Characterizing the pocketome of *Mycobacterium tuberculosis* and application in rationalizing polypharmacological target selection

Praveen Anand and Nagasuma Chandra[*]

Department of Biochemistry, Indian Institute of Science,

Bangalore-560012, India.

KEYWORDS

Polypharmacology, Binding site similarity network, Pocketome, Drug-target network, Drug repurposing.

Correspondence to:

Prof. Nagasuma Chandra
Department of Biochemistry
Indian Institute of Science
Bangalore, India, 560012
Tel:+91-80-22932892
Email: nchandra@biochem.iisc.ernet.in

Supplementary Information

Figure S1

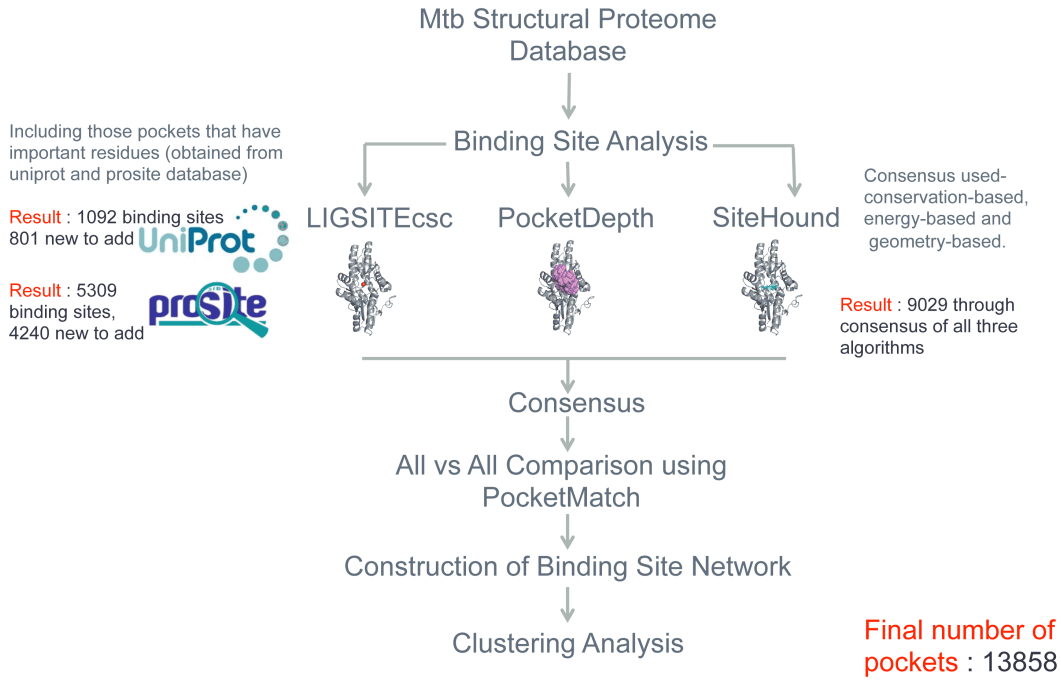


Figure S1. **A workflow for *Mtb*-pocketome analysis.** This workflow depicts the different steps involved in obtaining the pockets from protein structures using consensus predictions from three different algorithms for binding site prediction. All the PocketDepth[21] pockets within 5Å radius of predicted LIGSITEcsc[33] pockets were selected. For all pockets thus identified, SiteHound[34], an energy method that scans favorable interaction zones for a methyl probe within the protein was also used as a filter to extract out a set of consensus ligand binding sites.

