Indian Journal of Biochemistry & Biophysics
Vol. 63, September 2026, pp. 995-1012
DOI: 10.56042/ijbb.v63i9.30865

In silico characterization unveils *Helicobacter pylori* protein HP0973 as a novel EGFR-interacting virulence factor

Debabrata Dash[1], Debashis Barik[1,2], Aditya Kumar Panda[1,2] & Sushil Kumar Pathak[1,2]*

[1]Department of Biotechnology; & [2]Centre of Excellence on Bioprospecting of Ethno-pharmaceuticals of Southern Odisha (CoE-BESO), Berhampur University, Berhampur-760 007, Odisha, India

Received 17 April 2026; *revised 03 July 2026*

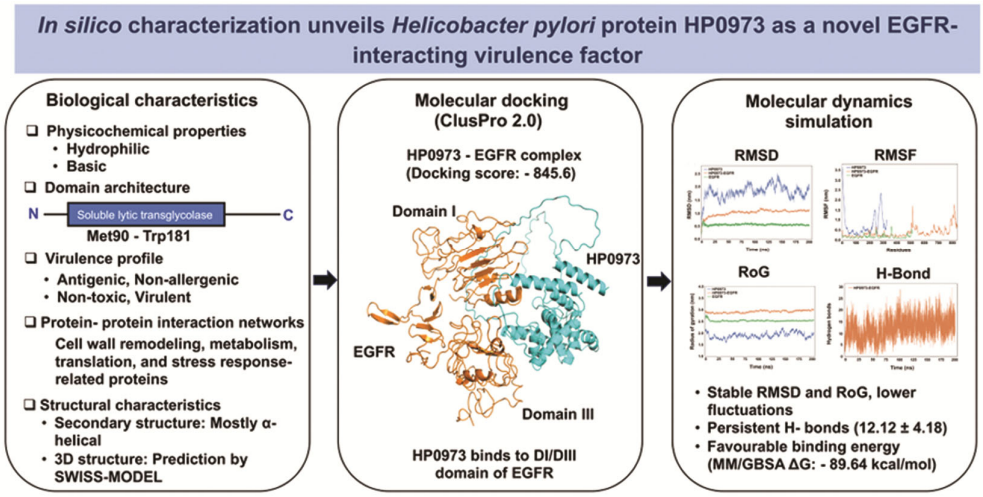
Helicobacter pylori infection induces gastric inflammation and carcinogenesis through multiple virulence factors. However, a substantial proportion of its genome encodes hypothetical proteins with unknown function. Characterizing these proteins is crucial to expanding our understanding of bacterial pathogenesis. This study presents comprehensive *in silico* characterization of the hypothetical secreted protein HP0973. Physicochemical and virulence profiling revealed HP0973 as basic, stable, non-toxic, and highly antigenic protein. Sequence alignment and phylogenetic analysis confirmed conservation among *H. pylori* strains with a soluble lytic transglycolase (SLT) domain (Met90 - Trp181). Protein-protein interaction networks positioned HP0973 centrally with metabolic, translational, and stress-response partners. Secondary structure mostly contained α-helical structure elements, whereas tertiary modelling favoured SWISS-MODEL structure (ERRAT 96.17%; QMEAN -1.00). Molecular docking revealed binding of HP0973 to DI/DIII domains of Epidermal Growth Factor Receptor (EGFR), competing with Epidermal Growth Factor (EGF), as confirmed by reciprocal docking. 200 ns MD simulations affirmed complex stability (RMSD: 1.015 ± 0.11 nm, RMSF 0.363 ± 0.24 nm; RoG 2.911 ± 0.04 nm), sustained H-bonds (12.12 ± 4.18), and favourable MM/GBSA binding free energy (-89.64 kcal/mol), dominated by van der Waals forces. These findings suggest HP0973 as a novel EGFR modulator in *H. pylori*-driven pathogenesis, warranting experimental validation for therapeutic targeting.

Keywords: *Helicobacter pylori*, HP0973, Hypothetical protein, Receptor tyrosine kinase, Protein structure modeling, Protein-protein interaction

Helicobacter pylori is a gram-negative microaerophilic bacterium that infects nearly half of the human population worldwide[1]. Persistent *H. pylori* infection is a major risk for the development of gastric cancer[2]. Several bacterial virulence factors contribute to its pathogenicity, including cytotoxin-associated antigen (Cag) pathogenicity island, vacuolating exotoxin (VacA), blood group antigen binding adhesions, BabA and SabA, duodenal ulcer-promoting gene (dupA), γ-glutamyl trans-peptidase (GGT), protease HtrA and many more[3]. Considering the general non-invasive nature of *H. pylori*, the pathogenicity of *H. pylori* is largely attributed to its diverse array of secreted proteins, which play crucial roles in evading the host's immune defences, maintaining a persistent infection, and potentially inducing disease progression[4]. *H. pylori* secretome is estimated to contain approximately 160 proteins, yet many of these secreted proteins have only been weakly characterized based on structural homology predictions, while others remain completely uncharacterized[4]. Structural and functional characterization of these proteins is essential to gain better insight into *H. pylori* pathogenesis.

Several *H. pylori* proteins are known to interact with pattern recognition receptors (PRRs), such as TLR2, TLR4, and TLR5, and activate pro-inflammatory signalling via NF-kB and MAPK pathways in epithelial cells and immune cells[5]. In addition to innate immune recognition, several integrin receptors and receptor tyrosine kinases (RTKs) facilitate *H. pylori* infection in host cells[6]. Epidermal growth factor receptor (EGFR), a subfamily of receptor tyrosine kinases, is frequently activated during bacterial and viral infections, leading to proliferation, epithelial cell survival, and inflammation, which are key to pathogenesis[7]. Several EGFR ligands have been identified such as epidermal growth factor (EGF), transforming growth factor-alpha (TGFA), heparin-binding EGF-like growth

*Correspondence:
E-mail: skpathak.su@gmail.com
Suppl. data available on respective page of NOPR



Graphical abstract

factor (HBEGF), betacellulin (BTC), amphiregulin (AREG), epiregulin (EREG), and epigen (EPGN), which activate the receptor through binding at the extracellular ligand-binding cleft, triggering receptor dimerization and downstream signalling cascade[8]. *H. pylori* infection is known to increase EGFR expression and induce its transactivation via the extracellular signal-regulated kinase (ERK) and Src pathways as well as through metalloproteinase-mediated shedding of EGFR ligands, which alter the balance of apoptosis and proliferation in gastric epithelial cells[9]. In addition to ligand-mediated transactivation, certain bacterial pathogens, such as *Staphylococcus aureus* employ a ligand-independent strategy, where its surface virulence factor protein A directly binds to EGFR, triggering receptor phosphorylation and downstream ERK activation without the requirement for canonical growth factor engagement[10]. EGFR activation also intersects with other RTKs, such as ErbB2 and c-Met, which have been implicated in *H. pylori*-driven signaling that promotes epithelial survival, DNA damage responses, and carcinogenic progression[11].

Given the multitude of host receptors involved in *H. pylori*-host interactions and the diverse mechanisms of receptor modulation, it is pertinent to investigate how uncharacterized bacterial proteins, such as HP0973, might interact with host receptors. HP0973, a hypothetical protein, has previously been detected in *H. pylori* culture supernatant and recognized as one of the 344 essential genes[12, 13]. HP0973 contains a nuclear-targeting sequence that may facilitate intracellular trafficking[14]. However, its function and potential host interactions remain largely unexplored. This study employed computational approaches, including sequence and structural annotation, protein-protein interaction network analysis, and molecular docking, to predict the potential interaction of HP0973 with EGFR, thereby providing insights into its possible role in bacterial pathogenesis.

Materials and Methods

Protein sequence retrieval

The protein sequence of HP0973 from *H. pylori* strain 26695 (accession number AAD08028.1) was retrieved from the NCBI protein database (https://www.ncbi.nlm.nih.gov/protein/) in the FASTA format. UniProt database (UniProtKB ID: O25625_HELPY) was used to obtain additional information. The presence of a signal peptide in HP0973 was evaluated using Deepsig[15] and SignalIP 6.0[16].

Physicochemical characterization of HP0973

The Expert Protein Analysis System (ExPASy) server was employed to analyze the physicochemical properties of HP0973 (https://web.expasy.org/protparam/), which provides theoretical metrics including molecular weight, amino acid composition, positive and negative residue counts, extinction coefficient, theoretical isoelectric point (pI), aliphatic index (AI), instability index (II), and the grand average of hydropathicity (GRAVY) score[17].

Multiple sequence alignment and phylogenetic analysis

NCBI-Blastp suite (https://blast.ncbi.nlm.nih.gov/) was used to identify HP0973 homologs. Nonredundant and RefSeq databases with default

parameters were selected for homology analysis. The Clustal Omega server was employed for multiple sequence alignment (https://www.ebi.ac.uk/jdispatcher/msa/clustalo). A phylogenetic tree was constructed to elucidate the evolutionary relationships of HP0973 protein from *H. pylori* 26695 with HP0973 orthologs from other *H. pylori* strains using the neighbor-joining method with 1000 bootstrap replicates (n=1000) in MEGA11 (https://www.megasoftware.net/).

Virulence factor analysis

VaxiJen 2.0 was used to predict the antigenicity of HP0973 (https://ddg-pharmfac.net/vaxijen/VaxiJen/ VaxiJen.html) with 0.4 threshold value[18]. Antigenic regions within the HP0973 sequence were identified using antigenic peptide prediction (http://imed.med.ucm.es/Tools/ antigenic.pl). The allergenicity of HP0973 was determined using Allertop 2.1 (https://ddg-pharmfac.net/allertop_test/)[19], whereas the virulence potential of the protein was evaluated using VirulentPred 2.0 (https://bioinfo.icgeb.res.in/virulent2/)[20].

