Journal of Proteins and Proteomics (2021) 12:93–104
https://doi.org/10.1007/s42485-021-00065-y

SHORT COMMUNICATION

Discovering mycobacterial lectins as potential drug targets and vaccine candidates for tuberculosis treatment: a theoretical approach

Shobana Sundar[1] · Lokesh Thangamani[1] · Shanmughavel Piramanayagam[1] · Jeyakumar Natarajan[2]

Received: 26 February 2021 / Revised: 16 April 2021 / Accepted: 10 May 2021 / Published online: 18 May 2021

Abstract

M. tuberculosis proliferates within the macrophages during infection and they are bounded by carbohydrates in the cell wall, called lectins. Despite their surface localization, the studies on exact functions of lectins are unexplored. Hence, in our study, using insilico approaches, 11 potential lectins of Mtb was explored as potential drug targets and vaccine candidates. Initially, a gene interaction network was constructed for the 11 potential lectins and identified its functional partners. A gene ontology analysis was also performed for the 11 mycobacterial lectins along with its functional partners and found most of the proteins are present in the extracellular region of the bacterium and belongs to the PE/PPE family of proteins. Further, molecular docking studies were performed for two of the potential lectins (Rv2075c and Rv1917c). A novel series of quinoxalinone and fucoidan derivatives have been made to dock against these selected lectins. Molecular docking study reveals that quinoxalinone derivatives showed better affinity against Rv2075c, whereas fucoidan derivatives have good binding affinity against Rv1917c. Moreover, the mycobacterial lectins can interact with the host and they are considered as potential vaccine candidates. Hence, immunoinformatics study was carried out for all the 11 potential lectins. B-cell and T-cell binding epitopes were predicted using insilico tools. Further, an immunodominant epitope ^{1062}SIPAIPLSVEV^{1072} of Rv1917c was identified, which was predicted to bind B-cell and most of the MHC alleles. Thus, the study has explored that mycobacterial lectins could be potentially used as drug targets and vaccine candidates for tuberculosis treatment.

Keywords *Mycobacterium tuberculosis* · Mycobacterial lectins · Gene network analysis · Molecular docking · Immunoinformatics

Abbreviations

TB	Tuberculosis
Mtb	*Mycobacterium tuberculosis*
PDB	Protein Data Bank
SMM	Stabilized Matrix Base method
IDE	Immunodominant epitope
TLR2	Toll-like receptor 2

✉ Shobana Sundar
sundar.shobana@gmail.com

[1] Computational Biology Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore, Tamilnadu, India

[2] Data Mining and Text Mining Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore, Tamilnadu, India

Introduction

Tuberculosis (TB) is an infectious disease caused by the rod-shaped bacilli *Mycobacterium tuberculosis* (Mtb) and is affected millions of people worldwide. In accordance with WHO, in the year 2018, nearly 10 million people around the world had developed TB and 1.5 million people had died from the disease. Many new chemotherapeutic drugs and vaccinations are been used to treat TB; however, the emergence of drug resistance strains increases the treatment burden.

Mtb proteome has abundance of hypothetical or putative or uncharacterized proteins whose function remains unexplored. This also includes the unknown functions of the mycobacterial lectins. Lectins are glycan or sugar-binding proteins and they bind diverse range of sugar molecules. They are involved in many biological processes such as cell–cell interactions, host–pathogen interactions, etc., initially, lectins are majorly found in plants and animals

(Vijayan and Chandra 1999). In bacteria, lectins occur as fimbriae and pili which are filamentous protein appendages projecting from their bacterial surface. The most extensively studied bacterial lectins are the mannose-specific FimH of type 1 fimbriae and the galabiose-specific PapG of P fimbriae of *Escherichia coli*. FimH is associated with urinary tract infections (Chen et al. 2009) and PapG suggested having a role in renal colonization (Lane and Mobley 2007).

In 1989, Kundu et al. isolated mycotin (a lectin) from *Mycobacterium smegmatis* and it was found to agglutinate human erythrocytes (Kundu et al. 1989). Later in 2007, Singh et al., by comprehensive bioinformatics analysis, identified eleven potential lectins in the Mtb genome (Singh et al. 2007). In 2012, Abhinav et al. performed a bioinformatics based homology search of lectin-encoding gene regions in 30 fully or partially sequenced mycobacterial genomes and identified 94 potential glycan-binding proteins (Abhinav et al. 2013).Three potential lectins of the Mtb genome identified by Singh et al. (2007) were also found in this study. This study further adds to the importance of mycobacterial lectins in the pathogenesis of the organism. Therefore, in this study through insilico approaches, 11 mycobacterial lectins found by Singh et al. (2007) was explored as potential drug targets and vaccine candidates which can be used for treatment of TB.

Methods

Gene network analysis

Protein–protein interaction network for all the 11 putative mycobacterial lectins were retrieved using stringApp tool in Cytoscape 3.7.2 (Smoot et al. 2011). The confidence score cutoff was set at 0.4 and the maximum additional interactors was set to 100. The functional enrichment data for all the nodes was also retrieved using the functional enrichment tool in Cytoscape. Clustering analysis was performed by using the MCODE app in Cytoscape. Default parameters were used in the clustering analysis.

Ab initio homology modeling

Two mycobacterial lectins, Rv1917c and Rv2075c which has 1459 and 487 amino acids were retrieved from UniprotKB in FASTA format. The three-dimensional model was generated using the I-TASSER server (Zhang 2008) which generates a 3D model of query sequence by multiple threading alignments and iterative structural assembly simulation. The conformation of the best 3D model (model 1) for the two selected lectin proteins was selected for further studies and the homology models were further validated by the Ramachandran plot and ERRAT scores (Colovos and Yeates 1993). Pymol was used to visualize the modeled structure.

Molecular docking studies

AutoDock Vina program was used to for molecular docking studies (Trott and Olson 2010). 54 fucoidan derivatives were retrieved from PubChem database (Kim et al. 2016) and made to dock with the Rv1917c homology model whereas 233 quinoxalinone derivatives retrieved from PubChem database and were made to dock with the Rv2075c homology model. The retrieved ligand structures were converted into PDB file format in OBABEL command line. The grid box was laid for the whole protein and the grid spacing was fixed at 1 Å. The coordinates were set to $126 \times 126 \times 126$ for Rv1917c whereas for Rv2075c it was fixed at $62 \times 60 \times 64$, respectively. The ligand docking scores were noted and the interactions were obtained using Discovery studio visualizer.

Immunoinformatics studies

Sequence based B-cell prediction

ABCpred server (Saha and Raghava 2007) was used to predict the B-cell epitopes using the target protein sequence. The window length was fixed at 20 and the default threshold of 0.5 was used for the prediction. The retrieved epitopes were filtered using their score.

Cytotoxic T-lymphocyte prediction

IEDB tools were used for T-cell epitope predictions and predominantly occurring 10MHC(Major Histocompatibility Complex) alleles in human population for HLA class I (A*01:01, A*02:01, A*03:01, A*11:01, A*24:02, B*07:02, B*08:01, B*35:0, B*40, B*44) were considered for predictions (Agrewala and Wilkinson 1999; Chodisetti et al. 2012; Verma et al. 2018). Stabilized Matrix Base method (SMM) (Peters and Sette 2005) was used for the prediction and the peptide length was fixed at 11 amino acids. The promising MHC I T-cell epitopes were predicted based on their IC50 values (< 50 nM) of the peptides. Proteasomal tools available at IEDB were also employed to predict T-cell epitopes which consider several important stages of the MHC degradation pathway: proteasome cleavage, TAP binding, MHC binding and epitope selection.

