Computational Simulations of Aptamers Targeting Epithelial Cell Adhesion Molecule (EpCAM)+ Breast Cancer Cells

AMRITHA SHIVAPRASATH[1,2] AND GAURAV SHARMA[2]

[1]Lago High School, CA, Folsom
[2]Eigen Sciences, NC, Apex

Published August, 2025

Breast cancer remains the most prevalent cancer among women, with an incidence of over 2.3 million new cases in the United States. The epithelial cell adhesion molecule (EpCAM) receptor is highly expressed on the surface of breast cancer cells, where it plays a pivotal role in tumor progression and metastasis. Since the EpCAM receptor gets overexpressed on the surface of the breast cancer cells, they are an attractive target for both diagnosis and therapeutic application in breast cancer management. Aptamers are synthetic nucleic acids that selectively and with high affinity bind to the target molecules. We hypothesize that selected DNA-based aptamers can bind to the extracellular domain of EpCAM and serve as candidates for targeted breast cancer diagnostics. The structure of the EpCAM receptor was predicted using the AlphaFold 3 software, which provides insights into its function and structure. The aptamer 3D structure modeling was performed using the Vfold2D and Vfold3D software, respectively. In the next step, the aptamers were docked on the EpCAM receptor using the HDOCK2.0 software to obtain the EpCAM-aptamer complex output structures. These output structures were further validated using the binding energy calculations, and several interactions were formed between the EpCAM and aptamer. Based on these analyses, aptamer AS1411was selected as the most appropriate candidate for EpCAM-mediated breast cancer cell detection. This work advances the potential of EpCAM-targeted aptamers as targeted therapeutic and diagnostic agents against breast cancer.

1. INTRODUCTION

Breast cancer is a serious issue that affects millions of people worldwide and is one of the leading causes of cancer deaths in women. [1] The impact of breast cancer goes far beyond the physical challenges of the disease itself. Patients often deal with not only physical but also emotional trauma as they traverse through complex treatments like surgery, chemotherapy, or radiation. [1] The financial burden can also be overwhelming since treatments, follow-up care, and medications are expensive, especially for those without access to affordable healthcare. [2] This can leave families struggling to keep up with medical bills while also managing the emotional stress of supporting and caring for a loved one. [2] Caregivers, such as family and friends, often feel the pressure as they take on extra responsibilities, leading to burnout and exhaustion. On a larger scale, breast cancer affects entire communities by increasing healthcare demands worldwide. To ease this burden, it's important to focus on raising awareness and accessibility. This can be done by raising funds while improving the science of early detection, making treatments more effective and accessible for everyone. [3] The bar graph in Figure 1 illustrates the number of new breast cancer cases reported in 2022 among women in the top 10 countries globally. China recorded the highest number of cases, followed by the United States and India. [4] These figures represent absolute case numbers and reflect the influence of population size, healthcare infrastructure, and diagnostic capabilities. While China and India have large populations contributing to high case counts, countries like the United States and France show high detection rates due to widespread screening and early diagnosis. [4] It's important to note that these values do not account for differences in population age or size, which is why age-standardized rates are also used in epidemiology to compare countries more accurately. Nonetheless, this data highlights the global burden of breast cancer and the need for improved awareness, early detection, and healthcare access worldwide.

EpCAM, or Epithelial Cell Adhesion Molecule, is a protein mainly found in epithelial tissues throughout the body. It helps cells stick together and plays a role in cell growth [5]. EpCAM is often overproduced in various cancers, therefore making it

