ISSN: 2634 - 8861

Journal of Infectious Diseases & Case Reports

SCIENTIFIC Research and Community

Research Article

Open Access

Computational Analysis of Anion Transporter Proteins Revealed Rv3679 and Rv2397c as Potential Drug Targets for *Mycobacterium Tuberculosis*

Vikas Jha[1]*, Arpita Singh[2] , Nimisha Rane[2], Vrushali Dhamapurkar[1], Kabir Thakur[1], Reetikesh Patel[1], Ashish Jhangiani[1], Arpita Marick[1], Darpan Kaur[1], Sankalp Kasabe[1] and Saloni Rasal[1]

[1]National Facility for Biopharmaceuticals, Guru Nanak Khalsa College of Arts, Science & Commerce, Mumbai, Maharashtra 400019, India

[2]School of Biotechnology and Bioinformatics, D Y Patil Deemed to be University, Navi Mumbai, Maharashtra 400614, India

ABSTRACT

Tuberculosis is still a serious, life-threatening disease all throughout the world. One of the greatest therapeutic challenges is the treatment of multidrug-resistant tuberculosis. The prevalence of multidrug resistant tuberculosis (MDR-TB) poses a danger to the ability of conventional control strategies to stem the global TB epidemic. Drug resistance in TB is primarily a result of poor prescribing practices, noncompliance with treatment regimens, erratic drug supply, and subpar drug quality. The cell wall composition of *Mycobacterium tuberculosis* makes it many times less permeable to chemotherapeutic drugs, the current vaccination only provides modest protection against the disease. The cell membrane, which is made up of proteins from the ATP binding cassette family, allows *Mycobacterium tuberculosis* to quickly alter its cellular membrane to adapt to an unfavourable environment. A greater understanding of such transport proteins may lead to earlier detection and more effective therapeutic drug targets. A thorough computational analyses of *Mycobacterium tuberculosis* anion transporter proteins was conducted. Bioinformatics tools were used to predict its physiochemical properties, secondary structure, functional analysis, protein-protein interaction, subcellular localization, and molecular docking was also carried out. From the selected proteins, Rv3679 and Rv2397c exhibited best results and can be considered as potential drug targets.

***Corresponding author**

Vikas Jha, National Facility for Biopharmaceuticals, Guru Nanak Khalsa College, Mumbai 400019 India. E-mail: vikasjjha7@gmail.com

Received: September 23, 2022; **Accepted:** September 29, 2022; **Published:** October 31, 2022

Keywords: *Mycobacterium Tuberculosis*, Anion Transporter, Molecular Docking, Physio-Chemical Analysis, Structure Validation, Homology Modelling, Drug Target

Introduction

Mycobacterium tuberculosis (Mtb) is one of the world's most persistent and deadly pathogens. It causes tuberculosis (TB) that remains a significant health concern due to high mortality rates worldwide. According to the annual TB report of World Health Organization, 10 million new cases and 1.4 million people died in 2019. Eight countries had two-thirds of the people who developed TB [1]. Though preventable, TB is a transmissible disease that majorly affects the lungs (pulmonary TB) and can also affect other sites (extrapulmonary TB).

Due to the plasticity of its cellular envelope, Mtb can easily modulate its cellular membrane to adapt to an unfavorable host environment and survive inside the host along with minimization of the immune response [2]. Cell membrane proteins such as proteins from the ATP-Binding cassette transporter family have a crucial role in defining these key features. The inner membrane is formed by a phospholipid bilayer and many integral proteins, including ATP-Binding Cassette (ABC) transporters [2]. ABC transporter components are encoded for in 2.5% of the entire genome. It is essential for the transport of amino acids, sugar, and many substrates involved in virulence, nutrition, pathogenesis, cell integrity, and cellular homeostasis, and hence influences Mtb physiology. The ABC transporter, which acts as an efflux pump, may contribute to Mtb antibiotic resistance [3, 4]. It also increases resilience to stress and aids in differentiation and communication. ABC has the critical property of functioning as an ATP-dependent entryway into the bacterial cell. This property can be used to efficiently distribute antimicrobials inside the Mtb cell.

Despite the abundance of ABC transporters throughout life, remarkably few systems have been biochemically, structurally, or mechanistically defined. As a result, broad generalizations about how these systems work are drawn from a few observations, and there is a compelling need to extend these characterizations to other transporters. Using bioinformatic analyses, previous research was able to classify Mtb ABC transporter components into various families. So far, three ABC importer systems have been found, namely Groups I, II, and III. The system engaged in anions, peptides, sugars, and amino acids is comprised of group I ABC importers [2]. Anions like phosphate, sulphur, and molybdate are essential for Mtb survival. They are involved in a variety of biological processes in addition to chemical production. Anion Transport proteins such as SubI, CysA1, CysW, and CysT are

involved in the production of cofactors such as mycothiol, which aids in the protection of oxygen-sensitive proteins, L-cysteine, which is involved in the first line of defense against oxidative stress imposed by macrophages, and methionine, and these proteins may also be potential drug targets [5]. The upregulation of genes responsible for L-cysteine biosynthesis under dormancy has shed light on the link between latent tuberculosis and L-cysteine necessity. Hence, targeting L-cysteine biosynthesis might prove to be a valuable route for antimicrobial drug discovery. Molybdate ions are an essential cofactor for series of enzymes performing nitrogen, carbon, and sulfur metabolism. MOD ABC is responsible for importing molybdate ions, which also helps in survival of Mtb under anaerobic conditions [6]. MOD C was found to be one of the potential biomarkers as this antigen was found in the urine of the patients suffering from active pulmonary TB [7].

Phosphorus is required for the synthesis of nucleotides, phospholipids, and high-energy metabolic intermediates. The PST uptake system of Mtb consists of PstS, which has a high affinity for inorganic phosphorous, and PstB, which provides energy to accelerate the element's uptake. Apart from these, PstA1 modulates gene expression in response to external phosphate availability and is responsible for virulence in the mouse aerosol mode of infection [8]. PstA2 aids Mtb resistance to acidic environments. In tuberculosis patients, PstS1 works as an immunodominant agent. Last but not least, the PstS3 import mechanism is critical for virulence [9].

Despite being an essential component of Mtb survival, the proteins involved in anion transport are poorly understood, both functionally and structurally. Because of the importance of this protein family, the current work is designed to assess its physicochemical features, secondary and tertiary structures, functional analyses, and motif prediction. The found features of individual proteins in this work will aid in the discovery of potential drug targets and further enhance our understanding of the Mtb pathogen, facilitating the development of innovative therapies in the future. Importantly, identifying these targets allows for medication modification with greater efficacy and lower cytotoxicity.

Materials and Methods

Retrieval of Nucleotide and Protein Sequence

Tuberculist was used to obtain the amino acid sequence of the protein [10]. Various criteria such as gene name, location etc. were used to obtain the required sequences [10]. Mycobrowser server was used to acquire information about the function of all the proteins [11]. Mycobrowser also known as Mycobacterial browser is a comprehensive proteomic and genomic data depository for pathogenic Mycobacterium. A genomic and proteomic study of these organisms is possible using manually-curated annotations to appropriate tools provided [12].

Physiochemical Characterization

ProtParam from ExPASy was used to determine the primary structure of 34 anion transporter proteins [13]. It provides information about the physical and biochemical properties including molecular weight (Mw), amino acid composition, extinction coefficients, isoelectric point, aliphatic index, instability index, number of negative and positive residues, half-life, and Grand average hydropathicity (GRAVY) [14].

Functional Analysis

Prediction of SS-bonding of cysteine residues and location of disulfide bridges in protein sequence was made using CYS_REC tool. The tool outputs the position of the cysteins, the total number of cysteins present, and, if present, the pattern of the pairs of cysteins within the protein sequence [15]. NetNGlyc server was used to predict the glycosylation sites. Using artificial neural networks that look at the context of Asn-Xaa-Ser/Thr sequons, the NetNglyc server predicts N-Glycosylation sites in human proteins. Inter-pro identifies protein families, domains, and functional sites to characterize the proteins and indicate essential locations [16, 17]. Psite was used to predict the motifs in the amino acid sequence [18].

Human and Gut Flora Non-Homology Analysis

The proteins with high similarity percentage to the host (Human) might carry undesirable cross-reactivity and can interfere with the binding between the active sites. Therefore, non-homology BLAST search was performed against Human Proteome. In addition to this, the anion transporter proteins of MTB were filtered against the proteins of gut flora bacteria using BLASTp search [19, 20].

Homology Modelling

Tertiary protein structures offer vital insight into the molecular mechanisms behind protein functions, enabling researchers to devise useful experiments like site-directed mutagenesis and analyses of disease-related mutation sites. The SWISS-MODEL and PHYRE2 servers were used to carry out 3D homology modeling of the proteins. SWISS-MODEL which is a completely automated server offers comparative automated modeling of 3D protein structures. Similarly, a 3D modeling server called PHYRE2 employs an Ab initio folding simulation called Poing2 to model regions of your proteins with no detectable homology to known structures [21, 22].

Structure Validation

SAVESv6.0 server was used in order to validate the 3D modelled structures. It provides various programs such as ERRAT, PROCHECK, VERIFY3D etc [23]. In this study, the ERRAT score and PROCHECK program were used for the authentication of the models. ERRAT (overall quality factor) interprets the nonbonded interactions among different types of atoms [24]. PROCHECK is a program that validates the stereochemistry of a protein structure, it analyzes residue by residue geometry as well as the overall geometry structure of the protein [25]. Therefore, to determine the overall quality of model (Z-score) of the predicted structure ProSA-web program is employed wherein the quality scores are displayed according to the reference of all known protein structures [26, 27]. QMEAN score was determined to assess the quality of the structure [28].

Ligand Retrieval

A total of 6 compounds including Viomycin, Ciprofloxacin, Clofazimine, Delpazolid, 4-Aminosalicylic Acid and Clarithromycin were selected as ligands, these compounds were procured from the public chemical structure repository, PubChem [29]. SDF files of these compounds were downloaded and further used for analysis.

Protein and Ligand Preparation

Protein

Rv3679 and Rv2397c proteins were shortlisted from the 34 anion transporter proteins. The 3D structures of these proteins were prepared for molecular docking analysis employing UCSF Chimera 1.15 tool through which Kollman charges and polar hydrogen atoms were added to the protein molecule, the water molecules were eliminated and the charged protein molecule was saved in PDB format [30].

Ligand

Viomycin, Ciprofloxacin, Clofazimine, Delpazolid, 4-Aminosalicylic Acid and Clarithromycin were chosen as ligands, and their structures were obtained from PubChem databank in SDF format. PyRx Virtual Screening Tool was used for generating structural variations in order to optimize and minimize energy of the selected ligands [31].

