Differential Gene Expression Analysis of Latent and Active Tuberculosis used for Machine Learning Guided Therapeutic Peptide Design

Baothman Othman A.[1,2]
1. Department of Biochemistry, Faculty of Science, King Abdulaziz University, Jeddah, SAUDI ARABIA
2. Center of Artificial Intelligence in Precision Medicines, King Abdulaziz University, Jeddah, SAUDI ARABIA
oabaothman@kau.edu.sa

Abstract
Tuberculosis (TB) constitutes a significant and escalating threat to global health. In this study, bioinformatics tools were used to find possible TB hub genes and in silico approaches based on structure and machine learning to target those genes. The study identifies crucial hub genes in three distinct sample types: latent tuberculosis infection (LTBI), active tuberculosis (ATB) and healthy cells, using the GSE62525 dataset from the GEO database. The upregulated genes were used to conduct gene enrichment analysis and construct a protein-protein interaction (PPI) network. Results from the network analysis showed the top ten hub genes. Interleukin-10 (IL10) was identified with promising therapeutic potential in TB.

The residue contact was analysed to understand the interaction between the selected crucial node and its receptor. Peptide was built based on the 20:18 residue interface to determine the nature of the interaction between IL10 and its receptor IL-10Rβ. Virtual screening confirmed the stability and interaction of two mutants out of 6,480 mutant peptides that showed significantly increased binding affinities to IL10. Both variant-I (CG_KYC) and variant-II (CV_RYC) peptides exhibited substantial binding to IL10, with variant-II showing the highest affinity, as seen by binding free energies of -68.13 and -95.64 kcal/mol respectively, post-500 ns MD simulation. The study identified active peptides that could lead to future therapies for TB.

Keywords: Tuberculosis, Network analysis, ML-based mutation, DeepPurpose, Active Peptides, Molecular dynamics simulation.

Introduction

Tuberculosis (TB) is a predominant global cause of mortality and disability, impacting individuals worldwide. *Mycobacterium tuberculosis*, often known as MTB or simply *M. tuberculosis*, is a pathogenic type of bacteria that causes TB. It belongs to the Mycobacteriaceae family and shares a relationship with Koch Bacillus. Airborne TB affects one-third of the global population and kills between one million and six million people annually[36]. Furthermore, there is a direct correlation between the various geographical areas and the seven phylogenetically separate lineages that make up the Mycobacterium tuberculosis complex (MTBC), which is responsible for TB in humans. Among the most geographically separated lineages, lineage 4 (sometimes called Euro-American) and lineage 2 (which includes Beijing) stand out.

Compared to more locally distributed lineages, these are more dangerous. Concomitantly, disease severity, transmission and pro-inflammatory host immune responses[8] are all impacted by this heightened virulence. In addition, TB epidemiology is characterized by its heterogeneity. But the awareness surrounding HPV might not be as widespread like other infectious diseases. Numerous factors, such as the characteristics of the infective host, the pathogen, the susceptible host, the environment and distal determinants, contribute to the diverse patterns in TB epidemiology. The total of these variables may increase or decrease heterogeneity[43]. A broad variety of clinical symptoms are caused by the heterogenicity of TB, which is influenced by factors such as pathogen traits and environmental conditions. From asymptomatic latent infections to active TB, these could range widely[5].

In a clinical context, MTB infection can manifest in many forms, ranging from asymptomatic latent tuberculosis infection (LTBI) to active tuberculosis (ATB). There are significant challenges in curing TB due to the complexity of the disease and the wide range of lesions that patients experience. Symptoms of acute TB include a persistent cough that produces phlegm, night sweats, weakness and loss of weight. Although LTBI controls the infection in the majority of individuals exposed to MTB, 5-10% of those exposed to ATB develop the disease [10,40]. It takes several years after contracting LTBI for TB to develop.

In its dormant condition, MTB enhances drug resistance in pathogenic bacteria by prolonging the generation time and reducing the likelihood of mutational drug resistance[7]. The majority of instances of TB outside the lungs are not associated with transmission from person to person[41]. Furthermore, a study by Mokrousov et al showed that the 14717- 15 cluster had the highest mortality rates (58.3%) in a location where TB is highly prevalent. In comparison, 31.4% of the isolates were from Beijing, while 15.2% were from outside Beijing. This study emphasizes the significance of taking medication resistance and TB strain pathogenicity into account during drug development[27].

Likewise, Telacebec, the antibiotic that Lee and Pethe[24] analyzed, has finished three clinical trials and a promising pre-clinical trial in which it inhibits MTB growth. Specifically, a phase 2a clinical trial found that telacebec dose was associated with a decrease in bacterial load in patient sputum. To fully assess its therapeutic potential, however, additional rigorous clinical trials are required[24].

The TB epidemic that swept North America and Europe in the 18th and 19th centuries gave the disease its moniker, "Captain Among These Men of Death." The discovery of streptomycin in 1944 and isoniazid in 1952, along with public health measures and the BCG vaccine, ushered in the modern era of TB management[9]. Previously, 113 implantations involving bone graft goods used in transplantation identified a TB epidemic in the United States. A troubling trend of 105 patients starting TB treatment and 8 patients dying is evident[34]. The discovery of effective medications in 1944 and the deployment of "triple therapy" in 1952 are other landmarks in the history of TB treatment. While innovations in the 1970s and 1980s shortened treatment durations, drug-resistant strains emerged, particularly in low-resource nations.

The development of new medications is essential in the battle against medication resistance[44] while intermittent regimens have demonstrated promise despite their difficulty. Identifying biomarkers is crucial for advancing the tuberculosis diagnostic and therapeutic pipeline[13]. A study identified 180 genes (DEGs) linked to myeloid leukocyte activation and cytokine production through bioinformatics analysis of two datasets[25]. Among the 98 differentially expressed genes and 4 hub genes found in a TB experiment, according to a bioinformatics study, are those mainly linked to different pathways such as cytokine-dependent signalling, cytokine-cytokine receptor interaction, beta-galactosidase activity, measles and JAK-STAT[48].

This work showcases the use of bioinformatics analysis to find TB genes that are differentially expressed. The focus is on latent TB, active TB and healthy cases. In order to pinpoint hub genes, the research relied on their physiological functions and the protein-protein interaction network. The goal is to find new ways to treat pulmonary TB by improving our knowledge of its molecular mechanisms.

Material and Methods

Microarray Data Collection: The GEO database [1] was used to gather microarray data from a single dataset that included active, latent and healthy cells. The dataset with the accession number GSE62525 was located using the filter "Expression Profiling by Array" and the following criteria: "Tuberculosis AND ("healthy" or "control") AND "latent" AND "active". A total of 42 samples were identified within this collection.

Recognition of Differentially Expressed Genes (DEGs): In this particular case, of the 42 samples, 14 cells were identified as having ATB, 14 as having LTB and 14 as healthy. First, we did a differential analysis on ATB in comparison to healthy cells (group 1) and second, we did the same thing in comparison to LTB (group 2). We constructed a bar graph to analyse the datasets and determine the sample mean. The GEO2R analysis[1] was applied to the samples. Differentially expressed genes were those with a log2 fold change (FC) value greater than or equal to 1. Simultaneously, using a corrected P-value of less than 0.05, a volcano plot was created to locate the DEGs that were upregulated and downregulated. Also, to look at the genes shared by both sets of people, a Venn diagram was made.

KEGG and GO Enrichment Analysis: A volcano plot and an adjusted P-value less than 0.05[31] were used in conjunction with GO and KEGG to identify the upregulated and downregulated DEGs. This analysis utilized the common elevated genes from both groups 1 and 2. Within the GO route (CC) are cellular components, biological processes (BP) and molecular functions (MF). KEGG's presentation of the metabolic pathways associated with the gene list makes understanding the disease possible.

Network Analysis and Identification of Hub Genes: We built a protein-protein interaction (PPI) network using the Search Tool for the Retrieval of Interacting Genes (STRING) [38] database to further study the genes that affected tuberculosis pathogenesis. After the hub genes were found using the degree technique. Cyctoscape[12] used the cytohubba plug-in[6] to identify them. After identifying the top 10 hub genes through the use of the degree technique, they were subjected to additional analysis.

The purpose of this study was to examine the function of the top hub genes by collecting information about them from different literature. The data allowed for the selection of the potential target protein.

Template Peptide and Variants Generation: PDBePISA (Proteins, Interfaces, Structures and Assemblies) was employed to identify the interface residue between the proteins[22]. We selected residues that exhibited a continuous association for the purpose of peptide design. In the peptide structure, the glycine (GLY) residues served as linkers. By measuring the distance between the terminal residues of neighbouring peptides, we were able to determine the optimal number of GLY to utilize. We mutated these particular residues to find the most efficient linker residues that may further enhance the peptide-protein interaction. What followed was an evaluation of each mutant's affinity for the target protein. A string of five GLY residues (GG--GGG) was produced by extracting each linker residue separately in order to carry out the mutation.

A total of 3,20,00,00 mutants were generated using its tool module in Python and the resultant glycine stretch was subjected to mutation using a user-defined function written in script[40]. The number of mutants increased to 20^5 (or

3,200,000) variants when twenty amino acids were swapped out at each of the five locations in the linker strand. Following the generation of the mutant peptide stretch, full mutant peptides were assembled by connecting the initial two residues of the stretch with the second set of residues. This led to additional analysis involving the produced peptides.

Clustering: The nature of the amino acids was used to further encode the variants that were created. There are a total of twenty amino acids and they were categorized as either non-polar, aromatic, polar, positively charged, or negatively charged. The peptide sequence is shown in table 1 using single-letter codes (B, J, O, U and X). Clustering the encoded sequences yielded unique sets of peptides. By applying the elbow method[37,46], we were able to ascertain the ideal cluster size. The peptides were grouped into different numbers of clusters using the elbow method and the WCSS was computed for each cluster size.

WCSS assesses the compactness of clusters for a defined number of clusters. The k-means function of the cluster module was utilized for clustering in the sklearn package of python[30] whereas the tfidf Vectorizer function of the feature extraction module was employed for feature extraction. The matplotlib module in Python [18] was used to generate all of the plots. After clustering, the centroid of each cluster, which stood in for the entire cluster, was extracted. Amino acid sequences were obtained by decoding the three extracted centroids. The four letters B, J, O, U and X were used to construct numerous peptide sequences because they each correspond to multiple residues.

