

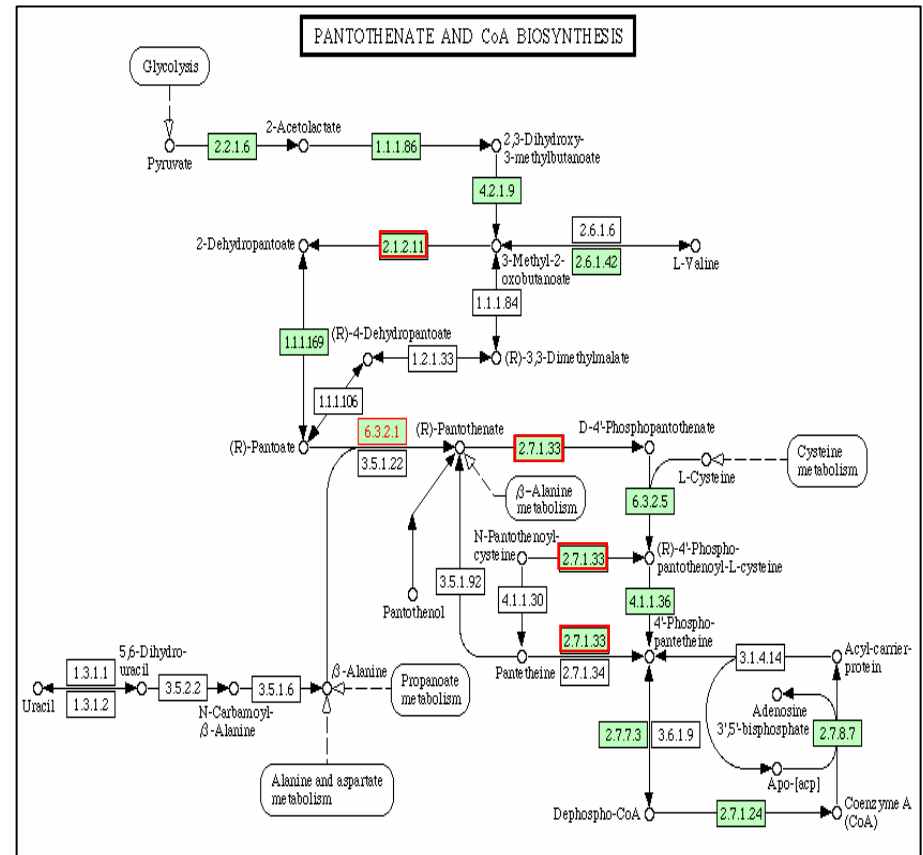
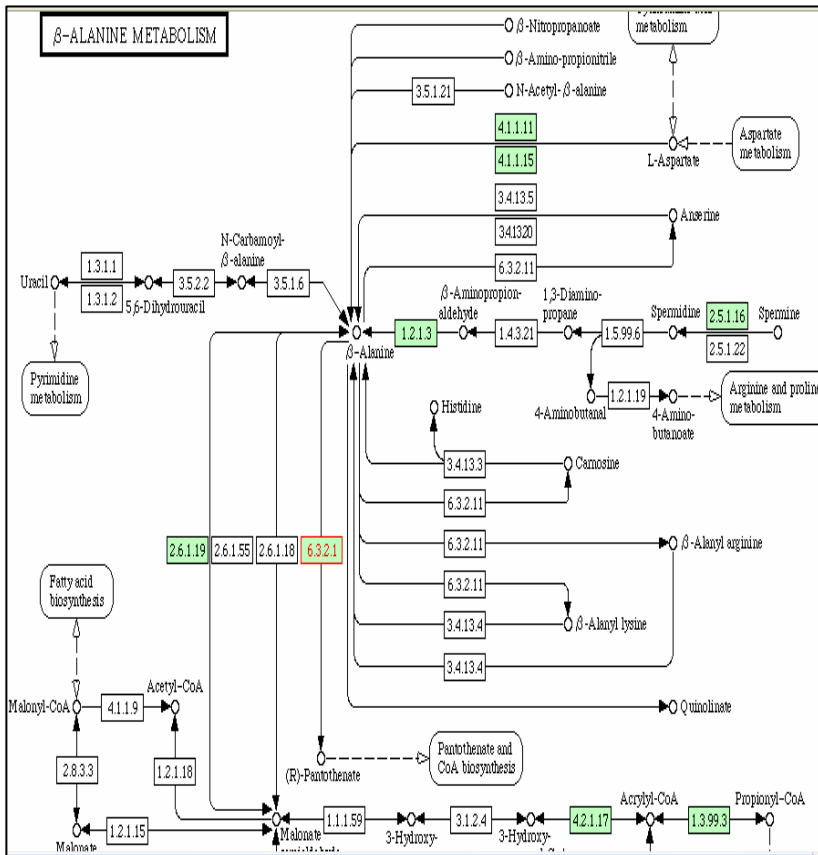
**PDB : 1ZTB**

