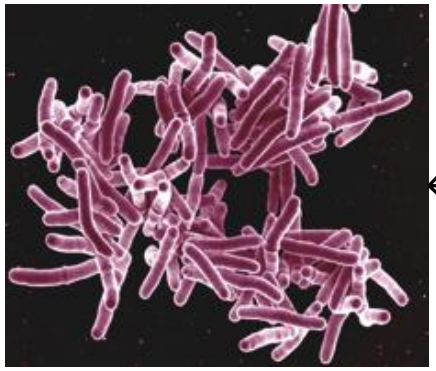
# *Target-Pathogen: a structural bioinformatic approach to prioritize drug targets in pathogens*

Darío Fernández Do Porto
**Argentine Consortia of Bioinformatics (BIA)**
Science School
University of Buenos Aires

# Are pathogens fighting back?



Pathogens

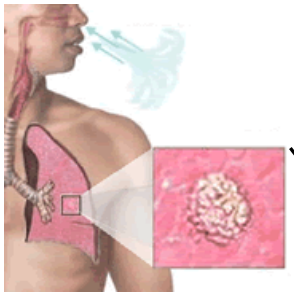
Antimicrobial resistance (AMR) threatens the effective prevention and treatment of an ever-increasing range of infections caused by bacteria, parasites, viruses and fungi.

The cost of health care for patients with resistant infections is higher than care for patients with non-resistant infections due to longer duration of illness, additional tests and use of more expensive drugs.

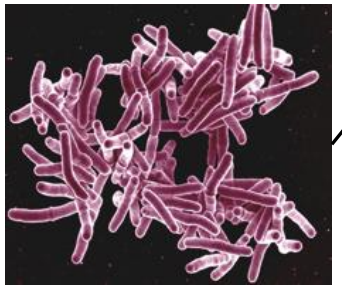
Globally, 480 000 people develop multi-drug resistant TB each year, and drug resistance is starting to complicate the fight against HIV and malaria, as well.

World Health Organization

# New Technologies and new paradigms



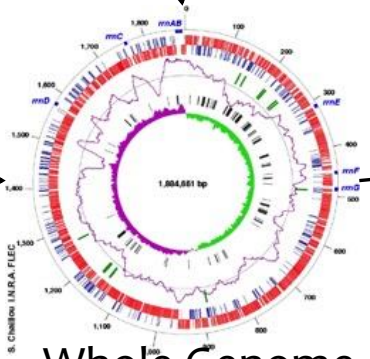
Patiens



Pathogens

Next Generation Sequencing

Multiple Strains



Whole Genome Sequence

Experimental Data

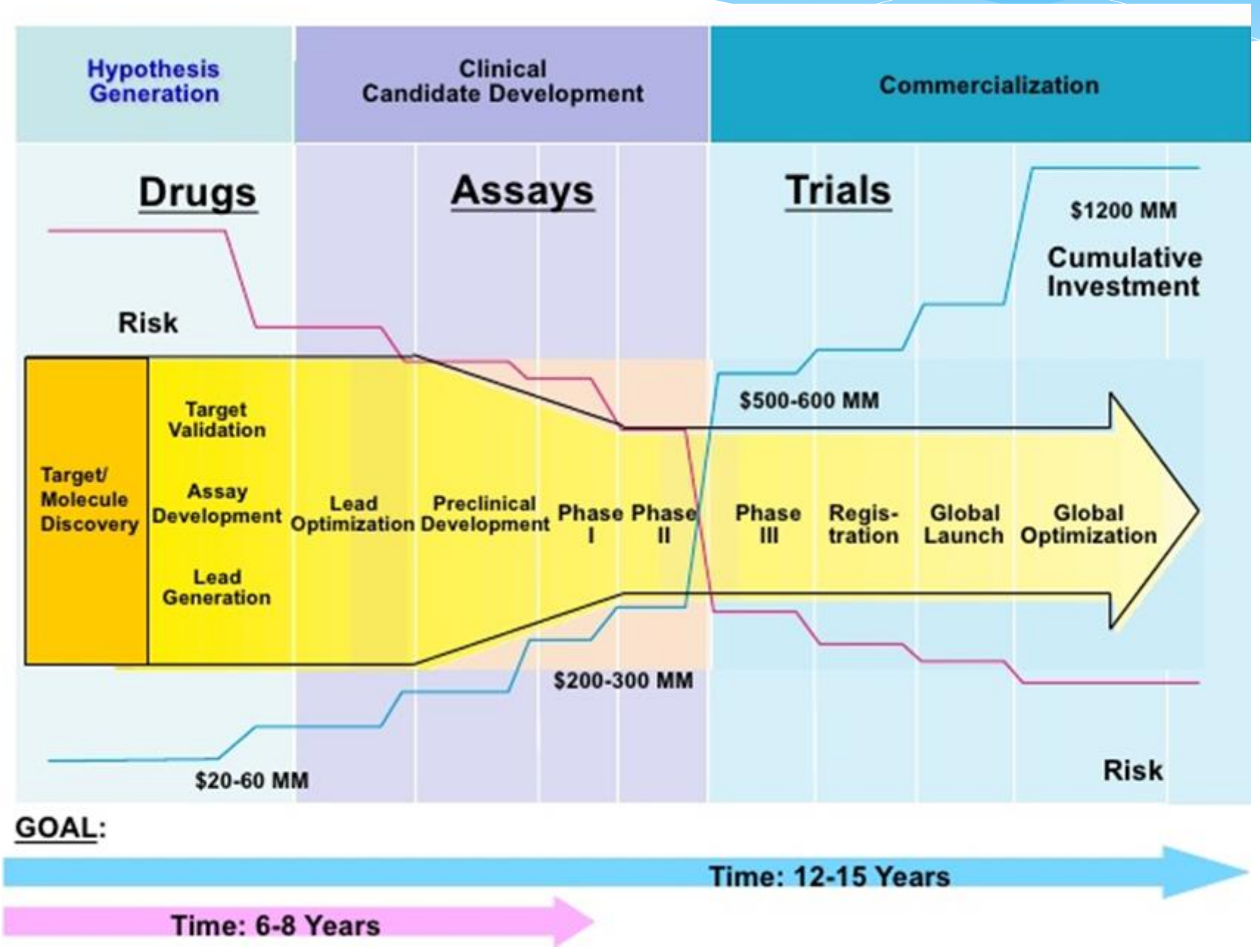
- Expression
- Proteomics
- Essensial
- Mutagenesis
- Resistance

## Bioinformatics

New Protein Targets?

New Drugs?

# Standard Drug discovery pipeline

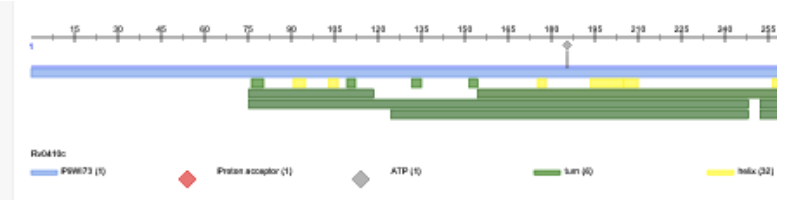


# target.sbg.qb.fcen.uba.ar

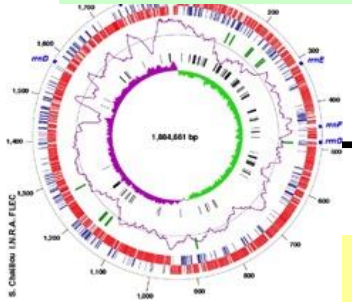
# Whole genome analysis and structurome prediction

## WG anotation of protein properties

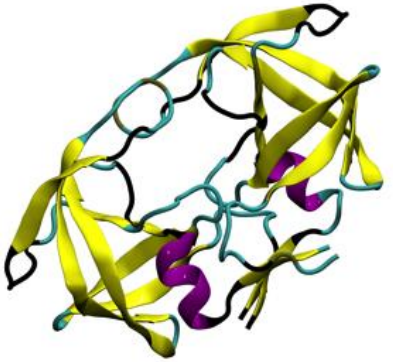
- Localization, Gene Ontology, KEGG, Relevant Residues, PFAM, EC Enzyme, etc…



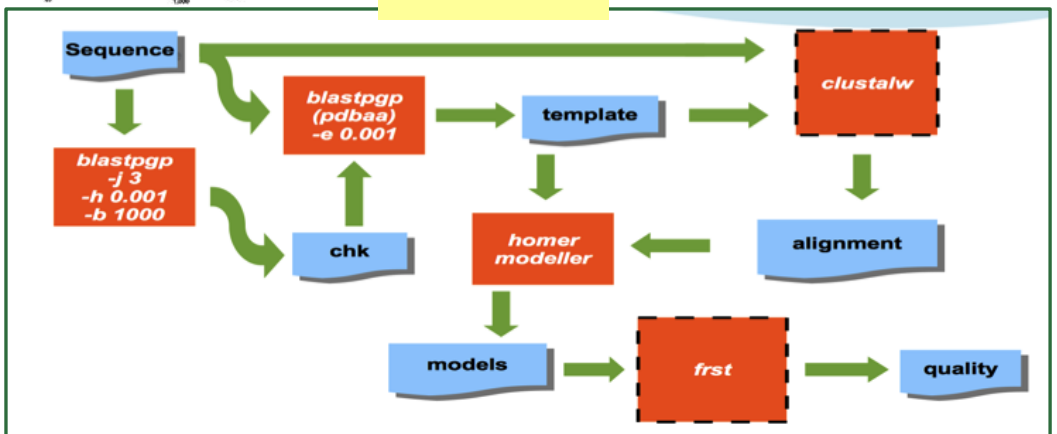
## WG protein structure prediction



PQITLWKRPIVTIKVEGQLREALLDTGADDTVLEDINLSGKWKPKII
GGI
RGFVKVKQYEDILIEICGHRAVGAVLVGPTPANIIGRNMLTQIGCTL
NF
PQITLWKRPIVTIKVEGQLREALLDTGADDTVLEDINLSGKWKPKII
GGI