Functional protein association networks

The protein-protein interaction network was identified using STRING v12.0 (https://string-db.org/)[21] and IntAct (https://www.ebi.ac.uk/intact/home)[22].

HP0973 secondary structure prediction

Secondary structure of HP0973 was predicted using PSIPRED 4.0 (http://bioinf.cs.ucl.ac.uk/psipred/) and SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa%20_sopma.html) servers. Protein sequence in FASTA format was used for these predictions. The PSIPRED workbench is based on the PSI-BLAST algorithm, whereas SOPMA computes the parameters and predicts the secondary structures based on the homology method. CDvist (http://cdvist.joulinelab.org/), a sequence-based tool for predicting protein domains, was used for domain prediction.

Tertiary structure prediction and quality assessment of HP0973

The tertiary structure of HP0973 was predicted by I-TASSER, (https://www.aideepmed.com/I-TASSER/), SWISS-MODEL (https://swissmodel.expasy.org/) and AlphaFold servers (https://alphafoldserver.com/). The 3D models predicted by I-TASSER were based on multiple threading alignments. The SWISS-MODEL server was used in the automated mode to predict the tertiary structure of HP0973 as no significant homology exists in the solved structure of the PDB database (https://swissmodel.expasy.org/interactive/). Models were refined using Galaxyrefine server ((https://galaxy.seoklab.org/cgi-bin/submit. cgi?type=REFINE) and visualized using PyMol. The refined structures were evaluated using PROCHECK[23] and ERRAT[24] module of SAVES (https://saves.mbi.ucla.edu/) server. The ProSA-web server[25] was used to determine the Z-score of the predicted structure. Models were also validated using the SWISS-MODEL structural evaluation tool.

Protein binding interface prediction

Putative protein-protein interaction sites on HP0973 were predicted using different tools such as Solvent Accessibility based Protein-Protein Interface Identification and Recognition (SPPIDER) (https://sppider.cchmc.org/), Meta Protein-Protein Interaction Site Predictor (MetaPPISP) (https://pipe.rcc.fsu.edu/meta-ppisp.html), and Protein structure transformer (PeSTo) servers. All the servers are structure based interface predictor. SPPIDER predicts the interface using alternative machine learning techniques like Support Vector Machines and Neural Networks[26]. MetaPPISP predicts protein interfaces by combining multiple machine-learning systems that use sequence and structural features to identify interface residues[27]. PeSTo predicts the interface residues based solely on the position in space and the type of atoms[28].

Molecular docking with EGFR

The HPInet module of InTxDB (https://bis.zju.edu.cn/InTxDB/predict/), a tool for detecting interactions between secreted proteins from gram-negative bacteria and host proteins using a transformer-based neural network, was employed to evaluate the interaction between HP0973 and the ectodomain of EGFR (PDB ID: 1IVO). Molecular docking was performed to investigate the interaction of HP0973 with human EGFR using ClusPro 2.0 web server[29]. The crystal structure of the extracellular domain of EGFR with its natural ligand EGF (PDB ID-1IVO) was used as receptor with chain A of EGFR extracted for docking. The structures were prepared by removing water molecules, non-interacting ligands, and EGF (chain C) to evaluate whether HP0973 binds at or near the EGF binding site. Protein-protein docking was performed in ClusPro2.0

in default balanced mode, which considers shape complementarity, electrostatics, and desolvation energy for global rigid-body docking. Based on the cluster size and interaction energy, the top-ranked model was selected and analyzed using PyMOL.

Molecular dynamics (MD) simulation

MD simulations provide crucial insights into model flexibility and conformational changes in biomolecules at the atomistic level over time. It enables the analysis of structural changes in biological macromolecules under physiological conditions. In this study, a 200 ns simulation was performed for the HP0973-EGFR complex using GROMACS 2022[30]. The topology of the protein complex was created using the CHARMM36 force field[31]. The protein complexes were solvated in a cubic box with TIP3P water molecules, and Na^+/Cl^- counterions with 0.15 M salt were employed to neutralize the system. Energy minimization was performed using the steepest descent algorithm with 10000 steps to eliminate undesirable connections. Equilibration of systems was performed with 300 ps NVT (constant number of particles, volume, and temperature) and 300 ps NPT (constant number of particles, pressure, and temperature) equilibration steps. Furthermore, a production MD of 200 ns was employed for trajectory analysis. The protein-protein complex stability and dynamic behaviour were assessed via Post MD simulation trajectory analysis, including Root mean square deviation (RMSD), Root mean square fluctuation (RMSF), Hydrogen bond analysis, Radius of gyration (RoG), and binding free energy analysis through MM/GBSA calculation.

Results

Retrieval of protein sequence and physicochemical properties analysis

From the NCBI protein database, the protein sequence of the hypothetical protein HP0973 was retrieved specifically for the *H. pylori* 26695 strain with Accession No. AAD08028.1. The sequence was then cross-referenced with the UniProt database. Various attributes, including protein definition, locus, and sequence in the FASTA format, were saved for analysis. The retrieved protein, classified as a hypothetical protein, consisted of 353 amino acids. The specific properties of the sequence are outlined in (Suppl. Table S1). The presence of a signal peptide from Met1 to Ala28 was detected using DeepSig and SignalIP 6.0.

The physicochemical properties of HP0973 were estimated using the Expasy-ProtParam tool (Table 1). The mature protein consists of 325 amino acids with a molecular weight of 36.8 kDa. The theoretical isoelectric point (pI) of 9.32 indicates that HP0973 is strongly basic, suggesting that the protein has a strong affinity for negatively charged biomolecules, such as nucleic acids, phospholipids, or acidic surface protein components, which is typical of proteins involved in regulation or binding functions. HP0973 has 56 positively charged residues (Arg + Lys) and 42 negatively charged residues (Asp + Glu), which indicates the net positive charge at physiological pH. This charge asymmetry provides further support for its role in protein-protein interactions or the formation of a macromolecular complex. Its instability index of 48.84 indicates that HP0973 is considered unstable *in vitro*, and probably needs interaction partners, or the cellular environment, to preserve its structural stability. The half-life of HP0973 in *E. coli* is greater than 10 h, while in yeast it is greater than 20 h, indicating relative *in vivo* stability, despite showing *in vitro* instability in terms of its biological relevance during bacterial growth or infection. The aliphatic index was 60.71, indicating a moderate abundance of aliphatic side chains and thermal stability. The grand average of hydropathicity

Table 1 — The physicochemical properties of HP0973 predicted by ExPASy server

ProtParam Parameters	Values (without Signal peptide)
Number of amino-acids	325
Molecular weight (Da)	36805.78
Theoretical pI	9.32
Total number of negatively charged residues (Asp + Glu)	42
Total number of positively charged residues (Arg + Lys)	56
Total Number of atoms	5165
Formula	$C_{1610}H_{2584}N_{450}O_{507}S_{14}$
Extinction coefficient	34630 Abs 0.1% (=1 g/l) 0.598, assuming all pairs of Cys residues form cystines 34380 Abs 0.1% (=1 g/l) 0.591, assuming all Cys residues are reduced
Estimated half-life	4.4 hour (mammalian reticulocytes, *in vitro*) >20 h (*Yeast, in vivo*) >10 h (*E. coli, in vivo*)
Instability index	48.84 (Unstable)
Aliphatic index	60.71
Grand Average of hydropathicity	-0.973

(GRAVY) of -0.973 suggests that HP0973 is highly hydrophilic and preferably soluble in aqueous intracellular environments. The extinction coefficient (34,630 M^{-1} cm^{-1}) indicates that HP0973 harbours a considerable number of aromatic residues (Trp, Tyr, and Phe), suggesting significant aromatic contributions to protein folding and interaction surfaces.

Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignment using Clustal Omega revealed strong conservation of HP0973 among *H. pylori* strains, indicating a potentially important functional role (Suppl. Fig. S1). Conserved regions were observed throughout the sequence, which was further supported by conservation analysis using the ConSurf server (Fig. 1a). The ConSurf results indicate the level of conservation on a colour scale, with dark magenta indicating regions of high conservation. Phylogenetic analysis suggested that HP0973 of *H. pylori* strain 26695 clustered with similar proteins from different *H. pylori* strains, forming a clade clearly defined by high bootstrap values (>90%) (Fig. 1b). The short branch lengths in this cluster imply a small difference in sequence between different *H. pylori* HP0973 homologs. By contrast, the homologs from other *Helicobacter* species, which comprise *H. acinonychis*, *H. cetorum*, and *H. canis*, form different and more distant branches, indicating an evolutionary divergence from *H. pylori*. Taken together, the phylogenetic pattern suggests that HP0973 is a conserved protein in *H. pylori* that may act in some species-specific biological functions.

Determination of virulence properties

Evaluating the antigenicity, allergenicity, toxicity, and virulence of secreted proteins is essential for elucidating their potential impact on host immune responses. The VaxiJen2.0 tool predicted HP0973 as a probable antigen with a score of 0.634 which exceeds the threshold value of 0.4. Further antigenic peptides were predicted using the Kolaskar and Tongaonkar method[32], which predicted 12 antigenic determinants within the HP0973 protein sequence (Suppl. Table S2). The AllerTOP classified the protein as a non-allergenic, while CSM-toxin predicted HP0973 as non-toxic in nature. The virulence of the protein was further assessed by the VirulentPred 2.0 programme, which employs advanced machine learning (ML) techniques, leading to higher predictive accuracy, better generalization, and improved usability. The protein HP0973 was predicted to be a virulent protein.

