

Target-Pathogen: A Structural bioinformatic approach to prioritize drug targets in pathogens.

Ezequiel J. Sosa, Germán Burguener, Esteban Lanzarotti, Lucas Defelipe, Leandro Radusky, Agustín M. Pardo, Marcelo Marti, Adrián G. Turjanski* & Darío Fernández Do Porto*

* To whom correspondence should be addressed. Email: dariofd@gmail.com

* Correspondence may also be addressed to adrian@qb.fcen.uba.ar

SUPPLEMENTARY MATERIAL

PROPERTIES FOR FILTERING AND SCORING PATHWAYS	2
TARGET-PATHOGEN METHODOLOGY	8
TARGET-PATHOGEN TUTORIAL	14

Score
Sorts all / the filtered proteins by calculating a numeric value o score. Score formula is a weighted linear sum of the protein properties.

Activity Biological Process Localization Pathways Structure Pocket Metadata Add new Properties

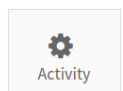
Name	Description	Coefficient	Norm.
X centrality	Shortest-path betweenness centrality (normalized) for a reaction graph.	2	0.25

Score = 2 * centrality + 2 * chokepoint + overexpression_stress + overexpression_starvation + overexpression_infection + overexpression_hypoxia

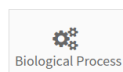
Protein Product	Size	Druggability	Pathways	Gene	Description	Properties
Rv3205c	393 aa	0.985	OC - PWY-7289	Rv3205c moeB1	Probable adenylyltransferase/sulfurtransferase MoeZ	centrality: 0.02, chokepoint: true, overexpression_stress: true, overexpression_hypoxia: true
Rv2245	417 aa	0.777	OC - PWY-321	Rv2245 kasA	3-oxoacyl-[acyl-carrier-protein] synthase 1	centrality: 0.02, chokepoint: true, overexpression_stress: True, overexpression_starvation: True, overexpression_infection: false, overexpression_hypoxia: True
Rv1285	333 aa	0.652	2PW	Rv1285 cysD	Sulfate adenylyltransferase subunit 2	centrality: 0.02, chokepoint: true, overexpression_stress: True, overexpression_starvation: true, overexpression_infection: false, overexpression_hypoxia: true
Rv3535c	262 aa	0.865				centrality: 0.02, chokepoint: true, overexpression_stress: True, overexpression_starvation: true, overexpression_infection: false, overexpression_hypoxia: true
Rv3130c	464 aa	0.706	Reaction DIACYLGLYCEROL-O-ACYLTRANSFERASE-RXN (no pw assigned) :-	Rv3130c tgs1	Probable diacylglycerol O-acyltransferase tgs1	centrality: 0.00, chokepoint: true, overexpression_stress: True, overexpression_starvation: false, overexpression_infection: True, overexpression_hypoxia: True

Figure S4. Scoring procedure scheme. After filtering out all non-druggable, non-essential and close homolog to human proteins, a Scoring Function ($SF = \frac{H+S+R+I}{4} + \frac{Ch+Cy}{2}$) was set in Target-Pathogen to obtain a particular list of ranked genes to combat latent tuberculosis. The first term of the equation integrates available expression data under different conditions mimicking infection. The selected conditions, which group different reports, comprise Hypoxia (%), Starvation (S), RNOS stress (&) and infection in mice (I). The second term focuses in metabolic context of the proteins. In this way *h* and *c* determines if the reaction associated to the protein is a chokepoint or central in the bacteria metabolism, respectively. Each variable takes the value of 1 if the protein comply the criteria and 0 if not. Different coefficients were applied to expression and metabolic parameters terms in order to assign the same weight to both components: Centrality and chokepoint =2, overexpression in hypoxia, starvation, infection and stress =4. Centrality parameter is highlight above. Top five ranked proteins are shown at the bottom of the page. Finally the structure of Rv1285 is shown with the druggable pocket overlapping with the crystallized drug binding site.

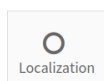
PROPERTIES FOR FILTERING AND SCORING PATHWAYS



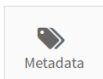
Activity
EC (ENZYME number) or GO (Gene ontology) Molecular Function terms. Results are sorted from particular to general



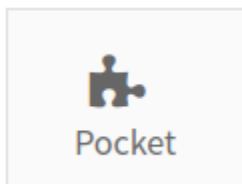
Biological Process
GO Biological Process terms. Results are sorted from particular to general



Localization
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

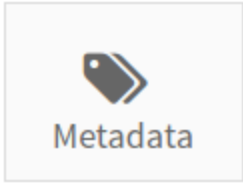
One of the protein structures has at least one pocket residue in contact with a metal (less than 3 Å)

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)

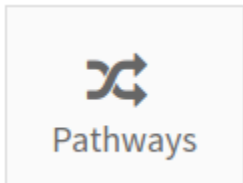


Human offtarget

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 $1 - \max(\text{alignment identity})$, 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 ($\text{human_offtarget} = 1 - 0.6$, 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 connecting them.

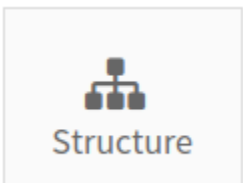
When centrality ≥ 0.1 the reaction is considered highly central

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

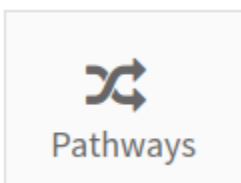
One of the protein structures has at least one residue in contact with a metal (less than 3 Å)

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

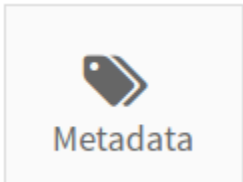
Chokepoint reactions/reactions in pathway ratio

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

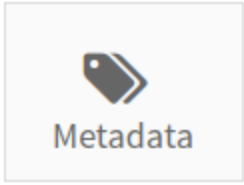
Overexpressed in model of starvation (<https://www.ncbi.nlm.nih.gov/pubmed/26791267>)

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% evaluate $1e-5$)

Gut microbiome norm

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

TARGET-PATHOGEN METHODOLOGY

All data present in Target Pathogen database is based either on the *in-silico* calculation of selected properties for each protein or, on the integration and meta-analysis of publicly available data. The pipeline-engine which we call Target Pathogen is schematically shown below.