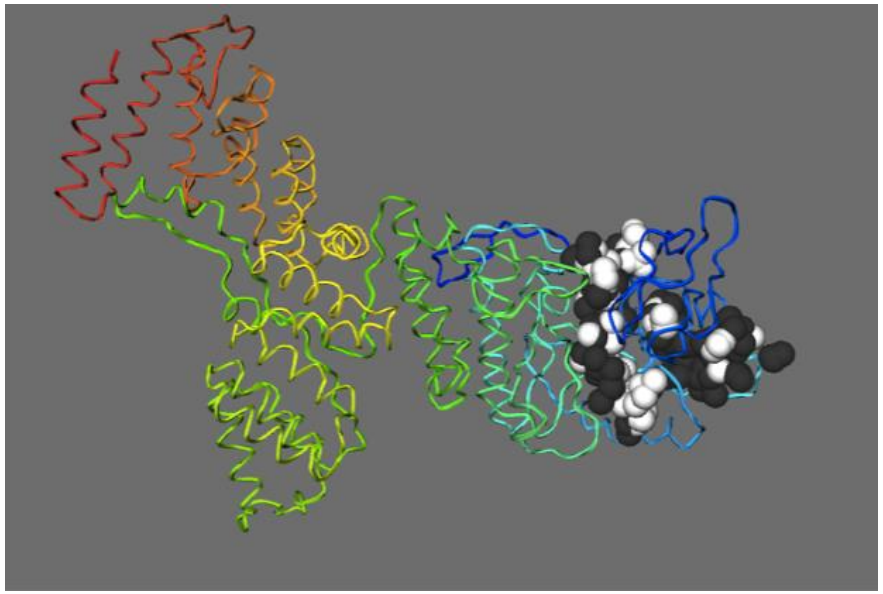
PIPELINE



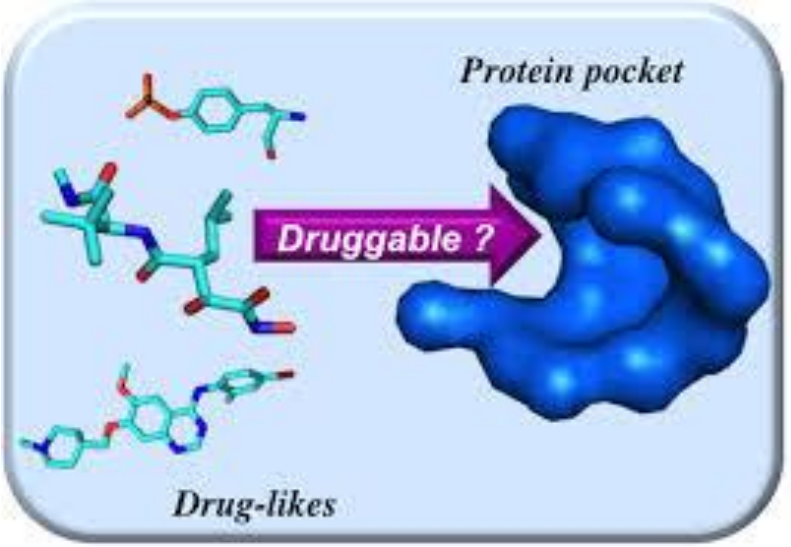

Structure With Quality Assesment for drug development

# How can we select a protein that binds a Drug like compound?

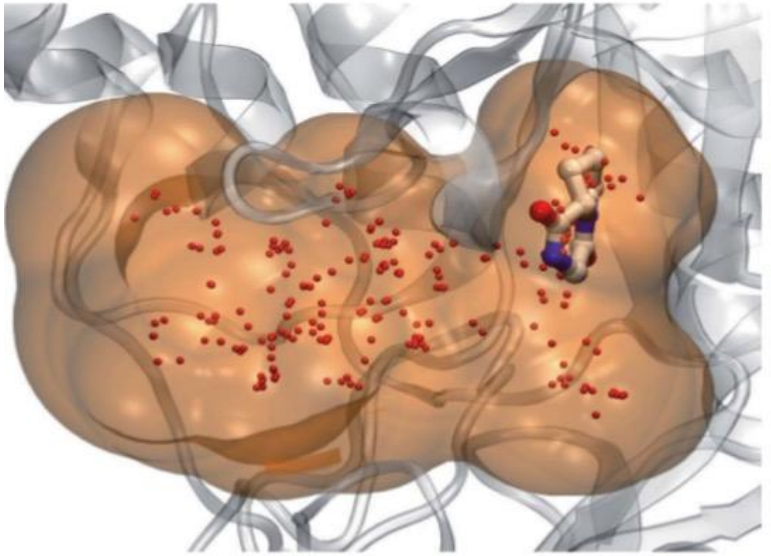
Find pockets?



Concept of Druggability

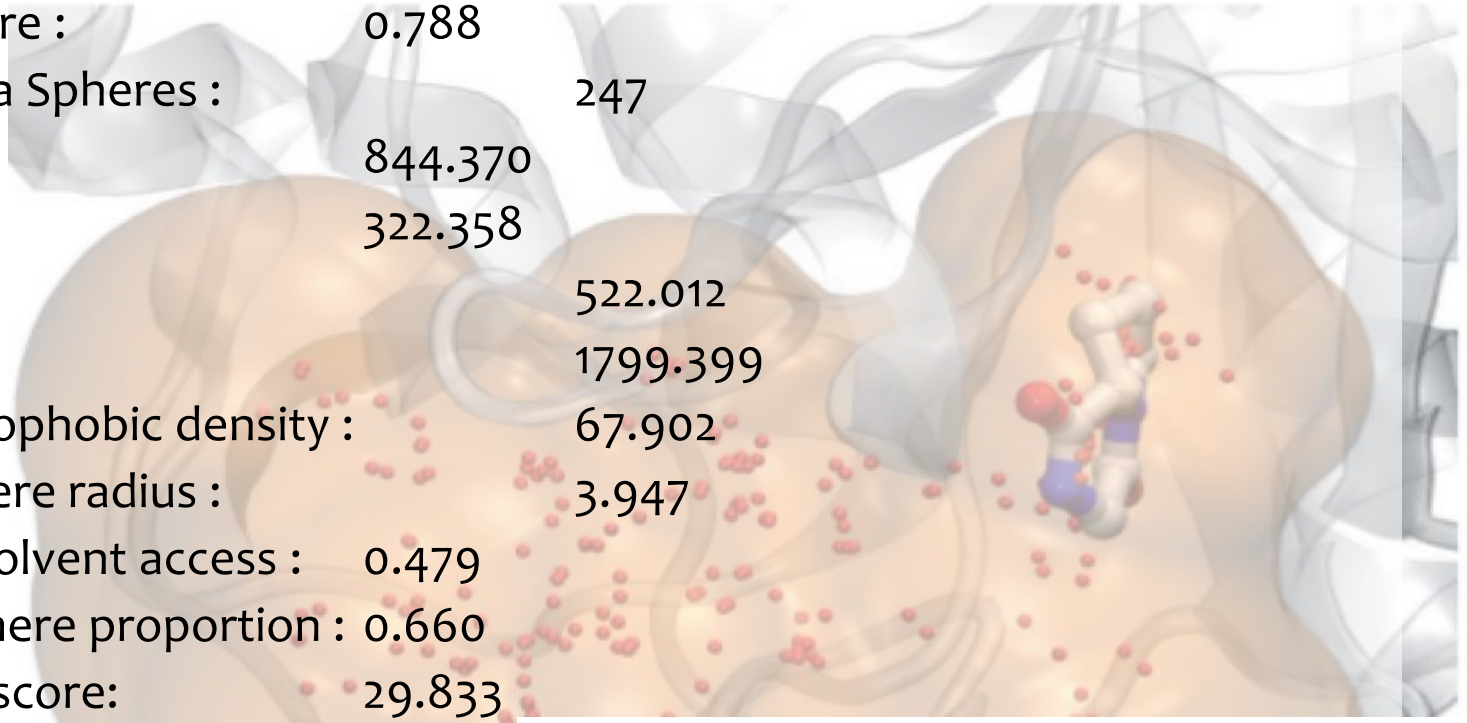



To identify a POCKET!
Fpocket: We implemented a pocket detector program
We estimated pocket properties and Determine druggability



# A pocket inside a protein

- Druggability Score : 0.788
- Number of Alpha Spheres : 247
- Total SASA : 844.370
- Polar SASA : 322.358
- Apolar SASA : 522.012
- Volume : 1799.399
- Mean local hydrophobic density : 67.902
- Mean alpha sphere radius : 3.947
- Mean alp. sph. solvent access : 0.479
- Apolar alpha sphere proportion : 0.660
- Hydrophobicity score: 29.833
- Aminoa Acid Composition
- Distances between Aminocids

