Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

In silico genome-scale analysis reveals a putative TssM-like deubiquitinase as a virulence-associated marker in Burkholderia pseudomallei strain CM000113

Supriyo Mukherjee
Manipal Academy of Higher Education
Rajarshi Bhattacharya
Manipal Academy of Higher Education
Thripthi Nagesh Shenoy
Manipal Academy of Higher Education
Chiranjay Mukhopadhyay
Manipal Academy of Higher Education
Somasish Ghosh Dastidar
Manipal Academy of Higher Education
Saugata Hazra
Indian Institute of Technology Roorkee
Abdul Ajees Abdul Salam
Manipal Academy of Higher Education
Ranita Ghosh Dastidar
ranita.gd@manipal.edu

Manipal Academy of Higher Education

Article

Keywords: Burkholderia pseudomallei, putative proteins, virulence factors, functional prediction, in silico approaches, melioidosis

Posted Date: February 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8672691/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Title: *In silico* genome-scale analysis reveals a putative TssM-like deubiquitinase as a virulence-associated marker in *Burkholderia pseudomallei* strain CM000113

Author: Supriyo Mukherjee[1,2*], Rajarshi Bhattacharya[2,3*], Thripthi Nagesh Shenoy[4], Chiranjay Mukhopadhyay[2,5,6], Somasish Ghosh Dastidar[2,3], Saugata Hazra[7], Abdul Ajees Abdul Salam[4], Ranita Ghosh Dastidar[1,2#]

Affiliation:

[1] Department of Biochemistry, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India

[2] Center for Emerging and Tropical Diseases, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India

[3] Centre for Molecular Neurosciences, Department of Anatomy, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India

[4] Manipal Institute of Applied Physics, Manipal Academy of Higher Education, Manipal, India

[5] Department of Microbiology, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India

[6] Manipal Institute of Virology, Manipal Academy of Higher Education, Manipal, India

[7] Department of Biosciences and Bioengineering, Indian Institute of Technology, Roorkee, India

*First Author: Supriyo Mukherjee & Rajarshi Bhattacharya both contributed equally

[#]Corresponding Author: Ranita Ghosh Dastidar[1,2#]

Email: ranita.gd@manipal.edu

ORCID: 0000-0001-6773-1113

Abstract:

Burkholderia pseudomallei produces a disease known as Melioidosis. Much of its genetic makeup remains undiscovered, including genes that code for what are called hypothetical and putative proteins. Uncharacterized proteins could uncover new mechanisms of virulence as well as reveal novel pathways and processes associated with pathogenesis. In this study, we present data on the *B. pseudomallei* CM000113 strain, which has 3,094 putative proteins (PPs), representing approximately 41% of its genome. Comparative sequence analysis using the Virulence Factor Database (VFDB) identified 20 PPs with >90% homology to known virulence determinants. Antigenicity profiling using VaxiJen v2.0 predicted a strong immunogenic potential for 18 of these PPs, underscoring their potential as vaccine candidates. Interestingly, one putative protein (IFOILCOA_05323) had a predicted structure that resembled TssM, a deubiquitinase. It also had several domain regions that were predicted to be associated with ubiquitin carboxyl-terminal hydrolase 14 open reading frame. Further *in silico* prediction of the protein sequences suggested that they participate in the bacterial intracellular immune evasion mechanism of the human host cell. It is predicted that these proteins will reside mainly in the periplasmic space and cytoplasm to fulfil their putative roles. Protein–protein docking, HADDOCK scoring, and molecular dynamics simulations demonstrated stable interactions with host HLA molecules, particularly HLA-DQB2, corroborated by host–pathogen interactome predictions. These integrative *in silico* analyses offer the first insights into the structural and functional attributes of these PPs, leading to their consideration for experimental analyses of vaccine and therapeutic strategies against melioidosis.

Keywords: *Burkholderia pseudomallei*, putative proteins, virulence factors, functional prediction, *in silico* approaches, melioidosis

1. Introduction:

Burkholderia pseudomallei is an intracellular, Gram-negative bacterium that causes melioidosis, a disease that can be very serious and even deadly in humans and animals. The bacterium is mostly found in the soil and surface waters of tropical and subtropical regions, making the disease a major public health threat[1]. Infections can occur by inhalation, transdermal contact, or ingestion of contaminated food[2-4]. The disease can be asymptomatic, but many cases are likely to be full-blown, including symptoms such as severe pneumonia and fulminant septicemia. When the disease is not treated in a timely and effective manner, it has a fatality rate of over 50%[4,5].

Melioidosis continues to pose a considerable and growing threat to public health at a global scale. There are about 165,000 infections and 89,000 deaths due to melioidosis each year[6]. Most of these cases occur in endemic areas, which include Southeast Asia and Northern Australia, and that is where most of the deaths occur[7]. This infection is emerging as a global threat because of globalization, international travel, and climate change which have all resulted in the diagnosis of melioidosis in new, previously non-endemic areas[8,9]. The treatment of infections caused by *Burkholderia pseudomallei* is problematic as the organism possesses innate resistance to many of the antibiotics that are commonly prescribed, and that includes the aminoglycosides and first and second generation cephalosporins[4, 10-12]. This organism also causes major barriers to the control and eradication of the disease as *Burkholderia pseudomallei* can survive in macrophages, form latent infections, and even causes disease relapse after prolonged antibiotic treatment[13-15].

Although we have gained considerable knowledge on the virulence mechanisms of *Burkholderia pseudomallei*, its molecular pathogenesis, host adaptation mechanisms, and antimicrobial resistance strategies remain poorly understood. One of the most promising, yet

most backward, of the above-listed scientific directions is the systematic study of the so-called putative proteins, which, although predicted from the genome, remain uncharacterized[16–18]. During the analysis of bacterial DNA sequence, a large part of the genome is described as putative uncharacterized proteins. In *Burkholderia pseudomallei*, almost half of the predicted proteome was annotated as hypothetical, putative and uncharacterized. Such proteins may be described as factors of uncharacterized virulence, mechanisms for immune evasion, or even metabolic enzymes that are necessary for the survival of the bacterium within the host[19–21]. Therefore, a systematic study of the above-mentioned putative protein is required to discover novel biomarkers, new therapeutic targets and vaccine candidates.

Recent innovations in bioinformatics and computational biology have provided the means to functionally characterize previously uncharacterized proteins. Enhanced sequence homology searches, structural domain prediction, protein–protein interaction network analysis, and molecular docking allow more detailed protein function predictions. The addition of subtractive proteomics, where candidate proteins are filtered based on essentiality, homology to host proteomics, and predicted subcellular localizations, has revealed more efficient mechanisms for determining novel therapeutics. However, detailed functional definitions of putative proteins (PPs) in *Burkholderia pseudomallei*, especially in clinically relevant isolates, are lacking.

This study details the PP repertoire of *B. pseudomallei* CM000113, an Indian clinical isolate identified in our previous study[22], by conducting a comprehensive *in silico* study. Using subtractive proteomics, we identified PPs that were vital for bacterial survival, had no human counterparts, and were substantially aligned with known virulence factors. Using these proteins, we performed domain analysis for structural and functional characterization, followed by homology modelling and active site prediction. To determine the potential biomedical utility of the proteins, we antigenically profiled them, identified vaccine candidates, and employed

molecular docking and molecular dynamics simulations to predict the protein-ligand interface druggability as a measure of potential therapeutic value. Overall, this study lays a foundation for investigating the role of PPs in the pathogenesis of *B. pseudomallei* and provides candidates for primary research focused on developing therapeutic and vaccine solutions for melioidosis.

2. Materials and Methods

2.1. Genome Assembly and Annotation

Genomic exploration of *Burkholderia pseudomallei* enables the systematic study of individual genes and fundamental investigations into the proteome of the organism. Importantly, these investigations into *B. pseudomallei* genetic architecture and protein biology provide insights into the proteins and proteomes that thermodynamically drive the organism's life cycle, mechanisms of virulence, immune modulation, and evasion, as well as adaptation to and persistence within the mammalian host. Indeed, these investigations into proteins remain vital for outlining the organism's proteomes and the associated formation of vaccines and therapeutically active agents. The Illumina-generated paired FASTQ sequence files of *B. pseudomallei* strain CM000113 were obtained from the NCBI Public Sequence Database (NCBI SRA: SRR36404674, Bioproject: PRJNA889557). The Sample was sequenced using the Illumina HiSeq platform and subsequently assembled with Unicycler v0.4.8[23], while annotation was performed using the Rapid Annotation using Subsystem Technology toolkit (RASTtk)[24]. The toolkit incorporates gene prediction, discovery and annotation of coding sequences (CDS), and assignment of functions from those CDS. The protein products of these genes have been identified and characterized; however, they still lack experimental validation. These were coupled to putative reads and subsequently assigned to various downstream computational analyses. This computational analysis anticipated the putative capacity of

proteins to enact pathobiological activities, serve as putative antigens, and withstand drug targeting.

2.2 Prediction of Potential Virulence Factors

Virulence Factors are important for understanding bacterial infections, as they enable bacteria to colonize, invade hosts, and evade the immune response. Identifying PPs that are similar to known virulence factors strongly suggests the relevance of PP to pathogenicity and helps narrow down the candidates most likely to be involved in the disease process. To analyze the presence of virulent PPs, sequence comparisons were conducted using the Virulence Factor Database (VFDB)[25] and the BLASTp[26] tool included in the BLAST2GO software[27]. Comparisons were filtered using an E-value criterion of 0.001 to increase confidence in the results. Potential virulent proteins which were chosen had greater than 90% sequence similarity to proteins of known virulence that had been validated through experimentation.

2.3 Prediction of Antigenicity

Antigenic proteins are important for vaccine development because they trigger an immune response. Computational identification of immunogenic proteins predicts potential candidates, thereby reducing the burden of experimentation. To assess antigenicity, the VaxiJen v2.0[28] server, which employs an alignment-independent method based on physicochemical properties, was utilized. For sequences of bacterial origin, a threshold score of 0.4 was applied, whereby proteins that exceeded the threshold were considered highly antigenic candidates that can be included in subunit vaccines.

2.4 Prediction of Druggability

Assessing the druggability of a particular structure is central to determining which PPs possess conserved patterns, motifs, and regions that interact with small drug-like ligands and have the potential to be developed further. As the first step, sequence similarity searches were performed

in DrugBank[29], a manually curated and maintained database of drug target molecules. BLAST2GO[27] was used to execute BLASTp[26] queries with a cutoff E-value of 0.001 to eliminate false-positive results. PPs that showed significant similarity to known drug target molecules were prioritized because they were more likely to possess the structural or functional motifs needed to interact with a putative drug target.

2.5 Prediction of Protein Family and Domain

The absence of close homologs suggests potential novel structural features and greater functional diversity within a particular domain, thus providing insight into the conserved molecular functions of the domain, whether enzymatic or binding. Several domain prediction tools were used to further refine the functional annotation of PPs. InterProScan[30] is a meta-predictor that compiles the results from various databases, including Pfam (v37.0)[31] SUPERFAMILY (v1.75)[32] and others. The NCBI Conserved Domain Database (CDD)[33] and GenomeNet Motif Search[34] were used to obtain additional predictors. Structural classification was further refined using the CATH database (v4.3)[35] at the superfamily level. This multilayered approach resulted in the prediction of various novel enzymatic, interaction, and structural functions.

2.6 Predicting Subcellular Localizations

The location of a protein within a cell can predict its putative function. For example, proteins found in the cytoplasm are usually involved in regulatory processes, whereas proteins in the membranes or periplasmic space are highly involved in host interactions or are part of the cell's antibiotic defensive response. PP localization predictions were made using CELLO v2.5[36,37], which uses support vector machines trained on n-peptide compositions, which are cellular peptides composed of a small number of amino acids. Using the One-Against-One (OAO)

approach tailored for Gram-negative bacteria, localization predictions were used to hypothesize potential roles in virulence and immune recognition.

2.7 Predictions of the Protein's Physicochemical Properties

A protein's physicochemical traits will determine the factors that influence the protein's expositional traits of solubility and structural stability. The physicochemical characteristics of a protein influence its structural stability and provide insight into its solubility. For instance, a protein's half-life can determine its instability index, while the hydrophilic content of the protein can affect its GRAVY score, which influences the likelihood of the protein being a candidate for drug development and vaccine formulation. Parameters such as amino acid composition, molecular weight, theoretical pI, instability index, aliphatic index, and GRAVY score were calculated using the ProtParam tool on the ExPASy server[38]. These metrics were weighed to ascertain the stability, solubility, and overall suitability of the protein for experimental applications.

2.8 Construction of Physical Maps

Mapping the location of putative proteins (PPs) on a genome enables the understanding of the function of these proteins in the context of nearby target features, such as operons, islands of pathogenicity, or resistance genes. The genome sequence of *B. pseudomallei* CM000113 was retrieved and analyzed using the Proksee platform[39]. A series of physical maps were autogenerated to display the features of interest, including the coding sequences, protein-coding genes, and other surrounding elements of the genome. Twenty-three putative proteins of interest were located and annotated using tBLASTn searches (E-value cutoff 1e-5)[40]. Microsoft PowerPoint 365 was used to edit the generated maps to highlight and zoom into specific genomic coordinates to better visualize and draw the focus on the regions of interest with surrounding elements containing predicted loci of interest to virulence.

2. 9 Secondary Structure Prediction

Molecular conformation is dictated at a fundamental level by specific elements, such as alpha helices and beta sheets, which contribute to the folding, stability, and possible interactions of proteins. Predicting such structures is crucial for advancing to the next phase of tertiary structure modelling. Therefore, PSIPRED[41,42], NetSurfP v3.0[43] and SOPMA[44] were used, which employ different algorithms to predict secondary structures using sequence information and neural network predictors with PSI-BLAST, NetSurfP v3.0 and a consensus approach using DSSP, respectively. These predictions were used to draw inferences regarding the flexibility of the protein, the degree to which the regions in the protein sequence contain conserved motifs, and the overall stability of the protein.

