

RESEARCH ARTICLE

Core-genome-mediated promising alternative drug and multi-epitope vaccine targets prioritization against infectious *Clostridium difficile*

Sara Aiman¹✉, Qurrat ul Ain Farooq¹✉, Zhongjie Han¹, Muneeba Aslam², Jilong Zhang¹, Asifullah Khan^{2*}, Abbas Ahmad³, Chunhua Li^{1*}, Yasir Ali^{4*}

1 Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing, China, **2** Department of Biochemistry, Abdul Wali Khan University Mardan, Mardan, Pakistan, **3** Department of Biotechnology, Abdul Wali Khan University Mardan, Mardan, KP, Pakistan, **4** School of Biomedical Sciences, Chinese University of Hong Kong, Hong Kong, Hong Kong

✉ These authors contributed equally to this work.

* asif@awkum.edu.pk (AK); chunhuali@bjut.edu.cn (CL); yasirali@cuhk.edu.hk (YA)



OPEN ACCESS

Citation: Aiman S, Farooq QuA, Han Z, Aslam M, Zhang J, Khan A, et al. (2024) Core-genome-mediated promising alternative drug and multi-epitope vaccine targets prioritization against infectious *Clostridium difficile*. PLoS ONE 19(1): e0293731. <https://doi.org/10.1371/journal.pone.0293731>

Editor: Sheikh Arslan Sehgal, The Islamia University of Bahawalpur Pakistan, PAKISTAN

Received: August 1, 2023

Accepted: October 18, 2023

Published: January 19, 2024

Copyright: © 2024 Aiman et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Prevention of *Clostridium difficile* infection is challenging worldwide owing to its high morbidity and mortality rates. *C. difficile* is currently being classified as an urgent threat by the CDC. Devising a new therapeutic strategy become indispensable against *C. difficile* infection due to its high rates of reinfection and increasing antimicrobial resistance. The current study is based on core proteome data of *C. difficile* to identify promising vaccine and drug candidates. Immunoinformatics and vaccinomics approaches were employed to construct multi-epitope-based chimeric vaccine constructs from top-ranked T- and B-cell epitopes. The efficacy of the designed vaccine was assessed by immunological analysis, immune receptor binding potential and immune simulation analyses. Additionally, subtractive proteomics and druggability analyses prioritized several promising and alternative drug targets against *C. difficile*. These include FMN-dependent nitroreductase which was prioritized for pharmacophore-based virtual screening of druggable molecule databases to predict potent inhibitors. A MolPort-001-785-965 druggable molecule was found to exhibit significant binding affinity with the conserved residues of FMN-dependent nitroreductase. The experimental validation of the therapeutic targets prioritized in the current study may worthy to identify new strategies to combat the drug-resistant *C. difficile* infection.

1.0. Introduction

Clostridium difficile is a gram-negative anaerobic bacterium causes pseudomembranous colitis and life-threatening diarrhea [1, 2]. *C. difficile* is associated with hospital-acquired infections with high rates of morbidity and mortality worldwide, leading to epidemic in certain parts of the world [3]. *C. difficile* infections (CDIs) mainly develop in elderly individuals because of

disruption of the protective colon microbiota after antimicrobial therapy [4, 5]. Multiple factors contribute to the increasing number and severity of CDIs. The main risk factors associated with CDIs include prior antibiotic treatment, which leads to disrupted intestinal flora [6, 7]. Hospitalized patients are at maximum risk after the first month of antibiotic therapy [8]. Other risk factors include proton-pump inhibitors (PPIs), chemotherapy [9], and lack of response to toxins, mostly in older individuals. Individuals colonized with *C. difficile* spores are likely to serve as infection reservoirs that contaminate the environment with *C. difficile* spores [8]. Bacteria with a pathogenic locus (Paloc) expressing a binary toxin (toxin A = enterotoxin, toxin B = cytotoxin) are toxigenic strains, that cause colonic mucosal damage, resulting in actin filament degradation [10, 11]. Mutations in the toxin repressor gene (*tcdC*) have been reported to produce 15 times more toxin A and B hypervirulent strains i.e. BI/NAP027 and ribotype 027, that exhibit resistance to fluoroquinolones [12–15]. Multi-component biofilm development is also a substantial contributor to *C. difficile* drug resistance. The multi-layered biofilm is a dense matrix composed of DNAs, proteins, and polysaccharides [16].

The frequent occurrence of CDIs has increased the rate of emergence of multiple antibiotic-resistant (MAR) and hypervirulent strains during the last two decades [17, 18]. The rapid increase in the MAR bacterial strains is a major global health concern in the 21st century. Currently, it is estimated that over 700,000 annual deaths are caused by MAR pathogens worldwide, and death toll predicted to reach up to 10 million by 2050 [19]. *C. difficile* is classified as an urgent threat according to 2019 Antibiotic Resistance Report of the Centers for Disease Control and Prevention (CDC) [19, 20]. *C. difficile* resistant is reported to commonly prescribed antibiotics, including erythromycin, penicillin, tetracycline, clindamycin, lincomycin, cephalosporins, aminoglycosides, clindamycin, gentamicin, imipenem, moxifloxacin, rifampicin, and fluoroquinolones [21, 22]. Multiple studies have reported an increase of up to 90% in the antibiotic resistance of *C. difficile* in the past decade [23]. Worldwide surveillance reported that the irrational use of antibiotics majorly contributes to the increase in *C. difficile* resistance which led to the occurrence and reoccurrence of CDIs. The high rate of CDI recurrence is a serious problem; approximately 25% of treated individuals develop recurrence, and approximately 40% to 60% of individuals experience subsequent recurrence after their first recurrence of CDI [24, 25]. Antibiotic resistance of *C. difficile* results in suboptimal clinical outcomes, leading to CDI treatment failure. Current preventive measures are insufficient to treat CDIs, and there is no commercially available vaccine that provides protection against CDIs. Therefore, novel therapeutic strategies are required to combat CDIs, and prevent its subsequent recurrence.

Microbial pan-genomics has sparked the interest of the scientific community owing to the availability of whole genome sequence data of pathogenic strains [26–28]. The core proteome data of pathogenic strains provide primary information to identify novel therapeutic targets in pathogenic species. Combined bioinformatics, chemo-informatics, and immunoinformatic techniques have rapidly increased the identification of potential drug and vaccine targets against pathogenic strains [29–31]. Advancements in genomics and proteomics have facilitated the identification of suitable targets for functional medicines. Subtractive proteomics and vaccinomics is a novel approach that facilitates the identification of potential therapeutic targets against specific pathogenic strains [32]. The purpose of this work was to use comparative proteomic techniques to identify therapeutic drug and vaccine candidate proteins by focusing on the core proteome of *C. difficile* drug-resistant strains. Reverse vaccinology and immunoinformatic strategies were implemented for the prediction of immunodominant epitopes from the prioritized vaccine candidates to design a highly immunogenic vaccine construct. The efficacy of proposed vaccine construct was evaluated via immune-simulation analysis. Subtractive

proteomics approach was utilized to predict potential drug target proteins. Further, pharmacophore-based modeling, virtual screening, molecular docking, and molecular dynamic simulation analyses were performed to identify novel inhibitors against the top-prioritized drug targets. *In silico* ADME studies were done to investigate the pharmacokinetic properties of the lead inhibitory compounds.

2.0. Methodology

All the steps and methodological layout of the current study is summarized in Fig 1.

2.1 Core proteome retrieval

The core proteome of *C. difficile* (NCBI Taxonomy ID: 1496) was retrieved from EDGAR version 3 web resource (<http://edgar3.computational.bio>) [33]. The genome of the *C. difficile* strain 630 was selected as a reference genome. Paralogous sequences were determined using the standalone version of CD-HIT resource [34] with a similarity index of 60% [35]. Redundant sequences were discarded to acquire non-paralogous sequences for further analysis. Besides, sequences with less than 100 amino acid residues were also discarded. The PSORTb v3.0.2 web tool (<http://www.psort.org/psortb>) was used for determining the subcellular localization of non-paralogous pathogen proteins [36].

2.2 Reverse vaccinology

2.2.1 Vaccine candidates' identification. Outer membrane and extracellular proteins were further prioritized based on their allergenicity, antigenicity, and toxicity parameters. The allergenicity assessment of human non-homologous pathogen proteins was performed by AllerTOP v2.0 webserver (<http://www.ddg-pharmfac.net/AllerTOP>) [37]. VaxiJen v2.0 webserver (<http://www.jenner.ac.uk/VaxiJen>) was used to evaluate the antigenicity level of the prioritized bacterial proteins with a threshold of >0.4 [38]. ToxinPred2 webserver (<https://webs.iitd.edu.in/raghava/toxinpred2/>) was utilized to determine the toxicity of the prioritized bacterial proteins [39].

2.2.2 Assessment of B-cell, MHC-I, and MHC-II epitopes. The prioritized bacterial proteins were further subjected to multiple immunoinformatics tools to predict the potential B- and T-cell epitopes that can induce strong immune responses in the host immune system. T-cell epitopes are represented by major histocompatibility complex (MHC) molecules. MHC class I epitopes are recognized by CD8+ and MHC class-II epitopes are recognized by CD4+ T cells [40]. The Immune Epitope Database (IEDB) analysis resource was used for the prediction of T-cell epitopes in the prioritized pathogen proteins [41]. Stabilized Matrix Method (SMM) prediction approach was utilized to identify 9-mer MHC-I epitopes [42] and the NetMHCIIpanEL 4.1 prediction algorithm was used to predict 15-mer MHC-II epitopes [43]. We used the entire collection of human HLA reference set. Top binding overlapping epitopes were prioritized with an IC₅₀ value of <200 nM [35, 44]. ABCPred webserver (<http://codes.bio/abcpred/>) was used to predict B-cell epitopes with a cut-off value of >0.5, and other default parameters. The server detects linear B-cell epitopes that induce humoral immune responses and stimulate B lymphocytes [45]. The effective synthesis of a vaccine construct is affected by the variations in the distribution and expression of human HLA alleles among different ethnic groups and various regions of the world [46]. The population coverage of the prioritized MHC-I and MHC-II epitopes was determined by the IEDB population coverage tool. The tool calculates population coverage based on the distribution of HLA-binding alleles of each prioritized epitope in multiple global regions [47]. The IEDB epitope conservancy tool was

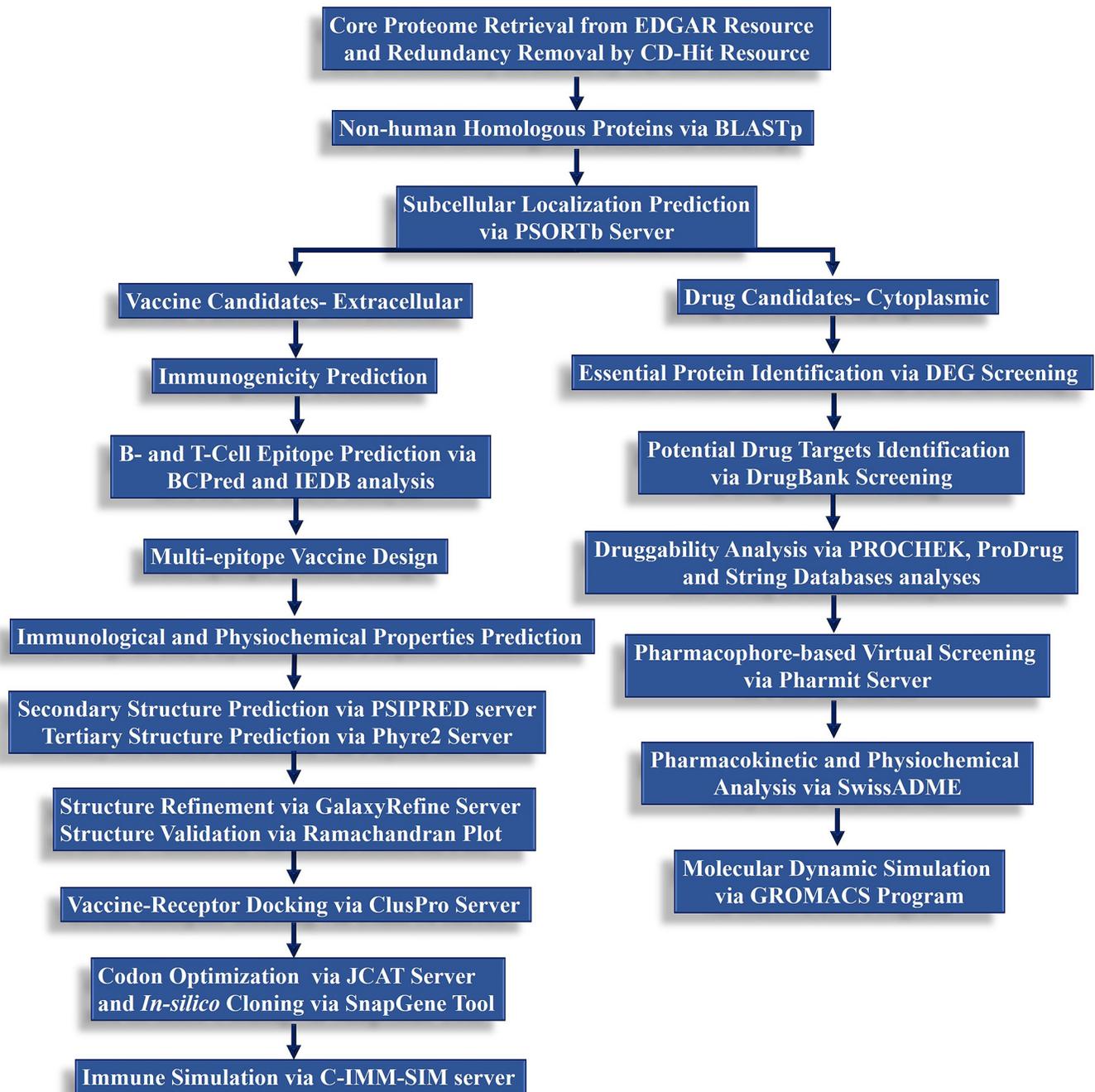


Fig 1. The methodological layout to prioritize drug and vaccine targets in the core proteome of *C. difficile* via comparative and subtractive proteogenomic analysis, immunoinformatics approaches, and druggability analyses.

