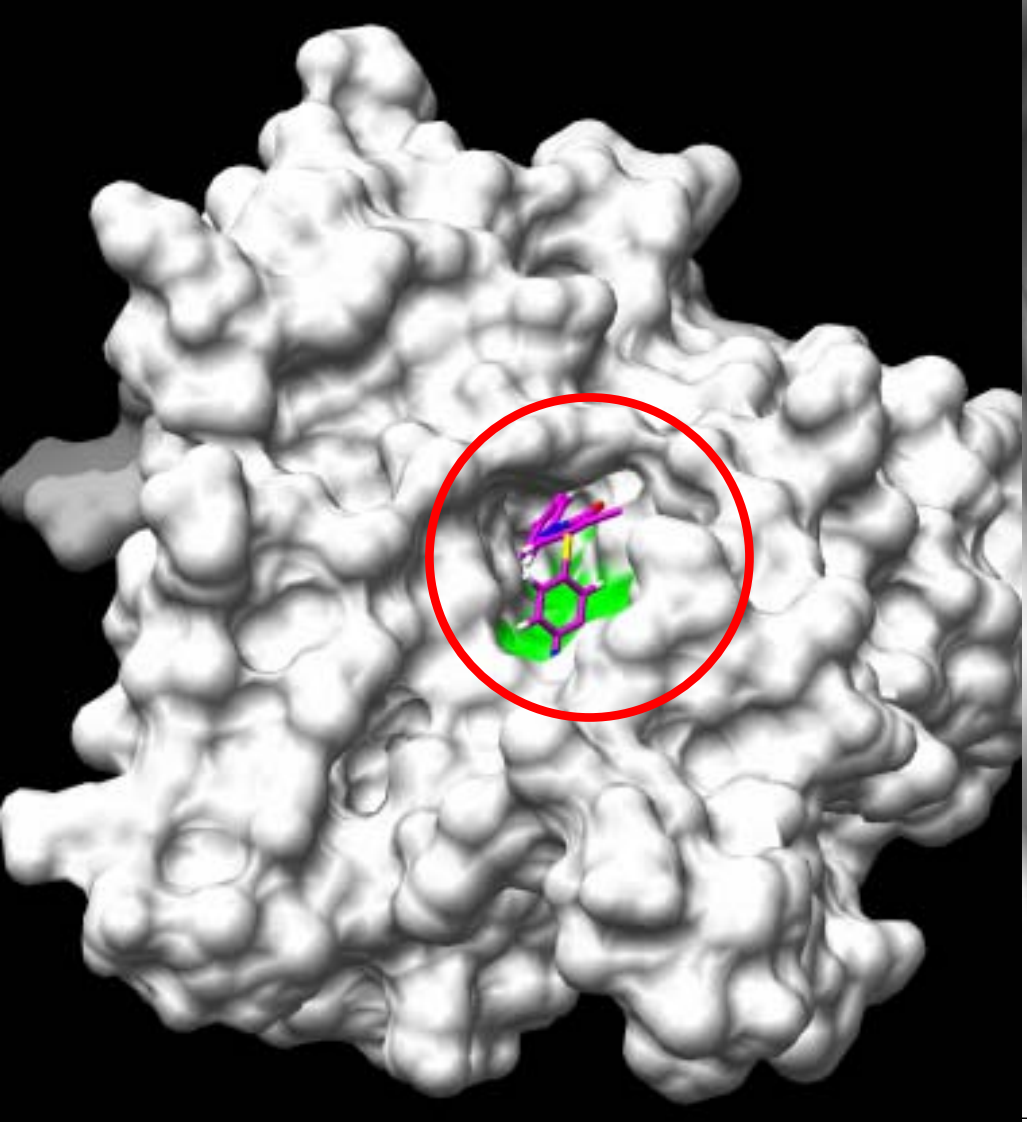
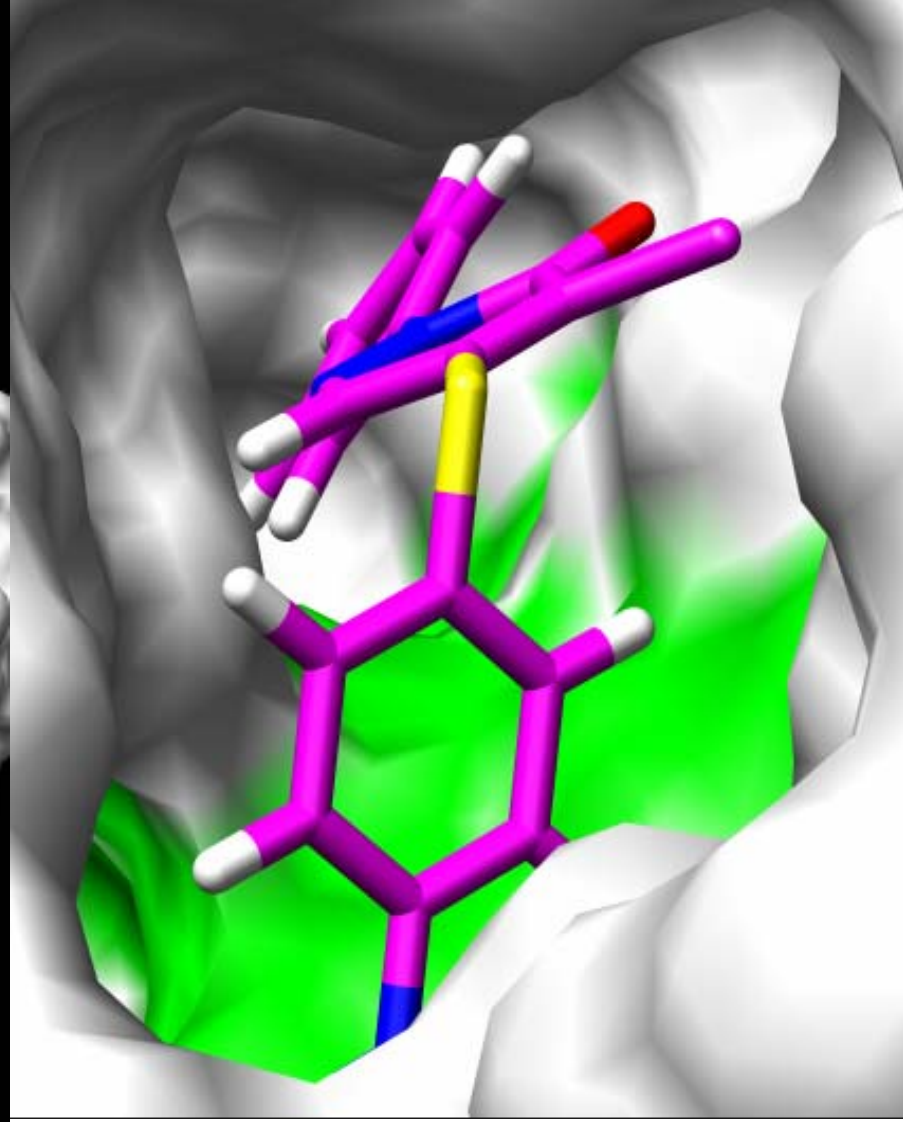




