefined:undefined:-1.00:-1.00
VDAFLGTWKLVDSKNFDDYMKSLGVGFATRQVASMTKPTTIIEKNGDILTLKTHSTFKNTEISFKLGVEFDETTA
DDRKVKSIVTLDGGKLVHLQKWDGQETTLVRELIDGKLILTLTHGTAVCTRTYEKE*
>P1;blbp
sequence:blbp:     : :     : : : : 0.00: 0.00
VDAFCATWKLTDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGEEFEETSI
DDRNCKSVVRLDGDKLIHVQKWDGKETNCTREIKDGKMVVTLTFGDIVAVRCYEKA*
```

Modeling of BLBP

STEP 1: Align blbp and 1hms sequences

Output

```
>P1;1hms
structureX:1hms:   1 : : 131 : :undefined:undefined:-1.00:-1.00
VDAFLGTWKLVDSKNFDDYMKSLGVGFATRQVASMTKPTTIIEKNGDILTLKTHSTFKNTEISFKLGVEFDETTA
DDRKVKSIVTLDGGKLVHLQKWDGQETTLVRELIDGKLILTLTHGTAVCTRTYEKE*
>P1;blbp
sequence:blbp:     : :     : : : : 0.00: 0.00
VDAFCATWKLTDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGEEFEETSI
DDRNCKSVVRLDGDKLIHVQKWDGKETNCTREIKDGKMVVTLTFGDIVAVRCYEKA*
```

Modeling of BLBP

STEP 1: Align blbp and 1hms sequences

Output

```
 _aln.pos          10        20        30        40        50        60
1hms      VDAFLGTWKLVDSKNFDDYMKSLGVGFATRQVASMTKPTTIIEKNGDILTLKTHSTFKNTEISFKLGV
blbp      VDAFCATWKLTDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGE
 _consrvd ****  **** ** *** *** **********   **** **   *      *  ******* * **

 _aln.p   70        80        90       100       110       120       130
1hms      EFDETTADDRKVKSIVTLDGGKLVHLQKWDGQETTLVRELIDGKLILTLTHGTAVCTRTYEKE
blbp      EFEETSIDDRNCKSVVRLDGDKLIHVQKWDGKETNCTREIKDGKMVVTLTFGDIVAVRCYEKA
 _consrvd ** **  ***  ** * *** ** * ***** **   **  ***   *** *  *  * ***
```

Modeling of BLBP

STEP 2: Model the blbp structure using the alignment from step 1.

Python script for model building

```
# Homology modelling by the automodel class
from modeller.automodel import *     # Load the automodel class
log.verbose()                        # request verbose output
env = environ()                      # create a new MODELLER environment

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'

a = automodel(env,
              alnfile  = 'blbp-1hms.ali',     # alignment filename
              knowns   = '1hms',              # codes of the templates
              sequence = 'blbp')              # code of the target
a.starting_model= 1                  # index of the first model
a.ending_model  = 1                  # index of the last model
                                     # (determines how many models to calculate)
a.make()                             # do the actual homology modelling
```

Run by typing `mod9v3 model.py` in the directory where you have the python file. MODELLER will produce a `align.log` file

Modeling of BLBP

STEP 2: Model the blbp structure using the alignment from step 1.

Python script for model building

```
# Homology modelling by the automodel class
from modeller.automodel import *     # Load the automodel class
log.verbose()                        # request verbose output
env = environ()                      # create a new MODELLER environment

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'

a = automodel(env,
              alnfile  = 'blbp-1hms.ali',     # alignment filename
              knowns   = '1hms',              # codes of the templates
              sequence = 'blbp')              # code of the target
a.starting_model= 1                  # index of the first model
a.ending_model  = 1                  # index of the last model
                                     # (determines how many models to calculate)
a.make()                             # do the actual homology modelling
```

Run by typing `mod9v3 model.py` in the directory where you have the python file. MODELLER will produce a `align.log` file

Modeling of BLBP

STEP 2: Model the blbp structure using the alignment from step 1.

Python script for model building

```
# Homology modelling by the automodel class
from modeller.automodel import *     # Load the automodel class
log.verbose()                        # request verbose output
env = environ()                      # create a new MODELLER environment