Helper T-lymphocyte prediction

MHC II binding epitopes were predicted using IEDB tools and predominantly occurringHLA class II (DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:02, DRB1*11:01, DRB1*13:02,

DRB1*15:01, HLA-DQA1*05:01|DQB1*02:01, HLA-DPA1*02:01|DPB1*01:01, HLA-DQA1*01:02|DQB1*06:02, HLA DQA1*03:01|DQB1*03:02) alleles were considered for predictions. SMM method was used for the prediction and the peptide length was set at 15 amino acids. The promising MHC II T-cell epitopes were predicted based on IC50 values (< 50 nM) of the peptides.

Antigenicity, population coverage and similarity for the human proteome

The antigenicity of the epitopes was assessed using VaxiJen server (Doytchinova and Flower 2007) and the default parameters were used. Population coverage of the epitopes predicted was analysed using the IEDB population coverage tool and it was predicted for population of South Asia, India and for the entire globe. In order to avoid cross reactivity with the human host, the predicted epitopes were also checked for their similarity with the human proteome by performing BLAST (Altschul et al. 1997) analysis.

Peptide allergenicity, transmembrane topology and solubility

The predicted epitopes were then scanned for allergenicity using ANTIGENpro (Magnan et al. 2010) and SORTALLER (Zhang et al. 2012) tools. Structural properties of the predicted epitopes including transmembrane topology and their solubility upon overexpression were further predicted by ABTMpro and SOLpro respectively, which are available at SCRATCH protein prediction server (http://scratch.proteomics.ics.uci.edu/).

Selection of immunodominantepitopes (IDEs)

The immunodominant epitope (IDE) can bind both B-cell and T-cell effectively and these regions can induce strong immune responses. Hence, epitopes which can bind both B-cell and major of T-cells were selected as IDEs (Verma et al. 2018).

Results and discussion

Gene network analysis

Protein–protein interaction network for the 11 putative mycobacterial lectins was obtained using the stringApp tool of Cytoscape. A network of 104 nodes and 1095 edges were retrieved and is given in Fig. 1. The functional enrichment data for all the 104 nodes were retrieved and classified using its gene ontology (GO) component, process, domain databases such as INTERPRO and PFAM. Figure 2 illustrates the functional enrichment data of all the 104 nodes used in this analysis. Gene classification based on GO component reveals that majority of mycobacterial lection network comprises of genes that are present in the extracellular region followed by cell surface and cell envelope. Classification based on GO process discloses that most of the genes are involved in the interspecies interaction between the species. Classification based on protein domains yielded that most of the genes belong to PE/PPE family of proteins.

Clustering analysis was performed on the mycobacterial lectin network and is depicted in Fig. 3. The network divided into eight clusters and its details are given in Table 1. The first cluster has a score about 28.933 and possesses about 31 nodes and 434 edges and most of the genes are uncharacterized proteins (Table S1). The exact roles of these uncharacterized proteins in TB pathogenesis remain unexplored.

Mycobacterial lectins could be used as potential drug targets for treating Mtb drug resistance

Interference of the mycobacterial lectin- host–cell surface carbohydrates interaction using carbohydrate ligand mimics (or an anti-adhesive drug) is a recent therapeutic approach which can aid to treat resistance in TB patients. In this sort, Mydock-McGrane et al. (2016) has developed mannosides as FimH antagonist which aids in the treatment of urinary tract infections in humans. Recently, Kuhaudomlarp et al. (2020) have found certain non-carbohydrate glycomimetics as Inhibitors of Calcium(II)-Binding Lectins, namely LecA for the treatment of biofilm-associated *P. aeruginosa* infections. Similarly, Šmak et al. (2021) through in-silico virtual screening studies have identified potential inhibitors for selectins which are cell surface lectins for treatment of COVID-19. Fucose-containing glycans were identified as specific ligands for the N-terminal part of the AlsI protein from *Candida albicans* (Donohue et al. 2011). However, fucose-containing ligands were found minimal in chemical compound databases. Therefore in our study, we did not perform molecular docking studies for Rv2082 and Rv1753 which belongs to Agglutinin-like sequences lectin family. Quinoxalinone inhibitors were used against the C-type lectin dendritic cell-specific intercellular adhesion molecule 3–grabbing non-integrin (DC-SIGN) (Mangold et al. 2012). Rv2075c of Mtb is believed to be a C-type lectin. Sulfated glycolipid, such as fucoidan, was found to strongly inhibit filamentous hemagglutinins like laminin and thrombo spondin, suggesting its anti-adhesive role (Roberts and Ginsburg 1988). Rv0355, Rv1917c, Rv3343 and Rv3350 of Mtb are characterised as filamentous hemagglutinins. Therefore, mycobacterial lectins could function as potential drug targets and certain lead-like molecules against lectin–carbohydrate interactions can be identified and used in TB treatment.