<https://doi.org/10.1371/journal.pone.0293731.g001>

employed to assess the conservation of the prioritized MHC-I and MHC-II epitopes among various *C. difficile* strains.

2.2.3 Multi-epitope-based vaccine engineering. Overlapping B- and T-cell epitopes were prioritized and further examined for their allergenicity, antigenicity, and toxicity parameters. The overlapping epitopes were also screened against human epitopes to avoid autoimmune responses. The non-homology threshold was set at an E-value >0.05 [48]. The top-lead

overlapping epitopes were used to design a multi-epitope chimeric vaccine, using different adjuvant and linker peptide sequences, to enhance the immunogenicity of the designed vaccine construct [49]. The Pan DR Helper T Epitope (PADRE) sequence was also incorporated in the vaccine construct to avoid polymorphisms of HLA-DR molecules in case of different populations [50]. The “HEYGAEALERAG” and “GGGS” linkers were used to conjugate the prioritized overlapping epitopes [42]. The prioritized overlapping B- and T-cell epitopes in the designed vaccine model ensure the generation of both cell-mediated and humoral immune responses against the antigenic peptide vaccine construct [44]. EAAAK linker was used to connect the adjuvant i.e., β -defensin at the N-terminus of the multi-epitope chimeric vaccine construct. A strong immunostimulatory adjuvant enhances the immunogenicity of the designed vaccine construct and activates long-term innate and adaptive immunity against pathogens [51].

2.2.4 Immunological and physicochemical properties assessment. The designed multi-epitope vaccine construct with various combinations of epitopes was subjected to immunogenic analysis. The AlgPred webserver (<https://webs.iitd.edu.in/raghava/algpred2/>) was used to assess the allergenic behavior of the designed vaccine construct [47]. The VaxiJen v2.0 [38] and ANTIGENpro (<http://scratch.proteomics.ics.uci.edu>) [52] webserver were used for antigenicity assessment of the vaccine construct. These servers evaluate the antigenicity based on the principal amino acid properties using auto cross covariance (ACC) transformation [38] and 10 fold cross validation of the peptide sequence against known protein data files [53]. Likewise, the solubility of the vaccine construct was calculated by SOLpro webserver [54]. The ProtParam program on the ExPASy server (<https://web.expasy.org/protparam/>) was used to determine the various physicochemical parameters of the vaccine construct [55].

2.2.5 Secondary and tertiary structures prediction, refinement, and validation. Secondary structure prediction of the designed vaccine construct was performed using PRISPRED 4.0 (<http://globin.bio.warwick.ac.uk/psipred/>) [56] and SOPMA (<http://www.ibcp.fr/predict.html>) [57] servers. The servers use position-specific scoring matrices to predict the transmembrane topology, transmembrane helix, and recognition of folds and domains in the protein sequence. The tertiary structures of the multi-epitope vaccine constructs were generated by Phyre2 web tool (<http://www.sbg.bio.ic.ac.uk/phyre2>) [58] and further refined by the GalaxyRefine web program (<http://galaxy.seoklab.org/refine>) [59]. Subsequently, the refined tertiary structure was validated by the ERRAT tool, Ramachandran plot of PROCHECK suite [60] (<https://saves.mbi.ucla.edu/>), and ProSA-Web [61] server (<https://prosa.services.came.sbg.ac.at>).

2.2.6 Vaccine-receptor docking analysis. ClusPro is a protein-protein docking server (<https://cluspro.org>) that was used to study molecular interactions between the designed vaccine construct with human HLA and TLR receptors to determine receptor-vaccine interactions [62]. The refined vaccine constructs were docked against HLA-A*11-01 (PDB ID: 5WJL), MHC-II allele HLA DRB1*04-01 (PDB ID: 5JLZ), TLR2 (PDB ID: 6NIG), and TLR4 (PDB ID: 3FXI) receptor molecules. The vaccine-receptor docked complex with the lowest docking score was prioritized for immunological validation. The molecular interactions between the residues in the vaccine-receptor complex were determined using the PDBsum programme (<https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>) [63].

2.2.7 In Silico immune simulation. The C-ImmSim web program (<http://www.cbs.dtu.dk/services/C-ImmSim-10.1/>) was used for the computational immune simulations of the multi-epitope vaccine to assess the immunogenic potential of the designed vaccine [64]. The server employs multiple machine learning methods to predict the potential stimuli of the host immune system and provides information on humoral and cellular responses to the antigen [65]. The standard clinical protocol calls for a four-week interval between two vaccine doses

[66]. We followed the protocol previously used by Aslam et al., 2021 to carry out the immune simulation of the designed chimeric vaccine construct [44]. The simulation was run for 1, 84, and 168 hours with the default settings. The immune system was simulated for a thousand times using the HLA-A*0101 and A*0201, HLA-B*0702 and B*3901, HLA-DRB1*0101, and DRB1*0401 antigens.

2.2.8 *In Silico* codon optimization and *in-silico* restriction cloning. The prioritized multi-epitope vaccine construct was subjected to codon optimization and *in-silico* restriction cloning. Java Codon Adaptation Tool (JCAT) (<http://www.prodoric.de/JCat>) server was utilized for reverse translation of the peptide sequence to cDNA. The codons were optimized to achieve the maximum expression of vaccine in the bacterial expression system [67]. The maximum possible expression potential of the cloned vaccine gene was calculated by the codon adaptation index (CAI) and percentage GC content. The optimal CAI value reported for favorable transcriptional and translational efficacy is 0.8–1, while the optimum GC content is 30%–70% [42, 68]. The optimized prioritized vaccine gene was computationally cloned in the *E. coli* vector using the Snapgene (<https://www.snapgene.com/>) software. The Addgene server (<https://www.addgene.org/>) was used to retrieve the *E. coli* plasmid pET28a_TIAL1 for computational restriction cloning [69].

2.3 Subtractive proteomics and druggability analysis. The human proteome (Taxonomy ID: 9606) and the human-gut proteome (retrieved from TiD database [70]) were searched for homologs of the non-paralogous pathogenic protein sequences using the standalone BLASTp tool [71], with an E-value cut off of 10^{-4} , bit score ≤ 100 , query coverage $\leq 35\%$, and sequence identity $\leq 35\%$. In-house python-based tool was used to filter out homolog protein sequences based on set parameters. The pathogen-specific protein sequences were further screened against DEG (Database of Essential Genes) [72] via BLASTp to identify the pathogen essential proteins involved in different metabolic pathways. Human non-homologous pathogen essential proteins in the cytoplasmic regions were subjected to druggability analysis to identify potential drug targets and further used to identify potent inhibitors against *C. difficile* infections. The DrugBank database was screened to identify novel drug target in the shortlisted proteins via BLASTp parameters set at an E-value cut off of 10^{-4} , Bit score < 100 , query coverage $\leq 35\%$, and sequence identity $\leq 35\%$. At the threshold values, the non-hit proteins were selected as potential drug targets. Protein Data Bank (PDB) database [73] was curated to determine experimentally validated 3D-structures of the shortlisted *C. difficile* proteins. The 3D-structures of the shortlisted *C. difficile* proteins were predicted by the SWISS-MODEL web tool [74] with $< 90\%$ sequence homology with the PDB database entries. 3D-structural validation was carried out using the PROCHECK suite of programs [60]. Putative active sites in these protein structures were determined using the PockDrug server [75]. STRING database v10.5 [76] was used to determine the protein-protein interaction (PPI) and hub proteins identification. Proteins with a high average node degree ($K \geq 5$) were considered as hub proteins. The hub proteins are of critical importance for the survivability of pathogens.

2.3.1 Pharmacophore-based virtual screening. The Pharmit webserver [77] was utilized for designing pharmacophore model, using the crystal structure of putative nitroreductase in complex with fmn (Flavin Mononucleotide) (cd3205) acquired from *C. difficile* 630 at 1.35 Å resolution (PDB: 3GFA). A pharmacophore is a spatial arrangement of essential features of an interaction within a molecular structure. The server is based on state-of-the-art sub-linear algorithm to screen millions of compounds in a large number of compound databases i.e., MolPort, ZINC, PubChem, and ChEMBL. Pharmit uses AutoDock Vina scoring function to evaluate pharmacophore models, molecular structures, and energy minimization [77, 78]. The pharmacophore model was constructed based on 8 pharmacophore features (2 hydrogen donors, 5 hydrogen acceptors, and 1 hydrophobic). The resultant hit compounds list was

further minimized to obtain the top most significant molecules among the millions of compounds available in the Pharmit repository.

2.3.2 Molecular docking and ADME profiling. The drug-like behavior of the top 10 hit compounds was evaluated by Lipinski's rule of five [79]. Molecular docking analysis determines the binding orientations and affinities of the drug-like compound with the receptor protein [80]. The top 10 hit drug-like compounds were subjected to molecular docking analysis with the target protein via CB-Dock server [81]. CB-Dock improves the accuracy of the predicted binding site of the target protein and binding poses of the query ligand using a novel curvature-based cavity detection approach (CurPocket) and performs docking via the Auto-Dock Vina program [81]. Protein-ligand interaction analysis and visualization were performed by Discovery Studio Visualizer v.4.5 Client version (Accelrys, San Diego, CA, USA). The ADME (Absorption, Distribution, Metabolism, and Excretion), physicochemical, and pharmacokinetic properties were examined using SwissADME web resource [82].

2.3.3 Molecular dynamic simulation. The highly effective ligand prioritized after ADME profiling was subjected to molecular dynamic (MD) simulations. MD simulation was performed to evaluate the stability, flexibility, hydrogen bond interaction, and inhibitory potential of small drug-like compounds. The GROMACS 2019.2 software was used for MD simulation [83]. The parameters used for simulations were GROMOS96 43a1 force-field, solvation with TIP4P water model, cubic boundary box type calculated through the buffer and NVT and NPT ensemble for 100 ns trajectory at 1 bar pressure and 300K temperature [50]. Finally, the trajectory analyses were calculated for root-mean-square distance (RMSD), root-mean-square fluctuation (RMSF), radius of gyration (Rg), and hydrogen bonds to study the interaction pattern of ligand and dynamic fluctuation in protein [84].

3.0 Results

3.1 Core proteome retrieval

The core proteome of *C. difficile* retrieved from EDGAR yielded a total of 42,700 protein sequences. The removal of redundant or paralogous sequences resulted in 2,113 non-paralogous pathogen core proteins (S1 File).

3.2 Subcellular localization

PSORTb categorized the non-paralogous pathogen proteins into their corresponding regions, including the cytoplasm, periplasm, intermembrane space, outer membrane, and extracellular spaces (Table 1). Cytoplasmic proteins were regarded as the best drug candidates, whereas surface proteins were prioritized for model vaccine constructs prioritization.

3.3 Reverse vaccinology

3.3.1 Vaccines candidates identification. Core proteins located on the outer membrane and extracellular regions of cells were prioritized for the vaccinomics study. The antigenicity

Table 1. Subcellular localization of *C. difficile*-specific non-redundant proteins.

S.No	Cellular localization	Number of Proteins
1.	Cytoplasmic	1226
2.	Cytoplasmic Membrane	502
3.	Extracellular	17
4.	Cell-wall	27
5.	Unknown	341

<https://doi.org/10.1371/journal.pone.0293731.t001>

Table 2. Vaccine candidate proteins finalized based on antigenicity, allergenicity, and toxicity parameters.

Protein IDs	Protein Names	Allergenicity	Antigenicity Vaxijen v2.0 (Threshold >0.4)	Toxicity
CD630_18220	thiol peroxidase	Non-Allergen	0.5455	Non-Toxin
CD630_27870	cell surface protein	Non-Allergen	0.4669	Non-Toxin
CD630_16310	superoxide dismutase	Non-Allergen	0.4338	Non-Toxin
CD630_10170	multidrug family ABC transporter ATP-binding protein/permease	Non-Allergen	0.4630	Non-Toxin

<https://doi.org/10.1371/journal.pone.0293731.t002>

of the prioritized proteins was predicted by the Vaxijen v2.0 webserver with 79–89% accuracy [38]. Four proteins (CD630_18220, CD630_27870, CD630_16310, and CD630_10170) were prioritized based on immunological parameters (Table 2). These proteins were subjected to B- and T-cell epitopes prediction for *in silico* chimeric vaccine designing. Vaccine candidate proteins were selected based on antigenicity, allergenicity, and toxicity parameters.