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'

a = automodel(env,
              alnfile  = 'blbp-1hms.ali',     # alignment filename
              knowns   = '1hms',              # codes of the templates
              sequence = 'blbp')              # code of the target
a.starting_model= 1                  # index of the first model
a.ending_model  = 1                  # index of the last model
                                     # (determines how many models to calculate)
a.make()                             # do the actual homology modelling
```

Run by typing `mod9v3 model.py` in the directory where you have the python file. MODELLER will produce a `align.log` file

Modeling of BLBP

STEP 2: Model the blbp structure using the alignment from step 1.

Python script for model building

PDB file

Can be viewed with Chimera

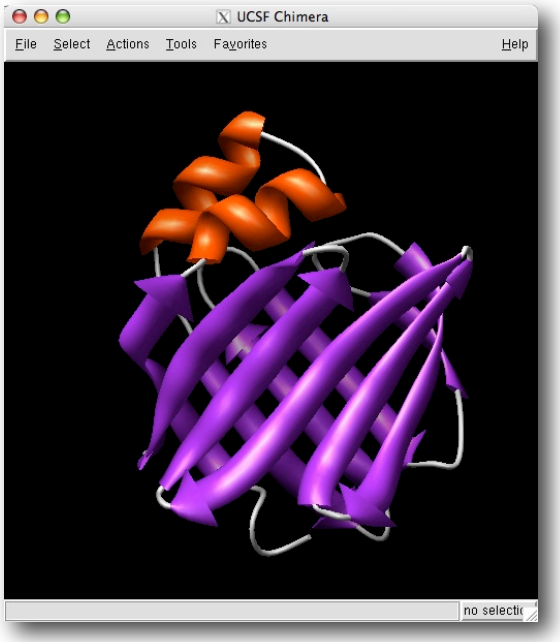
http://www.cgl.ucsf.edu/chimera/

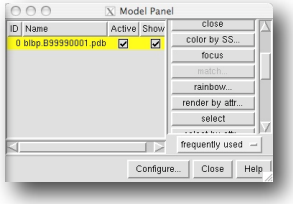
Rasmol

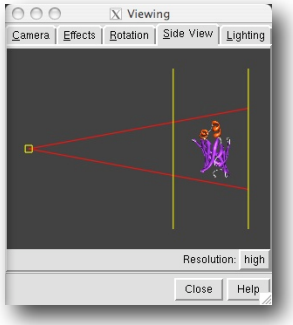
http://www.openrasmol.org

