Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 6s, https://carcinogenesis.com/

Precision Protein Engineering for Endodontic Applications: A Computational Approach

Nezar Boreak[1*], Aisha Abdu Taher Majrabi[1], Abdulaziz Ali Tohari[1], Amnah Ageeli[1], Razan Abdullah Mahdi[1], Renad Adel M Alsam[1], Afnan Hassan Sayed[1], and Asmaa Almohammad[2]

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

[2]Al Neelain University, 52nd St, Khartoum, Sudan.

*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

Protein-Sol and PSIPRED are crucial resources for developing endodontic research and treatment plans. Effective antimicrobial peptides and immune modulators can be designed more easily thanks to PSIPRED's assistance in predicting protein secondary structures related to immune responses or bacterial resistance within the root canal. Its predictions derived from machine learning aid in maximizing protein function and stability in the intricate oral environment. However, Protein-Sol offers essential information about protein solubility, guaranteeing the effective administration of biologics for endodontic treatment, infection management, and tissue regeneration. Protein-Sol helps customize therapeutic proteins to stay accessible and effective within the pulp tissue or root canal by comparing exposed and buried residues. Furthermore, the design of targeted therapeutics is further guided by the visualization of energy and charge heat maps, accessible surface areas, and ionizable groups. These work together to improve the creation of biocompatible and efficient biomaterials, which improves root canal therapy results, infection prevention, and endodontic tissue reconstruction.

Keywords: *PSIPRED, Protein-Sol, endodontics, machine learning, therapeutic, Surface Patches, Heat map..*

How to Cite: Nezar Boreak, Aisha Abdu Taher Majrabi, Abdulaziz Ali Tohari, Amnah Ageeli, Razan Abdullah Mahdi, Renad Adel M Alsam, Afnan Hassan Sayed, and Asmaa Almohammad., (2025) Precision Protein Engineering for Endodontic Applications: A Computational Approach, *Journal of Carcinogenesis*, *Vol.24, No.6s*, 515-522.

1. INTRODUCTION

Endodontics the area of dentistry that focuses on diagnosing and treating disorders and injuries affecting the dental pulp and surrounding tissues is essential to preserve the integrity and function of teeth. Recent developments in endodontics have moved toward more specialized, biomaterial-based therapies, in which therapeutic chemicals, proteins, and biologics are designed to improve tissue regeneration, infection prevention, and healing. Using computational techniques to develop more potent treatments has resulted in the need for a deeper comprehension of protein function, stability, and solubility. The two most valuable technologies for forecasting essential characteristics of proteins used in endodontic treatment plans are PSIPRED and Protein-Sol. Using machine learning methods and evolutionary data analysis, PSIPRED is a popular tool for protein secondary structure prediction (Jones DT., 1999). Because of its capacity to forecast alpha-helices, beta-sheets, and coils, it is crucial for creating proteins with optimal structures for immunological regulation and drug targeting in the intricate dental environment. However, in complex environments such as pulp tissue or the root canal, Protein-Sol calculates the solubility of proteins, essential in guaranteeing the biological activity and bioavailability of medicinal medicines (Hebditch M et al., 2017). Solubility is crucial when creating biologics like growth factors, regeneration proteins, or antimicrobial peptides for endodontic therapies. In a particular environment, the ratio of exposure to buried residues, the presence of hydrophilic or hydrophobic regions, & the total surface charge affects how soluble a protein. For instance, hydrophilic proteins tend to be more soluble in water, making them more appropriate for local or systemic distribution in dental procedures (Habibi N et al., 2014).

On the other hand, hydrophobic proteins frequently prone to aggregation can pose serious problems for therapeutic application. For root canal cleaning, tooth regeneration, and tissue healing, Protein-Sol assists researchers in optimizing protein formulations by forecasting solubility and structural stability (Santos R et al., 2017). Additionally, PSIPRED and Protein-Sol display important protein characteristics such as charge distributions, ionizable groups, and accessible surface areas (Agapito G et al., 2013). Understanding how a protein will interact with bacterial pathogens, host tissues, and other therapeutic agents is crucial for drug design (Berman HM et al., 2000; Lazar GA et al., 2003). Proteins with optimal binding characteristics are designed for enhanced therapeutic potential based on heat maps of charge and energy distributions produced by these methods (Wanat K et al., 2020; Hebditch M et al., 2019).

In this work, we investigate how PSIPRED and Protein-Sol can be used in tandem to improve infection control methods, promote tissue regeneration, and improve biologics transport and design, all of which can be used to advance endodontic therapies. By using these resources, dental researchers and endodontists can learn more about the behavior of proteins, which can help create new, focused treatments for various oral health conditions and dental diseases.

