ELSEVIER

Using molecular-mimicry-inducing pathways of pathogens as novel drug targets

Anjali Garg[1], Bandana Kumari[1], Neelja Singhal[2] and Manish Kumar[1]

[1] Department of Biophysics, University of Delhi South Campus, New Delhi 110021, India
[2] Department of Microbiology, University of Delhi South Campus, New Delhi 110021, India

Several microbial pathogens cause autoimmune diseases in humans by exhibiting molecular mimicry with the host proteins. However, the contribution of autoimmunity in microbial pathogenesis has not been evaluated critically. Clinical and experimental observations have supported and corroborated that autoimmunity was a fundamental process underlying pathology of human tuberculosis bacteria. In the current review, we propose novel drug targets based on a pathogen's molecular-mimicry-inducing proteins. The process for identification of drug targets has been explained using *Mycobacterium tuberculosis* as a model organism. The procedure described here can be applied for repurposing other known drugs and/or discovery of novel therapeutics against other pathogenic bacteria that exhibit molecular mimicry with the host's proteins.

Introduction

When macromolecules found on pathogens and in host tissues share structural, functional or immunological similarities it is called molecular mimicry [1]. Molecular mimicry can occur in the form of complete identity or homology at the protein level, or as similarity in sequences of amino acids and structure. Sequence-based molecular mimicry plays an important part in immune response to infection and in autoimmune diseases. To attribute an autoimmune disease with molecular mimicry, certain criteria should be met: (i) there should be similarity between an epitope of the host, microorganism or environmental agent; (ii) antibodies or T cells cross-reactive with both epitopes must be detected in patients with an autoimmune disease; (iii) there should be evidence of an epidemiological link between exposure to a microbe or an environmental agent and development of autoimmune disease; and (iv) an autoimmune disease should be able to develop in an animal model when sensitized with the epitopes, exposed to the environmental agent or infected with the microbe [2].

Many pathogens exhibit molecular mimicry with the host proteins and cause autoimmune diseases. These pathogens have been listed in Table S1 (see supplementary material online). A detailed and explicit study on the role of molecular mimicry in microbial pathogenesis has not been conducted for most of the pathogens, except for a few fragmentary studies. For example, it was reported that group A *Streptococcus* and group B *Neisseria meningitides* use molecular mimicry to prevent induction of a pathogen-specific immune response [3]. Autoantibodies responsible for Wegener's granulomatosis and systemic lupus erythematosus have been observed in nearly half of the patients suffering from tuberculosis (TB) [4]. A few other autoimmune diseases such as inflammatory bowel disease, Behçet's disease, ankylosing spondylitis, Crohn's disease, ulcerative colitis and sarcoidosis have been associated with pathogenesis of *Mycobacterium tuberculosis* [5]. In an analysis conducted on differential gene expression among TB patients and patients with autoimmune or infectious diseases, it was found that combination of infection and autoimmune disease signatures could explain 96.7% of the differentially expressed TB signatures [6]. Autoimmunity has not been considered as a crucial process in pathology of TB. It continues to be an overlooked event with fragmentary studies [5].

Blocking the metabolic chokepoint has been used as a successful strategy for identifying new drug targets against a particular or-

Corresponding author: Kumar, M. (manish@south.du.ac.in)

https://doi.org/10.1016/j.drudis.2018.10.010

ganism [7,8]. In the present review, we describe how blocking the chokepoint involved in production of a pathogen's mimicry proteins and their interaction partners can be used for discovery of novel targets against pathogens. In this review, this approach has been explained using *M. tuberculosis* as the model organism. The initial step in this process involves identification of interaction partners of pathogen proteins (IPPP) involved in molecular mimicry with the host proteins. The homologs of the host protein, which might be present in IPPP, are removed, and chokepoints of the metabolic pathway are identified. Finally, drug candidates targeting the chokepoint proteins are selected from the DrugBank database and their efficiency and suitability is assessed.

Schema of drug repurposing

The procedure adopted for the process is explained using *Mycobacterium* spp. as the model organism. In the present manuscript, epitopes of the pathogen and host proteins involved in molecular mimicry are referred to as path-memitope and host-memitope, respectively. Similarly, proteins carrying path-memitope and host-memitope are referred to as path-protein and host-protein, respectively. The steps involved in the process are shown in Fig. 1 and described in detail below.

Data extraction

The experimentally verified events in autoimmune diseases caused by molecular mimicry were obtained from a database developed by us earlier: miPepBase [9]. In brief, miPepBase is an indigenously developed, manually curated database containing information about proteins and peptides that exhibit molecular mimicry and autoimmune diseases. A keyword search in miPepBase using 'mycobacterium' displayed 25 entries and/or events related to mimicry (Table 1). In the 25 events, 20 distinct *Mycobacterium* proteins involved in molecular mimicry were identified. These proteins were responsible for seven different types of autoimmune diseases caused by cross-reactivity with 12 different types of host proteins. We observed that one protein of the pathogen (A0A040DMG3) was removed by UniProt, hence it was excluded from our further studies. The seven different types of autoimmune diseases caused by the remaining 24 molecular mimicry events were encephalomyelitis; leprosy; multiple sclerosis; primary biliary cirrhosis; rheumatoid arthritis; skin disease and type 1 diabetes (Table 1). Also, not all of the 24 molecular mimicry events were caused by the proteins of *M. tuberculosis*. One event was caused by proteins of *Mycobacterium avium*; six were due to proteins of *M. avium* subsp. *paratuberculosis*; four caused by proteins of *Mycobacterium leprae*; one was due to proteins of *Mycobacterium gordonae*; 11 were due to proteins of *M. tuberculosis* and one was caused by proteins of *Mycobacterium bovis*.