Table S1

Cut-off	Analysis	Justification	Remarks
PMAX ≥ 0.6	(Figure 1, Figure2)Ligand Associations, Clustering coefficient score for PPI index	Significant similarities observed at this cut-off and implies binding of similar ligands	Default cut-off used in this study
PMAX ≥ 0.6±0.1	(Figure 6) Drug bank compound association to sets, atleast one site within the cluster should have PMAX ≥0.6 and the rest of the sites with PMAX ≥ 0.5	A drugbank compound should have significant similarity (PMAX ≥0.6) to atleast one of the pockets in the set and part similarity (PMAX ≥ 0.5) to rest of the sites	Chosen compound is ensured to have bindings site similairty with all the sites within the sets
PMAX ≥ 0.7	Figure 3A and Figure 3B, Binding site similarity network and clustering of Mtb pocketome and MOAD dataset	The stringency of similarity is increased to pick up highly similar sites for polypharmacology	More confident similarities were sought (at the cost of missing some). Hence a higher cut-off. This implies that all the sites are significantly similar.
PMAX ≥ 0.6	Figure S4, Approved drug association	The stringency of similarity is increased to pick up highly similar sites for polypharmacology	More confident associations were sought
PMAX ≥ 0.4	Comparison to other studies and current literature	0.4 is the minimum required value to indicate significant hit for a pocket	The definition/boundary of binding site is not fixed, so the cut-off was lowered to overcome the sensitivity to variations in binding site definitions

Different cut-off used in this study. Various PocketMatch cut-offs that have been used in this study along with the justification and description has been mentioned here for each of the analysis carried out.

Table S2

Network Type	Network	Corresponding Figure in manuscript	Nodes	Edges	Purpose
1	Binding-Site similarity network of Mtb Pocketome at PMAX ≥ 0.7.	Figure 5A	Predicted binding sites from Mtb Pocketome	Similarity between binding sites detected using PocketMatch algorithm with a cutoff PMAX ≥ 0.7.	To obtain sets of similar binding sites in form of clusters using MCODE algorithm.
2	Binding-Site similarity network of known binding sites from PDB obtained from MOAD database	Figure 5B	Binding sites extracted from MOAD dataset of known Protein-ligand complexes from PDB	Similarity between binding sites using PocketMatch algorithm with same cut-off as above used for predicted sites	This network was constructed to validate the clusters obtained using MCODE algorithm.
3	Sub-network of Figure 5A , each binding site set showing binding site similarities to 'known drug-binding sites DB'	Figure 6	Two types of nodes in the network. Circular nodes – Binding sites of the corresponding sets Triangular nodes – Binding sites of drugs	Two kinds of edges. Plain edges – Binding site similarity between the pockets in the pocketome (b/w circular nodes) Dashed edges – Binding site similarity between the predicted site and the drug binding sites . (b/w circular and triangular nodes)	To visualize the similarity relationship between the binding site sets and the known drug binding sites.
4	Bipartite network of target proteins in Mtb and its association with approved drugs.	Figure S4	Two types of nodes Blue nodes– Approved drugs. Red nodes– Targets in Mtb proteome	Edges here represent the similarity between a binding site of the corresponding protein from the Mtb Pocketome and the binding site of approved drug. This is calculated using PocketMatch and the cut-off used for similarity is same as above.	This network was constructed to rank-list the targets with polypharmacological profile and the drugs with potential of repurposing.

Different network variants constructed in this study. The reference to the corresponding figures in the manuscript is also mentioned with the detailed description of the node and the edges in each of them. The purpose on such a network construction is also mentioned briefly.

Table S3

Experiment/Method	Tools/Dataset	Reference No.
1. Obtaining Structural Proteome	• Mtb Structural Proteome • Protein Data Bank • MODBASE	27
2. Deriving consensus binding site	• PocketDepth • LigsiteCSC • SiteHound • Prosite • Uniprot	22, 34, 35, 36, 37 and 38
3. Construction of Binding site similarity network	• PocketMatch • Cytoscape	23, 40
4. Deriving drug binding sites	• Drugbank • Drugport • TBDrugome	44, (http://www.ebi.ac.uk/thornton-srv/databases/drugport/) 61
5. Validation (Predicting pocket, Ligand association, Clustering, drug association)	• Procognate • KEGG • TM-Align • Autodock Vina • BindingMOAD	47, 51, 42, 50, 48
6. Extracting sets/groups of similar binding sites from binding Site similarity network (nodes as binding sites and edges with similarity relationship PMAX >= 0.7)	• MCODE	39
7. Construction of binding-site similarity network at PMAX ≥0.6 and obtaining CC for each node in the network	• igraph package in R	41
7. Clustering of bipartite network (with two types of nodes predicted Mtb pockets of target proteins and approved drug binding sites, edges being the binding site similarity >=0.7)	• TNET	60

Datasets and tools used in the study. Various datasets and tools used in this study has been explained with the brief description of the experiment/method and the corresponding reference number in the manuscript.

