Journal of Carcinogenesis
Covering Genesis to Therapeutics
Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 9s, https://carcinogenesis.com/

Domain Architecture and Functional Dynamics of Enterococcal Surface Protein (Esp) in Enterococcus faecium: Insights into Endodontic Virulence and Persistence

Hamed A. Alshawkani1, Nezar Boreak[1]*, Aishah haqawi[1], Alhanouf Ali Moafa[1], Ruba Muhammed Ahmed Hanani[1], Lujain S Alsawat[2], Ghazal Y Saati[3], Abed Mohammed Mathkor[1], Hitesh Chohan[1]

[1]Department of Restorative Dental Sciences, College of Dentistry, Jazan University, Jazan 45142, Saudi Arabia.

[2]College of Dentistry, King Abdulaziz University, Jeddah 80209, Saudi Arabia.

[3]Vision College, Jeddah 23643, Saudi Arabia

***Corresponding Author**

Nezar Boreak, Department of Restorative Dental Sciences, College of Dentistry,

Jazan University, Jazan 45142, Saudi Arabia

E-mail: nboraak@jazanu.edu.sa

ABSTRACT

Enterococcus faecium (E.f) is a durable pathogen often linked to persistent endodontic (root canal) infections. It frequently expressions strong resistance to ordinary treatments. The enterococcal surface protein (Esp), which start on the bacterial cell surface, is a key virulence factor. Esp contributes to biofilm formation, tissue attachment, & avoidance of the host immune system. This study presents a comprehensive in silico analysis of Esp, concentrating on its domain structure, physicochemical features, & its importance in endodontic infections.

Insilico analysis with InterProScan & SMART recognized more than a few structural domains within Esp. These include a YSIRK-type signal peptide, Rib domain, repetitive surface antigen motifs, bacterial immunoglobulin-like folds, & a C-terminal LPXTG anchor sequence. Collected, these domains suggest the most important roles in adhesion & host-pathogen interactions. 3-D structural modeling, created with AlphaFold & validated through Ramachandran plot valuation, presented a stable configuration. In total, 88.55% of residues were in energetically favored regions. Supplementary analyses covered hydropathy profiles, atomic & amino acid composition, & conserved motifs imperative for binding to the extracellular matrix & for bacterial persistence. These findings highlight the structural & functional importance of Esp in E. faecium disease. They suggest Esp could be a promising target for new therapies against difficult endodontic infections.

Keywords: *SMART, Esp, Enterococcus faecium, Ramchandran Plot, domain organization, Endodontic.*

How to Cite: Hamed A. Alshawkani, Nezar Boreak, Aishah haqawi, Alhanouf Ali Moafa, Ruba Muhammed Ahmed Hanani, Lujain S Alsawat, Ghazal Y Saati, Abed Mohammed Mathkor, Hitesh Chohan (2025) in Enterococcus faecium: Insights into Endodontic Virulence and Persistence, *Journal of Carcinogenesis, Vol.24, No.9s,* 607-616.

1. INTRODUCTION

During a routine dental examination, a patient, Mr. Johnson, presented with persistent pain in his previously treated tooth. Despite undergoing thorough mechanical instrumentation and chemical irrigation, the root canal infection continued to recur. These issues highlight the significant challenge persistent endodontic (root canal) infections pose in dental practice. *E. faecium* is an exceedingly resilient, opportunistic pathogen that often initiates in these challenging cases. Though *E. faecium* is isolated less frequently than *E. faecalis*, it tolerates harsh canal conditions. It also resists antibiotics. These traits make it tough to eliminate & contribute to treatment letdown. (Dahlen, 2000; Raza, 2018; Lins, 2013).

A key determinant of *E. faecium* virulence is the Enterococcal surface protein (Esp). Esp is a large cell wall–anchored adhesin. It facilitates biofilm development, host tissue adhesion, and immune evasion in several infection models (Sava, 2010). When the esp gene is inactivated or deleted, bacterial adherence and biofilm formation on abiotic substrates decrease. These principles weaken colonization & persistence of the bacteria. This reduction in biofilm mass can make the

bacteria more vulnerable to antimicrobial treatments. Yet, failure to totally eliminate the bacteria can result in retreatment failures, as the residual cells continue to proliferate. This checks Esp crucial role in colonization & persistence (Heikens, 2007). Esp expression is frequently higher in hospital-acquired *E. faecium* strains. This suggests it enhances survival rates below clinical pressures (Sava, 2010).

In the endodontic environment, nutrient levels are low. There is also exposure to antimicrobials and immune activity. Adherence, biofilm formation, and survival are essential (Love, 2001). *E. faecium is* a hardy Gram-positive bacterium. It is initiated more often in persistent root canal infections because it forms durable biofilms & resists ordinary treatments (Stuart, 2006). Esp is a main virulence factor. It encourages adhesion, immune evasion, & biofilm growth (Toledo-Arana, 2001). While its role is vibrant in other infections (Shankar, 2001), its influence in endodontics requires more investigation. Studying its structure and function may explain how E. faecium survives in treated canals. This research may help develop new strategies for resistant infections (Kayaoglu, 2006; Sadanandan, 2025).

This study uses insilico analysis to examine *E. faecium* Esp. Previous research has predominantly relied on wet-lab methods to study Esp. These studies emphasize its biochemical interactions & functional assays. Though our study extricates itself via exploiting computational techniques. We make available an exhaustive structural & physicochemical analysis of Esp. The areas are to: (i) map its domain architect, with signal peptides, repeat sequences, anchoring motifs, & Ig-like domains; (ii) evaluate its physicochemical properties, such as amino acid (aa) content, hydropathy, & molecular stability; & (iii) find conserved motifs connected to adhesion, extracellular matrix interaction, & stress endurance (Spiegelman, 2022; Jarva, 2020; Eaton, 2002). We use the available insilico tool InterProScan & SMART for domain annotation. InterProScan offers inclusive information on protein domain structures. This improves consideration of Esp functional roles. AlphaFold is used for structure prediction. This affords insights that are critical for designing peptide inhibitors. These inhibitors can interrupt biofilm formation & expand treatment effectiveness. Ramachandran analysis is used for validation. This ensures predicted structures are reliable and applicable in real-world clinical scenarios.