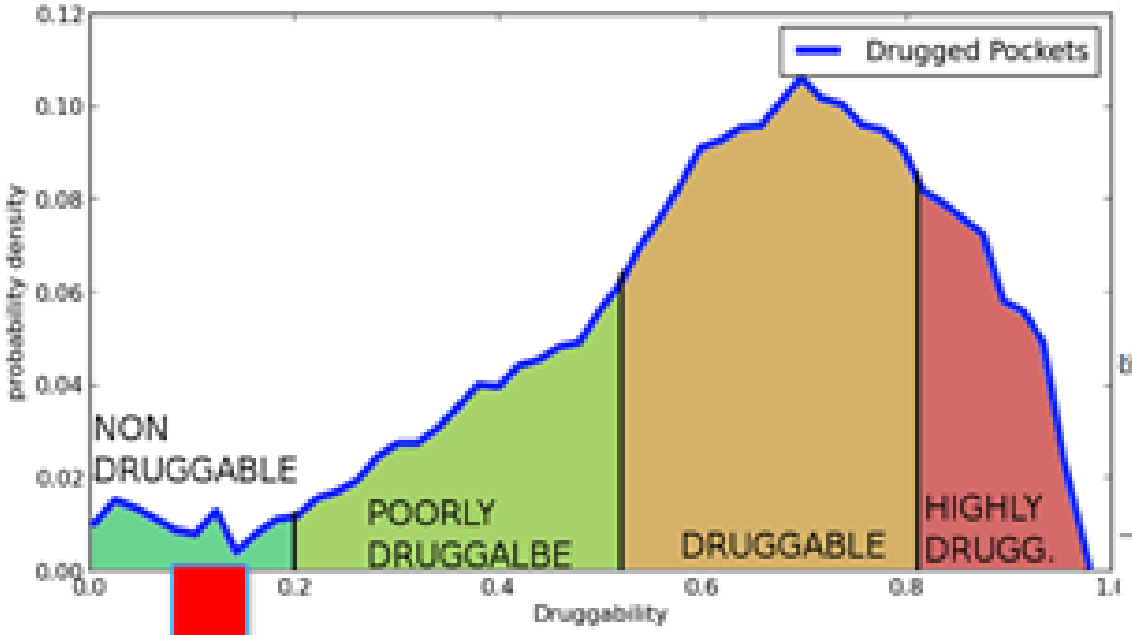


fpocket

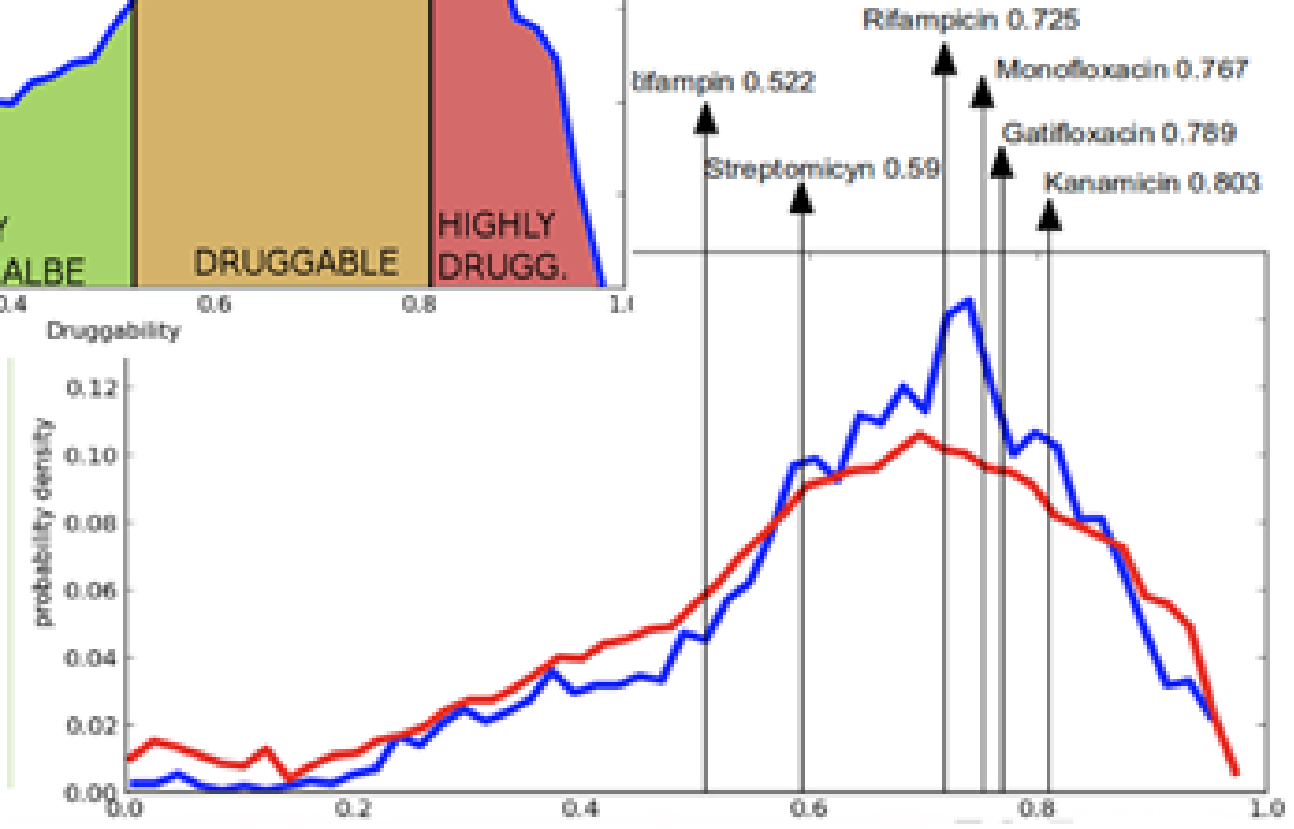
scalable high performance pocket detection

Relevant Information related to the protein pockets

# Druggability in patogens




Some Relevant Targets Are not Druggable!!



# How to select an attractive target from the metabolic point of view

Pathway Tools

# Manual Curation

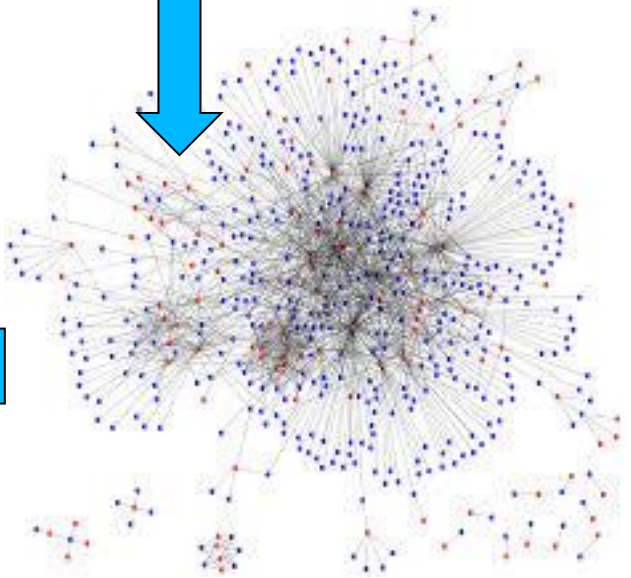
Pathway Tools

SBML™

.sif

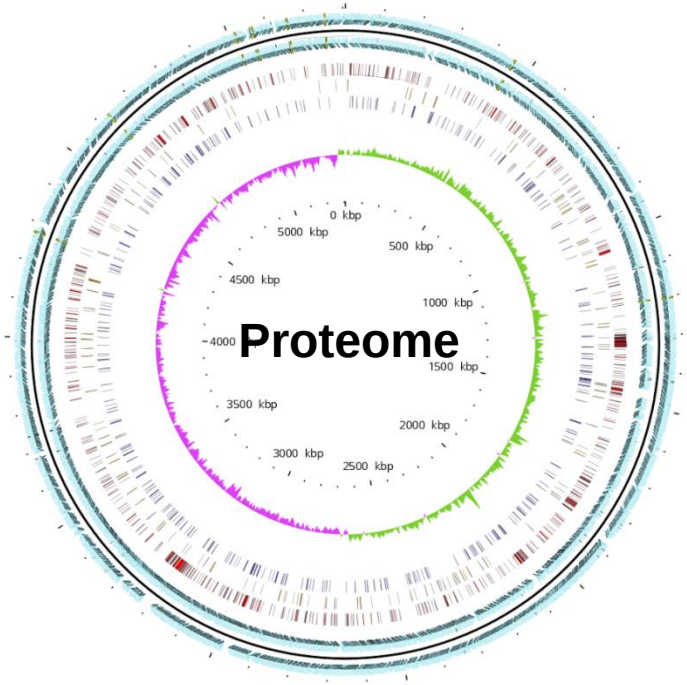
```
R1 linkedwith R2
R2 linkedwith R4
R4 linkedwith R3
```

