

Comparative protein structure modeling of genes, genomes and complexes

Marc A. Marti-Renom

Department of Biopharmaceutical Sciences
University of California, San Francisco

Comparative protein structure modeling of genes, genomes and complexes

**Comparative protein structure modeling of genes,
genomes and complexes**

Modelat de gens, genomes i complexos

**Comparative protein structure modeling of genes,
genomes and complexes**

Modelat de gens, genomes i complexos

**Modelado estructural de genes, genomas y complejos.
Aplicaciones biomédicas y biotecnológicas.**

**Ya po', modelando ene estructuras de
genes, genomas y complejos... cachai?**

Marc A. Marti-Renom

Department of Biopharmaceutical Sciences
University of California, San Francisco

Modelado por homología...

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

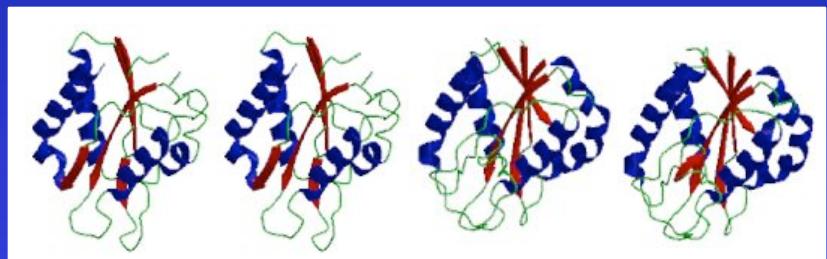
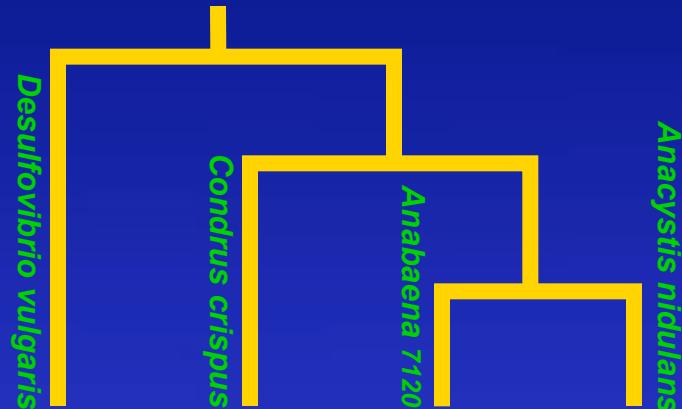
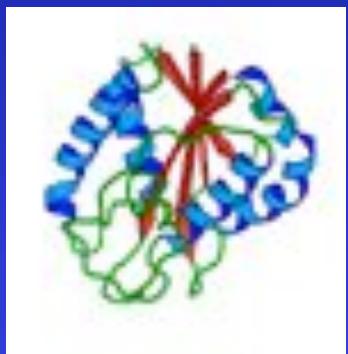
Theory



Experiment

Principles of Protein Structure

GFCHIKAYTRLIMV...



Folding

Ab initio prediction

Evolution

Threading
Comparative Modeling

Comparative Modeling by Satisfaction of Spatial Restraints (MODELLER)

3D GKITFYERGFQGHCYESDC-NLQP...
SEQ GKITFYERG---RCYESDCPNLQP...

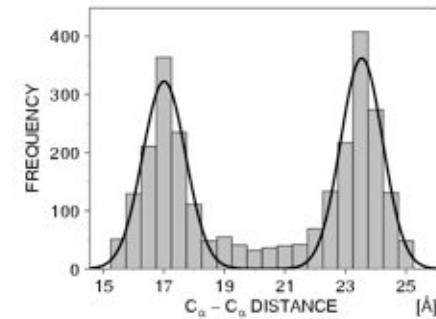
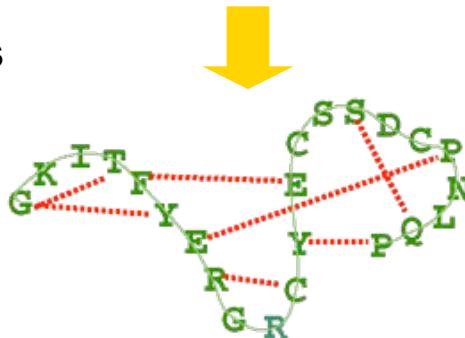
A. Šali & T. Blundell. *J. Mol. Biol.* **234**, 779, 1993.
J.P. Overington & A. Šali. *Prot. Sci.* **3**, 1582, 1994.
A. Fiser, R. Do & A. Šali. *Prot Sci.* **9**, 1753, 2000.

<http://salilab.org/modeller>

Comparative Modeling by Satisfaction of Spatial Restraints (MODELLER)

3D GKTFYERGFQGHHCYESDC-NLQP...
SEQ GKTFYERG---RCYESDCPNLQP...

1. Extract spatial restraints



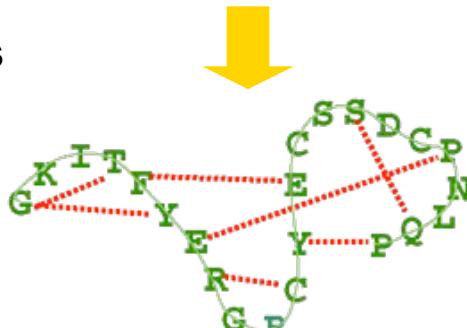
- A. Šali & T. Blundell. *J. Mol. Biol.* **234**, 779, 1993.
J.P. Overington & A. Šali. *Prot. Sci.* **3**, 1582, 1994.
A. Fiser, R. Do & A. Šali. *Prot Sci.* **9**, 1753, 2000.

<http://salilab.org/modeller>

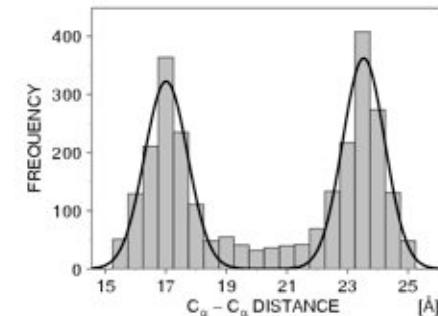
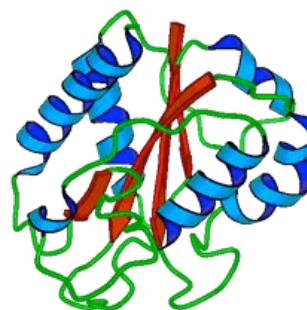
Comparative Modeling by Satisfaction of Spatial Restraints (MODELLER)

3D GKIFYERGFQGHHCYESDC-NLQP...
SEQ GKIFYERG---RCYESDCPNLQP...

1. Extract spatial restraints



2. Satisfy spatial restraints



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

- 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>

Steps in Comparative Protein Structure Modeling

START

TARGET

ASILPKRLFGNCEQTSDEGLK
IERTPLVPHISAQNVLKIDD
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



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

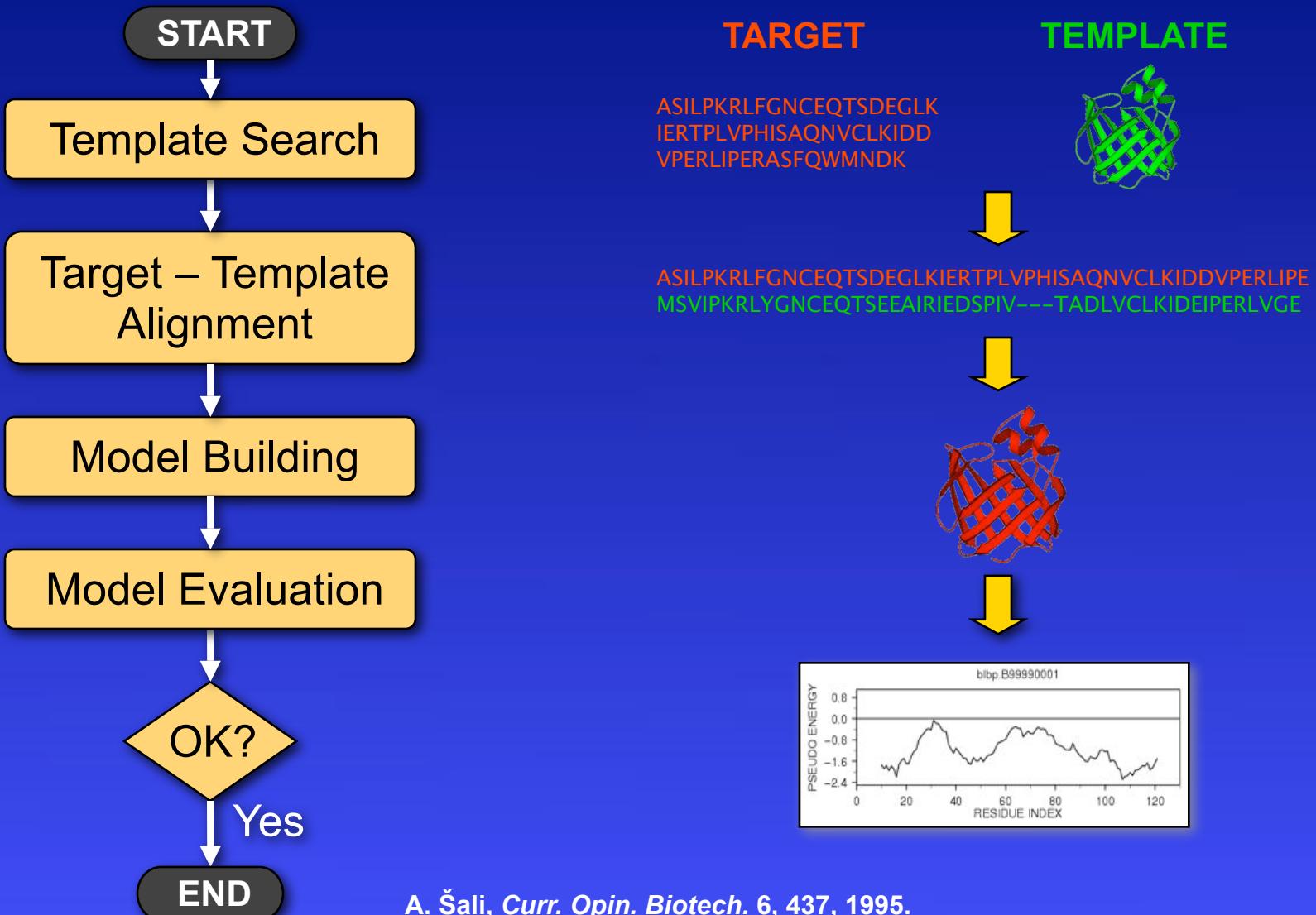


A. Šali, *Curr. Opin. Biotech.* 6, 437, 1995.

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

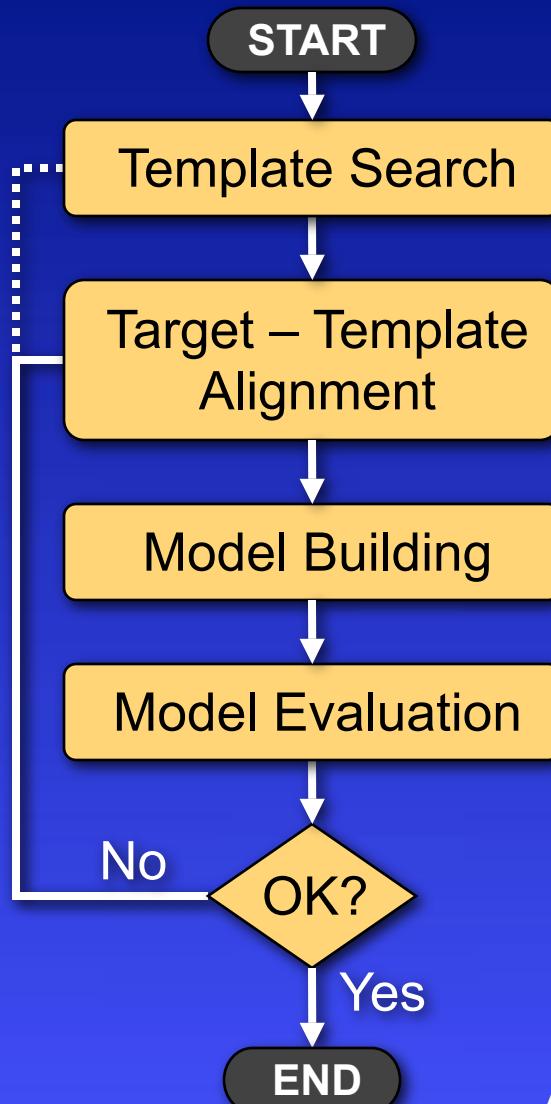


A. Šali, *Curr. Opin. Biotech.* 6, 437, 1995.

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

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

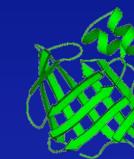
Steps in Comparative Protein Structure Modeling



