Research Article

CrossMark

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/modification pathway of *Mycobacterium tuberculosis* revealed the diagnostic potential of Rv0054 and Rv3644c

Abstract

Mycobacterium tuberculosis being a causative agent of *tuberculosis* is a powerful pathogen that has evolved to survive within the host. There are certain metabolic pathways that play a vital role in host-pathogen interaction, pathogenicity and virulence which is indicated by the pathophysiology of *Mycobacterium tuberculosis* (MTB). The pathways involve many proteins that are vital for MTB survival in the host. One such pathway is DNA replication, repair, recombination, and restriction/modification pathway. The study of DNA repair mechanisms in *Mycobacterium tuberculosis* has progressed more slowly than in other bacteria due to the technological challenges in dealing with a slow-growing pathogen. In this study, by utilizing immunoinformatic analysis & homology modelling approach, the evaluation of the proteins involved in this pathway was carried out which can lead to the discovery of potential drug targets, vaccine candidates as well as various diagnostic markers.

Keywords: *In-silco*, *Mycobacterium tuberculosis*, homology modelling, diagnostic markers, vaccine candidates

Vikas Jha,[1] Sathi Maiti,[2] Dattatray Sawant,[1] Darpan Kaur,[1] Sankalp Kasbe,[1] Abhishek Kumar,[1] Badal Saiya,[1] Shloka Shukla,[1] Simeen Rumani,[1] Mrunmayi Markam[1]
[1]National Facility for Biopharmaceuticals, G. N. Khalsa College, India
[2]Department of Five Years Integrated Course in Bioanalytical Sciences, GNIRD, G.N. Khalsa College, India

Correspondence: Vikas Jha, National Facility for Biopharmaceuticals, G. N. Khalsa College, Matunga, Mumbai, Maharashtra, India, Email vikasjjha7@gmail.com

Abbreviations: ACC, auto cross-covariance; CXC, colorectal cancer; MTB, *Mycobacterium tuberculosis;* MDR-TB, multidrug-resistant tuberculosis; INH, isoniazid; RFP, rifapim; RNI, Reactive nitrogen intermediates; GRAVY, grand average of hydropathicity index; HMM, hidden markov model; INF, interferon; IP-10, IFN-gamma-inducible protein 10; Th1, type 1 T helper

Introduction

Tuberculosis (TB) is a highly infectious disease caused by *Mycobacterium tuberculosis* (MTB) that has posed a constant threat throughout human history due to its severe potential implications. The genus Mycobacterium is believed to have originated more than 150 million years ago by Bazin. *Mycobacterium tuberculosis* is one of the leading infectious agents in most cases of tuberculosis occurring worldwide. MTB having a very ancient origin has survived for over 70,000 years and is currently infecting nearly 5.8 million people around the globe; with new cases of TB each year, nearly one-third of the world's population are carriers of the TB bacillus and are at a chance for creating the active infection at a global scale by WHO. Because of its infectious nature, vigorous immunological response, slow progressive development, and the need for long-term care combating and preventing tuberculosis have been a persisting problem in epidemic history, with the prevalence of multi-drug resistant forms, including its major social repercussions. Multidrug-resistant tuberculosis (MDR-TB), produced by isolates of MTB are resistant to at least two of the most effective anti-TB medications, isoniazid (INH) and rifapim (RFP), hence, still posing complications for TB eradication by Lin and Flynn.

With the continued growth of multi- and extensively resistant MTB strains, undermining the administration of this major catastrophe, new MTB treatments are urgently needed, and metabolic pathways present enticing and potentially powerful targets. There are certain metabolic pathways that play an essential role in host-pathogen interaction,

pathogenicity and virulence which is indicated by the pathophysiology of *Mycobacterium tuberculosis* (MTB). The pathways involve a large number of proteins that are vital for MTB survival in the host.[1] Further exploration of various proteins from these pathways, may aid in the development of vaccine candidates, identification of newer drug targets, and even diagnostic markers. One such pathway is DNA replication, repair, recombination and restriction/modification pathway.[2] MTB encodes a complex set of proteins that guarantees chromosomal DNA replication and repair. Generally, in bacterial systems, chromosomal replication is carried out by a massive multi-protein replisome that provides high effectiveness and precision in the synthesis of the leading and lagging DNA strands.[3] The helicase-primase core complex, and the clamp loader complex are the three catalytic centers that work together to accomplish this. The helicase-primase consists of the DNA-B helicase that uncoils the two DNA strands and the DNA-G primase.[4] On the lagging strand, it synthesizes short RNA primers, which the replicative DNA polymerase, Pol IIIa, uses to start replication.[5] Most of the constituent proteins in the replisome perform specialty tasks such as DNA unwinding, RNA primer synthesis, clamp loading, and DNA synthesis.[6] MTB being an intracellular pathogen is exposed to several highly DNA-damaging attacks in vivo, mainly from antimicrobial reactive oxygen and nitrogen intermediates produced by the host (RNI).[7] As a result, having DNA damage repair and reversing pathways that can effectively reverse the damaging impact of these problems is pertinent for bacterial survival. The study of DNA repair mechanisms in MTB has progressed more slowly than in other bacteria due to the technological challenges in dealing with such a slow-growing pathogen. Hence, rather than practical experiments, most conclusions regarding *Mycobacterium tuberculosis* DNA repair are still focused on *insilico* methodologies.[8] Studying the biochemistry of this pathway can strengthen the knowledge about this disease and can help us consider the DNA replication and repair machinery as a source of new targets for anti-TB drug development.

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al.   **191**

With over 3924 open reading frames, MTB has the second-largest bacterial genome sequence.[9] Moreover, the DNA replication, repair, recombination, and restriction/modification pathway is structurally, but not functionally reported in many studies. In this study, various computational tools have been used to generate biochemical, structural and functional information about all the proteins in this pathway and also to test the applicability of immunoinformatic analysis & homology modelling approach for MTB proteins involved in this pathway.

Methodology

Retrieval of the protein sequence

For the present study 69 amino acid sequences (Table 1) of *Mycobacterium tuberculosis* involved in DNA replication, repair, recombination, and restriction/modification pathway were retrieved in FASTA format using Mycobrowser database (https://mycobrowser.epfl.ch/).[10] Mycobrowser is an exhaustive genomic and proteomic information repository for pathogenic mycobacteria. It provides physically curated annotations and relevant tools to work with genomic and proteomic investigation of these organisms.

Prediction of antigenicity of the protein sequences: For the prediction of antigenic properties of the proteins VaxiJen server was used (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). VaxiJen server is based on auto cross-covariance (ACC) and auto transformation of protein sequences into invariable vectors of major amino acid properties which was used to evaluate the antigenicity of the protein sequence. The VaxiJen server's algorithm is based on the method of alignment of sequence and to analyse the physicochemical properties of the protein to identify them as antigenic or non-antigenic.

Physicochemical characterization and solubility prediction: The Physicochemical properties of the proteins such as the number of amino acids, pI value, molecular weight, molecular formula, number of atoms, extinction coefficients, estimated half-life, instability index, total number of positively and negatively charged residues, aliphatic index, and grand average of hydropathicity (GRAVY) were analyzed using Expasy's ProtParam server (https://web.expasy.org/protparam/).[11] The total length and solubility of the protein were predicted using the SOSUI server (https://harrier.nagahama-i-bio.ac.jp/sosui/). CYS_REC server (http://www.softberry.com/berry.phtml?topic=cys_rec) was used to analyze the presence of Cysteine residues in the proteins and their bonding patterns.[12]

Secondary structure prediction: For the prediction of Secondary structures based on the primary sequence of the protein SOPMA (Self-Optimized Prediction Method with Alignment) server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) was used. With the help of this server, information about the protein such as α-helix, 3_{10} helix, π helix, β bridge, Extended strands, β-turn, bend region and random coil was obtained. When an unknown protein is entered into this server, it searches for all the proteins with similar properties and evolution available in the database.[13]

Tertiary structure prediction & homology modelling: Homology modelling of the proteins was performed using the Phyre2 tool.[14] (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) This tool aids in generating automated 3D structures based on comparative methods. The protein sequences were submitted, and the most suitable 3D models were selected on the basis of the highest value of Ramachandran's favored region. Ramachandran Plot analysis was done using Structural Assessment tool of SWISS-MODEL server (https://swissmodel.expasy.org/interactive). Ramachandran plot analysis provides information on the total number of amino acid residues found in the favourable, allowed, and disallowed regions.[15] The accuracy and stereochemical quality of the models were analyzed using ERRAT, Verify3D, PROCHECK, and PROVE from PROCHECK's server (https://servicesn.mbi.ucla.edu/PROCHECK/). The predicted LG score, MaxSub score and Z-score of the proteins were anticipated using ProQ and ProsA server respectively. ProSA (Protein Structure Analysis) is a popular web tool that is used to verify 3D models of various proteins for possible errors. The output of this tool is an overall quality score in the form of a plot that describes the scores of all experimentally derived protein chains present in the protein data bank.[16] ProQ (Protein Quality Predictor) is a tool that is used to predict the quality of a protein model based on neural networks which utilises several structural features such as atom-atom contacts, residue-residue contacts, and solvent-accessible surface area. It differs from other model predictors in a way in which it is optimized to find correct models instead of native structures. MaxScore and LGscore are the two outcome quality measures of this tool where MaxScore has a range between 0 and 1 (insignificant and significant respectively) and LGscore is the negative log P value which indicates structural similarity.