PyMol

http://pymol.sourceforge.net/







Model file →

`blbp.B99990001.pdb`

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

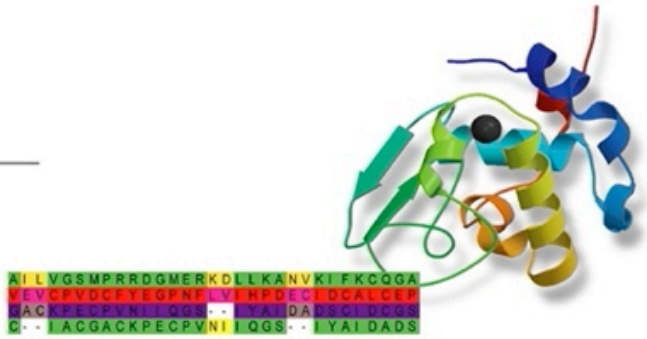
Tutorial

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

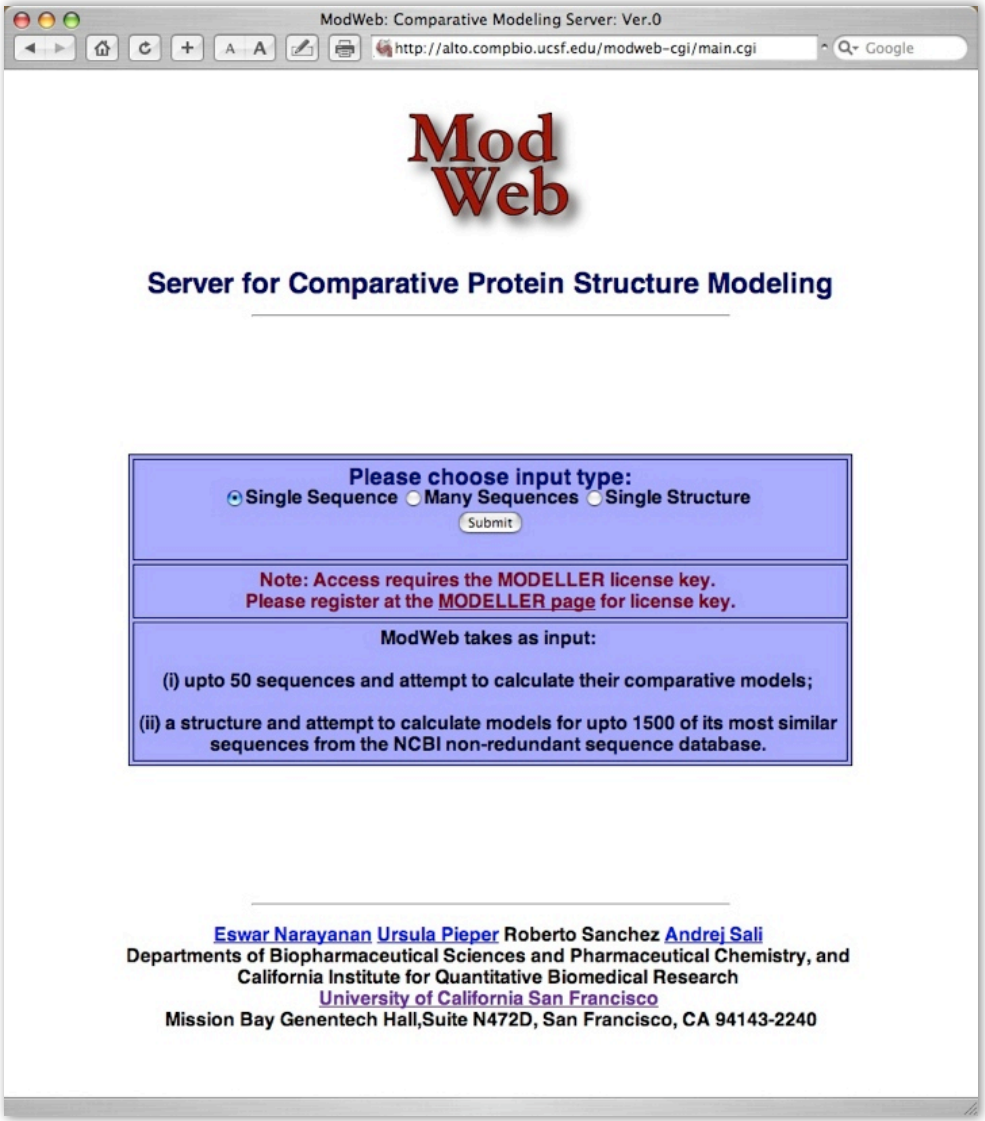
MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2,3). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms.

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

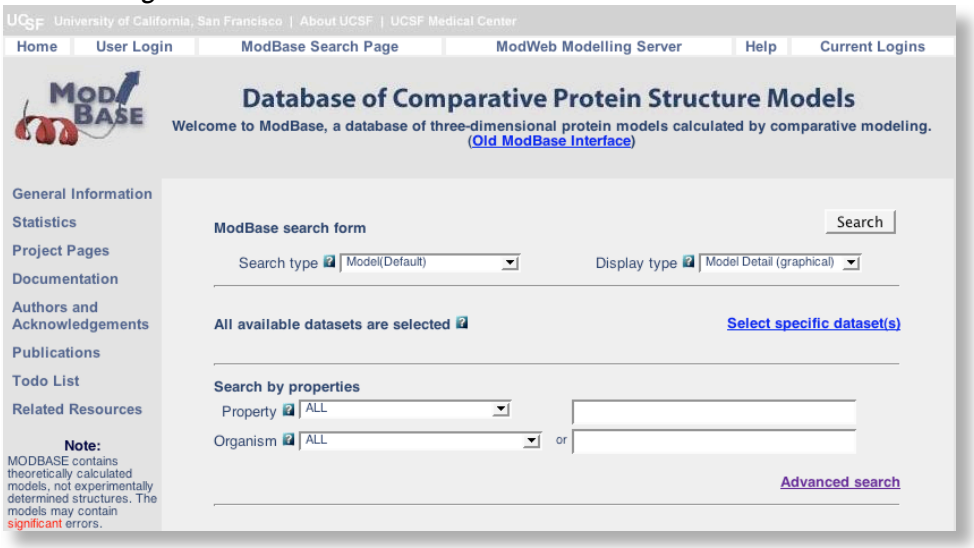
http://salilab.org/modweb



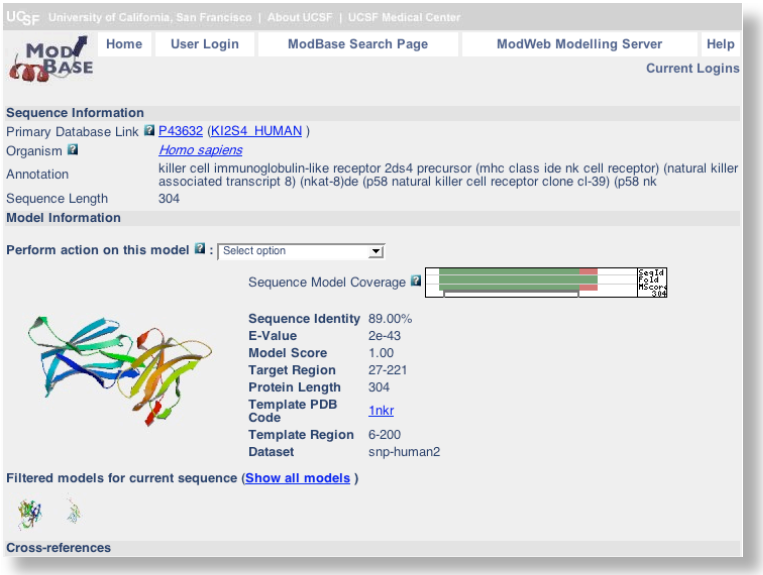
MODBASE

http://salilab.org/modbase

Search Page



Model Details



Sequence Overview

SeqId Fold MScore	Q8G8A6	hypothetical protein	*Pseudomonas aeruginosa*	3738
SeqId Fold MScore	Q8G9W1	hypothetical protein	*Escherichia coli*	1140
SeqId Fold MScore	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

Acknowledgments

Structural Genomics Unit (CIPF)
Marc A. Marti-Renom
Emidio Capriotti
Peio Ziarsolo Areitioaurtena

Comparative Genomics Unit (CIPF)