This study connects molecular sightings with practical endodontic treatment. Dental professionals may deliberate combining Esp-specific inhibitors or competing peptides into canal irrigants or intracanal bandages. These therapeutic agents could interrupt biofilm structures & help achieve more effective clearance. This integrative plan associates Esp-related molecular research with clinical submissions & highlights Esp as an auspicious therapeutic target for refining consequences in persistent *E. faecium* root canal infections.

2. METHODS AND MATERIALS

Domian Organizations

The Esp protein sequence from *Enterococcus faecium* was analyzed for domain architecture using InterProScan and SMART. InterProScan v5 (https://www.ebi.ac.uk/interpro/search/sequence/) participates in multiple databases such as Pfam, PROSITE, and SUPERFAMILY to expect protein domains & functional motifs (Philip Jones., 2014; Letunic I., 2018; Letunic I., 2012). SMART (http://smart.embl.de/) was used to perceive signaling domains, immunoglobulin-like regions, & repeats specific to bacterial surface proteins (Letunic I., 2012). The sequence (FASTA) was given to both platforms, & predicted domains were likened for consistency. AlphaFold-pre., dictated 3D structures, directed structural domain validation. Ramachandran plots were used to assess model quality. Functional domains known to be involved include YSIRK-type signal peptides, Rib domains, LPXTG motifs, and Ig-like repeats (Binbay FA., 2023).

Structure Assessment

The three-dimensional structure of the protein (Esp), predicted using AlphaFold, was confirmed through Ramachandran plot examination via the server (https://swissmodel.expasy.org) (Schwede T., 2003). This tool assesses model eminence by evaluating the φ (phi) & ψ (psi) angles (backbone dihedral). A high % of residues in favored areas points to a decent model value (Andrew M Waterhouse., 2024). The study completes the assessment of the stability & reliability of the predicted structure.

Protein Functional Analysis

The physico-chemical properties of the Esp protein, such as molecular weight, isoelectric point, amino acid composition, aliphatic index, and hydropathy (GRAVY), were analyzed using ProtParam and emboss_pepinfo on the ExPASy server (Gasteiger, 2005). Hydrophobicity, a key factor in protein folding and structural stability, was specifically examined. Analysis of hydrophobic residue distribution enables identification of membrane-spanning regions, aggregation tendencies, and folding patterns (Eisenberg, 1984; Wimley, 1996). Hydrophobicity profiles were generated using the Kyte–Doolittle hydropathy scale implemented in the *emboss_pepinfo* platform (http://emboss.open-bio.org/wiki/Appdocs).

In this approach, each amino acid receives a hydrophobicity value. These values are then plotted along the protein sequence (Kyte, 1982; Huang, 2014). A specified residue window smooths local fluctuations to help distinguish hydrophobic from

hydrophilic stretches. Positive peaks in the plot indicate hydrophobic regions, which may represent membrane-associated or surface-interactive domains. This analysis facilitates the prediction of transmembrane segments and surface-exposed regions that could influence the virulence and functional mechanisms of the Esp protein.

Amino acid and atomic composition

The amino acid & atomic composition of the protein Esp was determined via the ProtParam tool available through the ExPASy (Gasteiger E., 2005). ProtParam quantifies the number & % of each amino acid & computes the total atomic content, with carbon, hydrogen, nitrogen, oxygen, & sulfur. The examination was shown via the full-length Esp sequence (FASTA format), which covers 1975 amino acids. These limitations inform the biochemical properties, stability, & interaction possibilities of the protein (ExPASy server, 2003).

3. RESULTS

Study of the domain architecture of *E. faecium* Esp protein reveals a structure representative of Gram-positive surface adhesins. The Esp protein is induced with an N-terminal signal peptide, which directs secretion through the cell membrane via the Sec pathway. Subsequently, this is a non-repeated, variable N-terminal domain, often involved in primary adherence. Primarily, the N-terminal area covers YSIRK-type signal peptide (15-50 aa), Rib (584-651aa), Long Rib (655-753aa), Cathgene3d (1088-1172 aa). Repeat domains are located at the C-terminal area, such as Rib/alpha/Esp surface antigen repeat (1594-1673 aa), Bacterial Ig domain (1822-1900), & LPXTG anchor motif (1933-1974). The total length of the Esp is (1-1975 amino acids). In the direction of the C-terminal, Esp structures an LPXTG motif that allows covalent anchoring to the bacterial cell wall, vital for surface display. It also comprehends Ig-like domains near repeat areas, probably helping in host binding & molecular interactions. Prominent Rib domains & surface antigen repeats further enhance its host interaction, possibly, while the YSIRK signal domain/LPXTG anchor domain surface protein (Figure 1A). The SMART diagram above represents a N-terminal signal domain, an internal repeat, detected by the PROSPERO program, and the C-terminal region contains a repeat Rib domain, including an anchor protein (Figure 1B).