Protein–protein interaction search

The IPPP were found using the database STRING [10]. STRING contains information about protein interactions, established by experimental studies and by genomic analysis like domain fusion, phylogenetic profiling and gene neighborhood. We included only those interactions that scored ≥0.4 (i.e., the default value). Using STRING, of the 19 path-proteins, we were able to find interacting partners for 16 proteins. Among the 16 path-proteins, one protein (P9WQ90) was a homo-dimer whereas two proteins (P0A521 and Q49375) were oligomers. For those path-proteins (A0A045I964, A0A0E2WUC4 and Q53467), about which protein-interaction information could not be retrieved using STRING [11], a BLAST search against the UniProtKB database was used to find homologous proteins. The first hits retrieved after the BLAST search of A0A045I964 and A0A0E2WUC4 were I6XH73 and F5Z390, respectively. However, for path-protein Q53467 we did not find any hits with high sequence homology (Table S2, see supplementary material online). Hence, it was removed from further analysis. I6XH73 and F5Z390 were also mycobacterial proteins. The STRING search revealed that, for the 15 path-proteins, there were 148 interacting protein partners. In the present work, if IPPP had alignment identity <50% with alignment coverage <80% with a human protein, they were considered as non-homologous IPPP (nHIPPP). As per this guideline among 148 IPPP, five proteins were homologous IPPP. Hence, these were also excluded from further analysis (Table 2).

Pathway mapping and determination of chokepoints in mycobacterial metabolic pathways

The 143 nHIPPP belonged to *M. leprae*, *M. avium* subsp. *paratuberculosis* and *M. tuberculosis*. Each nHIPPP was mapped in their corresponding metabolic networks in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [12]. KEGG is a database resource that cross-integrates genomic, chemical and systemic functional information of an organism. Because of this, KEGG is widely used as a reference knowledge base for integration and interpretation of large-scale datasets generated by genome sequencing and other high-throughput experimental technologies. The number of pathways to which these proteins were mapped are: 12 for *M. leprae*, 14 for *M. avium* subsp. *paratuberculosis* and 18 for *M. tuberculosis*. The pathways were analyzed manually to find possible chokepoint reaction(s). Our analysis revealed that these 143 proteins were a part of 53 chokepoints.

Authentication of chokepoint targets and druggability of selected targets

The validation of essentiality of chokepoint proteins in mycobacterial metabolic pathways was done in two ways. Homologs of chokepoint proteins were searched in all known mycobacterial proteomes and a total of 45 mycobacterial reference proteomes were present in UniProtKB (in October 2017). If a chokepoint protein showed ≥50% identity over 80% of sequence length in a minimum of ten mycobacterial proteomes, it was considered as a part of the core proteome (Table S3, see supplementary material online). We found that 47 of the 53 chokepoint proteins were part of core proteins (Table 3). Alternatively, a chokepoint protein was considered as an essential protein if it had an alignment identity ≥50% with a protein contained in Database of Essential Genes (DEG), over 80% of sequence length. We also found that 31 of the 53 chokepoint proteins shared a close homolog with DEG proteins (Table 3). Proteins that could not qualify either criterion were removed from further analysis.

The potential drugs that can block the IPPP were searched using DrugBank, the most widely used database of drug molecules [13]. Currently, DrugBank contains ~8200 different categories of drugs, namely FDA-approved small-molecule drugs, FDA-approved bio-