Molecular Docking

A key approach for predicting the prevailing binding interactions of a chosen ligand with a protein having a known three-dimensional structure is molecular docking. Using Auto Dock Vina, the 6 chosen ligand structures were docked with Rv3679 and Rv2397c proteins of *Mycobacterium tuberculosis* [32]. The prediction of bound confirmation based on free binding energies, which were computed using the empirical force field, is carried out by the Auto Dock Vina software. The Auto Dock Vina docking via PyRx Virtual Screening Tool was used to carry out the docking analysis [31]. This analysis assisted in identifying candidate ligands with strong protein binding affinities as possible inhibitors.

Results and Discussion

Physiochemical Characterization

Table 1: Physicochemical Properties of Anion Transport Proteins as Predicted by Expasy's ProtParam Program

Protein Name	No. of Amino acids	Mol.wt	pI	negative residues	positive residues	Extinction Coefficients		Half Life			Instability Index	Aliphatic index	GRAVY
Rv2684	429	45155.71	5.66	27	22	49055	48930	100	>20	>10	22.31	139.32	1.074
Rv2685	428	45237.69	5.73	29	23	54555	54430	30	>20	>10	23.94	136.43	1.018
Rv3578	413	42524.95	11.22	12	28	42065	41940	5.5	3min	2min	38.61	139.69	1.066
Rv2643	498	53225.75	8.74	34	38	87110	86860	30	>20	>10	28.57	121.91	0.583
Rv2397c	351	37443.78	9.12	38	41	14440		30	>20	>10	26.72	99.97	0.037
Rv2399c	283	29682.94	9.38	19	21	34490		30	>20	>10	32.11	126.82	0.799
Rv2398c	272	29288.68	9.2	14	17	50420	50420	30	>20	>10	33.89	132.54	0.88
Rv1857	261	26544.36	9.03	15	19	8730	8480	100	>20	>10	36.66	94.64	0.348
Rv1858	264	276776.08	10.76	11	20	36440	36440	30	>20	>10	31.86	137.05	0.935
Rv1859	369	38610.2	8.95	34	37	19605	19480	30	>20	>10	38.17	103.36	0.117
Rv1860	325	32720.76	4.93	25	19	30940		30	>20	>10	62.72	67.23	-0.256
Rv2329c	515	55136.35	9.48	27	38	81735	81360	30	>20	>10	36.84	108.35	0.519
Rv1737c	395	41109.14	9.69	15	22	67505	67380	30	>20	>10	30.26	109.01	0.825
Rv0261c	469	49356.53	9.14	20	26	88600	88350	30	>20	>10	30.44	108.25	0.658
Rv0267	463	49133.54	9.32	21	28	97650	97400	30	>20	>10	34.8	105.21	0.701
Rv0934	374	38211.07	5.14	26	17	37025	36900	100	>20	>10	25.58	87.57	0.072
Rv0928	370	37935.07	5.77	28	26	35660	35410	5.5	3min	2min	20.47	75.51	-0.14
Rv0820	258	28029.11	6.37	29	28	13075	12950	100	>20	>10	40.02	97.87	-0.088
Rv3301c	221	24826.53	5.73	34	27	10095	9970	30	>20	>10	27.48	102.9	-0.073
Rv0821c	213	23513.84	5.24	33	21	9970	9970	30	>20	>10	33.48	101.36	0
Rv0545c	417	42713	10.18	15	25	69440	69440	100	>20	>10	27.14	117	0.69
Rv2281	552	58789.41	8.57	37	39	79410	79410	30	>20	>10	32.26	107.86	0.53
Rv0930	308	33466.01	10.51	14	24	54430		100	>20	>10	43.38	131.3	0.878
Rv0936	301	32235.95	9.28	15	21	74495	74370	30	>20	>10	47.92	121.5	0.654
Rv0933	276	29996.39	9.35	30	35	11585	11460	30	>20	>10	40.16	89.09	-0.127
Rv0935	338	34793.42	7.99	18	19	34950	34950	30	>20	>10	27.44	131.51	0.917
Rv0929	324	34229.47	10.51	17	27	27960		100	>20	>10	32.37	121.39	0.786
Rv0932c	370	37831.96	4.97	27	21	40130	39880	100	>20	>10	21.87	72.32	-0.132
Rv2400c	356	37418.32	6.31	34	32	53065	52940	30	>20	>10	29.83	88.82	-0.011
Rv0143c	492	50808.12	7.79	32	33	70785	70360	30	>20	>10	23.12	110.93	0.574
Rv1707	486	51171.12	9.43	31	36	21430	21430	100	>20	>10	27.27	126.77	0.696
Rv1739c	560	59366.88	9.59	39	47	51005	50880	30	>20	>10	49.79	120.11	0.636
Rv3679	340	35856.28	5.27	40	31	12950		30	>20	>10	32.4	114.79	0.177
Rv3680	386	41404.62	5.78	43	40	18575	18450	30	>20	>10	37.58	95.13	0.049

The physiochemical properties of 34 anion transporter proteins (refer supplementary file) were evaluated by ExPAsy ProtParam Server (Table 1). This server consists of analysis tools including Compute pI/Mw, employed in the prediction of isoelectric point (pI) and molecular weight (Mw) of the protein; ProtParam, a tool estimating various physicochemical parameters; PeptideMass, used in theoretically cleaving proteins and calculating the masses of their peptides as well as any known cellular or artifactual posttranslational

modifications; PeptideCutter, for predicting cleavage sites of proteases or chemicals present in the protein sequences; ProtScale, a tool utilized for amino acid scale representation, including hydrophobicity plots. The molecular weight of the proteins is closely related to the evaluation of biomolecule functionality, such as gene and metabolic regulation [13]. The analysis revealed that Rv1858 has the highest molecular weight of 276776.08, whereas Rv0821c has the lowest molecular weight with a value of 23513.84 (refer supplementary file). The isoelectric point (pI) is where the net charge on amino acid is zero as it produces an equal amount of positive and negative ions [33]. Acidic amino acids have a low pI value, whereas basic amino acids have a high pI value [34]. Rv3578 was found to be highly basic with a value of 11.22. Rv1860 and Rv932C were the most acidic proteins with the value of 4.93 and 4.97, respectively (refer supplementary file).

The instability index (II), in addition to pI, offers an estimate of the protein's stability in vitro and in vivo. A protein is stable if its instability index is below 40, and unstable if it is above 40 (refer supplementary file) [35]. Rv1860 is the most unstable protein as it has a value of 60; on the other hand, Rv0928, a protein involved in phosphate binding, is the most stable protein, with a score of 20. In total 6 (17.65%) proteins were unstable whereas 22 (82.35%) were stable. To describe the relative volume of a protein occupied by aliphatic side chains (valine, leucine, isoleucine, and alanine), the aliphatic index is calculated (refer supplementary file). It is a measure of thermostability of a protein. A Protein with a high aliphatic index is more thermally stable [36]. It can be observed that Rv3578 and Rv2684 are among the most stable proteins while Rv1860 is the most unstable. The average of Hydropathicity is determined through the GRAVY value. A positive GRAVY value points towards Hydropathicity and non-polar protein, whereas a negative GRAVY value indicates hydrophilicity and the protein being polar has better interaction with water [33]. It is evident that Rv2684 with the value 1.074, followed by Rv3578 and Rv2685 having the value of 1.066 and 1.018, respectively, were the most hydrophobic. Rv1860 with -0.256 is the most hydrophilic and hence, will have better interaction with water (refer supplementary file).

Functional Analysis

Table 2: Number of Cysteine Residues, their Position and Disulfide Bridge Predicted using CYS_REC Tool

Protein Name	No. of Cysteine	Positions	Prediction	Score
Rv2684	2	CYS 127	not SS-bounded	-25.5
		CYS 227	not SS-bounded	-48.7
Rv2685	3	CYS 30	not SS-bounded	-34.5
		CYS 127	not SS-bounded	-25.9
		CYS 227	not SS-bounded	-47.5
Rv3578	3	CYS 72	not SS-bounded	-38.8
		CYS 111	not SS-bounded	-33.1
		CYS 391	not SS-bounded	-22.3
Rv2643	5	CYS 132	not SS-bounded	-46.6
		CYS 144	not SS-bounded	-34.1
		CYS 371	not SS-bounded	-32.3
		CYS 445	probably not SS-bounded	-5
		CYS 449	probably not SS-bounded	-11.2
Rv2397c	No Cysteines			
Rv2399c	No Cysteines			
Rv2398c	1	CYS 147	not SS-bounded	-45.3
Rv1857	4	CYS 22	not SS-bounded	-20.8
		CYS 149	probably not SS-bounded	-14.6
		CYS 155	probably not SS-bounded	-13.1
		CYS 210	not SS-bounded	-23.1
Rv1858	1	CYS 65	not SS-bounded	-29.5
Rv1859		CYS 103	probably not SS-bounded	-12.6
		CYS 190	not SS-bounded	-35.2
	3	CYS 331	not SS-bounded	-36.9
Rv1860	No Cysteines			
Rv2329c	7	CYS 151	probably not SS-bounded	-13.9
		CYS 216	not SS-bounded	-33.5
		CYS 292	not SS-bounded	-26.9
		CYS 373	not SS-bounded	-31.1
		CYS 399	not SS-bounded	-18.7
		CYS 460	not SS-bounded	-23.9

