Study of biophysical mechanisms of aptamer-based biosensor for detection of Escherichia coli O157:H7

Muhammad Fakhrullah Mohamad Azmi[1], Wan Mardhiyana Wan Ayub[1], Muhammad Fadzlisyam Redzuan[1], Irfan Danial Ismadi[1], Mohd Ifwat Mohd Ghazali[1], Muhamad Arif Mohamad Jamali[1], Liyana Azmi[2], Nur Zaireena Zainal[2], Nazefah Abdul Hamid[2], Shahino Mah Abdullah[1*]

[1]*Faculty of Science and Technology, Universiti Sains Islam Malaysia, Bandar Baru Nilai, Nilai 71800, Negeri Sembilan, Malaysia*

[2]*Faculty of Medicine and Health Sciences, Universiti Sains Islam Malaysia, Bandar Baru Nilai, Nilai 71800, Negeri Sembilan, Malaysia*

Abstract

The increasing prevalence of pathogenic Escherichia coli (E. coli) in water and food sources poses a significant threat to public health, necessitating the development of rapid and accurate biosensor detection methods such as aptamer-based biosensors due to their high specificity and sensitivity. Aptamers are nucleic acids that can bind with high affinity and specificity to a range of target molecules. In this study, 1,000 shuffled variants of a known aptamer (PDB ID: 2AU4) were generated and evaluated for stability using RNAfold and RNALfoldz based on minimum free energy (MFE). The five most stable sequences were selected and analyzed for their secondary and tertiary structures using RNAComposer. The target protein, Shiga toxin (Stx, PDB ID: 1C48), was modeled with AlphaFold 3 and validated through Ramachandran plot analysis. Molecular docking using the HDOCK server revealed aptamer-protein binding interactions, offering insights into the structural features that influence binding specificity and stability. In conclusion, this research bridges theory for future applications, thereby establishing a theoretical framework to support the future development of aptamer-based biosensors targeting E. coli O157:H7.

Keywords**:** *Escherichia Coli; Aptamer; Protein; Molecular free energy; Molecular docking*

1.0 Introduction

Escherichia coli (E. coli) is a diverse group of bacteria found in the environment, foods, and intestines of humans and animals [1]. Enterohemorrhagic E. coli (EHEC), including the notorious O157:H7 strain can cause severe foodborne disease, leading to bloody diarrhea and potentially life-threatening conditions such as hemolytic uremic syndrome (HUS) [2,3]. Thus, the increasing prevalence of pathogenic E. coli in water and food sources poses a significant threat to public health, necessitating the development of rapid and accurate detection methods. Aptamer-based biosensors have emerged as a promising tool for pathogen detection due to their high specificity, sensitivity, and stability [4]. However, a comprehensive understanding of the structural dynamics underlying aptamer-target interactions remains limited. This study helps bridge that gap by using computational biophysical simulations to find stable aptamer candidates and predict how well they bind to E. coli Shiga toxin, bringing us a step closer to developing more effective next-generation biosensors.

This study will employ docking simulation to elucidate the interaction dynamics between specific aptamers and their target molecules which in this case is protein E. coli [5]. Docking simulations are very popular approaches able to assess the capacity of a given ligand to interact with a target. The goal of ligand-protein docking is to predict how a protein interacts with ligands of known three-dimensional structure [6]. There are three docking methods that can be classified: protein-small molecule docking, protein-nucleic acid docking, and protein-protein docking [7]. To evaluate a good docking process, a docking score will be calculated and analysed. Docking score is an algorithm designed to compute the binding affinity of a protein-ligand complex [8]. By simulating the binding processes at an atomic level,

we will gain insights into the conformational changes and binding affinities of the aptamer that govern the selectivity and sensitivity of these biosensors [9].

