

Comparative protein structure modeling of genes and genomes

Marc A. Marti-Renom

Department of Biopharmaceutical Sciences
University of California, San Francisco

Comparative Modeling...

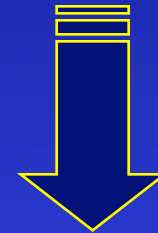
Why protein structure prediction?

	Y 2003	Y 2005
Sequences	1,000,000	millions
Structures	18,000	50,000

Why protein structure prediction?

	Y 2003
Sequences	1,000,000
Structures	18,000

Theory

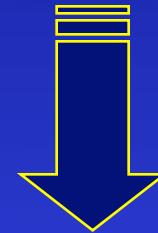


Experiment

Why protein structure prediction?

	Y 2003
Sequences	1,000,000
Structures	400,000

Theory



Experiment

[http://salilab.org/
modbase](http://salilab.org/modbase)

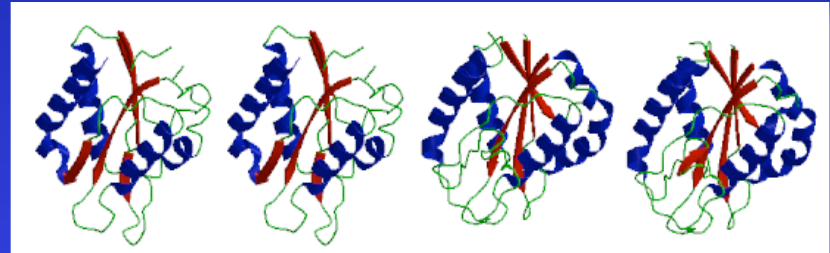
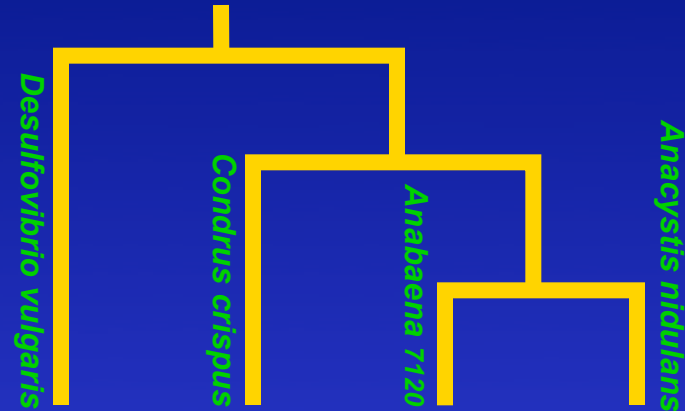
Principles of Protein Structure

GFCHIKAYTRLIMVG...



Folding

Ab initio prediction



Evolution

Threading
Comparative Modeling

Comparative Modeling by Satisfaction of Spatial Restraints (MODELLER)

3D GKITYFIERGFQGHCIYESDC-NLQP...
SEQ GKITYFIERG---RCYESDCPNLQP...

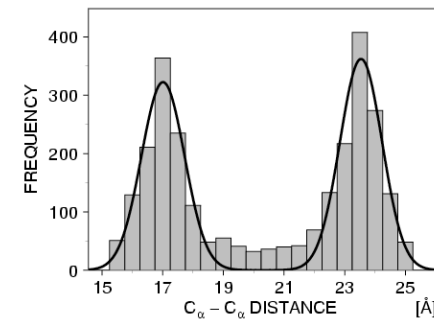
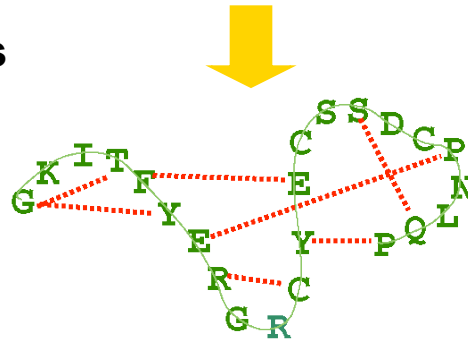
A. Šali & T. Blundell. *J. Mol. Biol.* **234**, 779, 1993.
J.P. Overington & A. Šali. *Prot. Sci.* **3**, 1582, 1994.
A. Fiser, R. Do & A. Šali. *Prot Sci.* **9**, 1753, 2000.

<http://salilab.org/modeller>

Comparative Modeling by Satisfaction of Spatial Restraints (MODELLER)

3D GKITYFIERGFQGHCEYSDC-NLQP...
SEQ GKITYFIERG---RCYESDCPNLQP...

1. Extract spatial restraints



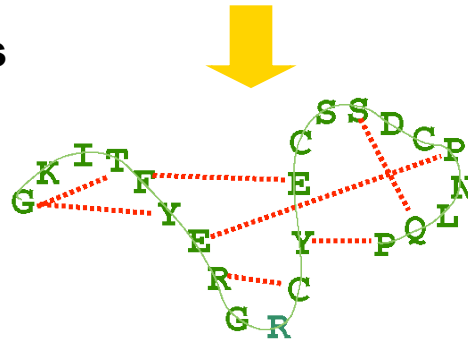
A. Šali & T. Blundell. *J. Mol. Biol.* **234**, 779, 1993.
J.P. Overington & A. Šali. *Prot. Sci.* **3**, 1582, 1994.
A. Fiser, R. Do & A. Šali. *Prot Sci.* **9**, 1753, 2000.

<http://salilab.org/modeller>

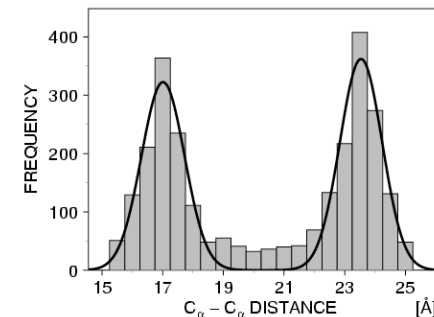
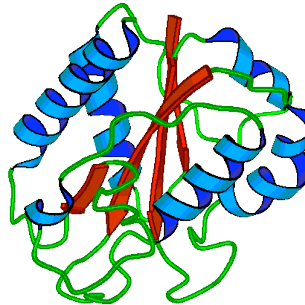
Comparative Modeling by Satisfaction of Spatial Restraints (MODELLER)

3D GKITFYERGFQGH CYESDC-NLQP...
SEQ GKITFYERG---RCYESDCPNLQP...

1. Extract spatial restraints



2. Satisfy spatial restraints



$$F(\mathbf{R}) = \prod_i p_i(f_i / I)$$

Steps in Comparative Protein Structure Modeling

START

TARGET

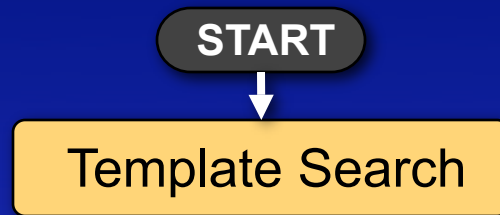
ASILPKRLFGNCEQTSDEGLK
IERTPLVPHISAQNVCLKIDD
VPERLIPERASFQWMNDK

A. Šali, *Curr. Opin. Biotech.* 6, 437, 1995.

R. Sánchez & A. Šali, *Curr. Opin. Str. Biol.* 7, 206, 1997.

M. A. Martí-Renom et al. *Ann. Rev. Biophys. Biomolec. Struct.*, 29, 291, 2000.

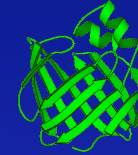
Steps in Comparative Protein Structure Modeling



TARGET

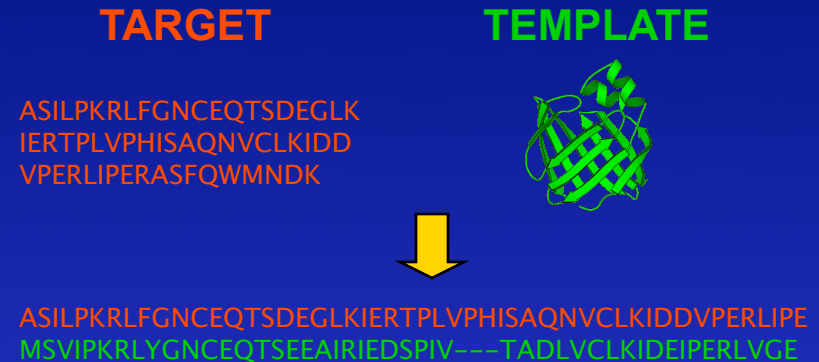
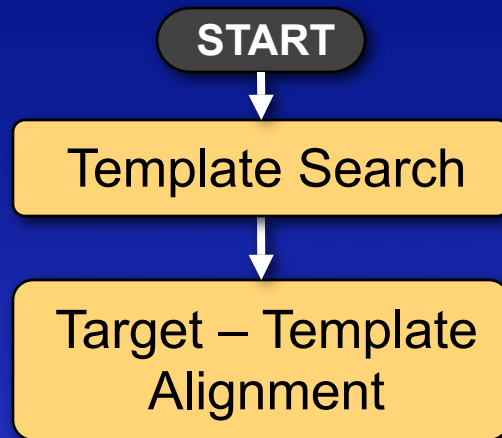
ASILPKRLFGNCEQTSDEGLK
IERTPLVPHISAQNVCLKIDD
VPERLIPERASFQWMNDK

TEMPLATE



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

Steps in Comparative Protein Structure Modeling

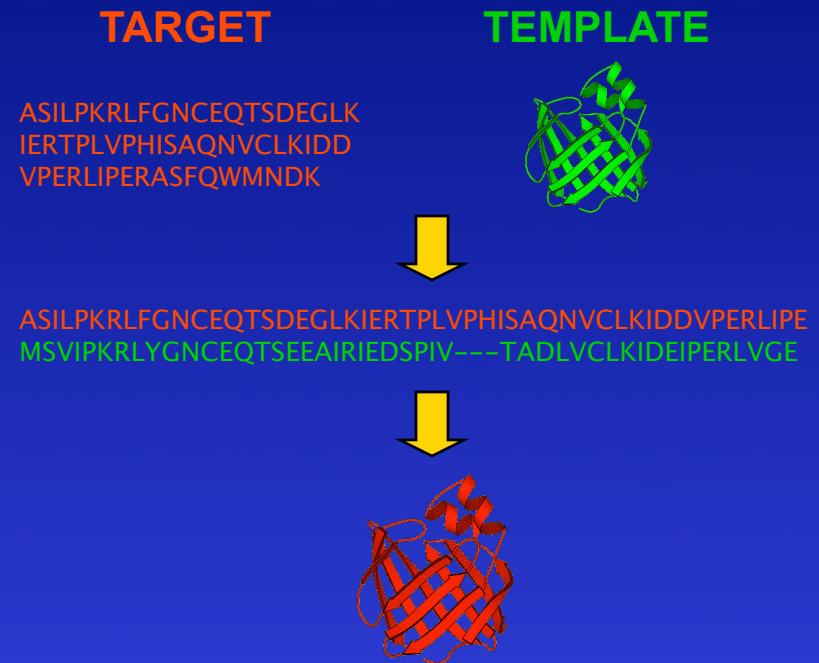
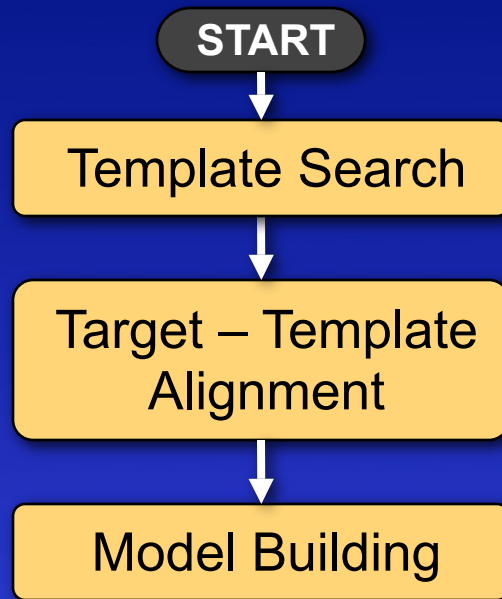


