

CHARACTERIZATION OF GASTRIC CANCER ASSOCIATED VIRULENCE FACTOR CAGPAI OF *HELICOBACTER PYLORI*: AN IN-SILICO STUDY

MUCCEE F^{1*}, RAZZAQ F², AYUBE N³, IQBAL R⁴, RAFIQUE F⁵

¹School of Biochemistry and Biotechnology, University of Punjab, Lahore 52254, Pakistan

²Institute of Molecular Biology and Biotechnology, The University of Lahore, Pakistan

³Department of Biochemistry and Biotechnology, The Islamia University of the Bahawalpur, Pakistan

⁴Department of Zoology, Government College University Lahore, Pakistan

⁵Faculty of Allied Health Sciences (FAHS), Superior University, Raiwind Road, Lahore, Pakistan

*Corresponding author email address: fatima.sbb@pu.edu.pk

(Received, 27th October 2023, Revised 20th December 2023, Published 23rd December 2023)

Abstract Gastric cancer (GC) is reported to be the sixth most prevalent cancer all over the world. *Helicobacter pylori* is one of its causative agents. Its virulence factor cytotoxin-associated gene pathogenicity island (cagPAI) induces GC via causing gastritis, DNA breakage, inhibiting p53 gene, and stimulating pathogenicity pathways. Present study has targeted this pathogenic factor. Sequences of cag1, cag2, cag5, cag6 and cag8 isoforms were retrieved from Uniprot database and subjected to in-silico tools for characterization. Tools included CELLU, ProtParam, AlphaFold database, MEME suite, and STRING tool. Analysis revealed all the cag isoforms to be extracellular. All were diverse for molecular weight, isoelectric point (pI), aliphatic index, instability index, and GRAVY. Most of the conserved motifs and the interacting proteins were common among most of the proteins. Deviation in three-dimensional (3D) configuration was also recorded. This study might help design drugs against *H. pylori* pathogenicity factors and inhibiting cagPAI genes' interactions with other proteins. Thus, might help in reducing GC incidence.

Keywords: gastric cancer; *Helicobacter pylori*; metastasis; conserved motif; physicochemical properties

Introduction

Gastric cancer (GC) initiates in the mucosal lining of the stomach, resulting in a malignant tumor in liver and lymph nodes (Ayyildiz et al., 2020). According to an estimation, 750,000 deaths have been caused by GC (Sung et al., 2021). Different stages of this cancer are Stage 0, also known as carcinoma in-situ, stage I (divided into stage IA and stage IB), stage II (stage IIA and stage IIB), stage III (stage IIIA, IIIB, and IIIC); and stage IV. Early stage is characterized by weight loss, poor appetite, abdominal pain, heartburn, vomiting, nausea and final stage symptoms include blood in stool, trouble in swallowing, stomach pain, vomiting, ascites, looseness of the bowels, and jaundice (Bhardwaj et al., 2022). Prognosis and early diagnosis of GC is challenging as the early symptoms are not prominent. Patients are diagnosed after metastasis in more than 50% cases (Li et al., 2023a; Li et al., 2023b). GC is caused by dysregulation of HCl secretion, salt-preserved and pickled food, tobacco and genetics (Crafa et al., 2023; Helisz et al., 2023). Additionally, the ecological niche of alimentary canal matters a lot for human health. It undergoes dysbiosis in pathological conditions. Pathogenic ones

replace the normal healthy bacteria. Hence, dysbiosis is one of the contributing factors. Several studies reported the presence of *Helicobacter pylori* in human gut in GC patients (Amalia et al., 2023; AMJAD et al., 2018; Han et al., 2023; Ishtiaq et al., 2019; Ji et al., 2023; Kesharwani et al., 2023; Murata-Kamiya and Hatakeyama, 2022; Reyes, 2023; Salvatori et al., 2023; Shirani et al., 2023; Usui et al., 2023). The bacterium is responsible for GC prevalence in under-developed and developed countries at the rate of 85-95% and 30-50%, respectively (Puculek et al., 2018).

H. pylori has also been declared a major cause of GC by International Agency for Research on Cancer (IARC) (Chiang et al., 2021; Stewart et al., 2020). It facilitates ulcer in stomach and adenocarcinoma by causing inflammation and superficial gastritis. This bacterium also causes double-stranded DNA breaks (DSB) and homologous recombination (HR) abnormalities. The HR defects results in non-HR regulated DSB repair (Murata-Kamiya and Hatakeyama, 2022). In addition to the above effects, it also causes oxidative and endoplasmic reticulum (ER) stress. It enhances inflammation through