Screening: The best peptide out of all the generated sequences was screened out using a machine learning model trained using the DeepPurpose framework[17]. An epoch of 150, a learning rate of 0.001, a batch size of 16 and a hidden layer of 64 by 32 dimensions make up the model's architecture. The 344 data points of protein-peptide complexes used to train this predictive model are from the Skempi v2.0[19] dataset. The training dataset contained information about proteins, peptides and their interactions. With 70% of the data going into training, 10% into validation and 20% into testing, this dataset was split into three parts. Using DeepPurpose's conjoint triad encoder, the sequences were encoded. Prior to moving further with the training, the affinity value was adjusted. By calculating negative logarithmic values for each and dividing the result by the matching peptide sequence length, the affinity values were normalized.

We trained the model using the encoded sequence and a normalised affinity score, then employing Pearson correlation to evaluate its performance. The trained model was given the target protein sequence and produced peptide sequences in order to determine which one was the best. We proceeded with additional analysis using the two peptides with the greatest predictions. The process of using ML to choose the mutant peptides is illustrated in fig. 1.

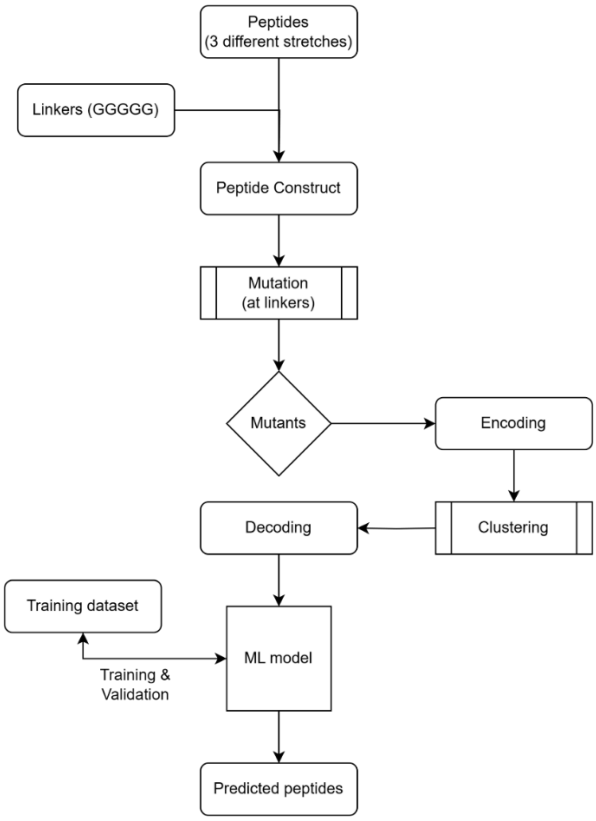


Fig. 1: Workflow of the peptide construction, mutation, clustering and screening of the peptides using ML.

Table 1
Single letter representation for each amino acid group

Amino acid group	Amino acids	Representing Single Letter Code
Non-polar	G, V, A, L, I, M	B
Non-polar Aromatic	F, Y, W	J
Polar uncharged	S, T, C, P, N, Q	O
Polar positively charged	K, R, H	U
Polar negatively charged	D, E	X

Molecular Docking: The HDock server server[49] was utilised for protein-peptide docking. Predicting the complex structure from the structures of the individual proteins is what this docking is all about. A protein's steric and physico-chemical complementarity at its interface is crucial to docking methods. This server employs a hybrid approach, integrating both template-based and template-free docking methods to predict the interactions between receptors and ligands. In this case, the docking for targeted docking with the 3D IL10 protein structure and the peptides brought attention to the binding site residues. The PepFold 3 server[23] was used to create the 3D structure of both the original peptide and the variant peptides. Additional molecular dynamics simulation research was conducted using the protein-peptide complexes following docking.

Molecular Dynamics Simulation: By employing the Gromacs 2022.4 software package[2], protein-ligand complex molecular dynamics (MD) simulations have been conducted. Top-docked poses were utilized throughout the 500 ns MD simulation. The molecular topology was generated prior to using the CHARMM36 force field[16] on the proteins and ligands. The electrostatic force over a given distance was then determined using the Particle Mesh Ewald (PME) method[47]. The system was placed in a cubic solvation box with a 1.0 nm buffer and then solvated with TIP3P water molecules[14]. Subsequently, by using Na^+ and Cl^- ions, the neutralization was performed. To remove the steric conflicts, the system 50,000 iterations of the steepest descent algorithm were performed.

The LINCS algorithm[15] was then employed to constrain bonds and ensure system stability. Subsequently, the system temperature was increased to 310 K over a 100 ps simulation in the NVT ensemble, using a timestep of 2 fs. Furthermore, the system was equilibrated in the NPT ensemble at 310 K and 1 atmosphere for 1 ns. The initial production run was 500 ns long. The Parrinello-Rahman pressure coupling method was employed to maintain constant pressure during the production run[26] while the velocity-rescaling approach[4] was used to couple the temperature. The post MD simulation analysis was performed on the visual platform called "Analogue" developed by Growdea Technologies[35,42].

MM/GBSA: The binding free energy of the protein-peptide complexes was determined using the GROMACS add-on tool gmx MM/PBSA[46]. Here, the last 30 ns of the MD simulation were used for the calculation of the binding free energy. The equations applied to compute the MM/GBSA are shown as follows:

$$\Delta G = G_{complex} - [\, G_{receptor} + G_{ligand} \,] \quad (1)$$
$$\Delta G_{binding} = \Delta H - T\Delta S \quad (2)$$
$$\Delta H = \Delta G_{GAS} + \Delta G_{SOLV} \quad (3)$$
$$\Delta G_{GAS} = \Delta E_{EL} + \Delta E_{VDWAALS} \quad (4)$$
$$\Delta G_{SOLV} = \Delta E_{GB} + \Delta E_{SURF} \quad (5)$$
$$\Delta E_{SURF} = \gamma . SASA \quad (6)$$

Here, equation 1 represents the change in Gibbs free energy (ΔG) of protein-ligand complex formation. The total free energies of the complex, free enzyme and ligand in solution are denoted by $G_{complex}$, G_{enzyme} and G_{ligand} respectively. The binding free energy (ΔG_{bind}) is the difference between the total free energy of the complex and the sum of the free energies of the unbound components. The enthalpy change (ΔH) includes contributions from gas-phase energy (ΔG_{gas}) and solvation free energy (ΔG_{solv}). The binding free energy also includes an entropy term (ΔS). Electrostatic energy (ΔE_{EL}) and van der Waals energy (ΔE_{vdw}) contribute to the total interaction energy. The solvation free energy is further divided into non-polar (ΔG_{GB}) and polar (ΔG_{surf}) components. The non-polar component is calculated based on the change in solvent-accessible surface area (SASA) and the solvent surface tension parameter (γ).

Results and Discussion

Recognition of DEGs: The GEO database's accession number, GSE62525, was used to get a gene expression profile for three types of samples: ATB (active tuberculosis), LTBI (latent tuberculosis infection) and healthy cells. It was divided into two groups for the purposes of analysis: ATB against healthy cells (group 1) and ATB versus LTBI (group 2). A comparison of the data set was conducted with the help of a box plot, which demonstrates that selected samples have comparable mean values (Fig. 2). A total of 8794 DEGs were identified in group 1, of which 371 genes were upregulated and 8423 were downregulated. Similarly, 6962 DEGs were identified in group 2, of which 504 genes were upregulated and 6458 were downregulated. In the volcano plot, the distribution of every DEG is illustrated.

The downregulated genes in both groups exhibit a greater degree of significance compared to the upregulated genes group 1 containing a greater quantity of DEGs than group 2 (Fig. 3).

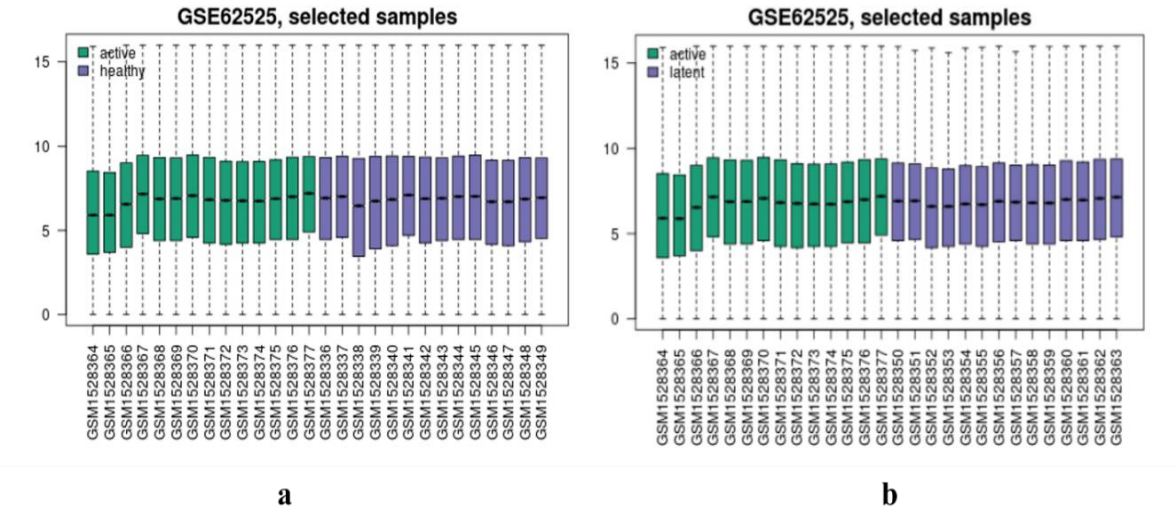


Fig. 2: Box plot indicating the average value of the samples taken (a: group 1; b: group 2)