Aptamer binding often indu Aptamer binding often induces conformational changes in both the aptamer and the target protein [10]. Aptamers are single-stranded oligonucleotides that fold into defined architectures and bind to targets such as proteins with high affinity and specificity [11]. Aptamer is built with three structures: primary, secondary and tertiary structures. Primary structures are built with a long sequence of nucleotides (A, T/U, C, G). The primary structures are fundamental for the secondary and tertiary structures. Secondary structures, generally a two-dimensional structure can form various secondary structures like hairpins, loops, bulges, and G-quadruplexes, which contribute to their overall 3D shape and binding capabilities. The evaluation stability of these structures is according to their molecular free energy (MFE) value. The lowest MFE value (negative MFE) determines the stability of that aptamer structure [12]. Tertiary structure or commonly known as three-dimensional structure enables aptamers to recognize and bind to their specific targets in molecular docking and molecular dynamic simulation [13].

2.0 Methodology

2.1. Software workflow

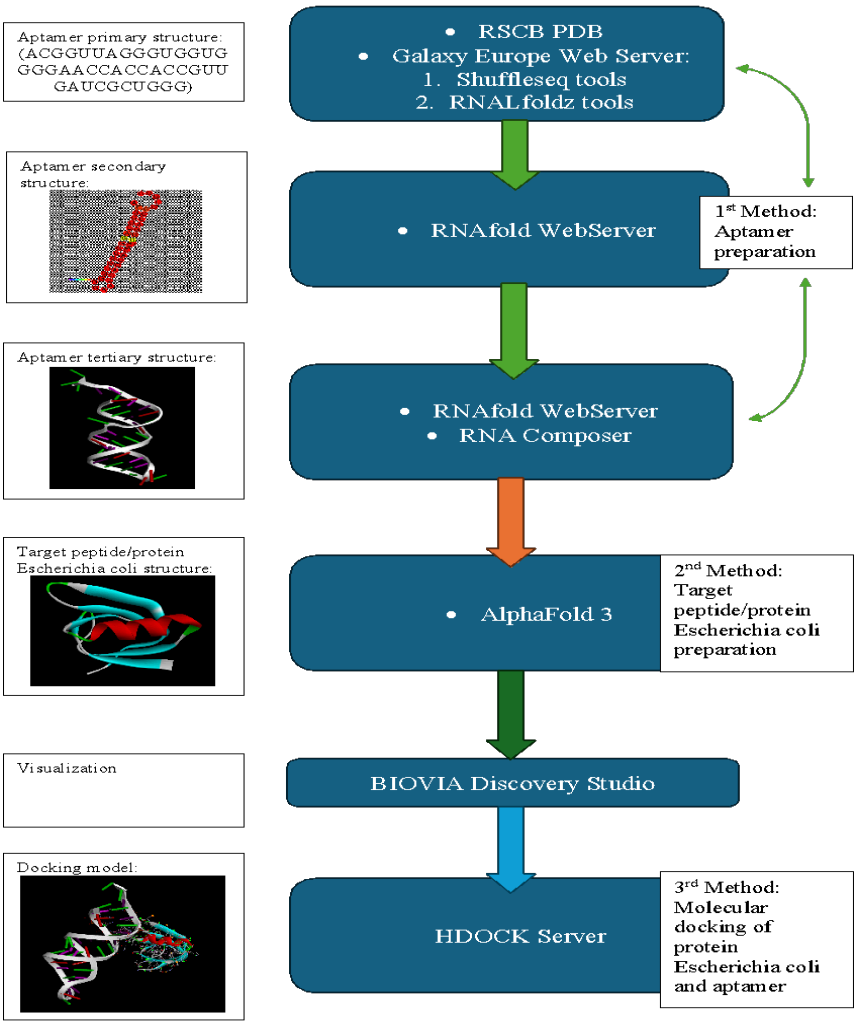

Fig. 1. Software workflow starting from aptamer structure, peptide structure and docking model

2.2. Aptamer preparation

The structure of aptamer was retrieved from Protein Data Bank with PDB code: 2AU4 as shown in Fig. 2.

