MedCrave
Step into the World of Research
Journal of Lung, Pulmonary & Respiratory Research

Research Article
Open Access
CrossMark

Rv2746c and Rv2881c, a potential drug target of *Mycobacterium tuberculosis* revealed by in-silico investigation of proteins involved in lipid biosynthesis

Abstract

Tuberculosis is a serious disease that requires a greater understanding of its pathophysiology to develop effective treatment strategies. To gain a better understanding of mycobacterial physiology, researchers are focusing on the key components associated with cell wall synthesis. Although mycolic and fatty acids are the primary lipid components of the mycobacterial cell envelope, understanding the proteins involved in the lipid biosynthesis pathway may open up new avenues for fundamental research. This research included a thorough computational examination of proteins from the fatty acid biosynthesis pathways. Rv2881c and Rv2764c are essential genes for lipid synthesis. It is a potential drug target because knocking out these genes has an impact on *Mtb* growth. The study's findings provide researchers with specific cues and concrete information that can be applied in a variety of biotechnological applications.

Keywords: *Mycobacterium tuberculosis*, lipid biosynthesis pathway, molecular docking, physio-chemical analysis, ligand interaction, drug target

Volume 9 Issue 3 - 2022

Vikas Jha,[1] Nimisha Rane,[2] Tanvi Kanade,[3] Dattatray Sawant,[1] Yukta Gharat,[1] Mrunmayi Markam,[1] Kruti Maroo,[3] Mehul Nair,[1] Leena Jagtap,[3] Namira Kazi[3]

[1]National Facility for Biopharmaceuticals, Guru Nanak Khalsa College of Arts, Science, India
[2]School of Biotechnology and Bioinformatics, D Y Patil Deemed to be University, India
[3]Department of Biological Sciences, SVKM's NMIMS Sunadan Divatia School of Science, India

Correspondence: Vikas Jha, National Facility for Biopharmaceuticals, Guru Nanak Khalsa College of Arts, Science & Commerce, Mumbai-400019, Maharashtra, India, Email vikasjjha7@gmail.com

Received: October 21, 2022 | **Published:** November 01, 2022

Abbreviations: *Mtb*, *Mycobacterium tuberculosis*; LAM, lipoarabinomannan; PDIM, phthiocerol dimycocerosate; DAT, diacyltrehaloses [GRAVY], grand average of hydropathicity; Mw, molecular weight; pI, isoelectric point; ESS, extended secondary structures; kcal/mol, kilocalorie per mole; ER, endoplasmic reticulum; cAMP, cyclic adenosine monophosphate; cGMP, guanosine 3',5'-cyclic monophosphate

Introduction

Tuberculosis is a centuries-old disease that has evolved alongside humans in recent years.[1] Despite numerous studies and new treatment modalities over the years, tuberculosis remains a threat to global public health. It is one of the top three infectious causes of death worldwide, according to the World Health Organization. Every year, approximately 10 million people contract tuberculosis (TB), with nearly 1.5 million dying as a result. However, the risk of developing active TB after infection is determined by several factors, one of which is the person's (host) immune response to it. Tuberculosis (TB) has also been observed to thrive wherever there is poverty, overcrowding, and chronic debilitating illness. Although tuberculosis remains one of the leading causes of death in resource-poor countries worldwide, the incidence of tuberculosis has increased in both developing and developed countries due to the emergence of drug-resistant strains.[2]

Mycobacterium tuberculosis (*Mtb*) pathophysiology suggests that various metabolic pathways are important in host-pathogen interaction, pathogenicity, and virulence.[3] These pathways contain a slew of proteins that are essential for *Mtb* survival in the host. *Mtb* has a remarkable ability to persist in the host for long periods with a modified metabolic status and phenotypic drug tolerance without causing symptoms of active tuberculosis.[4] *Mtb* cell wall, which is made up of mycolic acid, glycolipids such as diacyltrehaloses and polyacyltrehaloses, lipomannan, lipoarabinomannan (LAM), mannose-capped-LAM, sulfolipids, and trehalose-6,6'-dimycolate, gives it an advantage in the host.[5] The glycobiology of *Mtb* infection of frontline alveolar macrophages is still unknown. Mycolic acid, Trehalose-6,6'-dimycolate, sulfolipid, phthiocerol dimycocerosate (PDIM), Diacyltrehaloses (DAT), and cholesterol are lipids that contribute to immune response suppression. Glycolipids such as LM, LAM, and Man-LAM modulate the host immune system and aid pathogen survival. Drug resistance occurs as a result of gene mutations in *Mtb*, most commonly as a result of an anti-TB drug.

Mtb changes its fatty acid metabolism to survive in the host, as evidenced by altered cell wall composition. LAM has been identified as mojar virulence factor in tuberculosis because it is a major component of the cell wall and inhibits the activity of phosphatidyl-inositol-3-kinase, limiting phagosome maturation. According to legend, the *M. tuberculosis* H37Rv mutant strain that lacked phthiocerol dimycocerosates (PDIM) could not replicate and resulted in fewer tubercule formations. *Mtb* lipid homeostasis has a significant impact on host lipid metabolism. Lipids in the cell envelope contribute to virulence, survival, and persistence.[6] The lipid metabolism of *Mycobacterium tuberculosis* is quite complicated. It is made up of catabolic and anabolic processes. *Mtb* produces a vast variety of lipids, ranging from simple fatty acids to extraordinarily complex long-chain substances like mycolic acid. About 250 enzymes are specifically specialized for fatty acid metabolism.[7] Fatty acids are stored in the form of Triacylglycerol to counteract adverse situations such as nutrient deficiency.[8] Triacylglycerol is related to the host's ability to support *Mtb* over the long run.[9] Late diagnosis, improper or inadequate administration of effective medications, decreased availability of less toxic, less expensive, and effective medications, prolonged treatment duration, noncompliance with medication regimen, and development of drug resistance are the causes of the failure of TB therapy.

