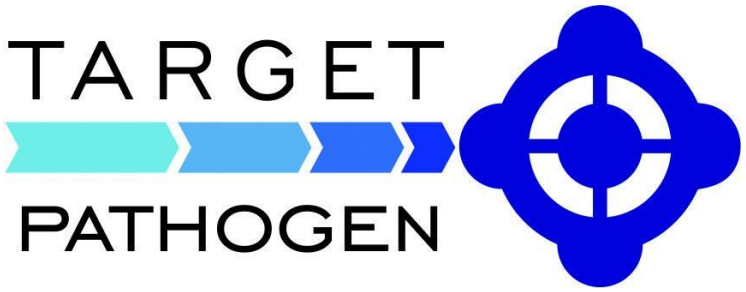


A Structural bioinformatic approach to prioritize drug targets in pathogens.

User Guide

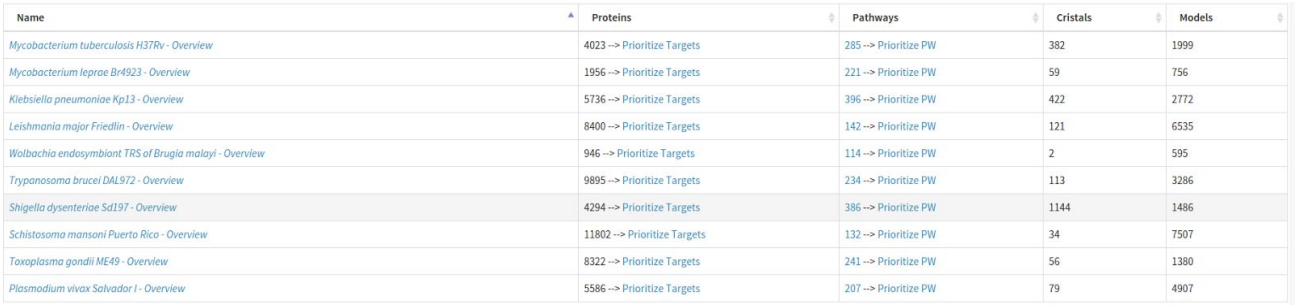
Using Target-Pathogen

The following sections will guide users with examples to browse available genomic information, to obtain a ranked list of putative drug targets and to choose promising pathways from the drug discovery point of view.

Browsing the available genomic information in Target-Pathogen.

The genome browser can be accessed and queried using the web interface at http://target.sbg.qb.fcen.uba.ar/patho/.

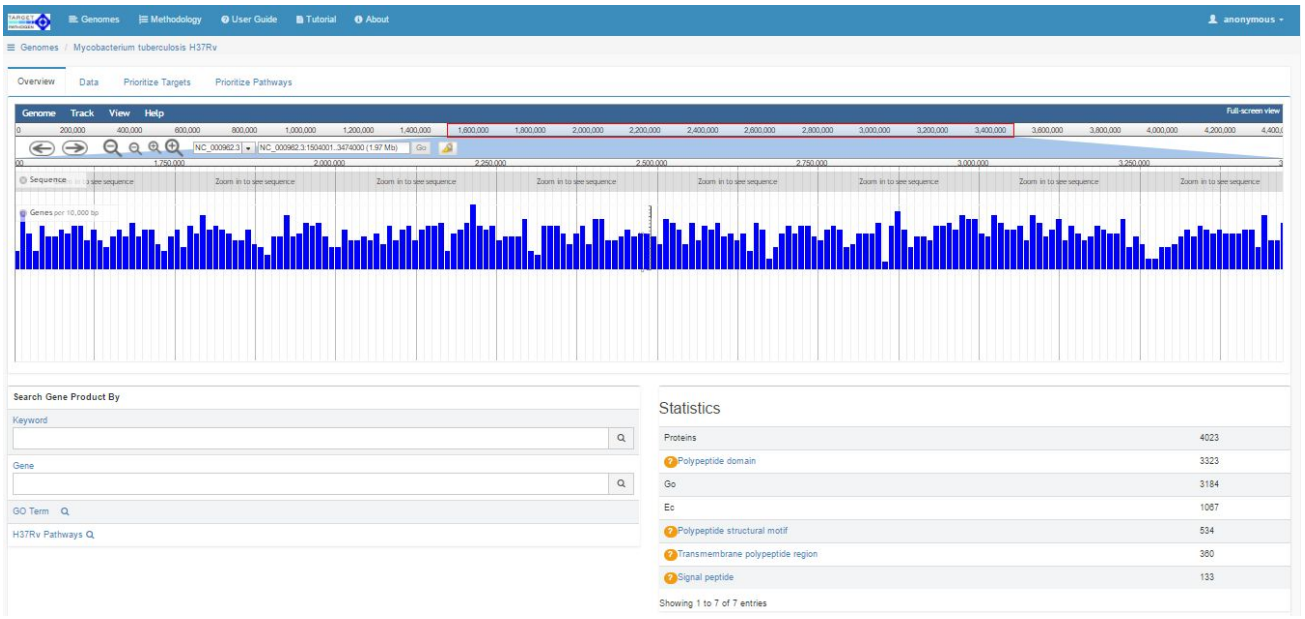
Here you have to choose one of the genomes already upload in Target-Pathogen by clicking Genomes.



Name	Proteins	Pathways	Cristals	Models
Mycobacterium tuberculosis H37Rv - Overview	4023 --> Prioritize Targets	285 --> Prioritize PW	382	1999
Mycobacterium leprae Br4923 - Overview	1956 --> Prioritize Targets	221 --> Prioritize PW	59	756
Klebsiella pneumoniae Kp13 - Overview	5736 --> Prioritize Targets	396 --> Prioritize PW	422	2772
Leishmania major Friedlin - Overview	8400 --> Prioritize Targets	142 --> Prioritize PW	121	6535
Wolbachia endosymbiont TRS of Brugia malayi - Overview	946 --> Prioritize Targets	114 --> Prioritize PW	2	595
Trypanosoma brucei DAL972 - Overview	9895 --> Prioritize Targets	234 --> Prioritize PW	113	3286
Shigella dysenteriae Sd197 - Overview	4294 --> Prioritize Targets	386 --> Prioritize PW	1144	1486
Schistosoma mansoni Puerto Rico - Overview	11802 --> Prioritize Targets	132 --> Prioritize PW	34	7507
Toxoplasma gondii ME49 - Overview	8322 --> Prioritize Targets	241 --> Prioritize PW	56	1380
Plasmodium vivax Salvador I - Overview	5586 --> Prioritize Targets	207 --> Prioritize PW	79	4907