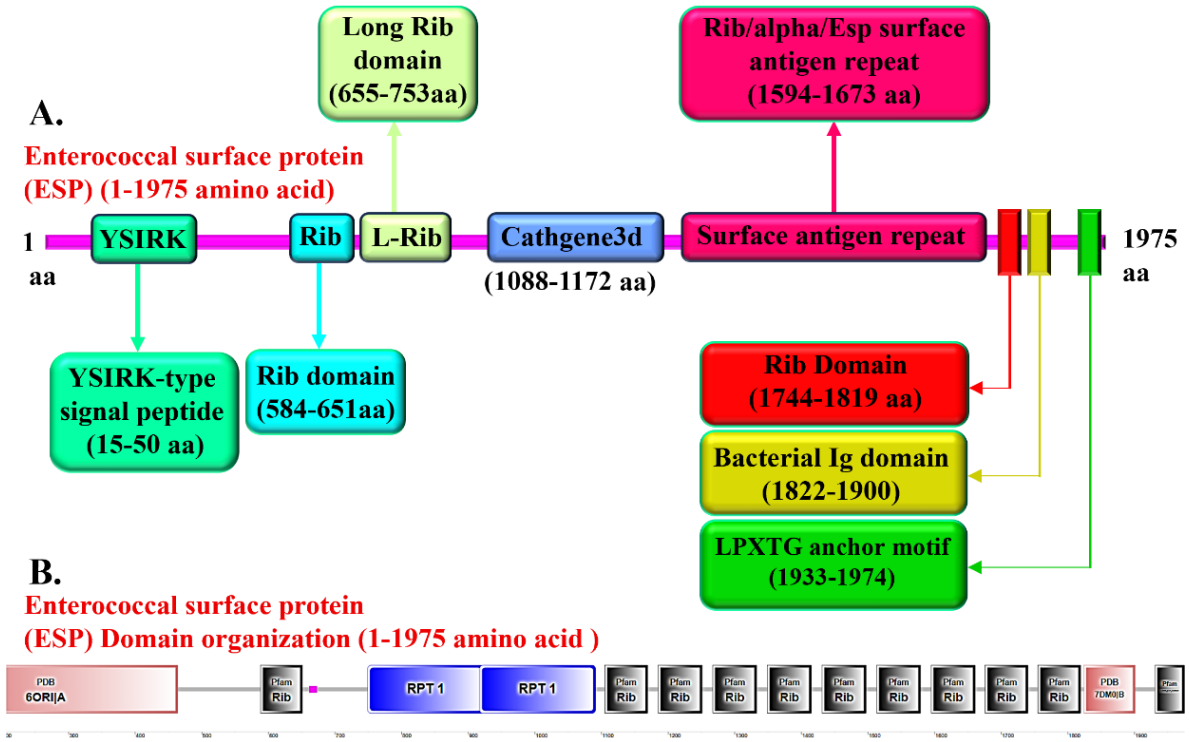

Fig. 1. (A) Schematic representation of the domain organization of Enterococcal surface protein (Esp). (B) Predicted domains, repeats, and motifs of Esp with high confidence.

Structural Validation and Quality Assessment of Esp Protein

Examination of the Ramachandran plot showed that most amino acids in the Esp structure occupy stable, energetically favorable positions. Of all residues, 88.55% were in the most favored regions, with only a few outliers. Studies of glycine, proline, and pre-proline residues also demonstrated proper alignment, supporting the model's structural integrity. These results confirm that the AlphaFold-predicted Esp protein model is reliable and stable (Figure 2A).

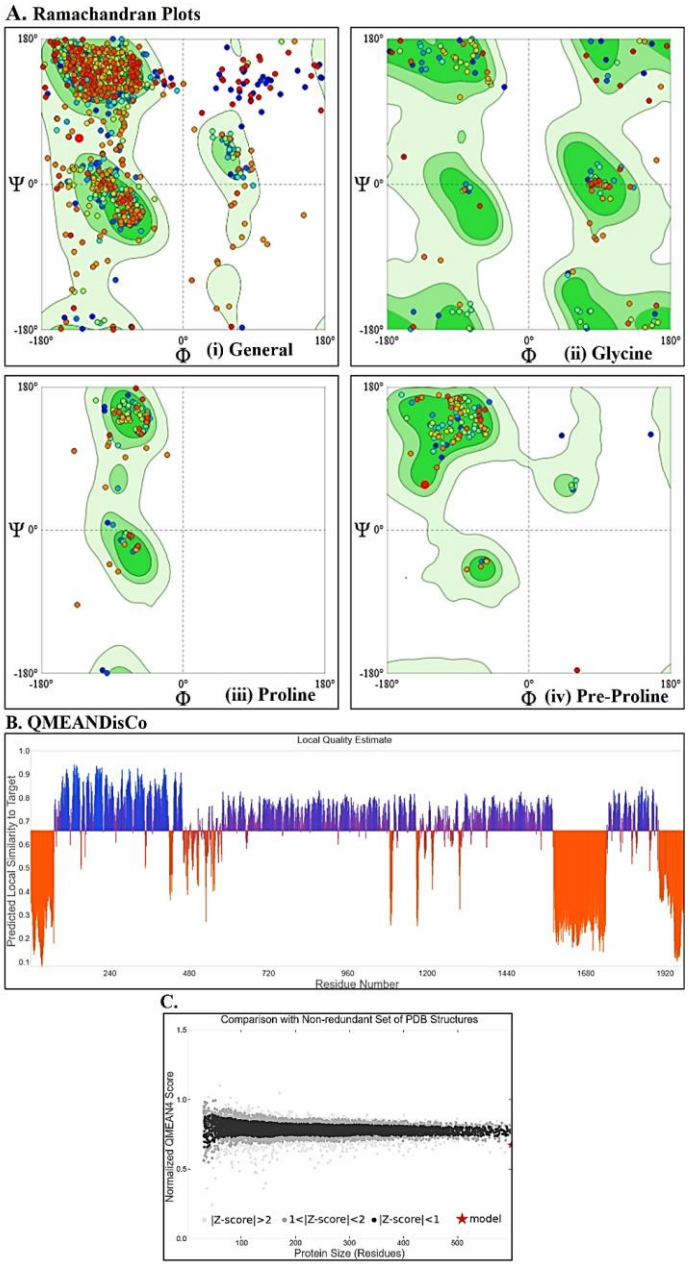

Fig. 2. (A) Ramachandran Plots (i) General (ii) Glycine (iii) Proline (iv) Pre-proline (B) Coordinates with QMEAN local scores in the B-factor column (C) Comparison with a non-redundant set of PDB structures.

