

Computational Investigation of the *Helicobacter pylori* CagF Protein and Identification of Repurposable Drug Candidates through Protein–Protein Interaction Network Analysis

Saikat Maity^{1,*}, Sabuj Khatua^{1,*}, Arnab Pramanik^{1,*}, Siddhanta Santra^{1,*}, Deepanjan Chattopadhyay^{2,*},
Tanima Saha^{2,#}

¹Department of Microbiology, Ramakrishna Mission Vivekananda Centenary College, Rahara, Khardaha, West Bengal, 700118, India.

²Department of Molecular Biology and Biotechnology, University of Kalyani, Kalyani 741235, Nadia, West Bengal, India.

*Contributed equally

#Corresponding Author

DOI: <https://doi.org/10.51584/IJRIAS.2026.11080155>

Received: 03 September 2026; Accepted: 09 September 2026; Published: 21 September 2026

ABSTRACT

CagF (Cag22) is an essential chaperone protein in *Helicobacter pylori* that plays a vital role in the stabilization and translocation of virulence factor CagA with the help of the type IV secretion system (T4SS). The protein contributes significantly to host-pathogen interactions. It is associated with the development of gastric diseases, along with gastric cancer. In this study, protein homology modeling, structure-based virtual screening of approved drugs, molecular docking and analysis of the CagF protein was performed. We employed a drug repurposing pipeline from which 10 compounds were shortlisted which helps to inhibit the CagA through CagF. Among the screened compounds, paliroden exhibited the most favourable predicted docking score (−9.3 kcal/mol) against the selected CagF binding pocket. The present study provides a computational pipeline for prioritizing candidate drug compounds for further investigation of CagF as a potential anti-virulence target.

Keywords: CagF, *Helicobacter pylori*, gastric cancer, drug repurposing, bioinformatics.

INTRODUCTION

Helicobacter pylori is a microaerophilic, Gram-negative bacteria that colonizes in gastric mucosa, specifically near pylorus antrum of the stomach (Marshall, 2002); causes gastric ulcer and gastric adenocarcinoma (Correa & Piazzuelo, 2011; Herrera & Parsonnet, 2009). *H. pylori* secrete the effector protein CagA from cytotoxin associated gene A through type IV secretion system (T4SS), involved in CagA translocation in *H. pylori* (Cover et al., 2020). T4SS have been associated with gastric carcinogenesis where T4SS pili of *H. pylori* is induced at the presence of host cells and delivers the effector protein CagA, followed by phosphorylation at EPIYA motifs through oncogenic Src and Abl tyrosine kinases (Hatakeyama, 2017). As CagA (Cytotoxin associated Gene A) interacts with the host cell proteins, an alternative strategy is used to inhibit CagA secretion by targeting an upstream protein of CagA in cag PAI (Akopyants et al., 1998; Olbermann et al., 2010). CagF is a cag pathogenicity island-encoded chaperone-like protein that directly interacts with CagA and binds close to the C-terminal (Pattis et al., 2007) and facilitates its translocation into the host cell (Bonsor et al., 2013; Couturier et al., 2006). *H. pylori* colonize in gastric epithelium of human and it is associated with chronic gastritis, gastric adenocarcinoma, peptic ulceration and MALT – lymphoma. *H. pylori* infection leads to approximately 76% of all gastric cancer cases in world wide. According to the IARC (International Agency for Research on Cancer) project of WHO stated that around 11.8 million lifetime stomach cancer cases out of an expected 15.6 million total cases are reported a particular age stratum (Jin Young Park, 2025; Kim et al., 2011). *H. pylori* dedicate

several genes for biosynthesis of urease to hydrolyse urea into NH_3 and CO_2 which buffer the cytosol and erect a neutral layer around the bacterium in acidic medium of stomach (Montecucco & Rappuoli, 2001). *H. pylori* possess a cag pathogenicity island which contains around 30 genes encoded Cag type IV secretion system and 17-18 genes involve in the translocation of CagA into gastric cells (Bonsor et al., 2013; Censini et al., 1996). When CagA is injected directly into gastric epithelial cells via a Type IV secretion system, it undergoes tyrosine phosphorylation at EPIYA motifs and acts as a molecular hub, aberrantly activating proliferation pathways like SHP2-ERK and PI3K/Akt, while disrupting cell polarity and suppressing tumor suppressors. CagA can induce malignant neoplasms in mammals. After delivery into host cell CagA disturbs multiple host signalling pathways by acting as a hub protein (Hatakeyama, 2014; Stein et al., 2013). While CagA delivery acts as a direct bacterial oncoprotein that dysregulates host cell signalling, *H. pylori* simultaneously fuel an indirect carcinogenic pathway through persistent inflammation and chronic tissue remodelling. Progression of gastric carcinogenesis through a hit-and-run mechanism in which pro-oncogenic actions of CagA are consecutively acquired by a series of genetic or epigenetic alterations compiled in cancer-predisposing cells during long-term infection with CagA-positive *H. Pylori* (Hatakeyama, 2014; Stein et al., 2013). It is an inner membranous or cytosolic type protein. Experimental disruption of CagF has been shown to impair CagA translocation, highlighting CagF as a potential anti-virulence target. To eradicate *H. pylori* infections, it requires a 10- to 14-day combination course of two or more antibiotics paired with a proton pump inhibitor (PPI) to lower stomach acid and help the tissue heal. Some drugs can resist *H. pylori* by multi-drug prescription treatment (Ierardi et al., 2019). The FDA (US Food and Drug Administration) approved Talicia for treatment of *H. pylori* infection in 2019, based on results of two phase III studies that disclosed eradication rates of 84.1% and 90.3%. Antibiotic resistance among *H. pylori* strains is the major cause of eradication failure. *H. Pylori* develop high rates of resistance against Amoxicillin, Clarithromycin, Metronidazole, Tetracycline, Levofloxacin nowadays because of their frequent usage, which causes failure of standard treatment plans. *H. Pylori* develop resistance through genetic mutations, efflux pumps, and biofilms, making standard therapy less effective worldwide. For this reason, drug repurposing is one of the easiest ways to resist *H. pylori* and cure diseases like Peptic Ulcer, Gastritis, Gastric Adenocarcinoma, Gastric MALT Lymphoma etc (Boyanova et al., 2023; Kalfus et al., 2020). Drug repurposing approach uses existing approved drugs with new targets to gear up the drug discovery method. *In-silico* drug repurposing process saves both cost and time. Although CagF is experimentally established as an important component of CagA translocation, its structural characteristics and potential ligand-binding sites remain insufficiently explored. Therefore, the main focus of the study is to check interaction between CagA and CagF and build the three-dimensional (3D) structure of CagF by using integrated system biology and structural biology approach. Also, we intended identify ligand-binding pockets of CagF, and to target the CagF by FDA approved drugs to disrupt the downstream signalling pathway to inhibiting the CagA by using computational drug repurposing approach.

MATERIALS AND METHODS

2.1. Protein-Protein Interaction (PPI) network construction:

Protein-protein interaction (PPI) network analysis was performed to visualize and characterize physical and functional associations among proteins. STRING database is used to reserve physical and functional protein-protein interactions information. Separately the protein-protein interactions network of CagA and CagF were built by using STRING plugin to Cytoscape v3.10.4 (confidence score of 0.400) and then both the PPIs were merged and visualized in the Cytoscape v3.10.4. The connections in a network were considered as 'unidirectional'. The interacting partners were denoted as nodes and their connections were denoted as edges in the network.

2.2. Network topology analyses:

The network topology analyses were done for identifying highly connected nodes within the network, which participate in interaction with CagA and CagF, which are responsible for development of gastric diseases, including gastric cancer. The 'Network analyzer' plugin in the Cytoscape v3.10.4 was used to perform network topology analysis. Degree, Closeness Centrality, Betweenness Centrality, and Clustering coefficient are several topological features that used for validation of the network quality.



2.3. Investigation of hub proteins

In a PPIs network, the most interactive protein which has more association with other proteins is known as hub proteins. Hub protein analysis was done by STRING plugin to Cytoscape v.3.10.4. Algorithms such as, Degree, Betweenness, and Bottleneck in Cytohubba were used to identify hub proteins. Subnetworks are interconnected via the hub proteins.

2.4. Modular analysis

In a Protein-Protein Interactions network, the connected sub-networks were determined by the MCODE plugin to Cytoscape v.3.10.4. To do modular analysis, the particular parameter was set as Haircut = Ok, Maximum depth = 100, Degree cutoff = 2, K-score = 2, Node score cutoff = 0.2.