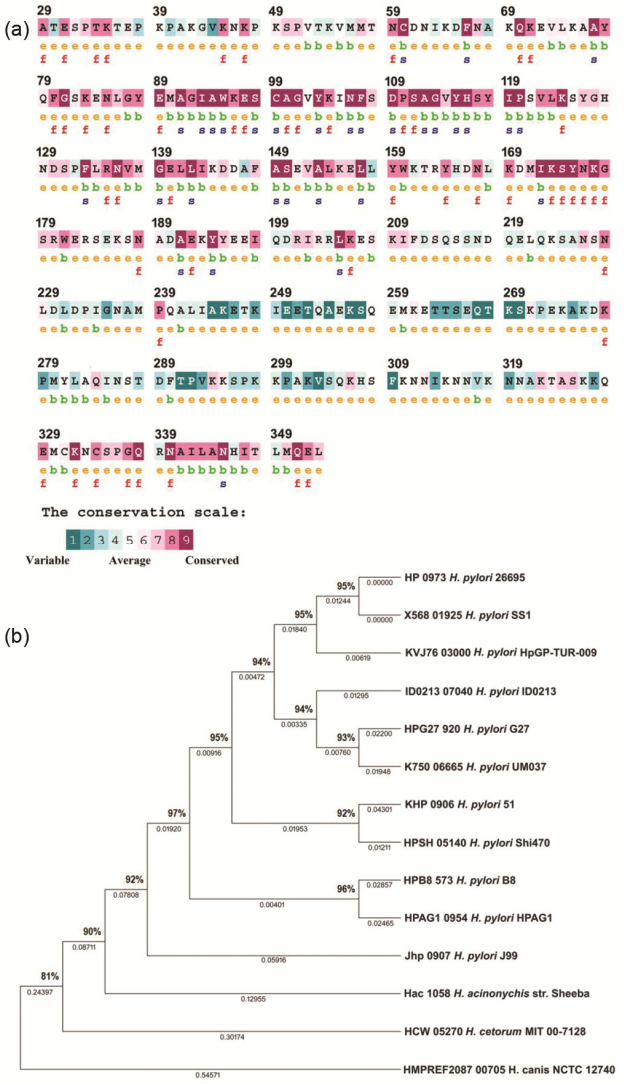


Fig. 1 — Sequence conservation of HP0973 and phylogenetic analysis. (a) Conserved residues analysis of HP0973 using the ConSurf server; residues are coloured from variable (cyan) to highly conserved (magenta); AND (b) Intraspecific phylogenetic tree showing evolutionary relationships among *H. pylori* strains based on HP0973 protein sequences using the Neighbor-Joining method with 1000 bootstrap replicates in MEGA11

Protein-Protein interaction analysis

A STRING-based protein-protein interaction analysis revealed that HP0973 is placed at the centre of a well-connected network of proteins involved in functions related to cell wall remodeling, metabolism, translation, and stress responses (Fig. 2a). Linkage with GpmI (HP0974) and GatC (HP0975), with a confidence score of 0.700, indicates the potential involvement of HP0973 in carbon metabolism and carbohydrate transport, while linking it to core translational functions through

association with GlyS (HP0972). Additionally, the predicted association with CopP (HP1073), a copper-binding stress response protein, suggests a possible role in bacterial adaptation to stress. Additionally, a BlastKOALA-based KEGG annotation of HP0973-interacting proteins revealed that the interacting partners included glycyl-tRNA synthetase (HP0972), tRNA aminotransferase subunit GatC (HP0975), and phosphoglycerate mutase (HP0974), suggesting the potential involvement of HP0973 in protein biosynthesis and metabolic coordination/carbohydrate metabolism. From IntAct database, two interacting partners have been detected to have a physical association with HP0973 as detected by Yeast two-hybrid assay[33]. The association with CopP suggests a potential link between HP0973 and metal stress response pathways, which are critical for *H. pylori* survival in the gastric environment (Fig. 2b). The interaction with HP0241 indicates that HP0973 may function in coordination with other conserved hypothetical proteins. Together, these experimentally supported interactions highlight the role of HP0973 in stress adaptation.

Structure prediction and quality assessment

The secondary structure of HP0973 was predicted using PSIPRED and SOPMA. The summarized prediction data are presented in (Table 2, and Fig. 3) provides a visual representation. Secondary structure analysis of HP0973 revealed a predominantly α-helical architecture (54.67%). In contrast, the presence of β-sheet elements was minimal, with extended strands alone contributing to 2.83%. Turn structures contributed 3.97%, whereas a substantial proportion of the protein (38.53%) was composed of random coil regions, suggesting the presence of flexible or intrinsically disordered segments. The secondary structure profile indicates that HP0973 is an α-helical protein with minimal β-structure and a moderate degree of conformational flexibility, as would be expected of proteins that function enzymatically or through protein-protein interactions. A soluble lytic transglycolase (SLT) domain was identified by the CDvist server, spanning over Met90-Trp181 amino acids.

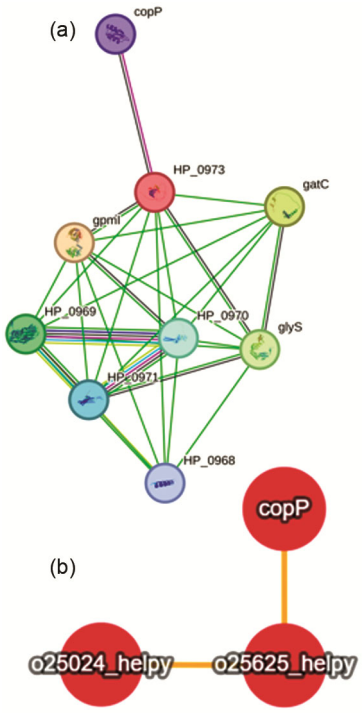


Fig. 2 — Protein interaction network of HP0973 analysed using STRING (a) and IntAct database (b). Nodes represent proteins, and edges indicate predicted functional associations by co-expression, genomic context, and curated database evidences.

The three-dimensional structure of HP0973 was predicted using the SWISS-MODEL, I-TASSER, and AlphaFold servers. SWISS-MODEL proposed two models from which the selected model had Global Model Quality Estimate (GMQE) value of 0.73. Using a multiple threading approach, three models were predicted by I-TASSER. The selected model had C-score of -4.59 with estimated TM-score of 0.24 and RMSD value of 18.1 Å. The structure predicted by the AlphaFold server exhibited a moderate global confidence, reflected by a predicted TM-score (pTM) of 0.55, suggesting a reliable fold. The central core region of the protein displayed high confidence (pLDDT >90) and adopted a compact α-helical fold, indicating a stable and well-defined structural domain. All three predicted structures were further refined using the GalaxyRefine server and subsequently visualized in PyMol (Fig. 4a). The refined structures were validated and evaluated by SAVES and ProSA-web server (Table 3, and Fig. 4b-d). Among the

Table 2 — The secondary structure of HP0973 predicted by SOPMA server

Structural parameters	Values	Percentage
Alpha helix (Hh)	193	54.67
3_{10} helices (Gg)	0	0.00
Pi-helix (Ii)	0	0.00
Beta bridge (Bb)	0	0.00
Extended strand (Ee)	10	2.83
Beta turn (Tt)	14	3.97
Bend region (Ss)	0	0.00
Random coil (Cc)	136	38.53
Ambiguous states	0	0.00

Fig. 3 — The secondary structure of HP0973 predicted by PSIPRED server. The yellow, pink and grey colour represents strand, helix and coil region of the protein, respectively

evaluated models, the SWISS-MODEL-derived HP0973 structure showed the highest overall structural quality, with an ERRAT score of 96.17% and 95.9% residues in the most favoured regions of the Ramachandran plot, indicating excellent stereochemical integrity. Its QMEAN value (-1.00), closest to zero among all models, and a QMEANDisCo score of 0.52 ± 0.05 further supported the reliability of both global and local structural features. In comparison, the I-TASSER model exhibited inferior quality, reflected by a low ERRAT score (63.76%), poor Ramachandran statistics (67.6%), and a highly negative QMEAN value (-10.70). Although the AlphaFold model showed acceptable stereochemistry (93.2% favoured residues; ERRAT 89.15%), its lower QMEAN (-2.00) and Verify3D score (53.85%) indicated reduced sequence-structure compatibility. Consequently, the SWISS-MODEL structure was selected for all downstream docking and simulation analyses.