Submit Manuscript | http://medcraveonline.com

©2022 Jha et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and build upon your work non-commercially.

Copyright: ©2022 Jha et al.

Several bioinformatics tools were used in this study to generate biochemical, structural, and functional information about all of the proteins involved in lipid biosynthesis. The goal of this study is to perform Insilco characterization on all of the proteins involved in this pathway and to elucidate their structural and functional details.

Materials and methods

Retrieval of nucleotide and protein sequence

The protein's amino acid sequence was obtained using a Tuberculist. To gather the necessary sequences, several criteria, including gene name, location, etc., were applied.[10] Information about each protein's function was gathered using the Mycobrowser server. A comprehensive proteome and genomic data repository for pathogenic Mycobacterium is called Mycobrowser, commonly known as the Mycobacterial browser. These organisms can be studied genomically and proteomically applying manually curated annotations to the given tools.[11]

Physiochemical characterization

The main structure of 25 proteins from lipid synthesis pathway was determined using ProtParam from ExPASy.[12] It offers details on the physical and biochemical characteristics, such as the 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).[13]

Functional analysis

The CYS REC tool was used to predict the SS-bonding of cysteine residues and the placement of disulphide bridges in the protein sequence. The tool returns the position of the cysteines, their overall quantity, and, if any pairs of cysteines are present, their pattern within the protein sequence.[14] The glycosylation sites were predicted using the NetNglyc service. The NetNglyc server predicts N-Glycosylation sites in human proteins by using artificial neural networks that consider the context of Asn-Xaa-Ser/Thr sequons. To describe the proteins and point out crucial areas, Inter-pro identifies protein families, domains, and functional sites.[15] The amino acid sequence motifs were predicted using Psite.[16]

Human and gut flora non-homology analysis

Proteins with a high degree of homology to the host (the human) may exhibit undesired cross-reactivity and may obstruct the binding of the active sites. So, a non-homology BLAST search against the Human Proteome was carried out. Additionally, using a BLASTp search, the *MTB* liqid synthesis pathway proteins were checked against the proteins of the bacteria in the gut flora.[17]

Homology modelling

Tertiary protein structures give researchers crucial insight into the molecular processes that underlie protein activities, allowing them to design practical experiments like site-directed mutagenesis and examinations of mutation sites that are associated with the disease. The proteins were modelled in three dimensions using the SWISS-MODEL and PHYRE2 servers. A completely automated server called SWISS-MODEL provides comparative automated modelling of 3D protein structures. Similar to this, a 3D modelling service called PHYRE2 uses an Ab initio folding simulation called Poing2 to simulate parts of your proteins that don't match any known structures.

Structure validation

The 3D-modelled structures were validated using the SAVESv6.0 server. It offers several tools, including VERIFY3D, PROCHECK.[18] The models in this study were authenticated using the PROCHECK tool. PROCHECK is a program that examines both the overall geometry of the protein as well as the geometry of each residue to validate the stereochemistry of a protein structure. Therefore, the ProSA-web application is used to calculate the predicted structure's overall model quality (Z-score), with the quality scores shown about all known protein structures. The QMEAN score was calculated to evaluate the structure's quality.[19]

Ligand retrieval

As ligands, Genistein, Coumestrol, Daidzein, Nyasol, and (Z)-hinokiresinol were chosen from the PubChem public chemical structure repository.[20] These compounds' SDF files were downloaded and evaluated.

Protein and ligand preparation

Rv2881c and Rv2764c proteins were chosen from a list of 25 lipid synthesis pathway proteins. The 3D structures of these proteins were prepared for molecular docking analysis using the UCSF Chimera 1.15 tool, which added Kollman charges and polar hydrogen atoms to the protein molecule, eliminated the water molecules, and saved the charged protein molecule in PDB format.[21]

Ligand

As ligands, Genistein, Coumestrol, Daidzein, Nyasol, and (Z)-hinokiresinol were chosen, and their structures were obtained in SDF format from the PubChem databank. The PyRx Virtual Screening Tool was used to create structural variations to optimize and minimise the energy of the chosen ligands.[22]

Molecular docking

Molecular docking is an important method for predicting the prevailing binding interactions of a chosen ligand with a protein with a known three-dimensional structure. The six chosen ligand structures were docked with *Mycobacterium tuberculosis* Rv2881c and Rv2764c proteins using Auto Dock Vina.[23] The Auto Dock Vina software performs bound confirmation prediction based on free binding energies computed using the empirical force field. The docking analysis was carried out using the Auto Dock Vina docking via PyRx Virtual Screening Tool.[24] This analysis aided in the identification of candidate ligands with high protein binding affinities as potential inhibitors.

Results and discussion

Physiochemical characterization

ExPAsy ProtParam Server was used to analyse the physiochemical parameters of 25 *Mtb* (protein involved in the process of lipid biosynthesis) Acyltransferase, Mycolytransferase, and Phospholipid biosynthesis proteins (Table 1). This service has analysis tools such as Compute pI/Mw, which is used to forecast the protein's isoelectric point (pI) and molecular weight (Mw); ProtParam, which is used to estimate numerous physicochemical characteristics; and PeptideMass, which is used to theoretically cleave proteins and calculate the masses of their peptides, as well as any known biological or artifactual posttranslational changes; PeptideCutter, which is used to anticipate cleavage sites of proteases or compounds present in protein sequences; ProtScale is a tool for representing amino acid scales, including hydrophobicity graphs. Protein molecular weight is directly related to the assessment of biomolecule functionality, such as gene and metabolic regulation.[25] Rv2482c has the maximum molecular weight of 88282.96, while Rv2261c has the lowest molecular weight