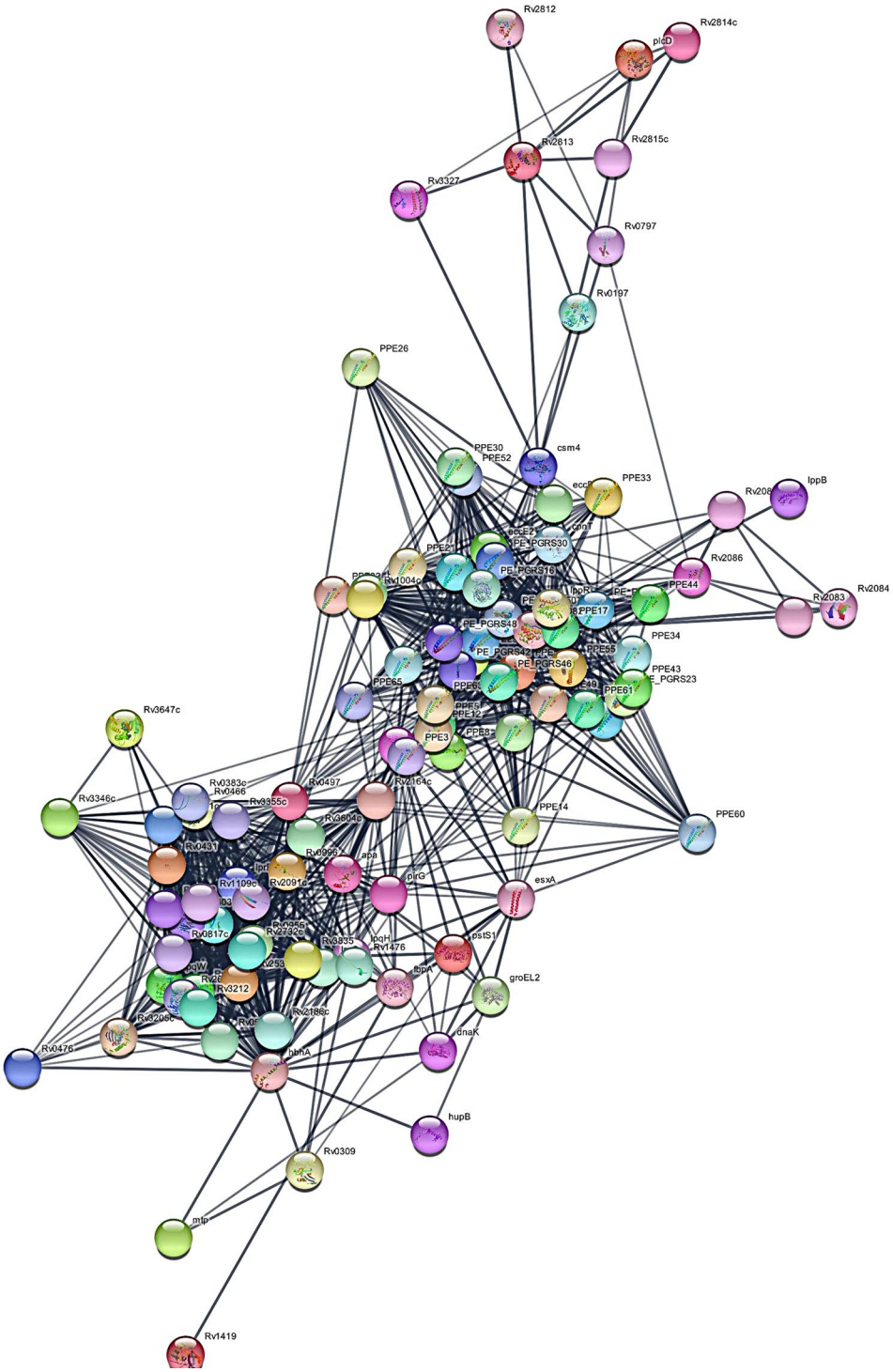

Fig. 1 Gene interaction network of the putative 11 mycobacterial lectins retrieved from Cytoscape. Nodes represent individual proteins in the network

Molecular docking studies of quinoxaline derivatives against the homology modelled lectin Rv2075c

The homology model of Rv2075c (Fig. 4a) was obtained using I-TASSER server. Model 1 was selected as best predicted models with C-score -3.19, TM score 0.36 ± 0.12, and RMSD 15.2 ± 3.4 Å. The validation of the model was checked using the Ramachandran plot (Supplementary Fig. S1), where it shows 59.0% residues in most favoured regions and 25.6% residues in additional allowed regions,i.e., the total of 84.6% residues in allowed region. Further, the homology model of R1917c has an ERRAT score of 78.7056%, which indicates the obtained model is a good

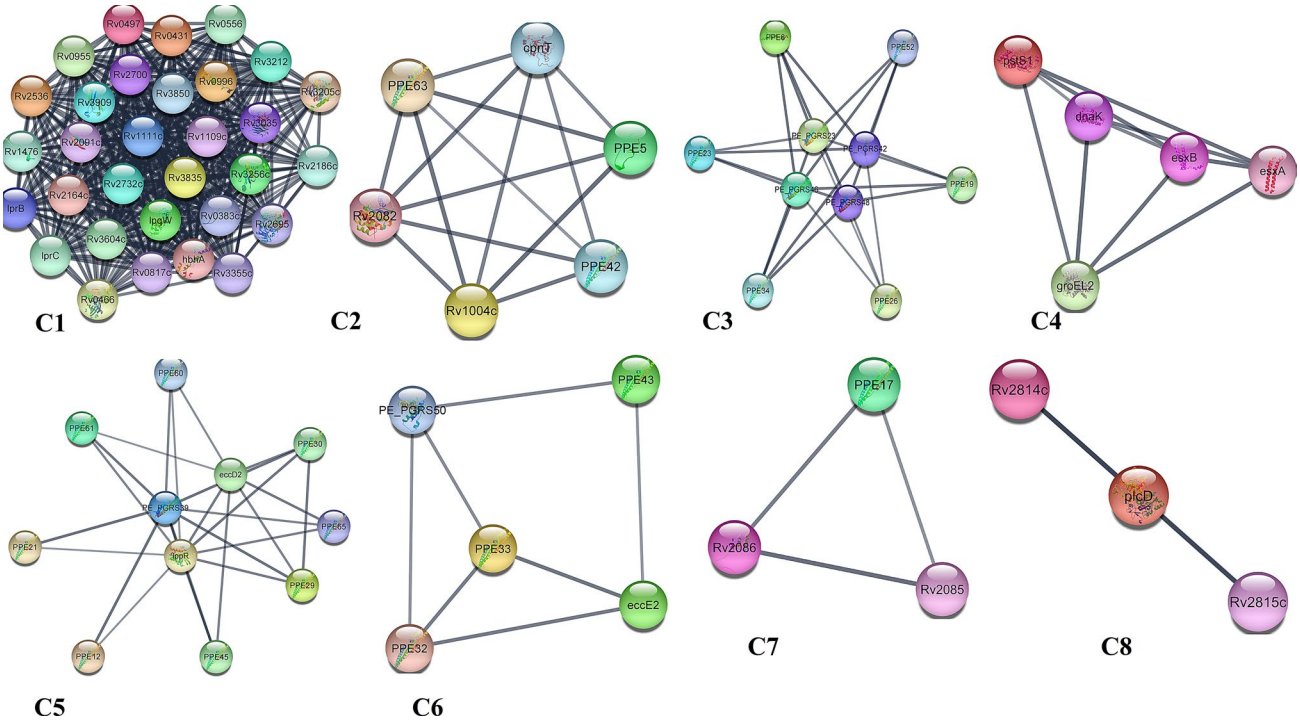

Fig. 2 Clustering analysis by MCODE app on the retrieved Cytoscape gene interaction network. Eight clusters were obtained and are marked as C1–C8. Nodes represent individual proteins in the network