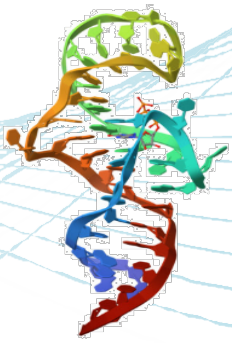

Fig. 2. 2AU4 Aptamer

The structure from PDB was exported in FASTA Sequence format because then the structure sequence needs to be shuffled to produce other clones of 2AU4 aptamer sequences. The shuffle process was using shuffleseq tools from usegalaxy.eu web server with 1000 number of shuffles. As a result, there were 1000 structures sequence of aptamer that can be validated in the next process. The validation of the aptamer's structures sequence was based on their molecular free energy (MFE) produced. This validation uses RNALfoldz tools from Galaxy Europe (https://usegalaxy.eu/) in FASTA sequence format. Then, the sequence was rearranged in MS Excel with their specific MFE value and from 1000 sequences, the best five sequences was chosen has the lowest MFE value as presented in Table 1. Next, for 3D structure analysis, RNA Composer is used, and the input format are in FASTA sequence and in dot-bracket format. This study uses RNAfold WebServer as it provides an interactive graphical output of the MFE structure or aptamer secondary structure as shown in Figure 3. From that, RNAfold WebServer was used to transform the FASTA sequence format into dot-bracket format.

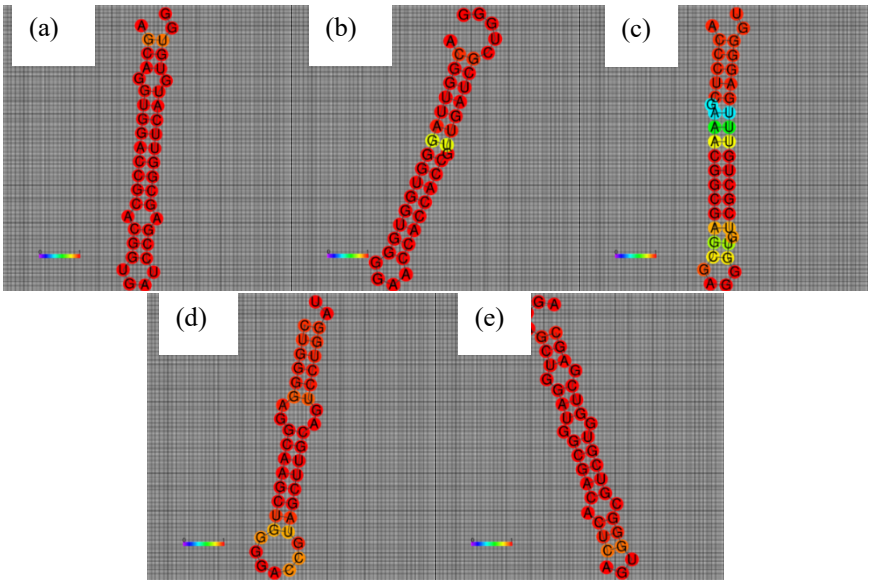

Fig. 3. MFE structure for (a) model 1; (b) model 2; (c) model 3; (d) model 4; (e) model 5.

Then, the five sequences of aptamer chosen are generated in RNAComposer software for three-dimensional structure in PDB file format as in Fig. 3 [14]. These five PDB files were analysed using BIOVIA Discovery Studio and ready to be docked with the target protein of Escherichia coli.

