

In Silico Functional Characterization of a Hypothetical Protein From *Pasteurella Multocida* Reveals a Novel S-Adenosylmethionine-Dependent Methyltransferase Activity

Bioinformatics and Biology Insights
Volume 17: 1–17
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11779322231184024



Md. Habib Ullah Masum^{1*}, Sultana Rajia^{1*}, Udit Paul Bristi¹,
Mir Salma Akter¹, Mohammad Ruhul Amin¹, Tushar Ahmed Shishir²,
Jannatul Ferdous³, Firoz Ahmed¹, Md. Mizanur Rahaman⁴
and Otun Saha¹

¹Department of Microbiology, Noakhali Science and Technology University, Noakhali, Bangladesh. ²Department of Mathematics and Natural Sciences, BRAC University, Dhaka, Bangladesh. ³Department of Medicine, Abdul Malek Ukil Medical College, Noakhali, Bangladesh. ⁴Department of Microbiology, University of Dhaka, Dhaka, Bangladesh.

ABSTRACT: Genomes may now be sequenced in a matter of weeks, leading to an influx of “hypothetical” proteins (HP) whose activities remain a mystery in GenBank. The information included inside these genes has quickly grown in prominence. Thus, we selected to look closely at the structure and function of an HP (AFF25514.1; 246 residues) from *Pasteurella multocida* (PM) subsp. *multocida* str. HN06. Possible insights into bacterial adaptation to new environments and metabolic changes might be gained by studying the functions of this protein. The PM HN06 2293 gene encodes an alkaline cytoplasmic protein with a molecular weight of 28352.60Da, an isoelectric point (pI) of 9.18, and an overall average hydrophobicity of around -0.565 . One of its functional domains, tRNA (adenine (37)-N6)-methyltransferase TrmO, is a S-adenosylmethionine (SAM)-dependent methyltransferase (MTase), suggesting that it belongs to the Class VIII SAM-dependent MTase family. The tertiary structures represented by HHpred and I-TASSER models were found to be flawless. We predicted the model's active site using the Computed Atlas of Surface Topography of Proteins (CASTp) and FTSite servers, and then displayed it in 3 dimensional (3D) using PyMOL and BIOVIA Discovery Studio. Based on molecular docking (MD) results, we know that HP interacts with SAM and S-adenosylhomocysteine (SAH), 2 crucial metabolites in the tRNA methylation process, with binding affinities of 7.4 and 7.5 kcal/mol, respectively. Molecular dynamic simulations (MDS) of the docked complex, which included only modest structural adjustments, corroborated the strong binding affinity of SAM and SAH to the HP. Evidence for HP's possible role as an SAM-dependent MTase was therefore given by the findings of Multiple sequence alignment (MSA), MD, and molecular dynamic modeling. These in silico data suggest that the investigated HP might be used as a useful adjunct in the investigation of *Pasteurella* infections and the development of drugs to treat zoonotic pasteurellosis.

KEYWORDS: *Pasteurella multocida*, Pasteurellosis, Novel SAM-dependent MTase, Hypothetical protein, Molecular docking, Molecular dynamics

RECEIVED: April 1, 2023. **ACCEPTED:** June 6, 2023.

TYPE: Original Research Article

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Otun Saha, Department of Microbiology, Noakhali Science and Technology University, University Road, Noakhali 3814, Bangladesh. Email: otun.saha@nstu.edu.bd

Introduction

Next-generation sequencing (NGS) has shortened the time it takes researchers to collect massive volumes of data.¹ The difficulty of attributing functions to genes is growing as the genomes of more and more species are sequenced. Among the all sequenced data, more than 30% of proteins in various animals are called “Hypothetical Proteins” (HPs) because their molecular activities are unknown.² The increased quantity of raw HP is compelling researchers to find ways to use them. *Characterizing hypothetical proteins in silico aids in the determination of their 3-dimensional (3D) structures, which may lead to the discovery of previously unknown domains, motifs, pathways, protein networks, etc.*^{3–5} *Potential biomarkers and pharmaceutical targets may potentially be uncovered by structural and functional annotation of HPs.*⁶ *One such example is the*

*newly discovered Shigella dysenteriae ATCC 12039 HP, which shows promise as a treatment against that particular bacteria.*¹ *Meanwhile, newly characterized M4, a bacterial metalloprotease, demonstrating their use in the development of antimicrobial vaccines and biotechnological enzymes.*⁷ In addition, an HP from *Orientia tsutsugamushi* str. Karp shows promise as a new antibacterial medication targeting the bacterium. The roles of putative proteins in several pathogenic bacteria have been effectively annotated using a number of bioinformatics databases and techniques.^{8–10} *Pasteurella multocida* (PM) is an example of a pathogenic bacterium; it is rod-shaped, gram-negative, facultative anaerobic, coagulase-negative, and causes several zoonotic diseases across a wide range of hosts and habitats.^{1–4} The bacterium is a common commensal or opportunistic pathogen that lives in the upper respiratory tracts of many different types of animals.¹¹ This includes dogs,^{12–15} cats,^{16–18} rabbits,^{19–21} cattle,^{22,23} goat,^{24–27} bison,²⁸

* Equal contribution.



swine,^{11,29-31} marine mammals,³² chimpanzee,³³ and komodo dragons.^{34,35} This implies that the PM may infect a wide variety of species and is responsible for a number of economically significant illnesses such as avian fowl cholera, bovine hemorrhagic septicemia, zoonotic pneumonia, and swine atrophic rhinitis.^{2,3,5,9} In humans, respiratory infections are uncommon, but those who suffer from chronic pulmonary sickness are particularly vulnerable.^{36,37} Severe consolidation pneumonia, epiglottitis, lymphadenopathy, and abscess formation are all possible symptoms of pasteurellosis in such situations.^{15,38} In the previous 30 years, the number of human cases of pasteurellosis has increased from 20 to 30, and this trend seems to be continuing. It has been estimated that more than 300 000 persons in the United States visit emergency rooms annually due to animal scratches or bites, with PM being the most often linked illness type.^{15,39} A total of 162 cases of *Pasteurella* infections were reported in Hungary during the years of 2002 and 2015.⁴⁰ Forty-four instances of *Pasteurella* infections were reported in the United States between 2000 and 2014, with 8 patients requiring intensive care unit (ICU) treatment.⁴¹ Invasive pasteurellosis, however, was associated with a 27.1% mortality rate in Hungary and a 21% mortality rate in the United States.^{40,41} Because of this, there is now a pressing need to investigate and study zoonotic pasteurellosis extensively to contain its spread. In this light, PM subsp. *multocida* str. HN06, the whole genome of which was just released. The National Center for Biotechnology Information (NCBI) database states that it encoding 2117 proteins (AFF25514.1). Nevertheless, expression and function data are missing for approximately 2000 predicted protein-encoding coding sequences. The word “hypothetical” has been used to these chains. These HPs account for almost half of all proteins in the genome (47.6%). For these HPs to discover their potential roles in the cell and provide light on novel structures and functions in this bacterium’s participation in the illness process, functional annotation is essential. Because of the potential importance of this organism’s genome to the success of a medication or vaccine still in development in labs, in silico examination of these putative proteins is crucial. In this work, we use many different bioinformatics programs to investigate the structure and function of a putative protein (accession no. AFF25514.1; 246) from PM subsp. *multocida* str. HN06.

In light of these considerations, the purpose of this study is to define a PM subsp. *multocida* str. HN06 HP and investigate its potential as a therapeutic target of the bacterium, which may be helpful in combating zoonotic pasteurellosis as well. Thus, we predicted the 3D structure of this bacterium’s HP, annotated its function, and described it using a variety of computational methods. Researchers also identified its role as an altered form of a protein essential to their replication machinery called *S*-adenosylmethionine (AdoMet or SAM)-dependent methyltransferase (MTase). The potential of this HP in preventing pasteurellosis was effectively identified.

Ultimately, this study may be used as a future hope for preventing and treating PM zoonosis.

Materials and Methods

Retrieval of protein sequence

By searching the NCBI Protein database (<https://www.ncbi.nlm.nih.gov/protein/>) for the phrase “Hypothetical proteins AND *Pasteurella multocida*” we were able to locate the 246-residue HP of PM subsp. *multocida* str. HN06. Among the hits found, we randomly selected an HP (accession no. AFF25514.1, GI|380873147|), and its sequence was acquired in FASTA format for further examination. A sequence-based peptide search was also performed in the UniProt database (<https://www.uniprot.org/peptidesearch/>) to determine whether or not the protein is redundant. The whole research plan is shown in Figure 1.

Physicochemical properties analysis

The chemical and physical attributes of the favored HP were assessed using the ProtParam tool on the ExPASy website (<https://web.expasy.org/protam/>). The analyzer provides theoretical metrics such as molecular mass, amino acid composition, totally positive and negative residue count, extinction coefficient, theoretical pH, aliphatic index (AI), instability index (II), and grand average of hydropathicity (GRAVY) score.⁴²

Annotation of functional domain

Functional annotation was applied to the HP to reveal its functions. Several publicly available tools and databases, including NCBI CDD (<https://www.ncbi.nlm.nih.gov/cdd/>),⁴³ InterProScan (<https://www.ebi.ac.uk/interpro/search/sequence/>),⁴⁴ and SUPERFAMILY (<https://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/>)⁴⁵ were used to annotate precisely the conserved and functional domain within HP. The default settings were considered in each case. These databases and other bioinformatics tools aid in the identification of conserved domains, which are then used to classify the proteins.