3.3.2 MHC-I, MHC-II, and B-cell epitopes prediction. The B- and T-cell epitopes were predicted from vaccine candidate proteins to design potential multi-epitope vaccine constructs against *C. difficile*. The lead MHC-I and MHC-II epitopes for the prioritized proteins were selected based on an IC₅₀ values <200 nM. BCpred scores greater than 0.8 and 75% specificity were used to predict overlapping B-cell epitopes. Overlapping lead epitopes were selected from each prioritized vaccine candidate protein based on non-allergenic, non-toxic, and high antigenic nature. Overall, 12 such epitopes were prioritized to design a highly immunogenic vaccine construct (S1 and S2 Tables). The main objective was to determine key overlapping epitopes capable of eliciting cell-mediated and humoral immune responses. Overlapping epitopes were scanned against human peptides to eliminate homologous epitopes based on the E-values. Conservation of the prioritized epitopes was ensured to be 100% in multiple *C. difficile* strains. The conserved epitopes are assumed to provide broader protection against various *C. difficile* strains [85]. The IEDB results revealed that the prioritized epitopes showed >97% coverage of the worldwide populations; however, the population coverage of some epitopes was high in various regions of the world (S1 Fig and S3 Table).

3.3.3 Multi-epitope chimeric vaccine designing. A multi-epitope-based chimeric vaccine was designed using the prioritized human non-homologous overlapping immunodominant epitopes. The GGGs and HEYGAEALERAG linkers were used to join the selected epitopes. These linkers are responsible for the structural stability to the vaccine construct and allow each epitope to independently perform its protective role in human body after the vaccine administration [86]. The epitopes were linked to beta-defensin peptide sequence at the N-terminus using EAAAK linkers to enhance immunogenic responses. The consequences of HLA-DR variation in various groups have been countered by incorporating PADRE peptide sequences into the vaccine design. It has been previously reported that PADRE peptide-containing vaccine models provide enhanced immunity and robust cytotoxic T lymphocyte (CTL) responses against antigens [87] (Fig 2A).

3.3.4 Immunogenicity and physicochemical properties assessment. The immunological properties determined a highly antigenic, non-allergenic, and non-toxic nature of the designed vaccine construct. Antigenicity scores of 0.882658 and 1.0433 calculated using ANTIGENpro and VaxiJen 2.0, respectively, indicate the substantial antigenic nature of the designed chimeric vaccine construct. An AlgPed score of -0.47191405 indicated that the designed chimeric vaccine construct exhibited non-allergenic behavior. The solubility score of 0.619589 predicts high solubility of the chimeric vaccine upon expression. The ProtParam web tool was employed to predict various physicochemical properties of the chimeric vaccine. The optimal molecular weight of the vaccine construct was calculated to be ~43 kDa which indicates easy purification of the vaccine. The grand average of hydropathicity (GRAVY) value was -0.376

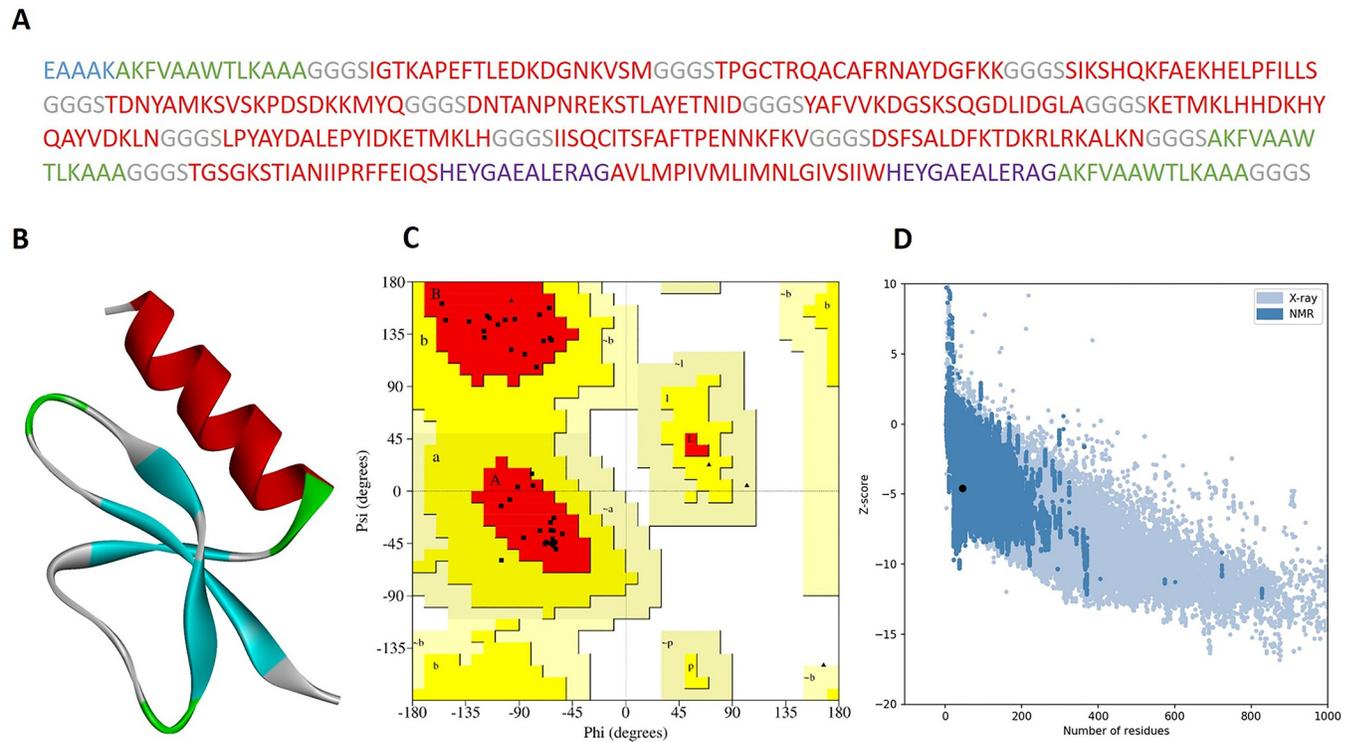


Fig 2. Vaccine design, 3D-structure prediction, and validation. (A) Multi-epitope chimeric vaccine construct. (B) Refined 3D-structure of the vaccine construct. (C) Ramachandran plot with 94.7% of residues in the favored region of the plot. (D) ProSA-web plot representing the z-score (-4.57) of the refined vaccine construct.

<https://doi.org/10.1371/journal.pone.0293731.g002>

and aliphatic index was 66.56, indicating hydrophilic nature of the chimeric vaccine construct. The Instability Index score of 32.70 predicts the stability of the construct. Immunological and physicochemical properties of the construct indicate the capabilities of the designed vaccine construct to elicit significant immunogenic responses.

3.3.5 Secondary and tertiary structures prediction, refinement, and validation. The secondary structure elements of the chimeric vaccine depicted 36.94% α -helices, 12.50% β -sheets, 29.17% coils, and 21.39% extended strands (S2 Fig). The presence of α -helical coil-coiled domains in the multi-epitope chimeric vaccine construct is critical in order to enable proper protein folding based on standard protein structures, and confers an effective humoral immunity in response to a specific pathogen [88]. The tertiary structures of the chimeric vaccine constructs were predicted by Phyre2 web tool, refined using the GalaxyRefine server (Fig 2B), and further validated by Ramachandran plot. The ERRAT quality score was calculated 100%, Ramachandran plot revealed that 94.7% of the residues appeared in the core region of the plot (Fig 2C), and the z-score of -4.57 of the refined vaccine constructs determined by ProSA-web plot represented a high-quality 3D structure of the vaccine construct (Fig 2D). The high structure validation scores of the proposed vaccine indicated a prediction with a topology of greater accuracy.

3.3.6 Vaccine-receptor docking analysis. Molecular docking analysis was carried out to evaluate the binding promiscuity of the designed vaccine with human HLA and TLR immune receptors. ClusPro webserver was used for rigid body protein-protein docking based on billions of confirmations sampling, RMSD-based clustering, and structural refinement based on energy minimization. The docking analysis revealed that the vaccine construct exhibited

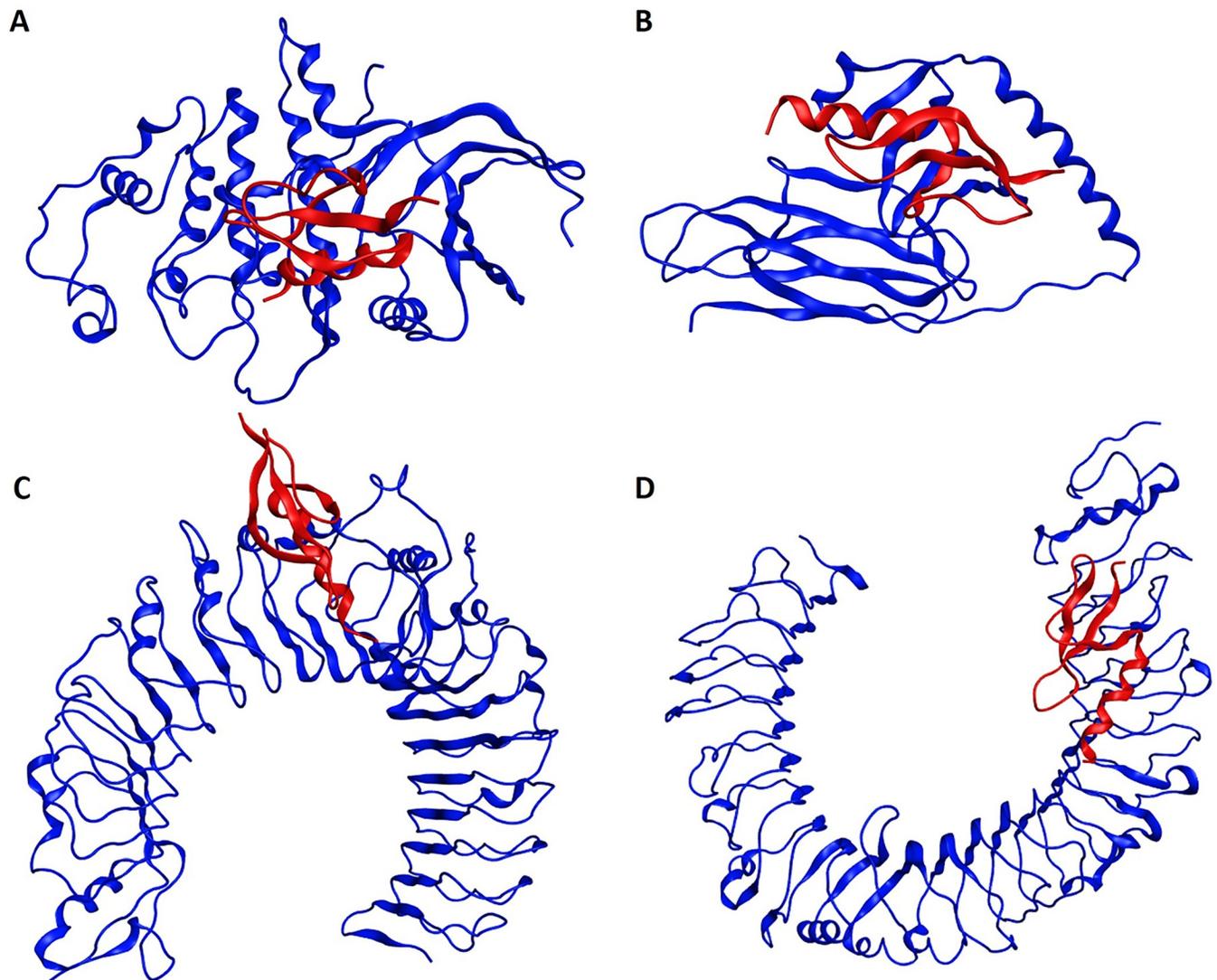


Fig 3. Molecular docking of the multi-epitope chimeric vaccine construct with the human HLA and TLR immune receptors. The blue color represents receptor molecule and the red color represents vaccine construct. (A) Vaccine construct and HLA (PDB ID: 5WJL) docking complex with a global energy of -780.2 kcal/mol. (B) Vaccine construct and HLA (PDB ID: 5JLZ) docking complex with a global energy of -1029.4 kcal/mol. (C) Vaccine construct and TLR2 (PDB ID: 6NIG) docking complex with a global energy of -949.6 kcal/mol. (D) Vaccine construct and TLR4 (PDB ID: 3FXI) docking complex with a global energy of -751.9 kcal/mol.

<https://doi.org/10.1371/journal.pone.0293731.g003>

lowest binding energy of -780.2 kcal/mol with HLA-A*11-01 (PDB ID: 5WJL) and -1029.4 kcal/mol with MHC-II allele HLA DRB1*04-01 (PDB ID: 5JLZ), -949.6 kcal/mol with TLR2 (PDB ID: 6NIG), and -751.9 kcal/mol with TLR4 (PDB ID: 3FXI) (Fig 3). The docking results demonstrated that the vaccine design was capable of developing several molecular interactions with HLA and TLR receptors binding sites. The lowest docking energy scores showed the highest binding affinity of vaccine construct with human immune receptors. These findings speculate the capability of vaccine to trigger potent immune responses when administered to a human host. The molecular interactions between the residues of the vaccine-MHC-I complex demonstrated 12 hydrogen bonds, 5 salt bridges, and 146 non-bonded interactions. Vaccine-MHC-II complex showed 7 hydrogen bonds, 3 salt bridges, and 148 non-bonded interactions. The interactions between vaccine-TLR2 complex revealed 7 hydrogen bonds, 3 salt bridges,

and 148 non-bonded contacts, whereas 9 hydrogen bonds, 1 salt bridge, and 117 non-bonded interactions were observed between the residues of vaccine-TLR4 complex (S4 Table and S3 Fig).