Cytoscape



Graph parameters

# Discarding side effects




CDS
rRNA
tRNA
Hypothetical
Putative
Hypothetical conserved
GC skew +
GC skew -

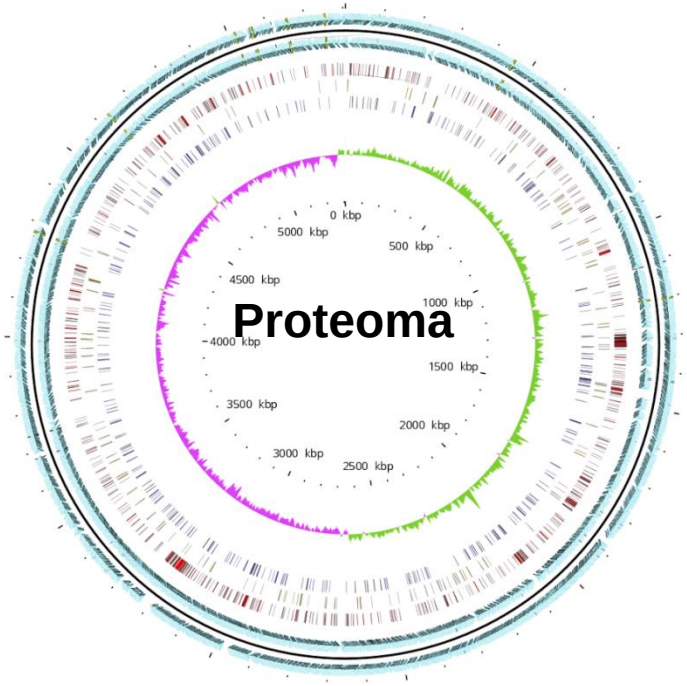
BLASTp

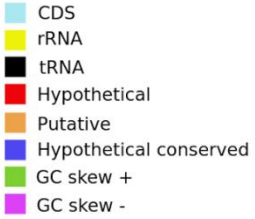
Identity >0.4

Score *off-target*: 1-(%Id) of the best hit

**Posible Interferencia**

# Metadata
## Essenciality



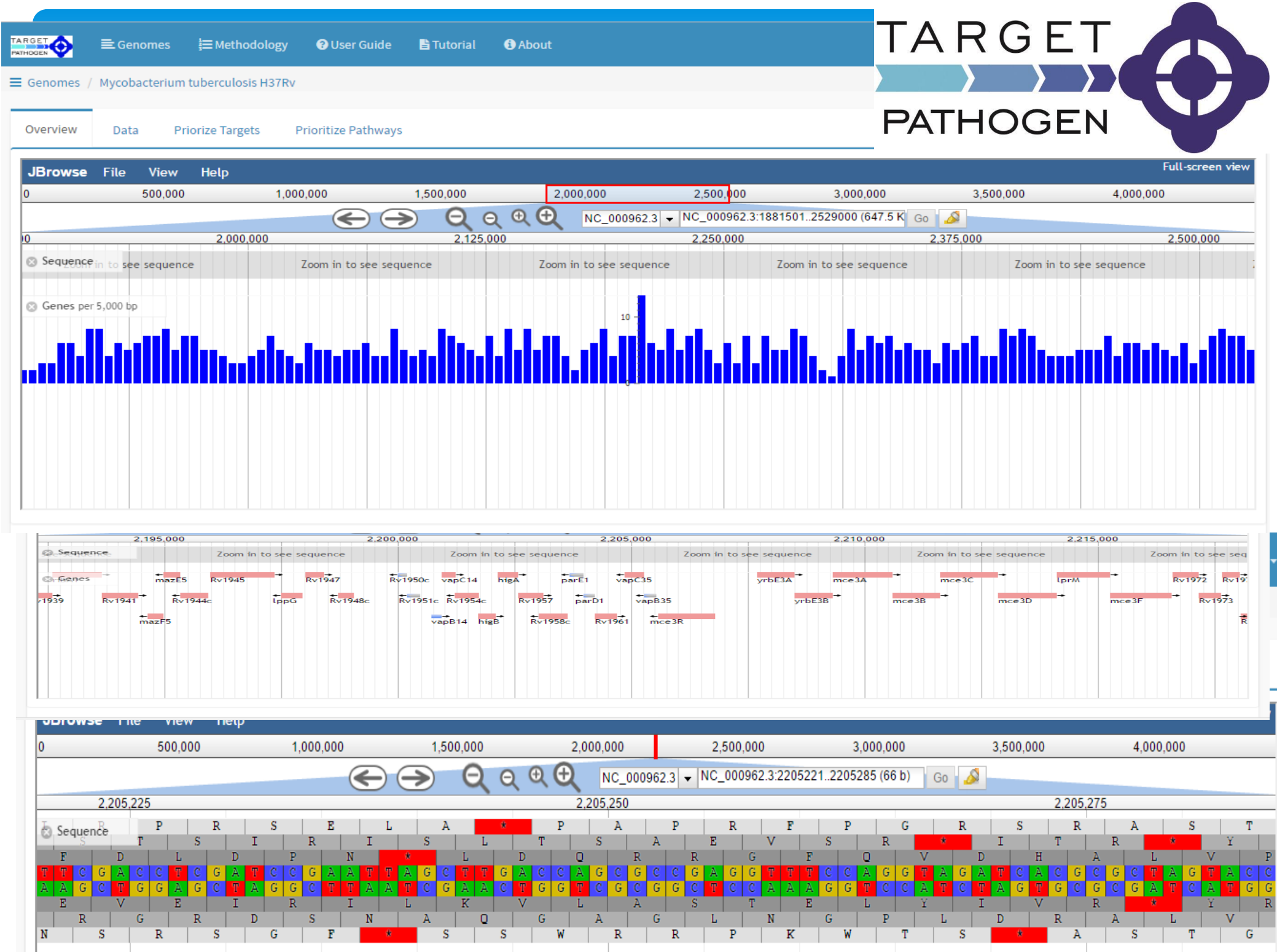




E-value < a $10^{-5}$

**Essenciality**

BBH (BLASTp)

DEG


TARGET PATHOGEN
Genomes
Methodology
User Guide
Tutorial
About
Genomes / Mycobacterium tuberculosis H37Rv
Overview
Data
Priorize Targets
Prioritize Pathways
JBrowse File View Help
Full-screen view
0 500,000 1,000,000 1,500,000 2,000,000 2,500,000 3,000,000 3,500,000 4,000,000
NC_000962.3
NC_000962.3:1881501..2529000 (647.5 K
Go
2,000,000 2,125,000 2,250,000 2,375,000 2,500,000
Sequence
Zoom in to see sequence
Genes per 5,000 bp
10
2,195,000 2,200,000 2,205,000 2,210,000 2,215,000
Genes
mazE5 Rv1945 Rv1947 Rv1950c vapC14 higA parE1 vapC35 yrbE3A mce3A mce3C lprM Rv1972
Rv1941 Rv1944c lppG Rv1948c Rv1951c Rv1954c Rv1957 parD1 vapB35 yrbE3B mce3B mce3D mce3F Rv1973
mazF5 vapB14 higB Rv1958c Rv1961 mce3R
NC_000962.3
NC_000962.3:2205221..2205285 (66 b)
2,205,225 2,205,250 2,205,275

TARGET PATHOGEN

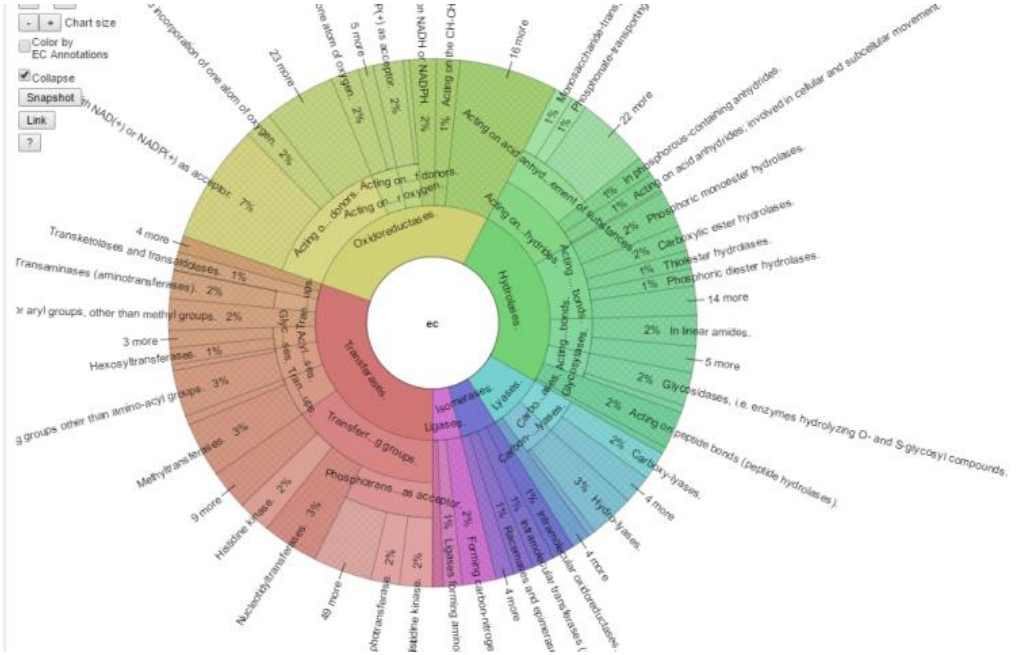
Search Gene Product By

Keyword

Gene

GO Term

H37Rv Pathways




## Statistics

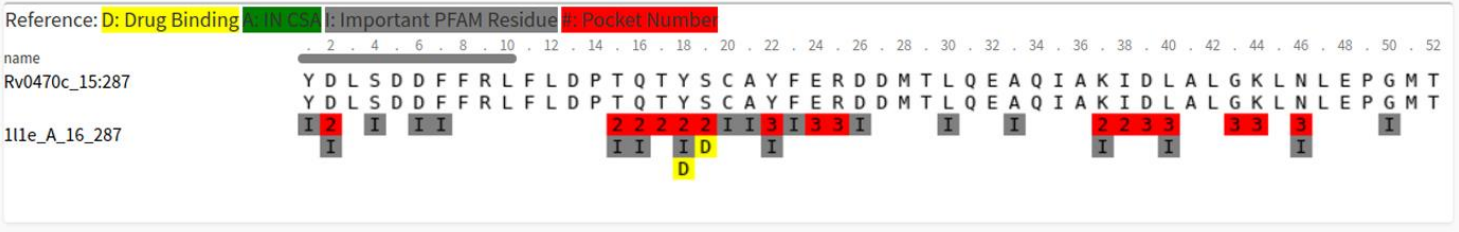
| | |
|---|---|
| Proteins | 4023 |
| Polypeptide domain | 3323 |
| Go | 3184 |
| Ec | 1067 |
| Polypeptide structural motif | 534 |
| Transmembrane polypeptide region | 360 |
| Signal peptide | 133 |