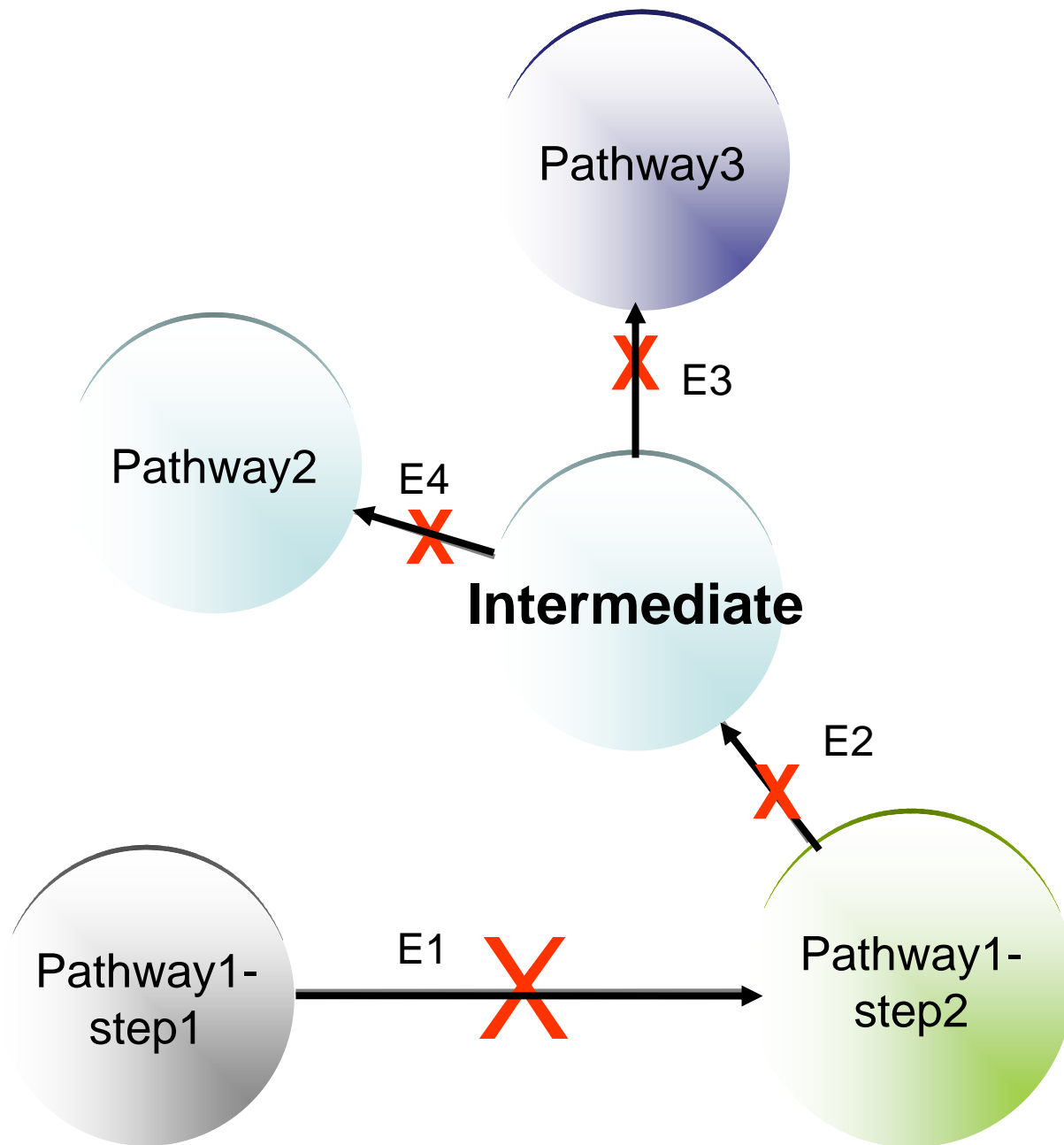
**Ligand : ZINC00367031**

# Pantothenate biosynthesis

Enzymes with invariant peptides :