3.3.7 Computational immune simulation. The immune simulation results predicted a substantial increase in the primary and secondary responses caused by the prioritized multi-epitope chimeric vaccine construct. Theoretically, this pattern is consistent with the generation of a real-time immune response. The primary simulated responses showed elevated IgM levels. The simulation of secondary and tertiary immune responses showed momentous escalation of B-cells population and increase in the levels of IgG1, IgG2, IgM, and IgM+IgG antibodies were observed, while antigen levels decreased (Fig 4A and 4B). These results suggest immunological memory generation, as evidenced by an increase in memory B-cell population and isotype switching. As a result, repeated exposures to chimeric antigen led to a precipitous drop in antigen levels (Fig 4C). After repeated antigen exposure, the number of cytotoxic (TC) and helper (TH) cells increased, and associated memory responses were predicted to initiate (Fig 4D and 4E). Additionally, macrophage, dendritic cell, and natural killer cell populations were

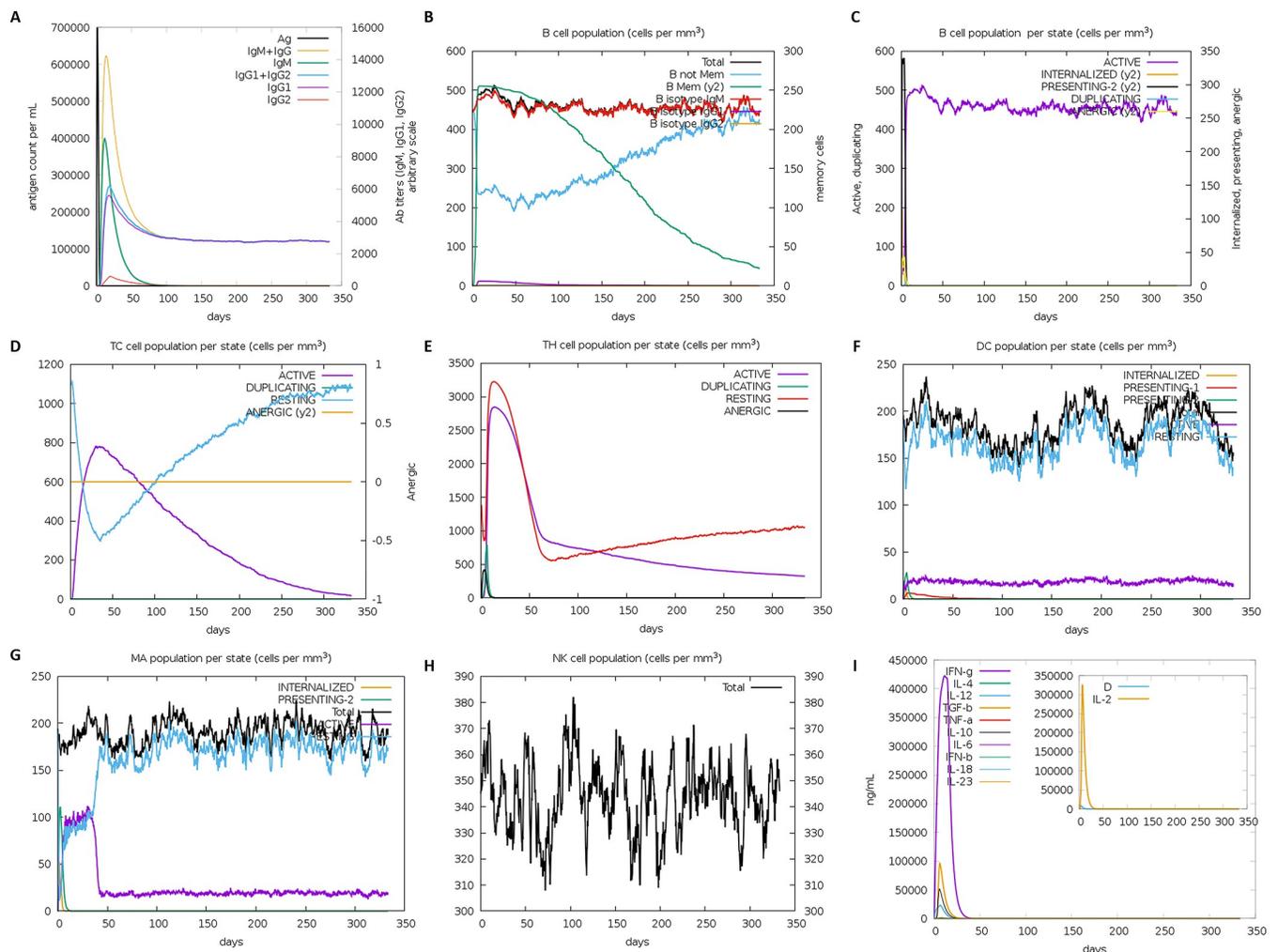


Fig 4. In-silico immune simulation of the chimeric vaccine peptide determined by C-ImmSim server. (A, B) A significant rise of the B-cell populations and high levels of immunoglobulin antibodies, with a reduction in antigen levels. (C) The rising B-cell population with repeated antigen exposure. (D, E) The increase in helper and cytotoxic T-cell population with repeated antigen exposure. Dendritic cell, macrophage, and natural killer cell populations increase during immunization period. (I) Increased cytokine concentrations following repeated antigen exposure.

<https://doi.org/10.1371/journal.pone.0293731.g004>

stimulated and kept at high levels throughout the immunization period (Fig 4F–4H). Moreover, higher levels of an interleukin like IL-2 and a cytokine like IFN- γ were observed (Fig 4I). These results indicated that the proposed vaccine construct may elicit significant immune responses against the pathogen.

3.3.8 Codon optimization and *in-silico* cloning. The top-ranked vaccine construct was subjected to codon optimization and *in silico* restriction cloning to evaluate the expression potential of the proposed vaccine model. JCAT revealed that the GC content of the optimized cDNA sequence of the prioritized vaccine construct was predicted to be 67.96% with a significant CAI value of 0.95 (S5 Table and S4 Fig). These values are in the optimum range i.e. GC content 30%–70% and CAI index of 0.8–1.0 [67], suggesting that the vaccine construct had a high expression potential. The codon sequence of the proposed vaccine was successfully inserted into the pET28a_TIAL1 vector to ensure heterologous cloning and expression of the vaccine in the *E. coli* expression system (S5 Fig).

3.4 Subtractive proteomics and druggability analysis

The non-paralogous proteins were subjected to additional screening against proteins from the human proteome and the human gut microbiome, which resulted in 610 non-homologous hits (S2 File). The proteins homologous to human and human gut microbiome proteins were discarded as they might cause immuno-metabolic disorders in humans. Screening of DEG resource yielded 385 *C. difficile* core proteins, showing significant sequence similarity with the essential proteins in the database (S3 File). These human non-homologous and *C. difficile* essential proteins might be promising drug and vaccine targets against *C. difficile*-mediated infection. Based on the subcellular localization (Table 1), the cytoplasmic proteins were prioritized as suitable drug targets. Finally, 184 proteins were prioritized as suitable drug candidate proteins during downstream druggability analysis. These proteins showed no homology with the DrugBank database entries. The PDB database was manually screened for the availability of 3D structures of these proteins. Among the 184 proteins, the putative nitroreductase in complex with fmn (cd3205) structural information was available (PDB: 3GFA). The PPI analysis via STRING v10.5 database predicted the putative nitroreductase as a hub protein with the node degree (K) ≥ 5 (Fig 5), representing the high number of molecular interactions. Among the rest of the proteins targets the top 100 proteins were prioritized based on PockDrug probability scores > 0.5 and their 3D structures were designed using the SWISS-MODEL server [74]. The 3D-structure evaluation indicated the high quality of these proteins based on the ERRAT tool, QMEAN Z-scores [89], and RAMPAGE scores [60] (S6 Table). These top drug targets were thought to have promising druggable pockets that could serve as anchors for small drug-like compounds and play critical roles as hub proteins in *C. difficile*.

3.4.1 Pharmacophore modeling, virtual screening, and molecular docking. Among the shortlisted drug-target candidate proteins, putative nitroreductase (PDB: 3GFA) was prioritized for pharmacophore-based virtual screening to identify small drug-like compounds. The obtained pharmacophore model based on the 3D-structure of a putative nitroreductase from Pharmit server showed 8 pharmacophore features i.e., two hydrogen donors, five hydrogen acceptors and one hydrophobic feature as shown in Fig 6. The top ten screened molecules were prioritized based on lowest binding energy after being docked repeatedly (S7 Table). The scores and Root Mean Square Distance (RMSD) values of the selected compounds are shown in Table 3. FMN-dependent nitroreductase is the member of a structurally homologous family. The enzyme bounded FMN cofactor mediates the electron transfer from NAD(P)H to proceed the first step of the ping-pong Bi-Bi reaction pathway. The crystal structure of this enzyme shows the binding position of the FMN cofactor is highly conserved i.e., residues 10–

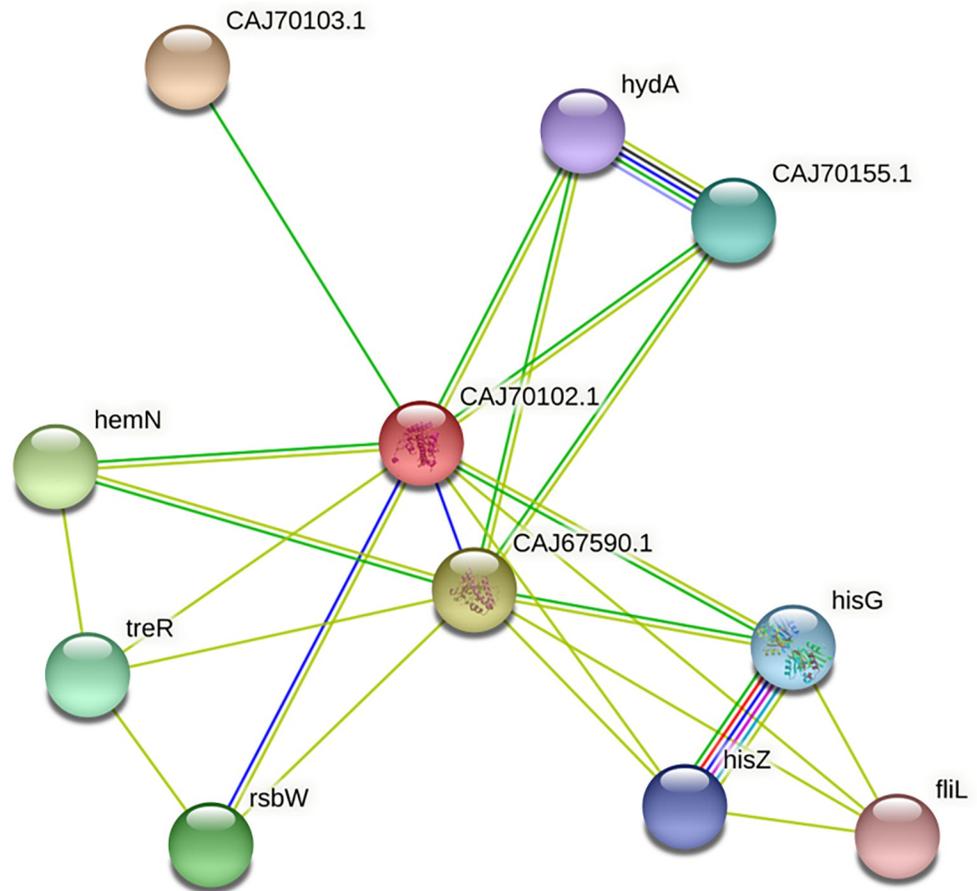


Fig 5. Protein-protein interaction of pathogen protein CD630_32050 (putative nitroreductase) with PDB ID: 3GFA acquired from STRING database. Red color indicated the hub protein.

<https://doi.org/10.1371/journal.pone.0293731.g005>

25 and 40–55. In the present investigation, this data was used in a pharmacophore-based computational screening procedure. The top hit compound i.e. MolPort-001-785-965 showed interactions with the residues in the conserved region, i.e. Asn17, Ser19, Thr50, Glu51, and Lys54 of the nitroreductase. Besides, the top hit compound was involved in interactions with the flexible region (residues 90–134) (Fig 6A), which is a part of cofactor and substrate binding site [90]. Moreover, the docking analysis showed that all the top 10 compounds exhibited interactions with the residue in the conserved region of the nitroreductase. The top 10 hit compounds were further subjected to molecular docking analysis to re-evaluate the binding conformation of the lead compounds within the active site of the receptor molecule (Fig 6B). These drug-like compounds might have a potential to combat the *C. difficile*-mediated infection and may be worthy of consideration in drug discovery and development [91].