increasing expression of genes Smad7, ROS, RNS, IL-17, IL-21, and NF- κ B (Salvatori et al., 2023). Virulence of *H. pylori* is contributed by several virulence factors, including cytotoxin associated gene pathogenicity island (cagPAI), urease, flagellum, vacuolating cytotoxin A (vacA), catalase, superoxide dismutase, lewis antigens, arginase, phospholipases, lipopolysaccharide, blood group antigen binding adhesion, sialic acid binding adhesion, outer inflammatory protein A, duodenal ulcer promoting gene A, adherence associated lipoprotein A and B, outer membrane protein Q, outer membrane protein Z, cholesteryl α -glucosyltransferase, γ -glutamyl transpeptidase, neutrophil-activating protein and heat shock proteins. These are frequently associated with protecting pathogens from the stomach environment, proliferation, colonization, and carcinogenesis. The cagPAI is a 40 kb DNA segment and a cluster of 27-31 genes that encodes for Type IV secretion system (T4SS) and cagA oncoprotein (Argent et al., 2008). With the help of this secretory system, cagA is injected into host cells. Once inside cells, it causes the previously mentioned abnormalities at DNA level (Reyes, 2023). Literature also reports that cagPAI in *H. pylori* accelerates GC-associated gene expression by inducing transcription (Ahmadzadeh et al., 2023; Amin et al., 2023; Fatima et al., 2023). These alterations in oncogenes are the hallmarks of GC. Considering this GC association of cagPAI, we designed present research project to characterize isoforms of its different genes i. e. cag1, cag2, cag5, cag6, and cag8. This might help in making strategies to control *H. pylori* through site-directed mutagenesis or by designing vaccines complementary to its most susceptible genes.

Methodology

Retrieving the sequences from Uniprot database

To characterize cagPAI variants, sequences were retrieved from Uniprot database (www.uniprot.org, accessed in Sep. 2023). Total five cagPAI genes i. e. cag1, cag2, cag5, cag6 and cag8 were selected. Two variants of each of these genes were considered for analysis. And for convenience, these were designated as A and B i. e. cag1A, cag1B, cag2A, cag2B,

cag5A, cag5B, cag6A, cag6B, cag8A and cag8B (Supplementary data Table 1).

Prediction of sub-cellular localization by CELLU tool

CELLU v.2.5: subcellular Localization predictor (cello.life.nctu.edu.tw, accessed in Sep. 2023) was consulted to assess the sub-cellular location of cagPAI variants documented in the present study.

Prediction of physicochemical properties by Protparam tool

To predict the physicochemical properties, PROTPARAM tool (<https://web.expasy.org/protparam/>, accessed in Sep. 2023) was used.

Prediction of conserved protein motif analysis by MEME suite

To determine the conserved domains in the cag variants of *H. pylori*, the online Multiple Em for Motif Elicitation (MEME) suite (<http://meme.sdsc.edu/meme/meme.html>, accessed in Sep. 2023) was consulted.

Prediction of 3D configuration by AlphaFold

For the determination of variations among the cagPAI isoforms at the level of 3D configuration, AlphaFold protein structure database (<https://alphafold.ebi.ac.uk>, accessed in Sep. 2023) developed by EMBL-EBI was used.

Prediction of interactions with proteins by STRING tool

Search tool for retrieval of interacting genes/proteins i. e. STRING (string91.embl.de/newstring/cgi/show_input_page.pl, accessed in Sep. 2023) was consulted to assess the possible proteins with which the isoforms of cagPAI might interact.

Results

Assessment of sub-cellular localization

Analysis of sub-cellular localization of cagPAI protein isoforms revealed that they were extracellular. Highest scores were observed in extracellular localization i. e. from 1.482 to 3.373. In addition to this, all the variants except cag6B were also found to localize in the outer-membrane because of the scores recorded (1.083 to 1.728). However, the score values were smaller than those for extracellular location (Table 1).

Table 1: Sub-cellular localization of cagPAI gene variants predicted using CELLU tool

cagPAI variants	Extracellular	Outer membrane	Periplasmic	Cytoplasmic	Inner membrane
cag1A	1.482	1.368	1.257	0.773	0.120
cag1B	1.872	1.083	0.967	0.560	0.518
cag2A	2.276	1.149	0.949	0.436	0.190
cag2B	1.503	1.465	1.360	0.440	0.232
cag5A	1.809	1.728	0.807	0.365	0.291
cag5B	1.963	1.095	0.863	0.767	0.312
cag6A	2.493	1.127	0.730	0.337	0.313
cag6B	3.373	0.511	0.481	0.465	0.170

[Citation Muccee, F., Razaq, F., Ayube, N., Iqbal, R., Rafique, F. (2023). Characterization of gastric cancer associated virulence factor cagPAI of *Helicobacter pylori*: An in-silico study. *Biol. Clin. Sci. Res. J.*, 2023: 617. doi: <https://doi.org/10.54112/bcsrj.v2023i1.617>]

cag8A	2.012	1.153	1.085	0.491	0.259
cag8B	1.531	1.380	1.042	0.806	0.242

Assessment of physicochemical properties

Protparam was used to compute the physicochemical attributes of isoforms. The analysis revealed variations in molecular weights, pI, aliphatic index, instability index, and GRAVY of the proteins. The pI ranged between 5.08 to 9.78. Aliphatic index was

highest (116.07) in cag6B and lowest in cag2B (67.17). Lowest (12.07) and highest (56.51) values for instability index were recorded in cag1B and cag1A, respectively. GRAVY value was negative except in cag5B and cag6B (Table 2).