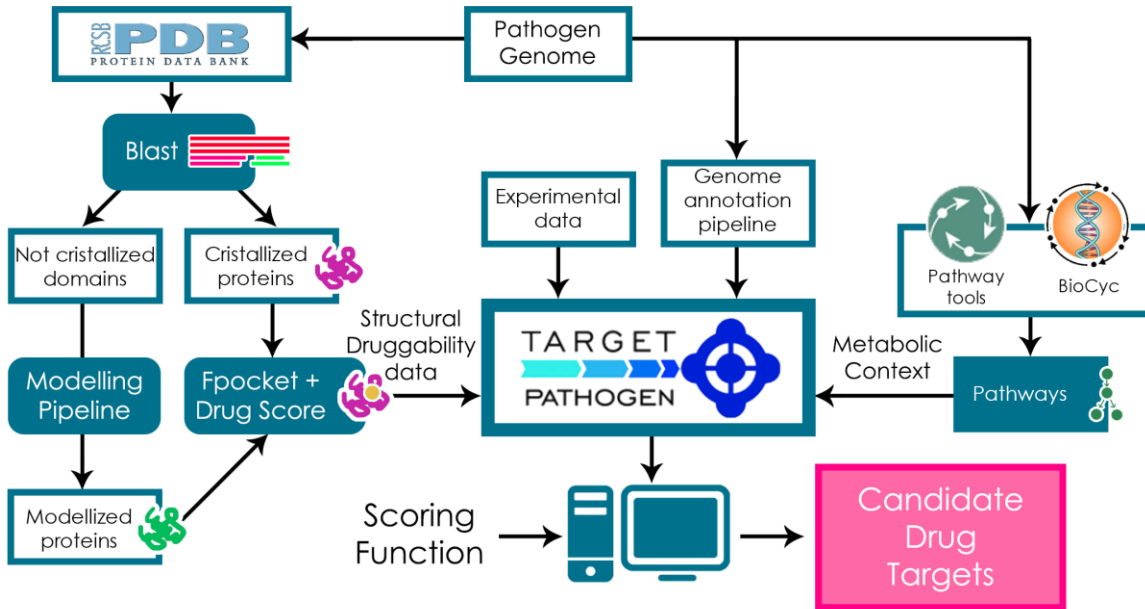


Figure 1. General sketch target prioritization pipeline. Structural druggability and metabolic analyses are integrated with available experimental and in silico data. After all data is integrated in Target- Pathogen, an user designed scoring function is used to weight different features in order to obtain a ranked list of candidate drug targets

Generation of Structural homology based models

For all ORFs in *Mycobacterium tuberculosis*, *Klebsiella pneumoniae* and *Shigella dysenteriae* genomes we attempted to build homology-based models using the following structural genomic pipeline. The first step consists in performing a psi-blast search against a template library, which includes all sequences from every individual protein chain in the PDB, grouped at 95% sequence identity threshold using CD-hit (Li and Godzik 2006). Then, every target structure was built with the MODELLER software (Eswar et al. 2008), using local alignment derived from the above-described psi-blast search (Altschul et al. 1997). For each target sequence, 5 different models were built, and their quality measures were assigned using the GA341 (Melo and Sali 2007) and QMEAN (Melo and Sali 2007; Benkert, Tosatto, and Schomburg 2008)

methods. The best model (max QMean Score) for each protein was kept. Previously, only the models with GA341 score above 0.7 and over 60% coverage were retained. Other models were obtained from Modbase database (Pieper et al. 2014).

For or all the structures (crystals and models) we then compute several structural properties like: i) the DS for each pocket using fpocket, ii) the similarity with human protein (to evaluate potential off-target effects), iii) The active site residues (if available) by using data from Catalytic Site Atlas (CSA) (Furnham et al. 2013) and iv) The PFAM conserved or family relevant residues.

Structural Assessment of Druggability

Structural druggability of each potential target was assessed by determining (and characterizing) the ability of putative pockets to bind a drug-like molecule by using the fpocket program (Schmidtke et al. 2010) and DrugScore (DS) index (Schmidtke et al. 2010; Schmidtke and Barril 2010). Briefly, the method is based on Voronoi tessellation algorithm to identify pockets and computes suitable physicochemical descriptors (hydrophobic density, polar and apolar surface area, hydrophobic and polarity score) that are combined to yield the DS, which ranges between 0 to 1. Figure 2 shows a histogram for the druggability score computed for those pockets present in all unique protein in the Protein Data Bank, which were crystallized in complex with a drug like compound that correspond effectively to the binding pocket is shown:

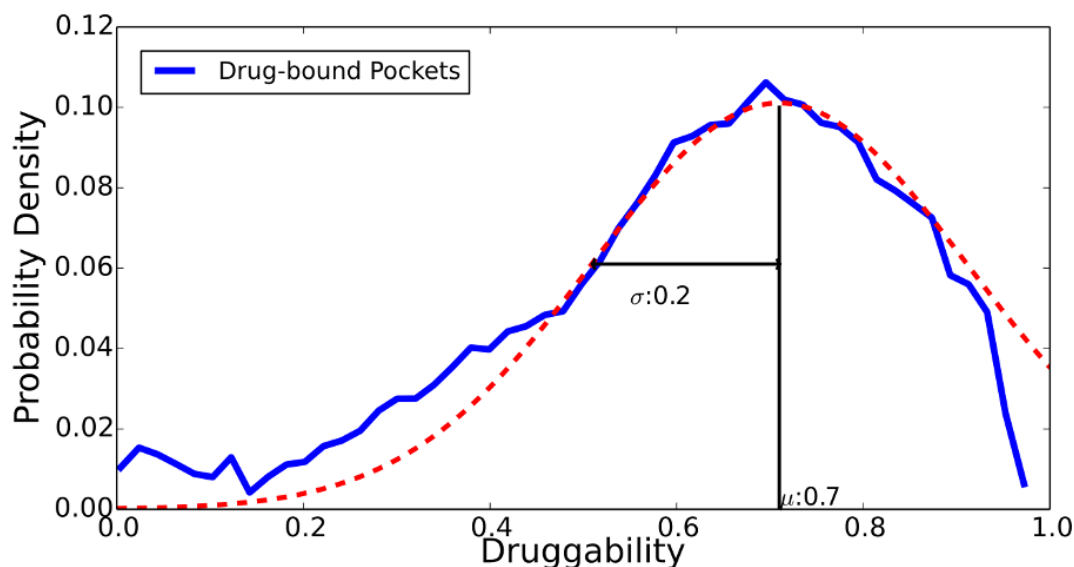


Figure 2: In blue, histogram for the druggability score computed for all those pockets present in all unique protein in the Protein Data Bank, which were crystallized in complex with a drug like compound inside the corresponding pocket. In red, the gaussian fitting of the pocket classification sets.

Fitting the resulting histogram to a gaussian distribution results in a mean of 0.7 with a standard deviation of 0.2. As expected the DS computed for all pockets in the PDB, except those having inside a drug like molecule, show a distribution that peaks at DS of zero, and falls rapidly (See Green plot in Figure 3)

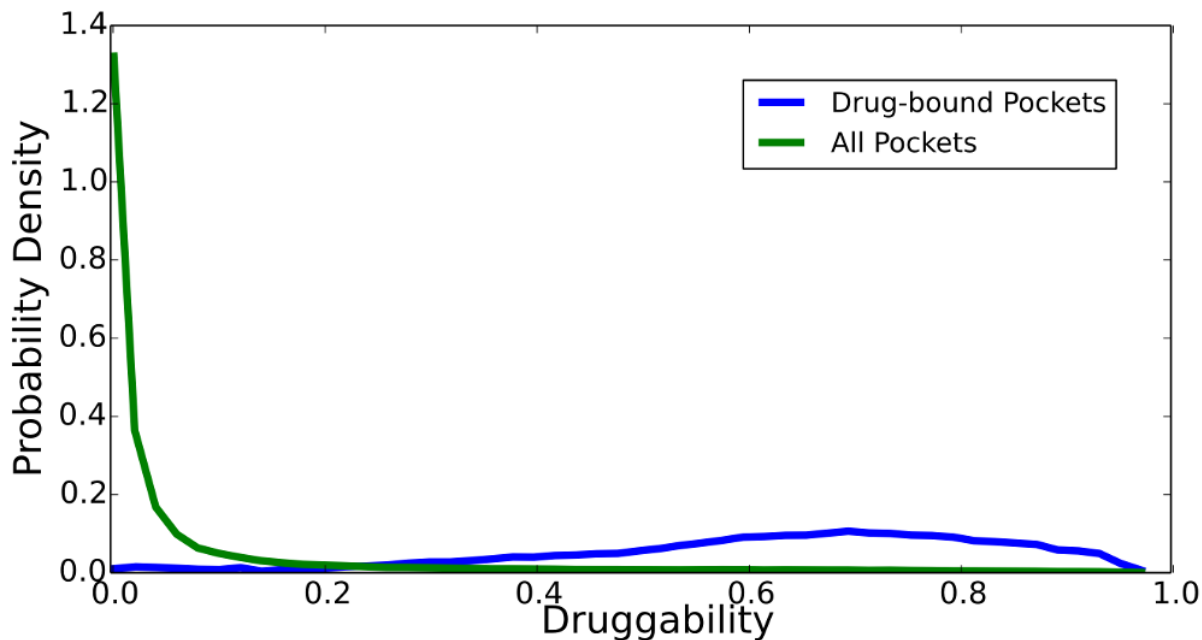


Figure 3: In green, the druggability score computed by fpocket software for all the pockets in all the structures of the PDB database except those having inside a drug like molecule which are shown in blue.