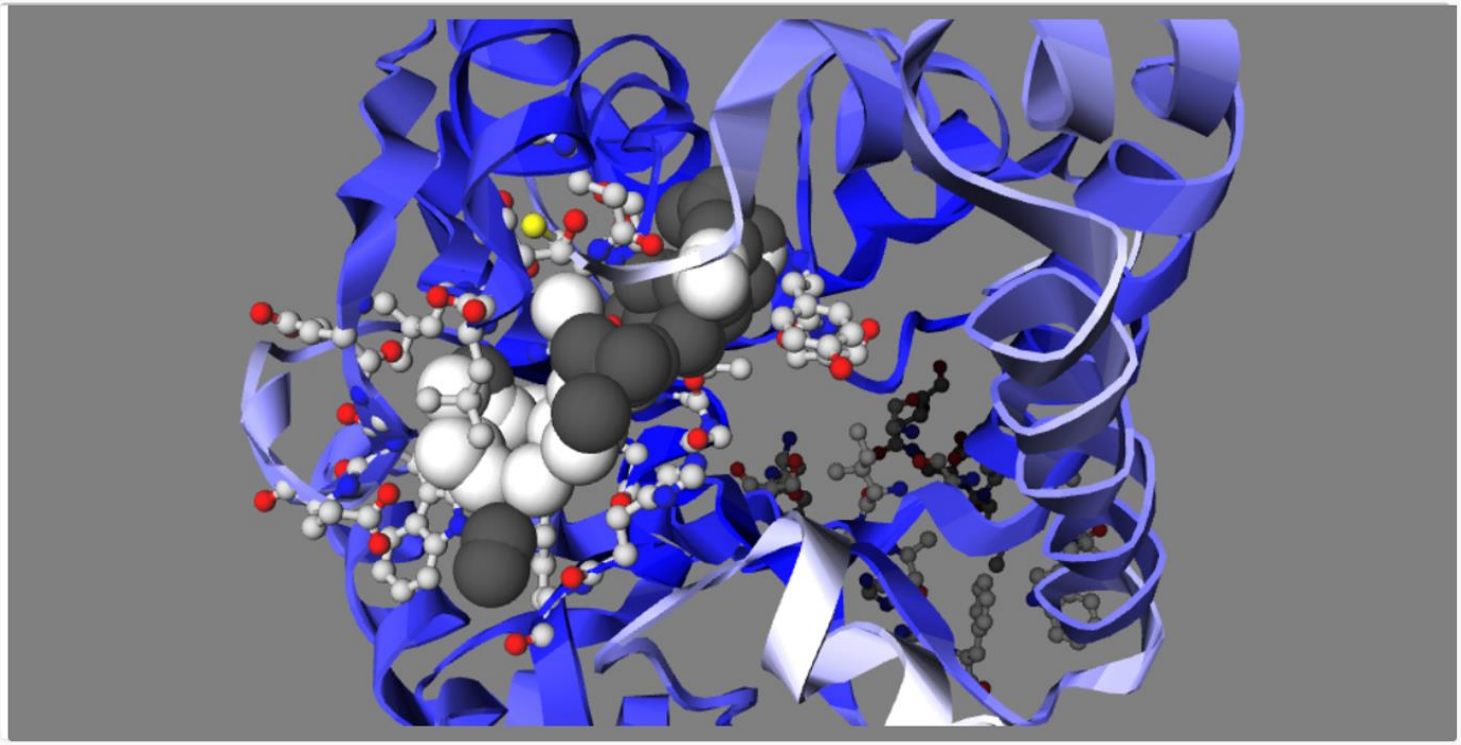
Showing 1 to 7 of 7 entries

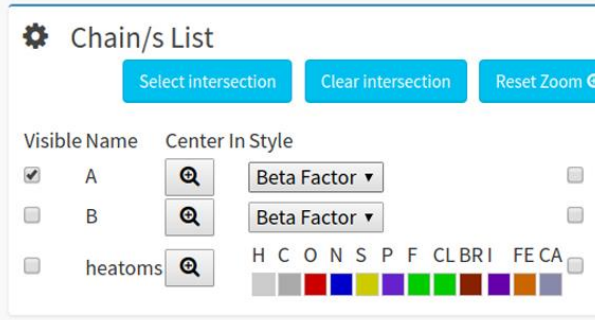
# OVERVIEW

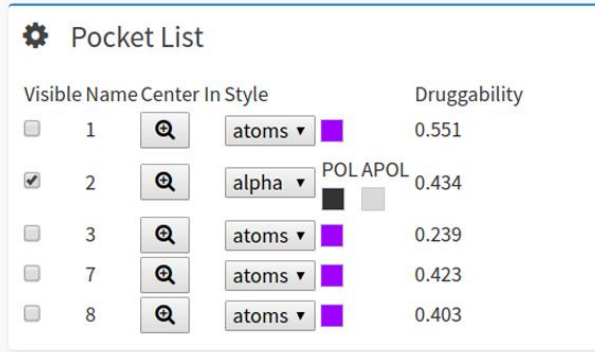
Genome Browser.
EC and GO searches

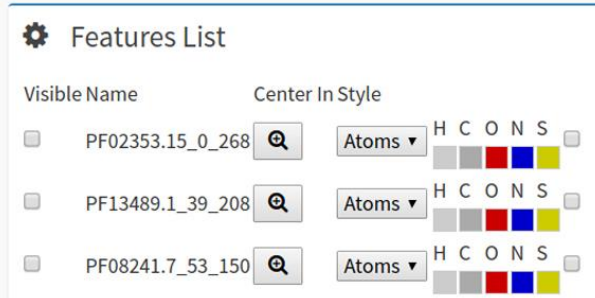
# Protein structure





TARGET PATHOGEN

# Filters

## Filter

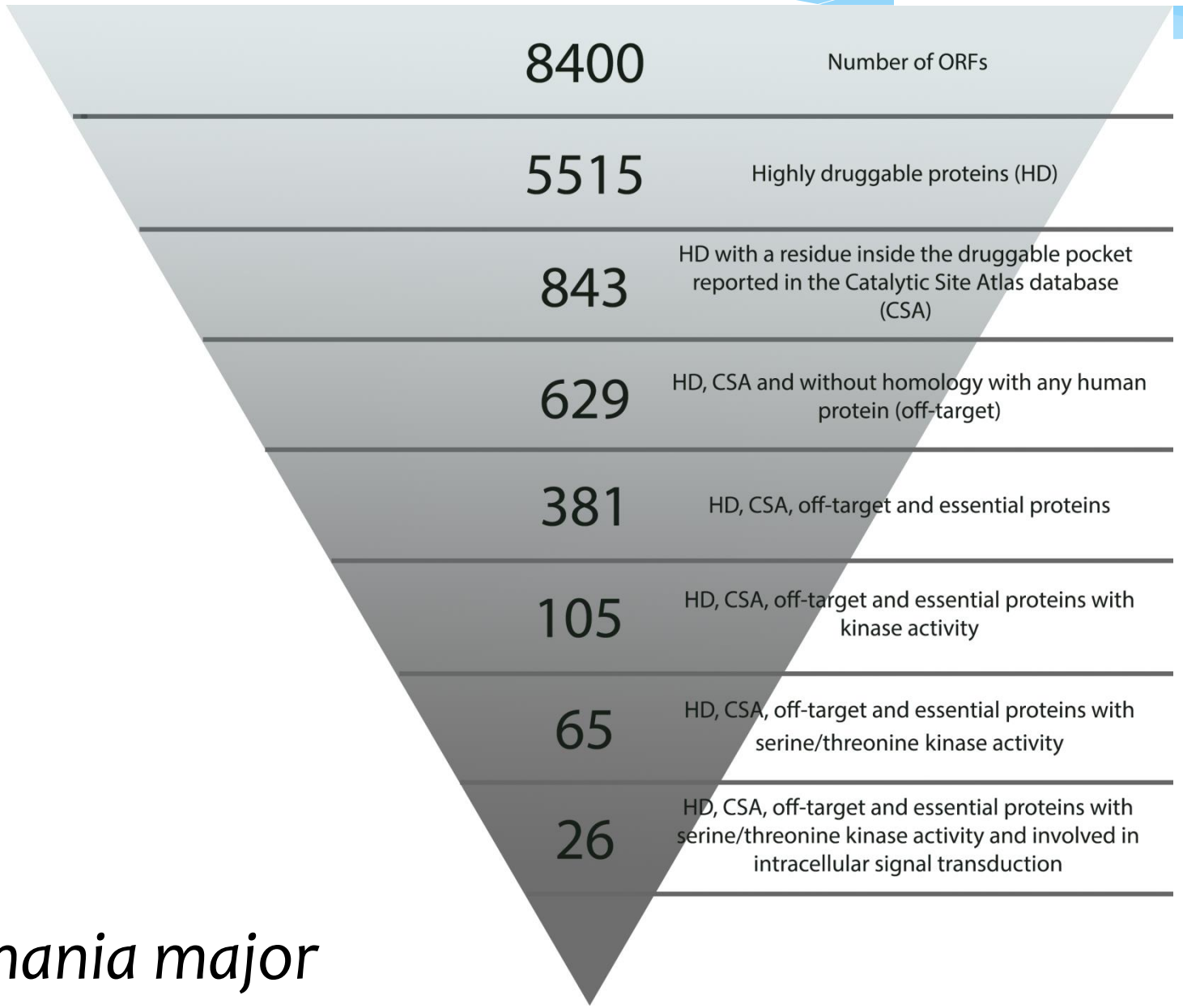
Removes the proteins that do not fullfill ALL the conditions