B epitope prediction and scanning protein for IFN epitopes: To anticipate linear B-cell epitopes based on antigen protein sequence features, we used a range of methods, including amino acid scales and HMMs (http://tools.iedb.org/bcell/result/). With a default threshold value of 0.350, the IEDB's Bepipred method was used to predict linear B-cell epitopes from the conserved region of the proteins.[17–19] The Immune Epitope Database (IEDB Emini's) surface accessibility prediction tool was used to predict surface epitopes from the conserved region using the default threshold value of 1.0.[20] The antigenic sites were detected using the Kolaskar and Tongaonkar antigenicity approach, with a default threshold value of 1.025.[21] BepiPred-2.0 web server was used for predicting B-cell epitopes from antigen sequences.[22] Parker Hydrophilicity Prediction[23] was used to determine which regions of a protein were on the surface and hence predict the antigenicity of the protein. Chou and Fasman Beta turn prediction was performed for the protein in order to obtain beta turn areas in the query protein, as beta turns play an important role in antigenicity production.[24] IFNepitope (https://webs.iiitd.edu.in/raghava/ifnepitope/scan.php) was used for predicting Interferon (INF) gamma inducing regions in the protein. The IFN gamma epitope server predicts IFN epitopes based on a machine learning algorithm called Support Vector Machine by building overlapping sequences. For in silico immunization development, the epitopes with positive outcomes for the IFN-γ reaction were chosen.[25]

Molecular docking analysis

The idea behind docking is to elicit an effective immune response between an antigen and an antigen receptor. H dock is a web server used for protein protein docking. In this study, both IFN gamma epitopes and B cell epitopes of Rv0054 & Rv3644c were docked with IP-10 protein (Crystal structure of mouse) using H-dock tool (http://hdock.phys.hust.edu.cn/). The workflow of the HDOCK web server is divided into four stages: (1) data input where input of either PDB structure or sequence in FASTA format is accepted (2) sequence similarity search is conducted against the PDB database for the receptor and ligand molecules and for protein input, the HH suite package is used (3) structure modeling is done using MODELLER in which the selected templates undergo sequence alignment using ClustalW. and (4) FFT-based global docking in which priority is given to user-input structures- Based on HDOCKlite, in which an improved shape wise scoring function is used based on putative binding modes by Yan. IP-10 (interferon-gamma-inducible protein) is a chemokine that belongs to the Colorectal cancer (CXC) family

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/ modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al.　192

and is involved in the pathophysiology of a variety of immunological and inflammatory responses. It also has antifibrotic properties and is a powerful angiostatic factor. IP-10's biological effects are mediated via interactions with the G-protein-coupled receptor CXCR3, which is found on Th1 lymphocytes. IP-10 is thus a potential candidate for anti-inflammatory molecule structure-based rational drug design by Jabeen.

Results and discussion

MTB has long been characterized by a greater mortality rate, and it is predicted to account for 1.4 million tuberculosis fatalities today, second only to the Human Immunodeficiency Virus (HIV) among infectious diseases. Computational studies have provided multiple key molecular-level insights into the repair of damaged DNA by monofunctional DNA glycosylases, the first enzymes to function in the base excision repair pathway that targets non bulky nucleobase modifications.[27] Despite of the advancements in medical technologies and extensive research, MTB remains a serious public health concern as currently there are no effective drugs or vaccines established for its treatment. The in-silico technique is the initial step in developing a vaccine, and it is critical that the target protein be chosen correctly. To improve our understanding of host-microbe interactions and obtain a better understanding of their mechanisms, we used in-silico analysis to create a potential new vaccine candidate. Immunoinformatics is a relatively new field that has the potential to speed up immunology research. Computational models are already playing an important role not just in steering the selection of key experiments, but also in the creation of novel testable hypotheses through extensive analysis of complicated immunology data that could not be accomplished using traditional methods alone. The assessment of the various proteins involved in the selected pathway was implemented using this approach which might lead to the discovery of prospective therapeutic targets, vaccine candidates as well as diagnostic indicators.

Prediction of antigenicity of the protein sequences

The sequences were studied for their antigenicity by VaxiJen server keeping the default threshold, and it was found that out of all the sequences, 62 amino acid sequences were antigenic, and they were marked on the basis of the antigenicity score (Table 1). The study revealed several MTB proteins which are potential candidates for the investigation of both cellular and humoral immune responses in an infected host. The antigenicity scores ranged from the highest being 0.7400 to the lowest being 0.2138.

Physicochemical characterization and Solubility prediction

The physicochemical characteristics of all the proteins involved in DNA replication, repair, recombination, and restriction/modification pathway were computed using Expasy's ProtParam and the resulting values for a number of amino acid residues, theoretical isoelectric point pI, instability index, aliphatic index, and Grand Average Hydropathicity (GRAVY) for each protein were evaluated (Table 2). The theoretical pI values for these proteins range from 4.51 to 10.66. The theoretical isoelectric point can be described as the pH at which a particular molecule carries no net electrical charge, however, protein carries a net negative charge above the isoelectric point and the protein carries a net positive charge, below isoelectrical point. For the pI values, we found that 60% of protein sequences were acidic (pI<7) and 40% of them were basic in nature (pI>7). The acidic nature is helpful as the pathogen is able to tolerate and survive the acidity of phagolysosomes during chronic infection inside the host.[28] This data can also be very advantageous for the advancement of the

cushion framework for the refinement of protein by an isoelectric centering strategy.[13] Further, an instability index of >40 indicates a stable protein and <40 indicates unstable protein.[29,30] In our study, 38 sequences (Table 2) had an instability index of less than 40 implying that those protein sequences were stable. The values of the instability index for the proteins range from 20.7 to 93.23. The aliphatic index of a protein is the relative volume occupied by the aliphatic side chains like Alanine (Ala), Valine (Val), Isoleucine (Ile), and Leucine (Leu). The proteins with high aliphatic index values are more thermally stable and in the case of globular proteins, it could be regarded as a positive factor for the increase of thermostability. The Aliphatic index of the protein's ranges from 64.33 to 133.69 (Table 2), which shows steadiness across a wide range of temperatures. It might very well be viewed as a positive factor for the expansion of the thermostability of globular proteins.[31,32] The values of the extinction coefficient specify the amount of light a particular protein can absorb at a certain wavelength and the values of the molar extinction coefficient can be procured if the amino acid composition of the protein is known by Conn. The extinction coefficient values obtained for proteins from ProtParam are of two different types one assuming all pairs of Cys residues form cystines and the other assuming all Cys residues are reduced. The two types of extinction coefficient values for these proteins are almost in the similar range and in some proteins both the values are same; the values range from 4470 to 57885. The grand average hydropathy (GRAVY) value for a protein is calculated as the sum of hydropathy values of all amino acids, divided by the number of residues in the sequence. The soluble nature of the protein helps in DNA packaging in the cell which forms the protein moiety of nucleoprotein. GRAVY values for these proteins lie within -0.474 to 1.023 (Table 2). From all the selected sequences, 63 protein sequences had the grand average hydropathicity (GRAVY) score less than 0 implying that these protein sequences were hydrophilic in nature and soluble in water and could be a good choice for drug designing as it has lower value (Table 2).[33,34] This information could clarify whether the protein is globular (hydrophilic) or membranous (hydrophobic) and might give insights into the localization of proteins.