Multiple sequence alignment and phylogenetic analysis

Sequence similarities with the studied HP were searched using NCBI’s Basic Local Alignment Search Tool (BLAST) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). We used NCBI’s BLASTp method⁴⁶ to search for matches in a unique protein database. Multiple protein sequences were initially retrieved from the NCBI protein database, all of which were assumed to have the same purpose. The Molecular Evolutionary Genetics Analysis X (MEGA X) program was then used to conduct the multiple sequence alignment (MSA) and phylogenetic analysis between the HP and recovered protein sequences.⁴⁷ The ClustalW method, which

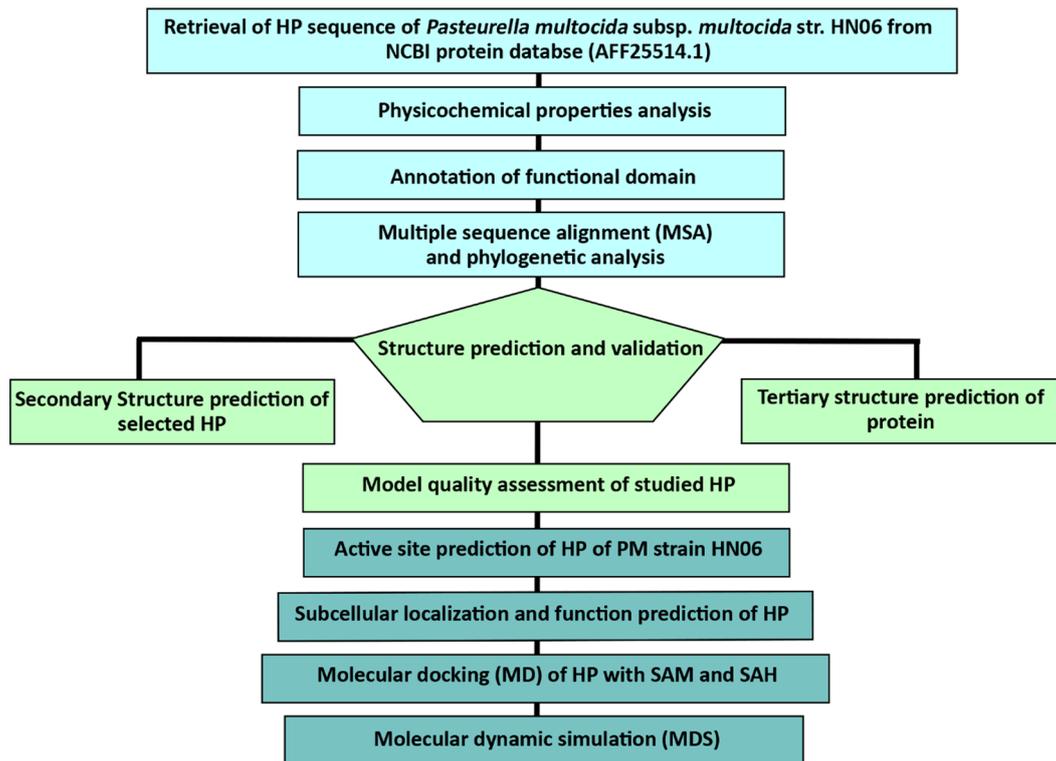


Figure 1. The study's overarching notion is shown in a flowchart. Cyan, light green, and blue boxes represent HP's sequence analysis, structural evaluation, and molecular interaction tests, respectively.

HP indicates hypothetical protein; NCBI, National Center for Biotechnology Information; PM, *Pasteurella multocida*; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine.

works in steps, was employed for the MSA analysis.⁴⁸ To further illustrate the evolutionary separation of the linked proteins, a phylogenetic tree was built by homologous sequence alignment. We used the standard settings (maximum likelihood, or ML, techniques) with 1000 replicates of the bootstrap.⁴⁹

Secondary structure prediction of selected hypothetical protein

The PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred>)^{50,51} and SOPMA (<https://npsa-prabi.ibcp.fr/cgi-bin/npsaautomat.pl?page=/NPSA/npsasopma.html>) servers were used to make predictions for the HP's secondary structure (2D). Comparatively, SOPMA predicts a protein's secondary structure by consulting the "DATABASE.DSSP," whereas the PSIPRED service employs feed-forward neural networks and the PSI-BLAST algorithm.^{50,51} Secondary structure prediction was performed in both instances using the HP's FASTA sequence.⁵²

Tertiary structure prediction of protein

The tertiary (3D) structure of the HP was predicted by the HHpred (<https://toolkit.tuebingen.mpg.de/tools/hhpred>)⁵³⁻⁵⁶ and I-TASSER (<https://zhanggroup.org/I-TASSER/>) servers.⁵⁷⁻⁵⁹ Using the MODELLER software developed at the Max Planck Institute for Developmental Biology,⁵³⁻⁵⁶ the HHpred predicts the 3D structure of a hitherto uncharacterized protein. In addition,

beginning with an amino acid sequence, I-TASSER generates 3D atomic models using multiple threading alignments and iterative structure assembly simulations.⁶⁰ For homology modeling, both HHpred and I-TASSER used their respective default values for all parameters. The 3D structures predicted by the HP were refined, and their energy was minimized using the YASARA energy minimization server⁶¹ (<http://www.yasara.org/minimizationserver.htm>). GalaxyRefine (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>)⁶² was then used to further enhance the refined 3D structures. GalaxyRefine generates several possible structures; the best quality and performance ones are hand-picked. PyMOL and BIOVIA Discovery Studio were then used to create 3D images of the HP's structures.

Model quality assessment of studied hypothetical protein

The energy-minimized and fine-tuned 3D structure of the HP was evaluated using the PROCHECK⁶³ and ERRAT⁶⁴ modules of the SAVES server (<https://saves.mbi.ucla.edu/>). The Z-score of the projected models was also forecasted using the ProSA server (<https://prosa.services.came.sbg.ac.at/prosa.php>).^{65,66} The HP 3D model was further validated using the SWISS-MODEL Structural Evaluation tool^{67,68} developed by the Swiss Institute of Bioinformatics (SIB). In the end, the highest-quality model was selected for future study.

Active site prediction of hypothetical protein of Pasteurella multocida strain HN06

The HP's active site and residues were determined with the use of the Computed Atlas of Surface Topography of Proteins (CASTp) (<http://sts.bioe.uic.edu/castp/calculation.html>)⁶⁹ and FTSite (<https://ftsitesite.bu.edu/>)⁷⁰ servers. Protein Data Bank (PDB),⁷¹ UniProt,⁷² and Structure Integration with Function, Taxonomy, and Sequence (SIFTS)⁷³ databases were also used. When a protein's structure and its sequence are correlated, as they are in the CASTp server, rapid residue-level annotations become possible.⁶⁹ The predicted active site and residues were further validated by molecular docking (MD) analysis. In addition, the docking investigation verified the anticipated active site and residues. Small organic compounds of varying sizes and polarities may bind to ligand-binding sites, as shown by the FTSite server's implementation of an algorithm verified by experimental data. Without employing evolutionary or statistical data, the program achieves near experimental accuracy.⁷⁰

Subcellular localization and function prediction of hypothetical protein

The spatial environment that governs a protein's interaction patterns and biological networks influences a protein's ability to function at its best.⁷⁴ For this context, the subcellular localization of the HP was predicted by multiple servers including PSLpred (<https://webs.iitd.edu.in/raghava/pslpred/>),⁷⁵ SOSUIGramN (https://harrier.nagahama-i-bio.ac.jp/sosui/sosuiagramn/sosuiagramn_submit.html),⁴² Gneg-PLoc (<http://www.csbio.sjtu.edu.cn/bioinf/Gneg-multi/>),^{76,77} DeepTMHMM 2.0 (<https://dtu.biolib.com/DeepTMHMM>),⁷⁸ and PSORTb (<https://www.psorth.org/psorth/>) servers.⁷⁹

Molecular docking of hypothetical protein with S-adenosylmethionine and S-adenosylhomocysteine

Molecular docking is frequently employed to investigate and evaluate the intermolecular interactions between ligands and macromolecules.⁸⁰ Hence, docking experiments were performed on the HP using both SAM and S-adenosylhomocysteine (SAH) as the ligands. The Structured Data File (SDF) formatted data files and the structures of both ligands were downloaded from the PubChem (<https://pubchem.ncbi.nlm.nih.gov/>)⁸¹ database and then converted to the PDB format using the PyMOL program. AutoDock Vina^{82,83} software and the SeamDock (<https://seamless.rpbs.univ-paris-diderot.fr/cloudless/instance/5208806/ctx/index.html>)⁸⁴⁻⁸⁶ server were then used to conduct a docking study between the HP and ligands. AutoDock Vina was used for both site-specific and blind docking, with the program being run with exhaustiveness = 24 and energy range = 4. Except for adjusting the exhaustiveness number to 24, the SeamDock server's default settings were used for the MD analysis. Even yet, this server merely underwent the blind docking method.

Molecular dynamic simulation

The stability and function of every protein complex depend on the atoms' mobility, which may be analyzed computationally using molecular dynamic simulation (MDS).⁸⁶⁻⁸⁸ For this reason, MDS was performed on the HP-ligand complexes, such as HP-SAM and HP-SAH, predicted by the AutoDock Vina, using the Internet server "WebGRO for Macromolecular Simulations" (<https://simlab.uams.edu/>).⁸⁹ The ligand topology files, which are required for the simulation run, were generated using the GlycoBioChem PRODRG2 Server (<http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrgr>).⁹⁰ Selecting "neutralize" and "add 0.15 M salt" and using the SPC⁹¹ box type of triclinic water model were other necessary parameters in addition to using the Gromos96 43a1⁹² force field on the Webgrow server. Moreover, the energy minimization settings⁹³ include a steepest descent integrator and 5000 steps. NVT/NPT (here, N-Constant number, V-Constant volume, T-Constant temperature, P-Constant pressure) equilibration, 300 K temperature, 1 bar pressure, 50 ns simulation period, and 1000 estimated frames per simulation are also recommended for MDS runs.⁹⁴ Finally, the results of the MDS analysis have been interpreted, and the stability and flexibility of the docked complexes have been assessed using metrics such as the root mean square deviation (RMSD) of the given structure over time, the root mean square fluctuation (RMSF) of each residue in the given structure, the average number of H-bonds in each frame over time, the radius of gyration (Rg) or structural compactness, and the solvent-accessible surface area (SASA).⁸⁹

Result and Discussion

Retrieval of protein sequence

The NCBI Protein database was queried at random, yielding the HP PMCN06 2293, which is the PM strain HN06 HP. The acquired sequence was then used to search UniProt, a public, free database of protein sequences and their functional annotations. For the sake of analysis, the HP's attributes have been saved. This includes the HP's locus, definition, accession, version, and version as well as the HP's total number of amino acids and FASTA sequence. There are a total of 246 amino acids in the HP, which has been labeled as PMCN06 2293 and assigned the locus, accession, and version numbers of AFF25514, AFF25514, and AFF25514.1 (Table 1).

Physicochemical properties analysis

Several physicochemical parameters of the HP PMCN06 2293 were analyzed using the ProtParam tool of the ExPASy service, and the findings are shown in Table 2. The server predicted that the HP has a 246 amino acid sequence and a molecular weight of 28352.60 Da. A theoretical pI value of -9.18 was calculated as well for the HP by the server, indicating that it is an alkaline protein with a high negative charge. Protein stability is a crucial factor in various biological processes. One way to determine the stability of a protein is by calculating its II. If the II of a protein is less than

Table 1. The properties of HP protein retrieved from NCBI protein database.

PROPERTIES	HYPOTHETICAL PROTEIN PMCN06_2293
Locus	AFF25514
Definition	Hypothetical protein PMCN06_2293 (<i>Pasteurella multocida</i> subsp. <i>multocida</i> str. HN06)
Accession	AFF25514
Version	AFF25514.1
Amino acid	246
Organism	<i>Pasteurella multocida</i> subsp. <i>multocida</i> str. HN06
FASTA sequence	>AFF25514.1 hypothetical protein PMCN06_2293 (<i>Pasteurella multocida</i> subsp. <i>multocida</i> str. HN06) MSDLSLQLHAIGIIHTPYKEKFSVPRQPNLVQDGTGILELLPPYNQAETVRGLEQFSLWLIFQFDRVATG KWRPTVRPPRLGGNQRVGVFASRSTHRPNPLGLSKVELRRVECQNGKVRLLHLGAVDLVDGTPIFDIKPYLAY ADSEPEAKSGFAQEKPCTLQVIFSEKAQNALQKIEKRRPHFKRFITEVIAQDPRPAYQKMQSLERVYGIIRLHEF NIRWKMETTEEQQARILDIEEVEKKKCD

Abbreviations: HP, hypothetical protein; NCBI, National Center for Biotechnology Information.