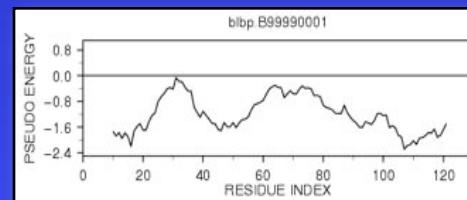
TARGET

ASILPKRLFGNCEQTSDEGLK
IERTPLVPHISAQNVLKIDD
VPERLIPERASFQWMNDK

TEMPLATE



ASILPKRLFGNCEQTSDEGLK**IERTPLVPHISAQNVLKIDDVPERLIP**
MSVIPKRLYGNCEQTSEEAIRIEDSPIV--TADLVVCLKIDEIPERLVGE

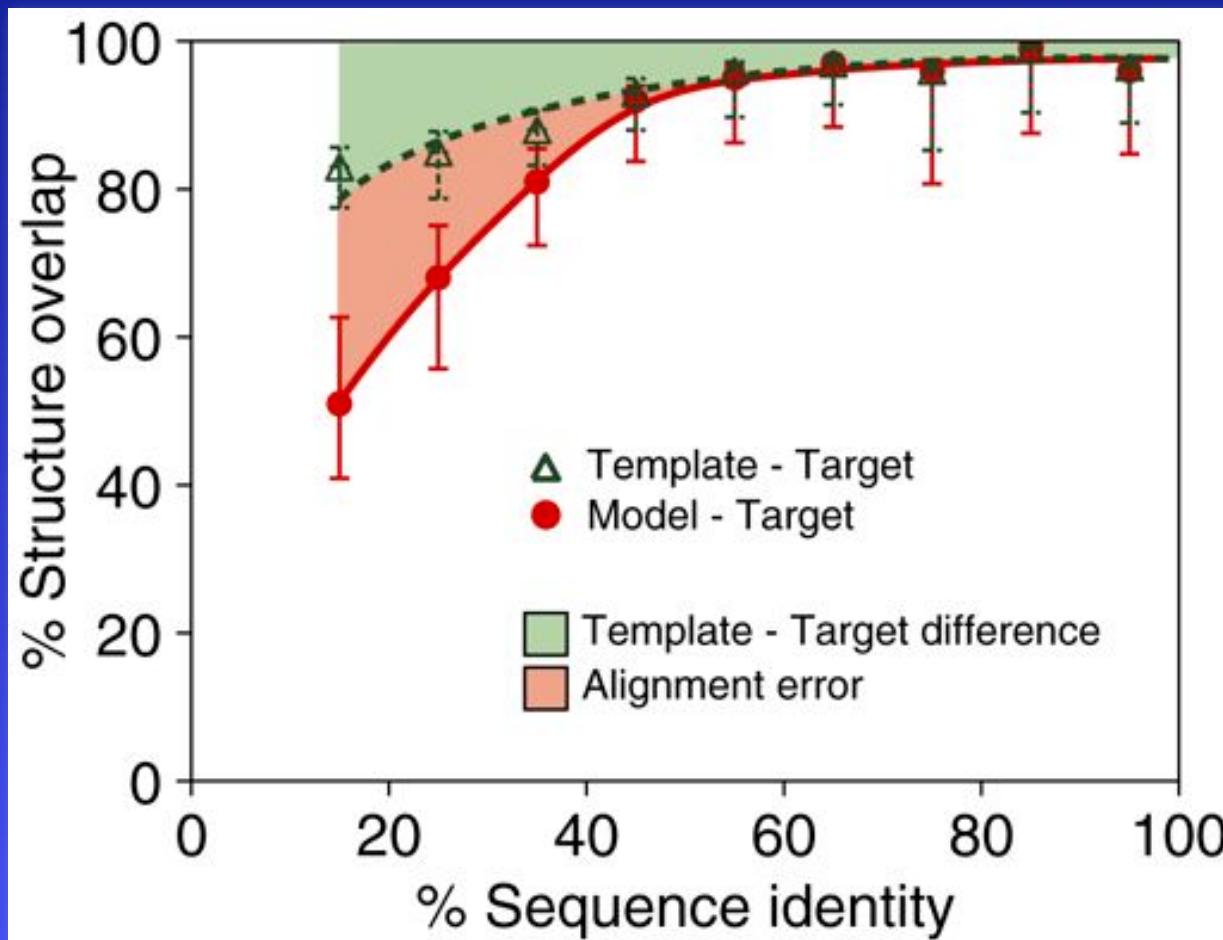


A. Šali, *Curr. Opin. Biotech.* 6, 437, 1995.

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

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

Model Accuracy as a Function of Target-Template Sequence Identity



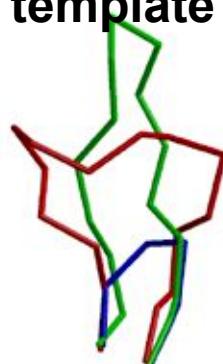
Typical Errors in Comparative Models

MODEL

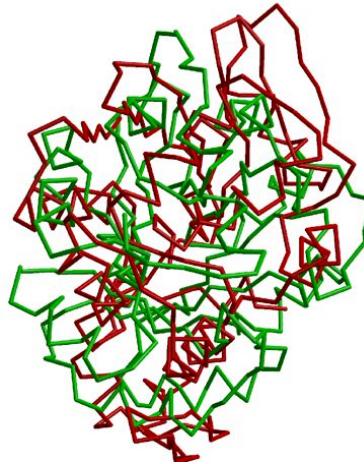
X-RAY

TEMPLATE

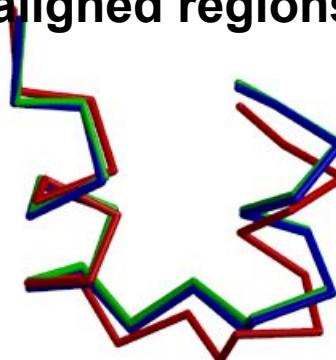
Region without a template



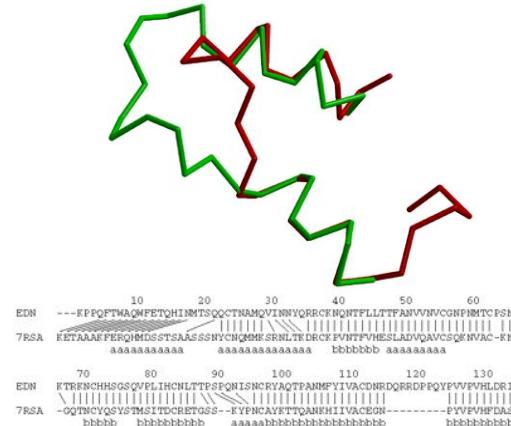
Incorrect template



Distortion in correctly aligned regions



Misalignment



Sidechain packing

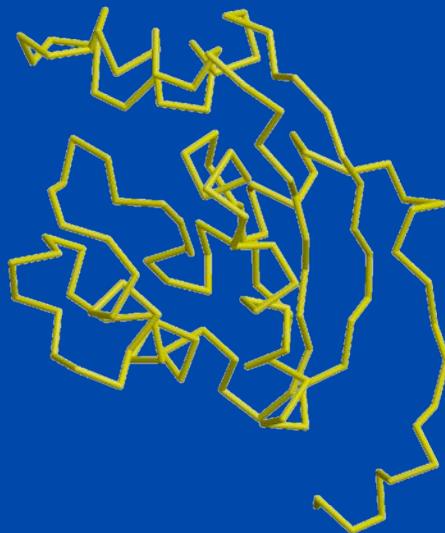


Model Accuracy

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

HIGH ACCURACY

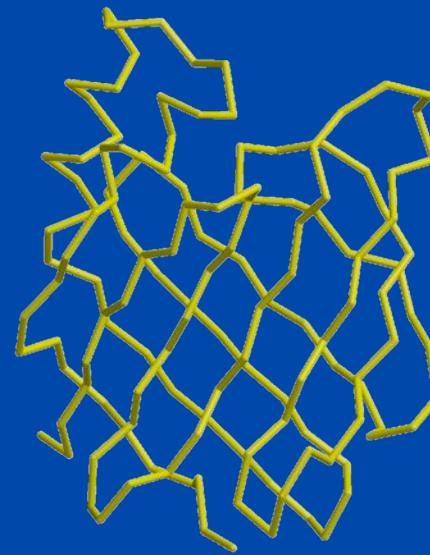
NM23
Seq id 77%



X-RAY

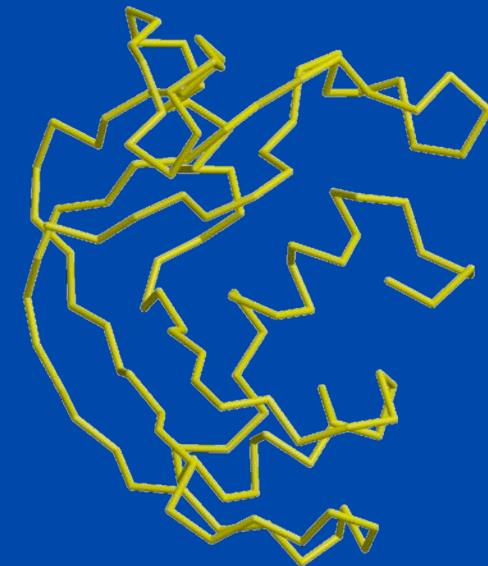
MEDIUM ACCURACY

CRABP
Seq id 41%



LOW ACCURACY

EDN
Seq id 33%

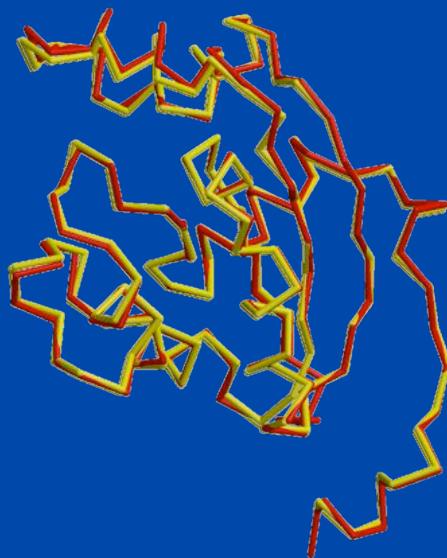


Model Accuracy

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

HIGH ACCURACY

NM23
Seq id 77%
 $C\alpha$ equiv 147/148
RMSD 0.41 Å

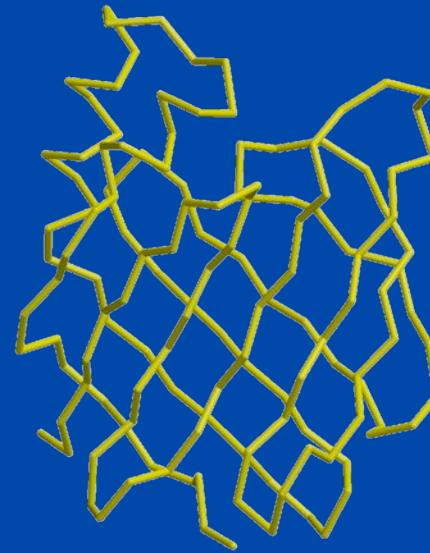


Sidechains
Core backbone
Loops

X-RAY / MODEL

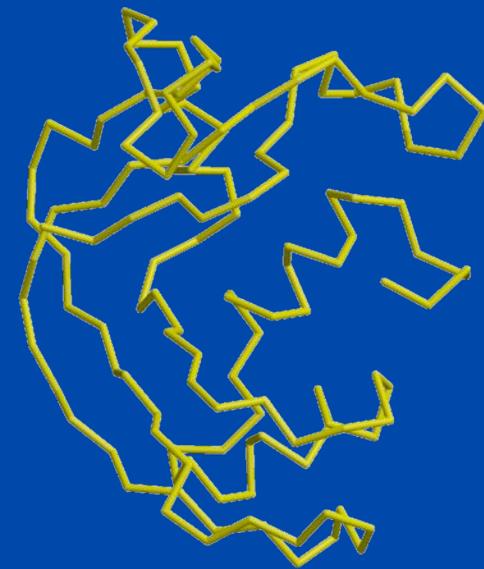
MEDIUM ACCURACY

CRABP
Seq id 41%



LOW ACCURACY

EDN
Seq id 33%

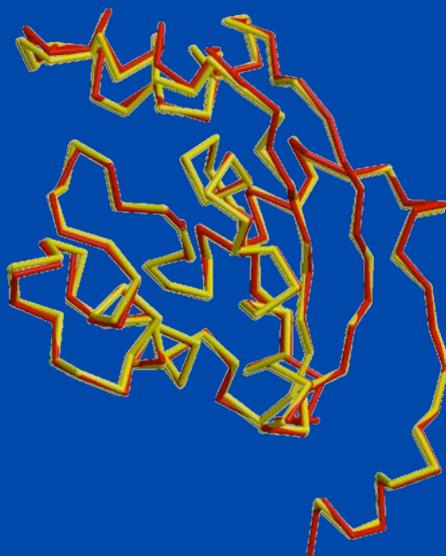


Model Accuracy

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

HIGH ACCURACY

NM23
Seq id 77%
 $C\alpha$ equiv 147/148
RMSD 0.41Å

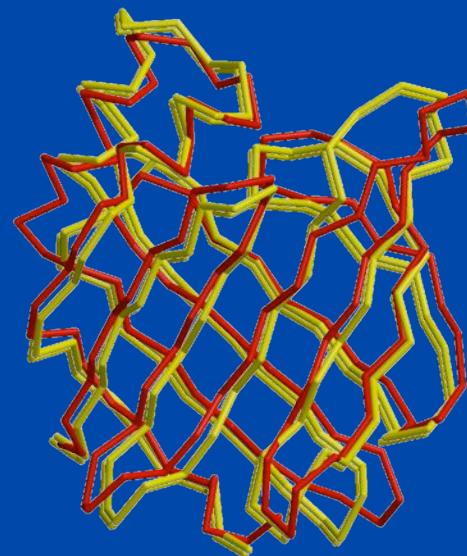


Sidechains
Core backbone
Loops

X-RAY / MODEL

MEDIUM ACCURACY

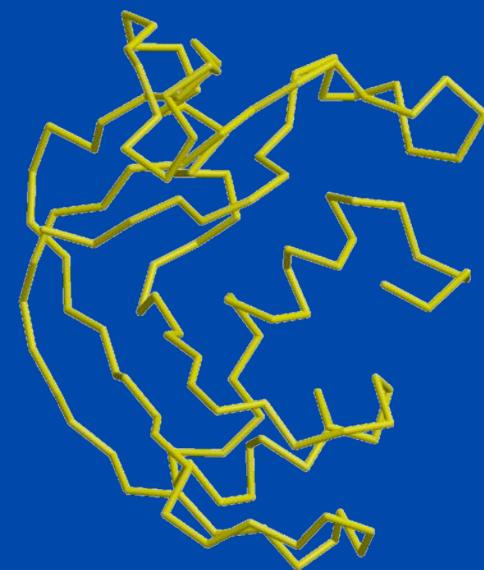
CRABP
Seq id 41%
 $C\alpha$ equiv 122/137
RMSD 1.34Å