Functional analysis of the proteins includes prediction of disulphide bonds and transmembrane region. After distinguishing between membrane and soluble proteins from amino acid sequences, the SOSUI server was used to predict the transmembrane helices for soluble proteins. Though there were cysteine residues present in some of the protein sequences no evidence was found for the presence of disulphide bonds. However, the gene Rv2836c (dinF) shows presence of 12 transmembrane regions among all selected proteins which is an important factor to be considered for the efficacy of drug and, disulphide bridges play an important role in determining thermostability of the protein molecule.[35]

Secondary structure prediction

On analyzing the proteins using the SOPMA tool (Table 3), the presence of alpha helix is obtained to be dominant in the structures, followed by random coil, extended strand, and beta turns (Mukesh, Prathap, and Sabitha 2013). The default parameters with window width set at 17; similarity threshold set at 8 and division factor set as 4 were considered for the secondary structure prediction.[9] In an alpha helix chain, the hydrogen bond forms between the hydrogen atom in the polypeptide backbone amino group of another amino acid that is four amino acids farther along the chain and the oxygen atom in the polypeptide backbone carbonyl group in one amino acid which holds the stretch of amino acids in a right-handed coil. In an alpha helix every helical turn shows presence of 3.6 amino acid residues. The side chains or R groups of the polypeptide protrude out from the α-helix

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/ modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al.　**193**

chain which are not involved in the H bonds which help to maintain the alpha helix structure. The models for properties of individual residues and short segments of a polypeptide chain in a random coil contributes a framework for interpreting experimental NMR data for non-native protein conformations.[36] Proteins typically have compact, globular shapes assembled by combination of beta sheets. However, they require reversals in the direction of their chain to obtain these compact shapes. The reverse turn is also known as the hairpin bend or beta turn that provides a common structure which satisfies the requirement of chain reversal. Another type of structure responsible for chain reversals which are more complicated than reverse turns are loops. Although they do not show the presence of any periodic structures like beta sheets and alpha helices, they are well defined most of the time and are rigid.[37]

Table I The following table depicts the genes involved in DNA replication, repair, recombination and restriction/modification pathway and prediction of Antigenicity Score using Vaxijen Server

Rv ID	Gene name	Description	Score
Rv1317c	alkA	DNA-3-methyladenine glycosidase II	0.4047
Rv2836c	dinF	DNA-damage-inducible protein F	0.503
Rv1329c	dinG	probable ATP-dependent helicase	0.4408
Rv3056	dinP	DNA-damage-inducible protein	0.4442
Rv1537	dinX	probable DNA-damage-inducible protein	0.4677
Rv0001	dnaA	chromosomal replication initiator protein	0.3864
Rv0058	dnaB	DNA helicase (contains intein)	0.4691
Rv1547	dnaE1	DNA polymerase III, α subunit	0.4199
Rv3370c	dnaE2	DNA polymerase III α chain	0.4387
Rv2343c	dnaG	DNA primase	0.5114
Rv0002	dnaN	DNA polymerase III, β subunit	0.6
Rv3711c	dnaQ	DNA polymerase III e chain	0.4202
Rv3721c	dnaZX	DNA polymerase III, γ (dnaZ) and τ (dnaX)	0.5522
Rv2924c	fpg	formamidopyrimidine-DNA glycosylase	0.5759
Rv0006	gyrA	DNA gyrase subunit A	0.5057
Rv0005	gyrB	DNA gyrase subunit B	0.6662
Rv2092c	helY	probable helicase, Ski2 subfamily	0.4458
Rv2101	helZ	probable helicase, Snf2/Rad54 family	0.4457
Rv2756c	hsdM	type I restriction/modification system DNA methylase	0.4363
Rv2755c	hsdS	type I restriction/modification system specificity determinant	0.4672
Rv3296	lhr	ATP-dependent helicase	0.41
Rv3014c	ligA	DNA ligase	0.5559
Rv3062	ligB	DNA ligase	0.5447
Rv3731	ligC	probable DNA ligase	0.5334
Rv1020	mfd	transcription-repair coupling factor	0.4799
Rv2528c	mrr	restriction system protein	0.4972
Rv2985	mutT1	MutT homologue	0.5417
Rv1160	mutT2	MutT homologue	0.5312
Rv0413	mutT3	MutT homologue	0.4964
Rv3589	mutY	probable DNA glycosylase	0.4257
Rv3297	nei	probable endonuclease VIII	0.4816
Rv3674c	nth	probable endonuclease III	0.3086
Rv1316c	ogt	methylated-DNA-protein-cysteine methyltransferase	0.4402
Rv1629	polA	DNA polymerase I	0.5243
Rv1402	priA	putative primosomal protein n' (replication factor Y)	0.4462
Rv3585	radA	probable DNA repair RadA homologue	0.519
Rv2737c	recA	recombinase (contains intein)	0.5066
Rv0630c	recB	exodeoxyribonuclease V	0.5005
Rv0631c	recC	exodeoxyribonuclease V	0.4653
Rv0629c	recD	exodeoxyribonuclease V	0.4881
Rv0003	recF	DNA replication and SOS induction	0.5034
Rv2973c	recG	ATP-dependent DNA helicase	0.5281
Rv1696	recN	recombination and DNA repair	0.516
Rv3715c	recR	RecBC-Independent process of DNA repair	0.4549
Rv2736c	recX	regulatory protein for RecA	0.647
Rv2593c	ruvA	Holliday junction binding protein DNA helicase	0.6785
Rv2592c	ruvB	Holliday junction binding protein	0.4449
Rv2594c	ruvC	Holliday junction resolvase, endodeoxyribonuclease	0.5832
Rv0054	ssb	single strand binding protein	0.7372
Rv1210	tagA	DNA-3-methyladenine glycosidase I	0.4588
Rv3646c	topA	DNA topoisomerase	0.6152
Rv2976c	ung	uracil-DNA glycosylase	0.2138

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/ modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al. **194**

Table Continued...

Rv ID	Gene name	Description	Score
Rv1638	uvrA	excinuclease ABC subunit A	0.5413
Rv1633	uvrB	excinuclease ABC subunit B	0.4409
Rv1420	uvrC	excinuclease ABC subunit C	0.4965
Rv0949	uvrD	DNA-dependent ATPase I and helicase II	0.3221
Rv3198c	uvrD2	putative UvrD	0.4615
Rv0427c	xthA	exodeoxyribonuclease III	0.5531
Rv0071	-	group II intron maturase	0.4251
Rv0861c	-	probable DNA helicase	0.3907
Rv0944	-	possible formamidopyrimidineDNA glycosylase	0.4262
Rv1688	-	probable 3-methylpurine DNA glycosylase	0.6476
Rv2090	-	partially similar to DNA polymerase I	0.4737
Rv2191	-	similar to both PolC and UvrC proteins	0.4231
Rv2464c	-	probable DNA glycosylase, endonuclease VIII	0.2796
Rv3201c	-	probable ATP-dependent DNA helicase	0.4516
Rv3202c	-	similar to UvrD proteins	0.4949
Rv3263	-	probable DNA methylase	0.4812
Rv3644c	-	similar in N-t	0.7400

Table 2 Physiochemical characteristics of all the proteins obtained from ExPasy's ProtParam

Rv ID	Molecular weight	Theoretical pI	Instability index	Aliphatic index	GRAVY
Rv1317c	53710.49	9.56	36.59	89.23	-0.097
Rv2836c	44737.86	10.41	22.31	133.69	1.023
Rv1329c	70135.47	6.1	34.9	95.5	0.038
Rv3056	37562.17	8.31	43.92	96.18	-0.037
Rv1537	49075.62	5.84	38.21	94.43	-0.062
Rv0001	56548.64	5.45	41.82	86.67	-0.381
Rv0058	96916.74	8.71	42.57	89.11	-0.308
Rv1547	129322.96	5.5	32.68	89.7	-0.199
Rv3370c	116483.67	7.09	39	87.83	-0.122
Rv2343c	69562.95	6.38	40.78	82.28	-0.213
Rv0002	42113.09	4.76	33.1	102.89	0.16
Rv3711c	35749.93	5.71	39.06	101.06	0.035
Rv3721c	61891.31	5.61	44.12	94.12	-0.119
Rv2924c	31950.7	9.98	41.1	90.42	-0.243
Rv0006	92274.31	5.41	38.09	95.68	-0.303
Rv0005	78439.74	6.18	31.31	85.36	-0.379
Rv2092c	99573.95	6.99	46.88	91.13	-0.26
Rv2101	111630.45	5.61	47.79	95.33	-0.2
Rv2756c	60084.11	5.31	41.39	78.44	-0.415
Rv2755c	39211.94	9.61	45.25	95.44	-0.034
Rv3296	161347.7	6.21	41.68	97.51	-0.011
Rv3014c	75257.1	5.42	34.66	91.94	-0.235
Rv3062	53704.57	9.18	30.8	100	0.07
Rv3731	40159.7	6.57	42.6	80.95	-0.337
Rv1020	132908.42	5.55	36.68	96.47	-0.101
Rv2528c	33648.17	5.53	30.75	91.57	-0.305
Rv2985	34748.34	9.28	42.31	82.46	-0.427
Rv1160	15160.28	5.79	21.3	105.25	-0.028
Rv0413	23481.16	5.04	48.69	80.97	-0.353
Rv3589	33684.45	8.85	46.08	86.68	-0.227
Rv3297	28525.6	9.09	33.23	93.69	-0.224
Rv3674c	26998.28	9.83	46.35	96	-0.06
Rv1316c	17858.23	5.92	25.77	91.15	-0.17
Rv1629	98439.98	5.01	33.73	93.23	-0.22
Rv1402	69839.07	9.8	49.68	99.15	-0.006
Rv3585	49881.01	6.74	37.27	100.6	0.132
Rv2737c	85389.06	6.01	28.47	91.89	-0.185
Rv0630c	118722.38	5.98	39.82	93.49	-0.166
Rv0631c	119501.23	6.15	39.71	96.34	-0.17
Rv0629c	61714.72	6.73	32.18	105.58	0.008