Based on this analysis we classified each pocket according to the following four categories, non druggable (ND), with DS between 0 and 0.2, poorly druggable (PD), with DS between 0.2 and 0.5, Druggable (D) with DS between 0.5 and 0.7, and highly druggable (HD) with DS between 0.7 and 1. In the web application, visualization is available for the first three categories but statistics are available for all the detected pockets regardless of its druggability score.

Active site pocket identification

To identify the active site pocket and/or determine the relevance of a given pocket to protein function, Target-Pathogen uses TuberQ `s methodology (Radusky et al. 2014). It consist in two different analyses, that rely on, (i) the information from the CSA (Catalytic Site Atlas) and (ii) a PFAM position site importance criteria

The data from CSA (downloaded from <http://www.ebi.ac.uk/thornton-srv/databases/CSA/>) consists of a list of PDB linked to a number of residues, which comprise the corresponding protein active site. To map the active sites to as many protein and/or domains as possible, each PDB in CSA was assigned to a protein with that PDB hit.

As an alternative approach to determine the relevance of a given pocket (or residue), we looked for residues of a given PFAM family/domain that are located in an important position and are well conserved. Important positions were defined as those positions in the corresponding HMMer model whose information content was larger than a defined importance cutoff value (*icov*). The nature of the conserved amino acids in the corresponding position was determined by comparing each residue type emission probability (*ep*) with *icov*. If the ratio between *ep* and *icov* was larger than a conserved type cutoff value (*ctcov*), the corresponding residue type was assumed to be conserved. Optimal values of *icov* and *ctcov* were 0.27 and 0.24, respectively.

Off-target and essentiality criteria

All proteins in the database were subjected to NCBI-BLASTp (*e-value* smaller than $1e-07$) against human proteome to identify non-host homologs targets. The criteria for regarding a protein as a human homologue were a sequence similarity of greater than 30% using a BLOSUM62 matrix, for a length of more than 30% of the bacterial query protein with an E-value less than 10^{-4} .

Furthermore, all proteomes were submitted to the Database of Essential Genes (Zhang 2004a; Luo et al. 2014)(DEG, which contains experimentally validated genes under different conditions in three domains of life) for homology analyses (Barh et al. 2013; Zhang 2004b). The BLASTp cut-off values used were: *e-value* = $1e-05$, *bit score* ≥ 100 , *identity* $\geq 35\%$ (Barh et al. 2011). Only Mycobacterium tuberculosis essential genes were defined as in previous works (Defelipe et al. 2016; Radusky et al. 2014)

Metabolic network construction

Metabolic networks (MN) were built by using the PathoLogic algorithm within Pathway Tools v. 19.0 (Karp et al. 2016) (or a previously existing PGDB was used when it was available). PathoLogic creates a Pathway/Genome Database (PGDB) containing the predicted metabolic pathways of a given organism using as input a Genbank file with the corresponding product annotations and gene coordinates along the genome. The Genbanks entries were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/>) and were used as initial input for MN reconstructions. The steps involved in the reconstruction include determining gene-protein-reaction associations, which are based in either the availability of the corresponding enzyme commission (EC) number or alternatively in the gene product annotation, using a custom dictionary within Pathway Tools which links products to reactions. The reconstructed metabolic network was exported in systems biology markup language (SBML) and format for downstream analyses. Reactions involving macromolecules (such as DNA, RNA and proteins, as per the BioCyc ontology) were filtered, and only the small-molecule complement of the MNs was considered. After MN reconstruction, we generated a list of all compounds present in the network, and we collected their frequency as reaction participants using a Python script. Those who most frequently appeared as

reaction participants are considered currency compounds (such as ATP, cofactors, water) and were disregarded from the network since they may create artificial links on the graph-based representation of the network as they are involved in many reactions which are not necessarily related.

Metabolic network analysis.

After MN reconstruction, we generated a reaction graph, where nodes represent reactions (i.e usually enzymes) and there is an edge between two nodes if the product of one reaction is used as substrate on the reaction that follows. Cytoscape v. 2.8.3 was used for data visualization and further MN analyses (Karp et al. 2016; Russell and Cohn 2012). Choke-point analysis was conducted in order to identify potential drug targets from the metabolic perspective. We also calculated the betweenness centrality of every node in MN, using the `betweenness_centrality` function in the NetworkX python package. The betweenness centrality of a given node v , $C_B(v)$, in the graph $G = (V, E)$, where V is a set of vertices or nodes and E a set of edges is given by: $C_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$ (Brandes 2001) where σ_{st} is the number of shortest paths from $s \in V$ to $t \in V$ and $\sigma_{st}(v)$ represents the number of shortest paths from s to t that some node $v \in V$ lies in.

Links to 3rd Party Databases

Target Pathogen provides links to other websites to clarify vocabulary terms, ontologies or database entries. To prevent broken links caused by changes in an external website domain or in url path for entry accession, we will query every month, random links of every site, and query for a correct response. The checked sites are: Uniprot, PDB, Biocyc, Sequence Ontology, Gene ontology, xFam/Pfam and ExPasy. In case of error, we will correct the links to the failing site.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs."

Nucleic Acids Research 25 (17): 3389–3402.

Barh, Debmalya, Krishnakant Gupta, Neha Jain, Gourav Khatri, Nidia León-Sicairos, Adrian Canizalez-Roman, Sandeep Tiwari, et al. 2013. "Conserved Host–pathogen PPIs : Globally Conserved Inter-Species Bacterial PPIs Based Conserved Host-Pathogen Interactome Derived Novel Target in C. Pseudotuberculosis, C. Diphtheriae, M. Tuberculosis, C. Ulcerans, Y. Pestis, and E. Coli Targeted by Piper Betel Compounds." *Integrative Biology* 5 (3): 495.

Barh, Debmalya, Neha Jain, Sandeep Tiwari, Bibhu Prasad Parida, Vivian D'Afonseca, Liwei Li, Amjad Ali, et al. 2011. "A Novel Comparative Genomics Analysis for Common Drug and Vaccine Targets in Corynebacterium Pseudotuberculosis and Other CMN Group of Human Pathogens." *Chemical Biology & Drug Design* 78 (1): 73–84.

Benkert, Pascal, Silvio C. E. Tosatto, and Dietmar Schomburg. 2008. "QMEAN: A Comprehensive Scoring Function for Model Quality Assessment." *Proteins* 71 (1): 261–77.

Brandes, Ulrik. 2001. "A Faster Algorithm for Betweenness Centrality*." *The Journal of Mathematical Sociology* 25 (2): 163–77.