2.10 Tertiary Structures Prediction

In the absence of experimental structures, computational modelling provides insight into the mechanisms at play. Several modelling approaches were applied to obtain the best accuracy. For homology modelling, MODELLER[45] was used; for threading and *ab initio* assembly, I-TASSER[46] was used and for hybrid comparative and *ab initio*, Robetta[47] was used. The models were polished in GalaxyWEB[48] and visualized using PyMOL software[49].

2.11 Quality Assessment of Predicted Models

For downstream processes to proceed, the predicted structures must be accurate and reliable. Quality assessment for the structures was obtained using PROCHECK[50] for Ramachandran analysis, ERRAT[51] for nonbonded interactions, and ProSA[52] for the global Z-scores. Only the highest confidence models were retained for further use.

2.12 Host–Pathogen Interactome Analysis

Interactions between hosts and pathogens demonstrate the virulence of certain bacterial proteins. Information was retrieved from previous interactome studies on *B. pseudomallei* strain K96243[21] **(Supplementary Table S3)**. MEGA 11[53] was used to perform evolutionary analyses to determine homologous PPs and their corresponding human protein counterparts. For some candidate PPs that interact with homologs, molecular docking studies were conducted to allow focused investigations of immunologically pertinent interactions.

2.13 Active Site prediction

Prediction of ligand-binding residues facilitates functional annotation and drug discovery processes. Active sites were predicted using CASTpFold[54], which combines geometric pocket detection and evolutionary conservation scoring. The sites were ranked by residue composition, depth of cavity, and volume, and the best were selected for docking and further analysis of their functional interpretation.

2.14 Protein-Protein Docking Analysis

Identification of putative interaction interfaces involved in host–pathogen communication was achieved through protein-protein docking analyses. Docking was performed using HADDOCK 2.4[55-57]. Active residues were defined as binding-site residues, whereas the others in the pocket were considered passive. No docking restraints were used, and the complexes were clustered and scored using the metrics and Z-scores from HADDOCK. The binding energies of the docked complexes were predicted using PRODIGY[58].

2.15 Molecular Dynamics Simulation

Using the Desmond module within Schrödinger, two 500 ns molecular dynamics simulation was performed on the docked complex[59]. The system was set in an orthorhombic SPC water box (10 Å edge lengths) and run using the OPLS4 force field[60] under NPT conditions at 300 K and 1 bar, pH 7.4. The simulation trajectory was recorded every 100 ps. Duplicate simulations

were performed, and the trajectory in each simulation was analyzed for RMSD, Rg, SASA, and RMSF. These were visualized using PyMOL v3.1.6.1[49].

2.16 Structural Similarity Analysis

The docked protein complex was used as a query for the DALI server[61] to search for structural homologs, which allows a user to compare a given structure to all structures in the Protein Data Bank (PDB) and retrieve the most similar structure. The results were ranked according to their Z-scores, which the server interprets as an importance indicator, thresholded at a value of 30 or above. These high Z-score entries were separate structures, and PyMOL 3.1.6.1[49] was used for structural alignment and comparison with the query protein.

3. Results

3.1. Identification and Selection of Putative Proteins

The entire genome of *Burkholderia pseudomallei* strain CM000113 (NCBI SRA: SRR36404674, BioProject: PRJNA889557) was acquired, and meticulous scrutiny revealed several putative proteins with unknown functions. The complete genome assembly contained 7,471 protein-coding sequences. 3,094 of these proteins (41.41%) had no known function. The protein-coding sequences and putative proteins were obtained using the RAST annotation pipeline and subsequently filtered using terms such as “hypothetical,” “uncharacterized,” “putative,” and “-like protein”, although NCBI PGAP analysis is showing only 879 sequences, termed as ‘hypothetical protein’. The final collection of all proteins with unknown function was selected for additional examination of their structure, function, and pathogenicity. This workflow is shown in **Figure 1**.

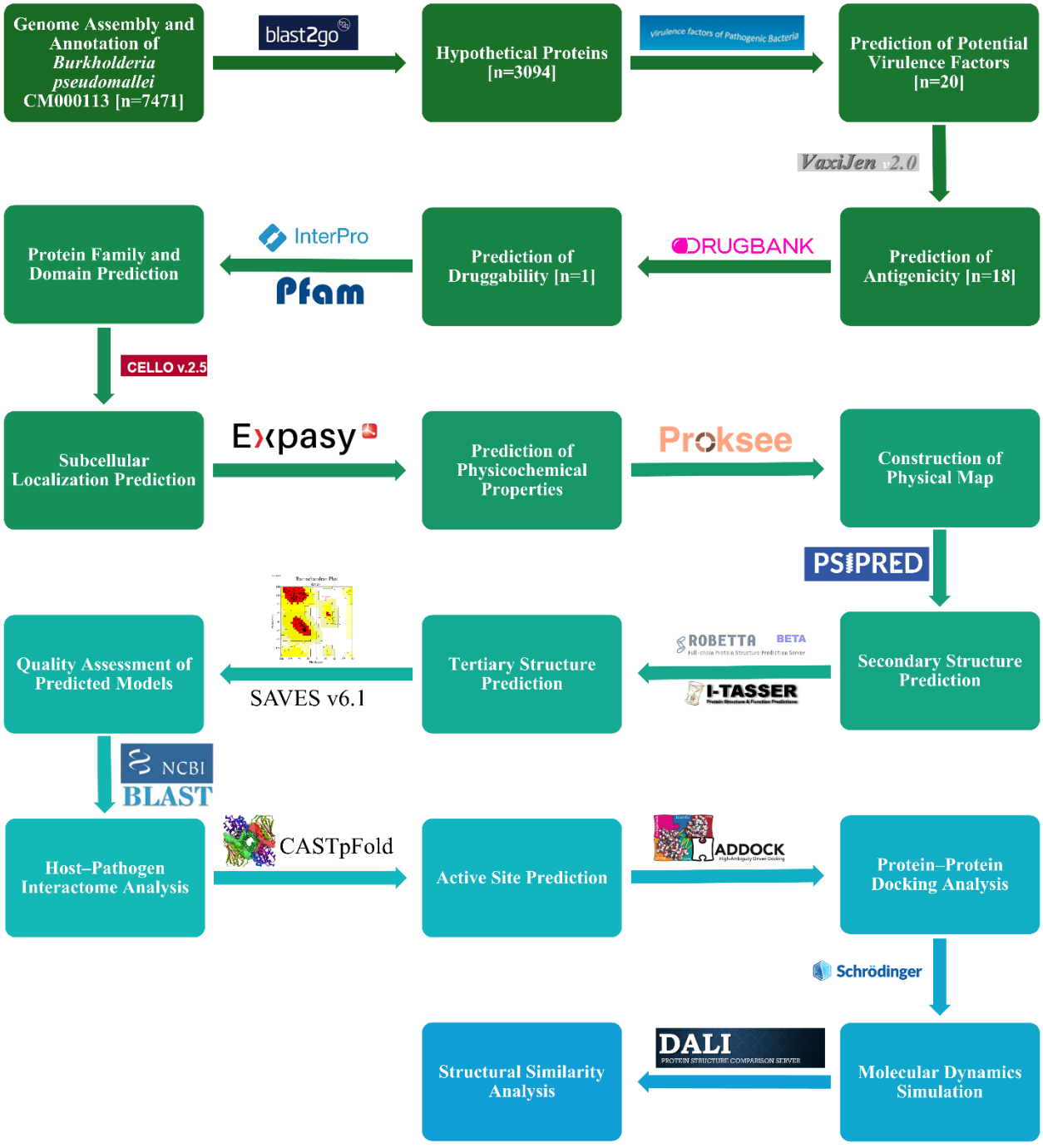


Figure 1: The workflow of the study includes all step-by-step processes and the utilized tools.

3.2. Potential Virulence Factors, Antigenicity and Druggable Targets Prediction

Comparative analysis against the VFDB database retrieved multiple uncharacterized proteins exhibiting virulence-associated sequences with high similarity (>90%). A BLASTp query of the VFDB complete protein sequence database revealed 979 of the 3094 uncharacterized proteins which is reported in **Supplementary Table S1, Supplementary Dataset S1**. Among

these, 20 sequences showed high similarity (>90%) to the database, and these uncharacterized proteins were identified as putative virulence factors of *B. pseudomallei*. Using the VaxiJen v2.0 Prediction Antigenicity tool, 18 of the 20 putative proteins had an antigenicity score of >0.4 **(Table 1)**. These proteins were flagged as potential contributors to the pathogenicity of *B. pseudomallei*. The antigenicity prediction score predicted that 18 of the 20 putative proteins showed a score of > 0.4 using VaxiJen V2.0 **(Table 1)**[28]. A BLASTp search of the DrugBank target sequences revealed uncharacterized proteins that were significantly similar to documented drug targets. Furthermore, these 20 uncharacterized proteins were filtered using BLASTp against the DrugBank database Drug Target Sequences, of which one sequence had similarity, which is shown in **Table 1**. Therefore, the putative protein IFOILCOA_05323 was put forward for a more detailed assessment of its druggability to identify new potential therapeutic targets **(Supplementary Table S2)**.

Table 1: Assessment for Virulence, Antigenicity, and Druggability

Name	Virulence Property (VFDB)	Antigenicity (VaxiJen)	Druggability (Drugbank)
IFOILCOA_01247	651	0.4292	NA
IFOILCOA_02611	184	0.5539	NA
IFOILCOA_04007	159	0.5863	NA
IFOILCOA_05320	1017	0.559	NA
IFOILCOA_05323	488	0.5265	49.25%
IFOILCOA_06028	174	0.7219	NA
IFOILCOA_02560	275	0.4582	NA
IFOILCOA_05646	333	0.6916	NA
IFOILCOA_01249	625	0.5097	NA
IFOILCOA_05314	298	0.6657	NA
IFOILCOA_05806	624	0.8184	NA
IFOILCOA_03994	447	0.7902	NA
IFOILCOA_05798	226	1.0474	NA
IFOILCOA_05783	217	0.6244	NA

IFOILCOA_02557	324	0.3641	NA
IFOILCOA_02850	281	0.3733	NA
IFOILCOA_03796	566	0.7266	NA
IFOILCOA_02724	348	0.641	NA
IFOILCOA_02608	313	0.4855	NA
IFOILCOA_03812	194	0.698	NA

3.3. Analysis of Protein Families and Domains

Functional enrichment and domain analysis using InterProScan, Pfam, and SUPERFAMILY showed a Big-5 domain (PF13205) within the putative protein IFOILCOA_05323 **(Supplementary Figure S1)**. Further information on the structural organization is provided by the classification of superfamilies based on the CATH. This protein is a member of the papain-like cysteine peptidase superfamily (IPR038765) **(Supplementary Figure S1)**. These proteins generally have one α-helix and four strands that form an antiparallel β-sheet containing the Cys-His-Asn catalytic triad. These results collectively provide indications of the potential biological functions of the proteins that were subjected to further examination.

3.4. Prediction of Sub-Cellular Localization

The putative proteins were subjected to CELLO v.2.5 to classify them according to their predicted subcellular localization based on the probable localization of each protein within a bacterium. The identified IFOILCOA_05323 protein was predicted to localize to the periplasmic space or cytoplasm with reliability scores of 1.890 and 1.310, respectively **(Supplementary Figure S2)**, whereas the other proteins were predicted to associate with the bacterial membrane.

3.5. Prediction of Physicochemical Properties

ProtParam results provided a comprehensive physicochemical analysis of the potential protein, including length, molecular weight, charge, aliphatic index, isoelectric point, and stability indices, which resulted in a comprehensive analysis of the protein's potential solubility, stability, and expressibility. IFOILCOA_05323 has a molecular weight of 30,590.35 Da and contains 284 amino acids, which, as a small protein, is ideal for recombinant expression and purification in prokaryotes, and is likely comprised of soluble domains, small multi-domain enzymes, or peripheral/monotopic proteins. The theoretical isoelectric point (pI) is 5.11, highlighting that this protein is likely to be acidic in nature and may contain a net negative charge under physiological conditions, which generally improves aqueous solubility. The total number of Asp and Glu residues was 33, which was significantly larger than the 24 Arg or Lys residues, and supported the net negative charge under neutral conditions. The grand average of hydropathicity (GRAVY) was -0.180, indicating that the protein is poorly hydrophobic and is therefore likely to be a soluble protein. The aliphatic index was 89.12. This is indicative of the side chain composition that aligns with an elevation in thermostability in globular proteins, which is ideal for further biophysical characterization once a suitable buffer formulation is determined. In contrast, the instability index was 48.04, which is higher than the empirical stability predictive threshold of 40, suggesting that the isolated protein is highly likely to be unstable *in vitro*. Taken together, the combination of a highly alkaline pI, a mildly negative GRAVY, and a high aliphatic index signifies a protein with hydrophilicity and thermal stability, but also an overall unpredictable and likely unstable protein that would need to be optimized with a constructed aid in folding for any useful biochemical application **(Table 2)**.

Table 2: Physicochemical properties of putative protein IFOILCOA_05323.

