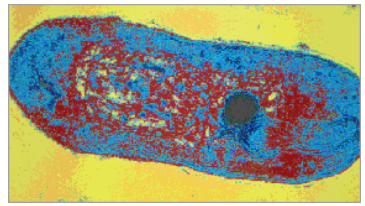
New approaches to identifying drug targets

Bhupesh Taneja
IGIB, CSIR, India

Indo-Russian Workshop
on Systems Biology and Genome Informatics,
October 12-14 2008

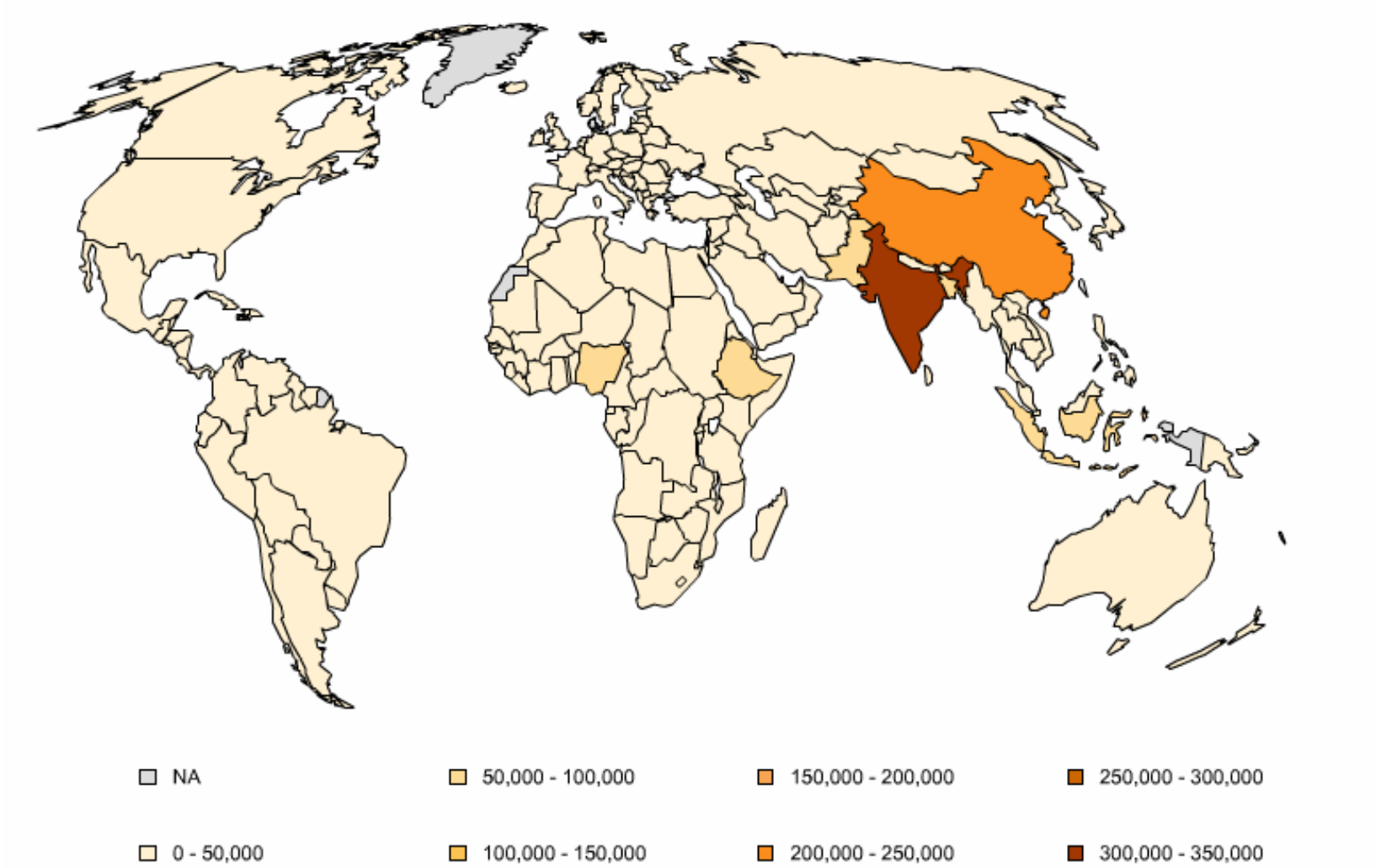
Potential targets of *M. tuberculosis*



EM of *M. tuberculosis*, Source: Institut Pasteur library

- *Mycobacterium tuberculosis*:
  - Causative agent of tuberculosis
  - Leading cause of death by any single infectious organism in the world
  - According to WHO estimates, there are approximately 9 million new infections and 2 million casualties of tuberculosis every year (WHO report, 2004)
  - Need for new and powerful drug and vaccine targets
  - The availability of the complete genome sequence of *M. tuberculosis* has been an important step that has helped in the renewed efforts to search for new and novel drug targets

Deaths due to Tuberculosis in 2005



Source: GlobalHealthFacts.org

Standard TB treatment Regimens: DOTS (Directly Observed Treatment Scheme)

Regimen	Combination Drugs	Duration	Dose
2(HRZE)	Isoniazid (**H**) Rifampicin (**R**) Pyrazinamide(**Z**) Ethambutol (**E**)	2 months	Daily
4(HR)$_3$		4 months	Thrice a week
2(HR)ZE	Isoniazid (**H**) Rifampicin (**R**) Pyrazinamide(**Z**) Ethambutol (**E**)	2 months	Daily
6(HR)		6 months	Daily