Defelipe, Lucas A., Dario Fernández Do Porto, Pablo Ivan Pereira Ramos, Marisa Fabiana Nicolás, Ezequiel Sosa, Leandro Radusky, Esteban Lanzarotti, Adrián G. Turjanski, and Marcelo A. Marti. 2016. "A Whole Genome Bioinformatic Approach to Determine Potential Latent Phase Specific Targets in Mycobacterium Tuberculosis." *Tuberculosis* 97 (March): 181–92.

Eswar, Narayanan, David Eramian, Ben Webb, Min-Yi Shen, and Andrej Sali. 2008. "Protein Structure Modeling with MODELLER." In *Methods in Molecular Biology*, 145–59.

Furnham, Nicholas, Gemma L. Holliday, Tjaart A. P. de Beer, Julius O. B. Jacobsen, William R. Pearson, and Janet M. Thornton. 2013. "The Catalytic Site Atlas 2.0: Cataloging Catalytic Sites and Residues Identified in Enzymes." *Nucleic Acids Research* 42 (D1): D485–89.

Karp, Peter D., Mario Latendresse, Suzanne M. Paley, Markus Krummenacker, Quang D. Ong, Richard Billington, Anamika Kothari, et al. 2016. "Pathway Tools Version 19.0 Update: Software for Pathway/genome Informatics and Systems Biology." *Briefings in Bioinformatics* 17 (5): 877–90.

Li, W., and A. Godzik. 2006. "Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences." *Bioinformatics* 22 (13): 1658–59.

Luo, Hao, Yan Lin, Feng Gao, Chun-Ting Zhang, and Ren Zhang. 2014. "DEG 10, an Update of the Database of Essential Genes That Includes Both Protein-Coding Genes and Noncoding Genomic Elements." *Nucleic Acids Research* 42 (Database issue): D574–80.

Melo, Francisco, and Andrej Sali. 2007. "Fold Assessment for Comparative Protein Structure Modeling." *Protein Science: A Publication of the Protein Society* 16 (11): 2412–26.

Pieper, Ursula, Benjamin M. Webb, Guang Qiang Dong, Dina Schneidman-Duhovny, Hao Fan, Seung Joong Kim, Natalia Khuri, et al. 2014. "ModBase, a Database of Annotated Comparative Protein Structure Models and Associated Resources." *Nucleic Acids Research* 42 (Database issue): D336–46.

Radusky, Leandro, Lucas A. Defelipe, Esteban Lanzarotti, Javier Luque, Xavier Barril, Marcelo A. Marti, and Adrián G. Turjanski. 2014. "TuberQ: A Mycobacterium Tuberculosis Protein Druggability Database." *Database: The Journal of Biological Databases and Curation* 2014 (0): bau035.

Russell, Jesse, and Ronald Cohn. 2012. *Cytoscape*. Book on Demand Limited.

Schmidtke, Peter, and Xavier Barril. 2010. "Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites." *Journal of Medicinal Chemistry* 53 (15): 5858–67.

Schmidtke, Peter, Vincent Le Guilloux, Julien Maupetit, and Pierre Tufféry. 2010. "Fpocket: Online Tools for Protein Ensemble Pocket Detection and Tracking." *Nucleic Acids Research* 38 (Web Server issue): W582–89.

Zhang, R. 2004a. "DEG: A Database of Essential Genes." *Nucleic Acids Research* 32 (90001): 271D – 272.

———. 2004b. "DEG: A Database of Essential Genes." *Nucleic Acids Research* 32 (90001): 271D – 272.

TARGET-PATHOGEN TUTORIAL

The goal of Target-Pathogen is to become a useful resource for researchers working in the field of drug discovery to translate biological questions in a computational tractable way by exploring, filtering and weighting the vast quantity of genomic-scale data sets that are now available in order to produce a shortlist of suitable targets for further investigation. The main feature of Target-Pathogen is to integrate data from different sources with structural druggability analysis and metabolic network reconstruction in a consistent and effective manner, contributing to a better selection of potential drug targets for screening campaigns and the analysis of targets for structure-based drug design projects.

The general purpose of this tutorial is to show users how to explore data present in Target-Pathogen and how to weight this information in order to identify and prioritize drug targets for pathogens.

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.

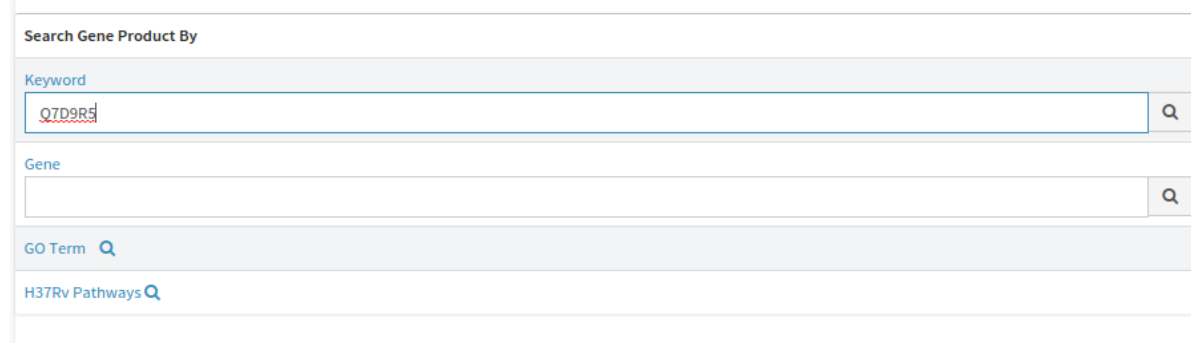
Target-Pathogen function as a database that allow users to rank and prioritize targets for drug development. A large amount of information of each gen and protein within genomes is actually present in the database. The genome browser can be accessed and queried using the web interface at <http://target.sbg.qb.fcen.uba.ar/patho/>. Here you have to choose one of the genomes already uploaded in Target-Pathogen by clicking Genomes. Suppose you are searching for a particular *Mycobacterium tuberculosis* protein, so you must select H37Rv genome.

The following example will take you through a trip around Target-Pathogen, showing its salient characteristics to search for a protein, that allows accession of the desired record in a fast and intuitive manner.

All searches start in the *main search page*, where you can use a keyword including Go terms, Uniprot ID (1), PFam ID (1, 2) or structure PDB ID (3) or you can specifically search a gene or pathway. Target-Pathogen also allows users to navigate the genome using JBrowse. Searches may return a single database entry (e.g. when searching by Gene) or multiple entries (e.g. Keyword and pathways). Finally,

genomes can be also easily explored hierarchically by EC number (4) or the different categories of Gene Ontology (5) by using Krona.

Let's assume, that in the present example, we already know our target protein ID, thus we simply type "Q7D9R5" in "Keyword", to retrieve all associated records.