Table 2: Physicochemical properties of cagPAI gene variants predicted using Protparam tool

Protein	Mol. Wt.	Isoelectric point (pI)	Aliphatic index	Instability index	GRAVY
cag1A	20450.20	8.47	69.10	56.51	-0.534
cag 1B	12424.32	8.58	92.35	12.07	-0.026
cag 2A	30609.78	9.33	74.20	35.60	-0.482
cag 2B	21024.59	9.30	67.17	35.05	-0.667
cag 5A	85819.21	8.62	86.42	33.01	-0.141
cag 5B	7183.68	9.78	107.33	18.99	0.030
cag 6A	23104.30	5.08	91.16	30.72	-0.403
cag 6B	9538.20	5.27	116.07	21.20	0.058
cag 8A	60481.34	9.43	81.27	42.74	-0.794
cag 8B	56230.26	9.34	80.77	45.57	-0.905

Assessment of conserved protein motifs

Conserved motif analysis performed using MEME suite, helped to identify six conserved motifs in each of the cagPAI genes documented in the present study. Total 62 motifs were recognized. Among these, CENTPIM, CEWTASM, FYGCSYD,

CEWTASM, PREYLP, GRKLNEIVQ, DDFGRK, GFNKGW, SMTDAAIDPNMTNSGLRWYRV, SPEDNSIELSPSDSAWRTNLV, and FDDGTFTYFGF motifs were shared by the different component genes of cagPAI (Table 3, Figure 1).

Table 3: Prediction of conserved protein motifs in cagPAI gene isoforms performed using MEME suite

#	Motifs identified		
cag1A			
1	LQSVDTYQSE	DNINFYMPYMNMA YWVFKKEFPSPEYE	DFIRRMQRYS
2	MADTINTAMRVLNASKHYTATGVREPD	ACTKSFKKSAMV	
3	QYSQSALNTN	HGAWGI LYFLTLVKL	
4	LGYLVSQ	NKPYGLKA IEILNAWANEL	
5	FPSPEYEDFI	RRMRQY SQSALNTNHG	
6	PDACTKSFKK	SAMVSY DLALGYLVSQ	
cag1B			
1	NAFVNFFKNS	LVDKRYDSL	LIGAGVLCCV
2	DSLGLIGAGV	LCCVLSGAMG	IVGIIFVAIG
3	ADTINTTETT	HETKKNPNA	FVNFFKNSLV
4	KQPKVATTVN	NETQKSQA	TSVTNEPTEA
5	LVKLVEKLSK	KQPKVA	TTVNNETQKS
6	TQKSQATSV	TNEPTEA	KETKD
cag2A			
1	NGSMKMVCLH	CENTPIM	EKVESGRGGA
2	FFLGCKNYPK	CEWTASM	DSQDLKCPKC
3	KRVRGYLICF	VCNTPKM	IQRGLNGISF
4	MIQRGLNGIS	FYGCSYDV	NKGDCCKGVL
5	VKANIKENSF	FLGCKNYP	KCEWTASMD
6	LLGFPSRKFT	PNEFFT	AVSLTLNAME
7	ADEEKVLLGF	PSRKFT	PNEFFTAVSL
cag2B			
1	ANIKENSFFL	GCENYPKCEWTASMD	SQDLKCPKCN
2	INGSMKMVCL	HCENTPIMEKVESGRG	GAYACKNCNR
			KFYFIDLAKQ

[Citation Muccee, F., Razaq, F., Ayube, N., Iqbal, R., Rafique, F. (2023). Characterization of gastric cancer associated virulence factor cagPAI of *Helicobacter pylori*: An in-silico study. *Biol. Clin. Sci. Res. J.*, 2023: 617. doi: <https://doi.org/10.54112/bcsrj.v2023i1.617>]