Citation: Jha V, Rane N, Kanade T, et al. Rv2746c and Rv2881c, a potential drug target of *Mycobacterium tuberculosis* revealed by in-silico investigation of proteins involved in lipid biosynthesis. *J Lung Pulm Respir Res.* 2022;9(3):66–73. DOI: 10.15406/jlprr.2022.09.00283

of 14923.09, according to the analysis. The isoelectric point (pI) of an amino acid is where the net charge is zero since it produces an equal quantity of positive and negative ions.[26] The pI value of acidic amino acids is low, whereas the pI value of basic amino acids is high.[27] Rv2746c was shown to be extremely basic with a value of 10.56, whereas Rv0914c was discovered to be acidic with a value of 4.99. In addition to the pI, the instability index (II) estimates the protein's stability in vitro and in vivo. If a protein's instability index is less than 40, it is stable; if it is greater than 40, it is unstable.[28] Rv2482c, a protein involved in phosphatidylinositol synthesis, is the most unstable protein, with a score of 53.08; on the other side, Rv2612c, a protein involved in phosphatidylinositol synthesis, is the most stable protein, with a score of 24.5. Six (24%) of the proteins chosen were unstable, while 19 (76%) were stable. The aliphatic index is used to indicate the relative volume of a protein inhabited by aliphatic side chains (valine, leucine, isoleucine, and alanine). It is a measure of a protein's thermostability. A high aliphatic index protein is more heat stable.[29] More over 40% of the proteins were thermally stable, with Rv2746c and Rv1822 being among the most stable and Rv3803c being the most unstable. The GRAVY value is used to calculate the average of Hydropathicity. A positive GRAVY value implies hydrophilicity and a non-polar protein, whereas a negative GRAVY value shows hydrophilicity and a polar protein with stronger water interaction.[26] Rv2746c, with a value of 0.83, was clearly the most hydrophobic. Rv3720 has the lowest hydrophilicity (-0.362) and hence will interact with water the best.

Functional analysis

The results of functional analysis using the CYS REC tool indicate that practically every protein possessed cysteine residues, with the exception of three proteins (Table 2): Rv3803c, Rv0045c, and Rv1627c. Cysteine is used in primary structure analysis because it is easily oxidised, resulting in a dimer with a disulphide bond between two cysteines. This property is required for the tertiary and quaternary structure to be stable.[30] Table 2 shows that a total of four residues, cys13, cys315 from protein Rv3804c and cys127, cys132 from Rv1886c, are SS bound. N-glycosylation is one of the most difficult and crucial chemical reactions. The NetNGlyc programme predicted that 8 proteins do not contain any N-glycosylation sites (Table 3). Rv0129c has the most such sites of any protein in the sample. The Psite server, like the NetNGlyc programme, predicts the motifs contained in the selected amino acid sequences. N-myristoylation sites were found in all of the protein sequences studied. Myristoylation is a protein posttranslational modification caused by the addition of myristic acid. The amount of myristoylation sites present relates proportionally to the protein's conformational stability for catalytic activity.[31] All of the proteins were found to have Microbodies C-terminal targeting signal locations. Phosphorylation sites for protein kinase C and Casein Kinase II were discovered for each protein, except Rv2612c. Proteins with a high number of phosphorylation sites are more likely to be regulated often.[32]

Table 1 Physicochemical properties of proteins involved in lipid biosynthesis as predicted by Expasy's protparam program

Protein name	Mol wt	No of Amino acids	pI	Negative residues	Positive residues	Extinction coefficients		Half life			Aliphatic index	Instability index	GRAVY
Rv2289	28576.44	260	6.14	31	29	28420	28670	100	>20	>10	94.56	45.59	-0.152
Rv2881c	32002.89	306	8.76	17	20	55710	55460	100	>20	>10	116.57	28.17	0.75
Rv3804c	35686.2	338	6.08	25	23	73005	72880	30	>20	>10	73.08	40.28	-0.161
Rv1886c	31890.91	300	5.69	21	19	68995	68870	30	>20	>10	73.3	46.47	-0.148
Rv0129c	36771.27	340	5.92	23	21	85830	85830	30	>20	>10	68.97	49.34	-0.253
Rv3803c	31088.89	299	6.13	18	15	83420		30	>20	>10	67.46	26.21	-0.049
Rv0564c	36153.43	341	6.72	34	33	24075	23950	30	>20	>10	93.78	33.88	0.072
Rv2982c	33992	334	5.49	30	25	24450	23950	30	>20	>10	102.37	43.66	0.296
Rv2612c	23251.54	217	9.49	15	21	34740	34490	30	>20	>10	108.34	24.5	0.634
Rv1822	22911.34	209	9.85	8	15	58565	58440	30	>20	>10	120.81	30.07	0.72
Rv2746c	22136.31	209	10.56	10	18	27960	27960	30	>20	>10	129.76	32.89	0.83
Rv1551	69191.98	621	6.1	84	78	77350	77350	100	>20	>10	101.58	40.21	-0.103
Rv2482c	88282.96	789	8.53	97	100	96385	96260	100	>20	>10	90.58	53.08	-0.225
Rv0437c	24227.95	231	9.48	20	25	10220	9970	100	>20	>10	100.61	41.65	0.148
Rv0436c	31218.2	286	10.28	14	31	61670	61420	30	>20	>10	111.29	50.89	0.454
Rv0045c	32099.06	298	5.24	39	25	35980		100	>20	>10	95.27	37.02	-0.106
Rv0914c	43874.9	412	4.99	56	37	52160	52160	30	>20	>10	84.17	31.01	-0.2
Rv1543	36820.88	341	9.07	33	37	17545	17420	30	>20	>10	92.14	31.71	-0.084
Rv1627c	42386	402	5.79	41	35	55265	54890	30	>20	>10	88.43	32.4	0.005
Rv1814	34882.39	300	9.69	21	31	110475	110350	30	>20	>10	95.3	32.1	-0.002
Rv1867	53560.49	494	5.58	60	50	79995	79870	30	>20	>10	85.77	34.58	-0.253
Rv2261c	14923.09	140	8.82	9	11	21095	20970	5.5	3	2	98.36	39.43	0.355
Rv2262c	37258.46	360	9.65	17	26	101325	100950	30	>20	>10	109.56	35.98	0.583
Rv3523	41421.29	394	6.2	40	37	60055	59930	30	>20	>10	83.1	36.58	0.001
Rv3720	46852.84	420	6.22	52	52	72100	71580	30	>20	>10	82.48	39.62	-0.362