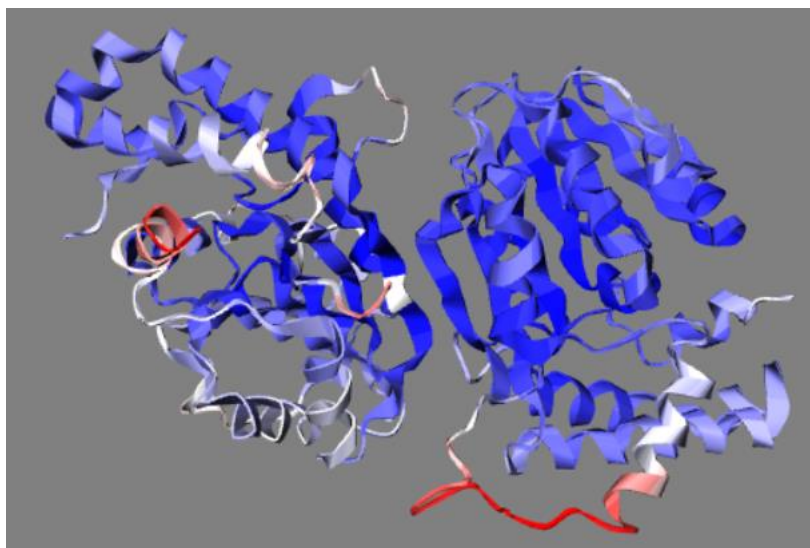
The resulting records are listed in the ranking page. For each record, size, druggability score, gene name and the number of pathways where the proteins are involved is presented.

Protein Product	Size	Druggability	Pathways	Gene	Description
Rv0470c	288 aa	0.551	 1PW	Rv0470c <i>pcaA</i>	Cyclopropane mycolic acid synthase 3

By clicking on the desired row, eight tabs of the corresponding record will be expanded. This tabs, will contain general information, metadata, protein's sequence, ontology and PFAM domains (2, 6), joined with structural and metabolic information.

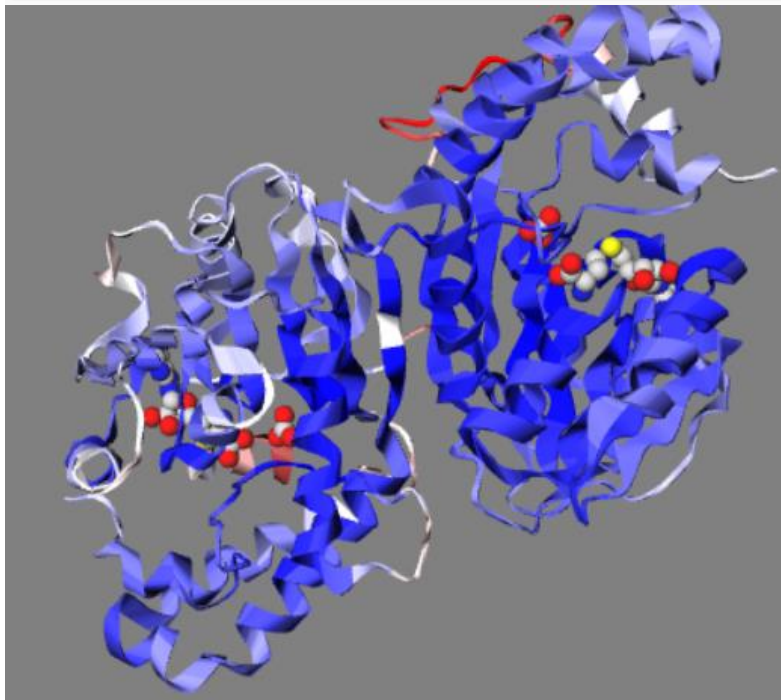
In the present example, our protein of interest has been crystallized (PDB ID=111e). To access the record, click in the in 111e structure. Six main tabs will be display.

The top tab is where the visualization of the protein can be downloaded for VMD. Clicking in the download button a compressed file is download. This is the visualization for the protein in our example in the GLMol web software (<http://webglmol.osdn.jp/index-en.html>) used in the server:



Other tabs presents the structure related data, including the interactive pocket visualization module. The visualization module allows i) to select which pocket to show (ticking the corresponding pocket Select field), ii) display present HETATMS (7), assigned CSA (8) or PFAM relevant residues, iii) Display the protein chains in different styles, iv) Display the pocket residues or the alpha spheres (With polar and apolar spheres). In the druggable pocket of the example shown below, we depict polar alpha spheres of pocket "1" in black while its apolar alpha spheres in white. The HETATMS found in the crystal structure are shown as balls and sticks in different colours.

Reference:	D: Drug Binding	I: Important PFAM Residue	P: Pocket Number
name			
Rv0470c_15:287			
	50	52	54
	56	58	60
	62	64	66
	68	70	72
	74	76	78
	80	82	84
111e_A_16:287			



Chain/s List

Select intersection Clear intersection

Visible Name	Center In Style	
<input checked="" type="checkbox"/> A	<input type="text" value="Q"/> Beta Factor	<input type="text" value="Beta Factor"/>
<input checked="" type="checkbox"/> B	<input type="text" value="Q"/> Beta Factor	<input type="text" value="Beta Factor"/>
<input checked="" type="checkbox"/> heatoms	<input type="text" value="Q"/> H C O N S P F CLBI	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

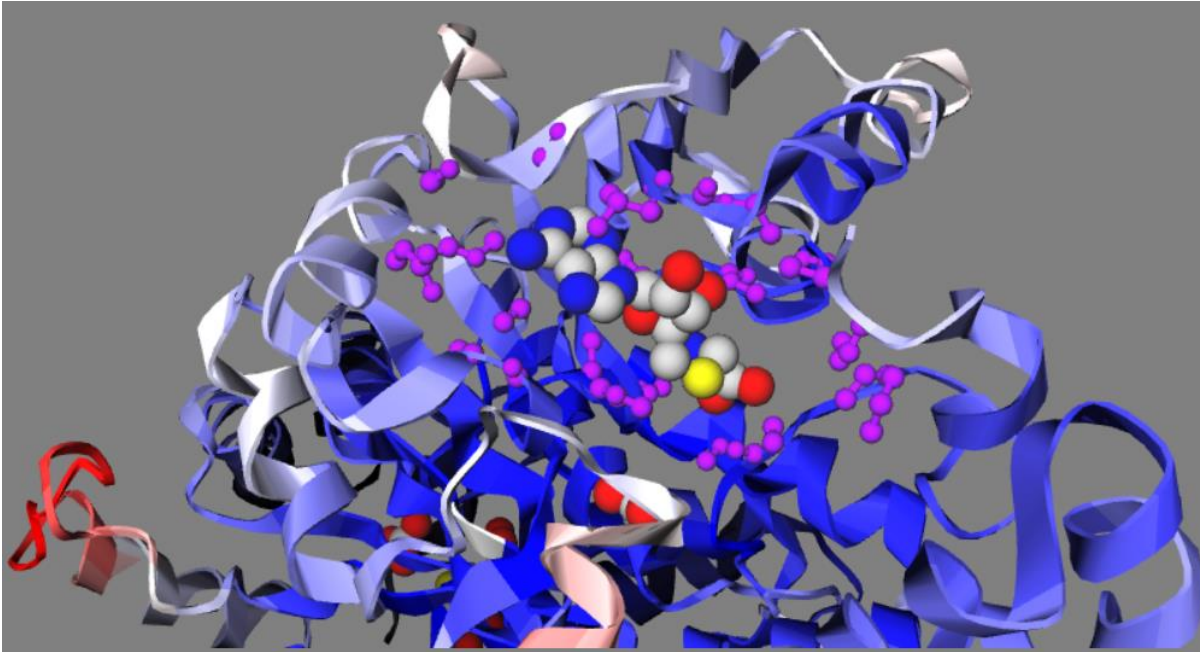
Pocket List

Visible Name	Center In Style	Druggability
<input type="checkbox"/> 1	<input type="text" value="Q"/> atoms	0.551
<input type="checkbox"/> 2	<input type="text" value="Q"/> atoms	0.434
<input type="checkbox"/> 3	<input type="text" value="Q"/> atoms	0.239
<input type="checkbox"/> 7	<input type="text" value="Q"/> atoms	0.423
<input type="checkbox"/> 8	<input type="text" value="Q"/> atoms	0.403