A. Šali, *Curr. Opin. Biotech.* 6, 437, 1995.

R. Sánchez & A. Šali, *Curr. Opin. Str. Biol.* 7, 206, 1997.

M. A. Martí-Renom et al. *Ann. Rev. Biophys. Biomolec. Struct.*, 29, 291, 2000.

Steps in Comparative Protein Structure Modeling

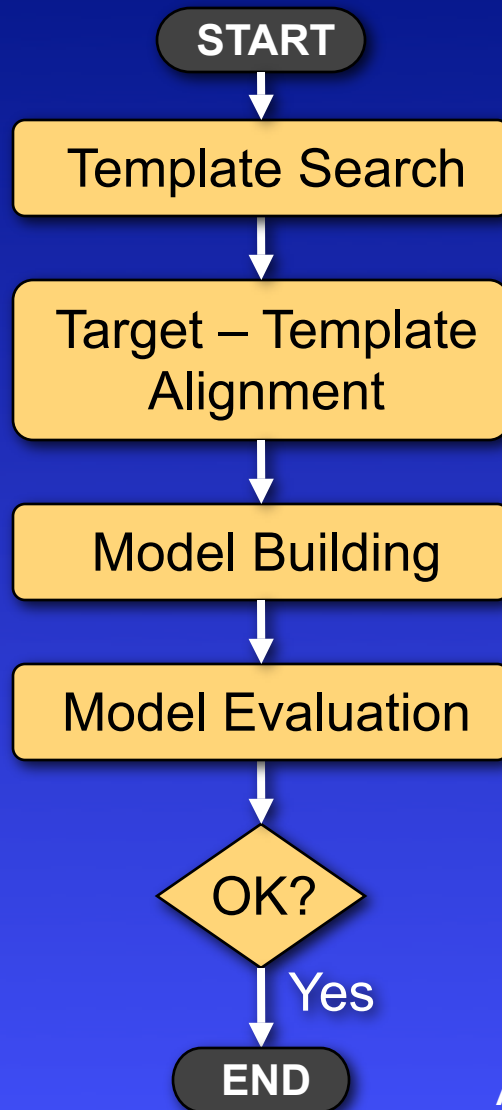


A. Šali, *Curr. Opin. Biotech.* 6, 437, 1995.

R. Sánchez & A. Šali, *Curr. Opin. Str. Biol.* 7, 206, 1997.

M. A. Martí-Renom et al. *Ann. Rev. Biophys. Biomolec. Struct.*, 29, 291, 2000.

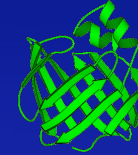
Steps in Comparative Protein Structure Modeling



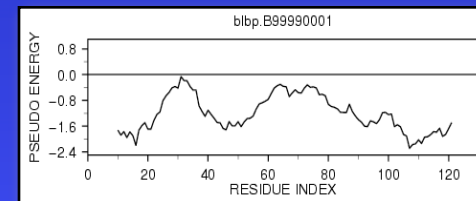
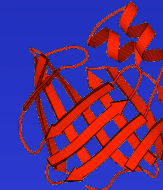
TARGET

ASILPKRLFGNCEQTSDEGLK
IERTPLVPHISAQNVCLKIDD
VPERLIPERASFQWMNDK

TEMPLATE



ASILPKRLFGNCEQTSDEGLKIERTPLVPHISAQNVCLKIDDVPERLIPE
MSVIPKRLYGNCEQTSEEAIRIEDSPIV---TADLVCLKIDEIPERLVGE

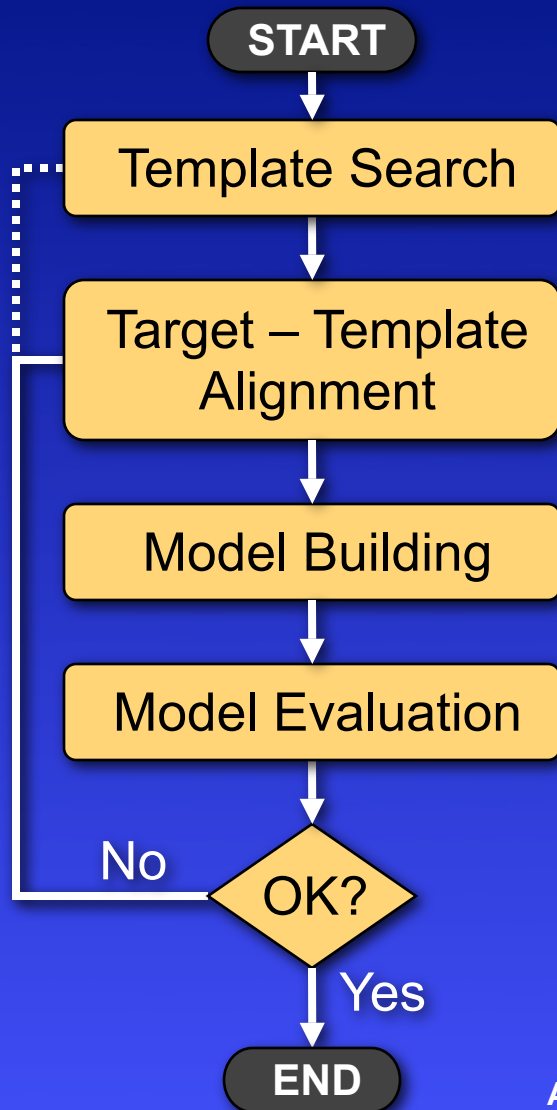


A. Šali, *Curr. Opin. Biotech.* 6, 437, 1995.

R. Sánchez & A. Šali, *Curr. Opin. Str. Biol.* 7, 206, 1997.

M. A. Martí-Renom et al. *Ann. Rev. Biophys. Biomolec. Struct.*, 29, 291, 2000.

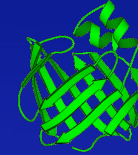
Steps in Comparative Protein Structure Modeling



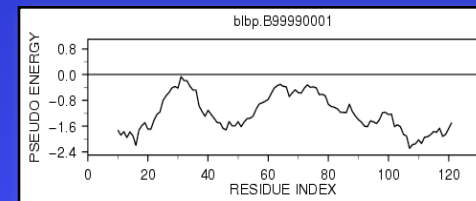
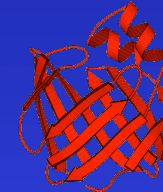
TARGET

