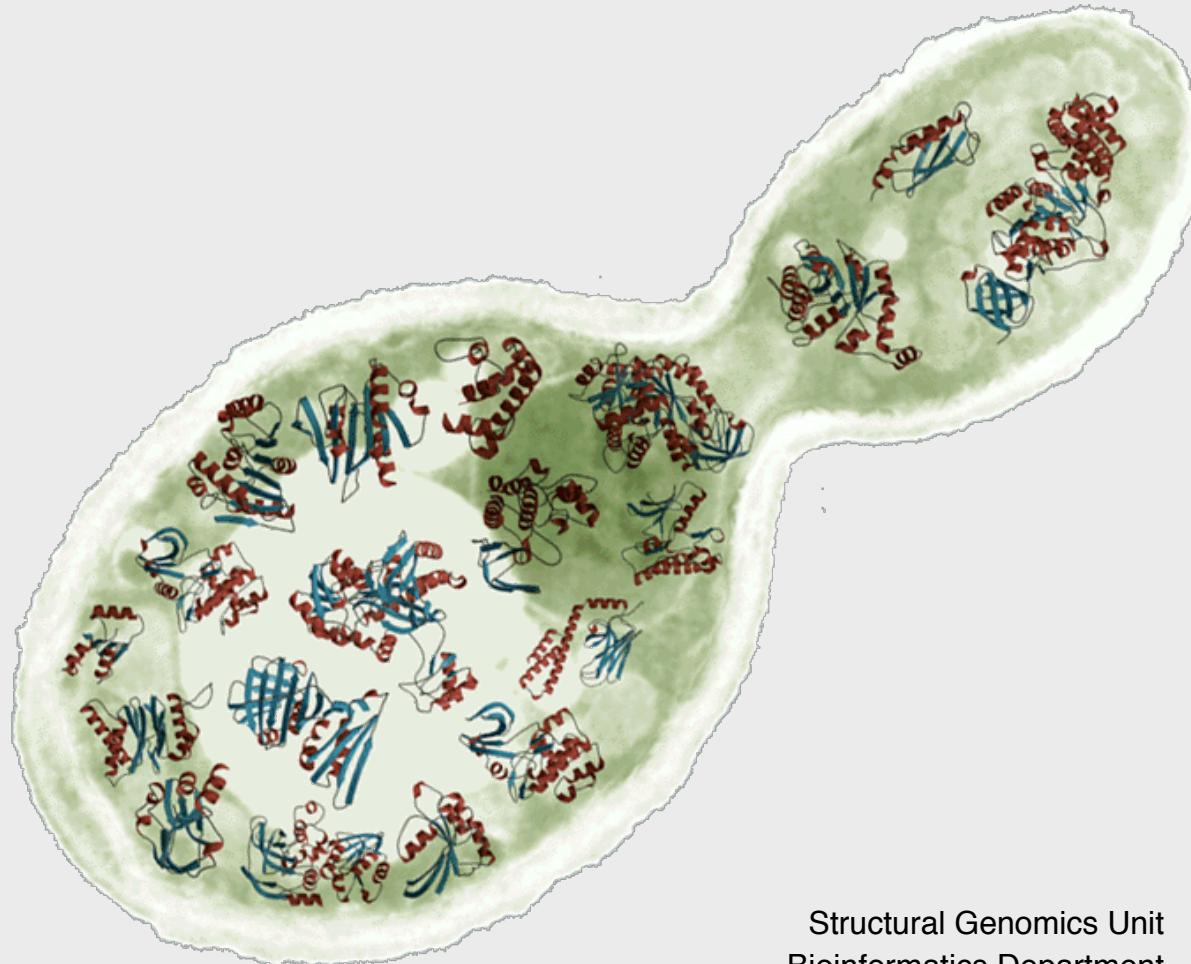


# Modeling the structure and function of proteins and macromolecular