3	VNKGDCCKGVL REINGSM KMVCLHCENT
4	TEKQKIKRLE RFILASV KANIKENSFF
5	MNSVL FYGCSDY VNKGDCCKGVL
6	RKKNFKNNEF FTATSLT LNIMEFCLYI
cag5A	
1	MIYSNLILPI HDPQCKRSC LMLMDEFTLC
2	TQKKVFKDKA NQPQKKKSF KEIIDGLKE
3	PKSDDDLFDI WVYAIQ DFPAYYFKAL
4	LKERVKTFGF WLQAIL LLSYSFITSG
5	ELIILENTLK PIKCHK ALYYDDPFPT
6	GLIALFFLYK FIKTQK KVFKDKANQP
cag5B	
1	MEDFLYNT LYFIEDYKLV
2	EDFLYNTLYF IEDYKLVV IFSFIGLIAL
3	IKAQKKALKD KANQPQ KKKTLKKS
4	IGLIALFFLY KFIKAQ KKALKDKANQ
5	YFIEDYKLVV IFSFIG LIALFFLYKF
6	LVVIFSFIGL IALFFL YKFIKAQKKA
cag6A	
1	EISDNFNNFM CDEVARISD LVASYLPREY
2	ESTIKNELTL GEFVAEISD NFNFMDEV
3	YYNYLNNLVL ASYNRCKQEKTF AESTIKNELT
4	DEVARISDLV ASYLPREYLPF IDGNMMGVAF
5	MGVAFQILGI DDFGRK LNEIVQDIGT
6	DFGRKLNEIVQDIGTKYIILSKNKTY
cag6B	
1	M CDEVAK ISDLVASYLP
2	MCDEVAKI SDLVAS YLPREYLPF
3	ILSKNKTYLTSLERAKLITQLKLNLE
4	PREYLPFFID GNMMGVAFQILGIDDFGRK
5	AFQILGIDDF GRKLNEIVQ DIGTKYIILS
6	KISDLVASYL PREYLP PFIDGNMMGV
7	IGTKYIILSK NKTYLT SLERAKLITQ
cag8A	
1	EQAFFKKIVN CFCLGY LFLSGVIEAA
2	KDETISYITT GFNKGW NIVPNSNHIF
3	IFVVQPDGKL SMTDAAIDPNMTNSGLRWYRV NEIAEKFKLI
4	ISIKTDKSQK SPEDNSIELSPSDSAWRTNLV VRTNKALYQF
5	RSKHIMPSEI FDDGTFTYFGF KNITLQPAIF
6	NLMFEKEAVN FALMTRDYQEF LKTKKLIVDA
cag8B	
1	YYQAPEKRSK HIMPSEIFDDGTFTYFGFKNITLQP AIFVVQPDGK
2	TVIQLEKDET ISYITTGFNKGWNIVPNSNHIFIQP KSVKSNLMFE
3	IFVVQPDGKL SMTDAAIDPNMTNSGLRWYRV NEIAEKFKLI
4	ISIKTDKSQK SPEDNSIELSPSDSAWRTNLV VRTNKALYQF
5	ASAYLTVKLE YPQRHEVSS VIEEELKKRE
6	KRNPVPRNYN YYQAPEKRS KHIMPSEIFD

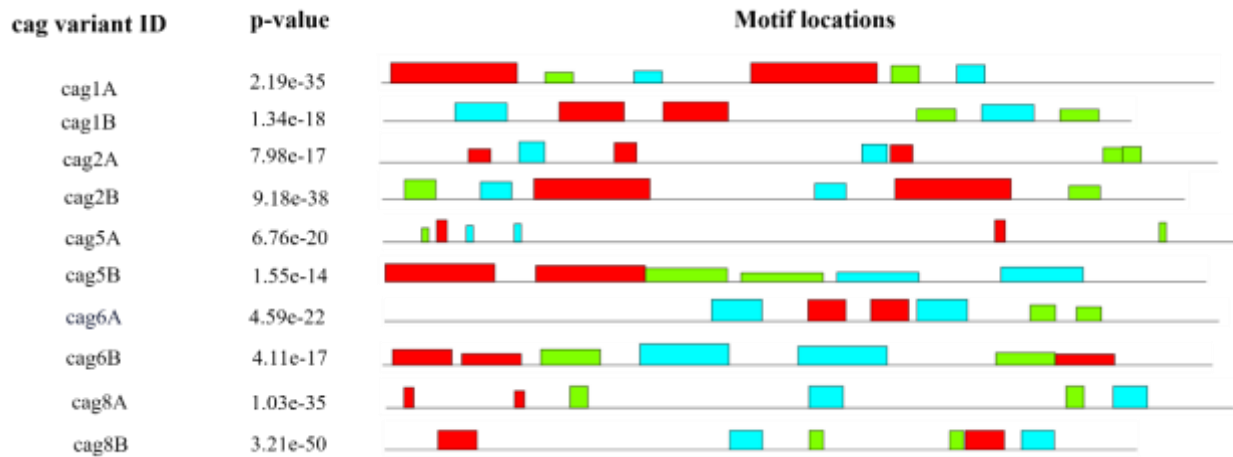


Figure 1: Conserved protein motifs identified in cagPAI gene variants using MEME suite

Assessment of 3D configuration

The comparison of 3D conformations of cagPAI protein isoforms revealed great variation among them. Simplest level of folding was observed in cag1B and cag5B. Highest complex folding was

observed in cag5A. The cag2A, cag2B, and cag8A showed slight similarity in 3D structure (Figure 2).