ASILPKRLFGNCEQTSDEGLK
IERTPLVPHISAQNVCLKIDD
VPERLIPERASFQWMNDK

TEMPLATE



ASILPKRLFGNCEQTSDEGLKIERTPLVPHISAQNVCLKIDDVPERLIPE
MSVIPKRLYGNCEQTSEEAIRIEDSPIV---TADLVCLKIDEIPERLVGE

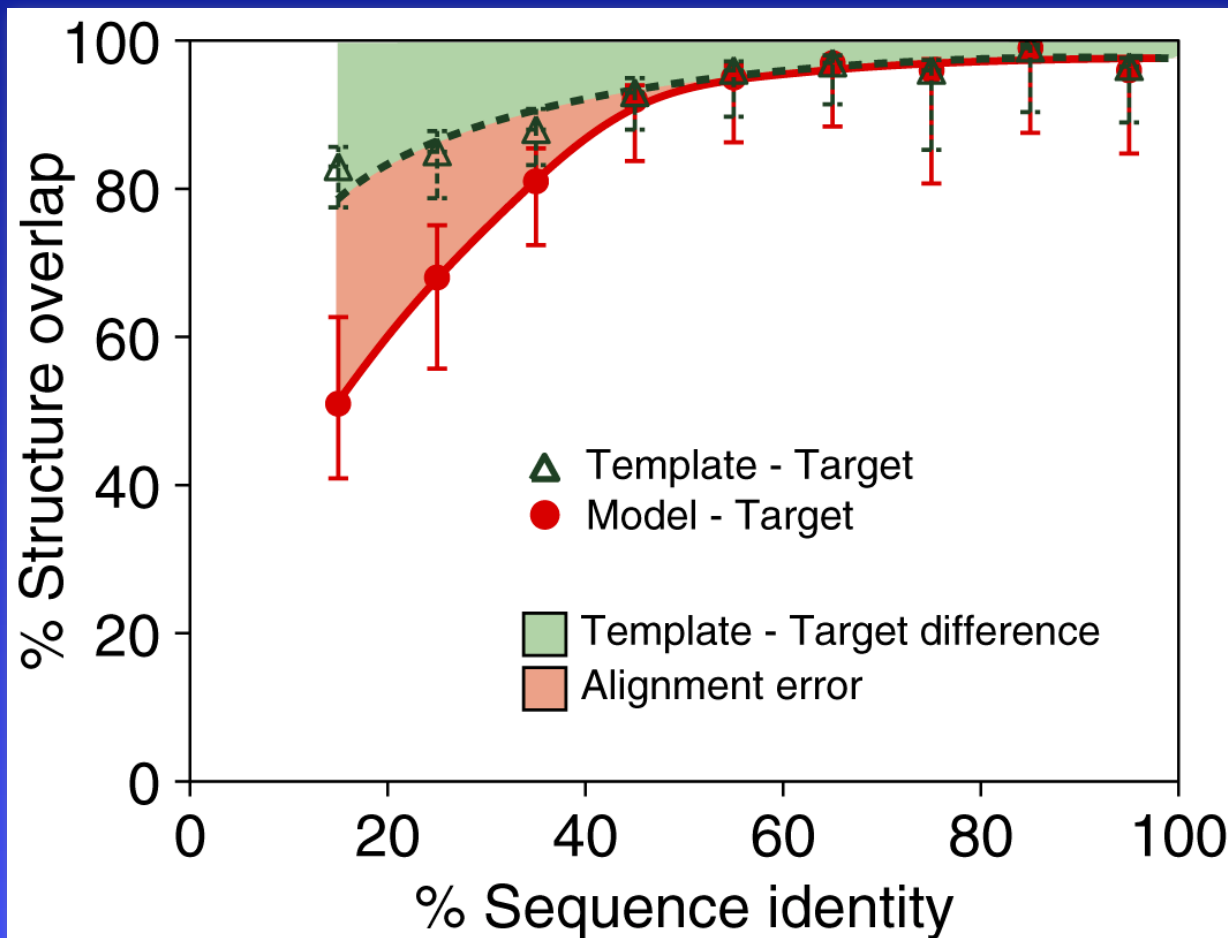


A. Šali, *Curr. Opin. Biotech.* 6, 437, 1995.

R. Sánchez & A. Šali, *Curr. Opin. Str. Biol.* 7, 206, 1997.

M. A. Martí-Renom et al. *Ann. Rev. Biophys. Biomolec. Struct.*, 29, 291, 2000.

Model Accuracy as a Function of Target-Template Sequence Identity



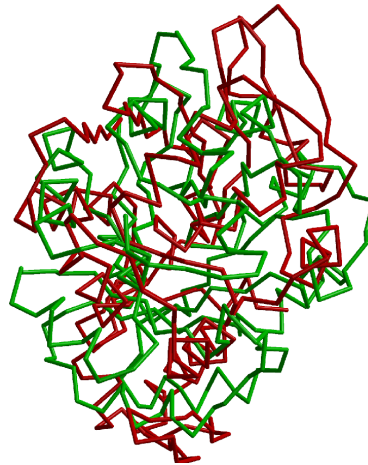
Typical Errors in Comparative Models

MODEL

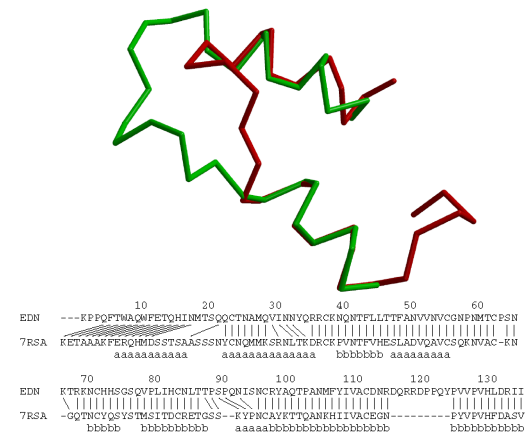
X-RAY

TEMPLATE

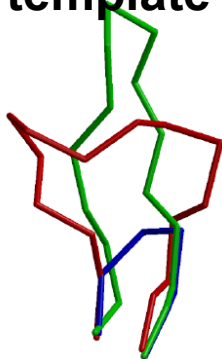
Incorrect template



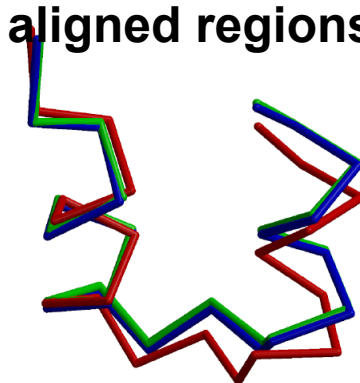
Misalignment



Region without a
template



Distortion in correctly
aligned regions



Sidechain packing

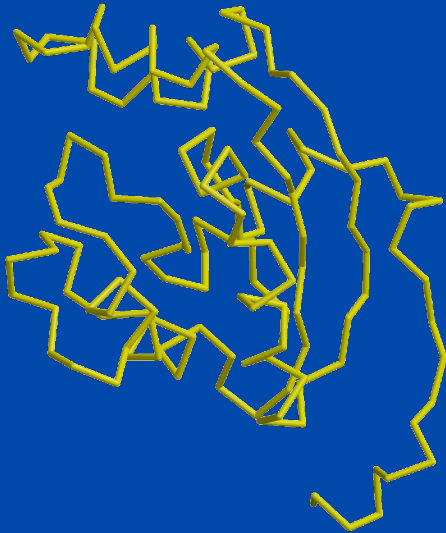


Model Accuracy

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

HIGH ACCURACY

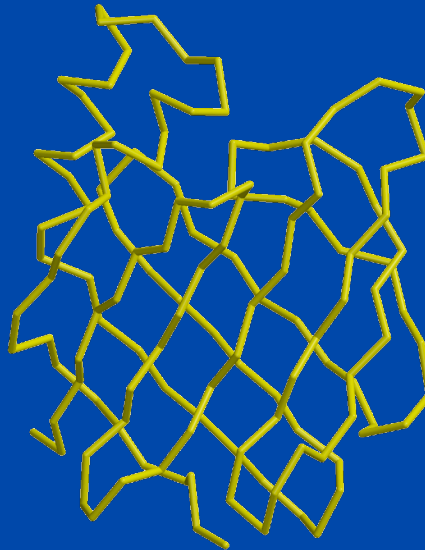
NM23
Seq id 77%



X-RAY

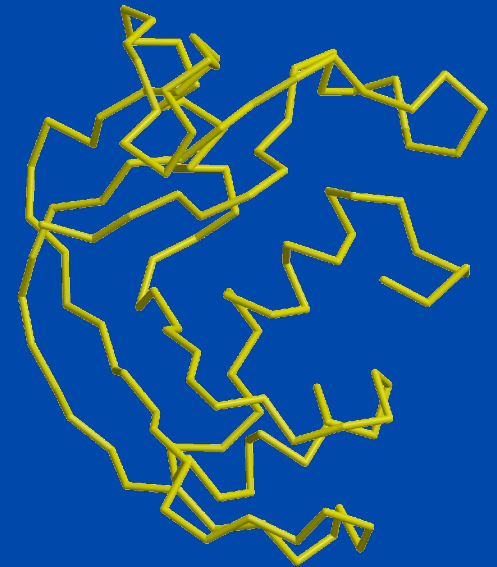
MEDIUM ACCURACY

CRABP
Seq id 41%



LOW ACCURACY

EDN
Seq id 33%

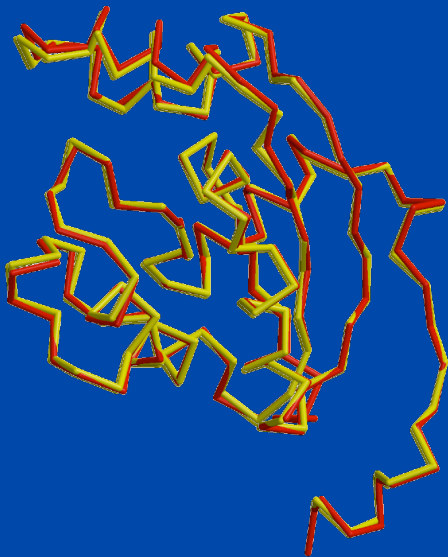


Model Accuracy

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

HIGH ACCURACY

NM23
Seq id 77%
C α equiv 147/148
RMSD 0.41Å

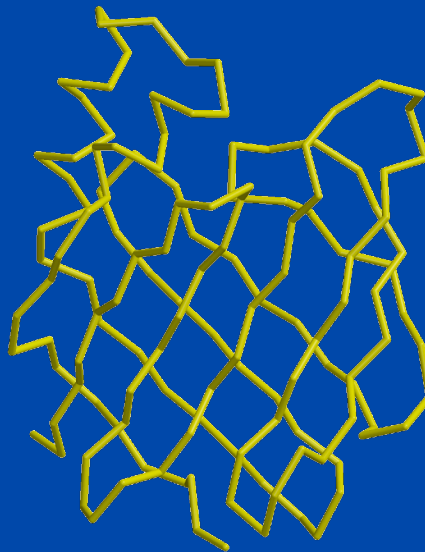


Sidechains
Core backbone
Loops

X-RAY / MODEL

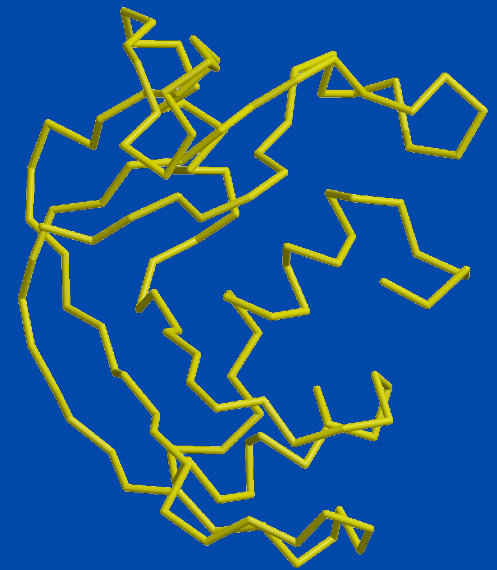
MEDIUM ACCURACY

CRABP
Seq id 41%



LOW ACCURACY

EDN
Seq id 33%

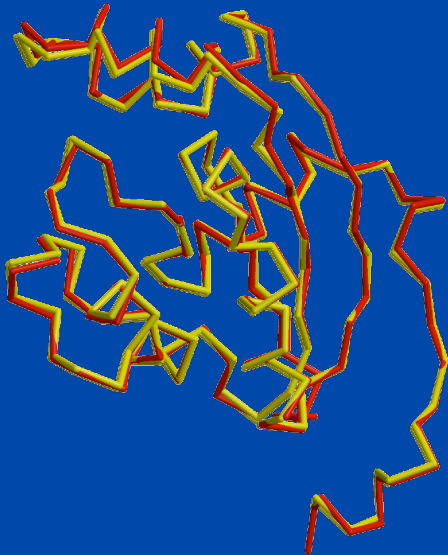


Model Accuracy

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

HIGH ACCURACY

NM23
Seq id 77%
C α equiv 147/148
RMSD 0.41Å

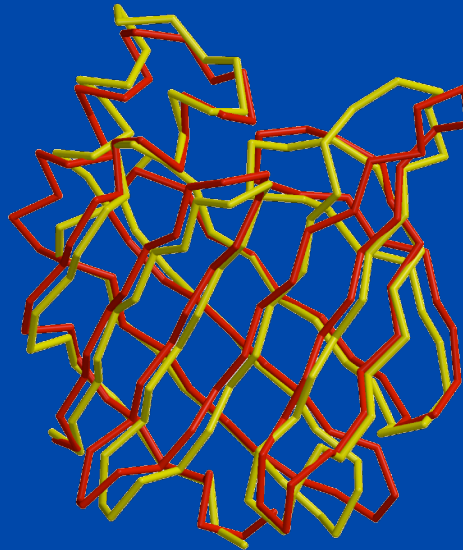


Sidechains
Core backbone
Loops

X-RAY / MODEL

MEDIUM ACCURACY

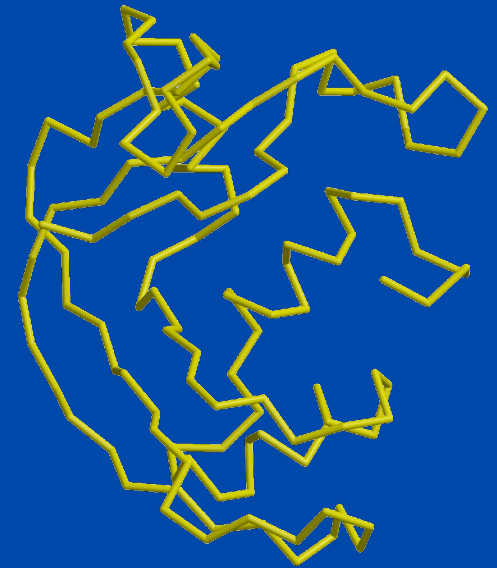
CRABP
Seq id 41%
C α equiv 122/137
RMSD 1.34Å



Sidechains
Core backbone
Loops
Alignment

LOW ACCURACY

EDN
Seq id 33%

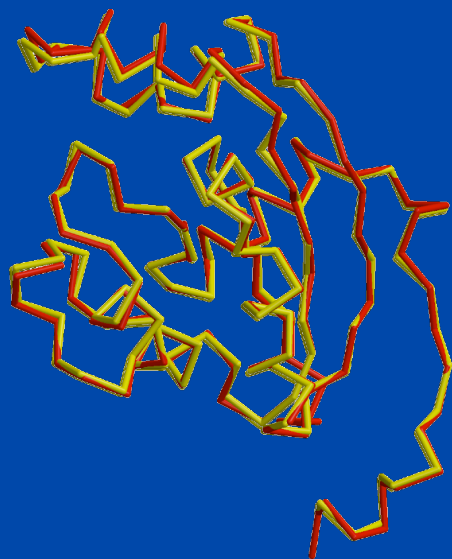


Model Accuracy

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

HIGH ACCURACY

NM23
Seq id 77%
C α equiv 147/148
RMSD 0.41Å

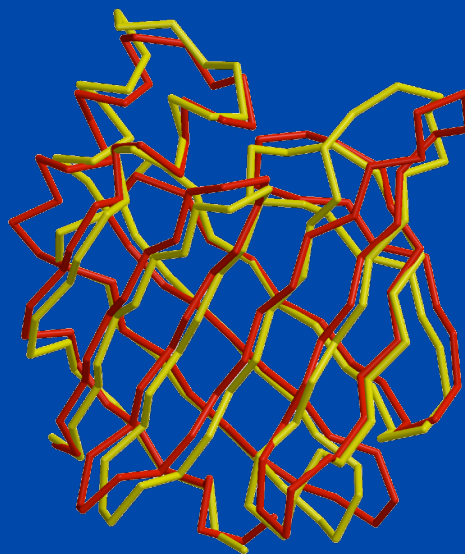


Sidechains
Core backbone
Loops

X-RAY / **MODEL**

MEDIUM ACCURACY

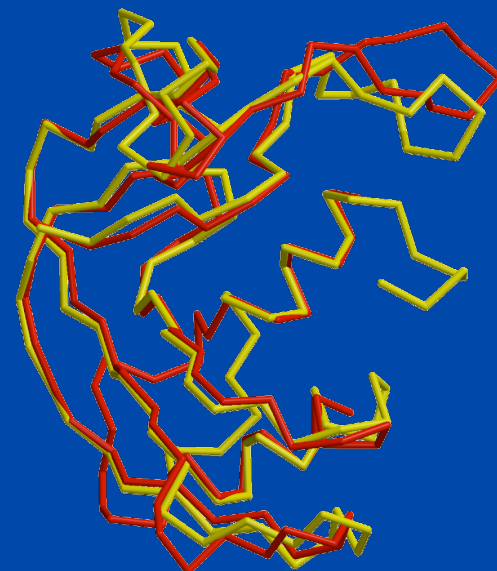
CRABP
Seq id 41%
C α equiv 122/137
RMSD 1.34Å