The QMEANDisCo profile showed that most regions of the model had resemblance scores above 0.7, indicating high prediction confidence. Lower scores were mostly found at the protein termini and in flexible, dynamic regions. Overall, the model maintained a consistent and robust structural framework (Figure 2B).

Comparison with homologous structures in the Protein Data Bank showed that the Esp model QMEAN score, marked by a red star, was within the acceptable range (|Z-score| < 1). This indicates strong agreement between the predicted and experimental structures. Together, these validation results confirm that the Esp model is robust and suitable for future studies of its function and interactions (Figure 2C).

Residue Property Distribution of Esp Protein

The analysis of the Esp protein from *Enterococcus faecium*, using EMBOSS-pepinfo, revealed a diverse and significant array of building blocks across its extensive amino acid sequence (1975 aa). Slight amino acids, such as glycine, alanine, & serine, were spread via the protein (Esp), showing pliable regions that might allow the protein to form loops or create shapes required to interact with host molecules.

Hydrophobic aliphatic amino acids, such as leucine, isoleucine, & valine, were found via the protein (Esp). They are obliged to keep the central portion of the protein stable and support its overall shape. Aromatic amino acids, including phenylalanine, tyrosine, & tryptophan, were short common but spread in a manner that may aid in protein binding or recognizing other molecules (Figure 3).

The Esp protein had considerable nonpolar residues, indicating it has large non-water-loving areas that enable it adhere to cell surfaces or bind with other proteins. There were rarer polar amino acids, but more positively charged ones, especially lysine and arginine. This positive charge might help the protein stick to negative surfaces, like those on host cells or teeth. A balanced mix of negatively charged amino acids helps keep the electric stability in different parts of the protein. These features show how the Esp protein can adapt in different situations and support its role in forming bacterial layers, avoiding the immune system, and staying in infections for a long time (Figure 3).

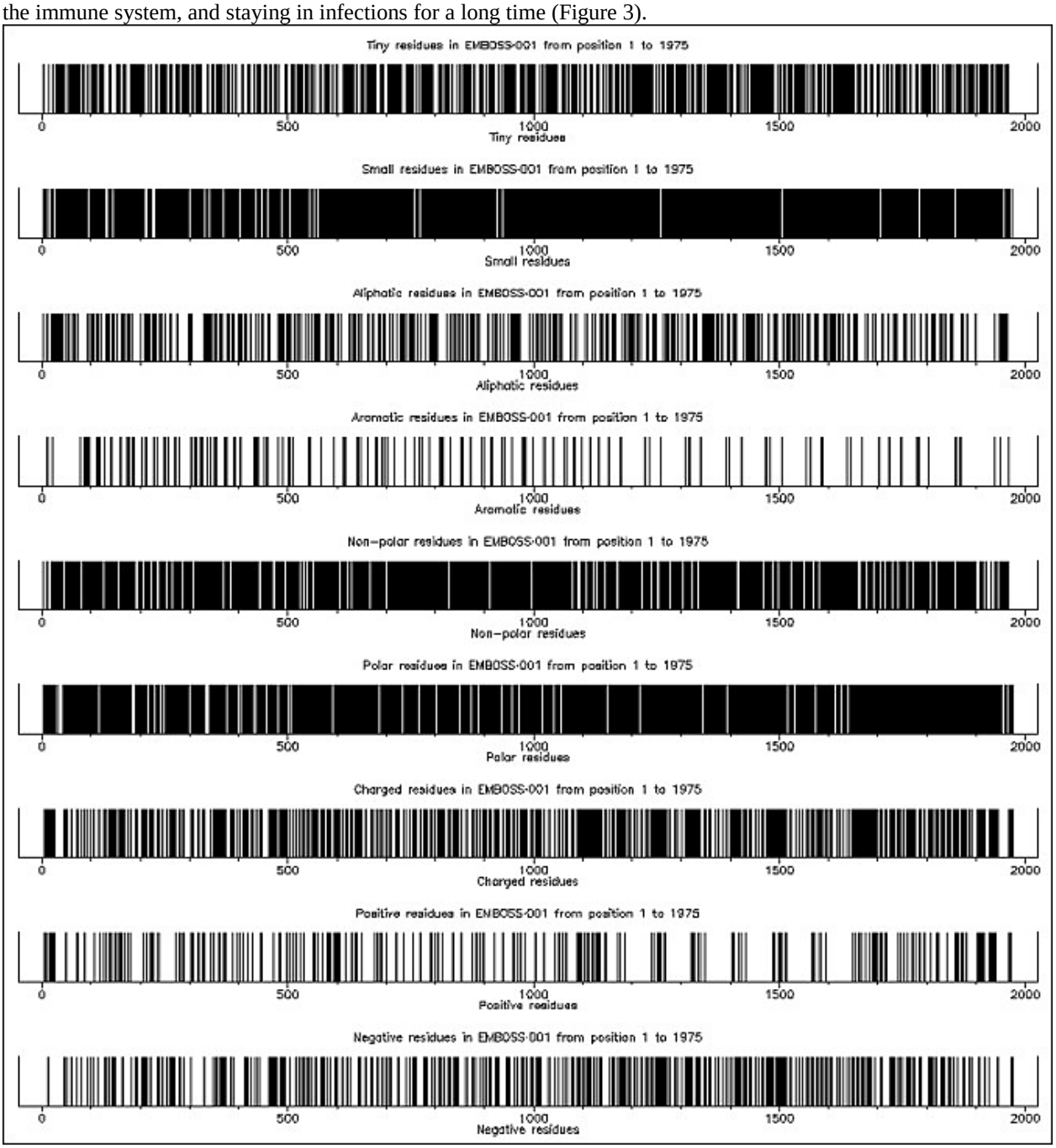

Fig. 3. The distribution of amino acid residue types in the EMBOSS-001 protein sequence from positions 1-975. The frequency of amino acid groups (miniature, small, aliphatic, aromatic, non-polar, polar, charged, positive, and negative) throughout the sequence is depicted in each chart.