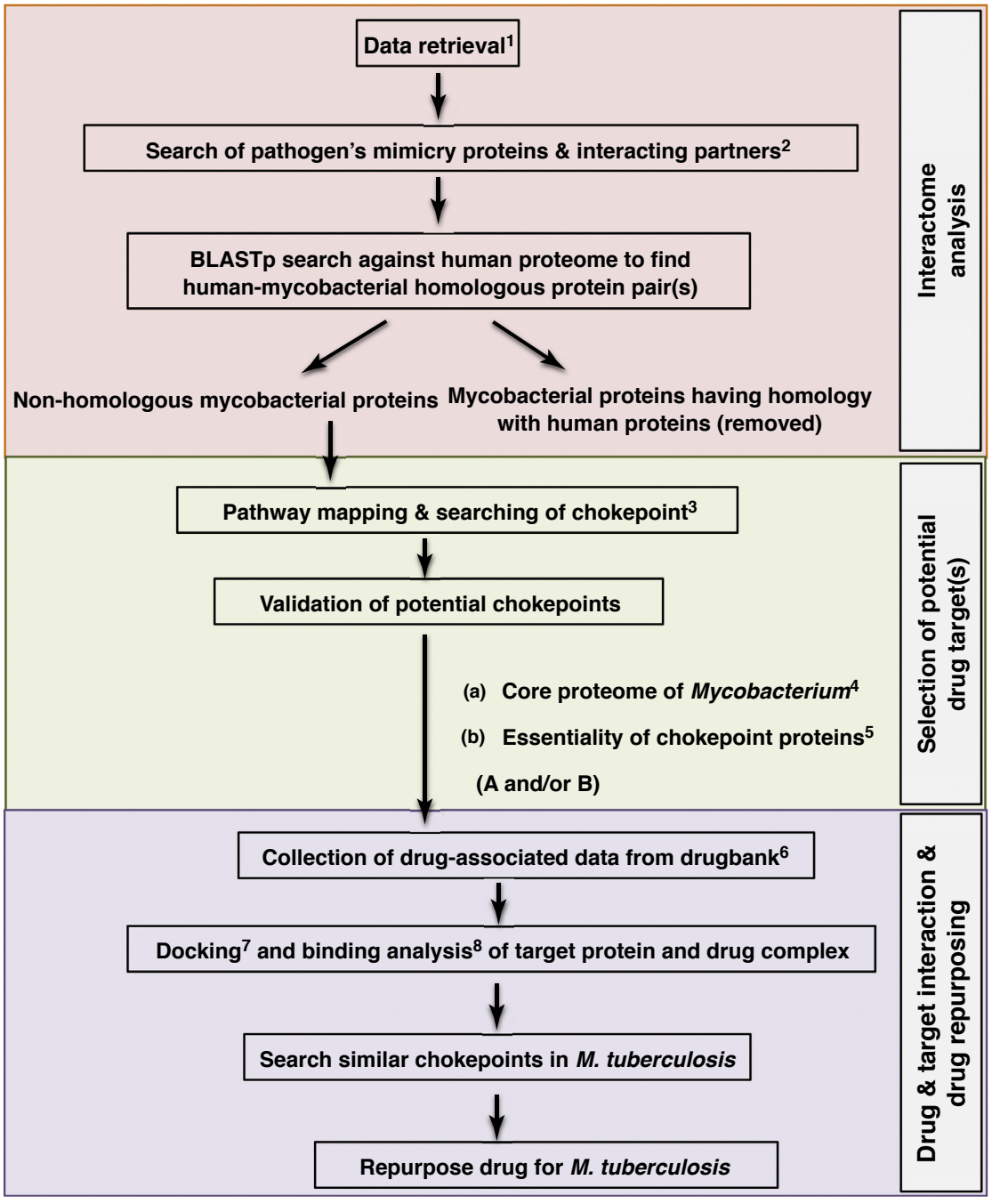

FIGURE 1

The scheme of drug repurposing proposed for *Mycobacterium tuberculosis*. In the figure we show the complete process which is clustered into three major sections. First, interactome analysis includes protein data retrieval, collection of interacting proteins and removal of path-proteins that are homologous to human protein(s). Second, selection of potential drug target(s) that include mapping of mycobacterial nHIPPP on mycobacterial metabolic pathway(s) and search of possible chokepoint protein(s). Chokepoint proteins pass through filters namely part of core proteome (A) or essential proteins (B). All chokepoint protein that crosses either filter is moved to the third step. Third, drug and target interaction and drug repurposing chokepoint proteins were searched for effective ligand(s) and their interaction was analyzed after docking. In the last step *M. tuberculosis* homolog was searched for each chokepoint protein. Superscript numbers reference a list of databases and servers used during the whole process: 1, miPepBase; 2, STRING; 3, KEGG; 4, UniProtKB; 5, DEG; 6, DrugBank; 7, PatchDock; 8, LigPlot+ v.1.4.

tech drugs, nutraceuticals and experimental drugs. To find the appropriate drug candidate, we downloaded sequences of all four types of targets: drug targets, drug enzymes, drug carriers and drug transporters, from DrugBank. Using BLAST we searched for homologs of chokepoint proteins among DrugBank target proteins. The drug molecule associated with the best hit of the DrugBank target protein was considered as a potential binder of homologous chokepoint proteins. Here too, a hit was considered as a homolo-

TABLE 1

List of *Mycobacterium spp.* proteins involved in molecular mimicry

No.	Pathogen protein entry (UniProt AC)	Mimicry peptide	Pathogen protein name	Pathogen name	Host name	Host protein entry (UniProt AC)	Host protein name	Host mimicry peptide	Autoimmune disease
1.	A0A040DMG3[a]	ACFTRPARWTL	Transmembrane protein	*M. tuberculosis*	Mouse	F6RT34	Myelin basic protein MBPAc	ASQKRPSQRSK	Encephalomyelitis
2.	A0A045I964	QRCRVHFMRNLYTAV	Transposase	*M. tuberculosis*	Human	P02686	Myelin basic protein	ENPVVHFFKNIVTPR	Multiple sclerosis
3.	A0A0E2WUC4	QRCRVHFLRNVLAQV	Transposase	*M. avium*	Human	P02686	Myelin basic protein	ENPVVHFFKNIVTPR	Multiple sclerosis
4.	A5U2C2	AAQHRQIVADF	UvrABC system protein C	*M. tuberculosis*	Mouse	F6RT34	Myelin basic protein MBPAc	ASQKRPSQRSK	Encephalomyelitis
5.	A5U956	AAQARPVKTVI	MYCTX transferase	*M. tuberculosis*	Mouse	F6RT34	Myelin basic protein MBPAc	ASQKRPSQRSK	Encephalomyelitis
6.	O32984	VSPWGKPEGRTRKPNKSSNK	50S ribosomal L2	*M. leprae*	Mouse	P02687	Myelin basic protein	VVHFFKNIVTPRTPPPSQGK	Leprosy
7.	O32984	EQANINWGKAGRMRWKGKRP	50S ribosomal L2	*M. leprae*	Mouse	P02687	Myelin basic protein	GAPKRGSGKDGHHAARTTHY	Leprosy
8.	P09239	NA	65 kDa heat shock protein	*M. leprae*	Human	P13645	Cytokeratin-10 of keratin	NA	Leprosy
9.	P0A521	AGKPLLIIAEDVEGE	HSP65	*M. bovis*	Human	P10809	HSP60	HRKPLVIIAEDVDGE	Rheumatoid arthritis
10.	P46861	NTLSAPTFVKDFPVETTPLT	Lysyl-tRNA synthetase	*M. leprae*	Mouse	P02687	Myelin basic protein	VVHFFKNIVTPRTPPPSQGK	Leprosy
11.	P9WG07	AYYGALPLIV	ABC transport	*M. tuberculosis*	Rabbit	P25274	Mid-region encephalitogen from myelin basic protein	TTHYGSLPQK	Multiple sclerosis
12.	P9WM57	ATQYRPDQLAK	Uncharacterized protein R	*M. tuberculosis*	Mouse	F6RT34	Myelin basic protein MBPAc	ASQKRPSQRSK	Encephalomyelitis
13.	P9WN15	ASMNRPNLVAL	Uncharacterized glycosyl hydrolase	*M. tuberculosis*	Mouse	F6RT34	Myelin basic protein MBPAc	ASQKRPSQRSK	Encephalomyelitis
14.	P9WPE7	STVKDLLPLL	65 kDa heat shock protein	*M. tuberculosis*	Rat	P02788	Human lactoferrin	SGQKDLLFKD	Rheumatoid arthritis
15.	P9WPE7	STVKDLLPLL	65 kDa heat shock protein	*M. tuberculosis*	Rat	P02787	Human transferrin	PHGKDLLFKD	Rheumatoid arthritis
16.	P9WPE7	VPGGGDMGG	65 kDa heat shock protein	*M. tuberculosis*	Human	P12035	Human keratin	GGYGGGMGG	Skin diseases
17.	P9WPE7	VPGGGDMGG	65 kDa heat shock protein	*M. tuberculosis*	Human	P10809	Human hsp65	GGMGGGMGG	Skin diseases
18.	P9WQ90	ASHQRQRAFAQ	Probable aspartate aminotransferase	*M. tuberculosis*	Mouse	F6RT34	Myelin basic protein MBPAc	ASQKRPSQRSK	Encephalomyelitis
19.	Q49375	GDL(IL)AE	65 kDa heat shock protein	*M. gordonae*	Human	P10515	Pyruvate dehydrogenase complex-E2	GDLIAE	Primary biliary cirrhosis
20.	Q53467	SHQIRPVCGQR	Putative transport protein	*M. avium* subsp. *paratuberculosis*	Mouse	F6RT34	Myelin basic protein MBPAc	ASQKRPSQRSK	Encephalomyelitis