2.5. Secondary Structure prediction, Physicochemical Properties determination and prediction of subcellular location

Secondary structure and physicochemical properties of CagF were analysed for the quality assessment of the predicted protein model and characterize the protein sequence. Secondary structure prediction was performed in Sopma (Geourjon&Deléage, 1995) (accessed by: https://npsa.lyon.inserm.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) which evaluate the structural integrity, stereochemistry and structural geometry based on the number of helices, sheets, turns, loops, random coils etc. and physicochemical properties are determined from Expasy (Gasteiger et al., 2005) (accessed by: <https://web.expasy.org/protparam/>) which dictate how protein fold, function and mostly the interaction within the living organisms. These are relied on the physical properties like molecular size and shape, colloidal nature solubility etc. and the chemical properties like amphoteric nature, isoelectric point, hydrolysis etc. Subcellular Localization prediction was done by the cello (Yu et al., 2014) (accessed by: <https://cello.life.nctu.edu.tw/>) server to find out the cellular compartmentalization of CagF protein in host cell.

2.6. Protein sequence retrieval

Uniport is one of the most standardised, widely and expertly curated protein database in the world. The amino acid sequence of CagF was retrieved in FASTA format from the UniProt database (accessed by: <https://www.uniprot.org/>, accession: P94835) (Wu, 2006) and used for protein structure prediction.

2.7. Protein structure prediction

The principle of protein structure prediction is that 3D structure of protein is conserved relative to its amino acid sequence. So, the small variation in the amino acid sequence of a protein causes minimal changes to its 3D structure. Protein structure prediction method like AlphaFold3 utilize this principle by incorporating other proteins, ligands, metal ion, etc into their model. This method can predict structure of multi- component complex and assess their performance using statistical performance data such as average RMSD. Advanced tools like AlphaFold3 predict protein structures with extreme precision that often matches laboratory results. The unavailability of crystallographic structure of a particular protein needs the prediction of the three-dimensional (3D) structure by comparative modeling with similar reference structure. As the x-ray crystallographic structure of CagF is not available in the RCSB PDB database, so we performed 3D structure prediction of CagF. The 3D structure of the CagF protein was predicted using computational protein structure prediction server AlphaFold3 (<https://alphafoldserver.com/>) (Abramson et al., 2024).

2.8. Structural validation

The predicted CagF models were subjected to refinement using GalaxyRefine server. The refined models were ranked based on the provided global quality scores by GalaxyRefine. The model showing the most favourable overall structural characteristics was selected for subsequent analyses (Heo et al., 2013)(accessed by: <https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>). The selected model was subsequently validated using the SAVES server (accessed by: <https://saves.mbi.ucla.edu/>), which incorporates several widely accepted

protein structure validation tools, such as ERRAT, PROCHECK, and Verify3D. The overall quality of each model was assessed based on the ERRAT quality factor, Verify3D compatibility score, and Ramachandran plot statistics generated by PROCHECK. The model having the highest validation scores and the most favourable stereochemical characteristics was selected for subsequent computational analyses.

2.9. Binding Site Detection

The binding sites of a protein are determined by the organization of its constituent amino acid sequence, pocket geometry, physicochemical properties. The CagF protein was analysed to identify potential ligand-binding pockets. Binding site prediction was performed using the DrugRep server (Gan et al., 2023) (accessed by: http://cao.labshare.cn:10180/DrugRep/php/screen_blinddock.php), which identifies putative binding cavities based on pocket geometry and binding scores. The drugs were screened on the basis of binding pocket (Grid volume= 470, Center - x= -14.8, y= -0.2, z= -3.0 and size- x=17, y=17, z=15, spacing- 1 Å) of CagF protein in Drug Rep server. The binding pocket with the highest binding score and suitable cavity volume was selected as the active site for subsequent molecular docking studies.

2.10. Retrieval of FDA approved drugs by structure based virtual drug screening

Identification of existing FDA approved drugs were screened and retrieved from MTiOpenScreen (accessed by: <https://bioserv.rpbs.univ-paris-diderot.fr/services/MTiOpenScreen/>) webserver against the target protein via structure based virtual screening. The MTiopenscreen webserver (Labbé et al., 2015) provides access to the Drug Bank (Knox et al., 2024) database to screen FDA approved drugs. No additional ADMET analysis was performed. Structure-Based Virtual Screening (SBVS) is an essential computational method in the drug screening process. This method utilizes the three-dimensional (3D) structure of a target protein to analyse its interactions with potential drug molecules (Cavasotto & W. Orry, 2007). In the present study, the predicted 3D structure of the CagF protein and its recognised binding site were used for virtual screening where 1500 FDA approved top-ranked molecules are obtained from the server. Through the screening process the top 100 compounds were shortlisted, and subsequently the 10 compounds with the most favourable predicted screening scores were selected for further analysis. Those ten compounds were then subjected for molecular docking.

2.11. Molecular Docking analysis

Molecular docking was performed to investigate the potential binding of the shortlisted compounds to the predicted CagF protein structure. The refined and structurally validated CagF was used for docking. Before docking, the receptor structure was prepared in AutoDockTools by removing non-essential molecules, adding the required hydrogen atoms, assigning atomic charges, and converting the structure into PDBQT format. The shortlisted ligand structures were similarly prepared by adding hydrogen atoms, assigning appropriate charges, and converting each ligand into PDBQT format using AutoDockTools. The ligand-binding pocket identified using the DrugRep server was selected as the docking site. The docking grid was centred at x = -14.8, y = -0.2, and z = -3.0 Å, with grid dimensions of 17 × 17 × 15 Å and a grid spacing of 1 Å. The same binding-site coordinates and grid dimensions were used for docking of all shortlisted compounds to ensure consistency between the docking runs. Molecular docking was performed using AutoDock Vina (V1.1.2) (Eberhardt et al., 2021) (accessed by: <https://vina.scripps.edu/>). The ten shortlisted compounds were compared based on their AutoDock Vina docking scores. The compound showing the most favourable predicted docking score among the shortlisted compounds was selected. The predicted binding pose and ligand-residue interactions of the selected compound were visualized using Pymol (Yuan et al., 2017) (accessed by: <https://www.pymol.org/>).

RESULTS AND DISCUSSION

3.1. Protein-Protein interactions study

The experimentally established CagA–CagF interaction was further examined within the broader *H. pylori* protein interaction network using STRING plugin to Cytoscape. To comprehend how complicated biological reactions work, we did protein-protein interactions network analysis. As a result, construction of PPI network in Cytoscape software was done with 19 nodes and 50 edges as represented in the Figure-1. The result demonstrated

interactions among several cag pathogenicity island-associated proteins, including CagA, CagF, CagT, CagE, CagN, CagL of *H. pylori* which interact with each other.

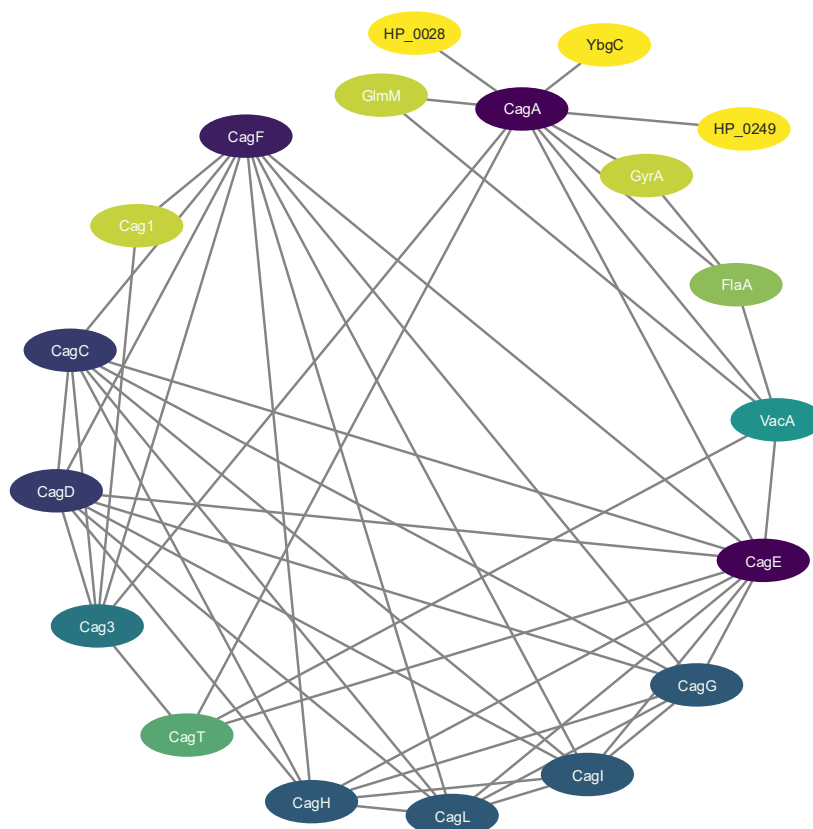


Figure 1: Protein-Protein Interactions (PPIs) Network of CagA and CagF of *H. pylori*.