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



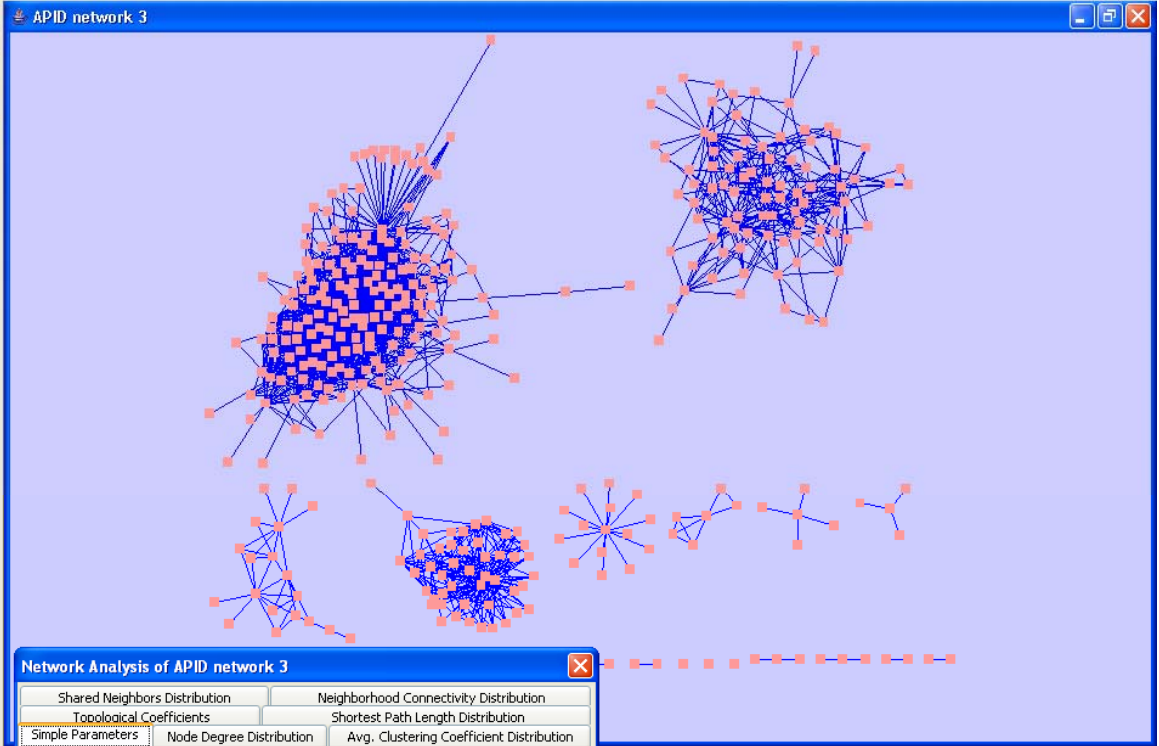


# Visualizing protein-protein interactions to identify hub proteins

- **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

Control Panel

Network	Nodes	Edges
APID network	1542(0)	10530(0)
APID network 3	423(0)	2238(0)
APID network 3	423(0)	2238(0)

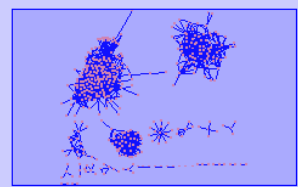


**Network Analysis of APID network 3**

Topological Coefficients	
Simple Parameters	Avg. Clustering Coefficient Distribution
Clustering coefficient : <b>0.335</b>	Number of nodes : <b>423</b>
Connected components : <b>25</b>	Number of edges : <b>2238</b>
Network diameter : <b>6</b>	Network density : <b>0.023</b>
Network radius : <b>1</b>	Network heterogeneity : <b>1.352</b>
Network centralization : <b>0.215</b>	Isolated nodes : <b>3</b>
Shortest paths : <b>47224 (26%)</b>	Number of self-loops : <b>149</b>
Characteristic path length : <b>2.355</b>	Multi-edge node pairs : <b>0</b>
Avg. number of neighbors : <b>9.877</b>	Analysis time (sec) : <b>1.797</b>

Help

Save Statistics



# Important hub proteins

- SecA1
  - PanC
  - GroEL2
- 
- Are these good targets?

# SecA1 - SecA Protein Translocation ATPase KEGG Pathway - Protein export

**KEGG** Protein export - *Mycobacterium tuberculosis* H37Rv

[Pathway menu](#) | [Pathway entry](#)

*Mycobacterium tuberculosis* H37Rv  Current selection

PROTEIN EXPORT

Sec dependent pathway

SecB	SecA	SecY	SecE	SecG
SecD	SecF	YajC	YidC	

Signal peptidase

SPase I	SPase II
---------	----------

SRP (signal recognition particle) dependent pathway

SRP9	SRF72	SRP19	SRPR	RNA
SRP14	SRP68	SRP54		4.5S

Tat (twin-arginine translocation) system

TatA	TatB	TatC
TatE		

# Docking of ligands against SecA1

<b>Decoy molecule</b>	<b>Rank</b>
<b>fenofibrate</b>	<b>71185</b>
<b>Methyl Dihydrogen_phosphate</b>	<b>3129</b>
<b>Guanosine_5_monophosphate</b>	<b>27619</b>
<b>ADP</b>	<b>59442</b>
<b>Total library size used for docking</b>	<b>&gt; 2 million</b>