assemblies



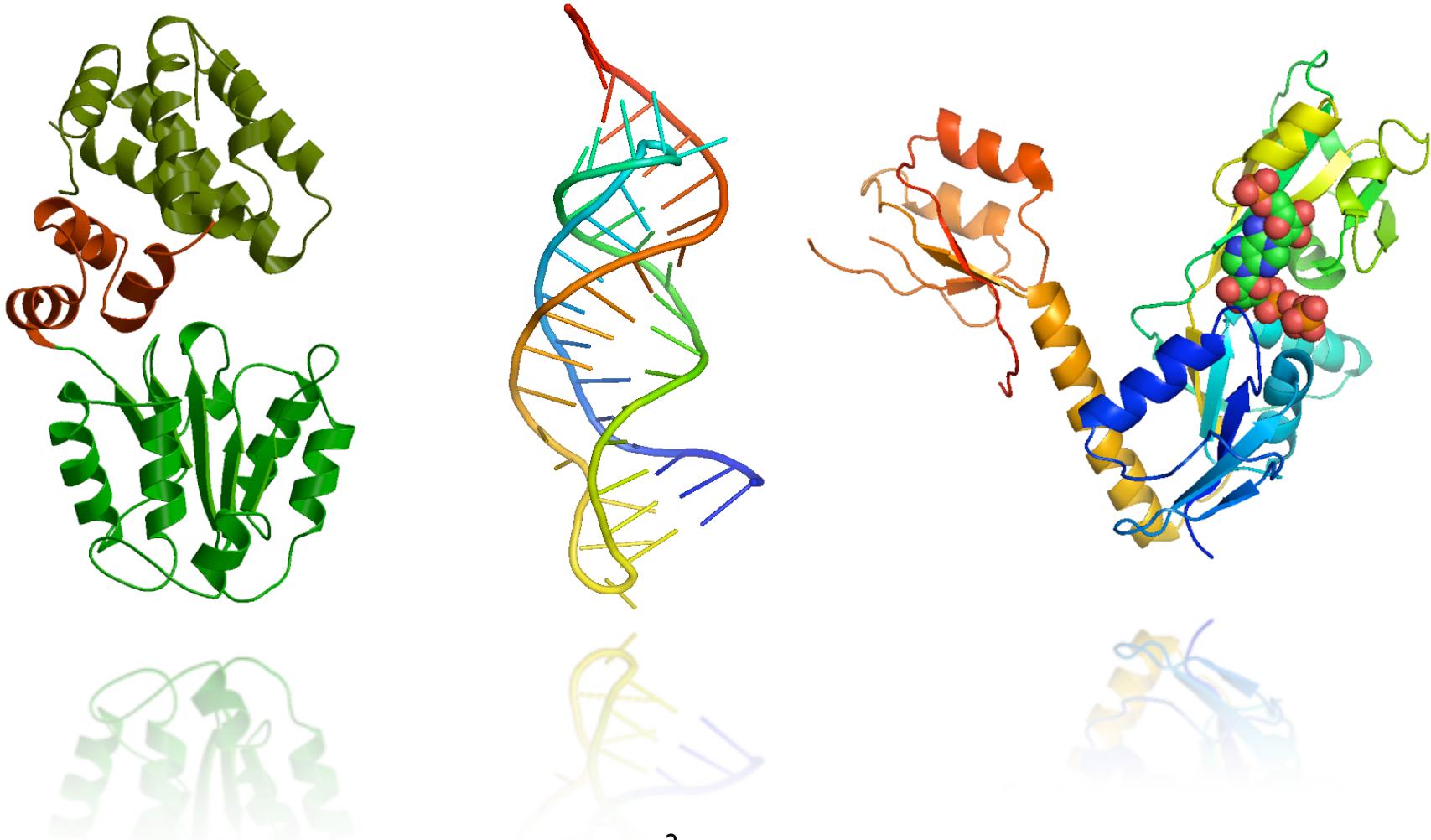
Marc A. Marti-Renom  
<http://bioinfo.cipf.es/sgu/>