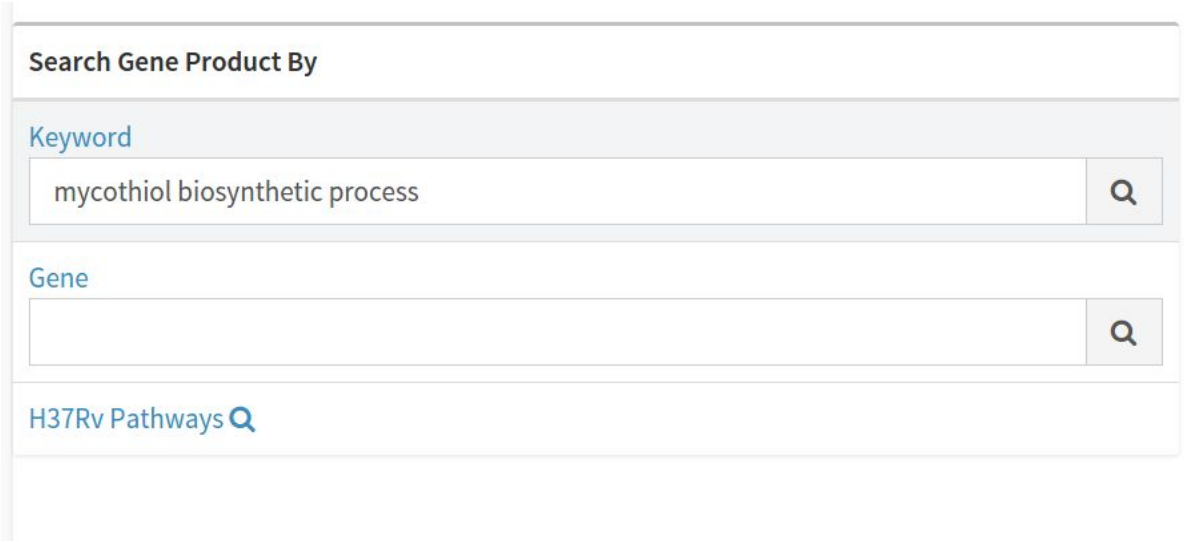
Once you select your genome, you will be redirected to the genome page. Here, you will find three ways to explore the genome. First, the genome can be navigated in a fast,

smooth scrolling and zooming way using JBrowse. Using this browser you can scale from a entire genome to a single base view. Another interface offers a main search menu with several options to retrieve the protein structural druggability records. The options include the use of (i) Keyword (e.g UniProt id, Protein Kinase PknB), (ii) Gene (NCBI locus tag or gene name) or (iii) Pathways. Searches may return a single database entry (e.g. when searching by Gene) or multiple entries (e.g. Keyword and pathways).



Also, genomes can be be easily explored hierarchically by EC number (1) or the different categories of Gene Ontology by using Krona.

Let's assume that you are searching for proteins annotated with the go term mycothiol biosynthetic process, thus simply type it in the keyword field.



The resulting records are listed in the ranking page below the Filter and Score tab. For each record, size, druggability score, gene name and the number of pathways where the proteins are involved is presented. By clicking on the desired row, eight tabs of the corresponding record will be expanded.

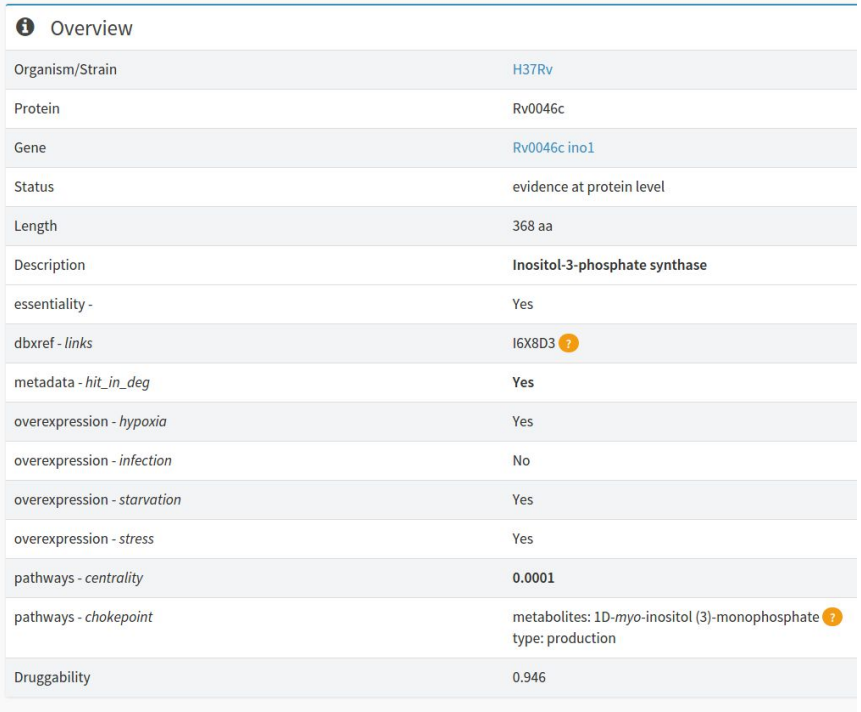
Protein Product	Size	Druggability	Pathways	Gene	Description
Rv0046c	368 aa	0.946	3PW	Rv0046c ino1	Inositol-3-phosphate synthase
Rv0486	481 aa	0.979	1PW	Rv0486 mshA	D-inositol 3-phosphate glycosyltransferase
Rv0819	316 aa	-	1PW	Rv0819 mshD	Mycothiol acetyltransferase
Rv1170	304 aa	0.781	1PW	Rv1170 mshB	1D-myo-inositol 2-acetamido-2-deoxy-alpha-D-glucopyranoside deacetylase
Rv2130c	415 aa	0.494	2PW	Rv2130c mshC	L-cysteine:1D-myo-inositol 2-amino-2-deoxy-alpha-D-glucopyranoside ligase
Rv2855	460 aa	0.974	1PW	Rv2855 mtr	Mycothione reductase
Rv3137	261 aa	0.868	No data	Rv3137 hisN	Histidinol-phosphatase

Showing 1 to 7 of 7 entries (filtered from 4,023 total entries)