Table S4

PM Score Cut-off	No. of Hits for PMIN	No. of Hits for PMAX
≥0.4	13547	13297
≥0.5	11601	11029
≥0.6	7612	6906
≥0.7	2385	1644
≥0.8	114	14
≥0.9	15	0
1	7	0

PDB ligand hit frequency against *Mtb* Pocketome. This table describes the number of hits found for Mtb Pocketome when compared to known binding sites present in PDB at different PMIN and PMAX cut-off scores reported by PM.

Supplementary Text 1

Mtb pockets obtained were first clustered at 0.6 PMAX. The binding site network obtained with this cut-off was then subjected to MCODE algorithm that resulted in 105 clusters. Around 7379 binding sites (nodes) were present in this network. Remaining (13858 - 7369) sites were unique singletons. These singletons could be considered as individual site types each. So the total number of site types now add upto = 105 + (13858 - 7369) = 6584 types.

The stringency of PMAX was increased to 0.7 and binding site similarity network now obtained, consisted of 698 nodes (binding sites) and MCODE algorithm clustered them into 29 sets. These could be considered further for polypharmacological applications.

Figure S2

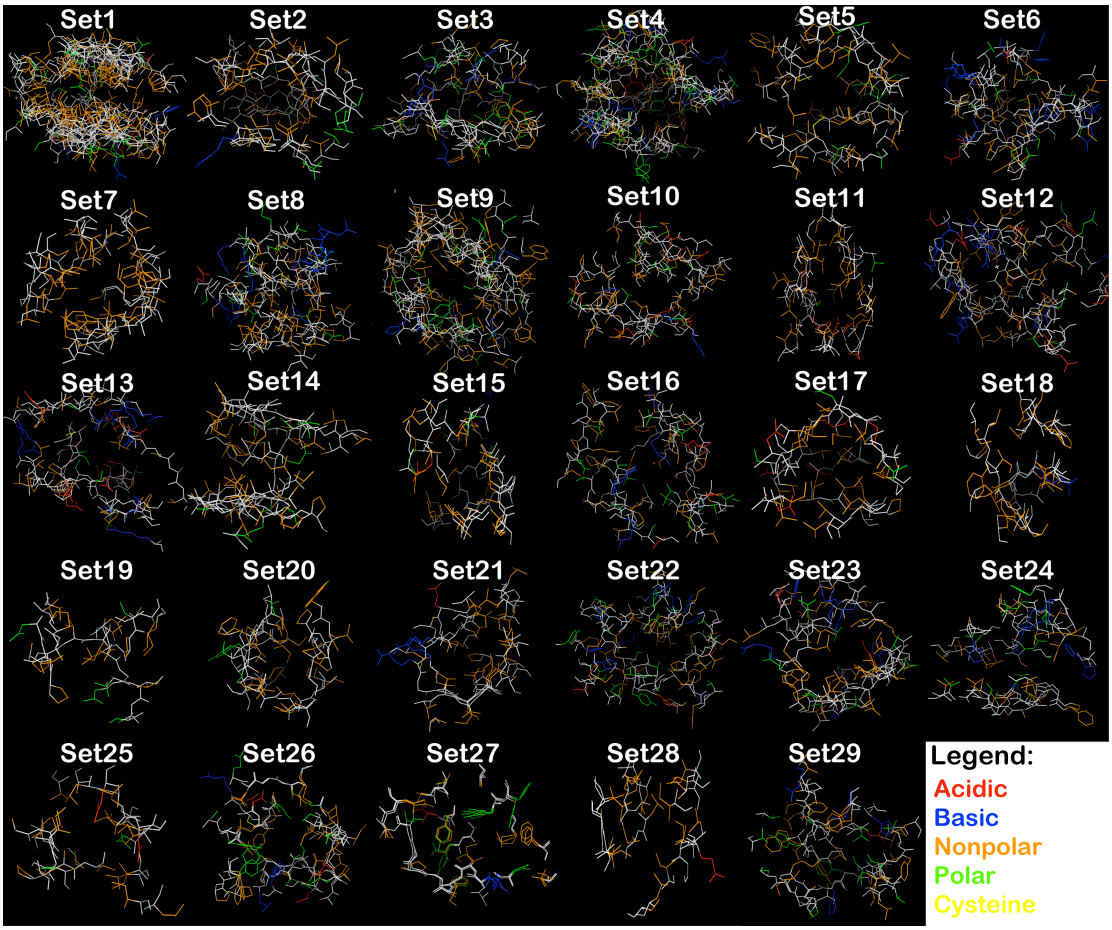


Figure S2. Superpositions of binding site sets. This figure depicts the superposition of similar sites present within a set. The residues are colored according to their chemical properties (provided as legend – right bottom corner). The backbone trace atoms are colored in white. Site superpositions were obtained using PocketAlign. The site having highest degree within the set was considered as the reference and all other sites were superposed onto it.

Table S5

Protein	Pocket	No. of Cofactor hits	No. of Drug hits	Clustering Coefficient (PMAX≥0.6)	PPI Index	Description	TDR Druggability Index	Normalized Degree