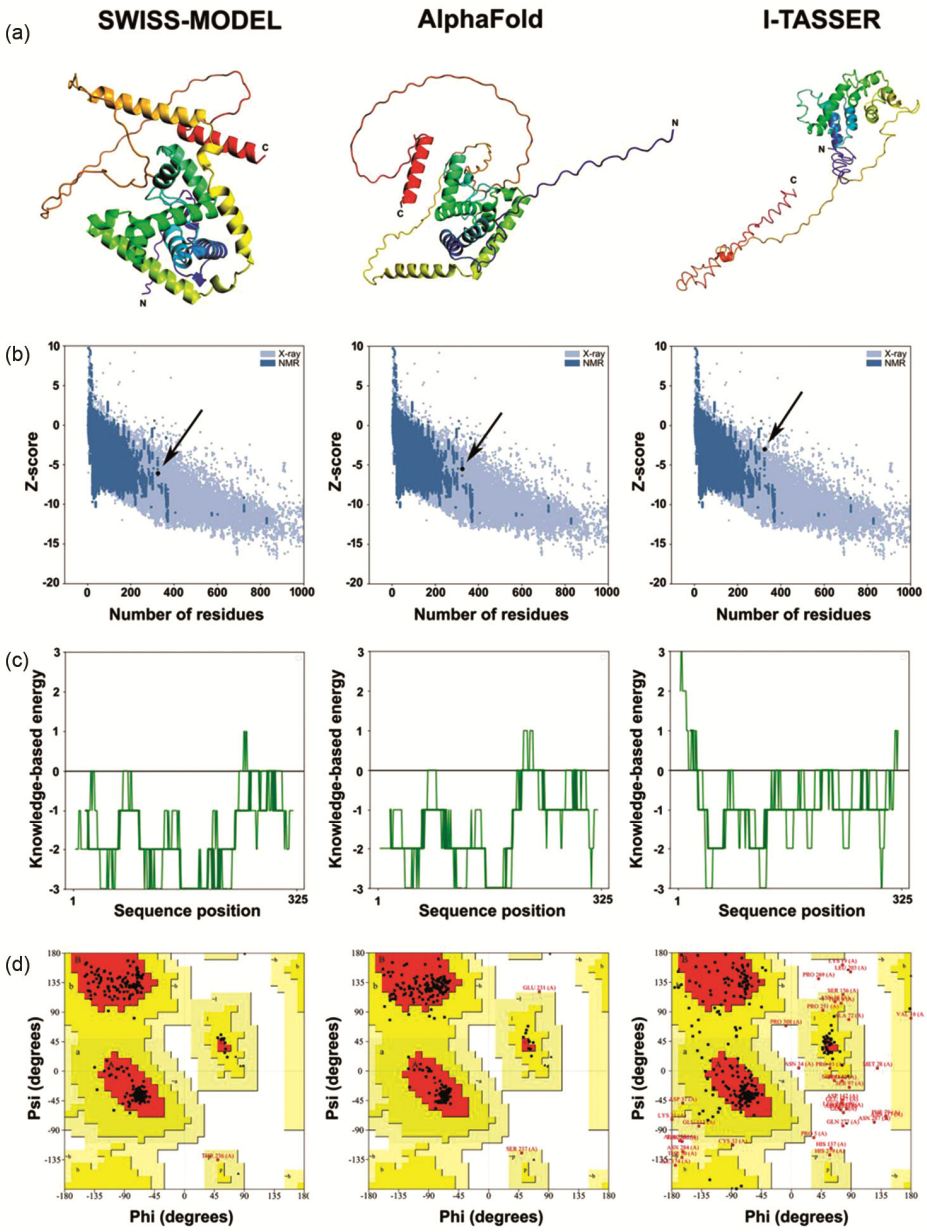


Fig. 4 — Predicted three-dimensional structure of HP0973 using SWISS-MODEL, AlphaFold and I-TASSER represented in cartoon format (a). N-terminal (Blue) and C-terminal (Red) labelled. ProSA-web Z-scores of the predicted models of HP0973 (b). The black dot (marked by arrow) indicated the Z-score, showing their placement within the range of experimentally determined X-ray and NMR structures, indicating overall model reliability (b). Knowledge-based energy profiles along the HP0973 sequence, illustrating residue-wise energy distributions and identifying energetically favourable and unfavourable regions (c). Ramachandran plot for predicted models showing stereochemical arrangements of residues in the structure (d)

Protein binding interface prediction

The protein-binding interface of HP0973 was predicted by multiple servers, including SPPIDER, MetaPPISP, and PeSTo, and the predicted binding residues are shown in (Table 4). SPPIDER server predicted a total of 113 residues at the protein-interaction interface, whereas MetaPPISP and PeSTo predicted 27 and 119 binding interface residues, respectively. The common residues identified by all three servers were LYS49, SER50, PRO51, VAL52, THR53, LYS54, VAL55, MET56, MET57, and TYR281.

Molecular docking analysis

HPINet, a host-pathogen protein-protein interaction prediction module within InTxDB, was used to

Table 3 — Structural validation parameters for refined HP0973 models

Server Name	Z-Score	ERRAT Quality Factor (%)	Ramachandran Most Favoured (%)	QMEAN Value	QMEANDisCo Value	Verify3D (%)
SWISS-MODEL	-6.08	96.17	95.9	-1.00	0.52± 0.05	62.46
I-TASSER	-3.00	63.763	67.6	-10.70	0.30 ± 0.05	55.38
AlphaFold	-5.47	89.150	93.2	-2.00	0.47± 0.05	53.85

Table 4 — Binding interface residues of HP0973

Server	Protein binding interface residues
SPPIDER	ALA29, THR30, GLU31, SER 32, PRO33, THR34, LYS35, THR36, GLU37, PRO38, LYS39, PRO40, ALA41, LYS42, GLY43, VAL44, LYS45, ASN46, LYS47, PRO48, LYS49, SER50, PRO51, VAL52, THR53, LYS54, VAL55, MET56, MET57, THR264, SER265, GLU266, GLN267, THR268, LYS269, SER270, LYS271, PRO272, GLU273, LYS274, ALA275, LYS276, ASP277, LYS278, PRO279, MET280, TYR281, LEU282, ALA283, GLN284, ILE285, ASN286, SER287, THR288, ASP289, PHE290, THR291, PRO292, VAL293, LYS294, LYS295, SER296, PRO297, LYS298, LYS299, PRO300, ALA301, LYS302, VAL303, SER304, GLN305, LYS306, HIE307, SER308, PHE309, LYS310, ASN311, ASN312, ILE313, LYS314, ASN315, ASN316, VAL317, LYS318, ASN319, ASN320, ALA321, LYS322, THR323, ALA324, SER325, LYS326, LYS327, GLN328, GLU329, MET330, CYS331, LYS332, ASN333, CYS334, SER335, GLN338, ARG339, ILE342, ASN345, HIE346, ILE347, THR348, LEU349, MET350, GLN351, GLU352, LEU353
MetaPPISP	ASN46, LYS47, PRO48, LYS49, SER50, PRO51, VAL52, THR53, LYS54, VAL55, MET56, MET57, THR58, ASN59, CYS60, ASN67, TRP95, GLU97, CYS99, GLY101, VAL102, TYR103, LYS104, ILE105, ASN106, PHE107, TYR281
PeSTo	ALA29, THR30, GLU31, SER 32, PRO33, THR34, LYS35, THR36, GLU37, PRO38, LYS39, PRO40, ALA41, LYS42, GLY43, VAL44, LYS45, ASN46, LYS49, SER50, PRO51, VAL52, THR53, LYS54, VAL55, MET56, MET57, THR58, VAL102, TYR103, ILE105, ASN106, PHE107, SER108, LYS124, GLY127, ASP102, GLU140, ILE143, LYS144, ARG180, ARG183, GLU255, LYS256, GLN258, GLU259, GLU262, THR263, THR264, SER265, THR268, LYS269, SER270, LYS271, PRO272, LYS274, ALA275, LYS276, ASP277, LYS278, PRO279, MET280, LEU282, ILE285, ASN286, SER287, THR288, ASP289, VAL293, LYS294, LYS295, SER296, PRO297, LYS298, PRO300, ALA301, LYS302, VAL303, SER304, GLN305, LYS306, HIE307, SER308, PHE309, LYS310, ASN311, ASN312, ILE313, LYS314, ASN315, ASN316, VAL317, LYS318, ASN320, ALA321, LYS322, THR323, ALA324, SER325, LYS326, LYS327, GLN328, GLU329, MET330, CYS331, LYS332, ASN333, CYS334, SER335, PRO336, GLN338, ARG339, ILE342, ASN345, HIE346, LEU349, MET350, GLU352, LEU353

predict the protein sequence-based interaction of HP0973 with the ectodomain of EGFR. The resulting residue-importance heat map for the HP0973-EGFR complex is presented in **(**Suppl. Fig. 2). A blind docking approach was employed for HP0973 and EGFR interaction as there was no direct evidence of interaction. The monomeric form of EGFR was prepared for docking analysis. ClusPro generated multiple docking conformations based on the cluster minimization score. The selected binding model had lowest energy score of -845.6 with a representative of 36 members in the cluster. Structural analysis of the selected model demonstrated that HP0973 binds to the ligand-binding cleft, which exists between the DI and DIII domains of EGFR, by establishing multiple contacts with residues from both domains (Fig. 5a & b). To compare this binding mode with the physiological ligand, EGF was separately docked with EGFR. As expected, EGF bound to EGFR at the interface of the DI–DIII domains, consistent with the reported EGFR–EGF complex (Fig. 5c). To access, HP0973 engagement to EGFR in the presence of its physiological ligand, HP0973 was docked with preformed EGFR-EGF complex (Fig. 5d). HP0973 failed to occupy the ligand-binding cleft instead localizing to a distinct surface of EGFR, suggesting that EGF binding restricts HP0973 binding to the cleft. Notably, when EGF was docked with the preformed HP0973-EGFR complex, EGF failed to occupy its canonical binding interface (Fig. 5e). The docking results suggest that HP0973 binds to EGFR in a functionally active region and may affect the process of ligand-dependent receptor activation. To study the atomic behaviour and long-term stability of the protein complex, a 200 ns MD simulation was performed using GROMACS with the CHARMM36 force field.