100 records per page

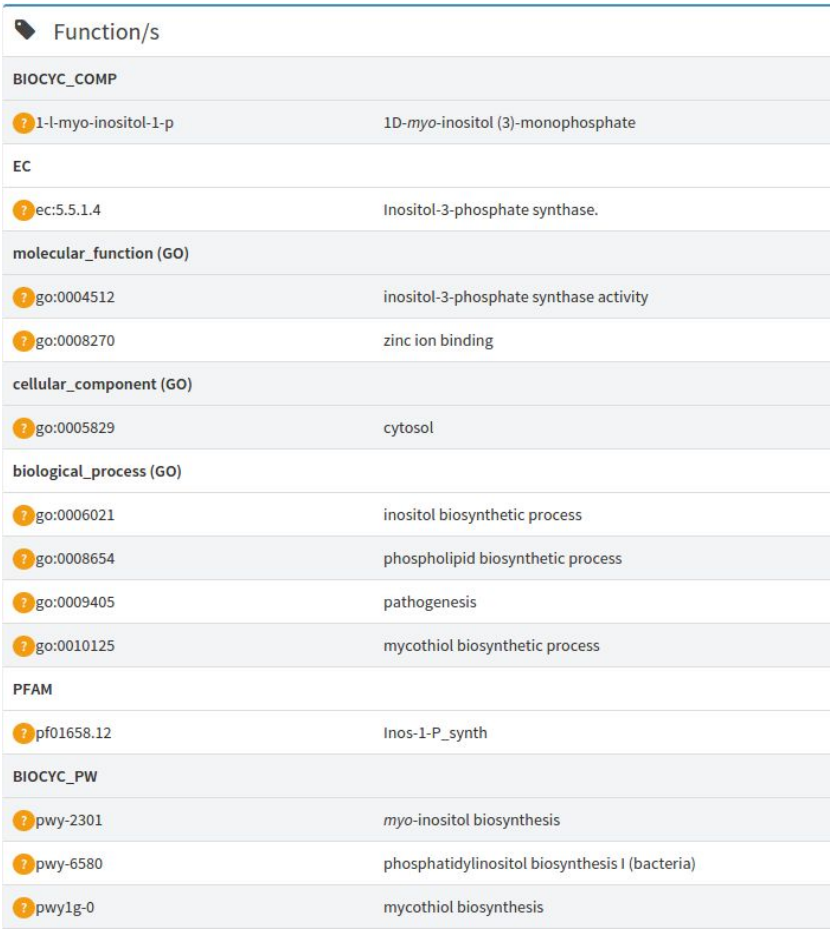
In the **Overview tab**, general information and some metadata (if available), are presented. This tab also shows the higher druggability score detected in a protein pocket.

Other tabs show data about the protein's sequence, pathways, ontology and families (PFAM)(2, 3).



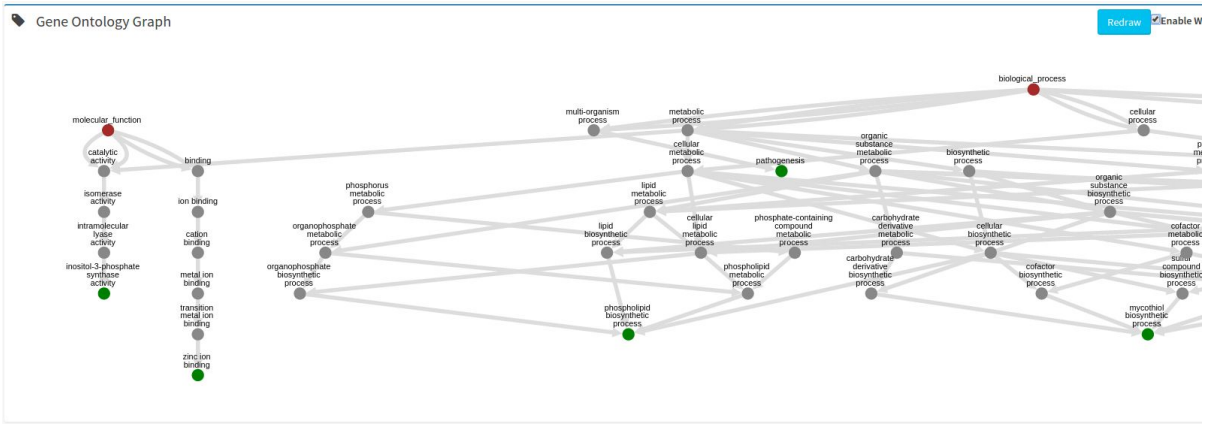
Overview

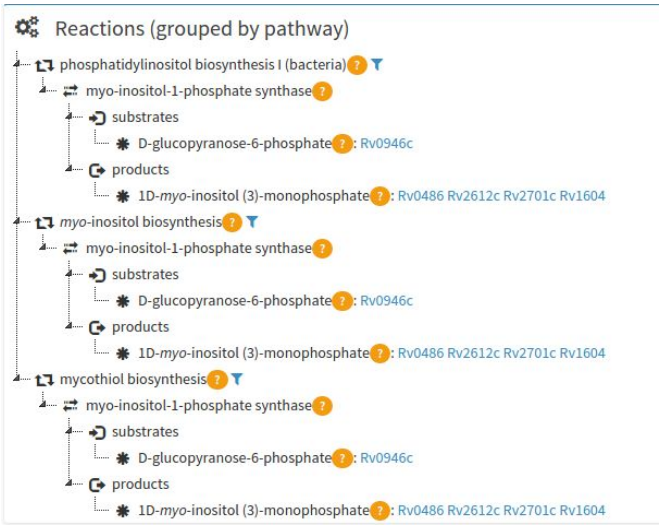
Organism/Strain	H37Rv
Protein	Rv0046c
Gene	Rv0046c ino1
Status	evidence at protein level
Length	368 aa
Description	**Inositol-3-phosphate synthase**
essentiality -	Yes
dbxref - *links*	I6X8D3
metadata - *hit_in_deg*	**Yes**
overexpression - *hypoxia*	Yes
overexpression - *infection*	No
overexpression - *starvation*	Yes
overexpression - *stress*	Yes
pathways - *centrality*	**0.0001**
pathways - *chokepoint*	metabolites: 1D-*myo*-inositol (3)-monophosphate type: production
Druggability	0.946