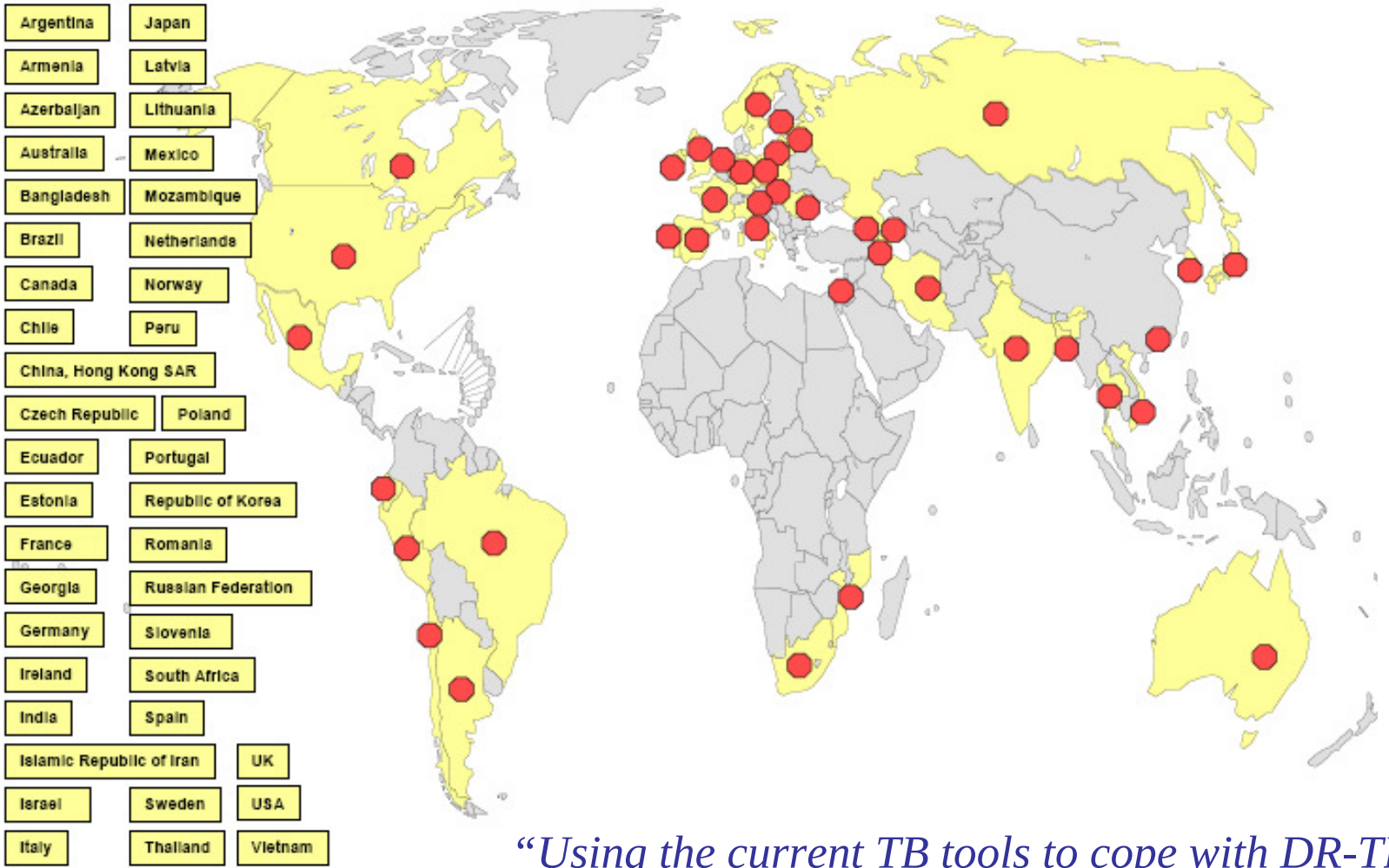
Targets of current TB drugs

Drug	Target
• Isoniazid	• Acyl carrier protein reductase
• Rifampicin	• RNA polymerase β subunit
• Pyrizinamide	• energy metabolism
• Ethambutol	• Arabinosyl transferase
• Streptomycin	• Ribosomal S12 protein and 16S rRNA
• Kanamycin	
• Quinolones	• DNA gyrase

Targets of TB drugs in pipeline

Drug	Target
• Moxifloxacin • Gatifloxacin	• DNA topoisomerase & DNA Gyrase
• PA-824	• Cell wall synthesis
• TMC- 207	• ATP synthase
• LL3858	• Unknown
• OPC 67683	• Mycolic acid inhibitors

41 Countries with XDR TB



Source: World Health Organisation

"Using the current TB tools to cope with DR-TB is like trying to put out a forest fire with a garden hose." *Françoise Louis, MSF TB Advisor.*

- **Need for new drug targets and/ or Novel approaches to drug discovery**

Open Source Drug Discovery

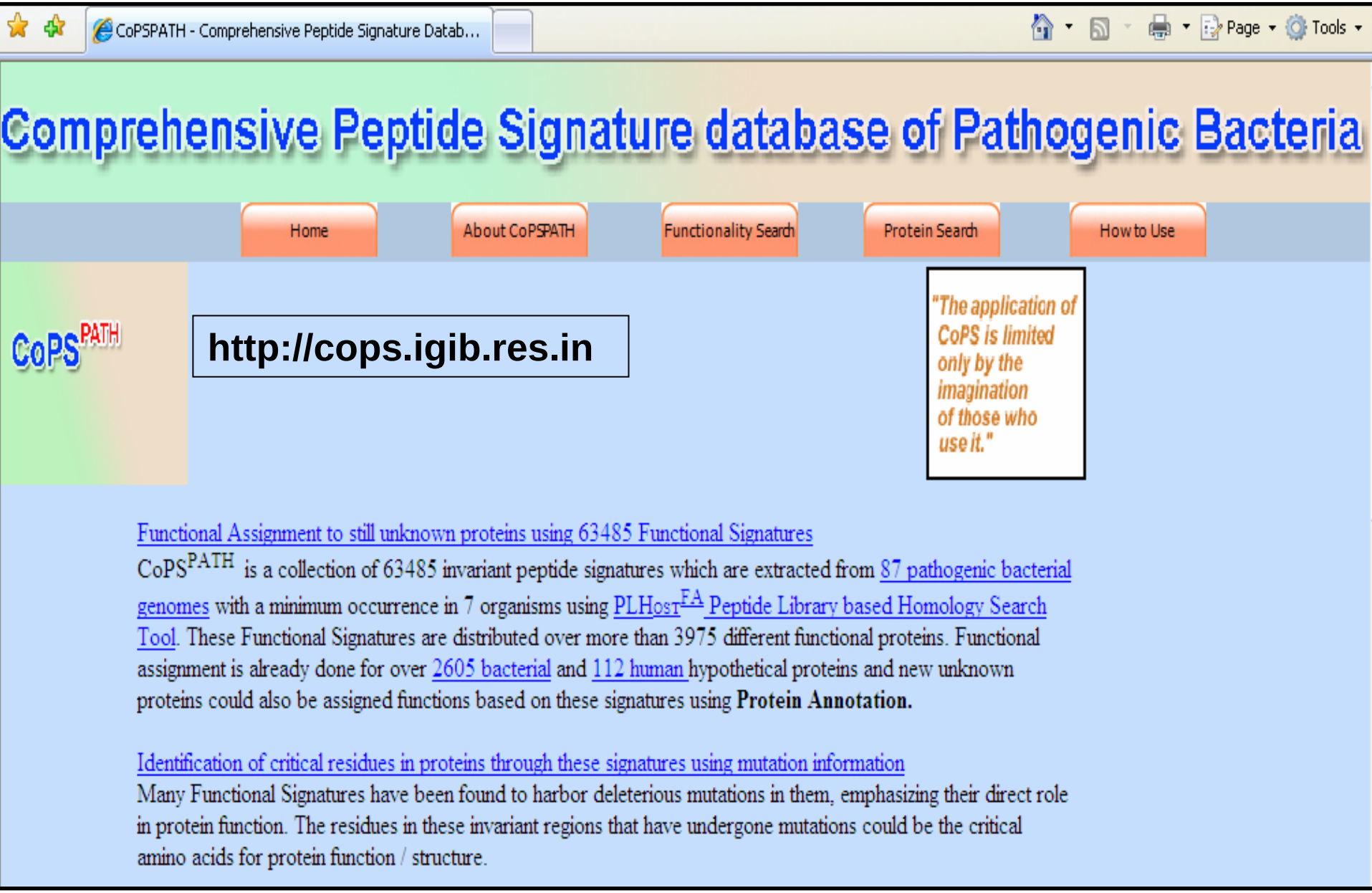
Work Packages

Phase- I	
WP-1	**Identification of Targets *(in Silico)***
WP-2	**Expression of Targets**
WP-3	**Validation of targets and Screen Development**
WP-4	**Identification of Chemical library**
WP-5	**Microarray gene expression**
WP-6	**Lead optimization on the non-toxic Hits**
WP-7	**Synthesis of analogues**
WP-8	**Identify non specific binding using Proteomics**
Phase-2	
WP-9	**Preclinical toxicity**
WP-10	**Clinical Trial**

Search for potential non-toxic drug targets of *M. tuberculosis*

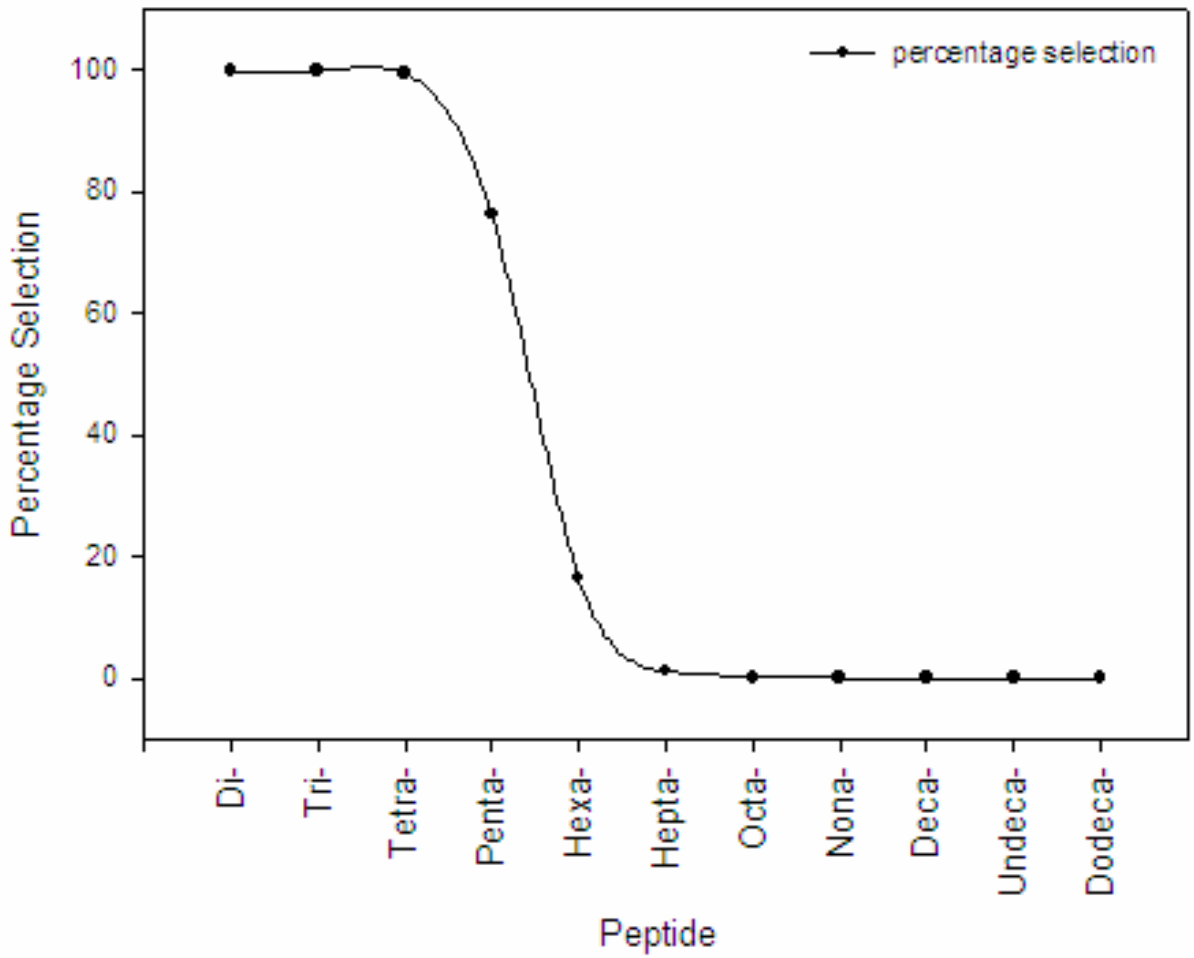
Search for potential non-toxic drug targets of *M. tuberculosis*