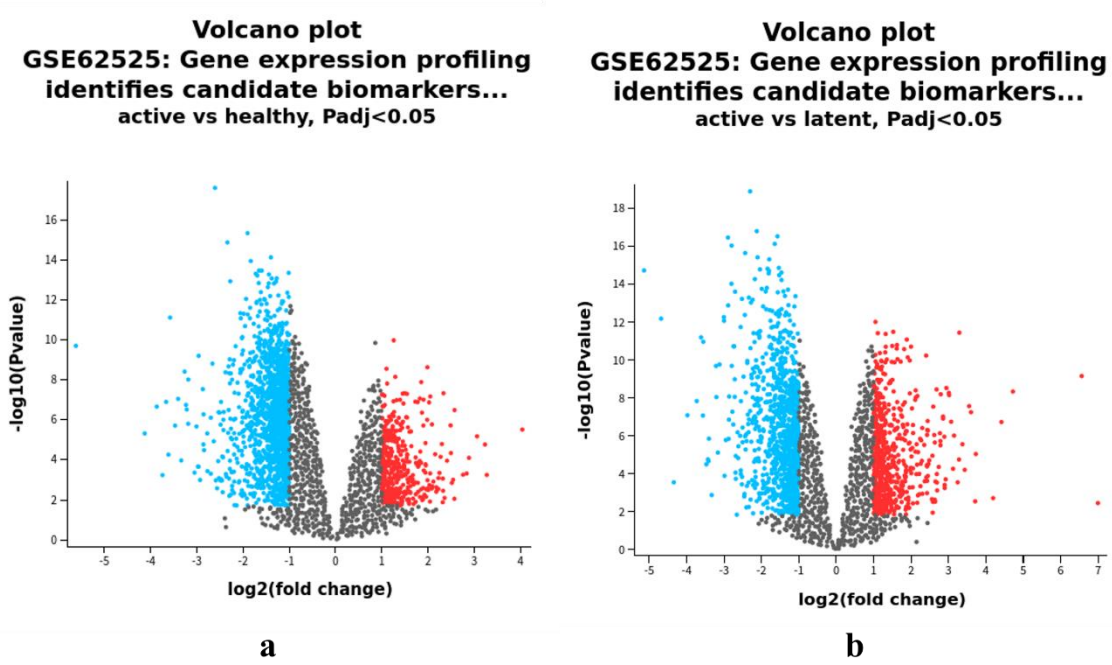


Fig. 3: Volcano plot showing the distribution of DEGs, Red colour indicates upregulated genes and blue indicates downregulated genes (a: group 1 and b: group 2)

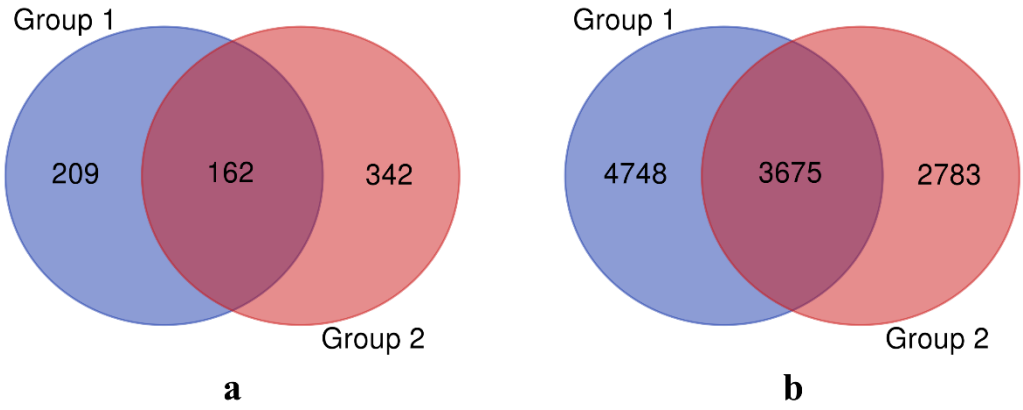
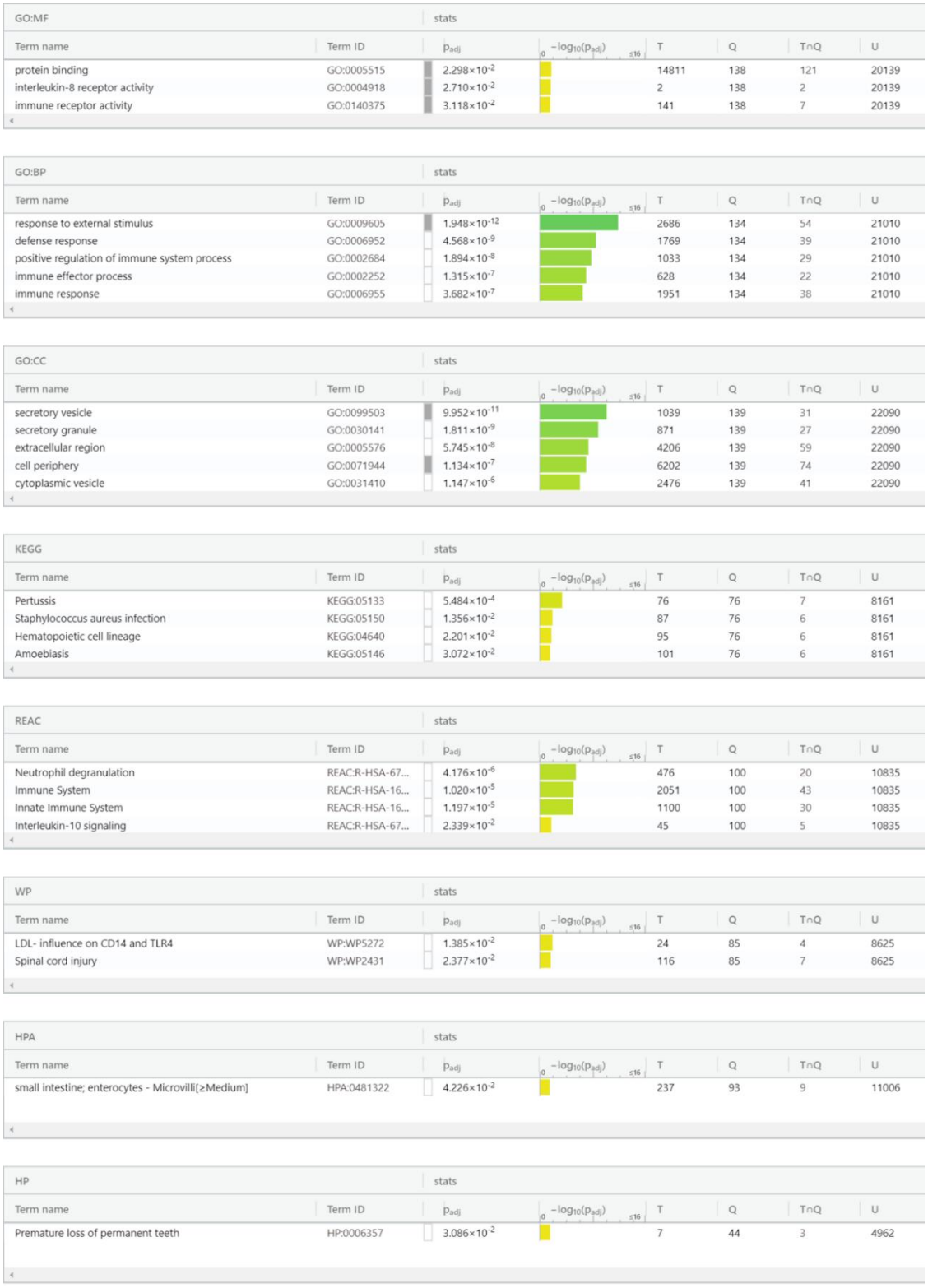


Fig. 4: Venn diagram showing the common genes (a: upregulated genes and b: downregulated genes)



GO:MF

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	T	Q	T∩Q	U
protein binding	GO:0005515	2.298×10^{-2}		14811	138	121	20139
interleukin-8 receptor activity	GO:0004918	2.710×10^{-2}		2	138	2	20139
immune receptor activity	GO:0140375	3.118×10^{-2}		141	138	7	20139

GO:BP

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	T	Q	T∩Q	U
response to external stimulus	GO:0009605	1.948×10^{-12}		2686	134	54	21010
defense response	GO:0006952	4.568×10^{-9}		1769	134	39	21010
positive regulation of immune system process	GO:0002684	1.894×10^{-8}		1033	134	29	21010
immune effector process	GO:0002252	1.315×10^{-7}		628	134	22	21010
immune response	GO:0006955	3.682×10^{-7}		1951	134	38	21010

GO:CC

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	T	Q	T∩Q	U
secretory vesicle	GO:0099503	9.952×10^{-11}		1039	139	31	22090
secretory granule	GO:0030141	1.811×10^{-9}		871	139	27	22090
extracellular region	GO:0005576	5.745×10^{-8}		4206	139	59	22090
cell periphery	GO:0071944	1.134×10^{-7}		6202	139	74	22090
cytoplasmic vesicle	GO:0031410	1.147×10^{-6}		2476	139	41	22090

KEGG

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	T	Q	T∩Q	U
Pertussis	KEGG:05133	5.484×10^{-4}		76	76	7	8161
Staphylococcus aureus infection	KEGG:05150	1.356×10^{-2}		87	76	6	8161
Hematopoietic cell lineage	KEGG:04640	2.201×10^{-2}		95	76	6	8161
Amoebiasis	KEGG:05146	3.072×10^{-2}		101	76	6	8161

REAC

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	T	Q	T∩Q	U
Neutrophil degranulation	REAC:R-HSA-67...	4.176×10^{-6}		476	100	20	10835
Immune System	REAC:R-HSA-16...	1.020×10^{-5}		2051	100	43	10835
Innate Immune System	REAC:R-HSA-16...	1.197×10^{-5}		1100	100	30	10835
Interleukin-10 signaling	REAC:R-HSA-67...	2.339×10^{-2}		45	100	5	10835

WP

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	T	Q	T∩Q	U
LDL- influence on CD14 and TLR4	WP:WP5272	1.385×10^{-2}		24	85	4	8625
Spinal cord injury	WP:WP2431	2.377×10^{-2}		116	85	7	8625

HPA

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	T	Q	T∩Q	U
small intestine; enterocytes - Microvilli[≥Medium]	HPA:0481322	4.226×10^{-2}		237	93	9	11006

HP

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	T	Q	T∩Q	U
Premature loss of permanent teeth	HP:0006357	3.086×10^{-2}		7	44	3	4962

Fig. 5: Gene enrichment of the common DEGs with initial five log values using gprofiler

A Venn diagram is produced to illustrate the common genes showing that there are 162 upregulated genes and 3675 downregulated genes (Fig. 4).