Rv0980c	Rv0980c.1_r1	160	215	1	0.65	PE-PGRS family protein PE_PGRS18	0.3	0.00763359
Rv0242c	Rv0242c.1_r1	112	266	0.64	0.51	3-ketoacyl-acyl carrier protein reductase	0.4	**0.633588**
Rv0080	Rv0080_r1	0	224	0.75	0.5	Conserved Hypothetical protein	Present in TDR	0.0553435
Rv0917	Rv0917_r1	542	324	0.52	0.5	Possible glycine betaine transport integral membrane protein BetP	NA	0.0324427
Rv2223c	Rv2223c_r1	118	239	0.65	0.47	Probable exported protease	NA	**0.370229**
Rv2688c	Rv2688c_r1	146	153	1	0.46	Antibiotic-transport ATP-binding protein ABC transporter	0.4	0.00763359
Rv0744c	Rv0774c_r1	34	352	0.4	0.42	Possible transcriptional regulatory protein	Present in TDR	0.0534351
Rv2321c	Rv2321c_r1	3	223	0.63	0.42	Probable ornithine aminotransferase (C-terminus part) RocD2	0.1	0.0324427
Rv1059	Rv1059_r1	263	244	0.57	0.41	Conserved Hypothetical protein	NA	**0.496183**
Rv1359	Rv1359_r7	14	141	0.97	0.41	Probable transcriptional regulatory protein	NA	0.0648855
Rv0242c	Rv0242c_r1	312	213	0.64	0.41	3-ketoacyl-acyl carrier protein reductase	0.4	**0.641221**
Rv3825c	Rv3825c_r1	430	251	0.54	0.41	Polyketide synthase Pks2	0.7	**0.503817**
Rv2836c	Rv2836c.2_r2	49	190	0.7	0.4	Possible DNA-damage-inducible protein F	0.1	**0.301527**
Rv0931c	Rv0931c_r2	0	139	0.95	0.4	Transmembrane serine/threonine-protein kinase D PknD	0.7	0.028626
Rv1902c	Rv1902c_r1	159	213	0.6	0.38	Probable sialic acid-transport integral membrane protein NanT	0.1	**0.53626**
Rv1475c	Rv1475c_r4	13	182	0.69	0.38	Probable iron-regulated aconitate hydratase Acn	0.2	0.0209924
Rv1266c	Rv1266c_r1	6	171	0.73	0.38	Probable transmembrane serine/threonine-protein kinase H PknH	0.7	0.0687023
Rv2766c	Rv2766c_r1	0	213	0.58	0.37	Probable short-chain type dehydrogenase/reductase	0.5	**0.704198**
Rv3085	Rv3085_r1	413	285	0.44	0.37	Probable short-chain type dehydrogenase/reductase	0.4	0.0763359