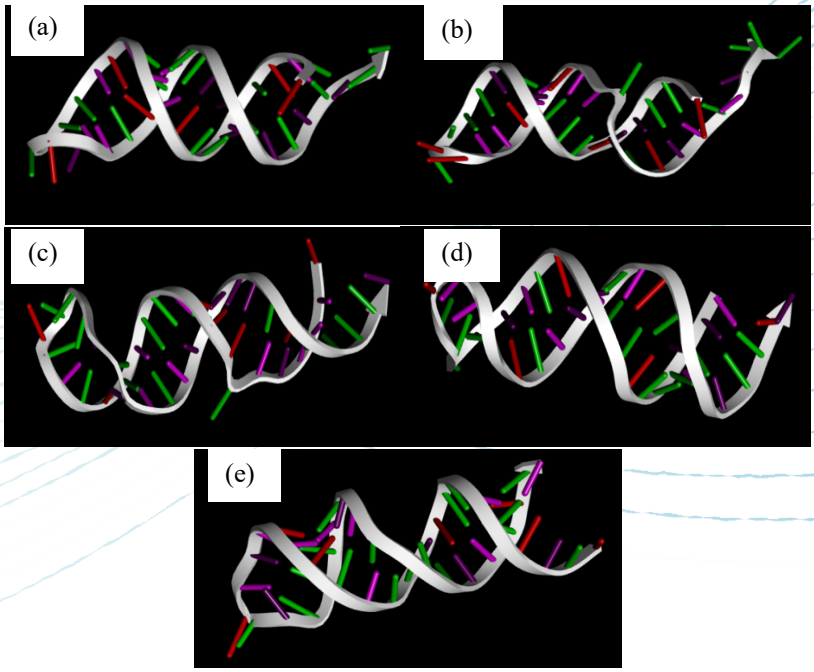

Fig. 4. 3D structures of 2AU4 Aptamer for (a) model 1; (b) model 2; (c) model 3; (d) model 4; (e) model 5.

2.3. Target protein Escherichia coli preparation

The crystal structure of target protein Escherichia coli was retrieved from PDB code: 1C48 called as Shiga Toxin (Stx). Shiga Toxin (Stx) is a potent cytotoxic protein that colonize human's colon and induce bloody diarrhea such as hemorrhagic colitis and hemolytic uremic syndrome that can lead to death [15]. There are many computationally approaches to generate accurate biomolecular structures predictions containing protein/peptide structures from amino acid sequences such as using PEP-FOLD4, AlphaFold 3, ESMFold, and OmegaFold [16,17]. Throughout this project, AlphaFold 3 were used as this software shows highly accurate prediction of the long peptide sequence structures [18]. The software operates the input sequence in FASTA format to read the peptide sequences. Next, the three-dimensional structures were visualized in Protein Data Bank (PDB) format using BIOVIA Discovery Studio software as shown in Fig. 5 and choose chart option to analyse the Ramachandran plot. The Ramachandran Plot was used to confirm the structure of peptide generate by these software [19]. Ramachandran Plot is a graphical representation of the dihedral angles, Phi (ϕ) and Psi (ψ) of amino acid residues in protein structures. There are four quadrants to explain the structure of protein as visualised in Fig. 6. In the quadrant-I that is the area of all conformations allowed, which it is the left-handed α-helix region. While quadrant-II, the biggest region, is the β-sheet region and quadrant-III is the right-handed α-helix region where both regions have better conditions for the conformation of atoms. Lastly, quadrant-IV has practically no framed locale [20].

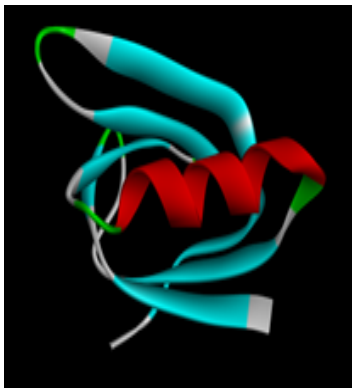

Fig. 5. 3D structures of Stx in the forms of solid ribbon

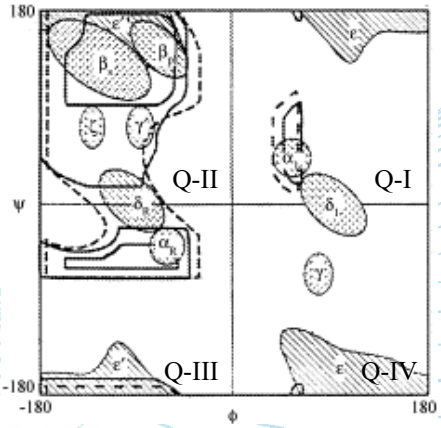

Fig. 6. Ramachandran Plot Quadrants [20]

2.4. Molecular docking of protein Escherichia coli and aptamer

Five chosen aptamers were docked with the protein Stx obtained from AlphaFold 3 as it shows stable structures as shown in Ramachandran plot, Fig. 7. The molecular docking process was using HDOCK server [21]. HDOCK server is a protein-protein and protein-DNA/RNA docking based on a hybrid algorithm of template-based modelling and ab initio free docking. The Stx protein as a ligand molecule and the chosen aptamer's PDB file was input as a receptor molecule. This was executed without any adding specification on the binding site. HDOCK results showed top ten possible docking residues prediction for each of five specific chosen aptamers and Stx protein in the PDB files format to visualize in the BIOVIA.