Sidechains
Core backbone
Loops
Alignment

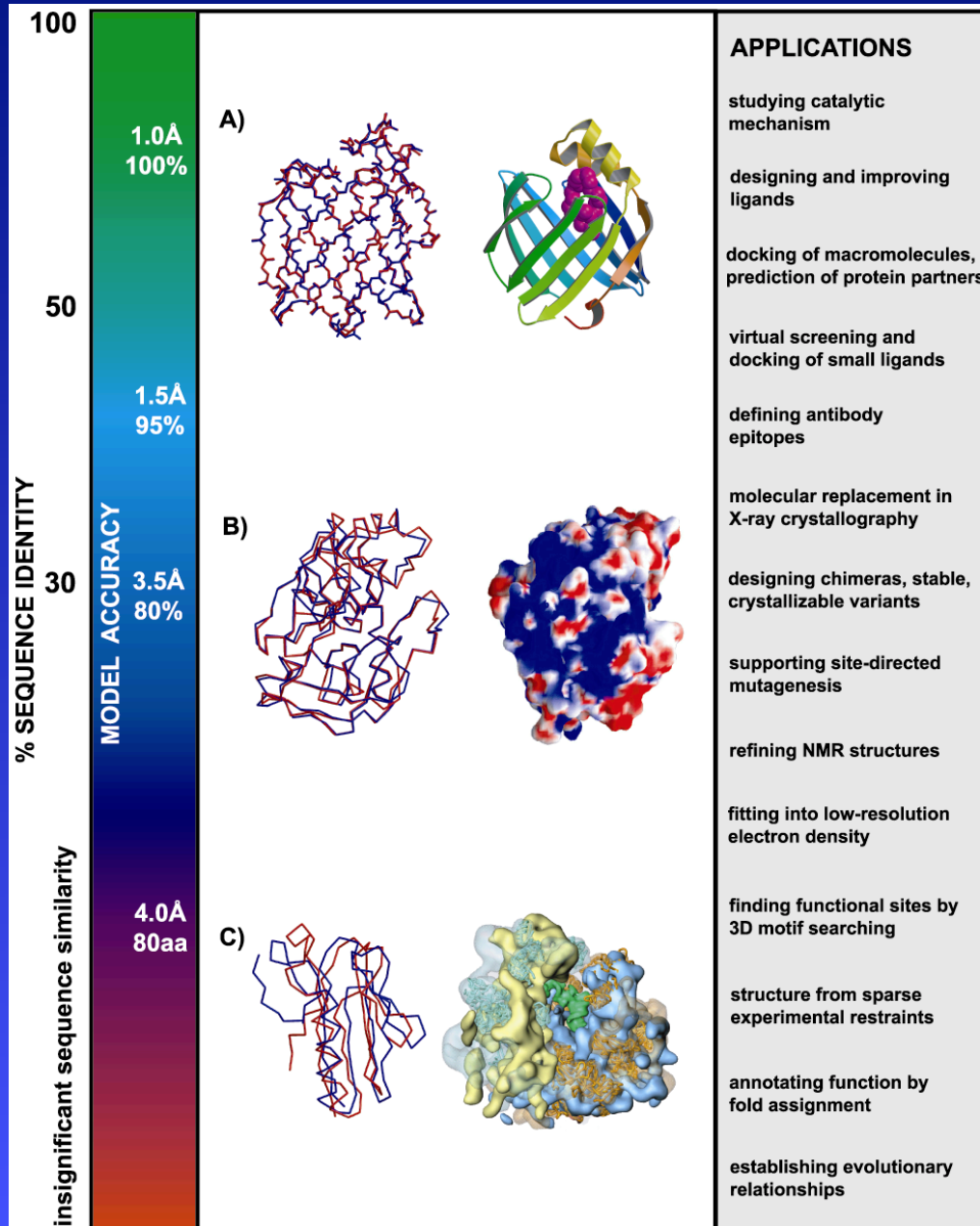
LOW ACCURACY

EDN
Seq id 33%
C α equiv 90/134
RMSD 1.17Å



Sidechains
Core backbone
Loops
Alignment
Fold assignment

Applications of Comparative Models



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

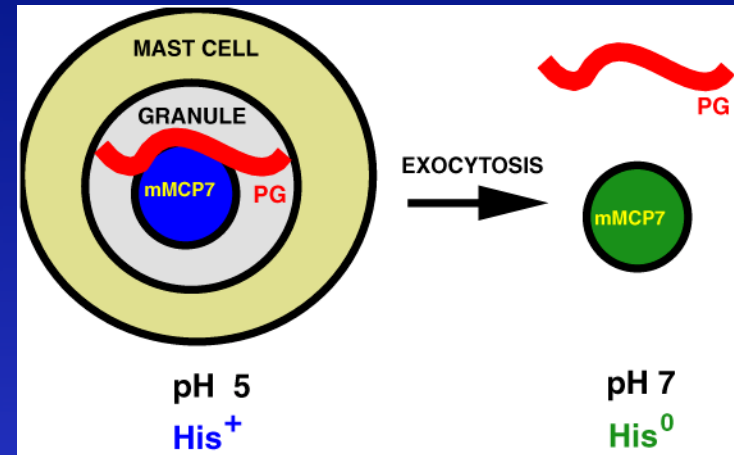
A. Sali & J. Kuriyan.
TIBS **22**, M20, 1999.

genes...

Do mast cell proteases bind proteoglycans? Where? When?

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

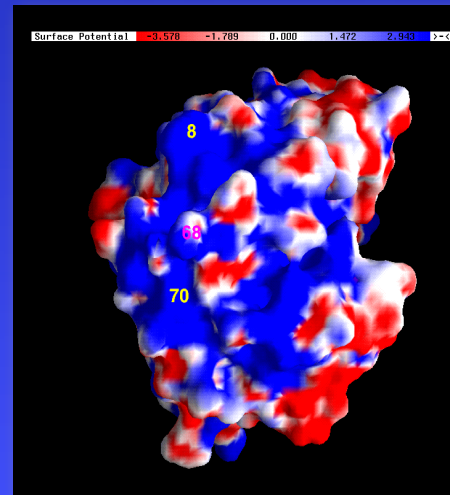
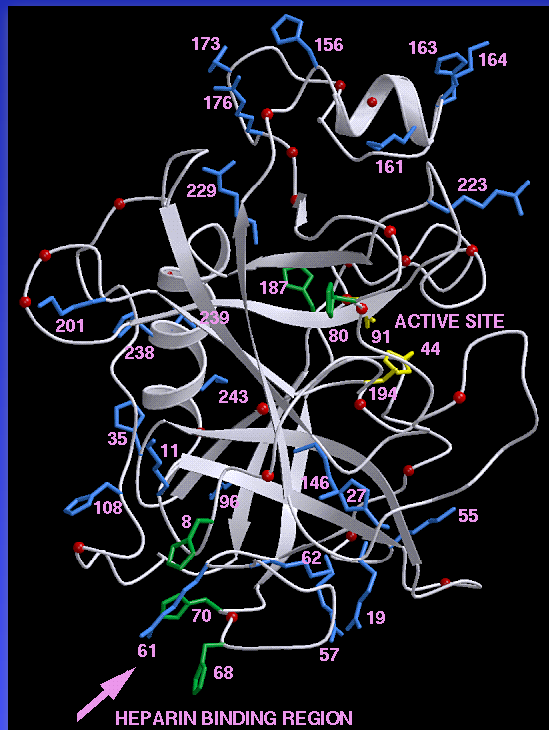
1. mMCPs bind negatively charged proteoglycans through electrostatic interactions?
2. Comparative models used to find clusters of positively charged surface residues.
3. Tested by site-directed mutagenesis.



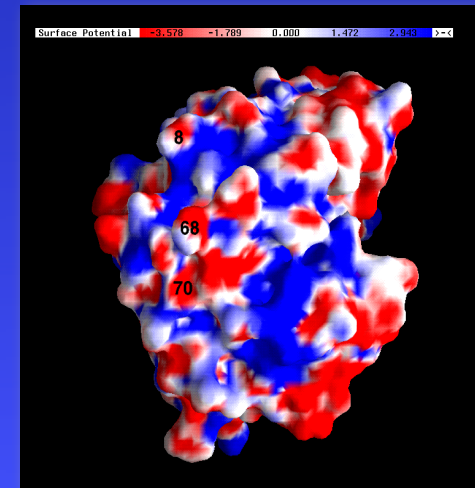
Huang et al. *J. Clin. Immunol.* **18**,169,1998.

Matsumoto et al. *J. Biol. Chem.* **270**,19524,1995.

Šali et al. *J. Biol. Chem.* **268**, 9023, 1993.



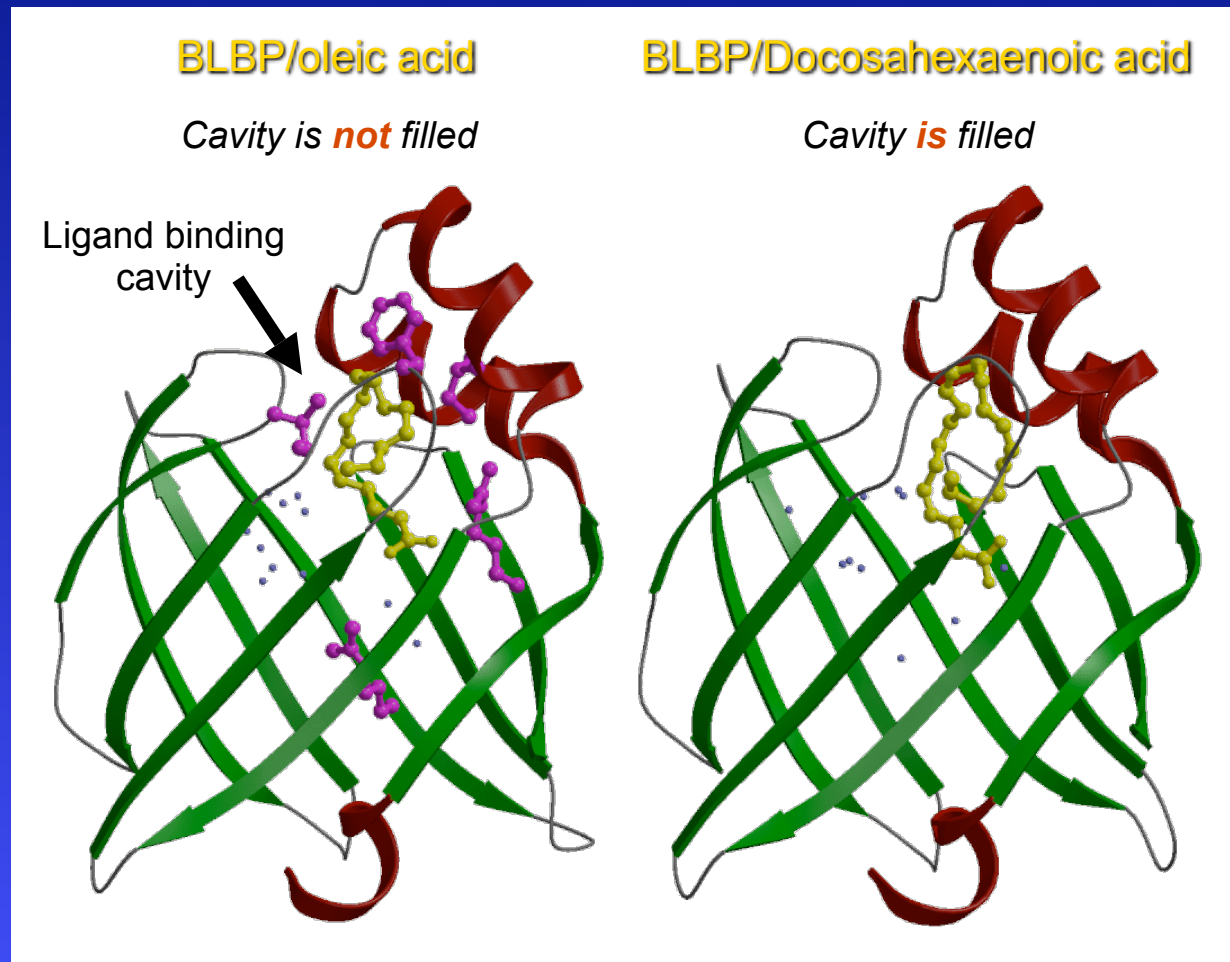
Native mMCP-7 at pH=5 (His⁺)



Native mMCP-7 at pH=7 (His⁰)

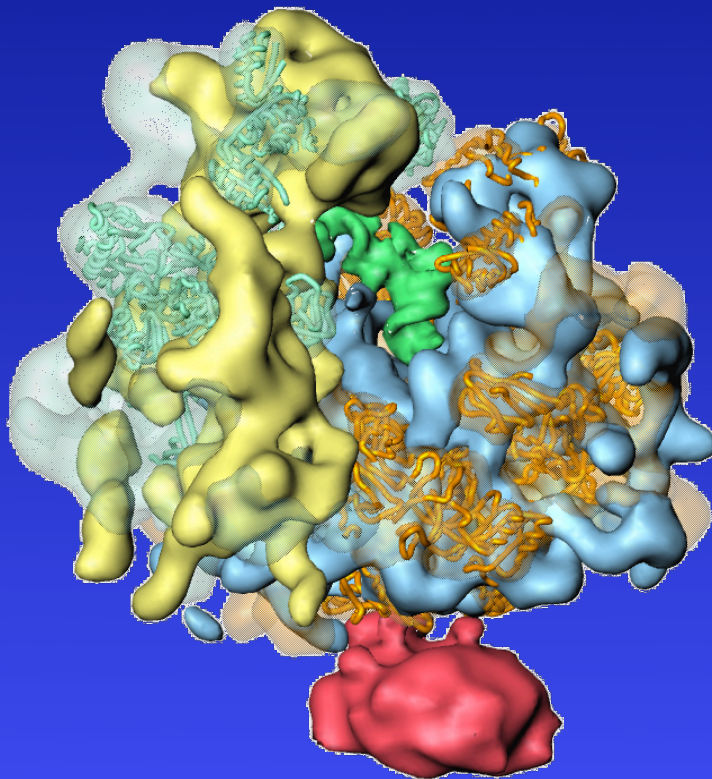
What is the physiological ligand of Brain Lipid-Binding Protein?