Table 2. The physicochemical properties of HP protein predicted by ExPASy server.

PROTPARAM PARAMETERS	VALUES
Number of amino acids	246
Molecular weight (MW)	28352.60
Theoretical pI	9.18
Total number of negatively charged residues (Asp + Glu)	31
Total number of positively charged residues (Arg + Lys)	37
Formula	$C_{1272}H_{2020}N_{362}O_{361}S_6$
Total number of atoms	4021
Extinction coefficients	25565 Abs 0.1% (=1 g/L) 0.902, assuming all pairs of Cys residues form cystines 25440 Abs 0.1% (=1 g/L) 0.897, assuming all Cys residues are reduced
Estimated half-life	30 hours (mammalian reticulocytes, in vitro) >20 hours (yeast, in vivo) >10 hours (<i>Escherichia coli</i> , in vivo)
Instability index	56.57
Aliphatic index	84
Grand average of hydropathicity (GRAVY)	-0.565

Abbreviations: HP, hypothetical protein; pI, isoelectric point.

40, it is anticipated to be stable. However, if the II is more than 40, the protein is expected to be unstable.⁹⁵ This predicts that HP is an unstable protein with a stability score of 56.57. The AI of a protein is the ratio of the volume occupied by its aliphatic side chains (alanine [Ala], valine [Val], isoleucine [Ile], and leucine [Leu]) to the overall volume of the protein.⁹⁶ Therefore, an AI of 84 is predicted for HP, indicating the protein's widened temperature stability. For each amino acid in the query sequence, its

hydropathy value is computed and then divided by the total number of residues to get the GRAVY score for the peptide or protein. The computed value for HP is -0.565, proving that it is a hydrophilic protein. According to the Beer-Lambert law, the extinction coefficient serves as a proportionality constant and measures the intensity of a certain wavelength of light absorbed by a protein.⁹⁷ Therefore, the extinction coefficient of the HP was calculated to be 25565. There are plenty of tyrosine, tryptophan, and cysteine

Table 3. The identical proteins with the HP, aligned by BLASTp algorithm, the NCBI.

ACCESSION	ORGANISM	DESCRIPTION	QUERY COVER (%)	PERCENT IDENTITY (%)	E-VALUE
WP_005754696.1	<i>Pasteurella multocida</i>	tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase	100	100.00	0.0
WP_101774491.1	<i>Pasteurella oralis</i>	tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase TrmO	97	76.67	2e-136
WP_100296032.1	<i>Cavibacterium pharyngocola</i>	tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase TrmO	96	73.53	5e-132
TCP93251.1	<i>Cricetibacter osteomyelitidis</i>	tRNA-Thr(GGU) m(6)t(6)A37 methyltransferase TsaA	96	73.95	2e-131
MBF1227901.1	<i>Haemophilus parainfluenzae</i>	tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase TrmO	99	71.84	1e-130
WP_005696892.1	<i>Haemophilus</i>	tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase TrmO	99	71.84	9e-130
MBN6711327.1	<i>Canicola haemoglobinophilus</i>	tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase TrmO	96	73.53	5e-129
WP_041639985.1	<i>Mannheimia succiniciproducens</i>	tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase TrmO	97	72.50	3e-128
WP_109128133.1	<i>Aggregatibacter segnis</i>	tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase TrmO	100	71.54	
MCI7353942.1	<i>Actinobacillus rossii</i>	tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase TrmO	97	72.08	1e-127

Abbreviations: HP, hypothetical protein; NCBI, National Center for Biotechnology Information.

around because of the high extinction coefficient.⁹⁵ However, Table 2 provides a comprehensive overview of the physicochemical properties of HP. These features will be helpful when working with the protein in future studies.

Annotation of functional domain

Predicted by the servers to be present in the HP is the well-known conserved domain of tRNA (adenine(37)-N6)-methyltransferase TrmO (Supplementary Table 1). *Escherichia coli yaeB* or tRNA (adenine(37)-N6)-methyltransferase TrmO is an SAM-dependent MTase variant that has also been identified.^{98,99} Moreover, this variant of SAM-dependent MTase has been classified as a unique AdoMet-dependent methyltransferase Class VIII.^{98,99} This organism's version of SAM-dependent MTase is responsible for the formation of N6-methyl-threonylcarbamoyl adenosine (m6t6A) by methylating t6A at position 37 of tRNA-Thr.^{98,99} It has been shown that the attenuation activity of the operon is considerably improved after N6 methylation of t6A to m6t6A, which is consistent with the effective decoding of ACC codon.^{98,99} In addition, most known MTases use SAM as a co-factor of methylation or a donor of the methyl group, which is

thereafter converted into SAH by cleavage of the CH₃ group.¹⁰⁰ Many computational methods have speculated that the HP's conserved domain functions like *E. coli*'s tRNA (adenine(37)-N6)-methyltransferase TrmO or a modified AdoMet-dependent methyltransferase of Class VIII.

Multiple sequence alignment and phylogenetic analysis

The NCBI protein database served as a BLASTp server, which returned HP values for the proteins that were found. In this instance, the software was run against a nonredundant protein database to return the microorganisms with the largest percentage of identical protein sequences, the lowest e-value, and the highest query coverage. These results suggest that the HP and tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase TrmO may have comparable purposes (Table 3). After that, the MEGA X program was used to do sequence alignment and phylogenetic tree building. For MSA and tree building, we used the MEGA X software's ClustalW algorithm and ML technique, respectively, for their iterative processes. The HP and *Pasteurella* tRNA (N6-threonylcarbamoyladenosine(37)-N6)-

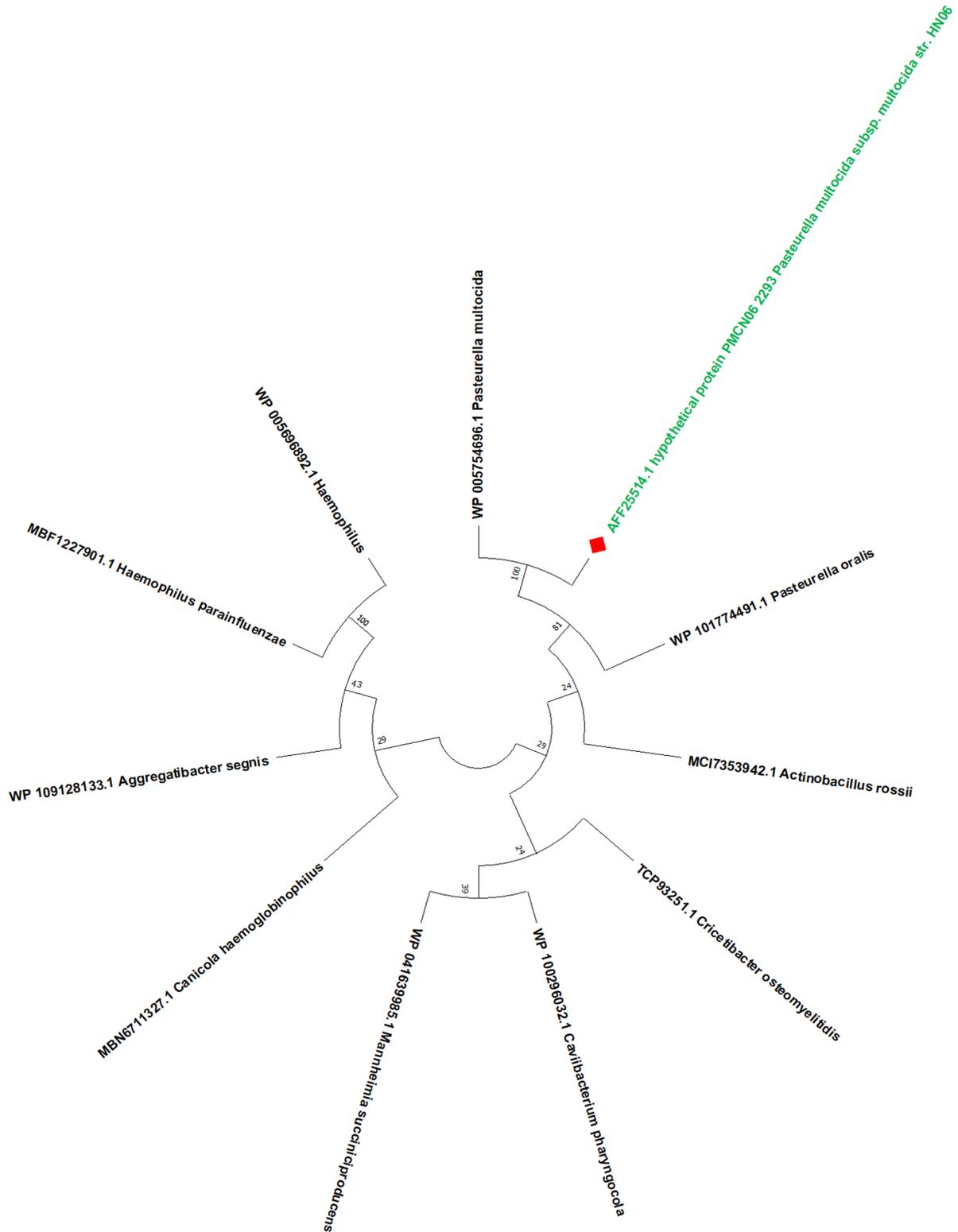


Figure 2. The evolution and ancestral relationship of the HP with the top aligned sequences. The red marked sequence represents the HP, whereas the tree nodes represent the ancestral relationship. HP indicates hypothetical protein.

methyltransferase are 100% identical in sequence, placing them in the same clade of the evolutionary tree (Figure 2). Nevertheless, HP has also revealed that *Pasteurella oralis* tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase has 81% sequence similarity with that of *Actinobacillus rossii*, the closest relative

outside of the subtree. As a tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase or Class VIII SAM-dependent MTase, HP was also shown to have sequence similarity with an unidentified ancestor's tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase (Figure 2).