3.2. Network Topology Analyses

As shown in Figure-1, the PPI network was acquired based on the node degree distribution; higher the degree, darker the colour of the node. Network density, network centralization and network heterogeneity were the various parameters used for the analyses of the PPI networks, with the values as 0.292, 0.294 and 0.584, respectively. The average clustering coefficient was 0.640, that indicates a relatively high tendency of neighbouring proteins to form interconnected groups within the constructed network. Important proteins have been identified via topological analysis of the PPI network; proteins with higher degrees, which indicate many connections, are considered as essential for a network.

3.3. Investigation of hub proteins

The top 10 hub proteins represented in Table-1 were selected on the basis of Degree and in Figure-2 some other hub proteins are represented, were identified using Bottleneck and Betweenness algorithms of CytoHubba. The PPI network showed that the proteins with the high betweenness centrality were CagA, CagE, Cag3, VacA, and CagF. These five proteins may act as the bridges between nodes to keep the network intact, indicating that they occupy central positions in the constructed network. Depending on the degree value, the interactome with the most interactions was found to be involving CagA, CagE, CagF, CagD, and CagC.

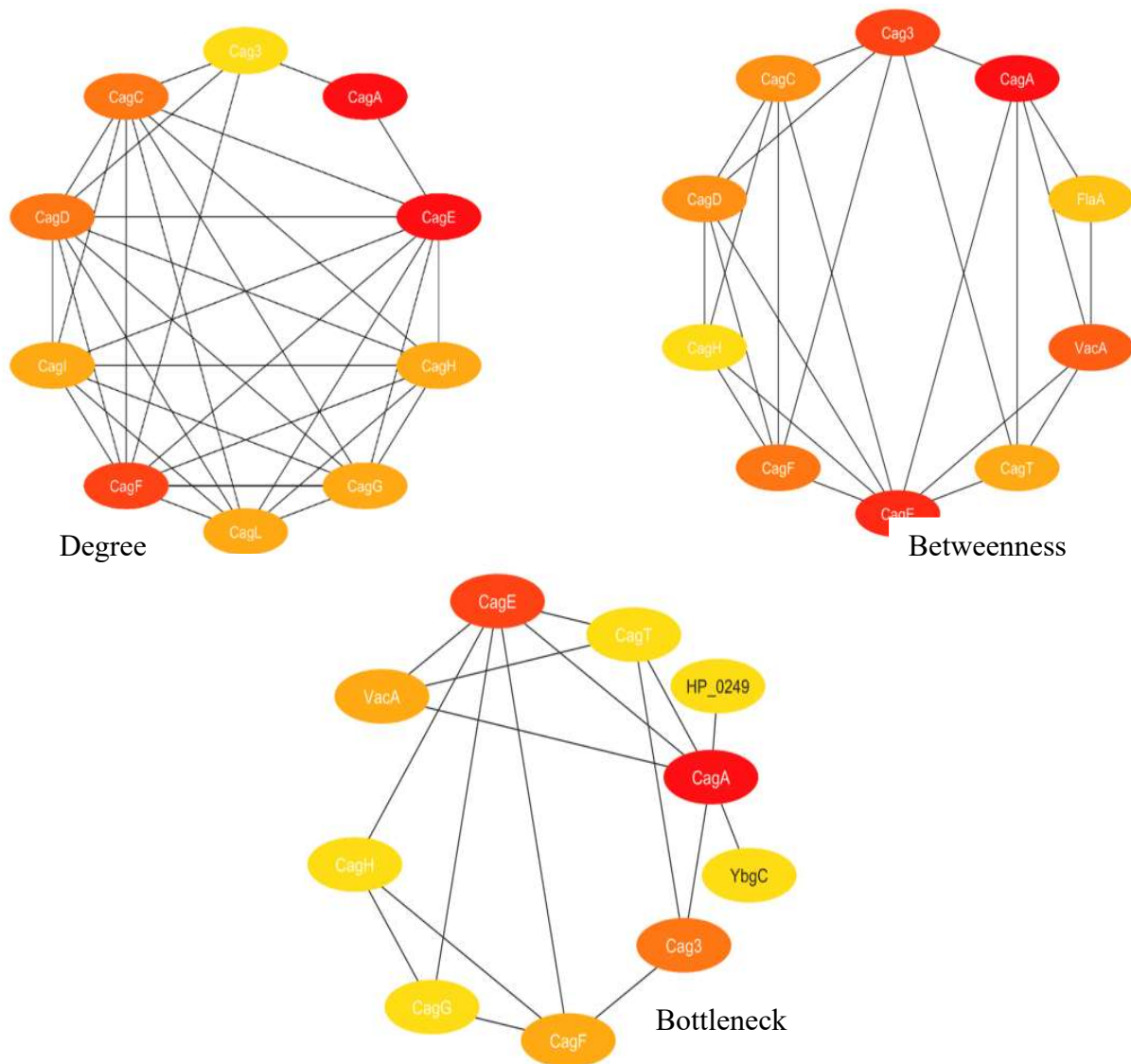


Figure 2: Hub and important proteins identified using different algorithms Degree, Betweenness, Bottleneck.

Name	Degree	BetweennessCentrality	ClosenessCentrality	ClusteringCoefficient
CagA	10	0.496949891	0.692307692	0.155555556
CagE	10	0.34204793	0.692307692	0.533333333
CagF	9	0.051416122	0.545454545	0.666666667
CagD	8	0.010021786	0.529411765	0.821428571
CagC	8	0.010021786	0.529411765	0.821428571
CagL	7	0	0.514285714	1
CagG	7	0	0.514285714	1
CagH	7	0	0.514285714	1
CagI	7	0	0.514285714	1
Cag3	6	0.135076253	0.6	0.333333333

Table 1: Topological characteristics of the ten highest-ranked proteins based on degree

Modular Analysis:

The MCODE plugin was used to produce sub-networks. MCODE identified two clusters (Figure-3) with scores of 8.0 and 3.33, containing 8 and 4 nodes and 28 and 5 edges, respectively. The result show that CagF, CagE, CagC, CagL, CagG, CagD, CagH, CagI formed cluster together. CagE may act as an intermediate to construct interaction network and may represent a network-associated component connecting CagA and CagF associated proteins. The hub proteins, CagE, CagF, CagD, CagC, CagL based on the degree score were reidentified in the Cluster-1 and CagA reidentified in the Cluster-2. MCODE modular analysis produced two major cluster of proteins that were intricately associated in the network. CagF emerged as one of the highest-scoring hub proteins, and considering its major role in *H. pylori* associated gastric carcinogenesis, we selected CagF for further analysis.

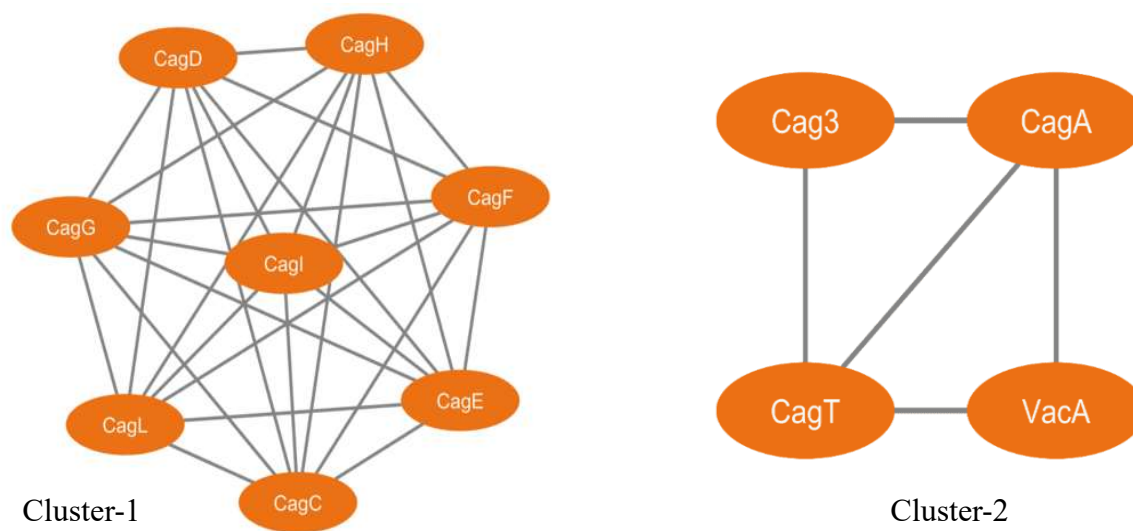


Figure 3: Sub-network clusters generated by MCODE.

