A genome-wide quest for drug discovery targets against tropical diseases.

Marc A. Marti-Renom
Genome Biology Group (CNAG)
Structural Genomics Group (CRG)

*iCrea
INSTITUCIÓ CATALANA DE
RECERCA I ESTUDIS AVANÇATS

cnag CRG
Centre
for Genomic
Regulation

Outline...

COMPARATIVE MODELING

EXAMPLES

THE TROPICAL DISEASE INITIATIVE

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.
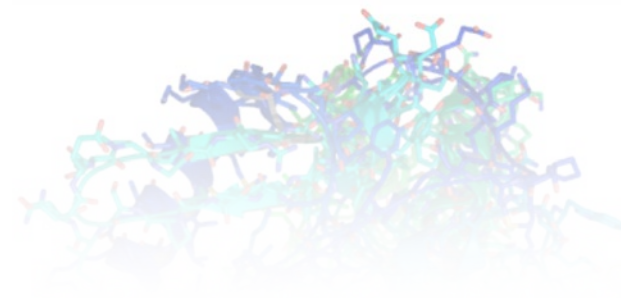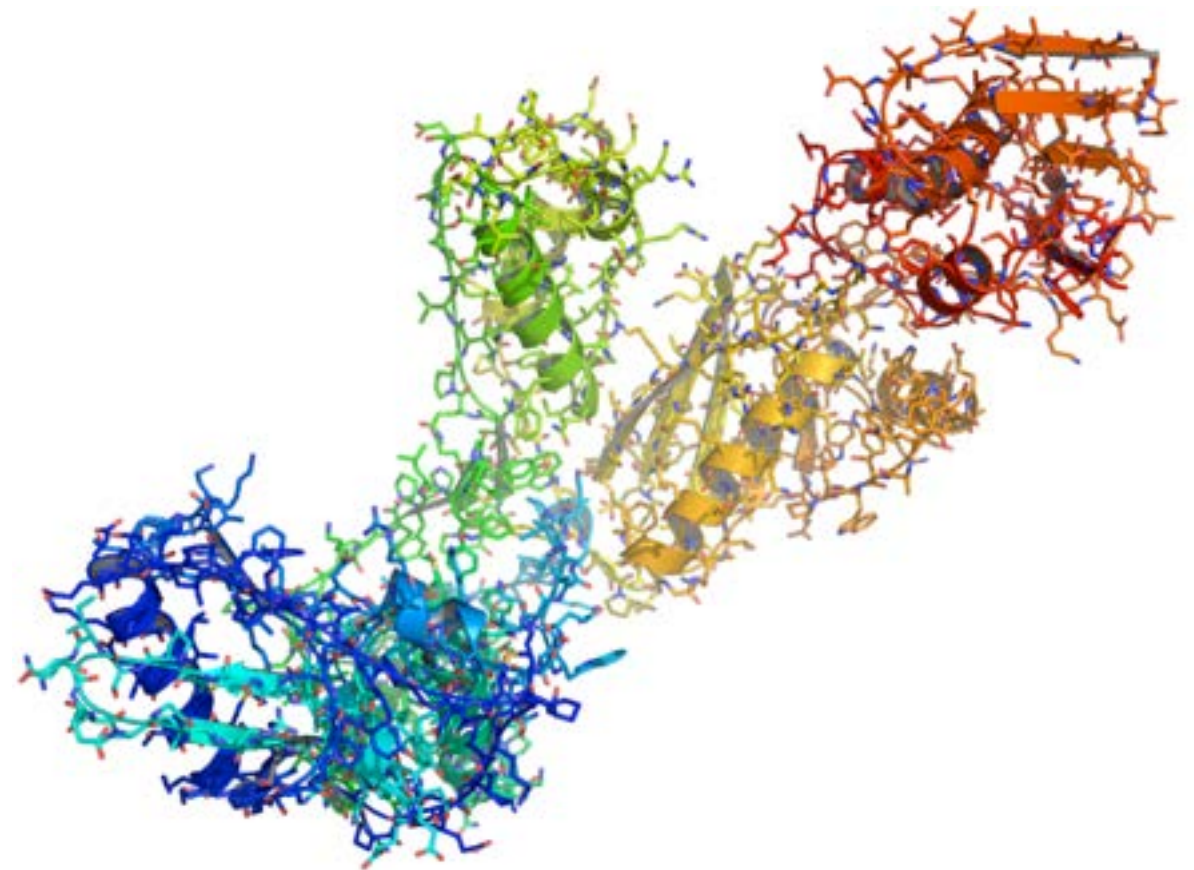
Nomenclature

Fold: Three dimensional conformation of a protein sequence (usually at domain level).

Domain: Structurally globular part of a protein, which may independently fold.

Secondary Structure: Regular sub-domain structures composed by alpha-helices, beta-sheets and coils (or loops).

Backbone: Protein structure skeleton composed by the carbon, nitrogen and oxygen atoms.

Side-Chain: Specific atoms identifying each of the 20 residues types.

protein prediction .vs. protein determination

X-Ray

NMR

Comparative Modeling

Threading

Ab-initio

Experimental data

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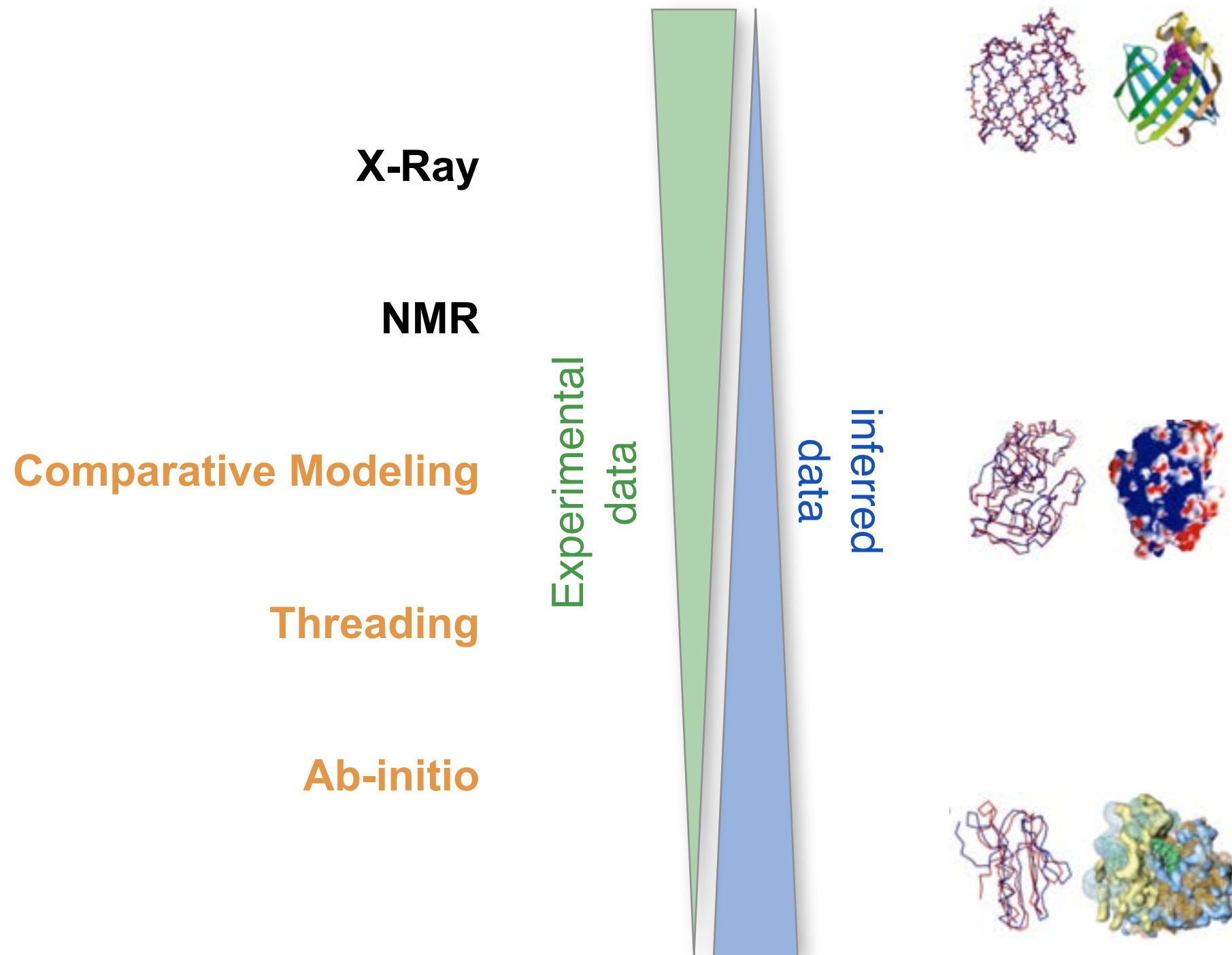
Why is it useful to know the structure of a protein, not only its sequence?

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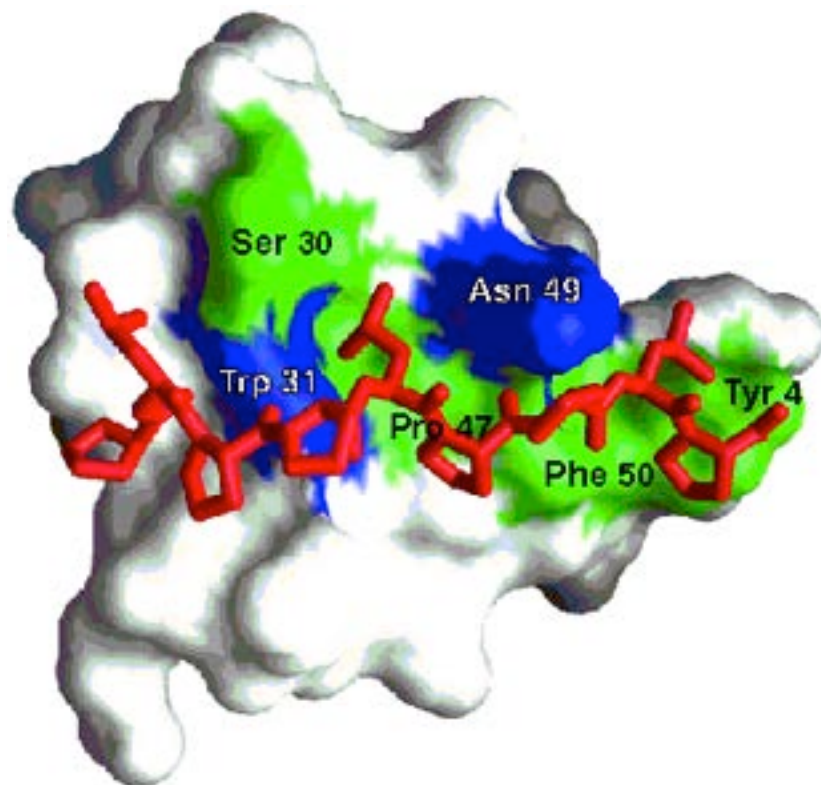
In addition, since evolution tends to conserve function and function depends more directly on structure than on sequence, **structure is more conserved in evolution than sequence**.

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

Principles of protein structure
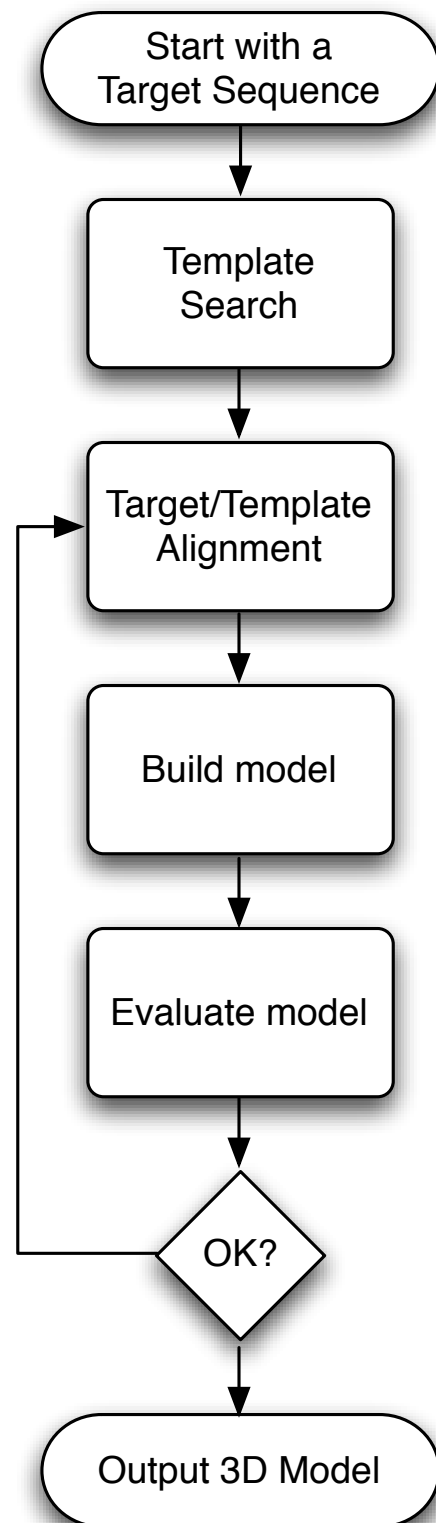
GFCHIKAYTRLIMVG...
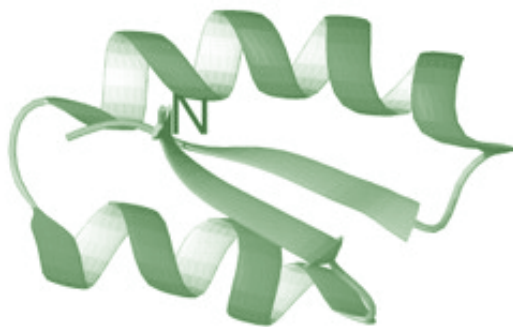
Folding (physics)

Ab initio prediction

Desulfovibrio vulgaris

Condrus crispus

Anabaena 7120

Anacystis nidulans

GFCHIKAYTRLIMVG...
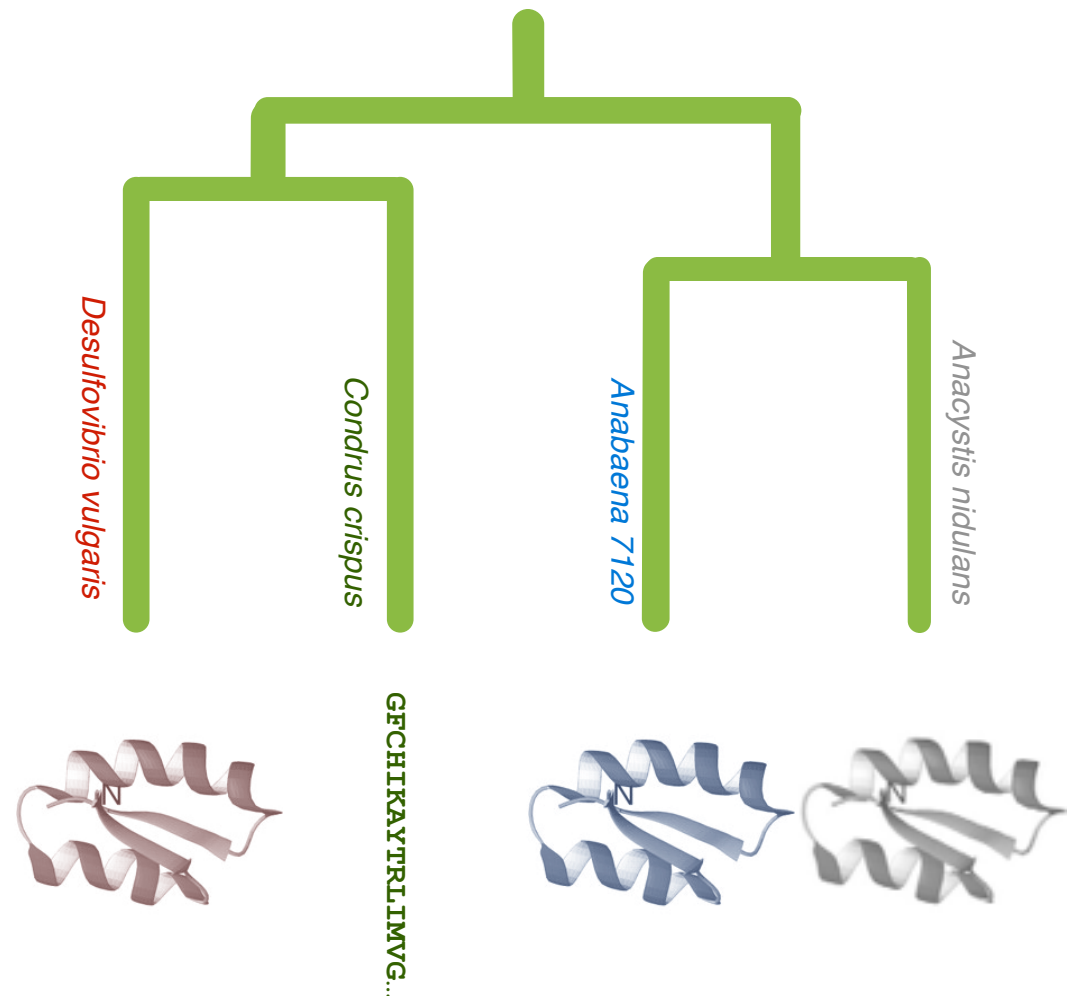
Evolution (rules)

Threading
Comparative Modeling

Comparative modeling by satisfaction of spatial restraints

Start with a
Target Sequence

Template
Search

Target/Template
Alignment

Build model

Evaluate model

OK?

Output 3D Model

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

MSVIPKR--GNCEQTSE
ASILPKRLFGNCEQTSD

$C_\alpha - C_\alpha$ DISTANCE [Å]

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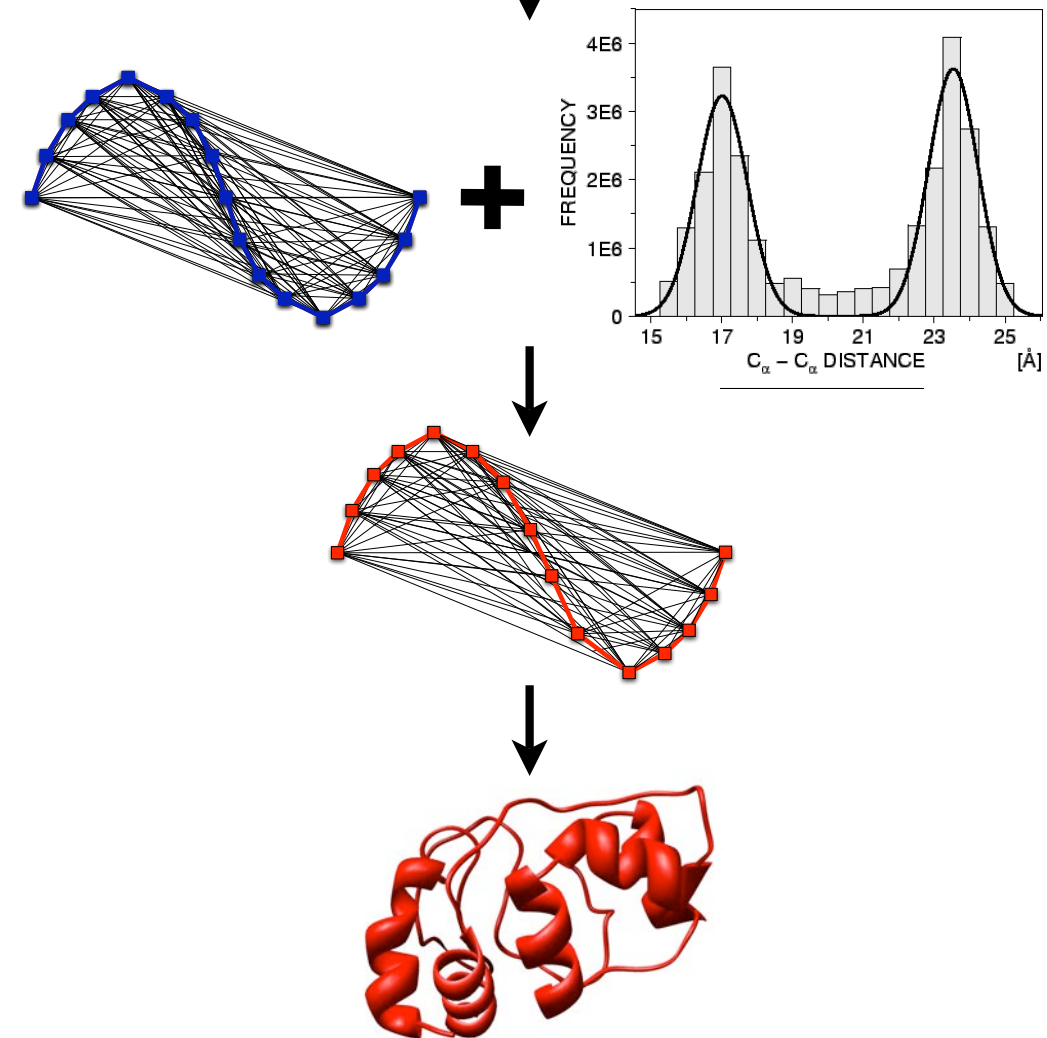
A. Šali & T. Blundell. J. Mol. Biol. 234, 779, 1993.
J.P. Overington & A. Šali. Prot. Sci. 3, 1582, 1994.
A. Fiser, R. Do & A. Šali, Prot. Sci., 9, 1753, 2000.