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



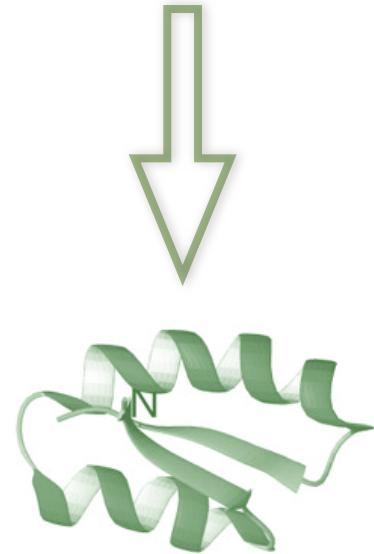
# Structural Genomics Unit

*Bioinformatics Department, CIPF*

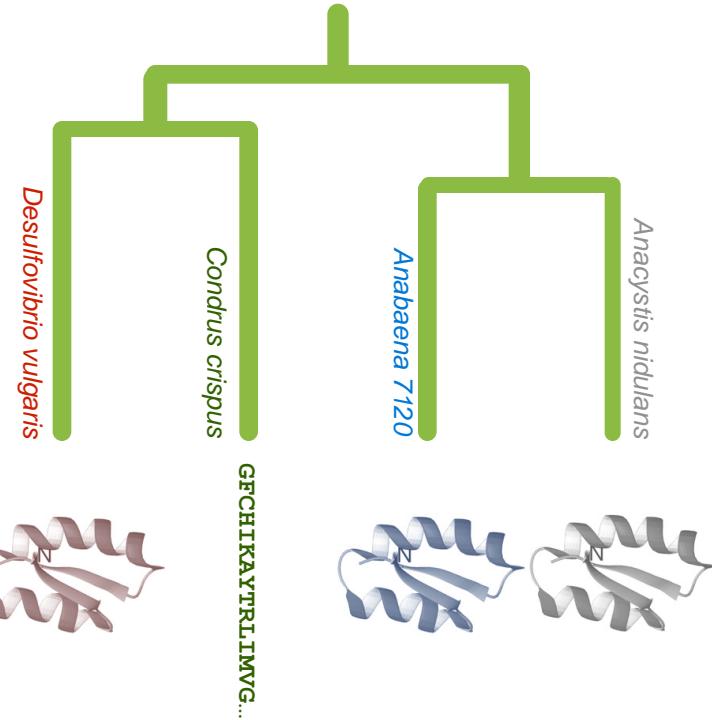


# Principles of protein structure

GFCHIKAYTRLIMVG...