		CYS 511	SS-bounded	19608474
Rv1737c		CYS 309	not SS-bounded	-34
	2	CYS 378	not SS-bounded	-31.5
Rv0261c	4	CYS 79	not SS-bounded	-26.3
		CYS 128	not SS-bounded	-27.8
		CYS 193	not SS-bounded	-35.5
		CYS 312	not SS-bounded	-28.2
Rv0267	4	CYS 93	not SS-bounded	-21.5
		CYS 133	not SS-bounded	-41.9
		CYS 231	not SS-bounded	-32.7
		CYS 450	not SS-bounded	-21.6
Rv0934	3	CYS 24	not SS-bounded	-22.1
		CYS 238	probably not SS-bounded	-11.6
		CYS 244	not SS-bounded	-20.3
Rv0928	5	CYS 23	not SS-bounded	-27.1
		CYS 45	probably not SS-bounded	-13.1
		CYS 74	not SS-bounded	-17.5
		CYS 119	not SS-bounded	-22
		CYS 319	not SS-bounded	-19.1
Rv0820	3	CYS 41	not SS-bounded	-36.7
		CYS 158	not SS-bounded	-22.7
		CYS 176	not SS-bounded	-21.4
Rv3301c	2	CYS 20	not SS-bounded	-28.6
		CYS 120	not SS-bounded	-31.7
Rv0821c	1	CYS 20	not SS-bounded	-30.7
Rv0545c	1	CYS 225	not SS-bounded	-27.2
Rv2281	1	CYS 451	not SS-bounded	-28.9
Rv0930	No Cysteines			
Rv0936	3	CYS 37	not SS-bounded	-28.5
		CYS 39	probably not SS-bounded	-10.9
		CYS 40	not SS-bounded	-17.3
Rv0933	2	CYS 3	not SS-bounded	-24
		CYS 173	not SS-bounded	-25.9
Rv0935	1	CYS 226	not SS-bounded	-21.6
Rv0929	No Cysteines			
Rv0932c	5	CYS 23	not SS-bounded	-18.2
		CYS 43	probably not SS-bounded	-0.3
		CYS 72	not SS-bounded	-18.9
		CYS 119	probably not SS-bounded	-4.6
		CYS 319	not SS-bounded	-15.2
Rv2400c	3	CYS 11	probably not SS-bounded	-13.3
		CYS 28	not SS-bounded	-30.4
		CYS 38	not SS-bounded	-24
Rv0143c	2	CYS 183	not SS-bounded	-22.9
		CYS 422	not SS-bounded	-33.1
Rv1707	1	CYS 101	not SS-bounded	-33.2
Rv1739c	3	CYS 119	not SS-bounded	-30.5
		CYS 299	not SS-bounded	-29.4
		CYS 462	not SS-bounded	-22.9

Rv3679	No Cysteines			
Rv3680	3	CYS 24	probably not SS-bounded	-13
		CYS 25	probably not SS-bounded	-4.3
		CYS 290	not SS-bounded	-30.9

Table 3: Motifs Present in Proteins determined using NetNGlyc and Psite Server

Protein	Rv2684	Rv2685	Rv3578	Rv2643	Rv2397c
N-glycosylation site	4	2	3	1	1
Protein kinase C phosphorylation site	2	2	3	4	3
Casein kinase II phosphorylation site	6	4	1	2	4
Tyrosine kinase phosphorylation site					1
N-myristoylation site	13	11	7	12	10
Prenyl group-binding site		1	1		
MicrobodiesC-terminal targeting signal	2	4	8	6	7
Leucine zipper			1		
cAMP- and cGMP-dependent protein kinase phosphorylation site.				1	
Glycosaminoglycan attachment site					1
Endoplasmic reticulum targeting sequence					1
Cell attachment					1
ATP/GTP-binding site motif A (P-loop)					1
ABC transporters					1
Binding-protein-dependent transport systems inner membrane comp					
Amidation site.					
Prokaryotic membrane lipoprotein lipid attachment site					
Growth factor and cytokines receptors family signature 2					
Regulator of chromosome condensation (RCC1) signature 2					

	Rv0820	Rv0928	Rv0934	Rv0267	Rv0261c	Rv1737c	Rv2329c	Rv1860	Rv1859	Rv1858	Rv1857	Rv2398c	Rv2399c
Rv0820		5	2	4	3		1	2	1			1	
Rv0928	5		2	2	2	3	6	2	4	2	1	2	2
Rv0934	2	2			7	2	11	2	8	2	2	2	1
Rv0267	4	2										1	
Rv0261c	3	2	7			3	6	2	4	2	1	2	2
Rv1737c		3	2		2		3	2	1	1			
Rv2329c	1	6	11		23	17		6	11	9	6	9	9
Rv1860	2	2	2		2		1		1				
Rv1859	1	4	8		2	4	6	8	7	3	3	3	5
Rv1858		2	2										
Rv1857		1	2										
Rv2398c	1	2	2										
Rv2399c		2	1										

Rv1707	Rv0143c	Rv2400c	Rv0932c	Rv0929	Rv0935	Rv0933	Rv0936	Rv0930	Rv2281	Rv0545c	Rv0821c	Rv3301c
2	2	2	5	1		1			2	1		
8	5	4	2	1		5	3	4	3	4	2	3
8	3	3	3	5	6	5	2	1	1		2	2
						1	1					
20	23	9	17	6	12	5	8	6	17	22	1	3
1	1	1					1			1	1	1
5	6	6	4	7	3	6	3	3	4	4	4	4
						2	1					
1	1		2	1	2				1	1		1
1					1							
						1						
						1						
1	1	1	1	1	1		1		1	1		1

Rv1739c		6	8		11	1	6					1		1					1
Rv3679	1	6	3		6		11						1			1			
Rv3680	1	3	7	1	6	1	5		1				1						

The result of functional analysis using CYS_REC tool suggest that almost every protein had cysteine residues, except 6 proteins (Table 2) including Rv2397c, Rv2399c, Rv1860, Rv0930, Rv0929 and Rv3679. Cysteine is involved in primary structure analysis as it can easily be oxidized, forming a dimer having a disulfide bridge between two cysteines. This feature is essential for the stabilization of the tertiary and quaternary structure [37]. From Table 2, it can be noted that only residue CYS511 of Rv2329c is SS bounded. N-glycosylation is one of the most chemically complex and important transformations. The result obtained from the NetNGlyc tool which is represented in (Table 3) predicted that 11 proteins have no N-glycosylation sites present in them. Rv0928 and Rv0932c have five such sites, the maximum among the sampled proteins. Like the NetNGlyc tool, the Psite server predicts the motifs present in the selected amino acid sequences. N-myristoylation site was observed in all the anion transport protein sequences under investigation. Myristoylation is a post-translational alteration in protein that occurs by the addition of myristic acid. The conformational stability for the protein's catalytic activity is directly proportional to the number of myristoylation sites present [38-40]. It was observed that all the proteins have Microbodies C-terminal targeting signal sites. Phosphorylation regions for protein kinase C and Casein Kinase II were found for each protein with Rv0935, Rv0545c, and Rv0267 as the only exceptions. Proteins having many phosphorylation sites suggest that they may be regulated frequently [41].

Human and Gut Flora Non-Homology Analysis

Host metabolism could be adversely affected by potential pharmacological targets that are homologous to the host. Hence, identifying non-homologous proteins is important step for a good drug target prediction. The 5 essential genes were subjected to non-homology analysis, and they were all identified as non-homologous to human proteome [19]. By digesting food particles and preventing harmful bacteria from colonising the intestine, gut micro flora play a crucial role in metabolism [42]. Co-involvement of the human host with gut microflora have crucial role in absorbing the poorly digestible dietary components, degradation of xenobiotics, and vitamin synthesis [43, 44]. They also provide resistance to opportunistic bacteria and colonization of pathogens [44, 45]. A deterioration of the healthy gastrointestinal microbiota population results in a poor first line of defense against invading bacteria. And it may also lead to the deficiency of nutrients in the host [44]. After the homology search, 17 proteins that were gut flora non-homologous were shortlisted.

Homology Modelling

SWISS-MODEL and PHYRE2 SERVER predicted the 3D structures of the sequences. To assess the predicted models, Q-mean, Z-score, ERRAT score, and Ramachandran plot analyses were carried out. Each server chooses the ideal modelling template based on how closely the query sequences resemble each other. The examination of the tertiary structures for the proteins under investigation was carried out (refer supplementary file). The Z-score and Q-mean score, which should be between -4 and +4, were used by SWISS-MODEL to predict successful models for 16 proteins. The same criteria was used by PHYRE to identify models for 13 proteins.

Comparison of the approved 3D models produced by SWISS-MODEL and PHYRE based on their respective ERRAT levels yield notable results. For instance, the ERRAT value for Rv0820 for PHYRE is 66% while for SWISS-MODEL it is 87.8049%. Other widely used structures have a comparable pattern. A greater ERRAT number indicates a higher level of accuracy in the projected model for non-bonded interactions. However, both servers' structures have high negative z-score values, which indicate more nativeness [46]. Three more protein sequences with large ERRAT values and negative z-scores, Rv1860, Rv1739, and Rv2399c, were also structurally represented by the SWISS-MODEL service. According to the aforementioned statistical findings, SWISS-MODEL outperformed PHYRE. (refer supplementary file).

The Ramachandran plot, used in further study of the acceptable models, reveal that, on average, 91.875% of amino acids were in the most favored zone, whereas 0.34% were in the prohibited region. If 90% of the residues are in the preferred region, tertiary structures are deemed to be of high grade [47]. Similar to this, an average ERRAT value of 93.21% is discovered. (refer supplementary file).

Structure Validation

Table 4: Predicted Secondary Structure Elements using SOPMA Software

Protein Name	Alpha helix	Extended strand	Random Coil
Rv2684	59.91	16.08	19.58
Rv2685	57.48	16.36	21.73
Rv3578	63.2	11.14	20.58
Rv2643	45.58	15.26	34.34
Rv2397c	36.18	21.94	35.61
Rv2399c	55.12	14.84	24.03
Rv2398c	57.72	15.81	20.22
Rv1857	34.1	20.69	36.78
Rv1858	46.12	17.42	29.92
Rv1859	31.71	22.76	38.75
Rv1860	20.31	14.15	62.15
Rv2329c	42.33	20.78	31.07
Rv1737c	56.71	11.9	25.82
Rv0261c	44.56	21.54	27.72
Rv0267	43.84	20.3	29.81
Rv0934	33.96	16.58	42.78
Rv0928	31.89	15.95	44.86
Rv0820	41.09	17.83	34.5
Rv3301c	83.26	2.26	14.48
Rv0821c	83.57	2.35	14.08
Rv0545c	52.04	12.95	28.3
Rv2281	53.44	18.3	23.19
Rv0930	50.65	14.94	29.22
Rv0936	46.51	17.28	31.23
Rv0933	44.93	17.39	31.88
Rv0935	52.37	15.68	26.33
Rv0929	53.09	14.51	29.63
Rv0932c	33.78	16.49	42.97
Rv2400c	38.2	17.13	37.36
Rv0143c	43.7	14.02	39.23
Rv1707	52.26	16.46	26.95
Rv1739c	57.14	11.61	26.43
Rv3679	46.47	14.41	31.18
Rv3680	59.07	9.84	26.94

The pattern of hydrogen bonding determines a protein's secondary structure, with alpha helices and beta sheets being the most important types of secondary structures. Water-soluble globular proteins frequently contain alpha-helices. Alpha - helices are commonly involved in protein interactions with other proteins, lipid interactions in the cell membrane, and nucleic acid interactions. Highly stable alpha helices have found use in pharmaceuticals [48]. The amino acid substitution, which is primarily associated with protein thermostability, is found in alpha helix. Rv0821c has a higher percentage of alpha - helix (83.57%), while Rv3301c has a percentage of 83.26% (Table 4). Random coils are regions that form an irregular secondary structure but are not characterized by hydrogen bonding. It is essential for protein enzymatic turnover, flexibility, and conformational changes [49]. They can range in length from 4 to 20 residues. It also increases protein flexibility, for example, during enzymatic turnover. They are more tolerant of mutation than other structures. Rv1859 has the highest percentage of random coil with a value of 22.76%. Extended secondary structures are formed by the accumulation of secondary structures that directly interact with atoms in the main chain (ESS). It aids in the stabilization of the original secondary structure and subsequent tertiary structures. Rv1860 has the highest percentage of ESS (62.15%) (refer supplementary file) [50]. According to the SOSUI tool, 64.71% of proteins were transmembrane, while 35.29% were soluble. Table 4 shows which proteins are soluble and which proteins are transmembrane.