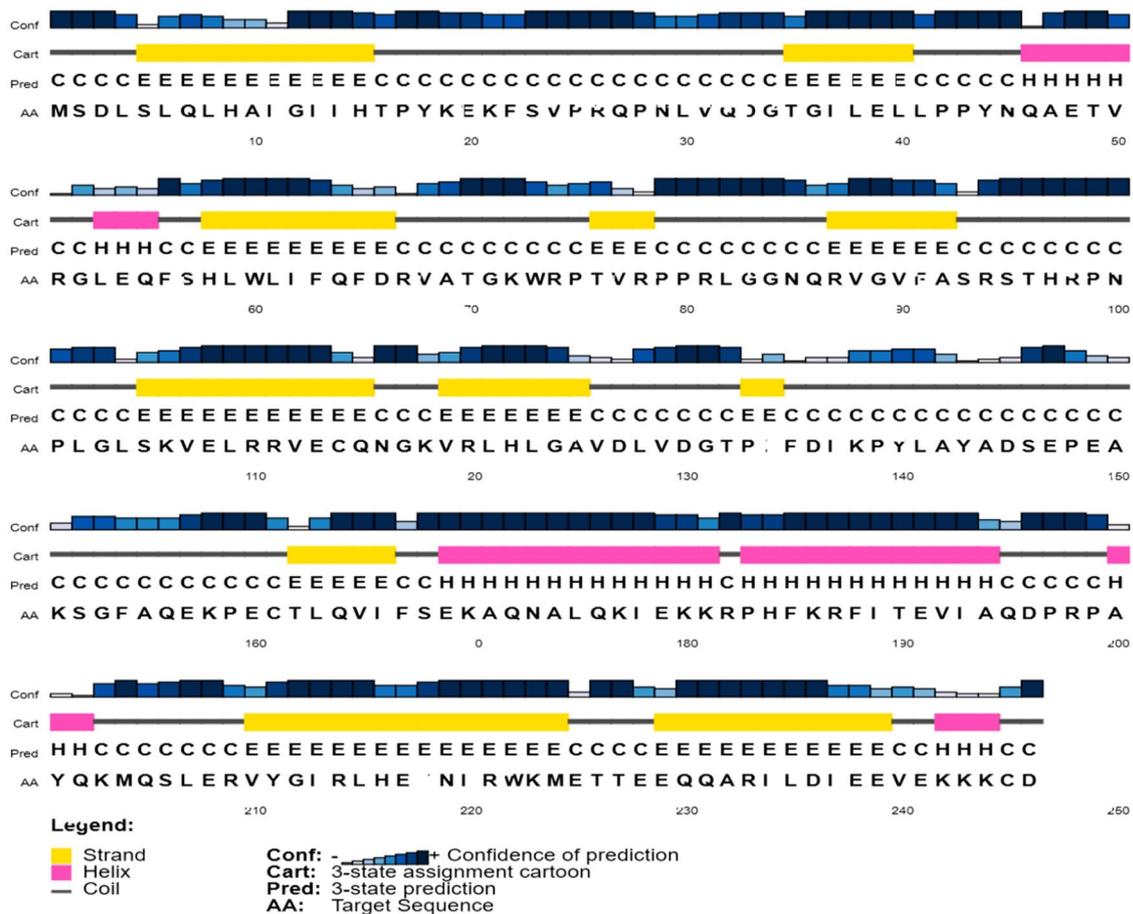


Figure 3. The secondary structure of the HP predicted by PSIPRED server. The strand, helix, and coil structures are depicted by the yellow, pink, and ash colors. HP indicates hypothetical protein.

Secondary structure prediction

The HP's secondary structure has been predicted using tools like PSIPRED and the SOPMA servers. As a quick summary, the PSIPRED server projected that the HP structure will include the most random coils, followed by prolonged strands, and finally an alpha-helix area (Figure 3). The SOPMA server agreed with the PSIPRED's assessment that the HP would have a greater proportion of random coil than extended stand or alpha helix (Table 4 and Supplementary Figure 1).

Tertiary structure prediction

For accurate HP model prediction, we used the HHpred and I-TASSER servers. The HHpred server determined an optimal 3D model of HP by comparing it to a database of known protein structures and picking a template that best fit the protein's structure. Using the criteria of a 100% success rate, an E -value of $6.7e-67$, and a secondary structure score of 28.4, the template 7BTU_B was selected as the template to aim toward (Figure 4A). In addition, I-TASSER predicted a total number of 29 models for the intended HP, and the model with a C -score of 0.00, an estimated TM-score of 0.710.11, and an estimated RMSD of 5.83.6 was selected

Table 4. The predicted secondary structure of the HP by SOPMA server.

STRUCTURAL PARAMETER	VALUE	PERCENTAGE
Alpha helix (Hh)	49	19.92
3_{10} helix (Gg)	0	0.00
Pi helix (Ii)	0	0.00
Beta bridge (Bb)	0	0.00
Extended strand (Ee)	59	23.98
Beta turn (Tt)	8	3.25
Bend region (Ss)	0	0.00
Random coil (Cc)	130	52.85
Ambiguous states (As)	0	0.00

Abbreviation: HP, hypothetical protein.

among all models predicted by I-TASSER (Figure 4B). Subsequently, the YASARA and GalaxyRefine servers have reduced and refined the anticipated models. PyMOL and BIOVIA Discovery Studio were used to examine and display the tertiary structures of the predicted and revised models.

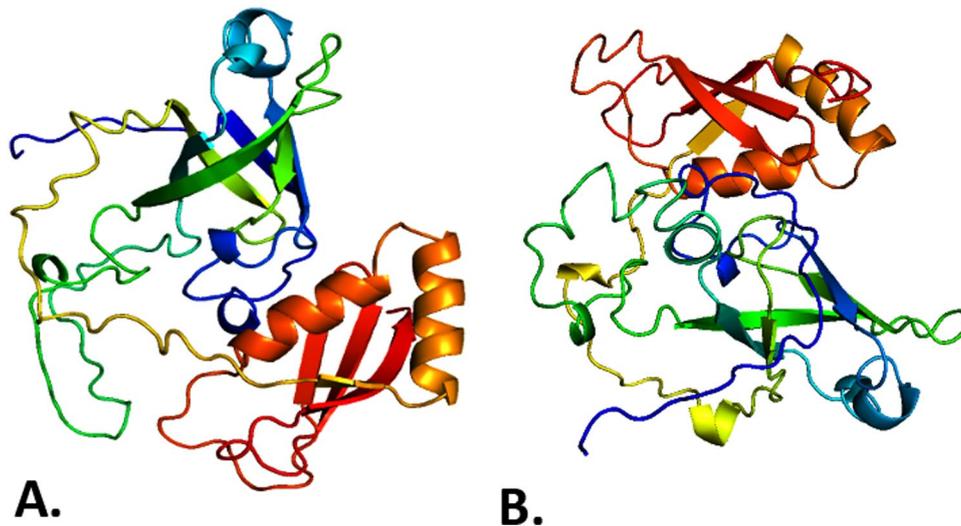


Figure 4. The tertiary structure of the HP predicted by HHpred (A) and I-TASSER (B) servers. The spiral and arrow ribbon represent alpha-helix and beta-sheet structures, whereas the line ribbon represents coil structure of the HP, respectively. HP indicates hypothetical protein.

Model quality assessment

The SAVES PROCHECK found that 89.9% of the amino acid residues in the HHpred-predicted model of the HP were located in the Ramachandran preferred area, but only 84.5% of the residues in the I-TASSER referenced model were located there (Figure 5). The ERRAT score is likewise greater in the HHpred-predicted model (87.5) compared with the I-TASSER-predicted model (85.3211) (Table 5 and Supplementary Figure 2). Both HHpred and I-TASSER provide a negative value for the HP model's projected Z-score: -4.76 and -4.71 (Table 5 and Supplementary Figure 3). The SWISS-MODEL predicts that the HP created by HHpred has a MolProbity score of 3.64, a Ramachandran preferred area of 90.76%, a QMEAN of -3.64, and a QMEANDisCo Global of 0.61 0.05. The server also came up with a MolProbity score of 3.64, a Ramachandran preferred area of 90.57%, a Qualitative model energy analysis (QMEAN) of -3.64, and a Qualitative model energy analysis-distance constraint (QMEANDisCo) Global value of 0.62 0.05 for the I-TASSER projected model (Table 5). We analyzed each anticipated model's structure and settled on the HHpred model for further study.

Active site prediction

The CASTp server predicted a total number of 75 amino acid residues within the active site of HP. However, the active site has been predicted to be covered a total surface area and surface volume of 1811.175 and 2510.612 Å², respectively (Figure 6A). In the meantime, the FTSite predicted 37 active amino acid residues within the active site of HP (Figure 6B). However, there are 27 common active amino acid residues reported from the servers including Lys-21, Phe-22, Ser-23, Val-24, Pro-25, Arg-26, Pro-28, Phe-63, Gln-64, Phe-65, Asp-66, Arg-94,

Thr-96, Gly-103, Leu-104, Ser-105, Asp-127, Leu-128, Val-129, Thr-132, Gln-195, Asp-196, Pro-197, Arg-198, Pro-199, Ala-200, and Tyr-201 (Figure 6C).

Prediction of subcellular localization

Numerous servers—such as PSLpred, SOSUIGramN, Gneg-PLoc, DeepTMHMM 2.0, and PSORTb—have made predictions on where in the cell the HP will be found. Different cellular locations are linked to various biological processes,¹⁰¹ therefore knowing where an HP is found inside the cell might provide light on its potential role. This knowledge might be useful in creating a medication that inhibits the functioning of the targeted protein.¹⁰¹ As a result, the authors hypothesized the HP is a cytoplasmic protein with comparable functions to other cytoplasmic proteins (Supplementary Table 2).

Molecular docking analysis

The MD study showed that the HP and ligands had several intermolecular interactions (SAM and SAH). Docking scores of -7.4 and 7.5 (kcal/mol) for the HP indicate that SAM and SAH, 2 ligands, have a strong affinity for the HP in site-specific docking (AutoDock Vina) (Table 6 and Figure 7A and B). With a docking score of -7.7 (kcal/mol), both SAM and SAH showed strong attraction for HP in blind docking (Table 6 and Figure 7C and D). Site-specific docking, however, reveals that the HP-SAM and HP-SAH-docked complexes include 16 and 19 interacting amino acid residues of the HP, respectively (Table 6 and Figure 8A and B). The HP-SAM- and HP-SAH-docked complexes have 6 conventional hydrogen bonds. The HP-SAM had 8 van der Waals and 2 carbon-hydrogen bonds, whereas the HP-SAH-docked complexes had 9 van der Waals and 3 carbon-hydrogen bonds. Hydrogen bonds are a vital aspect in determining the specificity of ligand binding. In