Root mean square deviation (RMSD) analysis

RMSD plot was analysed to evaluate the structural stability of HP0973, EGFR and their complex (Fig. 6a). Over the simulation period, Apo HP0973 exhibited large conformational fluctuations compared with an average RMSD value of 1.811 ± 0.24 nm,

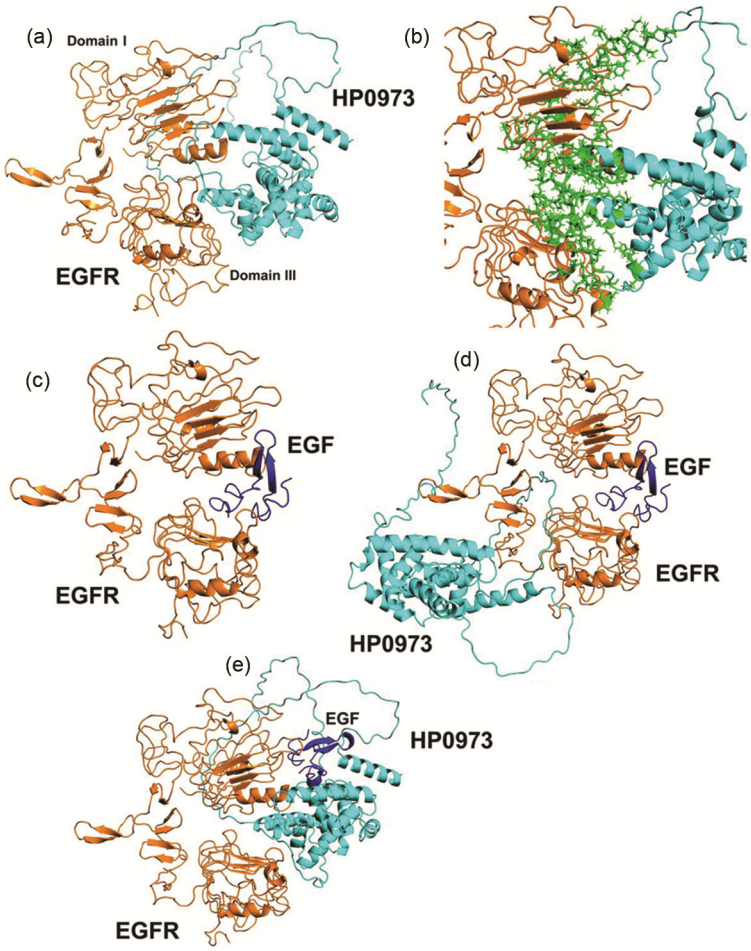


Fig. 5 — Structural interaction of HP0973 with EGFR. Docked structure of HP0973 (Cyan) bound to the extracellular domain of EGFR (orange) at the region of Domain I and Domain III interface (a). The HP0973-EGFR complex, showing interacting residues represented as stick models (green) (b). Docking model of EGF (blue) and EGFR (orange) demonstrating canonical binding pocket engagement (c). Docking of EGF (blue) onto the pre-formed HP0973-EGFR complex, showing displacement of EGF from its native binding pocket by HP0973 (d). Docking of EGF with preformed HP0973-EGFR complex showing disruption of EGF binding to EGFR (e)

indicating intrinsic flexibility in the absence of receptor binding, whereas apo EGFR displayed higher stability throughout the simulation with average RMSD values of 0.552 ± 0.03 nm. Interestingly, when HP0973 formed a complex with EGFR, the RMSD value was significantly reduced (1.015 ± 0.11 nm). The HP0973-EGFR complex became well stabilized after initial 5-10 ns, indicating the formation of a structurally stable complex over the simulation period. The conformational flexibility of HP0973 was restricted upon interaction with EGFR, resulting in a more stable structure.

Root mean square fluctuation (RMSF) analysis

Conformational fluctuations and stability at the residue level were assessed using RMSF analysis (Fig. 6b). A lowest residue-wise flexibility was shown by apo-EGFR (0.167 ± 0.09 nm), reflecting a stable unbound conformation. In the apo state, HP0973 (0.721 ± 0.58 nm), showed elevated peaks (>1.0 -1.6 nm) at multiple segments, mostly within the mid-region and C-terminal segments, indicating intrinsic disorder and conformational flexibility. In contrast, the HP0973-EGFR complex exhibited an average RMSF of 0.363 ± 0.24 nm, indicating a notable reduction in residue-level fluctuation upon EGFR binding. Specifically, the HP0973 region fluctuated less (~0.2-0.4 nm), confirming direct engagement in receptor binding.

Radius of Gyration (RoG) analysis for structural compactness

The structural organization and compactness of EGFR upon binding with HP0973 were assessed using RoG analysis (Fig. 6c). HP0973 alone exhibited a low RoG value (1.899 ± 0.11 nm), indicating a moderately flexible structure. Apo-EGFR, on the

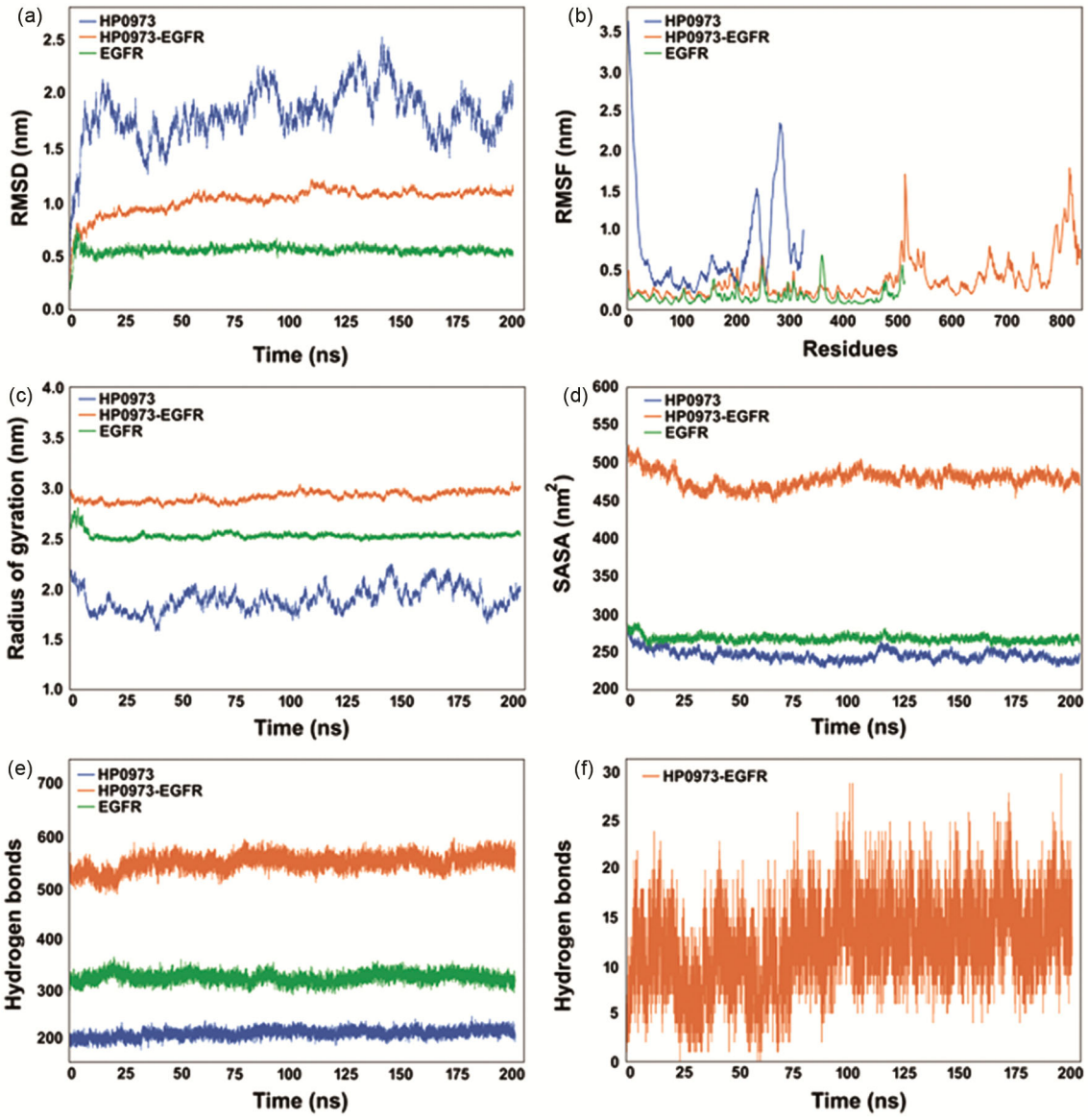


Fig. 6 — MD simulation analysis of HP0973 interaction with EGFR. Backbone RMSD profiles of HP0973, HP0973-EGFR complex, and EGFR over a 200 ns simulation, indicating structural stability of HP0973 upon complex formation (a). RMSF analysis showing reduced flexibility of HP0973 in the complex compared to the apo form, specifically at the interaction interface (b). RoG plots illustrating the compactness of HP0973, EGFR, and their complex during the simulation (c). An increased surface area of the complex was demonstrated by SASA profile (d). Total number of hydrogen bonds during the simulation period (e) and intermolecular hydrogen bonds between HP0973 and EGFR (f).

other hand, had a higher and more stable RoG value (2.526 ± 0.03 nm), which was favourably correlated with the size of the well-organized ectodomain. HP0973, upon complex formation with EGFR, showed a higher RoG value (2.911 ± 0.04 nm) with minimal fluctuation. Importantly, the narrow RoG distribution of the complex suggests that HP0973 binding does not destabilize the structure but rather facilitates the formation of a stable and well packed assembly. The absence of large fluctuations from the RoG indicates the formation of a rigid and sustained interaction interface between HP0973 and EGFR.