3.0 Results and Discussion

3.1. Structures of candidate aptamers and target protein Escherichia coli

Result from the Table 1, conclude that the aptamer model 1 with the sequence: AGCAGGUGGACCGCACGGUGAUCCGAGCGGUUCAUGUGUGG, has the lowest MFE values from RNAfoldz tools which is -24.0 kcal/mol and followed by model 2 which is -23.4 kcal/mol and model 3 is -22.6 kcal/mol. However, for models 4 and 5, the MFE values are decreased to -20.5 and -20.0 kcal/mol respectively. These results suggest that all the 2AU4 aptamer sequences exhibit lower MFE, suggesting structural stability [12].

From the Ramachandran plot, Fig. 7, Stx protein operated in AlphaFold 3 visualize that both regions, right-handed α-helix (quadrant III) and β-sheets (quadrant II) are the most prevalent structure in Stx. This suggest that Stx structures imparts stability to proteins and is crucial for their proper folding and function. From the analysis on both aptamer and protein, it is suggested to dock the top five model aptamer sequences with AlphaFold 3's structure.

Table 1. Molecular free energy (MFE) values for the best five aptamer models sequences from RNALfoldz tools

Aptamer Model	Aptamers sequences	MFE (kcal/mol)
1	AGCAGGUGGACCGCACGGUGAUCCGAGCGGUUCAUGUGUGG	-24.0
2	ACGGUUAGGGUGGUGGGGAACCACCACCGUUGAUCGCUGGG	-23.4
3	ACCCUCGAAACGGCGAGCGAGGGUGUCGCUGUUUGAGGGGU	-22.6
4	CUGGGGAGGCAAGCUGGGGACCGUAGCUUGCAGUCCUGGAU	-20.5
5	AGUAGCUGGAUGGCGACACUCAGUGGGCGUCGUGGUCGAGC	-20.0

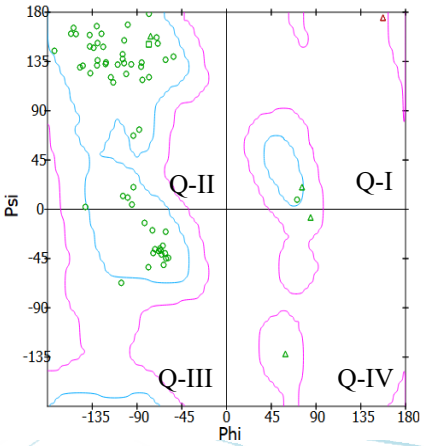

Fig. 7. Ramachandran plot for AlphaFold 3.

3.2. Molecular docking analysis

The analysis in 3.1 concludes that all of the 2AU4 aptamer structures are stable to be docking with AlphaFold 3's structures. Thus, molecular docking analysis was performed to estimate the binding affinity between the aptamer and Stx protein. There were 50 docking models produced from HDOCK server. The docking scores of the best prediction model in each of the five chosen aptamers with Stx are presented in Table 2, where the aptamers are ranked based on their docking scores. In general, the lower the rank, the better [22].

The scoring function that evaluates protein–aptamer binding is the most critical component of the docking method. A good dock score for a given aptamer signifies that it is potentially a good binder [23]. The results show that model 3 have the lowest docking scores of -267.50 as shown in Fig. 8. This is followed by model 2 and 5 with the binding energy of -250.90 and -244.88 respectively. However, even though model 1 has the lowest MFE values, it has higher docking scores of -232.13 compared to model 3, 2 and 5.