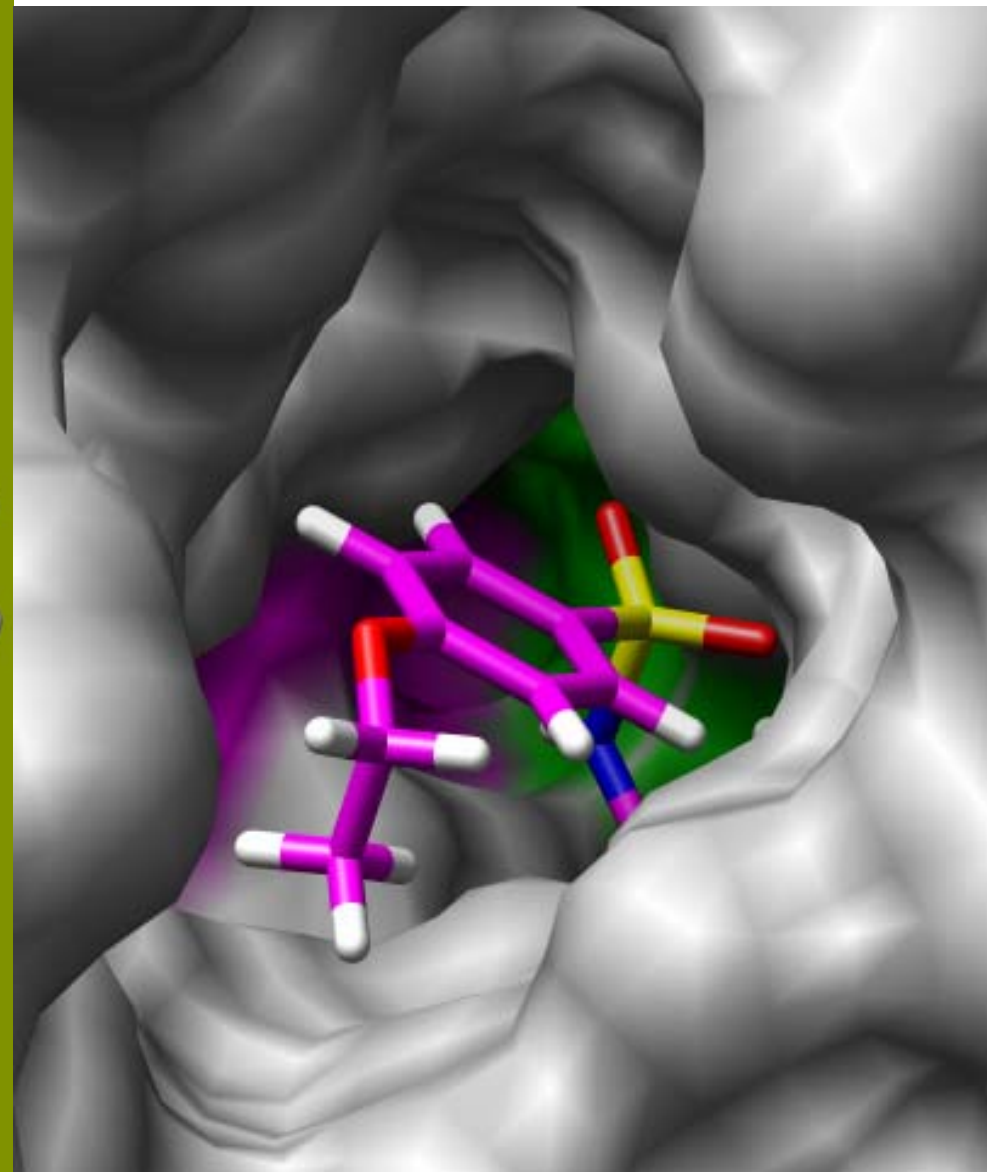
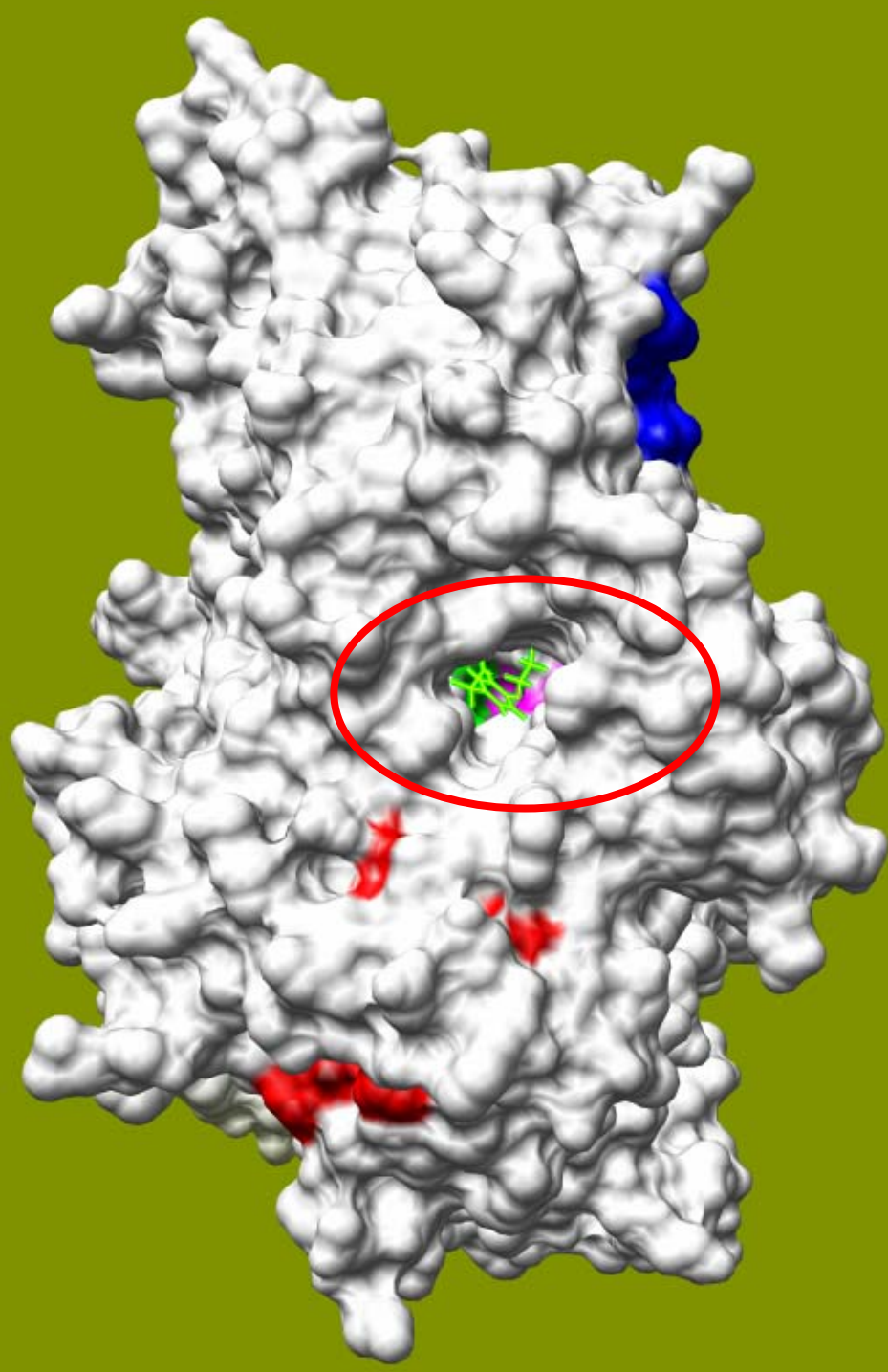