Molecular Docking

Table 5: Binding energy of Rv3679 protein with the selected 6 ligands

Sr. No	Ligand	Binding Energy n(ΔG) (kcal/mol)
1	Viomycin	-7
2	Ciprofloxacin	-6.7
3	Clofazimine	-7.5
4	Delpazolid	-6.4
5	4-Aminosalicylic Acid	-5.7
6	Clarithromycin	-6.8

Table 6: Binding energy of Rv2397c protein with the selected 6 ligands

Sr. No	Ligand	Binding Energy n(ΔG) (kcal/mol)
1	Viomycin	-6.4
2	Ciprofloxacin	-6.6
3	Clofazimine	-7.6
4	Delpazolid	-6.4
5	4-Aminosalicylic Acid	-5.4
6	Clarithromycin	-6.1

Table 7: Binding energy of Rv3679 protein with drugs currently used for treatment of Tuberculosis

Sr. No	Ligand	Binding Energy n(ΔG) (kcal/mol)
1	Isoniazid	-4.8
2	Pyrazinamide	-4.8
3	Fluoroquinolone	-7.1

Table 8: Binding energy of Rv2397c protein with drugs currently used for Tuberculosis treatment

Sr. No	Ligand	Binding Energy n(ΔG) (kcal/mol)
1	Isoniazid	-5.1
2	Pyrazinamide	-4.9
3	Fluoroquinolone	-7.1

A simulation technique called molecular docking explores the best way to attach a ligand to an active site on a target. In this method, the binding site in the target is chosen using 3D coordinates, and the binding affinity of the resulting orientation of the molecule within the binding site, which creates the complex, is calculated [51]. The significance and sensitivity of binding affinity values are determined by the largest magnitude negative number (highest binding affinity or lowest binding energy), which represents the most advantageous conformation of the complex formed when the involved ligand effectively binds with the active pockets of the target. Molecular docking with the Rv3679 and Rv2397c anion transporter proteins of *Mycobacterium tuberculosis* was done using 6 bioavailable ligands. The binding energies of these ligands for proteins Rv3679 and Rv2397c are shown in Table 5 and Table 6, respectively indicating that they have a high affinity for the target proteins. The binding affinity of complexes for Rv3679 protein was found to be between −5.7 and −7.5 kcal/mol whereas complexes of Rv2397c protein exhibited a binding affinity between -5.4 and -7.6 kcal/mol. From Table 7 and Table 8 it can be observed that the 6 selected ligands exhibit higher binding affinity towards the target proteins as compared to the drugs used previously for the treatment of *Mycobacterium tuberculosis* infections. Additional hydroxyl groups in ligands with lower binding energies may establish hydrogen bonds with the target protein, indicating a desired interaction. The pi-sigma bond imparts stabilising charges to the ligand, enabling it to intercalate into the binding sites of the receptor. Alkyl and pi-alkyl linkages also aid ligands in interacting hydrophobically in the binding pocket of the receptor.

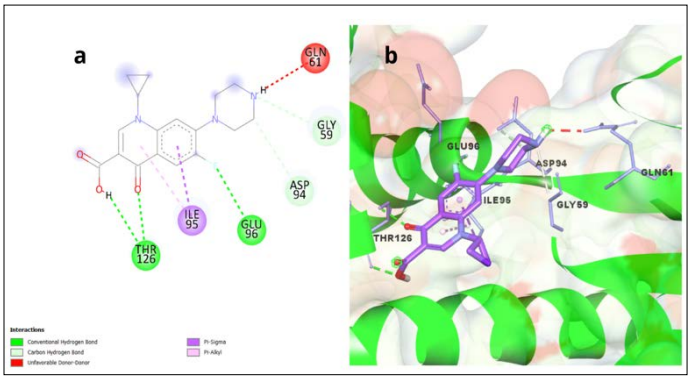


Figure 1: (a and b) 2D interaction plot of Ciprofloxacin docked in the binding pockets of Rv3679 protein. 3D representation showing the position of Ciprofloxacin within the binding site of Rv3679 protein

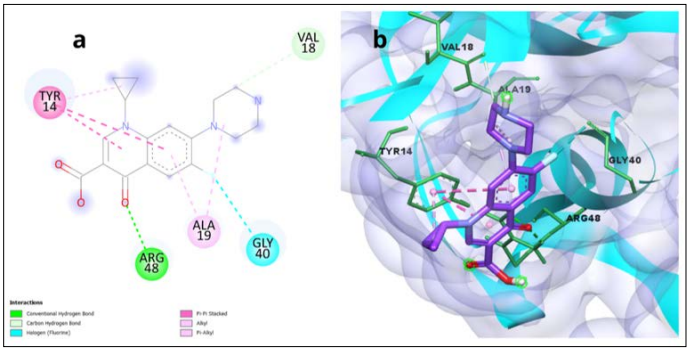


Figure 2: (a and b) 2D interaction plot of Ciprofloxacin docked in the binding pockets of Rv2397c protein. 3D representation showing the position of Ciprofloxacin within the binding site of Rv2397c protein

Ciprofloxacin belonging to fluoroquinolone showed a binding energy of −6.7 kcal/mol with Rv3679 protein and exhibited a variety of interactions including Conventional hydrogen bonds with THR126 and GLU96 residue, Carbon Hydrogen bonds with ASP94 and GLY 59, Pi-Sigma bond with ILE 95, Pi-Alkyl bond and Unfavourable Donor-Donor bond with ILE 95 and GLN 61, respectively and are presented in Figure 1. A similar binding score of -6.6 kcal/mol was observed for protein Rv2397c. Figure 2 shows various interactions consisting of ARG 48 Conventional hydrogen bond, a Carbon hydrogen bond with VAL 18, Halogen bond with GLY 40, Pi-Pi stacked bond with TYR 14, Alkyl and Pi- Alkyl bond with ALA 19 residue. It was demonstrated that this family of synthetic antibiotics could have a variety of anti-infective properties, including antiviral activity. An in silico study performed on interaction of SARS-CoV-2 targets: S-, E- and TMPRSS2 proteins revealed that ciprofloxacin could interact and potentially inhibit crucial SARS-CoV-2 proteins with a binding affinity of −6.8 kcal/mol which is similar to the results obtained in the current study [52].

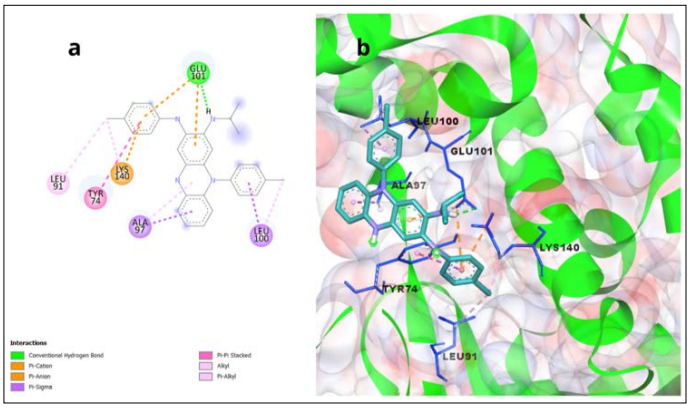


Figure 3: (a and b) 2D interaction plot of Clofazimine docked in the binding pockets of Rv3679 protein. 3D representation showing the position of Clofazimine within the binding site of Rv3679 protein

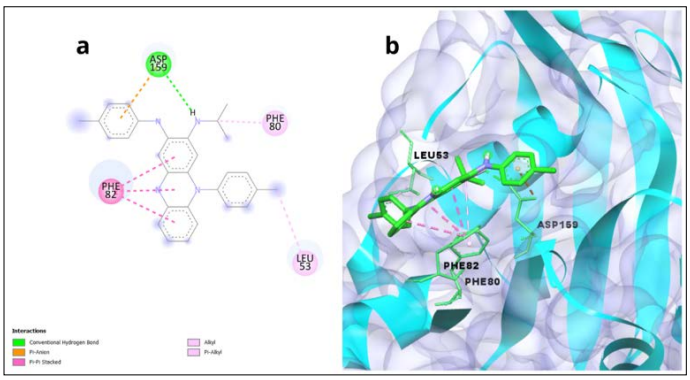


Figure 4: (a and b) 2D interaction plot of Clofazimine docked in the binding pockets of Rv2397c protein. 3D representation showing the position of Clof4azimine within the binding site of Rv2397c protein

Clofazimine docked with Rv3679 protein shows a binding affinity of -7.5 kcal/mol. From Figure 3 it can be seen that Clofazimine exhibited Convention hydrogen bond with GLU 101, Pi- Cation and Pi- Anion bonds with LYS 140 residue and Pi- Pi Stacked bond with TYR 74. Further, it has also shown Pi- Sigma bonds and Alkyl bonds with ALA 97 and LEU100 residues as well as Pi- Alkyl bonds with LEU 91. A binding affinity of -7.6 kcal/mol was observed when Clofazimine was docked against Rv2397c protein. As represented in Figure 4, the ligand Clofazimine formed Conventional hydrogen bond and Pi- Anion bond with ASP 159 residue, a Pi- Pi stacked bond with PHE 82 and Alkyl and Pi- Alkyl bonds with PHE80 and LEU 53 exhibiting varied interactions. According to a molecular docking study conducted by Ajmal et al., clofazimine was found to bind to the hydrophobic pocket near to the drug site II in human serum albumin (HAS). Clofazimine caused a rise in the protein's helical structure while very slightly altering its tertiary structure. It also interacted significantly with HSA, inducing secondary structure in the protein and minor changes in protein topology [53].