Solvent-accessible surface area (SASA) analysis

SASA was performed to examine the surface rearrangements and conformational exposure of HP0973, EGFR, and their complex (Fig. 6d). In the unbound form, HP0973 showed a mean SASA of 246.14 ± 7.18 nm², suggesting a compact solvent-

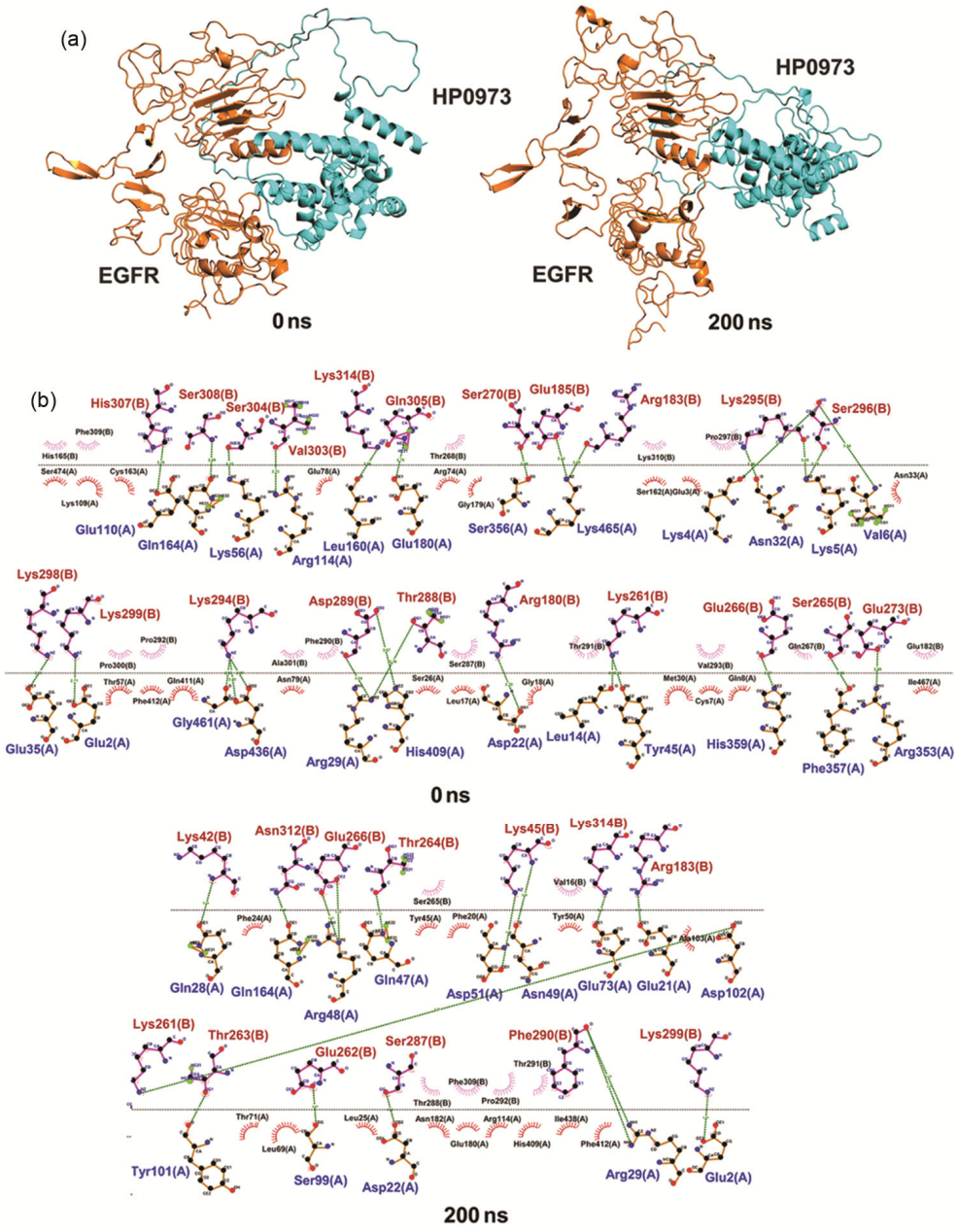


Fig. 7 — Snapshot of MD simulation at 0 ns and 200 ns (a). Ligplot^{+} interaction map depicting the interfacial hydrogen bonds (green dashed line), between HP0973 (red) and EGFR (blue) at each time point (b).

exposed fold, whereas the SASA of apo-EGFR was slightly higher, with lower variation (267.92 ± 4.16 nm²), consistent with the larger ectodomain and surface topology. The HP0973-EGFR complex demonstrated substantially higher SASA (478.83 ±11.31 nm²) reflecting the formation of an extended protein-protein interface and combined molecular surface. Following initial equilibration, the complex showed a stable SASA trajectory, which suggests that HP0973 binding formed a stable interaction surface.

Analysis of H-bonds

Intermolecular interactions and stability were evaluated using hydrogen bond analysis formed during the 200-ns simulation period. The HP0973-EGFR complex-maintained a higher number of intermolecular hydrogen bonds (547.62 ± 15.63) than the apo EGFR (321.78 ± 9.82) and apo HP0973 (212.63 ± 8.38) (Fig. 6e). The elevated H-bonds count throughout the simulation enhanced structural stabilization upon complex formation. In addition, the

HP0973-EGFR complex exhibited a stable average of 12.12 ± 4.18 interfacial hydrogen bonds over the simulation period, reflecting sustained intermolecular engagement and supporting the formation of a stable binding interface (Fig. 6f).

A structural snapshot of the HP0973-EGFR complex at the 0 and 200 ns frames revealed the preservation of the overall binding throughout the simulation (Fig. 7a). Minor loop and interfacial adjustments suggest conformational refinement. LigPlot+ software package was used for two-dimensional visualization and estimation of H-bonds and hydrophobic interactions. At 0 ns, the HP0973-EGFR complex was stabilized by 28 H-bonds and 17 hydrophobic contacts at the intermolecular interface, whereas at 200 ns, it maintained 16 H-bonds and 14 hydrophobic interactions (Fig. 7b). EGFR residues Glu2, Asp22, Arg29, and Gln164 stabilized the complex by forming persistent H-bonds with HP0973. In addition, HP0973 residues Arg183, Lys299, and Lys314 formed a stable H-bond at the interface with EGFR throughout the simulation trajectory. Hydrophobic contacts also facilitated the complex stabilization. HP0973 residues, Phe309, Thr291 maintained continuous hydrophobic interactions with EGFR. Additionally, Ser265 and Thr288 of HP0973 initially participated in H-bonding but later shifted to hydrophobic interactions with EGFR residues, reflecting adaptive interfacial rearrangement. Together, these observations suggest conformational adaptation without disruption of binding integrity, supporting the long-term stability of the HP0973-EGFR complex over the 200 ns simulation.

Binding energy calculation

The binding energy of the HP0973 complex with EGFR was estimated using the MM/GBSA method implemented by the HAWKDOCK server. The total binding energy was calculated as -89.64 kcal/mol indicating a strong and thermodynamically favourable interaction. Among the different energy components, the electrostatic energy components (ΔE_ele= - 150.22 kcal/mol) and van der Waals interactions (ΔE_vdW = -114.14 kcal/mol) exhibited the major contribution to complex formation. Furthermore, polar solvation energy (ΔG_pol = 214 kcal/mol) and non-polar solvation energies (ΔG_nonpol = -15.70 kcal/mol) collectively indicate that hydrophobic interactions and solvent-mediated interactions predominantly govern the stability of the complex.

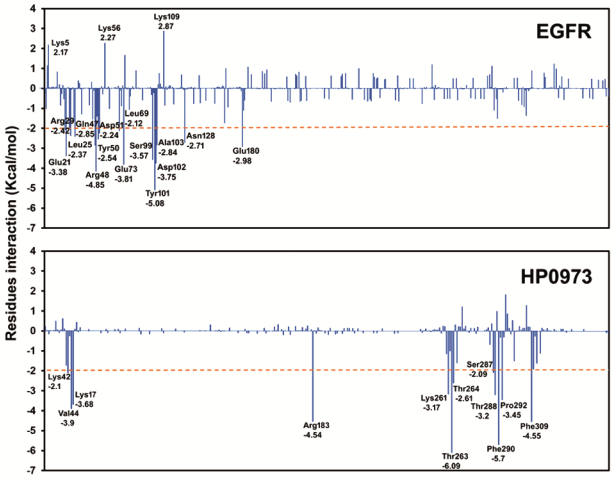


Fig. 8 — Decomposition of ΔG on a per-residue basis for HP0973 interacting with EGFR Only residues of | ΔG| ≥ -2.0 kcal/mol are shown in the figure.

Furthermore, the key residues governing the stability of the complex were identified by per-residue MM/GBSA energy decomposition analysis (Fig. 8). Residues contributing more than 2.0 kcal/mol were considered energetically significant. Several residues of EGFR and HP0973 contributed to favourable binding at the interface. Tyr101 (-5.08 kcal/mol), Arg48 (-4.85 kcal/mol), Glu73 (-3.81 kcal/mol), Asp102 (-3.75 kcal/mol), Ser99 (-3.57 kcal/mol), and Glu21 (-3.38 kcal/mol) emerged as dominant hotspot residues of EGFR with additional contributions from Glu180 (-2.98 kcal/mol), Gln47 (-2.85 kcal/mol), Ala103 (-2.84 kcal/mol), Asn128 (-2.71 kcal/mol), Tyr50 (-2.54 kcal/mol), Arg29 (-2.42 kcal/mol), and Leu25 (-2.37 kcal/mol), indicating a distributed electrostatic-polar interaction surface. Thr263 (-6.09 kcal/mol), Phe290 (-5.70 kcal/mol), Phe309 (-4.55 kcal/mol), and Arg183 (-4.54 kcal/mol) constituted the major stabilizing core of HP0973, supported by Val44 (-3.90 kcal/mol), Lys17 (-3.68 kcal/mol), Pro292 (-3.45 kcal/mol), Thr288 (-3.20 kcal/mol), Lys261 (-3.17 kcal/mol), Thr264 (-2.61 kcal/mol), Ser287 (-2.09 kcal/mol), and Lys42 (-2.10 kcal/mol). Collectively, the 200 ns decomposition profile suggested that complex stability is maintained by well-defined electrostatic and hydrophobic hotspots, reflecting a structurally optimized and energetically persistent interface.