TABLE 1 (*Continued*)

No.	Pathogen protein entry (UniProt AC)	Mimicry peptide	Pathogen protein name	Pathogen name	Host name	Host protein entry (UniProt AC)	Host protein name	Host mimicry peptide	Autoimmune disease
21.	Q73T54	MIAVALAGL	Uncharacterized protein	*M. avium* subsp. *paratuberculosis*	Human	Q8IWU4	Beta cell protein zinc transporter 8 (ZnT8)	MIIVSSCAV	Type 1 diabetes
22.	Q73T54	LAANFVVAL	Uncharacterized protein	*M. avium* subsp. *paratuberculosis*	Human	Q8IWU4	Beta cell protein zinc transporter 8 (ZnT8)	VAANIVLTV	Type 1 diabetes
23.	Q73WP1	WYIPPLSPVV	MAP_2619	*M. avium* subsp. *paratuberculosis*	Human	Q16653	Human myelin oligodendrocyte glycoprotein	MEVGWYRPPFSRVVHLYRNGK	Multiple sclerosis
24.	Q741P6	LKYGSLPLSF	SecD	*M. avium* subsp. *paratuberculosis*	Rabbit	P25274	Mid-region encephalitogen from myelin basic protein	TTHYGSLPQK	Multiple sclerosis
25.	Q745A5	PGRRPFTRKELQ	Uncharacterized protein	*M. avium* subsp. *paratuberculosis*	Human	P02686	Myelin basic protein	ENPVVNFFKNIVTP	Multiple sclerosis

Data sourced, with permission, from [9].
[a] Shows obsolete UniProtKB entry.

gous protein if it showed ≥50% identity over 80% of the sequence length. Using all DrugBank target and chokepoint protein pairs, for the five potential chokepoint proteins, we were able to identify 11 drug candidates. Proteins against which we could find drugs were mostly interaction partners of mimicry proteins responsible for multiple sclerosis. In the next stage, these probable drugs were further optimized according to Lipinski's Rule of Five scales: molecular weight ≤500, number of rotatable bonds ≤10, H-bond donors ≤5, H-bond acceptors ≤10 and logP ≤5 (Table S4, see supplementary material online). Additionally, half-life ≥60 min and toxicity information were also considered while evaluating a drug molecule. Those drug molecules that possessed a minimum of five of the seven parameters were considered as probable drugs. Drug-like compounds categorized by DrugBank as dietary supplements, micronutrients or vitamins were excluded.

After benchmarking on the basis of Lipinski's Rule of Five along with toxicity and half-life of drug molecules, we were finally left with four probable drug candidates. Of these four probable drug candidates, we noted that three were experimental approved drugs: DB08185, DB00759 and DB01930 against three chokepoint proteins of *M. leprae* rpsS, rpsC and panC respectively (Table 3). The fourth drug candidate was an experimentally verified drug: DB07349 against narH of *M. avium* subsp. *paratuberculosis*.