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al. **195**

Table Continued...

Rv ID	Molecular weight	Theoretical pI	Instability index	Aliphatic index	GRAVY
Rv0003	42180.2	6.75	40.51	108.7	-0.047
Rv2973c	80328.99	6.16	33.15	97.73	-0.114
Rv1696	62196.88	5.12	27.03	97.39	-0.132
Rv3715c	22119.35	4.99	20.7	105.62	-0.058
Rv2736c	19145.81	9.49	52.22	94.31	-0.399
Rv2593c	20189.23	6.42	26.54	109.59	0.308
Rv2592c	36626.94	5.35	38.65	100.41	0.048
Rv2594c	19753.76	9.22	22.43	99.31	0.122
Rv0054	17321.05	5.12	42.3	64.33	-0.474
Rv1210	22973.04	7.89	53.01	74.22	-0.449
Rv3646c	102317.51	8.23	37.66	82.91	-0.472
Rv2976c	24449.05	9.18	42.03	92.07	0.019
Rv1638	106099.45	6.45	33.9	91.67	-0.247
Rv1633	78038.32	5.05	44.26	93.48	-0.33
Rv1420	71582.11	8.5	39.92	86.53	-0.339
Rv0949	85049.88	5.36	41.51	92.62	-0.273
Rv3198c	75603.69	6.5	35.79	97.41	-0.104
Rv0427c	32108.26	5.15	39.67	80.82	-0.279
Rv0071	26891.83	9.55	40.8	92.04	-0.375
Rv0861c	59772.18	5.73	37.1	98.67	-0.14
Rv0944	16462.95	9.27	29.91	88.99	-0.047
Rv1688	21340.08	9.88	36.44	80.44	-0.231
Rv2090	41938.99	5.58	44.81	89.19	-0.204
Rv2191	69148.12	9.62	49.07	88.79	-0.088
Rv2464c	29681.92	9.88	41.72	82.54	-0.327
Rv3201c	116688.87	5.86	37.89	94.41	-0.02
Rv3202c	110729.42	8.72	44.14	96.31	-0.036
Rv3263	60673.5	8.35	35.3	92.31	-0.103
Rv3644c	41784.51	8.11	42.03	90.02	-0.057

Table 3 The following table depicts the presence of alpha helix, extended strand, beta turn, and random coil present in the proteins using SOPMA tool

Gene name	Alpha helix	Extended strand	Beta turn	Random coil
Rv1317c	49.40%	7.26%	5.85%	37.50%
Rv2836c	61.50%	12.30%	5.92%	20.27%
Rv1329c	48.64%	12.95%	4.82%	33.58%
Rv3056	40.46%	17.34%	5.49%	36.71%
Rv1537	39.74%	11.66%	5.18%	43.41%
Rv0001	49.31%	10.65%	3.94%	36.09%
Rv0058	46.57%	13.50%	6.18%	33.75%
Rv1547	47.21%	14.36%	6.76%	31.67%
Rv3370c	45.51%	11.68%	6.86%	35.96%
Rv2343c	51.33%	9.23%	7.67%	31.77%
Rv0002	28.86%	23.88%	5.72%	41.54%
Rv3711c	38.30%	12.46%	7.90%	41.34%
Rv3721c	48.44%	7.96%	5.71%	37.89%
Rv2924c	32.87%	17.99%	6.57%	42.56%
Rv0006	36.87%	21.24%	10.26%	31.62%
Rv0005	41.32%	16.81%	7.00%	34.87%
Rv2092c	52.43%	11.37%	6.95%	29.25%
Rv2101	46.40%	12.73%	4.74%	36.13%
Rv2756c	47.59%	9.26%	4.63%	38.52%
Rv2755c	35.99%	16.76%	5.77%	41.48%
Rv3296	44.42%	12.89%	6.94%	35.76%
Rv3014c	42.40%	13.17%	7.38%	37.05%
Rv3062	52.27%	14.00%	6.90%	26.82%
Rv2592c	50.29%	12.79%	5.52%	31.40%
Rv2594c	52.13%	18.09%	6.91%	22.87%
Rv0054	18.90%	18.29%	7.93%	54.88%
Rv1210	48.04%	3.92%	6.86%	41.18%
Rv3646c	43.25%	12.10%	6.21%	38.44%

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al. **196**

Table Continued...

Gene name	Alpha helix	Extended strand	Beta turn	Random coil
Rv2976c	41.41%	15.42%	7.05%	36.12%
Rv1633	52.29%	14.90%	8.60%	24.21%
Rv1420	44.58%	15.48%	4.95%	34.98%
Rv0427c	30.58%	15.81%	9.28%	44.33%
Rv3198c	48.43%	11.43%	34.86%	0.00%
Rv0071	39.15%	14.89%	7.23%	38.72%
Rv0861c	43.17%	18.63%	5.17%	33.03%
Rv0949	48.38%	13.62%	5.97%	32.04%
Rv0944	51.27%	9.49%	3.80%	35.44%
Rv1688	20.20%	25.12%	6.90%	47.78%
Rv2090	33.59%	8.65%	5.34%	52.42%
Rv2191	43.88%	10.54%	3.88%	41.71%
Rv2464c	32.84%	17.91%	5.97%	43.28%
Rv3201c	47.14%	10.72%	3.00%	39.15%
Rv3202c	49.86%	8.06%	4.45%	37.63%
Rv3263	45.75%	16.64%	4.70%	32.91%
Rv3644c	63.34%	9.73%	3.49%	23.44%
Rv1638	36.73%	19.44%	8.54%	35.29%
Rv3731	31.01%	16.48%	6.42%	46.09%
Rv1020	43.03%	13.94%	6.00%	37.03%
Rv2528c	44.44%	12.75%	6.86%	35.95%
Rv2985	31.55%	15.14%	4.73%	48.58%
Rv1160	34.04%	17.02%	6.38%	42.55%
Rv0413	21.66%	21.20%	7.37%	49.77%
Rv3589	50.66%	4.93%	6.25%	38.16%
Rv3297	34.12%	18.04%	5.49%	42.35%
Rv3674c	49.39%	9.39%	6.12%	35.10%
Rv1316c	27.88%	22.42%	8.48%	41.21%
Rv1629	55.53%	9.85%	4.98%	29.65%
Rv1402	35.27%	15.73%	4.73%	44.27%
Rv3585	36.67%	15.42%	8.96%	38.96%
Rv2737c	37.22%	20.25%	8.99%	33.54%
Rv0630c	47.71%	10.69%	4.39%	37.20%
Rv0631c	45.49%	10.03%	3.19%	41.29%
Rv0629c	52.00%	13.22%	5.04%	29.74%
Rv0003	49.09%	16.88%	4.42%	29.61%
Rv2973c	46.68%	13.98%	6.11%	33.24%
Rv1696	58.94%	10.05%	6.13%	24.87%
Rv3715c	40.89%	14.78%	7.39%	36.95%
Rv2736c	71.84%	0.57%	4.02%	23.56%
Rv2593c	47.45%	14.29%	8.67%	29.59%

Tertiary structure prediction & homology modelling

Homology modelling of amino acid sequences involved in DNA replication, repair, recombination, and restriction/modification pathway of *Mycobacterium tuberculosis* was carried out to predict the 3D structure of these sequences. The models for this study were generated using the Phyre2 tool and the models were validated using Ramachandran's plot analysis. The Ramachandran map is an efficient approach used to visualize the favored regions for backbone dihedral angles ψ(Psi) against φ(Phi) of amino acid residues. The method involves plotting the φ(Phi) and the ψ(Psi) scores on the X-axis and Y-axis respectively with angle spectrum ranging from − 180º to + 180º which predicts the secondary structure and possible conformation of the molecule (Abdullahi et al. 2021). In Ramachandran plot analysis, a good model is expected to have over 90 % of residues in the most favored regions which suggests a good quality of homology models. In this study, it was found that out of all the sequences, 7 protein structures had 100% of residues in the most favored regions and 58 protein structures had 90 % of the residues in the most favored regions (Table 4).