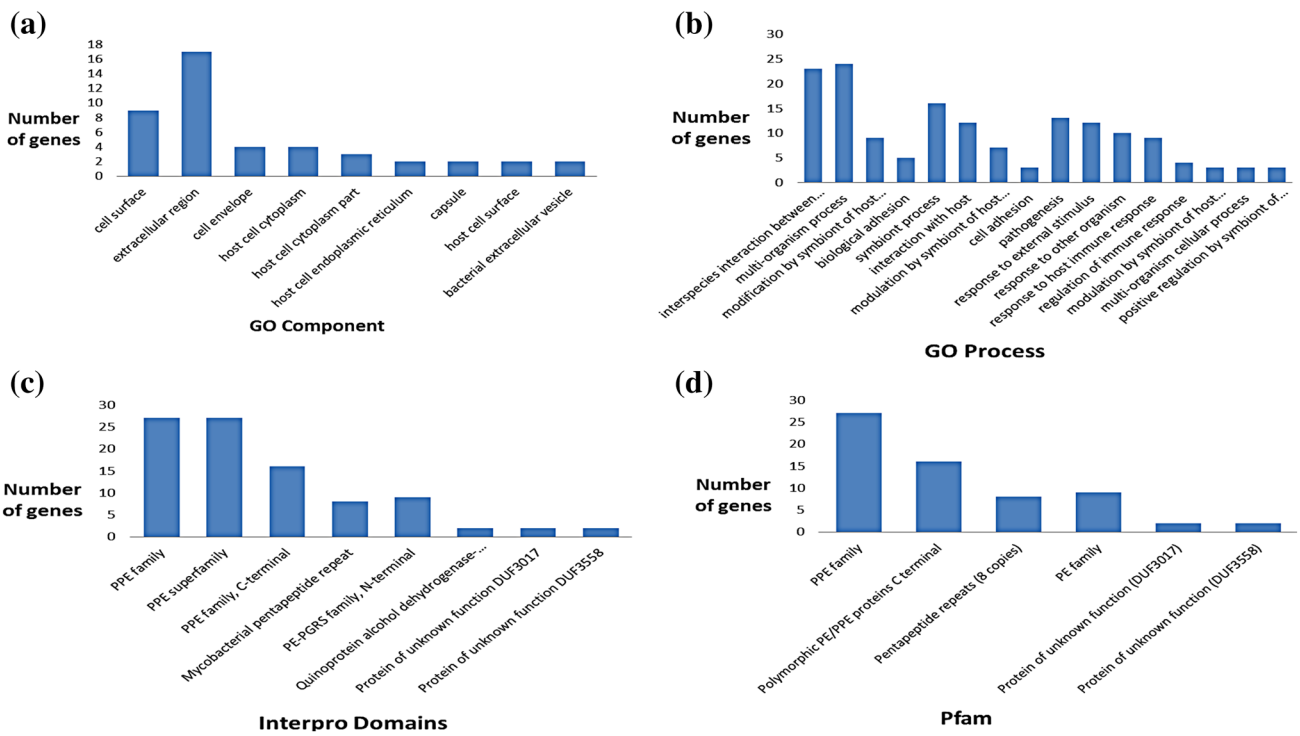

Fig. 3 Gene ontology analysis of the 11 mycobacterial lectins along with their functional partners. **a** Graph depicting the number of genes in different components of the cell. **b** Graph depicting the number of genes involved in different biochemical process. **c** Graph depicting the number of genes having different INTERPRO domains. **d** Graph depicting the number of genes having different PFAM domains

Table 1 Clustering analysis of the mycobacterial lectins network

Cluster	Score (density)	Nodes	Edges	Node ids
1	28.933	31	434	Rv3256c, Rv3604c, Rv0475, Rv0556, Rv1275, Rv1476, Rv3212, Rv2186c, Rv2732c, Rv3909, Rv3850, Rv1111c, Rv2164c, Rv0383c, Rv1274, Rv0431, Rv3205c, Rv2695, Rv2536, Rv3355c, Rv0996, Rv3035, Rv0817c, Rv2700, Rv2091c, Rv1109c, Rv3835, Rv0466, Rv0955, Rv1166, Rv0497
2	5.6	6	14	Rv2608, Rv0304c, Rv1004c, Rv3903c, Rv3539, Rv2082
3	5.333	10	24	Rv1706c, Rv1789, Rv2487c, Rv1243c, Rv2853, Rv0355c, Rv2634c, Rv1361c, Rv3144c, Rv1917c
4	5	5	10	Rv3875, Rv3874, .Rv0934, Rv0440, Rv0350
5	5	11	25	Rv1801, Rv0755c, .Rv3532, Rv2340c, Rv1548c, Rv3478, Rv2403c, Rv3887c, Rv2892c, Rv1802, Rv3621c
6	3.5	5	7	Rv3885c, Rv1809, .Rv3345c, Rv2768c, Rv1808
7	3	3	3	Rv1168c, Rv2086, Rv2085
8	3	3	3	Rv1755c, Rv2815c, Rv2814c

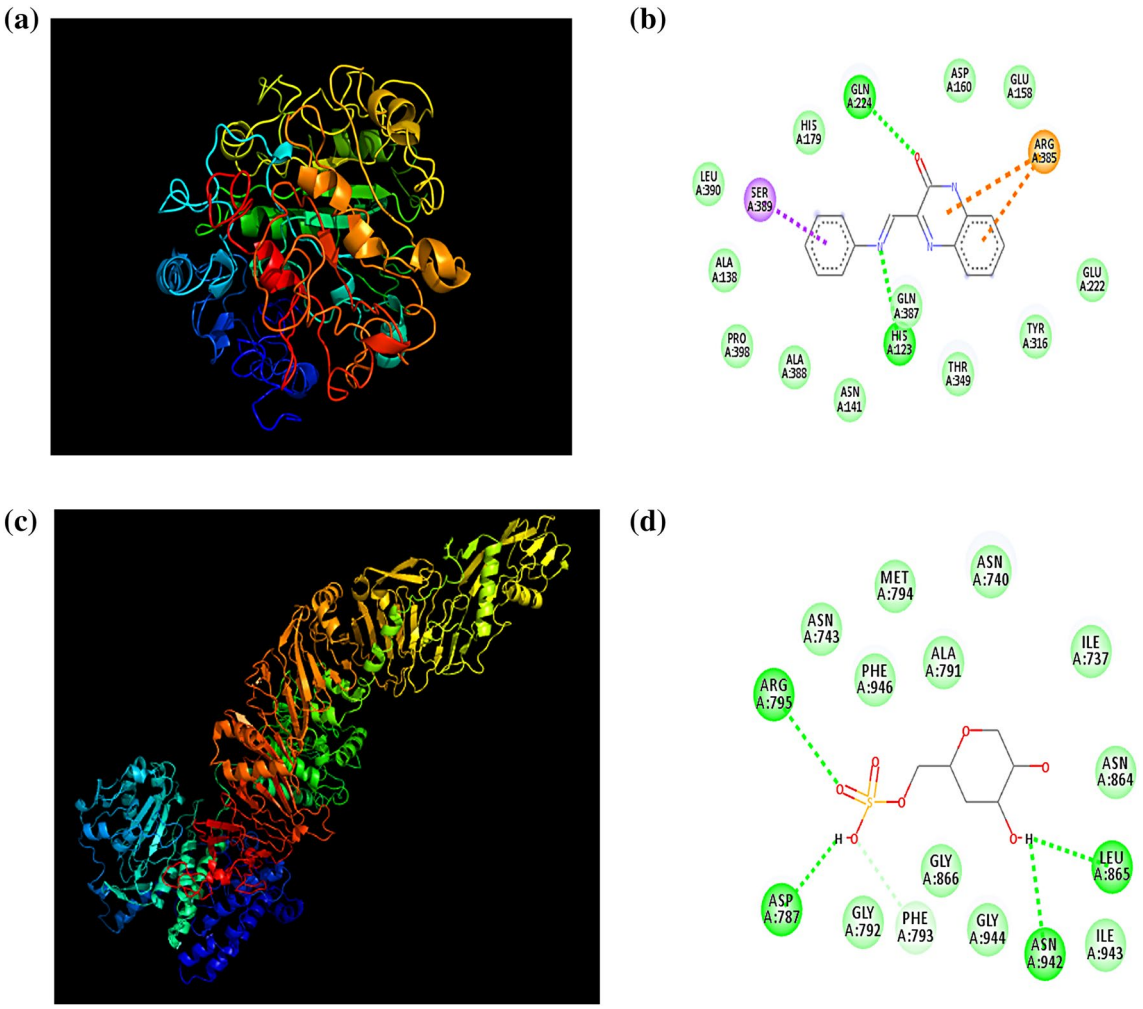

Fig. 4 Homology modelling and Molecular docking studies of Rv2075c and Rv1917c. **a** Homology model of Rv2075c. **b** Molecular docking of Quinoxaline derivative against Rv2075c.The hydrogen bonds are depicted in *green dotted lines* and the bond length is given. c. Homology model of Rv1917c. **d** Molecular docking of Fucoidan derivative against Rv1917c.The hydrogen bonds are depicted in *green dotted lines*

model which can be further used for various studies. 233 quinoxaline derivatives were retrieved from PubChem database and were made to dock against the modelled Rv2075c. The top three compounds with better docking scores are listed in Table 2. The top scored ligand interaction diagram is depicted in Fig. 4b.