- Specific aims: Drug Target Development Using In-Silico Biology
  - **In silico target discovery** for infectious disease research
  - **Comparative genomics** for identification of antibacterial drug targets
  - **In silico modeling** and active site prediction to assess targets
  - Creation of virtual chemical library, pharmacophore generation and optimization
  - Peptide and protein structure elucidation.



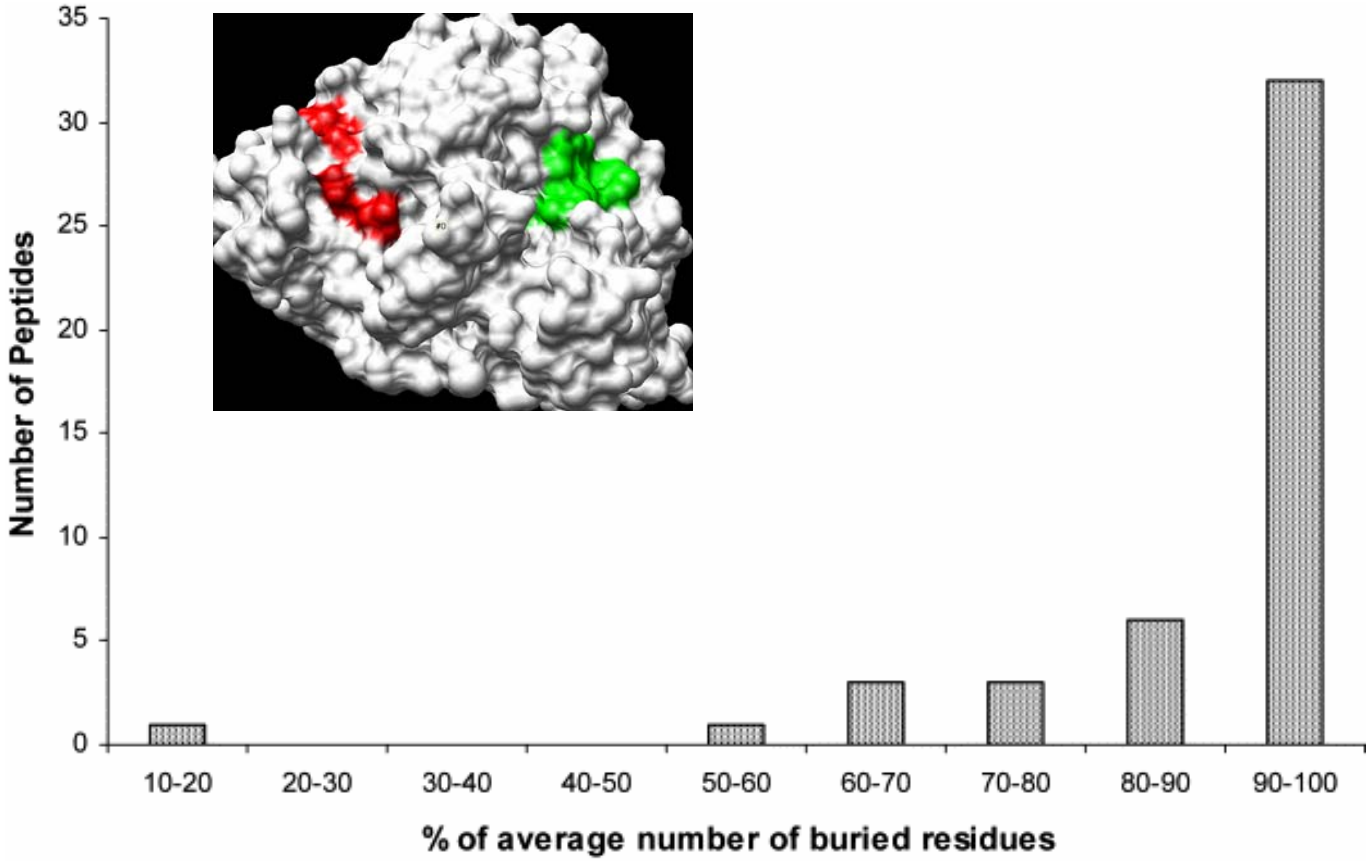
COPS-version 3.0: D. Dash et al., (2008), in preparation

How big are these peptides?



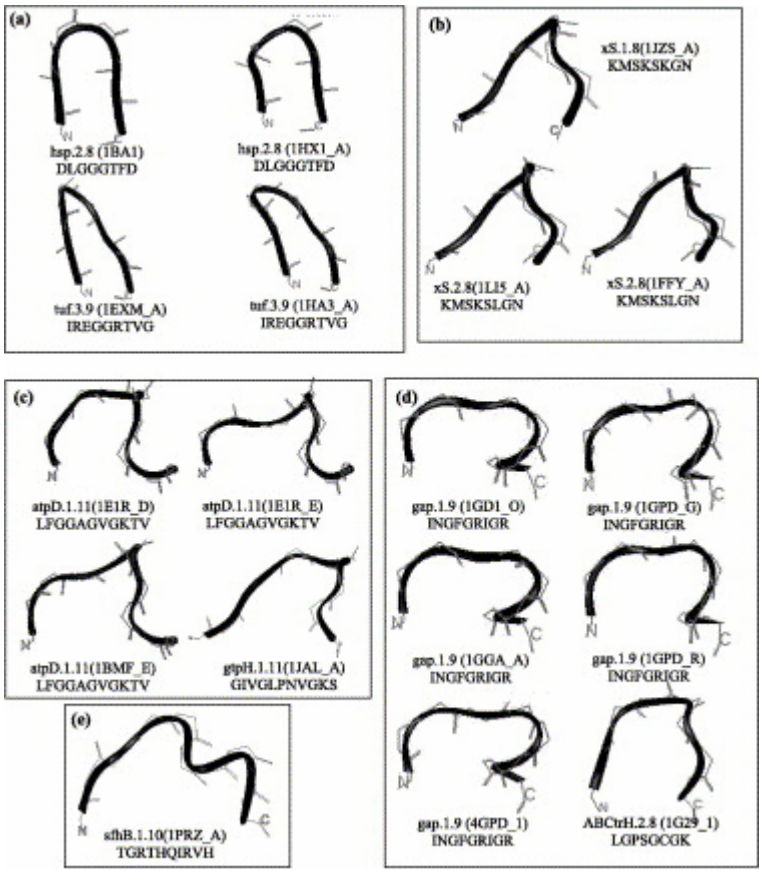
Where do they occur?