Hernán Dopazo
Leo Arbiza
Francisco García

Functional Genomics Unit (CIPF)
Joaquín Dopazo
Fátima Al-Shahrour
José Carbonell
Ignacio Medina
David Montaner
Joaquin Tárraga
Ana Conesa
Toni Gabaldón
Eva Alloza
Lucía Conde
Stefan Goetz
Jaime Huerta Cepas
Marina Marcet
Pablo Minguez
Jordi Burguet Castell

FUNDING
Prince Felipe Research Center
Marie Curie Reintegration Grant
STREP EU Grant
Generalitat Valenciana

Tropical Disease Initiative
Stephen Maurer (UC Berkeley)
Arti Rai (Duke U)
Andrej Sali (UCSF)
Ginger Taylor (TSL)
Barri Bunin (CDD)

STRUCTURAL GENOMICS
Stephen Burley (SGX)
John Kuriyan (UCB)
NY-SGXRC

MAMMOTH
Angel R. Ortiz

BIOLOGY
Jeff Friedman (RU)
James Hudsped (RU)
Partho Ghosh (UCSD)
Alvaro Monteiro (Cornell U)
Stephen Krilis (St.George H)

FUNCTIONAL ANNOTATION
Fatima Al-Shahrour
Joaquin Dopazo

COMPARATIVE MODELING
Andrej Sali
M. S. Madhusudhan
Narayanan Eswar
Min-Yi Shen
Ursula Pieper
Bino John
Maya Topf

FUNCTIONAL ANNOTATION
Andrea Rossi
Fred Davis

http://bioinfo.cipf.es
http://sgu.bioinfo.cipf.es