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

Wrong fold
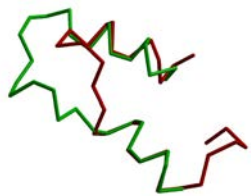
Miss alignments
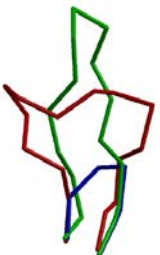
Loop regions
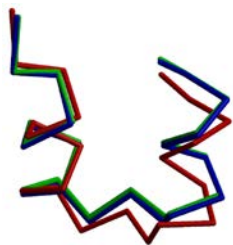
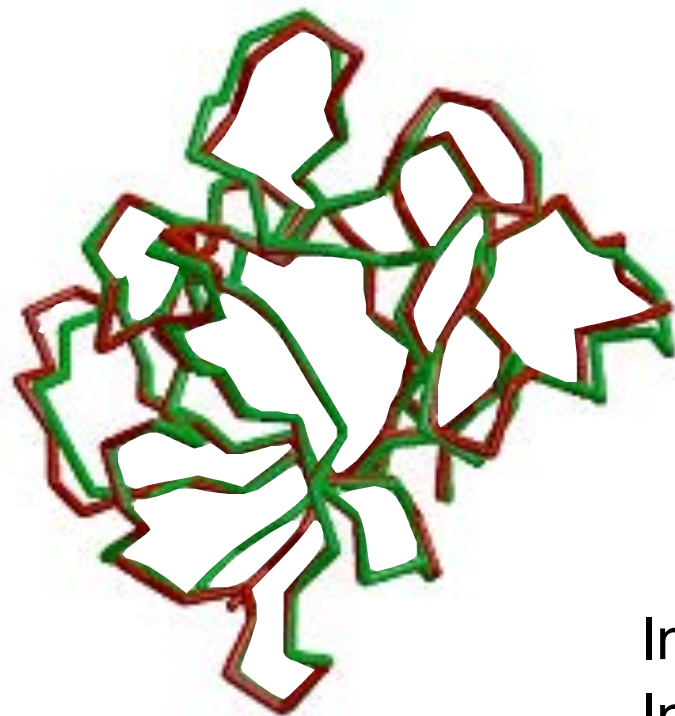**Rigid body distortions**
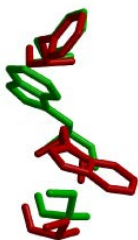
Side-chain packing

Marti-Renom etal. Ann Rev Biophys Biomol Struct (2000) 29, 291
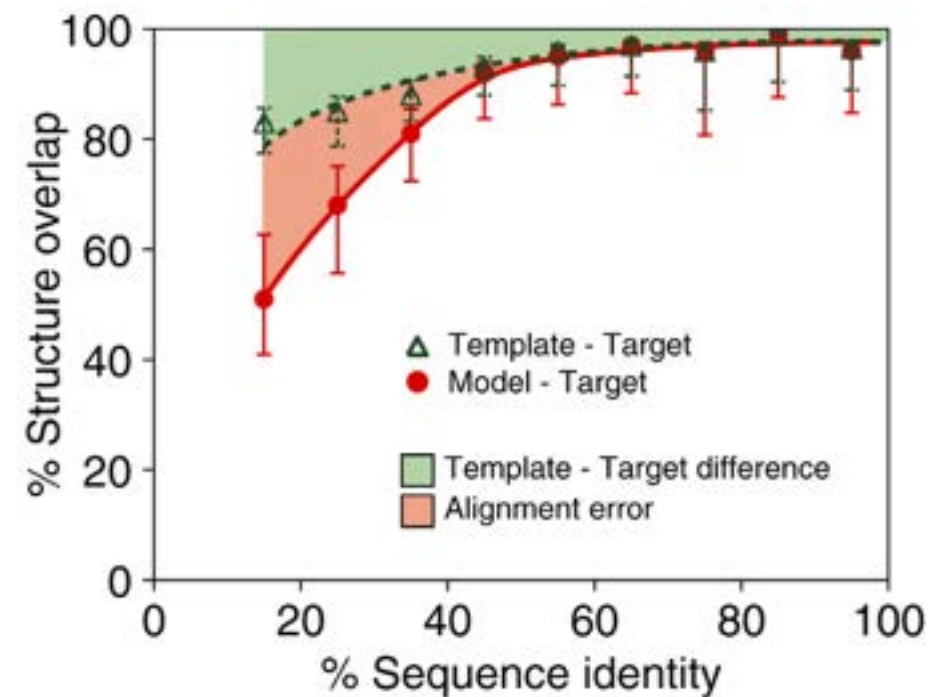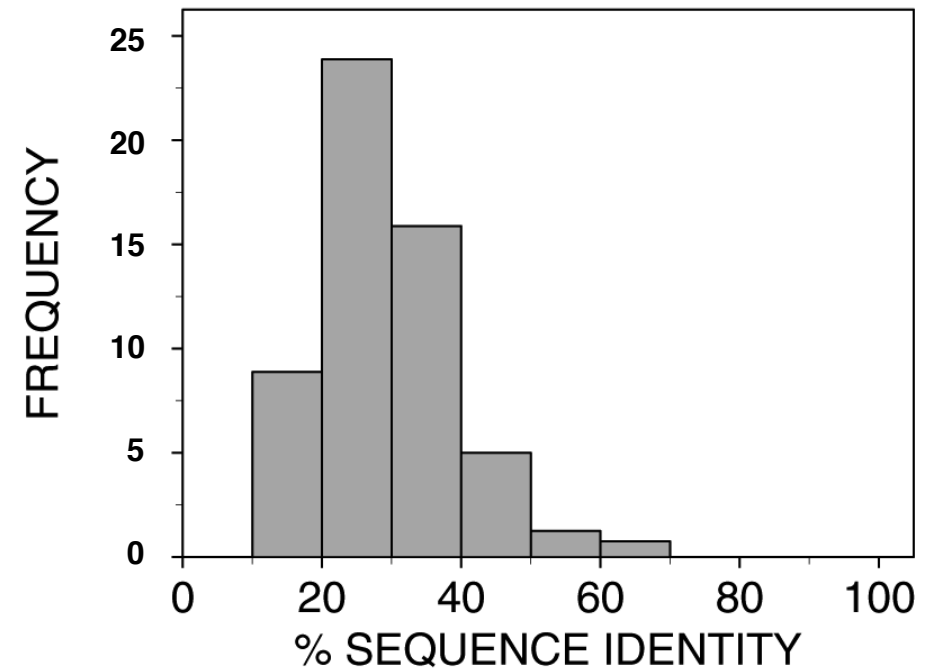
"Biological" significance of modeling errors

NMR – X-RAY
Erabutoxin 3ebx
Erabutoxin 1era
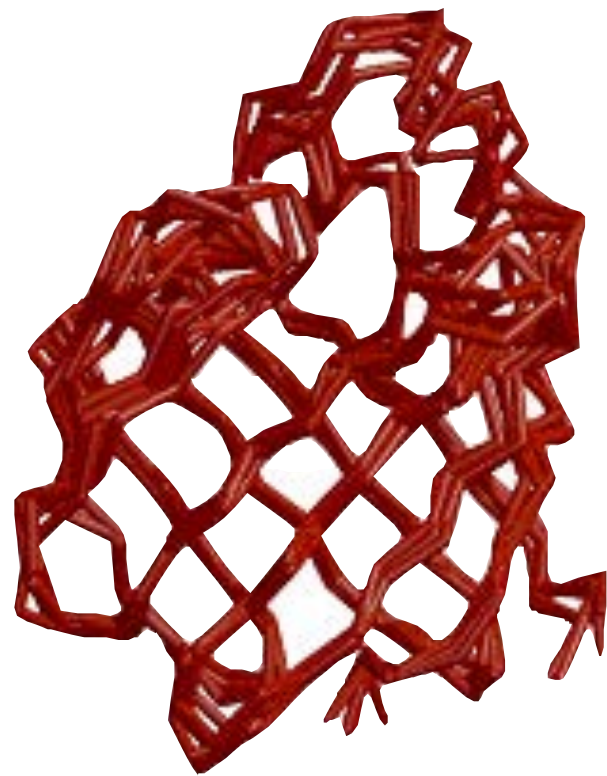
NMR
Ileal lipid-binding protein
1eal
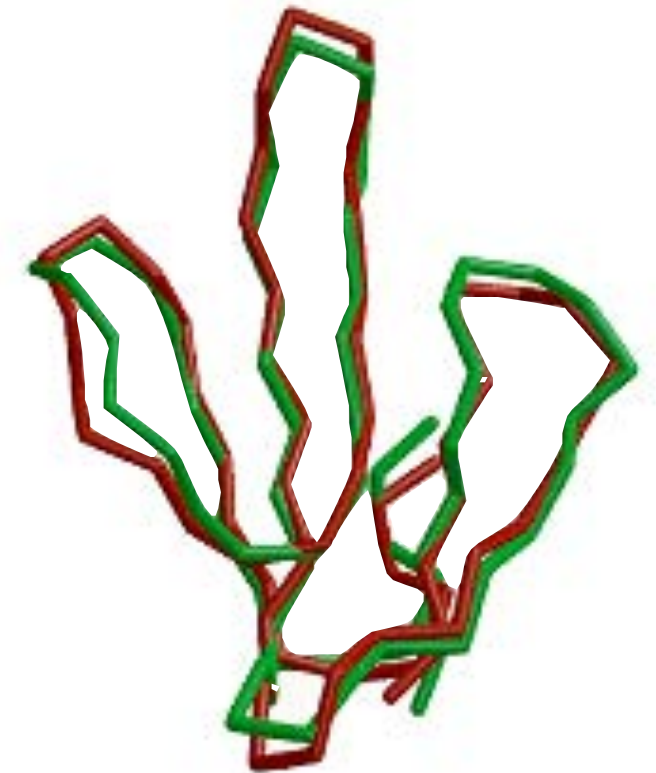
CRABPII 1opbB
FABP   1ftpA
ALBP   1lib
40% seq. id.

X-RAY
Interleukin 1β 41bi (2.9Å)
Interleukin 1β 2mib (2.8Å)
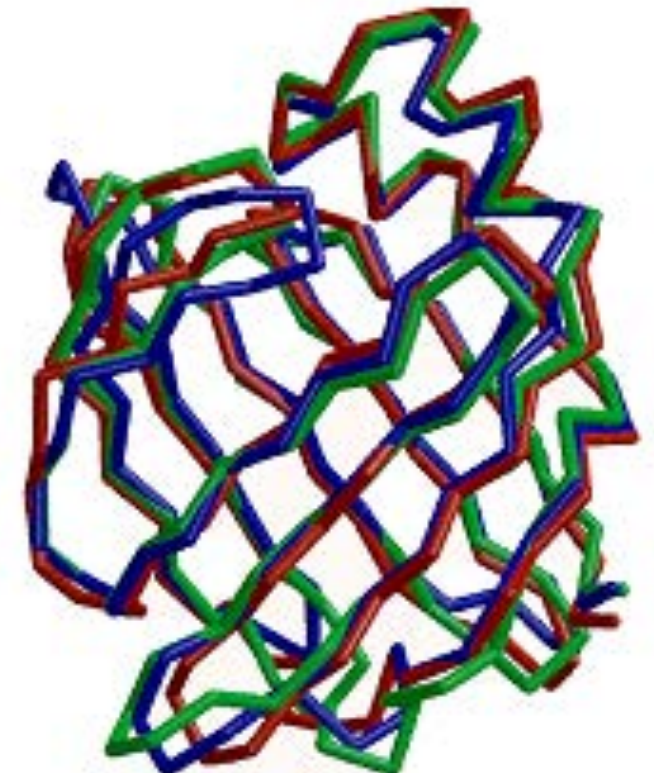
Model Accuracy

HIGH ACCURACY

NM23
Seq id 77%

Cα equiv 147/148
RMSD 0.41Å
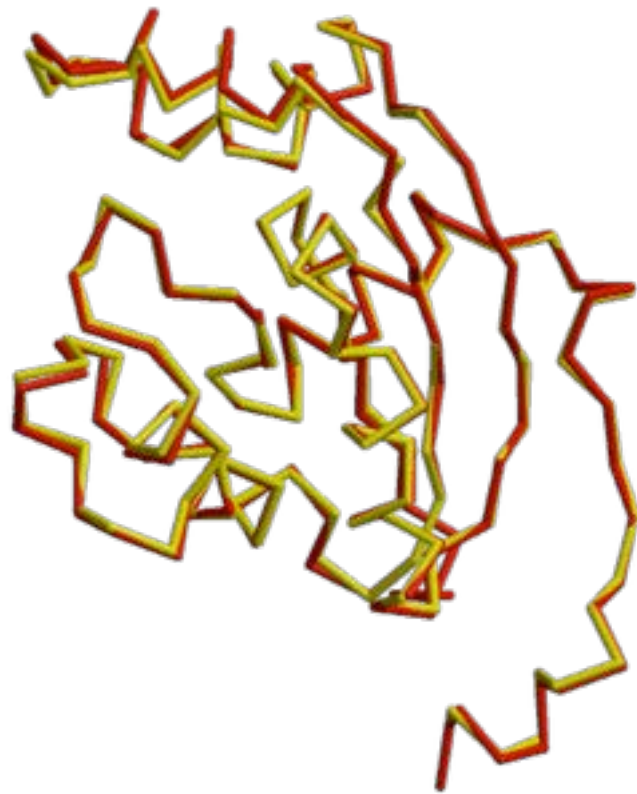
Sidechains
Core backbone
Loops

MEDIUM ACCURACY

CRABP
Seq id 41%

Cα equiv 122/137
RMSD 1.34Å
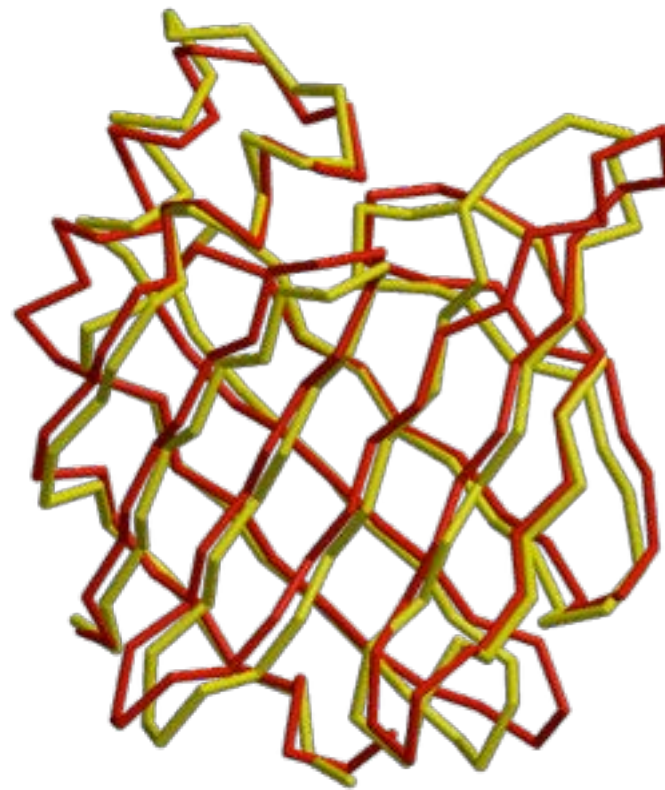
Sidechains
Core backbone
Loops
Alignment

LOW ACCURACY

EDN
Seq id 33%

Cα equiv 90/134
RMSD 1.17Å

Sidechains
Core backbone
Loops
Alignment
Fold assignment

X-RAY / MODEL

Utility of protein structure models, despite errors
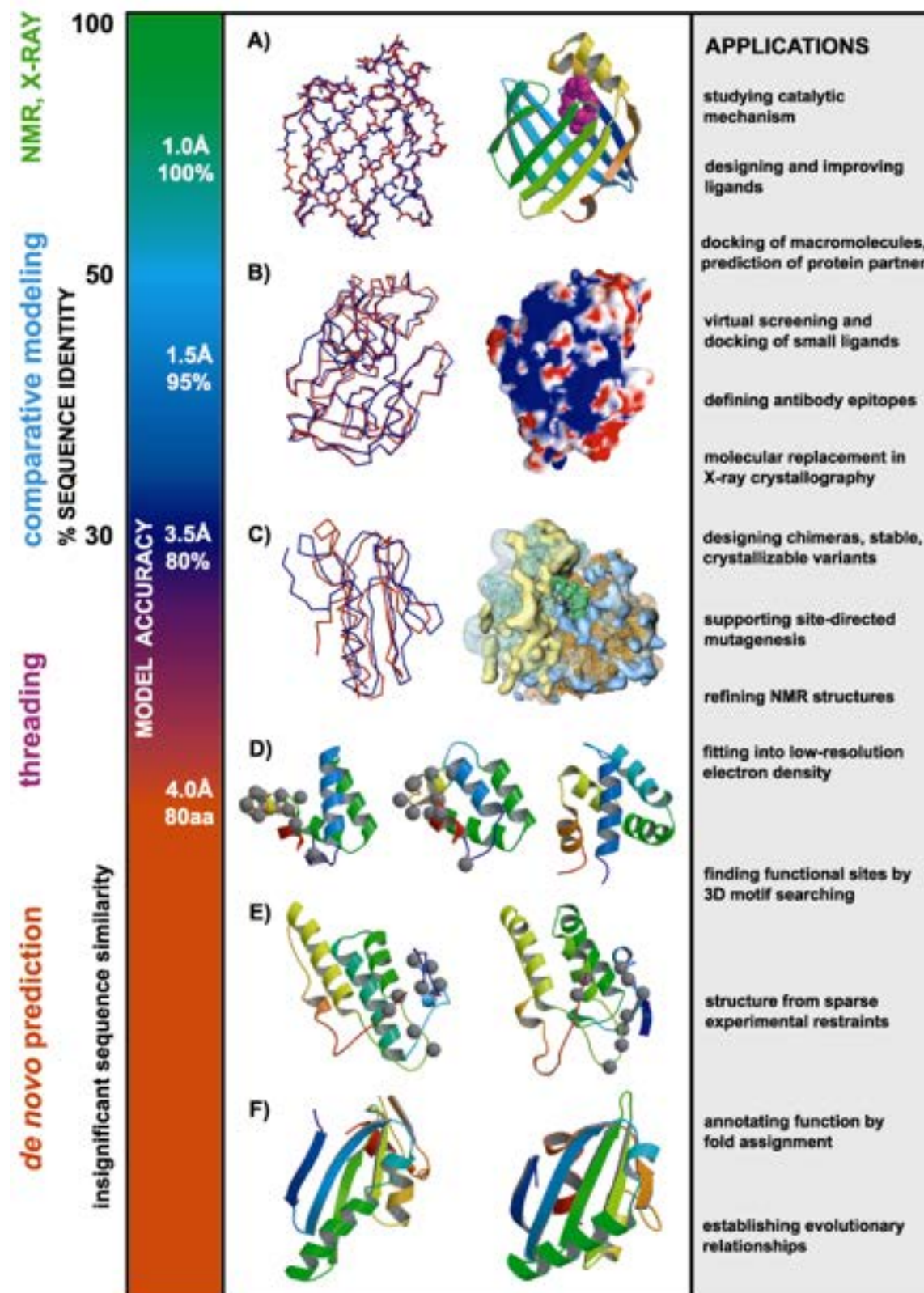
Structural analysis of missense mutations in human BRCA1 BRCT domains

Mirkovic et al. Structure-based assessment of missense mutations in human BRCA1: implications for breast and ovarian cancer predisposition. Cancer Res (2004) vol. 64 (11) pp. 3790-7

Structure-Based Assessment of Missense Mutations in Human BRCA1: Implications for Breast and Ovarian Cancer Predisposition

Nebojsa Mirkovic,[1] Marc A. Marti-Renom,[2] Barbara L. Weber,[3] Andrej Sali,[2] and Alvaro N. A. Monteiro[4,5]

[1]Laboratory of Molecular Biophysics, Pels Family Center for Biochemistry and Structural Biology, Rockefeller University, New York, New York; [2]Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, University of California at San Francisco, San Francisco, California; [3]Abramson Family Cancer Research Institute, University of Pennsylvania, Philadelphia, Pennsylvania; [4]Strang Cancer Prevention Center, New York, New York; and [5]Department of Cell and Developmental Biology, Weill Medical College of Cornell University, New York, New York

ABSTRACT

The *BRCA1* gene from individuals at risk of breast and ovarian cancers can be screened for the presence of mutations. However, the cancer association of most alleles carrying missense mutations is unknown, thus creating significant problems for genetic counseling. To increase our ability to identify cancer-associated mutations in *BRCA1*, we set out to use the principles of protein three-dimensional structure as well as the correlation between the cancer-associated mutations and those that abolish transcriptional activation. Thirty-one of 37 missense mutations of known impact on the transcriptional activation function of BRCA1 are readily rationalized in structural terms. Loss-of-function mutations involve non-conservative changes in the core of the BRCA1 C-terminus (BRCT) fold or are localized in a groove that presumably forms a binding site involved in the transcriptional activation by BRCA1; mutations that do not abolish transcriptional activation are either conservative changes in the core or are on the surface outside of the putative binding site. Next, structure-based rules for predicting functional consequences of a given missense mutation were applied to 57 germ-line BRCA1 variants of unknown cancer association. Such a structure-based approach may be helpful in an integrated effort to identify mutations that predispose individuals to cancer.

INTRODUCTION

Many germ-line mutations in the human *BRCA1* gene are associated with inherited breast and ovarian cancers (1, 2). This information has allowed clinicians and genetic counselors to identify individuals at high risk for developing cancer. However, the disease association of over 350 missense mutations remains unclear, primarily because their relatively low frequency and ethnic specificity limit the usefulness of the population-based statistical approaches to identifying cancer-causing mutations. To address this problem, we use here the three-dimensional structure of the human BRCA1 BRCT domains to assess the transcriptional activation functions of BRCA1 mutants. Our study is made possible by the recently determined sequences (3–6) and three-dimensional structures of the BRCA1 homologs (7, 8). In addition, we benefited from prior studies that attempted to rationalize and predict functional effects of mutations in various proteins (9–12), including those of BRCA1 (13, 14).