2. METHODS AND MATERIALS

Sequence Plot and secondary structure

We obtained the protein's amino acid sequence (227 aa) in FASTA format from UniProt (UniProt id: B1PCJ4) on the enterococci surface. To forecast the protein structure and sequence plot, we have entered our sequence into the PSIPRED Workbench, a bioinformatics-based tool (http://bioinf.cs.ucl.ac.uk/psipred/&uuid=6ccf3644-0390-11f0-a57b 00163e100466) database. One of the most popular and respected programs for predicting the secondary structure of proteins is PSIPRED. Based on machine learning methods, specifically neural networks, it has demonstrated a high degree of accuracy in predicting proteins' secondary structures. To help comprehend the structure and function of proteins, PSIPRED is frequently used to indicate the local secondary structure of a protein based on its amino acid sequence.

Predicted scaled solubility

Protein-Sol is a technique that uses the amino acid sequence of proteins to predict their solubility. For applications like protein manufacturing, purification, and drug development, solubility is crucial for comprehending how proteins behave in diverse settings (Kelley LA et al., 2009; Wu Z et al., 2024). Scaled solubility in this context refers to a predicted score that indicates a protein's relative solubility to other proteins; it is frequently scaled to a specific range or scale (https://protein-sol.manchester.ac.uk/results/solubility/run-4bbd72d8fb8c636b7ba2/results.html).

Methods for Visualization of Surface Patches

Using various computational methods and tools that mimic the three-dimensional (3D) structure of proteins and emphasize particular surface characteristics, like charge distribution, hydrophobicity, and the exposure of functional groups, surface patches on proteins can be seen. The following outlines the standard techniques for observing protein surface patches (https://protein-sol.manchester.ac.uk/results/pka/run-329cde094e5aee5f5c17/results.html).

Energy & charge Heat Map

We have presented a protein solubility prediction based on a sequence calibrated with experimental solubilities (https://protein-sol.manchester.ac.uk/). An Energy Heat Map shows how energy values are distributed throughout a molecule's surface, like a protein. By assisting in identifying high- or low-energy areas, which are frequently essential for molecular stability, binding interactions, and conformational changes, it offers insight into the energetic landscape of the molecule (Wu Z et al., 2024; Wade AD et al., 2019). A Charge Heat Map helps visualize areas of positive and negative charge distribution by showing the electrostatic potential of a molecule's surface. Knowing how a protein interacts with other molecules, such as substrates, ions, or other proteins, is made easier with the help of this map (Wang C et al., 2023).

Parade of the protein-solution

The demonstration begins with the user submitting protein data in the form of a primary amino acid sequence. This is the fundamental input for the application. Optionally, users can upload a protein structure file (PDB format) if they want a more detailed analysis of the protein's structural features (McGuffin LJ et al., 2000). The web application for the sequence-based algorithms are available on the protein-sol webserver, at https://protein-sol.manchester.ac.uk/abpred, with models and virtualisation software available at https://protein-sol.manchester.ac.uk/software.[14]

3. RESULTS

Structure & solubility prediction

It uses the amino acid content of a protein sequence to predict its secondary structure. PSIPRED combines neural networks with position-specific scoring matrices (PSSMs) to achieve these predictions. The sequence plot result shows that the coil

is gray, the helix is pink (AKES), and the helix is emphasized in yellow (Fig.1A). Generally speaking, these structures are divided into three primary groups which are The alpha-helix, Usually maintained by hydrogen bonds between the peptide chain's backbone; a right-handed spiral Beta-Sheet, A structure resembling a sheet that is created by hydrogen bonds between the protein's backbone; typically between strands that are parallel or antiparallel, And Coil or Undefined Coil: segments of the protein that can take on different conformations and do not consistently form repeating patterns (Fig.1B). Protein hydrophobicity, charged residue distribution, structural flexibility, and general stability are some variables that affect a protein's solubility in solution. According to Fig. 1C, the scaled solubility value (0.913) is the expected solubility, and the Pi is 4.340. According to the experimental solubility dataset's population average (PopAvrSol), any scaled solubility value above 0.45 is expected to be more soluble than the average soluble *E. coli* protein. In contrast, any protein with a lower-scaled solubility value is expected to be less soluble. The technique for protein-sol sequences computed 35 sequence characteristics. This comprises the following characteristics, which are computed over a sliding 21 amino acid sequence, as well as the conventional 20 amino acid composition and sequence length (len). KmR = K minus R, DmE = DminusE, KpR = K plus R, DpE = D+E, PmN = K+R-D-E, PpN = K+R+D+E, and aro = F + W + Y are the seven amino acid composite scores. Next, we compute seven more sequence features: pI, ent = sequence entropy, abs = absolute charge at pH 7, mem = Kyte-Doolittle hydropathy, dis = disorder propensity, bet = beta strand propensities, and fld = folding propensity (Figure 1D).