Keyword | Activity | Biological Process | Localization | Pathways | Structure | Pocket

Metadata | Add new Properties

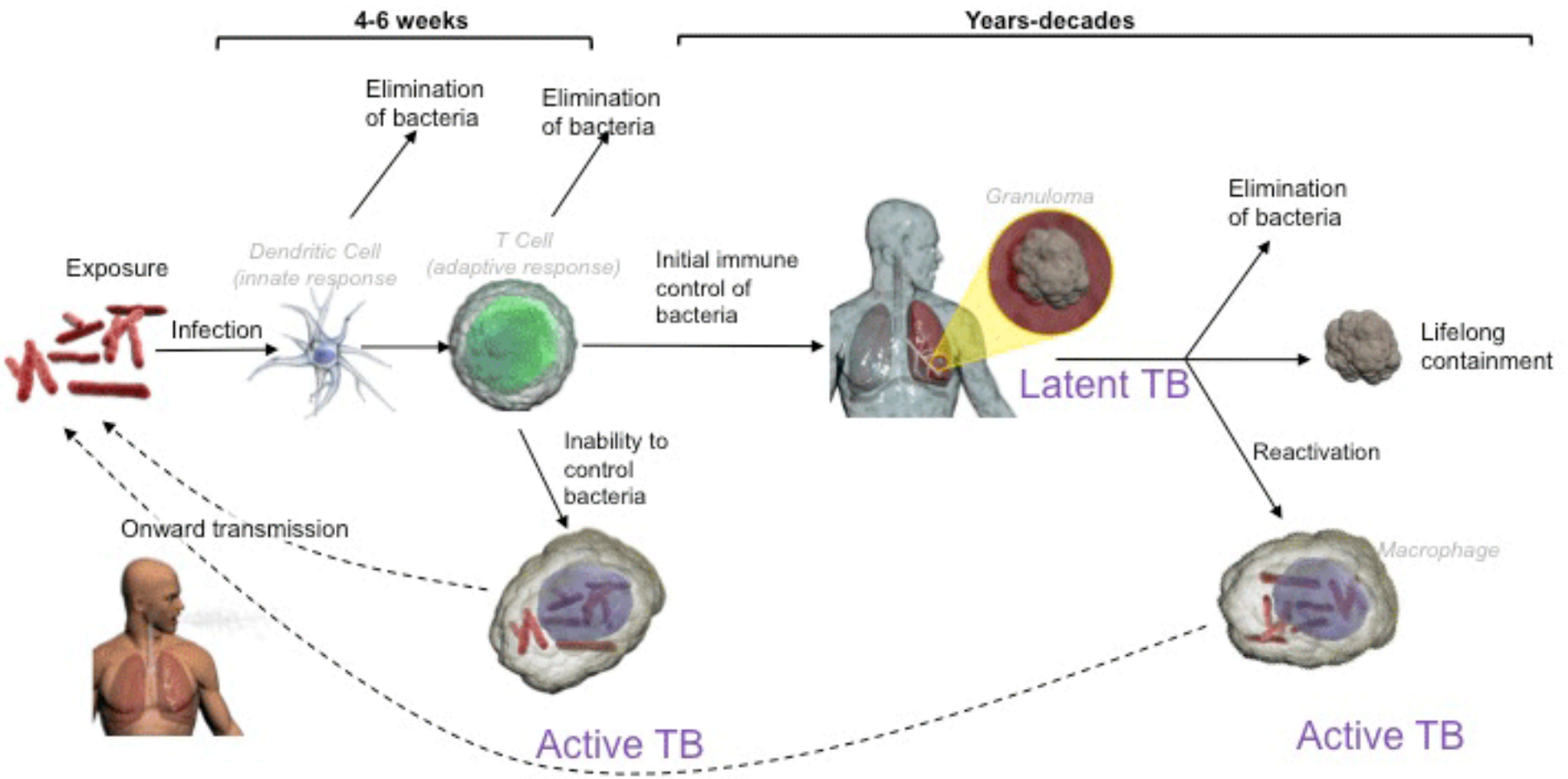
| | Name | Description | Operation | Value |
|---|---|---|---|---|
| X | druggability | Druggability score from the most druggable pocket. Druggable: druggability > 0.5 / Highly Druggable druggability > 0.7. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4014675/) | > | 0.5 |
| X | essentiality | Critical for the organism survival (https://www.ncbi.nlm.nih.gov/pubmed/26791267 | equal | true |
| X | human_offtarget | Max identity in a simple blast alignment with a human protein | < | 0.4 |

TARGET PATHOGEN




*Leishmania major*

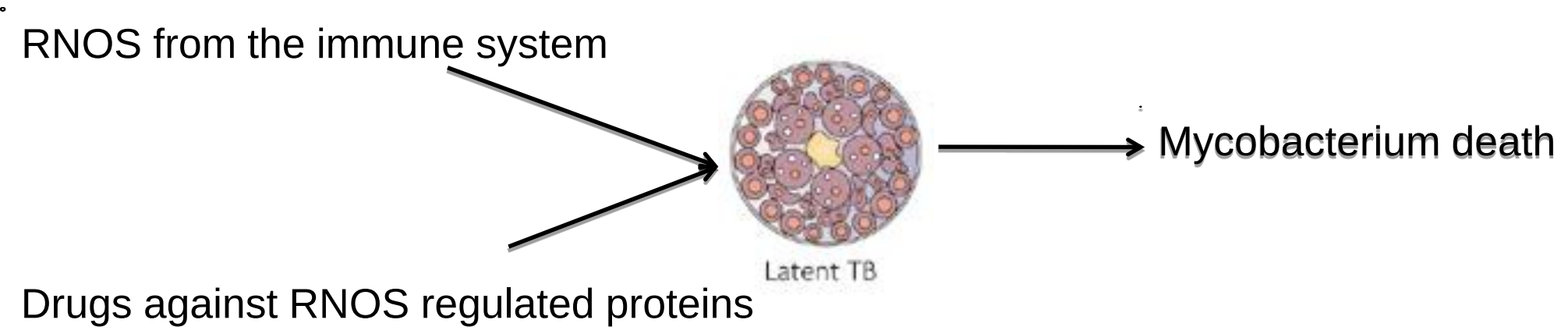
# Latent tuberculosis




- M. tuberculosis has the remarkable capacity to survive years within the hostile environment of the macrophage.
- Within the macrophage, tuberculosis bacilli is exposed to RNOS stress .
- There is not treatment for latent tuberculosis.

# How to kill latent *M. tuberculosis*

- Hipótesis:
  - if we know which proteins are targeted by RNOS and kill M. tuberculosis bacilli, we might be able to inhibit them with drugs, resulting in a synergistic bactericidal effect



# What features makes a protein a good target for laten tuberculosis drug selection?

Druggabilty

No side effects

Essenciality

Biologically Relevant

Important in the metabolic context

# Scoring function

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

## 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<br>Show distribution | Overexpressed in model of stress (https://www.ncbi.nlm.nih.gov/pubmed/26791267) | 0.25 | if is **equal** to true | 0.13 |
| X | overexpression starvation<br>Show distribution | Overexpressed in model of starvation (https://www.ncbi.nlm.nih.gov/pubmed/26791267) | 0.25 | if is **equal** to true | 0.13 |
| X | overexpression infection<br>Show distribution | Overexpressed in model of infection (https://www.ncbi.nlm.nih.gov/pubmed/26791267) | 0.25 | if is **equal** to true | 0.13 |

# Newly and Revalidated Mtb targets

Newly and revalidated *Mtb* targets found using structural druggability, metabolic importance analysis and expression data in infection mimicking conditions. Revalidated targets are taken from [25].