Secondary Structure Prediction, Physicochemical Properties Determination and Sub Cellular Localization:

The secondary structure investigation showed that the structure of CagF mostly contained alpha helices (67.16 %) followed by random coil (24.63 %), extended strand (8.21 %) and beta turn (0.00%) regions, respectively [Table-2 and Figure-4]. The high alpha helical content indicates a well folded and structurally stable protein architecture. The random coil regions come up with the conformational flexibility and functional interactions. The analysis of physico-chemical characteristics showed that the CagF protein has more negatively charged amino acid residues [(Asp + Glu) 46] than the positively charged amino acid residues [(Arg + Lys) 24] present. The presence of higher number of negatively charged amino acids in proteins is probably responsible for its acidic nature. The extinction coefficient was measured as $24870 \text{ M}^{-1} \text{ cm}^{-1}$ (Table-3), which calculates that the protein contains high amount of Trp, Tyr and Cys residues (Table-4). The Instability index (II) was computed to be 48.35, which falls above the conventional threshold of 40 and therefore suggests potential instability. The relative volume of side chains of aliphatic amino acids in a protein determines its aliphatic index which increases the thermal stability of protein. In our study, the CagF showed high aliphatic index 92.35 (Table-3), which suggested their thermostability. GRAVY score reflects the hydrophobicity or hydrophilicity of a protein. Positive GRAVY indicates that the protein is hydrophobic. On the other hand, negative GRAVY indicates the hydrophilic nature of a protein. The negative GRAVY scores (-0.444) (Table-3) of CagF suggests that CagF has an overall hydrophilic character. Sub-cellular analysis by CELLO predicted CagF to be predominantly cytoplasmic with the reliability score of 2.780.

Predicted secondary structure	Alpha helix (%)	Beta turn (%)	Extended strand (%)	Random coil (%)
CagF	67.16%	0.00%	8.21%	24.63%

Table 2: Predicted secondary structure composition of CagF.

Name of the protein	No. of Amino Acids	No. of positively charged Amino Acids	No. of negatively charged Amino Acids	Aliphatic Index	GRAVY Score	Instability Index	Extinction coefficient
CagF	268	24	46	92.35	-0.444	48.35	24870 M ⁻¹ cm ⁻¹

Table 3: Physico-chemical characteristics of CagF protein.

Amino Acid	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	Pro	Sec
Composition (%)	2.6	1.9	10.4	7.8	0	5.6	9.3	1.5	1.9	7.8	11.6	7.1	3.4	7.5	1.9	6.7	3	0.4	4.9	4.9	0	0

Table 4: Composition of Amino Acid in the CagF protein

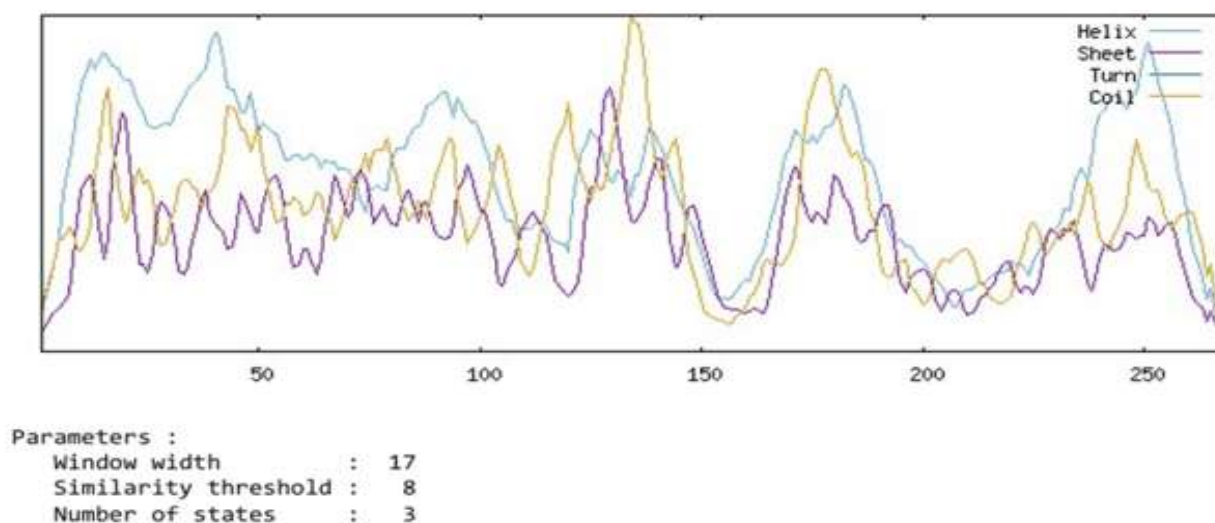


Figure 4: Predicted Secondary Structure of CagF.

Three-Dimensional Structure Prediction and Structure Validation

The three-dimensional structure of CagF protein (Figure-5) was generated in AlphaFold3, and best-fit model was chosen. In Ramachandran plot examination of the best selected model has 98.1% and 1.9% residues in most favoured region and additional allowed region, with no residues in the disallowed region respectively (Figure-6C). Whereas 0.0% was in the disallowed region. Score obtaining from the Ramachandran plot revealed a good quality model structure. The ERRAT score obtained was 95.3846 (Figure-6B), implying a favourable overall non-bonded interaction geometry. Those analyses imply that the stereo chemical parameters that the preferred model structure is fair and dependable and can be utilised for more drug studies. The verify 3D analysis indicated that 86.57% of the residues have averaged 3D-1D score ≥ 0.1 (Pass) and at least 80% of the amino acids have scored ≥ 0.1 in the 3D/1D profile which suggests that the model is reliable to build up a 3D structure (Figure-6A).

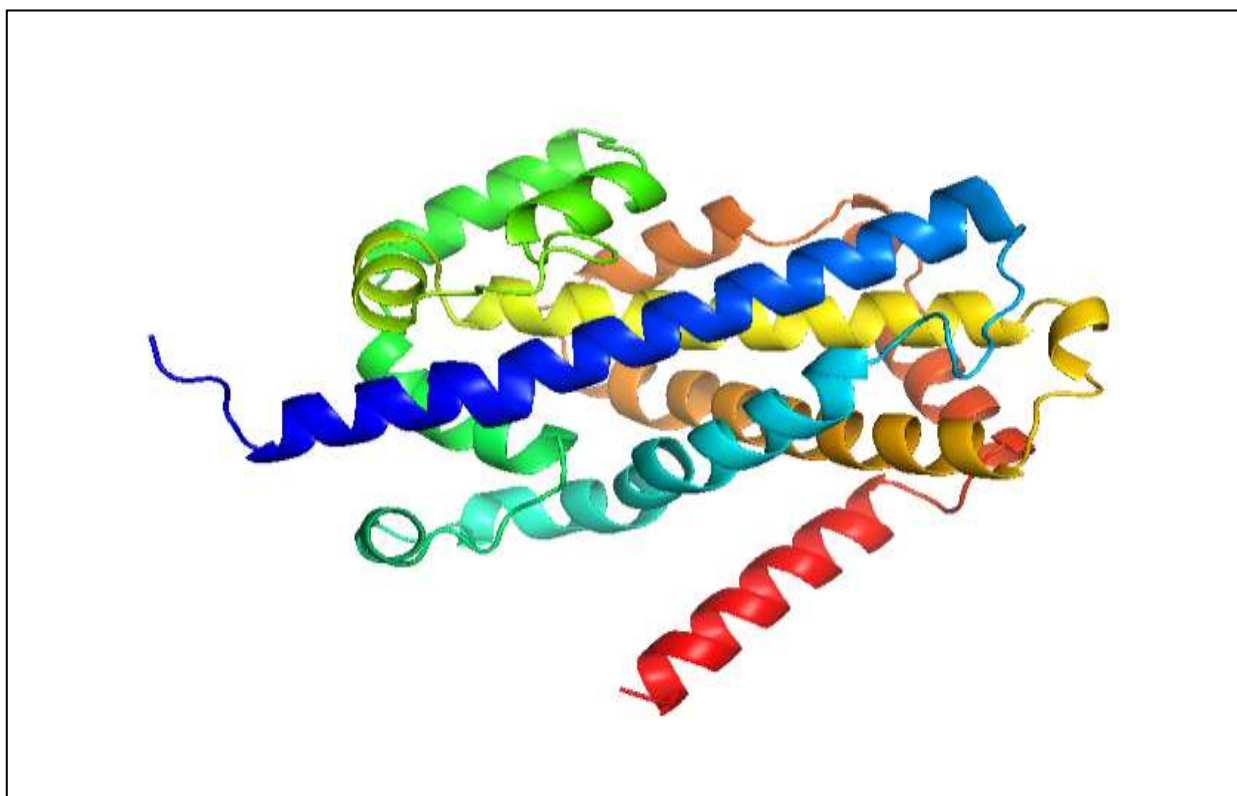


Figure 5 – Predicted three-dimensional protein structure of CagF.