Pathway Enrichment Analysis: Gene enrichment utilising gprofiler[31] yielded the KEGG and GO pathways. The analysis used the common genes that were upregulated (162 upregulated genes) and the initial five processes with the highest log values were selected (Fig. 5). The molecular functions (MF) domain emphasises significant functions such as protein binding, interleukin-8 receptor activity and immune receptor activity, showcasing the complex interactions of proteins in immune responses. The biological activities (BP) that have emerged, include response to

external stimuli, defence response, positive regulation of immune system activities, immune effector process and immunological response, illustrating the body's advanced mechanisms to identify and react to external threats.

The cellular components (CC) study identified key locations for immunological activity, such as secretory vesicles, secretory granules, the extracellular region, the cell periphery and cytoplasmic vesicles. KEGG pathways including pertussis, *Staphylococcus aureus* infection, haematopoietic cell lineage and amoebiasis, as well as REAC pathways such as neutrophil degranulation, the immune system, the innate immune system and interleukin-10 signalling, were discovered, demonstrating the range of diseases and immune responses under investigation. WikiPathways emphasised the impact of LDL on CD14 and TLR4 and the biological processes related to spinal cord injury. Meanwhile, the Human Protein Atlas (HPA) concentrated on the shape of enterocytes in the small intestine, particularly highlighting microvilli.

The Human Phenotype Ontology (HP) identified premature loss of permanent teeth as a manifestation of the genetic basis of this illness. This thorough research highlights the interrelationship between genetic expressions, pathways and physiological responses, providing profound insights into the intricate dynamics of health and disease.

Hub Gene Identification: The common upregulated genes were employed for protein-protein interaction (PPI) utilising STRING. The network has significantly more interactions than expected. Out of a total of 162 common genes, 149 formed connections with 149 nodes, resulting in the identification of 319 edges (Fig. 6). The nodes represent individual proteins or genes and the lines (edges) indicate the relationships or interactions between them. The different colours of the nodes and edges often represent various types of interactions or classifications such as activation, inhibition, or different functional groups. Further, the degree method was utilised to identify the top 10 hub genes using the Cytohubba plug-in of the Cytoscape.

The top 10 hub genes found in the Cytohubba were listed in Table 1, along with their scores determined by the degree method. The top score was observed for IL1B with a score of 42, while TLR4 and MMP9 had a score of 31 and IL10 had a score of 30. Fig. 7 shows the interactions between these top 10 genes. The interactions among various proteins, cytokines, chemokines and receptors that regulate immune responses were visually represented in this network. IL1B and IL10 are cytokines that exhibit contrasting effects during inflammation; MMP9 is an enzyme that aids in immune cell migration and degrades the extracellular matrix and TLR4 is a receptor involved in pathogen recognition.

FCGR1A receives the Fc segment of IgG antibodies as a receptor; CXCR1 and CXCR2 are receptors for chemokines. The protein CD274/PD-L1 is capable of impeding the immune response in order to avert tissue damage. The enzyme ARG1 can regulate immune responses through arginine depletion while the receptor TREM1 can enhance inflammatory responses. The top 10 hub genes (IL1B, TLR4, MMP9, IL10, FCGR1A, CXCR2, CXCR1, CD274, ARG1 and TREM1) were searched with the objective of identifying a target protein. Several genes among them exhibited potential as therapeutic targets or as prognostic indicators for TB determination.

Table 2
Top 10 hub genes ranked by Degree method with their scores

Rank	Gene	Score
1	IL1B	42
2	TLR4	31
2	MMP9	31
4	IL10	30
5	FCGR1A	24
6	CXCR2	21
6	CXCR1	21
8	CD274	19
9	ARG1	17
10	TREM1	14

Table 3
Binding free energy components for the protein-peptide complexes using MM/GBSA technique.

System	ΔVDWAALS	ΔEEL	ΔENPOLAR	ΔGGAS	ΔGSOLV	ΔTOTAL
Native	0.00 ± 0.00	-0.01± 0.02	-0.00 ± 0.00	-0.01 ± 0.02	-0.00 ± 0.00	-0.01 ± 0.02
Variant-I-CG_KYC	-69.73 ± 7.20	12.83 ± 14.11	-11.23 ± 0.36	-56.90 ± 13.55	-11.23 ± 0.36	-68.13 ± 13.68
Variant-II-CV_RYC	-104.30 ± 7.25	22.54 ± 14.61	-13.88 ± 0.40	-81.76 ± 11.96	-13.88 ± 0.40	-95.64 ± 11.86

Nevertheless, with regard to their functions in the immune response against Mycobacterium tuberculosis, IL-10 (interleukin-10) emerges as a notably promising target for therapeutic intervention. IL-10 is an anti-inflammatory cytokine that is essential for preventing tissue damage to the host by regulating the immune response[33]. IL-10 has the potential to inhibit immune responses in the context of tuberculosis, thereby enabling the bacteria to elude elimination and sustain a dormant infection[45]. IL-10 levels that are elevated, have been linked to a heightened vulnerability to tuberculosis and to more severe consequences of the disease.

A potential therapeutic approach for tuberculosis involves the modulation of IL-10 activity or its associated signalling pathways, as it could strengthen the immune system's capacity to eradicate the infection. Potentially, the efficacy of the host immune response against Mycobacterium tuberculosis could be enhanced through the reduction of IL10-mediated immunosuppression.

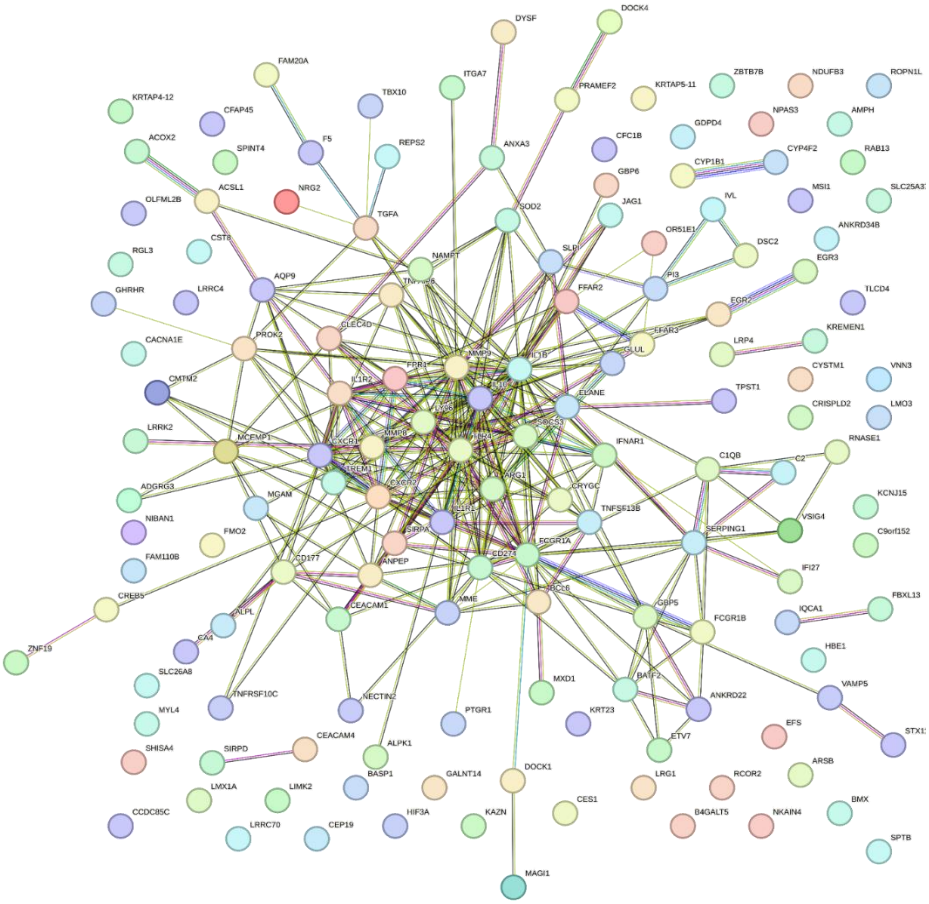


Fig. 6: Protein-Protein Interactions (PPI) network of shared upregulated DEGs.

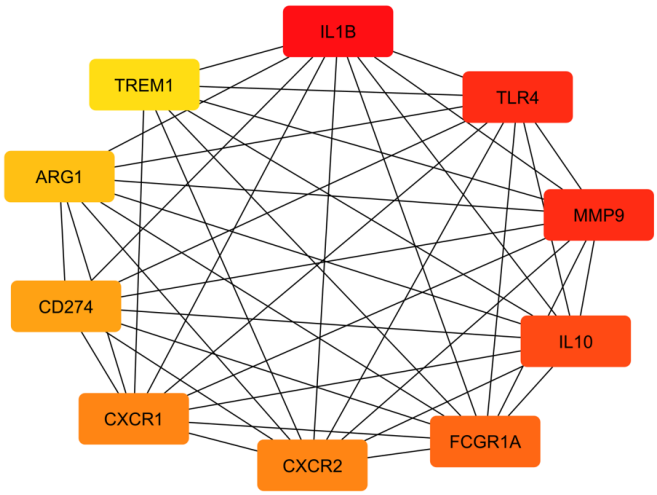


Fig. 7: Interaction of hub genes found from the cytohubba in the Cytoscape.