BRCA1 is a nuclear protein that activates transcription and facilitates DNA damage repair (15, 16). The tandem BRCT repeat at the COOH-terminus of BRCA1 are involved in several of its functions, including modulation of the activity of several transcription factors (15), binding to the RNA polymerase II holoenzyme (17), and activating transcription of a reporter gene when fused to a heterologous DNA-binding domain (18, 19). Importantly, cancer-associated mutations in the BRCT domains, but not benign polymorphisms, inactivate transcriptional activation and binding to RNA polymerase II (18–21). These observations suggest that abolishing the transcriptional activation function of BRCA1 leads to tumor development and provides a genetic framework for characterization of BRCA1 BRCT variants.

MATERIALS AND METHODS

The multiple sequence alignment (MSA) of orthologous BRCA1 BRCT domains from seven species, including *Homo sapiens* (GenBank accession number U14680), *Pan troglodytes* (AF207822), *Mus musculus* (U68174), *Rattus norvegicus* (AF036760), *Gallus gallus* (AF355273), *Canis familiaris* (U50709), and *Xenopus laevis* (AF416868), was obtained by using program ClustalW (22) and contains only one gapped position (Supplementary Fig. 1). According to PSI-BLAST (23), the latter six sequences are the only sequences in the nonredundant protein sequence database at National Center for Biotechnology Information that have between 30% and 90% sequence identity to the human BRCA1 BRCT domains (residues 1649–1859).

The multiple structure-based alignment of the native structures of the BRCT-like domains was obtained by the SALIGN command in MODELLER (Supplementary Fig. 2). It included the experimentally determined structures of the two human BRCA1 BRCT domains (Protein Data Bank code 1JNX; Refs. 8, 24), rat BRCA1 BRCT domains (1L0B; Ref. 7), human p53-binding protein (1KZY; Ref. 7), human DNA-ligase IIIα (1IMO; Ref. 25), and human XRCC1 protein (1CDZ; Ref. 13). Structure variability was defined by the root-mean-square deviation among the superposed Cα positions, as calculated by the COMPARE command of MODELLER. The purpose of these calculations was to gain insight into the variability of surface-exposed residues (*left panel* in Fig. 2). In conjunction with observed mutation clustering, these data may point to putative functional site(s) on the surface of BRCT repeats.

Comparative protein structure modeling by satisfaction of spatial restraints, implemented in the program MODELLER-6 (26), was used to produce a three-dimensional model for each of the 94 mutants. The crystallographic structure of the human wild-type BRCA1 BRCT domains was used as the template for modeling (8). The four residues missing in the crystallographic structure (1694 and 1817–1819) were modeled *de novo* (27). All of the models are available in the BRCA1 model set deposited in our ModBase database of comparative protein structure models (28).[6]

For the native structure of the human BRCT tandem repeat and each of the 94 mutant models, a number of sequence and structure features were calculated. These features were used in the classification tree in Fig. 3 (values for all 94 mutants are given in Supplementary Tables 1 and 2).

Buriedness. Accessible surface area of an amino acid residue was calculated by the program DSSP (29) and normalized by the maximum accessible surface area for the corresponding amino acid residue type. A residue was considered exposed if its accessible surface area was larger than 40Å^2 and if its relative accessible surface area was larger than 9% and buried otherwise. A mutation of a more exposed residue is less likely to change the structure and therefore its function.

Human BRCA1 and its two BRCT domains
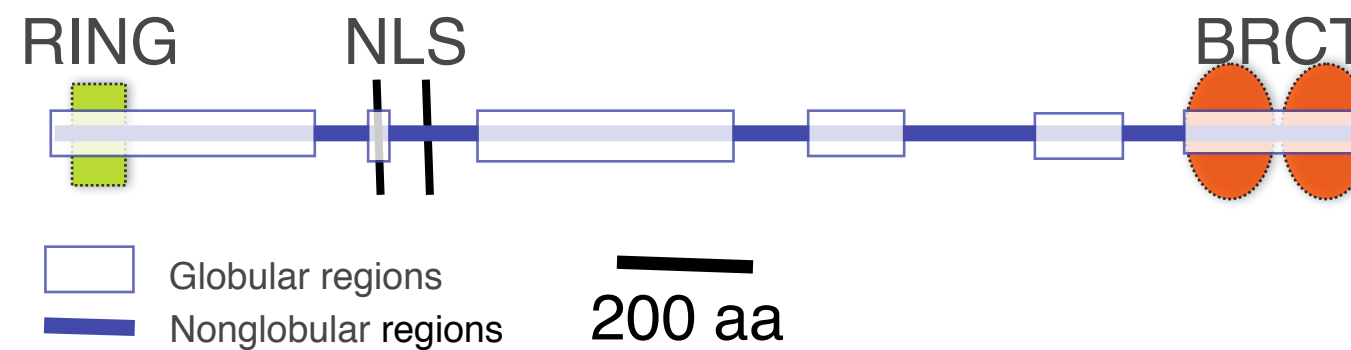
BRCA1 BRCT repeats, 1jnx

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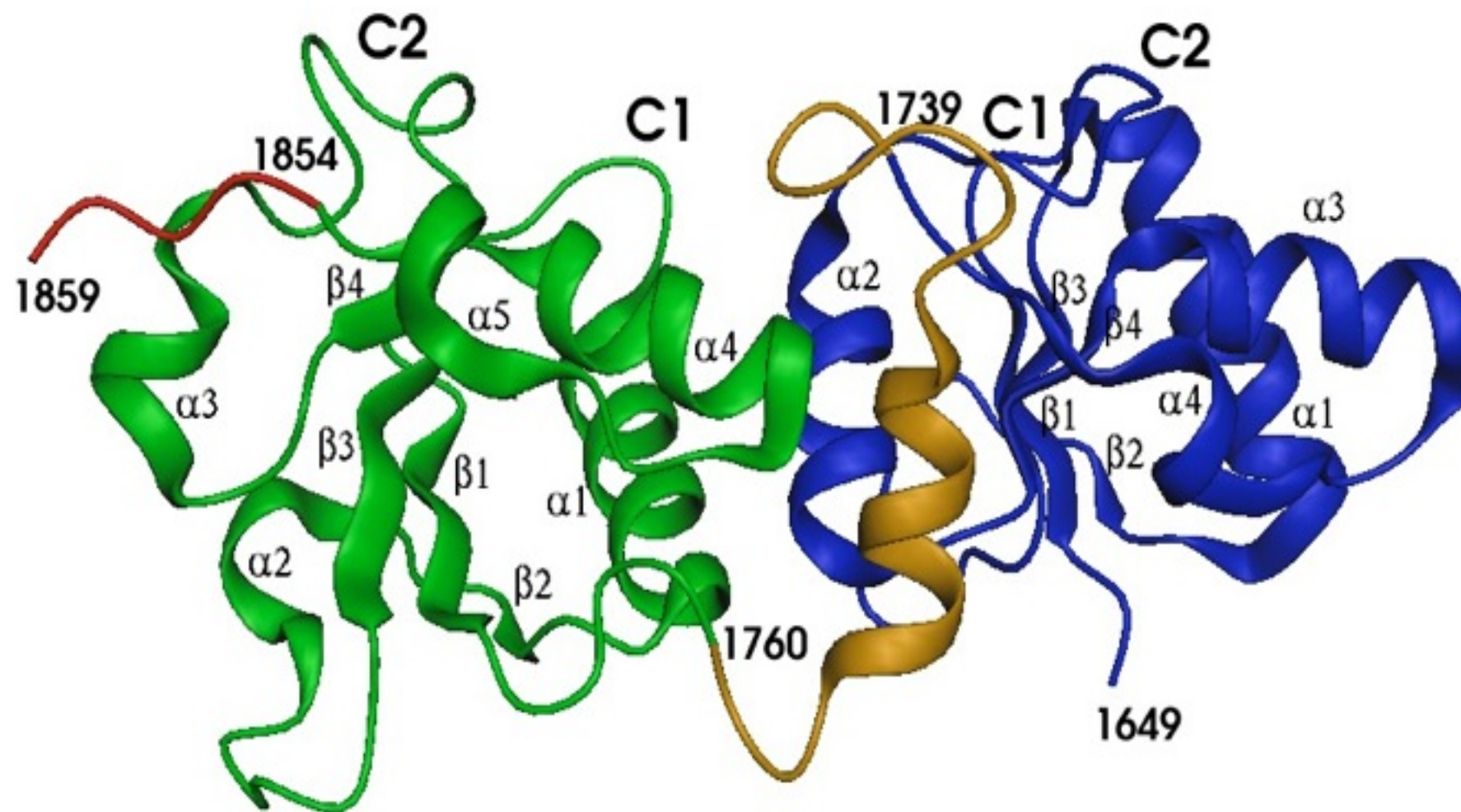
MYRIAD

BRACAnalysis ™

Comprehensive BRCA1-BRCA2 Gene Sequence Analysis Result

	SPECIMEN	PATIENT
Niecee Singer, MS Strang Cancer Prevention Center 428 E 72nd St New York, NY 10021	Specimen Type: Blood Draw Date: n/a Accession Date: Oct 27, 2000 Report Date: Nov 17, 2000	Name: Date of Birth: Feb 02, 1953 Patient ID: Gender: Female Accession #: 00019998 Requisition #: 56594

Physician: Fred Gilbert, MD

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:

Brian E. Ward, Ph.D.
Laboratory Director

Thomas S. Frank, M.D.
Medical Director

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 RA1752P F1761I		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

"Decision" tree for predicting functional impact of genetic variants

17
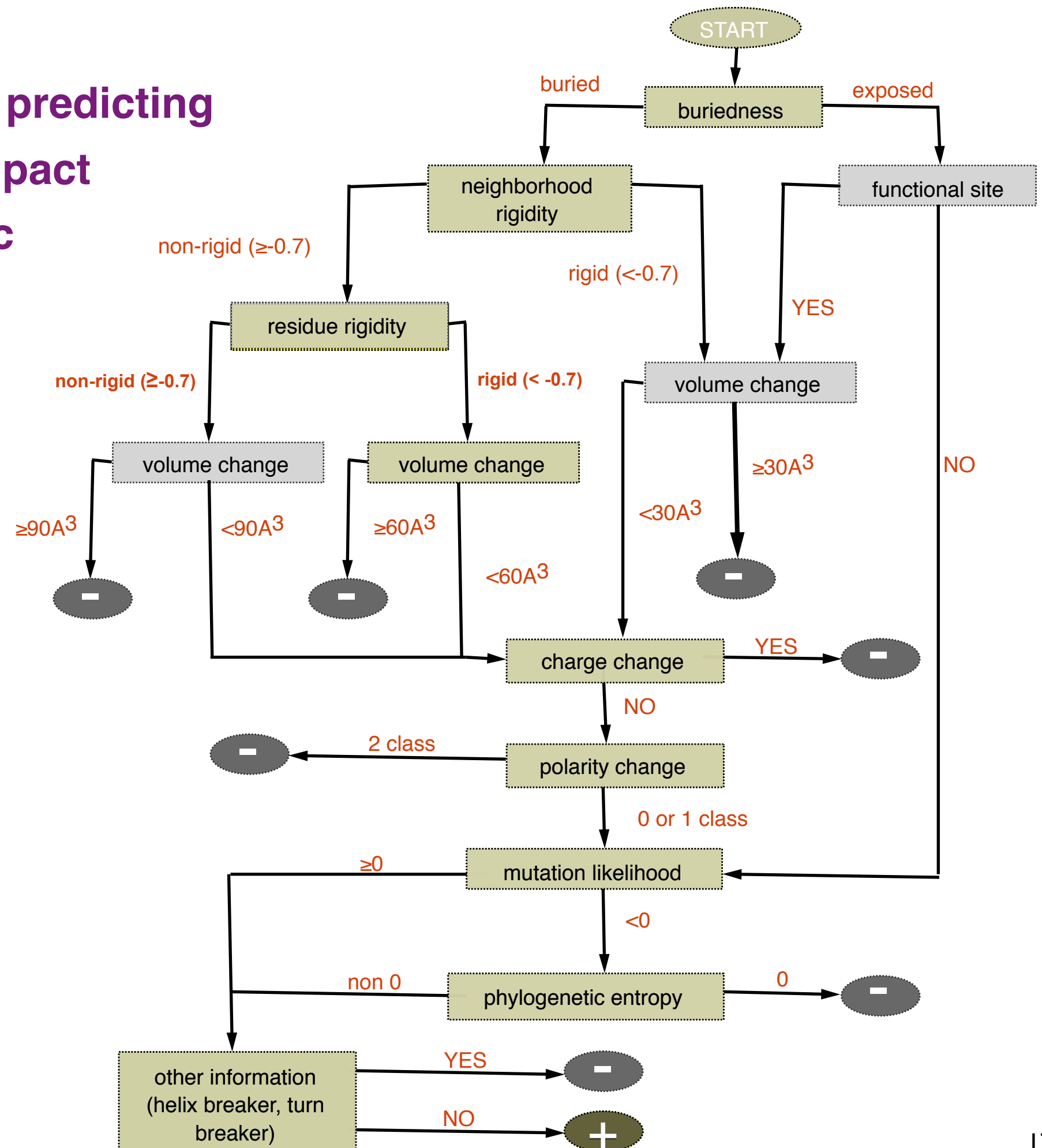
Putative binding site on BRCA1
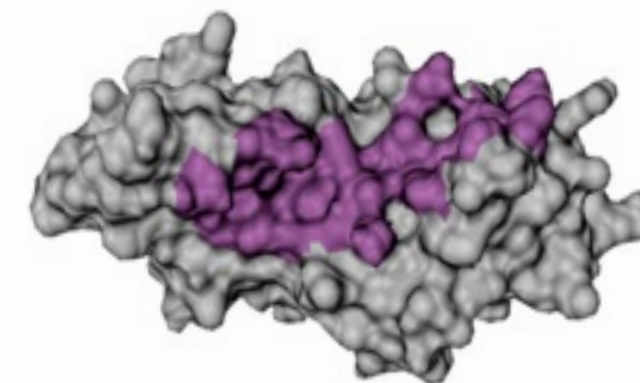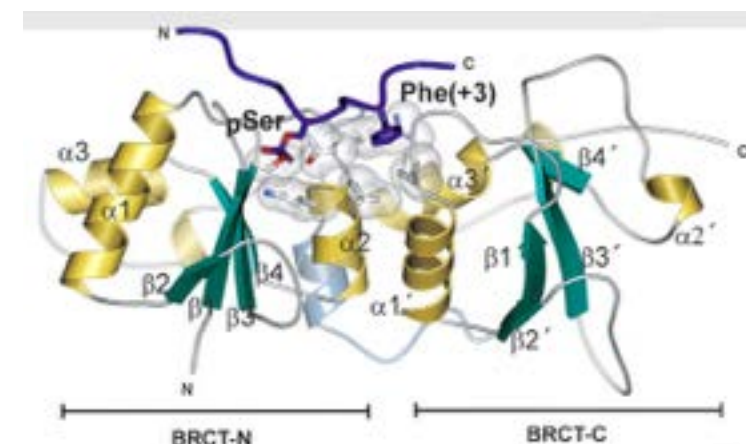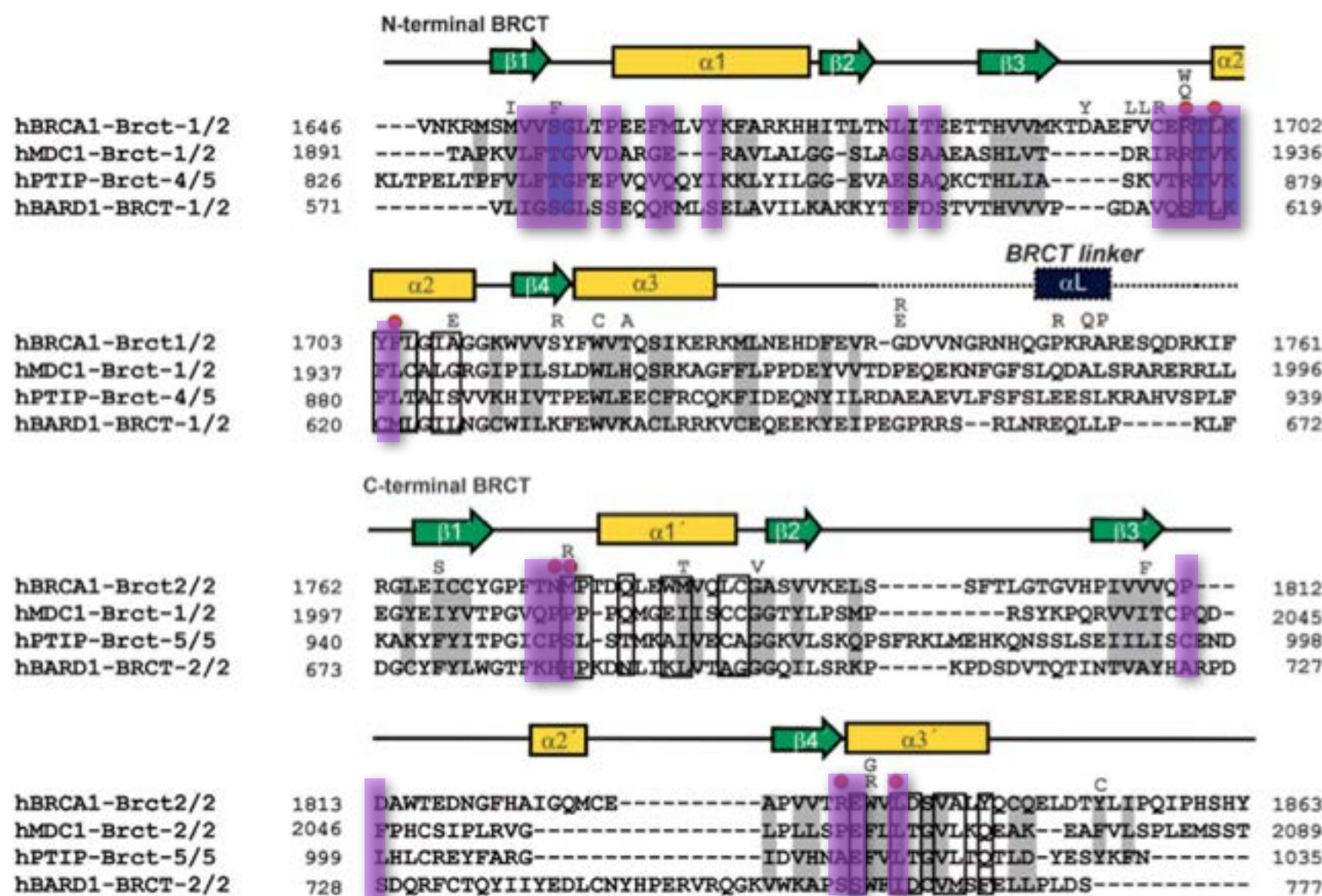
Putative binding site predicted in 2003 and accepted for publication on March 2004.