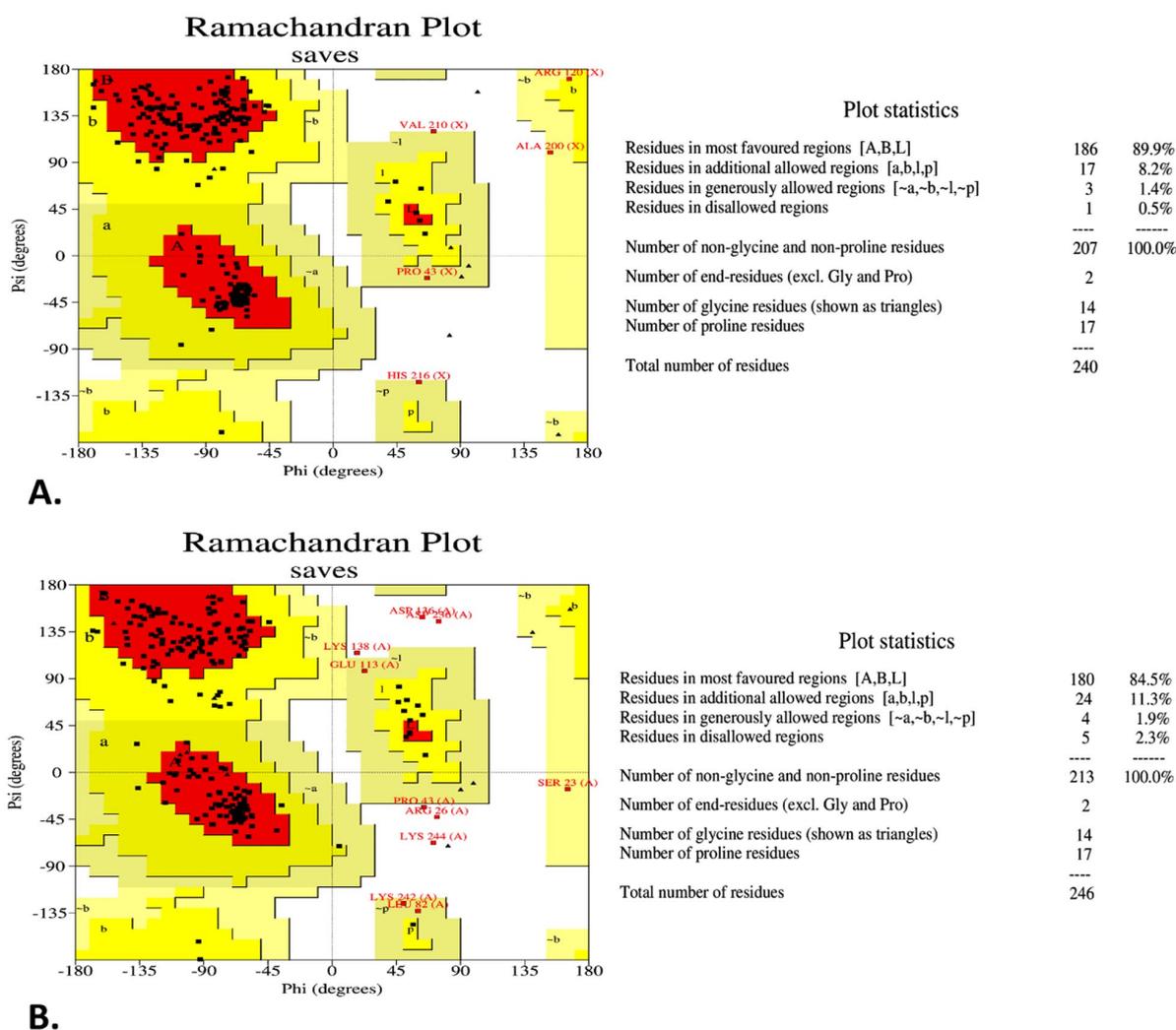


Figure 5. The Ramachandran plot of the predicted models by HHpred (A) and I-TASSER (B) server. The first represents the tertiary structure of the HP such as the beta-sheet region, where second and third quadrants represent the right-handed and the left-handed alpha-helix region, respectively. In addition, the red, yellow, gray, and white color regions depict the residues in most favored, additional allowed, generously allowed, and disallowed region, respectively. HP indicates hypothetical protein.

Table 5. The model quality assessment of the HP by SAVES, ProSA, and SWISS-MODEL structural assessment server.

SERVER	SAVES		PROSA	SWISS-MODEL STRUCTURE ASSESSMENT			
	PROCHECK (RAMACHANDRAN-FAVORED REGION)	ERRAT		MOLPROBITY SCORE	RAMACHANDRAN FAVORED REGION	QMEAN	QMEANDISCO GLOBAL
HHpred	89.9%	87.5	-4.76	3.64	90.76%	-3.64	0.61 ± 0.05
I-TASSER	84.5%	85.3211	-4.71	2.19	90.57%	-3.64	0.62 ± 0.05

Abbreviation: HP, hypothetical protein.

addition, blind docking showed that the HP has 18 interacting amino acid residues within the HP-SAM complex and 19 interacting amino acid residues within the HP-SAH complex (Table 6 and Figure 8C and D). There are a total of 6 conventional hydrogen bonds in the docked complexes of HP-SAM and HP-SAH. The HP-SAM contained 11 van der Waals and 1 carbon-hydrogen bonds, whereas HP-SAH docked

complexes had 8 van der Waals and 11 carbon-hydrogen bonds. Notably, the amino acid residues LYS-21, PHE-22, VAL-24, GLN-195, and ASP-196 are all documented in both site-specific and blind docking to interact with the SAM and SAH. Docking scores of -7.1 (kcal/mol) and -7.3 (kcal/mol) were obtained from the SeamDock server for the HP-SAM and HP-SAH complexes, respectively, validating the predictions of

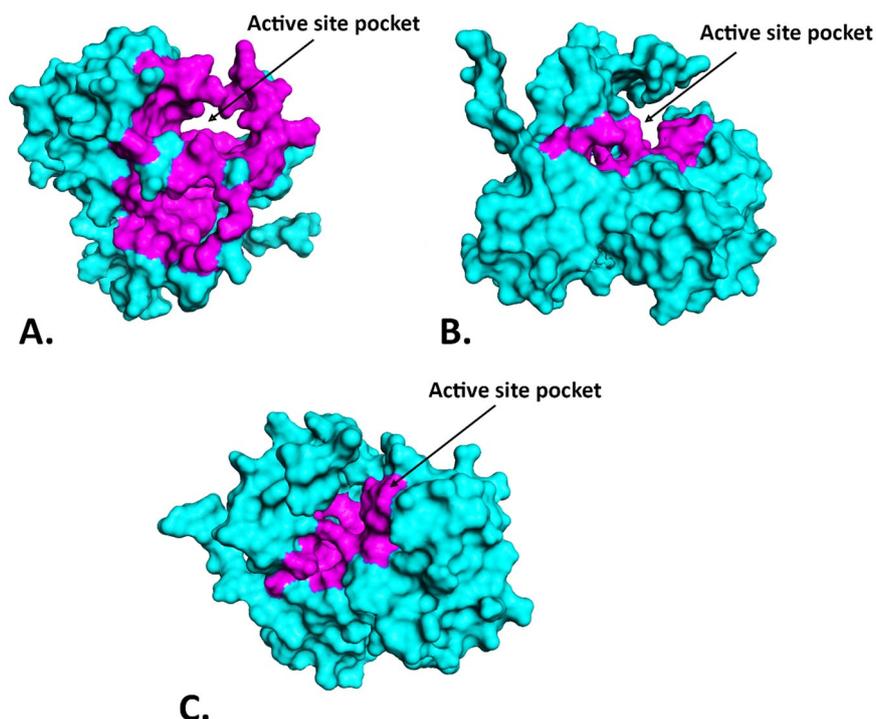


Figure 6. The predicted active sites and active amino acid residues by CASTp (A) and FTSite (B) server and common active residues (C) from these servers. The cyan color denotes the protein, whereas the purple color indicates the active amino acid residues. CASTp indicates Computed Atlas of Surface Topography of Proteins.

Table 6. The MD analysis of the HP with the SAM and SAH.

COMPLEX		AUTODOCK VINA			SEAMDOCK	
		DOCKING SCORE (KCAL/MOL)	RMSD	INTERACTING RESIDUES	DOCKING SCORE (KCAL/MOL)	INTERACTING RESIDUES
Site-specific docking	HP-SAM	-7.4	0.0	LYS-21, PHE-22, VAL-24, ARG-26, LYS-158, PRO-159, CYS-161, THR-162, LEU-163, GLN-195, ASP-196, PRO-197, ARG-198, PRO-199, ALA-200, and TYR-201		
	HP-SAH	-7.5	0.0	LYS-21, PHE-22, VAL-24, PRO-25, ARG-26, LYS-158, PRO-159, CYS-161, THR-162, LEU-163, ALA-194, GLN-195, ASP-196, PRO-197, ARG-198, PRO-199, ALA-200, TYR-201, and GLN-202		
Blind docking	HP-SAM	-7.7	0.0	LYS-21, PHE-22, VAL-24, PRO-25, ARG-26, LYS-158, PRO-159, CYS-161, THR-162, LEU-163, ALA-194, GLN-195, ASP-196, PRO-197, ARG-198, PRO-199, ALA-200, and TYR-201	-7.1	LYS-21, PHE-22, VAL-24, ARG-94, THR-96, GLN-195, and ASP-196
	HP-SAH	-7.7	0.0	LYS-21, PHE-22, VAL-24, PRO-25, ARG-26, GLN-27, LYS-158, PRO-159, CYS-161, THR-162, LEU-163, ALA-194, GLN-195, ASP-196, PRO-197, ARG-198, ALA-200, and TYR-201	-7.3	PHE-22, ARG-26, PRO-159, GLN-195, and TYR-201

Abbreviations: HP, hypothetical protein; MD, molecular docking; RMSD, root mean square deviation; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine.

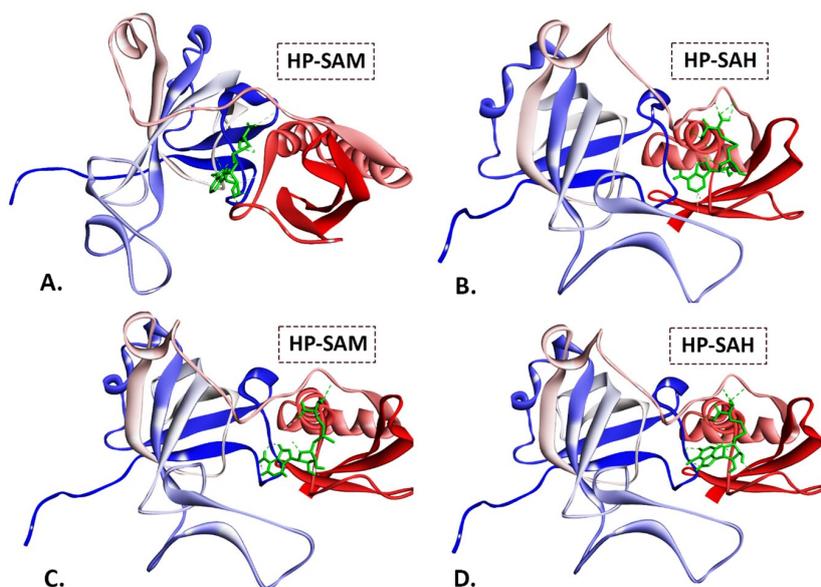


Figure 7. The molecular docking analysis of the HP with the SAM and SAH. The figure depicted both the site-specific (A and B) and blind docking (C and D) studies, where the ribbon indicates the HP and the sticks indicate the ligand (green color). HP indicates hypothetical protein; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine.