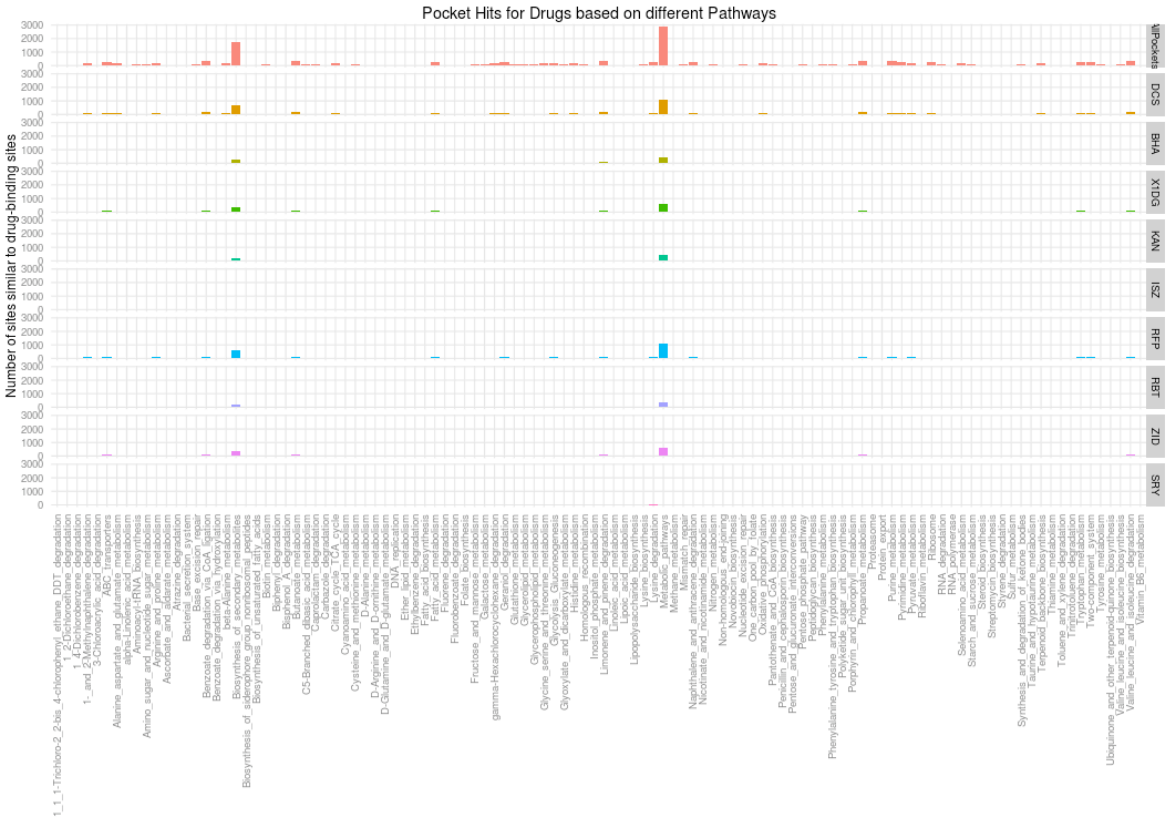
Additional PP targets found from Pocketome. This list of top 20 high scoring pockets along with the information on number of cofactor hits, number of drug hits, clustering coefficient, and TDR druggability score for each of them. Normalized degree is also reported for each node and the high degree values have been highlighted in bold.