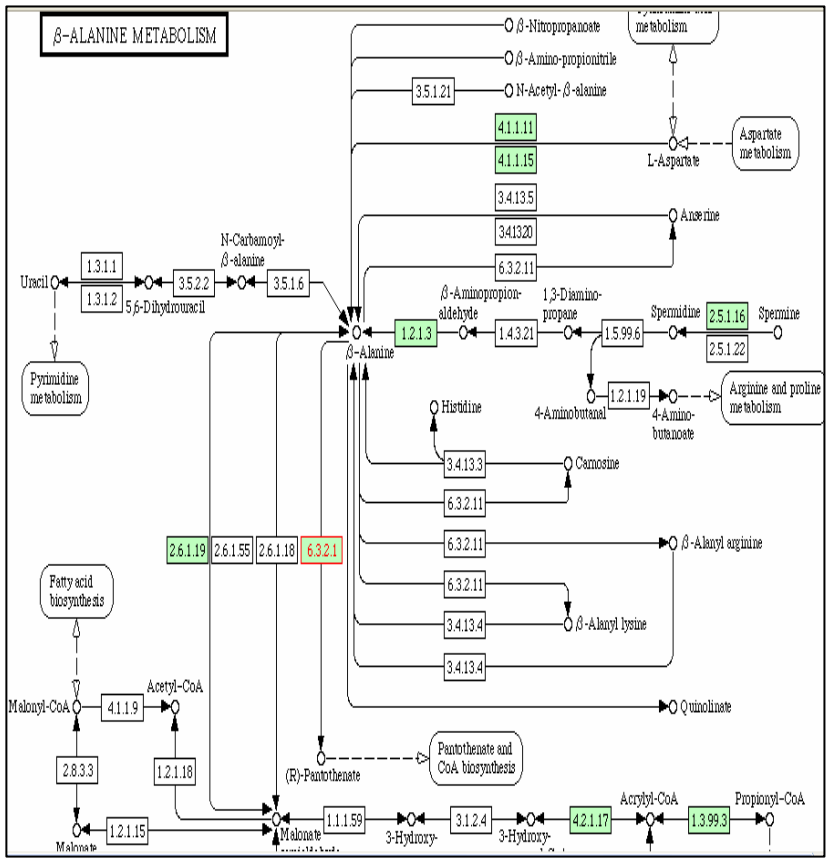
PDB : 1ZTB

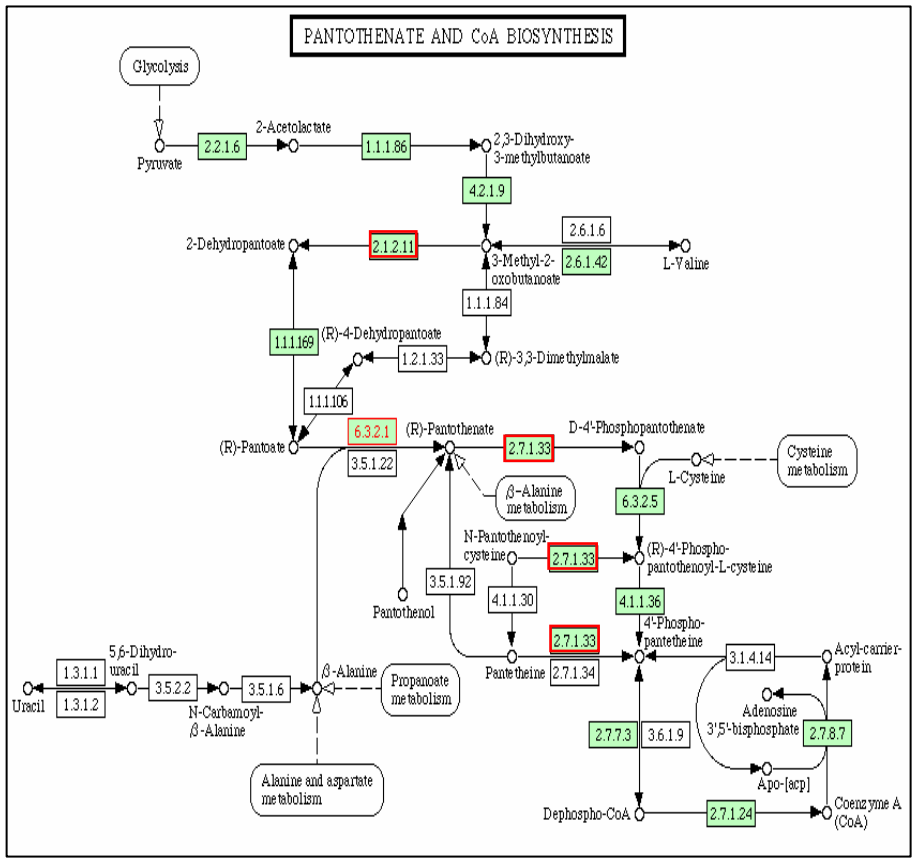
Ligand : ZINC00367031

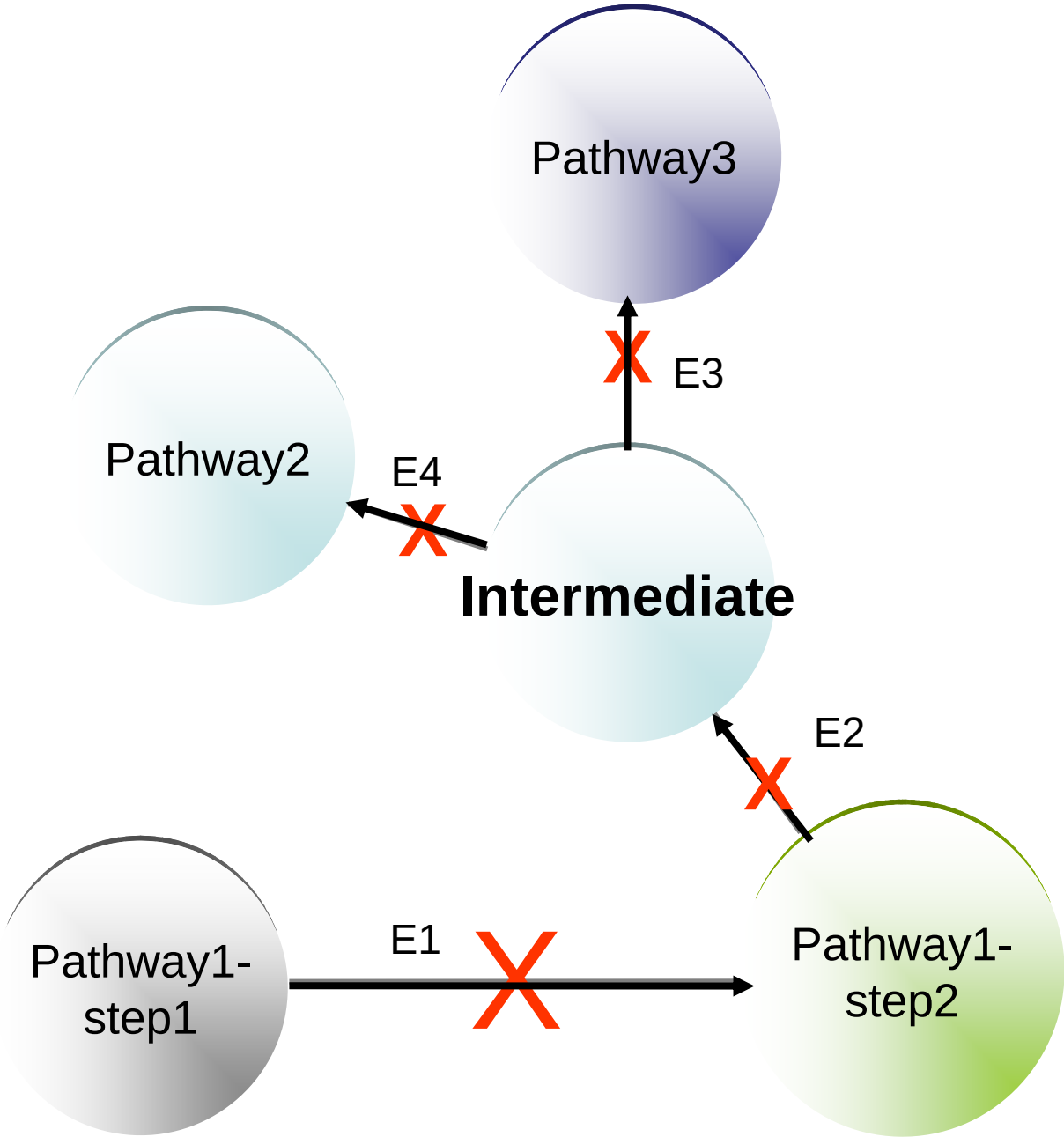
Pantothenate biosynthesis

Enzymes with invariant peptides :

Pantothenate kinase (CoaA),
Pantothenate Synthetase (PanC)
Methyltransferase (PanB)

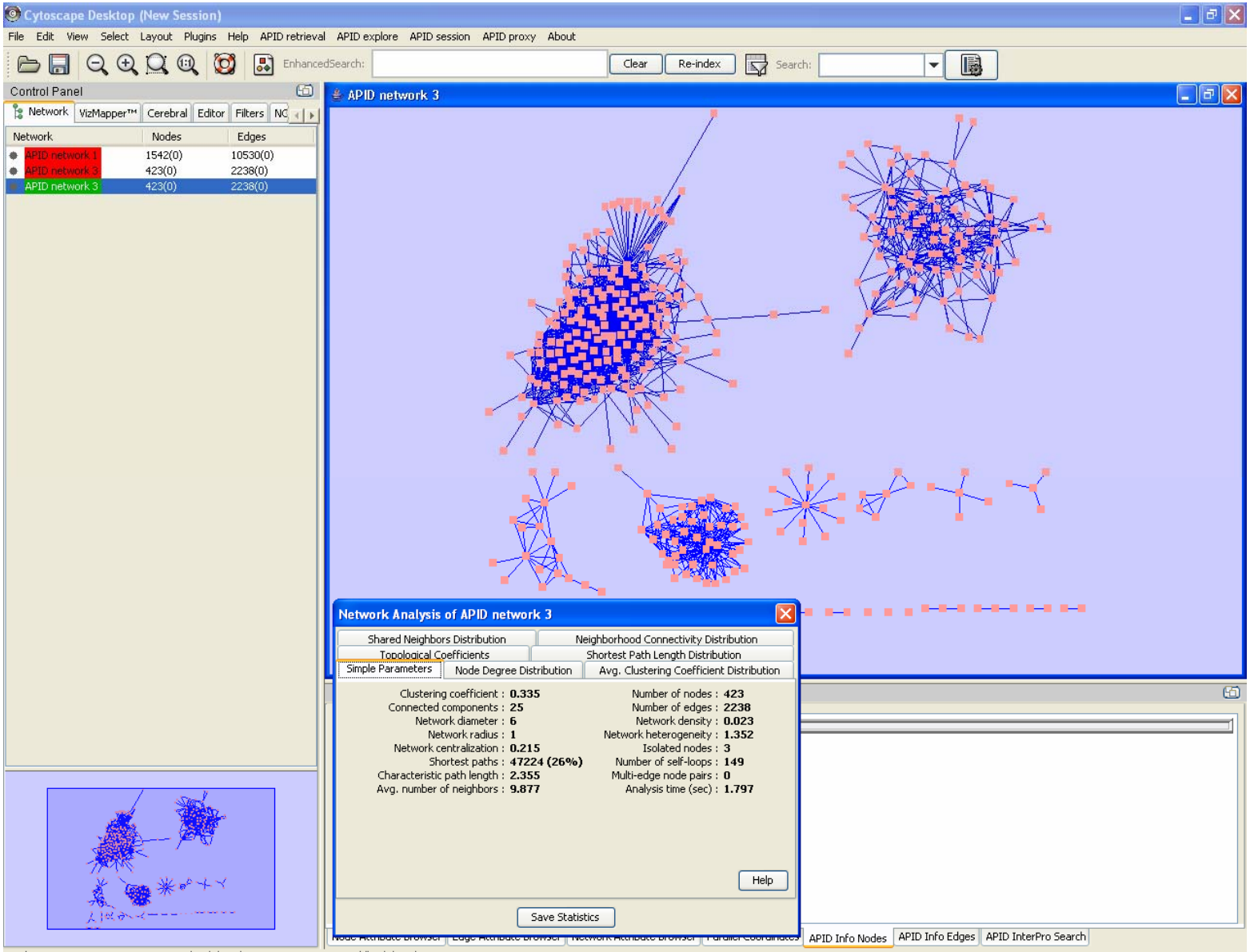





Pathway3
E3
Pathway2
E4
Intermediate
E2
Pathway1-
step1
E1
Pathway1-
step2

Visualizing protein-protein interactions to identify hub proteins

- ***Cytoscape***: an open source platform for ***visualizing*** molecular interaction networks and biological pathways and ***integrating*** these networks with annotations, gene expression profiles and other state data