Folding (physics)  
*Ab initio* prediction



Evolution (rules)  
Threading  
Comparative Modeling

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

# ModBase Statistics

Large-scale modeling of the TrEMBL-SWISSPROT databases

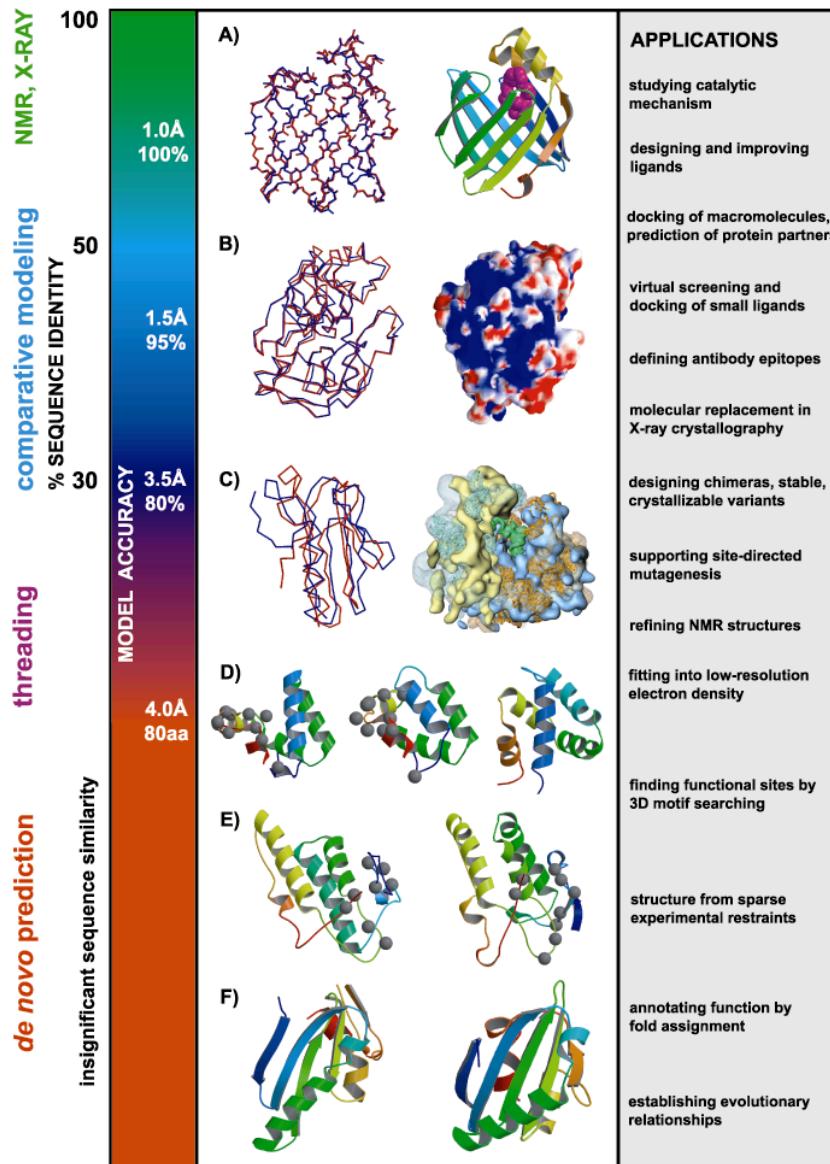
<http://www.salilab.org/modbase/>

Sequences (total)	1,930,692
Sequences (modeled)	1,084,784
Models	3,094,542

The screenshot shows the ModBase Search Page at [http://modbase.compbio.ucsf.edu/modbase-cgi-new/search\\_form.cgi](http://modbase.compbio.ucsf.edu/modbase-cgi-new/search_form.cgi). The page title is "Database of Comparative Protein Structure Models". It features a search form with fields for "Search type" (Model(Default)), "Display type" (Model Detail (graphical)), and "Property" (Database Accession Number). A note states: "MODBASE contains theoretical, calculated models, not experimentally determined structures. The models may contain significant errors." At the bottom, it says: "Users of Modbase are requested to cite this article in their publications: MODBASE, a database of annotated comparative protein structure models, and associated resources, Ursula Pieper, Narayanan Eswar, Hannes Braberg, M.S. Madhusudhan, Fred Davis, Ashley C. Stuart, Nebojsa Mirkovic, Andrea Rossi, Marc A. Marti-Renom, Andrej Sali, Ben Webb, Daniel Greenblatt, Conrad Huang, Tom Ferrin, Andrej Sali. Nucleic Acids Research 32, D217-D222, 2004 MODBASE is maintained by Ursula Pieper in the group of Andrej Sali, Departments of Pharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research Mission Bay Genentech Hall University of California San Francisco, San Francisco, CA 94143-2240. Please address all inquiries to [modbase@salilab.org](mailto:modbase@salilab.org).