Sidechains
Core backbone
Loops
Alignment

LOW ACCURACY

EDN
Seq id 33%

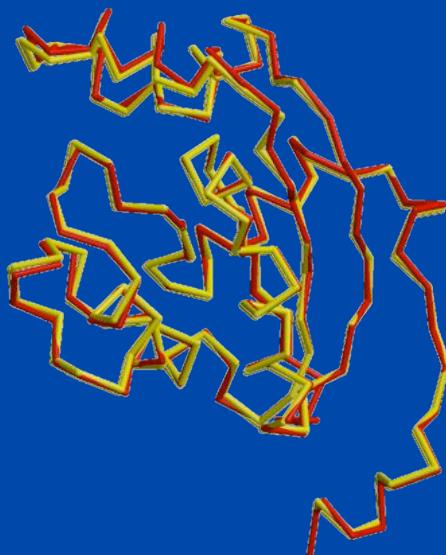


Model Accuracy

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

HIGH ACCURACY

NM23
Seq id 77%
 $C\alpha$ equiv 147/148
RMSD 0.41Å

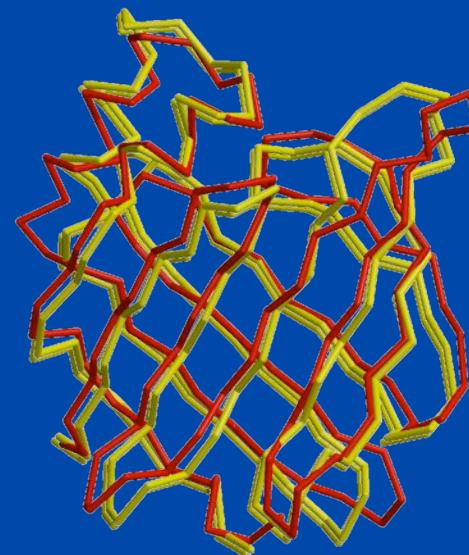


Sidechains
Core backbone
Loops

X-RAY / MODEL

MEDIUM ACCURACY

CRABP
Seq id 41%
 $C\alpha$ equiv 122/137
RMSD 1.34Å



Sidechains
Core backbone
Loops
Alignment

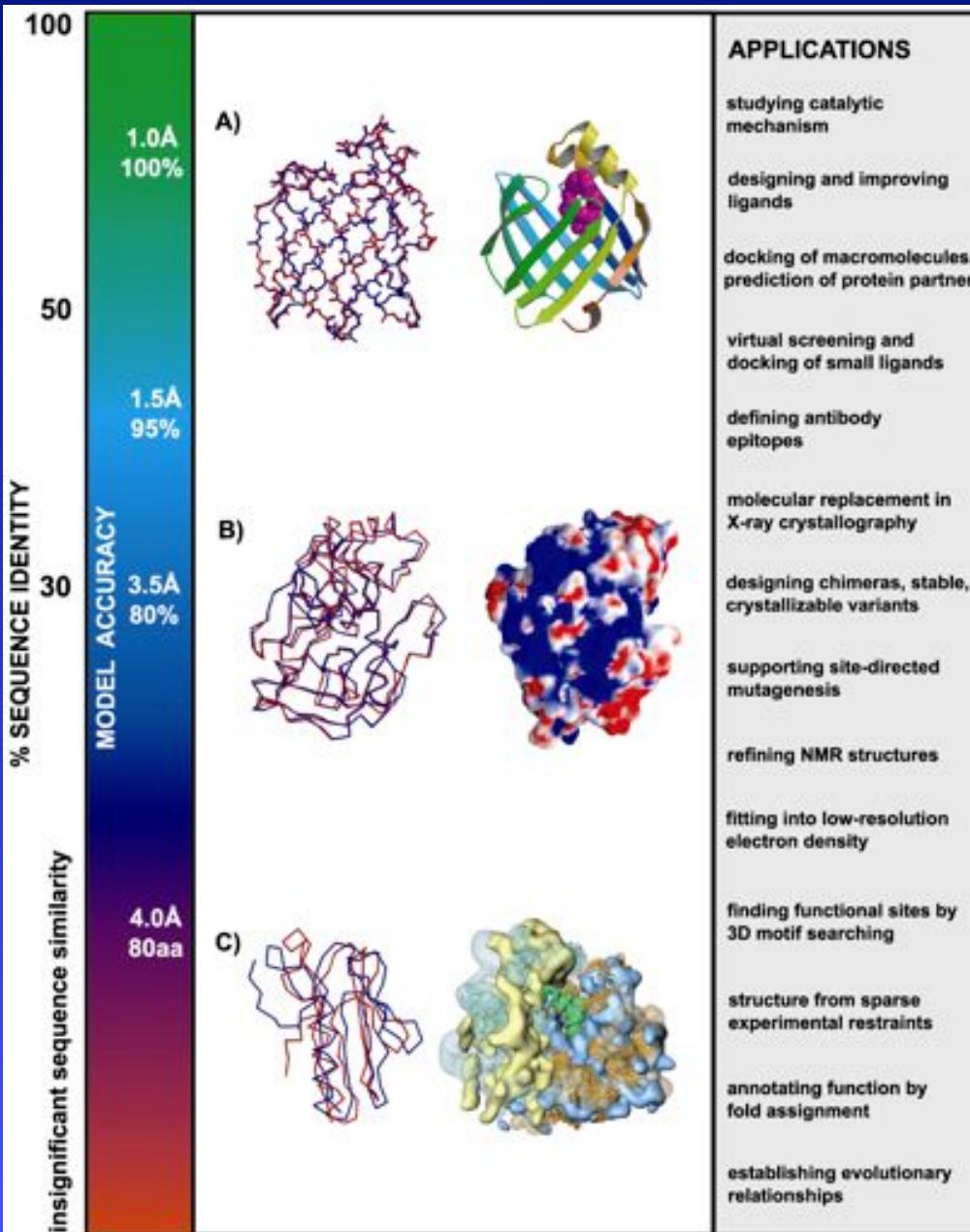
LOW ACCURACY

EDN
Seq id 33%
 $C\alpha$ 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.

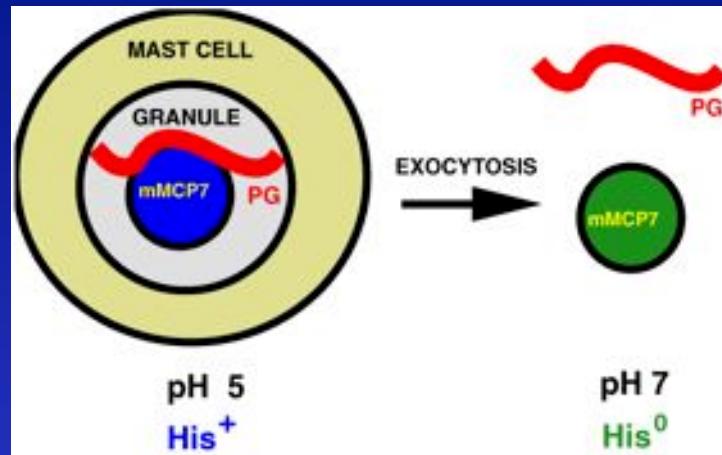
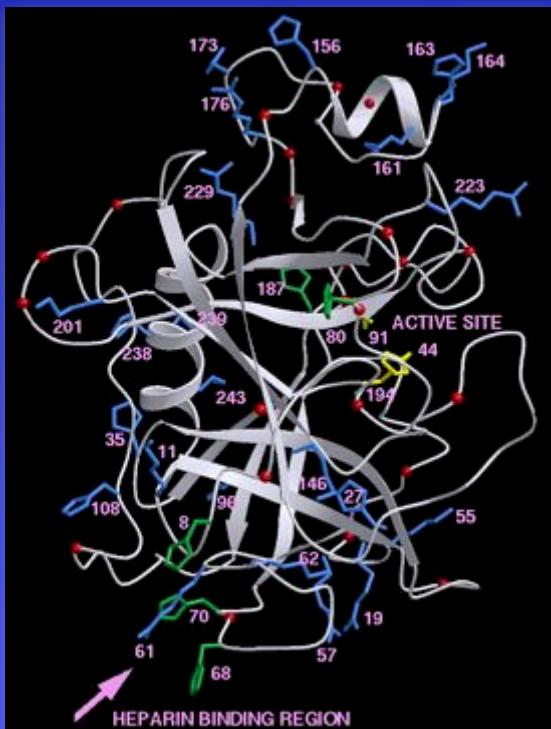
A. Sali & J. Kuriyan.
TIBS **22**, M20, 1999.

genes...

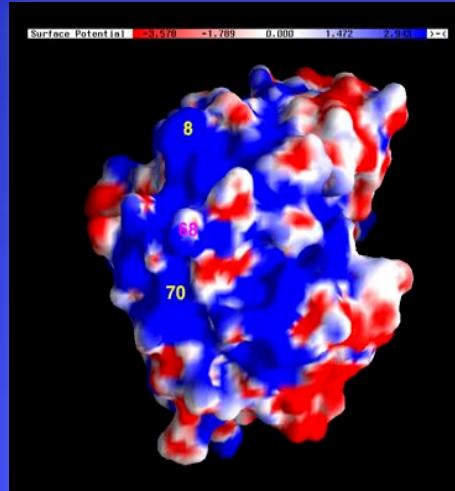
Do mast cell proteases bind proteoglycans? Where? When?

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

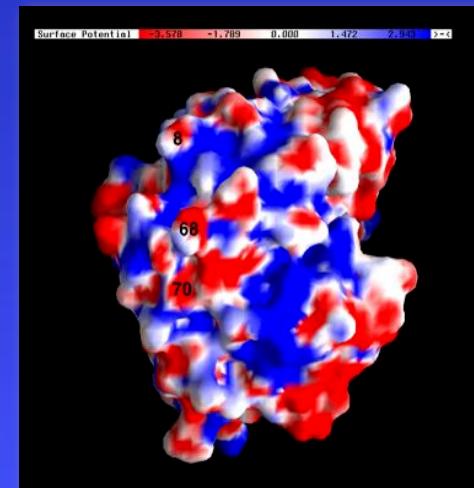
1. mMCPs bind negatively charged proteoglycans through electrostatic interactions?
 2. Comparative models used to find clusters of positively charged surface residues.
 3. Tested by site-directed mutagenesis.



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.