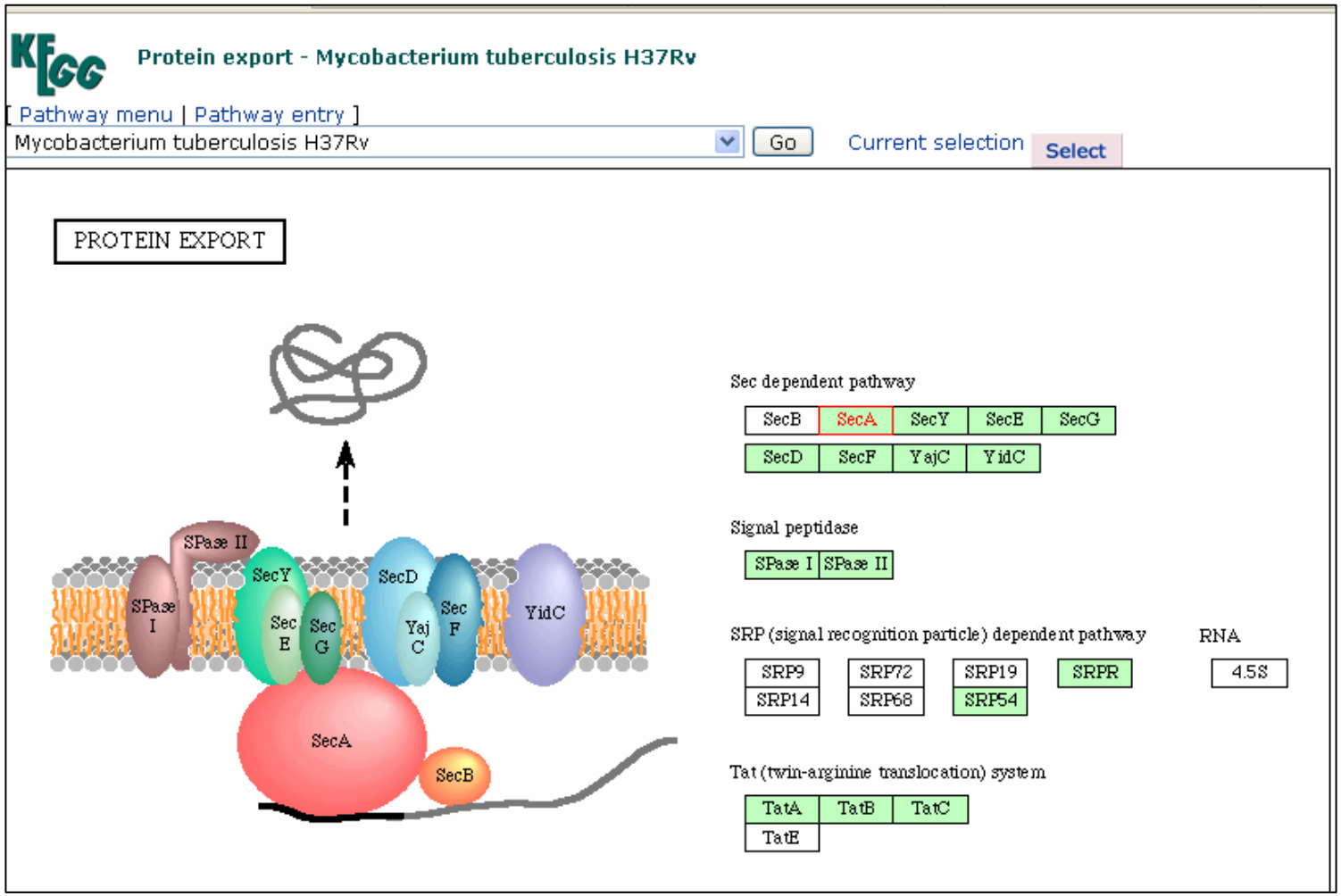

Cytoscape Desktop (New Session)
File Edit View Select Layout Plugins Help APID retrieval APID explore APID session APID proxy About
EnhancedSearch:
Clear
Re-index
Search:
Control Panel
Network VizMapper™ Cerebral Editor Filters
Network Nodes Edges
APID network 1 1542(0) 10530(0)
APID network 3 423(0) 2238(0)
APID network 3 423(0) 2238(0)
APID network 3
Network Analysis of APID network 3
Shared Neighbors Distribution
Neighborhood Connectivity Distribution
Topological Coefficients
Shortest Path Length Distribution
Simple Parameters
Node Degree Distribution
Avg. Clustering Coefficient Distribution
Clustering coefficient : 0.335
Connected components : 25
Network diameter : 6
Network radius : 1
Network centralization : 0.215
Shortest paths : 47224 (26%)
Characteristic path length : 2.355
Avg. number of neighbors : 9.877
Number of nodes : 423
Number of edges : 2238
Network density : 0.023
Network heterogeneity : 1.352
Isolated nodes : 3
Number of self-loops : 149
Multi-edge node pairs : 0
Analysis time (sec) : 1.797
Help
Save Statistics
APID Info Nodes
APID Info Edges
APID InterPro Search

Important hub proteins

- SecA1
- PanC
- GroEL2

- Are these good targets?