Next, averaging docking score proves informative method to account for multiple binding modes as computed by considering more proteins structures or more poses within a single structure [24]. The result from Table 3 suggests that model 2 aptamer has the lowest average docking scores of -227.73 for all the prediction models. However, model 1 only has an average energy of -216.93. This indicate that the docking score does not always directly correlate with MFE value. Even the aptamer itself is stable, there are also some effects that affect the docking score such as binding affinity between the aptamer and protein, and some aptamers undergo significant conformational changes upon binding.

Table 2. Docking scores for the best prediction model from five aptamer models sequences from HDOCK

Aptamer model	Aptamer sequence	Docking scores
3	ACCCUCGAAACGGCGAGCGAGGGUGUCGCUGUUU GAGGGGU	-267.50
2	ACGGUUAGGGUGGUGGGGAACCACCACCGUUGAU CGCUGGG	-250.90
5	AGUAGCUGGAUGGCGACACUCAGUGGGCGUCGUG GUCGAGC	-244.88
1	AGCAGGUGGACCGCACGGUGAUCCGAGCGGUUCA UGUGUGG	-232.13
4	CUGGGGAGGCAAGCUGGGGACCGUAGCUUGCAGU CCUGGAU	-220.23

Table 3. Average docking scores of ten models for the best five aptamer models sequences from HDOCK

Aptamer model	Aptamer sequence	Average docking scores
2	ACGGUUAGGGUGGUGGGGAACCACCACCGUUG AUCGCUGGG	-227.73
5	AGUAGCUGGAUGGCGACACUCAGUGGGCGUCG UGGUCGAGC	-222.05
1	AGCAGGUGGACCGCACGGUGAUCCGAGCGGUU CAUGUGUGG	-216.93
3	ACCCUCGAAACGGCGAGCGAGGGUGUCGCUGU UUGAGGGGU	-211.96
4	CUGGGGAGGCAAGCUGGGGACCGUAGCUUGCA GUCCUGGAU	-211.26

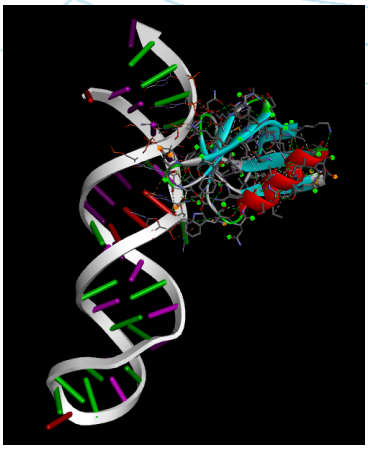

Fig. 8. 3D structures of best prediction model of model 3 2AU4 aptamer (white) binding with target protein Shiga Toxin 1C48 (red) in molecular docking process with -267.50 docking scores

4.0 Conclusion

In this study, in-silico methods have been used to investigate aptamer against the target protein of Escherichia coli using MFE calculation and molecular docking method. The analysis highlighted that the 2AU4 aptamer with the model 3 sequence and structure as in Fig. 8 displayed lower MFE values of -23.4 kcal/mol, lowest average docking score of -267.50 compared to others. However, as highlighted in Table 3, it is suggested that aptamer model 2 has the highest average docking scores value of its ten model predictions. This indicates that model 2's aptamer is stable and the interaction/binding strength between the AlphaFold 3's Shiga Toxin is higher. As a result, model 3 can be selected for the subsequent Molecular Dynamics (MD) simulation for advance confirmational stability such as Root Mean Square Deviation (RMSD) and Fluctuation (RMSF) plot of this protein-aptamer complex structure. In conclusion, because of the good result in docking analysis, we suggest that this aptamer model potentially is an effective detector that can be implemented in the future of aptamer-based biosensor technology such as colorimetric biosensor.