Function/s

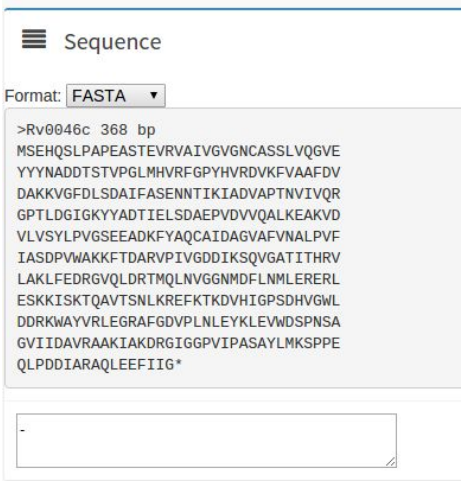
BIOCYC_COMP	
1-l-myo-inositol-1-p	1D-*myo*-inositol (3)-monophosphate
EC	
ec:5.5.1.4	Inositol-3-phosphate synthase.
molecular_function (GO)	
go:0004512	inositol-3-phosphate synthase activity
go:0008270	zinc ion binding
cellular_component (GO)	
go:0005829	cytosol
biological_process (GO)	
go:0006021	inositol biosynthetic process
go:0008654	phospholipid biosynthetic process
go:0009405	pathogenesis
go:0010125	mycothiol biosynthetic process
PFAM	
pf01658.12	Inos-1-P_synth
BIOCYC_PW	
pwy-2301	*myo*-inositol biosynthesis
pwy-6580	phosphatidylinositol biosynthesis I (bacteria)
pwy1g-0	mycothiol biosynthesis

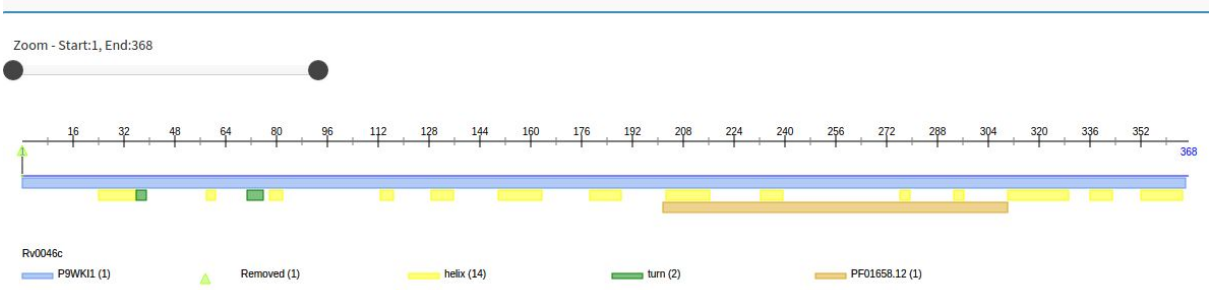




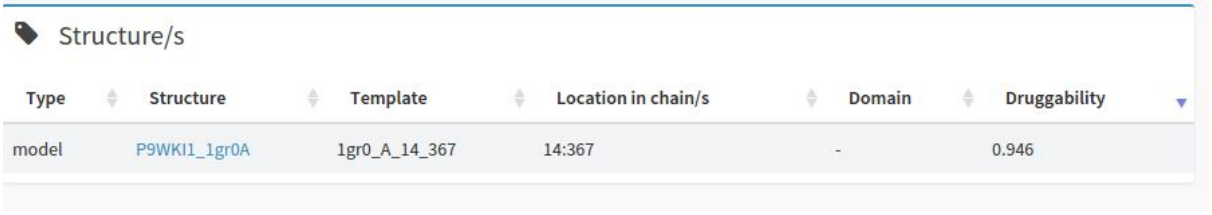
Features

Family/Domains	Segments (start:end)
Myo-inositol-1-phosphate synthase PF01658.12	202:311
Secondary Structure	**Segments** (start:end)
helix	24:36 - 58:61 - 78:8 more...
turn	36:39 - 71:76
Homologous	**Segments** (start:end)
P9WKI1	0:367
Post Traductional Modification	**Segments** (start:end)
Removed	0:1

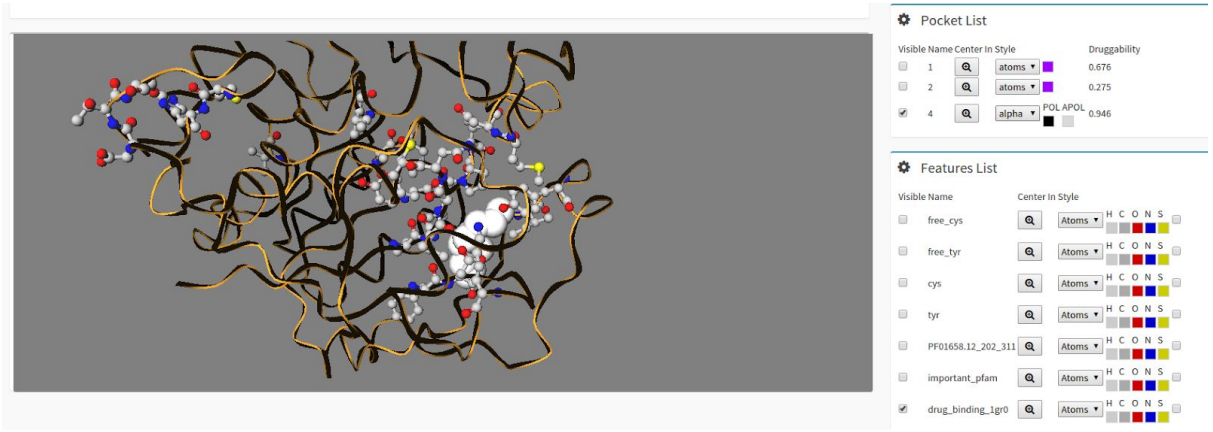




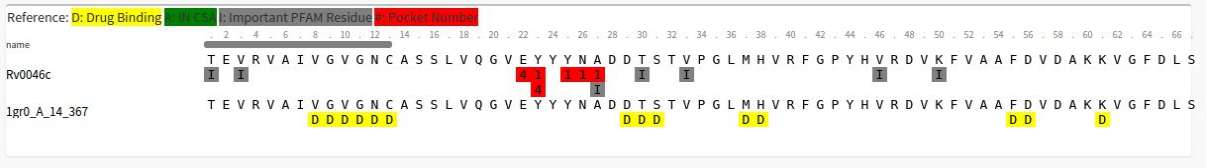
At last, by clicking one model or crystal in the **Structure-Tab**, you will be directed to a page that presents the structure-related data, including the interactive pocket visualization module



The visualization module allows the user (i) to select a pocket for graphical display, (ii) to display present heteroatoms, assigned PFAM or CSA relevant residues, (iii) to display the protein chains in different styles (iv) to display the alpha spheres or the pocket residues defining the pocket.



The displayed protein is available to download as a VMD file(4). Additional information about Model quality assessments and templates are shown at the top of this page. By clicking on the template chain, it's structure will be open.