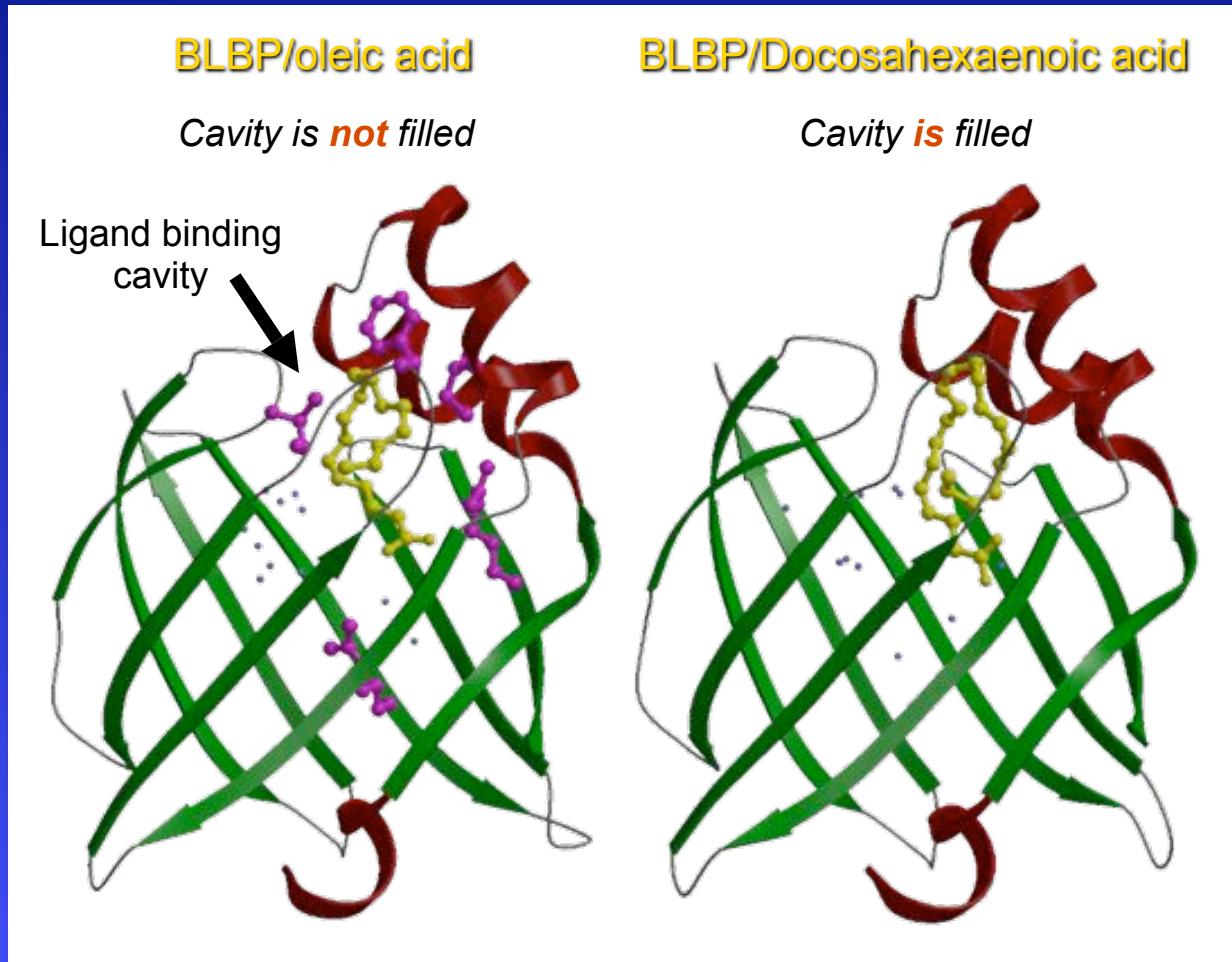
Native mMCP-7 at pH=5 (His⁺)



Native mMCP-7 at pH=7 (His⁰)

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

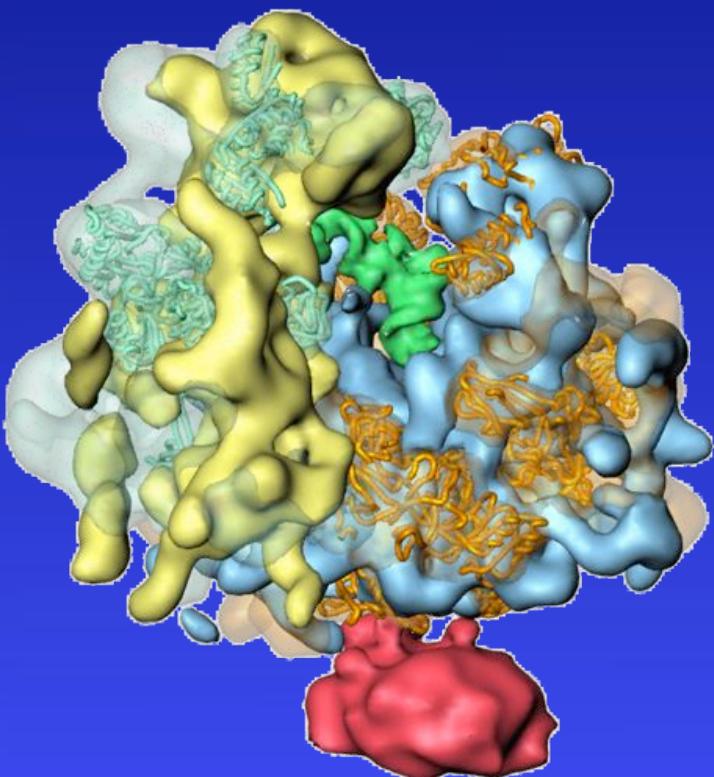
Predicting features of a model that are not present in the template



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

Some Models Can Be Used in Docking to Density Maps

(Yeast Ribosomal 40S subunit)

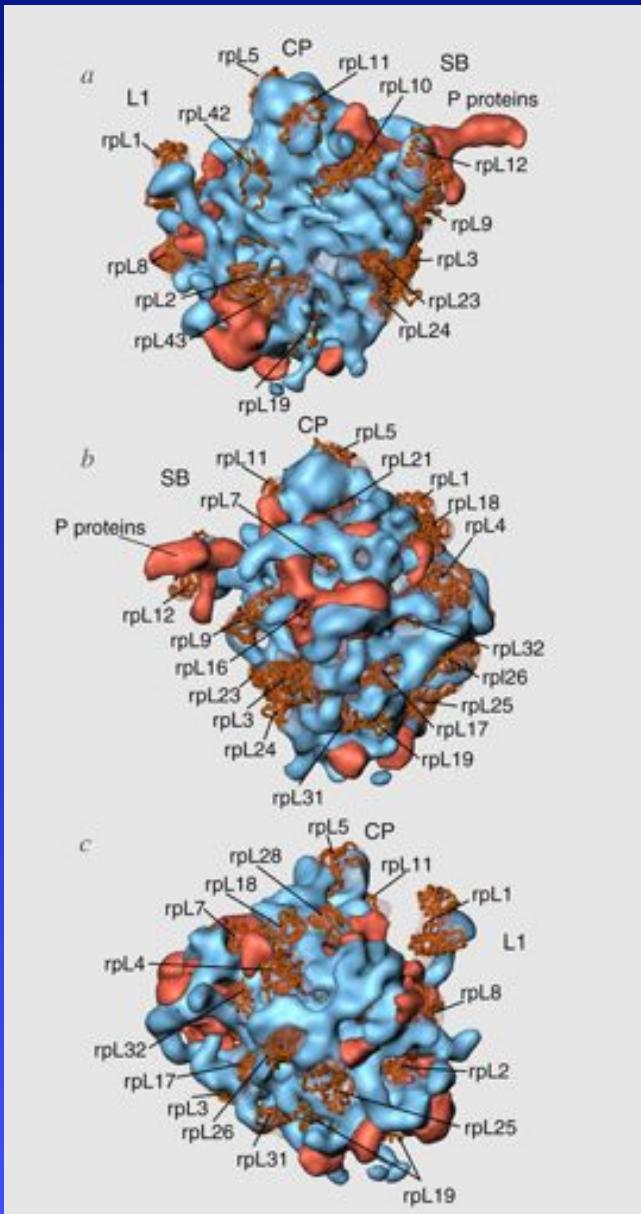


Small 30S subunit from *Thermus thermophilus*
Large 50S subunit from *Haloarcula marismortui*

Docking of comparative models into the cryo-EM map.

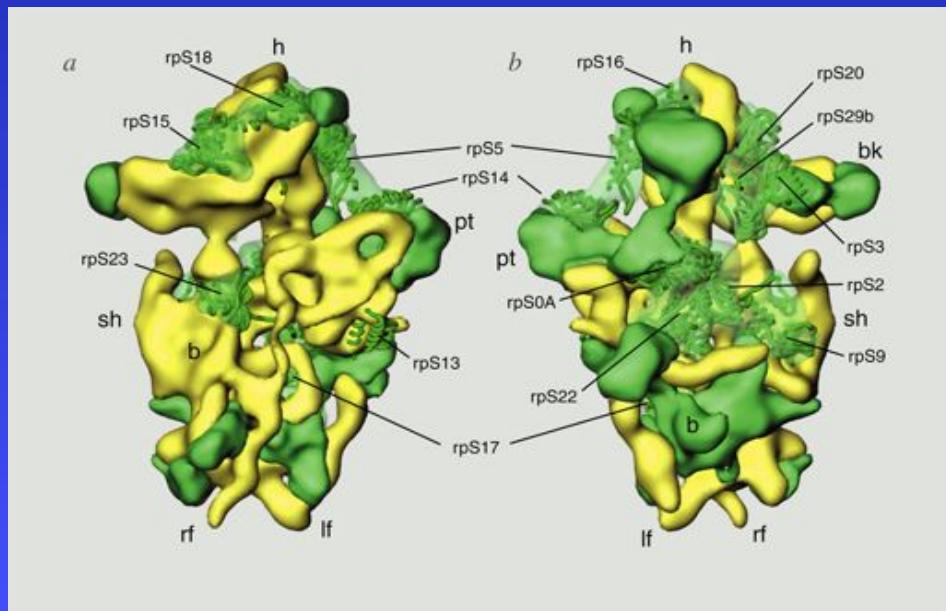
Spahn et al. 2001 Cell 107:373-386

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

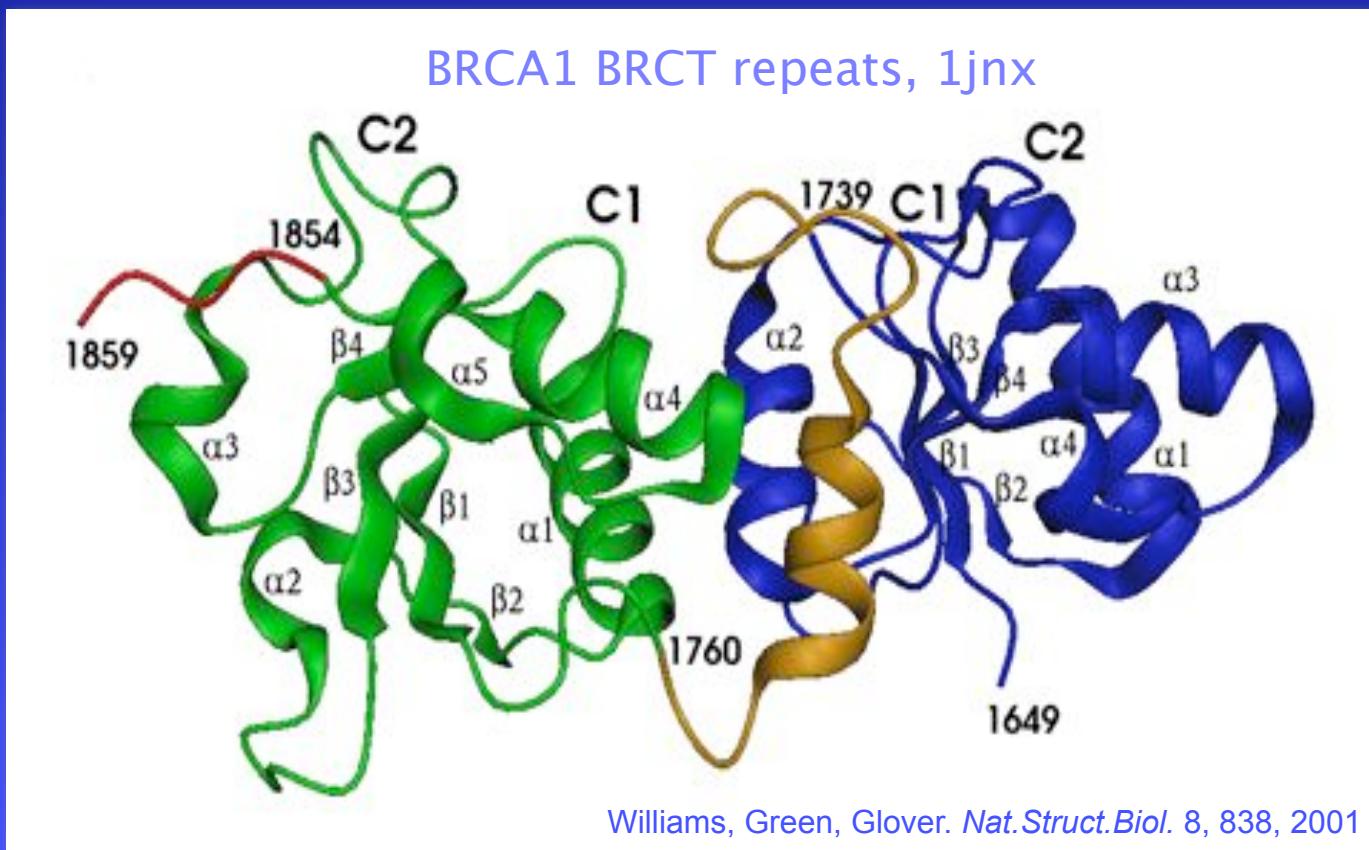
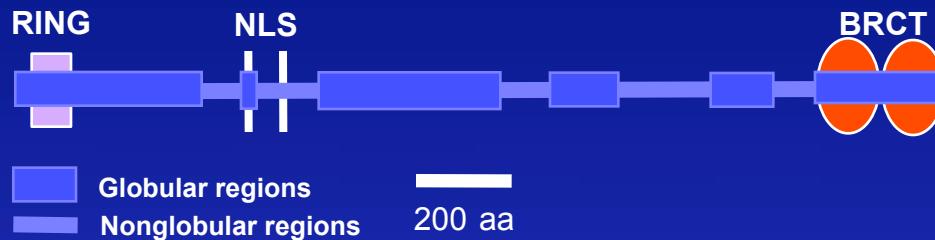


Structural analysis of missense mutations in human BRCA1 BRCT domains

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

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

Human BRCA1 and its two BRCT domains



~~CONFIDENTIAL~~

BRACAanalysis™

Comprehensive BRCA1-BRCA2 Gene Sequence Analysis Result

Strong Cancer Prevention Center
426 E 73rd St
New York, NY 10021

Specimen ID: 1000
Specimen Type: Blood
Draw Date: 10/08/2000
Assessment Date: Oct 10, 2000
Report Date: Nov 12, 2000

PATIENT
Name: Pauline H. 1000
Date of Birth: Feb 02, 1953
Patient ID: 1000
Gender: Female
Assessment #: 10000000
Report Date #: 1000H

Physician

Test Result

Gene Analyzed	Specific Genetic Variant
BRCA2	HQ116R
BRCA1	None Detected

Interpretation

GENETIC VARIANT OF UNCERTAIN SIGNIFICANCE

The BRCA2 variant HQ116R 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 Genetics Laboratories of this particular variant in an individual with a deleterious truncating mutation in BRCA2, however, reduces the likelihood that HQ116R is itself deleterious.