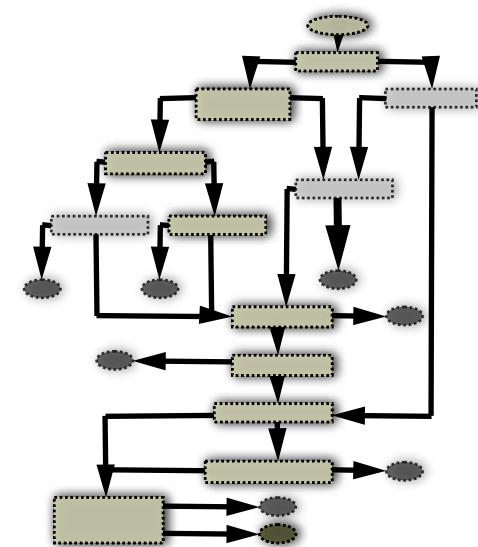


# Utility of protein structure models, despite errors

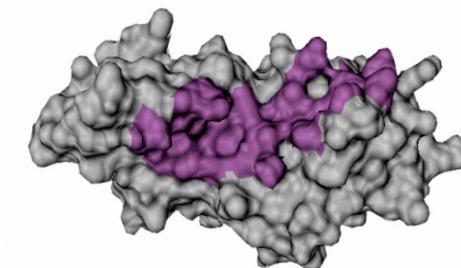
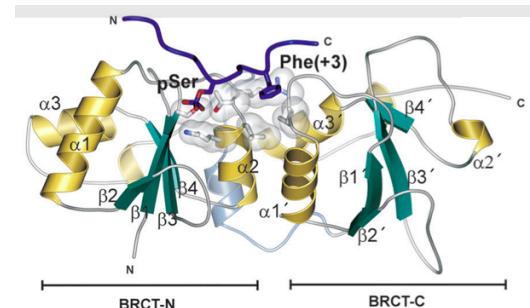
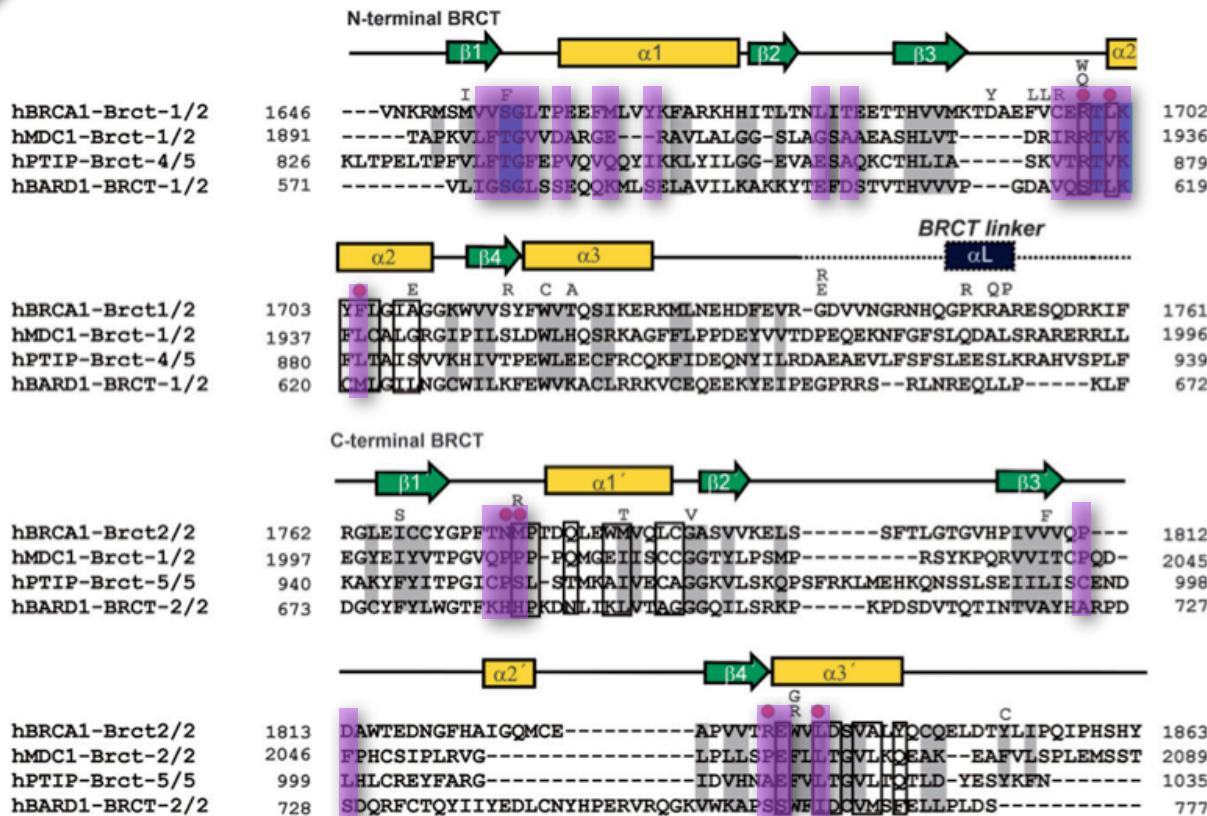


# Missense mutations in BRCT domains

	cancer associate	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

# For 20% protein structures function is *unknown*

	Structural Genomics*	Traditional methods
Annotated**	654	28,342
Not Annotated	506 (43.6%)	6,815 (19.4%)
Total deposited	1,160	35,157

\* annotated as STRUCTURAL GENOMICS in the header of the PDB file

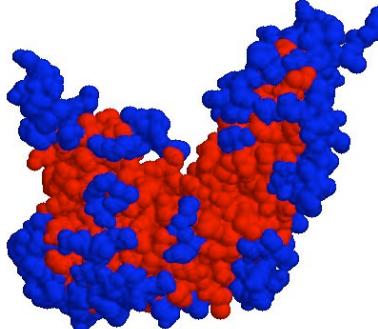
\*\*annotated with either CATH, SCOP, Pfam or GO terms in the MSD database  
36,317 protein structures, as of August 8th, 2006