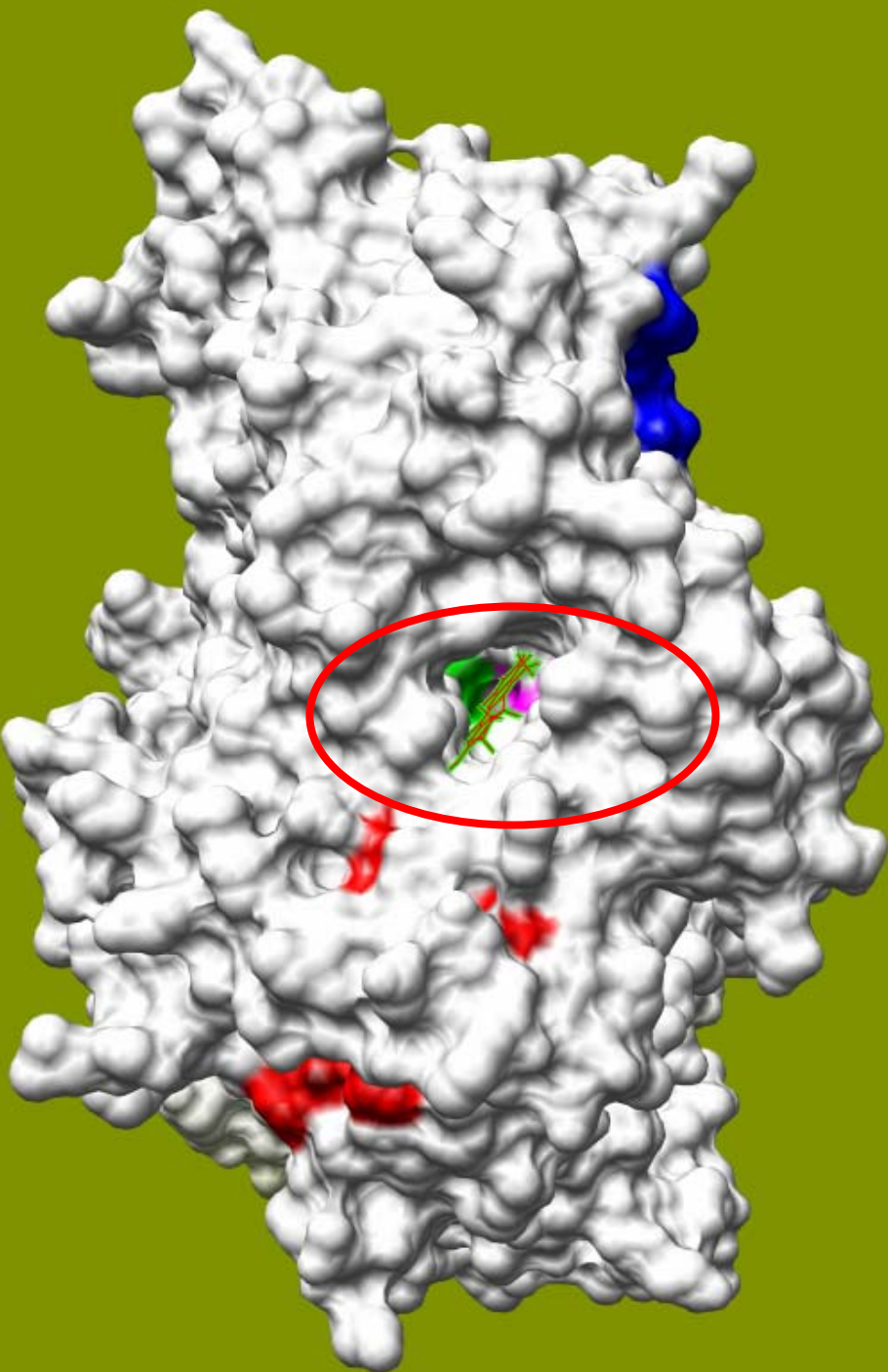
**PDB : 1NKT**

**Ligand : ZINC00038956**

**Green Surface: VDEVDSILIDEARTPLIISGPAD**

**Magenta Surface: GVHIVTVNDYLAKRD**





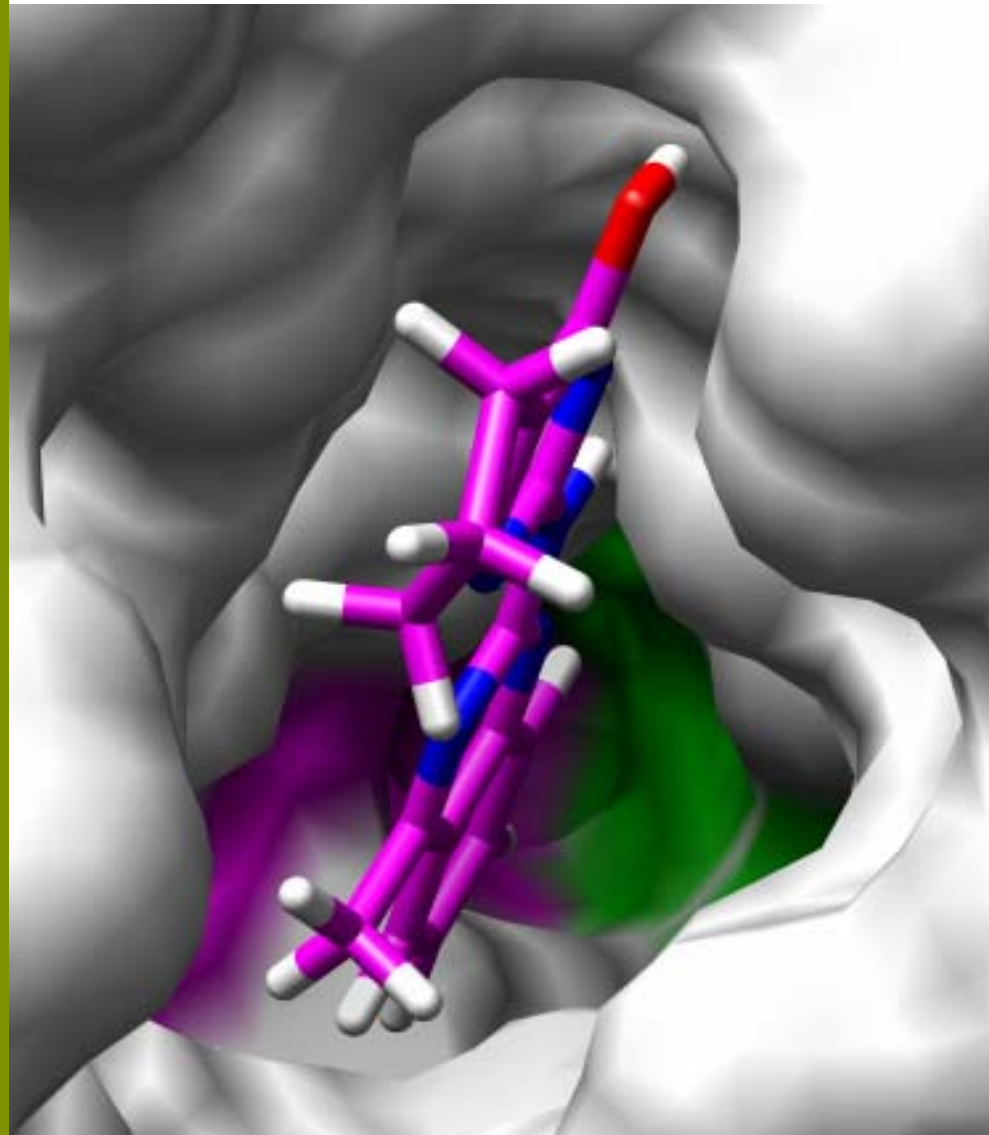
**PDB : 1NKT**

**Ligand : ZINC00094215**

**Green Surface:**

**VDEVDSILIDEARTPLIISGPAD**

**Magenta Surface: GVHIVTVNDYLAKRD**

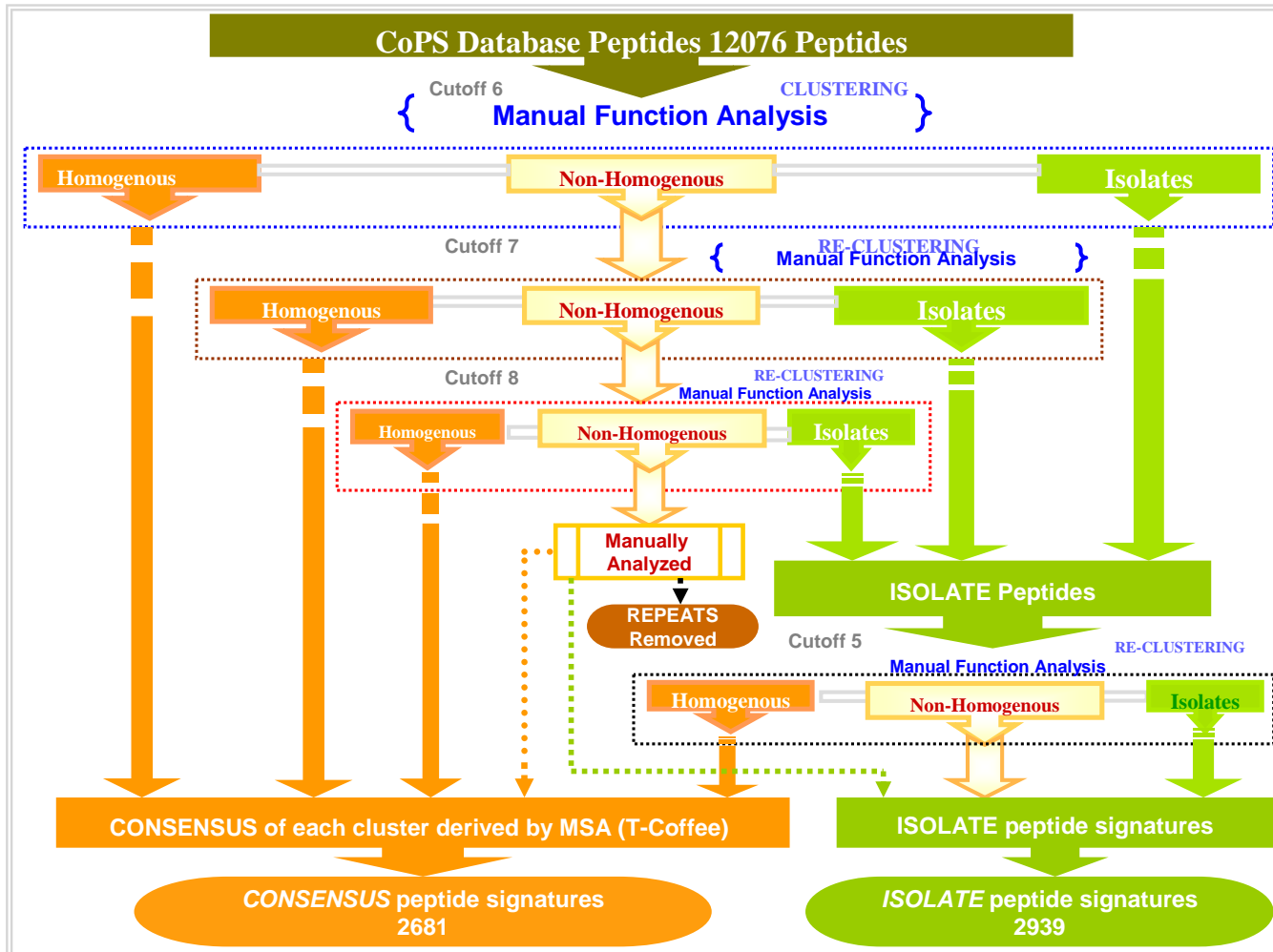


# Pepcluster: to identify invariant peptide regions (and annotate proteins)

- 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