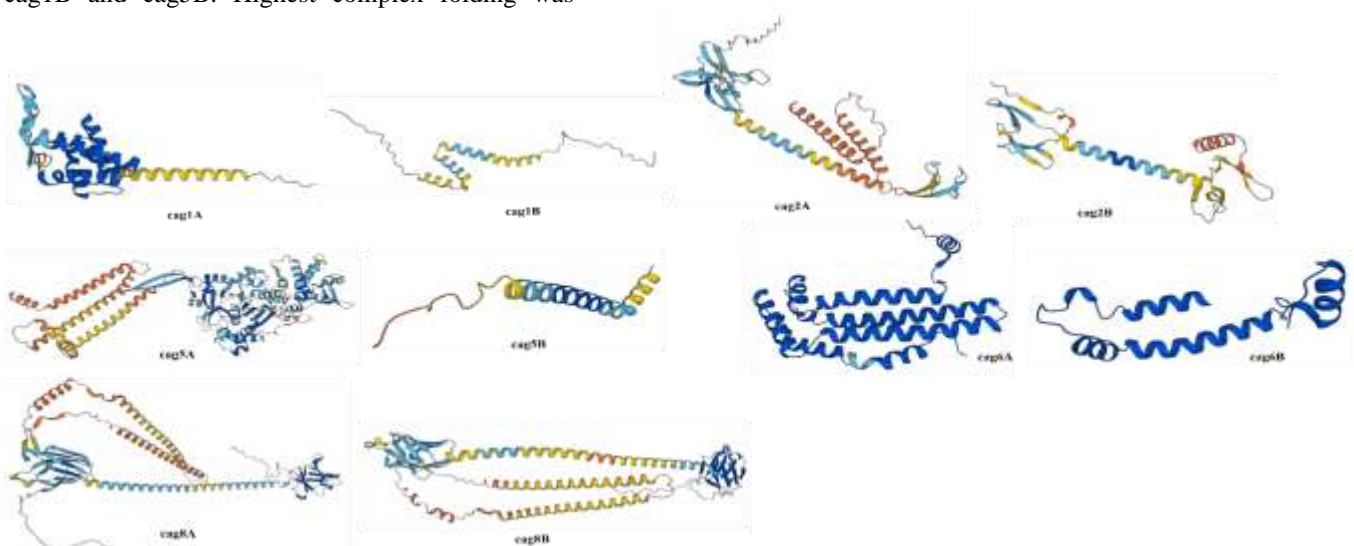


Figure 2: Three dimensional (3D) configuration analysis of cagPAI variants predicted using AlphaFold database

Interactions of cagPAI with other proteins

Interaction analysis using STRING tool revealed interactions of cag1A, cag1B, and cag2A with the same proteins i. e. cag3 (cag pathogenicity island proteins), cag4 (cage island protein), cagA (cytotoxicity immunodominant antigen), cagE (DNA transfer protein), vacA (vaculating cytotoxin), orf6 (putative vag island protein), rpoB (DNA directed RNA polymerase subunit beta) rpoA (DNA directed RNA polymerase subunit alpha) and HP_0333 (hypothetical protein). The cag2B and cag5B exhibited interaction with common proteins. i. e. cag alfa & cagE (cag island DNA transfer protein), 0527 (cag island protein 0527), comB3, comB2, and

verB4_2 (DNA transformation competency proteins) and cag8, cagV and cag6 (cag island proteins). The cag5A and cag6A have the same interaction pattern i. e. cag-alfa, cagE, cag7 (cag pathogenicity island), verB10_2 & verB10-1 (type 4 secretion proteins), comB10 and comB4 (competence protein), cag8, cag-V and cag6. Interaction with the same proteins was recorded in cag6B and cag8A i. e. caf-alfa, cag5, cag7, cag8, cag9, and cagV (cag pathogenicity island protein). The cag8B interacted with 0527, cagV, cagT, cag-alfa, cagA, cag5, cagM, cagE, verB11 and verB4_3 (Figure 3).