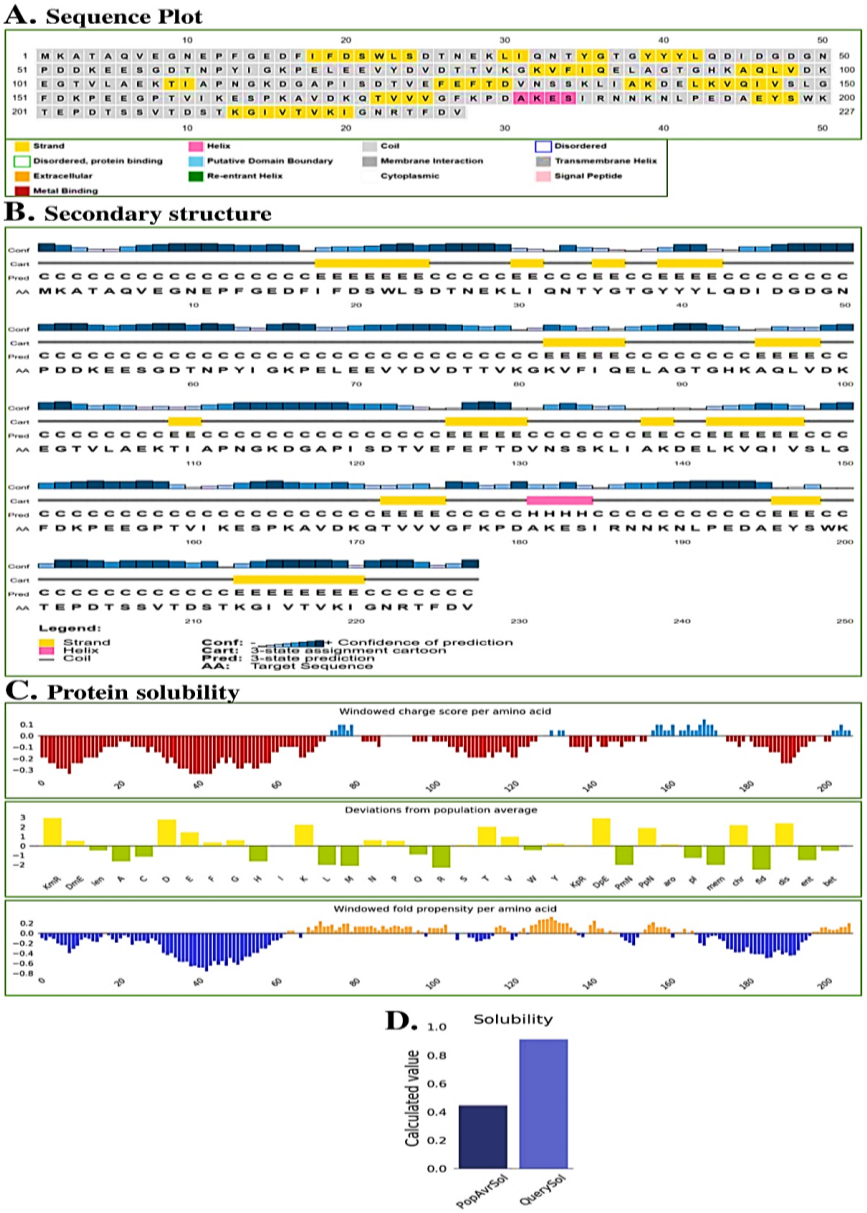

Fig. 1. (A) Amino acid sequence plot (B) Secondary structure prediction (C) Protein solubility (D) The Protein–Sol calculation.