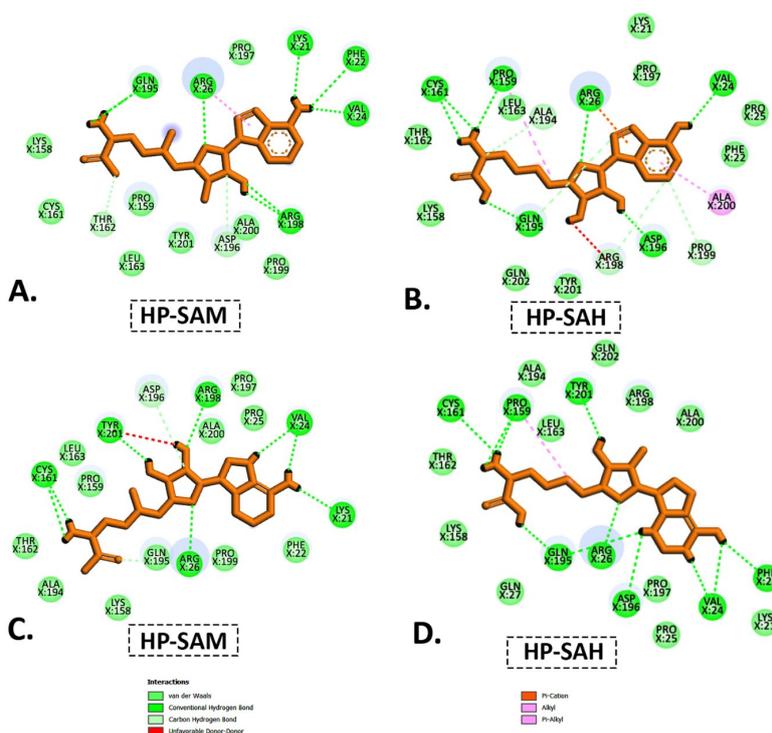


Figure 8. The interacting amino acid residues of the HP-ligand complexes, including HP-SAM (A and C) (site-specific and blind) and HP-SAH (B and D) (site-specific and blind) complexes predicted by AutoDock Vina software. The yellow color sticks depicted the ligands, whereas the disk represents the interacting amino acids.

HP indicates hypothetical protein; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine.

AutoDock Vina (Table 6 and Supplementary Figure 4). Results from the functional domain and MSA analyses suggested that the HP may act as a variant of SAM-dependent MTase; this hypothesis was confirmed by the following docking study. Therefore, we decided to conduct our molecular dynamic simulation research on docked complexes generated using site-specific docking (AutoDock Vina).

Molecular dynamic simulation

The stability and performance of the docked protein complexes have been assessed using an MDS to investigate the atomic dynamic movements inside the complexes. Using a time-dependent MDS at 50 ns with the Gromacs forcefield on the Webgrow server, we have assessed the anticipated stability and flexibility of docked complexes such as HP-SAM and HP-SAH

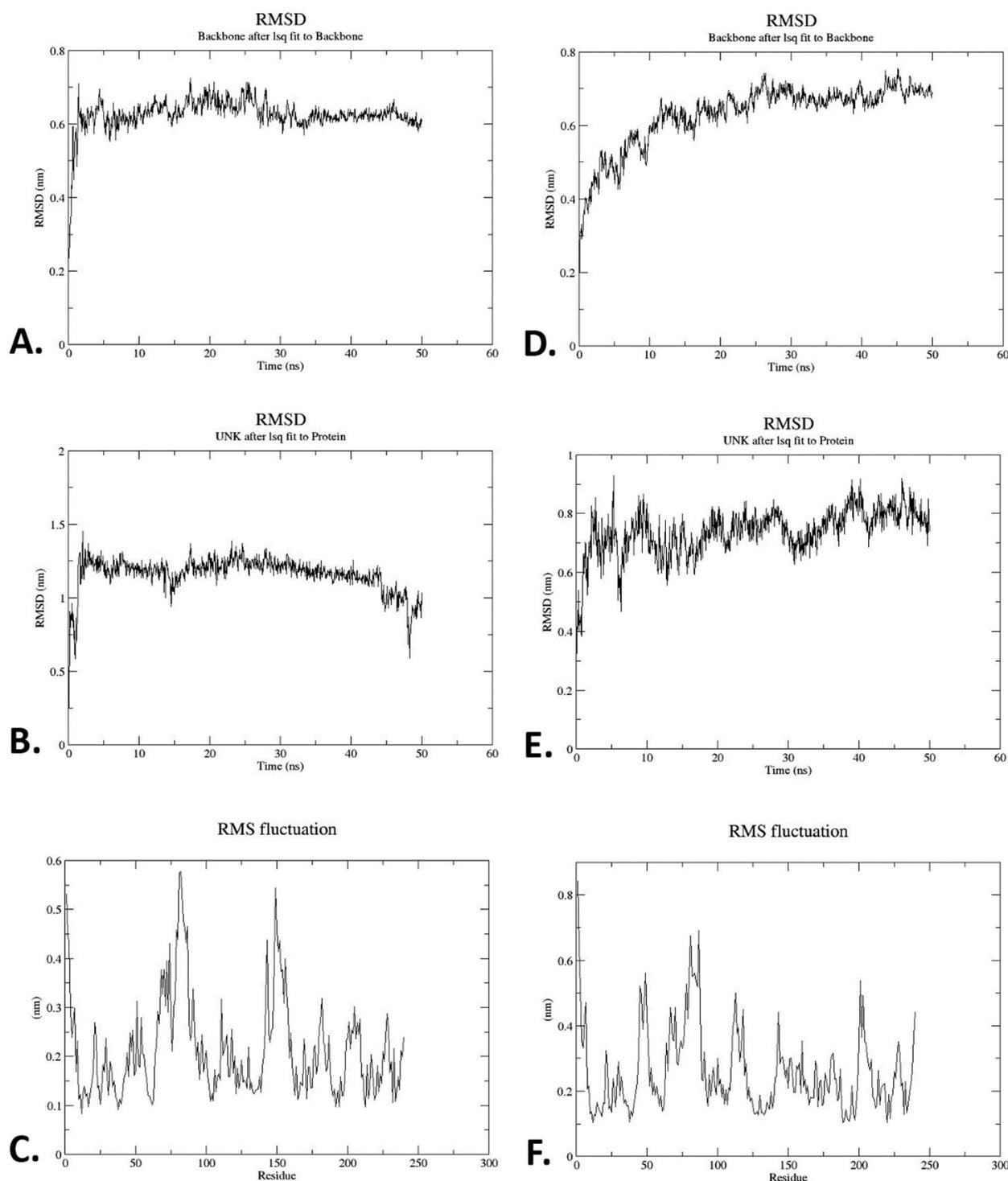


Figure 9. Molecular dynamic (MD) study of the HP-ligand complexes. The RMSD and RMSF of the HP-SAM (A, B, and C) and HP-SAH (D, E, and F) complexes were depicted as 50 ns run and up to 246 amino acid residues, respectively. HP indicates hypothetical protein; RMSD, root mean square deviation; RMSF, root mean square fluctuation; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine.

generated by AutoDock Vina. The RMSD and RMSF plots have been used to evaluate the complexes' residual fluctuations and changes. To assess the equilibrium and stability of the HP-SAM and HP-SAH complexes, we calculated their average potential. It has been calculated that the average potential energy of the HP-SAM is $-25\,4405$ kJ/mol, whereas that of the HP-SAH is $-25\,4919$ kJ/mol (Supplementary Figure 5). The

root mean square error, Rg, SASA, kinetic energy, enthalpy, volume, and density were all reported throughout the simulation. Changes in protein structure may be evaluated using RMSD by looking at how far C atoms deviate from the average orientation (Figure 9A to D). The average RMSF of all residues has also been counted to assess the local structural flexibility of the HP-SAM and HP-SAH (Figure 9E and F). Because of the

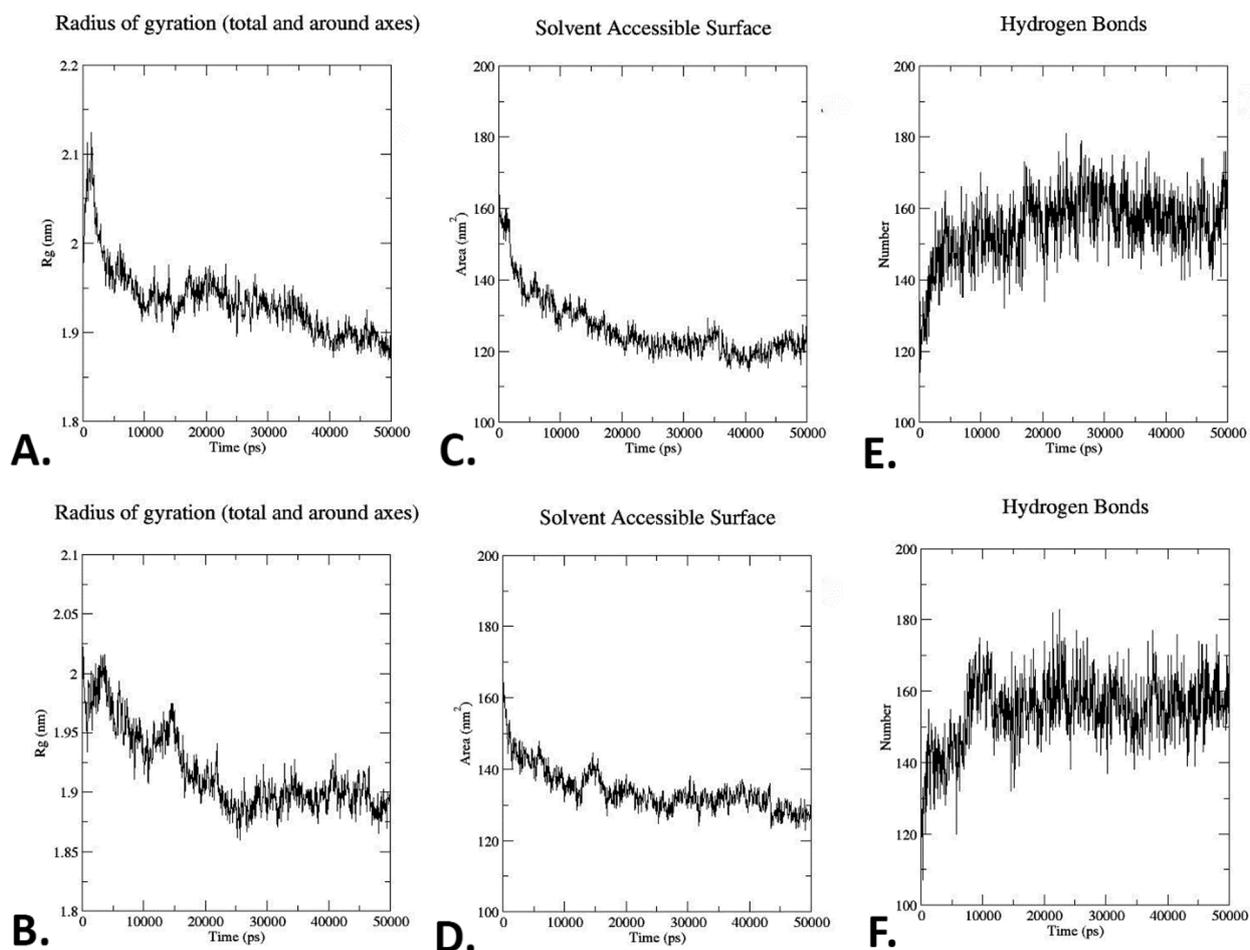


Figure 10. The Rg, SASA, and hydrogen bond analysis of the HP-SAM (A, C, and E) and HP-SAH (B, D, and E) docked complexes by MD simulation. HP indicates hypothetical protein; MD, molecular docking; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine; SASA, solvent-accessible surface area.