Features List

Visible Name	Center In Style	
<input type="checkbox"/> PF02353.15_0_268	<input type="text" value="Q"/> Atoms	H C O N S <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/> PF13489.1_39_208	<input type="text" value="Q"/> Atoms	H C O N S <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/> PF08241.7_53_150	<input type="text" value="Q"/> Atoms	H C O N S <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/> important_pfam	<input type="text" value="Q"/> Atoms	H C O N S <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/> PF02353.15_0_157	<input type="text" value="Q"/> Atoms	H C O N S <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/> PF13489.1_36_156	<input type="text" value="Q"/> Atoms	H C O N S <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/> PF02353.15_5_96	<input type="text" value="Q"/> Atoms	H C O N S <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/> free_cys	<input type="text" value="Q"/> Atoms	H C O N S <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/> free_tyr	<input type="text" value="Q"/> Atoms	H C O N S <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/> drug_binding	<input type="text" value="Q"/> Atoms	H C O N S <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

Another possible visualization of the same pocket, could be done by selecting to show the residues lining the pocket (not the alpha spheres) and the residues reported to be part of a domain in PFAM or a drug binding site. In the figure below superposition between the druggable pocket and one of the drug binding site of the crystallized protein is shown.



Obtaining a list of drug targets candidates for *Mycobacterium tuberculosis*.

We have previously defined two important features to select a gene product as a potential target for new drugs development to combat *Mycobacterium tuberculosis* (9). First, the role of the protein within the metabolism and second, it's ability to bind a drug-like molecule, which in turn inhibits its function. Target-Pathogen allows users to interactively visualize genomic data in order to explore these criteria in genes and proteins of ten genomes now available in our web server.

In this example we will guide you to obtain a ranked list of targets for drug development against latent *Mycobacterium tuberculosis*.

To obtain a short list of proteins that could be adequate candidates for drug targets you have to choose H37Rv genome and click "Prioritize Targets" in the protein column of the "Genomes" page. By doing that, you will be directed to a three tabs page, where you can filter and weight the data present in the database to display a set of proteins that fulfill the criteria defined by the user.

Once there, you can filter out all proteins without a druggable pocket and also present possible side effects with human host.

Filter
Removes the proteins that do not fulfill ALL the conditions

Keyword
 Activity
 Biological Process
 Localization
 Pathways
 Structure
 Pocket
 Metadata
 Add new Properties

Just by a click in Structure, a window containing all parameters related with the protein structure will be open.

Structure parameters ×

10 records per page Search:

check	Name	Description	Type
<input type="checkbox"/>	has_structure	Protein gas a 3d structure	value
<input type="checkbox"/>	structure_type	experimental or model	value
<input checked="" type="checkbox"/>	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/)	number
<input type="checkbox"/>	hydrophobicity	Hydrophobicity of the most druggable pocket	number
<input type="checkbox"/>	volume	Volume in cubic Å of the most druggable pocket	number
<input type="checkbox"/>	free_tyr	If any of the proteins structures has a tyr with his OH oxigen atom with no surronding atoms (more than cubic Å)	value
<input type="checkbox"/>	tyr	If any of the proteins structures has a tyr	value
<input type="checkbox"/>	free_cys	If any of the proteins structures has a cys with his SH sulfur atom with no surronding atoms (more than 3 Å)	value
<input type="checkbox"/>	cys	If any of the proteins structures has a cys with his SH sulfur atom with no surronding atoms (more than 3 Å)	value
<input type="checkbox"/>	csa	If any of the proteins cristals or model templates, has at least one residue reported in the Catalitic Site Atlas database	value

Showing 1 to 10 of 16 entries (filtered from 45 total entries)

If you check druggability you will filter proteins by the druggability score (10). If you want druggable and highly druggable proteins you must keep all proteins with DS>0.5.

Name	Description	Operation	Value
X druggability Show distribution	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/)	>	0.5

In a similar way, you can select “human_offtarget” in “Metadata” to filter out those protein with an human offtarget score > 0.6. By doing this you will retain 2047 records from a total of 4,023 proteins in *M. tuberculosis* genome.

The lack or inhibition of an essential protein will conduce to inhibit growth or to death of the pathogens. So, a key criteria to select a good group of targets in tuberculosis is the essentiality of the proteins. To use these criteria, we can click in “Metadata” in the Filter-Tab and select “essentiality”. In this case gene essentiality was defined as in previous works (11) (10). Another criteria to select a good group of targets is their lack of a close homolog in humans to prevent side effects (human offtarget property in “Metadata”) . By doing this you will keep the 762 druggable, essential and without close human homologous that are actually annotated for *M. tuberculosis* genome.

Filter

Removes the proteins that do not fulfill ALL the conditions

Filter options:

- Keyword
- Activity
- Biological Process
- Localization
- Pathways
- Structure
- Pocket
- Metadata
- Add new Properties


	Name	Description	Operation	Value	Duplicate in Score
X	druggability Show distribution	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/)	> ▼	0.5	Duplicate To Score
X	essentiality Show distribution	Critical for the organism survival (https://www.ncbi.nlm.nih.gov/pubmed/26791267)	equal ▼	true ▼	Duplicate To Score
X	human offtarget Show distribution	This score reflects the results of a blastp search of the pathogen protein in the human proteome database (ncbi	> ▼	0.6	Duplicate To Score


To further rank the putative targets to specifically combat latent tuberculosis, we use the right panel (Score) to define a scoring function as:


$$SF = \frac{H + S + R + I}{4} + \frac{Ch + Cy}{2}$$


Score


Sorts all / the filtered proteins by calculating a numeric value o score. Score formula is a weighted linear sum of the protein properties.



Activity



Biological Process



Localization


Pathways


Structure


Pocket


Metadata


Add new Properties

	Name	Description	Coefficient	Norm.
X	overexpression stress	Overexpressed in model of stress (https://www.ncbi.nlm.nih.gov/pubmed/26791267)	<input type="text" value="0.25"/>	if is equal to true ▾ 0.13
	Show distribution			
X	overexpression starvation	Overexpressed in model of starvation (https://www.ncbi.nlm.nih.gov/pubmed/26791267)	<input type="text" value="0.25"/>	if is equal to true ▾ 0.13
	Show distribution			
X	overexpression infection	Overexpressed in model of infection (https://www.ncbi.nlm.nih.gov/pubmed/26791267)	<input type="text" value="0.25"/>	if is equal to true ▾ 0.13
	Show distribution			
X	overexpression hypoxia	Overexpressed in model of hypoxia (https://www.ncbi.nlm.nih.gov/pubmed/26791267)	<input type="text" value="0.25"/>	if is equal to true ▾ 0.13
	Show distribution			
X	centrality	Shortest-path betweenness centrality (normalized) for reactions. In the used graph the nodes are the reactions and the edges the metabolites connecting them. When centrality ≥ 0.1 the reaction is considered highly central	<input type="text" value="0.5"/>	0.25
	Show distribution			
X	chokepoint	The protein catalyzes a chokepoint reaction	<input type="text" value="0.5"/>	if is equal to true ▾ 0.25
	Show distribution			

The first term of the equation integrates available expression data under different conditions mimicking infection. H, S, R, I are variables that defines overexpression in different experimental models: hypoxia, starvation, RNOS stress and mice infection models respectively (9). The second term focus in metabolic context of the proteins. In this way C_h and C_y determines if the reaction associated to the protein is a chokepoint or central in the bacteria metabolism. Note that expression and metabolic terms are divided by two and four respectively assigning the same weight to both components.