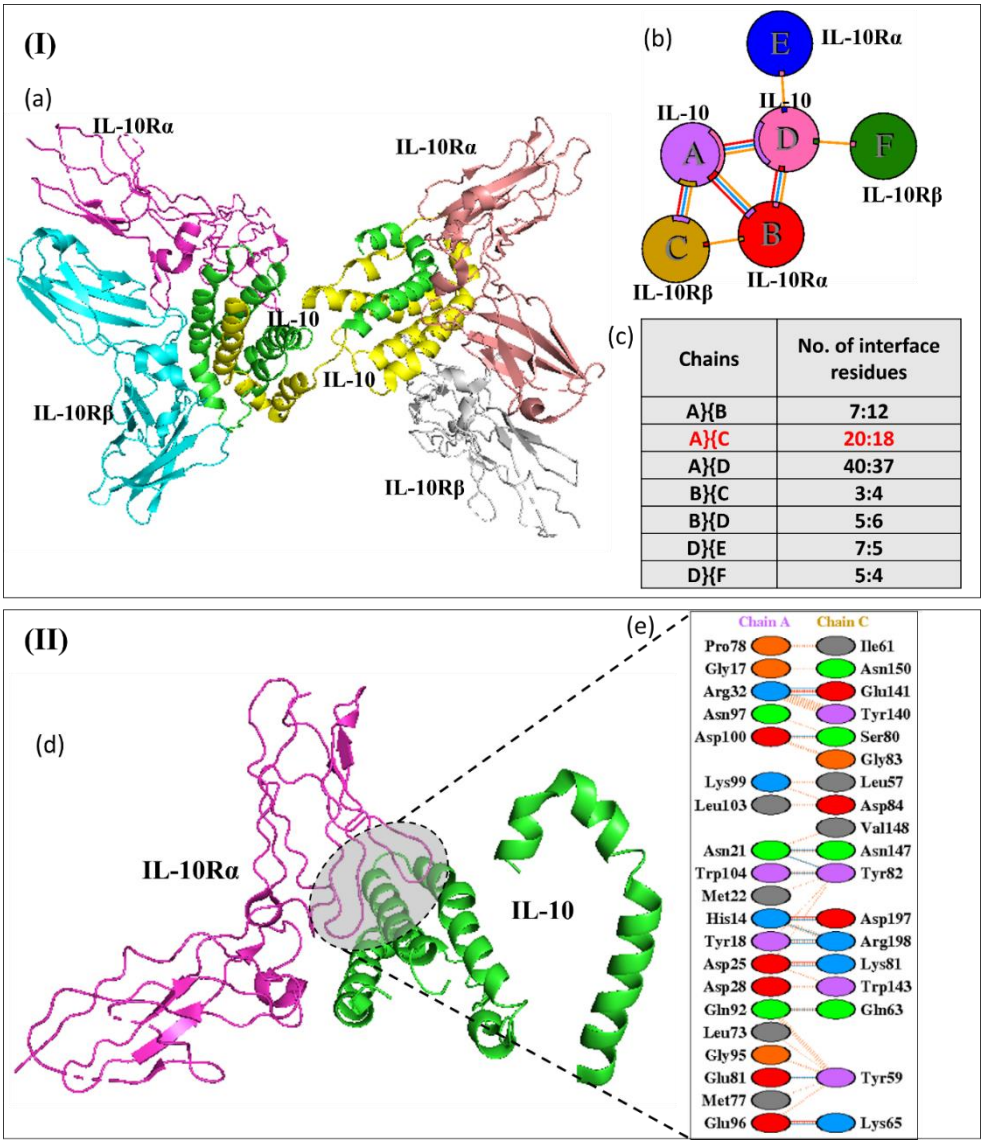


Fig. 8: (I) PDB structure visualization, (a) 3D structure of IL10 complexed with both IL-10Rα and IL-10Rβ (b) Interacting chains (c) Number of interface residues; (II) Interaction analysis; (d) of 3D structure of IL10 complexed with IL-10Rβ (e) Interacting residues between IL10 and IL-10Rβ.

The biological activity of IL-10, especially its interactions with the IL-10 receptor complex, depends on its dimeric form[20,21]. Two alpha units (IL-10Rα) that bind to the IL-10 dimer and two beta units (IL-10Rβ) required for signal transduction make up this receptor complex[28]. IL-10's interaction with its receptor complex sets off a series of intracellular signalling events that ultimately lead to the cytokine's immunoregulatory effects, which include immune cell activity modulation and suppression of inflammatory responses[11,28,32]. Preventing signal transduction by impeding IL-10's interaction with the IL-10 receptors might constitute a significant strategy. This may be accomplished via the development of antibodies that inhibit the receptor or molecules that mimic the binding sites of IL-10 on the receptor.

Interface Residue: The protein data bank (PDB) was queried with PDB ID 6X93 to acquire the three-dimensional crystal structure of the chosen protein (IL10 complex). Parts I and II of fig. 8 show the results of the thorough analysis of the structure. The observation of the 1L10 complex binding to both IL-10Rα and IL-10Rβ can be seen in fig. 8(a). Fig. 8(b) shows the results of an analysis of the PDBsum database that revealed the chains' interconnections. Following the IL10 dimer chain A-D in terms of the number of contacts with interface residues, chain A-C exhibited the largest number of interactions (Fig. 8(c)). Therefore, the interaction between IL-10 (chain A) and IL-10Rβ (chain C) was greatest with 20:18 residues. Fig. 8(d, e) shows the interaction residues between IL10 and IL-10Rβ, which were further examined using the A-C chain. The interaction between chain A (IL10) and chain C (IL-10Rβ) involved 18 residues.

After that, the PDBePISA tool was used to analyze the chain A-C complex and forecast the residues at the interface. For peptide design, we looked for and chose residues that formed a continuous structure. Interface chain C residues were measured by PDBePISA: 57–61, 63, 65, 78–85, 106, 108–

109, 138, 140–143, 147–150 and 197–198. The following sequences of amino acids were designed into peptides: 57–65, 78–85, 138–150, LSYRIFQDK, SLSKYGDH and NEYETWTMKNVYN, in that order. The three peptide distances (P1, P2 and P3) were plotted in fig. 9. The measurement of 6.1 Å was taken for the distance between P1 residues L and P2 residues H. The 8.9 Å distance was measured between P2's S and P3's N. Typically, the distance between neighbouring residues in a fully expanded polypeptide chain containing Cα atoms is approximately 3.8 Å.

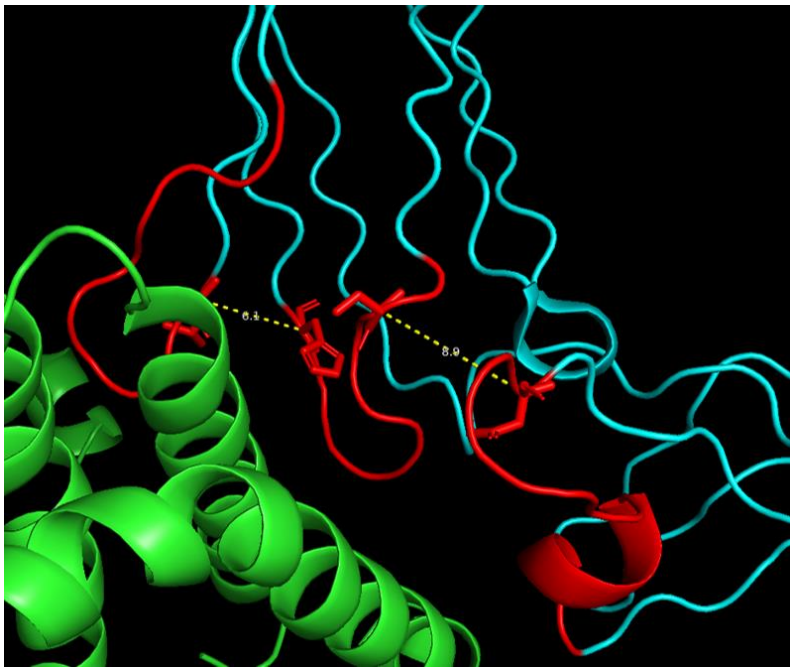


Fig. 9: Distance between the three peptides visualized in the Pymol visualizing tool.

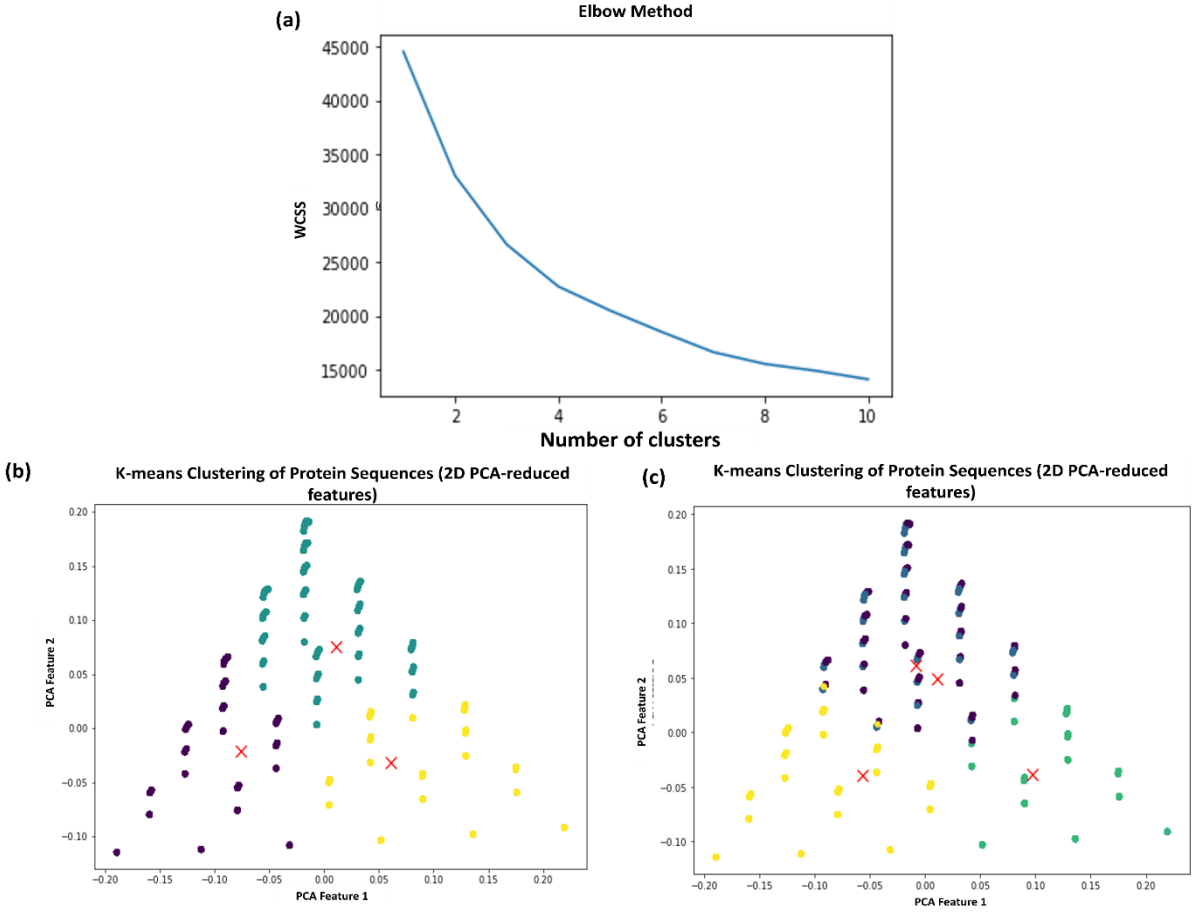


Fig. 10: (a) Elbow method plot (b) K-means clustering for 3 clusters (c) K-means clustering for 4 clusters

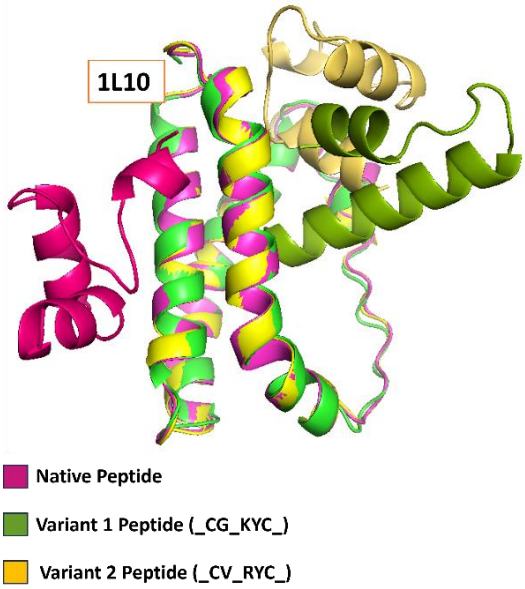


Fig. 11: Peptides (native, variant-I and variant-II) docked with the protein IL10 aligned with each other represented in 3D structure.