Authorized Signature:

Laboratory Director

Medical Director

This report 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 and physician that includes appropriate counseling. The accompanying technical specification fully describes the analytical method, participant characteristics, performance, and interpretive criteria used. This report may be considered consequential by some states. They can vary significantly among participants, particularly those not licensed by the National Committee for Clinical Laboratory Standard. Thus the basis discussed by the U.S. Food and Drug Administration. The results have been determined to be accurate or equivalent to that laboratory.

CONFIDENTIAL

BRCAanalysis™
Comprehensive BRCA1-BRCA2 Gene Sequence Analysis Result

Strong Cancer Prevention Center 426 E 73rd St New York, NY 10021	Specimen Type: Blood Draw Date: 10/08 Assessment Date: Oct 10, 2000 Report Date: Nov 12, 2000	PATIENT Name: Pauline H. 1983 Date of Birth: Feb 02, 1983 Patient ID: 12345678 Gender: Female Assessment #: 00000000 Reportation #: 000000
Physician	Test Result	
	Gene Analyzed BRCA2 BRCA1	Specific Genetic Variant H2116R 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

This report 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's physician who will interpret the findings in the context of the patient's clinical presentation. This report does not contain recommendations, interpretations, or conclusions of the physician. The physician may be consulted if interpretation is unclear. These reports were developed without participation of the American College of Medical Genetics. Thus the basic guidelines set by the U.S. Food and Drug Administration, The Ethics Committee of the American Society of Human Genetics, and the National Institutes of Health do not apply to these reports. The results have been determined to be accurate and appropriate to the laboratory.

Missense Mutations in BRCT Domains by Function

no transcription activation

transcription activation

?

cancer associated
not cancer associated

?

C1697R
R1699W
A1708E
S1715R
P1749R
M1775R

cancer associated
not cancer associated

M1652K
L1657P
E1660G
H1686Q
R1699Q
K1702E
Y1703H
F1704S

L1705PS
1715NS
1722FF
1734LG
1738EG
1743RA
1752PF
1761I

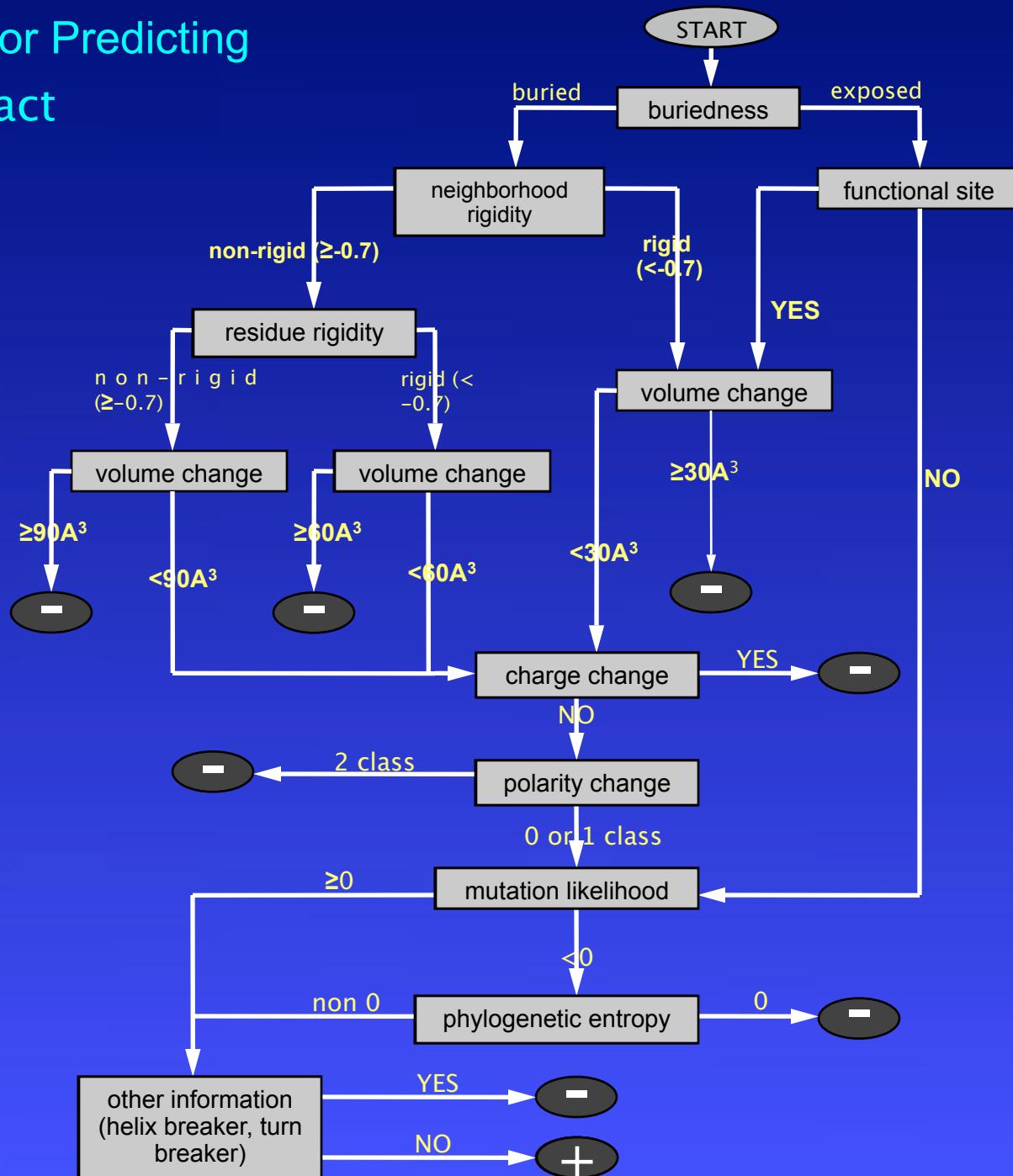
F1761S
M1775E
M1775K
L1780P
I1807S
V1833E
A1843T

M1652I
A1669S

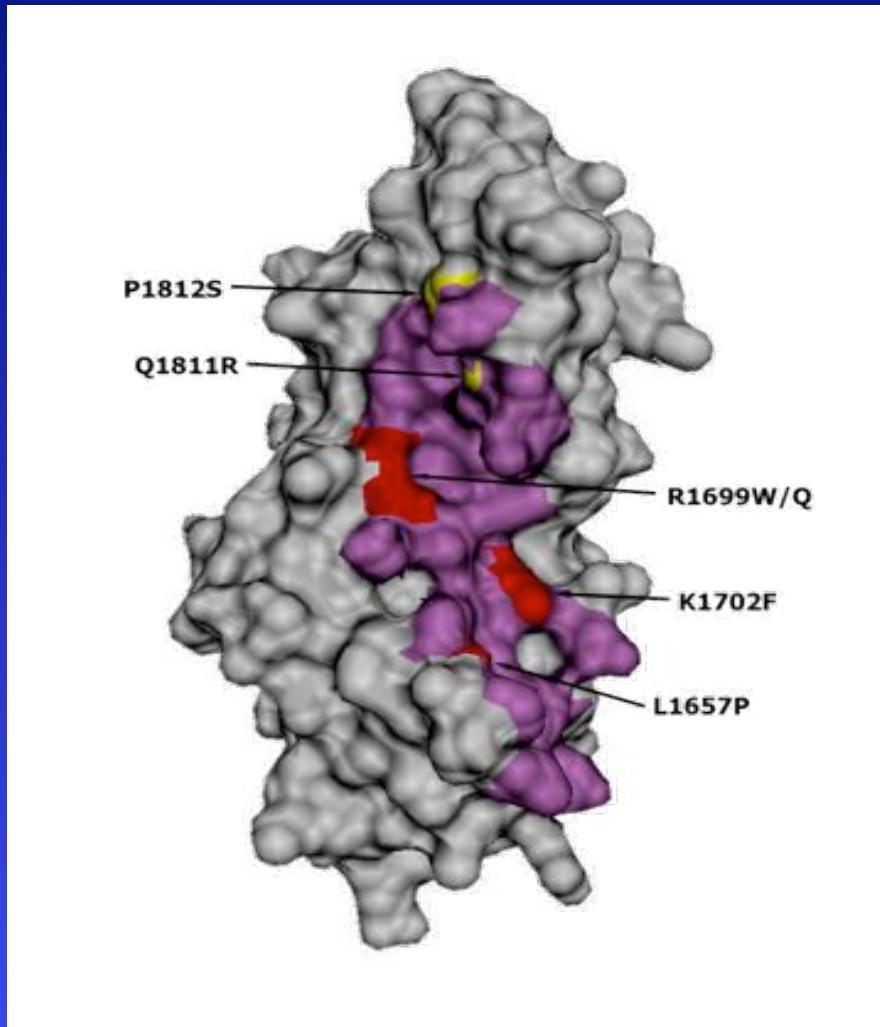
V1665M
D1692N
G1706A
D1733G
M1775V
P1806A

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 G1788 D G1788V G1803A V1804D V1808A V1809A V1809F V1810G Q1811R P1812S
A1823T V1833M W1837R W1837G S1841N A1843P T1852S P1856T P1859R N1819S

“Decision” Tree for Predicting Functional Impact of Genetic Variants



Putative Binding Site on BRCA1



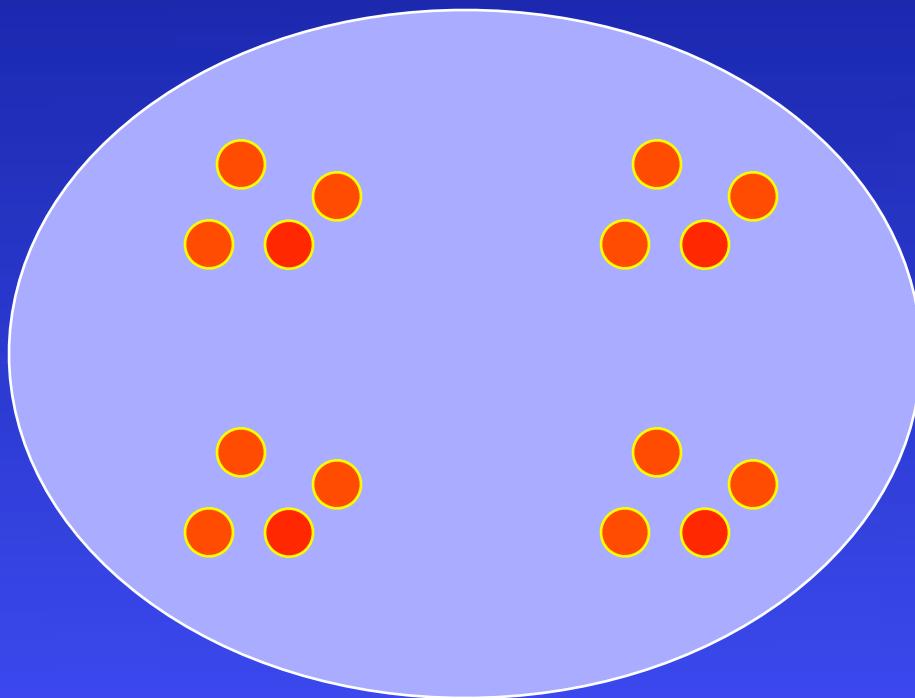
RMSMV**VSG**LTPEEFML**VY**KFARKHHITLT**N**LITEETTHVVMKTD~~A~~FVCERT**L**KYFLGIAGGKwVVSYFWVTQS~~I~~KERK
MLNEHDFEVRGDVVNGRNHQGPKRARESQDRKIFRGLEICCYGPFT**T**NMPTDQLEWMVQLCGASVVKE~~L~~SSFTLGTGVHP
IVVV**Q**PDAWTEDNGFHAIGQMCEAPVVTREW**V**LDSVALYQCQELDTYLIPQIP

genomas...