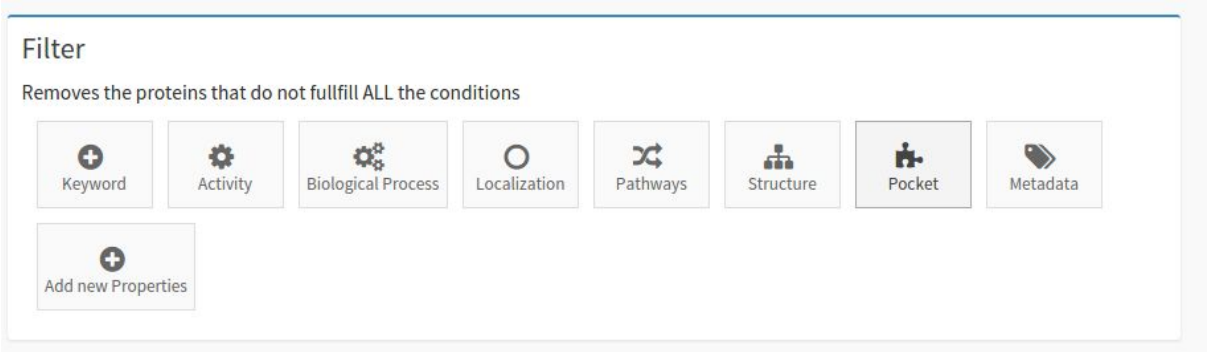


Pockets predicted by fpocket (5) with a druggability score greater than 0.2 are also accessible in the right margin of the page in the ‘Pocket list tab’. This tab shows all the pockets ordered by their druggability score. Finally, there is a Features List tab. By clicking in the different features you can stick out residues of interest as residues that bind crystallized ligands or conserved residues in PFAM.

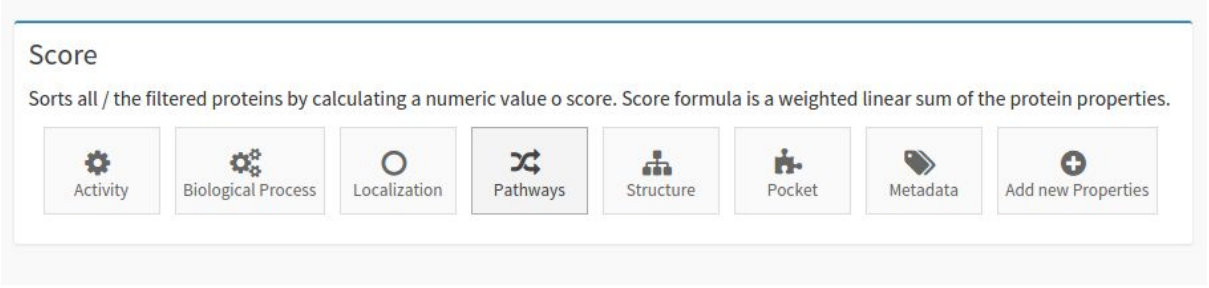
Querying Target-Pathogen to rank and prioritize drug targets.

To obtain a short list of proteins that could be good candidate drug targets you have to choose your genome and click in Prioritize targets.

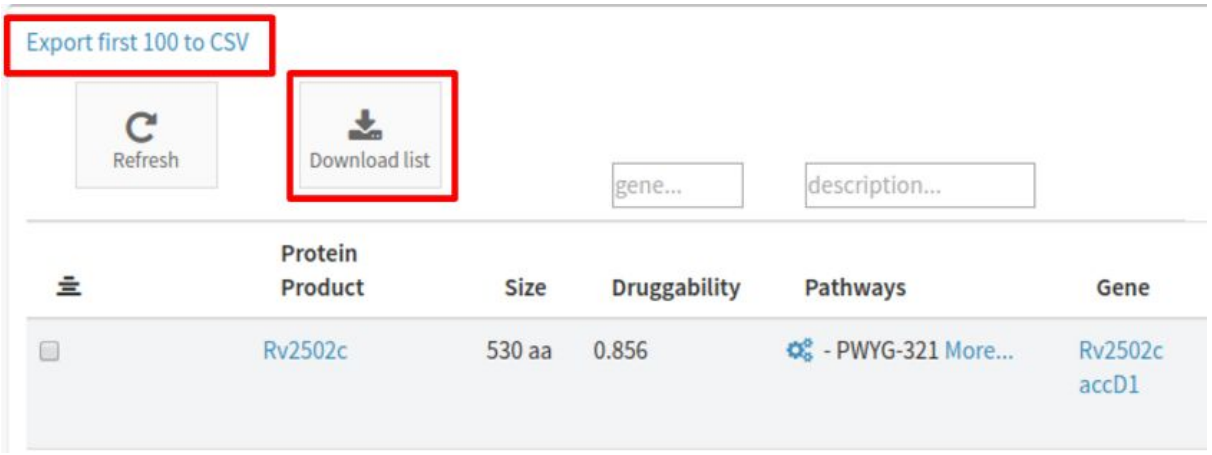
By doing that, you will be directed to a three tabs page. The first one, the “Filter tab”, allow researchers to filter molecular targets based on whether or not they fulfill a set of user defined criteria. For example if a protein is druggable or highly druggable.



The second one is the “Score tab”. This tab allows you to assign different weights to different features of the proteins to set an user defined score function (SF). In the SF each selected parameter must have a coefficient, to weight the importance of such value. It can be fractional or even negative. It is important to always verify the distribution of values of the property, for example, the “chokepoins” prop is the number of chokepoints in a pathway, and “chokepoints_norm” is the number of chokepoints in a pathway divided by the number of reactions.



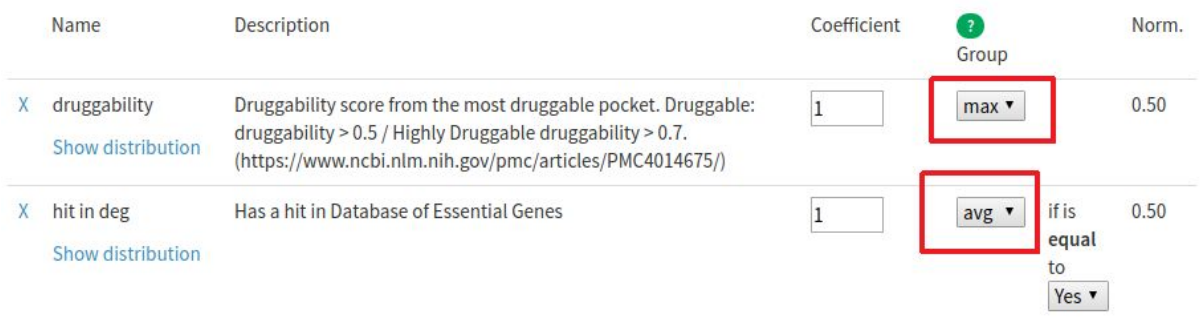
At last, you can combine different filters with the SF and Target-Pathogen will return a particular list of ranked genes according to an user predefined criteria. The third tab, at the bottom, is where it is displayed the proteins ranked by the criteria previously set. You can easily download this table in csv format by clicking on “Export first 100 results” (for a quick view) or the “Download list” button (for a complete filtered list of proteins with the selected scores).