Table 2 Number of cysteine residues, their position and disulfide bridge predicted using cys_rec tool

Protein name	No of cysteine	Positions	Prediction	Score
Rv2289	4	CYS 52	Not SS bounded	-20.8
		CYS 65	Probably no SS bounded	-9.4
		CYS 155	Not SS bounded	-22.3
		CYS 255	Not SS bounded	-23.1
Rv2881c	4	CYS 120	Not SS bounded	-29.7
		CYS 163	Not SS bounded	-29.1
		CYS 181	Not SS bounded	-27.1
		CYS 224	Not SS bounded	-20
Rv3804c	3	CYS 13	SS bounded	66.8
		CYS 135	SS bounded	58.5
		CYS 254	Not SS bounded	-46
Rv1886c	3	CYS 127	SS bounded	72.2
		CYS 132	SS bounded	64.9
		CYS 251	Not SS bounded	-52.3
Rv0129c	1	CYS 255	Not SS bounded	-55.5
Rv3803c	No Cysteines			
Rv0564c	3	CYS 26	Probably not SS bounded	-5.5
		CYS 77	Not SS bounded	-40
		CYS 253	Not SS bounded	-30.6
Rv2982c	8	CYS 15	Probably not SS bounded	-13.1
		CYS 159	Probably not SS bounded	-4.6
		CYS 194	Not SS bounded	-35.2
		CYS 202	Not SS bounded	-25.2
		CYS 253	Not SS bounded	-22
		CYS 280	Probably not SS bounded	-11.3
		CYS 289	Not SS bounded	-16.2
		CYS 313	Not SS bounded	-24.5
Rv2612c	5	CYS 59	Not SS bounded	-37.8
		CYS 92	Probably not SS bounded	-8.3
		CYS 102	Not SS bounded	-22.7
		CYS 124	Not SS bounded	-28
		CYS 192	Not SS bounded	-23.6
Rv1822	2	CYS 149	Not SS bounded	-40.7
		CYS 160	Not SS bounded	-27.3
Rv2746c	1	CYS 70	Not SS bounded	-27.3
Rv1551	1	CYS 548	Not SS bounded	-27.3
Rv2482c	1	CYS 395	Not SS bounded	-34.6
Rv0437c	4	CYS 59	Not SS bounded	-29.5
		CYS 86	Not SS bounded	-31.5
		CYS 197	Not SS bounded	-45.5
		CYS 230	Not SS bounded	-20.7
Rv0436c	4	CYS 26	Not SS bounded	-34.4
		CYS 112	Not SS bounded	-32.6
		CYS 210	Not SS bounded	-32.4
		CYS 231	Not SS bounded	-39.7
Rv0045c	No Cysteines			
Rv0914c	4	CYS 78	Not SS bounded	-21.1
		CYS 93	Not SS bounded	-18.7
		CYS 221	Not SS bounded	-29.3
		CYS 309	Not SS bounded	-39.4
Rv1543	2	CYS 112	Not SS bounded	-31.8
		CYS 218	Not SS bounded	-32.7
Rv1627c	No Cysteines			
Rv1814	2	CYS 11	Not SS bounded	-47.5
		CYS 165	Not SS bounded	-27.8
Rv1867	2	CYS 10	Not SS bounded	-23.5
		CYS 321	Not SS bounded	-19.5
Rv2261c	2	CYS 83	Not SS bounded	-27.1
		CYS 137	Not SS bounded	-15.4
Rv2262c	7	CYS 15	Not SS bounded	-19.1
		CYS 114	Not SS bounded	-29.9
		CYS 185	Not SS bounded	-16.6
		CYS 197	Probably not SS bounded	-8.9
		CYS 201	Probably not SS bounded	-13.8
		CYS 332	Probably not SS bounded	-12.5
		CYS 358	Not SS bounded	-18
Rv3523	3	CYS 207	Probably not SS bounded	-29.9
		CYS 214	Not SS bounded	-16.6
		CYS 386	Not SS bounded	-8.9
Rv3720	4	CYS 162	Probably not SS bounded	-6.1
		CYS 200	Not SS bounded	-54.5
		CYS 295	Probably not SS bounded	-10.8
		CYS 356	Not SS bounded	-34.5

Human and gut flora non-homology analysis

Potential pharmaceutical targets that are homologous to the host could have a negative impact on host metabolism. As a result, finding non-homologous proteins is a critical step in predicting therapeutic targets. Non-homology analysis was performed on the four critical genes, and they were all shown to be nonhomologous to the human proteome.[33] Gut microflora plays an important role in metabolism by digesting food particles and preventing dangerous bacteria from colonising the intestine.[34] The human host's interaction with gut microbiota is critical for absorbing poorly digested dietary components, xenobiotic breakdown, and vitamin production.[35] They also confer resistance to opportunistic bacteria and pathogen colonisation.[36] When the healthy gut microbiota population deteriorates, the body's first line of defence against invading microorganisms suffers. It may also result in nutritional deficiencies in the host.[37] Following the homology search, all proteins that were non-homologous to gut flora were shortlisted.