correlation between a protein's Rg and its SASA, the Rg of the HP-SAM and HP-SAH has been determined in the context of structural compactness evaluation (Figures 9C and D and 10A and B). The SASA of the HP-SAM is lower than that of the HP-SAH up to 50 ns. Structural stability has also been predicted for the HP-SAM and HP-SAH, based on their average intramolecular hydrogen bonds (Figure 10E and F). For energy, it is estimated that the HP-SAM has an average kinetic energy of 52481.3 kJ/mol and an average enthalpy of -609650 kJ/mol. The average kinetic energy and enthalpy of the HP-SAH, however, are much higher than those of the original SAH, coming in at 52575.6 and -610548 kJ/mol, respectively. The subsequent analysis of the docked complexes, HP-SAM and HP-SAH, revealed their stability and flexibility through the parameters such as RMSD, RMSF, Rg, SASA, and hydrogen bond analysis. The graphical depicts of all these parameters conveyed that the docking complexes are well stable and flexible, which imparts the HP to be a probable SAM-dependent MTase as well.

Conclusions

It has been established that the HP of PM strain HN06 is a valuable and stable protein, and one of the protein's functional domains is tRNA (adenine(37)-N6)-methyltransferase TrmO.

Surprisingly, the HP is an essential component in preventing the spread of pasteurellosis as it is a modified form of SAM-dependent MTase, namely, Class VIII SAM-dependent MTase. The biocomputational examination, in particular by MD and simulation studies, established the HP to be a Class VIII SAM-dependent MTase. It is possible to draw the conclusion that HP has great potential to progress research on *Pasteurella* infection, for example, by creating medications to treat this particular illness. We advise additional research into the protein's comprehensive characterization, in vitro and in vivo assessment to assess its potential as a new *Pasteurella* infection research tool.

Acknowledgements

The authors acknowledge the Department of Microbiology, Noakhali Science and Technology University and Research Cell, NSTU for providing the research facilities and fundings respectively.

Author Contributions

MHUM, OS, SR, and UPB carried out the studies (data collection, curation, molecular, and data analysis) and participated in drafting the manuscript. MHUM, MSA, MRA, TAS, JF,

FA, and MMR critically reviewed and drafted the manuscript. MHUM and OS visualized figures, interpreted data and results, and critically reviewed and edited the manuscript. OS developed the hypothesis, supervised the whole work, and helped to prepare and critically revise the manuscript. All authors read and approved the final manuscript

Data Availability

No data were used to support this study.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- Rabbi MF, Akter SA, Hasan MJ, Amin A. In silico characterization of a hypothetical protein from *Shigella dysenteriae* ATCC 12039 reveals a pathogenesis-related protein of the type-VI secretion system. *Bioinform Biol Insights*. 2021;15:11779322211011140. doi:10.1177/11779322211011140
- Shahbaaz M, Bisetty K, Ahmad F, Hassan MI. Current advances in the identification and characterization of putative drug and vaccine targets in the bacterial genomes. *Curr Top Med Chem*. 2016;16:1040-1069. doi:10.2174/1568026615666150825143307
- Nimrod G, Schushan M, Steinberg DM, Ben-Tal N. Detection of functionally important regions in “hypothetical proteins” of known structure. *Structure*. 2008;16:1755-1763. doi:10.1016/j.str.2008.10.017
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA*. 1999;96:4285-4288. doi:10.1073/pnas.96.8.4285
- Idrees S, Nadeem S, Kanwal S, et al. In silico sequence analysis, homology modeling and function annotation of *Ocimum basilicum* hypothetical protein G1CT28_OCIBA. *Int J Bioaut* 2012;16:111-118.
- Lubec G, Afjehi-Sadat L, Yang JW, John JP. Searching for hypothetical proteins: theory and practice based upon original data and literature. *Prog Neurobiol*. 2005;77:90-127. doi:10.1016/j.pneurobio.2005.10.001
- Hasan R, Rony MNH, Ahmed R. In silico characterization and structural modeling of bacterial metalloprotease of family M4. *J Gen Eng Biotech*. 2021;19:25. doi:10.1186/s43141-020-00105-y
- Naqvi AA, Anjum F, Khan FI, Islam A, Ahmad F, Hassan MI. Sequence analysis of hypothetical proteins from *Helicobacter pylori* 26695 to identify potential virulence factors. *Genomics Inform*. 2016;14:125-135. doi:10.5808/gi.2016.14.3.125
- Shahbaaz M, Hassan MI, Ahmad F. Functional annotation of conserved hypothetical proteins from *Haemophilus influenzae* Rd KW20. *PLoS ONE*. 2013;8:e84263. doi:10.1371/journal.pone.0084263
- Yang Z, Zeng X, Tsui SK. Investigating function roles of hypothetical proteins encoded by the *Mycobacterium tuberculosis* H37Rv genome. *BMC Genom*. 2019;20:394. doi:10.1186/s12864-019-5746-6
- Harper M, Boyce JD, Adler B. *Pasteurella multocida* pathogenesis: 125 years after Pasteur. *FEMS Microbiol Lett*. 2006;265:1-10. doi:10.1111/j.1574-6968.2006.00442.x
- Oehler RL, Velez AP, Mizrachi M, Lamarche J, Gompf S. Bite-related and septic syndromes caused by cats and dogs. *Lancet Infect Dis*. 2009;9:439-447. doi:10.1016/s1473-3099(09)70110-0
- Desai SS, Groves RJ, Glew R. Subacute *Pasteurella osteomyelitis* of the hand following dog bite. *Orthopedics*. 1990;13:653-656. doi:10.3928/0147-7447-19900601-09
- Rodríguez-Escot C, Hernández Medina E, Santana-Cabrera L, Sánchez-Palacios M. Severe *Pasteurella multocida* infection after a dog bite. *J Emerg Med*. 2012;43:717-718. doi:10.1016/j.jemermed.2011.07.041
- Wilson BA, Ho M. *Pasteurella multocida*: from zoonosis to cellular microbiology. *Clin Microbiol Rev*. 2013;26:631-655. doi:10.1128/cmr.00024-13
- Dendle C, Looker D. Review article: animal bites: an update for management with a focus on infections. *Emerg Med Australas*. 2008;20:458-467. doi:10.1111/j.1742-6723.2008.01130.x
- Griego RD, Rosen T, Orengo IF, Wolf JE. Dog, cat, and human bites: a review. *J Am Acad Dermatol*. 1995;33:1019-1029. doi:10.1016/0190-9622(95)90296-1
- Goldstein EJ. Bite wounds and infection. *Clin Infect Dis*. 1992;14:633-638. doi:10.1093/clinids/14.3.633
- Donnio PY, Le Goff C, Avril JL, Pouedras P, Gras-Rouzet S. *Pasteurella multocida*: oropharyngeal carriage and antibody response in breeders. *Vet Res*. 1994;25:8-15.
- Kawamoto E, Sawada T, Maruyama T. Prevalence and characterization of *Pasteurella multocida* in rabbits and their environment in Japan. *Nihon Juigaku Zasshi*. 1990;52:915-921. doi:10.1292/jvms1939.52.915
- DiGiacomo RF, Garlinghouse LE Jr, Van Hoosier GL Jr. Natural history of infection with *Pasteurella multocida* in rabbits. *J Am Veter Med Assoc*. 1983;183:1172-1175.
- Taylor JD, Fulton RW, Dabo SM, Lehenbauer TW, Confer AW. Comparison of genotypic and phenotypic characterization methods for *Pasteurella multocida* isolates from fatal cases of bovine respiratory disease. *J Vet Diagn Invest*. 2010;22:366-375. doi:10.1177/104063871002200304
- Dabo SM, Taylor JD, Confer AW. *Pasteurella multocida* and bovine respiratory disease. *Anim Heal Res Rev*. 2007;8:129-150. doi:10.1017/s1466252307001399
- Kumar AA, Shivachandra SB, Biswas A, Singh VP, Singh VP, Srivastava SK. Prevalent serotypes of *Pasteurella multocida* isolated from different animal and avian species in India. *Vet Res Commun*. 2004;28:657-667.
- Balsrud KJ. Atrophic rhinitis in goats in Norway. *Veter Rec*. 1987;121:350-353. doi:10.1136/vr.121.15.350
- Shafarin MS, Zamri-Saad M, Khairani BS, Saharee AA. Proliferation and transmission patterns of *Pasteurella multocida* B:2 in goats. *Trop Anim Health Prod*. 2008;40:335-340. doi:10.1007/s11250-007-9111-4
- Soriano-Vargas E, Vega-Sánchez V, Zamora-Espinosa JL, Acosta-Dibarrat J, Aguilar-Romero F, Negrete-Abascal E. Identification of *Pasteurella multocida* capsular types isolated from rabbits and other domestic animals in Mexico with respiratory diseases. *Trop Anim Health Prod*. 2012;44:935-937. doi:10.1007/s11250-011-9995-x
- Dyer NW, Ward AC, Weiser GC, White DG. Seasonal incidence and antibiotic susceptibility patterns of Pasteurellaceae isolated from American bison (*Bison bison*). *Can J Vet Res*. 2001;65:7-14.
- Cowart RP, Bäckström L, Brim TA. *Pasteurella multocida* and *Bordetella bronchiseptica* in atrophic rhinitis and pneumonia in swine. *Can J Vet Res*. 1989;53:295-300.
- Rhodes MB, New CW Jr, Baker PK, Hogg A, Underdahl NR. *Bordetella bronchiseptica* and toxigenic type D *Pasteurella multocida* as agents of severe atrophic rhinitis of swine. *Vet Microbiol*. 1987;13:179-187. doi:10.1016/0378-1135(87)90043-5
- Kielstein P. On the occurrence of toxin-producing *Pasteurella-multocida*-strains in atrophic rhinitis and in pneumonias of swine and cattle. *Zentralbl Veterinarmed B*. 1986;33:418-424. doi:10.1111/j.1439-0450.1986.tb00052.x
- Hansen MJ, Bertelsen MF, Christensen H, Bisgaard M, Bojesen AM. Occurrence of Pasteurellaceae bacteria in the oral cavity of selected marine mammal species. *J Zoo Wildl Med*. 2012;43:828-835. doi:10.1638/2011-0264r1.1
- Köndgen S, Leider M, Lankester F, et al. *Pasteurella multocida* involved in respiratory disease of wild chimpanzees. *PLoS ONE*. 2011;6:e24236. doi:10.1371/journal.pone.0024236
- Bull JJ, Jessop TS, Whiteley M. Deathly drool: evolutionary and ecological basis of septic bacteria in Komodo dragon mouths. *PLoS ONE*. 2010;5:e11097. doi:10.1371/journal.pone.0011097
- Montgomery JM, Gillespie D, Sastrawan P, Fredeking TM, Stewart GL. Aerobic salivary bacteria in wild and captive Komodo dragons. *J Wildl Dis*. 2002;38:545-551. doi:10.7589/0090-3558-38.3.545
- Klein NC, Cunha BA. *Pasteurella multocida* pneumonia. *Sem Resp Inf*. 1997;12:54-56.
- Talan DA, Citron DM, Abrahamian FM, Moran GJ, Goldstein EJ. Bacteriologic analysis of infected dog and cat bites. Emergency Medicine Animal Bite Infection Study Group. *New Engl J Med*. 1999;340:85-92. doi:10.1056/nejm199901143400202
- Myers EM, Ward SL, Myers JP. Life-threatening respiratory pasteurellosis associated with palliative pet care. *Clin Infect Dis*. 2012;54:e55-e57. doi:10.1093/cid/cir975
- Martin TCS, Abdelmalek J, Yee B, Lavergne S, Ritter M. *Pasteurella multocida* line infection: a case report and review of literature. *BMC Inf Dis*. 2018;18:420. doi:10.1186/s12879-018-3329-9
- Körmöndi S, Terhes G, Pál Z, et al. Human pasteurellosis health risk for elderly persons living with companion animals. *Emerg Infect Dis*. 2019;25:229-235. doi:10.3201/eid2502.180641
- Giordano A, Dincman T, Clyburn BE, Steed LL, Rockey DC. Clinical features and outcomes of *Pasteurella multocida* infection. *Medicine (Baltimore)*. 2015;94:e1285. doi:10.1097/md.0000000000001285
- Imai K, Asakawa N, Tsuji T, et al. SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in gram-negative bacteria. *Bioinformatics*. 2008;24:417-421. doi:10.6026/97320630002417