The hydropathy analysis of the EMBOSS-001 protein sequence

The Kyte-Doolittle hydropathy analysis was used to examine the distribution of hydrophobic and hydrophilic regions

within the Esp protein (Sequence). The consequent profile announced alternating hydrophobic and hydrophilic segments, indicating that Esp contains both buried internal residues and surface-exposed regions. This structural organization suggests Esp may interact with cell membranes or function on bacterial surfaces.

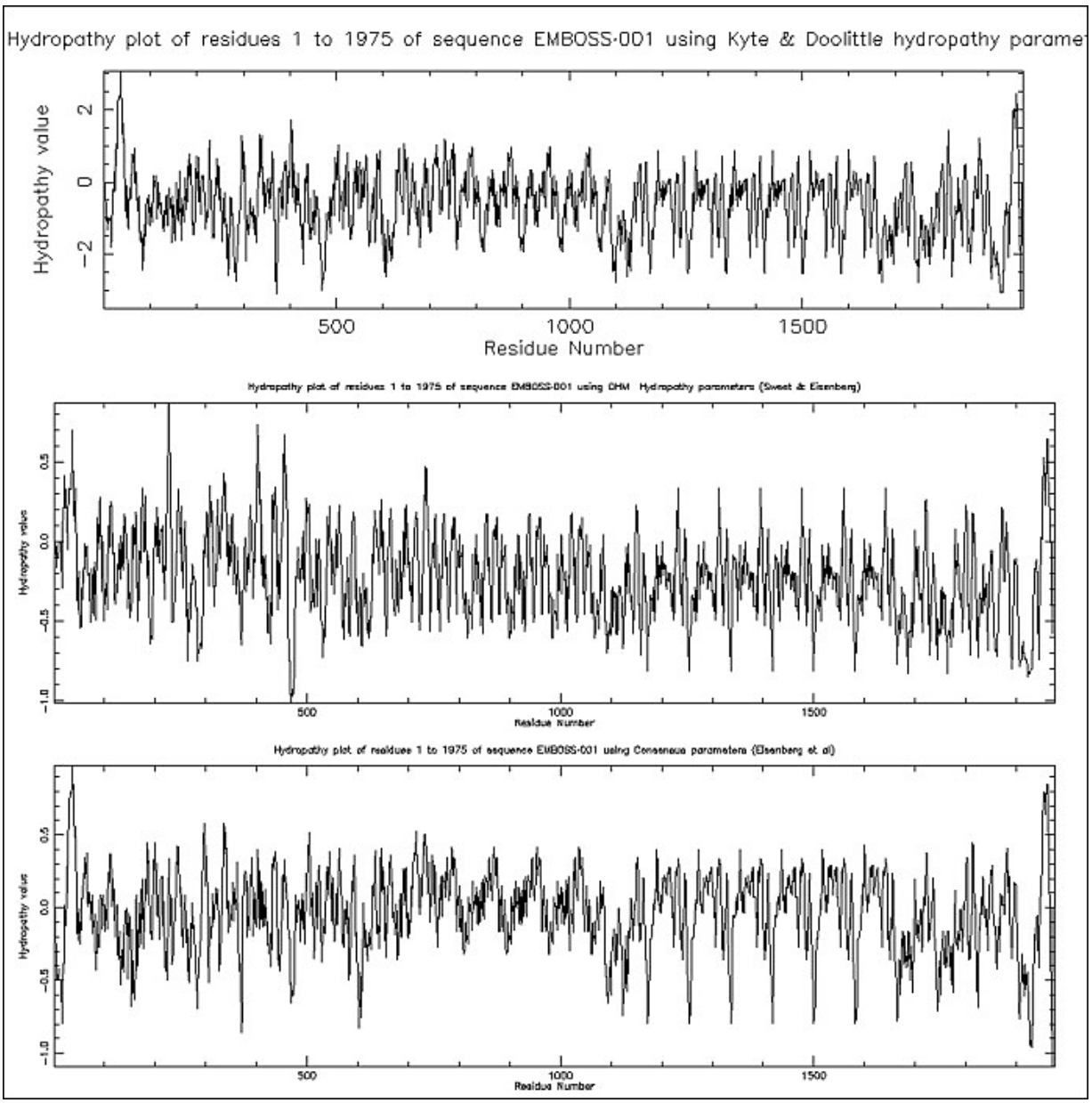

Fig. 4. Hydropathy plots of the EMBOSS-001 protein sequence were generated using Kyte & Doolittle. The plots show alternating hydrophobic and hydrophilic regions. This pattern indicates the presence of both surface-exposed and buried domains within the protein.

Supporting analyses using the Hphob scale and Consensus hydropathy model showed similar fluctuation patterns, backing up the main finding that Esp is mostly hydrophilic and suited for work in water-based environments. Consistently negative hydropathy scores across all three methods show that the EMBOSS-001 model has several water-loving regions with short water-repellent areas. These regions may explain Esp's role in sticking to surfaces, interacting with hosts, and forming biofilms in wet biological conditions (Figure 4).