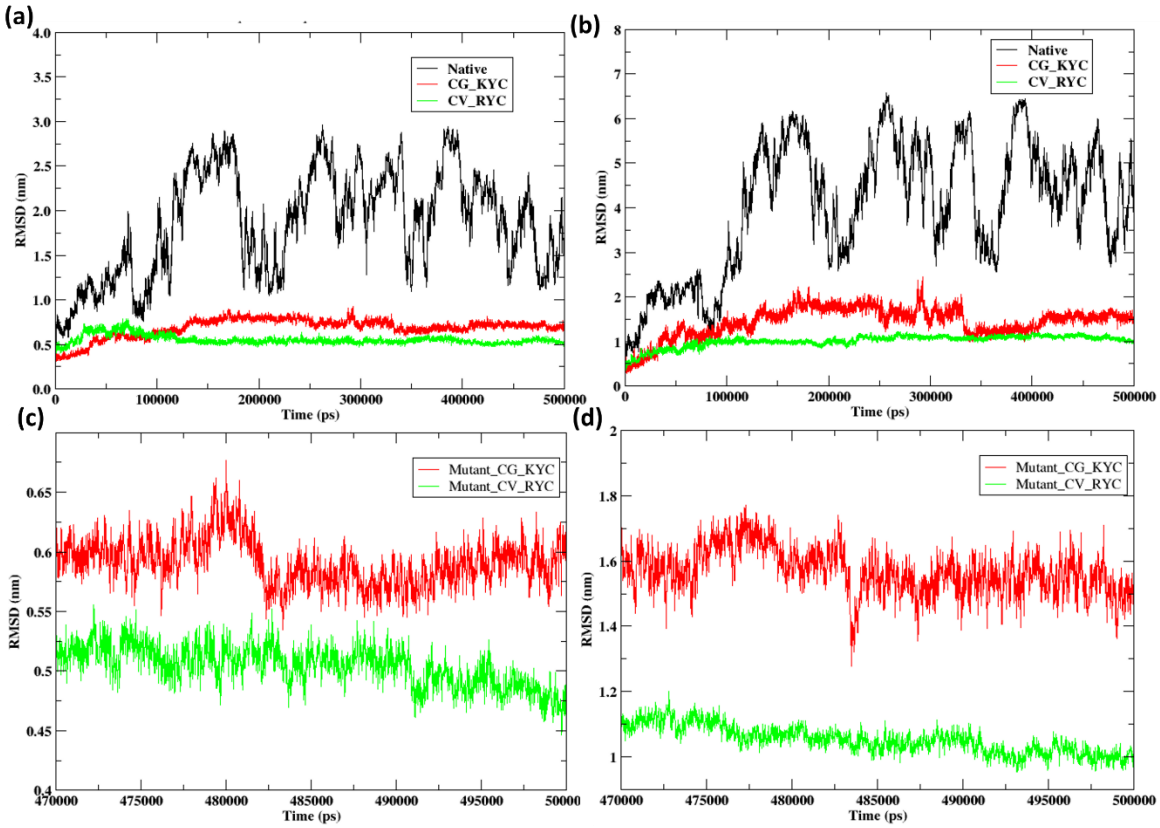


Fig. 12: RMSD of the protein-peptide complexes, (a) RMSD of the protein Cα atoms during the 500 ns simulation (b) RMSD of the ligands during the 500 ns simulation (c) RMSD of the protein Cα atoms for the last 30 ns simulation (d) RMSD of the ligands for the last 30 ns simulation

Therefore, the first gap was connected using two residues (GG), while the second gap was connected using GGG. The final product is a peptide with 35 residues, composed of 30 primary amino acids and 5 secondary amino acids that serve as linkers. Because of its tiny size and flexibility, glycine is frequently utilized as a linker in peptides and proteins. This is because it can help to ensure that functional domains are folded correctly or that different protein domains can bind to each other without causing much steric hindrance. Here is the final peptide sequence with the linkers added: KDQFIRYSL**GG**HDGYKSLS**GGG**NEYETWTMKNVYN. This sequence was considered the native peptide sequence that could bind to IL10. Further, this sequence was taken to generate mutations in the linkers for better binding with the target protein IL10.

Mutation and Clustering: The purpose of performing targeted modifications at these linker residues was to determine which ones would improve the peptide-protein interaction the most. The next thing to do was to see how well each mutant is bound to the target protein. Multiple mutations were carried out at each linker residue to alter its function. There were 20^5 (or 320,000) mutations in all. A process involving peptide conversion according to amino acid residues was used to create the mutant peptides. There are five distinct types of amino acid residues distinguished by their chemical properties: polar uncharged (pol), positively charged (pos), negatively charged (neg) and non-polar aromatic (NPAr). The side chain characteristics of amino acids, which impact their behaviour and interaction in proteins, are the basis for this classification.

The peptides' amino acid sequences were transformed using the one-letter codes according to the category that each residue belonged to. This transformation made the sequences easier to understand by classifying the amino acids into five separate groups according to their chemical characteristics. A k-means clustering analysis was subsequently performed on the reduced peptide sequences. A vector quantization approach seeks to group 'n' observations into 'k' groups, with the goal that each observation should be part of the group with the closest mean. This method finds groupings or clusters in the data, with a predetermined number of clusters (k).

The elbow method is utilised to determine the optimal number of clusters (k) for the k-means clustering algorithm. Calculating the WCSS for different values of k is integral to this procedure. The objective is to reduce the WCSS value, which quantifies the variance within each cluster. On the other hand, the WCSS tends to go down to zero as the number of clusters grows. The decline rate changes dramatically at the "elbow" position on the WCSS versus cluster number plot. Having a lesser number of clusters and minimizing the WCSS, are both achieved at this ideal location. There is a limit beyond which further clustering does not significantly increase the variance explained, a phenomenon known as diminishing returns.

To summarize, this approach of grouping and simplifying peptide sequences for cluster analysis is done quickly by applying k-means clustering. The elbow technique is beneficial for determining the optimal number of clusters by identifying the optimal balance between model complexity and clustering granularity.

Many people believe that the sweet spot for cluster size is when the curve begins to flatten out, creating an "elbow." The elbow plot indicated the possibility of 2-4 cluster formation, as demonstrated in fig. 10(a). Two clusters are insufficient; therefore, we created and plotted three and four clusters instead. The plots for 3 clusters with centroids and 4 clusters with centroids are displayed in fig. 10(b, c). The data was clustered using three clusters instead of the four projected clusters because two of them were found to be quite similar.

The centroids of these 3 clusters were extracted and the sequences are as follows:

UXOJBUJOBOBUXBJUOBOUJOOXJXOJOBUOBJO
UXOJBUJOBOBUXBJUOBOUJXOXJXOJOBUOBJO
UXOJBUJOBBBUXBJUOBOJBOOXJXOJOBUOBJO

The mutant region (position that can be mutated) is highlighted in yellow in each sequence. These centroids were decrypted/decoded back to get the amino acid sequences. Further, the mutated peptides were generated based on three centroids and the total number of sequences generated were 6480. These 6480 mutant peptides were used for virtual screening using machine learning.

Virtual Screening: All 6480 peptide sequences were screened to identify the most suitable peptide. DeepPurpose Architecture, which is a machine learning (ML) approach, was used to develop the model. The parameters were optimised and the screening proceeded. The best encoder was found to be Conjoint_triad, its R-squared on the test dataset was 0.966 and its Pearson correlation was 0.983; thus, it was used for the screening of 6480 peptides. The screening resulted in two best-predicted peptides: variant-I peptide – KDQFIRYSL**CG**HDGYKSLS**KYC**NEYETW TMKNVYN and variant-II peptide – KDQFIRYSL**CV**H DGYKSLS**RYC**NEYETWTMKNVYN. These two mutant peptides were used for molecular docking along with the native peptide – KDQFIRYSL**GG**HDGYKSLS **GGG**NE YETWTMKNVYN. Here, the highlighted red colour in the peptide sequences shows the mutated region.

Molecular Docking: Using the PepFold 3 server, we were able to predict the three-dimensional structures of both the native peptide and the constructed/mutant peptides. The Hdock server was used to dock both normal and mutant peptides with IL-10 chain A. The following residues were chosen for targeted docking: 13–14:A, 17–18:A, 20–22:A, 24–25:A, 28–29:A, 32–A, 73–74:A, 77–78:A, 81–92:A, 95–97:A, 99–100:A, 102–104:A, 107:A. The protein-peptide complexes that were docked, were shown to be positioned over each other in fig. 11. The docking scores for the three mutants were as follows: -118.28 kcal/mol for native, -226.43 kcal/mol for variant-I and -215.38 kcal/mol for variant-II.