Verify3D analysis indicated that 53 protein structures had a score greater than 0 which conveys that the predicted models were valid. Around 8 protein structures had the ERRAT value more than 95%. ERRAT is a verification algorithm for protein structures that is used for evaluating the quality of crystallographic model building and refinement. Generally, a score above 95% is considered as a good high-resolution structure which indicates that these 8 protein structures were credible and acceptable.[38] The ProsA-web server was used to calculate the Z-score of the protein models to determine if the protein model predicted falls within the range of high-quality experimental structures.[16,39] The requirement for ProsA-web server is only Cα atoms which helps in the evaluation of approximate models obtained in the structure determination process and low-resolution structures and can be compared against high-resolution structures. The variance of the total energy of the system from an energy distribution resulting from random conformations and shows overall model consistency which is indicated via z-score.[40,41] A z-score of -6.07 predicted by ProsA web server (Table 5) represents a good quality model (Prajapat, Bhattachar, and Kumar 2016). Taking that

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/ modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al. **197**

into account, in our study the two genes namely Rv3297 and Rv2593 had a score of -6.18 and -6.19 respectively which concludes that these models are of good quality. ProQ online server was used to forecast the quality of protein sequences used which depends on the neural system constructed apparatus that is based on the evaluation of the structural characters, there is the quality of a protein model, and it is efficient to discover local structures and to revise models. The quality estimates the LG score and MaxSub. The cutoff extent of LG score> 1.5 shows a very incredible model, > 2.5 extraordinary model and > 4 generally extraordinary model and there MaxSub score> 0.1 demonstrates amazingly extraordinary model, > 0.5 extraordinary model and > 0.8 incredibly incredible model. The study showed that all sequences had their LG scores and MaxSub scores as -0.835 and -0.113 (Table 5) respectively which indicates that the standard of all the protein structures is extremely good.[42–44] The quality of both global and local structures can be enhanced with this method.

Table 4 The following table represents the values for Ramachandran plot obtained for the selected proteins

Rv ID	Ramachandran favoured	Ramachandran outliers	Rotamer outliers
Rv1317c	95.70%	3.23	0
Rv2836c	100%	0	0
Rv1329c	89.76%	3.25	0.63
Rv3056	97.97%	0	1.08
Rv1537	95.69%	1.72	1.13
Rv0001	100%	0	0
Rv0058	97.22%	0.46	0.53
Rv1547	93.21%	0.22	7.19
Rv3370c	93.41%	2.69	1.09
Rv2343c	97.37%	0	0
Rv0002	96.67%	1.67	1.01
Rv3711c	91.67%	5.36	0
Rv3721c	93.72%	1.26	3.61
Rv2924c	100%	0	0
Rv0006	98.48%	0	1.04
Rv0005	95.06%	1.23	0
Rv2092c	93.83%	2.3	5.56
Rv2101	93.67%	1.78	1.22
Rv2756c	92.38%	2.42	2.28
Rv2755c	100%	0	0
Rv3296	96.56%	0.73	2.7
Rv3014c	96.55%	0	6.45
Rv3062	94.06%	1.78	2.63
Rv3731	88.65%	5.95%	1.29%
Rv1020	95.05%	0.81%	2.15%
Rv2528c	98.57%	0.00%	0.00%
Rv2985	95.56%	1.27%	2.72%
Rv1160	97.66%	0.00%	0.00%
Rv0413	96.00%	2.40%	0.00%
Rv3589	94.55%	1.98%	4.71%
Rv3297	93.68%	1.19%	1.92%
Rv3674c	95.65%	0.48%	0.00%
Rv1316c	95.29%	2.35%	0.00%
Rv1629	95.92%	0.35%	0.00%
Rv1402	90.17%	6.46%	2.33%
Rv3585	96.62%	0.68%	1.83%
Rv2737c	99.28%	0.00%	0.87%
Rv0630c	92.91%	3.54%	2.55%
Rv0631c	90.41%	3.93%	3.24%
Rv0629c	92.74%	3.07%	0.72%
Rv0003	93.97%	3.45%	0.70%

Table Continued...

Rv ID	Ramachandran favoured	Ramachandran outliers	Rotamer outliers
Rv2973c	91.62%	5.20%	2.15%
Rv1696	96.45%	0.76%	0.34%
Rv3715c	100.00%	0.00%	11.76%
Rv2736c	100.00%	0.00%	4.35%
Rv2593c	95.88%	1.55%	2.80%
Rv2592c	96.78%	1.17%	0.38%
Rv2594c	96.43%	0.00%	0.00%
Rv0054	89.83%	5.08%	4.62%
Rv1210	94.57%	3.26%	0.00%
Rv3646c	93.92%	1.34%	1.34%
Rv2976c	99.11%	0.00%	0.00%
Rv1638	95.49%	0.87%	1.46%
Rv1633	96.07%	1.28%	3.38%
Rv1420	97.67%	0.00%	3.90%
Rv0949	97.50%	0.00%	0.00%
Rv3198c	98.55%	0.00%	0.00%
Rv0427c	93.78%	2.90%	1.00%
Rv0071	94.48%	3.07%	0.71%
Rv0861c	94.07%	0.89%	2.46%
Rv0944	95.00%	1.43%	0.00%
Rv1688	100.00%	0.00%	12.50%
Rv2090	97.77%	0.32%	0.80%
Rv2191	91.25%	4.08%	8.82%
Rv2464c	97.67%	0.00%	0.00%
Rv3201c	90.95%	1.89%	0.83%
Rv3202c	91.59%	1.62%	1.70%
Rv3263	87.48%	3.33%	1.93%
Rv3644c	96.00%	1.09%	0.00%

Table 5 The following table represents the values and overall model quality extracted from webserver ProQ and ProSA

Rv ID	ProQ		ProSA
	Predicted LGscore	Predicted MaxSub	Z-Score
Rv1317c	-0.835	-0.113	-4.57
Rv2836c	-0.835	-0.113	-3.94
Rv1329c	-0.835	-0.113	-4.32
Rv3056	-0.835	-0.113	-4.32
Rv1537	-0.835	-0.113	-4.32
Rv0001	-0.835	-0.113	-4.32
Rv0058	-0.835	-0.113	-4.32
Rv1547	-0.835	-0.113	-4.32
Rv3370c	-0.835	-0.113	-4.32
Rv2343c	-0.835	-0.113	-4.32
Rv0002	-0.835	-0.113	-4.32
Rv3711c	-0.835	-0.113	-4.32
Rv3721c	-0.835	-0.113	-4.32
Rv2924c	-0.835	-0.113	-4.32
Rv0006	-0.835	-0.113	-4.32
Rv0005	-0.835	-0.113	-4.32
Rv2092c	-0.835	-0.113	-4.32
Rv2101	-0.835	-0.113	-4.32
Rv2756c	-0.835	-0.113	-4.32
Rv2755c	-0.835	-0.113	-4.32
Rv3296	-0.835	-0.113	-4.32
Rv3014c	-0.835	-0.113	-4.32
Rv3062	-0.835	-0.113	-4.32
Rv3731	-0.835	-0.113	-5.04
Rv1020	-0.835	-0.113	-15.24
Rv2528c	-0.835	-0.113	-5.9
Rv2985	-0.835	-0.113	-8.34

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/ modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al. **198**

Table Continued...