Of the four drugs, DB01930 is known to target the enzyme pantothenate synthetase of *M. tuberculosis* (https://www.drugbank.ca/drugs/DB01930). Pantothenate synthetase catalyzes the ATP-dependent condensation of pantoate and β-alanine to form pantothenate (vitamin B5) [14]. It is a known fact that pantothenate biosynthesis is essential for virulence of *M. tuberculosis* [15]. DB07349 is an experimental drug that targets narH and L. Several *in vivo* studies indicate that human lung granuloma, where *M. tuberculosis* resides during latency, is hypoxic and narH and L play an important part in bacterial survival in the hypoxic environment. This suggests that DB07349 can be an ideal drug candidate because it can kill *M. tuberculosis* residing in granuloma. Also, DB07349 can help in complete clearance of *M. tuberculosis* from the host, because long-term persistence of *M. tuberculosis* in the latent stage not only helps it in remaining unaffected during the antitubercular treatment but also helps the pathogen to develop resistance against currently used drugs [16]. The targets of the other two drugs (DB00759 and DB08185) are parts of the ribosomal protein complex. DB00759 (commonly known as tetracycline) is already an approved drug. It is being given to patients orally as well as by an ophthalmic ointment. These are also reported to inhibit the *M. tuberculosis* pathogen growth by binding to the 30S ribosomal subunit and blocking translation [17].

Drug–target interaction

Molecular docking is a useful tool for modeling the interaction between two biomolecules or a small molecule (could be a drug-like molecule) and a biomolecule at the atomic level. It allows us to model the behavior of binding partners in terms of binding affinity or interaction. To assess the binding potential of selected drug candidates with their target, PatchDock was used for docking drug molecules that passed the filtration criteria with their potential targets: gadB, rpsC, rpsS, panC and narH. PatchDock provides a list of receptor and ligand molecule complexes and their PatchDock

TABLE 2

List of pathogen mimicry protein, its interaction partners (IPPP), name of human homolog (if present), KEGG pathway ID to which IPPP belongs and chokepoint proteins

No.	Path-proteins	Interacting proteins of pathogen proteins (IPPP)	Human homolog of IPPP	Proteins could not be mapped on KEGG	KEGG pathway ID	Chokepoint proteins
1	A5U2C2	MRA_1028, MRA_1424, MRA_1430, MRA_1431, MRA_1432, MRA_1648, MtubH3_010100010416, mfd, uvrA, uvrB, uvrC	NA	MRA_1028, MRA_1424, MRA_1430, MRA_1431, MRA_1432, MRA_1648, MtubH3_010100010416	mtu03420	Mfd, uvrA, uvrB, uvrC
2	A5U956	MRA_0226, MRA_0381, MRA_1886, MRA_3014, MRA_3766, MRA_3767, MRA_3895, ethR	NA	NA	NA	NA
3	F5Z390	JDM601_3772, JDM601_3773, JDM601_3774	NA	NA	NA	NA
4	I6XH73	Rv3431c, gadB, nnr, Rv3434c, Rv3435c	NA	Rv3431c, nnr, Rv3434c, Rv3435c	mtu00410, mtu01100, mtu00650, mtu01120, mtu00250, mtu00430, mtu02024, mtu01110	mtu02024:gadB
5	O32984	rplB, rplC, rplD, rplF, rplN, rplP, rplV, rplW, rpmC, rpsC, rpsS	NA	NA	mle03010	rplB, rplC, rplD, rplF, rplN, rplP, rplV, rplW, rpmC, rpsC, rpsS
7	P09239	clpB, dnaJ1, dnaJ2, dnaK, groL2, groS, grpE, hrcA, htpG, mdh, pheT	dnaK, mdh	clpB, dnaJ1, dnaJ2 , groS, grpE, hrcA, htpG	mle03018, mle05152, mle00970	mle03018:groEL; mle05152:groEL; mle00970:pheT
6	P0A521	NA	NA	NA	NA	NA
8	P46861	argS, gltX, guaA, ileS, leuS, lysS, lysX (ML1393), metG, panC, pheT, proS	NA	NA	mle00970, mle01100, mle01110, mle00770, mle00230, mle00860, mle00450, mle00410, mle01120	mle00970:gltX, metG, leuS, ileS, lysS, argS, proS, pheT; mle00860:gltX; mle00410:panC
9	P9WG07	phoU1 (Rv3301c), phoU2 (Rv0821c), pstA1, pstA2, pstB1 or phoT (Rv0820), pstB2 or pstB (Rv0933), pstC1, pstS1, pstS2, pstS3, tcrX (Rv3765c)	NA	phoU1 (Rv3301c), phoU2 (Rv0821c), tcrX (Rv3765c)	mtu02010, mtu02020, mtu05152	mtu02010: pstB1, pstB2, pstA1, pstA2, pstC1, pstS1, pstS2, pstS3; mtu02020:pstS1, pstS2, pstS3; mtu05152:pstS1, pstS2, pstS3
10	P9WM57	Rv0184, Rv0336, Rv0515, Rv1128c, Rv1278, Rv1378c, Rv1765c, Rv2015c, Rv2100, Rv3074, Rv3776	NA	NA	NA	NA
11	P9WN15	Rv2006, Rv3400, Rv3401, aglA, glgB, glgE, glgX (Rv1564c), otsA, otsB (Rv3372), treS, treZ	NA	Rv3400, Rv3401	mtu01100, mtu00500, mtu01110, mtu00052	mtu00500:glgB; mtu00052:aglA
12	P9WPE7	Rv0312, Rv2264c, dnaJ1, dnaJ2, dnaK, groL2 (Rv0440), groS, hycE, metK, pheT, thrS	dnaK, metK	Rv0312, Rv2264c, dnaJ1, dnaJ2, groS, hycE	mtu00970, mtu05152, mtu03018	mtu00970:thrS, pheT; mtu05152:groL2; mtu03018:groL2
13	P9WQ90	NA	NA	NA	NA	NA
14	Q49375	NA	NA	NA	NA	NA
15	Q53467	NA	NA	NA	NA	NA
16	Q73T54	MAP_2073c, MAP_2138, MAP_2784, MAP_2925, MAP_3865c, MAP_3866c, MAP_3867c, atpA, ctpA, ctpC, nrdE	atpA	MAP_2073MAP_2138, MAP_2784, MAP_2925, MAP_3865c, MAP_3866c, MAP_3867c, ctpA, ctpC	mpa00230, mpa00240, mpa00190	NA