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

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

Common Evolutionary Origin of Coated Vesicles and Nuclear Pore Complexes

mGenThreader + SALIGN + MOULDER

yNup84 complex proteins
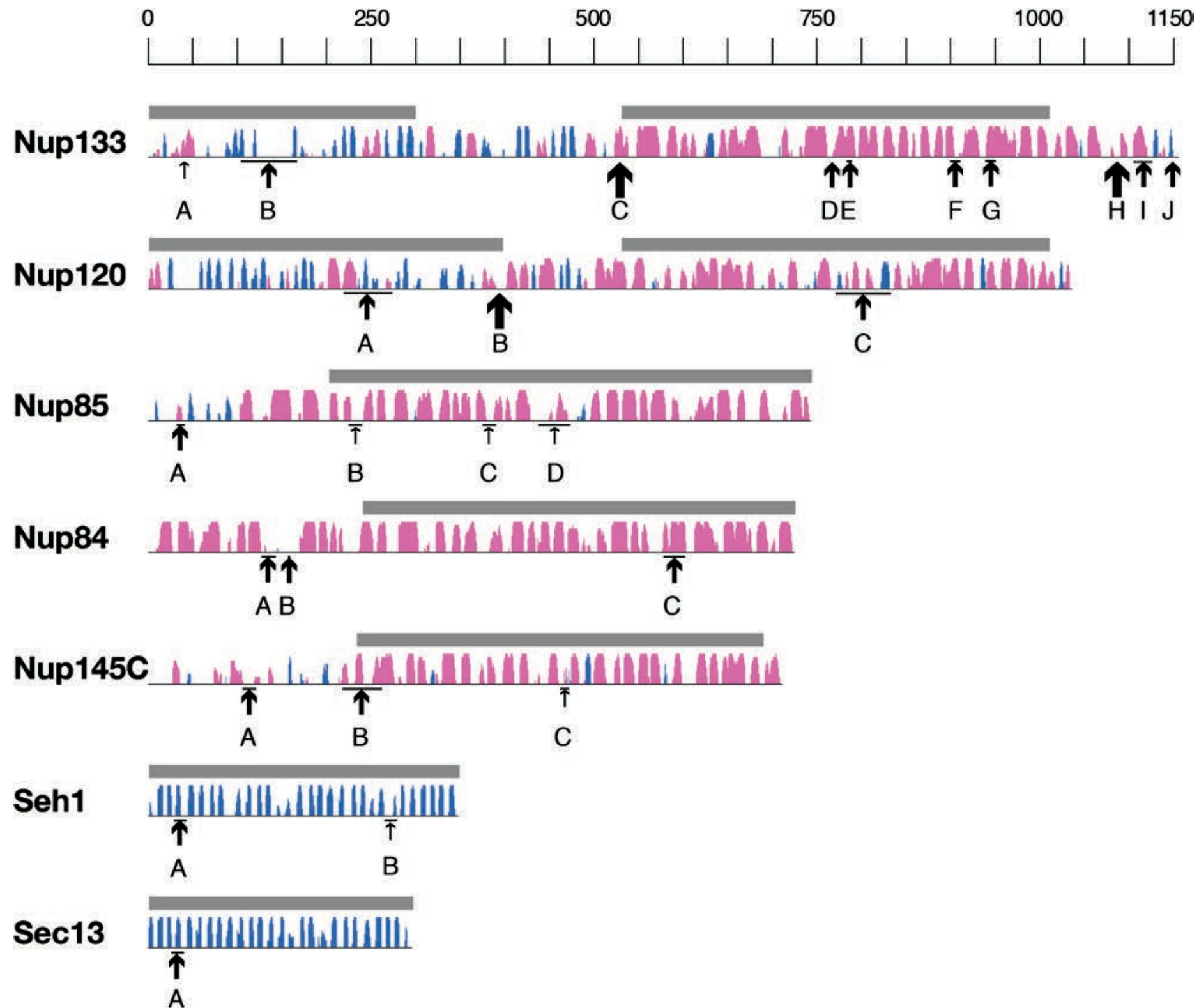
All Nucleoporins in the Nup84 Complex are Predicted to Contain β-Propeller and/or α-Solenoid Folds
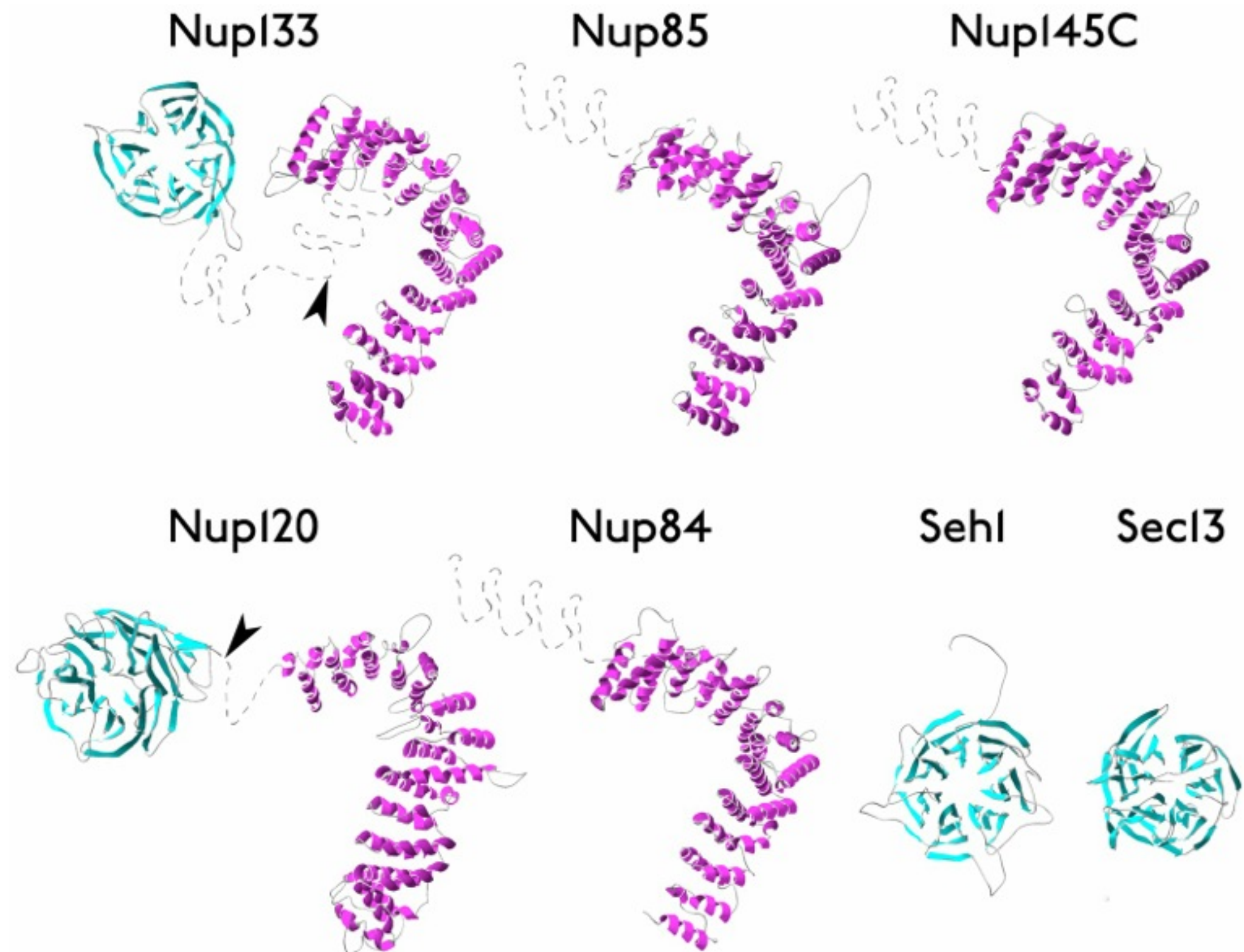
NPC and Coated Vesicles Share the β-Propeller and α-Solenoid Folds and Associate with Membranes
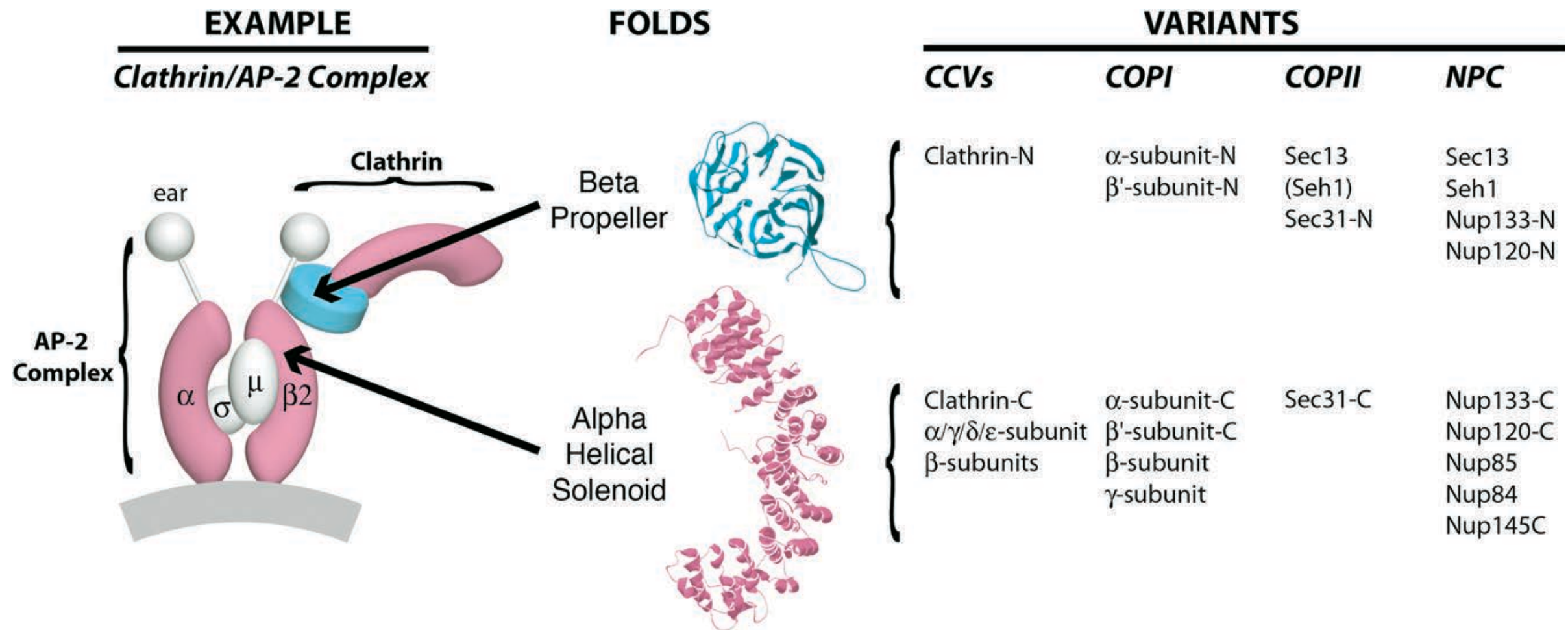
NPC and Coated Vesicles Both Associate with Membranes

Coated Vesicle

NPC model
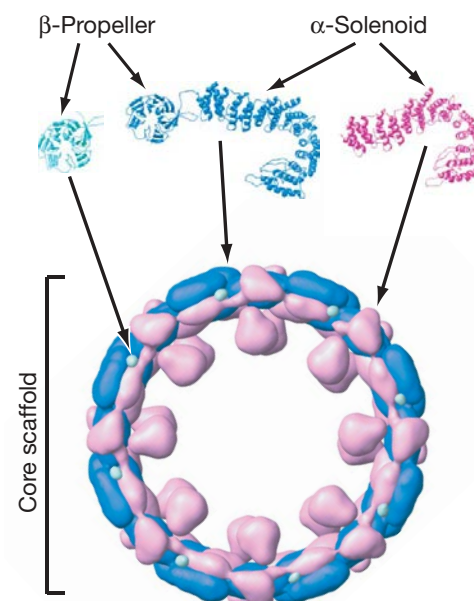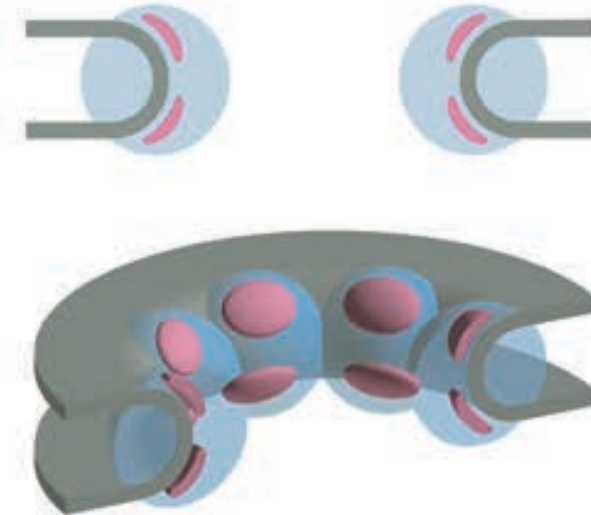
Nup 84 complex

1 Nup192, 2 Nup188, 3 Nup170, 4 Nup157, 5 Nup133,
6 Nup120, 7 Nup85, 8 Nup84, 9 Nup145C, 10 Seh1, 11 Sec13

β-Propeller   α-Solenoid

Core scaffold

Nuclear envelope

55°   35°

5 nm
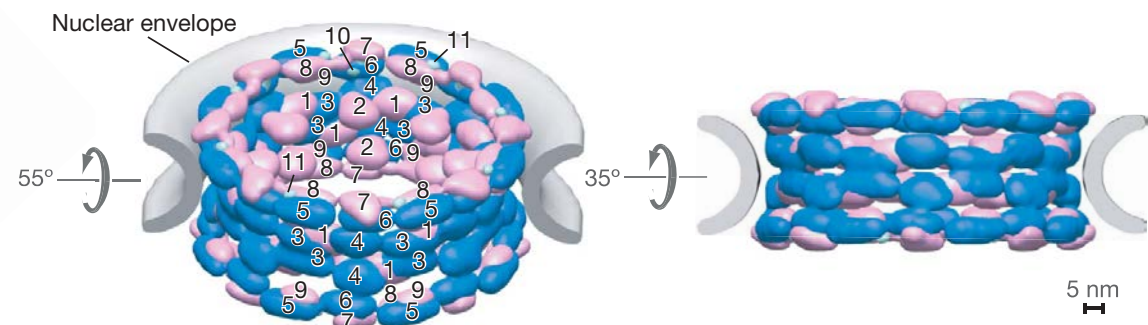
A Common Evolutionary Origin for Nuclear Pore Complexes and Coated Vesicles?
The proto-coatomer hypothesis
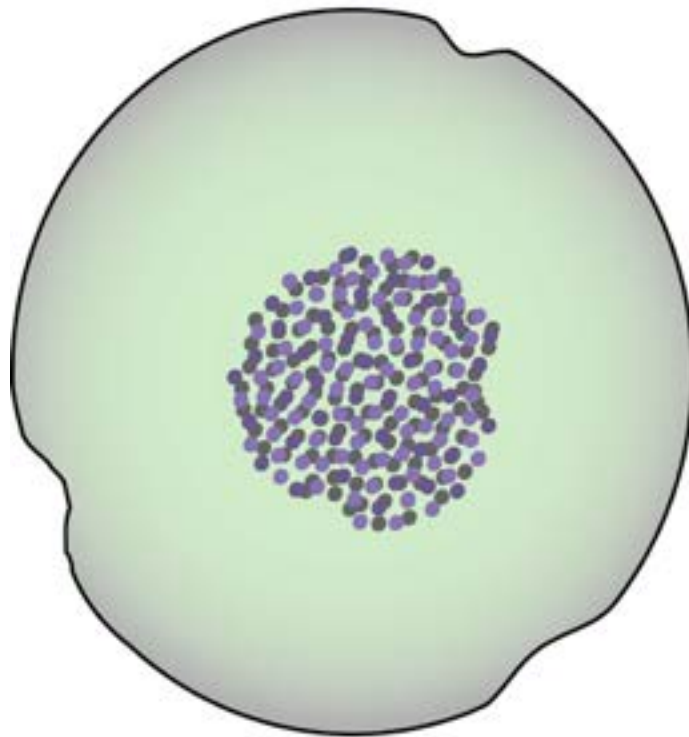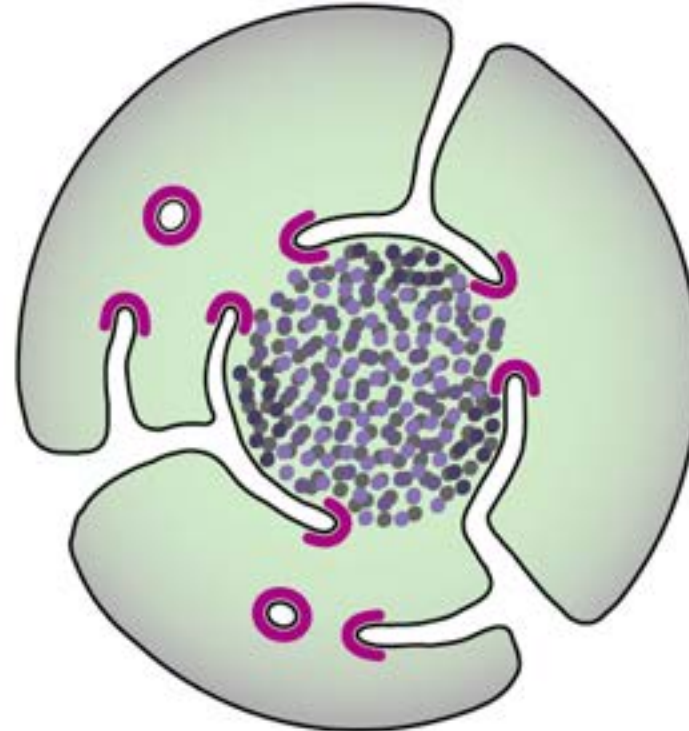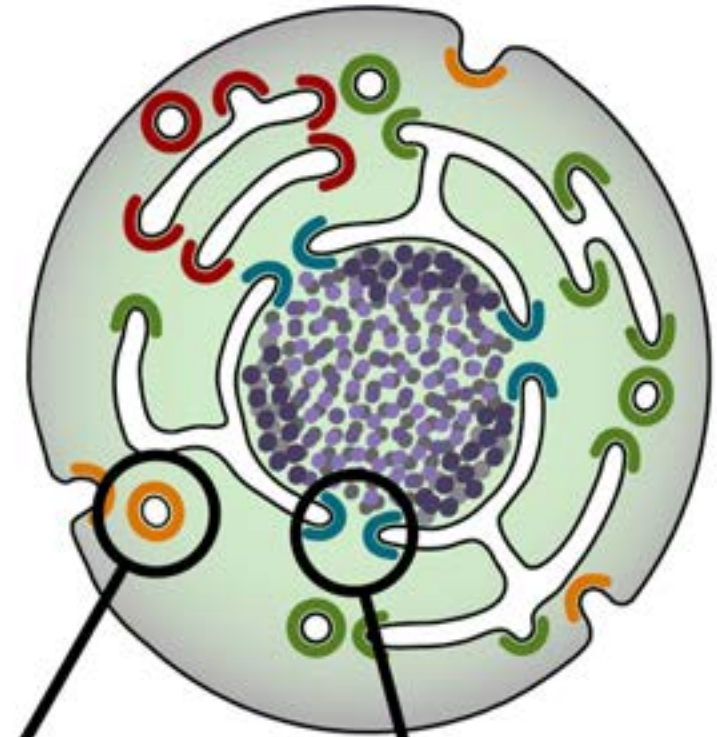
A simple coating module containing minimal copies of the two conserved folds evolved in proto-eukaryotes to bend membranes.

The progenitor of the NPC arose from a membrane-coating module that wrapped extensions of an early ER around the cell's chromatin.
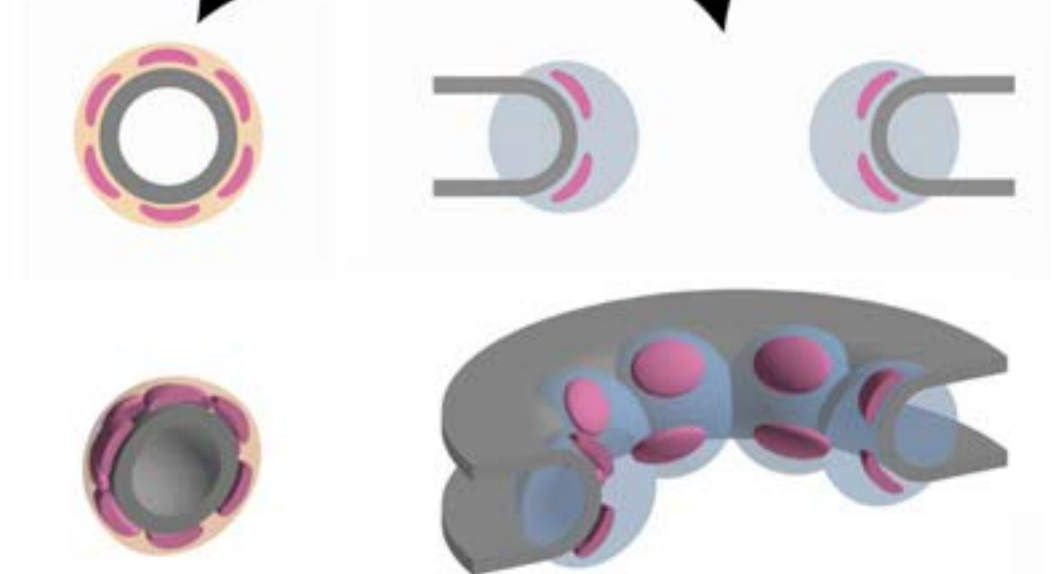
Tropical Disease Initiative (TDI)
Predicting binding sites in protein structure models.

http://www.tropicaldisease.org
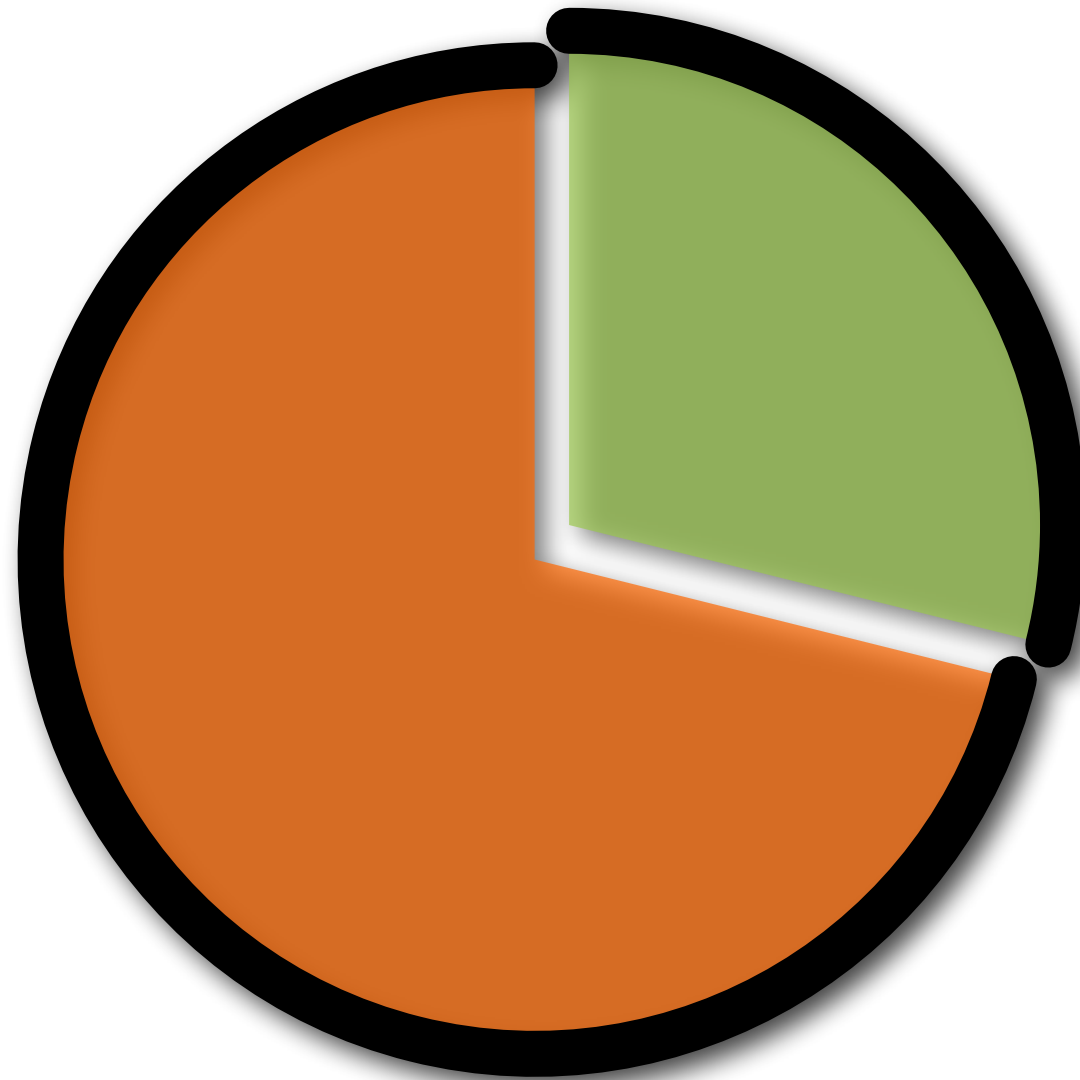
Need is High in the Tail

- DALY Burden Per Disease in Developed Countries
- DALY Burden Per Disease in Developing Countries

Heart diseases

Rare diseases

DALY

Disease

"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.