43. Marchler-Bauer A, Bo Y, Han L, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 2017;45:D200–D203. doi:10.1093/nar/gkw1129
44. Quevillon E, Silventoinen V, Pillai S, et al. InterProScan: protein domains identifier. *Nucleic Acids Res.* 2005;33:W116–W120. doi:10.1093/nar/gki442
45. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol.* 2001;313:903–919. doi:10.1006/jmbi.2001.5080
46. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008;36:W5–W9. doi:10.1093/nar/gkn201
47. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* 2021;38:3022–3027. doi:10.1093/molbev/msab120
48. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23:2947–2948. doi:10.1093/bioinformatics/btm404
49. Simpson MG. 2—phylogenetic systematics. In: Simpson MG, ed. *Plant Systematic*. 2nd ed. Academic Press; 2010:17–52.
50. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 1999;292:195–202. doi:10.1006/jmbi.1999.3091
51. Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. *Nucl Acids Res.* 2019;47:W402–W407. doi:10.1093/nar/gkz297
52. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci.* 1995;11:681–684. doi:10.1093/bioinformatics/11.6.681
53. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucl Acids Res.* 2005;33:W244–W248. doi:10.1093/nar/gki408
54. Zimmermann L, Stephens A, Nam S-Z, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol.* 2018;430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007
55. Hildebrand A, Remmert M, Biegert A, Söding J. Fast and accurate automatic structure prediction with HHpred. *Proteins.* 2009;77:128–132. doi:10.1002/prot.22499
56. Meier A, Söding J. Automatic prediction of protein 3D structures by probabilistic multi-template homology modeling. *PLoS Comput Biol.* 2015;11:e1004343. doi:10.1371/journal.pcbi.1004343
57. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 2015;43:W174–W181. doi:10.1093/nar/gkv342
58. Zheng W, Zhang C, Li Y, Pearce R, Bell EW, Zhang Y. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell Rep Methods.* 2021;1:100014. doi:10.1016/j.crmeth.2021.100014
59. Zhang C, Freddolino PL, Zhang Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucl Acids Res.* 2017;45:W291–W299. doi:10.1093/nar/gkx366
60. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010;5:725–738. doi:10.1038/nprot.2010.5
61. Krieger E, Joo K, Lee J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins.* 2009;77:114–122. doi:10.1002/prot.22570
62. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* 2013;41:W384–W388. doi:10.1093/nar/gkt458
63. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR.* 1996;8:477–486. doi:10.1007/bf00228148
64. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 1993;2:1511–1519. doi:10.1002/pro.5560020916
65. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007;35:W407–W410. doi:10.1093/nar/gkm290
66. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins.* 1993;17:355–362. doi:10.1002/prot.340170404
67. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics.* 2019;36:1765–1771. doi:10.1093/bioinformatics/btz828
68. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 2009;37:W510–W514. doi:10.1093/nar/gkp322
69. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* 2018;46:W363–W367. doi:10.1093/nar/gky473
70. Ngan C-H, Hall DR, Zerbe B, Grove LE, Kozakov D, Vajda S. FTSite: high accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics.* 2011;28:286–287. doi:10.1093/bioinformatics/btr651
71. Örmö M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* 1996;273:1392–1395. doi:10.1126/science.273.5280.1392
72. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2016;45:D158–D169. doi:10.1093/nar/gkw1099
73. Dana JM, Gutmanas A, Tyagi N, et al. SIFTS: updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. *Nucleic Acids Res.* 2018;47:D482–D489. doi:10.1093/nar/gky1114
74. Dönnies P, Höglund A. Predicting protein subcellular localization: past, present, and future. *Genom Proteom Bioinform.* 2004;2:209–215. doi:10.1016/s1672-0229(04)02027-3
75. Bhasin M, Garg A, Raghava GP. PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics.* 2005;21:2522–2524. doi:10.1093/bioinformatics/bti309
76. Xiao X, Cheng X, Chen G, Mao Q, Chou KC. PLoc_bal-mGpos: predict subcellular localization of Gram-positive bacterial proteins by quasi-balancing training dataset and PscAAC. *Genomics.* 2019;111:886–892. https://doi.org/10.1016/j.ygeno.2018.05.017
77. Shen H-B, Chou K-C. Gneg-mPLOC: a top-down strategy to enhance the quality of predicting subcellular localization of Gram-negative bacterial proteins. *J Theor Biol.* 2010;264:326–333. https://doi.org/10.1016/j.jtbi.2010.01.018
78. Hallgren J, Tsirigos KD, Pedersen MD, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. 2022:2022.04.08.487609. doi:10.1101/2022.04.08.487609
79. Yu NY, Wagner JR, Laird MR, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics.* 2010;26:1608–1615. doi:10.1093/bioinformatics/btq249
80. Aaftaab S, Khusbhoo J, Sasikala K, Mallika A. Molecular docking in modern drug discovery: principles and recent applications. In: Vishwanath G, Partha K, Ashit T, eds. *Drug Discovery and Development*. IntechOpen; 2019:pCh3.
81. Kim S, Thiessen PA, Bolton EE, et al. PubChem substance and compound databases. *Nucleic Acids Res.* 2016;44:D1202–D1213. doi:10.1093/nar/gkv951
82. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. *J Chem Inf Model.* 2021;61:3891–3898. doi:10.1021/acs.jcim.1c00203
83. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31:455–461. doi:10.1002/jcc.21334
84. Murail S, de Vries SJ, Rey J, Moroy G, Tufféry P. SeamDock: an interactive and collaborative online docking resource to assist small compound molecular docking. *Front Mol Biosci.* 2021;8:716466. doi:10.3389/fmolb.2021.716466
85. samuelmurail/docking_py: Docking_py, a python library for ligand protein docking [Internet]. *Zenodo.* 2020. https://doi.org/10.5281/zenodo.4506970
86. Childers MC, Daggett V. Insights from molecular dynamics simulations for computational protein design. *Mol Syst Des Eng.* 2017;2:9–33. doi:10.1039/c6me00083e
87. Patra P, Ghosh P, Patra B, Bhattacharya M. Biocomputational analysis and in silico characterization of an angiogenic protein (RNase5) in Zebrafish (*Danio rerio*). *Int J Pept Res Ther.* 2020;26:1687–1697. doi:10.1007/s10989-019-09978-1
88. Talukder A, Rahman M, Masum M. Biocomputational characterisation of MBO_200107 protein of *Mycobacterium tuberculosis* variant caprae: a molecular docking and simulation study [published online ahead of print August 30, 2022]. *J Biomol Struct Dyn.* doi:10.1080/07391102.2022.2118167
89. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 2015;1-2:19–25. doi:10.1016/j.softx.2015.06.001
90. Schüttelkopf AW, van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr.* 2004;60:1355–1363. doi:10.1107/s0907444904011679
91. Zielkiewicz J. Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water. *J Chem Phys* 2005;123:104501. doi:10.1063/1.2018637
92. Schuler LD, Daura X, van Gunsteren WF. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J Comput Chem* 2001;22:1205–1218. doi:10.1002/jcc.1078
93. Hratchian HP, Frisch MJ, Schlegel HB. Steepest descent reaction path integration using a first-order predictor-corrector method. 2010;133:224101. doi:10.1063/1.3514202
94. Van Gunsteren WF, Berendsen HJC. A leap-frog algorithm for stochastic dynamics. *Mol Simul.* 1988;1:173–185. doi:10.1080/08927028808080941

95. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, ed. *The Proteomics Protocols Handbook*. Humana Press; 2005:571-607.
96. Enany S. Structural and functional analysis of hypothetical and conserved proteins of *Clostridium tetani*. *J Infect Public Health*. 2014;7:296-307. doi:10.1016/j.jiph.2014.02.002
97. Herzog B, Schultheiss A, Giesinger J. On the validity of Beer-Lambert law and its significance for sunscreens. *Photochem Photobiol*. 2018;94:384-389. doi:10.1111/php.12861
98. Kimura S, Miyauchi K, Ikeuchi Y, Thiaville PC, Crécy-Lagard Vd, Suzuki T. Discovery of the β -barrel-type RNA methyltransferase responsible for N6-methylation of N6-threonylcarbamoyladenosine in tRNAs. *Nucleic Acids Res*. 2014;42:9350-9365. doi:10.1093/nar/gku618
99. Qian Q, Curran JF, Björk GR. The methyl group of the N6-methyl-N6-threonylcarbamoyladenosine in tRNA of *Escherichia coli* modestly improves the efficiency of the tRNA. *J Bacteriol*. 1998;180:1808-1813. doi:10.1128/jb.180.7.1808-1813.1998
100. Rudenko AY, Mariasina SS, Sergiev PV, Polshakov VI. Analogs of S-Adenosyl-L-methionine in studies of methyltransferases. *Mol Biol*. 2022;56:229-250. doi:10.1134/S002689332202011X
101. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Proteins*. 2006;64:643-651. doi:10.1002/prot.21018