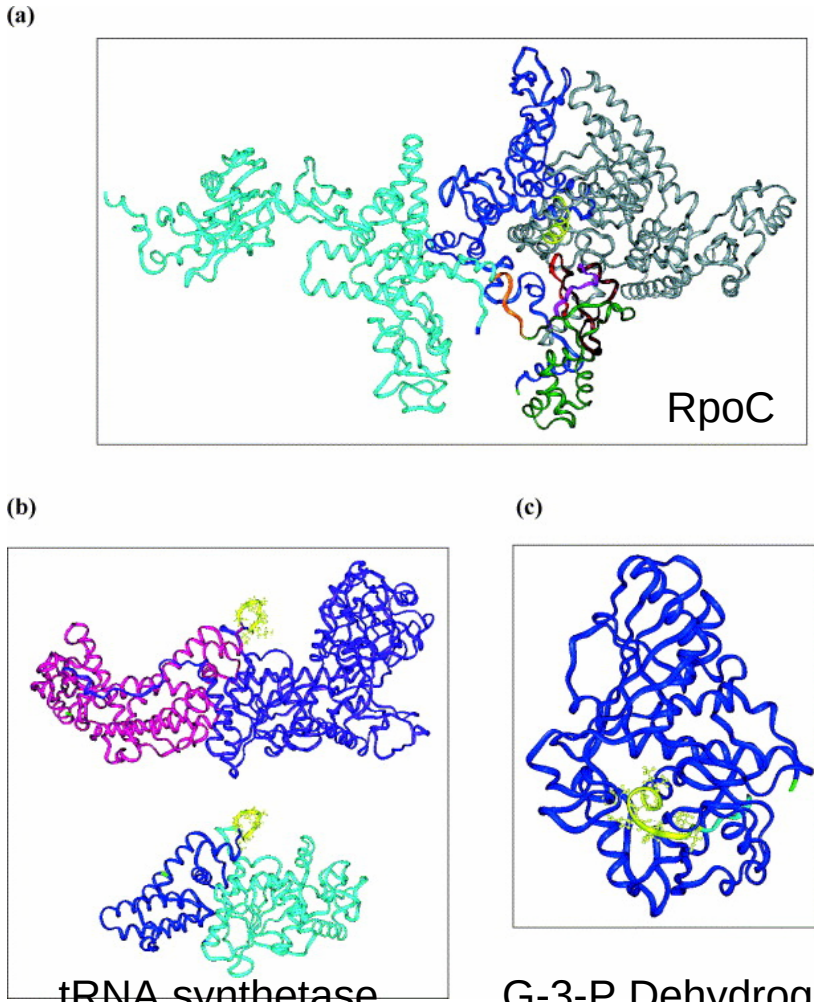
Invariant peptides are mostly buried



Prakash et al., 2005, JMB

COPS peptides: act as folding nuclei

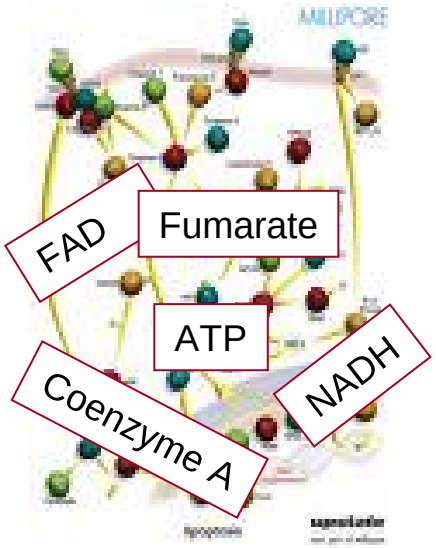




Prakash et al., (2004) Bioinformatics
Prakash et al., (2005), JMB

Mycobacterium tuberculosis SysBorg

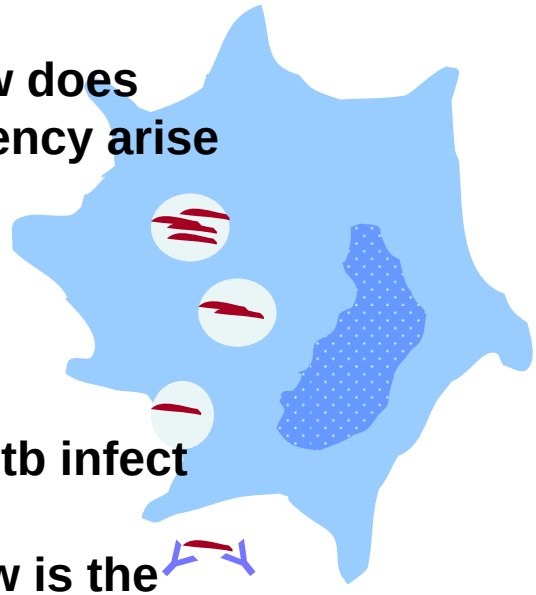
Mycobacterium tuberculosis SysBorg : A systems Biology platform for infectious diseases using Systems Biology of whole organism



Metabolic and Signalling Network

SysBorg

How does Latency arise



How does M.tb infect

How is the pathogen cleared

(CSIR Task force Network for in silico drug target discovery)