3.4.2 Drug likeliness and ADME analysis. SwissADME provides detailed information on the physicochemical properties, ADME, and medicinal potential of a compound. Drug likeliness is an intricate balance between structural and molecular properties of a compound that shows a certain degree of similarity with a known drug molecule. The drug likeliness of a compound is calculated based on different molecular parameters, such as molecular weight, hydrophobicity, hydrogen bonding, reactivity, electron distribution, pharmacophore entity, bioavailability, molecular stability, and toxicity [92]. Lipinski's rule of 5 is the most commonly

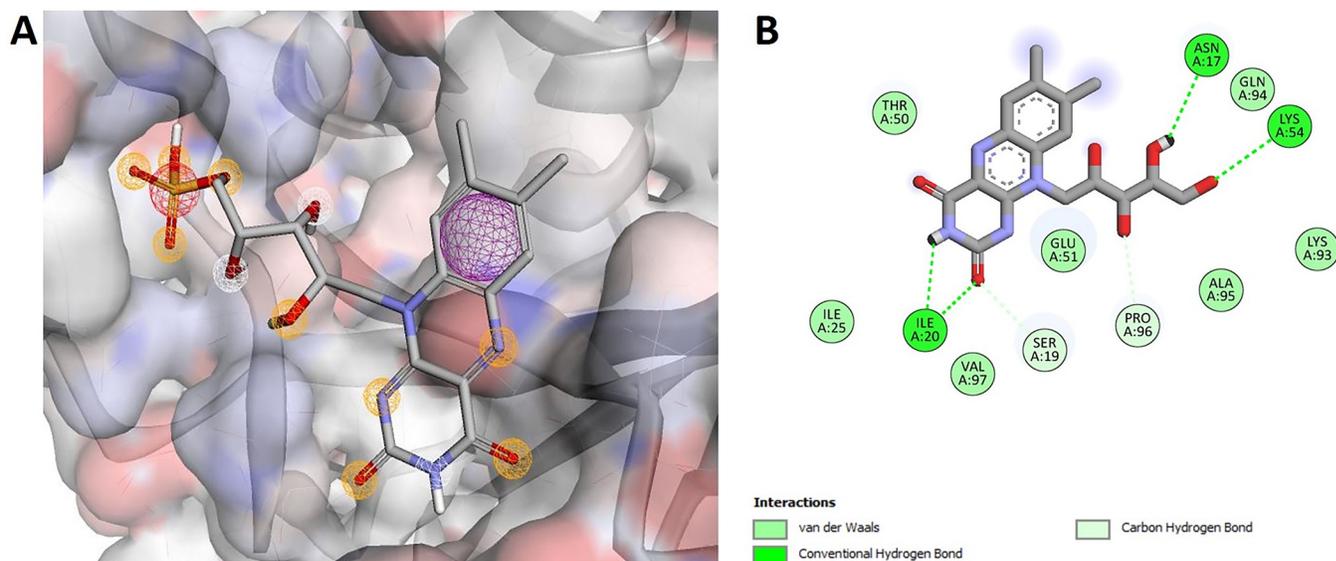


Fig 6. (A) The active site of the protein CD630_32050 was used to create a pharmacophore model. The characteristics are denoted by different colors. White represents a hydrogen-bond donor, yellow represents a hydrogen acceptor, green represents hydrophobic properties, and aromatic represents aromatic features (pink). (B) The molecular interactions of the top hit docked compound (CD630_32050) within the substrate-binding site. The nature of protein-ligand interactions is shown in different colors.

<https://doi.org/10.1371/journal.pone.0293731.g006>

used models to evaluate a therapeutic drug-like compound based on its solubility and permeability [79]. The four major pharmacokinetic parameters of a drug-like compound are absorption, distribution, metabolism, and excretion (ADME). The compound with the best molecular interaction and binding energy with the receptor protein may not be the most suitable drug. A reliable drug should be completely and rapidly absorbed in the gastrointestinal tract, directly distributed towards the respective target, metabolized in a way that does not affect its activity, and ultimately be eliminated without causing any harm [93]. The pharmacokinetic properties of a molecule can be calculated using chemical descriptors as there is a significant relationship between the chemical structure and physicochemical properties.

The pharmacokinetics and pharmacological properties of the top ten Pharmit hits were calculated to evaluate the efficacy of these compounds as drugs (S7 Table). The extremely hydrophobic compounds are poorly soluble in the gastrointestinal tract and solvate fat globules [94].

Table 3. Pharmit scores and RMSD values of the top 10 hit compounds obtained from pharmacophore-based virtual screening using the Pharmit server.

Compounds	(MolPort IDs)	Pharmit Score	RMSD Values
C1	MolPort-044-559-927	-8.57	1.1096
C2	MolPort-044-724-190	-8.57	1.065
C3	MolPort-003-939-021	-8.50	0.843
C4	MolPort-021-783-318	-8.00	0.803
C5	MolPort-039-136-733	-7.72	1.364
C6	MolPort-003-934-329	-7.36	0.807
C7	MolPort-001-785-965	-7.34	0.898
C8	MolPort-004-964-255	-6.92	1.135
C9	MolPort-003-666-643	-6.83	0.955
C10	MolPort-044-561-302	-6.68	0.941

<https://doi.org/10.1371/journal.pone.0293731.t003>

Based on physicochemical properties, all the compounds exhibited drug-like properties i.e., holding high solubility with the molecular weight of less than 500 Da. According to the ADME results, these compounds showed low gastrointestinal absorption and did not penetrate the blood-brain barrier (BBB). Compounds C3, C4, C5, and C8 are substrates for the p-glycoprotein, which is associated with pumping xenobiotics and detrimental components back into the gut lumen, hence pharmacokinetically decreasing the efficacy of being a potential drug [95]. ADME profiling ensured no mutagenicity or toxicity for all these drug-like compounds. Among these compounds, C1, C2, C3, C4, C5, C8, and C10 showed Lipinski violations, whereas C6, C7, and C9 did not show any Lipinski's rule violation, making them ideal drug-like molecules based on *C. difficile* nitroreductase inhibition. However, the rest of the lead compounds may require extra chemical transformation to have potential drug-like nature (S8 Table).

3.4.3 Molecular dynamic simulations. The MD simulation analyses were performed for the prioritized nitroreductase-C7 complex docked models to validate the stability of molecular interactions and flexibility of the complex [96]. The RMSD values putative nitroreductase-C7 complex rapidly increased from 0.9 nm in the starting and maintained the equilibrium at 0.6 nm between 25 to 45 ns. The system was observed to gain stability at 0.6 nm after 48 ns which remained constant throughout the remaining simulation time (Fig 7A). In this analysis, it was observed that the complex had some fluctuations in ligand up to 45 ns, while it was observed to be completely stable after 50 ns. According to the RMSF calculations, the complex fluctuated maximally up to 0.62 nm and did not exceed further because of the rigidity of the structure (Fig 7B). Most of the residues in the protein exhibited an RMSF value less than 0.4 nm, indicating that the ligand does not undergo any significant conformational change over the simulation time. Hydrogen bonding between a ligand and receptor is essential for stabilizing the ligand-protein complex. Fig 7C displays the total number of hydrogen bonds formed in the complex, calculated after 100 ns of simulation time. Analysis of the H-bonds revealed that the ligand in the complex had established up to 6 H-bonds. It can be concluded that the ligand bound effectively and tightly to the putative nitroreductase. The Rg enables the measurement of the compactness fluctuations of a ligand-protein complex [42]. The average Rg value of putative nitroreductase was calculated as approximately 1.7 most of the time during the simulation (Fig 7D). This result indicated that the value of Rg remained steady throughout the MD simulation, suggesting stable protein folding. Compared to the MD simulation results of the protein and the protein-ligand complex, the RMSF values indicated that the residues from 90–100 interacting with the C7 were showing lower and smoother fluctuations as compared to the corresponding part of the monomer protein (S6 Fig). The results are in good agreement with the binding modes of the complex structure, ensuring the stability of the complex.

4.0 Discussion

The emergence of antibiotic-resistant strains and the recurrence of CDIs represent an urgent threat to the clinical challenges in public health. The treatment options for moderate to severe cases are extremely limited. No commercial vaccine is available, and a limited number of drugs have been proven effective against CDIs. New therapeutic strategies are required to tackle multi-drug resistant strains of *C. difficile*. Active vaccination provides the best opportunity to prevent CDIs in high-risk individual. In the recent years, intense research led to the development of experimental vaccines against CDIs, which are currently in clinical trials. Recombinant peptide vaccines and toxoid-based vaccines have been proven promising in healthy adults; however, challenges associated with such vaccination strategy remained in elderly and immunocompromised individuals [97]. Moreover, the peptide sequence variability

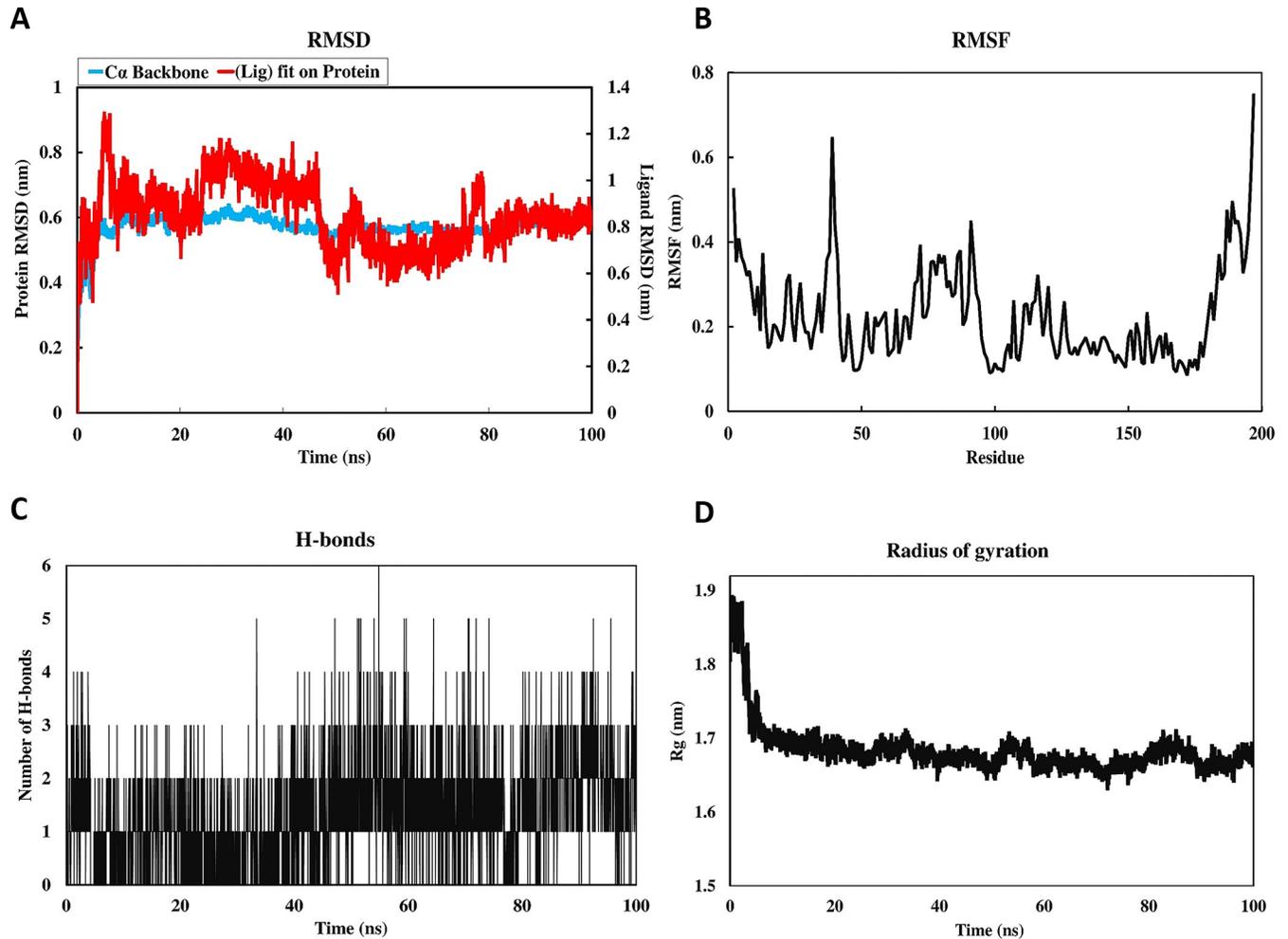


Fig 7. Molecular dynamics (MD) simulation results A) C7-putative nitroreductase RMSD analysis B) RMSF analysis of $C\alpha$ atoms C) H-bond estimation during 100 ns simulation D) Radius of gyration (Rg) analysis.

<https://doi.org/10.1371/journal.pone.0293731.g007>

from different *C. difficile* isolates raised the question of their ability to provide broader protection against CDIs. Therefore, more efforts are required to optimize the immunization strategies to enhance the efficacy of vaccines against CDIs. Core genome-mediated analysis is a promising approach for the identification of potential therapeutic targets in *C. difficile* as the core genome is conserved specie-wise. In this study, we utilized comparative genome analysis to identify pathogen-specific core drugs and vaccine target proteins. Furthermore, we used immunoinformatics and vaccinomics approaches to design a multi-epitope chimeric vaccine construct that can generate specific immunogenic responses against *C. difficile*.

We prioritized four proteins located in the outer membrane and extracellular regions as potent vaccine targets, based on high immunological properties. Highly conserved B-cell, MHC-I, and MHC-II epitopes were determined from the prioritized proteins to design a highly immunogenic vaccine construct against *C. difficile* following the reverse vaccinology techniques. Immune enhancers and adaptable adjuvant peptides were used to bind together overlapping B- and T-cell epitopes. To induce a targeted immune response against the pathogen while limiting negative responses towards the host, multi-epitope-based vaccine designing is a unique therapeutic method [98]. Immunological properties unveiled high antigenicity,

non-allergenic behavior, and non-toxic properties of the designed vaccine construct. The small molecular weight and hydrophilicity score of the vaccine construct determined highly solubility of the vaccine upon expression. Moreover, the stability score indicates that the expressed vaccine construct will be of high quality and ready for use. The tertiary structural analysis demonstrated a maximum number of residues in the favorable region of the Ramachandran plot, confirming the structural stability of the design vaccine construct. Immune simulation results validated the resemblance to actual cellular immune responses [29]. Strong immunogenic responses were triggered by repeated exposure to the same antigen. The clear development of memory B cells and T cells would last for several months, with the memory in the B cells. After repeated antigen exposure, the consistent high level of Ig production, T-cytotoxic, and T-helper cells indicated a humoral response. The molecular docking investigation further indicated that the design vaccine construct exhibited strong molecular interaction with human HLA and TLR immune receptors. These results indicate that the designed vaccine has the capability of inducing strong immunogenic responses against *C. difficile* infections and might be worthy of *in vitro* and *in vivo* investigations to confirm the findings of the current study.