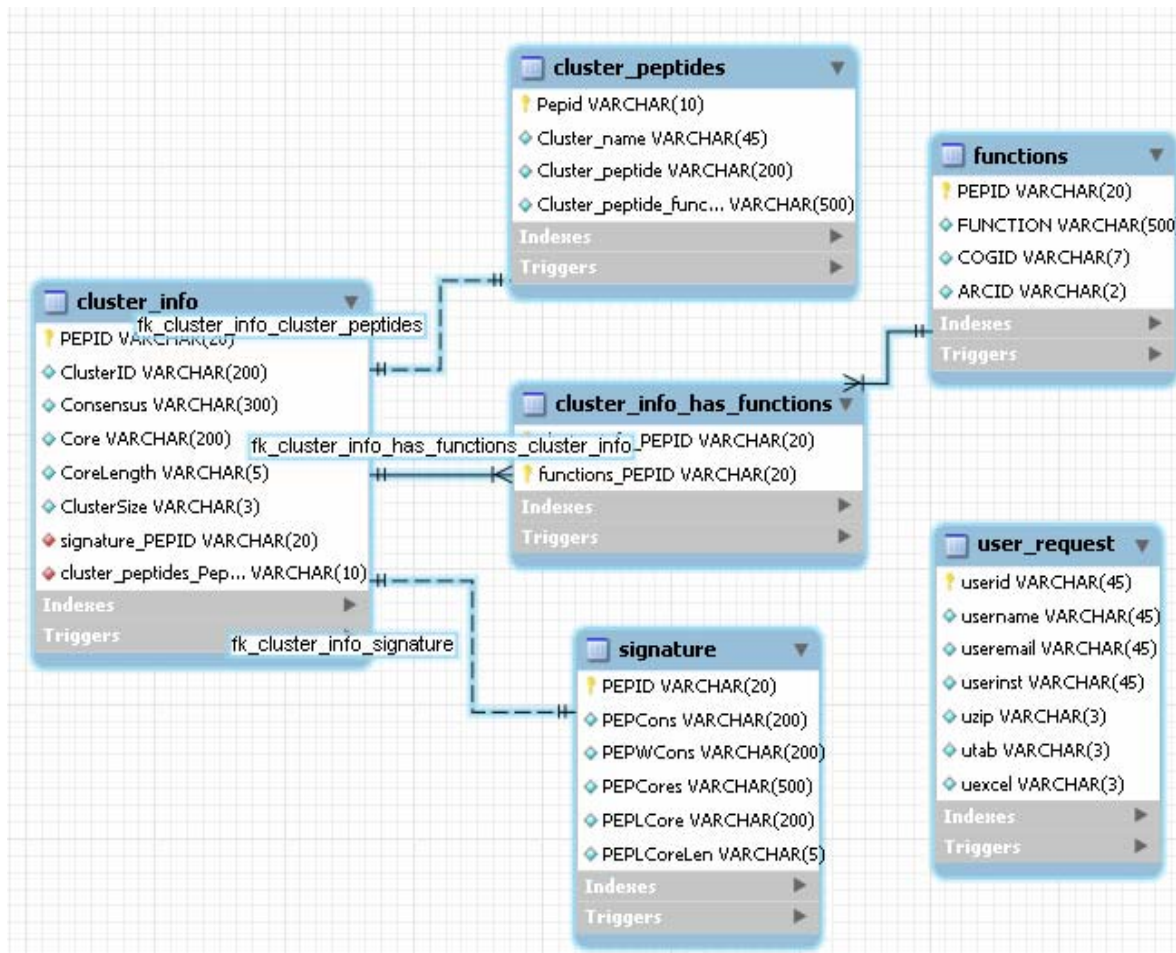
# PepCluster: Clustering Methodology

## CoPS Database Peptide Clustering Analysis (ver2)



**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

**Pep-Cluster: A Tool to predict Protein Function @ IGIB**

**Introduction:**

**Pep-Cluster** is a tool for prediction of the function of a protein based on the full or part of its amino acid sequence. A query sequence from a protein is matched against a collection of 5620 amino acid sequences in the Pep-Cluster library, that act as potential signatures for protein function. The signatures have been generated from short conserved sequences in the proteins of 52 bacteria and archaea, based on the evolutionary hypothesis that invariant or conserved peptides will be significantly correlated with protein function or domain function critical to the survival of the organism. Pep-Cluster predicts one or more possible functions for a query protein and assigns a score to each predicted function based on the number of matches and the degree of match to the signatures in the Pep-cluster library.