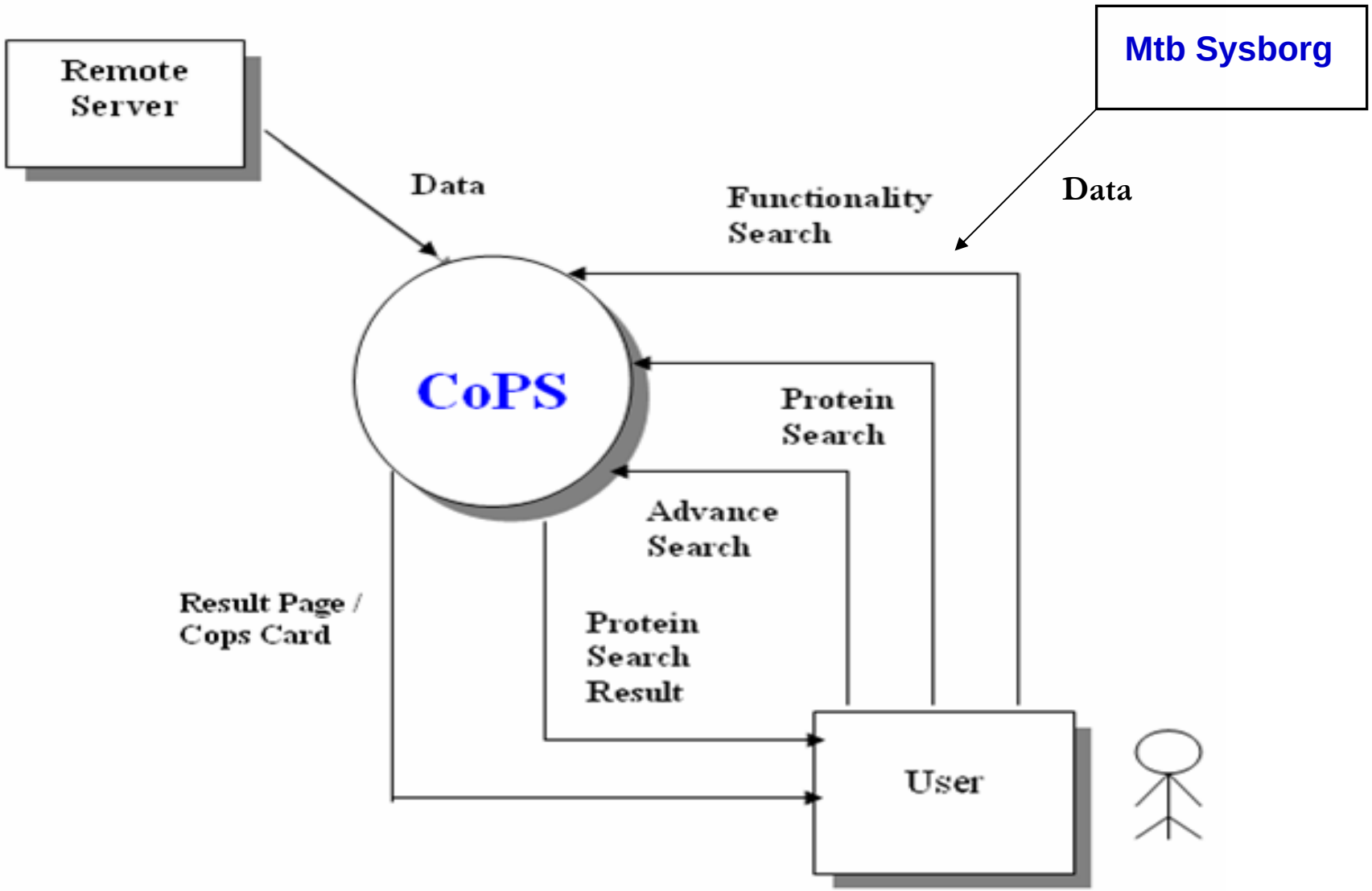
IGIB IMTECH CDRI IICB RRLJ

Other partners

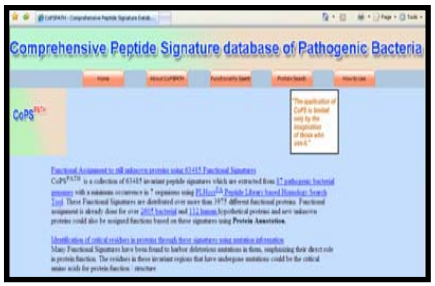
ACBR VPCI NII

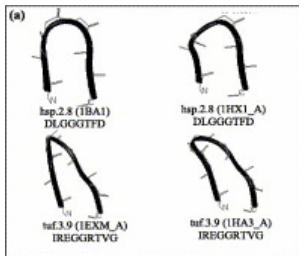
SOURCE OF PEPTIDES CLUSTERS: COPS

DATA FLOW DIAGRAM



Criteria for selection of potential non-toxic drug targets





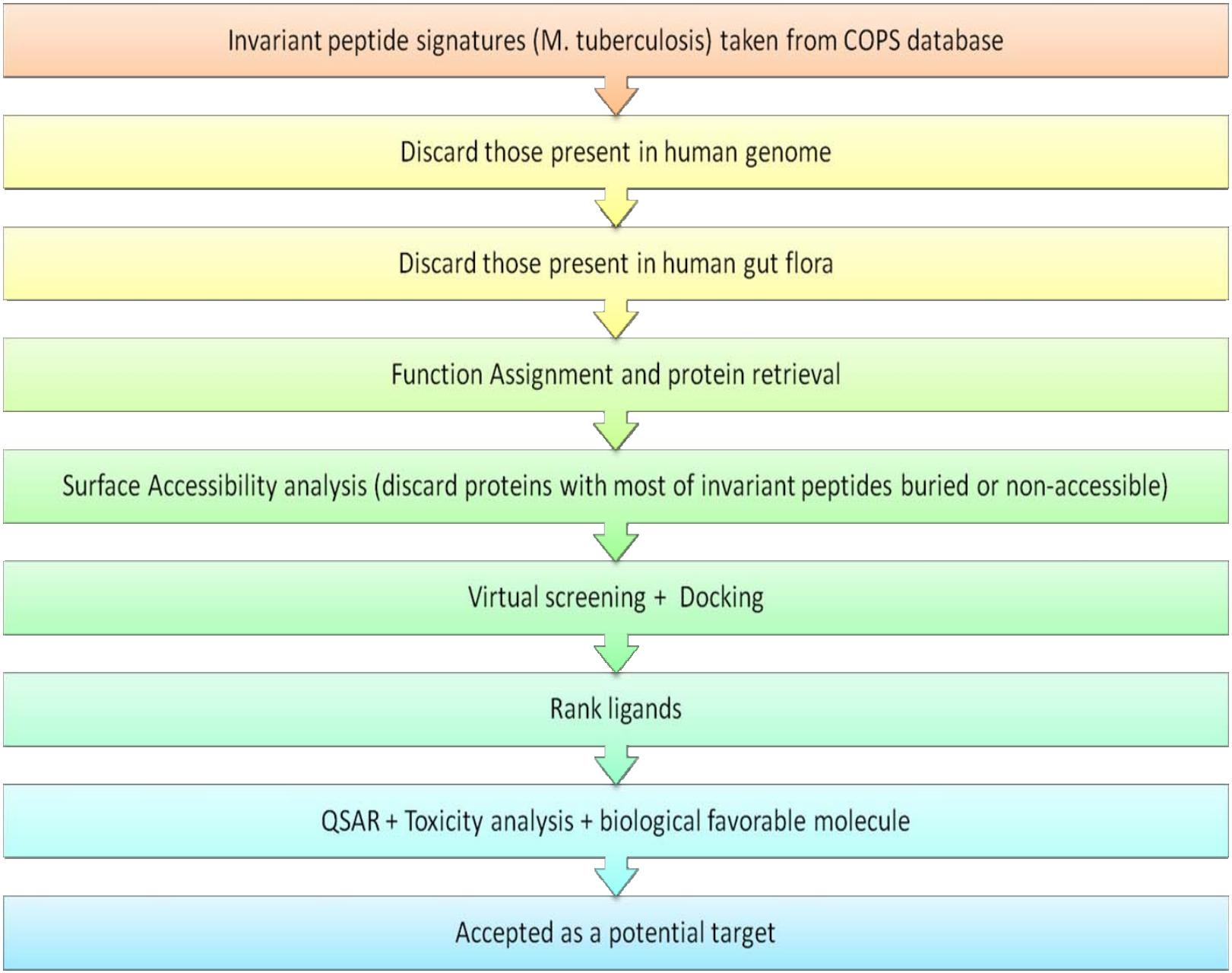
Prakash et al.,
J Mol Biol (2005)

Mycobacterium tuberculosis SysBorg

- ~64000 COPS-Path peptides serves as initial data
- Map peptides onto available PDBs (of *M. tuberculosis*)
- Surface analysis of peptides to check proximity to active site
- Peptides present in *M. tuberculosis*
  Protein Essential for survival
  (DEG & Sassetti *et al*., Mol Microbiol, 2003)
- Peptide absent in humans
- Peptide absent in gut flora (Gill *et al*, Science (2006), v. 312 pp. 1355)
- Initiate docking of ligand library with candidate target determined as above

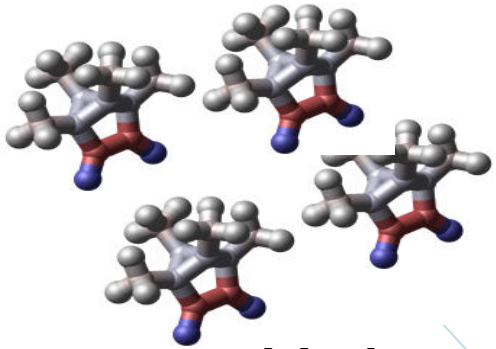
IGIB
Institute of Genomics & Integrative Biology

Summary: in silico steps for selection of potential target

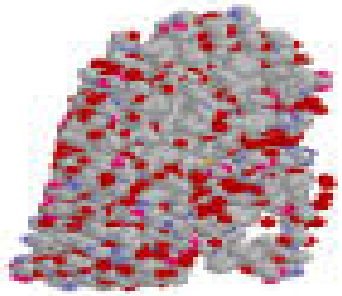


Target proteins shortlisted

- Enzymes involved in pantothenate biosynthesis (biosynthesis of CoA)
- Enzymes involved in the shikimate pathway (biosynthesis of aromatic amino acids)
- Cell division protein, FtsZ
- NAD-dependent DNA ligase
- SecA translocase



Ligand database



Target Protein

Molecular docking

Ligand docked into protein’s active site

A free database for virtual screening

ZINC

is not commercial

Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 4.6 million compounds in ready-to-dock, 3D formats. ZINC is provided by the Shoichet Laboratory in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF). To cite ZINC, please reference **Irwin and Shoichet, *J. Chem. Inf. Model. 2005;45(1):177-82*** **PDF**, **DOI**. We thank NIGMS for financial support (GM71896).

NEWS: Feb 4: The ZINC 8 release is coming - but still not ready. Read more
Dec 6: New Errata page for ZINC.
Nov 8: A new version of ZINC is scheduled for release in January 2008.
Jan 29: The 2007 ZINC release (ZINC7) is now the default version. To use the previous versions of ZINC please click here for ZINC 6 (2006). or click here for ZINC 5 (2005).

Caveat Emptor: We do not guarantee the quality of any molecule for any purpose and take no responsibility for errors arising from the use of this database. ZINC is provided in the hope that it will be useful, but you must use it at your own risk.

Please select:

- Vendors
- Documentation
- Search and Browse
- Property-filtered Subsets
- Upload
- Synthesis on Request
- Annotations
- User Subsets
- Uploaded Subsets

Compounds in ZINC may be ordered directly from vendors. Please visit their web sites and tell them you found it in ZINC! We thank these suppliers for making their catalogs available via collaborative agreements.