| Protein name | Rv | Status | Druggability | Pathway (importance) | Profile expression |
|---|---|---|---|---|---|
| Inositol-3-phosphate synthase | Rv0046c | New target | 0.719 | Myo-inositol biosynthesis (0.3871) L-1-phosphatidyl-inositol biosynthesis (Mycobacteria) (0.6063). mycothiol biosynthesis (0.5370) | Str, Hyp, Sta |
| 3-phosphoshikimate 1-carboxyvinyltransferase | Rv3227 | New target | 724 | Chorismate biosynthesis from 3-dehydroquinate (0.4828) | Str |
| O-acetylhomoserine aminocarboxypropyltransferase | Rv3340 | New target | 635 | Homocysteine biosynthesis (0.4681) | Hyp, Sta, Inf |
| 3-oxoacyl-[acyl-carrier-protein] synthase 2 | Rv2246 | New target | 709 | Mycolate biosynthesis (0.4517) fatty acid biosynthesis initiation II (0.3883) 8-amino-7-oxononanoate biosynthesis I (0.3765) | Str, Hyp |
| Octanoyltransferase | Rv2217 | New target | 703 | Lipoate biosynthesis and incorporation I (0.4529) | Sta |
| Bifunctional protein GlmU | Rv1018c | New target | 911 | UDP-N-acetyl-D-glucosamine biosynthesis I (0.4523) | Str, Hyp, Inf |
| Rv1465 | Rv1465 | New target | 926 | [2Fe–2S] iron-sulfur cluster biosynthesis | Str |
| 1D-myo-inositol 2-acetamido-2-deoxy-alpha-D-glucopyranoside deacetylase | RV1170 | Revalidated | 781 | Mycothiol biosynthesis (0.5370) | Str |
| Sulfate adenylyltransferase subunit 2 | Rv1285 | Revalidated | 891 | Selenate reduction (0.4579) sulfate activation for sulfonation (0.4326) | Str, Hyp, Sta |
| dTDP-glucose 4,6-dehydratase | Rv3464 | Revalidated | 676 | dTDP-L-rhamnose biosynthesis I (0.4459) | Str |
| Enoyl-[acyl-carrier-protein] reductase [NADH] | Rv1484 | Revalidated | 919 | 8-amino-7-oxononanoate biosynthesis I (0.3765) stearate biosynthesis II (bacteria and plants) (0.3700) | |
| 3-methyl-2-oxobutanoate hydroxymethyltransferase | Rv2225 | Revalidated | 937 | Phosphopantothenate biosynthesis I (0.4351) | Str, Hyp, Inf |
| Mycocyclosin synthase | Rv2276 | Revalidated | 887 | Mycocyclosin biosynthesis (0.4435) | Hyp |

# Prioritisize pathways

## 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. | 1 | | 0.50 |
| X | chokepoint | The protein catalyzes a chokepoint reaction | 1 | if is equal to true | 0.50 |

Score = centrality + chokepoint

SF=((Emgh+Edeg)/2+Cv+Cy +chk)/4 +Pb

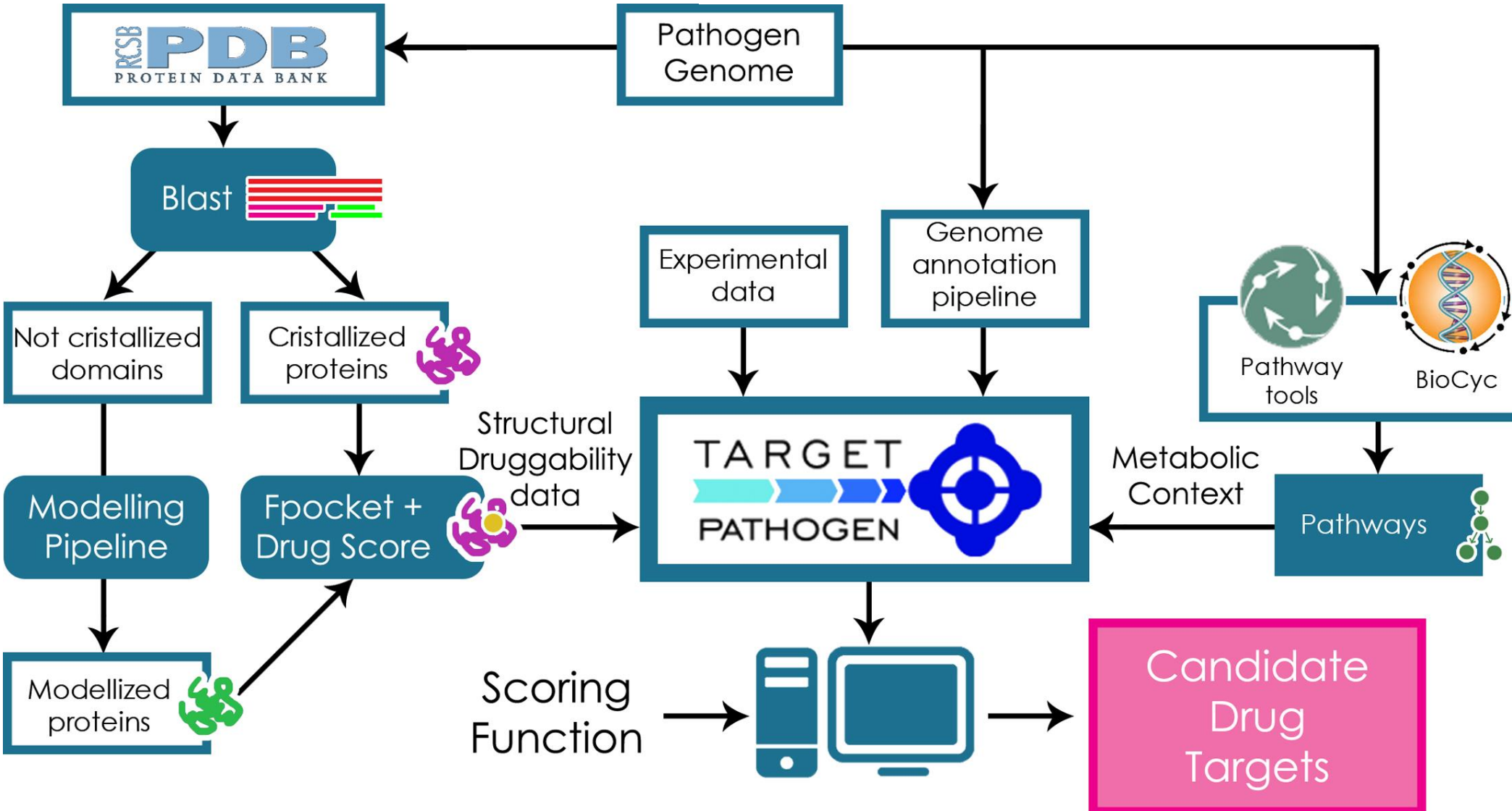
Nucleic Acids Research, 2017 1
doi: 10.1093/nar/gkx1015

# Target-Pathogen: a structural bioinformatic approach to prioritize drug targets in pathogens

**Ezequiel J. Sosa[1,2], Germán Burguener[1,2], Esteban Lanzarotti[1,2], Lucas Defelipe[1,2], Leandro Radusky[1,2], Agustín M. Pardo[3], Marcelo Marti[1,2,3], Adrián G. Turjanski[1,2,3,*] and Darío Fernández Do Porto[1,2,*]**

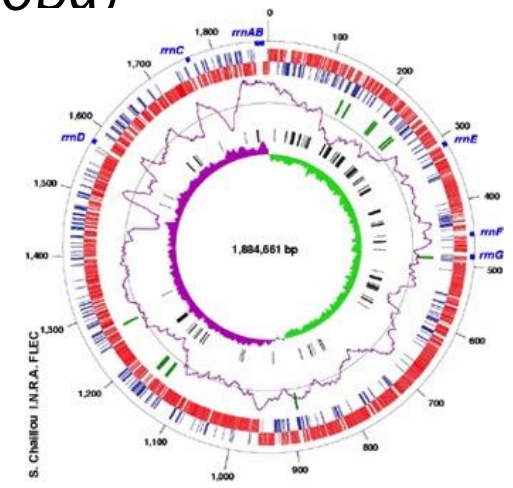
[1]IQUIBICEN-CONICET, Ciudad Universitaria, Pabellón 2, C1428EHA Ciudad de Buenos Aires, Argentina, [2]Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, C1428EHA Ciudad de Buenos Aires, Argentina and [3]Plataforma de Bioinformática Argentina (BIA), Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, C1428EHA Ciudad de Buenos Aires, Argentina

Received August 14, 2017; Revised September 29, 2017; Editorial Decision October 12, 2017; Accepted October 16, 2017



# Different Pathogens

- *Mycobacterium Tuberculosis (Marti, Piuri, UBA): Database 2014, Tuberculosis 2015*
- *Corynebacterium paratuberculosis (Acevedo, B. Horizonte): BMC Genomics, 2014; BMC Genomics, 2015, Frontiers in Genomics 2018*
- *Klebsiella pneumoniae (Nicolas, Rio de Janeiro): Scientific Reports 2018*
- *Leishmania Major (Ramos, UFB, Bahia)*
- *Bartonella bacilliformis (Abraham Espinosa,* University of São Paulo )
- *Trypanozoma Cruzi (Pablo Smircich, Montevideo)*
- *Staphylococcus aeurus (Dr.Bocco, Universidad de Córdoba)*



# Plataforma de Bioinformática Argentina

A Turjanski M Martí

**Microorganisms Genomics**
**Ing. Ezequiel Sosa**
**Dr. Germán Burguener**
**Lic. Agustín Pardo**
**Andrés Fernández Benevento**
**Federico Serral**

**Human Genomics**
**Lic. Jonathan Zayat**
**Dr. Sergio Nemirovsky**
**Dr. Juan Pablo Alracon**
**Sebastian Vishnopolska**
**Lic. Geronimo Dubra**

Ministerio de Ciencia, Tecnología e Innovación Productiva
Presidencia de la Nación