Comparative docking

Expansion

co-crystalized protein/ligand
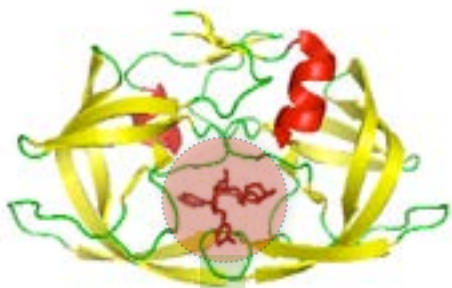
crystalized protein

2. Inheritance

model
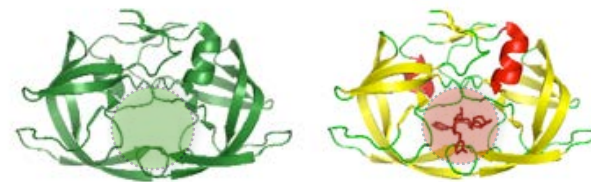
template
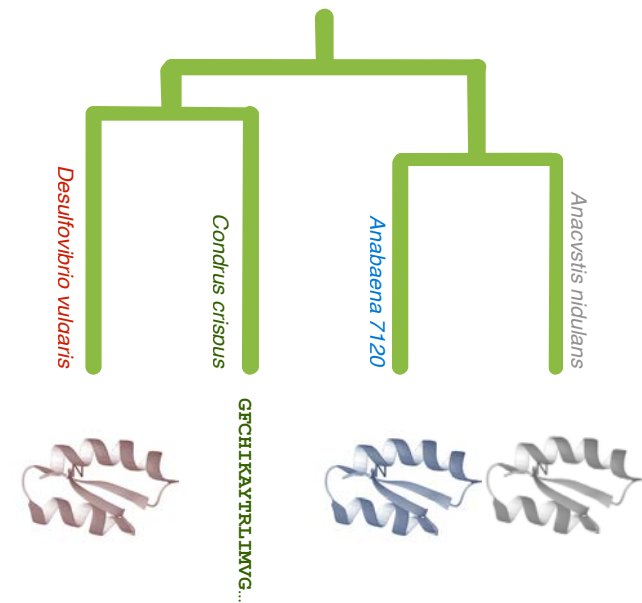
1. Modeling

Modeling Genomes

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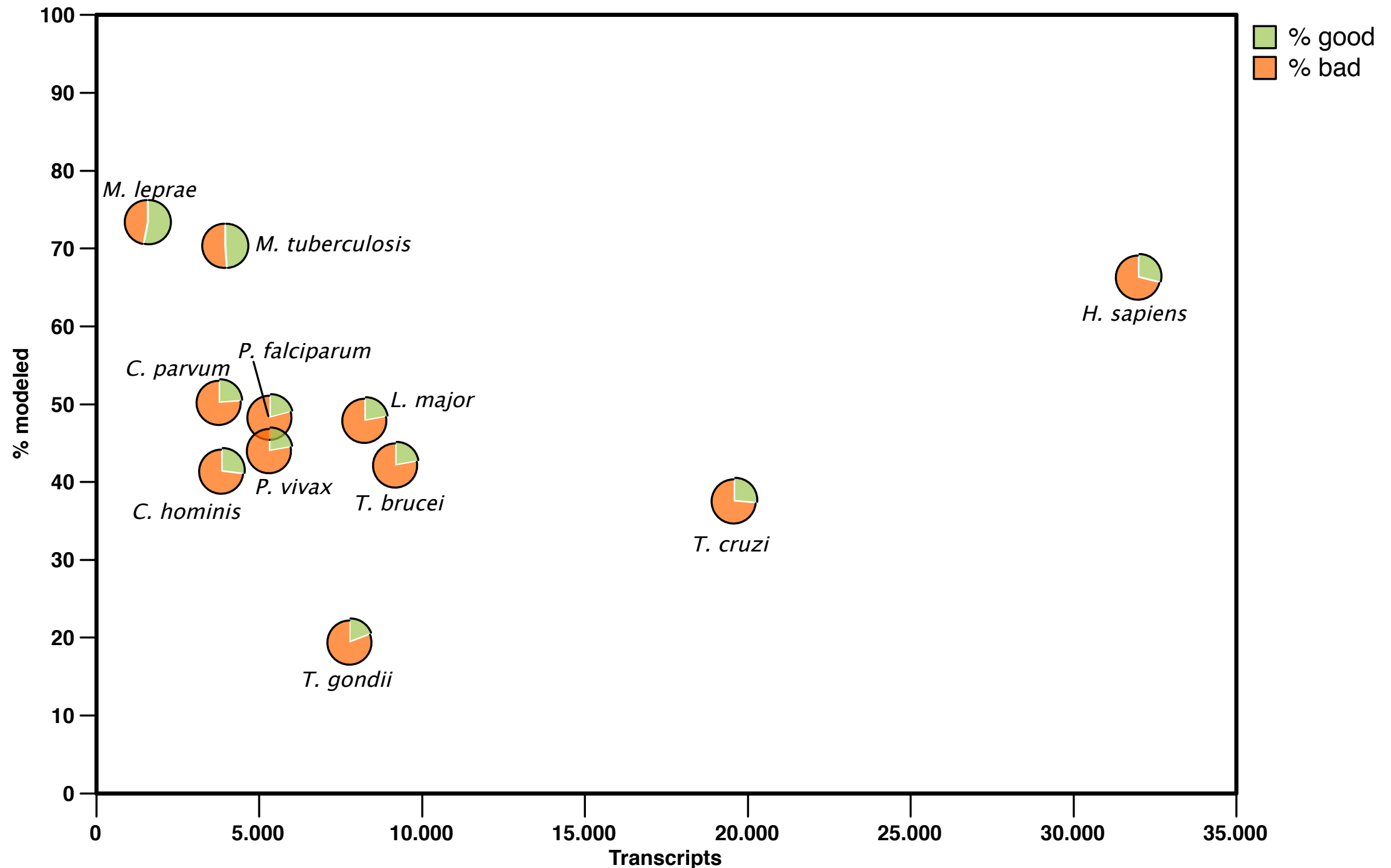
% good
% bad

A good model has MPQS of 1.0 or higher

Summary table
models with inherited ligands

29,271 targets with good models, 297 inherited a ligand/substance similar to a known drug in DrugBank

	Transcripts	Modeled targets	Selected models	Inherited ligands	Similar to a drug	Drugs
C. hominis	3,886	1,614	666	197	20	13
C. parvum	3,806	1,918	742	232	24	13
L. major	8,274	3,975	1,409	478	43	20
M. leprae	1,605	1,178	893	310	25	6
M. tuberculosis	3,991	2,808	1,608	365	30	10
P. falciparum	5,363	2,599	818	284	28	13
P. vivax	5,342	2,359	822	268	24	13
T. brucei	7,793	1,530	300	138	13	6
T. cruzi	19,607	7,390	3,070	769	51	28
T. gondii	9,210	3,900	1,386	458	39	21
TOTAL	**68,877**	**29,271**	**11,714**	**3,499**	**297**	**143**

L. major Histone deacetylase 2 + Vorinostat

Template 1t64A a human HDAC8 protein.
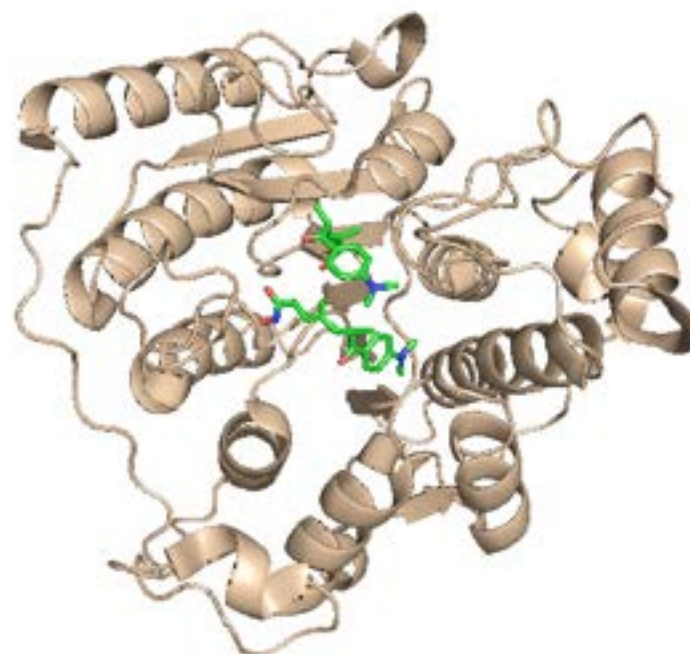
PDB	EO	Template	𝄞	Model	G→	Ligand	Exact	SupStr	SubStr	Similar
1c3sA	83.33/80.00	1t64A	36.00/1.47	LmjF21.0680.1.pdb	90.91/100.00	SHH	DB02546	DB02546	DB02546	DB02546
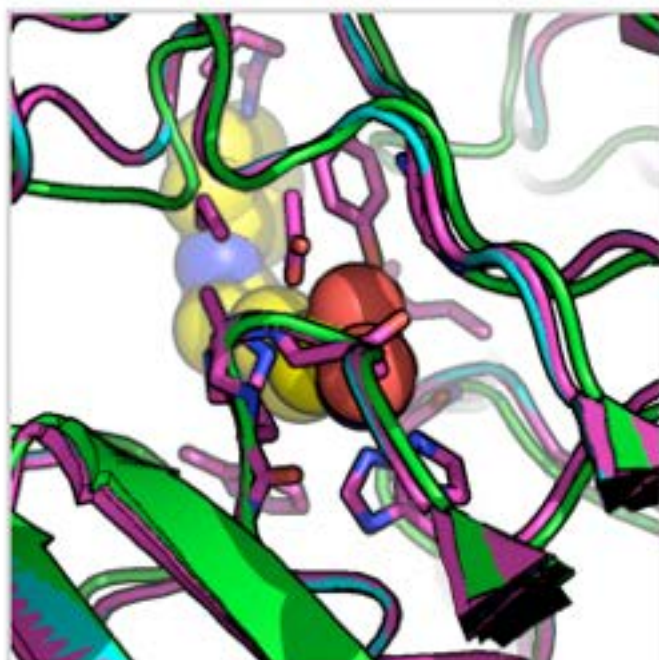
DB02546 Vorinostat

Small Molecule; Approved; Investigational

Drug categories:

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

Drug indication:

For the treatment of cutaneous manifestations in patients with cutaneous T-cell lymphoma who have progressive, persistent or recurrent disease on or following two systemic therapies.

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

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

P. falciparum tymidylate kinase + zidovudine

Template 3tmkA a yeast tymidylate kinase.
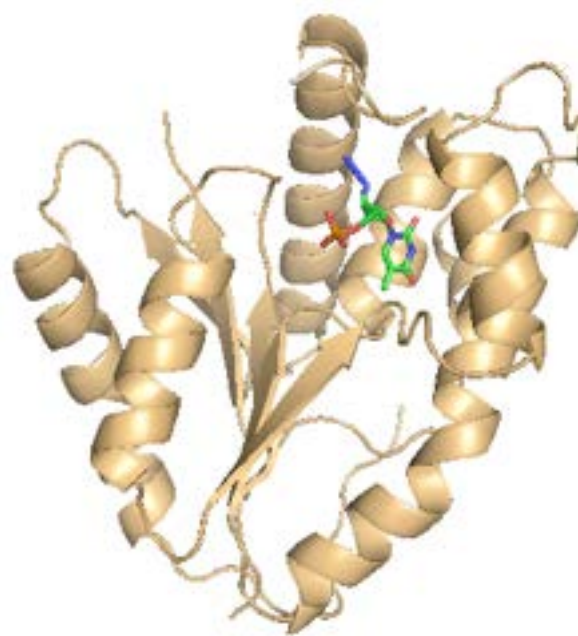
PDB	(O	Template	000	Model	G→	Ligand	Exact	SupStr	SubStr	Similar
2tmkB	100.00/100.00	3tmkA	41.00/1.49	PFL2465c.2.pdb	82.61/100.00	ATM		DB00495		DB00495
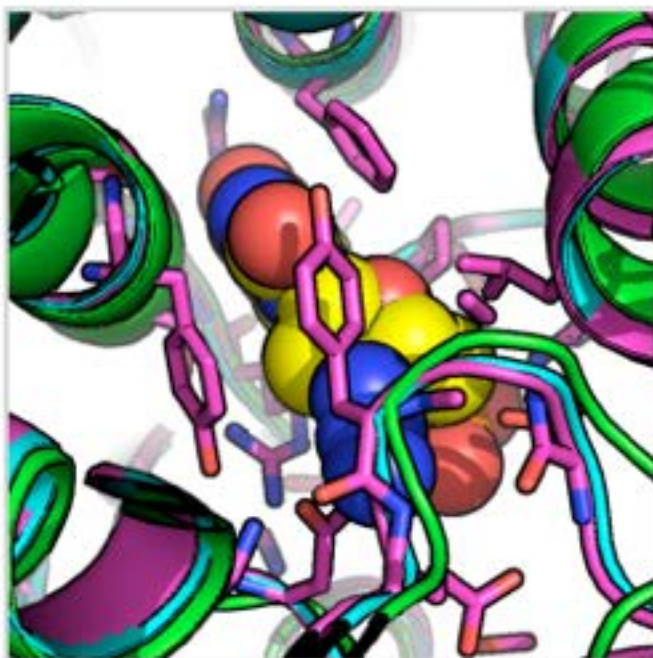
DB00495 Zidovudine

Small Molecule; Approved

Drug categories:

Anti-HIV Agents

Antimetabolites

Nucleoside and Nucleotide Reverse Transcriptase Inhibitors

Drug indication:

For the treatment of human immunovirus (HIV) infections.
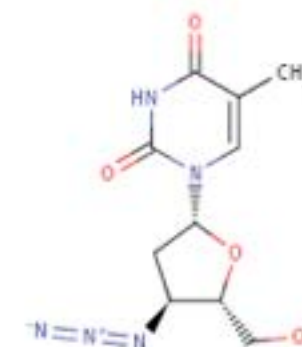
P. falciparum tymydilate kinase + zidovudine

NMR *Water-LOGSY* and *STD* experiments

Leticia Ortí, Rodrigo J. Carbajo, and Antonio Pineda-Lucena

TDI's kernel

http://tropicaldisease.org/kernel

TDI's kernel

http://tropicaldisease.org/kernel

L. Orti *et al.*, *Nat Biotechnol* **27**, 320 (Apr, 2009).

L. Orti *et al.*, *PLoS Negl Trop Dis* **3**, e418 (2009).

TCAMS-TB Target identification
Sali, Brown, Overington & Marti-Renom labs
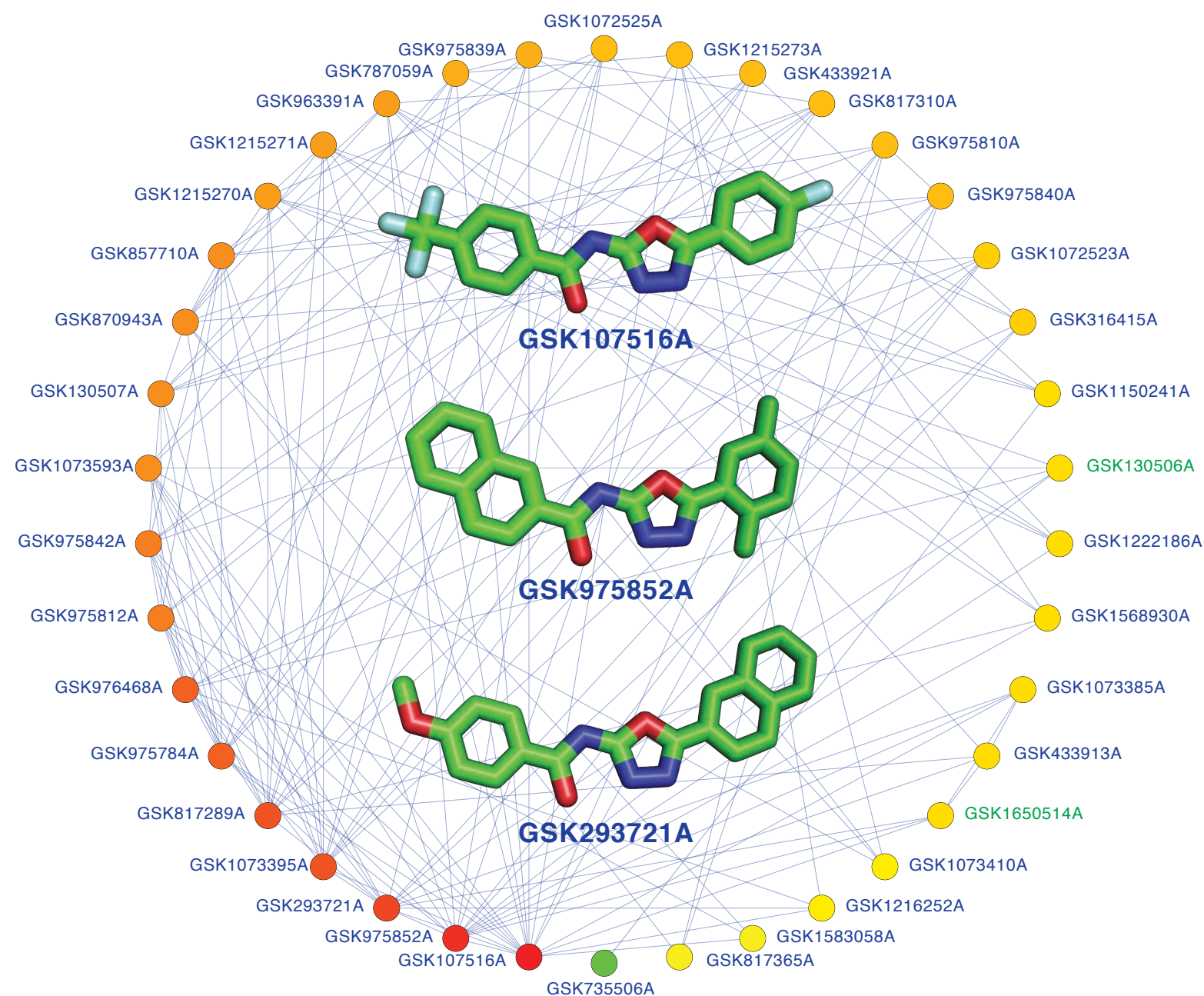
the TCAMS-TB dataset