Homology modelling

SWISS-MODEL and PHYRE2 SERVER calculated the sequence's 3D structures. Q-mean, Z-score, and Ramachandran plot analysis were performed to evaluate the predicted models. Each server selects the best modelling template depending on how similar the query sequences are. The tertiary structures of the proteins under research were investigated. SWISS-MODEL used the Z-score and Q-mean score, which should be between -4 and +4, to predict successful models for 8 proteins. PHYRE utilised the same criterion to identify models for 5 proteins. SWISSMODEL outperformed PHYRE, according to the aforementioned statistical data. The Ramachandran plot, which was employed in further investigation of the approved models, revealed that 91.875% of amino acids were in the most favourable zone, while 0.34% were in the prohibited zone. Tertiary structures are considered high grade if 90% of the residues are in the chosen region.[38]

Structure validation

A protein's secondary structure is determined by the pattern of hydrogen bonding, with alpha helices and beta sheets being the most essential types of secondary structures. Alpha-helices are typically seen in water-soluble globular proteins. Protein interactions with other proteins, lipid connections in the cell membrane, and nucleic acid interactions are all common uses for alpha helices. Pharmaceuticals have made use of highly stable alpha helices.[39] The amino acid change seen in alpha helix is mostly connected with protein thermostability. Rv2612c contains a larger proportion of alpha - helix (65.44%). (Table 4) Random coils are regions that have an uneven secondary structure but lack hydrogen bonding. Protein enzymatic turnover, flexibility, and structural changes are all dependent on it.[40] They can have lengths ranging from 4 to 20 residues. It also enhances protein flexibility during enzymatic turnover, for example. They can withstand mutations better than other structures. With a score of 47.69%, Rv2289 has the largest proportion of random coil, while Rv0129c has a value of 47.65%. The build-up of secondary structures that directly interact with atoms in the main chain results in the formation of extended secondary structures (ESS). It helps to stabilise the original secondary structure as well as future tertiary structures. The largest percentage of ESS (29.87%) is found in Rv0437.[41] The SOSUI tool determined that the proteins were 100% soluble. Table 4 illustrates which proteins are soluble.

Molecular docking

A simulation technique known as molecular docking studies the best way to attach a ligand to an active site on a target. This method selects the binding site in the target using 3D coordinates, and the binding affinity of the resulting orientation of the molecule within the binding site, which forms the complex, is determined.[42] The complex created when the involved ligand effectively binds with the active pockets of the target has the largest magnitude negative number (highest binding affinity or lowest binding energy).[5] bioavailable ligands were used in molecular docking with Mycobacterium tuberculosis Rv2746c and Rv2881c lipid synthesis proteins. The binding energies of these ligands for proteins Rv2746c and Rv2881c are shown in Tables 5 & 6, indicating that they have a high affinity for the target proteins. Complexes for Rv2746c protein exhibited binding affinities ranging from 6.8 to 8.1 kcal/mol, whereas complexes for Rv2881c protein had binding affinities ranging from -7.4 to -9.1 kcal/mol. In lower-binding-energy ligands, additional hydroxyl groups may establish hydrogen bonds with the target protein, indicating a favourable interaction. The pi-sigma bond stabilises the ligand, allowing it to intercalate into the binding sites of the receptor. Alkyl and pi-alkyl linkages also aid ligands in engaging hydrophobically in the binding pocket of the receptor.

Table 4 Predicted secondary structure elements using SOPMA Software

Protein	Alpha helix	Extended strand	Beta turn	Random coil	Solubility
Rv2289	36.15%	14.23%	1.29%	47.69%	Soluble
Rv2881c	47.71%	14.38%	7.84%	30.07%	Soluble
Rv3804c	37.87%	13.31%	7.40%	41.42%	Soluble
Rv1886c	37.54%	15.69%	8.62%	38.15%	Soluble
Rv0129c	34.71%	11.47%	6.18%	47.65%	Soluble
Rv3803c	34.78%	15.38%	10.70%	39.13%	Soluble
Rv0564c	48.68%	13.78%	7.92%	29.62%	Soluble
Rv2982c	51.50%	11.08%	6.59%	30.84%	Soluble
Rv2612c	65.44%	9.68%	6.45%	18.43%	Soluble
Rv1822	58.37%	11.00%	2.87%	27.75%	Soluble
Rv2746c	57.89%	11.96%	4.78%	25.36%	Soluble
Rv1551	60.23%	8.21%	6.67%	26.89%	Soluble
Rv2482c	56.53%	10.39%	4.56%	28.52%	Soluble
Rv0437c	24.24%	29.87%	6.49%	39.39%	Soluble
Rv0436c	54.90%	8.39%	4.20%	32.52%	Soluble
Rv0045c	41.28%	18.12%	5.03%	35.57%	Soluble
Rv0914c	43.45%	11.41%	7.04%	38.11%	Soluble
Rv1543	52.49%	10.85%	6.74%	29.91%	Soluble
Rv1627c	37.06%	12.44%	6.97%	43.53%	Soluble
Rv1814	45.33%	17.33%	3.33%	34.00%	Soluble
Rv1867	36.84%	15.38%	6.48%	41.30%	Soluble
Rv2261c	32.14%	21.43%	7.14%	39.29%	Soluble
Rv2262c	41.39%	17.22%	6.39%	35.00%	Soluble
Rv3523	41.62%	11.42%	5.84%	41.12%	Soluble
Rv3720	47.14%	13.81%	5.95%	33.10%	Soluble

Table 5 Binding energy of Rv2746c protein with the selected 5 ligands

Sr.No.	Ligand	Binding Energy n(ΔG) (kcal/mol)
1	Genistein	-7.8
2	Coumestrol	-8.1
3	Daidzein	-7.1
4	Nyasol	-6.8
5	(Z)-hinokiresinol	-7

Table 6 Binding energy of Rv2881c protein with the selected 5 ligands

Sr.No.	Ligand	Binding energy n(ΔG) (kcal/mol)
1	Genistein	-8.7
2	Coumestrol	-9.1
3	Daidzein	-8.5
4	Nyasol	-7.4
5	(Z)-hinokiresinol	-7.9

Genistein, a flavonoid, had a binding energy of 7.8 kcal/mol with Rv2746c protein and exhibited a variety of interactions, including

Conventional hydrogen bonds with THR153 and ARG134 residues, Carbon Hydrogen bonds with LYS150, Pi-Sigma bonds with ASP73 and ASP94, Pi-Alkyl bond, and Unfavourable Donor-Donor bond with PRO95, as shown in Figure 1. Protein Rv2881c has a comparable binding score of -8.7kcal/mol. Figure 2 depicts numerous interactions, including a SER214 Conventional hydrogen bond, a Carbon hydrogen bond with ASP190 and LYS213, an Amide-Pi stacking bond with GLY193 and GLY197, and an Alkyl and Pi-Alkyl bond with LYS212 and LYS213. This synthetic antibiotic family was shown to have a variety of anti-infective activities, including antiviral action. An in-silico investigation of DHFR protein interactions found that Genistein could interact with and potentially inhibit important Dihydrofolate reductase proteins with a binding affinity of 7.75kcal/mol, which is similar to the data obtained in the current study [genstin].