[Citation Muccee, F., Razaq, F., Ayube, N., Iqbal, R., Rafique, F. (2023). Characterization of gastric cancer associated virulence factor cagPAI of *Helicobacter pylori*: An in-silico study. *Biol. Clin. Sci. Res. J.*, 2023: 617. doi: <https://doi.org/10.54112/bcsrj.v2023i1.617>]

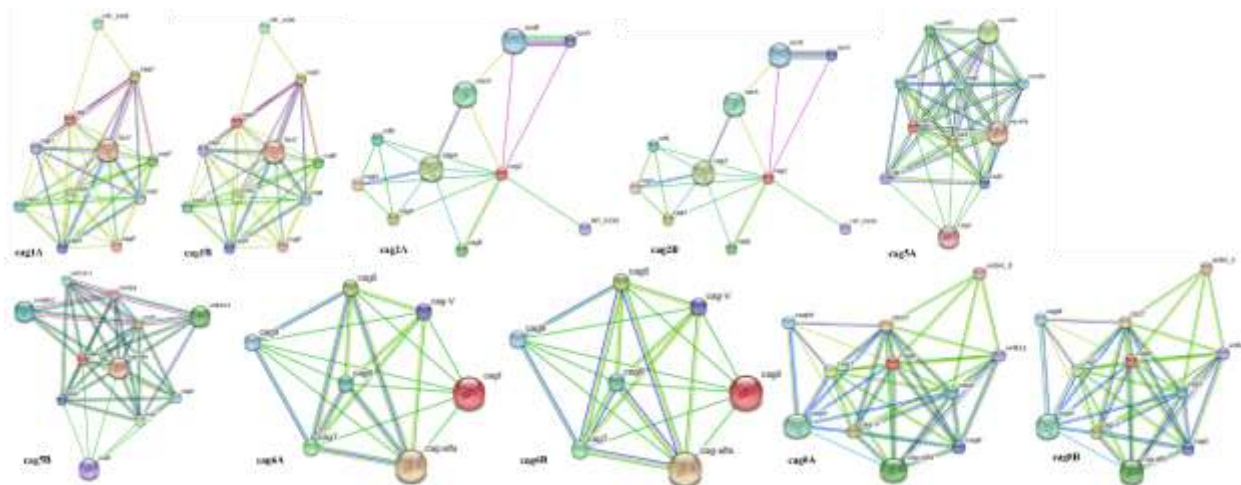


Figure 3: Interpretation of cagPAI gene variants interaction with other proteins using STRING tool

Discussion

Although several studies have characterized cagPAI island to identify the targets for drug designing (Demirci et al., 2021; Nammi et al., 2017; Taş et al., 2020). However, no one has characterized the component genes of cagPAI at the levels discussed in the present study. The pI indicates a protein's alkalinity or acidity status (Lautenbach et al., 2021). All the cagPAI isoforms were alkaline (pI = 8.47 to 9.78) except cag6 (pI = 5.08 & 5.27). Aliphatic index is proportional to protein's thermal stability (Sadeghi et al., 2006). According to present study findings, all the cagPAI variants were thermally stable, with the highest and lowest aliphatic index values of 116.07 and 67.17, respectively. According to the instability index, all the cag isoforms except cag1A, cag8A, and cag8B are stable *in-vitro* with values ranging between 12.07 to 35.60. Regarding the GRAVY, all the cag isoforms are non-polar except cag5B and cag6B (Babnigg and Joachimiak, 2010). The analysis of physicochemical properties showed that the cagPAI is a highly stable virulence factor, and it is impossible to denature it easily. Conserved motifs are the evolutionarily conserved signatures (Patel, 2017). These motifs can be tested for their immunogenicity status and vaccines can be designed against cagPAI isoforms. As per the findings, isoforms of cgaPAI shared the same conserved regions. So, targeting these regions for vaccination might inhibit multiple genes simultaneously.

Great variation has been observed in cag isoforms at the level of 3D structure. Most frequently GC-associated protein configuration might be confirmed by analyzing the *H. pylori* genomes from GC patients. This will help in identifying the specific genes for manipulation and vaccine designing. STRING tool helped in analyzing the proteins with which the cag isoforms interact to cause pathogenicity. These included cagV, cagT, cag-alfa, cagA, cag5, cagM, cagE, verB11 and verB4_3 etc. So, targeting the common interacting proteins for the

cag variants will help us to identify and then interfere with the most susceptible pathways for the inhibition of pathogenesis.

Conclusion

The conserved motifs shared by different cag isoforms can be exploited as novel drug targets. Present study findings of 3D configuration and protein motifs might also be helpful in the selection of regions for site-directed mutagenesis. This will lead to inhibition of pathogenic factors of *H. pylori*, thus reducing the GC incidence.