Rv ID	ProQ		ProSA
	Predicted LGscore	Predicted MaxSub	Z-Score
Rv1160	-0.835	-0.113	-5.51
Rv0413	-0.835	-0.113	-3.83
Rv3589	-0.835	-0.113	-6.93
Rv3297	-0.835	-0.113	-6.18
Rv3674c	-0.835	-0.113	-7.47
Rv1316c	-0.835	-0.113	-4.52
Rv1629	-0.835	-0.113	-11.34
Rv1402	-0.835	-0.113	-5.03
Rv3585	-0.835	-0.113	-5.98
Rv2737c	-0.835	-0.113	-5.91
Rv0630c	-0.835	-0.113	-3.02
Rv0631c	-0.835	-0.113	-12.31
Rv0629c	-0.835	-0.113	-5.26
Rv0003	-0.835	-0.113	-5.76
Rv2973c	-0.835	-0.113	-6.6
Rv1696	-0.835	-0.113	-5.5
Rv3715c	-0.835	-0.113	-0.85
Rv2736c	-0.835	-0.113	-2.88
Rv2593c	-0.835	-0.113	-6.19
Rv2592c	-0.835	-0.113	-9.24
Rv2594c	-0.835	-0.113	0.5
Rv0054	-0.835	-0.113	-4.16
Rv1210	-0.835	-0.113	-3.79
Rv3646c	-0.835	-0.113	-12.83
Rv2976c	-0.835	-0.113	-8.57
Rv1638	-0.835	-0.113	-8.67
Rv1633	-0.835	-0.113	-12
Rv1420	-0.835	-0.113	-4.81
Rv0949	-0.835	-0.113	-2.2
Rv3198c	-0.835	-0.113	-6.63
Rv0427c	-0.835	-0.113	-4.98
Rv0071	-0.835	-0.113	-2.32
Rv0861c	-0.835	-0.113	-6.67
Rv0944	-0.835	-0.113	-5.15
Rv1688	-0.835	-0.113	-1.27
Rv2090	-0.835	-0.113	-10.81
Rv2191	-0.835	-0.113	-5
Rv2464c	-0.835	-0.113	-5.55
Rv3201c	-0.835	-0.113	-5.22
Rv3202c	-0.835	-0.113	-9.51
Rv3263	-0.835	-0.113	-5.97
Rv3644c	-0.835	-0.113	-4.57

B-cell epitope prediction and scanning of proteins for IFN epitopes

B cell epitope-based prediction was performed for two genes with highest antigenic score Rv0054 & Rv3644c which could be valuable in planning and creating the epitope-based immunization against *Mycobacterium tuberculosis*. B-cells are an important part of the adaptive immune system because they can protect the body against pathogens and harmful molecules for a long time.[22] B-cell epitope assessment is essential for a variety of medical, immunological, and biological applications, including disease control, diagnostics, and vaccine development by Shirai. Intracellular pathogen evasion and recruitment of cytotoxic lymphocytes and natural killer cells are processes in which interferon gamma plays a very significant role.[25] The DNA damage pathway includes the recruitment of certain repair enzymes, and the initiation of sign transducers that direct cell cycle and cell survival by Brzostek-Racine. As per the results of B

cell prediction for genes with the highest antigen scores: Rv0054 & Rv3644c, IFN gamma inducing regions were predicted and were then proceeded further for molecular docking analysis.

Molecular docking analysis

Following the identification of epitope sequences of Rv0054 & Rv3644c genes, molecular docking was performed with IP-10 protein (Crystal structure of mouse) using H-dock. The docking scores were divided into different ranks of models_rank numbers. RMSD values below 2.0 Å are good docking scores. To this study, only rank 1 model were considered (Table 6). Since the output corresponds to all values below 2.0 Å, it is considered to share a good binding affinity by Ramírez.

Table 6 Molecular Docking analysis scores using Hdock

Rv3644c		Rv0054	
epitopes	Docking score	epitopes	Docking score
ALQCTSGGEPGCGRC	-145.17	AENVAESLTRGARVI	-137.7
CTSGGEPGCGRCRAC	-134.7	ENVAESLTRGARVIV	-128.84
TSGGEPGCGRCRACT	-98	NVAESLTRGARVIVS	-173.54
SGGEPGCGRCRACTT	-152.28	VAESLTRGARVIVSG	-123.27
GGEPGCGRCRACTTT	-157.13	AESLTRGARVIVSGR	-157.84
GEPGCGRCRACTTTL	-157.55	ESLTRGARVIVSGRL	-162.06
GRCRACTTTLAGTHA	-162.32	SLTRGARVIVSGRLK	-136.33
TTLAGTHADVRRVIP	-175.96	LTRGARVIVSGRLKQ	-159.27
VIPEGLSIGVDEMRA	-138.84	TRGARVIVSGRLKQR	-133.56
ANALLKVVEEPPPST	-155.65	RGARVIVSGRLKQRS	-141.54
NALLKVVEEPPPSTV	-157.65	GARVIVSGRLKQRSF	-142.39
ALLKVVEEPPPSTVF	-145.01	RVIVSGRLKQRSFET	-104.77
LLKVVEEPPPSTVFL	-149.37	VIVSGRLKQRSFETR	-137.58
LKVVEEPPPSTVFLL	-156.66	ETREGEKRTVIEVEV	-151.32
KVVEEPPPSTVFLLC	-175.5	EGEKRTVIEVEVDEI	-127.2
EEPPPSTVFLLCAPS	-138.64	VIEVEVDEIGPSLRY	-149.99
EPPPSTVFLLCAPSV	-187.81	VEVDEIGPSLRYATA	-171.98
PPPSTVFLLCAPSVD	-152.19	EVDEIGPSLRYATAK	-182
PSVDPEDIAVTLRSR	-135.32	VDEIGPSLRYATAKV	-162.87
SVDPEDIAVTLRSRC	-165.67	DEIGPSLRYATAKVN	-184.13
VDPEDIAVTLRSRCR	-137.47	EIGPSLRYATAKVNK	-161.33
DPEDIAVTLRSRCRH	-183.87	IGPSLRYATAKVNKA	-161.71
PEDIAVTLRSRCRHV	-170.52	GPSLRYATAKVNKAS	-118.74
EDIAVTLRSRCRHVA	-153.62	PSLRYATAKVNKASR	-143.1
DIAVTLRSRCRHVAL	-149.31	SLRYATAKVNKASRS	-155.24
IAVTLRSRCRHVALV	-180.32	LRYATAKVNKASRSG	-127.98
AVTLRSRCRHVALVT	-153.38	RYATAKVNKASRSGG	-165.53
VTLRSRCRHVALVTP	-170.93	TAKVNKASRSGGFGS	-100.59
TLRSRCRHVALVTPS	-168.09	GSGSRPAPAQTSSAS	-105.05
LRSRCRHVALVTPST	-137.93	SGSRPAPAQTSSASG	-144.57
RSRCRHVALVTPSTH	-113.98	GSRPAPAQTSSASGD	-108.57
SRCRHVALVTPSTHA	-161.97	SRPAPAQTSSASGDD	-120.45
RCRHVALVTPSTHAI	-161.29	SGGFGSGSRPAPAQT	-125.98
CRHVALVTPSTHAIA	-153.05	DDPWGSAPASGSFGG	-120.87
RHVALVTPSTHAIAQ	-134.31	DPWGSAPASGSFGGG	-167.58
LVTPSTHAIAQVLSD	-141.43	PWGSAPASGSFGGGD	-181.42
TANWAASVSGGHVGR	-129.85	WGSAPASGSFGGGDD	-95.74
EELRTALGAGGTGKG	-152.1		
ELRTALGAGGTGKGT	-149.44		
LRTALGAGGTGKGTG	-148.01		
RTALGAGGTGKGTGA	-135.56		
TALGAGGTGKGTGAA	-152.1		
LGAGGTGKGTGAALR	-112.96		
KGTGAALRGATGAMK	-152.21		

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/ modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al. **199**

Table Continued...