Compositional and Structural Characterization of the EMBOSS-001 Protein Sequence

The compositional & structural description of the EMBOSS-001 protein uncovered several distinctive molecular features. The study used the Phobius prediction tool (Figure 5A) a software tool that predicts the orientation of protein segments comparable to the cellular membrane —and showed alternating cytoplasmic (inside the cell) and non-cytoplasmic (outside the cell) segments. There was no evidence of prominent transmembrane helices, which are protein regions that generally span the cell membrane. This means the protein is primarily soluble & not embedded within the cell membrane.

The charge issuance analysis (Figure 5B) demonstrated an overall acidic profile, as the number of -ve charged residues (approximately 347) transcended the +ve charged residues (around 202). This inequality supports the hypothesis that EMBOSS-001 functions virtually in polar or aqueous environments, where electrostatic interactions may play a regulatory role.

The amino acid composition profile (Figure 5C) led a dominance of threonine (12.3%), valine (10.4%), & aspartic acid (9.5%), which contemplates a balance between polar & aliphatic residues that may enhance structural flexibility. In expansion, the absence of cysteine, pyrrolysine, & selenocysteine suggests a lack of disulfide crosslinking or specialized redox functions.

In terms of atomic composition (Figure 5D), hydrogen was the most abundant element, with 14,695 atoms, followed by carbon (9,309), oxygen (3,234), & nitrogen (2,417); sulfur atoms were scarce, with only 8 noticed. This elemental distribution reveals a hydrophilic & soluble nature, which is consistent with the marked acidic charge & amino acid profile. Taken together, these findings describe EMBOSS-001 as a non-membranous, water-soluble, & structurally stable protein optimized for activity in aqueous biological systems.

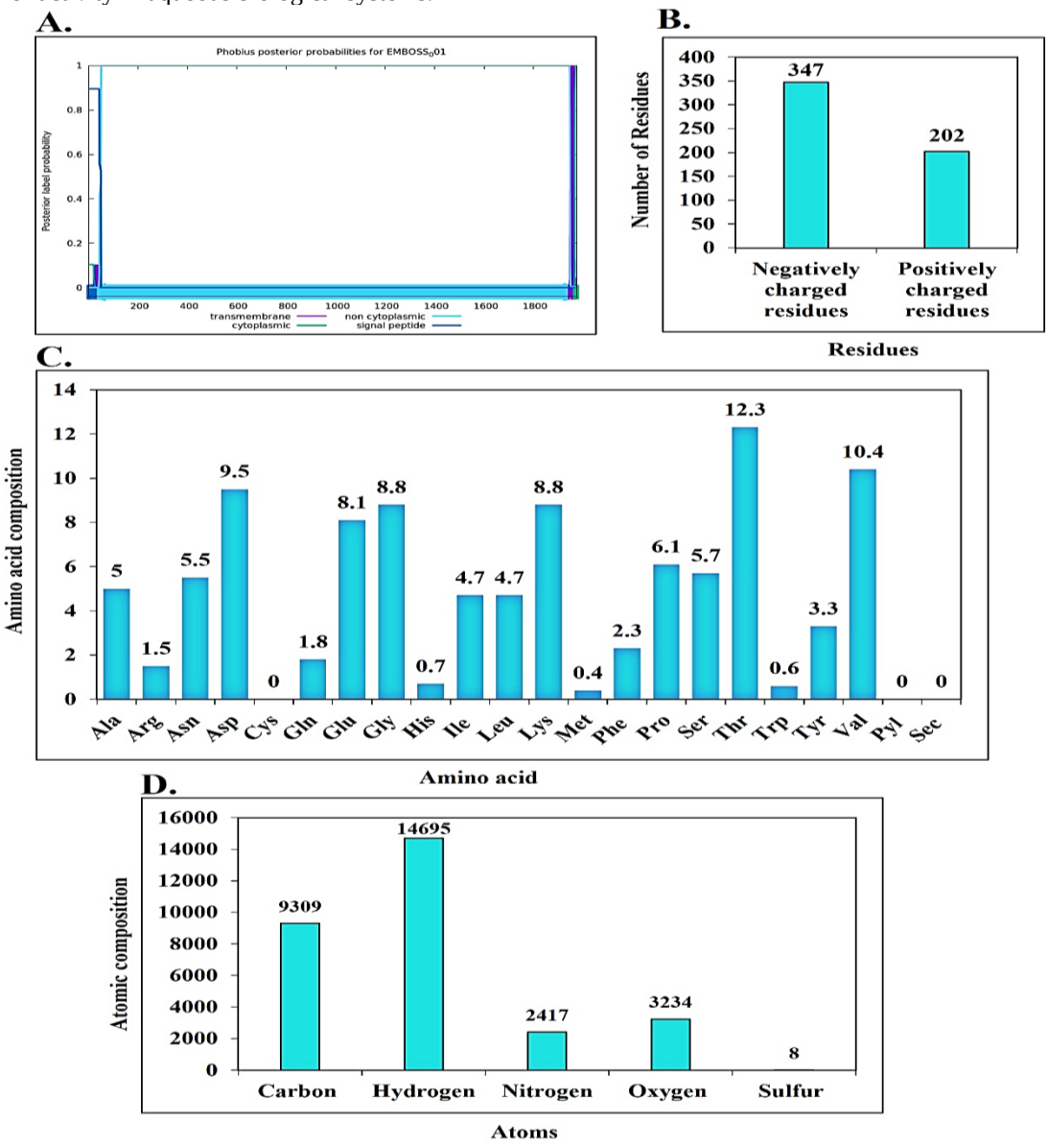

Fig. 5. Structural and compositional analysis of the EMBOSS-001 protein. (A) Phobius predicts transmembrane and signal peptide regions. (B) Distribution of charged residues. (C) Amino acid composition. (D) Atomic composition of the protein.