Discussion

H. pylori relies on a broad arsenal of secreted and surface-exposed factors to remodel the gastric niche, evade immunity, and promote oncogenic

transformation. Within this context, the *H. pylori* secretome still contains numerous hypothetical proteins whose functions remain unknown. The present work advances this field by providing an integrated *in silico* characterization of HP0973, combining sequence analysis, structural modelling, protein–protein interaction mapping and atomistic simulations with EGFR. HP0973 was identified as a conserved, secreted, and putatively virulent protein of *H. pylori* that can form a stable complex with EGFR, thereby suggesting a previously unrecognized mechanism by which *H. pylori* may modulate host signalling.

Physicochemical characterization revealed that mature HP0973 is a strong basic protein with a theoretical pI of 9.32. Proteins with higher pI values enriched with positive residues are mainly associated with negatively charged biomolecules, such as nucleic acids, phospholipid membranes, or acidic protein surfaces[34]. The higher proportion of positively charged to negatively charged residues in HP0973 may indicate its possible involvement in macromolecular complex formation. Such electrostatic complementarity is particularly relevant in host-pathogen systems, where bacterial effector or surface-associated proteins often exploit charge-driven interactions to engage host receptors or signalling molecules[35]. The instability index (48.84) classified HP0973 as unstable, whereas the long half-life and moderate aliphatic index indicated reasonable thermal stability, suggesting that the protein is likely stabilized *in vivo*, potentially through interactions with binding partners or localization within specific cellular compartments. In addition, a high hydropathy (GRAVY) value reflects the strong hydrophilic nature of HP0973.

Virulence profiling tools converged on the view that HP0973 is antigenic yet non-allergenic and non toxic, but likely to be virulent. VaxiJen 2.0 predicted HP0973 as a probable protective antigen, whereas the Kolaskar-Tongaonkar method identified several antigenic determinants within HP0973, which may be involved in the recognition by the host immune system. At the same time, AllerTOP and CSM-toxin indicated low allergenic and toxic potential, features that are often favoured in vaccine candidates. Importantly, VirulentPred 2.0, which applies updated machine-learning models to large datasets of experimentally validated virulent and non-virulent proteins, classified HP0973 as virulent, suggesting that this protein contributes to pathogenic processes rather than acting as a benign surface antigen. Together, these observations position HP0973 as a potentially exploitable antigen that may both participate in virulence and serve as a target for immunological intervention.

Multiple sequence alignment and phylogenetic analyses revealed that HP0973 is highly conserved among *H. pylori* strains but more divergent in other *Helicobacter* species, implying a species-specific function possibly linked to human gastric colonization. The conservation of hypothetical proteins within *H. pylori* has previously been considered as an indicator of essential or niche-adapted roles, particularly when such proteins cluster with genes involved in cell envelope biogenesis or stress responses[36]. Phylogenetic analysis showed that HP0973 from *H. pylori* strain 26695 clusters tightly with homologs from other *H. pylori* strains, with high bootstrap values and low branch lengths. In contrast, homologs from other *Helicobacter* species have diverged into more distant clades, suggesting evolutionary differentiation. Such a pattern suggests that HP0973 may have evolved species-specific functions that contribute to the adaptation, persistence, or virulence of *H. pylori* in a unique manner in the human gastric niche. The identification of an SLT domain in HP0973 supports a probable role in peptidoglycan remodelling or cell wall turnover, while also raising the possibility of a function in host interaction, as reported for several bacterial proteins containing SLT-like domains[37].

Functional association analysis placed HP0973 at the centre of a dense interaction network enriched for proteins involved in cell wall remodelling, carbohydrate metabolism, translation and stress responses. STRING annotation highlighted its association with phosphoglycerate mutase (HP0974), GatC (HP0975), and glycyl-tRNA synthetase (HP0972), suggesting a role in coordinating cell metabolic activity with protein biosynthesis. Notably, HP0973 was also associated with CopP, a copper-binding stress response protein, suggesting its potential role in adaptation to metal stresses, one of the key survival strategies of *H. pylori* under gastric conditions[38].

Studies of the secondary and tertiary structures showed that HP0973 mainly consisted of an α-helical fold with limited β-structure and some moderately flexible regions, a folding pattern often observed in

proteins involved in enzymatic regulation and protein–protein interactions. The application of comparative structure prediction using SWISS-MODEL, I-TASSER, and AlphaFold indicated that model quality differed significantly between the modelling pipelines. Among these, the SWISS-MODEL predicted structure demonstrated better stereochemical integrity following refinement and validation. The predominance of α helical content, limited β□sheets, and substantial coil regions mirrors the architecture of many secreted enzymes and binding proteins that combine folded cores with flexible segments for partner recognition[39]. Interface prediction algorithms (SPPIDER, MetaPPISP, and PeSTo) consistently highlighted a set of N terminal residues as likely interaction hotspots, providing candidate regions that may mediate binding to host or bacterial partners. These residues overlapped with conserved segments from the ConSurf analysis, providing further support for their functional importance.

A central finding of this work is the predicted interaction between HP0973 and the extracellular domain of EGFR. EGFR is a well-characterized receptor tyrosine kinase that is activated by several endogenous ligands, including EGF, (TGF-α), amphiregulin, epiregulin, and heparin-binding EGF-like growth factor (HB-EGF)[8]. Binding of these ligands to the extracellular domains I and III of EGFR induces a conformational rearrangement that enables receptor dimerization and autophosphorylation of intracellular tyrosine residues and activation of downstream signalling pathways linked to gastric carcinogenesis, particularly in the context of chronic *H. pylori* infection[40]. In contrast, direct binding of a secreted *H. pylori* protein to the ligand binding cleft of EGFR has not been described. The present analysis identified HP0973 as a plausible EGFR interacting protein. Fujiwara et al., have reported that *H. pylori* culture supernatant, which has been dialyzed to remove molecules smaller than 10 kD, significantly reduced binding of EGF to its receptor and also reduced EGF-induced gastric epithelial cell proliferation[41]. Protein A, a surface virulence factor of *S. aureus*, has been reported to directly bind to EGFR, triggering receptor phosphorylation and downstream ERK activation without the requirement for canonical growth factor engagement[10]. Molecular docking of HP0973 with the ectodomain of EGFR chain A in the monomeric form, in the absence of EGF, showed a well-defined interaction interface stabilized by approximately 20 polar contacts, implying a strong protein–protein interaction. The binding interface involved several positively charged and polar residues of HP0973, namely Lys233, Lys267, Lys270, Lys271, Lys282, Lys286, His279 and Gln277 interacting with EGFR key residues Glu2, Asp22, Asn32, Tyr45, Arg353, Ser356, Phe357, His409, and Lys465. Both DI and DIII domain of EGFR were involved in the interaction which are known to mediate ligand binding and receptor activation by growth factors such as EGF and transforming growth factor-α (TGF-α)[42]. Docking of EGF to apo EGFR recapitulated the experimentally determined EGFR-EGF crystal structure, validating the docking protocol. Competition docking experiments revealed that when EGFR was pre bound to EGF, HP0973 no longer accessed the cleft but instead associated with an alternative EGFR surface, whereas pre binding of HP0973 hindered subsequent EGF engagement with its native interface. These observations suggest that HP0973 may act as a modulator of EGFR ligand binding. By occupying or perturbing the ligand binding cleft, HP0973 could either mimic ligand induced activation, block physiological ligands, or induce non canonical receptor conformations, any of which could have profound consequences for downstream signalling.

The MD simulations provided additional mechanistic insights into the stability and dynamics of the HP0973-EGFR complex. Over 200 ns, the complex displayed a reduced RMSD value compared to apo HP0973, indicating that EGFR binding constrains the intrinsic flexibility of HP0973 and promotes a more ordered conformation. RMSF analysis showed a marked reduction in residue level fluctuations at the predicted interface, consistent with tight packing and sustained contacts during the simulation. The long-term stability of HP0973 with the EGFR complex was further supported by a reduced RoG value and a significant number of sustained hydrogen bonds (12.126 ± 4.18 nm). MM/GBSA analysis confirmed favourable binding, yielding a total free energy of -89.64 kcal/mol, mostly driven by electrostatic and van der Waals interactions. Per-residue decomposition identified several hotspot residues on both EGFR and HP0973 that contribute strongly to binding, including Arg29 and Gln28 on EGFR and Val244, Pro292, and Lys318 on HP0973. Together, the MD and energy decomposition results

support a model in which HP0973 is capable of forming a durable, thermodynamically favourable complex with EGFR.

From a pathophysiological perspective, the ability of a secreted *H. pylori* protein to engage EGFR directly raises several intriguing possibilities. First, HP0973 binding could modulate EGFR activation kinetics, either by acting as an agonist like ligand that triggers receptor dimerization and downstream signalling or by serving as an antagonist that blocks normal ligand binding, thereby altering epithelial turnover and repair. EGFR dysregulation has been implicated in gastric epithelial hyperproliferation, metaplasia, and cancer[43]. Inhibition of EGF-induced EGFR activation disrupts the balance between proliferation and apoptosis and promotes inflammatory signalling in the gastric epithelium during *H. pylori* infection. Prolonged impairment of EGFR signalling increases susceptibility to chronic inflammation and carcinogenic progression[44]. A virulent secreted protein, such as HP0973, that targets EGFR could therefore contribute to carcinogenic signalling in a manner complementary to that of CagA and VacA. Second, the predicted association of HP0973 with stress response pathways and cell wall metabolism suggests a dual role whereby HP0973 coordinates bacterial adaptation to hostile gastric conditions while simultaneously manipulating host cell signalling to favour persistence and tissue damage. Third, the antigenicity and surface exposure of HP0973 make it an attractive candidate for vaccine or therapeutic antibody development, especially if future studies confirm its expression during infection and its contribution to disease severity.