1. Ballell, L. *et al.* Fueling Open-Source Drug Discovery: 177 Small-Molecule Leads against Tuberculosis. *ChemMedChem* n/a–n/a (2013). doi:10.1002/cmdc.201200428

the TCAMS-TB dataset

1. Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S. & Hopkins, A. L. Quantifying the chemical beauty of drugs. 1–9 (2012). doi:10.1038/nchem.1243
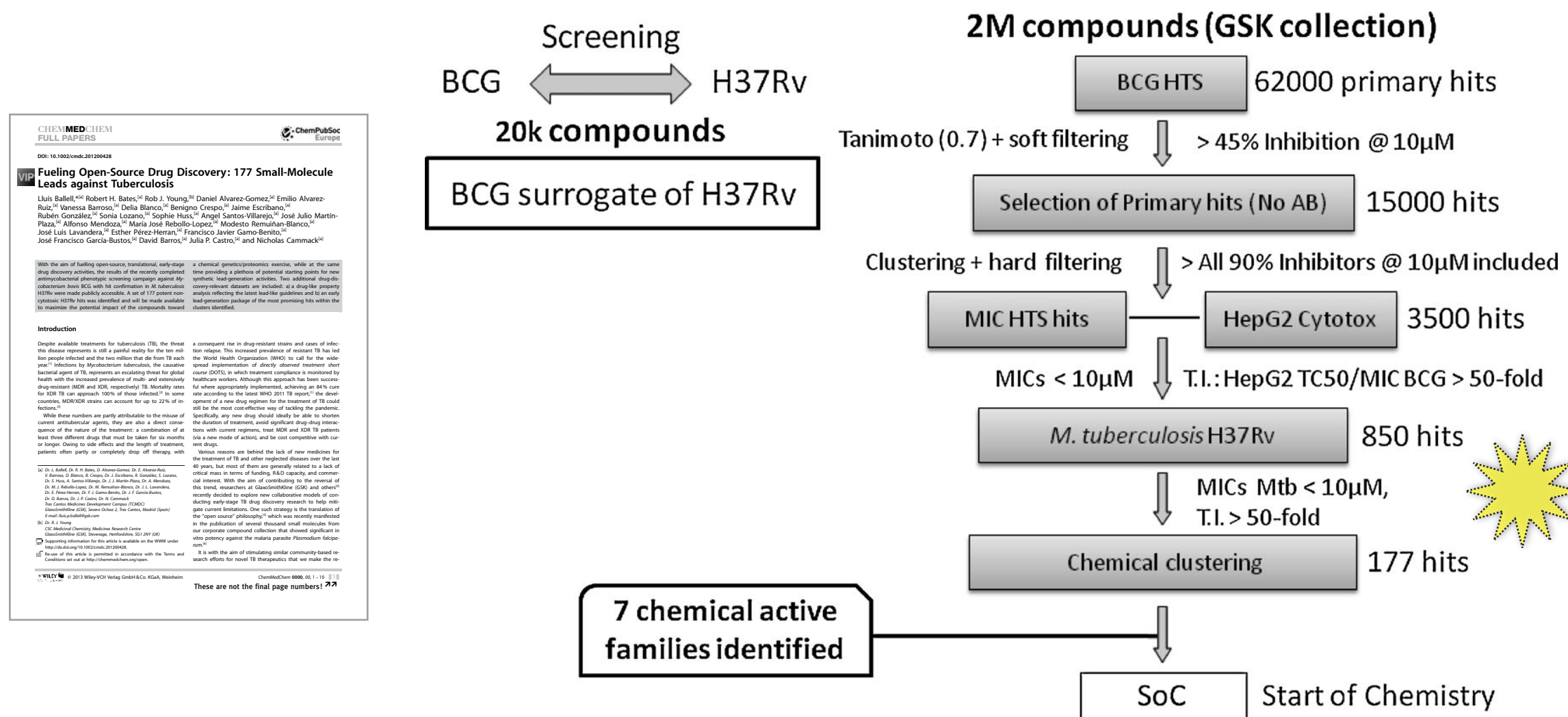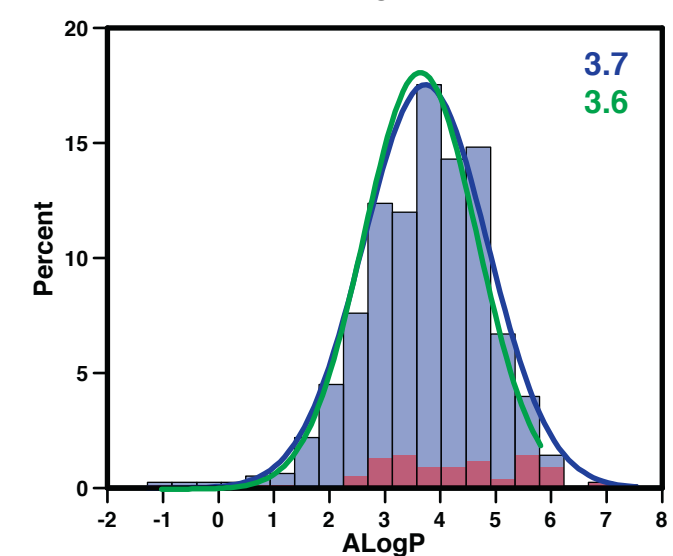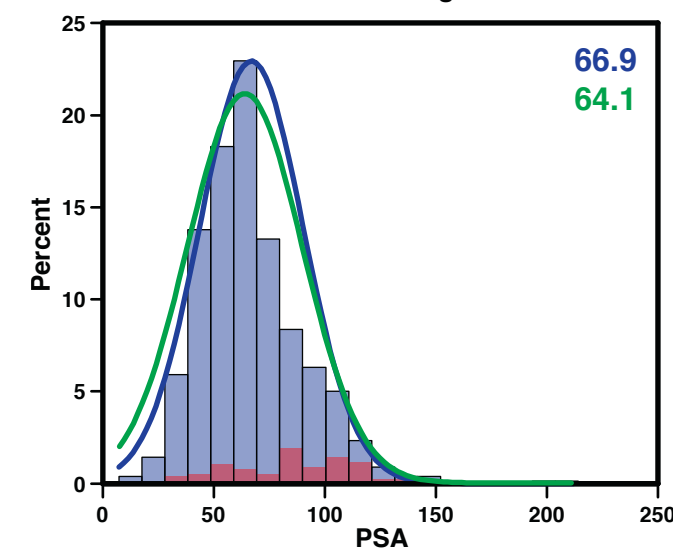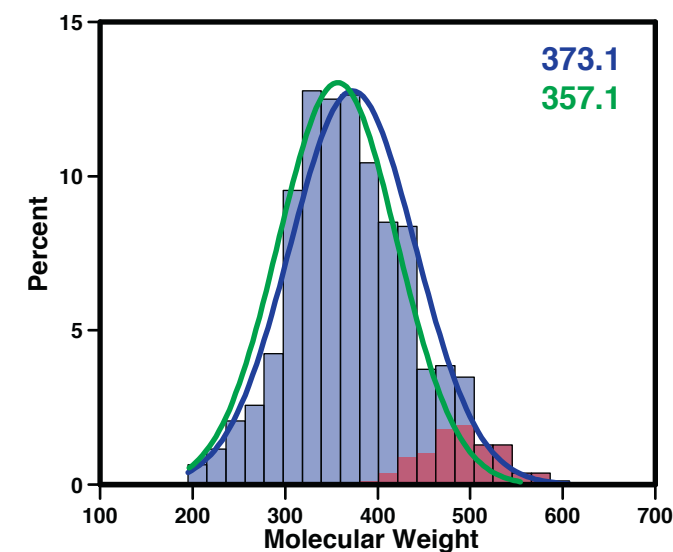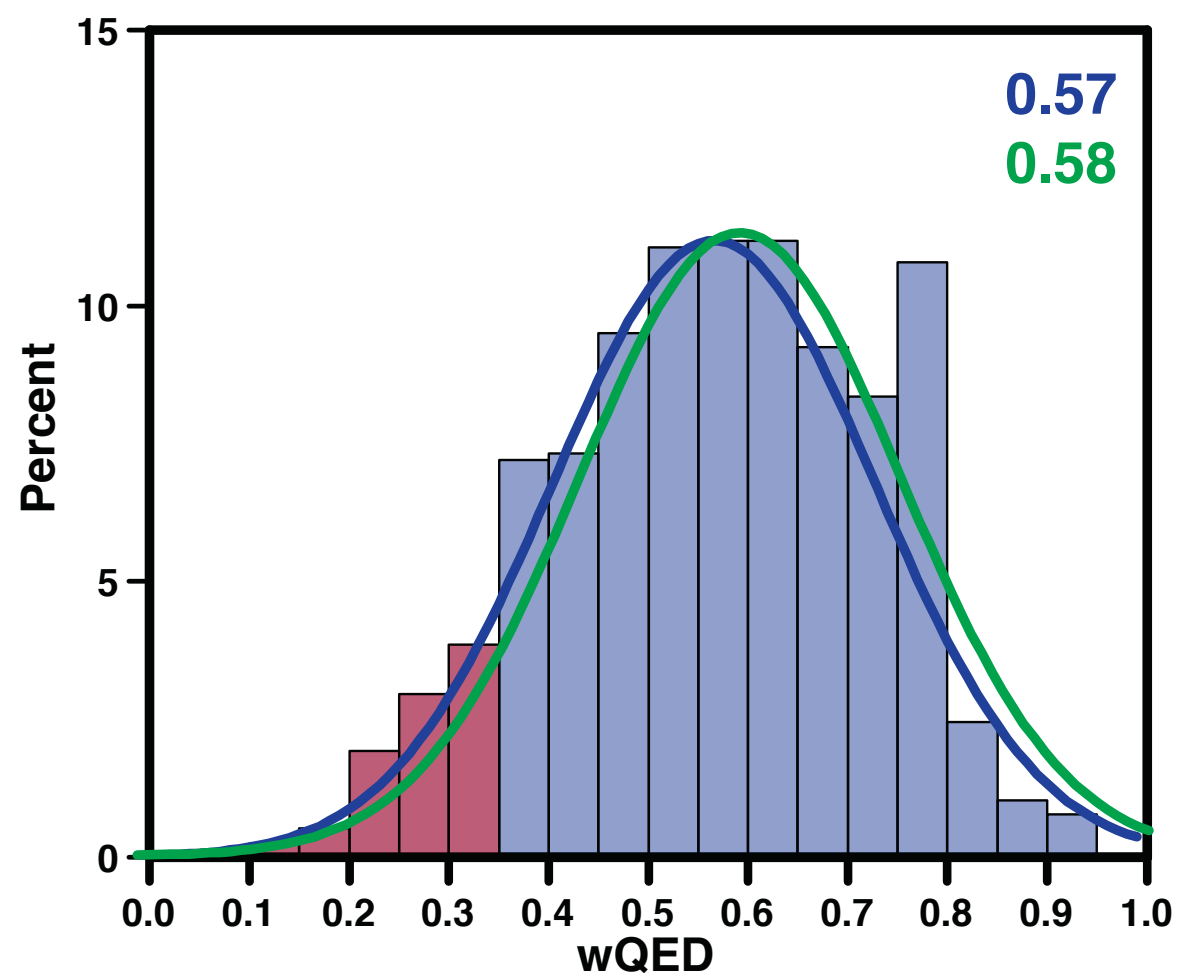
the TCAMS-TB dataset
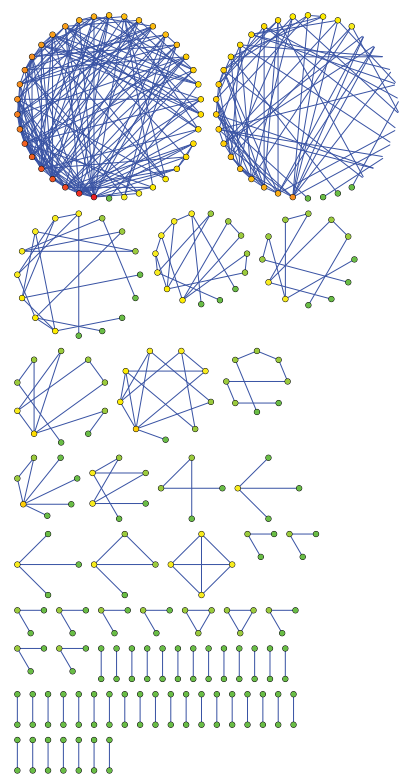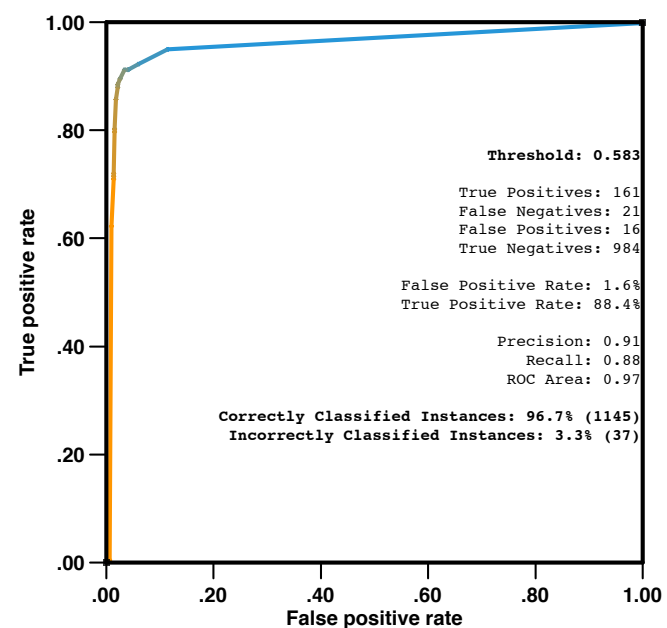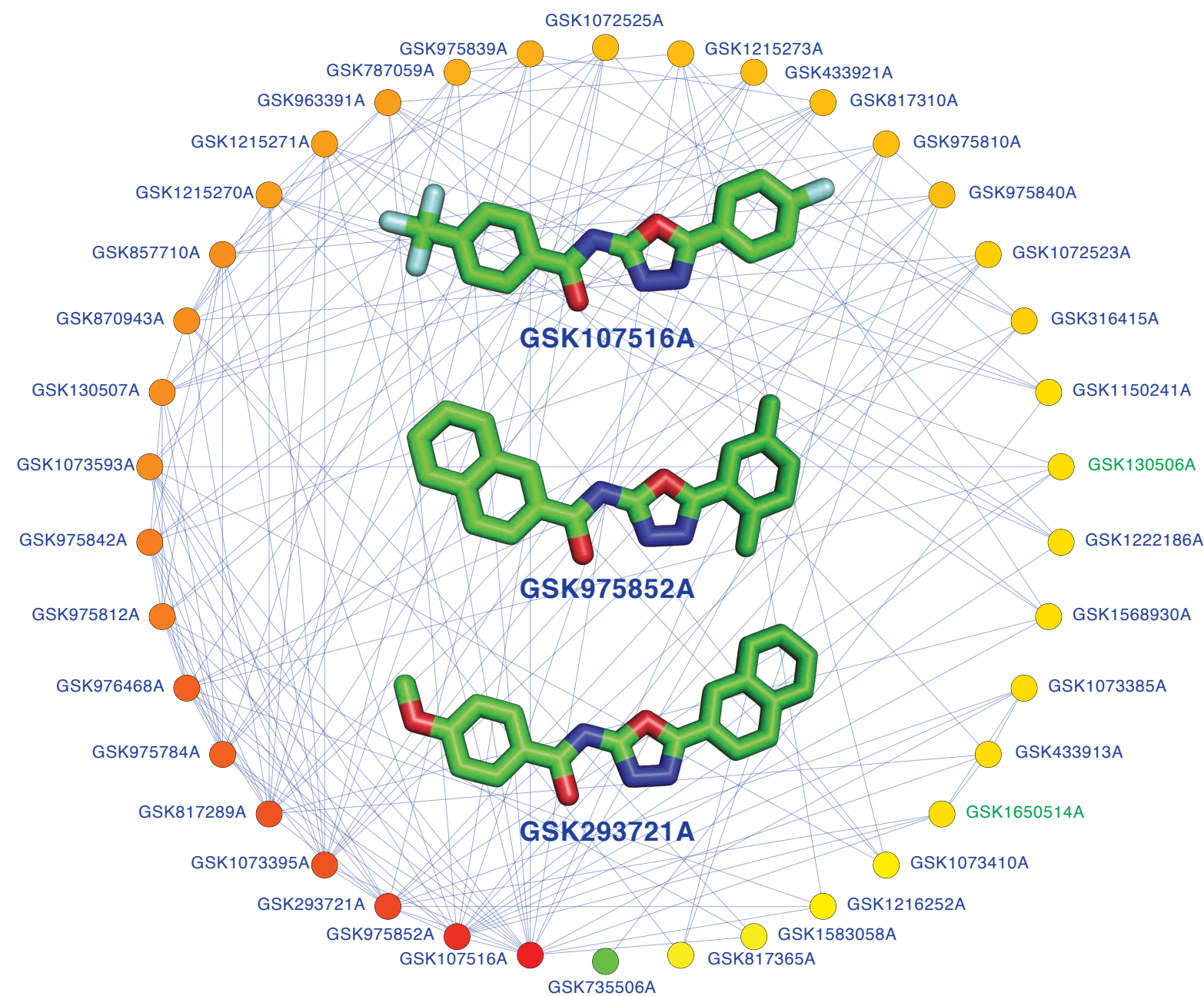
CNAG's nAnnoLyze (STR)

Marti-Renom lab

~2,500 PDB ligands

~776 GSK ligands

~5,000 MTB models

Sali lab

~16,000 PDB Chains
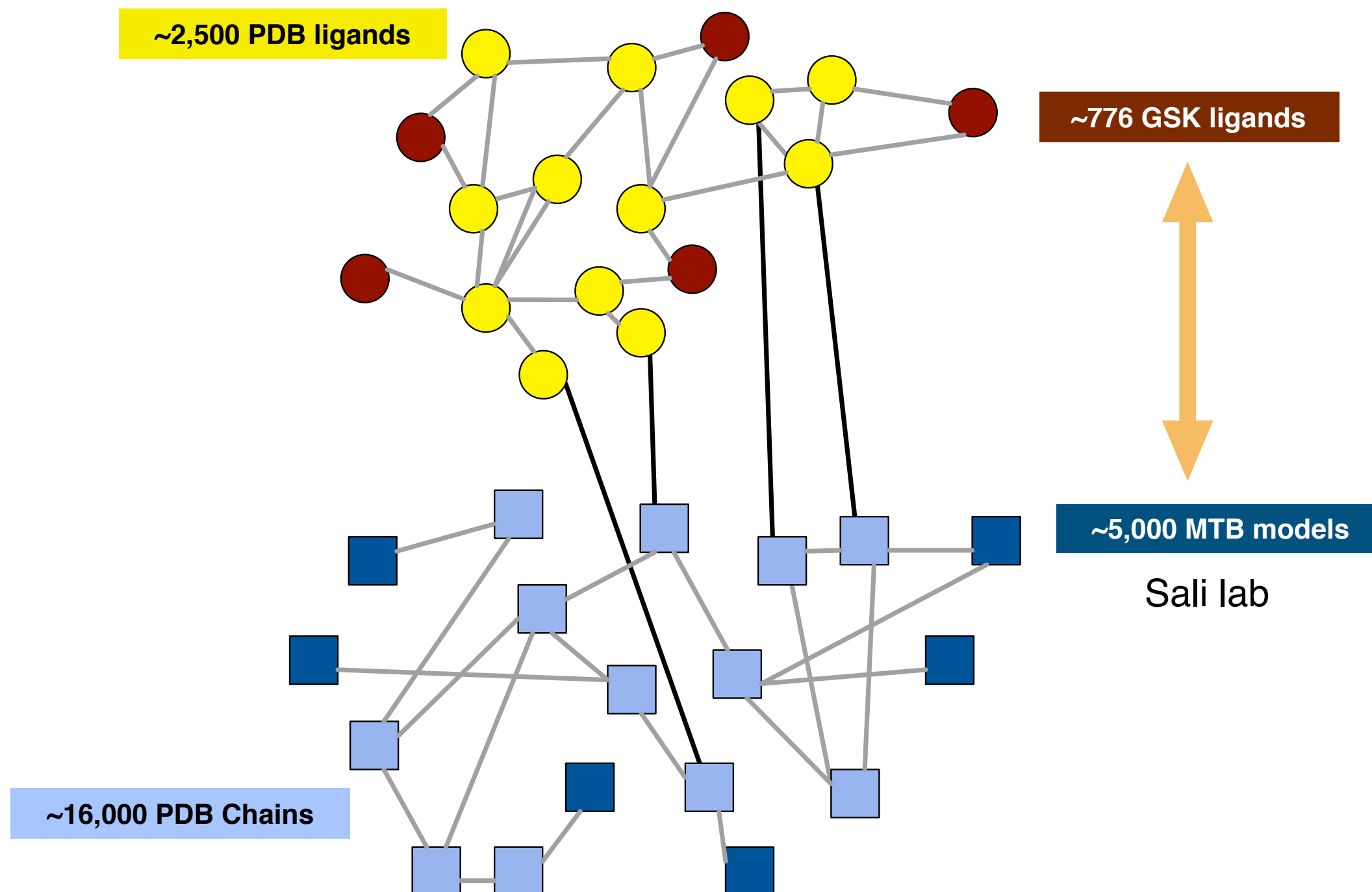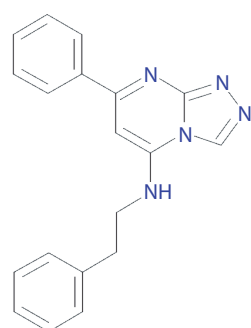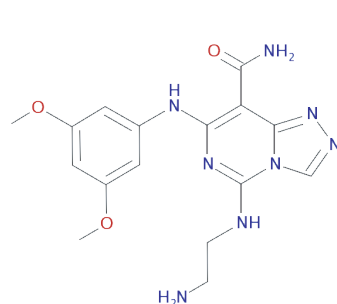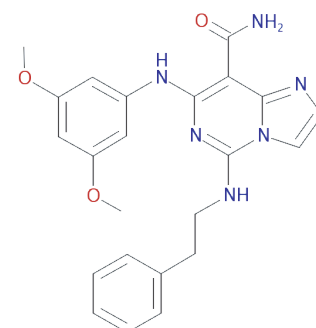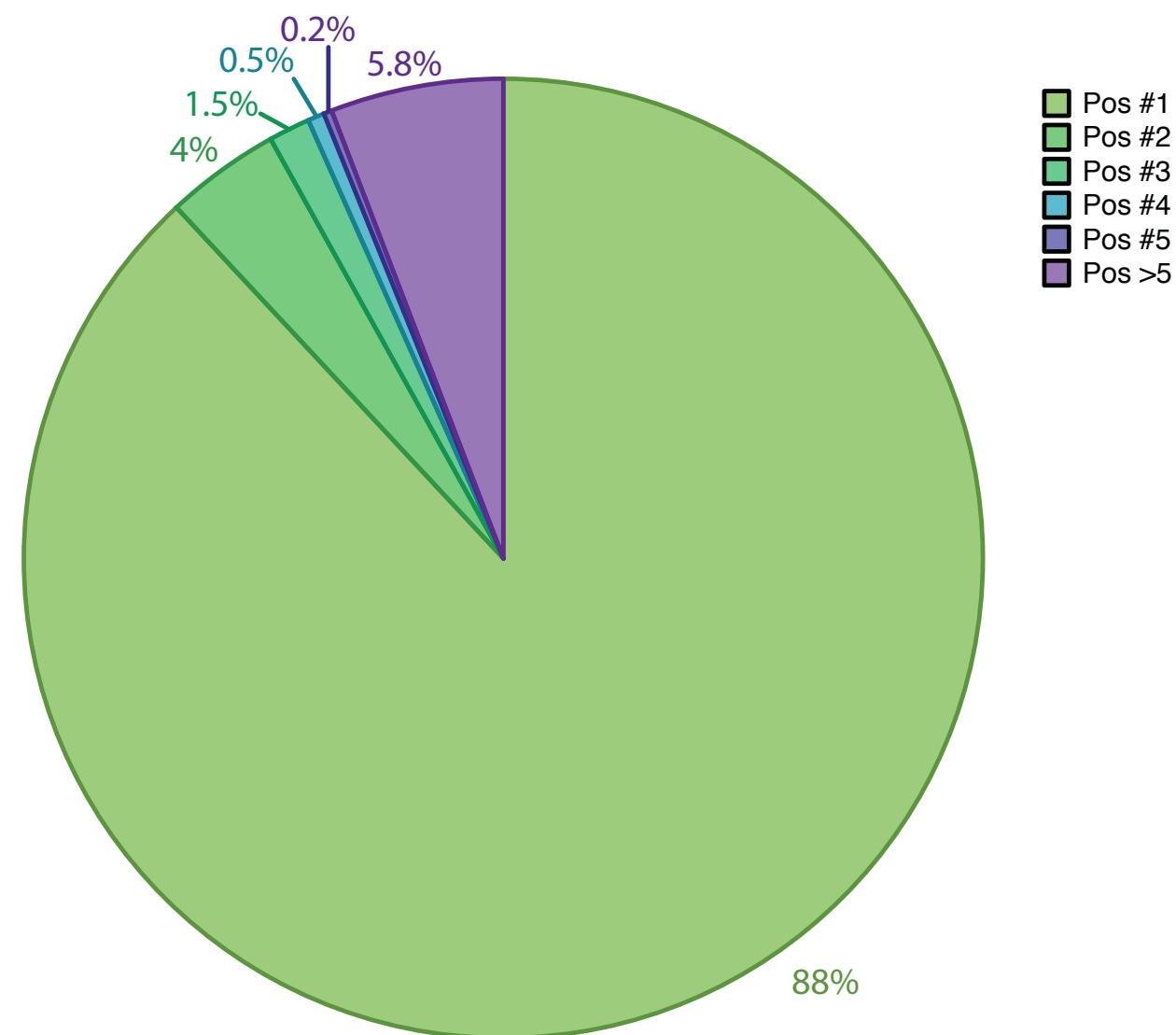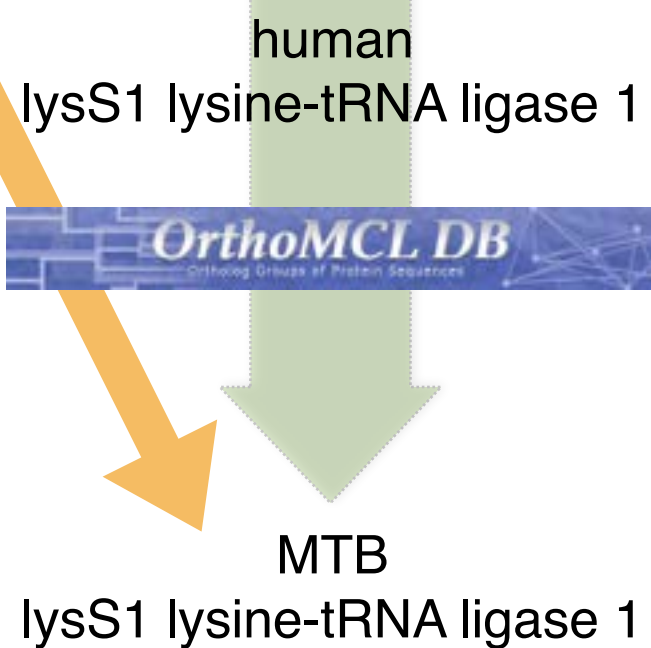
GSK (HIST)