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



1. BLBP binds fatty acids.
2. Build a 3D model.
3. Find the fatty acid that fits most snugly into the ligand binding cavity.

Some Models Can Be Used in Docking to Density Maps (Yeast Ribosomal 40S subunit)

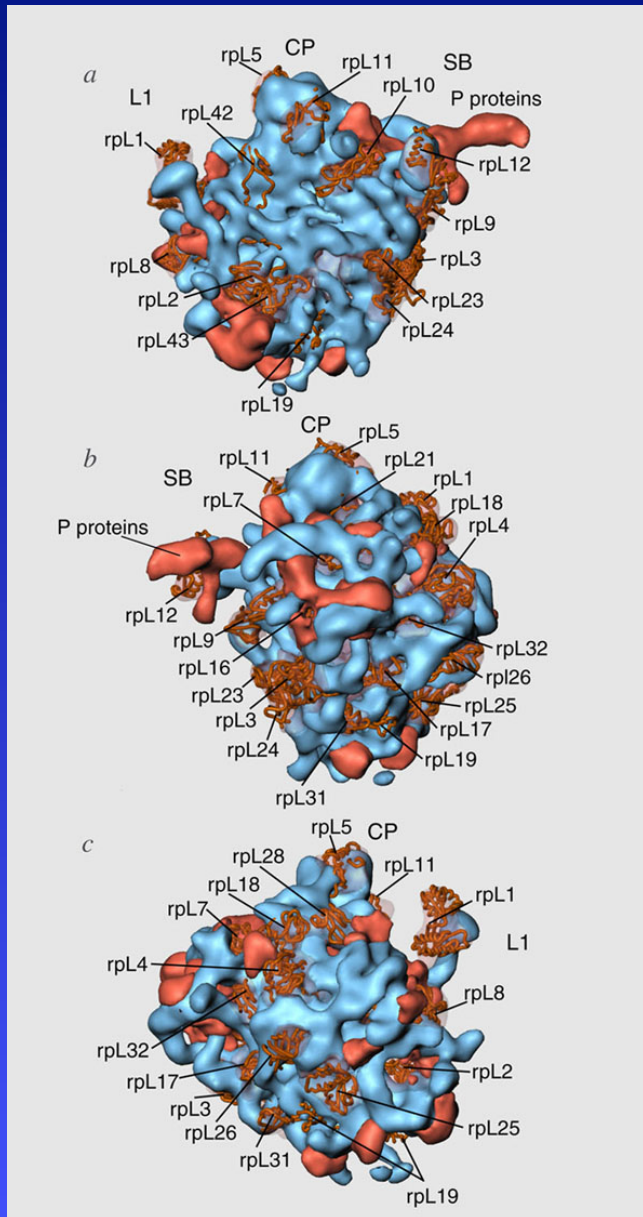


**Docking of comparative
models into the cryo-EM
map.**

Spahn *et al.* 2001 Cell **107**:373-386

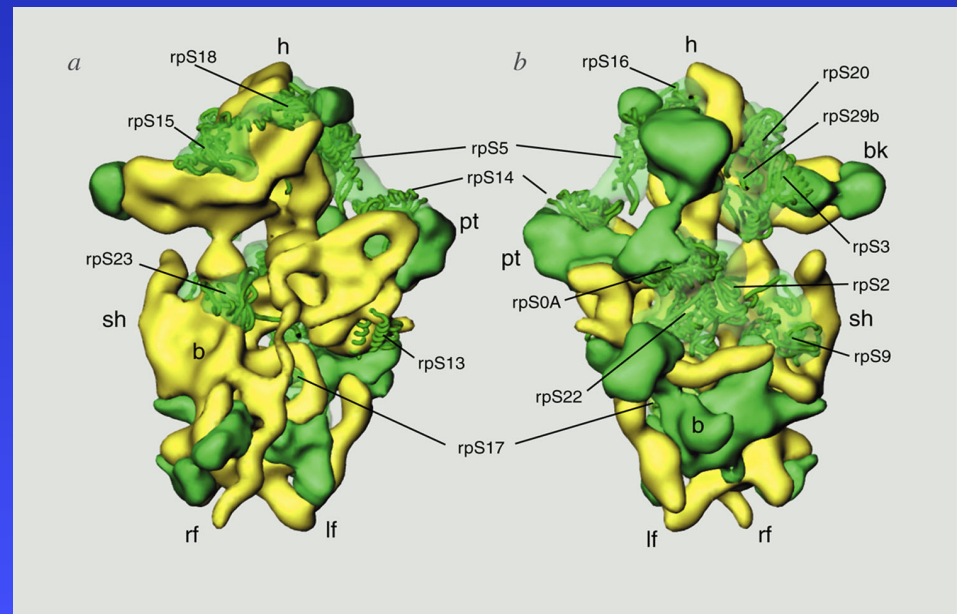
Small 30S subunit from *Thermus thermophilus*
Large 50S subunit from *Haloarcula marismortui*

60S Subunit



- 43 proteins could be modeled on 20-56% seq.id. to a known structure.
- The coverage of the models ranges from 34-99%.
- Models were manually docked into the 15Å cryo-electron density map.
- The solid orange in the 60S subunit and the solid green in the 40S subunit correspond to proteins without known bacterial homologs.

40S Subunit

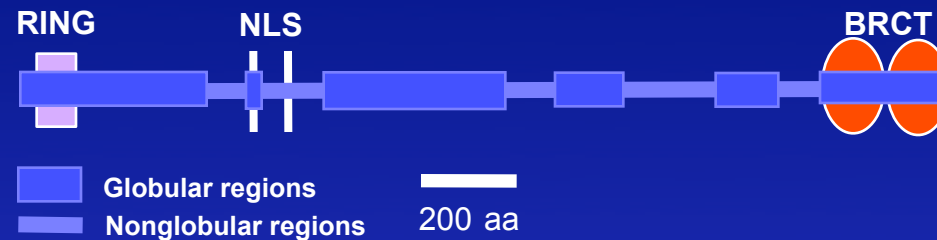


Structural analysis of missense mutations in human BRCA1 BRCT domains

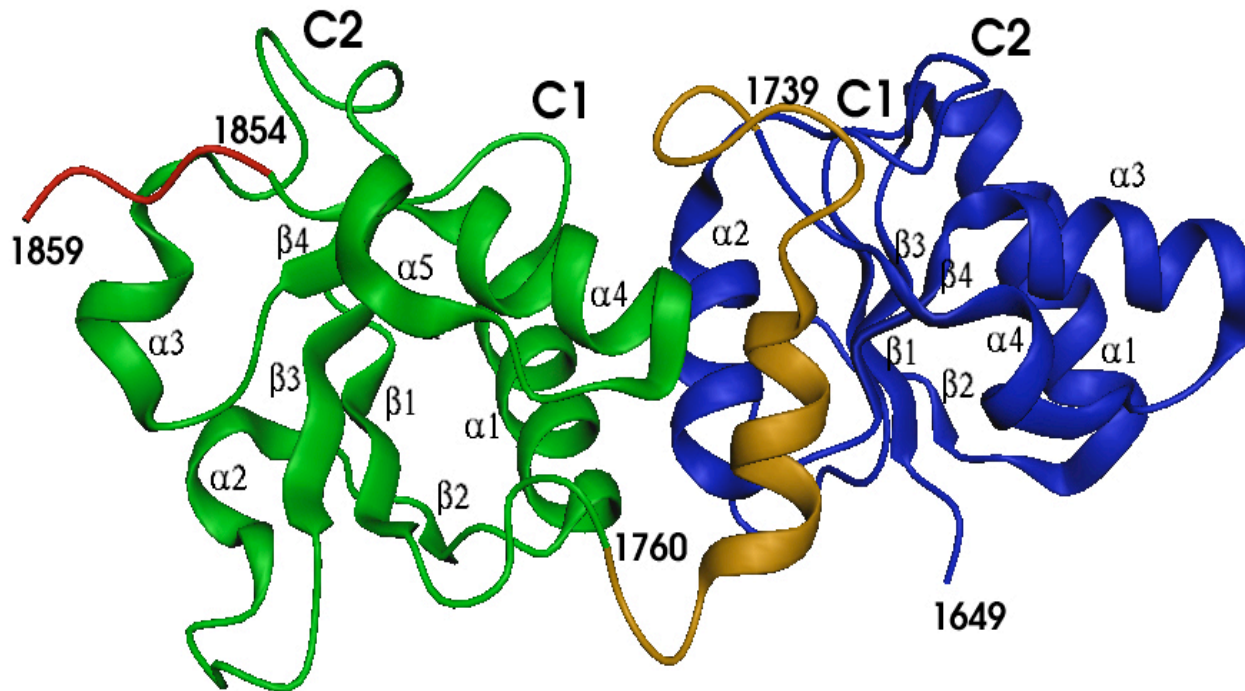
Nebojsa Mirkovic, Marc A. Marti–Renom, Andrej Sali

Alvaro N.A. Monteiro (Sprang Center, Cornell U.)

Human BRCA1 and its two BRCT domains



BRCA1 BRCT repeats, 1jnx



Williams, Green, Glover. *Nat. Struct. Biol.* 8, 838, 2001

CONFIDENTIAL

BRACAnalysis™
Comprehensive BRCA1-BRCA2 Gene Sequence Analysis Result

Strang Cancer Prevention Center
428 E 72nd St
New York, NY 10021

SPECIMEN
Specimen Type: Blood
Draw Date: n/a
Accession Date: Oct 27, 2000
Report Date: Nov 17, 2000

PATIENT
Name:
Date of Birth: Feb 02, 1953
Patient ID:
Gender: Female
Accession #: 00019998
Requisition #: 56694

Physician:

Test Result

<u>Gene Analyzed</u>	<u>Specific Genetic Variant</u>
BRCA2	H2116R
BRCA1	None Detected

Interpretation

GENETIC VARIANT OF UNCERTAIN SIGNIFICANCE

The BRCA2 variant H2116R results in the substitution of arginine for histidine at amino acid position 2116 of the BRCA2 protein. Variants of this type may or may not affect BRCA2 protein function. Therefore, the contribution of this variant to the relative risk of breast or ovarian cancer cannot be established solely from this analysis. The observation by Myriad Genetic Laboratories of this particular variant in an individual with a deleterious truncating mutation in BRCA2, however, reduces the likelihood that H2116R is itself deleterious.

Authorized Signature:

Laboratory Director

Medical Director

These test results should only be used in conjunction with the patient's clinical history and any previous analysis of appropriate family members. It is strongly recommended that these results be communicated to the patient in a setting that includes appropriate counseling. The accompanying Technical Specifications summary describes the analysis, method, performance characteristics, nomenclature, and interpretive criteria of this test. This test may be considered investigational by some states. This test was developed and its performance characteristics determined by Myriad Genetic Laboratories. It has not been reviewed by the U.S. Food and Drug Administration. The FDA has determined that such clearance or approval is not necessary.

CONFIDENTIAL

BRACAnalysis™
Comprehensive BRCA1-BRCA2 Gene Sequence Analysis Result

Strang Cancer Prevention Center
428 E 72nd St
New York, NY 10021

SPECIMEN
Specimen Type: Blood
Draw Date: n/a
Accession Date: Oct 27, 2000
Report Date: Nov 17, 2000

PATIENT
Name:
Date of Birth: Feb 02, 1953
Patient ID:
Gender: Female
Accession #: 00019998
Requisition #: 56694

Physician:

Test Result

Gene Analyzed	Specific Genetic Variant
BRCA2	H2116R
BRCA1	None Detected

Interpretation

GENETIC VARIANT OF UNCERTAIN SIGNIFICANCE

The BRCA2 variant H2116R results in the substitution of arginine for histidine at amino acid position 2116 of the BRCA2 protein. Variants of this type **may or may not** affect BRCA2 protein function. Therefore, the **contribution of this variant to the relative risk of breast or ovarian cancer cannot be established** solely from this analysis. The observation by Myriad Genetic Laboratories of this particular variant in an individual with a deleterious truncating mutation in BRCA2, however, reduces the likelihood that H2116R is itself deleterious.