Coumestrol docked with Rv2746c protein exhibits a binding affinity of -8.1 kcal/mol, From Figure 3 it can be noted that Coumestrol demonstrated Convention hydrogen bond with THR153, ARG134, LYS99, Pi- Anion bonds with ASP73, ASP94, ASP98 residue and Pi- alkyl Stacked bond with PRO95. Further, it has also showed Pi-Sigma bonds with LYS150 residues. A binding affinity of -9.1 kcal/mol was reported when Coumestrol was docked against Rv2881c protein. As depicted in Figure 4, the ligand Coumestrol established Conventional hydrogen bond with SER214 and ASP190 residue, Amide - Pi stacking bond with GLY193 and Pi- Alkyl bonds with LYS212 residue. LYS213 residue forms a pi-cation link Zafar et al.[43] used in silico techniques to investigate the influence of structural stability on selective binding of coumestrol to ER-alpha and not to ER-beta.

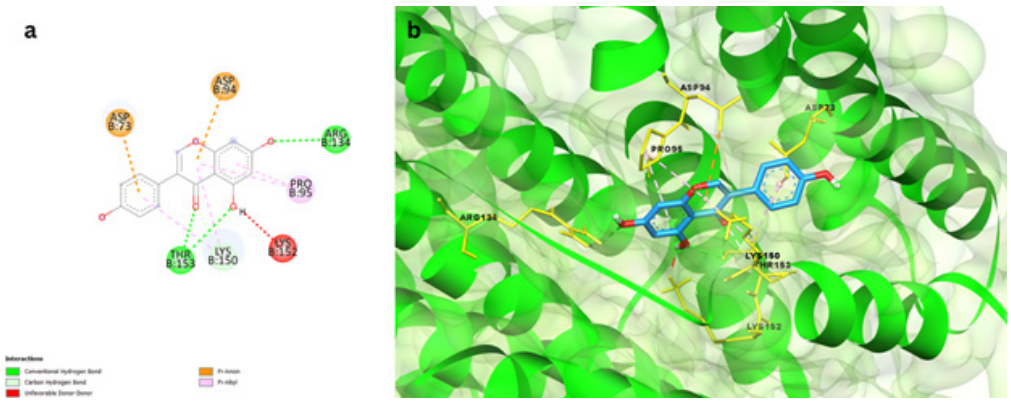

Figure 1 a) 2D interaction plot of Genistein docked in the binding pockets of Rv2746c protein, b) 3D representation showing the position of Genistein within the binding site of Rv2746c protein.

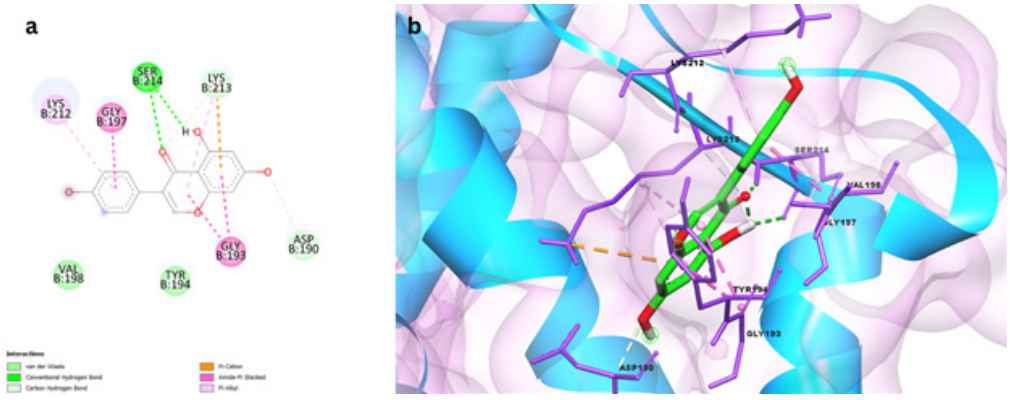

Figure 2 a) 2D interaction plot of Genistein docked in the binding pockets of Rv2881c protein b) 3D representation showing the position of Genistein within the binding site of Rv2881c protein.

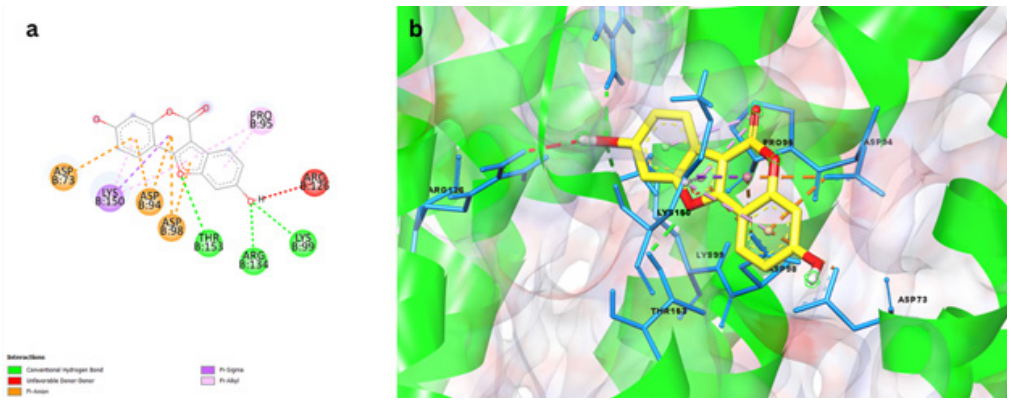

Figure 3 a) 2D interaction plot of Coumestrol docked in the binding pockets of Rv2746c protein, b) 3D representation showing the position of Coumestrol within the binding site of Rv2746c protein.