Protein Product	Size	Druggability	Pathways	Gene	Description
Rv3048c	325 aa	0.682	- PWY-7220 More...	Rv3048c nrdF2	Ribonucleoside-diphosphate reductase subunit beta nrdF2
Rv0116c	252 aa	0.727	Reaction RXN0-5402 (no pw assigned) -.-	Rv0116c ldtA	L,D-transpeptidase 1
Rv1169c	101 aa	0.523	No data	Rv1169c lipX	protein with domain: PE family

When scoring pathways, properties can belong to the proteins (for example "pocket with csa") or to the pathway (for example "completeness"). When using the later, there is a direct map between the property and the SF. However when protein properties are used, there are many proteins in one pathway, so there are many ways to combine the values. The group column assigns an operator:

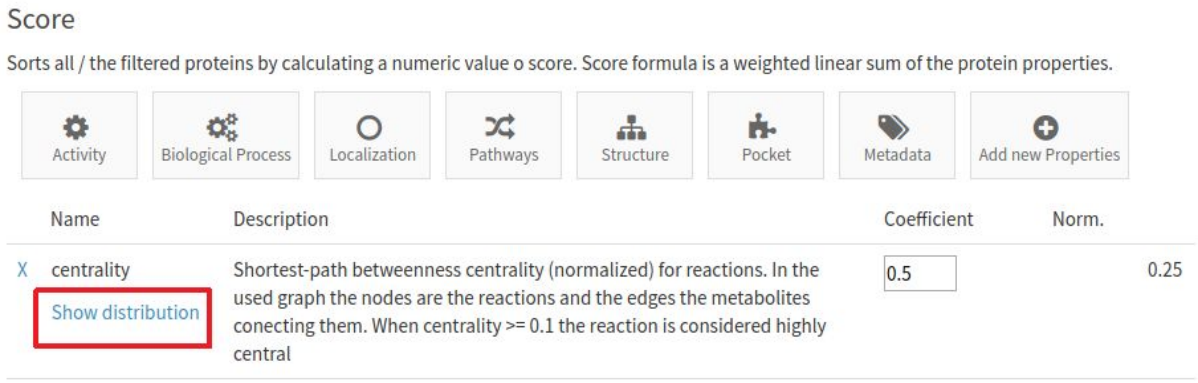
- max (Ex: druggability of the most druggable protein of the pathway)
- sum (Ex: protein count in the pathway with binding activity)
- min (Ex: protein with the lowest centrality)
- avg (Ex: hit in deg proteins / total proteins in the pw)

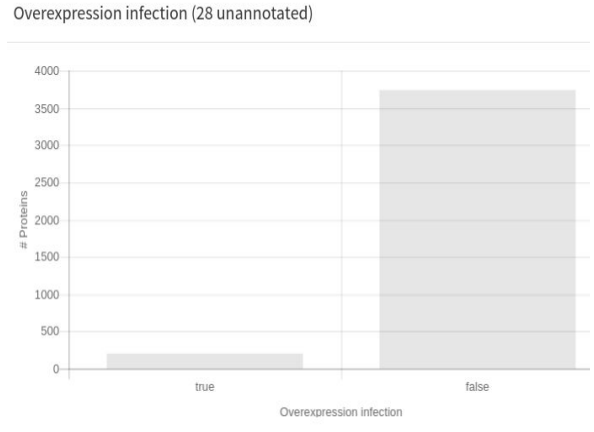


Score = druggability + hit_in_deg

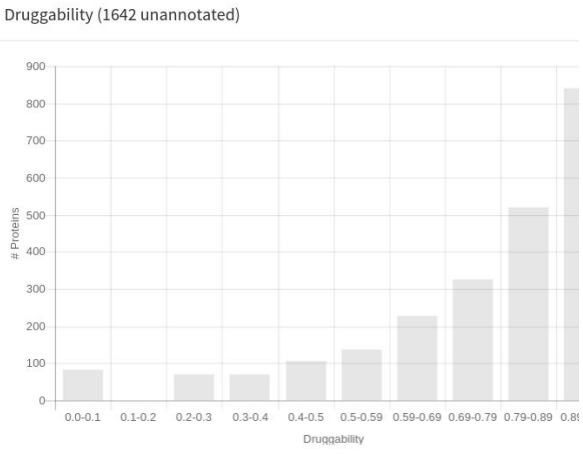
Properties Distribution.

You can access the histogram of Target-Pathogen built in properties(for the selected genome), through the links in the filter o score table of the selected parameters:



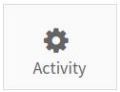


Categorical Property

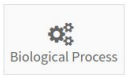


Numeric Property

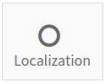
Properties for filtering and scoring proteins



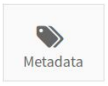
EC (ENZYME number) o GO (Gene ontology) Molecular Function terms. Results are sorted from particular to general



GO Biological Process terms. Results are sorted from particular to general



GO Cellular Component terms. Results are sorted from particular to general



General protein data. This category allows the use of the uploaded properties. A user can only view his/her own properties.

Pocket with free tyr

One of the protein structures has a tyr inside a pocket without atoms surrounding the oxygen of the OH group (more than 3 cubic Å)

Pocket with tyr

One of the protein structures has a tyr inside a pocket

Pocket with free cys

One of the protein structures has a cys inside a pocket without atoms surrounding the S atom of the SH group (more than 3 Å)

Pocket with cys

One of the protein structures has a cys residue inside a pocket

Pocket with csa

One of the protein structures has at least one residue inside a pocket reported in the Catalitic Site Atlas database

Pocket with domain extended

One of the protein crystal has a pfam domain with residues inside a pocket near a drug or cofactor

Pocket with pfam imp residue

One of the protein structures, has at least one residue, inside a pocket and a pfam domain, in contact (less than 3 Å) with a drug or cofactor

Pocket with drug binding

One of the protein structures has at least one pocket residue in contact with a drug (less than 3 Å)

Pocket with lipid binding

One of the protein structures has at least one pocket residue in contact with a lipid (less than 3 Å)

Pocket with metal binding

One of the protein structures has at least one pocket residue in contact with a metal (less than 3 Å)

Pocket with nucleotide binding

One of the protein structures has at least one pocket residue in contact with a nucleotide (less than 3 Å)

Pocket with sugar binding

One of the protein structures has at least one pocket residue in contact with asugar (less than 3 Å)

Human offtarget

This score reflects the results of a blastp search of the pathogen protein in the human proteome database (ncbi accession GCF_000001405.36) with the scale 1 - max(alignment identity), so when a protein has no hit in the human proteome, the value is 1, and if it has 2 hits, one with an identity of 0.4 and other with 0.6, the score is 0.4 (human_offtarget = 1 - 0.6, uses the max identity).