Brown lab

Human Target Class	No. of Compounds		Putative *Mycobacterium* Target		Accession No.		Essentiality
	BCG	H37Rv	Gene	Product	M. bovis BCG	MTB H37Rv	
Kinase	35	8	pknA	transmembrane serine/threonine-protein kinase A	YP_976148.1	NP_214529.1	Essential
			pknB	transmembrane serine/threonine-protein kinase B	YP_976147.1	NP_214528.1	Essential
			pknD	Ser/Thr protein kinase	YP_977078.1	NP_215446.1	NE
			pknH	putative transmembrane serine/threonine-protein kinase H	YP_977417.1	NP_215782.1	NE
			pknJ	putative transmembrane serine/threonine-protein kinase J	YP_978197.1	NP_216604.1	NE
			pknL	putative transmembrane serine/threonine-protein kinase L	YP_978280.1	NP_216692.1	NE
		None	pknF	anchored-membrane serine/threonine-protein kinase F	YP_977877.1	NP_216262.1	NE
		None	pknK	putative serine/threonine-protein kinase transcriptional regulatory protein K	YP_979189.1	NP_217596.1	NE
Other Enzyme	1	0	aao	Putative D-amino acid oxidase	YP_978034.1		NE
	5	2	amiB2	amidase	YP_977414.1	NP_215779.1	NE
	2	1	aofH	putative flavin-containing	YP_979728.1	NP_217686.1	NE

Compounds

776 compounds
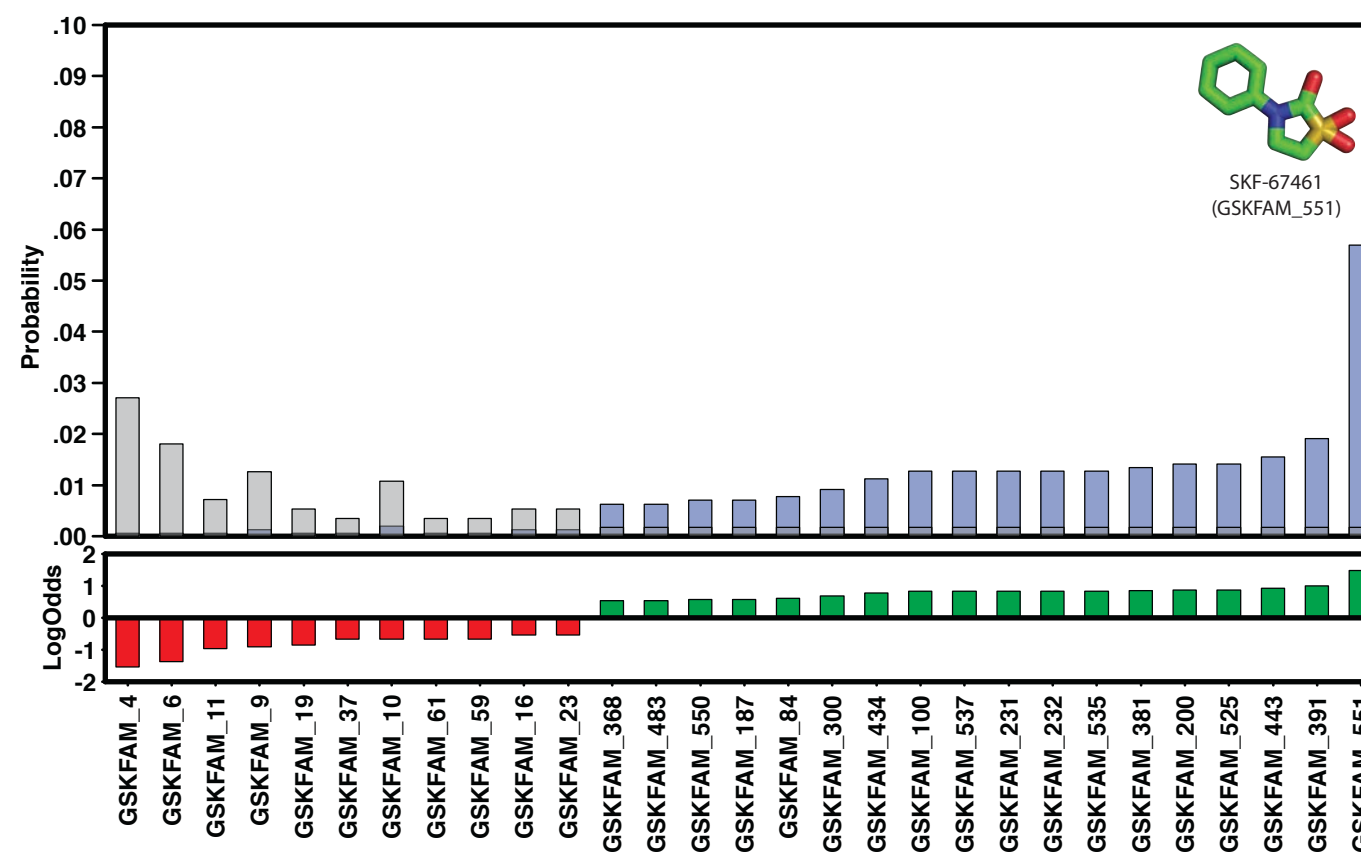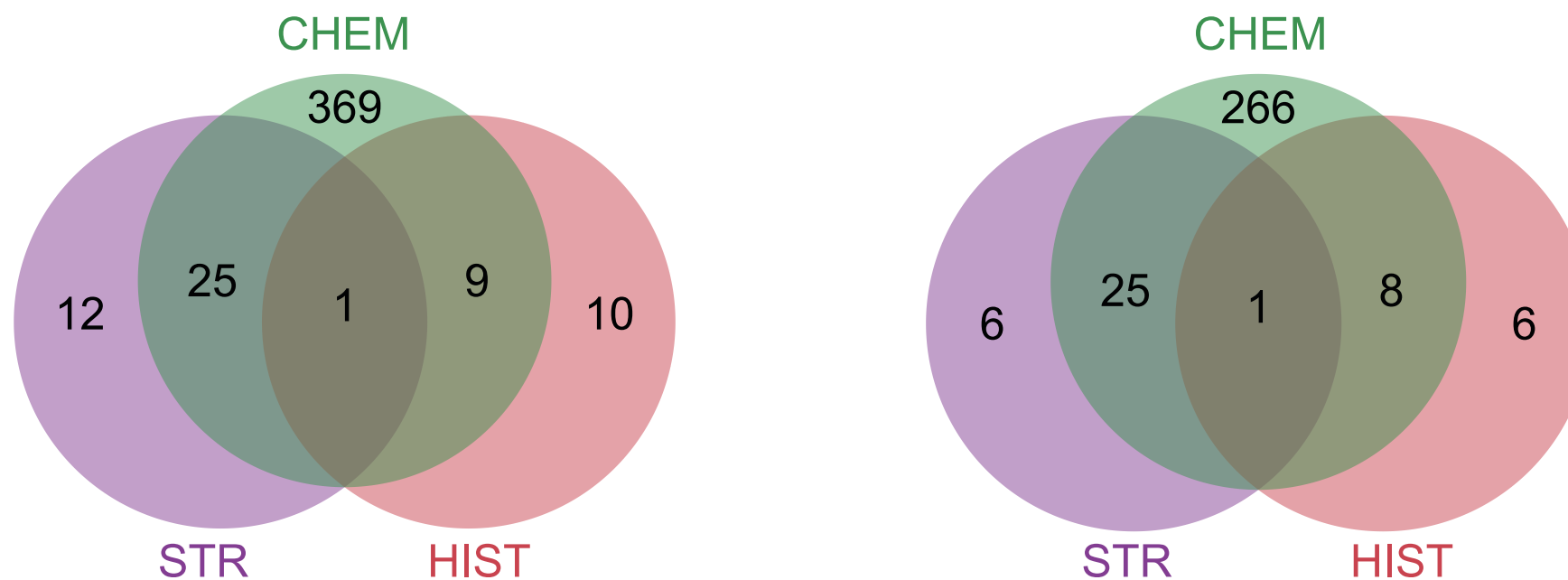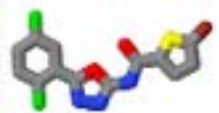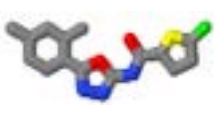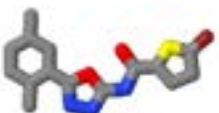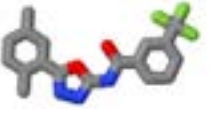
Pathways

1,044 unique targets in 112 (KEGG) pathways
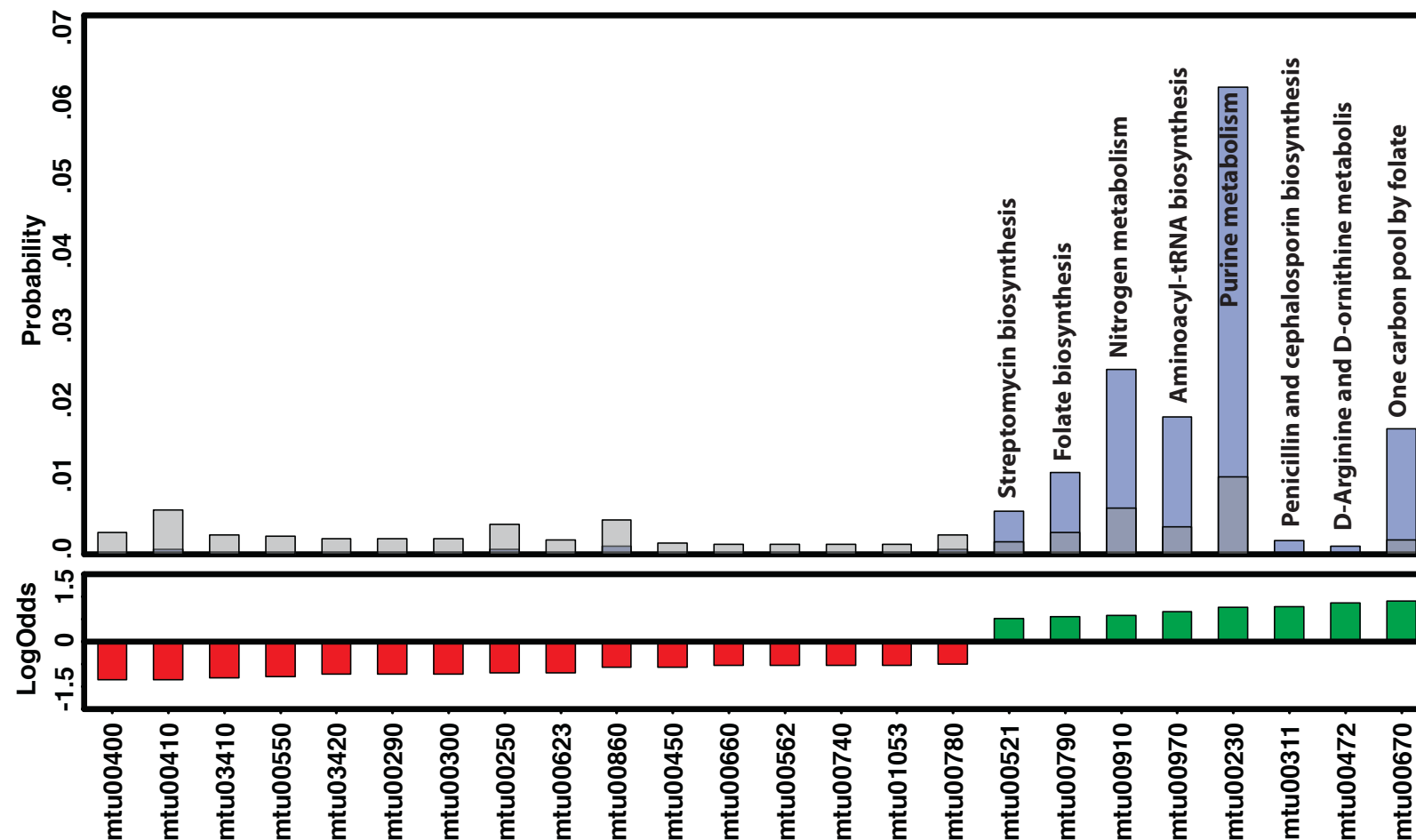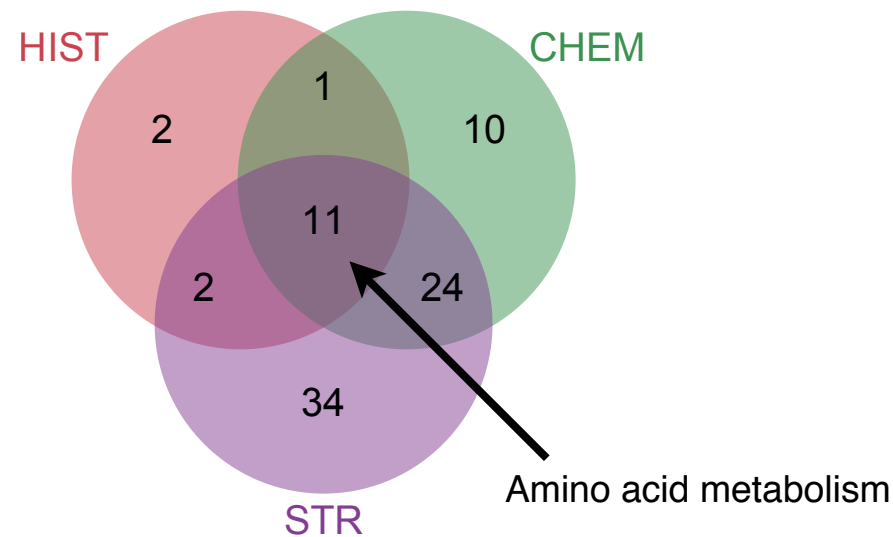
Compounds-pathways
8 compound family "statistically" associated to 13 pathways

GSK Family	Compound	Target	Pathways
1	GSK975784A	*Rv2182c*	**Glycerolipid metabolism (mtu00561)** **Glycerophospholipid metabolism (mtu00564)**
		Rv2483c	No Pathway
	GSK975810A	*Rv2182c*	**Glycerolipid metabolism (mtu00561)** **Glycerophospholipid metabolism (mtu00564)**
		Rv2483c	No Pathway
	GSK975839A	*Rv2182c*	**Glycerolipid metabolism (mtu00561)** **Glycerophospholipid metabolism (mtu00564)**
		Rv2483c	No Pathway
		Rv2299c	No Pathway
	GSK975840A	*Rv2182c*	**Glycerolipid metabolism (mtu00561)** **Glycerophospholipid metabolism (mtu00564)**
		Rv2483c	No Pathway
	GSK975842A	*Rv2182c*	**Glycerolipid metabolism (mtu00561)** **Glycerophospholipid metabolism (mtu00564)**
		Rv2483c	No Pathway
		Rv2045c	No Pathway
		Rv2139	Pyrimidine metabolism (mtu00240)
		Rv2299c	No Pathway
		Rv2483c	No Pathway

	GSK547490A	Rv0194	**ABC transporters (mtu02010)**
	GSK547491A	Rv0194	**ABC transporters (mtu02010)**
	GSK547499A	Rv0194	**ABC transporters (mtu02010)**
	GSK547500A	Rv0194	**ABC transporters (mtu02010)**
	GSK547511A	Rv0194	**ABC transporters (mtu02010)**

8 compound families with significant links to pathways

All available @ http://www.tropicaldisease.org/TCAMSTB

46

pknB target

GSK1598164A

GSK number	pIC50	Gene	Score$_{2PZI}$*	Score$_{3F69}$*
GW623128X	7.3	CNR2	-8.83	-8.63
GSK1519001A	5.7	NPSR1	-9.09	-8.65
GSK547481A	6.1	HTR4	-8.93	-9.27
GSK2043267A	7.5	CYP2C19	-8.8	-8.56
GSK381407A	5.6	P2RY14	-9.02	-8.71
GSK547543A	6.2	GPR55	-9.03	-10.37
GSK547511A	6.2	GPR55	-9.01	-8.58
GSK1598164A	8	IKBKB	-9.19	-8.96
GSK1635139A	5.6	CHRNA7	-8.6	-9.59
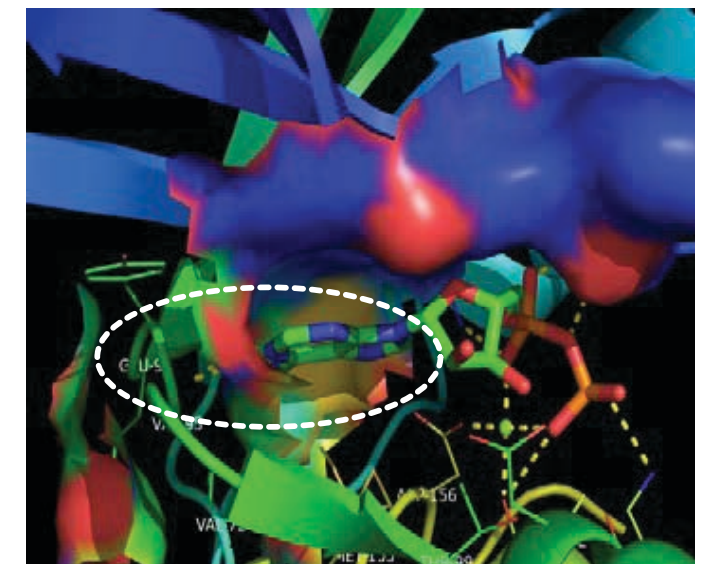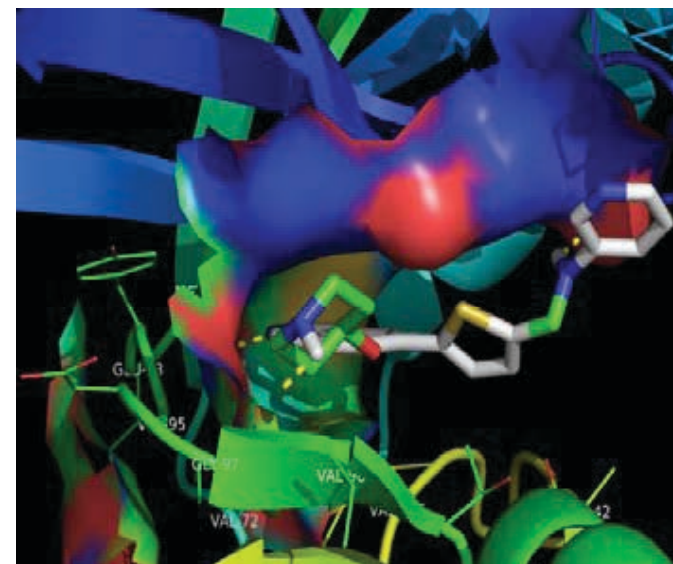
PknB kinase docking to GSK1598164A. A) Multiple sequence alignment of *Mycobacterium* PknB kinase with selected human kinases. Human kinases were selected on the criteria of having available PDB structures and top Psi-BLAST scores to *M. bovis* transmembrane serine/threonine-protein kinase B (pknB). First sequence in the alignment (gene name; PDB identifier) is *M. tuberculosis* transmembrane serine/threonine-protein kinase B (PknB; 3F69), which is 99% identical to *M. bovis* PknB and was used in compound docking models. Other sequences are CAMK2D (2EWL), MARK3 (2QNJ), MARK2 (3IEC), AKT2 (1GZK) and SGK1 (2R5T). Residues known to interact with ADP in pknB are highlighted in red. The amino acids aligned with Glu93, which may be essential for the binding of the GSK1132084A, are highlighted in green. **B)** Binding models of the GSK1598164A and ADP within pknB binding site (left and right panels, respectively).