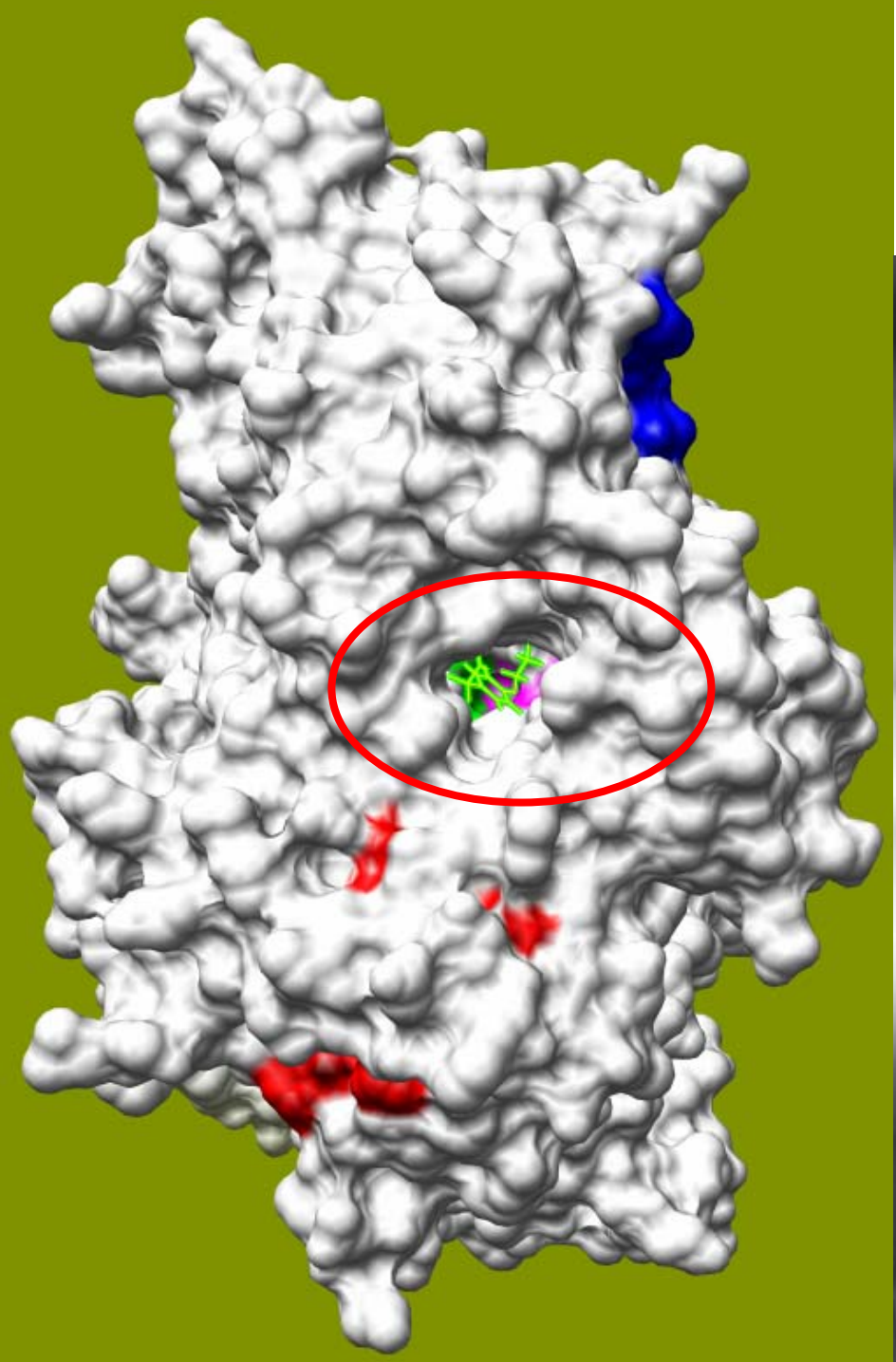
SecA1 - SecA Protein Translocation ATPase
KEGG Pathway - Protein export