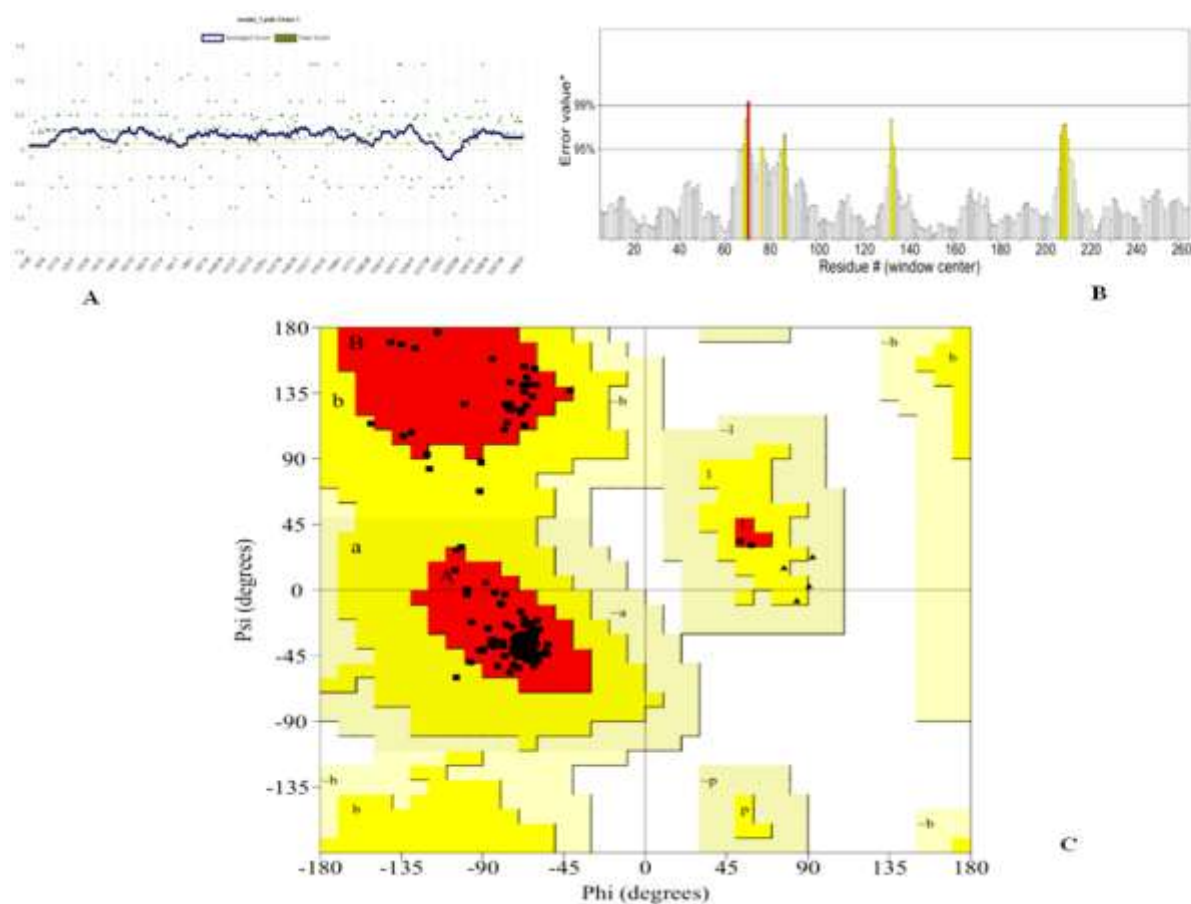


Figure 6: (A) Verify3D plot for modelled CagF structure, (B) ERRAT plot of modelled CagF structure, (C) Ramachandran plot for modelled CagF structure.

Binding pocket detection

Potential ligand-binding pockets in CagF were predicted using the DrugRep server. Five putative pockets were identified. Pocket 1 was selected for subsequent virtual screening and molecular docking based on its cavity characteristics. The binding pocket-1 contained amino acids such as VAL68 LEU258 TYR201 PHE198 ILE250 GLU71 GLU63 LYS261 TYR70 ASN67 LYS265 ASN197 MET257 VAL194 GLN195 ASN262 THR254 (Table- 5, Figure-7).

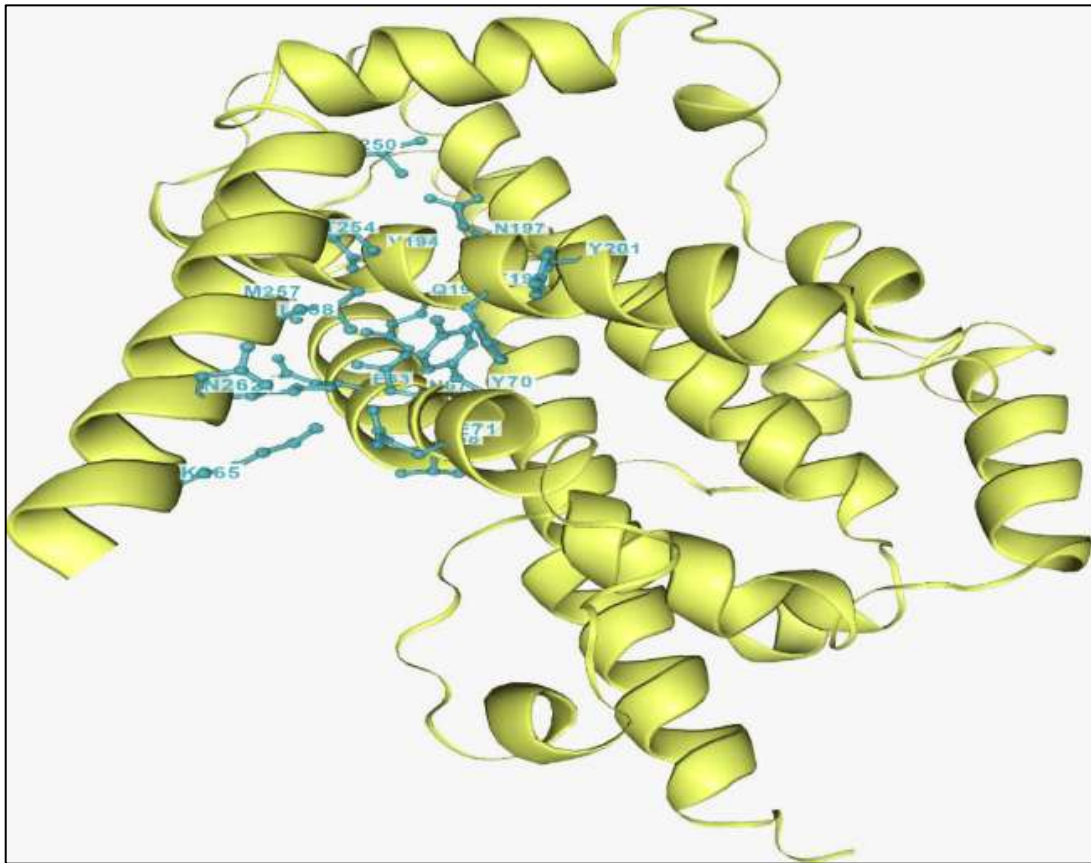


Figure 7: Detected Binding Site (Pocket-1)

Binding Pocket	Amino acid residues
Pocket 1	VAL68 LEU258 TYR201 PHE198 ILE250 GLU71 GLU63 LYS261 TYR70 ASN67 LYS265 ASN197 MET257 VAL194 GLN195 ASN262 THR254
Pocket 2	LEU96 ILE207 TYR155 TYR214 TYR202 ASN211 ALA156 PHE69 HSD97 LYS100 LYS206 ILE76 LYS210 VAL152 PHE148 VAL73 ASN93
Pocket 3	ASP17 ILE13 PHE98 ARG6 LYS9 ASN16 LEU5 ILE89 HSD83 GLU12 GLY87 ASN95 TYR85 LEU84
Pocket 4	LYS33 TYR157 ARG171 PRO226 TRP228 ARG168 LYS175 GLU161 ILE229 LEU172 GLN165 MET164 TYR167 ILE224 GLU40
Pocket 5	VAL203 VAL235 MET164 TYR208 MET217 GLN204 HSD218 MET221 GLU200 GLU236 ILE229 ILE163 LEU160 TYR167 ILE225

Table 5: Binding Pocket Compositions of CagF.

Virtual screening for potential drug compounds

Virtual screening was performed in MTiOpenScreen server. According to their minimum binding energy affinity by the server, 1500 drug molecules were screened and 100 compounds were shortlisted as per containing drug



in authorized drug library. The initial MTiOpenScreen screening generated predicted initial virtual screening scores. The ten highest-ranked compounds were subsequently subjected to AutoDock Vina docking. Among the ten shortlisted compounds, paliroden produced the most favourable AutoDock Vina docking score (-9.3 kcal/mol). Out of 100 compounds; 10 compounds were selected based on their lowest binding energy represented in Table-6.