Conception of Surface Patches

Protein surface patches are areas of the protein structure exposed to the environment and are essential for interactions with other molecules, including ions, substrates, and other proteins. Understanding the physicochemical characteristics of a protein's surface can be effectively accomplished by imaging surface patches. Analysis of the protein's charge distribution, polar/nonpolar ratios, and ionizable groups can provide important information on its stability and behavior and its capacity to interact with other molecules, including ions, ligands, and other proteins. The charge distribution on the surface of the Fab protein is diverse. Fig. 2A shows that positively charged areas are represented by blue regions and negatively charged areas by red regions. The Nonpolar to Polar Ratio (NPP) study finds areas of the Fab protein surface where nonpolar (hydrophobic) or polar (hydrophilic) residues predominate. Fig. 2B shows that green areas imply a higher NPP ratio, emphasizing nonpolar zones, which are generally hydrophobic, while purple areas reflect low NPP ratios, suggesting more polar regions.

The Nonpolar Ratio image shows protein surface areas with a high concentration of hydrophobic residues. High nonpolar ratio sections are green, indicating hydrophobic regions usually deep within the protein's core. Regions with low nonpolar ratios are shown in different colors, signifying that they are more hydrophilic and will likely interact with the solvent around them. The red area is visible. Fab antigen interface-I is the blue area with the fab surface, and the green area is highlighted to illustrate the fab chain-to-chain interface (Fig. 2C). Ionizable groups are visualized on the protein surface, indicating that they have been exposed to the solvent. The accessible surface area (ASA) indicates the degree of solvent exposure of these residues; the charge and general behavior of the protein in solution are influenced by the number of exposed groups (Fig. 2D).