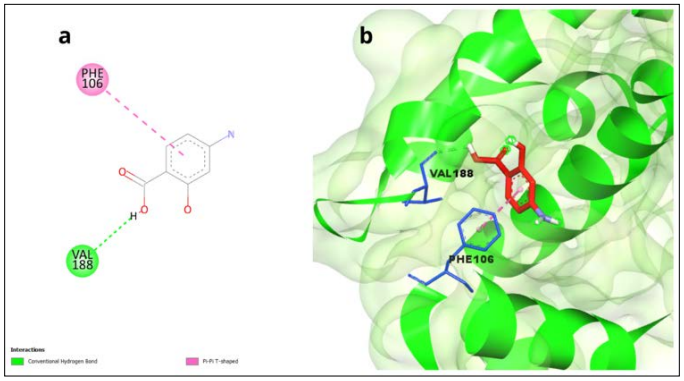


Figure 5: (a and b) 2D interaction plot of 4- Amino Salicylic acid docked in the binding pockets of Rv3679 protein. 3D representation showing the position of 4- Amino Salicylic acid within the binding site of Rv3679 protein

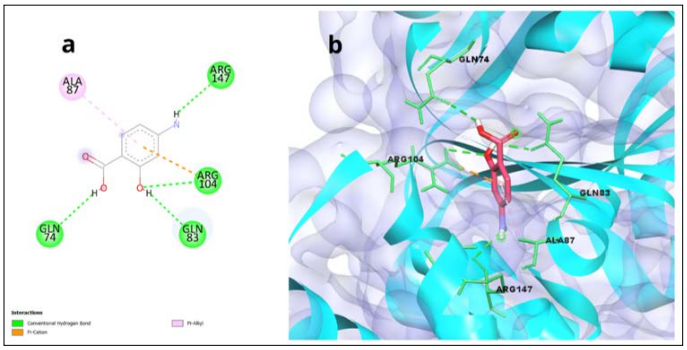


Figure 6: (a and b) 2D interaction plot of 4- Amino Salicylic acid docked in the binding pockets of Rv2397c protein. 3D representation showing the position of 4- Amino Salicylic acid within the binding site of Rv2397c protein

4- Amino Salicylic acid is an aminobenzoic acid and antitubercular agent exhibiting a binding score of -5.7 kcal/mol when docked against Rv3679 protein of *Mycobacterium tuberculosis*. Figure 5 represents the 2D and 3D interactions of the ligand including a Conventional hydrogen bond and Pi- Pi T- shaped bond with VAL 188 and PHE 106 amino acid residues, respectively. It showed a binding affinity of -5.4 kcal/mol with protein Rv2397c and formed Convention hydrogen bond with GLN 74, GLN 83 and ARG 104. In addition to this, Pi- Cation and Pi- Alkyl bonds were observed with ARG 104 and ALA 87, respectively. The above-mentioned interactions are represented in Figure 6.

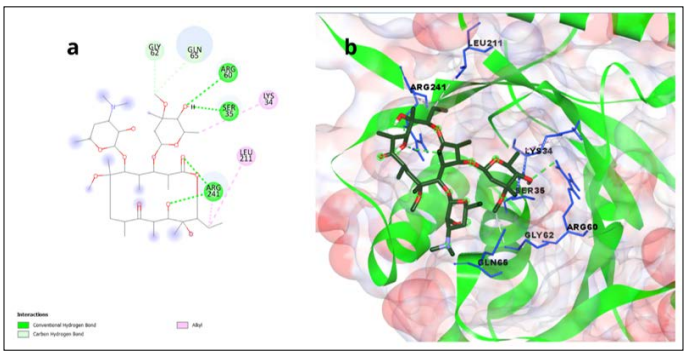


Figure 7: (a and b) 2D interaction plot of Clarithromycin docked in the binding pockets of Rv3679 protein. 3D representation showing the position of Clarithromycin within the binding site of Rv3679 protein

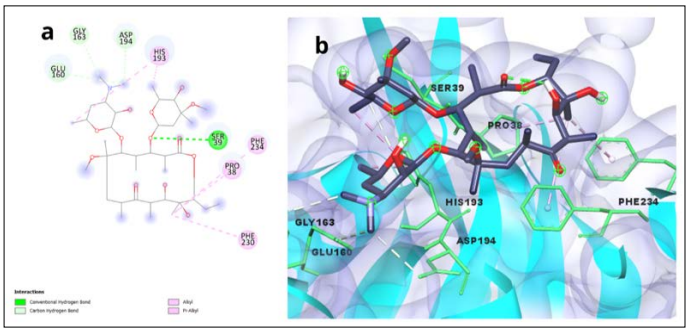


Figure 8: (a and b) 2D interaction plot of Clarithromycin docked in the binding pockets of Rv2397c protein. 3D representation showing the position of Clarithromycin within the binding site of Rv2397c protein

Molecular docking analysis of Clarithromycin showed a binding energy of -6.8 kcal/mol with Rv3679 protein. Figure 7 represents the 2D and 3D interaction of the ligand displaying interactions consisting of Conventional hydrogen bonds, Carbon hydrogen bond and Alkyl bond. Conventional hydrogen bonds were formed with ARG 60, ARG 241 and SER 35 residues, Alkyl bonds with LYS 34 and LEU 211 and Carbon hydrogen bonds with GLY 62 and GLN 65. Similarly, Clarithromycin presented a binding affinity of -6.1 kcal/mol with Mycobacterial protein Rv2397c. As represented in Figure 8, Clarithromycin formed Conventional hydrogen bond with SER 39, Carbon hydrogen bonds with GLY 163, GLU 160 and ASP 194 residues. Further, interactions such as Alkyl and Pi- Alkyl bonds with PHE 230, PRO 38, PHE 234 and HIS 193 residues were also observed. Clarithromycin falls under the drug macrolide antibiotic. In a molecular docking analysis of anti-helicobacterial pylori antibiotics against a proton pump inhibitor, Clarithromycin showed a binding affinity of -6.9 kcal/mol which is similar to the results obtained in the current study indicating that it can be a good candidate for anti-tubercle medications [54].

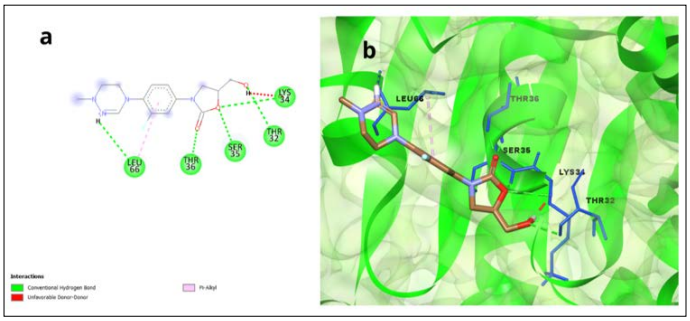


Figure 9: (a and b) 2D interaction plot of Delpazolid docked in the binding pockets of Rv3679 protein. 3D representation showing the position of Delpazolid within the binding site of Rv3679 protein

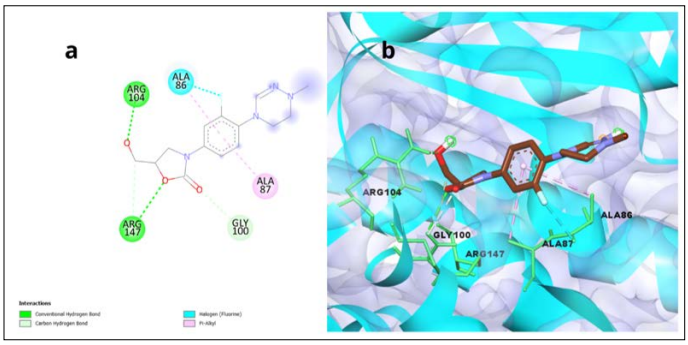


Figure 10: (a and b) 2D interaction plot of Delpazolid docked in the binding pockets of Rv2397c protein. 3D representation showing the position of Delpazolid within the binding site of Rv2397c protein

Delpazolid when docked with Mycobacterial protein Rv3679 displayed a binding affinity of -6.4 kcal/mol. According to Figure 9, Delpazolid formed Conventional hydrogen bonds with LEU 66, THR36, SER35, THR 32 and LYS 34 residues. It also exhibited a Pi- Alkyl and Unfavourable Donor- Donor bond with LEU 66 and LYS 34 residues, respectively. Further, it represented a binding energy of -6.4 kcal/mol after docking with another protein Mycobacterial protein Rv2397c. The interaction presented by Delpazolid here include a Convention hydrogen bond with ARG 147 AND ARG 104, Carbon Hydrogen bond with GLY 100, a Halogen (Fluorine) and Pi- Alkyl bond with ALA 86 and a Pi- Alkyl bond with ALA 87 residue as shown in Figure 10. Delpazolid, an Oxazolidinone has also been included as an established target (repurposing) in Phase 2 clinical trials of treatment for Tuberculosis and the result of the current study is in accordance with this as good binding affinity was exhibited by Delpazolid against Rv 3679 and Rv 2397c proteins of *Mycobacterium tuberculosis* [55].

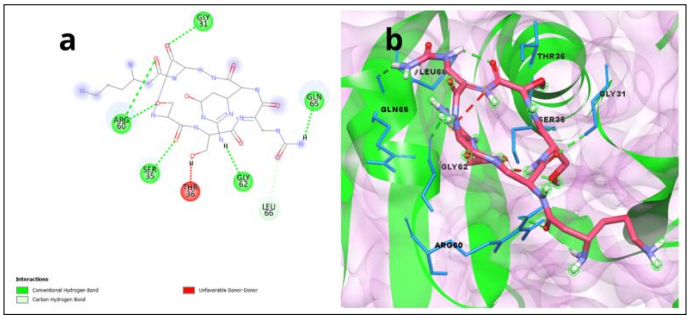


Figure 11: (a and b) 2D interaction plot of Viomycin docked in the binding pockets of Rv3679 protein. 3D representation showing the position of Viomycin within the binding site of Rv3679 protein

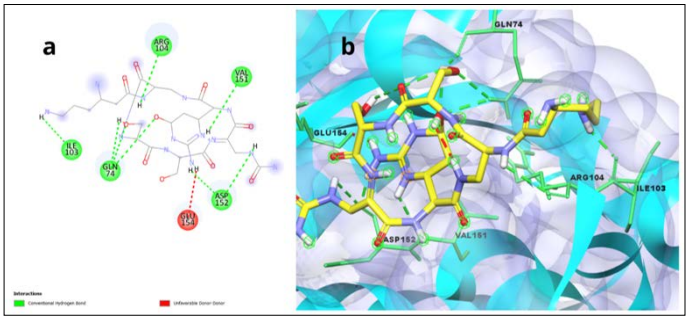


Figure 12: (a and b) 2D interaction plot of Viomycin docked in the binding pockets of Rv2397c protein. 3D representation showing the position of Viomycin within the binding site of Rv2397c protein

Viomycin presented a binding score of -7 kcal/mol with Rv3679 protein of *Mycobacterium tuberculosis*. As shown in Figure 11, Viomycin forms Convention hydrogen bonds with GLN 65, GLY 62, SER 35, ARG 60 and GLY 31 residues. A Carbon hydrogen bond and an Unfavorable Donor- Donor bond was formed with LEU 66 and THR 36, respectively. Post docking analysis of Viomycin with Rv2397c protein displayed a binding energy of -6.4 kcal/mol. Figure 12 represents interactions including Conventional hydrogen bonds with ARG 104, VAL 151, ILE 103, GLN 74 and ASP 152 residues and also an Unfavorable Donor- Donor bond with GLU 154.