In addition, we employed subtractive genome analysis, druggability analysis, and virtual screening approaches to identify novel inhibitors of *C. difficile* drug targets. Cytoplasmic proteins were prioritized as suitable drug candidates and were further scanned against the entries of DrugBank database for potential therapeutic drug targets identification. Non-hit proteins indicate that they have not yet been established as drug targets for *C. difficile*. According to the centrality-lethality law, inhibiting the activity of such proteins may be critical for the survival of pathogens [99, 100]. The prioritized druggable proteins were non-homologous to both human and human gut proteomes. In addition, objectives were set using rigorous threshold criteria for fundamental druggability characteristics. Drug designing for such specific targets will only affect the pathogen without disrupting the biological mechanism of the host. Among the prioritized druggable targets, the crystal structure of nitroreductase (i.e. CD630_07560) is available in the PDB database (PDB: 3GFA). Therefore, it was prioritized for the virtual screening of a drug-like molecule to identify novel drug-like compounds that inhibit the activity of this enzyme. FMN-dependent nitroreductases belong to a group of flavoenzymes, that are oxygen-insensitive and are essential for the reduction of nitro compounds in the presence of NAD(P)H [101]. Numerous bacterial species have been shown to contain nitroreductases. The physiological functions of these enzymes are unclear, although nitroreductases are believed to play an important role in the responses to a variety of antibiotics, specific oxidative stress conditions, and environmental chemical threats [102]. The expression of nitroreductase genes might be controlled by the MarRA and SoxRS regulatory systems. Studies have reported that nitroreductases have conserved domains for FMN binding, NAD(P)H electron transfer, and nitroaromatic-substrate interactions. Moreover, nitroreductases are consistent with the enzyme's ping-pong bi-bi catalytic mechanism, which explains the broader substrate specificity of these enzymes and establishes the groundwork for improving their activity for medicinal applications [103]. Nitroreductase is also involved in antibiotic resistance and bacterial pathogenesis [104]. For instance, nitroreductase has previously been reported in *Enterobacter cloacae* and *salmonella enterica* to be involved in resistance against chloramphenicol [104]; however, no such reports are available concerning *C. difficile*.

The top ten hit compounds in this study were riboflavin derivatives, among which C6, C7, and C9 followed all the Lipinski's rule of five. The results of virtual screening against the Mol-Prot database were refined based on RMSD and energy minimization. Based on the docking scores and ADME-profiling, these compounds showed effective hydrogen-bond interactions with the receptor in the conserved region. Riboflavin is a chemical analog of flavin

mononucleotide (FMN) that shows antimicrobial activity against gram-positive bacteria. Riboflavin directly binds to the peptides of the FMN riboswitch and suppresses FMN riboswitch-lacZ receptor gene expression in *Bacillus subtilis*. Targeting the FMN riboswitch to disrupt the expression of receptor genes with riboflavin results in antimicrobial activity against resistant bacteria and may lead to the development of novel antimicrobial drugs [105, 106]. These inhibitors showed strong binding interactions and safe drug profiles. Furthermore, the constant binding free energy along the MD trajectories implies the possibility of antibacterial medication [107]. However, additional *in vivo* and *in vitro* validation experiments should be conducted to determine their pharmacological efficacy, biocompatibility, and role as effective inhibitors.

Antibiotic resistance in bacteria has reached to an extremely dangerous level, and strains resistant to most of the commonly used antibiotics are now reported at alarming rates in many countries around the world [108]. Core genome-mediated analysis in combination with comparative and subtractive genome and/or proteome analysis identifies pathogen specific novel therapeutic drug and vaccine targets against such resistant pathogens. Reverse vaccinology approaches are novel strategies for the designing and developing broad-spectrum potential vaccines with better safety and increased potency. Druggability and virtual screening analyses serve as a starting point for the development of novel inhibitors, targeting pathogen-specific proteins. The *in silico* approaches used in this study could pave the way for identifying novel therapeutic targets among different pathogens and the development of specie-specific potent vaccines and drugs that aid in the elimination of diseases caused by multidrug-resistant pathogens. Experimental and clinical assays are required for further validation of the results acquired in this study.

5.0. Conclusions

In this study, the core genome of *C. difficile* was utilized to identify potential drug and vaccine targets against *C. difficile* infections. Subtractive proteomic analysis was performed to select pathogen-essential proteins that were non-homologous to the human and human gut proteomes. A multi-epitope vaccine was designed utilizing lead overlapping B- and T-cell epitopes from the prioritized four vaccine targets via reverse vaccinology. Highly immunogenic adjuvant and linker peptides were used to conjugate the prioritized epitopes in a vaccine construct that can generate strong host immunogenic response. The *in silico* restriction cloning analysis predicted effective expression potential of the proposed vaccine construct in an *E. coli* system. Furthermore, molecular docking and immune simulation analyses revealed that the vaccine has the capability to generate strong cell-mediated and humoral immune responses against *C. difficile* through significant binding affinities towards human immune receptors. Comparative genome analysis and multiple *in silico* druggability methods have prioritized several potential drug targets in *C. difficile* which have not been previously reported. FMN-dependent nitroreductases was prioritized for pharmacophore-based virtual screening to identify small drug-like compounds. The pharmacokinetic properties based on ADME profiling identified the top hit compounds. Docking studies and molecular dynamic simulation analysis identified potential binding affinities and stable interactions between drug-like compounds and putative nitroreductase which may lead to new strategies in drug development against multidrug-resistant *C. difficile* strains. Experimental and clinical assays pursuit are required to validate the finding of this study.

Supporting information

S1 File. *Clostridium difficile* non-paralogous proteins.
(DOCX)

S2 File. *Clostridium difficile* essential proteins non-homologous to the human host.
(DOCX)

S3 File. *Clostridium difficile* proteins homologous to DEG entries.
(DOCX)

S1 Fig. Population coverage of MHC-I and MHC-II alleles of the selected epitopes.
(TIF)

S2 Fig. Secondary structure prediction of vaccine construct.
(TIF)

S3 Fig. Molecular interactions between receptor proteins (Chain A) and vaccine constructs (Chain B).
(TIF)

S4 Fig. Codon optimization index graph of vaccine construct determined by JCAT server.
(TIF)

S5 Fig. *In silico* restriction cloning of the vaccine gene sequence (highlighted in red) into the *E. coli* pET28(+) expression vector.
(TIF)

S6 Fig. The individual RMSF graphs of protein (A) and protein-C7 complex (B).
(TIF)

S1 Table. Overlapping B-cell, MHC-I, and MHC-II epitopes prediction using IEDB analysis resource and ABCPred server.
(DOCX)

S2 Table. Immunogenicity analysis of non-human homologous overlapping epitopes.
(DOCX)

S3 Table. Prioritized MHC-I and MHC-II epitopes with the respective alleles.
(DOCX)

S4 Table. Molecular interactions between receptor molecules (Chain A) and vaccine construct (Chain B).
(DOCX)

S5 Table. Codon optimization of vaccine construct.
(DOCX)

S6 Table. Druggability analysis of shortlisted non-human homologous pathogenic essential proteins.
(DOCX)

S7 Table. Top 10 hit molecules obtained from virtual screening using Pharmit server.
(DOCX)

S8 Table. CB-Dock energy, physicochemical, pharmacokinetics, and medicinal chemistry properties prediction of the molecules obtained from virtual screening using CB-Dock and SwissADME servers.
(DOCX)

Acknowledgments

The authors thank Dr. Chunhua Li from Beijing University of Technology, China and Dr. Asifullah Khan from Abdul Wali Khan University Mardan, Pakistan for providing technical support and supervising this research.

Author Contributions

Conceptualization: Sara Aiman, Jilong Zhang, Asifullah Khan.

Data curation: Sara Aiman, Qurrat ul Ain Farooq, Zhongjie Han, Muneeba Aslam, Asifullah Khan, Abbas Ahmad, Chunhua Li.

Formal analysis: Sara Aiman, Qurrat ul Ain Farooq, Zhongjie Han, Jilong Zhang, Abbas Ahmad.

Funding acquisition: Chunhua Li.

Investigation: Qurrat ul Ain Farooq, Jilong Zhang, Asifullah Khan, Abbas Ahmad, Chunhua Li.

Methodology: Sara Aiman, Qurrat ul Ain Farooq, Zhongjie Han, Abbas Ahmad.

Project administration: Asifullah Khan, Chunhua Li.

Resources: Muneeba Aslam, Asifullah Khan, Yasir Ali.

Software: Qurrat ul Ain Farooq.

Supervision: Asifullah Khan, Chunhua Li.

Validation: Muneeba Aslam, Asifullah Khan, Chunhua Li, Yasir Ali.

Visualization: Muneeba Aslam, Asifullah Khan, Abbas Ahmad, Chunhua Li, Yasir Ali.

Writing – original draft: Sara Aiman, Qurrat ul Ain Farooq, Asifullah Khan.

Writing – review & editing: Zhongjie Han, Muneeba Aslam, Jilong Zhang, Asifullah Khan, Chunhua Li.

References

1. Lawson PA, Citron DM, Tyrrell KL, Finegold SM. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe*. 2016; 40:95–9. Epub 2016/07/03. <https://doi.org/10.1016/j.anaerobe.2016.06.008> PMID: 27370902.
2. Bartlett JG. Narrative review: the new epidemic of *Clostridium difficile*-associated enteric disease. *Ann Intern Med*. 2006; 145(10):758–64. Epub 2006/11/23. <https://doi.org/10.7326/0003-4819-145-10-200611210-00008> PMID: 17116920.
3. Ananthakrishnan AN. *Clostridium difficile* infection: epidemiology, risk factors and management. *Nat Rev Gastroenterol Hepatol*. 2011; 8(1):17–26. Epub 2010/12/02. <https://doi.org/10.1038/nrgastro.2010.190> PMID: 21119612.
4. Rupnik M, Wilcox MH, Gerding DN. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol*. 2009; 7(7):526–36. Epub 2009/06/17. <https://doi.org/10.1038/nrmicro2164> PMID: 19528959.
5. Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ. *Clostridium difficile* infection. *Nat Rev Dis Primers*. 2016; 2:16020. Epub 2016/05/10. <https://doi.org/10.1038/nrdp.2016.20> PMID: 27158839; PubMed Central PMCID: PMC5453186.
6. Deshpande A, Pasupuleti V, Thota P, Pant C, Rolston DD, Sferra TJ, et al. Community-associated *Clostridium difficile* infection and antibiotics: a meta-analysis. *The Journal of antimicrobial chemotherapy*. 2013; 68(9):1951–61. Epub 2013/04/27. <https://doi.org/10.1093/jac/dkt129> PMID: 23620467.