TABLE 2 (*Continued*)

No.	Path-proteins	Interacting proteins of pathogen proteins (IPPP)	Human homolog of IPPP	Proteins could not be mapped on KEGG	KEGG pathway ID	Chokepoint proteins
17	Q73WP1	MAP_0368, MAP_2102c (narK3_1), MAP_3636, MAP_3707c (narK3_2), MAP_4101c, fdhF, narG, narH, narI, narJ, narU	NA	MAP_3636, MAP_4101c, fdhF	mpa00910, mpa01120, mpa02020, mpa01100, mpa00630, mpa00680, mpa01200	mpa00910: narK3_1, narK3_2, narU, narH, narG
18	Q741P6	MAP_1042, MAP_1045, apt, dnaG, relA, secD, secE, secF, secG, secY, tatC	NA	NA	mpa03070, mpa03060, mpa02024, mpa00230, mpa01100, mpa03030	mpa03070: MAP_1042, secD, secE, secF, secG, secY; mpa03060:MAP_1042, secD, secE, secF, secG, secY; mpa02024: MAP_1045, MAP_1042, secE, secG, secY; mpa03030:dnaG
19	Q745A5	MAP_0105c, MAP_0106c, MAP_1410, MAP_2148, MAP_2752, MAP_2963c, MAP_3314c, ftsK, ogt, parB, topA	NA	NA	NA	NA

The table shows information related to pathogen protein involved in molecular mimicry (column 2), IPPP collected from STRING database at default parameters (column 3), HIPPP among the IPPP (column 4), IPPP which could not be mapped on KEGG (column 5), KEGG pathway ID in which IPPP mapped (column 6) and chokepoint proteins found after manual survey of KEGG pathway IDs listed in column 6 (column 7).

scores. The protein–ligand complex with the highest docked score was selected for further analysis. The structures of four potential drugs: DB08185, DB00759, DB01930 and DB07349, were downloaded from DrugBank. We observed that, among all chokepoint proteins, 3D structure of only panC was available in PDB. Hence, the 3D structures of the remaining proteins were obtained from Swiss-model (rpsS and rpsC) and modBase (narH). The intermolecular interactions and strengths, H-bonding, hydrophobic interactions and atom accessibilities are shown in Table S5 (see supplementary material online).

Drug repurposing for *M. tuberculosis*

Molecular mimicry plays an important part in primary establishment of *M. tuberculosis* inside the host. Hence, if *M. tuberculosis* mimicry-inducing proteins can be blocked, the pathogen can be eliminated, well-before it establishes itself inside the host. The steps described above can also be used to propose novel drugs against *M. tuberculosis*. As explained earlier, the 53 chokepoint proteins identified belong to three different species of mycobacteria. Hence, their homologs were searched in the proteome of *M. tuberculosis*. We observed that, of the 53 chokepoints, homologous proteins for 47 chokepoint proteins (14 of *M. tuberculosis*, 20 of *M. leprae* and 13 of *M. avium* subsp. *paratuberculosis*) were present in the proteome of *M. tuberculosis* (Table 3). Hence, we anticipate that these four drugs (DB08185, DB00759, DB01930 and DB07349) might be useful in the treatment of *M. tuberculosis*.

Prospects for the current approach

A lot of research has been done to discover novel drug targets and potent drugs against TB [18–24]. The current approach is different from earlier approaches, because our target here is not an active physiological process or protein(s), which helps in establishing TB bacteria inside the host. Our target is a protein(s) (and interacting partners) that is responsible for eliciting autoimmunity inside the host. Here, the authors propose to target and/or disrupt proteins of *M. tuberculosis* that evoke autoimmune diseases (using drugs or chemical compounds) as a prophylactic measure, before the onset of active TB infection. It would be pertinent to mention here that recent research proposed that mycobacterial infections might have driven autoimmunity as an evolutionary strategy and proteins involved in molecular mimicry are produced in the host long-before the appearance of the symptoms of TB [5]. Thus, our approach might be useful in devising novel prophylactic or vaccination measures against TB.

Another prospective use for our approach is that it can be used as a follow-up remedy after a patient is cured from TB. The drug molecules identified in our current study would disrupt the growth of latent bacteria residing inside the host, which will ultimately lead to clearance of TB bacilli from the host. The other advantage of our approach is that it is in-line with the therapy used for treatment of autoimmune diseases. Tumor necrosis factor (TNF)-blocker therapy is an effective treatment for many autoimmune diseases but it also significantly increases the risk of progression of latent TB to active TB. Thus, before commencing the TNF-blocker therapy for curing autoimmune diseases, patients are first tested for TB infection. Hence, use of a drug that does not involve the use of a TNF-blocker can lead to significant improvement in treatment of pathogen-induced autoimmunity.