Number of amino acids	Molecular weight	Theoretical pI	Total number of negatively charged residues	Total number of positively charged residues	Grand average of hydropathicity (GRAVY)	Aliphatic index

			(Asp + Glu)	(Arg + Lys)		
284	30590.35	5.11	33	24	-0.180	89.12

3.6. Physical Map

The putative protein, which is the subject of the present analysis, was located, as expected, in the physical map of *Burkholderia pseudomallei* CM000113, in a region that is approximately 6300 to 6315 kb, bracketed by other annotated coding sequences (for example, TssM, a Type VI secretion system virulence factor), and other putative proteins, as shown in **Figure 2** and **Supplementary Datasets S2 & S3**. This also suggests that the putative protein is virulent and is likely involved in host–pathogen interplay.

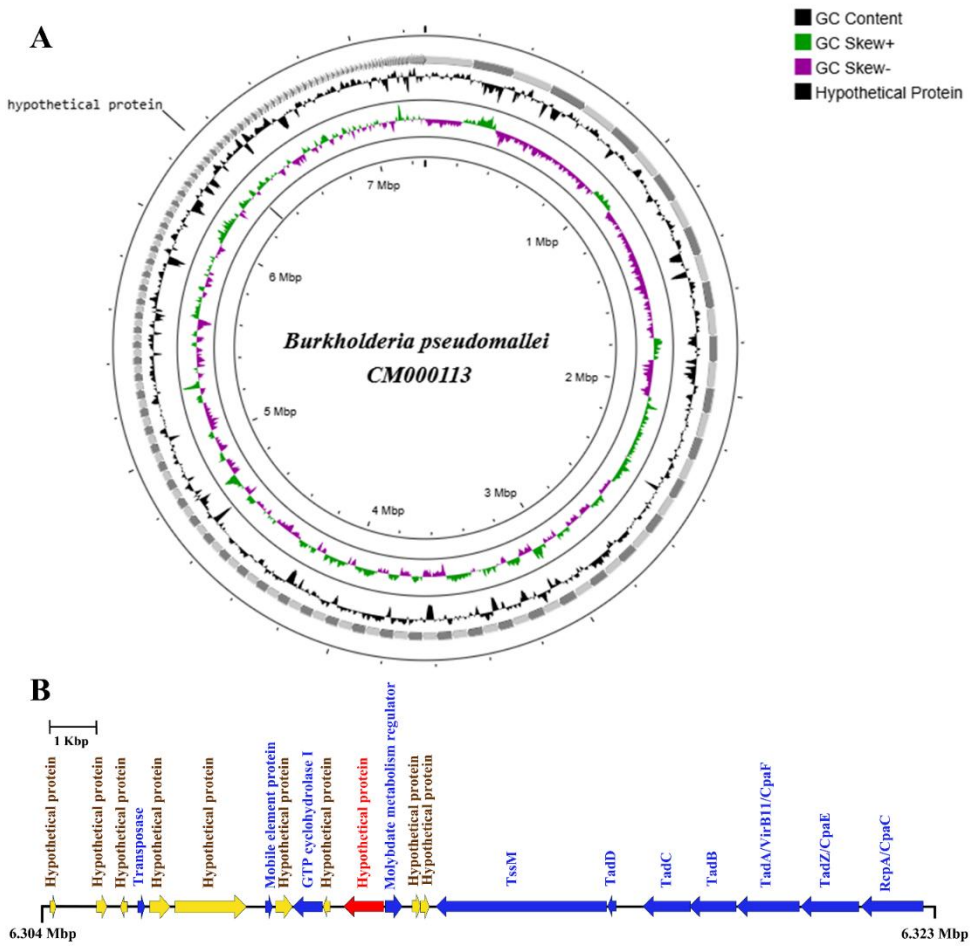


Figure 2: Genome overview and locus context of a prioritized putative protein in *Burkholderia pseudomallei* CM000113. (A) circular map showing GC content (black), GC skew+ (green), GC skew− (magenta), and the distribution of putative proteins; (B) genomic neighbourhood highlighting the putative protein of interest in red, annotated/known genes in blue, and other putative genes in brown, adjacent to tad pilus biogenesis genes (scale bar, 1 kb).

3.7. Multiple Sequence Alignment and Phylogenetic Study

Multiple sequence alignment (MSA) of the putative proteins and their homologs showed the papain-like catalytic triad (Cys–His–Asn) and neighboring active site residues, and the β-strand/core-defining residues of the Big-5 domain, which suggest enzymatic activity and an Ig-like flexible scaffold for binding, were all conserved. The phylogenetic tree placed the putative protein within a clade containing closely related proteins of *Burkholderia* virulence with high bootstrap support (≥70% on relevant internal nodes; 1,000 replicates), thereby supporting the functional prediction of domain conservation **(Figure 3A)**.

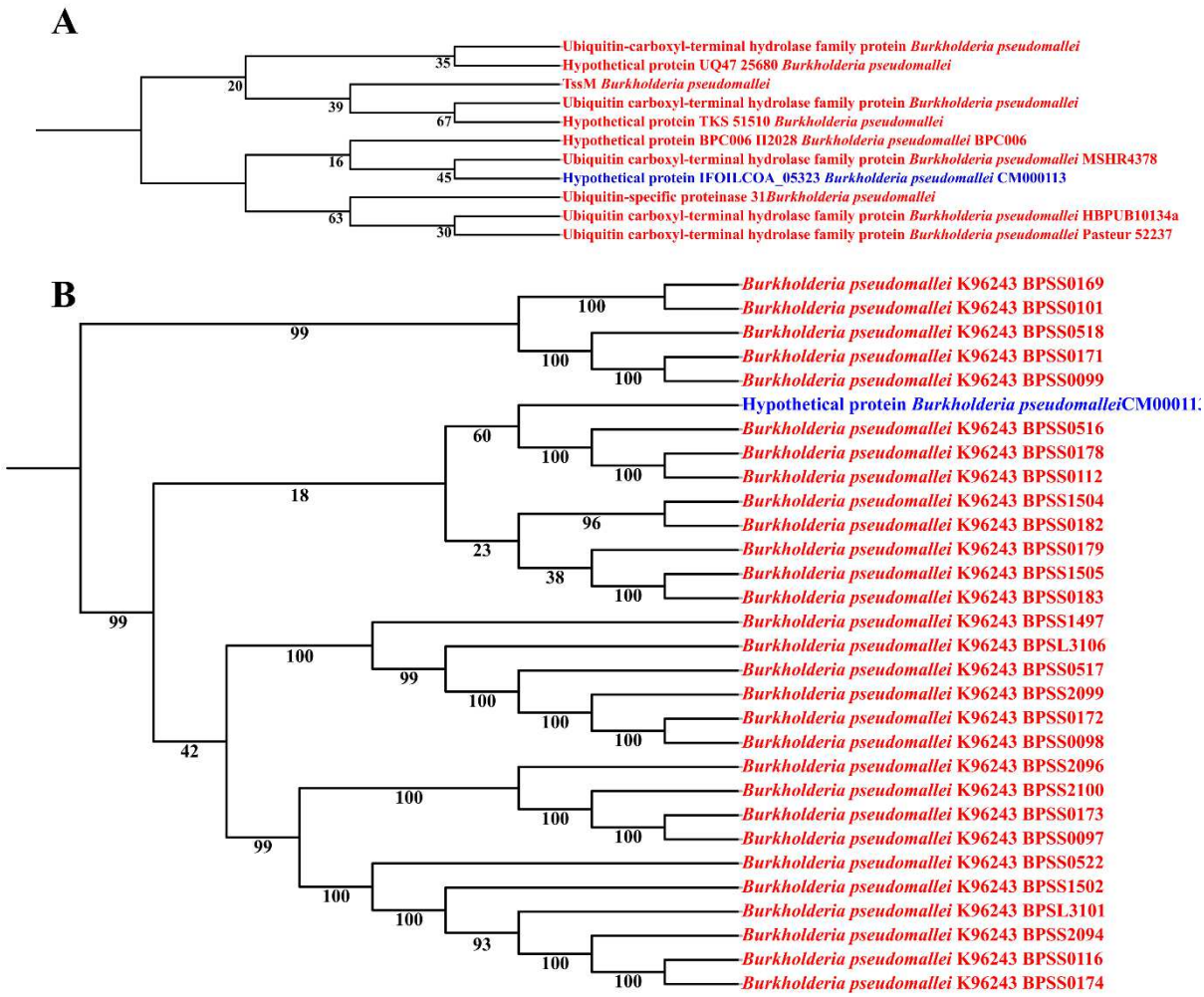


Figure 3: Phylogenetic analyses of a prioritized putative protein in *Burkholderia pseudomallei* CM000113. (A) Maximum-likelihood phylogeny of 11 proteins inferred under the Whelan–Goldman model with discrete Gamma rates (+G, five categories; α=0.1000) and a proportion of invariant sites (+I=9.16%) from 415 aligned positions in MEGA11; node labels show bootstrap support from 1,000 replicates and branches with support <50% are collapsed, with the focal putative protein highlighted in blue and other homologs in red. (B) Genome-wide paralog clustering depicting the placement of the putative protein among closely related proteins of *B. pseudomallei* K96243, with branch lengths proportional to pairwise amino-acid distances and tip labels reporting locus tags and residue spans.

3.8. Secondary and Tertiary Structure Prediction and Model Refinement

Molecular conformation is dictated at a fundamental level by specific elements, such as alpha helices and beta sheets, which contribute to the folding, stability, and possible interactions of proteins. Predicting such structures is crucial for advancing to the next phase of tertiary structure modelling. PSIPRED[41,42], NetSurfP v3.0[43] and SOPMA[44] predicted that some of the putative proteins included secondary structures, such as alpha helices, beta strands, and random coils. SOPMA predicted a random coil dominance of 48.94%, followed by alpha helix 27.82%, extended strand 17.96%, and beta-turn 5.28% **(Figure 4)**. This secondary structural information highlights the potential functional regions and structural motifs.

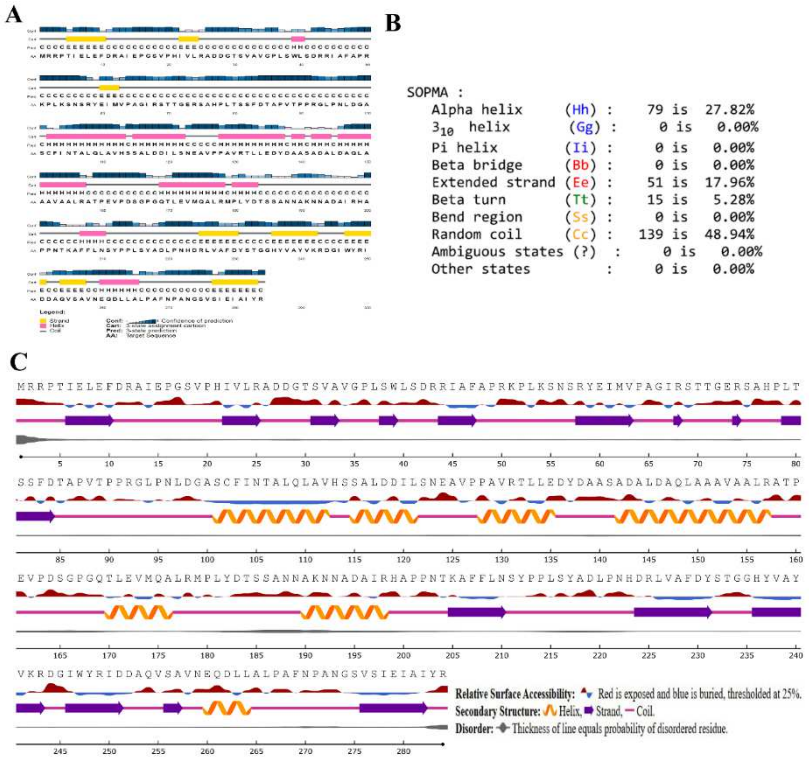


Figure 4: Residue-level structural features of the prioritized putative protein from *Burkholderia pseudomallei* CM000113. (A) PSIPRED per-residue secondary-structure prediction with confidence track, colored helix (magenta), strand (yellow) and coil (blue). (B) SOPMA composition summary showing alpha helix 27.82% (79 residues), extended strand 17.96% (51 residues), beta turn 5.28% (15 residues) and random coil 48.94% (139 residues). (C) Integrated sequence annotation using NetSurfP v3.0 depicting relative surface accessibility (red exposed; blue buried; 25% threshold), predicted secondary structure (orange helix; purple strand; gray coil), and intrinsic disorder (line thickness proportional to disorder probability).

MODELLER[45], I-TASSER[46], and Robetta[47] created and continued to refine tertiary structures in GalaxyWEB[48]. In the case of structure homology modelling, MODELLER[45] concentrated on one template and, in one case, predicted TssM (PDB ID:8q00A) as having a sequence identity of 99% and an RMSD of 0.648. In contrast, an I-TASSER[46] analysis, where the C-score was 0.93, indicated that the TM-score of the final predicted model of the putative protein was approximately 0.84±0.08 with an RMSD estimate of 4.2±2.8 Å **(Supplementary Figure S3 – S7)**. Subsequently, the Robetta[47] server performed *ab initio* modelling using the RoseTTAFold modelling approach. The GalaxyWEB[48] refinement tool was used on the structures predicted by Robetta for further refinement. The final 3D model was visualized using PyMOL[49], and the model showed significant structural features **(Figure 5A)**. Further structural evaluations confirmed that the refined models exhibited good quality and stability.

3.9. Quality Assessment of Predicted Models

The use of PROCHECK[50], ERRAT[51], and ProSA[52] in the models validated their structures. For the validation comparison of the 3D protein structures provided by the MODELLER[45], I-TASSER[46], and Robetta[47] servers, PROCHECK[50] Ramachandran plots were evaluated **(Table 3)**. The 3D structure was obtained from the Robetta server, and the structural refinement using GalaxyWEB[48] was of the best quality. The Ramachandran plot from the final model indicated that of the amino acid residues that were in the most favored region, 94.7% and in a minimum disallowed region were 0.4% (proline and glycine residues were evaluated with residue-specific criteria and excluded from general residue tallies due to their distinct φ,ψ distributions to avoid bias in outlier tallies) **(Figure 5 B, C, D)**. The final 3D structure had a Z-score of -9.85 and an overall quality factor of 92.963. With the reported Z-score, the model met other validation metrics as a final model to which further structural and functional analyses could be performed.