Table S6

Pocket	Isoniazid Adduct Binding Site	PMAX Score
2763c_1df7_A_r2	2nq8_ZID:B:450	0.661247
1059_r10	2idz_ZID:A:300	0.649248
0753c_r18	2ieb_ZID:A:300	0.601528
2858c_r1	2nq8_ZID:A:550	0.591792
1484_r1	2ie0_ZID:A:300	0.580556
2623_r1	2ie0_ZID:A:300	0.56821
1996_r1	1zid_ZID:A:300	0.542005
0155_r2	2idz_ZID:A:300	0.538327
2971_r8	2nq8_ZID:A:550	0.531915
2671.1_r3	2nq8_ZID:A:550	0.508963

Isoniazid Secondary Hits. These are the isoniazid secondary target hits picked up with high PMAX score. The first column represents the Mtb protein (Rv No.) separated by '_' and pocket identifier, whereas the second column represents the isoniazid adduct binding site obtained from PDB.

Figure S3



Known tubercular drug binding site hits in Mtb Pocketome. The topmost panned represents the distribution of all the pockets of pocketome across known 'Tuberculist' pathways. Each of the subsequent panels below correspond to the binding site similarity hits obtained to pockets from the pathway of the respective anti-tubercular drugs.

Table S7

Protein	Pocket	Drug binding site	Tuberculist Pathway	PMIN	PMAX
Rv0013	Rv0013_r2	1pbc_BHA:A:396	Folate_biosynthesis	0.567677	0.519889
Rv2964	Rv2964_r17	1pbc_BHA:A:396	One_carbon_pool_by_folate	0.489429	0.467677
Rv1207	Rv1207_r9	1pbc_BHA:A:396	Folate_biosynthesis	0.484144	0.462626
Rv2124c	Rv2124c_r1	1pbc_BHA:A:396	One_carbon_pool_by_folate	0.503876	0.459596
Rv0812	Rv0812_r27	1pbf_BHA:A:396	Folate_biosynthesis	0.542857	0.459459
Rv1093	Rv1093_r21	1pbf_BHA:A:396	One_carbon_pool_by_folate	0.496444	0.447436
Rv3608c_1eye_A	Rv3608c_1eye_A_r2	1pbc_BHA:A:396	Folate_biosynthesis	0.482835	0.440404
Rv3608c_1eye_A	Rv3608c_1eye_A_r2	1pbf_BHA:A:396	Folate_biosynthesis	0.564723	0.439646
Rv1005c	Rv1005c_r20	1pbc_BHA:A:396	Folate_biosynthesis	0.448485	0.428986
Rv2763c	Rv2763c_r2	1pbf_BHA:A:396	One_carbon_pool_by_folate	0.534759	0.426743
Rv2763c	Rv2763c_r2	1pbf_BHA:A:396	Folate_biosynthesis	0.534759	0.426743
Rv2211c	Rv2211c_r4	1pbc_BHA:A:396	One_carbon_pool_by_folate	0.486643	0.423232
Rv0956	Rv0956_r3	1pbf_BHA:A:396	One_carbon_pool_by_folate	0.558712	0.41963
Rv3608c_1eye_A	Rv3608c_1eye_A_r15	1pbf_BHA:A:396	Folate_biosynthesis	0.442943	0.41963
Rv2211c	Rv2211c_r3	1pbc_BHA:A:396	One_carbon_pool_by_folate	0.622222	0.414815
Rv0957	Rv0957_r7	1pbc_BHA:A:396	One_carbon_pool_by_folate	0.621212	0.414141
Rv1093	Rv1093_r31	1pbf_BHA:A:396	One_carbon_pool_by_folate	0.41394	0.41394
Rv3608c_1eye_A	Rv3608c_1eye_A_r1	1pbf_BHA:A:396	Folate_biosynthesis	0.45377	0.408974
Rv2124c	Rv2124c_r1	1pbf_BHA:A:396	One_carbon_pool_by_folate	0.522048	0.406423
Rv0013	Rv0013_r2	1pbf_BHA:A:396	Folate_biosynthesis	0.624467	0.406105
Rv1093	Rv1093_r21	1pbc_BHA:A:396	One_carbon_pool_by_folate	0.515385	0.406061
Rv1207	Rv1207_r9	1pbf_BHA:A:396	Folate_biosynthesis	0.543386	0.403805
Rv2447c	Rv2447c_r8	1pbc_BHA:A:396	Folate_biosynthesis	0.537374	0.401207
Rv2964	Rv2964_r17	1pbf_BHA:A:396	One_carbon_pool_by_folate	0.539118	0.400634

PAS secondary hits. All the hits in the folate metabolism pathwasy (PMAX ≥0.4) obtained for PAS (PDB HETATM code :BHA) binding site in the Mtb pocketome.