# Representation

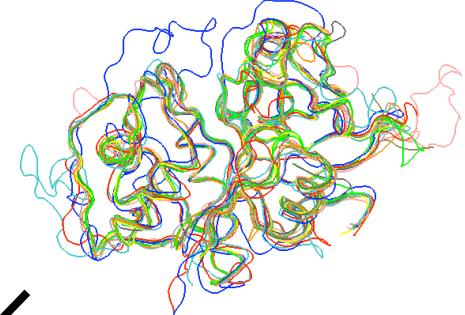
## Sequence conservation

CATB : | S P W Q T D I K M G H V C E R H S Y S N Q P R I F P C H  
CATD : - A - F R I L V Q V  
CATF : - V - V - V X  
CATG : - A - V  
CATK : - S - T V  
CATL : - V - V  
CATM : - V - V  
CATW : - V - V  
  
CATB : 33 S P W Q T D I K M G H V C E R H S Y S N Q P R I F P C H  
CATD : 47 - A - F R I L V Q V  
CATF : 51 V  
CATG : 54 S - T V  
CATK : 57 V  
CATL : 61 V  
CATM : 65 V  
CATW : 69 V  
  
CATB : 110 H V V V H R D P C P C Y T Q S D Q D - - - R K C R P Y T R P - - - - -  
CATD : 92 Q R R S - M Y -  
CATF : 94 V D S E R - K Y -  
CATG : 95 D R V V V D V C G - 7 - C H P X X C D A T T V C G V C G -  
CATK : 97 S - T V  
CATL : 100 V  
CATM : 104 S - T V  
CATW : 108 D S H V D P P K X Y Q V V V V V V V V V V V V V V V V  
  
CATB : 133 W - K S -  
CATD : 115 D P P S C D P M P M M V  
CATF : 119 V a s d c - A -  
CATG : 123 V a s d c - A -  
CATK : 127 D R V V V D V C G - 7 - C H P X X C D A T T V C G V C G -  
CATL : 136 G C -  
CATM : 139 E D T T T V  
CATW : 143 D S H V D P P K X Y Q V V V V V V V V V V V V V V V V V  
  
CATB : 202 I L I D -  
CATD : 147 - X D -  
CATF : 151 E G -  
CATG : 155 S - T V  
CATK : 159 D R V V V D V C G - 7 - C H P X X C D A T T V C G V C G -  
CATL : 163 V  
CATM : 167 V  
CATW : 171 V  
  
CATB : 244 V V A S I P R S D -  
CATD : 216 -  
CATF : 220 -  
CATG : 224 -  
CATK : 228 -  
CATL : 232 -  
CATM : 236 -  
CATW : 240 -

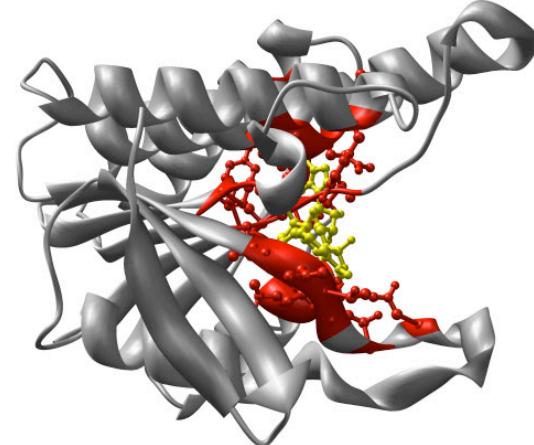
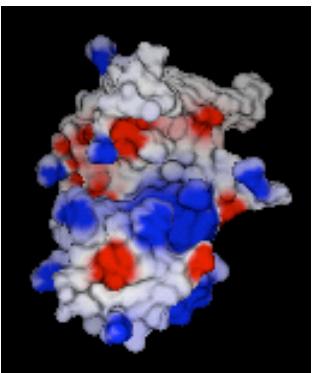
## Surface geometry