serS target

GSK1402290A
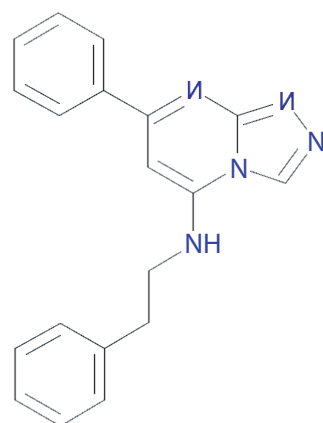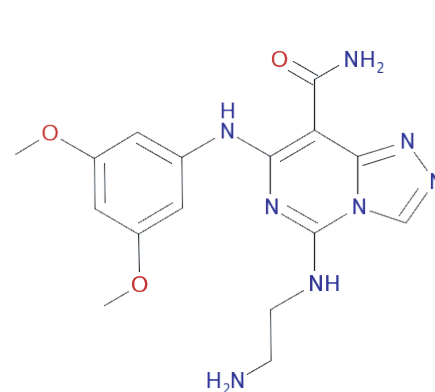
| GSK1402290A | CHEMBL474582 | CHEMBL508242 |
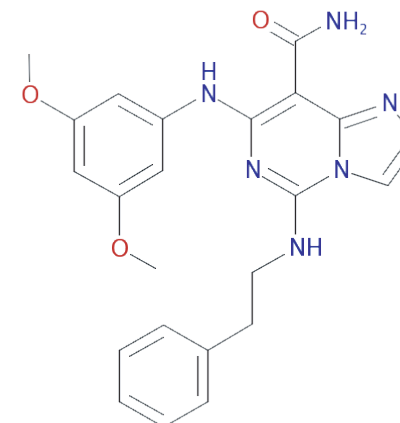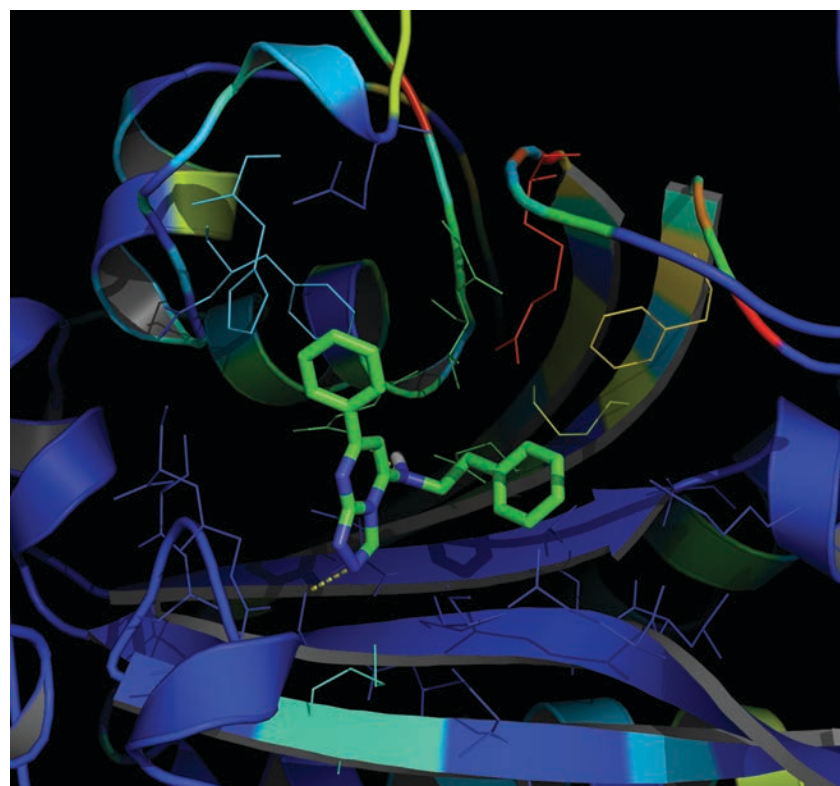
Targeting the aminoacyl-tRNA biosynthesis pathway. CHEM results show that GSK1402290A shared several substructural features with compounds reported as potent lysyl-tRNA synthetase inhibitors in the ChEMBL database (e.g., CHEMBL474582 and CHEMBL508242). STR results predicted the serS as a target of GSK1402290A with its binding site including residues F205, H209, G225, T226, E228, R257, F276, K278, and E280, which are conserved in the PFAM family PF00587 (tRNA synthetase class II core domain). The image shows the pose for GSK1402290A predicted by AutoDock and the binding site residues (i.e., within 6Å from the compound) colored from low sequence conservation (blue) to high sequence conservation (red).

Web Server

Article

http://www.tropicaldisease.org/TCAMSTB

Martínez-Jiménez, et al. (2013). PLoS CB, 9(10), e1003253

PLOS | COMPUTATIONAL BIOLOGY

Target Prediction for an Open Access Set of Compounds Active against *Mycobacterium tuberculosis*

Francisco Martínez-Jiménez[1,2], George Papadatos[3], Lun Yang[4], Iain M. Wallace[3], Vinod Kumar[4], Ursula Pieper[5], Andrej Sali[5], James R. Brown[4]*, John P. Overington[3]*, Marc A. Marti-Renom[1,2]*

1 Genome Biology Group, Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Spain, 2 Gene Regulation Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Barcelona, Spain, 3 European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, 4 Computational Biology, Quantitative Sciences, GlaxoSmithKline, Collegeville, Pennsylvania, United States of America, 5 Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California, United States of America

Abstract

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), infects an estimated two billion people worldwide and is the leading cause of mortality due to infectious disease. The development of new anti-TB therapeutics is required, because of the emergence of multi-drug resistance strains as well as co-infection with other pathogens, especially HIV. Recently, the pharmaceutical company GlaxoSmithKline published the results of a high-throughput screen (HTS) of their two million compound library for anti-mycobacterial phenotypes. The screen revealed 776 compounds with significant activity against the *M. tuberculosis* H37Rv strain, including a subset of 177 prioritized compounds with high potency and low *in vitro* cytotoxicity. The next major challenge is the identification of the target proteins. Here, we use a computational approach that integrates historical bioassay data, chemical properties and structural comparisons of selected compounds to propose their potential targets in *M. tuberculosis*. We predicted 139 target - compound links, providing a necessary basis for further studies to characterize the mode of action of these compounds. The results from our analysis, including the predicted structural models, are available to the wider scientific community in the open source mode, to encourage further development of novel TB therapeutics.

Introduction

One third of the world's population is infected with *Mycobacterium tuberculosis* (MTB), the causative agent of tuberculosis [1]. Approximately 95% of infected individuals are thought to have persistent, latent MTB infections that remain dormant until activated by specific environmental and host response events. Approximately 10% of latent infections eventually progress to active disease, which, if left untreated, kills more than half of the infected patients [2]. Moreover, there is an increasing clinical occurrence of MTB strains with extensive multi-drug-resistance (*eg*, MTB MDR and MTB XDR), where mortality rates can approach 100% [3]. In some countries, the MTB MDR and XDR strains may account for up to 22% of infections [1]. In addition, current TB therapeutic regimes involve a combination of antibiotics, administered at regular intervals over a 6-month period, which makes patient compliance an issue, especially in developing countries [1,2].

The discovery and development of new antibiotics is widely recognized as one of the major global health emergencies, yet it is also a major pharmaceutical challenge. Most currently used antibiotics were discovered during the golden era from the 1940s

to 1960s through large scale screening of compound collections for anti-bacterial activity – the so-called whole cell or phenotypic screens [4]. The emergence of bacterial molecular genomics technologies and the availability of whole genome sequences in the 1990s led to dramatic changes in anti-bacterial drug discovery, where the emphasis was placed on screening essential targets for inhibitory compounds. However, despite intensive efforts, target-based screening has been largely unsuccessful in producing clinical candidate molecules [5]. As a result, a return to whole cell screening has been widely advocated, in combination with novel technologies and bioinformatics to rapid identify targets associated with a compound's mechanism of action (MOA) [4,6].

Recently, the pharmaceutical company GlaxoSmithKline (GSK) completed an anti-mycobacterial phenotypic screening campaign against *M. bovis* BCG, a non-virulent, vaccine *Mycobacterium* strain, with a subsequent secondary screening in *M. tuberculosis* H37Rv (MTB H37Rv) for hit confirmation [7]. A total of 776 potent compound hits (including 177 MTB H37RV hits with limited human cell line toxicity) were made openly available to the wider scientific community through the ChEMBL database (http://dx.doi.org/10.6019/CHEMBL2095176). The aim of this release was to stimulate mechanism of action analyses using

50

Fast Sequencing for TB treatment
Köser & Peacock

Whole-Genome Sequencing for Rapid Susceptibility Testing of *M. tuberculosis*

TO THE EDITOR: Efforts to contain drug-resistant tuberculosis depend on the rapid detection and effective treatment of cases, together with public health interventions to prevent and investigate ongoing transmission. The necessary laboratory support for these activities includes the identification of the *Mycobacterium tuberculosis* complex, antimicrobial susceptibility testing, and bacterial genotyping. However, even in well-resourced countries, it typically takes 1 to 2 months to achieve all these goals because of the slow growth rate of the *M. tuberculosis* complex.[1,2] Moreover, phenotypic susceptibility testing can be unreliable and is not performed for some agents. Molecular techniques have accelerated some of these diagnostic functions, but they only interrogate a small part of the microbial genome and do not provide all the clinically relevant information.[1-3] Whole-genome sequencing has not been used as a diagnostic tool for tuberculosis, in part because of the need to culture *M. tubercu-losis* complex for several weeks, until sufficient DNA can be extracted.[2,4]

Here we report the use of rapid whole-genome sequencing to investigate the case of a patient with extensively drug-resistant (XDR) tuberculosis (the case history is provided in the Supplementary Appendix, available with the full text of this letter at NEJM.org). His first sputum sample became culture-positive after 3 days in the mycobacterial growth indicator tube (MGIT) culture system. DNA was extracted directly from the MGIT tube and sequenced with the use of the Illumina MiSeq platform. Two distantly related Beijing strains of *M. tuberculosis* were identified (in a ratio of 7:3) (Fig. 1B). Mixed infection was not apparent when standard genotyping was performed on three additional samples from this patient by means of mycobacterial interspersed repetitive-unit–variable-number tandem-repeat assay, which probably identified the majority strain. These findings have important implications for

Claudio U. Köser, Ph.D.
Public Health England
Cambridge, United Kingdom

Josephine M. Bryant, B.Sc.
Wellcome Trust Sanger Institute
Hinxton, United Kingdom

Jennifer Becq, Ph.D.
Illumina (Cambridge)
Little Chesterford, United Kingdom

M. Estée Török, Ph.D.
University of Cambridge
Cambridge, United Kingdom

Matthew J. Ellington, D.Phil.
Public Health England
Cambridge, United Kingdom

Marc A. Marti-Renom, Ph.D.
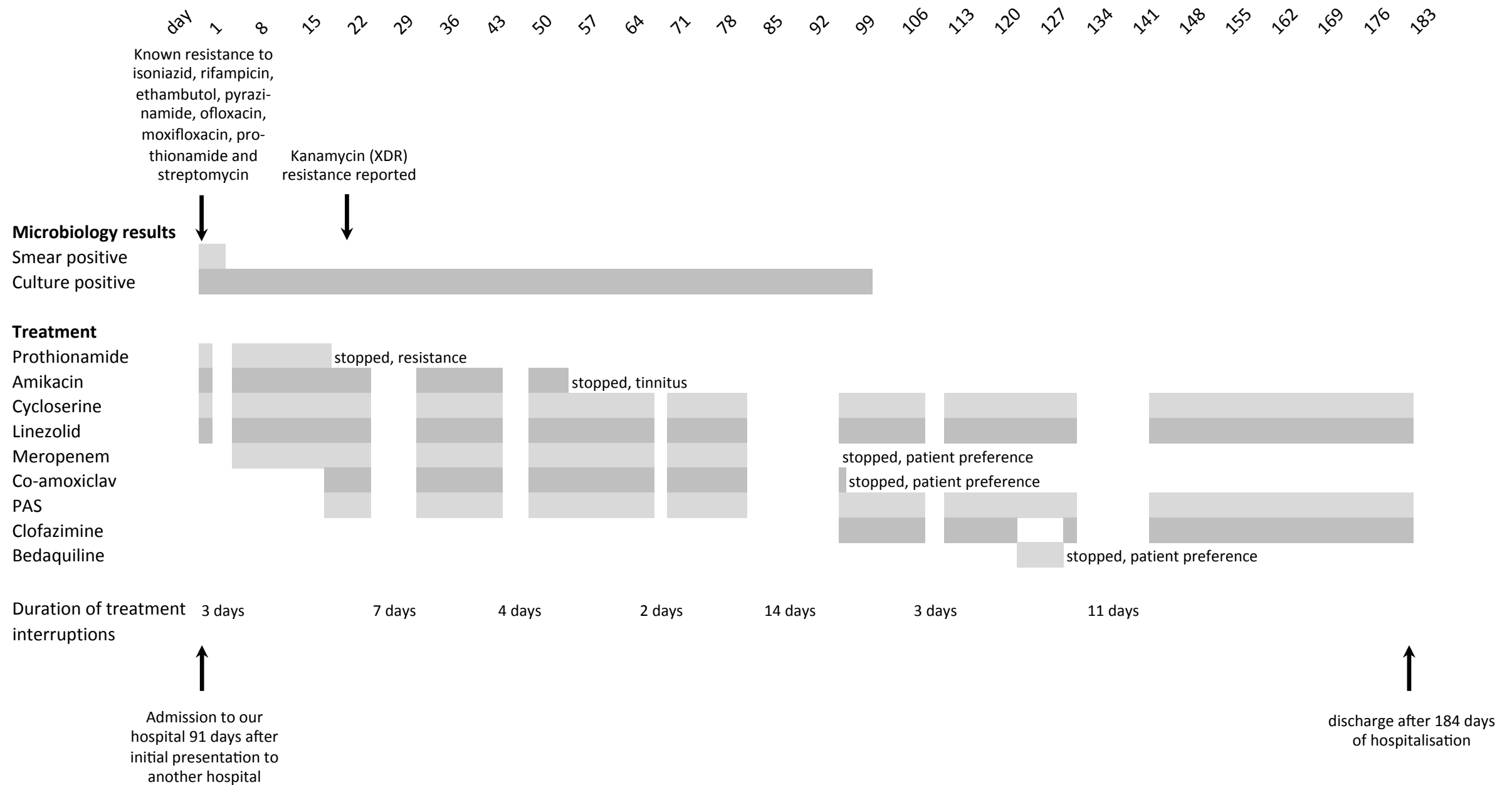Centre Nacional d'Anàlisi Genòmica
Barcelona, Spain

Andrew J. Carmichael, Ph.D.
Cambridge University Hospitals National Health Service
    Foundation Trust
Cambridge, United Kingdom

Julian Parkhill, Ph.D.
Wellcome Trust Sanger Institute
Hinxton, United Kingdom

Geoffrey P. Smith, Ph.D.
Illumina (Cambridge)
Little Chesterford, United Kingdom

Sharon J. Peacock, Ph.D.
University of Cambridge
Cambridge, United Kingdom
sjp97@medschl.cam.ac.uk

Microbiology Results, Phenotypic Drug Susceptibility Results and Anti-tuberculosis Therapy During Hospitalisation


day 1 8 15 22 29 36 43 50 57 64 71 78 85 92 99 106 113 120 127 134 141 148 155 162 169 176 183

Known resistance to isoniazid, rifampicin, ethambutol, pyrazinamide, ofloxacin, moxifloxacin, prothionamide and streptomycin

Kanamycin (XDR) resistance reported

Microbiology results
Smear positive
Culture positive

Treatment
Prothionamide — stopped, resistance
Amikacin — stopped, tinnitus
Cycloserine
Linezolid
Meropenem — stopped, patient preference
Co-amoxiclav — stopped, patient preference
PAS
Clofazimine
Bedaquiline — stopped, patient preference

Duration of treatment interruptions

3 days   7 days   4 days   2 days   14 days   3 days   11 days

Admission to our hospital 91 days after initial presentation to another hospital

discharge after 184 days of hospitalisation

52

Phylogenetic Analysis and the Distribution of Drug-Resistance Mutations.

a Distribution of SNPs and Resistance Mutations
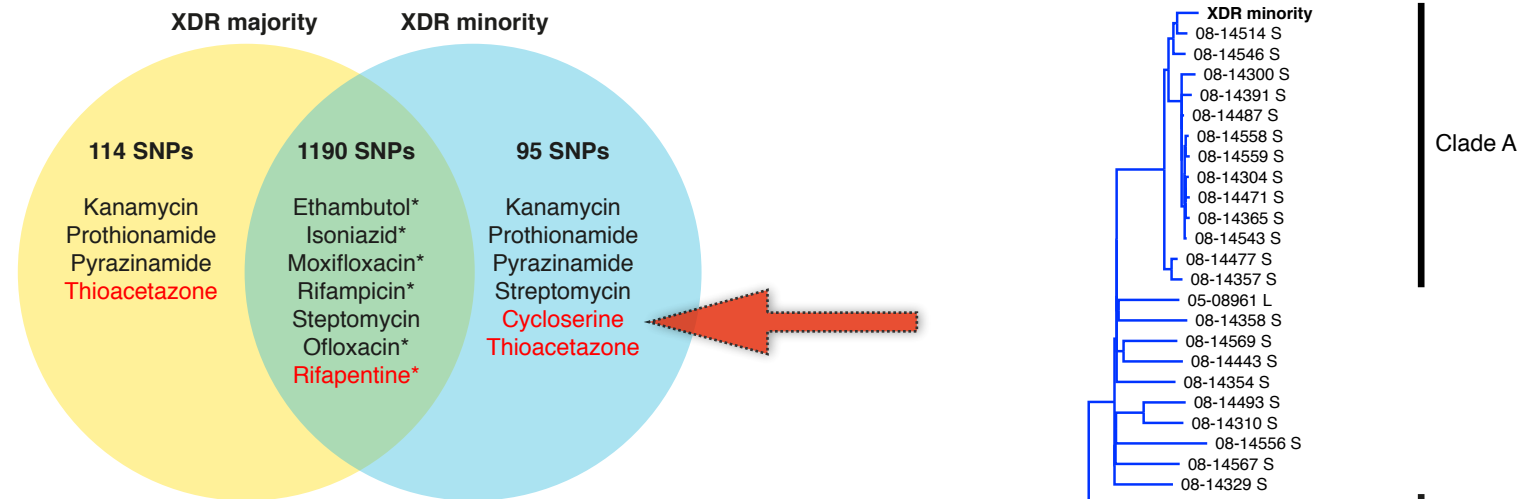
b Phylogenetic Analysis

315 SNPs
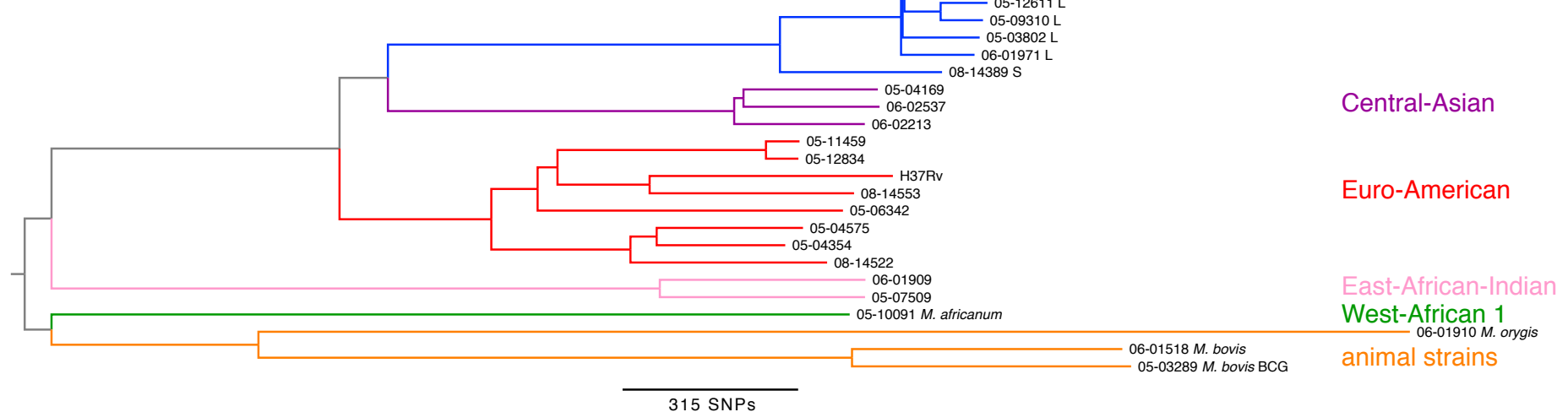
Alanine Racemase Conservation

Cycloserine

Alanine Racemase **M343T** Modeling

Cycloserine

Acknowledgments