Conclusion

Over the last decade or two, there has been a resurgence of interest in Tuberculosis medication research, which has resulted in some notable scientific discoveries. The present TB-drug pipeline incorporates unique chemical scaffolds and a wide range of targets. Despite these breakthroughs in chemotherapeutic treatment, TB eradication remains a global concern. As a result, in addition to the previously described scientific activities, it is critical to identify new TB targets that are not only vital during host infection but are also susceptible to pharmacological suppression. Tuberculosis is a potentially fatal disease caused by several strains of *Mycobacterium tuberculosis*. For several cellular activities, this bacterium significantly relies on transport via the cell membrane. As previously indicated, anions are essential for many biological activities, including Mtb survival. The goal of this work was to offer detailed information regarding proteins involved in anion transport mechanisms. This network has around 34 such proteins. Physiochemical metrics such as molecular weight, instability index, GRAVY, and so on were calculated using various online tools, functional motifs and 3D structural information were obtained, which may provide light on the biological function of the proteins. Based on various physiochemical and biological characteristics such as structural and functional analysis it was revealed that that proteins Rv2397c and Rv3679 showed the best qualifying results. Furthermore, Molecular docking of these two proteins was carried out with six selected ligands and with drugs currently used for treatment of Tuberculosis. Results showed remarkable binding energy of the selected ligands when compared to the drugs used for treatment, high number of conventional hydrogen bonding was also observed with these two proteins corresponding to promising intermolecular protein and drug interactions. Factors such as number of cysteine residues, pI, Aliphatic index, Instability Index, GRAVY and structural folds suggest that these proteins can be selected as reliable drug targets. In the future, we should uncover more reasonable and effective therapeutic targets to produce new drug candidates with higher efficacy.

Data Availability Statement

All data generated or analyzed during this study are included in this published article.

Author Contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical statement

No animals were harmed during this study.

References

1. Tuberculosis Reports (1916) The Lancet 188: 1120-1121.
2. de Oliveira MCB, Balan (2020) The atp-binding cassette (Abc) transport systems in mycobacterium tuberculosis: Structure, function, and possible targets for therapeutics. Biology (Basel) 9: 1-25.
3. Braibant M, Gilot P, Content J (2000) The ATP binding cassette (ABC) transport systems of Mycobacterium tuberculosis.[erratum appears in FEMS Microbiol Rev 2002 Mar;26(1):109]. FEMS Microbiol Rev 24: 449-467.
4. Szumowski JD, Adams KN, Edelstein PH, Ramakrishnan L (2012) Antimicrobial Efflux Pumps and Mycobacterium tuberculosis Drug Tolerance: Evolutionary Considerations 81-108.
5. Choi GE, Eom SH, Jung KH, Son JW, Shin AR, et al. (2010) CysA2: A candidate serodiagnostic marker for Mycobacterium tuberculosis infection. Respirology 15: 636-642.
6. Camacho LR, Ensergueix D, Perez E, Gicquel B, Guilhot C (1999) Identification of a virulence gene cluster of Mycobacterium tuberculosis by signature-tagged transposon mutagenesis. Mol Microbiol 34: 257-267.
7. Kim SH, Lee NE, Lee JS, Shin JH, Lee JY, et al. (2016) Identification of mycobacterial antigens in human urine by use of immunoglobulin g isolated from sera of patients with active pulmonary tuberculosis. J Clin Microbiol 54: 1631-1637.
8. Forrellad MA, Klepp LI, Gioffré A, García JS, Morbidoni HR, et al. (2013) Virulence factors of the mycobacterium tuberculosis complex. Virulence 4: 3-66.
9. Rengarajan J, Bloom BR, Rubin EJ (2005) Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. Proc Natl Acad Sci U S A 102: 8327-8332.
10. Lew JM, Kapopoulou A, Jones LM, Cole ST (2011) TubercuList – 10 years after. Tuberculosis 91: 1-7.
11. https://mycobrowser.epfl.ch/
12. Kapopoulou A, Lew JM, Cole ST (2011) The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis 91: 8-13.
13. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) The Proteomics Protocols Handbook. Proteomics Protoc Handb 571-608.
14. Kaur A, Pati PK, Pati AM, Nagpal AK (2020) Physico-chemical characterization and topological analysis of pathogenesis-related proteins from Arabidopsis thaliana and Oryza sativa using in-silico approaches. PLoS One 1-15.

15. Roy S, Maheshwari N, Chauhan R, Sen NK, Sharma A (2011) Structure prediction and functional characterization of secondary metabolite proteins of Ocimum. Bioinformation 6: 315-319.
16. Gupta R, Brunak S (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput 322: 310-322.
17. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, et al. (2021) The InterPro protein families and domains database: 20 years on. Nucleic Acids Res 49: D344-354.
18. Yang H, Chi H, Zhou WJ, Zeng WF, Liu C, et al. (2018) PSite: Amino Acid Confidence Evaluation for Quality Control of de Novo Peptide Sequencing and Modification Site Localization. J Proteome Res 17: 119-128.
19. Sudha R, Katiyar A, Katiyar P, Singh H, Prasad P (2019) Identification of potential drug targets and vaccine candidates in Clostridium botulinum using subtractive genomics approach. Bioinformation 15: 18-25.
20. Chakkyarath V, Shanmugam A, Natarajan J (2021) Prioritization of potential drug targets and antigenic vaccine candidates against Klebsiella aerogenes using the computational subtractive proteome-driven approach. J Proteins Proteomics 12: 201-211.
21. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, et al. (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res 46: W296-303.
22. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10: 845-858.
23. Saikia A, Palherkar D, Hiremath L (2021) In silico analysis and structural prediction of a hypothetical protein from Leishmania major. Biomed Biotechnol Res J 5: 320-320.
24. Colovos C, Yeates TO (1993) Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci 2: 1511-1519.
25. Kim S, Chen J, Cheng T, Gindulyte A, He J, et al. (2021) PubChem in 2021: New data content and improved web interfaces. Nucleic Acids Res 49: D1388-1395.
26. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25: 1605-1612.
27. Dallakyan S, Olson AJ (2015) Small-molecule library screening by docking with PyRx. Methods Mol Biol 1263: 243-250.
28. Morris GM, Lim-Wilby M (2008) Molecular docking. Methods Mol Biol 443: 365-382.
29. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26: 283-291.
30. Shehadi IA, Rashdan HRM, Abdelmonsef AH (2020) Homology modeling and virtual screening studies of antigen MLAA-42Protein: Identification of noveldrugcandidates against leukemia-an in silico approach. Comput Math Methods Med 16: 2020:8196147. https://pubmed.ncbi.nlm.nih.gov/32256683/#article-details
31. Wiederstein M, Sippl MJ (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35: 407-410.
32. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27: 343-350.
33. Pramanik K, Ghosh PK, Ray S, Sarkar A, Mitra S, et al. (2017) An in silico structural, functional and phylogenetic analysis with three dimensional protein modeling of alkaline phosphatase enzyme of Pseudomonas aeruginosa. J Genet Eng Biotechnol 15: 527-537.
34. Pradeep NV, Vidyashree KG, Lakshmi P (2012) In silico Characterization of Industrial Important Cellulases using Computational Tools 4: 8-15.
35. Sahay A, Piprodhe A, Pise M (2020) In silico analysis and homology modeling of strictosidine synthase involved in alkaloid biosynthesis in catharanthus roseus. J Genet Eng Biotechnol 18(1). https://jgeb.springeropen.com/articles/10.1186/s43141-020-00049-3.
36. Ikai A (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88: 1895-1898.
37. Khoo KK, Norton RS (2011) Role of Disulfide Bonds in Peptide and Protein Conformation. Amin Acids, Pept Proteins Org Chem 5: 395-417.
38. Podell S, Gribskov M (2004) Predicting N-terminal myristoylation sites in plant proteins. BMC Genomics 5: 37.
39. Zheng J, Knighton DR, Taylor SS, Xuong N□H, Sowadski JM, et al. (1993) Crystal structures of the myristylated catalytic subunit of cAMP□dependent protein kinase reveal open and closed conformations. Protein Science. Cambridge University Press 2: 1559-1573.
40. Olsen HB, Kaarsholm NC (2000) Structural effects of protein lipidation as revealed by Lys (B29)-myristoyl, des(B30) insulin. Biochemistry 39: 11893-11900.
41. Ardito F, Giuliani M, Perrone D, Troiano G, Muzio L Lo (2017) The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review). Vol. 40, International Journal of Molecular Medicine. Spandidos Publications 271-280.
42. Rabizadeh S, Sears C (2008) New horizons for the infectious diseases specialist: How gut microflora promote health and disease. Curr Infect Dis Rep 10: 92-98.
43. Neish AS (2009) Microbes in Gastrointestinal Health and Disease. Vol., Gastroenterology. W.B. Saunders 136: 65-80.
44. Hooper LV, Bry L, Falk PG, Gordon JI (1998) Host-microbial symbiosis in the mammalian intestine: Exploring an internal ecosystem. BioEssays 20: 336-43.
45. Guarner F, Malagelada JR (2003) Gut flora in health and disease. Vol., Lancet. Elsevier Limited 361: 512-519.
46. Alaidarous M (2020) In silico structural homology modeling and characterization of multiple N-terminal domains of selected bacterial Tcps. PeerJ 3: 8:e10143.
47. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, et al. (2019) STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47: D607-613.
48. Barbosa MA, Martins MCL (2018) Peptides and proteins as biomaterials for tissue regeneration and repair. Woodhead Publishing.
49. Buxbaum E (2015) Fundamentals of protein structure and function, second edition. Fundam Protein Struct Funct Second Ed Jan 1: 1-521.
50. Degrève L, Fuzo CA, Caliri A (2014) Extended secondary structures in proteins. Biochim Biophys Acta - Proteins Proteomics 1844: 384-388.
51. Murad HAS, Alqurashi TMA, Hussien MA (2022) Interactions of selected cardiovascular active natural compounds with CXCR4 and CXCR7 receptors: a molecular docking, molecular dynamics, and pharmacokinetic/toxicity prediction study. BMC Complement Med Ther 22: 1-22.
52. Hasan MR, Chowdhury SM, Aziz MA, Shahriar A, Ahmed H, et al. (2021) In silico analysis of ciprofloxacin analogs as inhibitors of DNA gyrase of Staphylococcus aureus.