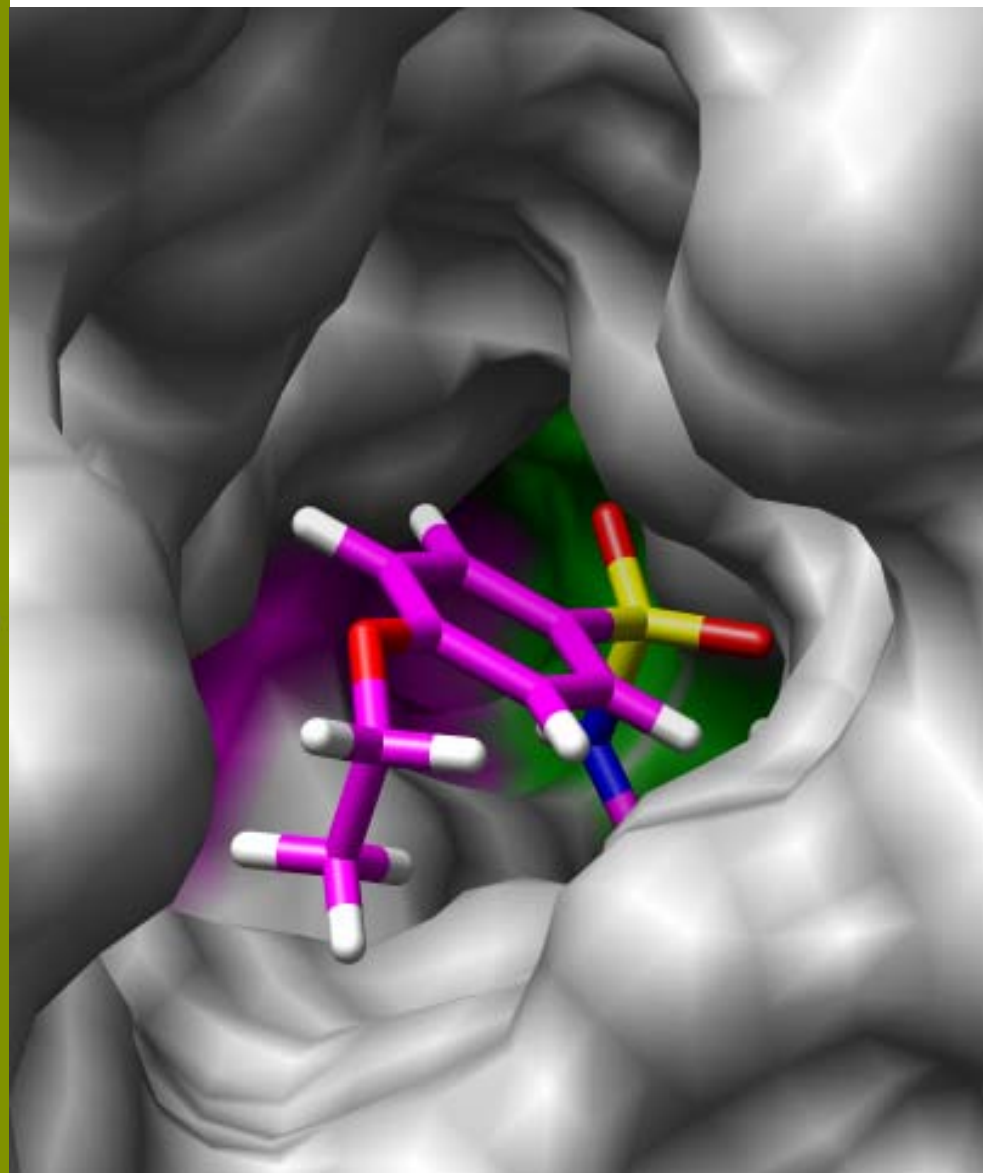


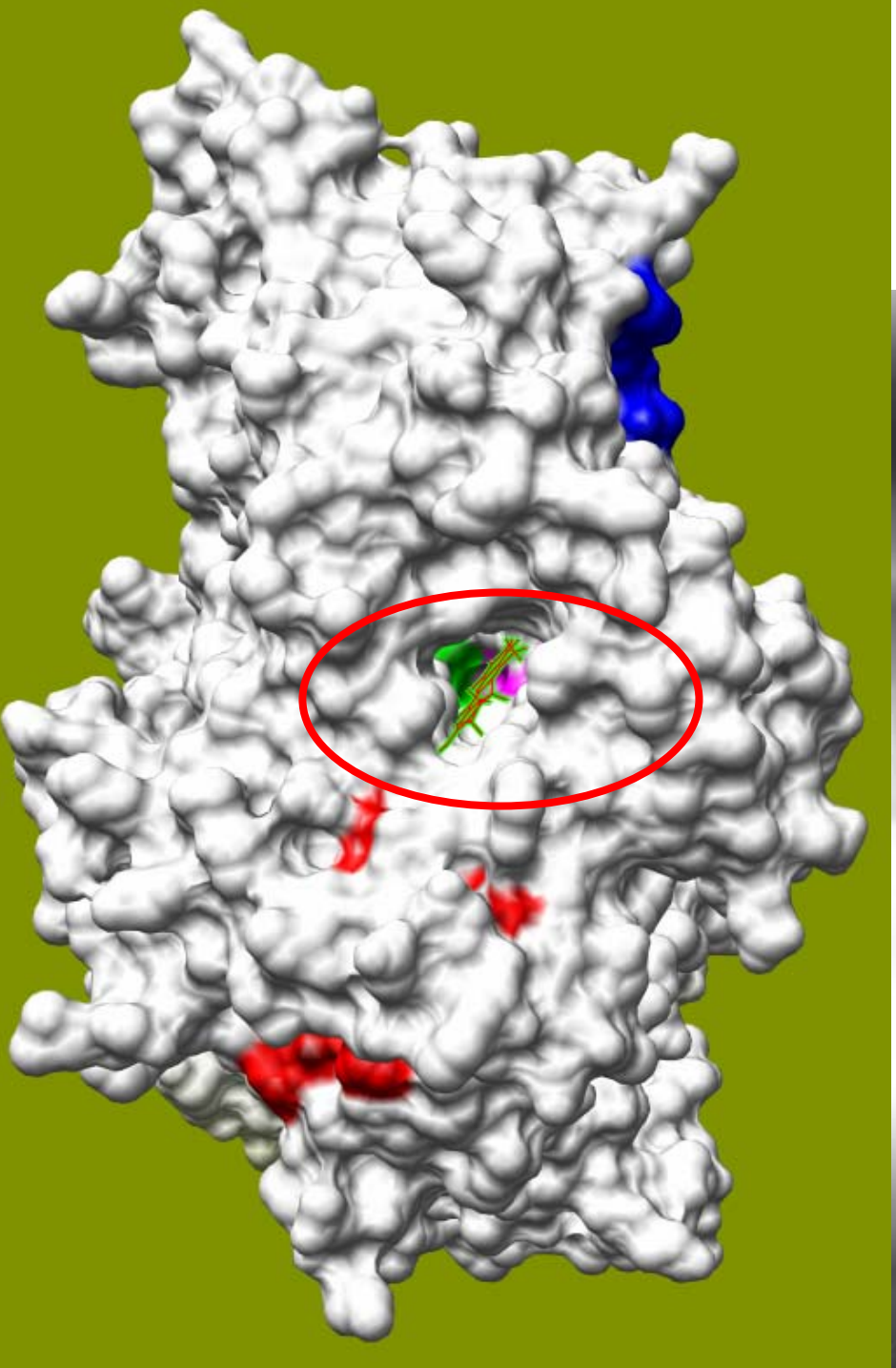
Docking of ligands against SecA1

Decoy molecule	Rank
fenofibrate	71185
Methyl Dihydrogen_phosphate	3129
Guanosine_5_monophosphate	27619
ADP	59442
Total library size used for docking	> 2 million

PDB : 1NKT
Ligand : ZINC00038956
Green Surface: VDEVDSILIDEARTPLIISGPAD
Magenta Surface: GVHIVTVNDYLAKRD







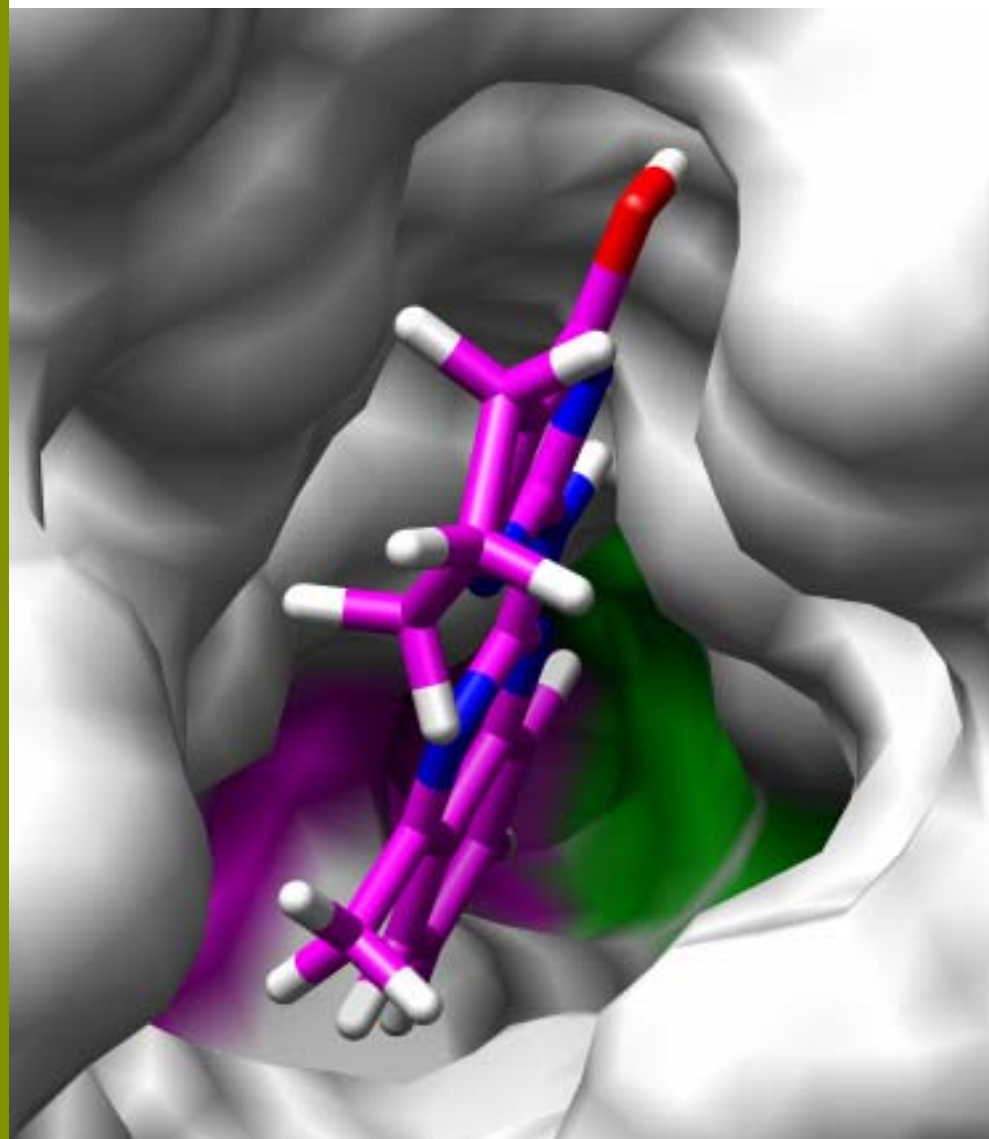
PDB : 1NKT

Ligand : ZINC00094215

Green Surface:

VDEVDSILIDEARTPLIISGPAD

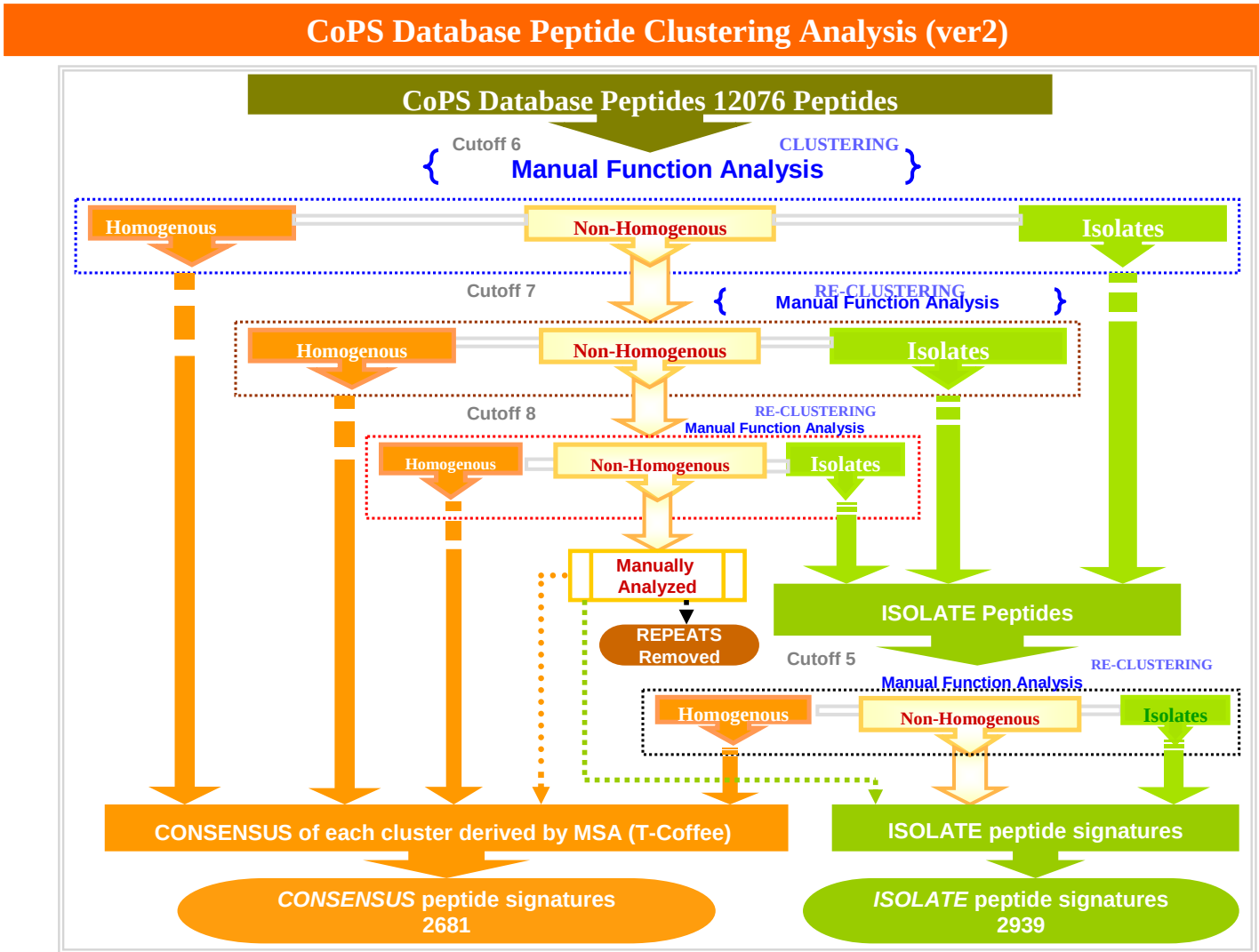
Magenta Surface: GVHIVTVNDYLAKRD



Pepcluster: to identify invariant peptide regions (and annotate proteins)

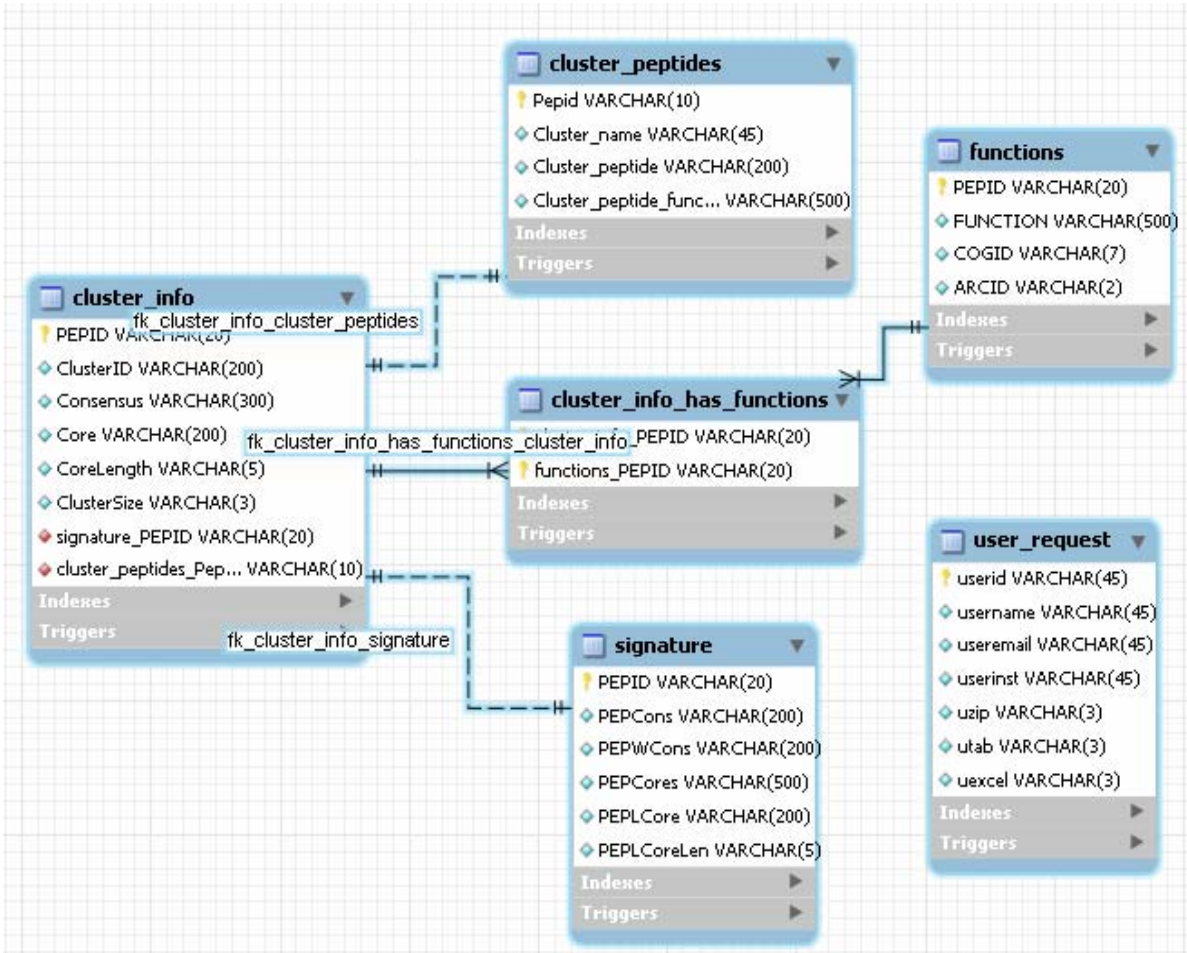
- PepCluster is a tool for rapid detection and identification of invariant peptide signatures of high functional importance
- PepCluster is derived from a initial peptide library of CoPS vers 2.0 containing 12076 peptides
  - Consists of 5620 invariant peptide signatures
  - offers a highly useful tool for assigning function to unannotated proteins based on signature sequences.
  - While assigning function to the unknown protein, provides detailed information on the sequence of signature peptide, its position and the functional category to which it belongs