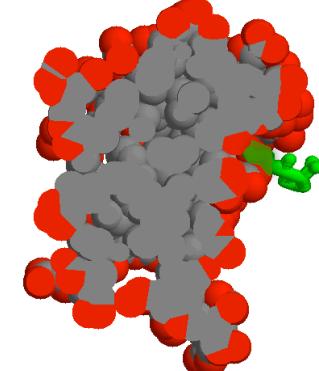
## Structure conservation



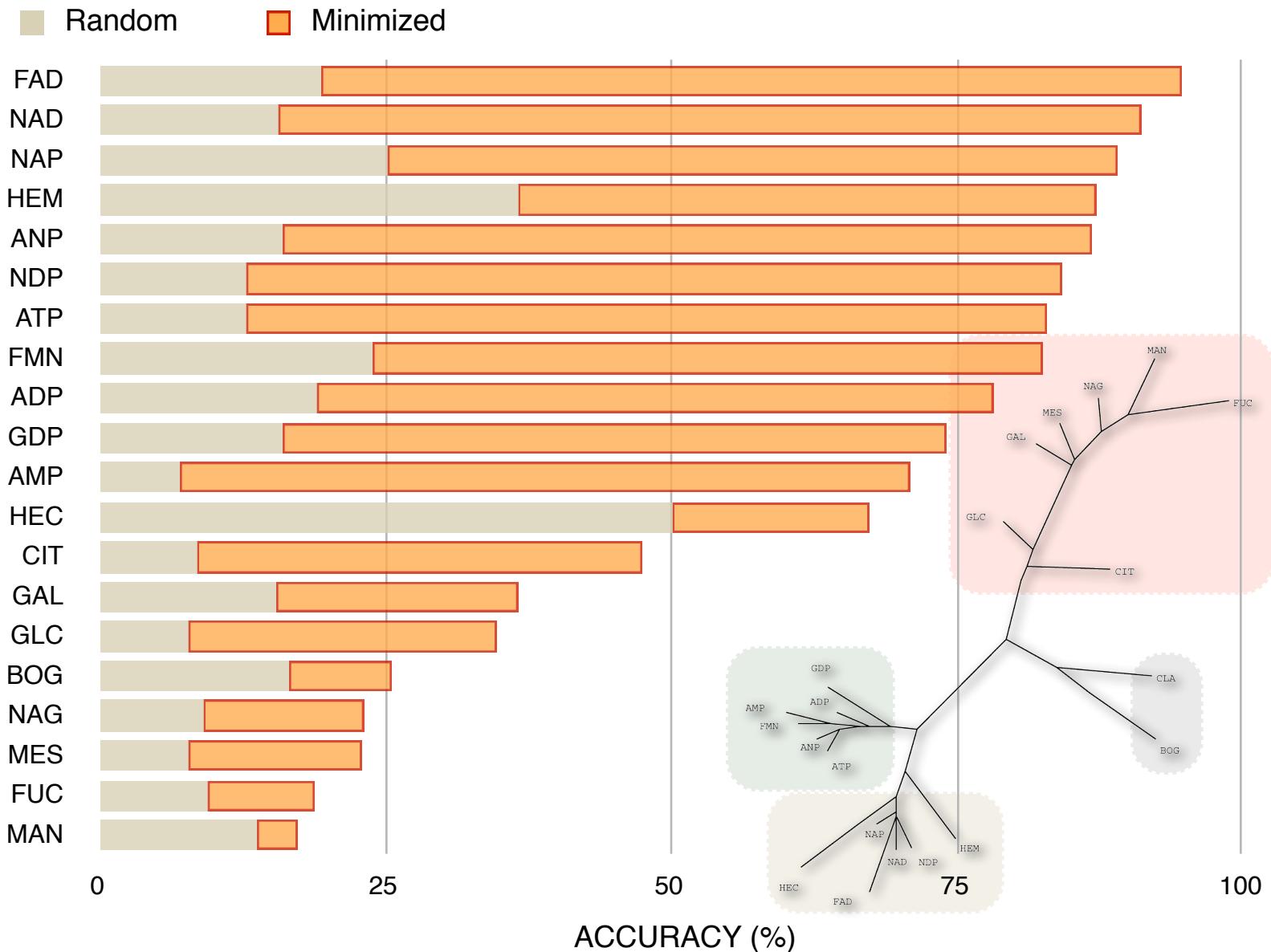
## Electrostatics



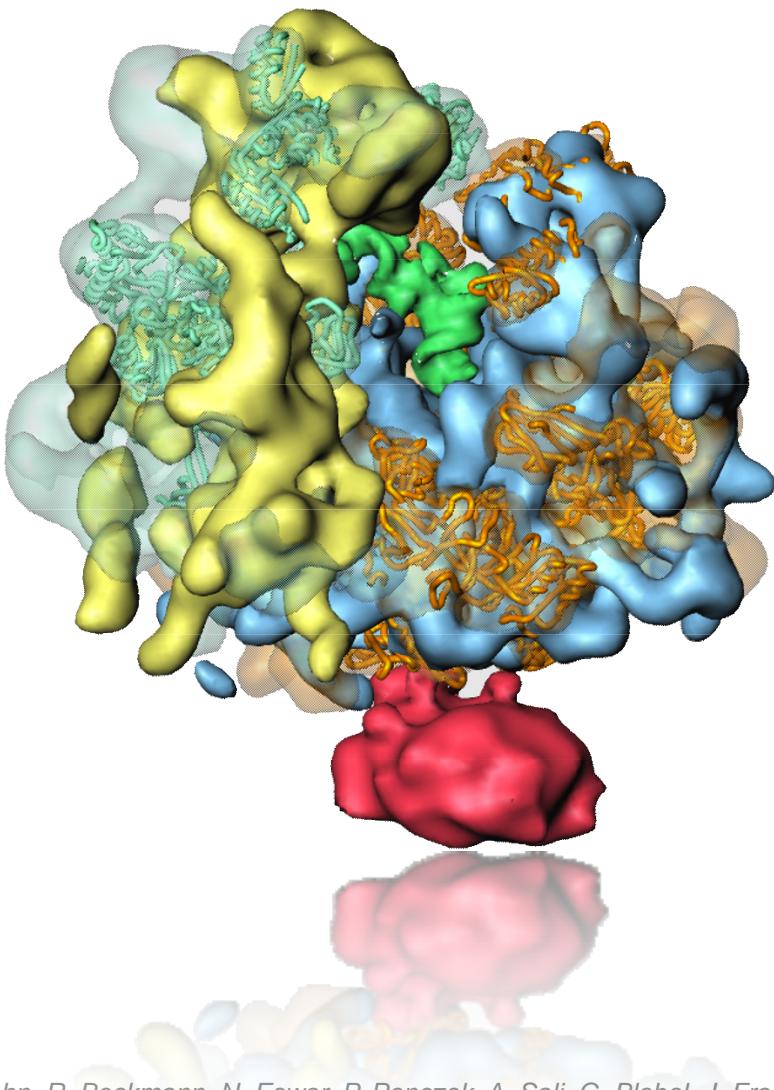
## Solvent accessibility



# Prediction accuracy



# *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%.

# Acknowledgments



## COMPARATIVE MODELING

**Andrej Sali**  
M. S. Madhusudhan  
**Narayanan Eswar**  
Min-Yi Shen  
Damien Devos  
**Neboja Mirkovic**  
**Ursula Pieper**

## MODEL ASSESSMENT

Francisco Melo (CU)  
Alejandro Panjkovich (CU)

## Tropical Disease Initiative

Stephen Maurer (UC Berkeley)  
Arti Rai (Duke U)  
Andrej Sali (UCSF)  
Thomas Kepler (Duke U)  
Ginger Taylor (TSL)

## STRUCTURAL GENOMICS

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

**EVA**  
Burkhard Rost (Columbia U)  
Alfonso Valencia (CNIO)

## CAMP

Xavier Aviles (UAB)  
Hans-Peter Nester (SANOFI)  
Ernst Meinjohanns (ARPIDA)  
Boris Turk (IJS)  
Markus Gruetter (UE)  
Matthias Wilmanns (EMBL)  
Wolfram Bode (MPG)

## MODELING ASSEMBLIES

Frank Alber  
Fred P. Davis  
Maya Topf

## FUNCTIONAL ANNOTATION

Fatima Al-Shahrour  
Joaquin Dopazo

## FUNCTIONAL ANNOTATION

**Andrea Rossi**  
Fred P. Davis

Prince Felipe Research Center  
Marie Curie Reintegration Grant  
STREP grant

## BIOLOGY

Jeff Friedman (RU)  
James Hudsped (RU)  
Partha Ghosh (UCSD)  
**Alvaro Monteiro (Cornell U)**  
Steven Krillis (St.George H)



**PRINCIPE FELIPE**  
CENTRO DE INVESTIGACION