Despite the structural and energetic support for the HP0973–EGFR interaction, this study has limitations inherent to *in silico* methodologies. Molecular docking and dynamics simulations depend on static or semi-flexible protein models and predefined force fields, and these models fail to represent the complete range of protein conformational changes, membrane environments, post-translational modifications, and ligand competition that occur in living organisms. Thus, although the computational analyses provide strong evidence for a biologically relevant HP0973-EGFR interaction, these findings require experimental validation.

Conclusion

This integrative analysis proposes HP0973 as a previously uncharacterized virulent component of the *H. pylori* secretome that is conserved across strains, antigenic and capable of forming a strong and dynamically favourable complex with EGFR. By linking bacterial stress adaptation, cell wall associated functions, and host growth factor receptor modulation, HP0973 may represent a missing piece in the complex puzzle of *H. pylori* driven gastric pathogenesis. Future experimental work validating HP0973-EGFR interaction, dissecting downstream signalling effects and assessing HP0973's contribution to colonization and oncogenesis will be crucial to determine whether this protein can serve as a biomarker or therapeutic target in *H. pylori* associated disease.

Acknowledgement

Financial support under 'Ramalingaswami Re-Entry Fellowship' (BT/RLF/Re-Entry/65/2017), Department of Biotechnology, Government of India and Mukhyamantri Research and Innovation Program 2024 (278/249/OSHEC), Government of Odisha is gratefully acknowledged. Debabrata Dash is a recipient of Innovation in Science Pursuit for Inspired Research (INSPIRE) Fellowships, Department of Science and Technology, Government of India. Infrastructure support from the World Bank-OHEPEE (Odisha Higher Education Program for Excellence and Equity), Department of Higher Education, Govt of Odisha to the "Centre of Excellence on Bioprospecting of Ethno-pharmaceuticals of Southern Odisha (CoE-BESO)" to the Department of Biotechnology, Berhampur University is sincerely acknowledged.

Conflict of interest

All authors declare no conflicts of interest.

References

1 Malfertheiner P, Camargo MC, El-Omar E, Liou JM, Peek R, Schulz C, Smith SI & Suerbaum S, *Helicobacter pylori* infection. *Nat Rev Dis Primers*, 9 (2023) 19.

2 Chen YC, Malfertheiner P, Yu HT, Kuo CL, Chang YY, Meng FT, Wu YX, Hsiao JL, Chen MJ, Lin KP, Wu CY, Lin JT, O'Morain C, Megraud F, Lee WC, El-Omar EM, Wu MS & Liou JM, Global Prevalence of *Helicobacter pylori* Infection and Incidence of Gastric Cancer Between 1980 and 2022. *Gastroenterology*, 166 (2024) 605.

3 Clyne M & T OC, Pathogenicity and virulence of *Helicobacter pylori*: A paradigm of chronic infection. *Virulence*, 16 (2025) 2438735.

4 Zanotti G & Cendron L, Structural and functional aspects of the *Helicobacter pylori* secretome. *World J Gastroenterol*, 20 (2014) 1402.

5 Smith SM, Role of Toll-like receptors in *Helicobacter pylori* infection and immunity. *World J Gastrointest Pathophysiol*, 5 (2014) 133.

6 Kwok T, Zabler D, Urman S, Rohde M, Hartig R, Wessler S, Misselwitz R, Berger J, Sewald N, König W & Backert S, Helicobacter exploits integrin for type IV secretion and kinase activation. *Nature*, 449 (2007) 862.

7 Arienti C, Pignatta S & Tesei A, Epidermal Growth Factor Receptor Family and its Role in Gastric Cancer. *Front Oncol*, 9 (2019) 1308.

8 Harris RC, Chung E & Coffey RJ, EGF receptor ligands. *Exp Cell Res*, 284 (2003) 2.

9 Keates S, Keates AC, Katchar K, Peek RM, Jr. & Kelly CP, *Helicobacter pylori* Induces Up-Regulation of the Epidermal Growth Factor Receptor in AGS Gastric Epithelial Cells. *J Infect Dis*, 196 (2007) 95.

10 Gómez MI, Seaghdha MO & Prince AS, Staphylococcus aureus protein A activates TACE through EGFR-dependent signaling. *Embo J*, 26 (2007) 701.

11 Chaturvedi R, Asim M, Piazuelo MB, Yan F, Barry DP, Sierra JC, Delgado AG, Hill S, Casero RA, Jr., Bravo LE, Dominguez RL, Correa P, Polk DB, Washington MK, Rose KL, Schey KL, Morgan DR, Peek RM, Jr. & Wilson KT, Activation of EGFR and ERBB2 by *Helicobacter pylori* results in survival of gastric epithelial cells with DNA damage. *Gastroenterology*, 146 (2014) 1739.

12 Kim N, Weeks DL, Shin JM, Scott DR, Young MK & Sachs G, Proteins released by *Helicobacter pylori in vitro*. *J Bacteriol*, 184 (2002) 6155.

13 Salama NR, Shepherd B & Falkow S, Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. *J Bacteriol*, 186 (2004) 7926.

14 Muller SA, Findeiss S, Pernitzsch SR, Wissenbach DK, Stadler PF, Hofacker IL, von Bergen M & Kalkhof S, Identification of new protein coding sequences and signal peptidase cleavage sites of *Helicobacter pylori* strain 26695 by proteogenomics. *J Proteomics*, 86 (2013) 27.

15 Savojardo C, Martelli PL, Fariselli P & Casadio R, DeepSig: deep learning improves signal peptide detection in proteins. *Bioinformatics*, 34 (2018) 1690.

16 Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G & Nielsen H, SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol*, 40 (2022) 1023.

17 Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD & Hochstrasser DF, Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*, 112 (1999) 531.

18 Doytchinova IA & Flower DR, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8 (2007) 4.

19 Dimitrov I, Bangov I, Flower DR & Doytchinova I, AllerTOP v.2--a server for in silico prediction of allergens. *J Mol Model*, 20 (2014) 2278.

20 Sharma A, Garg A, Ramana J & Gupta D, VirulentPred 2.0: An improved method for prediction of virulent proteins in bacterial pathogens. *Protein Sci*, 32 (2023) e4808.

21 Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ & von Mering C, The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*, 49 (2021) D605.

22 Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, Margalit H, Armstrong J, Bairoch A, Cesareni G, Sherman D & Apweiler R, IntAct: an open source molecular interaction database. *Nucleic Acids Res*, 32 (2004) D452.

23 Laskowski RA, Rullmannn 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*, 8 (1996) 477.

24 Colovos C & Yeates TO, Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*, 2 (1993) 1511.

25 Wiederstein M & Sippl MJ, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*, 35 (2007) W407.

26 Porollo A & Meller J, Prediction-based fingerprints of protein-protein interactions. *Proteins*, 66 (2007) 630.

27 Qin S & Zhou HX, meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics*, 23 (2007) 3386.

28 Krapp LF, Abriata LA, Cortes Rodriguez F & Dal Peraro M, PeSTo: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. *Nat Commun*, 14 (2023) 2175.

29 Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D & Vajda S, The ClusPro web server for protein-protein docking. *Nat Protoc*, 12 (2017) 255.

30 Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B & Lindahl E, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2 (2015) 19.

31 Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, Wei S, Buckner J, Jeong JC, Qi Y, Jo S, Pande VS, Case DA, Brooks CL, 3rd, MacKerell AD, Jr., Klauda JB & Im W, CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput*, 12 (2016) 405.

32 Kolaskar AS & Tongaonkar PC, A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*, 276 (1990) 172.

33 Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V, Chemama Y, Labigne A & Legrain P, The protein-protein interaction map of *Helicobacter pylori*. *Nature*, 409 (2001) 211.

34 Tokmakov AA, Kurotani A & Sato KI, Protein pI and Intracellular Localization. *Front Mol Biosci*, 8 (2021) 775736.

35 Zhou HX & Pang X, Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem Rev*, 118 (2018) 1691.

36 Kim KK & Kim HB, Protein interaction network related to *Helicobacter pylori* infection response. *World J Gastroenterol*, 15 (2009) 4518.

37 Chaput C, Labigne A & Boneca IG, Characterization of *Helicobacter pylori* lytic transglycosylases Slt and MltD. *J Bacteriol*, 189 (2007) 422.

38 Park SJ, Jung YS, Kim JS, Seo MD & Lee BJ, Structural insight into the distinct properties of copper transport by the *Helicobacter pylori* CopP protein. *Proteins*, 71 (2008) 1007.

39 Gazi AD, Kokkinidis M & Fadouloglou VE, alpha-Helices in the Type III Secretion Effectors: A Prevalent Feature with Versatile Roles. *Int J Mol Sci*, 22 (2021).

40 Costa AM, Ferreira RM, Pinto-Ribeiro I, Sougleri IS, Oliveira MJ, Carreto L, Santos MA, Sgouras DN, Carneiro F, Leite M & Figueiredo C, *Helicobacter pylori* Activates Matrix Metalloproteinase 10 in Gastric Epithelial Cells *via* EGFR and ERK-mediated Pathways. *J Infect Dis*, 213 (2016) 1767.

41 Fujiwara Y, Wyle F, Arakawa T, Domek MJ, Fukuda T, Kobayashi K & Tarnawski A, *Helicobacter pylori* culture supernatant inhibits binding and proliferative response of human gastric cells to epidermal growth factor: implications for H.pylori interference with ulcer healing? *Digestion*, 58 (1997) 299.

42 Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim JH, Saito K, Sakamoto A, Inoue M, Shirouzu M & Yokoyama S, Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell*, 110 (2002) 775.

43 Du Y, Karatekin F, Wang WK, Hong W & Boopathy GTK, Cracking the EGFR code: Cancer biology, resistance mechanisms, and future therapeutic frontiers. *Pharmacol Rev*, 77 (2025) 100076.

44 Polk DB & Peek Jr. RM, *Helicobacter pylori*: gastric cancer and beyond. *Nat Rev Cancer*, 10 (2010) 403.