Supplementary Text 2

Kinnings *et.al.* had proposed TB-drugome to understand the interactions of Mtb proteins with the currently approved drugs by constructing a drug-target network through similarities at the binding site. In their study, SMAP was used to obtain the similarity between predicted pockets in Mtb structural proteome and the drug-binding sites of the approved drugs. A systematic comparison was carried out at two levels – firstly the prediction of pockets, followed by the drug association obtained through binding site similarity.

The pockets were extracted from the ehits docking poses obtained from the TBdrugome studies. All the 1097 predicted pockets from the MODBASE models reported from TBdrugome studies were compared to pockets predicted from our approach. Overlap of atleast one residue was considered for the pockets detected by both the methods. Out of 1097 cases, 662 pockets detected had an overlap of atleast one residue, thus covering 60% of the pockets reported already. PocketMatch was also run to check all the drug associations reported by TBdrugome study and an average PMAX score of 0.40 was obtained. Data can be accessed at http://proline.biochem.iisc.ernet.in/mtbpocketome/methods.phpunder 'Comparison to TB-Drugome' section.

Figure S4

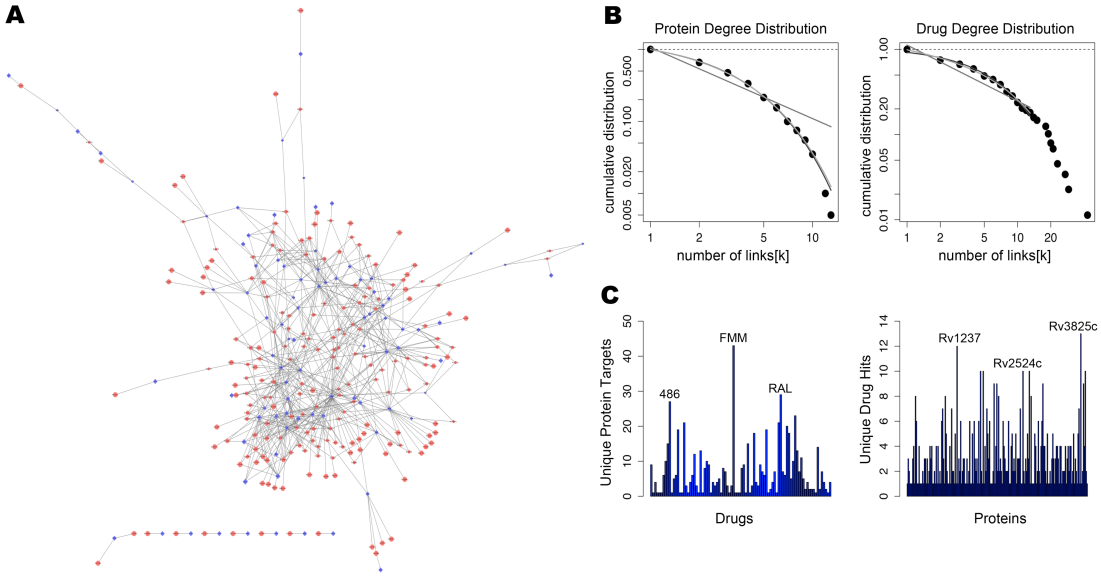


Figure S4. Bipartite network of high confident drug target and approved drugs. (A) Bipartite Network of high confident targets and approved drugs. The red nodes represent the high confidence targets obtained from targetTB and the blue nodes represent the approved drugs. The edge represents significant site similarity (Pvalue ~1e-04). (B) Cumulative Degree distribution of proteins and the drug in the bipartite network. (C) Drugs and proteins with highest number of connections in the bipartite network.