Sl.no.	Drug Bank ID	Common Name	Chemical Class	Disease Treating & Managing	Mechanism of Action	Initial virtual-screening binding affinity (kcal/mol)	AutoDock Vina docking binding affinity (kcal/mol)	Interacting Amino acid(s)
1	DB12411	Bemcentinib	Pyridazine Derivative	Adenocarcinoma of Lung, Colorectal Neoplasms, Central nervous system Neoplasms	Tyrosine protein kinase receptor UFO inhibitor	-10.9	-6.7	A TYR 70 A LEU 74 A ASN 197 A PHE 198 A TYR 201 A ASP 251 A THR 254 A ASN 255 A LEU 258 A ASN 262 A LYS 265
2	DB06393	Xaliproden	Tetrahydro pyridine derivative	Alzheimer disease	Serotonin 1a(5-HT1a) receptor agonist	-10.6	-8.3	A GLU 63 A ASN 67 A TYR 70 A GLU 71 A GLU 190 A ILE 191 A VAL 194 A ASN 197 A PHE 198 A GLU 200 A TYR 201 A GLN 204 A ASN 205 A ILE 250 A VAL 253 A THR 254 A ASN 255 A LYS 256 A MET 257 A LEU 258 A ASP 259 A GLN 260 A LYS 261 A ASN 262 A ASN 264 A LYS 265
3	DB12185	Exatecan	Mesylate dehydrate derivative	Breast Neoplasms, Fallopian tube Neoplasms Leukaemia	DNA topoisomerase -I inhibitor	-10.6	-7.6	A TYR 70 A ASN 197 A PHE 198 A TYR 201 A ASP 251 A THR 254 A ASN 255 A LEU 258 A ASN 262
4	DB11894	Efatutazone	Thiazolidinedione derivative	Colorectal Neoplasms, Lipocarcinoma,	Peroxisome proliferator-activated receptor	-10.3	-7.4	A ASN 67 A TYR 70 A GLU 71 A VAL 194 A GLN 195



				Multiple Myeloma	gamma agonist			A ASN 197 A PHE 198 A TYR 201 A THR 254 A ASN 255 A MET 257 A LEU 258 A ASN 262 A LYS 265
5	DB06435	Tarazepide	Carboxamide derivative	Gastrointestinal diseases and disorders	Inhibition of CCK-A receptor	-10.2	-7.8	A TYR 70 A GLU 71 A LEU 74 A ASN 197 A PHE 198 A TYR 201 A ASP 251 A THR 254 A ASN 255 A LEU 258 A ASP 259 A ASN 262 A LYS 265
6	DB09183	Dasabuvir	Methanesulfonamide derivative	Hepatitis C	RNA directed RNA polymerase inhibitor	-10.2	-6.6	A TYR 70 A VAL 194 A ASN 197 A PHE 198 A GLU 200 A TYR 201 A GLN 204 A ILE 250 A THR 254 A ASN 255 A LEU 258 A ASP 259 A ASN 262
7	DB15583	Fluazuron	Difluorobenzamide	Tick - bornemdis ease	Pharmacologically active on Fibroblast growth factor receptor 3	-10.1	-8.4	A ASN 67 A TYR 70 A GLU 71 A LEU 74 A VAL 194 A GLN 195 A ASN 197 A PHE 198 A TYR 201 A ILE 250 A THR 254 A ASN 255 A MET 257 A LEU 258 A ASP 259 A LYS 261 A ASN 262 A LYS 265
8	DB12875	Mavatrep	Benzimidazole derivative	Osteoarthritis, Knee	Vanilloid receptor antagonist	-10.1	-7.7	A ASN 67 A TYR 70 A GLU 71 A LEU 74 A VAL 194 A GLN 195 A ASN 197 A PHE 198 A TYR 201 A ASP 251 A THR 254



								A ASN 255 A MET 257 A LEU 258 A ASP 259 A LYS 261 A ASN 262 A LYS 265
9	DB11791	Capmatinib	Benzamide derivative	Breast Neoplasms , Carcinoma, Non-small cell lung cancer	Hepatocyte growth factor receptor inhibitor	-10.0	-9.0	A ASN 67 A TYR 70 A GLU 71 A LEU 74 A VAL 194 A GLN 195 A ASN 197 A PHE 198 A TYR 201 A ILE 250 A ASP 251 A THR 254 A ASN 255 A MET 257 A LEU 258 A LYS 261 A ASN 262 A LYS 265
10	DB05454	Paliroden	Tetrahydro pyridine	Alzheimer Disease, Parkinson Disease	Irreversible MAOB inhibitor	-10.0	-9.3	A ASN 67 A TYR 70 A GLU 71 A LEU 74 A VAL 194 A GLN 195 A ASN 197 A PHE 198 A TYR 201 A ILE 250 A ASP 251 A THR 254 A ASN 255 A MET 257 A LEU 258 A ASP 259 A LYS 261 A ASN 262 A LYS 265

Table 6: Top 10 Drugs found from virtual screening.

Molecular docking studies:

Molecular docking studies were performed using selected ten drugs in Auto Dock Vina. According to the AutoDock vina docking score, the Paliroden was selected as the best fit molecule having lower binding affinity – 9.3 kcal/ mol and adopted the best orientation of the drug compound when interacting with CagF (Figure-8). The Paliroden was previously investigated for neurogenerative disorders specifically in Alzheimer’s disease and Parkinson’s disease, was repurposed for targeting the CagF protein. The further wet-lab experiments can be done after knowing the intermolecular interaction (hydrogen bond, hydrophobic interaction etc.) between ligand and target protein. Detailed view of the binding interaction between Paliroden and CagF protein is represented in Figure- 9A and 9B.