ZINC Subsets

Popular ZINC subsets are available for download below. ZINC may be used free of charge for research by individuals and institutions. **Whereas you are free to share the results of a ZINC search or a screen of molecules from ZINC, you may not redistribute major portions of ZINC without the express written permission of John Irwin.** Additional usage notes may be found below the table.

Subset Click to download	Compounds Click to browse	Last Update	Selection criteria and notes	Only one source	T<.9	T<.8	T<.7	T<.6	Creator / Sponsor
lead-like (#1)	972608	2007-01-20	p.xlogp<4 and p.xlogp> -2 and p.mwt < 350 and p.n_h_donors <= 3 and p.n_h_acceptors <= 6 and p.mwt > 150; Teague, Davis, Leeson, Oprea. Angew Chem Int Ed Engl. 1999 Dec 16:38(24):3743-3748.	478853	222130	83331	28337	9279	jji at cgl.ucsf.edu
fragment-like (#2)	62175	2007-01-20	p.xlogp <=3 and -2 <= p.xlogp and p.mwt <=250 and 150 <= p.mwt and p.rb <=3 and p.n_h_donors <=2 and p.n_h_acceptors <=4; Carr RA, Congreve M, Murray CW, Rees DC. Drug Discov Today. 2005 Jul 15:10(14):987	12998	29457	14945	7448	3465	jji at cgl.ucsf.edu
drug-like (#3)	2066906	2006-05-02	p.xlogp <= 5 and p.mwt <= 500 and p.mwt > 150 and p.rb < 8 and p.psa < 150 and p.n_h_acceptors <= 10; Lipinski, J Pharmacol Toxicol Methods. 2000 Jul-Aug:44(1):235-49.	867704	N/A	128085	40053	12158	jji at cgl.ucsf.edu
greasy-leads (#4)	713314	2006-05-02	p.xlogp <6 and p.xlogp>2 and p.mwt<350 and p.mwt>=150; Some targets seem to demand greasier compounds	273547	N/A	N/A	N/A	N/A	jji at cgl.ucsf.edu
big-n-greasy (#5)	577555	2006-05-02	p.xlogp<6 and p.xlogp>2 and p.mwt<600 and p.mwt>300; Some targets seem to demand larger, greasier compounds	873635	N/A	N/A	N/A	N/A	jji at cgl.ucsf.edu
all-purchasable (#6)	2667437	2006-05-02	; Purchasable chemical space to 400 Daltons	1161935	N/A	N/A	N/A	N/A	jji at cgl.ucsf.edu
newton-hit-like (#7)	643959	2006-05-02	p.xlogp>1 and p.xlogp<3 and p.mwt>200 and p.mwt<350; Roger Newton's (Maybridge) informed tweak of Teague/Oprea's Lead-like concept (ref Lecture at UCSF Dec 05)	37516	23707	12027	5766	2429	jji at cgl.ucsf.edu
vernalis-leads (#8)	643959	2006-05-02	p.xlogp<4 and p.xlogp> -2 and p.mwt < 350 and p.n_h_donors <= 3 and p.n_h_acceptors <= 6 and p.mwt > 150; Subset #1 above, with group for facile derivitization as per Hubbard et al (Vernalis) J Chem Inf Comput Sci. 2004 Mar-Apr:44(2):643-51.	70136	N/A	N/A	N/A	N/A	jji at cgl.ucsf.edu
vernalis-frags (#9)	49134	2006-05-02	p.xlogp <=3 and -2 <= p.xlogp and p.mwt <=250 and 150 <= p.mwt and p.rb<=3 and p.n_h_donors <=2 and p.n_h_acceptors <= 4; Subset #2 above, with group for facile derivitization as per Hubbard et al (Vernalis) J Chem Inf Comput Sci. 2004 Mar-Apr:44(2):643-51.	1130	1282	983	732	470	jji at cgl.ucsf.edu
everything (#10)	5627809	2007-03-06	; Purchasable and non-purchasable	3093415	N/A	N/A	N/A	N/A	jji@cgl.ucsf.edu
clean-leads (#11)	643959	2006-05-02	p.xlogp<4 and p.xlogp> -2 and p.mwt < 350 and p.n_h_donors <= 3 and p.n_h_acceptors <= 6 and p.mwt > 150; As Subset #1, but without 'yuck' compounds.	191429	N/A	N/A	N/A	N/A	jji at cgl.ucsf.edu
clean-fragments	49134	2006-	p.xlogp <=3 and -2 <= p.xlogp and p.mwt <=250 and 150 <= p.mwt and p.rb<=3 and p.n_h_donors <=2 and	4625	12657	5988	2947	1383	jji at

DrugBank Homepage - Windows Internet Explorer

http://www.drugbank.ca/

File Edit View Favorites Tools Help

DrugBank Homepage

DrugBank

-Home- Browse PharmaBrowse ChemQuery TextQuery SeqSearch Data Extractor Download and Stats

This project is supported by Genome Alberta & Genome Canada, a private, non-profit corporation whose mandate is to develop and implement a national strategy in genomics and proteomics research for the benefit of all Canadians. For this purpose, it has received $600 million in funding from the Canadian government. This project is also supported in part by GenomeQuest, Inc., an enterprise genomic information company serving the life science community.

Search DrugBank for: Submit Reset

☑ Common Name ☑ Synonym and Brand Name ☐ All Text Fields

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains nearly 4800 drug entries including >1,480 FDA-approved small molecule drugs, 128 FDA-approved biotech (protein/peptide) drugs, 71 nutraceuticals and >3,200 experimental drugs. Additionally, more than 2,500 non-redundant protein (i.e. drug target) sequences are linked to these FDA approved drug entries. Each DrugCard entry contains more than 100 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

Please cite: Wishart DS et al., *DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. 2007 Dec 11*

Users may query DrugBank in any number of ways. The simple text query (above) supports general text queries of the entire textual component of the database. Clicking on the Browse button (on the DrugBank navigation panel above) generates a tabular synopsis of

Decoy molecules from DrugBank

	A	B	C	D
1				
2	PDB ID	Protein Name	PDB Decoy	Drug Bank
3				
4	2GES	Pantothenate Kinase (CoaA)	COK_2GES / [(2R,3S,4R,5R)-5-(6-AMINO-9H-PURIN-9-YL)- 4-HYDROXY-3-(PHOSPHONOO	Pyruvic acid
5			COK_2GET	Bezafibrate
6			COK_2GEU	Dihydroxy-Beta-Alanine
7			COK_2GEV	Coenzyme A
8				Adenosine-5'-Diphosphate
9				Phosphoaminophosphonic Acid
10				BIOTINOL-5-AMP
11				
12	1MOP	Pantothenate Synthetase (PanC)	APC_1N2G (DIPHOSPHOMETHYLPHOSPHONIC ACID ADENOSYL ESTER)	Ethylene Glycol
13			PAJ_1N2H (PANTOYL ADENYLATE)	2,4-Dihydroxy-3,3-Dimethyl-Butyrate
14			APC_1N2B (DIPHOSPHOMETHYLPHOSPHONIC ACID ADENOSYL ESTER)	Alpha,Beta-Methyleneadenosine-5'-Triphosphate
15			PAJ_1N2I (PANTOYL ADENYLATE)	Pantoyl Adenylate
16				Beta-Alanine
22	1ZTB	Chorismate Synthase (AroF)	EPS_Ecoli_1ZTB / 5-[(1-CARBOXYVINYL)OXY]-4-HYDROXY-3-(PHOSPH	Sulfanilamide
23				Sulfacetamide
24				Ethylene Glycol
25				Formic Acid
26				Riboflavin Monophosphate
27				Cobalt Hexammine Ion
28				Pyridoxyl-N,O-Cycloserylamide-5-Monophosphate
29				