Molecular docking studies of fucoidan derivatives against the homology modelled lectin Rv1917c

The homology model of Rv1917c (Fig. 4c) was obtained using I-TASSER server. Model 1 was selected as best predicted models with C-score 0.55, TM score 0.79 ± 0.09 and RMSD 8.5 ± 4.5 Å. The validation of the model was checked using the Ramachandran plot (Supplementary Fig. S2), where it shows 75.4% residues in most favoured regions and 18.2% residues in additional allowed regions, i.e., the total of 93.6% residues in allowed regions which indicates a good quality model. Further, the homology model of R1917c has an ERRAT score of 76.8116%, which indicates the obtained model is a good model which can be further used for various studies. 54 fucoidan derivatives were retrieved from PubChem database and were made to dock against the modelled Rv1917c. The top three compounds with better docking scores are listed in Table 3. The top scored ligand interaction diagram is depicted in Fig. 4d.

Mycobacterial lectins as potential vaccine candidates: an immunoinformatics study

As lectins are present on the mycobacterial cell surface, they can be easily recognized by our immune cells and lectins are used as potential vaccine candidates. FimH, an E.coli lectin, have been identified as a potential vaccine candidate in the treatment of urinary tract infection in humans (Zandi et al. 2020). Similarly, Pseudomonas lectins have also been used as potential vaccine candidates in treating cystic fibrosis patients (Day et al. 2019).

The potential MHC–I and MHC-II T-cell binding epitopes for the mycobacterial lectins are given in Tables 4 and 5. The peptides having IC50 > 50 nm, Vaxijen score < 0.4 and > 80% identity to human proteome are excluded for

Table 2 Molecular docking results of quinoxaline derivatives against modelled Rv2075c

S.no	PubChem CID	Ligand IUPAC Name	Binding affinity	Hydrogen bond	Hydrophobic interactions
1	136315423	3-(phenyliminomethyl)-H-quinoxalin-2-one	−7.2	HIS123, GLN224	SER389, ARG385
2	91407857	2,5,9,12-tetrazabicyclo (11.4.0) heptadeca-1(17),4,9,13,15-pentaene-3,6,8,11-tetrone	−7.1	PRO487, GLY375	PHE376, PRO113, VAL112, PRO487
3	66778485	Methyl 6-amino-2-cyclopropyl-5-((1E)-prop-1-enyl) pyrimidine-4-carboxylate	−7.1	TYR323	ALA321, ALA357, ALA365, LYS362, PRO348, ASP320, TYR323

Table 3 Molecular docking results of fucoidan derivatives against modelled Rv1917c

S.no	PubChem CID	Ligand IUPAC Name	Binding affinity	Hydrogen bond	Hydrophobic interactions
1	60022987	((2R)-4,5-Dihydroxyoxan-2-yl) methyl hydrogen sulfate	−7.5	ASN942, LEU865, ASP787, ARG795	PHE793
2	121308958	((4R)-4-Hydroxyoxan-3-yl) hydrogen sulfate	−7.0	GLY944	GLY866
3	59941957	((3S)-3,5-Dihydroxy-2-methyl-oxan-4-yl) hydrogen sulfate	−6.9	LEU1247 GLY1250 VAL1261	ILE681, SER1260, GLY1250

Table 4 Potential MHC-IT-cell binding epitopes of the putative mycobacterial lectins

Gene	Allele	Peptide	IC50 (nM)	Vaxijen score	Identity (%)
Rv2075c	HLA-A*02:01	LSLAQQLDIDV	22.85	1.2999	63.64
	HLA-B*07:02	GPKNANLGCTV	37.31	0.9200	66.67
Rv0355	HLA-A*02:01	PLLNFSLNIPV	21.82	1.9060	60
Rv1917c	HLA-A*02:01	LHVPIFLDIPV	46.76	1.7648	69.23
	HLA-A*02:01	SLPAIPLGIDV	25.28	0.9600	77.78
	HLA-A*02:01	SIPAIPLSVEV	48.41	1.0625	77.78
Rv3350c	HLA-A*02:01	VLPAIPLNVDV	42.55	1.6734	66.67
	HLA-A*02:01	PPMDFTLGLPV	27.60	1.2447	77.78

Table 5 Potential MHC-II T-cell binding epitopes of the putative mycobacterial lectins

Gene	Allele	Peptide	IC50 (nM)	Vaxijen Score	Identity
Rv1753c	HLA-DQA1*01:02/DQB1*06:02	AAWLAAAAVQAEQTA	24	0.5495	52.38
	HLA-DRB1*11:01	DVFLSTPRITVPAFG	24	0.5061	72.73
	HLA-DRB1*11:01	WDVFLSTPRITVPAF	24	0.8420	64.29
	HLA-DQA1*01:02/DQB1*06:02	PYAAWLAAAAVQAEQ	25	0.4940	52.38
	HLA-DQA1*01:02/DQB1*06:02	YAAWLAAAAVQAEQT	25	0.5512	52.38
	HLA-DQA1*01:02/DQB1*06:02	APYAAWLAAAAVQAE	26	0.4795	55
	HLA-DQA1*01:02/DQB1*06:02	AWLAAAAVQAEQTAA	26	0.6754	50
	HLA-DQA1*01:02/DQB1*06:02	AEQTAAQAAAMIAEF	42	0.6036	61.11
	HLA-DRB1*04:01	QMWAADVSAMSAYHA	47	0.4075	72.73
	HLA-DRB1*04:01	TYEQMWAADVSAMSA	48	0.4127	72.73
Rv2082	HLA-DRB1*01:01	VVDWMLLAAINALIA	5	0.8382	66.67
	HLA-DRB1*01:01	VDWMLLAAINALIAG	5	0.7113	66.67
	HLA-DRB1*01:01	TVVDWMLLAAINALI	7	0.8361	75
	HLA-DRB1*01:01	ATVVDWMLLAAINAL	9	0.8239	75
	HLA-DRB1*01:01	DWMLLAAINALIAGD	9	0.6032	66.67
	HLA-DRB1*01:01	WMLLAAINALIAGDQ	9	0.6229	66.67
	HLA-DRB1*13:02	KAAQLKRNEANDLRN	33	0.6501	69.23
	HLA-DRB1*13:02	AAQLKRNEANDLRNE	33	0.7623	69.23
	HLA-DRB1*13:02	RKAAQLKRNEANDLR	34	0.8492	69.23
	HLA-DRB1*13:02	AQLKRNEANDLRNER	34	0.8705	66.67
Rv2075c	HLA-DRB1*11:01	ELDLHYLPRLEGHGA	16	1.1678	66.67
	HLA-DRB1*11:01	LELDLHYLPRLEGHG	17	1.3206	66.67
	HLA-DRB1*11:01	LDLHYLPRLEGHGAP	17	0.8393	66.67
	HLA-DRB1*11:01	RALELDLHYLPRLEG	18	0.8866	72.73
	HLA-DRB1*11:01	ALELDLHYLPRLEGH	18	1.1554	66.67
	HLA-DRB1*13:02	GALAVVLITAAPVAA	26	0.4654	64.71
	HLA-DRB1*13:02	ALAVVLITAAPVAAD	26	0.5567	64.71
	HLA-DQA1*01:02/DQB1*06:02	LAVVLITAAPVAADA	26	0.5805	77.78
	HLA-DQA1*01:02/DQB1*06:02	AVVLITAAPVAADAY	26	0.4259	77.78
	HLA-DQA1*01:02/DQB1*06:02	VVLITAAPVAADAYQ	26	0.4392	77.78
	HLA-DQA1*01:02/DQB1*06:02	ALAVVLITAAPVAAD	27	0.5567	64.71
	HLA-DQA1*01:02/DQB1*06:02	VLITAAPVAADAYQV	27	0.6535	77.78
	HLA-DRB1*03:01	ALACTAIGADFTLPR	48	0.7486	66.67
	HLA-DRB1*03:01	LACTAIGADFTLPRT	48	0.5830	66.67
	HLA-DRB1*03:01	TAIGADFTLPRTGNQ	49	0.6611	75
Rv0355	HLA-DRB1*01:01	AAFEAALAATVHPAI	8	0.7897	66.67
	HLA-DRB1*07:01	QIPLLNFSLNIPVNI	13	1.7521	75
	HLA-DRB1*07:01	IPLLNFSLNIPVNIP	13	1.8709	75
	HLA-DRB1*07:01	PLLNFSLNIPVNIPI	14	1.7811	52.38
	HLA-DRB1*07:01	VPQIPLLNFSLNIPV	16	1.7174	69.23
	HLA-DRB1*07:01	PQIPLLNFSLNIPVN	16	1.6478	75
	HLA-DRB1*13:02	QIPLLNFSLNIPVNI	17	1.7521	75
	HLA-DRB1*07:01	ALFVSLVVSNLLGQN	17	0.4195	64.71
	HLA-DRB1*13:02	IPLLNFSLNIPVNIP	18	1.8709	75
	HLA-DRB1*13:02	PLLNFSLNIPVNIPI	18	1.7811	52.38
	HLA-DRB1*07:01	LFVSLVVSNLLGQNA	18	0.4992	52.38
	HLA-DRB1*13:02	LLNFSLNIPVNIPIH	22	1.6652	52.38
	HLA-DRB1*13:02	FNTSTVNLSTPANVS	33	0.6273	75
	HLA-DRB1*13:02	TSTVNLSTPANVSGL	33	0.7200	75
	HLA-DRB1*13:02	STVNLSTPANVSGLN	33	0.8467	75