PepCluster: Clustering Methodology



PepCluster contains a final set of 5620 signature peptides

Database Architecture



PepID specifies the Primary Key for all signatures

Pepcluster Webserver

http://pepcluster.igib.res.in

Search Utility and Output Display

```
>gi|1612|ref|conserved protein
LILISPAKTLDYQSPLTTTRYTLPELLDNTP
PQISTLMRISDLILISPAKTLDPPPPPPPPP
PADFOIUASDFOIUQERLKJQWEROIUASDF
LKJASDFOIUAAAAAAAAAAPGOIDDDDDIII
IASDFIOUQWERRRRRRRRRRRRRQWERKLJ
```

Results for Your Job:(test) [Tue Dec 4 13:07:48 2007]

"Your Input Fasta Sequence Details"

Title :: gi|1612|ref|conserved protein

Length :: 190 amino acids.

Signature Type Selected :: Core

"INVARIANT Peptide Signature Matches Found and Proposed Protein Annotations"

PEPCLUSTER ID	CORE SIGNATURE	CONSENSUS SIGNATURE	START POSITION	STOP POSITION	PROPOSED FUNCTION	COG's ID	ARC's ID
PEP0044	AGRAGR	QxAGRAGR	4	10	ATP-DEPENDENT DNA HELICASE RECQ	L	I
PEP0044	AGRAGR	QxAGRAGR	180	186	ATP-DEPENDENT DNA HELICASE RECQ	L	I
PEP1784	GRAGRD	qeiGRAGRDglp	5	11	ATP-DEPENDENT DNA HELICASE RECQ	L	I
PEP1784	GRAGRD	qeiGRAGRDglp	181	187	ATP-DEPENDENT DNA HELICASE RECQ	L	I
PEP2570	pppppppp	pppppppp	73	81	membrane protein	M	C

Functional annotations using PepCluster:

Control set: Organisms from original set of 52 bacterial genomes in CoPSv2

	M. tuberculosis		*E. coli*		*H. influenzae*	
	Frequency	**Percentage**	**Frequency**	**Percentage**	**Frequency**	**Percentage**
Total Unique Peptides Mapped	824	14.66 of total	2019	35.93 of total	1523	27.099 of total
Total Unique Proteins	494	12.36 of total	850	20.58 of total	557	33.61 of total
Total Matches	899		2175		1573	
Total Functional Mismatches	72	8.01 of mapped	140	6.44 of mapped	93	5.91 of mapped
Proteome size	3998		4131		1657	
Number of Hypo-theticals Annotated	24		11		56	

Examples of hypotheticals annotated

Signature Matched	Mtb Protein	Mtb H37Rv Annotation	Proposed function by Matched Signature
IPVIAAGGI	Rv0021c	Rv0021c: 322 aa – CONSERVED HYPOTHETICAL PROTEIN	HYPOTHETICAL PROTEIN, SIMILAR TO 2-NITROPROPANE DIOXYGENASE
PGTGKTTI	Rv0282	Rv0282: 631 aa - CONSERVED HYPOTHETICAL PROTEIN	ATPase , AAA family
ASASIAQVH	Rv0647c	Rv0647c: 488 aa - CONSERVED HYPOTHETICAL PROTEIN	UBIQUINONE BIOSYNTHESIS PROTEIN AARF, PUTATIVE
sTGPHLHfev	Rv0950c	Rv0950c: 332 aa - CONSERVED HYPOTHETICAL PROTEIN	cell wall-binding protein
IKVDREERPD	Rv1084	Rv1084: 673 aa - CONSERVED HYPOTHETICAL PROTEIN	Protein containing domain related to cellulase catalitic domain and thioredoxin domain
vphfEKMLYDna	Rv1084	Rv1084: 673 aa - CONSERVED HYPOTHETICAL PROTEIN	Protein containing domain related to cellulase catalitic domain and thioredoxin domain
nvYNLVTKRa	Rv1461	Rv1461: 846 aa -CONSERVED HYPOTHETICAL PROTEIN	putative iron-regulated ABC transporter
RYTTIQNW	Rv1461	Rv1461: 846 aa - CONSERVED HYPOTHETICAL PROTEIN	iron-regulated ABC-type transporter
GVGTGRFA	Rv2003c	Rv2003c: 285 aa - CONSERVED HYPOTHETICAL PROTEIN	SAM-dependent methyltransferase
RIAVNDEL	Rv2165c	Rv2165c: 396 aa - CONSERVED HYPOTHETICAL PROTEIN	methyltransferase
GPPGSGKTT	Rv2559c	Rv2559c: 452 aa - CONSERVED HYPOTHETICAL ALA, LEU, VAL RICH PROTEIN	REPLICATION FACTOR C SUBUNIT

Acknowledgements

- Prof. SK Brahmachari
- Dr. Aparna Basu
- GN Ramachandran Computing facility

Funding Resources

- CSIR, India
- DST-RFBR for Workshop

Students

- Amit Katiyar
- Bharati Dutta
- Suchir
- Vikram Kumar
- Nitin Singh

- **Спасибо!**