Authorized Signature:

Laboratory Director

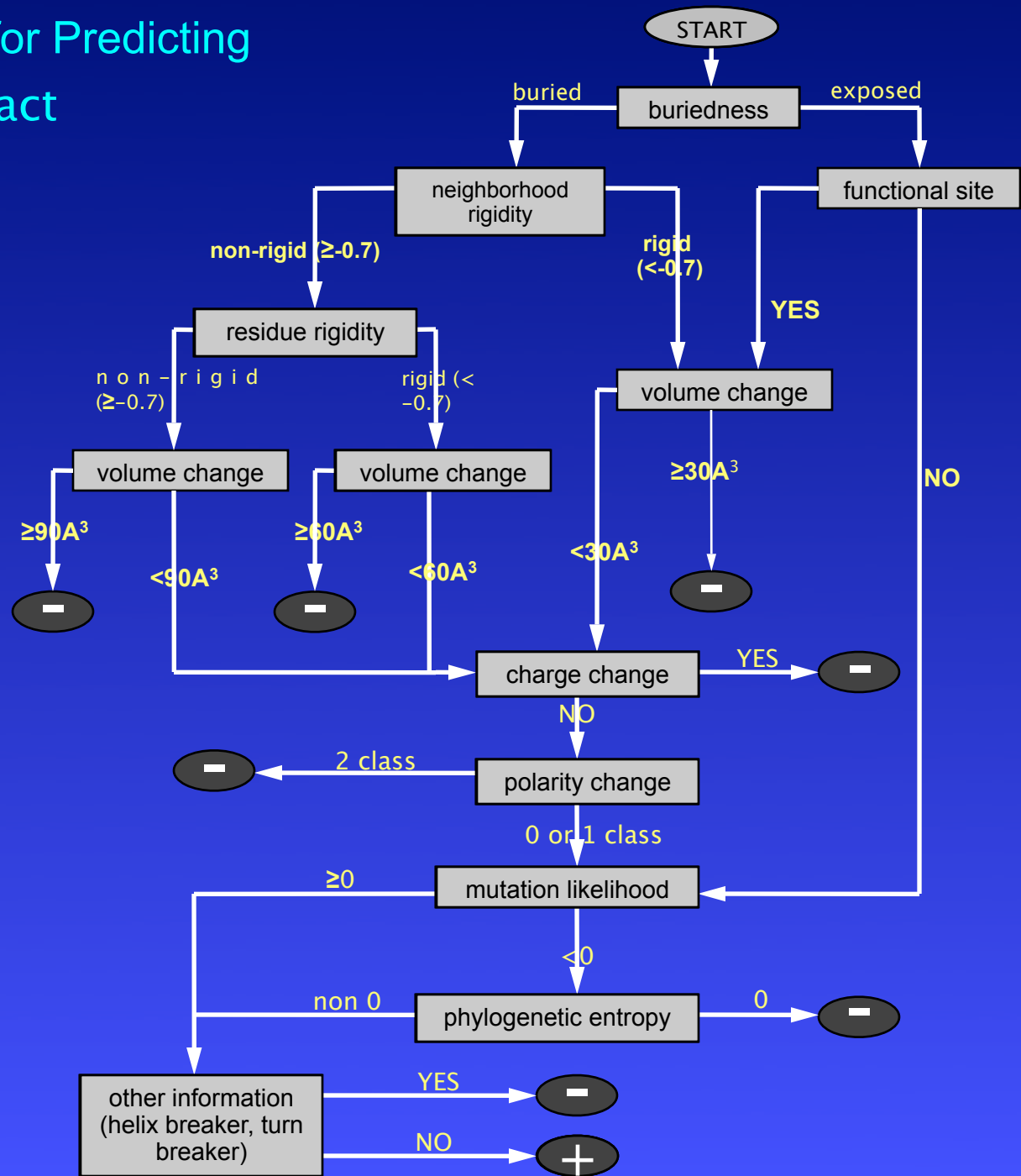
Medical Director

These test results should only be used in conjunction with the patient's clinical history and any previous analysis of appropriate family members. It is strongly recommended that these results be communicated to the patient in a setting that includes appropriate counseling. The accompanying Technical Specifications summary describes the analysis, method, performance characteristics, nomenclature, and interpretive criteria of this test. This test may be considered investigational by some states. This test was developed and its performance characteristics determined by Myriad Genetic Laboratories. It has not been reviewed by the U.S. Food and Drug Administration. The FDA has determined that such clearance or approval is not necessary.

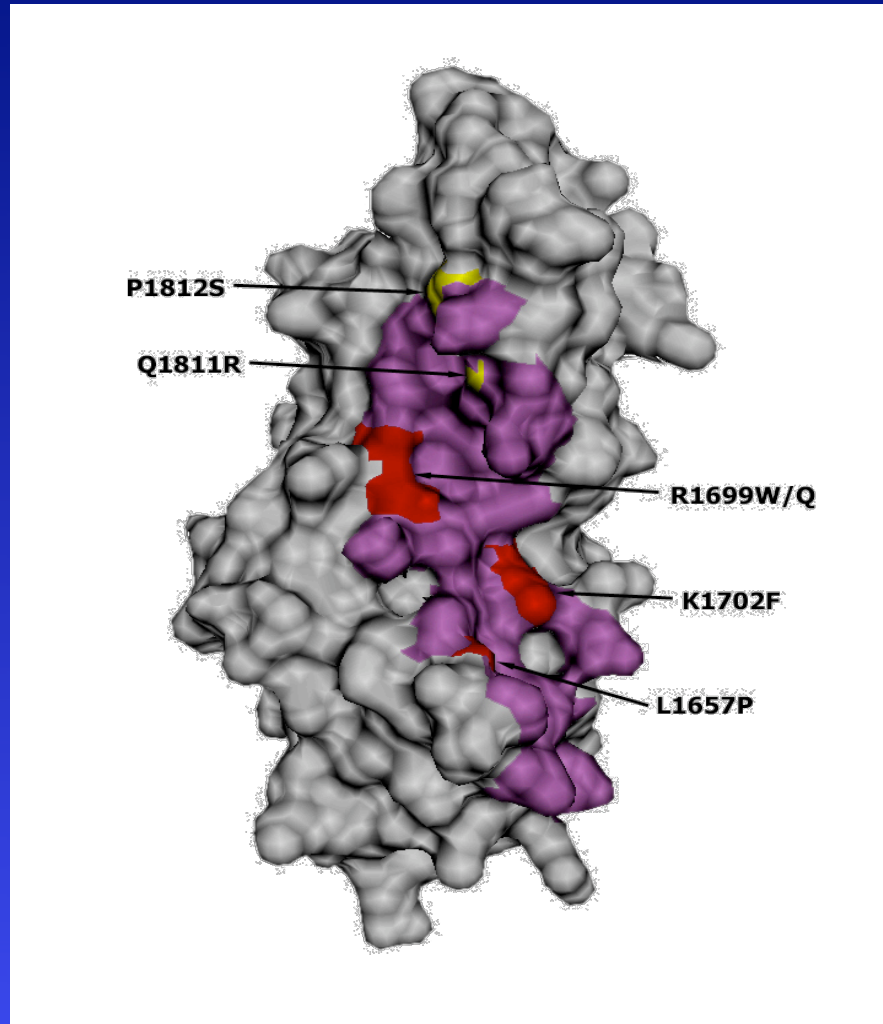
Missense Mutations in BRCT Domains by Function

	cancer associated	not cancer associated	?
no transcription activation	C1697R R1699W A1708E S1715R P1749R M1775R		M1652K L1705PS F1761S L1657P 1715NS M1775E E1660G 1722FF M1775K H1686Q 1734LG L1780P R1699Q 1738EG I1807S K1702E 1743RA V1833E Y1703H 1752PF A1843T F1704S 1761I
transcription activation		M1652I A1669S	V1665M D1692N G1706A D1733G M1775V P1806A
?			M1652T W1718S R1751P C1787S A1823T V1653M T1720A R1751Q G1788 V1833M L1664P W1730S R1758G D W1837R T1685A F1734S L1764P G1788V W1837G T1685I E1735K I1766S G1803A S1841N M1689R V1736A P1771L V1804D A1843P D1692Y G1738R T1773S V1808A T1852S F1695L D1739E P1776S V1809A P1856T V1696L D1739G D1778N V1809F P1859R R1699L D1739Y D1778G V1810G G1706E V1741G D1778H Q1811R W1718C H1746N M1783T P1812S N1819S

“Decision” Tree for Predicting Functional Impact of Genetic Variants



Putative Binding Site on BRCA1



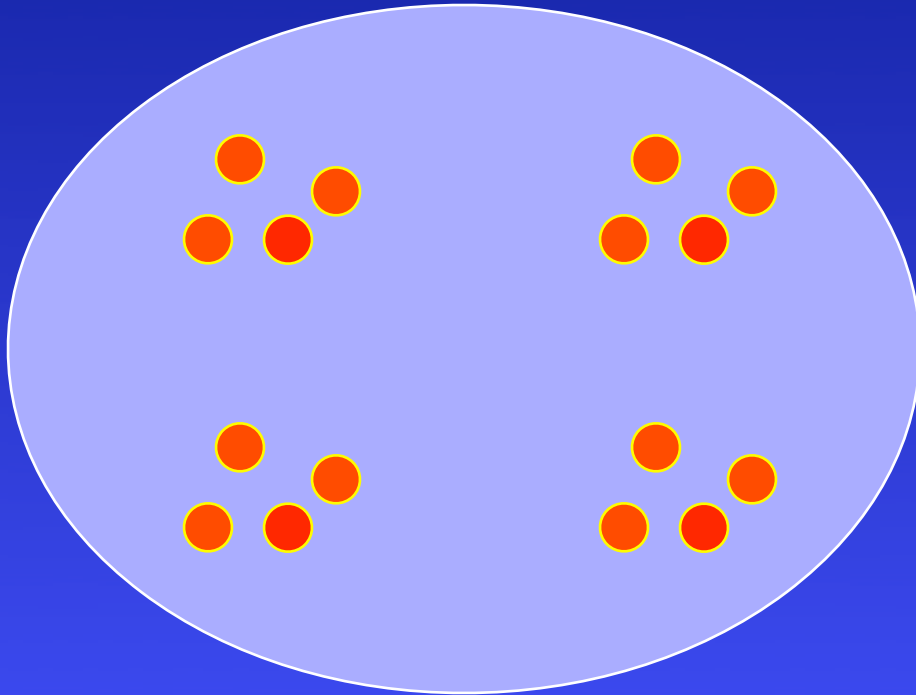
RMSMVVSGLTPEEFMLVYKFARKHHITLTNLITEETTHVVMKTDAEFVCERTLK^YFLGIAGGKWVVS^YFWVTQSIKERK
MLNEHDFEVRGDVVNGRNHQGPKRARESQDRKIFRGLEICCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTLGTGVHP
IVVVQPDAWTE^DNGFHAIGQMCEAPVVTREW^LDSVALYQCQELDTYLIPQIP

genomes...

Structural Genomics

Sali. *Nat. Struct. Biol.* **5**, 1029, 1998.
Sali *et al.* *Nat. Struct. Biol.*, **7**, 986, 2000.
Sali. *Nat. Struct. Biol.* **7**, 484, 2001.
Baker & Sali. *Science* **294**, 93, 2001.

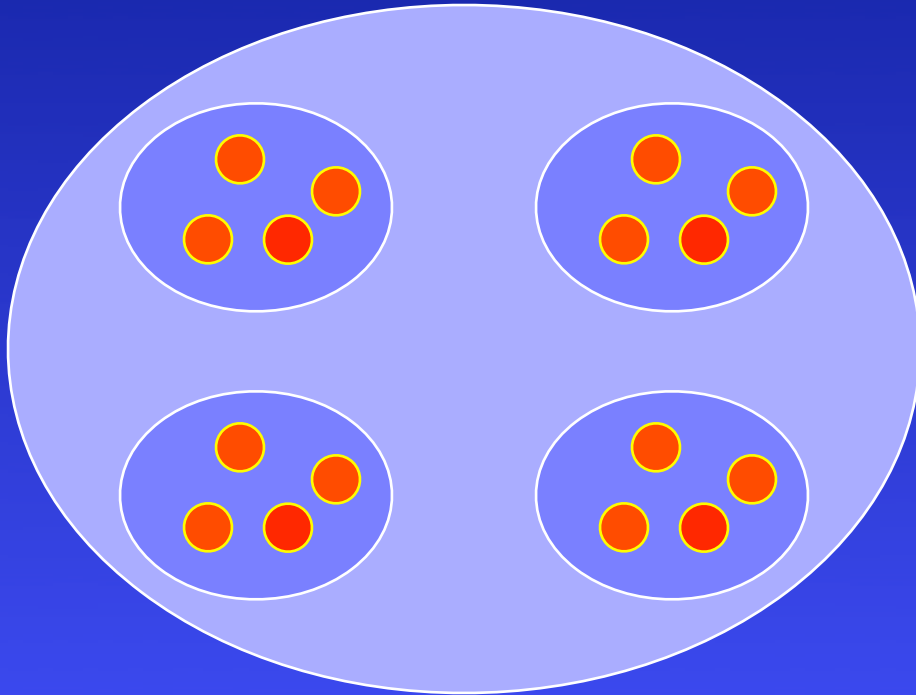
Characterize most protein sequences based on related known structures.