Citation: Jha V, Rane N, Kanade T, et al. Rv2746c and Rv2881c, a potential drug target of *Mycobacterium tuberculosis* revealed by in-silico investigation of proteins involved in lipid biosynthesis. *J Lung Pulm Respir Res.* 2022;9(3):66–73. DOI: 10.15406/jlprr.2022.09.00283

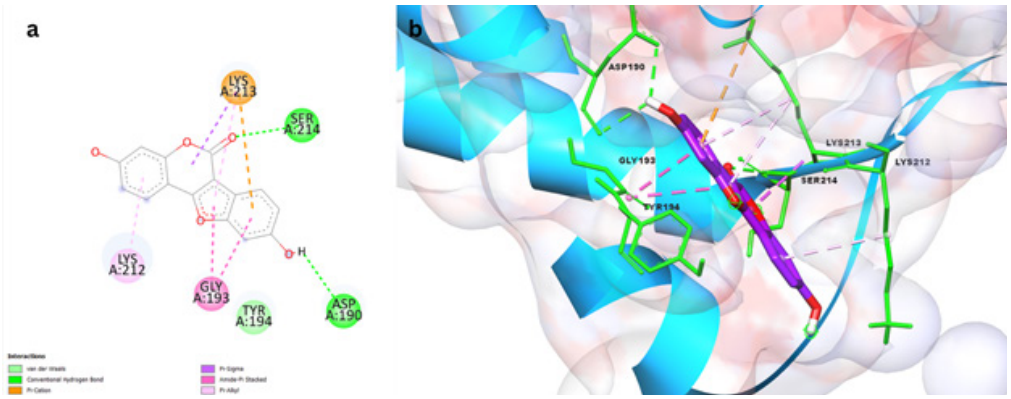

Figure 4 a) 2D interaction plot of Coumestrol docked in the binding pockets of Rv2881c protein, b) 3D representation showing the position of Coumestrol within the binding site of Rv2881c protein.

Conclusion

Over the last decade or so, there has been a renaissance of interest in Tuberculosis medicine research, which has resulted in some important scientific discoveries. The current TB therapeutic pipeline includes novel chemical scaffolds and a diverse set of targets. Despite these advances in chemotherapy, TB eradication remains a global concern. As a result, in addition to the previously mentioned scientific endeavours, it is crucial to uncover additional TB targets that are not only important during host infection but also vulnerable to pharmacological inhibition. Tuberculosis is a potentially lethal disease caused by *Mycobacterium tuberculosis* strains. This bacteria heavily relies on transport through the cell membrane for a variety of biological processes. As previously shown, triacylglycerol are required for many biological activities, including *Mtb* survival. The purpose of this study was to provide thorough information about proteins involved in lipid synthesis pathways. This network has around 25 such proteins. Using several online tools, physiochemical metrics such as molecular weight, instability index, GRAVY, and so on were computed, and functional motifs and 3D structural information were collected, which may provide light on the biological function of the proteins. Proteins Rv2746c and Rv2881c provided the best qualifying findings based on several physiochemical and biological properties such as structural and functional studies. Molecular docking of these two proteins was also performed with five selected ligands and medicines currently used to treat tuberculosis. When compared to the medications used for therapy, the selected ligands had a remarkable binding energy. A high number of traditional hydrogen bonding was also detected with these two proteins, indicating promising intermolecular protein and drug interactions. Number of cysteine residues, pI, Aliphatic index, Instability Index, GRAVY, and structural folds indicate that these proteins can be chosen as reliable therapeutic targets. In the future, we should identify more plausible and effective therapeutic targets in order to develop new drug candidates with more efficacy.

Data availability statement

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

Ethical statement

No animals were harmed during this study.

Acknowledgments

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

Conflicts 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.

References

1. Hirsh AE, Tsolaki AG, Deriemer K, et al. Stable association between strains of Mycobacterium tuberculosis and their human host populations. 2004.
2. World Health Organisation. Rapid Communication: on forthcoming changes to the programmatic management of tuberculosis preventive treatment. 2020;5.
3. Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*. 1998;393(6685):537–544.
4. Maurya RK, Bharti S, Krishnan MY. Triacylglycerols: Fuelling the Hibernating Mycobacterium tuberculosis. *Front Cell Infect Microbiol.* 2019;8:450.
5. Sonawane A, Mohanty S, Jagannathan L, et al. Role of glycans and glycoproteins in disease development by Mycobacterium tuberculosis. *Crit Rev Microbiol*. 2012;38(3):250–266.
6. Ghazaei C. Mycobacterium tuberculosis and lipids: Insights into molecular mechanisms from persistence to virulence. *J Res Med Sci*. 2018;23(1):63.
7. Crellin PK, Luo CY, Morita YS. Metabolism of Plasma Membrane Lipids in Mycobacteria and Corynebacteria. *Lipid Metabolism*. 2013.
8. Deb C, Lee CM, Dubey VS, et al. A novel in vitro multiple-stress dormancy model for mycobacterium tuberculosis generates a lipid-loaded, drug-tolerant, dormant pathogen. *PLoS One*. 2009;4(6):e6077.
9. Maurya RK, Bharti S, Krishnan MY. Triacylglycerols: Fuelling the hibernating mycobacterium tuberculosis. *Front Cell Infect Microbiol*. 2019;9(8):450.
10. Kapopoul A, Lew JM, Cole ST. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis (Edinb)*. 2011;91(1):8–13.
11. Sanoussi CN, Coscolla M, Anyinam BO, et al. Mycobacterium tuberculosis complex lineage 5 exhibits high levels of within-lineage genomic diversity and differing gene content compared to the type strain h37rv. *Microb Genom*. 2021;7(7):000437.