Each variable takes the value of 1 if the protein comply the criteria and 0 if not. A high value means that the protein fulfill most of the criteria that defines a promising drug target. The third tab (show below), 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 just by clicking on "Download List".

Showing 1 to 100 of 762 entries (filtered from 4,023 total entries)
[Export first 100 to CSV](#)

Protein Product	Size	Druggability	Pathways	Gene	Description	Properties	Score
Rv3206c	393 aa	0.985	1PW	Rv3206c moeB1	Probable adenylyltransferase/sulfurtransferase MoeZ	overexpression_stress: true, overexpression_starvation: true, overexpression_infection: false, overexpression_hypoxia: true, centrality: 0.02, chokepoint: true	1.26
Rv2245	417 aa	0.777	7PW	Rv2245 kasA	3-oxoacyl-[acyl-carrier-protein] synthase 1	overexpression_stress: True, overexpression_starvation: True, overexpression_infection: false, overexpression_hypoxia: True, centrality: 0.02, chokepoint: true	1.26
Rv1285	333 aa	0.652	2PW	Rv1285 cysD	Sulfate adenylyltransferase subunit 2	overexpression_stress: True, overexpression_starvation: True, overexpression_infection: false, overexpression_hypoxia: True, centrality: 0.01, chokepoint: true	1.25
Rv1286	615 aa	0.542	2PW	Rv1286 cysNC	Bifunctional enzyme CysN/CysC	overexpression_stress: True, overexpression_starvation: True, overexpression_infection: false, overexpression_hypoxia: True, centrality: 0.01, chokepoint: true	1.25
Rv2225	282 aa	0.768	1PW	Rv2225 panB	3-methyl-2-oxobutanoate hydroxymethyltransferase	overexpression_stress: True, overexpression_starvation: false, overexpression_infection: True, overexpression_hypoxia: True, centrality: 0.00, chokepoint: true	1.25

Uploading users' data

A key feature that distinguishes Target-Pathogen from other target prioritization software is that users can upload their own data in an easy way, just by simply uploading a tsv format archive (tab separated values). The file must be plain text, with columns separated by tabs and UTF-8 encoding. There are 2 types of columns, "tags" or "numeric". Tag columns must have no more than 20 different tags. Numeric columns must have "," as decimal separator and no other symbol. To represent the absence of value in a numeric column, the string "NaN" must be used. As an example we show a tsv archive with some antibiotics resistance related genes.

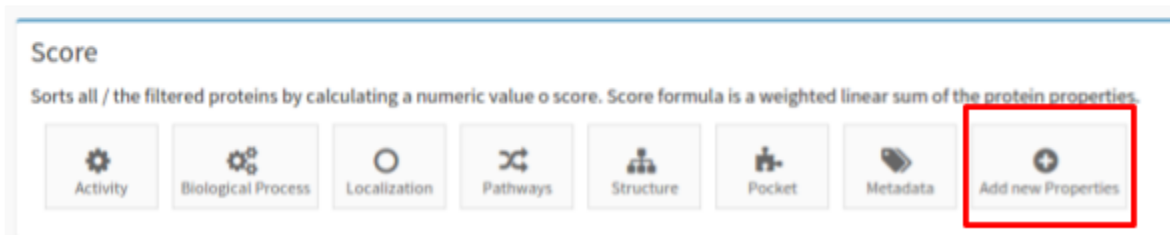
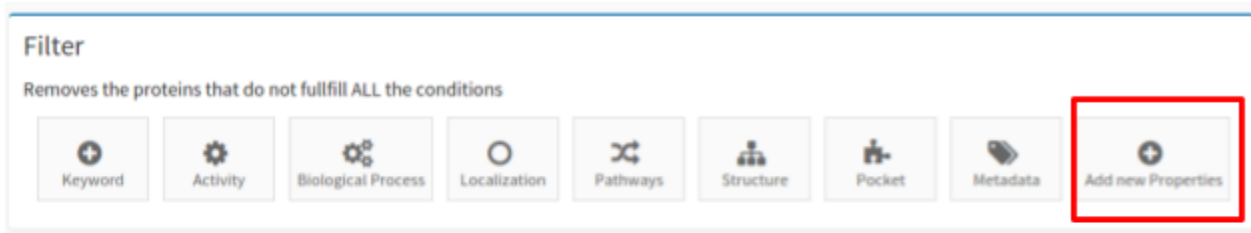
id	resistance	AMI	PAS	EMB	FLQ	INH	SM	RIF	ETH	PZA
Rv0846c	no	no	no	no	no	no	no	no	no	no
Rv0203	no	no	no	no	no	no	no	no	no	no
Rv0343	yes	no	no	yes	no	yes	no	no	no	no
Rv0341	yes	no	no	yes	no	yes	no	no	no	no
Rv0342	yes	no	no	yes	no	yes	no	no	no	no
Rv0483	no	no	no	no	no	no	no	no	no	no
Rv1433	no	no	no	no	no	no	no	no	no	no
Rv1804c	no	no	no	no	no	<u>no</u>	no	no	no	no

You can also download examples in here [Download example1](#) [Download example2](#).

The first column in the tsv must be the genes id of the corresponding genomes and must be strictly named "id".

Then you can add as many columns as you wish with different values that can be either numeric or strings. In this example each column represents first and second line antibiotics against *M tuberculosis*. For each combination of genes and antibiotics there is a "yes" if the gene has a genetic polymorphisms associated with the respective drug and a "no" if hasn't.

To upload your properties you must click in "Add new Properties" in the "Filter" or the "Score" tab.



Once you click "Add new properties", you will be directed to a new page where you will be asked to the login. To do this, you have to click on "Log in to upload!".

Uploads

Log in to upload!

After login in, just by clicking in "Upload your properties" you will be direct to the tab where you can upload a tsv archive containing the properties of your interest. There you have to choose and upload your file.

Upload Metadata ×

The file must be plain text, with columns separated by tab and UTF-8 encoding. There is one mandatory field called "id", with the locus tag or Uniprot code. There are 2 types of columns, "values" or "numeric". Value columns must have no more than 20 different values. Numeric columns must have "." as decimal separator and no other symbol. To represent the absence of value in a numeric column, the string "NaN" must be used.

[Download Example 1](#) [Download Example 2](#)

Example

id	essential	chokepoint	stress fold change
Rv0064	yes	product	0.3
Rv1363	no	no	6
Rv1364	no	unknown	NaN

No file chosen
Select TSV File

Once uploaded this data you can use it to filter or to calculate a new score.

Then new data will be added in the "Metadata" button in order to use it to obtain a personalized ranking of drug targets.

Choosing promising pathways as putative targets of new drugs.

Numerous genomic sequencing projects have provided a nearly complete list of the components that are present in an organism, so post-genomic projects now focus on understanding metabolic and signaling networks, large multimeric complexes or even whole organisms. This emerging field of systems biology

provides a key framework for understanding cellular metabolism under different conditions, facilitating the discovery of new drugs. Due to this, the reconstruction, through bioinformatic tools, of pathogens metabolic networks is key to explore possible molecular targets (proteins) of novel drugs. As said before, Target-Pathogen allow users to select and study proteins not only according to properties such as the essential role in the metabolism (essentiality) and / or feasibility of being inhibited (druggable), but also to its contextual role (contextuality) in metabolic pathways. Moreover, it allows you to rank pathways with 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).