Informatics Med 26: 100748.
53. Ajmal MR, Zaidi N, Alam P, Nusrat S, Siddiqi MK, et al. (2016) Insight into the interaction of antitubercular and anticancer compound clofazimine with human serum albumin: spectroscopy and molecular modelling 35: 46-57.
54. Uthansingh K, Kumari R, Pati GK, Behera MK, Sahu MC, et al. (2021) Molecular Docking of Anti Helicobacter pylori Antibiotics and Proton Pump Inhibitor: A Single Center survey. J Pure Appl Microbiol 15: 2103-2116.
55. Ejalonibu MA, Ogundare SA, Elrashedy AA, Ejalonibu MA, Lawal MM, et al. (2021) Drug discovery for mycobacterium tuberculosis using structure-based computer-aided drug design approach. Int J Mol Sci 22: 13259.

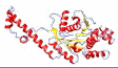
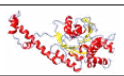
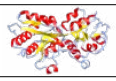
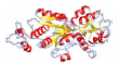
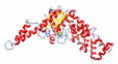
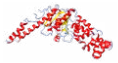
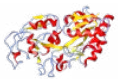
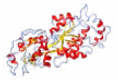
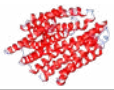
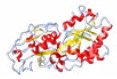
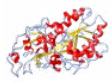
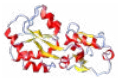
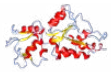
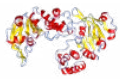
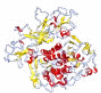
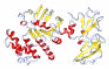
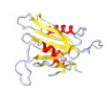
Supplementary file

Anion Transport Proteins and their functions

Protein Identifier	Protein	Functions
Rv2684	arsA	Transport of arsenic compounds across the membrane (export).
Rv2685	arsB	
Rv3578	arsB2	
Rv2643	arsC	
Rv2397c	cysA	Active transport across the membrane of multiple sulfur-containing compounds.
Rv2399c	cyst	
Rv2398c	cysW	
Rv1857	modA	Active transport of molybdenum into the cell across the membrane (import).
Rv1858	modB	Part of the binding-protein-dependent transport system modABC for molybdenum; responsible for energy coupling to the transport system.
Rv1859	modC	
Rv1860	modD	Alanine and proline rich secreted protein Apa (fibronectin attachment protein) (immunogenic protein MPT32) (antigen MPT-32) (45-kDa glycoprotein) (45/47 kDa antigen)
Rv2329c	narK1	Excretion of nitrite produced by the dissimilatory reduction of nitrate.
Rv1737c	narK2	
Rv0261c	narK3	
Rv0267	narU	
Rv0934	phoS1	Active transport of inorganic phosphate across the membrane (import). This is one of the proteins required for binding-protein-mediated phosphate transport.
Rv0928	phoS2	
Rv0820	phoT	
Rv3301c	phoY1	Transcriptional regulation of active transport of inorganic phosphate across the membrane.
Rv0821c	phoY2	
Rv0545c	pitA	Low-affinity inorganic phosphate transport across the membrane.
Rv2281	pitB	Phosphate transport.
Rv0930	pstA1	Active transport of inorganic phosphate across the membrane (import); Has ATP-binding ability and ATPase activity.
Rv0936	pstA2	
Rv0933	pstB	
Rv0935	pstC	Active transport of inorganic phosphate across the membrane (import).
Rv0929	pstC2	
Rv0932c	pstS	
Rv2400c	subI	Active transport across the membrane of multiple sulfur-containing compounds, including sulfate and thiosulfate (import).
Rv0143c	-	-
Rv1707	-	-
Rv1739c	-	Sulphate transport across the membrane.

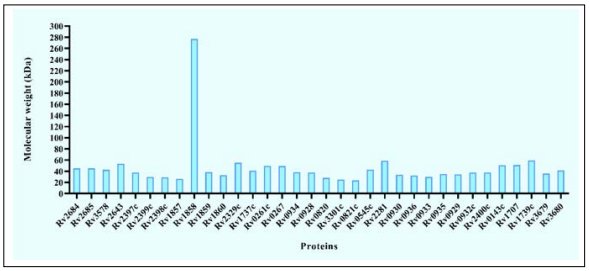
Rv3679	-	Anion-transporting ATPase; supposedly catalyzes the extrusion of undetermined anions [catalytic activity: ATP + H(2)O + undetermined anion(in) = ADP + phosphate + undetermined anion(out)].
Rv3680	-	

Comparison between 3D models generated by SWISS-MODEL and PHYRE servers

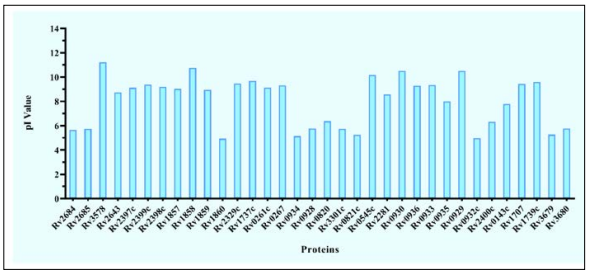
Protein	SWISS-MODEL	PHYRE
Rv3679		
Rv2400c		
Rv0934		
Rv3680		
Rv2397c		
Rv0928		
Rv0821c		
Rv0932c		
Rv3301c		
Rv1857		
Rv0820		
Rv0933		
Rv1859		
Rv1739c		
Rv2399c		
Rv1860		

Protein Name	Tools	Most Favoured Region	Additional allowed regions	Generously allowed regions	Disallowed regions	No. of non-gly & non-pro residues	No. of End residues	No. of glycine residues	No. of proline residues	Total no. of residues	ERRAT	PROSA (Z-score)	Q-mean
Rv2684	Phyre	334 (88.1%)	36 (9.5%)	6 (1.6%)	3 (0.8%)	379	2	33	15	429	72.1429	-5.46	-6.98
	SwissModel	678(91.6%)	48(6.5%)	11(1.5%)	3(0.4%)	740	4	66	30	840	92.0455	-6.37	-5.16
Rv2685	Phyre	335(89.1%)	29(7.7%)	7(1.9%)	5(1.3%)	376	1	35	16	428	67.619	-5.3	-7.77
	SwissModel	665(89.9%)	71(9.6%)	2(0.3%)	2(0.3%)	740	4	68	32	844	91.2302	-6.13	-5.45
Rv3578	Phyre	327(90.3%)	20(5.5%)	8(2.2%)	7(1.9%)	362	2	28	21	413	66.9136	-5.92	-7.24
	SwissModel	640(92.5%)	40(5.8%)	4(0.6%)	8(1.2%)	692	4	54	42	792	90.8016	-6.69	-4.48
Rv2643	Phyre	381(89.9%)	33(7.8%)	6(1.4%)	4(0.9%)	424	2	46	26	498	60.4082	-7.02	-7.03
	SwissModel	260(90.9%)	25(8.7%)	0	1(0.3%)	286	2	32	15	335	90.0312	-5.51	-4.53
Rv2397c	Phyre	277(92.0%)	20(6.6%)	2(0.7%)	2(0.7%)	301	2	33	15	351	75.4491	-7.46	-2.65
	SwissModel	531(93.2%)	37(6.5%)	1(0.2%)	1(0.2%)	570	2	58	28	658	97.1186	-7.96	-0.59
Rv2399c	Phyre	196(80.3%)	34(13.9%)	5(2.0%)	9(3.7%)	244	2	24	13	283	62.0438	-2.94	-9.49
	SwissModel	196(93.3%)	13(6.2%)	1(0.5%)	0	210	2	20	9	241	94.3723	-4.8	-3.59
Rv2398c	Phyre	214(91.5%)	13(5.6%)	0	7(3.0%)	234	1	25	12	272	84.0909	-1.91	-4.9
	SwissModel	197(90.8%)	18(8.3%)	1(0.5%)	1(0.5%)	217	2	24	11	254	93.9024	-1.62	-5.43
Rv1857	Phyre	210(93.3%)	10(4.4%)	3(1.3%)	2(0.9%)	225	1	22	13	261	65.6126	-6.95	-3.82
	SwissModel	183(93.4%)	11(5.6%)	0	2(1.0%)	196	1	16	10	223	89.7193	-7.44	-2.3
Rv1858	Phyre	200(90.5%)	17(7.7%)	2(0.9%)	2(0.9%)	221	2	22	19	264	87.5	-2.99	-4.32
	SwissModel	170(93.4%)	9(4.9%)	2(1.1%)	1(0.5%)	182	2	18	13	215	99.5025	-3.69	-4.26
Rv1859	Phyre	272(88.9%)	27(8.8%)	2(0.7%)	5(1.6%)	306	2	36	25	369	77.8393	-7.72	-2.62
	SwissModel	512(85.0%)	73(12.1%)	13(2.2%)	4(0.7%)	602	4	72	50	728	86.8263	-7.94	-3.36
Rv1860	Phyre	102(42.9%)	92(38.7%)	25(10.5%)	19(8.0%)	238	2	21	64	325	4.90196	-1.8	-21.34
	SwissModel	130(92.9%)	9(6.4%)	0	1(0.7%)	140	1	18	19	178	86.9822	-6.12	0.59
Rv2329c	Phyre	356(81.5%)	58(13.3%)	17(3.9%)	6(1.4%)	437	2	55	21	515	42.7146	0.6	-13.47
	SwissModel	364(95.0%)	16(4.2%)	1(0.3%)	2(0.5%)	283	2	50	15	450	96.3048	-5.71	-4.67
Rv1737c	Phyre	300(90.6%)	21(6.3%)	4(1.2%)	6(1.8%)	331	2	39	23	395	75.969	-3.85	-7.2
	SwissModel	290(91.2%)	23(7.2%)	3(0.9%)	2(0.6%)	318	2	38	22	380	92.5208	-4.21	-6.05
Rv0261c	Phyre	341(86.8%)	37(9.4%)	10(2.5%)	5(1.3%)	393	2	57	17	469	72.9847	-2.74	-10.55
	SwissModel	349(93.6%)	19(5.1%)	4(1.1%)	1(0.3%)	373	3	53	15	444	89.4614	-5	-5.29
Rv0267	Phyre	335(86.6%)	39(10.1%)	9(2.3%)	4(1.0%)	387	2	49	25	463	65.4185	-3.86	-8.86
	SwissModel	329(91.6%)	27(7.5%)	1(0.3%)	2(0.6%)	359	2	46	22		94.6078	-6.52	-4.14
Rv0934	Phyre	269(89.1%)	32(10.6%)	0	1(0.3%)	302	2	43	27	374	84.1226	-8.68	-1.42