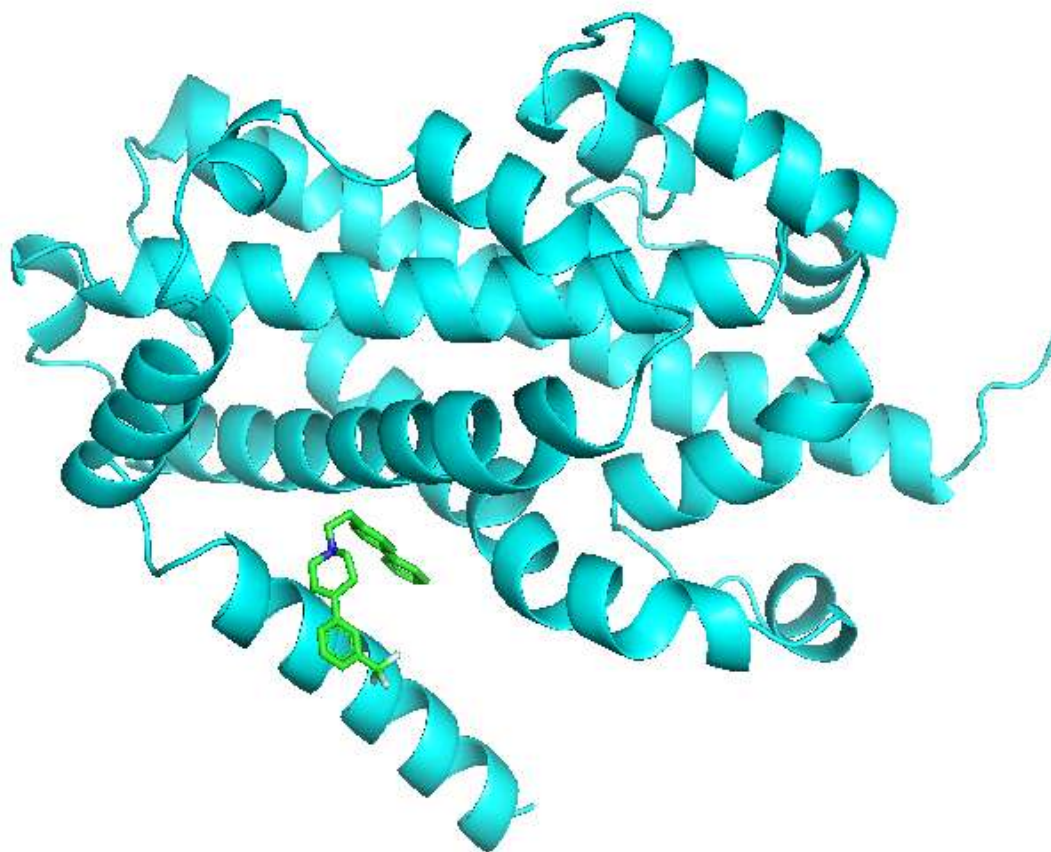
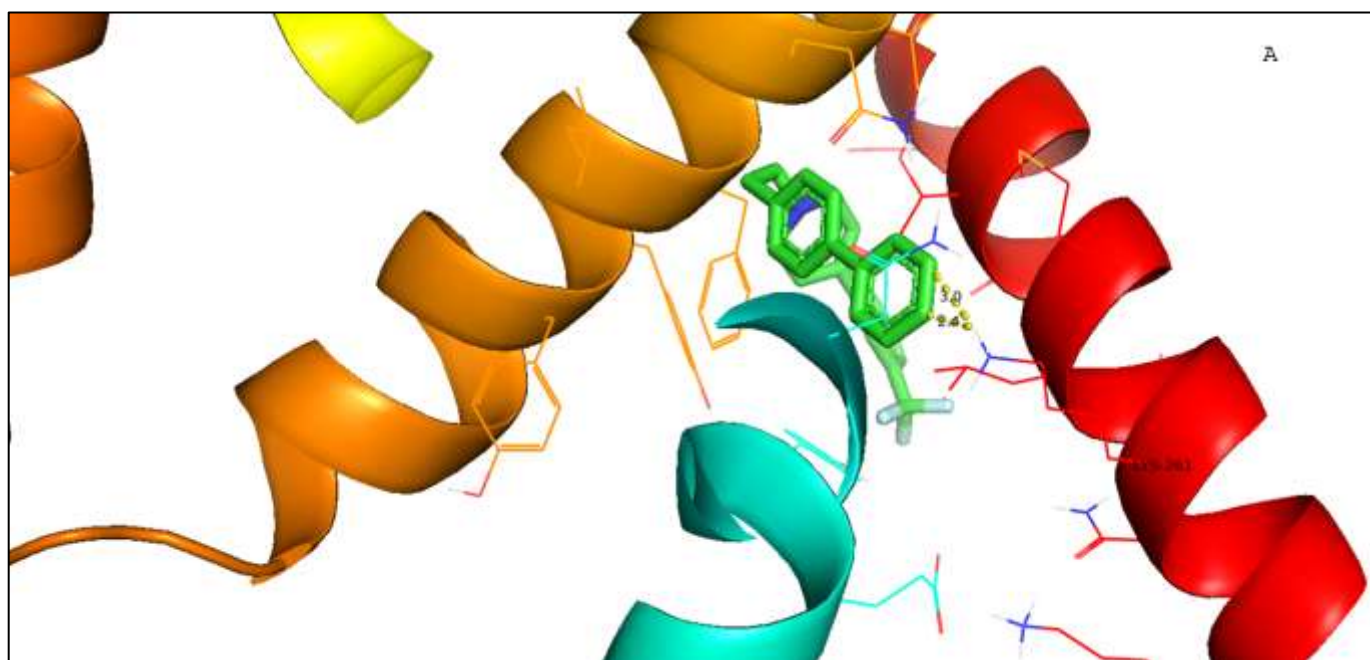


Figure 8: Predicted binding pose of Paliroden within the selected CagF binding pocket obtained using molecular docking.



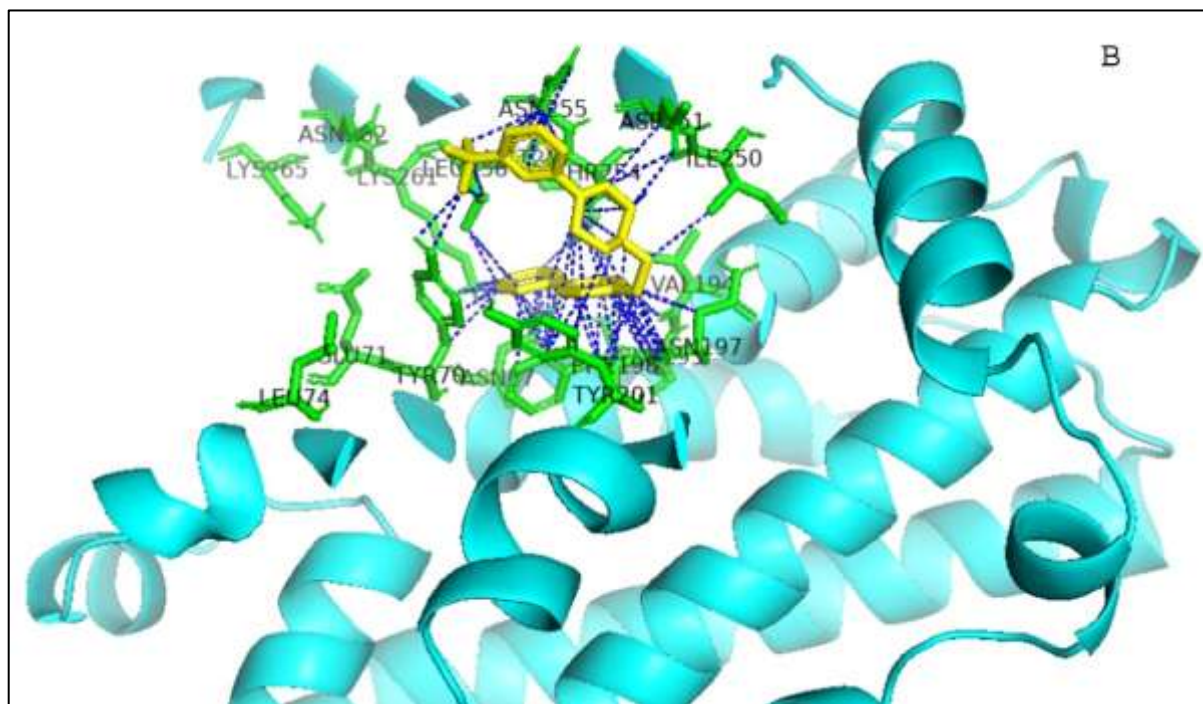


Figure 9A and 9B: Detailed visualization of predicted interactions between Paliroden and residues lining the CagF binding pocket.

In the absence of the crystal structure of CagF protein in *H. pylori*, reliable three-dimensional structural model was generated by using the alpha fold server and validated in saves webserver to analyse the Ramachandran plot, verify-3D, ERRAT value to confirm the most reliable protein model. The structural validation of the CagF model with 98.1% of residue in Ramachandran plot and ERRAT score of 95.38, and favourable verify3D and these validation metrics support the structural plausibility of the predicted structure which help to validate the 3D model and help to find out best 3D homology model. These validation parameters establish model utility for virtual screening. Secondary structure analysis show that the CagF protein consists of alpha helices (67.16%) with CagF chaperon function, and physicochemical properties show that acidic nature, which may help in electrostatic interaction. The binding site prediction identified a pocket comprising 17 amino acid residues particularly the presence of tyrosine residues is especially significant because their hydroxyl groups can participate in critical hydrogen bonding interactions with drug molecules, enhancing binding affinity. According to favourable binding energies within 100 drugs 10 drugs are chosen. Paliroden is chosen among the ten screened drugs based on predicted AutoDock Vina docking score -9.3 Kcal/mol. Paliroden is primarily used in neurogenerative disorders (Alzheimer's disease, Parkinson's disease etc.). The predicted binding pose indicates that the compound can occupy the selected CagF pocket and interact with several residues within the predicted cavity. After the molecular docking, paliroden drug bound more efficiently with binding pocket of CagF and may inhibit the CagF of *H. pylori*.

CONCLUSION

In this study, an integrated computational approach was employed to characterize CagF. PPI and network analyses identified CagF within a highly interconnected network of cag-associated proteins. Structural prediction and validation provided a computationally plausible three-dimensional model of CagF, which was subsequently used for binding-pocket prediction and structure-based virtual screening. It is observed that the drug Paliroden was identified as the most favourable compound based on the AutoDock Vina docking score (-9.3 kcal/mol) against the selected CagF binding pocket. The predicted interaction suggests that paliroden may interact with CagF; however, molecular docking alone cannot establish inhibition of CagF activity or disruption of the CagF interaction with other cag-pathogenicity island proteins. The findings suggest that the predicted interaction requires validation using biochemical, biophysical and cellular experiments to determine the effects of paliroden on CagF function. High quality of CagF model serves as a template for future structure-based research. Future studies may investigate whether Paliroden can be combined with established *H. pylori* therapies. Further



laboratory and preclinical studies are required to evaluate the pharmacological suitability, safety, and efficacy of Paliroden.

Conflicts of Interest

The authors have no conflicts of interest in financial or commercial issues of the present research work or its publication.

ACKNOWLEDGEMENTS

TS would like to acknowledge the Faculty Research Grant and RUSA Grant at University of Kalyani for the funding necessary for the research work.