TABLE 3

Drug target validation

Path-protein	Chokepoint proteins	Chokepoint proteins that were part of essential gene database	Chokepoint proteins that were part of core proteome	Homolog of chokepoint proteins in *M. tuberculosis* proteome	Chokepoint proteins included as drug-target in DrugBank	Potential drug molecule as per DrugBank against target proteins	Drugs follow at least 5 of 7 drug-like properties
A5U2C2	mfd, uvrA, uvrB, uvrC	uvrC	mfd, uvrA, uvrB, uvrC	mfd, uvrA, uvrB, uvrC	NA	NA	–
I6XH73	gadB	NA	gadB	gadB	gadB	gadB: DB03553	–
O32984	rplB, rplC, rplD, rplF, rplN, rplP, rplV, rplW, rpmC, rpsC, rpsS	rplB, rplC, rplD, rplF, rplN, rplP, rplW, rpmC, rpsC, rpsS	rplB, rplC, rplD, rplF, rplN, rplP, rplW, rplV rpmC, rpsC, rpsS	rplB, rplC, rplD, rplF, rplN, rplP, rplW, rpmC, rpsC, rpsS	rpsC, rpsS	rpsC: DB00759; DB09093 rpsS: DB08185; DB00560; DB00759; DB09093	DB08185 (2-methylthio-N6-isopentenyl-adenosine-5′-monophosphate), DB00759 (tetracycline)
P09239	mle03018:groEL; mle05152: groEL; mle00970:pheT	groL2, pheT	groL2, pheT	groL2, pheT	NA	NA	–
P46861	mle00970:gltX, metG, leuS, ileS, lysS, argS, proS, pheT; mle00860:gltX; mle00410: panC	gltX, metG, leuS, ileS, lysS, argS, pheT, panC	gltX, metG, leuS, ileS, lysS, argS, pheT, panC	gltX, metG, leuS, ileS, lysS, argS, pheT, panC	panC	panC: DB01930; DB02596; DB02694; DB03107	DB01930 ((1S)-2-{[{[[(2S)-2,3-dihydroxypropyl]oxy}(hydroxy) phosphoryl]oxy}-1-[(pentanoyloxy)methyl]ethyl octanoate)
P9WG07	mtu02010: pstA1, pstA2, pstB1, pstB2, pstC1, pstS1, pstS2, pstS3; mtu02020: pstS1, pstS2, pstS3; mtu05152:pstS1, pstS2, pstS3	NA	pstA1, pstB1,pstS2, pstS3	pstA1, pstB1,pstS2, pstS3	NA	NA	–
P9WN15	mtu00500:glgB; mtu00052: aglA	glgB	glgB, aglA	glgB, aglA	NA	NA	–
P9WPE7	mtu00970:thrS, pheT; mtu05152:groL2; mtu03018: groL2	thrS, pheT, groL2	thrS, pheT, groL2	thrS, pheT, groL2	NA	NA	–
Q73WP1	mpa00910:narK3_1, narK3_2, narU, narH, narG	NA	narK3_1, narK3_2, narH, narG	narK3_1, narK3_2, narU, narH, narG	narH	narH: DB04464; DB07349	DB07349 (2,4-dihydroxy-3,3-dimethyl-butyrate)
Q741P6	mpa03070:MAP_1042, secD, secE, secF, secG, secY; mpa03060:MAP_1042, secD, secE, secF, secG, secY; mpa02024:MAP_1045, MAP_1042, secE, secG, secY; mpa03030:dnaG	secD, secE, secF, secG, secY, dnaG	MAP_1042, MAP_1045, dnaG, secD, secE, secF, secG, secY	MAP_1042, MAP_1045, dnaG, secD, secE, secF, secG, secY	NA	NA	–

The table shows information of path-proteins (column 1), potential chokepoint found in KEGG metabolic network (column 2), chokepoint proteins which were part of essential genes (column 3) and core proteins (column 4), homologous of chokepoint proteins in *M. tuberculosis* proteome (column 5) and chokepoint protein listed as drug target in DrugBank database (column 6). Column 7 has potential drug molecule as per DrugBank target protein and column 8 contains the drugs that qualified the filter of drug candidate filter.

Concluding remarks

Computational methods and integrated omics approaches, encompassing genomics, proteomics and metabolomics, have proved a valuable tool in drug discovery. Comparative and subtractive genomics proved helpful for prediction and identification of potential therapeutic targets and vaccine candidate proteins in numerous pathogenic bacteria and fungi [25–29]. In the current review, we have described a novel approach to discover new drug targets and drug molecules using a pathogen's molecular-mimicry-inducing proteins. The identification has been done by employing a rigorous systems biology approach. The process and the workflow for identification of drug targets have been explained in detail using *M. tuberculosis* as the model organism. Our systematic analysis revealed that interacting proteins of mimicry-inducing proteins of mycobacteria contain several chokepoint proteins, which can serve as potential drug targets. Inhibitors of the chokepoint proteins were searched from DrugBank employing several stringent filters. The DrugBank search revealed three drug compounds enlisted in the experimental group and one in the approved group, which might be effective against *M. tuberculosis*. Interaction between target(s) and their cognate drug molecule(s) was further confirmed by molecular docking. The drug candidates identified during the course of this study are FDA-approved drug molecules, with proven efficacy against many microbial pathogens. The proposed drug candidates might be tested *in vitro* for assessing their efficacy against *M. tuberculosis* clinical isolates. Thus, instead of developing new chemotherapeutics, our approach helps in repurposing the known drugs against TB.

Using the interaction partners of mimicry proteins, the authors were able to discover only four drug candidates against TB. The trivial number of drugs might be because only one database was used to search drug molecules: DrugBank. DrugBank was preferred over other databases because it provides detailed information about the properties and mechanisms-of-action of ~12000 marketed or experimental drugs. However, the number of probable drug candidates would have increased if data from other relevant databases were also included in the study. For example, databases such as ChEMBL [30], PubChem [31] and ChemBank [32] could be used to provide a comprehensive collection of biological activity, whereas ZINC database [33] could be used for virtual screening. Similarly, incorporation of additional data for example protein–chemical interactions from the Therapeutic Target Database [34] and STITCH [35] can also increase the number of drug targets and candidates. Nevertheless, the authors believe that the scheme described in the current review can be applied for repurposing the known drugs and discovery of novel therapeutics against other pathogenic bacteria that exhibit molecular mimicry with a host's proteins.