Table 5 (continued)

Gene	Allele	Peptide	IC50 (nM)	Vaxijen Score	Identity
	HLA-DRB1*13:02	VPQIPLLNFSLNIPV	33	1.7174	69.23
	HLA-DRB1*13:02	PQIPLLNFSLNIPVN	33	1.6478	75
	HLA-DRB1*13:02	NTSTVNLSTPANVSG	34	1.0521	75
Rv1917c	HLA-DRB1*01:01	AALVAEFEAVRAAMV	4	0.4524	72.73
	HLA-DRB1*01:01	ALVAEFEAVRAAMVQ	4	0.5194	72.73
	HLA-DRB1*01:01	VAEFEAVRAAMVQPA	4	0.4867	72.73
	HLA-DRB1*01:01	LVAEFEAVRAAMVQP	5	0.5239	72.73
	HLA-DRB1*01:01	AEFEAVRAAMVQPAL	5	0.6552	72.73
	HLA-DRB1*07:01	ASAAASFNSVTSGLV	7	0.5286	76.92
	HLA-DRB1*07:01	SAAASFNSVTSGLVG	8	0.5709	75
	HLA-DRB1*07:01	AALVAEFEAVRAAMV	9	0.4524	72.73
	HLA-DRB1*07:01	ALVAEFEAVRAAMVQ	9	0.5194	72.73
	HLA-DRB1*07:01	LVAEFEAVRAAMVQP	9	0.5239	72.73
	HLA-DRB1*07:01	VAEFEAVRAAMVQPA	9	0.4867	72.73
	HLA-DRB1*07:01	AEFEAVRAAMVQPAL	9	0.6552	72.73
	HLA-DQA1*01:02/DQB1*06:02	AEFEAVRAAMVQPAL	25	0.6552	72.73
	HLA-DQA1*01:02/DQB1*06:02	VAEFEAVRAAMVQPA	26	0.4867	72.73
	HLA-DQA1*01:02/DQB1*06:02	FEAVRAAMVQPALVA	28	0.5115	69.23
	HLA-DQA1*01:02/DQB1*06:02	EFEAVRAAMVQPALV	29	0.7414	69.23
	HLA-DQA1*01:02/DQB1*06:02	EAVRAAMVQPALVAA	29	0.6378	69.23
Rv3343c	HLA-DRB1*01:01	NTSVLGLGAPALVSG	6	0.6365	76.92
	HLA-DRB1*01:01	TSVLGLGAPALVSGL	6	0.4444	76.92
	HLA-DRB1*01:01	LFNTSVLGLGAPALV	7	0.4582	61.11
	HLA-DRB1*01:01	FNTSVLGLGAPALVS	7	0.5524	61.11
	HLA-DRB1*01:01	EILPFTVLLSSLGVT	7	0.6799	76.92
	HLA-DRB1*01:01	ILPFTVLLSSLGVTH	7	0.7525	76.92
	HLA-DRB1*01:01	LPFTVLLSSLGVTHL	7	0.6532	76.92
	HLA-DRB1*13:02	IHILPDIPININETL	30	1.1691	41.67
	HLA-DRB1*13:02	ILPDIPININETLYL	32	0.9765	43.48
	HLA-DRB1*13:02	IEPIHILPDIPININ	36	1.5830	45
	HLA-DRB1*13:02	EPIHILPDIPININE	36	1.2046	40.91
	HLA-DRB1*13:02	TIEPIHILPDIPINI	37	1.4947	45
	HLA-DRB1*13:02	LIGPIHINTGFSIPV	37	0.4028	52.94
Rv3350c	HLA-DRB1*01:01	RGDYQGLAGFAVGYT	4	0.4371	75
	HLA-DRB1*01:01	RGGYEGLVGVRVGPT	7	1.0763	55.56
	HLA-DRB1*01:01	DGSVYVLASSIPLIN	7	0.4524	52.94
	HLA-DRB1*01:01	GSVYVLASSIPLINI	7	0.6895	52.94
	HLA-DRB1*07:01	GSVYVLASSIPLINI	8	0.6895	52.94
	HLA-DRB1*07:01	SVYVLASSIPLINIP	8	0.7677	52.94
	HLA-DRB1*01:01	GVFEAALAATVDPAL	8	0.5914	66.67
	HLA-DRB1*01:01	ADGSVYVLASSIPLI	8	0.6687	52.94
	HLA-DRB1*01:01	SVYVLASSIPLINIP	8	0.7677	52.94
	HLA-DRB1*07:01	ADGSVYVLASSIPLI	9	0.6687	52.94
	HLA-DRB1*07:01	DGSVYVLASSIPLIN	9	0.4524	52.94
	HLA-DRB1*07:01	ITIDAFQVVAGVLFL	10	0.6166	75
	HLA-DRB1*07:01	PLDAISLSESIPIRI	11	1.2241	75
	HLA-DRB1*03:01	IPAIVLDRILLDLHA	12	1.1047	71.43
	HLA-DRB1*07:01	LDAISLSESIPIRIP	12	1.3028	75
	HLA-DRB1*07:01	DAISLSESIPIRIPV	12	1.1725	64.29
	HLA-DRB1*03:01	PVDIPAIVLDRILLD	13	0.9334	75

Table 5 (continued)