Hit in deg

Has a hit in Database of Essential Genes

(Protein pathway properties)

Centrality

Shortest-path betweenness centrality (normalized) for reactions.

In the used graph the nodes are the reactions and the edges the metabolites conecting them.

When centrality >= 0.1 the reaction is considered highly central

Chokepoint

The protein catalyzes a chokepoint reaction

Chokepoint type

Chokepoint reaction type (consume, production or double)

Has structure

Protein has a 3D structure

Structure type

Experimentally obtained or modelled

Druggability

Druggability score from the most druggable pocket. Druggable: druggability > 0.5 / Highly Druggable druggability > 0.7. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4014675/)

Hydrophobicity

Hydrophobicity of the most druggable pocket

Volume

Volume in cubic Å of the most druggable pocket

Free tyr

One of the protein structures has a tyr without atoms surrounding the oxygen of the OH group (more than 3 cubic Å)

Tyr

One of the protein structures has a tyr residue

Free cys

One of the protein structures has a cys without atoms surrounding the S atom of the SH group (more than 3 Å)

Cys

One of the protein structures has a cys residue

Csa

One of the protein structures has at least one residue reported in the Catalitic Site Atlas database

Domain extended

One of the protein cristals has a pfam domain with residues near a drug or cofactor

Drug binding

One of the protein structures has at least one residue in contact with a drug (less than 3 Å)

Lipid binding

One of the protein structures has at least one residue in contact with a lipid (less than 3 Å)

Metal binding

One of the protein structures has at least one residue in contact with a metal (less than 3 Å)

Nucleotide binding

One of the protein structures has at least one residue in contact with a nucleotide (less than 3 Å)

Sugar binding

One of the protein structures has at least one residue in contact with a sugar (less than 3 Å)

(Pathway properties)

Reactions

Number of reactions in the pathway

Norm reactions

Number of reactions normalized by the highest pathway (with more reactions)

Reactions with gene

Number of reactions in the pathway with at least one known protein that catalize them

Completeness

Proportion of reactions in the pathway with at least one known protein that catalize them

Chokepoints

Number of chokepoints reactions in the pathway

Norm chokepoint

Chokepoint reactions/reactions in pathway ratio

Druggable

The pathway has at least one druggable protein

Max centrality

Maximum betweenness centrality of all the reactions in the pathway, normalized by the reaction with max betweenes centrality in the whole graph

Only H37Rv

Essentiality

Critical for the organism survival (https://www.ncbi.nlm.nih.gov/pubmed/26791267)

Overexpression stress

Overexpressed in model of stress (https://www.ncbi.nlm.nih.gov/pubmed/26791267)

Overexpression starvation

Overexpressed in model of starvation (https://www.ncbi.nlm.nih.gov/pubmed/26791267)

Overexpression infection

Overexpressed in model of infection (https://www.ncbi.nlm.nih.gov/pubmed/26791267)

Overexpression hypoxia

Overexpressed in model of hypoxia (https://www.ncbi.nlm.nih.gov/pubmed/26791267)

Only Kp13

Essential in mgh78578

Hits with an essential gene of Klebsiella pneumoniae MGH78578

Conserved pathogen

Hit count in different pathogen Klebsiella subspecies

Conserved no pathogen

Hit count in different non pathogen Klebsiella subspecies

Overexpressed in polymyxin

Overexpressed in polymyxin B resistance induction (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5088521/)

Overexpressed in polymixin all conditions

Overexpressed in presence of Polymyxin in different conditions (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5088521/)

Conserved pathogen norm

Hit count in different pathogenic Kp strains divided by the total number of compared bacteria (39 pathogenic Kp).

Gut microbiome

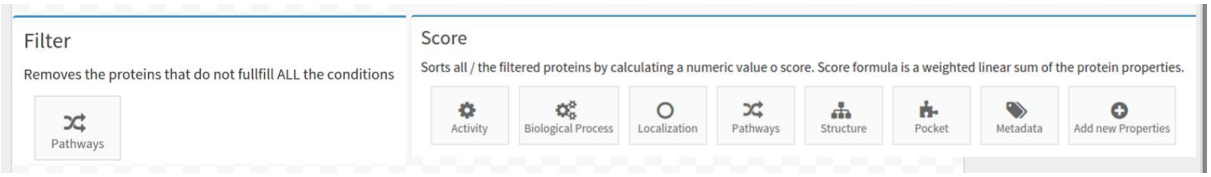
Number of gut microbiome organisms that have at least one hit with the Kp13 protein (blast: identity > 40% evalue 1e-5)

Gut microbiome norm

gut_microbiome normalized by the total number of compared bacteria (226)

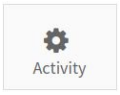
Choosing promising pathways as putative targets of new drugs.

Target Pathogen allow user to rank pathways with a user-defined criteria in order to prioritize entire pathways as good candidates for novel therapies. One fundamental advantage of studying the metabolic context of putative targets is that results are expected to allow the design of possible combined therapies (targeting more than one target from the same metabolic pathway). To Prioritize pathways you can combine different filters with the SF similar to the prioritization of proteins.



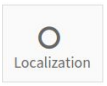
A high value would mean that most genes in the pathway and the pathway itself is attractive from the drug discovery point of view.

Properties for filtering and scoring pathways

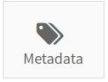


EC (ENZYME number) o GO (Gene ontology) Molecular Function terms. Results are sorted from particular to general

GO Biological Process terms. Results are sorted from particular to general



GO Cellular Component terms. Results are sorted from particular to general



General protein data. This category allows the use of the uploaded properties. A user can only view his/her own properties.