Acknowledgements

The authors gratefully acknowledge the huge support concerning the in-silico method from several experts, Aisyah Isrofi from USIM and Ahmed G Soliman from Ain Shams University. The authors also recognize the significant assistance from Dr Nazri Bahari for suggesting the aptamer design and Associate Professor Shinya Ikeno from Kyutech, Kyushu Institute of Technology as the principal investigator that evaluates the result of this research. This research was supported by Universiti Sains Islam Malaysia under USIM Research Grant (PPPI/USIM/FST/USIM/112623). The research was also funded by the Ministry of Higher Education, Malaysia under the Fundamental Research Grant Scheme (FRGS/1/2024/STG07/USIM/02/1).

References

[1] A. Rani, V. Ravindran, A. Surapaneni, E. Shahsavari, N. Haleyur, N. Mantri, A. Ball. (2022). Evaluation and comparison of recombinase polymerase amplification coupled with lateral-flow bioassay for Escherichia coli O157:H7 detection using different genes. *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–12. doi: 10.1038/s41598-021-81312-6.

[2] Z. Bai, X. Xu, C. Wang, T. Wang, C. Sun, S. Liu, D. Li. (2022). A comprehensive review of detection methods for Escherichia coli O157:H7. *TrAC Trends in Analytical Chemistry*, vol. 152, p. 116646. doi: 10.1016/J.TRAC.2022.116646.

[3] Y. Liu, H. Thaker, C. Wang, Z. Xu, M. Dong. (2022). Diagnosis and Treatment for Shiga Toxin-Producing Escherichia coli Associated Hemolytic Uremic Syndrome. *Toxins (Basel)*, vol. 15, no. 1. doi: 10.3390/TOXINS15010010.

[4] V. Léguillier, B. Heddi, J. Vidic. (2024). Recent Advances in Aptamer-Based Biosensors for Bacterial Detection. *Biosensors (Basel)*, vol. 14, no. 5, p. 210. doi: 10.3390/BIOS14050210/S1.

[5] J. B. Kaushal, P. Raut, S. Kumar. (2023). Organic Electronics in Biosensing: A Promising Frontier for Medical and Environmental Applications. *Biosensors 2023, Vol. 13, Page 976*, vol. 13, no. 11, p. 976. doi: 10.3390/BIOS13110976.

[6] M. Mohanty, P. S. Mohanty. (2023). Molecular docking in organic, inorganic, and hybrid systems: a tutorial review. *Monatsh Chem*, vol. 154, no. 7, pp. 683–707. doi: 10.1007/S00706-023-03076-1/FIGURES/19.

[7] K. Roy, S. Kar, R. N. Das. (2015). Other Related Techniques. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, pp. 357–425. doi: 10.1016/B978-0-12-801505-6.00010-7.

[8] C. Yang, E. A. Chen, Y. Zhang. (2022). Protein–Ligand Docking in the Machine-Learning Era. *Molecules*, vol. 27, no. 14, p. 4568. doi: 10.3390/MOLECULES27144568.

[9] H. Wang, Y. Zhao, S. Bie, T. Suo, G. Jia, B. Liu, R. Ye, Z. Li. (2019). Development of an Electrochemical Biosensor for Rapid and Effective Detection of Pathogenic Escherichia coli in Licorice Extract. *Applied Sciences 2019, Vol. 9, Page 295*, vol. 9, no. 2, p. 295. doi: 10.3390/APP9020295.

[10] S. Siddiqui, J. Yuan. (2021). Binding Characteristics Study of DNA based Aptamers for E. coli O157:H7. *Molecules*, vol. 26, no. 1. doi: 10.3390/MOLECULES26010204.

[11] C. A. Almazar, M. V. Mendoza, W. L. Rivera. (2023). In Silico Approaches for the Identification of Aptamer Binding Interactions to Leptospira spp. Cell Surface Proteins. *Trop Med Infect Dis*, vol. 8, no. 2, p. 125, doi: 10.3390/TROPICALMED8020125/S1.

[12] M. M. Alkriz, D. Joujeh. (2024). COMPUTATIONAL SELECTION AND OPTIMIZATION OF RNA APTAMER AGAINST ALPHA-FETOPROTEIN- AN IN SILICO STUDY. *Bulletin of Pharmaceutical Sciences Assiut University*, vol. 47, no. 2, pp. 989–1008. doi: 10.21608/BFSA.2024.299474.2176.