REFERENCES

1. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., ... Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016), 493–500. <https://doi.org/10.1038/s41586-024-07487-w>
2. Akopyants, N. S., Clifton, S. W., Kersulyte, D., Crabtree, J. E., Youree, B. E., Reece, C. A., Bukanov, N. O., Drazek, E. S., Roe, B. A., & Berg, D. E. (1998). Analyses of the cag pathogenicity island of *Helicobacter pylori*. *Molecular Microbiology*, 28(1), 37–53. <https://doi.org/10.1046/j.1365-2958.1998.00770.x>
3. Bonsor, D. A., Weiss, E., Iosub-Amir, A., Reingewertz, T. H., Chen, T. W., Haas, R., Friedler, A., Fischer, W., & Sundberg, E. J. (2013). Characterization of the Translocation-competent Complex between the *Helicobacter pylori* Oncogenic Protein CagA and the Accessory Protein CagF. *Journal of Biological Chemistry*, 288(46), 32897–32909. <https://doi.org/10.1074/jbc.M113.507657>
4. Boyanova, L., Hadzhiyski, P., Gergova, R., & Markovska, R. (2023). Evolution of *Helicobacter pylori* Resistance to Antibiotics: A Topic of Increasing Concern. *Antibiotics*, 12(2), 332. <https://doi.org/10.3390/antibiotics12020332>
5. Cavasotto, C., & W. Orry, A. (2007). Ligand Docking and Structure-based Virtual Screening in Drug Discovery. *Current Topics in Medicinal Chemistry*, 7(10), 1006–1014. <https://doi.org/10.2174/156802607780906753>
6. Censini, S., Lange, C., Xiang, Z., Crabtree, J. E., Ghiara, P., Borodovsky, M., Rappuoli, R., & Covacci, A. (1996). cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proceedings of the National Academy of Sciences*, 93(25), 14648–14653. <https://doi.org/10.1073/pnas.93.25.14648>
7. Correa, P., & Piazuelo, M. B. (2011). *Helicobacter pylori* Infection and Gastric Adenocarcinoma. *US Gastroenterology & Hepatology Review*, 7(1), 59–64.
8. Couturier, M. R., Tasca, E., Montecucco, C., & Stein, M. (2006). Interaction with CagF is required for translocation of CagA into the host via the *Helicobacter pylori* type IV secretion system. *Infection and Immunity*, 74(1), 273–281. <https://doi.org/10.1128/IAI.74.1.273-281.2006>
9. Cover, T. L., Lacy, D. B., & Ohi, M. D. (2020). The *Helicobacter pylori* Cag Type IV Secretion System. *Trends in Microbiology*, 28(8), 682–695. <https://doi.org/10.1016/j.tim.2020.02.004>
10. Eberhardt, J., Santos-Martins, D., Tillack, A. F., & Forli, S. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling*, 61(8), 3891–3898. <https://doi.org/10.1021/acs.jcim.1c00203>
11. Gan, J., Liu, J., Liu, Y., Chen, S., Dai, W., Xiao, Z.-X., & Cao, Y. (2023). DrugRep: an automatic virtual screening server for drug repurposing. *Acta Pharmacologica Sinica*, 44(4), 888–896. <https://doi.org/10.1038/s41401-022-00996-2>
12. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (pp. 571–607). Humana Press. <https://doi.org/10.1385/1-59259-890-0:571>

13. Geourjon, C., & Deléage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*, 11(6), 681–684. <https://doi.org/10.1093/bioinformatics/11.6.681>
14. Hatakeyama, M. (2014). *Helicobacter pylori* CagA and Gastric Cancer: A Paradigm for Hit-and-Run Carcinogenesis. *Cell Host & Microbe*, 15(3), 306–316. <https://doi.org/10.1016/j.chom.2014.02.008>
15. HATAKEYAMA, M. (2017). Structure and function of *Helicobacter pylori* CagA, the first-identified bacterial protein involved in human cancer. *Proceedings of the Japan Academy, Series B*, 93(4), 196–219. <https://doi.org/10.2183/pjab.93.013>
16. Heo, L., Park, H., & Seok, C. (2013). GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Research*, 41(W1), W384–W388. <https://doi.org/10.1093/nar/gkt458>
17. Herrera, V., & Parsonnet, J. (2009). *Helicobacter pylori* and gastric adenocarcinoma. *Clinical Microbiology and Infection*, 15(11), 971–976. <https://doi.org/10.1111/j.1469-0691.2009.03031.x>
18. Ierardi, E., Losurdo, G., Fortezza, R. F. La, Principi, M., Barone, M., & Leo, A. Di. (2019). Optimizing proton pump inhibitors in *Helicobacter pylori* treatment: Old and new tricks to improve effectiveness. *World Journal of Gastroenterology*, 25(34), 5097–5104. <https://doi.org/10.3748/wjg.v25.i34.5097>
19. Jin Young Park. (2025). Population-Based *Helicobacter pylori* Screen-and-Treat Strategies for Gastric Cancer Prevention: Guidance on Implementation (J. Y. Park, Ed.). International Agency for Research on Cancer. <https://www.ncbi.nlm.nih.gov/books/NBK615777/>
20. Kalfus, I. N., Graham, D. Y., Riff, D. S., & Panas, R. M. (2020). Rifabutin-Containing Triple Therapy (RHB-105) for Eradication of *Helicobacter pylori*: Randomized ERADICATE Hp Trial. *Antibiotics*, 9(10), 685. <https://doi.org/10.3390/antibiotics9100685>
21. Kim, S. S., Ruiz, V. E., Carroll, J. D., & Moss, S. F. (2011). *Helicobacter pylori* in the pathogenesis of gastric cancer and gastric lymphoma. *Cancer Letters*, 305(2), 228–238. <https://doi.org/10.1016/j.canlet.2010.07.014>
22. Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A., Pon, A., Cox, J., Chin, N. E. (Lucy), Strawbridge, S. A., Garcia-Patino, M., Kruger, R., Sivakumaran, A., Sanford, S., Doshi, R., Khetarpal, N., Fatokun, O., Doucet, D., Zubkowski, A., ... Wishart, D. S. (2024). DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Research*, 52(D1), D1265–D1275. <https://doi.org/10.1093/nar/gkad976>
23. Labbé, C. M., Rey, J., Lagorce, D., Vavruša, M., Becot, J., Sperandio, O., Villoutreix, B. O., Tufféry, P., & Miteva, M. A. (2015). MTiOpenScreen: a web server for structure-based virtual screening. *Nucleic Acids Research*, 43(W1), W448–W454. <https://doi.org/10.1093/nar/gkv306>
24. Marshall, B. (2002). *Helicobacter pylori*: 20 years on. *Clinical Medicine*, 2(2), 147–152. <https://doi.org/10.7861/clinmedicine.2-2-147>
25. Montecucco, C., & Rappuoli, R. (2001). Living dangerously: how *Helicobacter pylori* survives in the human stomach. *Nature Reviews Molecular Cell Biology*, 2(6), 457–466. <https://doi.org/10.1038/35073084>
26. Olbermann, P., Josenhans, C., Moodley, Y., Uhr, M., Stamer, C., Vauterin, M., Suerbaum, S., Achtman, M., & Linz, B. (2010). A Global Overview of the Genetic and Functional Diversity in the *Helicobacter pylori* cag Pathogenicity Island. *PLoS Genetics*, 6(8), e1001069. <https://doi.org/10.1371/journal.pgen.1001069>
27. Pattis, I., Weiss, E., Laugks, R., Haas, R., & Fischer, W. (2007). The *Helicobacter pylori* CagF protein is a type IV secretion chaperone-like molecule that binds close to the C-terminal secretion signal of the CagA effector protein. *Microbiology*, 153(9), 2896–2909. <https://doi.org/10.1099/mic.0.2007/007385-0>
28. Stein, M., Ruggiero, P., Rappuoli, R., & Bagnoli, F. (2013). *Helicobacter pylori* CagA: From Pathogenic Mechanisms to Its Use as an Anti-Cancer Vaccine. *Frontiers in Immunology*, 4. <https://doi.org/10.3389/fimmu.2013.00328>
29. Wu, C. H. (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Research*, 34(90001), D187–D191. <https://doi.org/10.1093/nar/gkj161>



30. Yu, C.-S., Cheng, C.-W., Su, W.-C., Chang, K.-C., Huang, S.-W., Hwang, J.-K., & Lu, C.-H. (2014). CELLO2GO: A Web Server for Protein subCELLular Localization Prediction with Functional Gene Ontology Annotation. PLoS ONE, 9(6), e99368. <https://doi.org/10.1371/journal.pone.0099368>
31. Yuan, S., Chan, H. C. S., & Hu, Z. (2017). Using PyMOL as a platform for computational drug design. WIREs Computational Molecular Science, 7(2). <https://doi.org/10.1002/wcms.1298>