References

- Ahmadzadeh, A., Rashidi, M., Mohsenifar, Z., Faeghi, F., Rezaei-Tavirani, M., Ahmadi, N., Zali, M. R., Alebouyeh, M., Feizi, M., and Ahmadzadeh, Z. (2023). Examining the effect of *Helicobacter pylori* cagPAI variety on gene expression pattern related to gastric cancer. *Hormone Molecular Biology and Clinical Investigation*.
- Amalia, R., Panenggak, N. S. R., Doohan, D., Rezkitha, Y. A. A., Waskito, L. A., Syam, A. F., Lubis, M., Yamaoka, Y., and Miftahussurur, M. (2023). A comprehensive evaluation of an animal model for *Helicobacter pylori*-associated stomach cancer: Fact and controversy. *Helicobacter* **28**, e12943.
- Amin, F., Hassan, N., Bashir, K., Khan, A., Bibi, H., Irshad, M., Khan, S., Nawaz, K., and Ullah, Z. (2023). Antimicrobial susceptibility profile of various bacteria isolated from respiratory tract infection. *Bulletin of Biological and Allied Sciences Research* **2023**, 48-48.
- Amjad, N., Rehman, M., Khalid, M., and AMJAD, I. (2018). Construction of the infectious molecule of beta satellite associated with ageratum yellow vein disease of *ageratum conyzoides*. *Bulletin of Biological and Allied Sciences Research* **2018**, 13-13.
- Argent, R. H., Thomas, R. J., Letley, D. P., Rittig, M. G., Hardie, K. R., and Atherton, J. C.

- (2008). Functional association between the *Helicobacter pylori* virulence factors VacA and CagA. *Journal of medical microbiology* **57**, 145-150.
- Ayyildiz, O., Aydin, Z., Yilmaz, B., Karaçavuş, S., Şenkaya, K., Icer, S., Taşdemir, E. A., and Kaya, E. (2020). Lung cancer subtype differentiation from positron emission tomography images. *Turkish Journal of Electrical Engineering and Computer Sciences* **28**, 262-274.
- Babnigg, G., and Joachimiak, A. (2010). Predicting protein crystallization propensity from protein sequence. *Journal of structural and functional genomics* **11**, 71-80.
- Bhardwaj, P., Bhandari, G., Kumar, Y., and Gupta, S. (2022). An investigational approach for the prediction of gastric cancer using artificial intelligence techniques: a systematic review. *Archives of Computational Methods in Engineering* **29**, 4379-4400.
- Chiang, T. H., Chang, W. J., Chen, S. L. S., Yen, A. M. F., Fann, J. C. Y., Chiu, S. Y. H., Chen, Y. R., Chuang, S. L., Shieh, C. F., and Liu, C. Y. (2021). Mass eradication of *Helicobacter pylori* to reduce gastric cancer incidence and mortality: a long-term cohort study on Matsu Islands. *Gut* **70**, 243-250.
- Crafa, P., Franceschi, M., Rodriguez-Castro, K. I., Franzoni, L., Russo, M., Brandimarte, G., Tursi, A., Rugge, M., and Di Mario, F. (2023). PPIs and gastric cancer: any causal relationship? *Acta Bio Medica: Atenei Parmensis* **94**.
- Demirci, M., Özge, Ü., and Kocazeybek, B. (2021). In silico analysis of virulence, resistance genes and phylogeny of *Helicobacter pylori* strains from different continents. *Experimed* **11**, 170-178.
- Fatima, S., CHEEMA, K., Shafiq, M., Manzoor, M., Ali, Q., Haider, M., and Shahid, M. (2023). The genome-wide bioinformatics analysis of 1-aminocyclopropane-1-carboxylate synthase (acs), 1-aminocyclopropane-1-carboxylate oxidase (aco) and ethylene overproducer 1 (eto1) gene family of *fragaria vesca* (woodland strawberry). *Bulletin of Biological and Allied Sciences Research* **2023**, 38-38.
- Han, Z., Liu, J., Zhang, W., Kong, Q., Wan, M., Lin, M., Lin, B., Ding, Y., Duan, M., and Li, Y. N. (2023). Cardia and non-cardia gastric cancer risk associated with *Helicobacter pylori* in East Asia and the West: A systematic review, meta-analysis, and estimation of population attributable fraction. *Helicobacter* **28**, e12950.
- Helisz, P., Gwioździk, W., Krupa-Kotara, K., Grajek, M., Głogowska-Ligus, J., and Słowiński, J. (2023). Risk factors and prevention of gastric cancer development—What do we know and what can we do? *Oncology* **3**, 26-42.
- Ishtiaq, M., Atif, M., Manzoor, M., Sarwar, M., and Rafaqat, N. (2019). Analysis of different allelopathic plant extracts and fungal metabolites on rice to control rice grain discoloration. *Bulletin of Biological and Allied Sciences Research* **2019**, 28-28.
- Ji, X., He, G., Wang, K., Zhang, Y., Yin, J., and Wang, K. (2023). Estimation of gastric cancer burden attributable to *Helicobacter pylori* infection in Asia. *Journal of Public Health* **45**, 40-46.
- Kesharwani, A., Dighe, O. R., and Lamture, Y. (2023). Role of *Helicobacter pylori* in Gastric Carcinoma: A Review. *Cureus* **15**.
- Lautenbach, V., Hosseinpour, S., and Peukert, W. (2021). Isoelectric point of proteins at hydrophobic interfaces. *Frontiers in Chemistry* **9**, 712978.
- Li, Y.-N., Xie, B., Zhang, Y., He, M. H., Xing, Y., Mu, D. M., Wang, H., and Guo, R. (2023a). Advances and key focus areas in gastric cancer immunotherapy: A comprehensive scientometric and clinical trial review (1999-2023). *World Journal of Gastroenterology* **29**, 5593-5617.
- Li, Y. N., Xie, B., Zhang, Y., He, M. H., Xing, Y., Mu, D. M., Wang, H., and Guo, R. (2023b). Advances and key focus areas in gastric cancer immunotherapy: A comprehensive scientometric and clinical trial review (1999-2023). *World Journal of Gastroenterology* **29**, 5593-5617.
- Murata-Kamiya, N., and Hatakeyama, M. (2022). *Helicobacter pylori*-induced DNA double-stranded break in the development of gastric cancer. *Cancer Science* **113**, 1909-1918.
- Nammi, D., Yarla, N. S., Chubarev, V. N., Tarasov, V. V., Barreto, G. E., Corolina Pasupulati, A. M., Aliev, G., and Reddy Neelapu, N. R. (2017). A systematic in-silico analysis of *Helicobacter pylori* pathogenic islands for identification of novel drug target candidates. *Current Genomics* **18**, 450-465.
- Patel, S. (2017). Pathogenicity-associated protein domains: The fiercely-conserved evolutionary signatures. *Gene Reports* **7**, 127-141.
- Puculek, M., Machlowska, J., Wierzbicki, R., Baj, J., Maciejewski, R., and Sitarz, R. (2018). *Helicobacter pylori* associated factors in the development of gastric cancer with special reference to the early-onset subtype. *Oncotarget* **9**, 31146.
- Reyes, V. E. (2023). *Helicobacter pylori* and Its Role in Gastric Cancer. *Microorganisms* **11**, 1312.