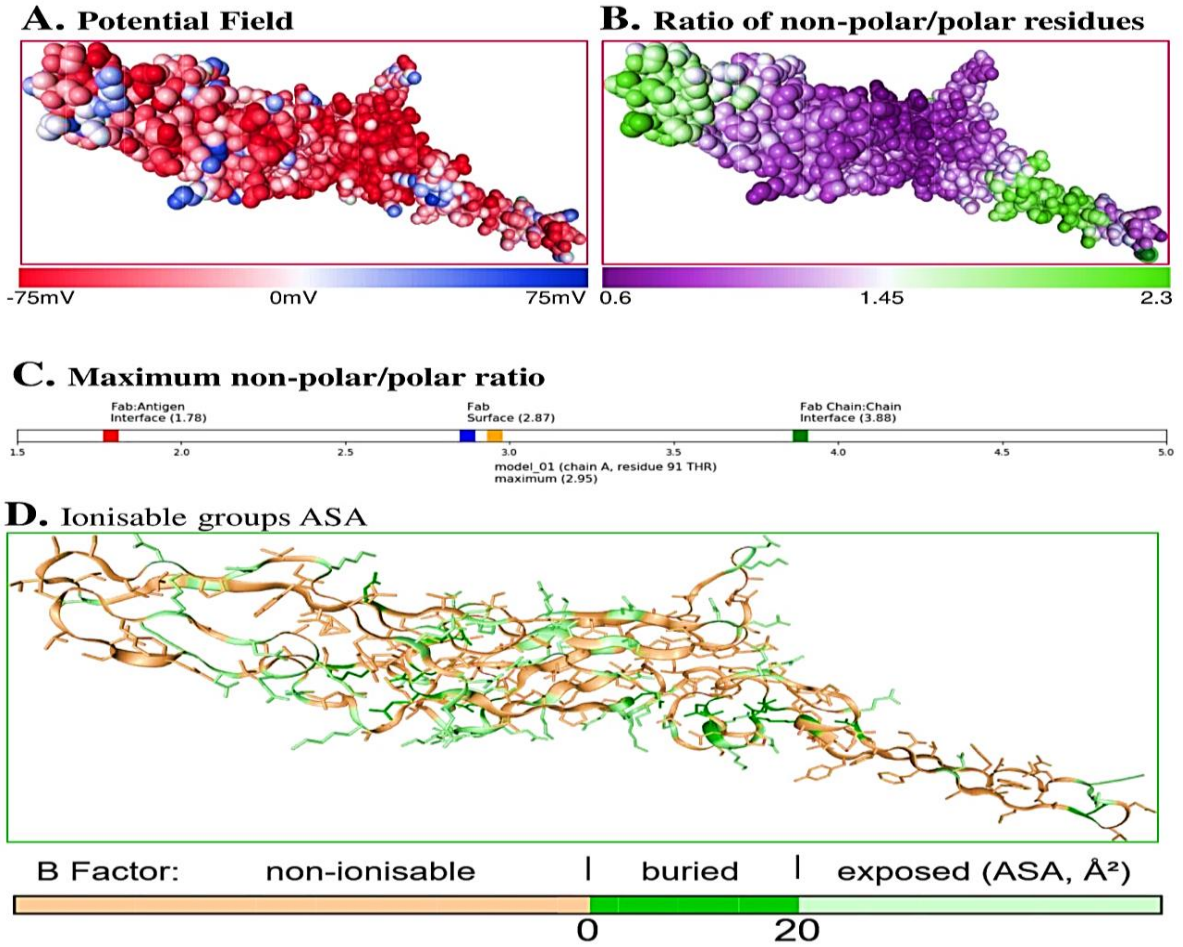

Fig. 2. Surface patches visualized (A) The color scheme of the fab is red for negative charge and blue for positive charge. (B) To indicate areas with different polar and non-polar residue ratios, the Fab structure is color-coded from low NPP ratio (purple) to high NPP ratio (green). (C) To indicate areas with different ratios of polar and non-polar residues, the Fab structure is color-coded from low NPP ratio (purple) to high NPP ratio (green). (D) By displaying ionizable groups and their corresponding accessible surface areas (ASA), this panel provides information about how buried or exposed these functional groups are on the protein's surface.

Heat map

Protein-sol offers surface visualization and heatmaps for the pH and ionic strength dependency on folded state protein stability. These are derived from pKa calculations and the Debye-Hückel (DH) approach for interactions between ionizable groups. Fig. 3 shows the expected charge (in units of e per amino acid) and the predicted pH-dependent contribution to stability (in units of J per amino acid) in two different heatmaps. Unfavorable interactions, buried hydrophobic residues exposed to the solvent, or charged residues arranged in a destabilizing configuration are frequently the locations of high-energy patches.

Additionally, these sites might be hotspots for binding or interactions, where minor conformational changes or binding events could reduce the system's energy (Fig. 3A). Important areas that could interact with metal ions or other biomolecules are highlighted in charge heat maps, which show how electrostatic interactions support protein stability (Fig. 3B).