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.

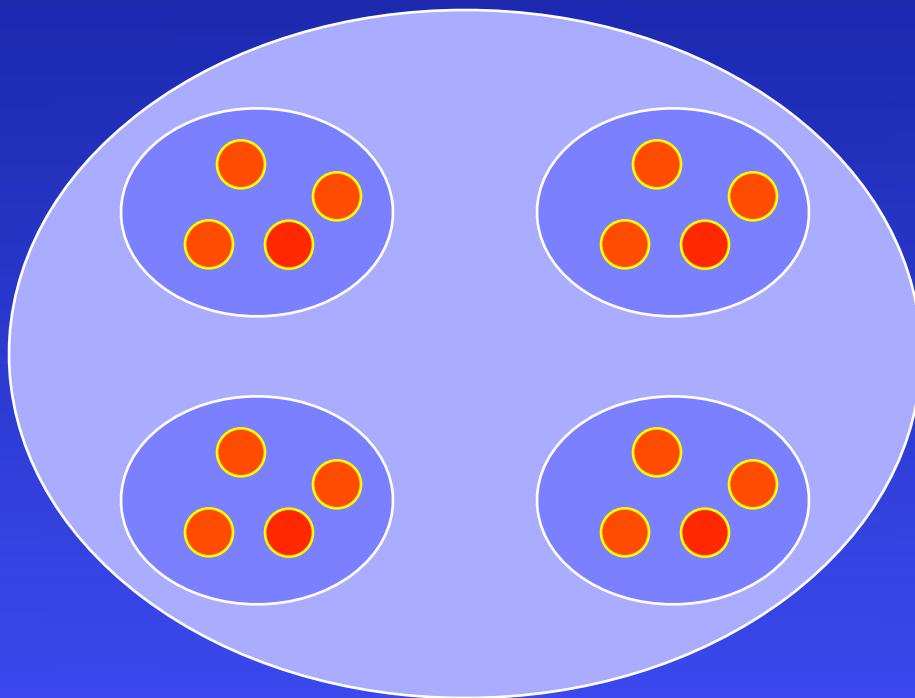
Characterize most protein sequences based on related known structures.



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.

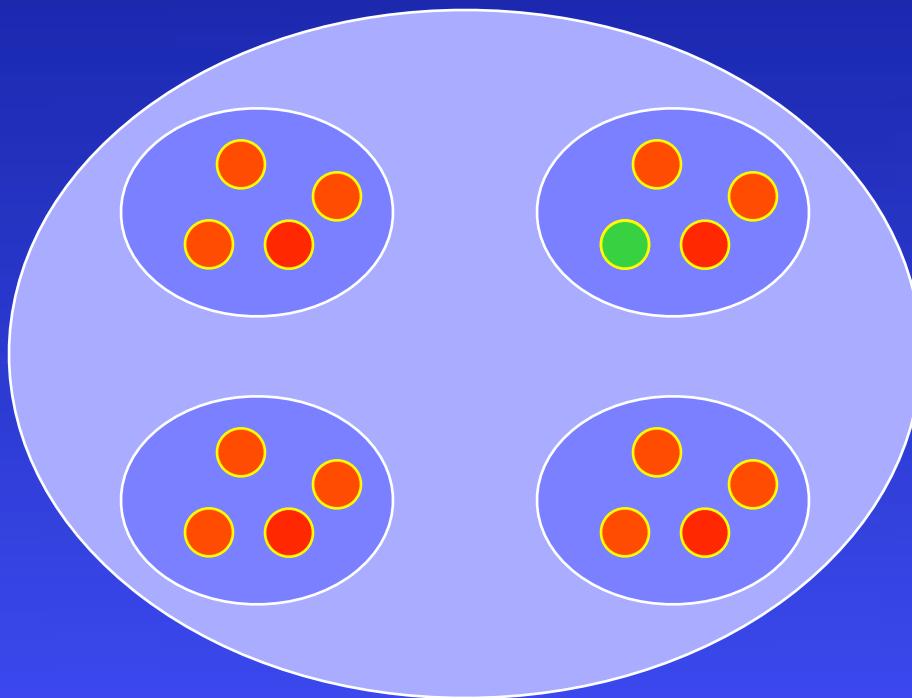
Characterize most protein sequences based on related known structures.



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.

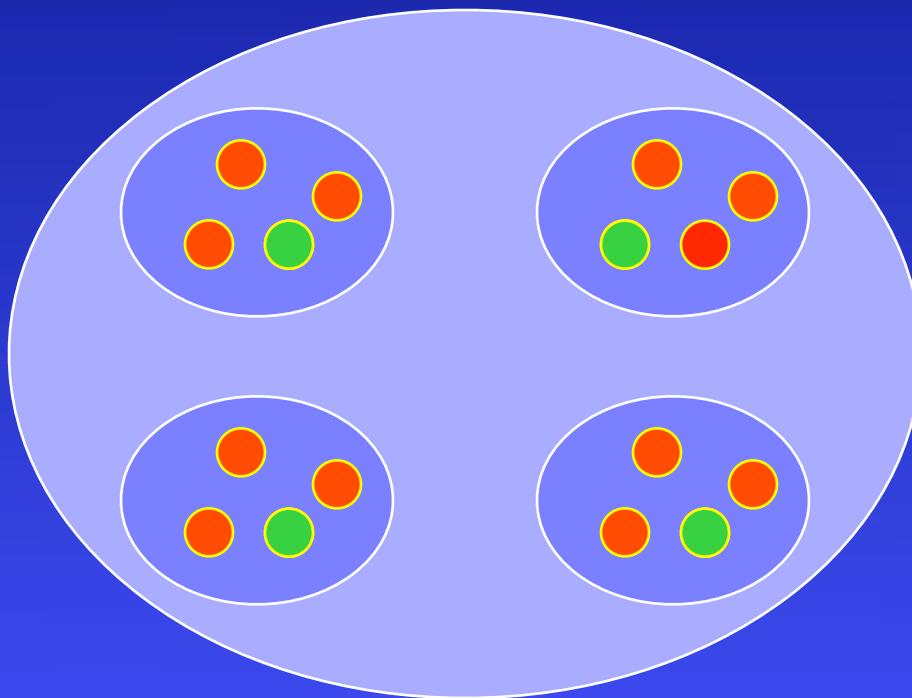
Characterize most protein **sequences** based on related known structures.



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.

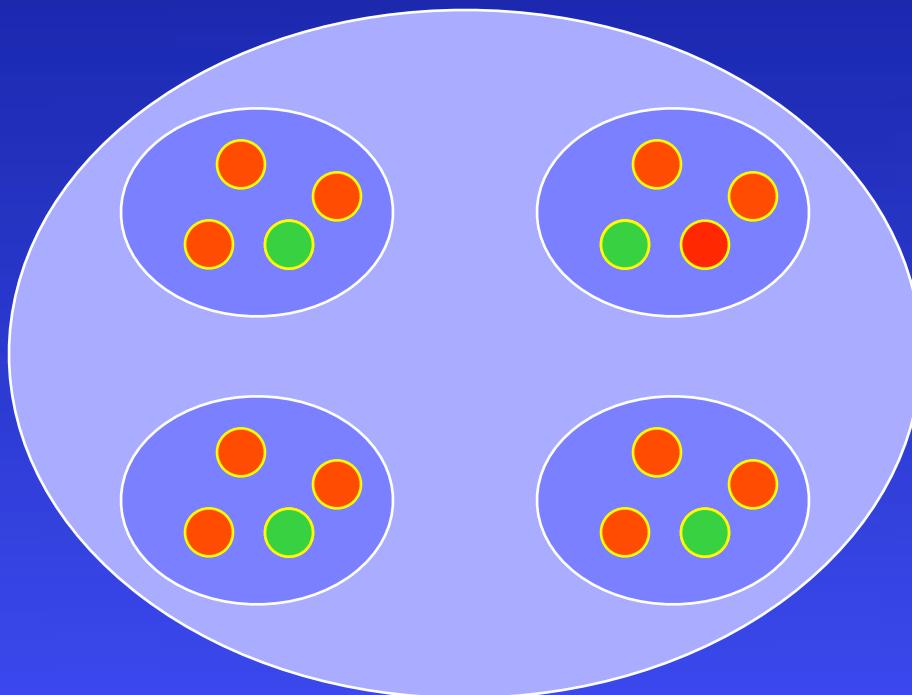
Characterize most protein **sequences** based on related known structures.



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.



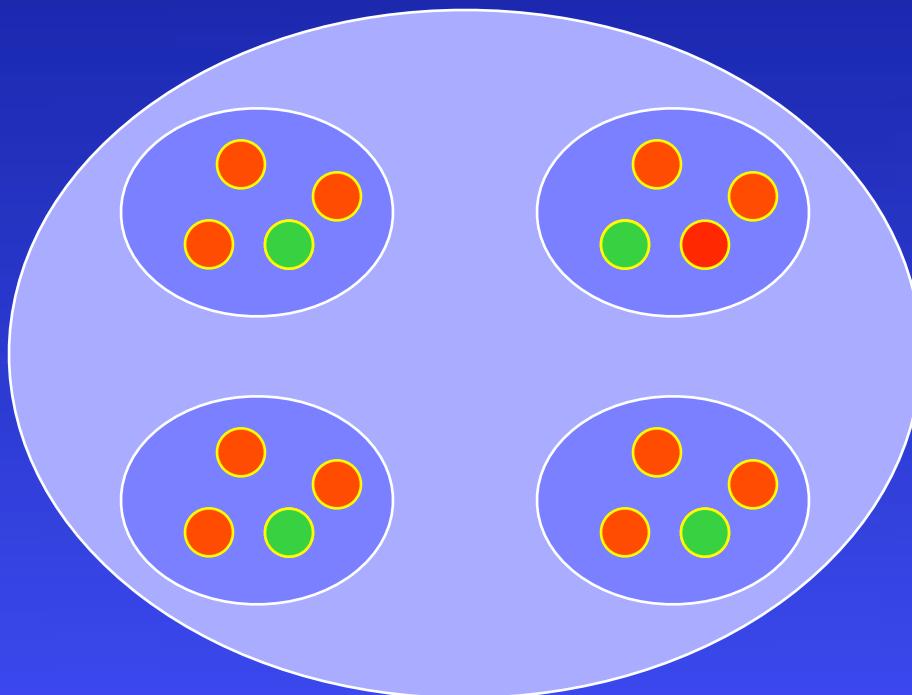
The number of “families” is much smaller than the number of proteins.

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/](http://salilab.org/modweb/)



**Mod
Web**

Server for Comparative Modeling of Genes and Genomes

Please choose input type:

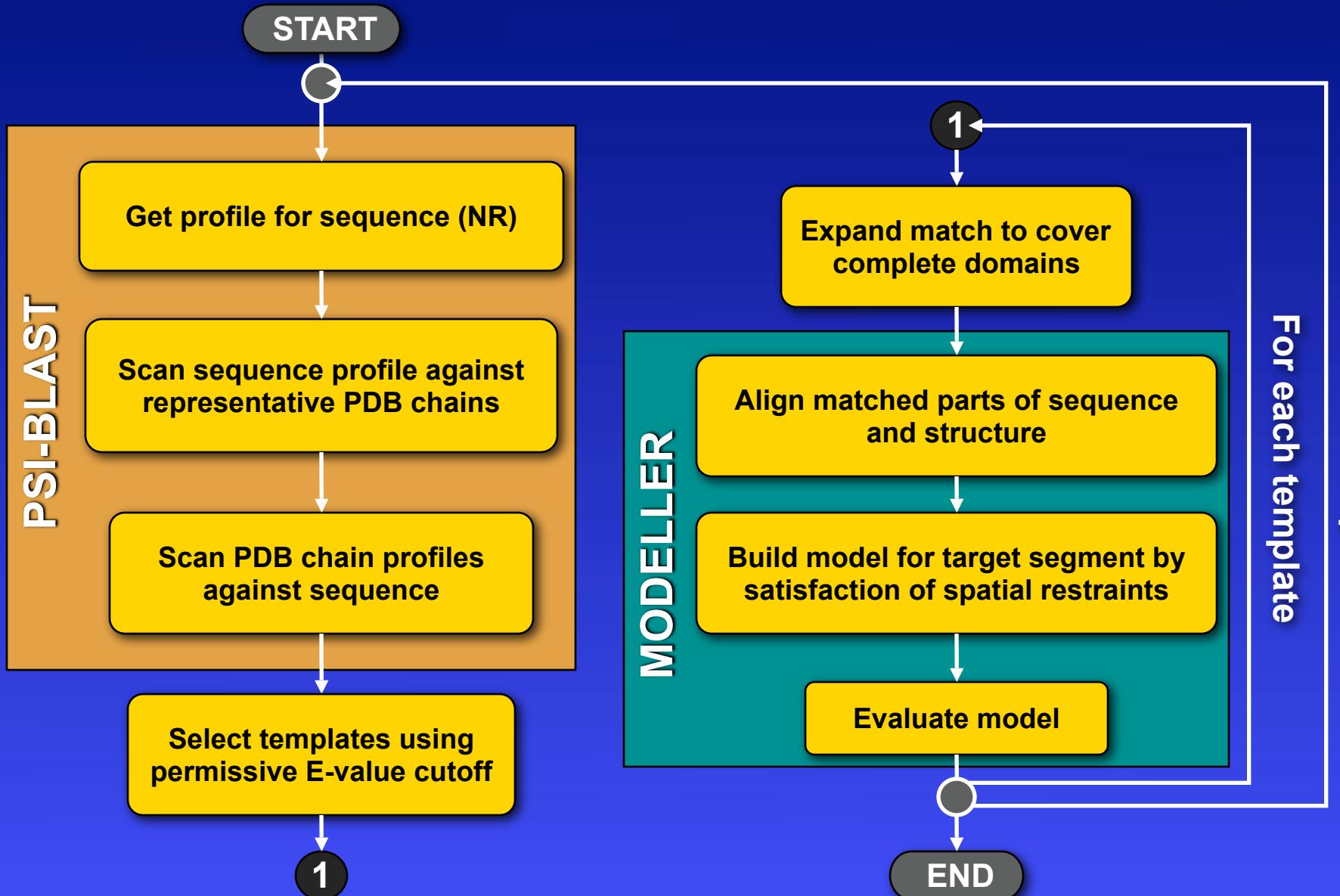
Sequence Enter

Structure

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	# Models <30% Seq. ID
Yeast hypothetical protein	P30187	230	96	21	45	0	35	11	
PhP oxidase	P38075	205	43	18	58	1	33	9	
Yeast hypothetical protein	P49954	271	239	15	54	4	40	190	
T. maritima L-threonine acetaldehyde-lyase	GI_48603902	342	2069	9	51	2	27	2069	
Hypothetical esterase	P30363	288	157	9	48	0	13	144	
Hypothetical protein	P40165	214	35	19	86	1	6	28	
mevalonate diphosphate decarboxylase	GI_1292990	391	121	10	68	2	18	101	
yeast glutathione synthetase	GI_2108534	419	26	30	42	0	26	0	

MODPIPE: Large-Scale Comparative Protein Structure Modeling



Comparative modeling of the TrEMBL database

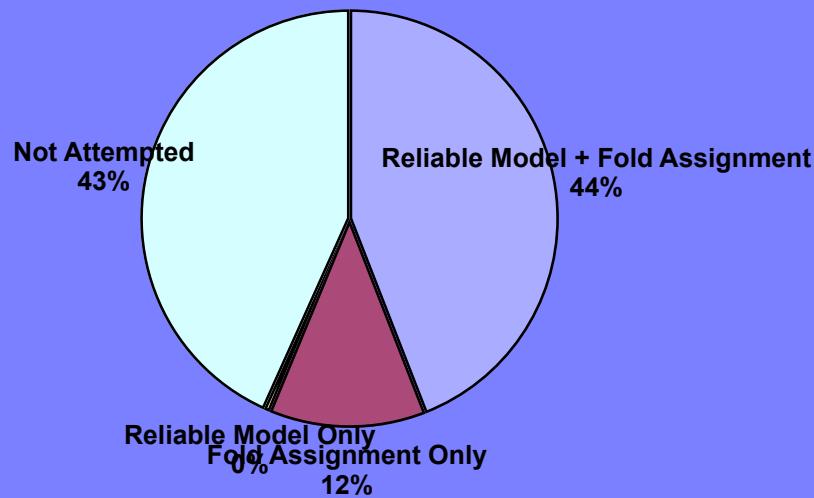
Unique sequences processed: 733,239

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

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

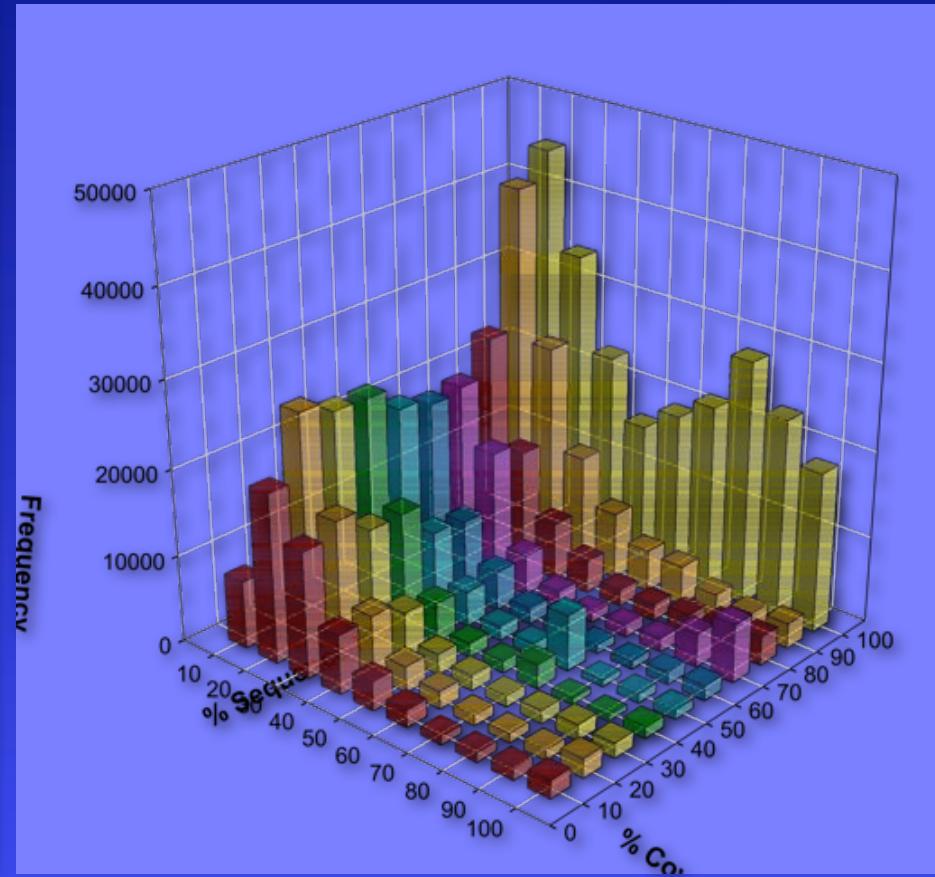
On average, only a domain per protein is modeled
(an “average” protein has 2.5 domains of 175 aa).

Modeling Coverage of the Sequence Space



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

Reliable Model: **Model Score ≥ 0.7**



Netscape ModBase Login Form

File Edit View Go Communicator Back Forward Reload Home Search Netscape Print Security Shop Stop Bookmarks Location http://pipe.rockefeller.edu/modbase-cgi/index.cgi What's Related

WebMail Calendar Radio People Yellow Pages Download Customize

Mod BASE 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. Ursula Pieper, Nanyang Xie, Bo Zhou, David A. Baker, Andrej Sali. *Natl Acad Sci USA* 99, 259-260.

MODBASE is maintained by Ursula Pieper in the group of Andrej Sali, Laboratories of Molecular Biophysics, Fels Family Center for Biochemistry and Structural Biology, The Rockefeller University, 1230 York Ave, New York, NY 10021, USA. E-mail: sali@rockefeller.edu.

© The Rockefeller University

MODBASE Contents

837,892 Reliable Models or PSI-BLAST Fold Assignments for domains in 415,337 proteins. Last Update on 04/03/02. MODBASE statistics.

Search for Models

Enter SwissProt/TBEMBL/GenBank/POB identifier or descriptor:

Advanced Search

Login

Academic login User login Logout

Current login: modbase

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

Notes

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

100%

Bookmarks Location http://pipe.rockefeller.edu/modbase-cgi/query_results.cgi?pub WebMail Calendar Radio People Yellow Pages Download Customize

Mod BASE

K-Mer Sequence Alignment Window

Similarity : 1

1ydrA : 2-157 ELVVALAENRVIQDGELPVPSTIPLAIVLTVVIAIDDPVWVLQVITVNRDDLPQSAQIVMRSERSEVETAHRAASVQAVVKAASLDATAYVIG6
model : 2-156 KVSIIAAAKAANVIGDPLVTSVSIKTVVNGVLLVQGQVLPVPAQVAVYVGSNTSNNVVVQVTAQVAVVHIVVSG6

Search Summary

SUMMARY		Search Criteria		
Keywords	dhfr			
Category	-			
Properties	(% Seq. Ident. and Model Size and Model Score)			
Ranges (min-max)	>30			
Values	8.00 82 0.01			
Average	25	181 0.79		
Maximum	35.00	432 1.00		

* Star indicates does not result in a significant match.

Signed by: Unsigned classes from local hard disk

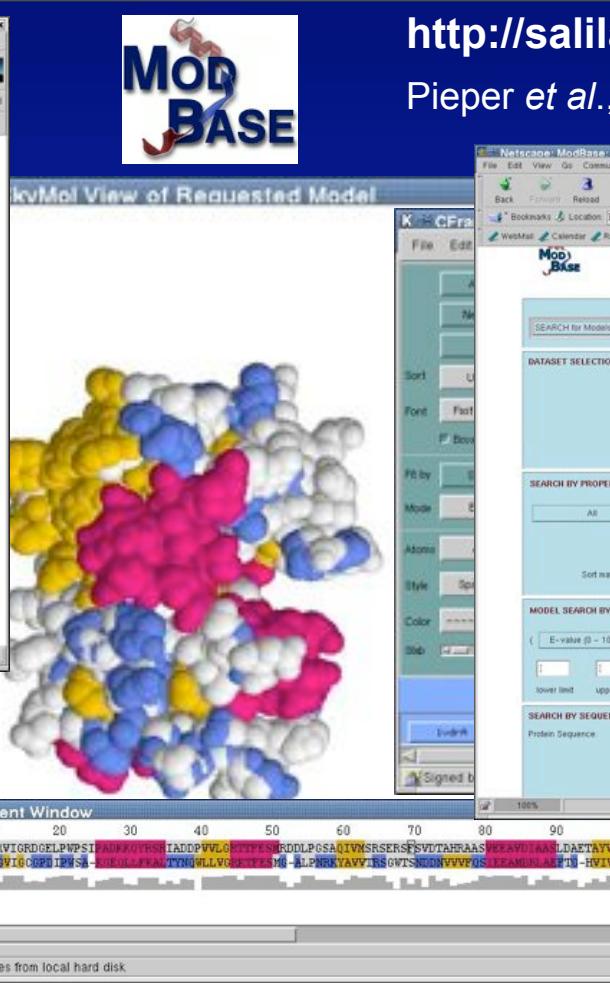
View sequences that match the search criteria but could not be modeled.

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

TARGET		MODEL DATA				TEMPLATE									
Model/Field Reliability	Sequence based View	Select Sequence Database Links	Database Description	Organism	Protein Size	Modeled Segment	Size	Seq.M (%)	E-value	Model Score	PSI-BLAST	Template based View	Segment	Description	
+	+	SR_021152	DIHYDROFOLATE REDUCTASE TYPE III (EC 1.5.1.1) DIHYDRO FOLATE REDUCTASE (DHFR) FROM E. COLI	Escherichia coli (String ID: SR_021152)	169	1-165	165	30.00	4e-32	1.00	0.07	+	+	1-158	DIHYDROFOLATE REDUCTASE
+	+	SR_022384	BIFUNCTIONAL DIHYDROFOLATE REDUCTASE-THYMIDYLIC ACID PHOSPHOTransFERASE (EC 1.5.1.11) DHFR-THF-DPase	Escherichia coli	520	24-231	206	30.00	2e-49	1.00	1.81	+	+	1-166	DIHYDROFOLATE REDUCTASE (EC 1.5.1.1) THYMIDYLIC ACID PHOSPHOTransFERASE (EC 2.7.7.1)
+	+	SR_022334	BIFUNCTIONAL DIHYDROFOLATE REDUCTASE-THYMIDYLIC ACID PHOSPHOTransFERASE (EC 1.5.1.11) DHFR-THF-DPase	Plasmodium chabaudi	583	25-241	221	30.00	3e-39	1.00	1.81	+	+	2-180	EC 1.5.1.1 DHYDROFOLATE REDUCTASE
+	+	SR_022304	BIFUNCTIONAL DIHYDROFOLATE REDUCTASE-THYMIDYLIC ACID PHOSPHOTransFERASE (EC 1.5.1.11) DHFR-THF-DPase	Plasmodium vivax	623	35-237	203	30.00	1e-37	1.00	1.81	+	+	14-203	EC 1.5.1.1 DHYDROFOLATE REDUCTASE

Sequence based View Select Sequence Database Links Database Description Organism Protein Size Modeled Segments-Schema

+	SR_027452	DIHYDROFOLATE REDUCTASE TYPE VIII (EC 1.5.1.3)(DHFR TYPE IIIC) DHFR-THF-DPase	Escherichia coli Shigella sonnei	169	
+	TR_030000	DHFR2 PROTEIN FRAGMENT DHFR-THF-DPase	Homo sapiens	121	