[Citation Muccee, F., Razaq, F., Ayube, N., Iqbal, R., Rafique, F. (2023). Characterization of gastric cancer associated virulence factor cagPAI of *Helicobacter pylori*: An in-silico study. *Biol. Clin. Sci. Res. J.*, **2023**: 617. doi: <https://doi.org/10.54112/bcsrj.v2023i1.617>]

- Sadeghi, M., Naderi-Manesh, H., Zarrabi, M., and Ranjbar, B. (2006). Effective factors in thermostability of thermophilic proteins. *Biophysical chemistry* **119**, 256-270.
- Salvatori, S., Marafini, I., Laudisi, F., Monteleone, G., and Stolfi, C. (2023). *Helicobacter pylori* and gastric cancer: pathogenetic mechanisms. *International Journal of Molecular Sciences* **24**, 2895.
- Shirani, M., Pakzad, R., Haddadi, M. H., Akrami, S., Asadi, A., Kazemian, H., Moradi, M., Kaviar, V. H., Zomorodi, A. R., and Khoshnood, S. (2023). The global prevalence of gastric cancer in *Helicobacter pylori*-infected individuals: a systematic review and meta-analysis. *BMC Infectious Diseases* **23**, 543.
- Stewart, O. A., Wu, F., and Chen, Y. (2020). The role of gastric microbiota in gastric cancer. *Gut Microbes* **11**, 1220-1230.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **71**, 209-249.
- Taş, S. K., Kırkık, D., Öztürk, K., and Tanoğlu, A. (2020). Determination of B- and T-cell epitopes for *Helicobacter pylori* cagPAI: An in silico approach. *The Turkish Journal of Gastroenterology* **31**, 713.
- Usui, Y., Taniyama, Y., Endo, M., Koyanagi, Y. N., Kasugai, Y., Oze, I., Ito, H., Imoto, I., Tanaka, T., and Tajika, M. (2023). *Helicobacter pylori*, homologous-recombination genes, and gastric cancer. *New England Journal of Medicine* **388**, 1181-1190.

Statements and Declarations

Funding

N/A

Acknowledgements

We acknowledge the School of Biochemistry and Biotechnology (SBB), University of the Punjab, Lahore.

Author contributions

FM, wrote the manuscript, FR perceived the idea, NA and RI retrieved data from the database, designed methodology and performed analysis. All authors approved final version of manuscript to publish.

Informed consent

N/A

Competing interests

The authors have no competing interests.

Data availability statement

The sequences of proteins documented in the present study are available at Uniprot database (www.uniprot.org).

Submission declaration and verification

The work is not been published previously, and it is not under consideration for publication elsewhere.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. © The Author(s) 2023