7. Lubbert C, John E, von Muller L. Clostridium difficile infection: guideline-based diagnosis and treatment. Dtsch Arztebl Int. 2014; 111(43):723–31. Epub 2014/11/19. <https://doi.org/10.3238/arztebl.2014.0723> PMID: 25404529; PubMed Central PMCID: PMC4238315.
8. Hensgens MP, Goorhuis A, Dekkers OM, Kuijper EJ. Time interval of increased risk for Clostridium difficile infection after exposure to antibiotics. The Journal of antimicrobial chemotherapy. 2012; 67(3):742–8. Epub 2011/12/08. <https://doi.org/10.1093/jac/dkr508> PMID: 22146873.
9. Figueroa Castro CE, Munoz-Price LS. Advances in Infection Control for Clostridioides (Formerly Clostridium) difficile Infection. Current Treatment Options in Infectious Diseases. 2019; 11(1):12–22. <https://doi.org/10.1007/s40506-019-0179-y>
10. Scardina T, Labuszewski L, Pacheco SM, Adams W, Schreckenberger P, Johnson S. Clostridium difficile infection (CDI) severity and outcome among patients infected with the NAP1/BI/027 strain in a non-epidemic setting. Infection control and hospital epidemiology. 2015; 36(3):280–6. Epub 2015/02/20. <https://doi.org/10.1017/ice.2014.45> PMID: 25695169.
11. Bauer KA, Johnston JEW, Wenzler E, Goff DA, Cook CH, Balada-Llasat JM, et al. Impact of the NAP-1 strain on disease severity, mortality, and recurrence of healthcare-associated Clostridium difficile infection. Anaerobe. 2017; 48:1–6. Epub 2017/06/25. <https://doi.org/10.1016/j.anaerobe.2017.06.009> PMID: 28645479.
12. Zilberberg MD, Shorr AF, Kollef MH. Increase in adult Clostridium difficile-related hospitalizations and case-fatality rate, United States, 2000–2005. Emerg Infect Dis. 2008; 14(6):929–31. Epub 2008/05/30. <https://doi.org/10.3201/eid1406.071447> PMID: 18507904; PubMed Central PMCID: PMC2600276.
13. Loo VG, Poirier L, Miller MA, Oughton M, Libman MD, Michaud S, et al. A predominantly clonal multi-institutional outbreak of Clostridium difficile-associated diarrhea with high morbidity and mortality. The New England journal of medicine. 2005; 353(23):2442–9. Epub 2005/12/03. <https://doi.org/10.1056/NEJMoa051639> PMID: 16322602.
14. McDonald LC, Killgore GE, Thompson A, Owens RC Jr., Kazakova SV, Sambol SP, et al. An epidemic, toxin gene-variant strain of Clostridium difficile. The New England journal of medicine. 2005; 353(23):2433–41. Epub 2005/12/03. <https://doi.org/10.1056/NEJMoa051590> PMID: 16322603.
15. Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, et al. Toxin production by an emerging strain of Clostridium difficile associated with outbreaks of severe disease in North America and Europe. Lancet. 2005; 366(9491):1079–84. Epub 2005/09/27. [https://doi.org/10.1016/S0140-6736\(05\)67420-X](https://doi.org/10.1016/S0140-6736(05)67420-X) PMID: 16182895.
16. Đapa T, Leuzzi R, Ng YK, Baban ST, Adamo R, Kuehne SA, et al. Multiple Factors Modulate Biofilm Formation by the Anaerobic Pathogen *Clostridium difficile*. 2013; 195(3):545–55. <https://doi.org/10.1128/JB.01980-12> J Journal of Bacteriology. PMID: 23175653
17. Freeman J, Bauer MP, Baines SD, Corver J, Fawley WN, Goorhuis B, et al. The changing epidemiology of Clostridium difficile infections. Clinical microbiology reviews. 2010; 23(3):529–49. Epub 2010/07/09. <https://doi.org/10.1128/CMR.00082-09> PMID: 20610822; PubMed Central PMCID: PMC2901659.
18. Vindigni SM, Surawicz CM. C. difficile Infection: Changing Epidemiology and Management Paradigms. Clin Transl Gastroenterol. 2015; 6(7):e99. Epub 2015/07/15. <https://doi.org/10.1038/ctg.2015.24> PMID: 26158611; PubMed Central PMCID: PMC4816260.
19. Andersson JA, Sha J, Kirtley ML, Reyes E, Fitts EC, Dann SM, et al. Combating Multidrug-Resistant Pathogens with Host-Directed Nonantibiotic Therapeutics. 2018; 62(1):e01943–17. <https://doi.org/10.1128/AAC.01943-17> J Antimicrobial Agents and Chemotherapy. PMID: 29109161
20. Prevention CfDca. 2019 Antibiotic Resistance Threats Report 2019. Available from: <https://www.cdc.gov/drugresistance/biggest-threats.html#:~:text=In%20addition%2C%20223%2C900%20cases%20of,at%20least%2012%2C800%20people%20died.>
21. Spigaglia P. Recent advances in the understanding of antibiotic resistance in Clostridium difficile infection. Ther Adv Infect Dis. 2016; 3(1):23–42. Epub 2016/02/11. <https://doi.org/10.1177/2049936115622891> PMID: 26862400; PubMed Central PMCID: PMC4735502.
22. Isidro J, Menezes J, Serrano M, Borges V, Paixão P, Mimoso M, et al. Genomic Study of a Clostridium difficile Multidrug Resistant Outbreak-Related Clone Reveals Novel Determinants of Resistance. Frontiers in microbiology. 2018; 9:2994–. <https://doi.org/10.3389/fmicb.2018.02994> PMID: 30574133.
23. Peng Z, Jin D, Kim HB, Stratton CW, Wu B, Tang YW, et al. Update on Antimicrobial Resistance in Clostridium difficile: Resistance Mechanisms and Antimicrobial Susceptibility Testing. J Clin Microbiol. 2017; 55(7):1998–2008. Epub 2017/04/14. <https://doi.org/10.1128/JCM.02250-16> PMID: 28404671; PubMed Central PMCID: PMC5483901.

24. Johnson S. Recurrent *Clostridium difficile* infection: a review of risk factors, treatments, and outcomes. *J Infect*. 2009; 58(6):403–10. Epub 2009/04/28. <https://doi.org/10.1016/j.jinf.2009.03.010> PMID: 19394704.
25. Cornely OA, Miller MA, Louie TJ, Crook DW, Gorbach SL. Treatment of first recurrence of *Clostridium difficile* infection: fidaxomicin versus vancomycin. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2012; 55 Suppl 2(Suppl 2):S154–61. Epub 2012/07/07. <https://doi.org/10.1093/cid/cis462> PMID: 22752865; PubMed Central PMCID: PMC3388030.
26. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, et al. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biology*. 2010; 11(10):R107. <https://doi.org/10.1186/gb-2010-11-10-r107> PMID: 21034474
27. Hao P, Zheng H, Yu Y, Ding G, Gu W, Chen S, et al. Complete Sequencing and Pan-Genomic Analysis of *Lactobacillus delbrueckii* subsp. *bulgaricus* Reveal Its Genetic Basis for Industrial Yogurt Production. *PLoS one*. 2011; 6(1):e15964. <https://doi.org/10.1371/journal.pone.0015964> PMID: 21264216
28. Hizbullah, Nazir Z, Afridi SG, Shah M, Shams S, Khan A. Reverse vaccinology and subtractive genomics-based putative vaccine targets identification for *Burkholderia pseudomallei* Bp1651. *Microbial Pathogenesis*. 2018; 125:219–29. <https://doi.org/10.1016/j.micpath.2018.09.033> PMID: 30243554
29. Aiman S, Ali Y, Malik A, Alkholief M, Ahmad A, Akhtar S, et al. Immunoinformatic-guided novel mRNA vaccine designing to elicit immunogenic responses against the endemic Monkeypox virus. *Journal of Biomolecular Structure and Dynamics*. 2023:1–15. <https://doi.org/10.1080/07391102.2023.2233627> PMID: 37424185
30. Aiman S, Ahmad A, Khan AA, Alanazi AM, Samad A, Ali SL, et al. Vaccinomics-based next-generation multi-epitope chimeric vaccine models prediction against *Leishmania tropica* - a hierarchical subtractive proteomics and immunoinformatics approach. 2023;14. <https://doi.org/10.3389/fimmu.2023.1259612> PMID: 37781384
31. Farzan M, Farzan M, Mirzaei Y, Aiman S, Azadegan-Dehkordi F, Bagheri N. Immunoinformatics-based multi-epitope vaccine design for the re-emerging monkeypox virus. *International Immunopharmacology*. 2023; 123:110725. <https://doi.org/10.1016/j.intimp.2023.110725> PMID: 37556996
32. Thomson KS. Extended-spectrum-beta-lactamase, AmpC, and Carbapenemase issues. *J Clin Microbiol*. 2010; 48(4):1019–25. Epub 2010/02/26. <https://doi.org/10.1128/JCM.00219-10> PMID: 20181902; PubMed Central PMCID: PMC2849556.
33. Blom J, Kreis J, Spänig S, Juhre T, Bertelli C, Ernst C, et al. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic acids research*. 2016; 44(W1):W22–W8. <https://doi.org/10.1093/nar/gkw255> PMID: 27098043
34. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics (Oxford, England)*. 2010; 26(5):680–2. <https://doi.org/10.1093/bioinformatics/btq003> PMID: 20053844
35. Aslam M, Shehroz M, Hizbullah, Shah M, Khan MA, Afridi SG, et al. Potential druggable proteins and chimeric vaccine construct prioritization against *Brucella melitensis* from species core genome data. *Genomics*. 2020; 112(2):1734–45. Epub 2019/11/05. <https://doi.org/10.1016/j.ygeno.2019.10.009> PMID: 31678593.
36. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics (Oxford, England)*. 2010; 26(13):1608–15. Epub 2010/05/18. <https://doi.org/10.1093/bioinformatics/btq249> PMID: 20472543; PubMed Central PMCID: PMC2887053.
37. Dimitrov I, Flower DR, Doytchinova I. AllerTOP—a server for in silico prediction of allergens. *BMC bioinformatics*. 2013; 14 Suppl 6(Suppl 6):S4. Epub 2013/06/14. <https://doi.org/10.1186/1471-2105-14-S6-S4> PMID: 23735058; PubMed Central PMCID: PMC3633022.
38. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007; 8(1):4. Epub 2007/01/09. <https://doi.org/10.1186/1471-2105-8-4> PMID: 17207271; PubMed Central PMCID: PMC1780059.
39. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP. Peptide toxicity prediction. *Methods in molecular biology (Clifton, NJ)*. 2015; 1268:143–57. Epub 2015/01/06. https://doi.org/10.1007/978-1-4939-2285-7_7 PMID: 25555724.
40. Ahmed RKS, Maeurer MJ. T-Cell Epitope Mapping. In: Schutkowski M, Reineke U, editors. *Epitope Mapping Protocols: Second Edition*. Totowa, NJ: Humana Press; 2009. p. 427–38.
41. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. 2017; 8(278). <https://doi.org/10.3389/fimmu.2017.00278> PMID: 28352270

42. Aiman S, Ali F, Zia A, Aslam M, Han Z, Shams S, et al. Core genome mediated potential vaccine targets prioritization against *Clostridium difficile* via reverse vaccinology-an immuno-informatics approach. 2022;29. <https://doi.org/10.26262/jbrt.v29i0.8481>
43. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic acids research*. 2020; 48(W1):W449–w54. Epub 2020/05/15. <https://doi.org/10.1093/nar/gkaa379> PMID: 32406916; PubMed Central PMCID: PMC7319546.
44. Rahman N, Ali F, Basharat Z, Shehroz M, Khan MK, Jeandet P, et al. Vaccine Design from the Ensemble of Surface Glycoprotein Epitopes of SARS-CoV-2: An Immunoinformatics Approach. *Vaccines (Basel)*. 2020; 8(3):423. Epub 2020/08/01. <https://doi.org/10.3390/vaccines8030423> PMID: 32731461; PubMed Central PMCID: PMC7565012.
45. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*. 2006; 65(1):40–8. Epub 2006/08/09. <https://doi.org/10.1002/prot.21078> PMID: 16894596.
46. Adhikari UK, Tayebi M, Rahman MM. Immunoinformatics Approach for Epitope-Based Peptide Vaccine Design and Active Site Prediction against Polyprotein of Emerging Oropouche Virus. *J Immunol Res*. 2018; 2018:6718083. Epub 2018/11/08. <https://doi.org/10.1155/2018/6718083> PMID: 30402510; PubMed Central PMCID: PMC6196980.
47. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2006; 7:153. Epub 2006/03/21. <https://doi.org/10.1186/1471-2105-7-153> PMID: 16545123; PubMed Central PMCID: PMC1513259.
48. Fadaka AO, Sibuyi NRS, Martin DR, Goboza M, Klein A, Madiehe AM, et al. Immunoinformatics design of a novel epitope-based vaccine candidate against dengue virus. *Sci Rep*. 2021; 11(1):19707. Epub 2021/10/07. <https://doi.org/10.1038/s41598-021-99227-7> PMID: 34611250; PubMed Central PMCID: PMC8492693.
49. Li W, Joshi MD, Singhania S, Ramsey KH, Murthy AK. Peptide Vaccine: Progress and Challenges. *Vaccines (Basel)*. 2014; 2(3):515–36. Epub 2014/01/01. <https://doi.org/10.3390/vaccines2030515> PMID: 26344743; PubMed Central PMCID: PMC4494216.
50. Aslam M, Shehroz M, Ali F, Zia A, Pervaiz S, Shah M, et al. Chlamydia trachomatis core genome data mining for promising novel drug targets and chimeric vaccine candidates identification. *Computers in Biology and Medicine*. 2021; 136:104701. <https://doi.org/10.1016/j.combiomed.2021.104701> PMID: 34364258
51. Nezafat N, Karimi Z, Eslami M, Mohkam M, Zandian S, Ghasemi Y. Designing an efficient multi-epitope peptide vaccine against *Vibrio cholerae* via combined immunoinformatics and protein interaction based approaches. *Comput Biol Chem*. 2016; 62:82–95. Epub 2016/04/24. <https://doi.org/10.1016/j.combiolchem.2016.04.006> PMID: 27107181.
52. Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, et al. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics (Oxford, England)*. 2010; 26(23):2936–43. Epub 2010/10/12. <https://doi.org/10.1093/bioinformatics/btq551> PMID: 20934990; PubMed Central PMCID: PMC2982151.
53. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. *Nucleic acids research*. 2005; 33(suppl_2):W72–W6. <https://doi.org/10.1093/nar/gki396> PMID: 15980571
54. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics (Oxford, England)*. 2009; 25(17):2200–7. Epub 2009/06/25. <https://doi.org/10.1093/bioinformatics/btp386> PMID: 19549632.
55. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods in molecular biology (Clifton, NJ)*. 1999; 112:531–52. Epub 1999/02/23. <https://doi.org/10.1385/1-59259-584-7:531> PMID: 10027275.
56. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic acids research*. 2019; 47(W1):W402–W7. <https://doi.org/10.1093/nar/gkz297> PMID: 31251384
57. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*. 1995; 11(6):681–4. Epub 1995/12/01. <https://doi.org/10.1093/bioinformatics/11.6.681> PMID: 8808585.
58. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*. 2015; 10(6):845–58. <https://doi.org/10.1038/nprot.2015.053> PMID: 25950237
59. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic acids research*. 2013; 41(Web Server issue):W384–8. Epub 2013/06/06. <https://doi.org/10.1093/nar/gkt458> PMID: 23737448; PubMed Central PMCID: PMC3692086.