For example, if we want to determine which pathways are relevant for develop new therapies for polymyxin B-resistant *Klebsiella pneumoniae*, we should be interested in a scoring function as defined in equation 2 in order to assign a score to each pathway:


$$SF = C_x + Chk + C_y + H + E + C_{Kp} + Pb$$


Where C_x = (*pathways.completeness*) is the ratio between the total number of reactions of a pathway associated with a gene and the total number of enzymatic reactions present in the pathway. Chk (*pathways.norm_chokepoint*) is the proportion of reactions that are actually chokepoints in the pathway. C_y (*pathways.max_centrality*) is the ratio between the node centrality and the node with the biggest centrality in the entire metabolism. C_{Kp} reflects the presence of the different proteins belonging to pathway in pathogenic *Klebsiella pneumoniae* (*metadata.conserved_pathogen_norm*), H is where off-target criteria (*metadata.human_offtarget*) analysis is taking place and E defines essentiality of the pathway (*metadata.hit_in_deg*, *metadata.essential in mgh78578*) and, at last, P (*overexpressed in polymixin*) is the ratio between the genes present in the pathway and the overexpressed genes in polymyxin B-induced transcriptomic response (12).


Score


Sorts all / the filtered proteins by calculating a numeric value o score. Score formula is a weighted linear sum of the protein properties.



Activity



Biological Process



Localization


Pathways


Structure


Pocket


Metadata


Add new Properties


	Name	Description	Coefficient	Group		Norm.
X	completeness Show distribution	Proportion of reactions in the pathway with at least one known protein that catalize them	<input type="text" value="1"/>			0.13
X	max centrality Show distribution	Maximum betweenes centrality of all the reactions in the pathway, normalized by the reaction with max betweenes centrality in the whole graph	<input type="text" value="1"/>			0.13
X	human offtarget Show distribution	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).	<input type="text" value="1"/>	avg ▼		0.13
X	hit in deg Show distribution	Has a hit in Database of Essential Genes	<input type="text" value="1"/>	avg ▼	if is equal to <input style="width: 40px;" type="text" value="Yes"/>	0.13
X	essential in mgh78578 Show distribution	Hits with an essential gene of Klebsiella pneumoniae MGH78578	<input type="text" value="1"/>	avg ▼	if is equal to <input style="width: 40px;" type="text" value="Yes"/>	0.13
X	overexpressed in polymyxin Show distribution	Overexpressed in polymyxin B resistance induction (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5088521/)	<input type="text" value="1"/>	avg ▼	if is equal to <input style="width: 40px;" type="text" value="YES"/>	0.13
X	conserved pathogen norm Show distribution	Hit count in different pathogenic Kp strains divided by the total number of compared bacteria (39 pathogenic Kp).	<input type="text" value="1"/>	avg ▼		0.13
X	norm chokepoint Show distribution	Chokepoint reactions/reactions in pathway ratio	<input type="text" value="1"/>			0.13

Overall, a high value would mean that most genes in the pathway and the hole pathway itself fulfill most of the user defined criteria and therefore are attractive from the drug discovery point view. At last, we

defined a pathway as druggable if at least one of the proteins involved is druggable and rule out non-druggable pathways (in the Filter, at the left part of the screen, property " *druggable*" was added as a filter).

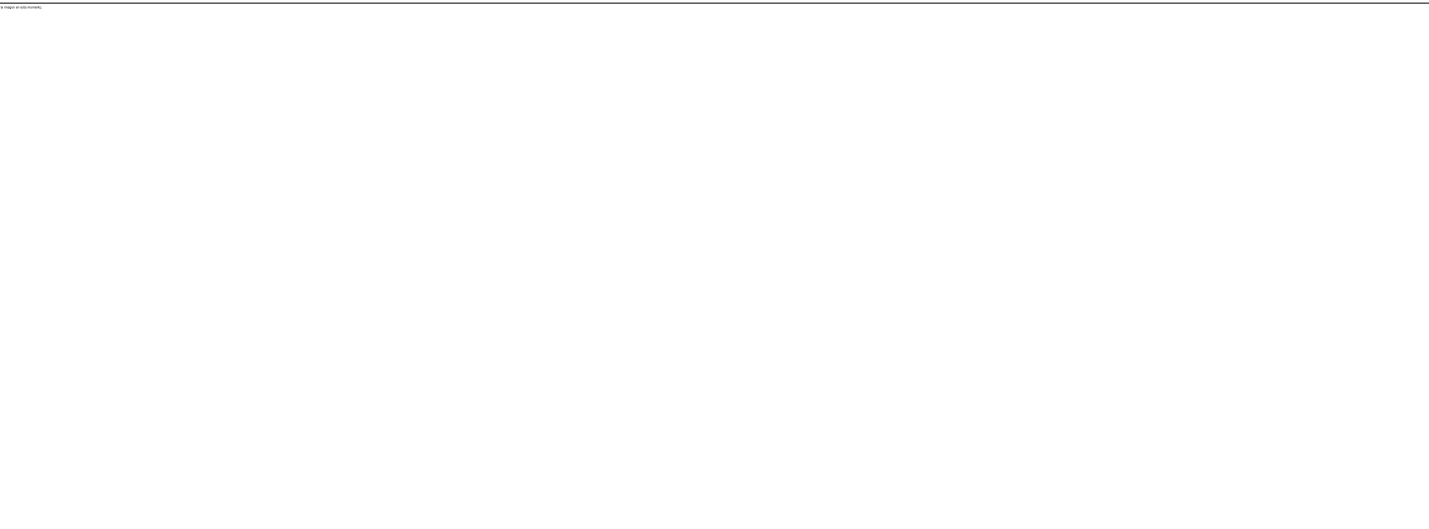
Filter

Removes the proteins that do not fulfill ALL the conditions


Pathways

Name	Description	Operation	Value
X druggable Show distribution	The pathway has at least one druggable protein	equal ▼	Yes ▼

Top five pathways are shown below.



1. The UniProt Consortium (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.*, **37**, D169–D174.
2. Finn,R.D. (2005) Pfam: the protein families database. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*.
3. Burley,S.K., Berman,H.M., Kleywegt,G.J., Markley,J.L., Nakamura,H. and Velankar,S. (2017) Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. *Methods Mol. Biol.*, **1607**, 627–641.
4. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.
5. Gene Ontology Consortium and Gene Ontology Consortium (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32**, 258D–261.
6. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.
7. Guruprasad,K., Savitha,S. and Babu,A.V.N. (2005) Computational tools for the analysis of heteroatom groups and their neighbours in protein tertiary structure. *Int. J. Biol. Macromol.*, **37**, 35–41.
8. Furnham,N., Holliday,G.L., de Beer,T.A.P., Jacobsen,J.O.B., Pearson,W.R. and Thornton,J.M. (2014) The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.*, **42**, D485–9.
9. 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.
10. 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.
11. 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.
12. Ramos,P.I.P., Custódio,M.G.F., Quispe Saji,G.D.R., Cardoso,T., da Silva,G.L., Braun,G., Martins,W.M.B.S., Girardello,R., de Vasconcelos,A.T.R., Fernández,E., *et al.* (2016) The polymyxin B-induced transcriptomic response of a clinical, multidrug-resistant *Klebsiella pneumoniae* involves multiple regulatory elements and intracellular targets. *BMC Genomics*, **17**, 737.