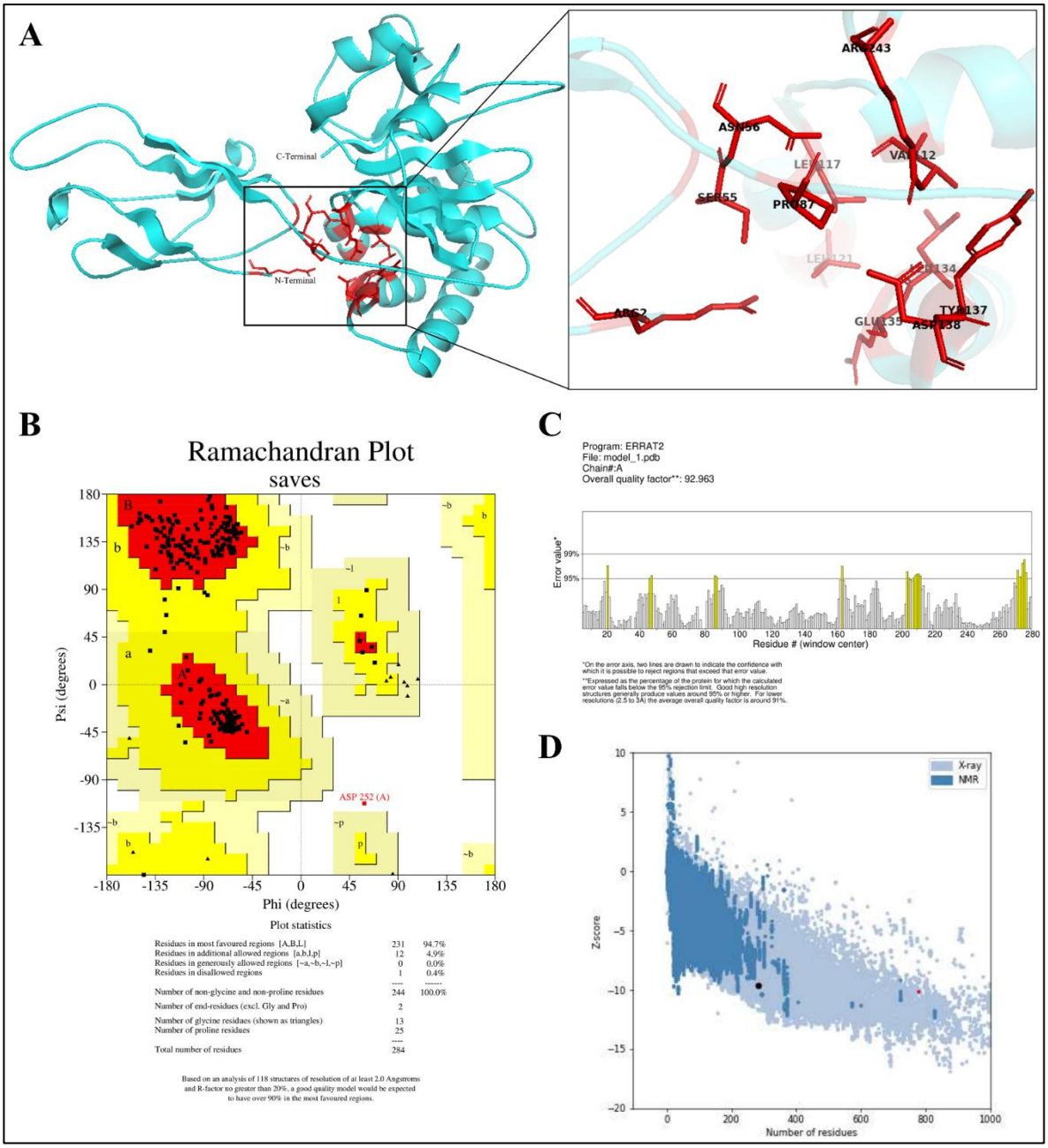


Figure 5: Structural model and validation of the prioritized putative protein from *Burkholderia pseudomallei* CM000113. (A) Homology model rendered in PyMOL 3.1.6.1 with a cartoon representation and inset highlighting the putative binding pocket and contacting residues. (B) Ramachandran plot from SAVES v6.1 showing residue distributions in favoured, additionally allowed, generously allowed, and disallowed regions. (C) ERRAT error-value profile across the sequence with the overall quality factor summarizing non-bonded interaction statistics. (D) ProSA Z-score positioned within the size-dependent distributions of X-ray and NMR structures, indicating the model's overall quality.

405 **Table 3:** Ramachandran plot summary of putative protein IFOILCOA_05323.

Modelling Tool	Most favored region (Residues)	Additional allowed Region (Residues)	Generously allowed region (Residues)	Disallowed region (Residues)
GalaxyWEB (Robetta predicted structure)	94.7%	4.9%	0.0%	0.4%
MODELLER	94.3%	5.3%	0.0%	0.4%
Robetta	88.9%	7.8%	0.8%	2.5%
I-TASSER	82.8%	14.3%	0.4%	2.5%

406

407 **3.10. Molecular Docking**

408 Using the information from the phylogenetic tree, along with multiple sequence alignment with
409 the assumed protein performed with ClustalW, the structural similarities with BPSS0516 were
410 documented **(Figure 3B, Supplementary Figure S8)**. Consistent with the interactions reported
411 in[21], several additional associated entities, including DOB, DQB1, DPB1, DRB3, X5D2U9,
412 BTBD9, FRIH, DQB2, PP4R1, FHL17, 2B1F, FRIL, IL16, DRB5, DRB4, and SARAF were
413 identified **(Table 4, Supplementary Table S3 & S4, Supplementary Figure S9 & S10)**.
414 Therefore, these were extracted to serve the purpose of the docking study.

415 **Table 4:** Bacterial putative protein IFOILCOA_05323 interaction with human host proteins

Bacterial Proteins	Host Protein (PDB)	Host protein name
BPSS0516	DQB2_HUMAN	HLA class II histocompatibility antigen, DQ beta 2 chain
BPSS0516	DRB4_HUMAN	HLA class II histocompatibility antigen, DR beta 4 chain
BPSS0516	DOB_HUMAN	HLA class II histocompatibility antigen, DO beta chain
BPSS0516	DRB3_HUMAN	HLA class II histocompatibility antigen, DR beta 3 chain
BPSS0516	FRIL_HUMAN	Ferritin light chain
BPSS0516	FRIH_HUMAN	Ferritin heavy chain

BPSS0516	DQB1_HUMAN	HLA class II histocompatibility antigen, DQ beta 1 chain
BPSS0516	2B1F_HUMAN	HLA class II histocompatibility antigen, DRB1 beta chain
BPSS0516	DPB1_HUMAN	HLA class II histocompatibility antigen, DP beta 1 chain
BPSS0516	SARAF_HUMAN	Store-operated calcium entry-associated regulatory factor
BPSS0516	X5D2U9_HUMAN	HLA class II histocompatibility antigen, DR beta 4 chain
BPSS0516	IL16_HUMAN	Pro-interleukin-16
BPSS0516	DRB5_HUMAN	HLA class II histocompatibility antigen, DR beta 5 chain
BPSS0516	BTBD9_HUMAN	BTB/POZ domain-containing protein 9
BPSS0516	PP4R1_HUMAN	Serine/threonine-protein phosphatase 4 regulatory subunit 1
BPSS0516	FHL17_HUMAN	Ferritin heavy polypeptide-like 17

The active sites of the proteins were predicted using CASTpFold[54]. CASTpFold[54] predicted that the most active binding sites were in spatial regions of 188.601 $Å^2$ surface area and 121.412 $Å^3$ volume. The predicted active amino acid residues were Met1, Arg2, Leu53, Lys54, Ser55, Asn56, Thr85, Ala86, Pro87, Val112, His113, Leu117, Leu121, Leu134, Glu135, Asp136, Tyr137, Asp138, Ala139, Lys242, Arg243, and Arg284 **(Figure 5A)**. In addition, the predicted active binding sites of human proteins are reported in **Supplementary Figures S11-S26**. The active residues were used for molecular docking. Molecular docking analysis was performed using High Ambiguity Driven protein-protein DOCKing (HADDOCK) 2.4[55], and the best model was selected based on the lowest Z-score and HADDOCK scores. The score was computed using the equation HADDOCK score = $1.0 * E_{vdw} + 0.2 * E_{elec} + 1.0 * E_{desol} + 0.1 * E_{air}$, where E_{air} is the distance restraint energy, E_{vdw} is the intermolecular van der Waals energy, E_{elec} is the intermolecular electrostatic energy, and E_{desol} is the empirical desolvation energy. The best HADDOCK model was selected from all human host proteins based on the Z-score and binding affinity. The Z-score reflects how many standard deviations a score is from the mean of the cluster, with lower values indicating a stronger negative performance. The Z-scores

of PP-DQB2, PP-DQB1, PP-DPB1, PP-DRB3, PP-X5D2U9, PP-BTBD9, PP-FRIH, PP-DQB2, PP-PP4R1, PP-FHL17, PP-2B1F, PP-FRIL, PP-IL16, PP-DRB5, PP-DRB4, and PP-SARAF's Z-scores were recorded to be -2.1, -1.5, -1.8, -2.1, -1.4, -1.9, -1.5, -1.6, -2.2, -1.4, -2.1, -1.4, -1.7, -1.6, -1.5, -2.0, and -1.4 respectively **(Figure 6, Table 5, Supplementary Table S5, Supplementary Figure S27)**. Therefore, the PP-DQB2 docked cluster exhibited the most negative Z-score, with a HADDOCK score of -117.0 +/- 9.0 **(Figure 7)**. These data led to the decision to continue with the molecular dynamics simulations.

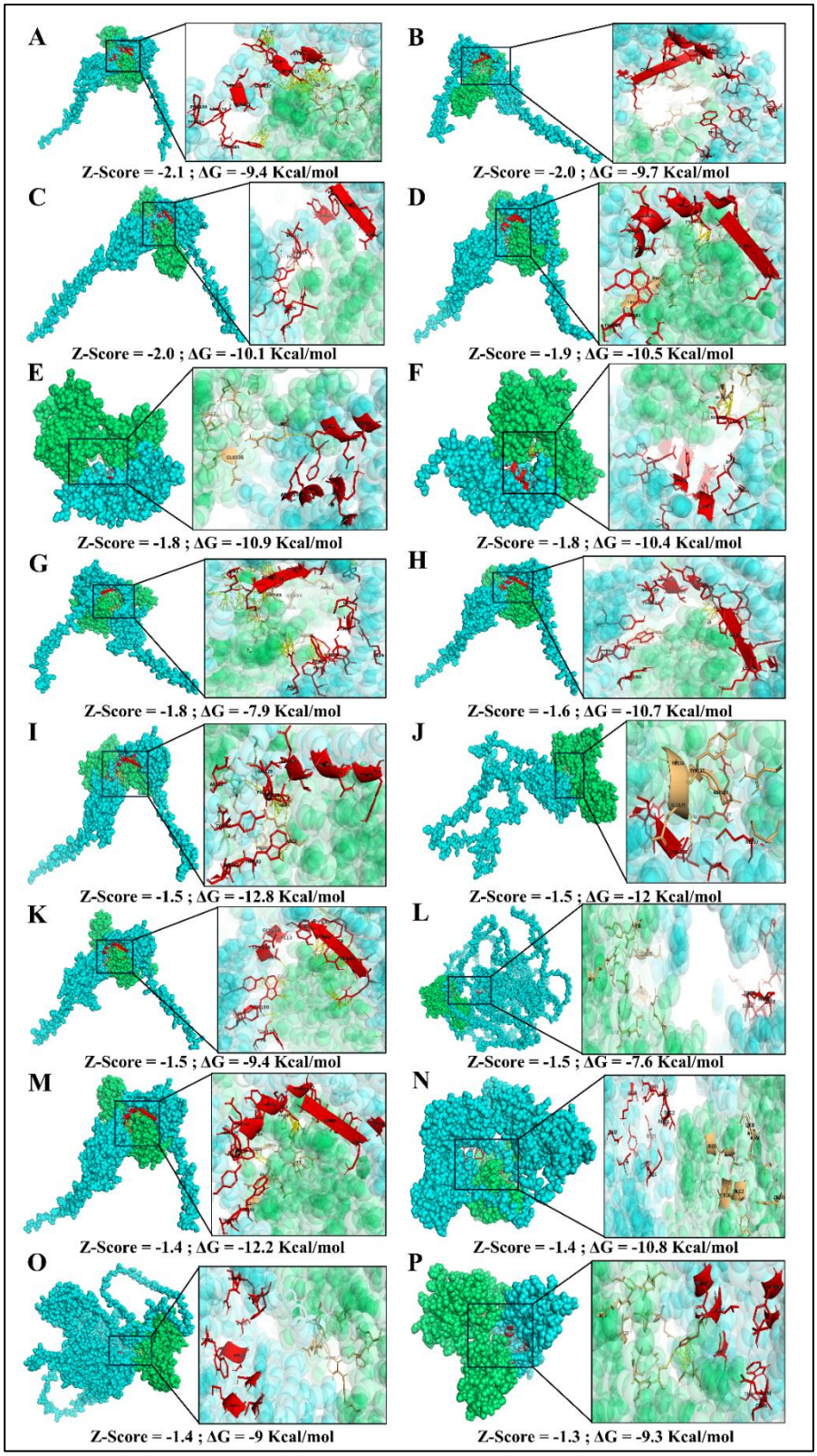


Figure 6: Docked complexes between the prioritized putative protein and human proteins generated with HADDOCK 2.4 and rendered in PyMOL; each panel shows the full complex and a magnified binding interface annotated with the HADDOCK cluster Z-score and predicted binding free energy ΔG (kcal/mol), where more negative Z-scores indicate clusters scoring better than the run average. A–P correspond to (A) DQB2-PP (B) DRB4-PP (C) DOB-PP (D) DRB3-PP (E) FRIL-PP (F) FRIH-PP (G)

DQB1-PP (H) 2B1F-PP (I) DPB1-PP (J) SARAF-PP (K) X5D2U9-PP (L) IL16-PP (M) DRB5-PP (N) BTBD9-PP (O) PP4R1-PP (P) FHL17-PP.