Gene	Allele	Peptide	IC50 (nM)	Vaxijen Score	Identity
	HLA-DRB1*03:01	VDIPAIVLDRILLDL	13	1.1140	69.23
	HLA-DRB1*03:01	DIPAIVLDRILLDLH	13	0.9887	71.43
	HLA-DRB1*07:01	GDHEGLFSLFYSLDV	15	0.6337	62.50
	HLA-DRB1*07:01	PIPLDAISLSESIPI	16	0.8321	75
	HLA-DRB1*07:01	IPLDAISLSESIPIR	16	1.1123	75
	HLA-DQA1*01:02/DQB1*06:02	AVTITGTTISAIPLG	30	0.4235	72.73
	HLA-DQA1*01:02/DQB1*06:02	IPAVTITGTTISAIP	31	0.6347	72.73
	HLA-DQA1*01:02/DQB1*06:02	PAVTITGTTISAIPL	31	0.8172	72.73
	HLA-DQA1*01:02/DQB1*06:02	VTITGTTISAIPLGF	31	0.6290	72.73
	HLA-DQA1*01:02/DQB1*06:02	DIPAVTITGTTISAI	32	0.7129	75
	HLA-DRB1*13:02	PLDAISLSESIPIRI	36	1.2241	75
	HLA-DRB1*13:02	GTILIGDIPIPIIDV	37	0.4577	60
	HLA-DRB1*13:02	TILIGDIPIPIIDVP	37	0.7000	66.67
	HLA-DRB1*13:02	LDAISLSESIPIRIP	37	1.3028	75
	HLA-DRB1*13:02	DAISLSESIPIRIPV	37	1.1725	64.29
	HLA-DRB1*04:01	TTRFLLDVNISGGLP	42	0.6965	64.29
	HLA-DRB1*13:02	IPLDAISLSESIPIR	42	1.1123	75
	HLA-DQA1*01:02/DQB1*06:02	GAAAAAMAAAAAPYA	42	0.5580	76.47
	HLA-DRB1*13:02	PIPLDAISLSESIPI	43	0.8321	75
	HLA-DRB1*04:01	TRFLLDVNISGGLPA	44	0.7053	64.29
	HLA-DPA1*02:01/DPB1*01:01	IPQIATTRFLLDVNI	44	0.7218	45.45
	HLA-DRB1*15:01	GDHEGLFSLFYSLDV	45	0.6337	62.50
Rv2813	HLA-DRB1*01:01	VNNLALQALVAAFAA	9	0.5759	62.50
	HLA-DRB1*01:01	NNLALQALVAAFAAD	9	0.4950	62.50
	HLA-DRB1*01:01	NLALQALVAAFAADK	9	0.4072	62.50
Rv3659c	HLA-DRB1*01:01	RAALHSQLAAAVQVL	9	0.4469	72.73
	HLA-DRB1*03:01	AAALVADIVTARLAF	41	0.6356	75
	HLA-DRB1*03:01	AALVADIVTARLAFL	41	0.5509	75
	HLA-DRB1*03:01	PAAAALVADIVTARL	42	0.5035	57.14
	HLA-DRB1*03:01	AAAALVADIVTARLA	42	0.5507	75
	HLA-DRB1*03:01	DPAAAALVADIVTAR	44	0.4446	75
	HLA-DRB1*13:02	EGGAGTVHANNPGEV	44	1.3955	64.29
	HLA-DPA1*02:01/DPB1*01:01	LHSQLAAAVQVLLHV	44	0.5424	72.73
	HLA-DRB1*13:02	AGTVHANNPGEVPAR	45	0.7984	58.82
	HLA-DRB1*13:02	GGAGTVHANNPGEVP	46	1.0540	64.29
	HLA-DRB1*13:02	GAGTVHANNPGEVPA	46	0.7021	58.82
	HLA-DRB1*13:02	GTVHANNPGEVPARM	46	0.7591	62.50

further analysis. B-cell binding epitopes were predicted for all the 11 putative mycobacterial lectins as tabulated in Supplementary Table S2. Further, we intend to identify certain immunodominant regions of the mycobacterial lectins which share both B-cell and T-cell binding epitopes. Through this analysis, we found that Rv1917c has an IDE ^{1062}SIPAIPLSVEV^{1072} which can bind both B-cell and T-cell commendably. Besides previous studies suggests that, Rv1917c can interact with human Toll-like receptor 2 (TLR2) and elicit functional maturation of human dendritic cells (DCs), which leads to secretion of interleukins from $CD4^+$ T cells and induction of Th2 immune response (Bansal et al. 2010). Further, we modelled the IDE and performed a molecular docking study with TLR2 receptor using Cluspro protein–protein interaction server. The docked IDE with the TLR2 receptor along with its interactions obtained using Ligplot software available at PDBsum server is depicted in Fig. 5. In-vitro and in-vivo studies need to be performed to prove the efficacy of the predicted peptide as a potential vaccine candidate.

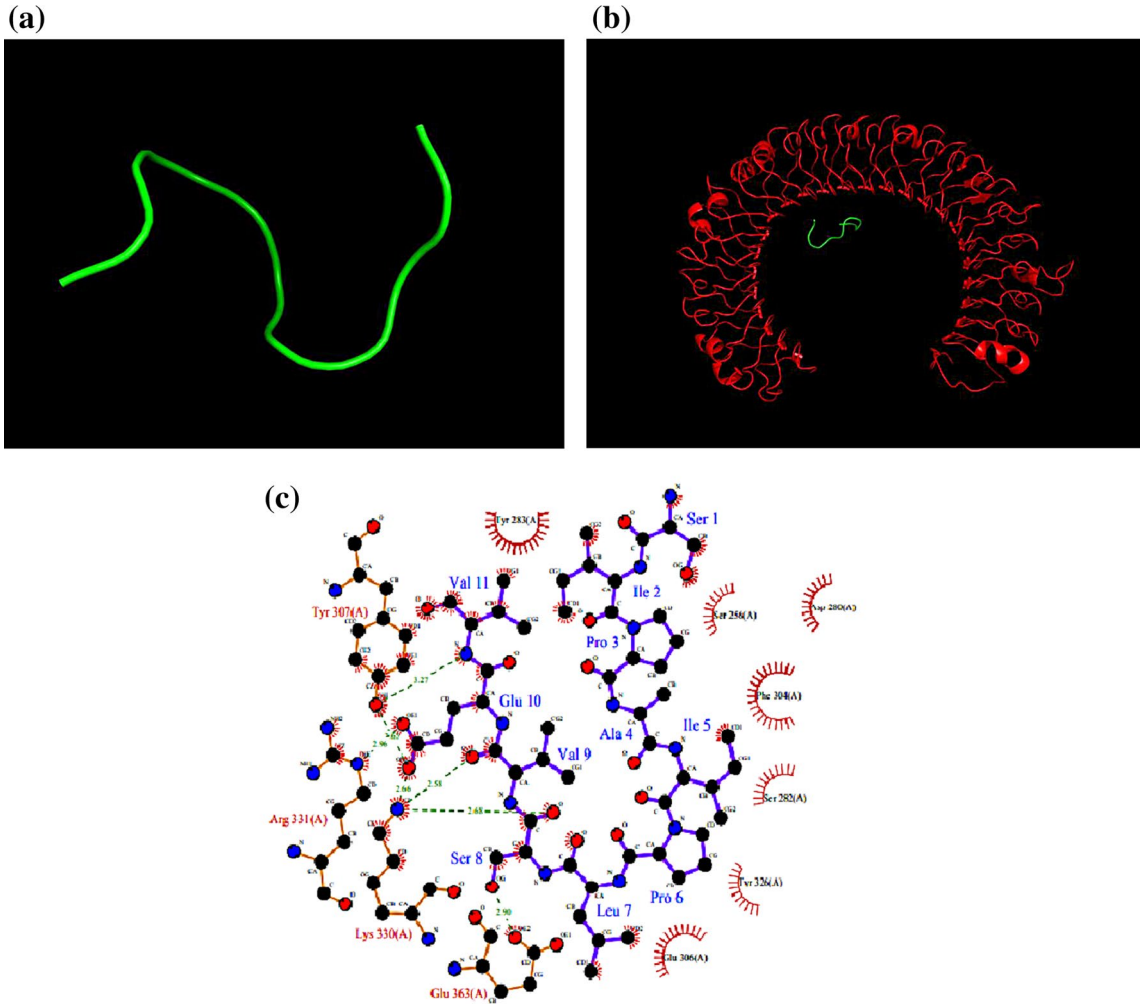

Fig. 5 Immunoinformatics studies. **a** Homology model of the peptide[1062]SIPAIPLSVEV[1072]. **b** Molecular docking of the peptide [1062]SIPAIPLSVEV[1072]with TLR2 receptor. **c** Peptide [1062]SIPAIPLSVEV[1072]interactions with the TLR2 receptor