Database Summary for this Sequence (100% Sequence Identity)

TBEMBL	SR_021152	Salmonella typhimurium	Dihydrofolate reductase	+
GI	1232727	Plasmid pLM0229	DHFR product (AA 1 - 157)	-
GI	1230950	Salmonella typhimurium	Dihydrofolate reductase	-
GI	56795	Escherichia coli plasmid pLM0229	EC 1.5.1.3 type I - Escherichia coli plasmid pLM0229	-

(157 Residues)

1-158 DIHYDROFOLATE REDUCTASE - CATH 2.40.430.10.5.1.2 (99%) Subject: SPTR-2001

1ydrA (1-158) DIHYDROFOLATE REDUCTASE - CATH 2.40.430.10.5.1.2 (99%) Subject: SPTR-2001

100%

<http://salilab.org/modbase>

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

8/9/02

Netscape ModBase Search Form

File Edit View Go Communicator Back Forward Reload Home Search Netscape Print Security Shop Stop Bookmarks Location http://pipe.rockefeller.edu/modbase-cgi/search_form.cgi What's Related

WebMail Calendar Radio People Yellow Pages Download Customize

Mod BASE Database of Comparative Protein Structure Models

User: Academic User Change User

SEARCH for Models **SEARCH for Sequences** **RESET**

DATASET SELECTION

Datasets: SPTR-2002 SPTR-2001 nysgrc_1M nysgrc_1Md nysgrc_1Sa nysgrc_1Ba

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

HELP

Model Details

PSI-BLAST Fold Assignments (left half) and Relative 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

Details of relative Models and Fold Assignments of this sequence using the current search parameters

Model Data

Field Model Reliability	Size	Seq.M (%)	E-value	Model Score	PSI-BLAST	LINKS
154	29.00	7e-43	1.00	0.07	+	3D View
155	26.00	2e-37	1.00	0.07	+	3D View

LINKS

Field Model Reliability	Size	Seq.M (%)	E-value	Model Score	PSI-BLAST	LINKS
154	29.00	7e-43	1.00	0.07	+	3D View
155	26.00	2e-37	1.00	0.07	+	3D View

(157 Residues)

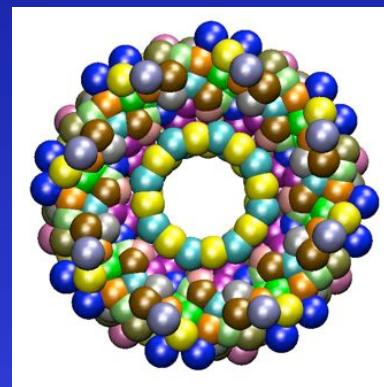
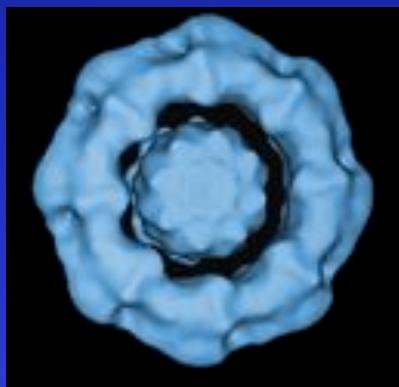
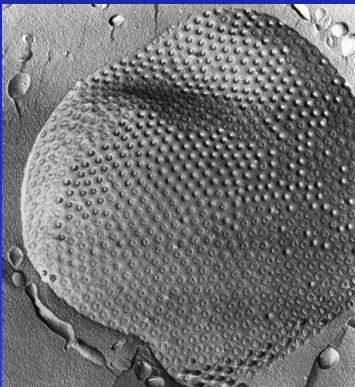
1-158 DIHYDROFOLATE REDUCTASE - CATH 2.40.430.10.5.1.2 (99%) Subject: SPTR-2001

1ydrA (1-158) DIHYDROFOLATE REDUCTASE - CATH 2.40.430.10.5.1.2 (99%) Subject: SPTR-2001

100%

complejos...

Modeling the Yeast Nuclear Pore complex by satisfaction of spatial restraints



**Andrej Sali
Frank Alber, Damien Devos**

Depts. Of Biopharmaceutical Sciences and Pharmaceutical Chemistry
California Institute for Quantitative Biomedical Research
University of California at San Francisco

UCSF

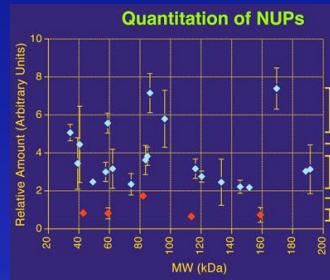
<http://salilab.org/>

**Mike Rout,
T. Suprapto, J. Kipper, L. Veenhoff,
S. Dokudovskaya**

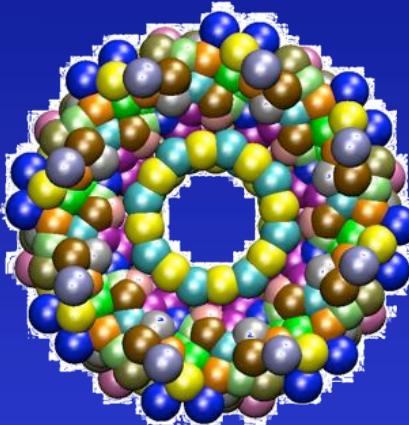
**Brian Chait,
W. Zhang**

The Rockefeller University,
1230 York Avenue, New York

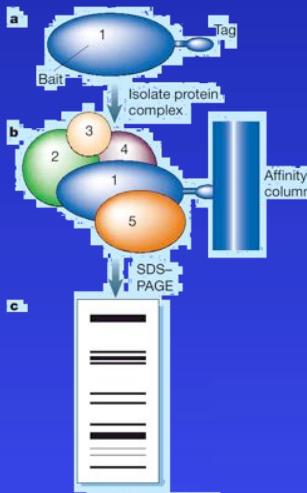
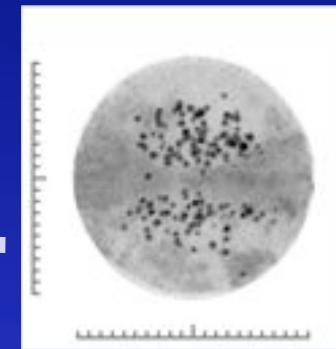
Integrate spatial information



NUP
Stoichiometry



NUP
Localization

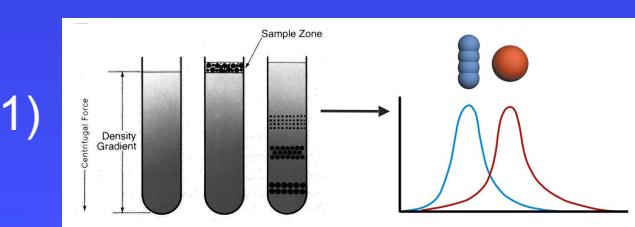


NUP– NUP
Interactions

NUP Shape



Symmetry
Global shape

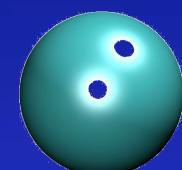
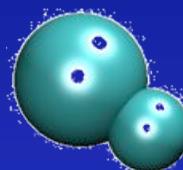
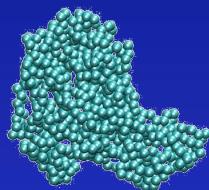
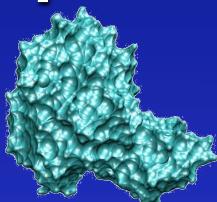
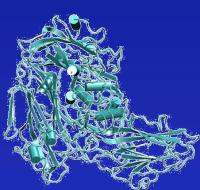


3)



Modeling scheme

■ Protein representation



■ Scoring Function: A sum of spatial restraints

- Excluded volume of proteins
- Symmetry of NPC (EM).
- Radial and axial localization of proteins (IEM)
- Protein-protein contacts (immuno-purification).

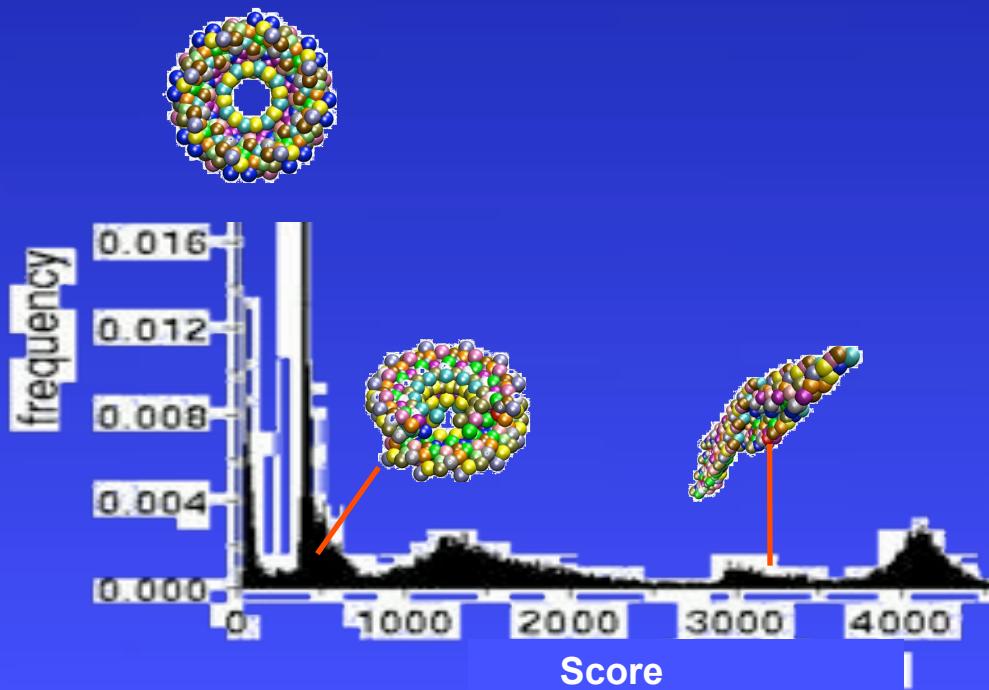
■ Optimization

Minimize violations of input restraints by conjugate gradients and molecular dynamics with simulated annealing.

Obtain an “ensemble” (~100,000) of many independently calculated models, starting from random configuration of protein centers.

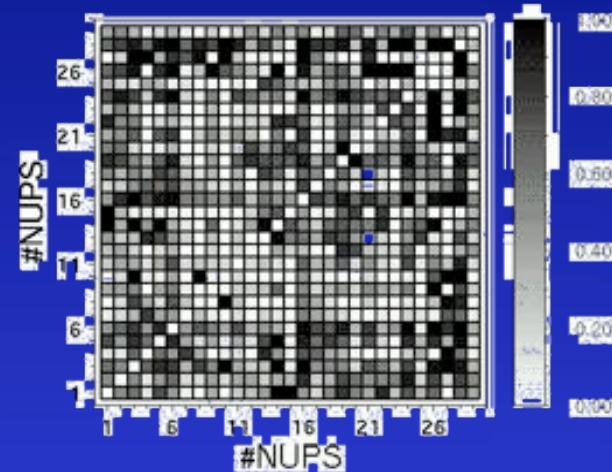
Analysis

- Assessing the well scoring models
 - ◆ How similar are the models to each other?
 - ◆ Do the models make sense given other data?
 - ◆ Using “toy” models as benchmarks.

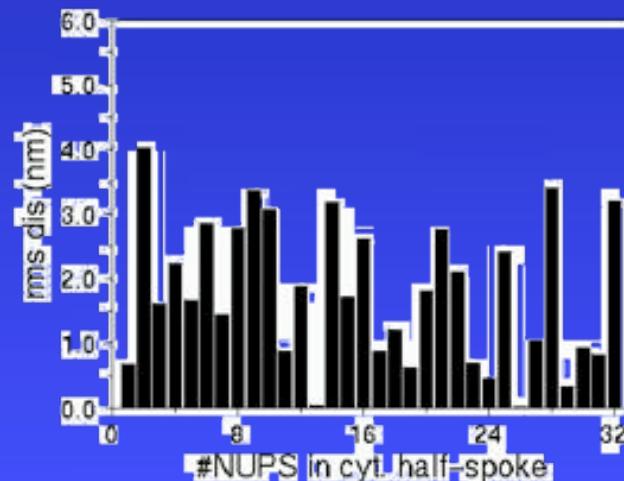
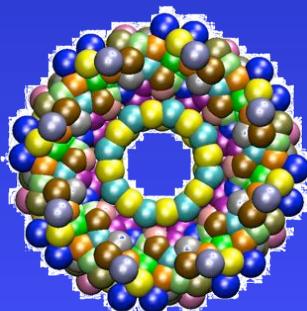


Analysis

- Search conservation of :
 - ◆ Protein-protein contacts



- ◆ Structural features



Conclusiones

- ✓ Hoy en día se pueden modelar alrededor de el 55% de las secuencias de proteínas (un ~25% de los dominios de proteínas).
- ✓ Aplicación a problemas biológicos.
- ✓ La Genómica Estructural intenta determinar o predecir el mayor número posible de estructuras de proteínas.
- ✓ Los modelos de baja resolución de complejos nos ayudan a entender las interacciones de proteínas las células.

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