12. Gasteiger E, GattikeA, Hoogland C, et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003;31(13):3784–3788.

13. Kaur A, PK Pati, AM Pati, et al. Physico-chemical characterization and topological analysis of pathogenesis-related proteins from Arabidopsis thaliana and Oryza sativa using in-silico approaches. *PLoS One*. 2020;15(9):e0239836.

14. Skinnider MA, Johnston CW, Gunabalasingam M, et al. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *nature communications*. 2020;11(1).

15. Lam PVN, Goldman R, Karagiannis K, et al. Structure-based Comparative Analysis and Prediction of N-linked Glycosylation Sites in Evolutionarily Distant Eukaryotes. *Genomics Proteomics Bioinformatics*. 2013;11(2):96–104.

16. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35(SUPPL 2):W407–410.

17. Anis Ahamed N, Panneerselvam A, Arif IA, et al. Identification of potential drug targets in human pathogen Bacillus cereus and insight for finding inhibitor through subtractive proteome and molecular docking studies. *J Infect Public Health*. 2021;14(1):160–168.

18. Bagaria A, Jaravine V, Huang YJ, et al. Protein structure validation by generalized linear model root-mean-square deviation prediction. *Protein Sci*. 2012;21(2):229–238.

19. Benkert P, Schwede T, Tosatto SC. QMEANclust: Estimation of protein model quality by combining a composite scoring function with structural density information. *BMC Struct Biol*. 2009;9:35.

20. Kim S, Thiessen PA,Bolton EE, et al. PubChem substance and compound databases. *Nucleic Acids Res*. 2016;4;44(D1):D1202-1213.

21. Madhavi Sastry G, Adzhigirey M, Day T, et al. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des*. 2013;27(3):221–234.

22. Sudhana Saddala M, Jangampalli Adi P. Discovery of small molecules through pharmacophore modeling, docking and molecular dynamics simulation against Plasmodium vivax Vivapain-3 (VP-3). *Heliyon*.2018;4(5):e00612.

23. Forli S, Huey R, Pique ME, et al. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc*. 2016;11(5):905–919.

24. Ferrari IV, Patrizi P. Article development and validation molecular docking analysis of Human serum albumin (HSA).

25. Alberts B, Johnson A, Lewis J, et al. Molecular biology of the cell. *Garland Science*. 2002.

26. Pramanik K, Ghosh PV, Ray S, et al. An in silico structural, functional and phylogenetic analysis with three dimensional protein modeling of alkaline phosphatase enzyme of Pseudomonas aeruginosa. *J Genet Eng Biotechnol*. 2017;15(2):527–537.

27. Kozlowski LP. IPC 2.0: Prediction of isoelectric point and pKadissociation constants. *Nucleic Acids Res*. 2021;49(W1):W285–W292.

28. Gamage DC, Gunaratne A, Periyannan GR, et al. Applicability of Instability Index for *In vitro* Protein Stability Prediction. *Protein Pept Lett*. 2019;26(5):339–347.

29. Prabhu D, Rajamanikandan S, Anusha SB, et al. In silico Functional Annotation and Characterization of Hypothetical Proteins from Serratia marcescens FGI94. *Biol Bull Russ Acad Sci*. 2020;47(4):319–331.

30. Feige MJ, Braakman I, Hendershot LM. CHAPTER 1.1. Disulfide Bonds in Protein Folding and Stability. 2018;1–33.

31. Azevedo R, Silva AMN, Reis CA, et al.In silico approaches for unveiling novel glycobiomarkers in cancer. *J Proteomics*. 2018;171:95–106.

32. Chou MF, Prisic S, Lubner JM, et al. Using Bacteria to Determine Protein Kinase Specificity and Predict Target Substrates. *PLoS One*. 2012;7(12):e52747.

33. Sudha R, Katiyar A, Katiyar P, et al. Identification of potential drug targets and vaccine candidates in Clostridium botulinum using subtractive genomics approach. *Bioinformation*. 2019;15(1):18–25.

34. Pickard JM, Zeng MY, Caruso R, et al. Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunol Rev*.2017;279(1):70–89.

35. Dahiya D, Nigam PS. the gut microbiota influenced by the intake of probiotics and functional foods with prebiotics can sustain wellness and alleviate certain ailments like gut-inflammation and colon-cancer. *Microorganisms*. 2022;10(3):665.

36. Pant A, Maiti TK, Mahajan D, et al. Human Gut Microbiota and Drug Metabolism. *Microb Ecol*. 2022;23:1–15.

37. Saim A, Sakat M. Structure Prediction and Characterization of Uncharacterized ABC Transporter ATP-Binding Protein Rv0986 of Mycobacterium tuberculosis (Strain ATCC 25618 / H37Rv). *BioRxiv*. 2020;5(23):112680.

38. Ullah M, Hira J, Ghosh T, et al. A bioinformatics approach for homology modeling and binding site identification of triosephosphate isomerase from Plasmodium falciparum 3D7. *J Young Pharma*. 2012;4(4):261–266.

39. Oklejas V, Zong C, Papoian GA, et al. Protein structure prediction: Do hydrogen bonding and water-mediated interactions suffice? *Methods*. 2010;52(1):84–90.

40. Der Lee RV, Buljan M, Lang B, et al. Classification of intrinsically disordered regions and proteins. *Chem Rev*. 2014;114(13):6589–6631.

41. Degrève L, Fuzo CA, Caliri A. Extended secondary structures in proteins. *Biochim Biophys Acta*. 2014;1844(2):384–388.

42. Baptista R, Bhowmick S, Shen J, et al. Molecular docking suggests the targets of anti-mycobacterial natural products. *Molecules*. 2021;26(2):475.

43. Zafar A, Ahmad S, Naseem I. Insight into the structural stability of coumestrol with human estrogen receptor α and β subtypes: A combined approach involving docking and molecular dynamics simulation studies. *RSC Advances*. 2015;5(99):81295–81312.