Conclusion

In this study, a gene interaction network of the 11 potential mycobacterial lectins was retrieved and their functional partners along with the gene ontology analysis. Out of 520 functional interactions, about 30 genes are found to be responsible for cellular activity. Mycobacterial lectins can be used as potential drug targets which can be used for anti mycobacterial drug resistance. Further, molecular docking studies were performed for Rv2075c and Rv1917c with quinoxalinone and fucoidan derivatives, respectively. Good binding affinity was observed for quinoxalinone derivatives against Rv2075c and fucoidan-based derivatives against Rv1917c. Mycobacterial lectins can be used as potential vaccine candidates. Therefore, potential B-cell and T-cell binding epitopes for all of the 11 potential mycobacterial lectins were predicted. Further, an epitope [1062]SIPAIPLSVEV[1072] of Rv1917c, which can bind B-cell and majority of MHC alleles was identified in our study. In conclusion, we have theoretically explored through insilico approaches that mycobacterial lectins can be used as potential drug targets and vaccine candidates against TB treatment.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42485-021-00065-y.

Acknowledgements The authors thank the DBT-Bioinformatics centre at the Department of Bioinformatics, Bharathiar University for the computational facilities. Shobana Sundar acknowledges the financial support received through the award of ICMR—Research Associate (Award Letter No: ISRM/11(13)/2019). Lokesh Thangamani acknowledges the financial support received through the award of ICMR—Senior Research Fellowship (Award Letter No: BMI/11(01)/2020)

Declarations

Conflict of interest All the authors declare that they have none conflict of interest.

References

Abhinav KV, Sharma A, Vijayan M (2013) Identification of mycobacterial lectins from genomic data. Proteins Struct Funct Bioinform 81(4):644–657

Agrewala JN, Wilkinson RJ (1999) Influence of HLA-DR on the phenotype of CD4+ T lymphocytes specific for an epitope of the 16-kDa α-crystallin antigen of *Mycobacterium tuberculosis*. Eur J Immunol 29(6):1753–1761

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402

Bansal K, Sinha AY, Ghorpade DS, Togarsimalemath SK, Patil SA, Kaveri SV, Balaji KN, Bayry J (2010) Src homology 3-interacting domain of Rv1917c of *Mycobacterium tuberculosis* induces selective maturation of human dendritic cells by regulating PI3K-MAPK-NF-κB signaling and drives Th2 immune responses. J Biol Chem 285(47):36511–36522

Chen SL, Hung CS, Pinkner JS, Walker JN, Cusumano CK, Li Z, Bouckaert J, Gordon JI, Hultgren SJ (2009) Positive selection identifies an in vivo role for FimH during urinary tract infection in addition to mannose binding. Proc Natl Acad Sci 106(52):22439–22444

Chodisetti SB, Rai PK, Gowthaman U, Pahari S, Agrewala JN (2012) Potential T cell epitopes of *Mycobacterium tuberculosis* that can instigate molecular mimicry against host: implications in autoimmune pathogenesis. BMC Immunol 13(1):13

Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2(9):1511–1519

Day CJ, Hartley-Tassell LE, Seib KL, Tiralongo J, Bovin N, Savino S, Masignani V, Jennings MP (2019) Lectin activity of *Pseudomonas aeruginosa* vaccine candidates PSE17-1, PSE41-5 and PSE54. Biochem Biophys Res Commun 513(1):287–290

Donohue DS, Ielasi FS, Goossens KV, Willaert RG (2011) The N-terminal part of Als1 protein from *Candida albicans* specifically binds fucose-containing glycans. Mol Microbiol 80(6):1667–1679

Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform 8(1):4

Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J (2016) PubChem substance and compound databases. Nucleic Acids Res 44(D1):D1202–D1203

Kuhaudomlarp S, Siebs E, Shanina E, Topin J, Joachim I, da Silva Figueiredo Celestino Gomes P, Varrot A, Rognan D, Rademacher C, Imberty A, Titz A (2020) Non-carbohydrate glycomimetics as inhibitors of calcium (II)-binding lectins. Angew Chem Int Ed 60(15)

Kundu M, Basu J, Chakrabarti P (1989) Purification and characterization of an extracellular lectin from *Mycobacterium smegmatis*. FEBS Lett 256(1–2):207–210

Lane MC, Mobley HL (2007) Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic *Escherichia coli* (UPEC) in the mammalian kidney. Kidney Int 72(1):19–25

Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, Baldi P (2010) High-throughput prediction of protein antigenicity using protein microarray data. Bioinformatics 26(23):2936–2943

Mangold SL, Prost LR, Kiessling LL (2012) Quinoxalinone inhibitors of the lectin DC-SIGN. Chem Sci 3(3):772–777

Mydock-McGrane LK, Cusumano ZT, Janetka JW (2016) Mannose-derived FimH antagonists: a promising anti-virulence therapeutic strategy for urinary tract infections and Crohn's disease. Expert Opin Therap Patents. 26(2):175–197

Peters B, Sette A (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinform 6(1):132

Roberts DD, Ginsburg V (1988) Sulfated glycolipids and cell adhesion. Archiv Biochem Biophys 267(2):405–415

Saha S, Raghava GP (2007) Prediction methods for B-cell epitopes. In: Flower DR (ed) Immunoinformatics. Humana Press, Totowa, NJ, pp 387–394

Singh DD, Chandran D, Jeyakani J, Chandra N (2007) Scanning the genome of *Mycobacterium tuberculosis* to identify potential lectins. Protein Peptide Lett. 14(7):683–691

Šmak P, Chandrabose S, Tvaroška I, Koča J (2021) Pan-selectin inhibitors as potential therapeutics for COVID-19 treatment: in silico screening study. Glycobiology

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27(3):431–432

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455–461

Verma S, Sugadev R, Kumar A, Chandna S, Ganju L, Bansal A (2018) Multi-epitope DnaK peptide vaccine against *S. Typhi*: an in silico approach. Vaccine 36(28):4014–4022

Vijayan M, Chandra N (1999) Lectins. Curr Opin Struct Biol 9:707–714

Zandi M, Fallah Mehrabadi J, Mahdavi M, Irani S (2020) Construction and development of FimH lectin domain for rising immune response after injection by uropathogenic *E. coli*. Hum Antibodies 28(2):169–178

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinform 9(1):40

Zhang L, Huang Y, Zou Z, He Y, Chen X, Tao A (2012) SORTALLER: predicting allergens using substantially optimized algorithm on allergen family featured peptides. Bioinformatics 28(16):2178–2179