4. DISCUSSION

The Esp protein is essential for E. faecalis virulence. It mediates biofilm formation, surface adhesion, & and evades host immune responses. It's part of the architecture (1-1975 aa) and is modular, typical of Gram-positive bacterial surface proteins. The N-terminal area has a YSIRK-type signal peptide (15-50 aa) to direct export across the cytoplasmic membrane. The Rib & L-Rib parts (584-753 aa) support adhesion & immune evasion. These domains share homology with Group B *Streptococcus* surface antigens (Lindahl et al., 2005).

The main functional region has Cathgene3d & surface antigen repeat parts. These parts likely permit flexibility & repeated antigen shows, which are implicated in immune modulation. The Rib/alpha/Esp surface antigen repetition (1594-1673 aa) is very meaningful. It collects host cell interaction and increases Esp immunogenicity (Shankar et al., 2001). The C-terminal domain possesses a ribosomal (Rib) domain (1744-1819 aa) & a bacterial Ig-like area (1822-1900 aa). These influence adhesion & structural support. The LPXTG anchor domain (1933-1974 aa) advances cell wall attachment via sortase in G+ve bacteria (Navarre & Schneewind, 1999).

Several studies validate the model structure. Ramachandran plots (Figure 2A) reveal that most residues are in favored regions, indicating good stereochemical stability. QMEANDisCo profile (Figure 2B) says high local grade for most residues, implying dedicated geometry. Deviations appear in flexible or loop regions. The QMEAN4 Z-score (Figure 2C) matches high-quality PDB structures and upholds the model's reliability. These studies confirm the predicted Esp structure's accuracy (Lovell SC., 2003).

Hydropathy analysis utilizing the Kyte-Doolittle algorithm shows many hydrophobic regions. This suggests potential transmembrane domains. A prominent hydrophobic prime from residues X to Y hints at at least one likely alpha-helix spanning the membrane. Hydrophobic stretches are flanked by hydrophilic regions. These are likely cytosolic or extracellular, similar to receptors or transporters. Still, experimental validation such as X-ray crystallography is a must to confirm membrane integration & structure. This *in silico* analysis entitles us to hypothesize about Esp's localization and function.

High threonine & charged residues suggest hydrogen bonding & electrostatic interactions. These stabilize tertiary structure & may support binding or catalysis. The absence of cysteine & selenocysteine means there are no disulfide bridges, supporting a cytoplasmic role. Panel (D) indicates atomic makeup: primarily hydrogen (14,695), carbon (9,309), oxygen (3,234), & nitrogen (2,417). This stoichiometry is usual for proteins. Exhaustive, Esp is an acidic, cytoplasmic, soluble protein. It may be examined in intracellular enzymatic or regulatory pathways (Zhang CT., 1992; Chou PY., 1989; Chun-Ting Zhang., 1995).

5. CONCLUSION

In instantaneous, this computational study explains the molecular basis of the Enterococcal Surface Protein (Esp) from *E. faecium.* Esp plays a key role in pathogenic persistence and virulence in endodontic infections. The analysis showed that Esp has a sophisticated domain organization. It contains a YSIRK-type signal peptide, which directs proteins to the cell surface. It also has Rib domains that help bacteria stick to surfaces. Esp includes bacterial immunoglobulin-like motifs, which resemble immune system proteins. Finally, Esp has an LPXTG anchoring sequence that attaches it to the bacterial cell wall. Each part contributes to Esp’s adhesive power, immune evasion, and ability to form biofilms.

AlphaFold modeling predicted the Esp protein's structure. Ramachandran validation checked the angles of its protein backbone. This confirmed the correct 3-D structure & durable shape of Esp. These findings support Esp reliable function in biological systems. The Esp protein contains both hydrophilic & electrically charged amino acids in its sequence. This allows Esp to adapt to different environments. As a result, Esp can interact with host tissues and persist in the root canal.

In conclusion, Esp is a multifunctional virulence factor. It combines adhesion, surface localization, & immune modulation. Its structure and chemistry make it a possible target for new treatments. These properties could help design inhibitors or vaccines to control E. faecium infections caused by biofilms. This study offers a theoretical basis for future experiments and medicine development. It supports new anti-virulence approaches against antibiotic-resistant enterococcal pathogens.

ACKNOWLEDGEMENT
The authors gratefully acknowledge the funding of the Deanship of Graduate Studies and Scientific Research, Jazan University, Saudi Arabia, through Project Number: GSSRD-24.

CONFLICT OF INTEREST
The author state that they have nothing to reveal.

REFERENCES

[1] Dahlen, G., Samuelsson, W., Molander, A., & Reit, C. (2000). Identification and antimicrobial susceptibility of enterococci isolated from the root canal. *Oral Microbiology and Immunology*, 15(5), 309-312.

[2] Raza, T., Ullah, S. R., Mehmood, K., & Andleeb, S. (2018). Vancomycin resistant Enterococci: A brief review. *Journal of the Pakistan Medical Association*, 68(5), 768-772.

[3] Lins, R. X., de Oliveira Andrade, A., Hirata Junior, R., Wilson, M. J., Lewis, M. A., Williams, D. W., & Fidel, R. A. (2013). Antimicrobial resistance and virulence traits of *Enterococcus faecalis* from primary endodontic infections. *Journal of Dentistry*, 41(9), 779-786.

[4] Sava, I. G., Heikens, E., & Huebner, J. (2010). Pathogenesis and immunity in enterococcal infections. *Clinical Microbiology and Infection*, 16, 533-540.