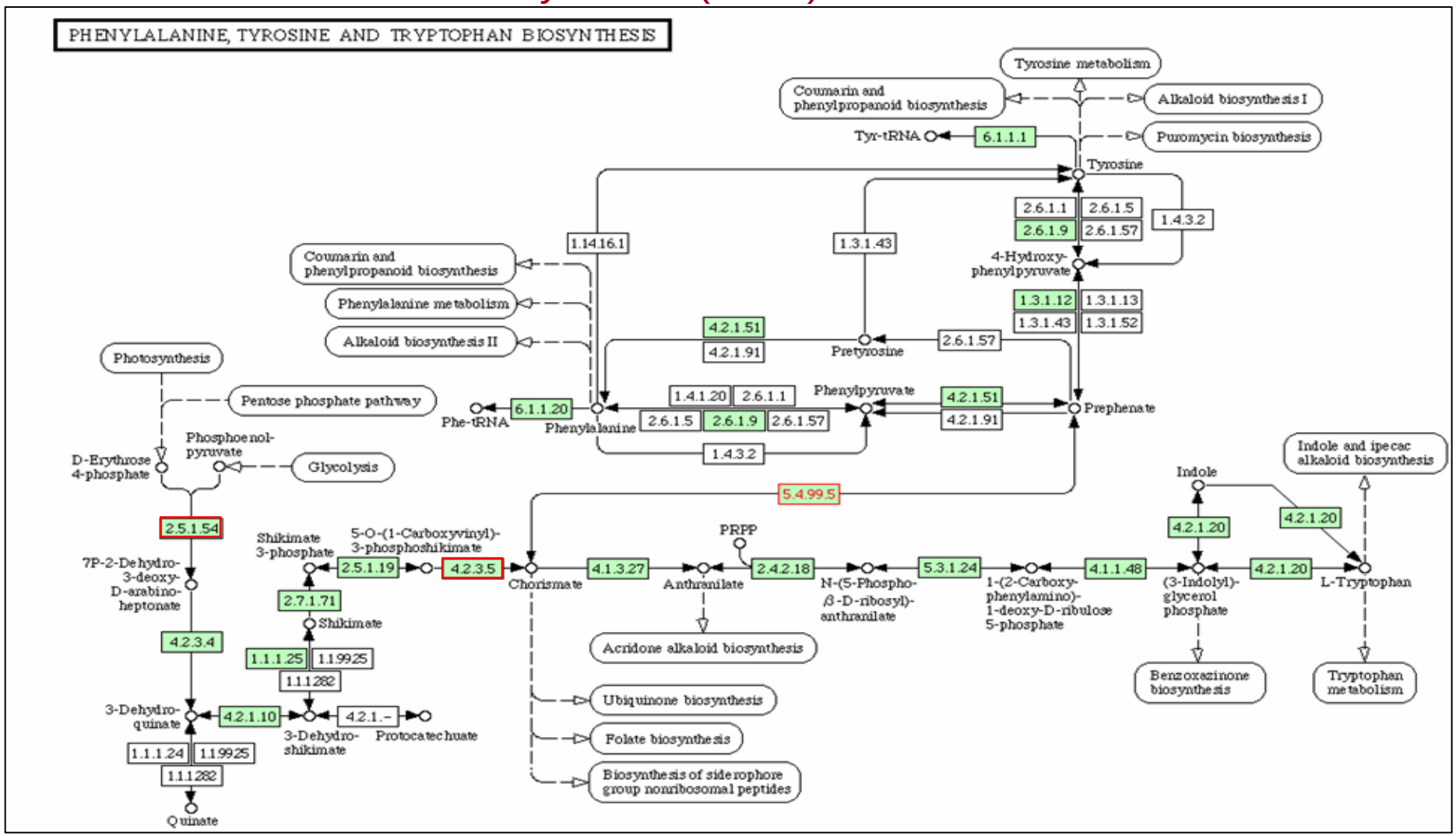
Evaluation of various docking and virtual screening softwares on HPC

- **FRED (OpenEye):**
- **Dock 6: MPI version**:
- **Autodock 3.05:**
- **Sybyl7.3(Surflex-Dock)**
- **eHits**

Biosynthesis of aromatic amino acids

Enzymes with invariant peptides :

Chorismate mutase,
Chorismate Synthase (AroF)
Synthase (AroG)



Docking of ligands with shikimate pathway proteins

AroF

Decoy molecule	Rank
Sulfanilamide	474
Sulfacetamide	2856
Ethylene Glycol	250
Formic Acid	231
EPS	20584
Total library size used for docking	> 2 million

AroG

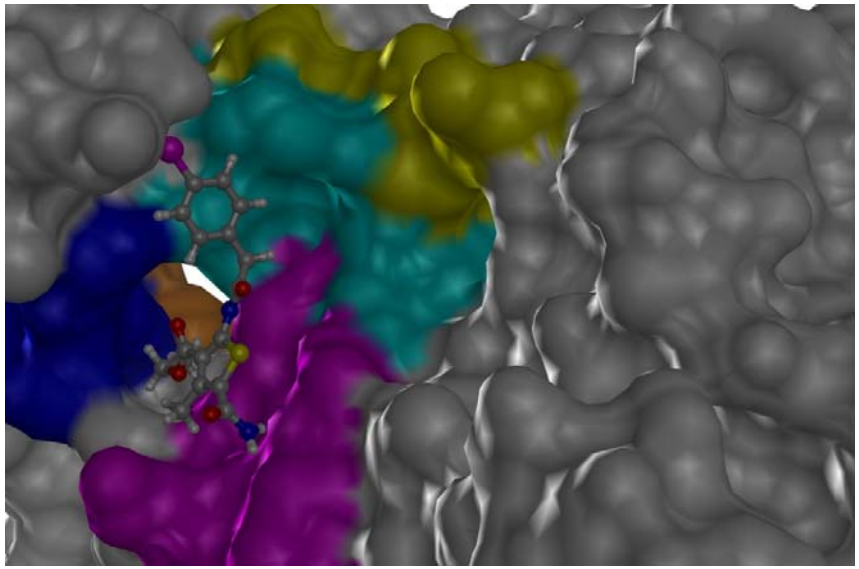
Decoy molecule	Rank
Propanol	367
Ethylene Glycol	371
2-sulfhydryl-Ethanol	377
Phospho-enol- Pyruvate	647
Total library size used for docking	> 2 million

Docking Results for Chorismate synthase (AroF) using FRED

1	Name	Consensus Score	Shapegauss	PLP	Chemgauss3	OEChemscore	Screenscore
2	ZINC08454842	87	-270.29	-33.15	-27.79	-22.10	-56.36
3	ZINC08454838	87	-268.20	-31.61	-37.93	-20.71	-68.56
4	ZINC00373283	79	-243.94	-28.13	-51.94	-20.49	-61.47
5	EPS.mol2	78	-315.05	-29.16	-35.84	-22.66	-73.95
6	ZINC08454844	65	-333.48	-36.49	-38.77	-23.21	-85.93
7	ZINC00372344	63	-289.57	-37.94	-46.18	-23.19	-78.60
8	ZINC08454846	60	-314.10	-37.02	-46.66	-24.84	-71.07
9	ZINC08454840	57	-337.40	-36.67	-40.16	-26.17	-81.30
10	ZINC00375133	57	-313.24	-42.38	-39.38	-23.86	-90.62
11	ZINC00651008	53	-346.11	-43.75	-40.49	-22.04	-92.27
12	ZINC00372523	48	-314.52	-41.71	-48.37	-26.06	-83.76
13	ZINC00689124	44	-357.36	-40.76	-45.41	-22.71	-96.92
14	ZINC01002757	33	-391.75	-47.56	-41.48	-25.29	-87.38
15	ZINC01002753	28	-384.97	-47.24	-44.79	-25.79	-97.24
16	ZINC00372455	27	-355.42	-46.49	-55.92	-30.76	-86.55
17	ZINC00375134	18	-379.91	-47.31	-53.08	-31.57	-95.27
18	ZINC08453733	17	-382.21	-46.23	-60.39	-27.49	-106.23
19	ZINC00675980	14	-388.36	-57.94	-50.09	-29.29	-96.88
20	ZINC00696019	7	-398.38	-55.43	-54.81	-27.88	-122.29