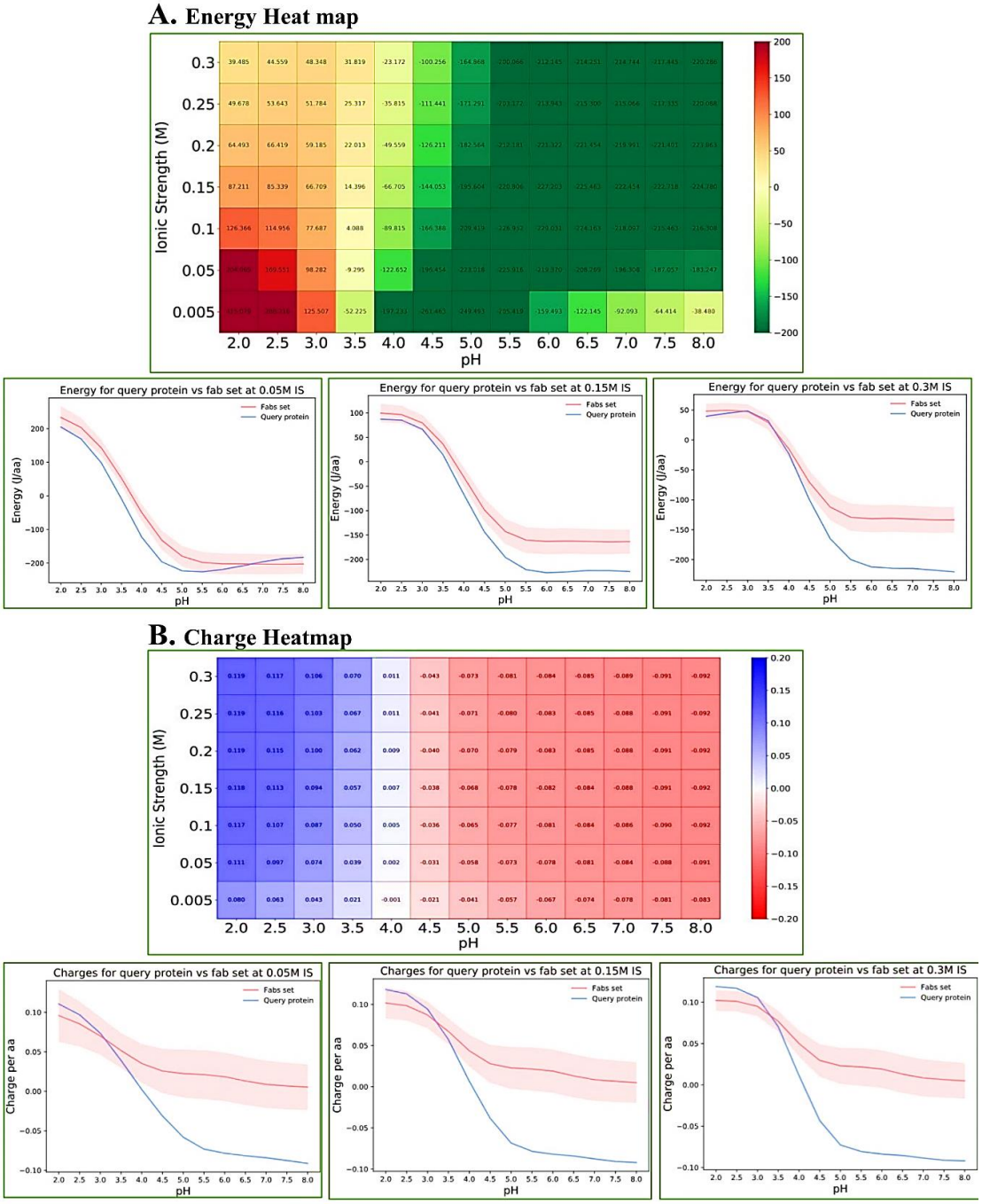

Fig. 3. (A) Energy heatmap: Energy heat maps make available a comprehensive representation of the energy landscape (B) Charge heat map: Charge heat maps emphasis on the electrostatic possible of a protein's surface.

Demonstration of the protein-sol

Protein-sol now offers all twelve machine learning algorithms. The user can enter A candidate Fv sequence into the online application, and it will be processed using the same methods as this study. The y-axis of the scatter plot shows the appropriate protein predictions, while the x-axis shows the initial experimental values or their mathematical adjustments. While the user-submitted protein is highlighted in orange, the mAb137 dataset is green, with three hues signifying the various FDA clearance stages.

Fig. 4A shows the associated heat map, shaded green if the candidate protein's prediction is within the permissible range and red if it exceeds the threshold. A graphical representation illustrates the analysis (Fig. 4B). Each scFV's color in the heatmap is determined by the threshold value. This threshold value is determined using the worst 10% cutoff for the anticipated Jain dataset values. The matching square is colored green if the projected result is below the experiment's threshold value and red otherwise. The forecasts for that category appear in the scatter graph when you hover over the heatmap. A rating of the potential antibody against the Jain dataset is also provided.

The original Jain dataset is ranked from best to worst results to determine the meta score, and the candidate sequence's position within each biophysical platform's ranking is then determined. The ranks for the biophysical platforms are then aggregated and averaged as follows:
Group X: BVP, ELISA ACC STAB, CIC, CSI, PSR, and ELISA
HIC and SMAC in Group Y
The closer the candidate is to the origin (0,0), the better we anticipate their behavior on the platforms; the lower the rating, the better for each group.
The X-axis of the scatter plot above displays the experimental findings for the Jain clinical stage treatments. In contrast, the Y-axis displays our prediction for the same protein based on the existing scFV sequence. When you hover over each data point in the Jain dataset, the actual experimental and algorithm-predicted values are shown.