**Background:**

The potential signatures in Pep-cluster were generated using the **PLHOSIT**<sup>®</sup> Peptide Library based homology search tool, at IGIB, which searches for octapeptide sequences that are invariant in some significant proportion of proteins of the same function. The output of PLHOSIT for 52 bacterial genomes have been collected together in version 2 of the **CoPS Database** (Comprehensive Peptide Signature). In Pep-cluster, the 12076 invariant peptides in CoPSv2 have been gathered together into functionally homogeneous clusters, based on homology, with a corresponding consensus signatures for each cluster. The total number of cluster signatures and motifs (peptides that do not fall into a cluster) constitute the **5620 potential signatures** for protein function in the Pep-Cluster library. The cluster members capture variations in conserved sequences in proteins of the same function that arise from mutations in different organisms along the evolutionary chain. [Clustering Detail](#)..

**Citation:**  
"Pep-Cluster: A Tool for Prediction of Protein function"  
[http://dx.doi.org/10.1093/bioinformatics/btt288]

**Download Invariant Peptide Cluster Signature Database (Make Request)**

Username:

University/Research Lab:

Email ID:

Download files Separated by Function as FASTA format:  Check to Download

Download whole Database as Tab-Delimited Text files:  Check to Download

Download whole Database as EXCEL Workbook:  Check to Download

\*Submit to get a license academic copy of signatures on making a request \*

Prof. J.K. Bhadramani (bhadra@igib.res.in), Dr. Aparna Rana (rana@igib.res.in), Dr. Bhupesh Kanya (kanya@igib.res.in), Sachin (sachin@igib.res.in)

**Invariant Peptide Clusters Signature Annotation Tool**

Job Title:

Input/Paste Single Protein Sequence to Assign Function:

Invariant Peptide Signature Mapping

Browse file to upload Single FASTA Protein Sequence:

Select Functional Classification Category (Select Classification):  GPCR Classification  mC Classification

Select Consensus/Core Signature:

Prof. J.K. Bhadramani (bhadra@igib.res.in), Dr. Aparna Rana (rana@igib.res.in), Dr. Bhupesh Kanya (kanya@igib.res.in), Sachin (sachin@igib.res.in)

**PEPCLUSTER Support 1**

**Pep-Cluster Tool FAQ's:**

Q) What can Pep-cluster do ?  
A) Pep-Cluster is a tool for annotating the unknown protein sequence based on consensus signatures peptide obtained by iterative clustering methodology. It assigns a predicted function from the identified potential new signatures from clustering of original CoPS peptides.

Q) How to use Pep-Cluster ?  
A) It is really easy to use Pep-Cluster, just enter your protein sequence or input it by cut-paste or upload the protein sequence file in fasta format. Pep-Cluster will match all the possible peptide signatures and display the output results on submitting the query.

Following is a step by step visual representation of achieving the same:

Step1: Type in a Job Title

Job Title:

Step2: Paste your sequence OR Upload FASTA File

Input/Paste Single Protein Sequence to Assign Function:

Invariant Peptide Signature Mapping

Browse file to upload Single FASTA Protein Sequence:

Prof. J.K. Bhadramani (bhadra@igib.res.in), Dr. Aparna Rana (rana@igib.res.in), Dr. Bhupesh Kanya (kanya@igib.res.in), Sachin (sachin@igib.res.in)

# Search Utility and Output Display



```
>gi|1612|ref|conserved protein
LILISPAKTLDYQSPLTTTRYTLPELLDNTP
PQISTLMRISDLILISPAKTLDP PPPPPPPP
PADFOIUASDFOIUQERLKJQWEROIUASDF
LKJASDFOIUAAAAAAAAA PGOIDDDDDII
IASDFIOUQWERRRRRRRRRRRRRQWERKIJ
```

Results for Your Job:(test) [Tue Dec 4 13:07:48 2007]							
<b>"Your Input Fasta Sequence Details"</b>							
		Title	::	gi 1612 ref conserved protein			
		Length	::	190 amino acids.			
		Signature Type Selected	::	Core			
<b>"INVARIANT Peptide Signature Matches Found and Proposed Protein Annotations"</b>							
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	GRAGR	qeIGRAGRdglp	5	11	ATP-DEPENDENT DNA HELICASE RECQ	L	I
PEP1784	GRAGR	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

	<i>M. tuberculosis</i>		<i>E. coli</i>		<i>H. influenzae</i>	
	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 Hypotheticals Annotated	24		11		56	



# Examples of hypotheticals annotated

<b>Signature Matched</b>	<b>Mtb Protein</b>	<b>Mtb H37Rv Annotation</b>	<b>Proposed function by Matched Signature</b>
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!**