Pocket with free tyr

One of the protein structures has a tyr inside a pocket without atoms surrounding the oxygen of the OH group (more than 3 cubic Å)

Pocket with tyr

One of the protein structures has a tyr inside a pocket

Pocket with free cys

One of the protein structures has a cys inside a pocket without atoms surrounding the S atom of the SH group (more than 3 Å)

Pocket with cys

One of the protein structures has a cys residue inside a pocket

Pocket with csa

One of the protein structures has at least one residue inside a pocket reported in the Catalitic Site Atlas database

Pocket with domain extended

One of the protein crystal has a pfam domain with residues inside a pocket near a drug or cofactor

Pocket with pfam imp residue

One of the protein structures, has at least one residue, inside a pocket and a pfam domain, in contact (less than 3 Å) with a drug or cofactor

Pocket with drug binding

One of the protein structures has at least one pocket residue in contact with a drug (less than 3 Å)

Pocket with lipid binding

One of the protein structures has at least one pocket residue in contact with a lipid (less than 3 Å)

Pocket with metal binding

One of the protein structures has at least one pocket residue in contact with a metal (less than 3 Å)

Pocket with nucleotide binding

One of the protein structures has at least one pocket residue in contact with a nucleotide (less than 3 Å)

Pocket with sugar binding

One of the protein structures has at least one pocket residue in contact with asugar (less than 3 Å)

Human offtarget

This score reflects the results of a blastp search of the pathogen protein in the human proteome database (ncbi accession GCF_000001405.36) with the scale 1 - max(alignment identity), so when a protein has no hit in the human proteome, the value is 1, and if it has 2 hits, one with an identity of 0.4 and other with 0.6, the score is 0.4 (human_offtarget = 1 - 0.6, uses the max identity).

Hit in deg

Has a hit in Database of Essential Genes

(Protein pathway properties)

Centrality

Shortest-path betweenness centrality (normalized) for reactions.

In the used graph the nodes are the reactions and the edges the metabolites conecting them.

When centrality >= 0.1 the reaction is considered highly central

Chokepoint

The protein catalyzes a chokepoint reaction

Chokepoint type

Chokepoint reaction type (consume, production or double)

Has structure

Protein has a 3D structure

Structure type

Experimentally obtained or modelled

Druggability

Druggability score from the most druggable pocket. Druggable: druggability > 0.5 / Highly Druggable druggability > 0.7. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4014675/)

Hydrophobicity

Hydrophobicity of the most druggable pocket

Volume

Volume in cubic Å of the most druggable pocket

Free tyr

One of the protein structures has a tyr without atoms surrounding the oxygen of the OH group (more than 3 cubic Å)

Tyr

One of the protein structures has a tyr residue

Free cys

One of the protein structures has a cys without atoms surrounding the S atom of the SH group (more than 3 Å)

Cys

One of the protein structures has a cys residue

Csa

One of the protein structures has at least one residue reported in the Catalitic Site Atlas database

Domain extended

One of the protein cristals has a pfam domain with residues near a drug or cofactor

Drug binding

One of the protein structures has at least one residue in contact with a drug (less than 3 Å)

Lipid binding

One of the protein structures has at least one residue in contact with a lipid (less than 3 Å)

Metal binding

One of the protein structures has at least one residue in contact with a metal (less than 3 Å)

Nucleotide binding

One of the protein structures has at least one residue in contact with a nucleotide (less than 3 Å)

Sugar binding

One of the protein structures has at least one residue in contact with a sugar (less than 3 Å)

(Pathway properties)

Reactions

Number of reactions in the pathway

Norm reactions

Number of reactions normalized by the highest pathway (with more reactions)

Reactions with gene

Number of reactions in the pathway with at least one known protein that catalize them

Completeness

Proportion of reactions in the pathway with at least one known protein that catalize them

Chokepoints

Number of chokepoints reactions in the pathway

Norm chokepoint

Chokepoint reactions/reactions in pathway ratio

Druggable

The pathway has at least one druggable protein

Max centrality

Maximum betweenness centrality of all the reactions in the pathway, normalized by the reaction with max betweenes centrality in the whole graph

Only H37Rv

Essentiality

Critical for the organism survival (https://www.ncbi.nlm.nih.gov/pubmed/26791267)

Overexpression stress

Overexpressed in model of stress (https://www.ncbi.nlm.nih.gov/pubmed/26791267)

Overexpression starvation

Overexpressed in model of starvation (https://www.ncbi.nlm.nih.gov/pubmed/26791267)

Overexpression infection

Overexpressed in model of infection (https://www.ncbi.nlm.nih.gov/pubmed/26791267)

Overexpression hypoxia

Overexpressed in model of hypoxia (https://www.ncbi.nlm.nih.gov/pubmed/26791267)

Only Kp13

Essential in mgh78578

Hits with an essential gene of Klebsiella pneumoniae MGH78578

Conserved pathogen

Hit count in different pathogen Klebsiella subspecies

Conserved no pathogen

Hit count in different non pathogen Klebsiella subspecies

Overexpressed in polymyxin

Overexpressed in polymyxin B resistance induction (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5088521/)

Overexpressed in polymixin all conditions

Overexpressed in presence of Polymyxin in different conditions (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5088521/)

Conserved pathogen norm

Hit count in different pathogenic Kp strains divided by the total number of compared bacteria (39 pathogenic Kp).

Gut microbiome

Number of gut microbiome organisms that have at least one hit with the Kp13 protein (blast: identity > 40% evalue 1e-5)

Gut microbiome norm

gut_microbiome normalized by the total number of compared bacteria (226)

1. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.

2. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.

3. Finn,R.D. (2005) Pfam: the protein families database. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*.

4. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.

5. Schmidtke,P., Le Guilloux,V., Maupetit,J. and Tufféry,P. (2010) fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res.*, **38**, W582–9.

6. Defelipe,L.A., Do Porto,D.F., Pereira Ramos,P.I., Nicolás,M.F., Sosa,E., Radusky,L., Lanzarotti,E., Turjanski,A.G. and Marti,M.A. (2016) A whole genome bioinformatic approach to determine potential latent phase specific targets in Mycobacterium tuberculosis. *Tuberculosis* , **97**, 181–192.

7. Radusky,L., Defelipe,L.A., Lanzarotti,E., Luque,J., Barril,X., Marti,M.A. and Turjanski,A.G. (2014) TuberQ: a Mycobacterium tuberculosis protein druggability database. *Database* , **2014**, bau035.

8. Hassan,S.S., Tiwari,S., Guimarães,L.C., Jamal,S.B., Folador,E., Sharma,N.B., de Castro Soares,S., Almeida,S., Ali,A., Islam,A., *et al.* (2014) Proteome scale comparative modeling for conserved drug and vaccine targets identification in Corynebacterium pseudotuberculosis. *BMC Genomics*, **15 Suppl 7**, S3.