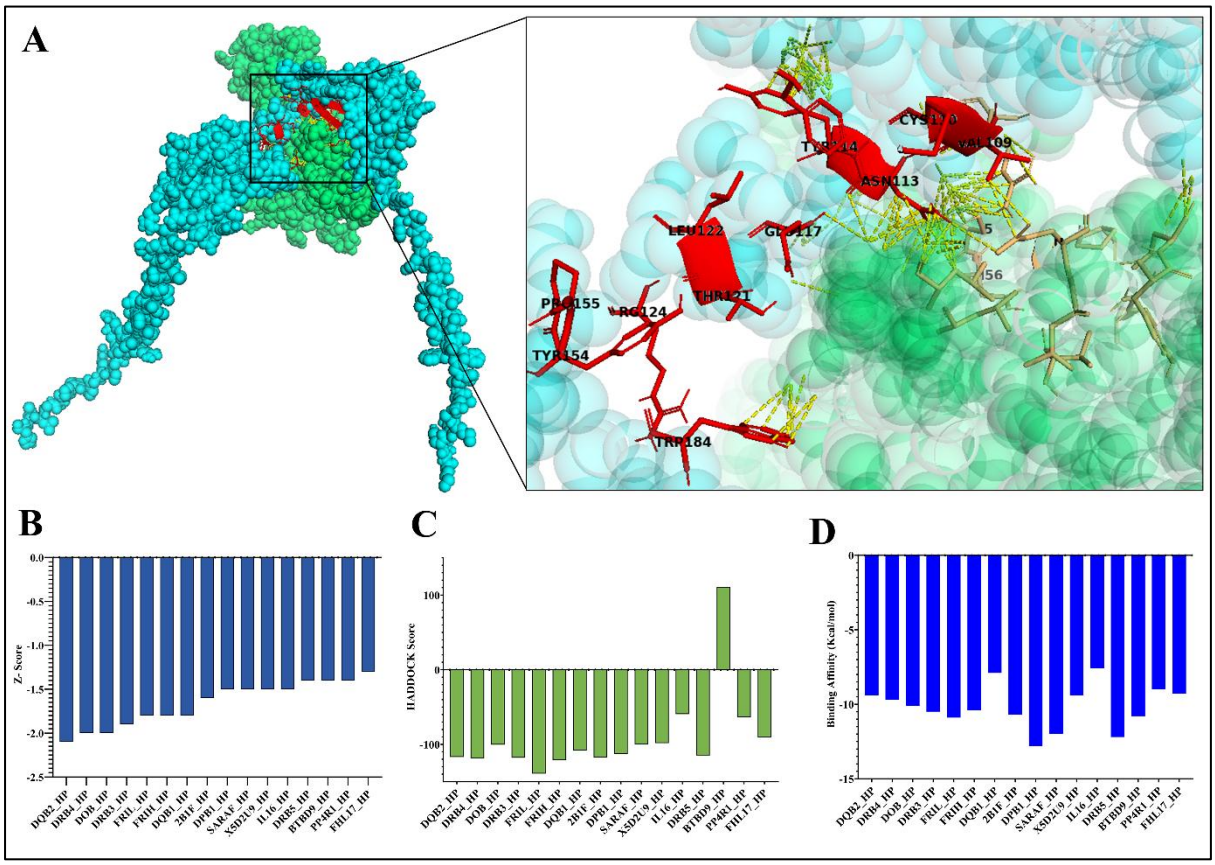


Figure 7: Top-ranked docked complex and summary metrics for interactions between the prioritized putative protein and human partners. (A) The highest-ranked complex from HADDOCK 2.4 is shown with an inset of the binding interface and contacting residues, rendered in PyMOL. (B) Bar plot of HADDOCK cluster Z-scores (C) Bar plot of HADDOCK scores (D) Estimated binding affinities ΔG in kcal mol^{-1}.\

Table 5: Docking results of putative protein IFOILCOA_05323 with human host proteins

Human Protein	Z-Score	HADDOCK Score	Binding free energy (Kcal/mol)
DQB2_HUMAN	-2.1	-117.0 +/- 9.0	-9.4
DRB4_HUMAN	-2.0	-119.0 +/- 7.8	-9.7
DOB_HUMAN	-2.0	-100.5 +/- 5.0	-10.1
DRB3_HUMAN	-1.9	-118.0 +/- 1.6	-10.5
FRIL_HUMAN	-1.8	-139.3 +/- 5.0	-10.9
FRIH_HUMAN	-1.8	-121.4 +/- 8.3	-10.4
DQB1_HUMAN	-1.8	-108.3 +/- 3.5	-7.9

2B1F_HUMAN	-1.6	-117.8 +/- 5.2	-10.7
DPB1_HUMAN	-1.5	-113.1 +/- 3.6	-12.8
SARAF_HUMAN	-1.5	-100.4 +/- 11.2	-12
X5D2U9_HUMAN	-1.5	-98.3 +/- 6.0	-9.4
IL16_HUMAN	-1.5	-59.5 +/- 5.0	-7.6
DRB5_HUMAN	-1.4	-114.9 +/- 3.6	-12.2
BTBD9_HUMAN	-1.4	111.1 +/- 13.9	-10.8
PP4R1_HUMAN	-1.4	-63.7 +/- 10.3	-9
FHL17_HUMAN	-1.3	-90.5 +/- 6.3	-9.3

3.11. Analysis of Molecular Dynamics Simulations

The docked structure led to the configuration of the protein complex structure consisting of two polypeptide chains: one belonging to the MHC-II molecule and the other to an uncharacterized protein (IFOILCOA_05323). A total of 500 ns separate replicates of the molecular dynamics simulations were performed on the MHC-II (green) and uncharacterized protein (red) complexes **(Supplementary Video S1)**. The MD trajectory was analyzed for the root mean square deviation (RMSD), radius of gyration (Rg), solvent-accessible surface area (SASA), and root mean square fluctuation (RMSF) **(Figure 8)**. Each of the two simulations (run 1 - blue; run 2 - red) for the protein-protein complex over the 500 ns period are presented, with the black line showing the average of all simulations run in the case of the RMSD analysis.

All parts of this model of the MHC-II protein experienced the same initial shifts in position, but only the C-terminal region experienced a protein folding transition to a more stable conformation, a three-turn alpha helix. This contributed to global and MHC-II structural stabilization, as it retained this conformation for the rest of the simulation **(Supplementary Video S1)**. This structural stabilization is consistent with the MHC protein exhibiting a relatively high RMSD (~25-27 Å) for the first 50 ns **(Figure 8A)** of the simulation, indicative of a conformational change. This change was validated in the static RMSD profile for the first

50 ns of the simulation, contributing an increasingly stable and lower range of RMSD (~ 18-21 Å) about the average RMSD for a static profile (black) exhibiting a stable RMSD for more than 100 ns **(Figure 8A)**. Overall, this indicates that the system as a whole reached a position where it was stable (with a consistent range of conformational states) and was not likely to continue to undergo significant conformational changes.

The radius of gyration is a widely accepted parameter for gauging the compactness of protein structures during MD simulations. A lower radius of gyration indicates a more compact structure, whereas a higher radius of gyration indicates a more extended structure. Thus, the radius of gyration can be used to provide insights into the structural changes (folding/unfolding) that occur during the simulation. **Figure 8B** shows the radius of gyration (R_g) from two separate simulation runs (run 1 in blue and run 2 in red), with the average R_g shown in black. From the graph, the average R_g was high (approximately 36-38 Å), suggesting that the structure was not densely packed. Additionally, R_g experienced a rapid decrease in the early phase of the simulations (~10 ns), an outcome of the protein folding process at the N-terminus **(Supplementary Video S1)**. Following the early decline in R_g, R_g values from both runs of the simulations appeared to plateau and at approximately 200-250 ns, Rg was found to be 30 Å for both runs and the average **(Figure 8B)**, indicating that the system had achieved a stable (or equilibrium) state.

The solvent-accessible surface area (SASA) is another key metric for understanding the extent of protein folding, changes, and interaction with the solvent during molecular dynamics (MD) simulations. Once the Proteins reached a stable conformation, the SASA measurements did not differ. **Figure 8C** shows the SASA values for all individual simulation runs in red and blue, and the average is shown in black. The data show that SASA values did not differ throughout the simulation duration. The first 150 ns had higher average SASA values in the 30,000-35,000

494 Å^2 range **(Figure 8C)**. However, at and beyond 150 ns, the average SASA value was ~28,000
495 Å^2, suggesting that the protein reached a stable conformation **(Figure 8C)**.

496 During a molecular dynamics (MD) simulation, the root mean square fluctuation (RMSF) is a
497 measure of the scope of variability and the individual flexibility and mobility of every amino
498 acid residue. Fluctuation is a measure of the amplitude of movement over time, and RMSF
499 identifies the most variable section of a protein. **Figure 8D** shows the coordinates for the RMSF
500 of the protein complex for 500 ns of simulation time, while the individual MHC-II chains
501 **(Figure 8E)** and Putative Protein **(Figure 8F)** also have 500 ns of simulation time. The
502 trajectories are color-coded, with red and blue representing the trajectory simulations and the
503 black line denoting the mean of the trajectories. The more variable residues were determined
504 to be the N-terminal region by the analysis, while the RMSF estimates of the C-terminal region
505 were determined to be more stable.

506 RMSF analysis indicated notable flexibility in the putative protein and MHC-II complex. The
507 MHC-II component exhibited higher RMSF values in the putative protein-MHC-II complex
508 **(Figure 8E)**, greater fluctuations in the MHC-II protein indicate that it is more adaptable to the
509 bound putative protein. This increase in adaptability is suggested to arise from the formation
510 of the three α-helix coils (C-terminal) at the three turn-α helix structure. Regarding the MHC-
511 II protein, the putative protein showed greater RMSF, implying that it interacts more loosely
512 with the MHC protein. The N- and C-termini of MHC-II also demonstrated higher fluctuations.
513 Increased flexibility was observed among specific residues **(Figure 8F)**, namely, Asn141,
514 Asn165, Ala171, Gly182, Arg198, Ser225, Ser232, Ile234, Ile242, Ile249, Gly254, Gly260,
515 and Leu266, along with His268. The N-terminal of the putative protein was the least stable
516 region, as reflected by the higher RMSF.

The average RMSF 8 to 10 angstroms were observed for residues Ala13, Glu15, Ser18, Pro20, Ser31, Asp42, Arg50, Ala65, Thr71, Ser75, Glu124, Pro128, Ala158, Asp164, Ala190, Pro213, Pro214, Ala218, Gly234, Asp244, Leu264, Asn273. **Figure 8F** highlights this data and the average RMSF. These residues reflect the flexible regions of the molecular complex in its most stable conformation.