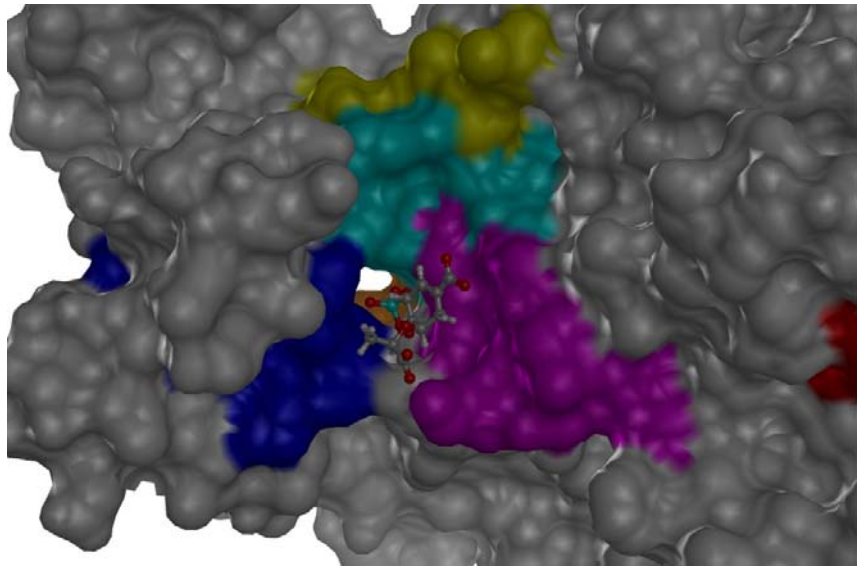
Protein-Ligand interactions of Chorismate synthase

Ligand: ZINC08454842 (CS: 87)

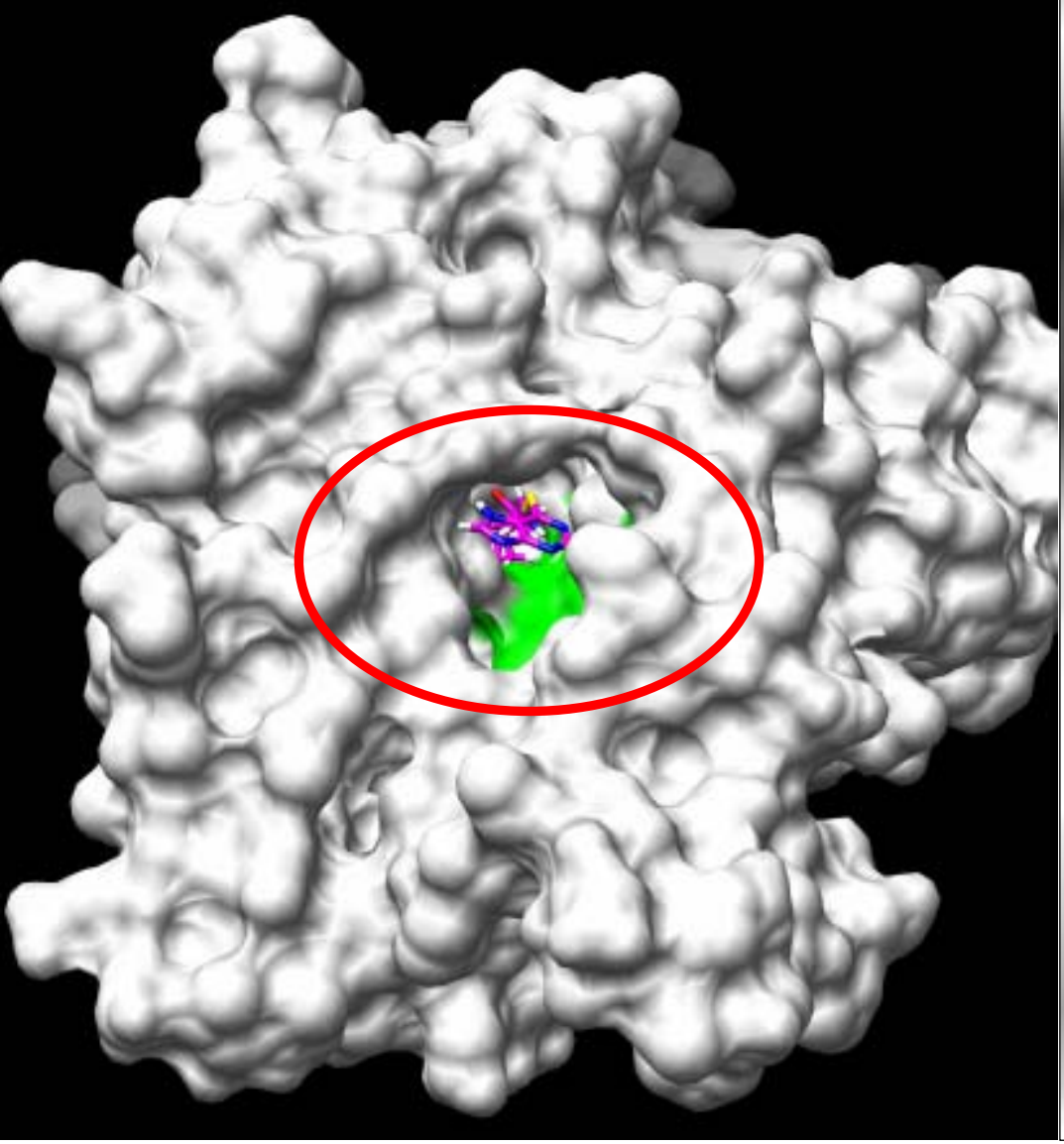


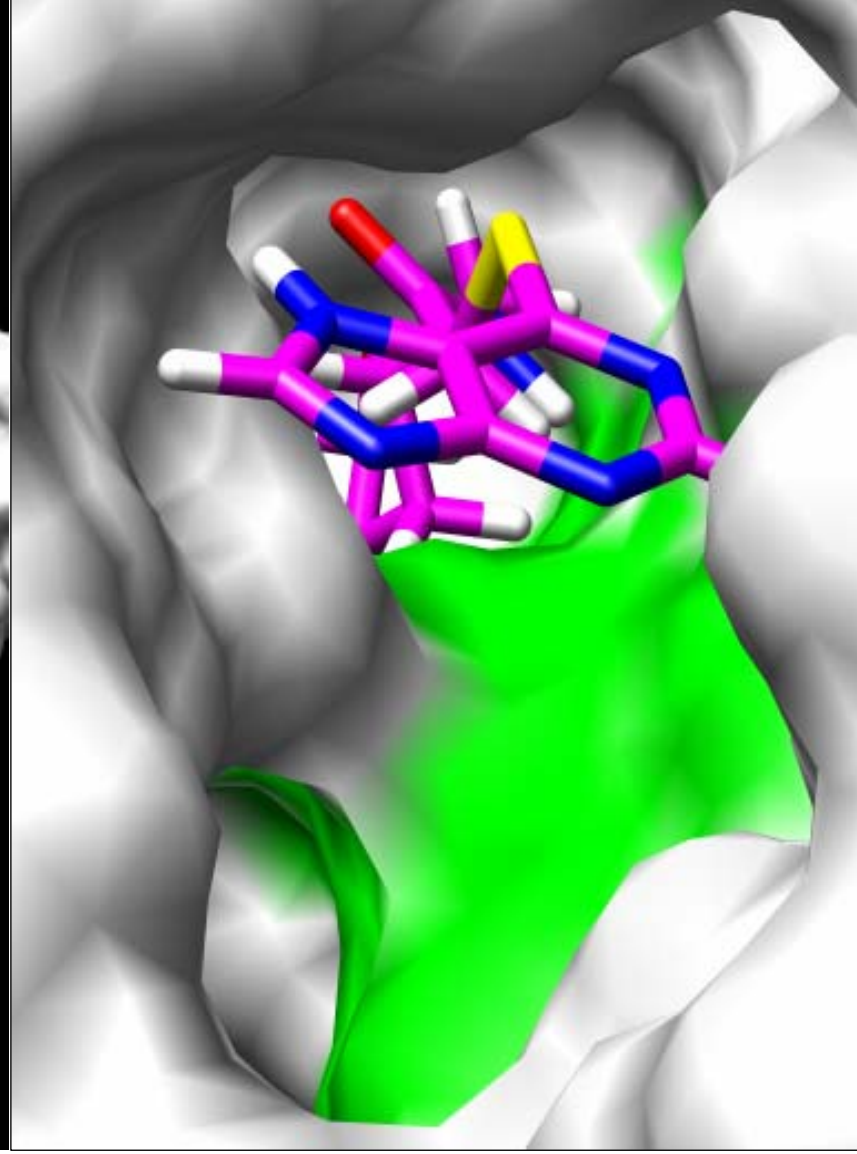
Chorismate synthase (AroF)

Ligand:EnolPyruvylShikimate-3-phosphate (CS: 78)



Chorismate synthase (AroF)





PDB : 1ZTB

Ligand : ZINC00033507