Acknowledgments

A.G. is thankful to ICMR for granted scholarship [3/1/3 J.R.F.-2016/LS/HRD-(32262)] and B.K. is thankful to ICMR for granted scholarship [Grant no. BIC/11(33)/2014].

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.drudis.2018.10.010.

References

1 White Cunningham, M. (2009) Molecular mimicry. *eLS* doi: https://doi.org/10.1002/9780470015902.a0000958.pub2
2 Peterson, L.K. and Fujinami, R.S. *et al.* (2007) Molecular mimicry. In *Autoantibodies* (2nd ed.) (Shoenfeld, Y., ed.), pp. 13–19, Elsevier
3 Nisini, R. (2015) Targeting immune-evasion mechanisms as a possible new approach in the fight against tuberculosis. *Immunol. Disord. Immunother.* 1, 101
4 Kakumanu, P. *et al.* (2008) Patients with pulmonary tuberculosis are frequently positive for anti-cyclic citrullinated peptide antibodies, but their sera also react with unmodified arginine-containing peptide. *Arthritis Rheum.* 58, 1576–1581
5 Elkington, P. *et al.* (2016) Tuberculosis: an infection-initiated autoimmune disease? *Trends Immunol.* 37, 815–818
6 Clayton, K. *et al.* (2017) Gene expression signatures in tuberculosis have greater overlap with autoimmune diseases than with infectious diseases. *Am. J. Respir. Crit. Care Med.* 196, 655–656
7 Rahman, S.A. and Schomburg, D. (2006) Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. *Bioinformatics* 22, 1767–1774
8 Gupta, M. *et al.* (2017) Identification of phosphoribosyl-AMP cyclohydrolase, as drug target and its inhibitors in *Brucella melitensis* bv. 1 16M using metabolic pathway analysis. *J. Biomol. Struct. Dyn.* 35, 287–299
9 Garg, A. *et al.* (2017) miPepBase: a database of experimentally verified peptides involved in molecular mimicry. *Front. Microbiol.* 8, 2053
10 Szklarczyk, D. *et al.* (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368
11 Apweiler, R. *et al.* (2004) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 32, D115–119
12 Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30
13 Wishart, D.S. *et al.* (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 36, D901–906
14 Jonczyk, R. *et al.* (2008) Pantothenate synthetase is essential but not limiting for pantothenate biosynthesis in *Arabidopsis*. *Plant Mol. Biol.* 66, 1–14
15 Hung, A.W. *et al.* (2016) Optimization of inhibitors of *Mycobacterium tuberculosis* pantothenate synthetase based on group efficiency analysis. *ChemMedChem* 11, 38–42
16 Gomez, J.E. and McKinney, J.D. (2004) *M. tuberculosis* persistence, latency, and drug tolerance. *Tuberculosis* 84, 29–44
17 Hentschel, J. *et al.* (2017) The complete structure of the *Mycobacterium smegmatis* 70S ribosome. *Cell Rep.* 20, 149–160
18 Lamichhane, G. (2011) Novel targets in *M. tuberculosis*: search for new drugs. *Trends Mol. Med.* 17, 25–33
19 Anishetty, S. *et al.* (2005) Potential drug targets in *Mycobacterium tuberculosis* through metabolic pathway analysis. *Comput. Biol. Chem.* 29, 368–378
20 Singh, V. and Mizrahi, V. (2017) Identification and validation of novel drug targets in *Mycobacterium tuberculosis*. *Drug Discov. Today* 22, 503–509
21 Duncan, K. (2004) Identification and validation of novel drug targets in tuberculosis. *Curr. Pharm. Des.* 10, 3185–3194
22 Ioerger, T.R. *et al.* (2013) Identification of new drug targets and resistance mechanisms in *Mycobacterium tuberculosis*. *PLoS One* 8, e75245
23 Mdluli, K. and Spigelman, M. (2006) Novel targets for tuberculosis drug discovery. *Curr. Opin. Pharmacol.* 6, 459–467
24 Raman, K. *et al.* (2008) targetTB: a target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. *BMC Syst. Biol.* 2, 109
25 Volker, C. and Brown, J.R. (2002) Bioinformatics and the discovery of novel antimicrobial targets. *Curr. Drug Targets Infect. Disord.* 2, 279–290
26 Johri, A.K. *et al.* (2006) Group B *Streptococcus*: global incidence and vaccine development. *Nat. Rev. Microbiol.* 4, 932–942
27 Doyle, M.A. *et al.* (2010) Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes. *BMC Genomics* 11, 222
28 Chong, C.E. *et al.* (2006) *In silico* analysis of *Burkholderia pseudomallei* genome sequence for potential drug targets. *In Silico Biol.* 6, 341–346

29 Amineni, U. *et al*. (2010) *In silico* identification of common putative drug targets in Leptospira interrogans. *J. Chem. Biol* 3, 165–173
30 Gaulton, A. *et al*. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. 40, D1100–1107
31 Kim, S. *et al*. (2016) PubChem substance and compound databases. *Nucleic Acids Res*. 44, D1202–D1213
32 Seiler, K.P. *et al*. (2008) ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res*. 36, D351–359
33 Irwin, J.J. and Shoichet, B.K. (2005) ZINC-a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model*. 45, 177–182
34 Li, Y.H. *et al*. (2018) Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res*. 46, D1121–1127
35 Szklarczyk, D. *et al*. (2016) STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res*. 44, D380–D384