Rv3644c		Rv0054	
epitopes	Docking score	epitopes	Docking score
IDLATYFRDALLVAA	-174.77		
AAHAGGVRANHPDMA	-151.49		
AHAPPERLLRCIEAV	-180.52		
HAPPERLLRCIEAVL	-194.87		
APPERLLRCIEAVLA	-158.55		
PPERLLRCIEAVLAC	-146.37		
EALAVNVKPKFAVDA	-146.73		

In this study, the selected best proteins have a good immune response to mice protein IP-10. The lead proteins show satisfactory physiochemical properties, antigenicity, secondary and tertiary structures, and molecular docking scores. Therefore, these proteins can be considered effective against MTB. We believe the findings will benefit in the development of conventional medicine based therapeutic approaches as well as the advancement of better research for future treatment of MTB.[45–64]

Conclusion

Tuberculosis is a life-threatening disease and a global health challenge. There is an urgent need for potent diagnostic marker against this deadly disease. For this study a total of 69 amino acid sequences involved in the DNA replication, repair, recombination and restriction/ modification pathway of *Mycobacterium tuberculosis* was taken into consideration. Retrieval of the amino acid sequences was done using Tuberculist tool and Mycobrowser. VaxiJen server was used to study the antigenicity of the protein sequences. The physicochemical characterization was done using various computational tools and servers based on different parameters. The distinct parameters were isoelectronic point, molecular weight, instability index, aliphatic index, GRAVY and also the positive & negative residues. SOPMA was used for the analysis of Secondary structure prediction where the alpha helix, 3_{10}Helix, Pi helix, Beta bridge, Extended strand, Beta turn, Bend region, random coil, Ambiguous states and other states were predicted. ProtParam was used to study the amino acid composition. Three-dimensional structures were predicted using the Phyre tool. Ramchandran plot maps were analyzed using Swiss model server. ProsA and ProQ servers were used to study the Z-score and LGscore & MaxSub score respectively. This study concludes that the two proteins Rv0054 & Rv3644c can be considered to play a potential role as a diagnostic agent for *Mycobacterium Tuberculosis*. Computational analysis and homology modelling of *Mycobacterium tuberculosis* involved in DNA replication, repair, recombination, and restriction/modification pathway provides a basis for analysis of these proteins. This research is believed to set a course for positive outcomes and potential diagnostic markers using immunoinformatic based tools that will aid in the development of remedy against *Mycobacterium tuberculosis*.

Acknowledgments

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

References

1. Reiche Michael A, Digby F Warner, Valerie Mizrahi. Targeting DNA Replication and Repair for the Development of Novel Therapeutics against Tuberculosis. *Frontiers in Molecular Biosciences*. 2017;4:75.

2. Cole ST, R Brosch, J Parkhill, et al. Deciphering the Biology of Mycobacterium Tuberculosis from the Complete Genome Sequence. *Nature*. 1998;393(6685):537–544.

3. Beattie Thomas R, Rodrigo Reyes–Lamothe. A Replisome's Journey through the Bacterial Chromosome. *Frontiers in Microbiology*. 2015;6:562.

4. Yao Nina, Mike O'Donnell. Bacterial and Eukaryotic Replisome Machines. *JSM Biochemistry and Molecular Biology*. 2016;3(1).

5. Mestre Olga, Tao Luo, Tiago Dos Vultos, et al. Phylogeny of Mycobacterium Tuberculosis Beijing Strains Constructed from Polymorphisms in Genes Involved in DNA Replication, Recombination and Repair. *PloS One*. 2011;6(1):e16020.

6. Ditse Zanele, Meindert H Lamers, Digby F Warner. DNA Replication in Mycobacterium Tuberculosis. In Tuberculosis and the Tubercle Bacillus. John Wiley & Sons, Ltd. 2017. p. 581–606.

7. Vultos Tiago Dos, Olga Mestre, Tone Tonjum, et al. DNA Repair in Mycobacterium Tuberculosis Revisited. *FEMS Microbiology Reviews*. 2009;33(3):471–487.

8. Mizrahi Valerie, Susan J Andersen. DNA Repair in Mycobacterium Tuberculosis. What Have We Learnt from the Genome Sequence? *Molecular Microbiology*. 1998;29(6):1331–1339.

9. Saikat Abu Saim Mohammad. Structure Prediction and Characterization of Uncharacterized ABC Transporter ATP–Binding Protein Rv0986 of <Em>Mycobacterium Tuberculosis</Em> (Strain ATCC 25618 / H37Rv). *BioRxiv*. 2020.

10. Kapopoulou Adamandia, Jocelyne M Lew, Stewart T Cole. The MycoBrowser Portal: A Comprehensive and Manually Annotated Resource for Mycobacterial Genomes. *Tuberculosis (Edinburgh, Scotland)*. 2011;91(1):8–13.

11. Wilkins MR, E Gasteiger, A Bairoch, et al. Protein Identification and Analysis Tools in the ExPASy Server. *Methods in Molecular Biology*. 1999;112:531–552.

12. Tran Ngoc Tuan, Ivan Jakovlić, Wei–Min Wang. In Silico Characterisation, Homology Modelling and Structure–Based Functional Annotation of Blunt Snout Bream (Megalobrama Amblycephala) Hsp70 and Hsc70 Proteins. *Journal of Animal Science and Technology*. 2015;57(1):44.

13. Verma Devvret, Neema Tufchi, Kumud Pant, et al. Computational Analysis and Homology Modeling of Potential Target. *Proteins of Mycobacterium Tuberculosis: An In–Silico Approach*. 2020;1:5–8.

14. Kelley Lawrence A, Stefans Mezulis, Christopher M Yates, et al. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nature Protocols*. 2015;10(6):845–858.

15. Oduselu Gbolahan O, Olayinka O Ajani, Yvonne U Ajamma, et al. Homology Modelling and Molecular Docking Studies of Selected Substituted Benzo[d]Imidazol–1–Yl)Methyl)Benzimidamide Scaffolds on Plasmodium Falciparum Adenylosuccinate Lyase Receptor. *Bioinformatics and Biology Insights*. 2019;13:1177932219865533.

16. Wiederstein Markus, Manfred J Sippl. ProSA–Web: Interactive Web Service for the Recognition of Errors in Three–Dimensional Structures of Proteins. *Nucleic Acids Research*. 2007;35(Web Server issue):W407–410.

17. Ponomarenko Julia V, Philip E Bourne. Antibody–Protein Interactions: Benchmark Datasets and Prediction Tools Evaluation. *BMC Structural Biology*. 2007;7:64.

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/ modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al.   **200**

18. Haste Andersen, Pernille, Morten Nielsen, Ole Lund. Prediction of Residues in Discontinuous B–Cell Epitopes Using Protein 3D Structures. *Protein Sci*. 2006;15(11):2558–2567.

19. Larsen Jens Erik Pontoppidan, Ole Lund, Morten Nielsen. Improved Method for Predicting Linear B–Cell Epitopes. *Immunome Research*. 2006;2:2.

20. Emini EA, JV Hughes, DS Perlow, et al. Induction of Hepatitis A Virus–Neutralizing Antibody by a Virus–Specific Synthetic Peptide. *Journal of Virology*. 1985;55(3): 836–839.

21. Kolaskar AS, PC Tongaonkar. A Semi–Empirical Method for Prediction of Antigenic Determinants on Protein Antigens. *FEBS Letters*. 1990;276 (1–2):172–174.

22. Jespersen Martin Closter, Bjoern Peters. BepiPred–2.0: Improving Sequence–Based B–Cell Epitope Prediction Using Conformational Epitopes. *Nucleic Acids Research*. 2017;45(W1):W24–29.

23. Parker JMR, D Guo, RS Hodges. New Hydrophilicity Scale Derived from High–Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and x–Ray–Derived Accessible Sites. *Biochemistry*. 1986;25(19):425–432.

24. Hasan Md, Md Arif Khan, Amit Datta, et al. A Comprehensive Immunoinformatics and Target Site Study Revealed the Corner–Stone toward Chikungunya Virus Treatment. *Molecular Immunology*. 2015;65:189–204.

25. Bibi Shaheen, Inayat Ullah, Bingdong Zhu, et al. In Silico Analysis of Epitope–Based Vaccine Candidate against Tuberculosis Using Reverse Vaccinology. *Scientific Reports*. 2021;11(1):1249.

26. Barberis I, NL Bragazzi, L Galluzzo, et al. The History of Tuberculosis: From the First Historical Records to the Isolation of Koch's Bacillus. *Journal of Preventive Medicine and Hygiene*. 2017;58(1):E9–12.

27. Kaur Rajwinder, Dylan J Nikkel, Stacey D Wetmore. Computational Studies of DNA Repair: Insights into the Function of Monofunctional DNA Glycosylases in the Base Excision Repair Pathway. *WIREs Computational Molecular Science*. 2020;10(5):e1471.

28. Vandal Omar H, Carl F Nathan, Sabine Ehrt. Acid Resistance in Mycobacterium Tuberculosis. *Journal of Bacteriology*. 2009;191(15):4714–4721.