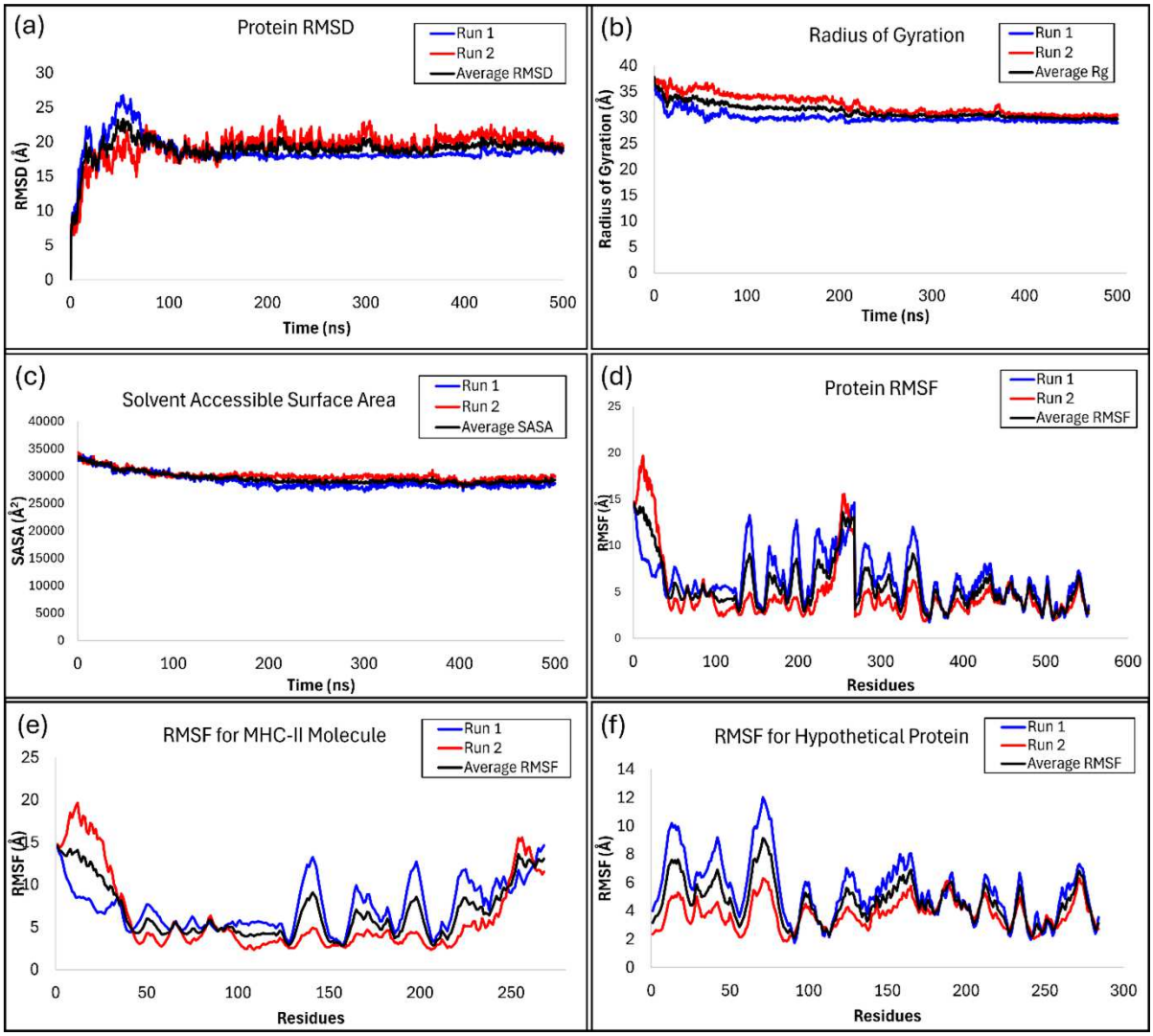


Figure 8: Analysis of RMSD, R_g, SASA, RMSF of the protein-protein complex over a 500 ns molecular dynamics simulation. The RMSF for (e) MHC-II molecule and (f) Putative protein is also depicted. The plots are shown for two independent runs, run 1 (blue) and run 2 (red). The average plots across both runs are represented in black. (a) Root-mean-square deviation (RMSD), (b) Radius of

Gyration (R_g), (c) Solvent Accessible Surface Area (SASA), and Root Mean Square Fluctuation (RMSF)

3.12. DALI for the docked complex of the protein

The DALI server[61] was used to perform a structural similarity assessment to obtain a preliminary guess of the identity and putative biological function of the putative protein. DALI returned a total of 999 protein structures with positive and significant (Z-score > 2) structural likenesses to the putative protein. These 999 structures formed a subset of approximately 341 unique protein entries. To obtain a better idea of the possible function of the putative protein in the docked complex, protein structures with a high positive structural resemblance (Z-score > 30) were chosen. Six chains achieved Z-scores greater than 30. Structural overlays of the putative protein and the other three proteins are shown in **Figure 9A**, while overlays with the TssM-UB-PA complex (PDB; 8Q00, chains A and C) and TssM-UB (PDB; 8S5I, chains A and C) are shown in **Figure 9A and D and Figure 9C and E**, respectively. The proteins of interest are members of the Type VI secretion system, particularly the TssM protein, a deubiquitinase.

While we recognize that these data strongly support a TssM-like deubiquitinase, its function still needs to be confirmed if there exists a catalytic triad residue conservation, along with Ub-binding pockets, and additional biochemical validation to ascertain the presence of deubiquitination. Three PDB entries provided six similar structures, 8Q00, 8S5I, and 8PZ3, along with their corresponding chains A and B.

Structural superimposition of the given putative protein with the proteins obtained from the DALI analysis **(Figure 9A)** showed 0.436 – 1.096 Å root mean square deviation (RMSD) scores, which indicated similarity with known deubiquitinating enzymes from *Burkholderia pseudomallei*. For example, superposition with the TssM-ubiquitin complex (PDB ID: 8SSI) yielded values of 0.558 and 0.740 Å for chains A and C, respectively, which also showed close

structural similarity. For the TssM-Ub-PA complexes (PDB IDs: 8Q00 and 8PZ3), superposition showed structural similarity to chains A and C (0.436 Å and 1.049 Å) for 8Q00 and chains A and B (1.096 Å and 0.809 Å) for 8PZ3. Such low RMSD values testify to the structural similarity between the putative protein and deubiquitinases. If the putative protein is similar to TssM proteins, its role as a deubiquitinase and involvement in the relevant processes of this protein are highly probable. A structural similarity search of MHC class II molecules yielded 999 proteins with structural similarities corresponding to the α1/β1 platform and α2/β2 Ig-like domains of MHC-II molecules.

The Z-scores for these hits ranged from 20 to 16.5, indicating a considerable level of structural similarity, consistent with the recognized identity and function of the MHC-II molecule. As indicated, functional inference is exceedingly structural, and highly TssM is concordant, but functional assignment is definitive conservation and requires a catalytic site assay, which is designed to follow up on experiments.

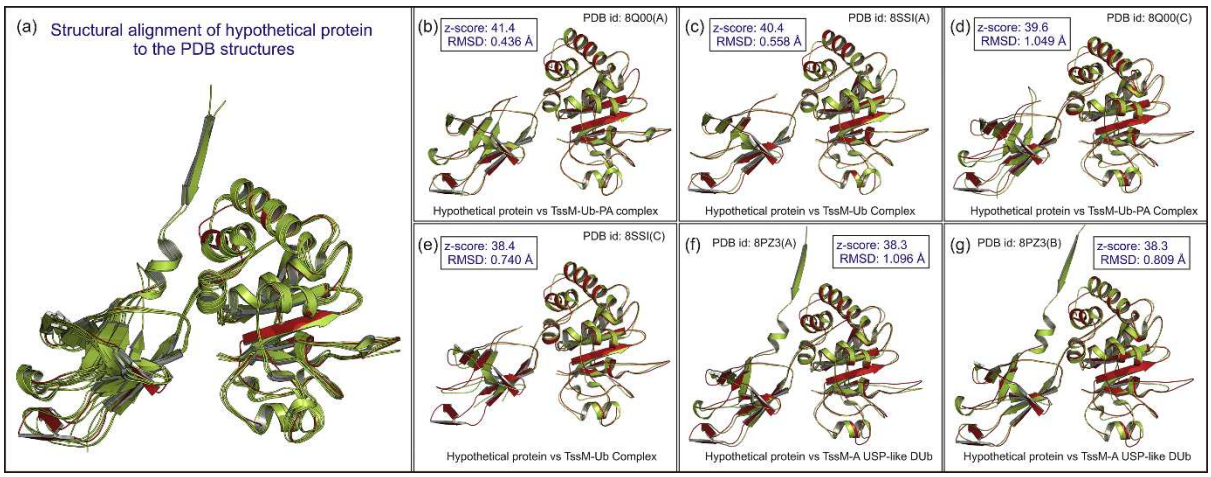


Figure 9: Structural similarity of *Burkholderia pseudomallei* deubiquitinase domains with the docked putative protein. (a) Structural alignment of the putative protein with reference PDB structures. (b–e) Structural superimposition of six chains from *Burkholderia pseudomallei* deubiquitinases. The putative protein is shown in red, while the structurally similar chains identified using the DALI server are shown in green. RMSD values after alignment and the corresponding DALI z-scores are provided in the respective insets.

Discussion

The current research provides comprehensive *in silico* results of the putative proteins (PP) of *Burkholderia pseudomallei* CM000113 and seeks to determine their functionality regarding pathogenicity, antigenicity, and therapeutic application **(Figure 1)**. Interestingly, we discovered that almost 41.41% of the proteome was devoid of functional annotation, which has plagued genomic research of *Burkholderia* sp. for some time, where such coding regions remain labeled as putative **(Figure 2**). This has previously been documented to occur with *Burkholderia pseudomallei* and other related *Burkholderia* species, as these organisms consistently contain a PP set far exceeding that of well-characterized bacterial genomes[17,62–65]. This demonstrates the lack of knowledge about the biology of this pathogen and reinforces the idea that PPs contain unrecognized value in the form of latent knowledge related to virulence, host-pathogen interactions, and drug discovery.

From our comparative analyses utilizing the Virulence Factor Database (VFDB), we identified 20 PP that exhibited over 90 % of the documented virulence factors, indicative of functional conservation. Furthermore, the majority of these proteins[19] were predicted to possess antigenic properties (i.e., PPs with the potential to be immunogenic). This is consistent with our previous study[22], which showed that epitopes derived from the PPs of *Burkholderia pseudomallei* are eligible for inclusion within multi-epitope vaccine frameworks **(Table 1)**[66]. Taken together, these findings bolster the argument that PP is a largely uncharacterized subset of virulence factor constituents that could prove valuable for vaccine formulation.

Some PPs have also been connected to domains via functional annotations associated with virulence. For instance, the presence of domains of the papain-like cysteine peptidase superfamily suggests a possible function in active immunosuppression or degradation of host proteins required for intracellular parasite survival[67]. Likewise, the bacterial immunoglobulin-

like (Big-5) domain, known to facilitate immune evasion and promote adhesion[68], is in line with the suggestion of its participation in host-pathogen interrelations. This is consistent with the fact that *B. pseudomallei* has been documented to cross a host range using a sizable and multifunctional suite of secreted and cell-surface proteins to subvert host defense mechanisms[69-71].

The fact that most of the putative PPs (VaxiJen score ≥ 0.4) are predicted to elicit an immune response helps bolster the argument for their inclusion as likely candidates for incorporation into a vaccine. This is especially true considering that the entry criterion has been applied to other bacteria, such as *Mycobacterium tuberculosis* and *Klebsiella pneumoniae*, to facilitate the identification of proteins that are likely to elicit an immune response[72-77]. The identification of putative epitopes predicted to be nontoxic and nonallergenic is equally reassuring to overcome the longstanding challenge in vaccine formulation associated with the antigenicity and safety of potential vaccine candidates **(Table 1)**.

The predicted interaction between one PP and HLA-DQB2 further emphasizes this immunological relevance **(Figure 7)**. High-affinity interactions from molecular docking were confirmed using long-timescale molecular dynamics simulations. The stability of host–pathogen complexes suggests that these complexes can manipulate the pathways involved in antigen presentation. *B. pseudomallei* has been shown to manipulate host immune signalling by disrupting NF-κB and type I interferon signalling[10,17,78]. These findings suggest that PPs may also be involved in immune subversion through direct interactions with MHC molecules. This hypothesis has been further supported by computational methods.

One PP showed a clear similarity with a DrugBank target, ubiquitin carboxyl-terminal hydrolase 14 **(Figure 3A)**. This aligns with the findings of recent structural studies on *B. pseudomallei* TssM, a putative deubiquitinase located in the Type VI secretion system[78–80]. The

DALI structural comparisons also confirmed the close similarity between the modeled PP and TssM-associated complexes, which adds to the accumulating evidence suggesting that this PP may act as a deubiquitinase **(Figure 9)**. This class of enzymes is increasingly recognized as an important virulence determinant of intracellular pathogens.

Deubiquitinases remain understudied; however, they are becoming potential drug targets, as they can neutralize host immunity mediated by ubiquitin[81-86]. Pathogens can blunt host defense by deubiquitinating immune substrates to perturb immune defense responses. Inhibiting bacterial deubiquitinases may attenuate the virulence of pathogens without bactericidal pressure and possibly minimize selective resistance[82,87]. Thus, we suggest that PP should be the target of further studies, as it may assist in the design of targeted structural inhibitors.

The generation and validation of accurate structural models using GalaxyWEB and PROCHECK, ProSA, and ERRAT are attributed to the integrated use of MODELLER, I-TASSER, and Robetta software. More than 94% of the final structure had Ramachandran favored residues, while the Z-score was within the accepted range of Z-scores of structures from the same model deposition **(Figure 5, Table 3)**. The validation results achieved are equal to and in some cases surpass even the thresholds accepted to be valid in the literature for *de novo* predicted bacterial proteins[88,89]. The predicted secondary structure elements of the proteins were random coil and alpha-helix dominant, which predicts flexibility, a characteristic common among effector proteins that need to be versatile in structure in response to diverse host targets[90]. The predicted active sites also demonstrated spatial clustering and, like papain-like proteases, were around a catalytic triad of residues conserved across the family, Cys-His-Asn **(Supplementary Figure S1)**. The same conservation was noted for *Burkholderia* proteins, where virulence attenuation and loss of proteolytic activity were attributed to mutations at the catalytic cysteine residues[91,92]. Therefore, the results of this computational study provide predictive evidence for the functions of these proteins.

In silico host-pathogen interactome mapping predicted interactions between PP and human proteins, including HLAs, ferritins, and signalers such as PP4R1 **(Table 4)**. Docking analyses identified the PP-HLA-DQB2 interaction as the strongest, with a HADDOCK score of -117 kcal/mol and low RMSD across 500 ns **(Figure 7 & 8, Table 5)**. The role of HLA molecules is significant because they play a role in pathways central to adaptive immunity by presenting antigens. Bacterial effectors that bind HLA may inactivate host responses by blocking peptide loading, preventing surface expression, or inhibiting T-cell activation, which may underlie host response branching.

Similar mechanisms have been reported in other intracellular bacteria. *For* example, *Mycobacterium tuberculosis* uses PE/PPE proteins to manipulate HLA expression and antigen processing[93–95]. These results suggest that *B. pseudomallei* may use PPs to compromise host antigen presentation. These findings could ignite primary immunological studies to establish the effect of PP binding on MHC-II function, even if they remain to be validated experimentally based on evidence from computational studies.