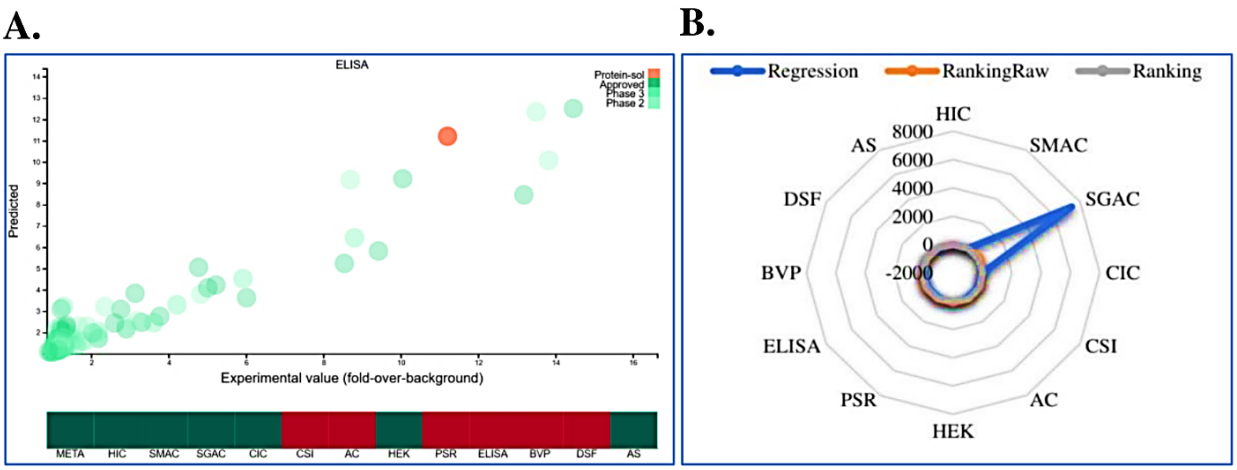

Fig. 4. (A) An all-inclusive overview of how the Protein-Sol web application aids in expecting solubility, visualizing surface properties, & evaluating protein structure (B) Analysis of machine learning algorithms.

4. DISCUSSION

Developing effective treatments for root canal infections, tooth regeneration, and tissue healing in endodontics is a complex challenge. Recent advances in bioinformatics and computational tools have provided novel approaches to protein-based therapeutics, offering the potential to optimize drug design, biomaterials, and targeted therapies. Among the most influential tools in this area are PSIPRED and Protein-Sol, which are pivotal in understanding protein structures, optimizing protein solubility, and guiding the design of more biocompatible, stable, and effective therapeutic agents. By leveraging these tools, endodontists and researchers can significantly improve clinical outcomes of root canal therapy, infection control, and tissue regeneration. PSIPRED can help identify which protein regions will form alpha-helices or beta-sheets, which are essential for the antibacterial properties of AMPs. By predicting these structural motifs, researchers can design AMPs with enhanced efficacy against bacteria, particularly those in the oral cavity, where environmental factors such as pH and ionic strength may affect protein stability (Khavani M et al., 2024; Huan Y et al., 2020; Kanai C et al., 2021; Jiang Z et al., 2008).

Protein-Sol addresses this by analyzing the exposed and buried residues within a protein's structure, helping researchers identify potential aggregation sites or regions prone to instability. This insight allows for modifying protein structures to enhance their solubility without compromising biological activity (Monti A et al., 2023) Ionizable groups, such as carboxyl and amino groups, play a critical role in the pH sensitivity of proteins. The presence and distribution of ionizable groups on the surface of a protein influence its solubility and binding properties under various pH conditions. This is particularly important in applications such as endodontics, where proteins must function effectively in the acidic root canal environment (Chen Z et al., 2023; Pace CN et al., 2000).

Recent studies have shown that heat maps of protein stability generated by pH and ionic strength variation, using methods like Debye-Hückel and pKa calculations, can guide the design of biotherapeutic proteins with optimized stability and solubility (Mwangi J et al., 2023; Buchan DWA et al., 2024). This predictive modeling is instrumental in designing proteins that retain their therapeutic properties, such as those used in infection control or tissue regeneration in endodontic treatments. Protein-Sol predicts the solubility of proteins based on key factors such as the ratio of exposed versus buried hydrophobic residues and the overall balance of polar and nonpolar regions (Hebditch M et al., 2019). Protein-Sol's ability to visualize solubility-related properties provides critical data that can guide experimental design, reducing the need for costly and time-consuming laboratory trials. The tool's predictive power is complemented by its accessibility, as it is freely available online, making it a valuable resource for researchers across various fields, including drug discovery and structural genomics (Buchan DWA et al., 2024; Burley SK et al., 2023; Madeira F et al., 2022).

5. CONCLUSION

PSIPRED and Protein-Sol are crucial tools in endodontics for understanding protein behavior in root canal infections, tooth regeneration, and tissue healing. PSIPRED predicts secondary protein structure, aiding in the design of antimicrobial peptides and immune modulators. Protein-Sol provides insights into protein solubility, enabling the delivery of therapeutic proteins in endodontic treatments. These tools also help visualize accessible surface areas and ionizable groups, guiding the design of targeted therapies. By leveraging PSIPRED and Protein-Sol, researchers can develop biocompatible, effective, and targeted biomaterials and therapeutics, optimizing outcomes in root canal therapy, infection control, and tissue regeneration.