Structural Genomics

Sali. *Nat. Struct. Biol.* **5**, 1029, 1998.
Sali *et al.* *Nat. Struct. Biol.*, **7**, 986, 2000.
Sali. *Nat. Struct. Biol.* **7**, 484, 2001.
Baker & Sali. *Science* **294**, 93, 2001.

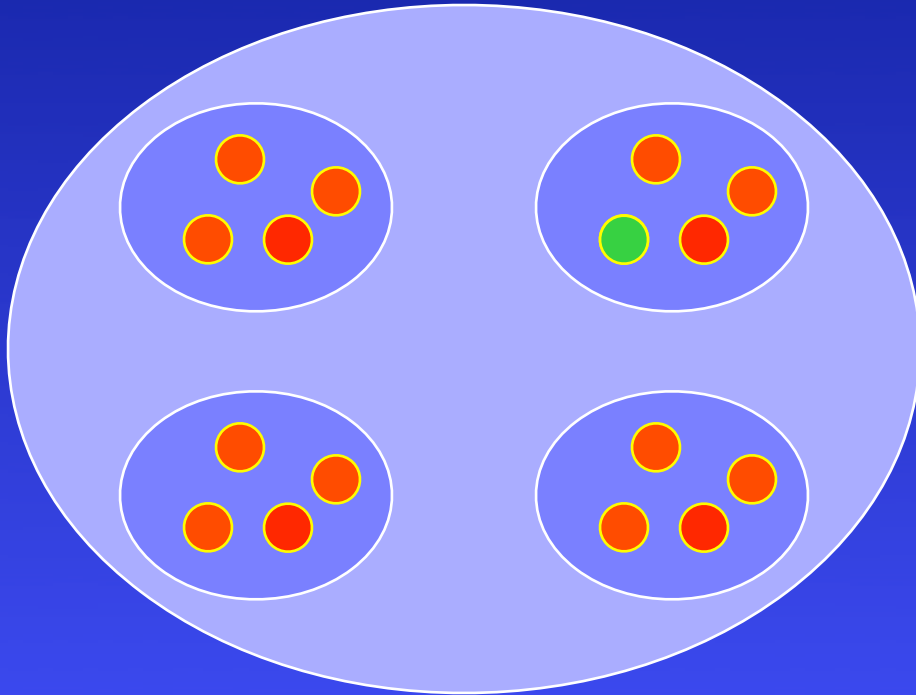
Characterize most protein sequences based on related known structures.



Structural Genomics

Sali. *Nat. Struct. Biol.* **5**, 1029, 1998.
Sali *et al.* *Nat. Struct. Biol.*, **7**, 986, 2000.
Sali. *Nat. Struct. Biol.* **7**, 484, 2001.
Baker & Sali. *Science* **294**, 93, 2001.

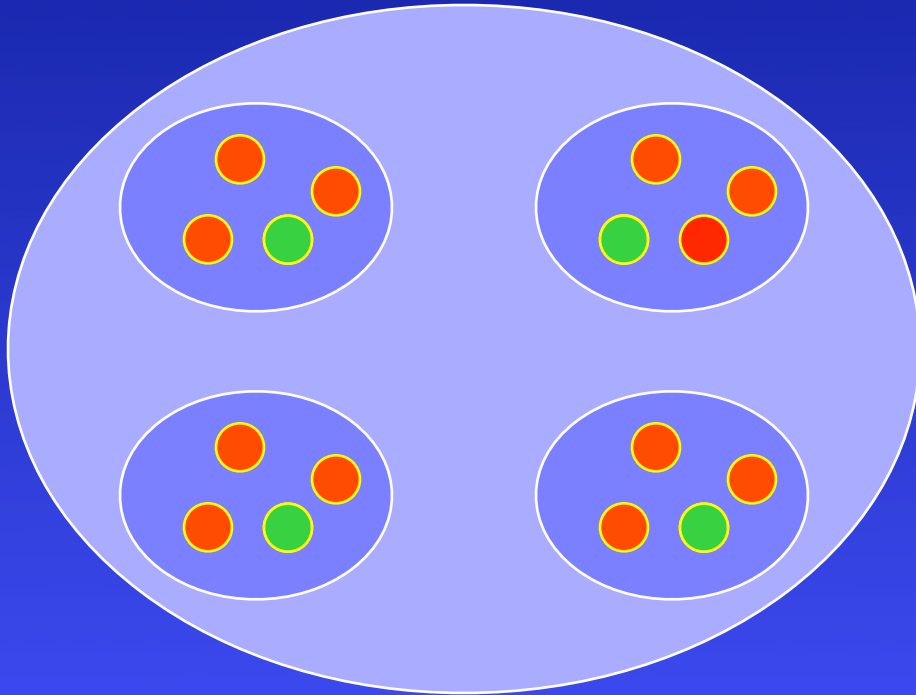
Characterize most protein **sequences** based on related known **structures**.



Structural Genomics

Sali. *Nat. Struct. Biol.* **5**, 1029, 1998.
Sali *et al.* *Nat. Struct. Biol.*, **7**, 986, 2000.
Sali. *Nat. Struct. Biol.* **7**, 484, 2001.
Baker & Sali. *Science* **294**, 93, 2001.

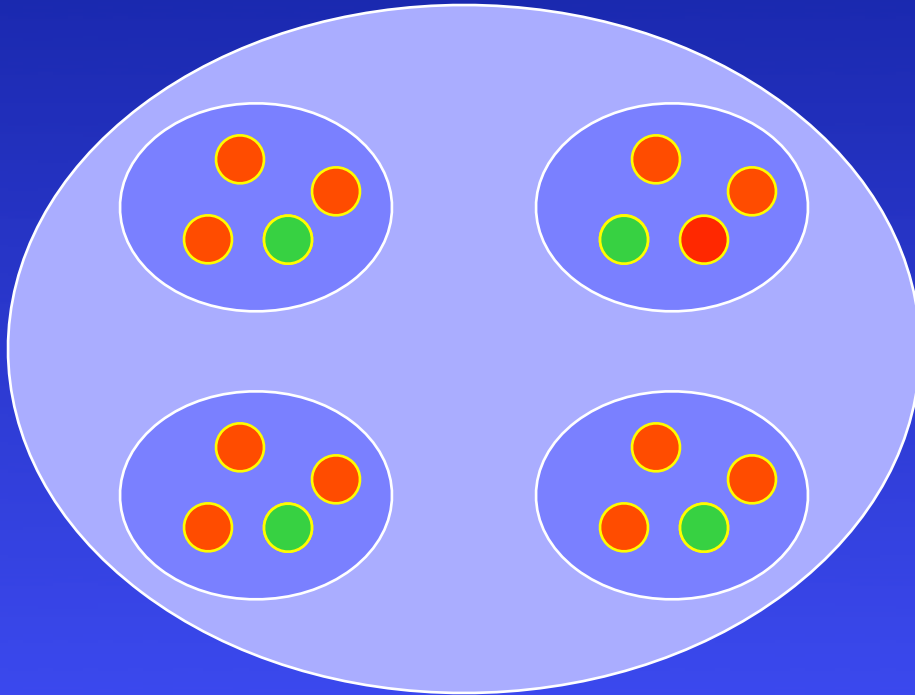
Characterize most protein **sequences** based on related known **structures**.



Structural Genomics

Sali. *Nat. Struct. Biol.* **5**, 1029, 1998.
Sali *et al.* *Nat. Struct. Biol.*, **7**, 986, 2000.
Sali. *Nat. Struct. Biol.* **7**, 484, 2001.
Baker & Sali. *Science* **294**, 93, 2001.

Characterize most protein **sequences** based on related known **structures**.



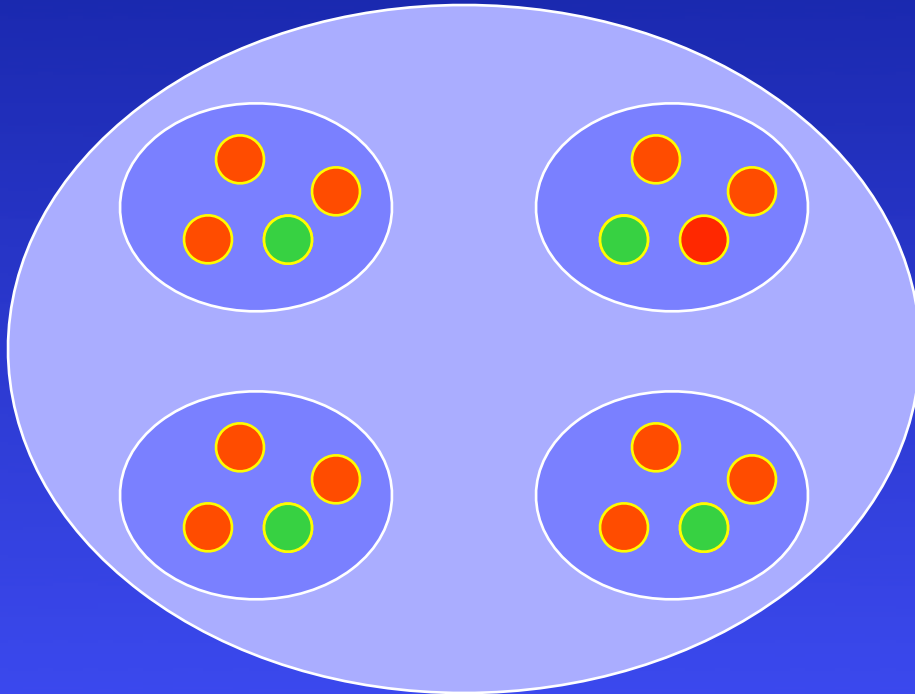
The number of “families” is much smaller than the number of proteins.

Any one of the members of a family is fine.

Structural Genomics

Sali. *Nat. Struct. Biol.* **5**, 1029, 1998.
Sali *et al.* *Nat. Struct. Biol.*, **7**, 986, 2000.
Sali. *Nat. Struct. Biol.* **7**, 484, 2001.
Baker & Sali. *Science* **294**, 93, 2001.

Characterize most protein **sequences** based on related known **structures**.



The number of “families” is much smaller than the number of proteins.

Any one of the members of a family is fine.

There are **~16,000** 30% seq id families (90%)
(Vitkup *et al.* *Nat. Struct. Biol.* **8**, 559, 2001)

Modeling with NY-SGRC structures

<http://salilab.org/modweb/>





ModWeb

Server for Comparative Modeling of Genes and Genomes

Please choose input type:

Sequence

Structure

Enter

June 2001

Protein Name	GI or Swissprot Code	Length	# Acceptable Models	Min. Seq. ID	Max. Seq. ID	# Models >50% Seq. ID	# Models 30-50% Seq. ID	# Models <30% Seq. ID
Yeast hypothetical protein	P38197	230	46	21	45	0	35	11
PNP oxidase	P38075	205	43	18	58	1	33	9
Yeast hypothetical protein	P49954	271	234	15	54	4	40	190
T. maritima L-threonine acetaldehyde-lyase	GI 4982322	342	2098	9	51	2	27	2069
Hypothetical esterase	P40363	288	157	9	48	0	13	144
Hypothetical protein	P40165	214	35	19	86	1	6	28
mevalonate diphosphate decarboxylase	GI 1292890	391	121	10	68	2	18	101
yeast glutathione synthetase	GI 2198534	419	26	30	42	0	26	0

Comparative modeling of the TrEMBL database

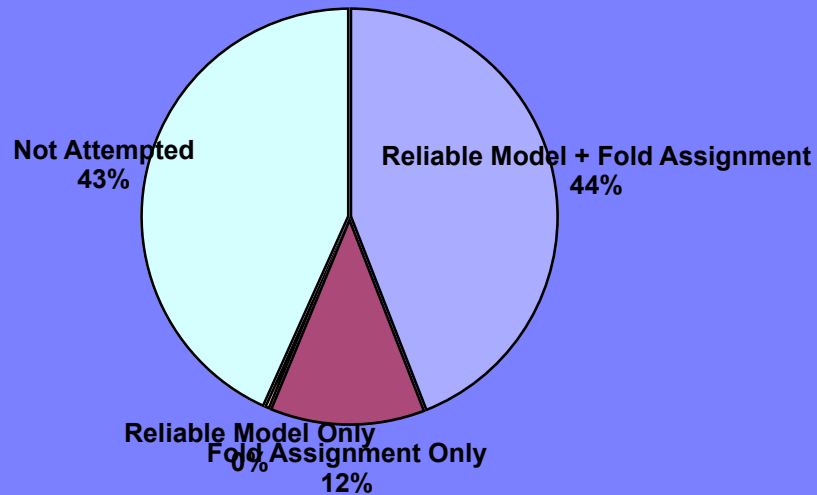
Unique sequences processed: 733,239

Sequences with fold assignments or models: 415,937 (57%)

70% of models based on <30% sequence identity to template.