	SwissModel	249(91.9%)	22(8.1%)	0	0	271	2	37	23	333	99.0741	-8.89	-0.01
Rv0928	Phyre	265(87.2%)	30(9.9%)	8(2.6%)	1(0.3%)	304	2	48	16	370	81.5341	-6.2	-3.05
	SwissModel	241(91.6%)	21(8.0%)	1(0.4%)	0	263	2	38	16	319	95.7237	-7	-1.54
Rv0820	Phyre	194(87.4%)	17(7.7%)	9(4.1%)	2(0.9%)	222	1	21	14	258	66	-6.38	-2.99
	SwissModel	194(89.0%	23(10.6%)	1(0.5%)	0	218	2	20	14	254	87.8049	-6.95	-2.5
Rv3301c	Phyre	194(96.0%)	7(3.5%)	1(0.5%)	0	202	2	12	5	221	90.1408	-6.48	-1.34
	SwissModel	358(92.7%)	26(6.7%)	0	2(0.5%)	386	4	24	10	424	97.2973	-7	-2.16
Rv0821c	Phyre	188(95.4%)	6(3.0%)	2(1.0%)	1(0.5%)	197	1	10	5	213	98.5366	-6.7	-1.31
	SwissModel	372(96.4%)	10(2.6%)	1(0.3%)	3(0.8%)	386	4	20	8	418	99.7506	-6.64	-1.55
Rv0545c	Phyre	281(81.9%)	41(12.0%)	13(3.8%)	8(2.3%)	343	1	49	24	417	45.7213	-1.37	-11.4
	SwissModel	257(93.5%)	17(6.2%)	1(0.4%)	0	275	1	42	10	328	93.4169	-4.29	-5.48
Rv2281	Phyre	477(88.6%)	39(8.1%)	4(0.8%)	12(2.5%)	482	2	54	14	552	50.4621	-5.51	-8.83
	SwissModel	361(85.3%)	54(12.8%)	4(0.9%)	4(0.9%)	423	3	47	11	484	88.5281	-5.37	-6.36
Rv0930	Phyre	224(83.6%)	28(10.4%)	8(3.0%)	8(3.0%)	268	2	21	17	308	64	-2.97	-7.16
	SwissModel	208(91.2%)	16(7.0%)	3(1.3%)	1(0.4%)	228	2	21	12	263	86.6935	-4.34	-4.63
Rv0936	Phyre	209(82.0%)	32(12.5%)	7(2.7%)	7(2.7%)	255	2	28	16	301	70.2055	-2.57	-9.06
	SwissModel	159(95.8%)	7(4.2%)	0	0	166	1	23	8	198	95.7895	-2.76	-4.67
Rv0933	Phyre	206(86.9%)	28(11.8%)	2(0.8%)	1(0.4%)	237	2	24	13	276	69.685	-6.72	-3.61
	SwissModel	384(90.6%)	34(8.0%)	5(1.2%)	1(0.2%)	424	3	36	26	489	84.4828	-7.57	-2.97
Rv0935	Phyre	237(84.9%)	32(11.5%)	5(1.8%)	5(1.8%)	279	2	38	19	338	71.6463	-0.71	-8.02
	SwissModel	185(91.1%)	15(7.4%)	1(0.5%)	2(1.0%)	203	2	23	14	242	88.3621	-2.53	-5.62
Rv0929	Phyre	246(87.9%)	24(8.6%)	3(1.1%)	7(2.5%)	280	1	27	16	324	68.1529	-2.66	-6.16
	SwissModel	188(89.5%)	19(9.0%)	3(1.4%)	0	210	2	19	10	241	87.4459	-2.15	-5.78
Rv0932c	Phyre	260(87.0%)	28(9.4%)	5(1.7%)	6(2.0%)	299	2	48	21	370	64.8256	-5.89	-3.99
	SwissModel	237(90.8%)	21(8.0%)	2(0.8%)	1(0.4%)	261	2	37	21	321	88.6364	-6.54	-1.59
Rv2400c	Phyre	215(90.0%)	24(10.0%)	0	0	239	2	25	21	287	98.5663	-8.82	0.1
	SwissModel	221(92.5%)	18(7.5%)	0	0	239	2	25	21	287	99.2832	-8.58	0.21
Rv0143c	Phyre	341(87.4%)	36(9.2%)	9(2.3%)	4(1.0%)	390	2	71	29	492	61.25	-6.41	-7.04
	SwissModel	628(88.5%)	70(9.9%)	5(0.7%)	7(1.0%)	710	4	131	50	895	86.8966	-6.75	-4.53
Rv1707	Phyre	358(85.9%)	49(11.8%)	7(1.7%)	3(0.7%)	417	2	48	19	486	74.4641	-4.35	-7.69
	SwissModel	366(91.7%)	27(6.8%)	4(1.0%)	2(0.5%)	399	2	48	19	468	91.704	-5.44	-5.22
Rv1739c	Phyre	439(90.3%0	39(8.0%)	4(0.8%)	4(0.8%)	486	2	44	28	560	73.6842	-7.04	-4.65
	SwissModel	880(91.3%)	72(7.5%)	6(0.6%)	6(0.6%)	964	4	88	54	1110	86.8545	-7.92	-3.46
Rv3679	Phyre	261(89.7%)	19(6.5%)	2(0.7%)	9(3.1%)	291	2	31	16	340	86.4458	-8.73	-2.05
	SwissModel	267(94.3%)	16(5.7%)	0	0	283	2	29	16	330	98.7539	-9.33	0.86
Rv3680	Phyre	307(91.4%)	26(7.7%)	0	3(0.9%)	336	2	29	19	386	88.0952	-8.67	-1.55
	SwissModel	296(91.1%)	28(8.6%)	1(0.3%)	0	325	5	26	19	375	98.615	-9.39	-0.27

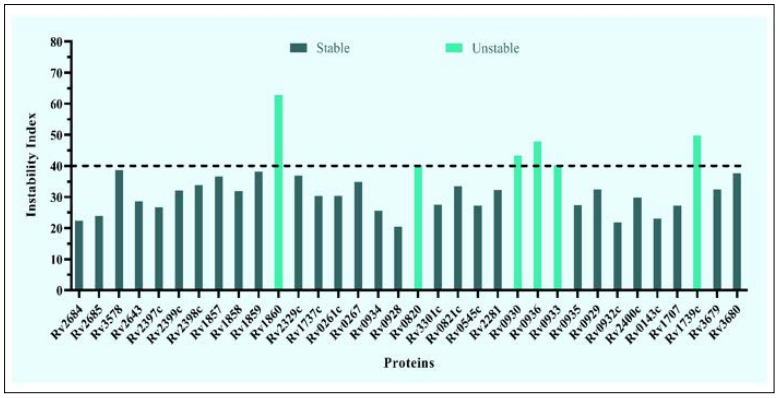


Comparison of Molecular weights (kDa) of all the proteins involved in anion transport network.

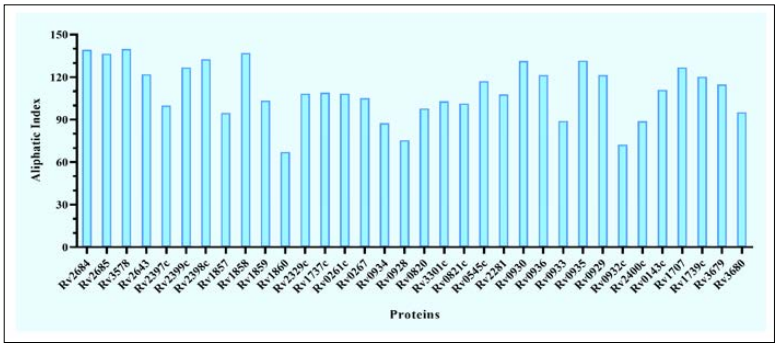


Graphical representation of Theoretical Isoelectric Point (pI) values of the proteins.

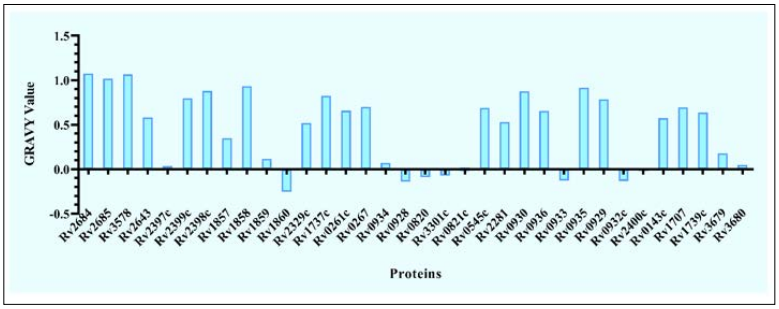
Citation: Vikas Jha, Arpita Singh, Nimisha Rane, Vrushali Dhamapurkar, Kabir Thakur, et al. (2022) Computational Analysis of Anion Transporter Proteins Revealed Rv3679 and Rv2397c as Potential Drug Targets for *Mycobacterium tuberculosis*. Journal of Infectious Diseases & Case Reports. SRC/JIDSCR-193.
DOI: doi.org/10.47363/JIDSCR/2022(3)169



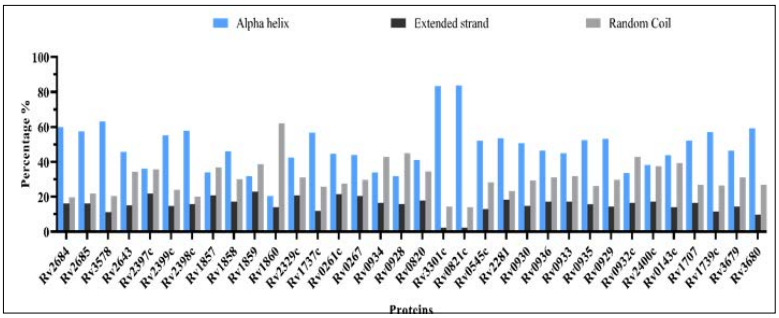
Graphical representation of Instability Index of the proteins. Proteins having Instability Index above 40 are unstable.



Aliphatic Index of the proteins as predicted by ProtParam tool.



Comparison of different GRAVY values of the proteins.



Graphical representation of percentage of alpha helixes, extended strands and random coils of anion transport proteins of Mycobacterium Tuberculosis.