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] Agapito, G., Guzzi, P. H., & Cannataro, M. (2013). Visualization of protein interaction networks: Problems and solutions. BMC Bioinformatics, 14(Suppl 1), S1.
[2] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. Nucleic Acids Research, 28(1), 235–242.
[3] Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chao, H., Chen, L., Craig, P. A., Crichlow, G. V., Dalenberg, K., Duarte, J. M., et al. (2023). RCSB Protein Data Bank (RCSB.org): Delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. Nucleic Acids Research, 51(D1), D488–D508.
[4] Buchan, D. W. A., Moffat, L., Lau, A., Kandathil, S. M., & Jones, D. T. (2024). Deep learning for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research, 52(W1), W287–W293.
[5] Chen, Z., Wang, X., Chen, X., Huang, J., Wang, C., Wang, J., & Wang, Z. (2023). Accelerating therapeutic protein design with computational approaches toward the clinical stage. Computational and Structural Biotechnology Journal, 21, 2909–2926.
[6] Hebditch, M., Carballo-Amador, M. A., Charon, M., & Warwicker, J. (2017). Protein-Sol: Predicting protein solubility from sequence. Bioinformatics, 33(7), 1114–1116.
[7] Habibi, N., Mohd Hashim, S. Z., Norouzi, A., & others. (2014). A review of machine learning methods to predict the solubility of overexpressed recombinant proteins in Escherichia coli. BMC Bioinformatics, 15, 134.
[8] Hebditch, M., & Warwicker, J. (2019). Charge and hydrophobicity are key features in sequence-trained machine learning models for predicting the biophysical properties of clinical-stage antibodies. PeerJ, 7, e8199.
[9] Huan, Y., Kong, Q., Mou, H., & Yi, H. (2020). Antimicrobial peptides: Classification, design, application and research progress in multiple fields. Frontiers in Microbiology, 11, 582779.
[10] Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. Journal of

Molecular Biology, 292(2), 195–202.
[11] Jiang, Z. (2008). Computational analysis of the interaction between ligand-receptor pairs. Current Pharmaceutical Design, 14(6), 588–592.
[12] Lazar, G. A., Marshall, S. A., Plecs, J. J., Mayo, S. L., & Desjarlais, J. R. (2003). Designing proteins for therapeutic applications. Current Opinion in Structural Biology, 13(5), 513–518.
[13] Kelley, L. A., & Sternberg, M. J. E. (2009). Protein structure prediction on the web: A case study using the Phyre server. Nature Protocols, 4, 363–371.
[14] Khavani, M., Mehranfar, A., & Mofrad, M. R. K. (2024). Antimicrobial peptide interactions with bacterial cell membranes. Journal of Biomolecular Structure and Dynamics. Advance online publication.
Kanai, C., Kawasaki, E., Murakami, R., Morita, Y., & Yoshimori, A. (2021). Computational prediction of compound–protein interactions for orphan targets using CGBVS. Molecules, 26(17), 5131.
[15] McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. Bioinformatics, 16(4), 404–405.
[16] Monti, A., Vitagliano, L., Caporale, A., Ruvo, M., & Doti, N. (2023). Targeting protein–protein interfaces with peptides: The contribution of chemical combinatorial peptide library approaches. International Journal of Molecular Sciences, 24(9), 7842.
[17] Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Research, 50(W1), W276–W279.
[18] Mwangi, J., Kamau, P. M., Thuku, R. C., & Lai, R. (2023). Design methods for antimicrobial peptides with improved performance. Zoological Research, 44(6), 1095–1114.
[19] Pace, C. N., Grimsley, G. R., & Scholtz, J. M. (2009). Protein ionizable groups: pK values and their contribution to protein stability and solubility. Journal of Biological Chemistry, 284(20), 13285–13289.
[20] Santos, R., Ursu, O., Gaulton, A., et al. (2017). A comprehensive map of molecular drug targets. Nature Reviews Drug Discovery, 16(1), 19–34.
[21] Wu, Z., Chen, S., Wang, Y., Li, F., Xu, H., Li, M., Zeng, Y., Wu, Z., & Gao, Y. (2024). Current perspectives and trend of computer-aided drug design: A review and bibliometric analysis. International Journal of Surgery, 110, 3848–3878.
[22] Wade, A. D., & Huggins, D. J. (2019). Optimization of protein–ligand electrostatic interactions using an alchemical free-energy method. Journal of Chemical Theory and Computation, 15(11), 6504–6512.
[23] Wang, C., & Zou, Q. (2023). Prediction of protein solubility based on sequence physicochemical patterns and distributed representation information with DeepSoluE. BMC Biology, 21, 12.
[24] Wanat, K. (2020). Biological barriers, and the influence of protein binding on the passage of drugs across them. Molecular Biology Reports, 47, 3221–3231.
[25] Warwicker, J. (2022). The physical basis for pH sensitivity in biomolecular structure and function, with application to the spike protein of SARS-CoV-2. Frontiers in Molecular Biosciences, 9, 834011.