AGENCIA

CONICET

UNIVERSIDAD DE BUENOS AIRES

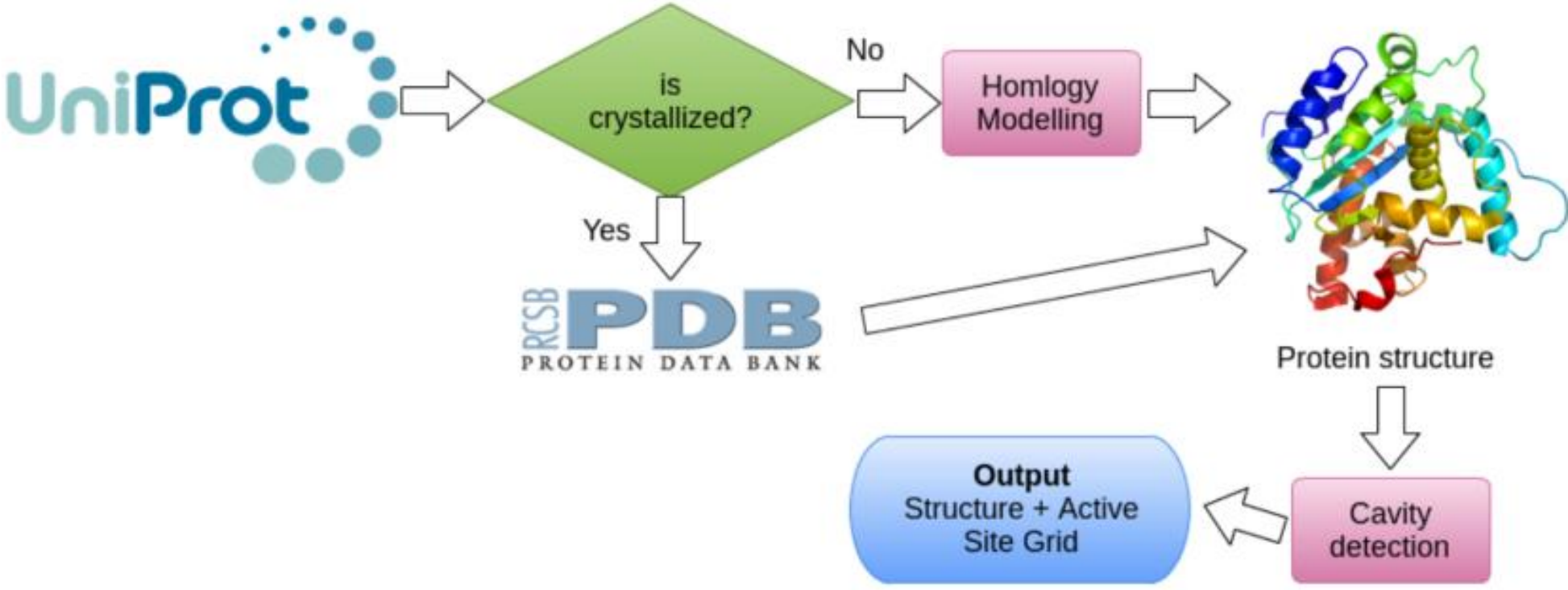
# Argentina

dariofd@gmail.com

# THANKS

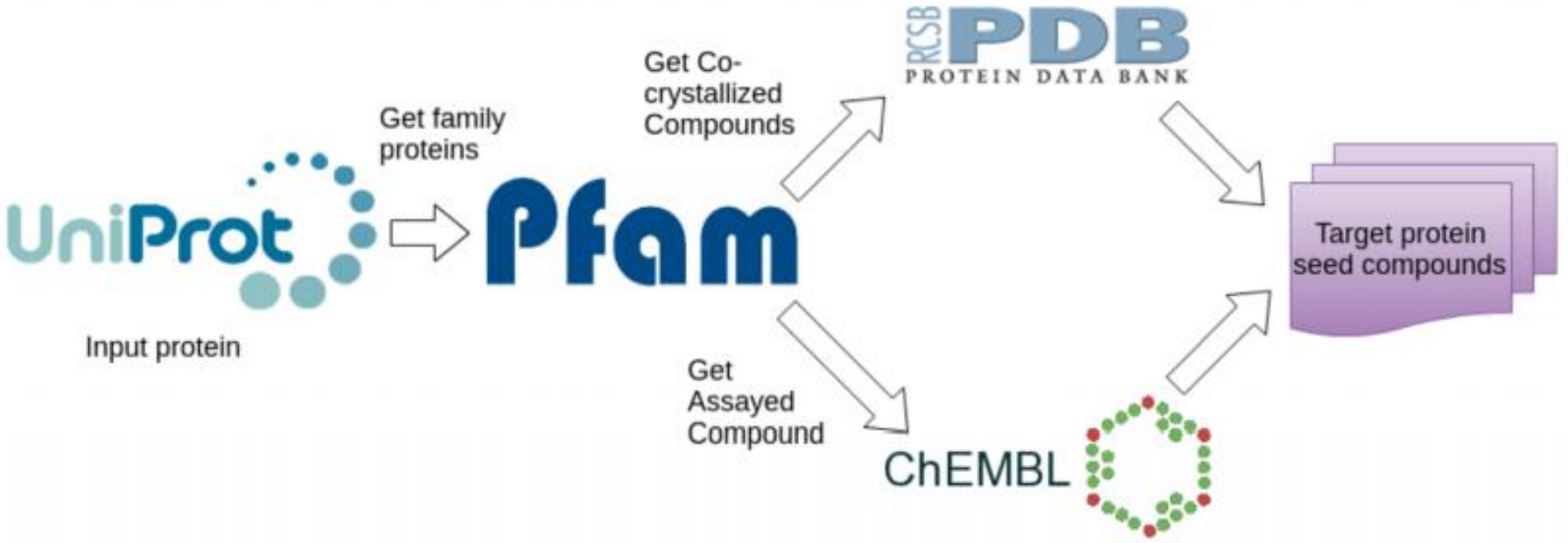
# LigQ
http://ligq.qb.fcen.uba.ar/

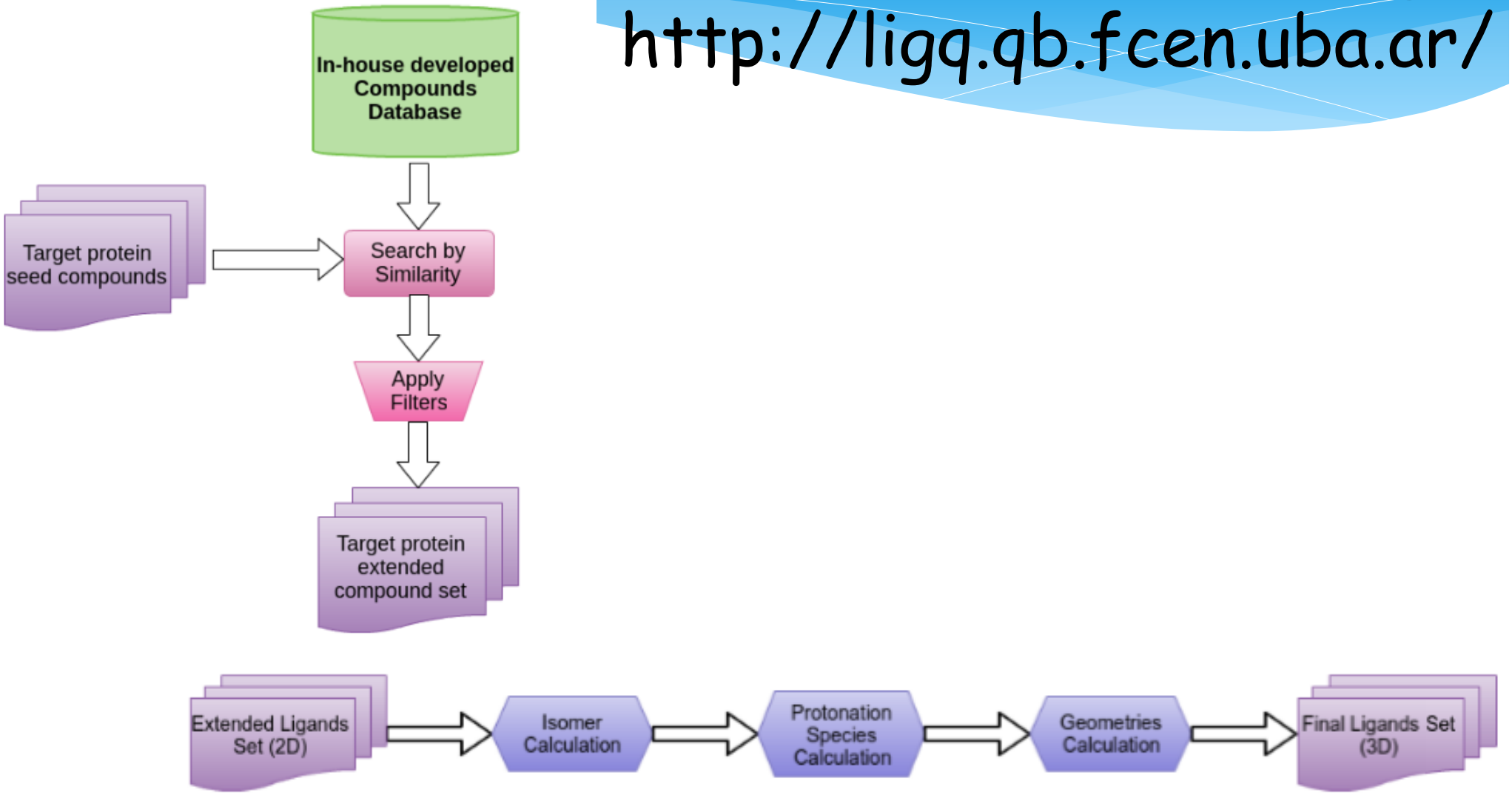
## Pocket Detection Module



# LigQ
# http://ligq.qb.fcen.uba.ar/

## Módulo de detección ligandos



# LigQ

http://ligq.qb.fcen.uba.ar/