The protein-protein complexes included in the Protein Data Bank typically have docking scores of approximately -200 kcal/mol or higher, as per hdock docking. In this case, the dock score was higher for the mutant peptides compared to the normal peptide. This suggests that the synthetic peptide has a higher binding affinity for the target protein IL10 compared to the natural peptide. Additional supplemental fig. S1 shows the LigPlot+ visualization of the protein-peptide interaction.

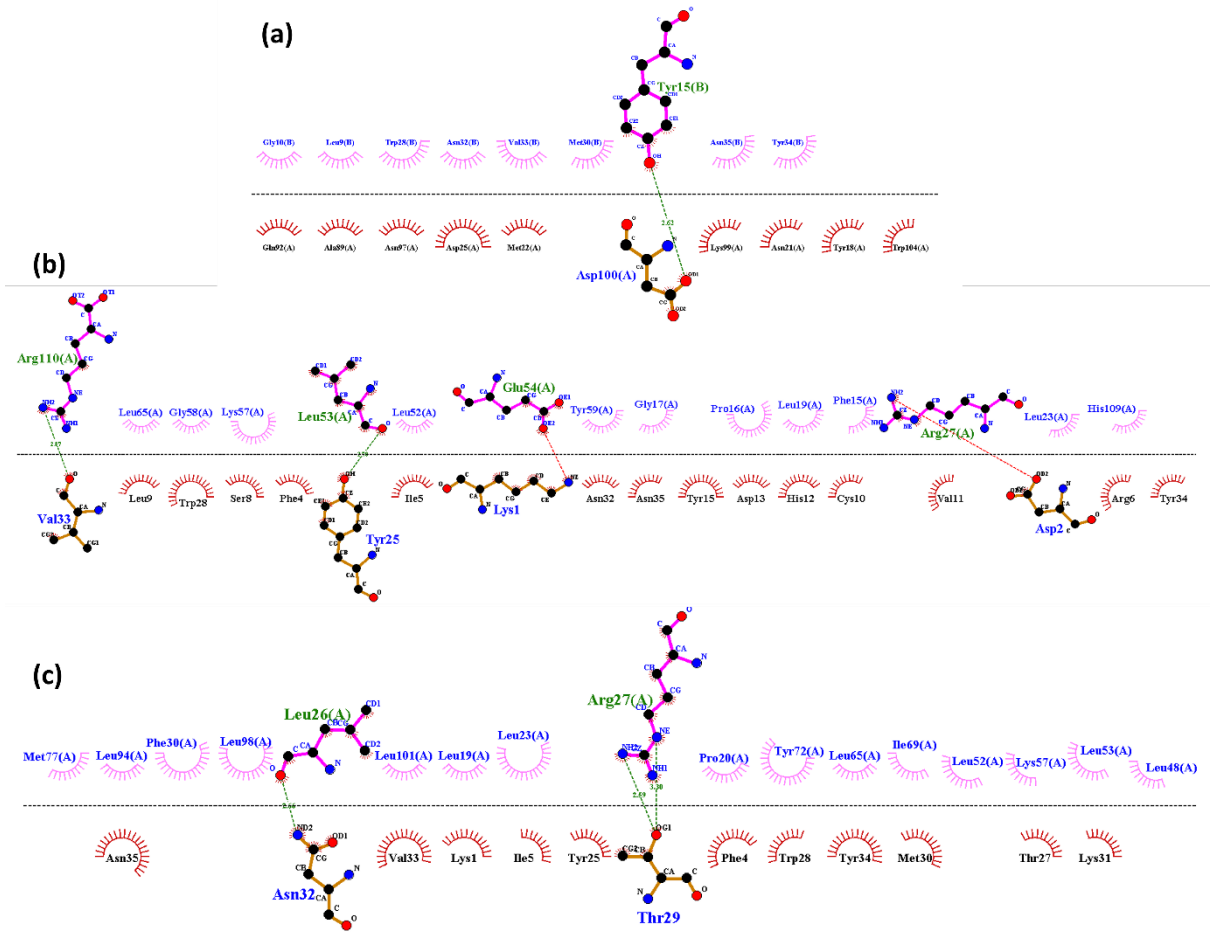


Figure S1: 2D interaction between the protein-peptide complexes for (a) Native peptide, (b) Mutant 1 peptide and (c) Mutant 2 peptide

One hydrogen bond with IL10 (Asp100) was found in the natural peptide (Tyr15). Asp2-Arg27, Lys1-Glu54, Tyr25-Leu53 and Val33-Arg110 were the peptide-protein residues that variant-I interacted with in IL10. In variant-II, IL10's Arg27 and Leu26 interacted with Thr29 and Asn32 respectively. Additionally, the docked protein-peptide complexes were simulated using molecular dynamics software.

Molecular Dynamics Simulation: Molecular dynamics simulations were conducted to examine the stability and flexibility of the protein-peptide complexes. The root mean square deviation (RMSD) was calculated to assess the conformational change that happens upon the binding of proteins and peptides. As illustrated in fig. 12(a, c), the relative mechanical strain (RMSD) of the Cα atoms of the protein when attached to the peptides is displayed. Fig. 12(b, d) shows the results of reporting RMSD for the peptides, while the protein molecule was utilized for fitting and aligning the structures. It is acceptable to have RMSD values between 0.1 and 0.3 nm, or 1-3 Å, particularly for small and globular proteins. Significant changes in structural conformation are indicated by deviations from this range.

The stability of the protein-ligand complex is positively connected with a reduced variation in the RMSD during the molecular dynamics simulation whereas a higher fluctuation implies a less stable protein-ligand complex [3,29]. The native peptide-bound protein exhibited a high RMSD of up to 2-3 nm throughout the 500 ns simulation. The RMSD of the protein bound to the two mutant peptides was significantly lower than that of the native (Fig. 12a). When attached to the two mutant peptides, both proteins displayed RMSD values between 0.5 and 0.75 nm, suggesting that they were quite stable. During the 500 ns simulation, the RMSD of the native peptides was 5-6 nm whereas the other peptides followed a similar pattern.

Fig. 12(b) shows that in comparison to the normal peptide, the variant peptides exhibited steady protein binding with RMSD values ranging from 1 nm to 1.5 nm. For an in-depth study, fig. 12(c, d) displayed the RMSD protein and mutant peptides for the last 30 ns of the simulation. It was noted that the protein-peptide bond remained stable for the majority of the remaining 30 ns.

Fig. 12(c) shows that in contrast, over the last 30 ns of the simulation, the RMSD of the protein attached to the variant-I peptide dropped from 0.65 nm to 0.6 nm and then remained steady and constant. In the last 30 ns of the simulation, the RMSD of the protein bound to variant-II peptide decreased from 0.5 nm to 0.45 nm without any notable fluctuations.

In the last 30 ns of the simulation, the RMSD of the variant-I peptide decreased from 1.6 nm to 1.5 nm, as seen in fig. 12(d). Variant-II peptide showed a RMSD decrease of 1.1 nm to 1 nm for the last 30 ns, suggesting a more stable conformation on binding to the protein than variant-I. Overall, when bound to the IL10 protein, the mutant peptides were more stable than the normal peptide.

MM/GBSA: The MM/GBSA method was also used to determine the protein-peptide complexes' binding free energies. All of the protein-peptide complex binding free energy components are given in table 2. The stronger is the binding affinity between the molecules involved, the lower the binding free energy value must be; a negative value indicates an advantageous and spontaneous binding process. A weak interaction with the protein was indicated by the binding free energy of -0.01 ± 0.02 kcal/mol that the natural peptide displayed. A high score usually means that the binding process demands energy; this suggests that it may have an unfavourable interaction with IL10.

In contrast to the original protein, the binding free energy of the two mutant peptides was significantly higher. Binding free energies of -68.13 ± 13.68 kcal/mol and -95.64 ± 11.86 kcal/mol were observed for variant-I peptide (CG_KYC) and variant-II peptide (CV_RYC) respectively in this context. These numbers point to spontaneously higher binding affinities to IL10, with variant-II displaying the higher affinity of the two.

It has been found that mutations can boost binding affinity which means that peptide sequences can be accurately modified to improve target protein interaction. When it comes to developing therapies, this could have major ramifications because high-affinity peptides are powerful regulators of protein function. Further research into therapies targeting IL10 or similar pathways may find the mutant peptides to be intriguing candidates due to their strong binding capabilities.

Conclusion

Researchers found that peptide linker alterations significantly increased binding affinity to the IL10 protein which could be used as a therapeutic intervention in tuberculosis. The work conducted molecular dynamics simulations and virtual screening to validate the predictions of the efficacy of mutant peptides that were made using machine learning techniques. The results show that after interacting with IL10, the mutant peptides were more stable and had higher binding affinities than the normal peptide. The maximum affinity and stability were observed in the variant-II peptide (CV_RYC), suggesting that it could be a promising option for future therapeutic research.

New tuberculosis medicines can be developed by merging network biology, computational biology, machine learning and molecular dynamics simulations. The study demonstrates the usefulness of this approach in discovering and improving physiologically active peptides.

Acknowledgement

Authors would like to express their gratitude to Growdea Technologies Pvt. Ltd. for their invaluable contributions in providing key insights into the presented research study.