[13] G. Liang, J. Zhao, Y. Gao, T. Xie, J. Zhen, L. Pan, W. Gong. (2024). Application and evaluation of molecular docking for aptamer and small molecular interaction - A case study with tetracycline antibiotics. *Talanta*, vol. 266, p. 124942. doi: 10.1016/J.TALANTA.2023.124942.

[14] J. Sarzynska, M. Popenda, M. Antczak, M. Szachniuk. (2023). RNA tertiary structure prediction using RNAComposer in CASP15. *Proteins: Structure, Function, and Bioinformatics*, vol. 91, no. 12, pp. 1790–1799. doi: 10.1002/PROT.26578.

[15] I. Henrique, F. Sacerdoti, R. Ferreira, C. Henrique, M. Amaral, R. Piazza, D. Luz. (2022). Therapeutic Antibodies Against Shiga Toxins: Trends and Perspectives. *Frontiers in Cellular and Infection Microbiology*, vol. 12, p. 825856. doi: 10.3389/FCIMB.2022.825856/PDF.

[16] A. Elofsson. (2023). Progress at protein structure prediction, as seen in CASP15. *Curr Opin Struct Biol*, vol. 80, p. 102594. doi: 10.1016/J.SBI.2023.102594.

[17] J. Rey, S. Murail, S. De Vries, P. Derreumaux, P. Tuffery. (2023). PEP-FOLD4: a pH-dependent force field for peptide structure prediction in aqueous solution. *Nucleic Acids Res*, vol. 51, no. W1, pp. W432–W437. doi: 10.1093/NAR/GKAD376.

[18] J. Abramson, J. Adler, J. Dunger, et. al. (2024). Accurate structure prediction of biomolecular

interactions with AlphaFold 3. *Nature 2024 630:8016*, vol. 630, no. 8016, pp. 493–500. doi: 10.1038/s41586-024-07487-w.

[19] S. Agnihotry, R. K. Pathak, D. B. Singh, A. Tiwari, I. Hussain. (2022). Protein structure prediction. *Bioinformatics: Methods and Applications*, pp. 177–188. doi: 10.1016/B978-0-323-89775-4.00023-7.

[20] P. Chakrabarti, D. Pal. (2001). The interrelationships of side-chain and main-chain conformations in proteins. *Prog Biophys Mol Biol*, vol. 76, no. 1–2, pp. 1–102. doi: 10.1016/S0079-6107(01)00005-0.

[21] Y. Yan, H. Tao, J. He, S. Y. Huang. (2020). The HDOCK server for integrated protein–protein docking. *Nature Protocols 2020 15:5*, vol. 15, no. 5, pp. 1829–1852, doi: 10.1038/s41596-020-0312-x.

[22] K. B. Abd Halim, N. A. Samin, M. A. M. Yussof. (2022). COMPUTATIONAL STUDIES OF POTENTIAL EBOLA VP40 INHIBITORS USING BIOACTIVE COMPOUNDS FROM MEDICINAL PLANTS OF MALAYSIA. *Jurnal Teknologi (Sciences & Engineering)*, vol. 84, no. 4, pp. 183–190. doi: 10.11113/JURNALTEKNOLOGI.V84.17954.

[23] C. Blanes-Mira, P. Fernández-Aguado, J. de Andrés-López, A. Fernández-Carvajal, A. Ferrer-Montiel, G. Fernández-Ballester. (2022). Comprehensive Survey of Consensus Docking for High-Throughput Virtual Screening. *Molecules*, vol. 28, no. 1, p. 175. doi: 10.3390/MOLECULES28010175.

[24] G. Vistoli, A. Mazzolari, B. Testa, A. Pedretti. (2017). Binding Space Concept: A New Approach To Enhance the Reliability of Docking Scores and Its Application to Predicting Butyrylcholinesterase Hydrolytic Activity. *J Chem Inf Model*, vol. 57, no. 7, pp. 1691–1702, doi: 10.1021/ACS.JCIM.7B00121.