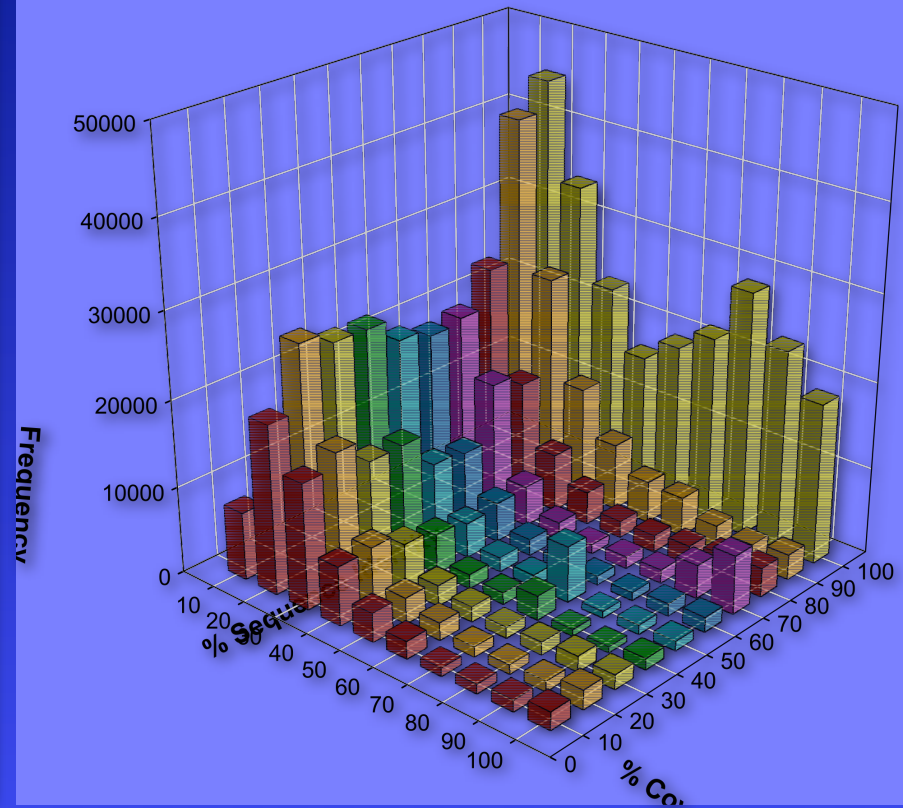
On average, only a domain per protein is modeled
(an “average” protein has 2.5 domains of 175 aa).

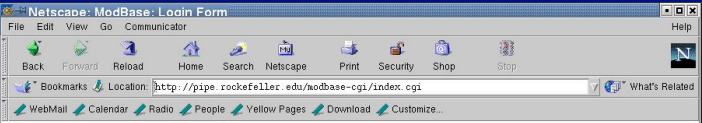
Modeling Coverage of the Sequence Space



Fold assignment: **PSI-BLAST E-value $\leq 10^{-4}$**

Reliable Model: **Model Score ≥ 0.7**







Database of Comparative Protein Structure Models

Welcome to MODBASE, a database of three-dimensional protein models calculated by [comparative modeling](#).

About MODBASE

- [General Information](#)
- [Glossary](#)
- [Authors and acknowledgements](#)
- [Publications](#)
- [Related resources](#)

Users of ModBase are requested to cite this article in their publications:
[MODBASE: a database of annotated comparative protein structure models](#),
Ulrich Pieper, Narayan Eswar, Andrew C. Stuart, Martin A. Björn, Anders Sal, *Nucl. Acids Res.* **30**, 255-259, 2002.

MODBASE is maintained by [Ulrich Pieper](#) in the group of [Anders Sal](#),
Laboratory of Molecular Biophysics, Fels Family Center for Biotechnology
and Structural Biology, The Rockefeller University, 1230 York Ave., New
York, NY 10021. Please address all inquiries to
modbase@rockefeller.edu.

© The Rockefeller University

MODBASE Contents

837,698 [Reliable Models](#) or [PSI-BLAST Fold Assignments](#)
for domains in 415,937 proteins. Last Update on 04/3/02.
[MODBASE statistics](#)

Search for Models

Enter SwissProt/TrEMBL/GenBank/PDB identifier or
descriptor.

[Advanced Search](#)

Login

[Academic login](#) [User login](#) [Logout](#)

Current login: modbase

Some datasets are accessible freely without a login (ie,
the "public" model set). Some datasets are available to
academic users only (ie, our "SP/TR" model set). And
some datasets require a specific username and password.
For commercial access to the models, please contact
[Structural Genomic Inc.](#)

Notes

MODBASE contains theoretically calculated models, not
experimentally determined structures. The models may
contain significant errors.

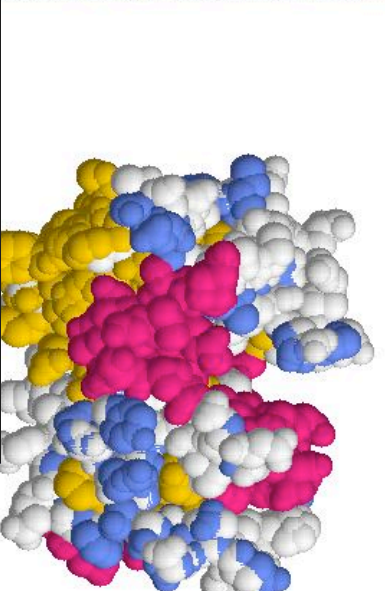


<http://salilab.org/modbase>

Pieper et al., Nucl. Acids Res. 2002.

8/9/02

kvMol View of Requested Model





Search Summary

SUMMARY		Search Criteria	
Keywords	dhfr		
Category	-		
Properties	(% Seq. Ident. and Model Size and Model Score)		
Ranges (min-max)	-30 - -		
Values			
Minimum	6.00	82	0.01
Average	25	181	0.79
Maximum	30.00	492	1.00

K+ Sequence Alignment Window

1vdra...2-157: ELVSAALANRVIRGIDDELPPSPIDAKRRTITSTIADDPVFLCKITFEIRDDLPSSARIVMSRSERSVUTTHAAASLEAAVDAIASDAETAVVIG
Model : 2-156: KYSLIAAXKGVICQDTPSAKQKOLLFLATVYQULLVQKRTESMQLPMPKRYAVVIRSGWISNDNVVVFQSEFAMDLATFTHVIVSGL

Similarity::1

Signed by: Unsigned classes from local hard disk

100%

NetScape: ModBase-Search Form

Database of Comparative Protein Structure Models

User: Academic User [Change User](#)

DATASET SELECTION

Datasets: SP/TR-2002, SP/TR-2001, nysrgc_tm1, nysrgc_tm2, nysrgc_v1a, nysrgc_tm4

SEARCH BY PROPERTIES

All

Organism: ALL or

Sort matching models by: Sequence identity

MODEL SEARCH BY PROPERTY RANGES

(E-value (0 - 100) and Model Size and Model Score (0.0-1.0))

lower limit upper limit lower limit upper limit lower limit upper limit

SEARCH BY SEQUENCE SIMILARITY

Protein Sequence:

View sequences that match the search criteria but could not be modeled.

29 matches were found using the specified search criteria. Click on the links in the table header to resort your output.

TARGET						MODEL DATA					TEMPLATE			
Model/Fold Reliability	Sequence based View	Select Sequence Database Links	Database Description	Organism	Protein Size	Modeled Segment	Size	Seq Id (%)	E-value	Model Score	PDB code	Template based View	Segment	Description
		# 68 027452	DHYDROFOLATE REDUCTASE TYPE III (EC 1.5.1.3) (DHFR TYPE III) Subset: SPFR-2001 ZFAM PRODOM	<i>Escherichia coli</i> <i>Shigella sonnei</i>	169	1-165	165	30.00	4e-32	1.00	1dfr &		1-158	DHYDROFOLATE REDUCTASE
		# 68 827282	BIFUNCTIONAL DHYDROFOLATE REDUCTASE-THYMIDYLATE SYNTHASE (DHFR-156 INCLUDES DHYDROFOLATE REDUCTASE (EC 1.5.1.3) Subset: SPFR-2001 ZFAM PRODOM	<i>Leishmania major</i>	520	24-231	208	30.00	2e-49	1.00	1dfr		1-186	DHYDROFOLATE REDUCTASE (EC 1.5.1.3) COMPLEX WITH FOLATE IDRF
		# 68 287715	BIFUNCTIONAL DHYDROFOLATE REDUCTASE-THYMIDYLATE SYNTHASE (DHFR-156 INCLUDES DHYDROFOLATE REDUCTASE (EC 1.5.1.3) Subset: SPFR-2001 ZFAM PRODOM	<i>Plasmodium chabaudi</i>	583	21-241	221	30.00	3e-39	1.00	1dfr		2-193	IDFR 5 DHYDROFOLATE REDUCTASE
		# 68 026264	BIFUNCTIONAL DHYDROFOLATE REDUCTASE-THYMIDYLATE SYNTHASE (DHFR-156 INCLUDES DHYDROFOLATE REDUCTASE (EC 1.5.1.3) Subset: SPFR-2001 ZFAM PRODOM	<i>Plasmodium vivax</i>	623	35-237	203	30.00	1e-37	1.00	1dfr		14-203	IDFR 5 DHYDROFOLATE REDUCTASE
Sequence based View		Database Description			Organism			Protein Size	Modeled Segments - Schema					
		# 68 027452	DHYDROFOLATE REDUCTASE TYPE III (EC 1.5.1.3) (DHFR TYPE III) Subset: ZFAM PRODOM			<i>Escherichia coli</i> <i>Shigella sonnei</i>			169					
		# TR Q9U0M0	DHFRP2 PROTEIN (FRAGMENT) Subset: ZFAM PRODOM			<i>Homo sapiens</i>			121					

Database Synonyms for this Sequence (100% Sequence Identity)			
TrEMBL	Q93V12	Salmonella typhimurium	Dihydrofolate reductase
GI	1333722	Plasmod pLMQ29	dhfr1 product (AA 1 - 157)
GI	6870050	Salmonella typhimurium	dihydrofolate reductase
GI	92755	Escherichia coli plasmid pLMQ29	S11706 dihydrofolate reductase (EC 1.5.1.3) type I - Escherichia coli plasmid pLMQ29

Model Details

PSI-BLAST Fold Assignments (left half) and [Reliable Models](#) (right half) are indicated in green.

* Indicates an E-value from an unfiltered PSI-BLAST search when a filtered search does not result in a significant match.

This Table displays all Models/Folds of this sequence
Display all reliable Models and Fold Assignments of this sequence
using the correct search parameters.

[Full View Download Metadata](#)

MODEL DATA					LINKS				
Fold/Model Reliability	Size	Seq id (%)	E-value	Model Score	Ras mol	3D Coord	PIR	PAP	Mod View

(157 Residues)

156

156 (2-158) DHYDROFOLATE REDUCTASE - CATH 3.40.430.10.1.1 (99%) Subset: SP/TR-2001

156

156 (1-156) DHYDROFOLATE REDUCTASE - CATH 3.40.430.10.5.1.2 (99%) Subset: SP/TR-2001

100%

Conclusions

- ✓ At present, useful 3D models can be obtained for domains in ~ 55% of the proteins (25% of domains).
- ✓ Sampling at >30% sequence identity level.
- ✓ Completeness in structural coverage.
- ✓ Application to biological problems.

Acknowledgments



Andrej Sali

Frank Alber

Fred Davis

Damien Devos

Narayanan Eswar

Rachel Karchin

Libusha Kelly

Michael F. Kim

Dmitry Korkin

M. S. Madhusudhan

Nebosja Mirkovic

Ursula Pieper

Andrea Rossi

Min-yi Shen

Maya Topf

<http://www.salilab.org>