29. Kohli Sakshi, Yadvir Singh, Khushbu Sharma, et al. Comparative Genomic and Proteomic Analyses of PE/PPE Multigene Family of Mycobacterium Tuberculosis H37Rv and H37Ra Reveal Novel and Interesting Differences with Implications in Virulence. *Nucleic Acids Research*. 2012;40(15):7113–7122.

30. Gasteiger Elisabeth, Christine Hoogland, Alexandre Gattiker, et al. Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook. In: John M Walker, editor. Totowa, NJ: Humana Press. 2005. p. 571–607.

31. Botto Marina, Philip N Hawkins, Maria CM Bickerstaff, et al. Amyloid Deposition Is Delayed in Mice with Targeted Deletion of the Serum Amyloid P Component Gene. *Nature Medicine*. 1997;3(8):855–859.

32. Saleem Afnan, Shiveeli Rajput. Insights from the in Silico Structural, Functional and Phylogenetic Characterization of Canine Lysyl Oxidase Protein. *Journal of Genetic Engineering and Biotechnology*. 2020;18(1):20.

33. Kyte J, RF Doolittle. A Simple Method for Displaying the Hydropathic Character of a Protein. *Journal of Molecular Biology*. 1982;157(1):105–132.

34. Prajapat Rajneesh, Ijen Bhattachar, Anoop Kumar. Homology Modeling and Structural Validation of Type 2 Diabetes Associated Transcription Factor 7–like 2 (TCF7L2). *Trends in Bioinformatics*. 2016;9:23–29.

35. Ferrè F, P Clote. Disulfide Connectivity Prediction Using Secondary Structure Information and Diresidue Frequencies. *Bioinformatics*. 2005;21(10):2336–2346.

36. Smith Lorna J, Klaus M Fiebig, Harald Schwalbe, et al. The Concept of a Random Coil: Residual Structure in Peptides and Denatured Proteins. *Folding and Design*. 1996;1(5):R95–106.

37. Scott W Robinson, Avid M Afzal, David P Leader. Bioinformatics: Concepts, Methods, and Data. Handbook of Pharmacogenomics and Stratified Medicine. 2014. p. 259–287.

38. Qiu Juanjuan, Shizhu Zang, Yufang Ma, et al. Homology Modeling and Identification of Amino Acids Involved in the Catalytic Process of Mycobacterium Tuberculosis Serine Acetyltransferase. *Mol Med Rep*. 2017;15(3):1343–1347.

39. Cloete Ruben, Erika Kapp, Jacques Joubert, et al. Molecular Modelling and Simulation Studies of the Mycobacterium Tuberculosis Multidrug Efflux Pump Protein Rv1258c. *PLOS ONE*. 2018;13(11):e0207605.

40. Sippl Manfred J. Recognition of Errors in Three–Dimensional Structures of Proteins. *Proteins: Structure, Function, and Bioinformatics*. 1993;17(4):355–362.

41. Manfred Jsippl. Knowledge–Based Potentials for Proteins. *Current Opinion in Structural Biology*. 1995;5(2):229–235.

42. Cristobal Susana, Adam Zemla, Daniel Fischer, et al. A Study of Quality Measures for Protein Threading Models. *BMC Bioinformatics*. 2001;2(1):5.

43. Nath Onkar, Shailesh Kumar, Sumit Govil, et al. Computational 3D Structure Prediction, Evaluation and Analysis of Pyruvate Dehydrogenase an Effective Target for Filarial Infection by Brugia Pahangi Using Homology Modeling Approach. *International Journal of Pharmaceutical Sciences and Drug Research*. 2014;6:120–123.

44. Amjad Beg, Shivangi, Fareeda Athar, et al. Structural And Functional Annotation Of Rv1514c Gene Of Mycobacterium Tuberculosis H37Rv As Glycosyl Transferases. 2018.

45. Abdullahi Mustapha, Shola Elijah Adeniji, David Ebuka Arthur, et al. Homology Modeling and Molecular Docking Simulation of Some Novel Imidazo[1,2–a] Pyridine–3–Carboxamide (IPA) Series as Inhibitors of Mycobacterium Tuberculosis. *J Genet Eng Biotechnol*. 2021;19(1):12.

46. Daniel Thomas M. The History of Tuberculosis. *Respiratory Medicine*. 2006;100(11):1862–1870.

47. Dimitrov Ivan, Ivan Bangov, Darren R Flower, et al. AllerTOP v.2—a Server for in Silico Prediction of Allergens. *Journal of Molecular Modeling*. 2014;20(6):2278.

48. Dimitrov Ivan, Darren R Flower, Irini Doytchinova. AllerTOP – a Server for in Silico Prediction of Allergens. *BMC Bioinformatics*. 2013;14(6): S4.

49. Dimitrov Ivan, Lyudmila Naneva, Irini Doytchinova, et al. AllergenFP: Allergenicity Prediction by Descriptor *Fingerprints. Bioinformatics*. 2014;30(6):846–851.

50. Doytchinova Irini A, Darren R Flower. VaxiJen: A Server for Prediction of Protective Antigens, Tumour Antigens and Subunit Vaccines. *BMC Bioinformatics*. 2007a;8(1):4.

51. Irini A Doytchinova, Darren R Flower. Identifying Candidate Subunit Vaccines Using an Alignment–Independent Method Based on Principal Amino Acid Properties. *Vaccine*. 2007b;25(5):856–866.

52. Fiebig Klaus M, Harald Schwalbe, Matthias Buck, et al. Toward a Description of the Conformations of Denatured States of Proteins. Comparison of a Random Coil Model with NMR Measurements. *The Journal of Physical Chemistry*. 1996;100(7):2661–2666.

53. Gazi MA, M Kibria, M Mahfuz, et al. Functional, Structural and Epitopic Prediction of Hypothetical Proteins of Mycobacterium Tuberculosis H37Rv: An in Silico Approach for Prioritizing the Targets. *Gene*. 2016;591(2):442–455.

Immunoinformatic analysis of proteins from DNA replication, repair, recombination, and restriction/modification pathway of Mycobacterium tuberculosis revealed the diagnostic potential of Rv0054 and Rv3644c

Copyright:
©2022 Vikas et al.   **201**

54. Käll Lukas, Anders Krogh, Erik LL Sonnhammer. A Combined Transmembrane Topology and Signal Peptide Prediction Method. *Journal of Molecular Biology*. 2004;338(5):1027–1036.

55. Lukas Käll, Anders Krogh, Erik LL, et al. Advantages of Combined Transmembrane Topology and Signal Peptide Prediction—the Phobius Web Server. *Nucleic Acids Research*. 2007;35(Web Server issue):W429–32.

56. Krogh A, B Larsson, G von Heijne, et al. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *Journal of Molecular Biology*. 2001;305(3):567–580.

57. Kumari Kriti, Uttam Gunjan, Kumari Ankita, et al. In–Silico Studies on Virulence Factors of Cryptococcus Species: Phylogenetic Analysis and B–Cell Epitope Prediction. *Biointerface Research in Applied Chemistry*. 2021;11(6).

58. Mostowy Serge, Marcel A Behr. The Origin and Evolution of Mycobacterium Tuberculosis. *Clinics in Chest Medicine*. 2005;26(2):207–216.

59. Mukesh M, M Prathap, M Sabitha. Structural Model of the Alpha Phosphoglucomutase: A Promising Target for the Treatment of Mycobacterium Tuberculosis. *International Journal of Pharmacy and Pharmaceutical Sciences*. 2013;5:107–114.

60. Prajapati Chirag, Chintan Bhagat. In–silico analysis and homology modeling of target proteins for *Clostridium Botulinum*. *Journal of Pharmaceutical Sciences and Research*. 2012;3:2050–2056.

61. Shen Hong–Bin, Kuo–Chen Chou. Virus–PLoc: A Fusion Classifier for Predicting the Subcellular Localization of Viral Proteins within Host and Virus–Infected Cells. *Biopolymers*. 2007;85(3):233–240.

62. Hong–Bin Shen, Kuo–Chen Chou. Virus–MPLoc: A Fusion Classifier for Viral Protein Subcellular Location Prediction by Incorporating Multiple Sites. *Journal of Biomolecular Structure & Dynamics*. 2012;28(2):175–186.

63. Warner Digby. The Role of DNA Repair in M. Tuberculosis Pathogenesis. *Drug Discovery Today: Disease Mechanisms*. 2010;7.

64. Yao Nina Y, Mike E O'Donnell. Evolution of Replication Machines. *Critical Reviews in Biochemistry and Molecular Biology*. 2016;51(3):135–149.