References

1. Barrett T., Wilhite S.E., Ledoux P., Evangelista C., Kim I.F., Tomashevsky M., Marshall K.A., Phillippy K.H., Sherman P.M., Holko M. and Yefanov A., NCBI GEO: archive for functional genomics data sets—update, *Nucleic Acids Research*, **41**, D991–D995 **(2013)**

2. Bauer P., Hess B. and Lindahl E., GROMACS 2022.4 Manual, [accessed 2023 Oct 14], https://zenodo.org/record/7323409, doi:10.5281/ZENODO.7323409 **(2022)**

3. Bhowmick S., AlFaris N.A., ALTamimi J.Z., ALOthman Z.A., Aldayel T.S., Wabaidur S.M. and Islam M.A., Screening and analysis of bioactive food compounds for modulating the CDK2 protein for cell cycle arrest: Multi-cheminformatics approaches for anticancer therapeutics, *Journal of Molecular Structure*, **1216**, 128316 **(2020)**

4. Bussi G., Donadio D. and Parrinello M., Canonical sampling through velocity rescaling, *The Journal of Chemical Physics*, **126**, 014101 **(2007)**

5. Cadena A.M., Fortune S.M. and Flynn J.L., Heterogeneity in tuberculosis, *Nature Reviews, Immunology*, **17(11)**, 691–702 **(2017)**

6. Chen S.H., Chin C.H., Wu H.H., Ho C.W., Ko M.T. and Lin C.Y., Cyto-Hubba: A Cytoscape plug-in for hub object analysis in network biology, In 20th international conference on genome informatics Citeseer **(2009)**

7. Colangeli R., Gupta A., Vinhas S.A, Chippada Venkata U.D., Kim S., Grady C., Jones-López E.C., Soteropoulos P., Palaci M, Marques-Rodrigues P. and Salgame P., Mycobacterium tuberculosis progresses through two phases of latent infection in humans, *Nature Communications*, **11(1)**, 4870 **(2020)**

8. Coscolla M. and Gagneux S., Consequences of genomic diversity in Mycobacterium tuberculosis, In Seminars in Immunology, **26(6)**, 431–444 **(2014)**

9. Daniel T.M., The history of tuberculosis, *Respiratory Medicine*, **100(11)**, 1862–1870 **(2006)**

10. Dheda K. et al, The epidemiology, pathogenesis, transmission, diagnosis and management of multidrug-resistant, extensively drug-resistant and incurable tuberculosis, *The Lancet Respiratory Medicine*, **5(4)**, 291–360 **(2017)**

11. Finbloom D.S. and Winestock K.D., IL-10 induces the tyrosine phosphorylation of tyk2 and Jak1 and the differential assembly of STAT1 alpha and STAT3 complexes in human T cells and monocytes, *Journal of Immunology*, **155(3)**, 1079–1090 **(1995)**

12. Franz M., Lopes C.T., Fong D., Kucera M., Cheung M., Siper M.C., Huck G., Dong Y., Sumer O. and Bader G.D., Cytoscape.js 2023 update: a graph theory library for visualization and analysis, *Bioinformatics*, **39(1)**, btad031 **(2023)**

13. Goletti D., Petruccioli E., Joosten S. and Ottenhoff T., Tuberculosis Biomarkers: From Diagnosis to Protection, *Infectious Disease Reports*, **8(2)**, 6568 **(2016)**

14. Harrach M.F. and Drossel B., Structure and dynamics of TIP3P, TIP4P and TIP5P water near smooth and atomistic walls of different hydroaffinity, *The Journal of Chemical Physics*, **140(17)**, 174501 **(2014)**

15. Hess B., Bekker H., Berendsen H.J.C. and Fraaije J.G.E.M., LINCS: A linear constraint solver for molecular simulations, *Journal of Computational Chemistry*, **18(12)**, 1463–1472 **(1997)**

16. Huang J. and MacKerell A.D., CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data, *Journal of Computational Chemistry*, **34(25)**, 2135–2145 **(2013)**

17. Huang K., Fu T., Glass L.M., Zitnik M., Xiao C. and Sun J., DeepPurpose: a deep learning library for drug–target interaction prediction, *Bioinformatics*, **36(22–23)**, 5545–5547 **(2021)**

18. Hunter J.D., Matplotlib: A 2D Graphics Environment, *Computing in Science & Engineering*, **9(3)**, 90–95 **(2007)**

19. Jankauskaite J., Jiménez-García B., Dapkunas J., Fernández-Recio J. and Moal I.H., SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation, *Bioinformatics*, **35(3)**, 462–469 **(2019)**

20. Josephson K., Logsdon N.J. and Walter M.R., Crystal structure of the IL-10/IL-10R1 complex reveals a shared receptor binding site, *Immunity*, **15(1)**, 35–46 **(2001)**

21. Kotenko S.V., Krause C.D., Izotova L.S., Pollack B.P., Wu W. and Pestka S., Identification and functional characterization of a second chain of the interleukin-10 receptor complex, *The EMBO Journal*, **16(19)**, 5894–5903 **(1997)**

22. Krissinel E. and Henrick K., Inference of macromolecular assemblies from crystalline state, *Journal of Molecular Biology*, **372(3)**, 774–797 **(2007)**

23. Lamiable A., Thévenet P., Rey J., Vavrusa M., Derreumaux P. and Tufféry P., PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex, *Nucleic Acids Research*, **44(W1)**, W449–W454 **(2016)**

24. Lee B.S. and Pethe K., Telacebec: an investigational antibacterial for the treatment of tuberculosis (TB), *Expert Opinion on Investigational Drugs*, **31(2)**, 139–144 **(2022)**

25. Li L., Lei Q., Zhang S., Kong L. and Qin B., Screening and identification of key biomarkers in hepatocellular carcinoma: Evidence from bioinformatic analysis, *Oncology Reports*, **38(5)**, 2607–2618 **(2017)**

26. Martoňák R., Laio A. and Parrinello M., Predicting Crystal Structures: The Parrinello-Rahman Method Revisited, *Physical Review Letters*, **90(7)**, 075503 **(2003)**

27. Mokrousov I., Pasechnik O., Vyazovaya A., Yarusova I., Gerasimova A., Blokh A. and Zhuravlev V., Impact of pathobiological diversity of Mycobacterium tuberculosis on clinical features and lethal outcome of tuberculosis, *BMC Microbiology*, **22(1)**, 50 **(2022)**

28. Moore K.W., de Waal Malefyt R., Coffman R.L. and O'Garra A., Interleukin-10 and the interleukin-10 receptor, *Annual Review of Immunology*, **19**, 683–765 **(2001)**

29. Patel H.M., Ahmad I., Pawara R., Shaikh M. and Surana S., In silico search of triple mutant T790M/C797S allosteric inhibitors to conquer acquired resistance problem in non-small cell lung cancer (NSCLC): a combined approach of structure-based virtual screening and molecular dynamics simulation, *Journal of Biomolecular Structure and Dynamics*, **39(4)**, 1491–1505 **(2021)**

30. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V. and Vanderplas J., Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, **12**, 2825–2830 **(2011)**

31. Raudvere U., Kolberg L., Kuzmin I., Arak T., Adler P., Peterson H. and Vilo J., g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update), *Nucleic Acids Research*, **47(W1),** W191–W198 **(2019)**

32. Saraiva M. and O'Garra A., The regulation of IL-10 production by immune cells, *Nature Reviews Immunology*, **10(3)**, 170–181 **(2010)**

33. Saxton R.A., Tsutsumi N., Su L.L., Abhiraman G.C., Mohan K., Henneberg L.T., Aduri N.G., Gati C. and Garcia K.C., Structure-based decoupling of the pro- and anti-inflammatory functions of interleukin-10, *Science*, **371(6535)**, eabc8433 **(2021)**

34. Schwartz N.G., Hernandez-Romieu A.C., Annambhotla P., Filardo T.D., Althomsons S.P., Free R.J., Li R., Wilson W.W., Deutsch-Feldman M., Drees M. and Hanlin E., Nationwide tuberculosis outbreak in the USA linked to a bone graft product: an outbreak report, *The Lancet Infectious Diseases*, **22(11)**, 1617–1625 **(2022)**

35. Sim(Ana), Analogue Release 2024, Growdea Technologies Pvt. Lt., v1.1, https://growdeatech.com/Analogue/ **(2024)**

36. Singh R., Dwivedi S.P., Gaharwar U.S., Meena R., Rajamani P. and Prasad T., Recent updates on drug resistance in *Mycobacterium tuberculosis*, *Journal of Applied Microbiology*, **128(6)**, 1547–1567 **(2020)**

37. Syakur M.A., Khotimah B.K., Rochman E.M.S. and Satoto B.D., Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster, IOP Conference Series: Materials Science and Engineering, **336(1)**, 012017 **(2018)**

38. Szklarczyk D., Kirsch R., Koutrouli M., Nastou K., Mehryary F., Hachilif R., Gable A.L., Fang T., Doncheva N.T., Pyysalo S. and Bork P., The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest, *Nucleic Acids Research*, **51(D1)**, D638–D646 **(2023)**

39. Talukdar Prasenjit, Akhter Sazia, Dey Kishan Kumar and Hazarika Nabanita, A comprehensive review on Exploration and Production scenario of Natural Gas Hydrate, *Res. J. Chem. Environ.*, **28(3)**, 104-109 **(2024)**

40. The Python Language Reference, Python documentation, https://docs.python.org/3/reference/index.html, [accessed 2024 Feb 29] **(2024)**

41. Toth A., Fackelmann J., Pigott W. and Tolomeo O., Tuberculosis prevention and treatment, *Canadian Nurse*, **100(9)**, 27-30 **(2004)**

42. Trajecta(Ana), Analogue Release 2024, Growdea Technologies Pvt. Lt., v1.1, https://growdeatech.com/Analogue/ **(2024)**

43. Trauer J.M., Dodd P.J., Gomes M.G.M., Gomez G.B., Houben R.M.G.J., McBryde E.S., Melsew Y.A., Menzies N.A., Arinaminpathy N., Shrestha S. and Dowdy D.W., The Importance of Heterogeneity to the Epidemiology of Tuberculosis, *Clinical Infectious Diseases*, **69(1)**, 159–166 **(2019)**

44. Turner J., Gonzalez-Juarrero M., Ellis D.L., Basaraba R.J., Kipnis A., Orme I.M. and Cooper A.M., *In vivo* IL-10 production reactivates chronic pulmonary tuberculosis in C57BL/6 mice, *Journal of Immunology*, **169(11)**, 6343–6351 **(2002)**

45. Umargono E., Suseno J. and Gunawan S.K., K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula, In The 2nd International Seminar on Science and Technology, 121-129 **(2020)**

46. Valdés-Tresanco M.S., Valdés-Tresanco M.E., Valiente P.A. and Moreno E., gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS, *Journal of Chemical Theory and Computation*, **17(10)**, 6281–6291 **(2021)**

47. Wang H., Gao X. and Fang J., Multiple Staggered Mesh Ewald: Boosting the Accuracy of the Smooth Particle Mesh Ewald Method, *Journal of Chemical Theory and Computation*, **12(11)**, 5596–5608 **(2016)**

48. Yan Y., Zhang D., Zhou P., Li B. and Huang S.Y., HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy, *Nucleic Acids Research*, **45(W1)**, W365–W373 **(2017)**

49. Zhang T., Rao G. and Gao X., Identification of Hub Genes in Tuberculosis via Bioinformatics Analysis, *Computational and Mathematical Methods in Medicine*, **2021(1)**, 8159879 **(2021)**.

(Received 01st October 2024, accepted 05th December 2024)