60. Lovell SC, Davis IW, Arendall III WB, de Bakker PIW, Word JM, Prisant MG, et al. Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins: Structure, Function, and Bioinformatics*. 2003; 50(3):437–50. <https://doi.org/10.1002/prot.10286>.
61. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*. 2007; 35(Web Server issue):W407–10. Epub 2007/05/23. <https://doi.org/10.1093/nar/gkm290> PMID: 17517781; PubMed Central PMCID: PMC1933241.
62. Kozakov D, Hall DR, Xia B, Porter KA, Padhorna D, Yueh C, et al. The ClusPro web server for protein–protein docking. *Nature Protocols*. 2017; 12(2):255–78. <https://doi.org/10.1038/nprot.2016.169> PMID: 28079879
63. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. *Protein Sci*. 2018; 27(1):129–34. Epub 2017/09/07. <https://doi.org/10.1002/pro.3289> PMID: 28875543; PubMed Central PMCID: PMC5734310.
64. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational Immunology Meets Bioinformatics: The Use of Prediction Tools for Molecular Binding in the Simulation of the Immune System. *PLoS one*. 2010; 5(4):e9862. <https://doi.org/10.1371/journal.pone.0009862> PMID: 20419125
65. Nain Z, Abdulla F, Rahman MM, Karim MM, Khan MSA, Sayed SB, et al. Proteome-wide screening for designing a multi-epitope vaccine against emerging pathogen *Elizabethkingia anophelis* using immunoinformatic approaches. *Journal of Biomolecular Structure and Dynamics*. 2020; 38(16):4850–67. <https://doi.org/10.1080/07391102.2019.1692072> PMID: 31709929
66. Ezeanolue E, Harriman K, Hunter P, Kroger A, Pellegrini C. General best practice guidelines for immunization: best practices guidance of the Advisory Committee on Immunization Practices (ACIP). National Center for Immunization and Respiratory Diseases. 2019.
67. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic acids research*. 2005; 33(Web Server issue):W526–W31. <https://doi.org/10.1093/nar/gki376>
68. Ismail S, Abbasi SW, Yousaf M, Ahmad S, Muhammad K, Waheed Y. Design of a Multi-Epitopes Vaccine against Hantaviruses: An Immunoinformatics and Molecular Modelling Approach. *Vaccines (Basel)*. 2022; 10(3). Epub 2022/03/27. <https://doi.org/10.3390/vaccines10030378> PMID: 35335010; PubMed Central PMCID: PMC8953224.
69. Kamens J. The Addgene repository: an international nonprofit plasmid and data resource. *Nucleic acids research*. 2015; 43(D1):D1152–D7. <https://doi.org/10.1093/nar/gku893> PMID: 25392412
70. Gupta R, Pradhan D, Jain AK, Rai CS. TiD: Standalone software for mining putative drug targets from bacterial proteome. *Genomics*. 2017; 109(1):51–7. <https://doi.org/10.1016/j.ygeno.2016.11.005> PMID: 27856224
71. Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic acids research*. 2008; 36(suppl_2):W5–W9. <https://doi.org/10.1093/nar/gkn201> PMID: 18440982
72. Luo H, Lin Y, Gao F, Zhang C-T, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic acids research*. 2014; 42(D1):D574–D80. <https://doi.org/10.1093/nar/gkt1131> PMID: 24243843
73. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic acids research*. 2000; 28(1):235–42. <https://doi.org/10.1093/nar/28.1.235> PMID: 10592235.
74. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*. 2014; 42(W1):W252–W8. <https://doi.org/10.1093/nar/gku340> PMID: 24782522
75. Borrel A, Regad L, Xhaard H, Petitjean M, Camproux A-C. PockDrug: A Model for Predicting Pocket Druggability That Overcomes Pocket Estimation Uncertainties. *Journal of Chemical Information and Modeling*. 2015; 55(4):882–95. <https://doi.org/10.1021/ci5006004> PMID: 25835082
76. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*. 2015; 43(D1):D447–D52. <https://doi.org/10.1093/nar/gku1003> PMID: 25352553
77. Sunseri J, Koes DR. Pharmit: interactive exploration of chemical space. *Nucleic acids research*. 2016; 44(W1):W442–W8. Epub 2016/04/19. <https://doi.org/10.1093/nar/gkw287> PMID: 27095195.
78. Ali F, Wali H, Jan S, Zia A, Aslam M, Ahmad I, et al. Analysing the essential proteins set of *Plasmodium falciparum* PF3D7 for novel drug targets identification against malaria. *Malaria journal*. 2021; 20(1):335. <https://doi.org/10.1186/s12936-021-03865-1> PMID: 34344361
79. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*.

- 2001; 46(1–3):3–26. Epub 2001/03/22. [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0) PMID: 11259830.
80. Rahul A, Ashutosh S, Subhabrata S. Role of Molecular Docking in Computer-Aided Drug Design and Development. In: Siavoush D, Maryam H-M, Babak S, editors. Applied Case Studies and Solutions in Molecular Docking-Based Drug Design. Hershey, PA, USA: IGI Global; 2016. p. 1–28.
 81. Liu Y, Grimm M, Dai W-t, Hou M-c, Xiao Z-X, Cao Y. CB-Dock: a web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacologica Sinica*. 2020; 41(1):138–44. <https://doi.org/10.1038/s41401-019-0228-6> PMID: 31263275
 82. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*. 2017; 7(1):42717. <https://doi.org/10.1038/srep42717> PMID: 28256516
 83. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. 2015; 1:19–25.
 84. Ali Y, Ahmad F, Ullah MF, Haq NU, Haq MIU, Aziz A, et al. Structural Evaluation and Conformational Dynamics of ZNF141(T474I) Mutation Provoking Postaxial Polydactyly Type A. *Bioengineering* (Basel, Switzerland). 2022; 9(12). Epub 2022/12/24. <https://doi.org/10.3390/bioengineering9120749> PMID: 36550955; PubMed Central PMCID: PMC9774408.
 85. Bui HH, Sidney J, Li W, Füsseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC bioinformatics*. 2007; 8:361. Epub 2007/09/28. <https://doi.org/10.1186/1471-2105-8-361> PMID: 17897458; PubMed Central PMCID: PMC2233646.
 86. Chen H, Wu B, Zhang T, Jia J, Lu J, Chen Z, et al. Effect of Linker Length and Flexibility on the Clostridium thermocellum Esterase Displayed on Bacillus subtilis Spores. *Appl Biochem Biotechnol*. 2017; 182(1):168–80. Epub 2016/12/10. <https://doi.org/10.1007/s12010-016-2318-y> PMID: 27933482.
 87. Wu C-Y, Monie A, Pang X, Hung C-F, Wu TC. Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. *Journal of Biomedical Science*. 2010; 17(1):88. <https://doi.org/10.1186/1423-0127-17-88> PMID: 21092195
 88. Corradin G, Villard V, Kajava AV. Protein structure based strategies for antigen discovery and vaccine development against malaria and other pathogens. *Endocr Metab Immune Disord Drug Targets*. 2007; 7(4):259–65. Epub 2008/01/29. <https://doi.org/10.2174/187153007782794371> PMID: 18220946.
 89. Benkert P, Tosatto SCE, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*. 2008; 71(1):261–77. <https://doi.org/10.1002/prot.21715> PMID: 17932912
 90. Parkinson GN, Skelly JV, Neidle S. Crystal structure of FMN-dependent nitroreductase from Escherichia coli B: a prodrug-activating enzyme. *J Med Chem*. 2000; 43(20):3624–31. Epub 2000/10/06. <https://doi.org/10.1021/jm000159m> PMID: 11020276.
 91. Abi Hussein H, Geneix C, Petitjean M, Borrel A, Flatters D, Camproux A-C. Global vision of druggability issues: applications and perspectives. *Drug discovery today*. 2017; 22(2):404–15. <https://doi.org/10.1016/j.drudis.2016.11.021> PMID: 27939283
 92. Ertl P, Rohde B, Selzer P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *Journal of Medicinal Chemistry*. 2000; 43(20):3714–7. <https://doi.org/10.1021/jm000942e> PMID: 11020286
 93. Hodgson J. ADMET—turning chemicals into drugs. *Nature biotechnology*. 2001; 19(8):722–6. Epub 2001/08/02. <https://doi.org/10.1038/90761> PMID: 11479558.
 94. Palm K, Stenberg P, Luthman K, Artursson P. Polar Molecular Surface Properties Predict the Intestinal Absorption of Drugs in Humans. *Pharmaceutical Research*. 1997; 14(5):568–71. <https://doi.org/10.1023/a:1012188625088>:1012188625088. PMID: 9165525
 95. Basharat Z, Jahanzaib M, Rahman N. Therapeutic target identification via differential genome analysis of antibiotic resistant Shigella sonnei and inhibitor evaluation against a selected drug target. *Infect Genet Evol*. 2021; 94:105004. Epub 2021/07/20. <https://doi.org/10.1016/j.meegid.2021.105004> PMID: 34280580.
 96. Zahid S, Ali Y, Rashid S. Structural-based design of HD-TAC7 PROteolysis TArgeting chimeras (PROTACs) candidate transformations to abrogate SARS-CoV-2 infection. *Journal of Biomolecular Structure and Dynamics*. 2023:1–16. <https://doi.org/10.1080/07391102.2023.2183037> PMID: 36841549
 97. Leuzzi R, Adamo R, Scarselli M. Vaccines against Clostridium difficile. *Human vaccines & immunotherapeutics*. 2014; 10(6):1466–77. Epub 2014/03/19. <https://doi.org/10.4161/hv.28428> PMID: 24637887; PubMed Central PMCID: PMC5396221.

98. Moise L, Gutierrez A, Kibria F, Martin R, Tassone R, Liu R, et al. iVAX: An integrated toolkit for the selection and optimization of antigens and the design of epitope-driven vaccines. *Human vaccines & immunotherapeutics*. 2015; 11(9):2312–21. Epub 2015/07/15. <https://doi.org/10.1080/21645515.2015.1061159> PMID: 26155959; PubMed Central PMCID: PMC4635942.
99. Zoraghi R, Reiner NE. Protein interaction networks as starting points to identify novel antimicrobial drug targets. *Current Opinion in Microbiology*. 2013; 16(5):566–72. <https://doi.org/10.1016/j.mib.2013.07.010> PMID: 23938265
100. He X, Zhang J. Why Do Hubs Tend to Be Essential in Protein Networks? *PLOS Genetics*. 2006; 2(6): e88. <https://doi.org/10.1371/journal.pgen.0020088> PMID: 16751849
101. Koder RL, Miller AF. Steady-state kinetic mechanism, stereospecificity, substrate and inhibitor specificity of *Enterobacter cloacae* nitroreductase. *Biochimica et biophysica acta*. 1998; 1387(1–2):395–405. Epub 1998/09/28. [https://doi.org/10.1016/s0167-4838\(98\)00151-4](https://doi.org/10.1016/s0167-4838(98)00151-4) PMID: 9748656.
102. Roldán MD, Pérez-Reinado E, Castillo F, Moreno-Vivián C. Reduction of polynitroaromatic compounds: the bacterial nitroreductases. *FEMS Microbiology Reviews*. 2008; 32(3):474–500. <https://doi.org/10.1111/j.1574-6976.2008.00107.x> PMID: 18355273
103. Johansson E, Parkinson GN, Denny WA, Neidle S. Studies on the nitroreductase prodrug-activating system. Crystal structures of complexes with the inhibitor dicoumarol and dinitrobenzamide prodrugs and of the enzyme active form. *J Med Chem*. 2003; 46(19):4009–20. Epub 2003/09/05. <https://doi.org/10.1021/jm030843b> PMID: 12954054.
104. Crofts TS, Sontha P, King AO, Wang B, Bidy BA, Zanolli N, et al. Discovery and Characterization of a Nitroreductase Capable of Conferring Bacterial Resistance to Chloramphenicol. *Cell chemical biology*. 2019; 26(4):559–70.e6. Epub 2019/02/21. <https://doi.org/10.1016/j.chembiol.2019.01.007> PMID: 30799223.
105. Martins SA, Combs JC, Noguera G, Camacho W, Wittmann P, Walther R, et al. Antimicrobial efficacy of riboflavin/UVA combination (365 nm) in vitro for bacterial and fungal isolates: a potential new treatment for infectious keratitis. *Invest Ophthalmol Vis Sci*. 2008; 49(8):3402–8. Epub 2008/04/15. <https://doi.org/10.1167/iovs.07-1592> PMID: 18408193.
106. Lee ER, Blount KF, Breaker RR. Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA biology*. 2009; 6(2):187–94. Epub 2009/04/30. <https://doi.org/10.4161/rna.6.2.7727> PMID: 19246992.
107. Ali Y, Imtiaz H, Tahir MM, Gul F, Sadozai UAK, Ur Rehman A, et al. Fragment-Based Approaches Identified Tecovirimat-Competitive Novel Drug Candidate for Targeting the F13 Protein of the Monkeypox Virus. *Viruses*. 2023; 15(2). Epub 2023/03/01. <https://doi.org/10.3390/v15020570> PMID: 36851785; PubMed Central PMCID: PMC9959752.
108. Hutchings MI, Truman AW, Wilkinson B. Antibiotics: past, present and future. *Current Opinion in Microbiology*. 2019; 51:72–80. <https://doi.org/10.1016/j.mib.2019.10.008> PMID: 31733401