[5] Heikens, E., Bonten, M. J., & Willems, R. J. (2007). Enterococcal surface protein Esp is important for biofilm formation of *Enterococcus faecium* E1162. *Journal of Bacteriology*, 189(22), 8233-8240.

[6] Love, R. M. (2001). *Enterococcus faecalis* a mechanism for its role in endodontic failure. *International Endodontic Journal*, 34(5), 399-405.

[7] Stuart, C. H., Schwartz, S. A., Beeson, T. J., & Owatz, C. B. (2006). *Enterococcus faecalis*: its role in root canal treatment failure and current concepts in retreatment. *Journal of Endodontics*, 32(2), 93-98.

[8] Toledo-Arana, A., Valle, J., Solano, C., Arrizubieta, M. J., Cucarella, C., Lamata, M., Amorena, B., Leiva, J., Penadés, J. R., & Lasa, I. (2001). The enterococcal surface protein, Esp, is involved in *Enterococcus faecalis* biofilm formation. *Applied and Environmental Microbiology*, 67(10), 4538-4545.

[9] Kayaoglu, G., & Orstavik, D. (2004). Virulence factors of *Enterococcus faecalis*: relationship to endodontic disease. *Critical Reviews in Oral Biology and Medicine*, 15(5), 308-320.

[10] Shankar, N., Lockatell, C. V., Baghdayan, A. S., Drachenberg, C., Gilmore, M. S., & Johnson, D. E. (2001). Role of *Enterococcus faecalis* surface protein Esp in the pathogenesis of ascending urinary tract infection. *Infection and Immunity*, 69(7), 4366-4372.

[11] Sadanandan, B., & Yogendraiah, K. M. (2025). *Enterococcus faecalis* Biofilm: A clinical and environmental hazard. *Medical Sciences Forum*, 35(1), 5.

[12] Spiegelman, L., Bahn-Suh, A., Montaño, E. T., Zhang, L., Hura, G. L., Patras, K. A., Kumar, A., Tezcan, F. A., Nizet, V., Tsutakawa, S. E., & Ghosh, P. (2022). Strengthening of enterococcal biofilms by Esp. *PLoS Pathogens*, 18(9), e1010829.

[13] Jarva, M. A., Hirt, H., Dunny, G. M., & Berntsson, R. P. (2020). Polymer adhesin domains in Gram-positive cell surface proteins. *Frontiers in Microbiology*, 11, 599899.

[14] Eaton, T. J., & Gasson, M. J. (2002). A variant enterococcal surface protein Esp(fm) in *Enterococcus faecium*: distribution among food, commensal, medical, and environmental isolates. *FEMS Microbiology Letters*, 216(2), 269-275.

[15] Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), 1236-1240.

[16] Letunic, I., & Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*, 46(D1), D493-D496.

[17] Letunic, I., Doerks, T., & Bork, P. (2012). SMART 7: Recent updates to the protein domain annotation resource. *Nucleic Acids Research*, 40(Database issue), D302-D305.

[18] Binbay, F. A., Rathod, D. C., George, A. A. P., & Imhof, D. (2023). Quality assessment of selected protein structures derived from homology modeling and AlphaFold. *Pharmaceuticals (Basel)*, 16(12), 1662.

[19] Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research*, 31(13), 3381-3385.

[20] Waterhouse, A. M., Studer, G., Robin, X., Bienert, S., Tauriello, G., & Schwede, T. (2024). The structure assessment web server: for proteins, complexes, and more. *Nucleic Acids Research*, 52(W1), W318-W323.

[21] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). In J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (pp. 571–607). Humana Press.

[22] Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1), 105-132.

[23] Huang, F., Oldfield, C. J., Xue, B., Hsu, W. L., Meng, J., Liu, X., Shen, L., Romero, P., Uversky, V. N., & Dunker, A. (2014). Improving protein order-disorder classification using charge-hydropathy plots. *BMC*

Bioinformatics, 15(Suppl 17), S4.

[24] Eisenberg, D., Weiss, R. M., & Terwilliger, T. C. (1984). The hydrophobic moment detects periodicity in protein hydrophobicity. *Proceedings of the National Academy of Sciences*, 81(1), 140-144.

[25] Wimley, W. C., & White, S. H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology*, 3(10), 842-848.

[26] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31, 3784-3788.

[27] Lindahl, G., Stålhammar-Carlemalm, M., & Areschoug, T. (2005). Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens. *Clinical Microbiology Reviews*, 18(1), 102-127.

[28] Shankar, N., Lockatell, C. V., Baghdayan, A. S., Drachenberg, C., Gilmore, M. S., & Johnson, D. E. (2001). Role of *Enterococcus faecalis* surface protein Esp in the pathogenesis of ascending urinary tract infection. *Infection and Immunity*, 69(7), 4366-4372.

[29] Navarre, W. W., & Schneewind, O. (1999). Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Reviews*, 63(1), 174-229.

[30] Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by Cα geometry: ϕ, ψ, and Cβ deviation. *Proteins*, 50(3), 437-450.

[31] Zhang, C. T., & Chou, K. C. (1992). An optimization approach to predicting protein structural class from amino acid composition. *Protein Science*, 1(3), 401–408.

[32] Chou, P. Y. (1989). Prediction of protein structural classes from amino acid compositions. In G. D. Fasman (Ed.), *Prediction of Protein Structure and the Principles of Protein Conformation* (pp. 549–586). Springer, Boston, MA. https://doi.org/10.1007/978-1-4613-1571-1_12

[33] Zhang, C.-T., Chou, K.-C., & Maggiora, G. M. (1995). Predicting protein structural classes from amino acid composition: Application of fuzzy clustering. *Protein Engineering*, 8(5), 425-435.