The 500 ns MD simulations exhibited important dynamic features of the system. Supporting evidence lies in the significant stabilization of the PP-HLA complex, as revealed by the fluctuations in the radius of gyration and RMSD. The decreasing SASA in the trajectory of SASA over the course of the simulations is indicative of surface burial, which leads to the conclusion that protein complexes are stable and buried, interactively hydrophobic regions of surface residues. Furthermore, the residue RMSF indicated that the N-terminal domains contained the largest RMSF values; therefore, the N-terminal domain likely carries out functions, such as facilitating binding via the N-terminal domains, which greatly changes their position **(Figure 9)**.

669 The presence of rearrangements that led to stable conformations of the complex was observed,
670 as in several other MD simulations of complexes between bacterial effectors and their hosts[96].
671 Notably, stabilization of the C-terminal helix in MHC II was evident. Furthermore, other
672 conformational changes in the complex were documented for antigen-presenting molecules, as
673 they also undergo changes in their C-terminus upon binding to effectors[97,98]. All this evidence
674 contributed to the conclusion that PP-HLA are likely to be physically and biologically linked.

675 While our findings align with and broaden the scope of the virulence factor studies conducted
676 on *B. pseudomallei*[22], there are still unexplored aspects that are yet to be investigated. The
677 studies noted the numerous and uncharacterized components of the secretion systems and their
678 associated proteins found in Indian clinical isolates. The current study also provides additional
679 evidence of the presence of an PP located next to TssM in the genome, reinforcing its putative
680 function in the Type VI secretion system **(Figure 2)**. In addition, some recurrent HLA
681 molecules[21] were predicted as probable targets in their computational modeling of the host-
682 pathogen interaction framework for *B. pseudomallei*. For structural predictions, our docking
683 and MD analyses offered further specificity with respect to PP-mediated interactions.

684 Relying on bioinformatics for predictions means that particular care should be taken in their
685 interpretation. While the overall structural approximation and docking scores are quite
686 impressive, some of the functionalities will need to be corroborated with wet lab studies, which
687 could include, for example, deubiquitinase activity, host protein pull-down assay, and
688 subsequently an infection assay to evaluate putative virulence attenuation. In addition,
689 antigenicity predictions need to be corroborated with immunological studies, for example,
690 through ELISA with the sera of infected patients or T-cell proliferation assays.

691 The endeavors in the forecasted future must be oriented toward the following objectives: (i)
692 unambiguously cloning and expressing the prioritized putative protein (PP); (ii) undertaking

693 an assortment of canonical functional assays to demonstrate the presence of deubiquitinase
694 activity; (iii) obtaining the 3D structure using X-ray crystallography and/or cryo-electron
695 microscopy; and (iv) measuring the interactions with different variants (alleles) of Human HLA
696 (immunological assays on HLA interactions). If achieved, the outcomes of these additional
697 computational and experimental steps will constitute the first of many steps necessary for
698 further progress in various fields, particularly in the development of instruments for disease
699 detection, design of new vaccines, and expanded options for therapeutics.

700 **Conclusions**

701 This study meticulously examined the proteins from *Burkholderia pseudomallei* strain
702 CM000113 for their ability to act as virulence factor or potential candidates for vaccines, and
703 therapeutic drugs. One high-priority protein had structural similarities to TssM deubiquitinase,
704 a catalytic triad of a potential substrate, and a stable fit for HLA-DQB2, implying that it might
705 modulate the immune system. These proteins also have sufficient potential for inclusion in
706 vaccines. The *B. pseudomallei* structural and docking models, in conjunction with molecular
707 dynamics, broadened the understanding of the pathogen's mechanisms, the findings of which
708 should be adapted for laboratory experimental purposes and then used to create a vaccine to
709 treat diseases caused by pathogens resistant to multiple drugs, particularly melioidosis.

710 **Declarations**

711 **Ethical approval**

712 The authors declare that their study does not need approval by an appropriate institutional
713 committee because the research in this paper does not involve experimentation with human
714 subjects or animals.

715 **Data Availability**

716 The genome assembled during this study are available in the NCBI Genome database under
717 BioProject PRJNA889557 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA889557/) and

BioSample SAMN31248483 (https://www.ncbi.nlm.nih.gov/biosample/SAMN31248483/). All data generated or analyzed during this study are included in the manuscript and its supplementary information file.

Author contributions

Supriyo Mukherjee: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Rajarshi Bhattacharya:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Thripthi Nagesh Shenoy:** Methodology, Visualization, Writing – review & editing. **Chiranjay Mukhopadhyay:** Resources, Supervision, Writing – review & editing. **Somasish Ghosh Dastidar:** Funding acquisition, Investigation, Supervision, Writing – original draft, Writing – review & editing. **Saugata Hazra:** Supervision, Writing – original draft, Writing – review & editing. **Abdul Ajees Abdul Salam:** Data curation, Formal Analysis, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ranita Ghosh Dastidar:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review & editing.

Conflicts of Interest

All the authors mentioned above mutually agree to submit this article to Scientific Reports. Furthermore, all authors confirm that there is no financial or non-financial conflict of interest among any of the authors.

Funding information

This work is supported by ICMR Adhoc grant (Project ID: 2021-13710; File No.: OMI/02/2022/ECD), awarded to RGD and DBT-Ramalingaswami Fellowship (SAN No: 102/IFD/SAN/2549/2019–20), awarded to SGD.

Acknowledgement

The authors gratefully acknowledge Kasturba Medical College, Manipal, Manipal Academy of Higher Education (MAHE) for Infrastructure and Resources for providing the research facilities and the Indian Council of Medical Research (ICMR) and Department of Biotechnology (DBT), India for financial support.

References:

1. Tullius, M. V. *et al.* LVS ΔcapB-vectored multiantigenic melioidosis vaccines protect against lethal respiratory *Burkholderia pseudomallei* challenge in highly sensitive BALB/c mice. *mBio* **15**, (2024).
https://doi.org/10.1128/mbio.00186-24

2. Petras, J. K. *et al.* Locally Acquired Melioidosis Linked to Environment — Mississippi, 2020–2023. *N Engl J Med* **389**, 2355–2362 (2023).
https://doi.org/10.1056/nejmoa2306448

3. Gee, J. E. *et al.* Multistate Outbreak of Melioidosis Associated with Imported Aromatherapy Spray. *N Engl J Med* **386**, 861–868 (2022).
https://doi.org/10.1056/nejmoa2116130

4. Wiersinga, W. J. *et al.* Melioidosis. *Nat Rev Dis Primers* **4**, (2018).
https://doi.org/10.1038/nrdp.2017.107

5. Hussin, A. *et al.* Improving the clinical recognition, prognosis, and treatment of melioidosis through epidemiology and clinical findings: The Sabah perspective. *PLoS Negl Trop Dis* **17**, e0011696 (2023).
https://doi.org/10.1371/journal.pntd.0011696

6. Limmathurotsakul, D. *et al.* Predicted global distribution of *Burkholderia pseudomallei* and burden of melioidosis. *Nat Microbiol* **1**, (2016).
https://doi.org/10.1038/nmicrobiol.2015.8

7. Mohapatra, P. R. & Mishra, B. Burden of melioidosis in India and South Asia: Challenges and ways forward. *The Lancet Regional Health - Southeast Asia* **2**, 100004 (2022).
https://doi.org/10.1016/j.lansea.2022.03.004

8. Webb, J. R. *et al.* Myanmar *Burkholderia pseudomallei* strains are genetically diverse and originate from Asia with phylogenetic evidence of reintroductions from neighbouring countries. *Sci Rep* **10**, (2020).
https://doi.org/10.1038/s41598-020-73545-8

9. Dawson, P. *et al.* Human Melioidosis Caused by Novel Transmission of *Burkholderia pseudomallei* from Freshwater Home Aquarium, United States1. *Emerg. Infect. Dis.* **27**, 3030–3035 (2021).
https://doi.org/10.3201/eid2712.211756

10. Mariappan, V. *et al.* Hijacking of the Host's Immune Surveillance Radars by *Burkholderia pseudomallei*. *Front. Immunol.* **12**, (2021).

https://doi.org/10.3389/fimmu.2021.718719
11. Pakkulnan, R., Thonglao, N. & Chareonsudjai, S. DNase I and chitosan enhance efficacy of ceftazidime to eradicate *Burkholderia pseudomallei* biofilm cells. *Sci Rep* **13**, (2023).
https://doi.org/10.1038/s41598-023-27790-2
12. Bugrysheva, J. V. *et al.* Antibiotic Resistance Markers in *Burkholderia pseudomallei* Strain Bp1651 Identified by Genome Sequence Analysis. *Antimicrob Agents Chemother* **61**, (2017).
https://doi.org/10.1128/aac.00010-17
13. Terreni, M., Taccani, M. & Pregnolato, M. New Antibiotics for Multidrug-Resistant Bacterial Strains: Latest Research Developments and Future Perspectives. *Molecules* **26**, 2671 (2021).
https://doi.org/10.3390/molecules26092671
14. Harris, M., Fasolino, T., Ivankovic, D., Davis, N. J. & Brownlee, N. Genetic Factors That Contribute to Antibiotic Resistance through Intrinsic and Acquired Bacterial Genes in Urinary Tract Infections. *Microorganisms* **11**, 1407 (2023).
https://doi.org/10.3390/microorganisms11061407
15. Al-Khikani, F. H. & Ayit, A. S. *Pseudomonas Aeruginosa* a Tenacious Uropathogen. *Biomed Biotechnol Res J* **6**, 311–318 (2022).
https://doi.org/10.4103/bbrj.bbrj_256_21
16. Holden, M. T. G. *et al.* Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14240–14245 (2004).
https://doi.org/10.1073/pnas.0403302101
17. Heacock-Kang, Y. *et al.* The *Burkholderia pseudomallei* intracellular 'TRANSITome'. *Nat Commun* **12**, (2021).
https://doi.org/10.1038/s41467-021-22169-1
18. Deshazer, D. Genomic diversity of *Burkholderia pseudomallei* clinical isolates: subtractive hybridization reveals a *Burkholderia mallei*-specific prophage in *B. pseudomallei* 1026b. *J Bacteriol* **186**, 3938–3950 (2004).
https://doi.org/10.1128/jb.186.12.3938-3950.2004
19. Gan, Y. *et al.* Characterization of *Burkholderia pseudomallei* infection and identification of novel virulence factors using a *Caenorhabditis elegans* host system. *Molecular Microbiology* **44**, 1185–1197 (2002).
https://doi.org/10.1046/j.1365-2958.2002.02957.x
20. Wiersinga, W. J. & Van Der Poll, T. Immunity to *Burkholderia pseudomallei*. *Current Opinion in Infectious Diseases* **22**, 102–108 (2009).

https://doi.org/10.1097/qco.0b013e328322e727
21. Loaiza, C. D., Duhan, N., Lister, M. & Kaundal, R. *In silico* prediction of host-pathogen protein interactions in melioidosis pathogen *Burkholderia pseudomallei* and human reveals novel virulence factors and their targets. *Briefings in Bioinformatics* **22**, (2020).
https://doi.org/10.1093/bib/bbz162
22. Varshith, M. R. *et al.* Virulome and phylogenomic profiling of a novel *Burkholderia pseudomallei* strain from an Indian clinical isolate. *Mol Genet Genomics* **299**, (2024).
https://doi.org/10.1007/s00438-024-02188-5
23. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**, e1005595 (2017).
https://doi.org/10.1371/journal.pcbi.1005595
24. Brettin, T. *et al.* RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* **5**, (2015).
https://doi.org/10.1038/srep08365
25. Zhou, S., Liu, B., Zheng, D., Chen, L. & Yang, J. VFDB 2025: an integrated resource for exploring anti-virulence compounds. *Nucleic Acids Research* **53**, D871–D877 (2024).
https://doi.org/10.1093/nar/gkae968
26. Konerding, D. An Essential Guide to the Basic Local Alignment Search Tool: BLAST. *Briefings in Bioinformatics* **5**, 93–94 (2004).
https://doi.org/10.1093/bib/5.1.93
27. Götz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**, 3420–3435 (2008).
https://doi.org/10.1093/nar/gkn176
28. Doytchinova, I. A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* **8**, (2007).
https://doi.org/10.1186/1471-2105-8-4
29. Knox, C. *et al.* DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Research* **52**, D1265–D1275 (2023).
https://doi.org/10.1093/nar/gkad976
30. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
https://doi.org/10.1093/bioinformatics/btu031
31. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412–D419 (2020).

https://doi.org/10.1093/nar/gkaa913
32. Wilson, D. *et al.* SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Research* **37**, D380–D386 (2008).
https://doi.org/10.1093/nar/gkn762
33. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Research* **43**, D222–D226 (2014).
https://doi.org/10.1093/nar/gku1221
34. Wang, J. *et al.* The conserved domain database in 2023. *Nucleic Acids Research* **51**, D384–D388 (2022).
https://doi.org/10.1093/nar/gkac1096
35. Sillitoe, I. *et al.* CATH: increased structural coverage of functional space. *Nucleic Acids Research* **49**, D266–D273 (2020).
https://doi.org/10.1093/nar/gkaa1079
36. Yu, C., Lin, C. & Hwang, J. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Science* **13**, 1402–1406 (2004).
https://doi.org/10.1110/ps.03479604
37. Yu, C., Chen, Y., Lu, C. & Hwang, J. Prediction of protein subcellular localization. *Proteins* **64**, 643–651 (2006).
https://doi.org/10.1002/prot.21018
38. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods in molecular biology (Clifton, N.J.)* **112**, 531–552 (1999).
https://doi.org/10.1385/1-59259-584-7:531
39. Grant, J. R. *et al.* Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research* **51**, W484–W492 (2023).
https://doi.org/10.1093/nar/gkad326
40. Gertz, E. M., Yu, Y.-K., Agarwala, R., Schäffer, A. A. & Altschul, S. F. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biol* **4**, (2006).
https://doi.org/10.1186/1741-7007-4-41
41. Mcguffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
https://doi.org/10.1093/bioinformatics/16.4.404
42. Buchan, D. W. A., Moffat, L., Lau, A., Kandathil, S. M. & Jones, D. T. Deep learning for

the PSIPRED Protein Analysis Workbench. *Nucleic acids research* **52**, W287–W293 (2024). https://doi.org/10.1093/nar/gkae328

43. Høie, M. H. *et al.* NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Research* **50**, W510–W515 (2022). https://doi.org/10.1093/nar/gkac439

44. Geourjon, C. & Deléage, G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* **11**, 681–684 (1995). https://doi.org/10.1093/bioinformatics/11.6.681

45. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *CP in Bioinformatics* **54**, (2016). https://doi.org/10.1002/cpbi.3

46. Zhou, X. *et al.* I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nat Protoc* **17**, 2326–2353 (2022). https://doi.org/10.1038/s41596-022-00728-0

47. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* **32**, W526–W531 (2004). https://doi.org/10.1093/nar/gkh468

48. Ko, J., Park, H., Heo, L. & Seok, C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Research* **40**, W294–W297 (2012). https://doi.org/10.1093/nar/gks493

49. The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC. https://pymol.org/

50. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* **26**, 283–291 (1993). https://doi.org/10.1107/S0021889892009944

51. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science* **2**, 1511–1519 (1993). https://doi.org/10.1002/pro.5560020916

52. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of

errors in three-dimensional structures of proteins. *Nucleic Acids Research* **35**, W407–W410 (2007).
https://doi.org/10.1093/nar/gkm290
53. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* **38**, 3022–3027 (2021). https://doi.org/10.1093/molbev/msab120
54. Ye, B., Tian, W., Wang, B. & Liang, J. CASTpFold: Computed Atlas of Surface Topography of the universe of protein Folds. *Nucleic acids research* **52**, W194–W199 (2024). https://doi.org/10.1093/nar/gkae415
55. Honorato, R. V. *et al.* The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nat Protoc* **19**, 3219–3241 (2024).
https://doi.org/10.1038/s41596-024-01011-0
56. Honorato, R. V. *et al.* Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* **8**, (2021).
https://doi.org/10.3389/fmolb.2021.729513
57. Bhattacharya, R., Gupta, A. M., Mitra, S., Mandal, S. & Biswas, S. R. A natural food preservative peptide nisin can interact with the SARS-CoV-2 spike protein receptor human ACE2. *Virology* **552**, 107–111 (2020).
https://doi.org/10.1016/j.virol.2020.10.002
58. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).
https://doi.org/10.1093/bioinformatics/btw514
59. Bowers, K. J. *et al.* Molecular dynamics---Scalable algorithms for molecular dynamics simulations on commodity clusters. in **84** (*Association for Computing Machinery*, 2006).
https://doi.org/10.1145/1188455.1188544
60. Lu, C. *et al.* OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **17**, 4291–4300 (2021).
https://doi.org/10.1021/acs.jctc.1c00302
61. Holm, L. Dali server: structural unification of protein families. *Nucleic acids research* **50**, W210–W215 (2022).
https://doi.org/10.1093/nar/gkac387

947 62. Lee, S.-H. *et al. Burkholderia pseudomallei* suppresses *Caenorhabditis elegans* immunity
948 by specific degradation of a GATA transcription factor. *Proc. Natl. Acad. Sci. U.S.A.* **110**,
949 15067–15072 (2013).
950 https://doi.org/10.1073/pnas.1311725110
951 63. Paksanont, S. *et al.* Effect of temperature on *Burkholderia pseudomallei* growth, proteomic
952 changes, motility and resistance to stress environments. *Sci Rep* **8**, (2018).
953 https://doi.org/10.1038/s41598-018-27356-7
954 64. Borlee, G. I. *et al.* Genome-scale analysis of the genes that contribute to *Burkholderia*
955 *pseudomallei* biofilm formation identifies a crucial exopolysaccharide biosynthesis gene
956 cluster. *PLoS Negl Trop Dis* **11**, e0005689 (2017).
957 https://doi.org/10.1371/journal.pntd.0005689
958 65. Seng, R. *et al.* Genetic diversity, determinants, and dissemination of *Burkholderia*
959 *pseudomallei* lineages implicated in melioidosis in Northeast Thailand. *Nat Commun* **15**,
960 (2024).
961 https://doi.org/10.1038/s41467-024-50067-9
962 66. Bazmara, H. *et al.* Antigenic Properties of Iron Regulated Proteins in *Acinetobacter*
963 *baumannii*: An *In Silico* Approach. *Int J Pept Res Ther* **25**, 205–213 (2017).
964 https://doi.org/10.1007/s10989-017-9665-6
965 67. Ozhelvaci, F. & Steczkiewicz, K. Identification and classification of papain-like cysteine
966 proteinases. *Journal of Biological Chemistry* **299**, 104801 (2023).
967 https://doi.org/10.1016/j.jbc.2023.104801
968 68. Chatterjee, S., Basak, A. J., Nair, A. V., Duraivelan, K. & Samanta, D. Immunoglobulin-
969 fold containing bacterial adhesins: molecular and structural perspectives in host tissue
970 colonization and infection. *FEMS Microbiology Letters* **368**, (2020).
971 https://doi.org/10.1093/femsle/fnaa220
972 69. Nan, D. *et al. Burkholderia pseudomallei* BipD modulates host mitophagy to evade killing.
973 *Nat Commun* **15**, (2024).
974 https://doi.org/10.1038/s41467-024-48824-x
975 70. Custódio, R. *et al.* Characterization of secreted sphingosine-1-phosphate lyases required
976 for virulence and intracellular survival of *Burkholderia pseudomallei*. *Molecular Microbiology*
977 **102**, 1004–1019 (2016).
978 https://doi.org/10.1111/mmi.13531
979 71. Bzdyl, N. M., Moran, C. L., Bendo, J. & Sarkar-Tyson, M. Pathogenicity and virulence of
980 *Burkholderia pseudomallei*. *Virulence* **13**, (2022).

https://doi.org/10.1080/21505594.2022.2139063
72. Phogat, S., Yadav, J., Chaudhary, D., Jaiwal, R. & Jaiwal, P. K. Synthesis of an Adjuvant-Free Single Polypeptide-Based Tuberculosis Subunit Vaccine that Elicits *In Vivo* Immunogenicity in Rats. *Mol Biotechnol* (2025)
https://doi.org/10.1007/s12033-025-01431-7
73. Lundberg, U., Senn, B. M., Schüler, W., Meinke, A. & Hanner, M. Identification and characterization of antigens as vaccine candidates against *Klebsiella pneumoniae*. *Human Vaccines & Immunotherapeutics* **9**, 497–505 (2013).
https://doi.org/10.4161/hv.23225
74. Moodley, A. *et al.* Reverse vaccinology approach to design a multi-epitope vaccine construct based on the *Mycobacterium tuberculosis* biomarker PE_PGRS17. *Immunol Res* **70**, 501–517 (2022).
https://doi.org/10.1007/s12026-022-09284-x
75. Andongma, B. T. *et al. In silico* design of a promiscuous chimeric multi-epitope vaccine against *Mycobacterium tuberculosis*. *Computational and Structural Biotechnology Journal* **21**, 991–1004 (2023).
https://doi.org/10.1016/j.csbj.2023.01.019
76. Rahmat Ullah, S., Majid, M., Rashid, M. I., Mehmood, K. & Andleeb, S. Immunoinformatics Driven Prediction of Multiepitopic Vaccine Against *Klebsiella pneumoniae* and *Mycobacterium tuberculosis* Coinfection and Its Validation via *In Silico* Expression. *Int J Pept Res Ther* **27**, 987–999 (2020).
https://doi.org/10.1007/s10989-020-10144-1
77. Jiang, F. *et al.* A comprehensive approach to developing a multi-epitope vaccine against *Mycobacterium tuberculosis*: from *in silico* design to *in vitro* immunization evaluation. **Front. Immunol. 14**, (2023).
https://doi.org/10.3389/fimmu.2023.1280299
78. Tan, K. S. *et al.* Suppression of Host Innate Immune Response by *Burkholderia pseudomallei* through the Virulence Factor TssM. *The Journal of Immunology* **184**, 5160–5171 (2010).
https://doi.org/10.4049/jimmunol.0902663
79. Burtnick, M. N. *et al.* The Cluster 1 Type VI Secretion System Is a Major Virulence Determinant in *Burkholderia pseudomallei*. *Infect Immun* **79**, 1512–1525 (2011).
https://doi.org/10.1128/iai.01218-10
80. Burtnick, M. N., Brett, P. J. & Deshazer, D. Proteomic analysis of the *Burkholderia*

pseudomallei type II secretome reveals hydrolytic enzymes, novel proteins, and the deubiquitinase TssM. *Infect Immun* **82**, 3214–3226 (2014).
https://doi.org/10.1128/iai.01739-14
81. Zhang, X. *et al.* Extracellular fibrinogen-binding protein released by intracellular *Staphylococcus aureus* suppresses host immunity by targeting TRAF3. *Nat Commun* **13**, (2022).
https://doi.org/10.1038/s41467-022-33205-z
82. Wan, M. *et al.* A bacterial effector deubiquitinase specifically hydrolyses linear ubiquitin chains to inhibit host inflammatory signalling. *Nat Microbiol* **4**, 1282–1293 (2019).
https://doi.org/10.1038/s41564-019-0454-1
83. Manzanillo, P. S. *et al.* The ubiquitin ligase parkin mediates resistance to intracellular pathogens. *Nature* **501**, 512–516 (2013).
https://doi.org/10.1038/nature12566
84. Kayagaki, N. *et al.* A Deubiquitinase That Regulates Type I Interferon Production. *Science* **318**, 1628–1632 (2007).
https://doi.org/10.1126/science.1145918
85. Ma, K. *et al.* Bacterial ubiquitin ligases hijack the host deubiquitinase OTUB1 to inhibit MTORC1 signaling and promote autophagy. *Autophagy* **20**, 1968–1983 (2024).
https://doi.org/10.1080/15548627.2024.2353492
86. Song, J. *et al.* The deubiquitinase OTUD1 enhances iron transport and potentiates host antitumor immunity. *EMBO Reports* **22**, (2021).
https://doi.org/10.15252/embr.202051162
87. Vozandychova, V., Stojkova, P., Hercik, K., Rehulka, P. & Stulik, J. The Ubiquitination System within Bacterial Host-Pathogen Interactions. *Microorganisms* **9**, 638 (2021).
https://doi.org/10.3390/microorganisms9030638
88. Saxena, S., Abdullah, M., Sriram, D. & Guruprasad, L. Discovery of novel inhibitors of *Mycobacterium tuberculosis* MurG: homology modelling, structure based pharmacophore, molecular docking, and molecular dynamics simulations. *Journal of Biomolecular Structure and Dynamics* **36**, 3184–3198 (2017).
https://doi.org/10.1080/07391102.2017.1384398
89. Fakhar, Z. *et al.* A comparative modeling and molecular docking study on *Mycobacterium tuberculosis* targets involved in peptidoglycan biosynthesis. *Journal of Biomolecular Structure and Dynamics* **34**, 2399–2417 (2016).

https://doi.org/10.1080/07391102.2015.1117397
90. Zhang, Z.-M. *et al.* Structure of a pathogen effector reveals the enzymatic mechanism of a novel acetyltransferase family. *Nat Struct Mol Biol 23*, 847–852 (2016).
https://doi.org/10.1038/nsmb.3279
91. Yao, Q. *et al.* A bacterial type III effector family uses the papain-like hydrolytic activity to arrest the host cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3716–3721 (2009).
https://doi.org/10.1073/pnas.0900212106
92. Zeiler, E. *et al.* Structural and functional insights into caseinolytic proteases reveal an unprecedented regulation principle of their catalytic triad. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11302–11307 (2013).
https://doi.org/10.1073/pnas.1219125110
93. Zhang, Z., Dong, L., Li, X., Deng, T. & Wang, Q. The PE/PPE family proteins of *Mycobacterium tuberculosis*: evolution, function, and prospects for tuberculosis control. *Front. Immunol.* **16**, (2025).
https://doi.org/10.3389/fimmu.2025.1606311
94. D'Souza, C., Kishore, U. & Tsolaki, A. G. The PE-PPE Family of *Mycobacterium tuberculosis*: Proteins in Disguise. *Immunobiology* **228**, 152321 (2022).
https://doi.org/10.1016/j.imbio.2022.152321
95. Guo, F. *et al.* Immunological effects of the PE/PPE family proteins of *Mycobacterium tuberculosis* and related vaccines. *Front. Immunol.* **14**, (2023).
https://doi.org/10.3389/fimmu.2023.1255920
96. Iyer, S. & Das, C. The unity of opposites: Strategic interplay between bacterial effectors to regulate cellular homeostasis. *Journal of Biological Chemistry* **297**, 101340 (2021).
https://doi.org/10.1016/j.jbc.2021.101340
97. Wieczorek, M. *et al.* Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Front. Immunol.* **8**, (2017).
https://doi.org/10.3389/fimmu.2017.00292
98. Loi, M., Gannagé, M. & Münz, C. ATGs help MHC class II, but inhibit MHC class I antigen presentation. *Autophagy* **12**, 1681–1682 (2016).
https://doi.org/10.1080/15548627.2016.1203488

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryVideoS1.mp4
- SupplementaryFile.docx