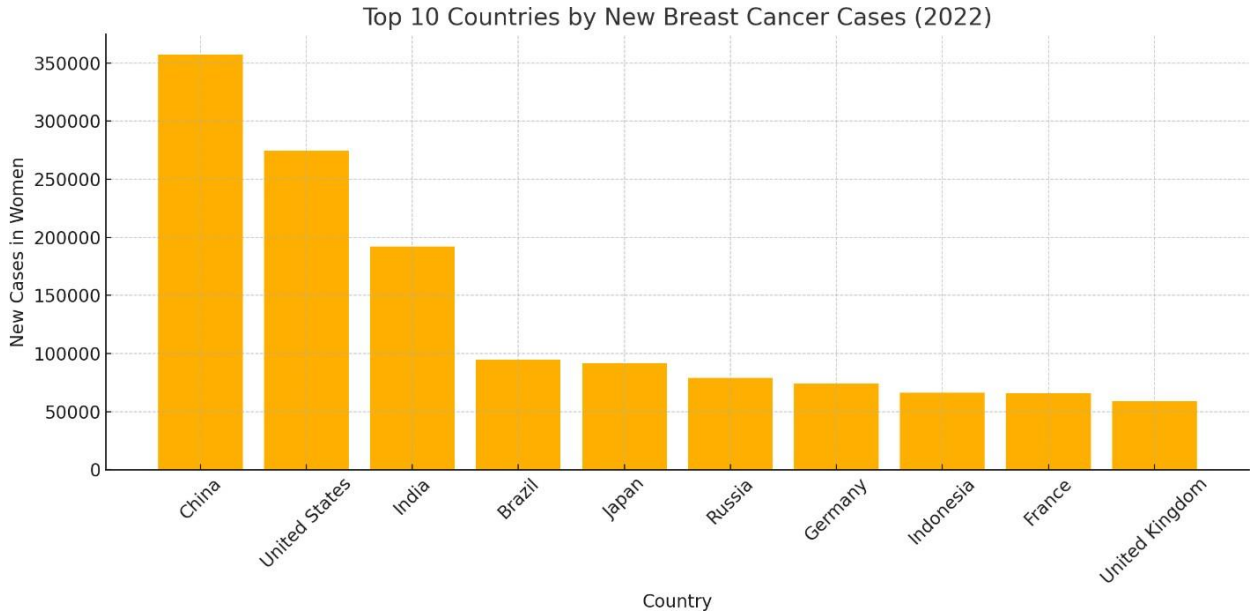

Fig. 1. New Breast Cancer Cases by Country (2022): This bar graph illustrates the number of new breast cancer cases reported in 2022 among women in the top 10 countries globally. China leads with the highest number of new cases, followed by the United States and India. These values reflect absolute case numbers and do not account for population size or age distribution.

a crucial target for diagnosis and treatment. Researchers are exploring therapies such as monoclonal antibodies targeting EpCAM specifically [6]. Additionally, its presence in tumor cells makes it a biomarker for detecting and monitoring cancer. By understanding EpCAM and how it works, scientists could lead to more efficient and improved cancer therapies. Figure 1 shows a strategy for treating breast duct cancer by targeting EpCAM. In normal breast duct tissue, cells have EpCAM on their surface. However, in cancerous tissue, cells display more EpCAM. Overexpression of EpCAM facilitates targeted therapeutic strategies, leading to apoptotic cell death in breast cancer tissues.

Aptamers are short RNA or DNA molecules that can bind to different things, such as proteins, molecules, and cells [7, 8]. Aptamers are specifically used to identify these targets in testing or drug discovery [9]. Most scientists prefer them because of their small sizes and low manufacturing costs. Aptamers can also be chemically modified to adapt to different conditions, such as extreme pH or temperatures. This versatility makes them valuable in diverse research and diagnostic applications. Also, aptamers can be explicitly modified to bind to reduce off-target effects in therapeutic settings [10].

Molecular docking is a technique that is used to identify molecules being bound to other molecules to form a stable complex (equilibrium) [11]This is crucial for designing new and effective drugs. Molecular docking is used mainly in pharmaceutical research, screening patients to predict how well different agents fit into an active site. [11]. The technique is still being studied for protein-protein interactions and enzyme mechanisms. Advancements in computational power and algorithms continue to improve the accuracy of docking predictions; consequently, molecular docking is becoming increasingly important in medicine and the development of various therapies. We hypothesize that selected DNA-based aptamers can bind to the extracellular domain of EpCAM and serve as candidates for targeted breast cancer diagnostics. The current research demonstrates that we have conducted molecular docking simulations to understand the aptamer-EpCAM interaction, which will help me select the most appropriate aptamer candidate. Based on our research, we have identified that the aptamer AS1411 is the best. The research will help in designing novel aptamer-based therapeutic strategies against breast cancer.

2. METHOD

In this research work, we have modeled the EpCAM and aptamers 3D structure to perform molecular docking simulations. we have used the following steps to model the aptamer 3D structures. First, using the UNAfold website, we uploaded the aptamer sequence and downloaded the .ct file after saving the PNG image of the aptamer 2D structures [12]. In the next step, we used the RNA structure server to convert the .ct file into dot and bracket notation [13]. Finally, on the VFOLD 3D website, copy the sequence and dot-bracket notation, set the number of clusters to 1, and submit the job to get the aptamers 3D structure. To perform the molecular docking simulations, we have used the HDOCK2.0 software [14]. **HDOCK** is an online platform for predicting protein docking on other elements using advanced computer algorithms to model and predict the interactions as accurately as possible. **P2Rank**, on the other hand, focuses more on predicting an exact binding position between proteins and/or DNA [15]. The predictions are based on sequence data, and machine learning is used to make the predictions. PDA-PRED is a tool that predicts protein-DNA by calculating the binding energy between protein and DNA sequences [16].In this research work, we have modeled the EpCAM and aptamers 3D structure to perform molecular docking simulations. we have used the following steps to model the aptamer 3D structures. First, using the UNAfold website, we uploaded the aptamer sequence and downloaded the .ct file after saving the PNG image of the aptamer 2D structures [12]. In the next step, we used the RNA structure server to convert the .ct file into dot and bracket notation [13]. Finally, on the VFOLD 3D website, copy the sequence

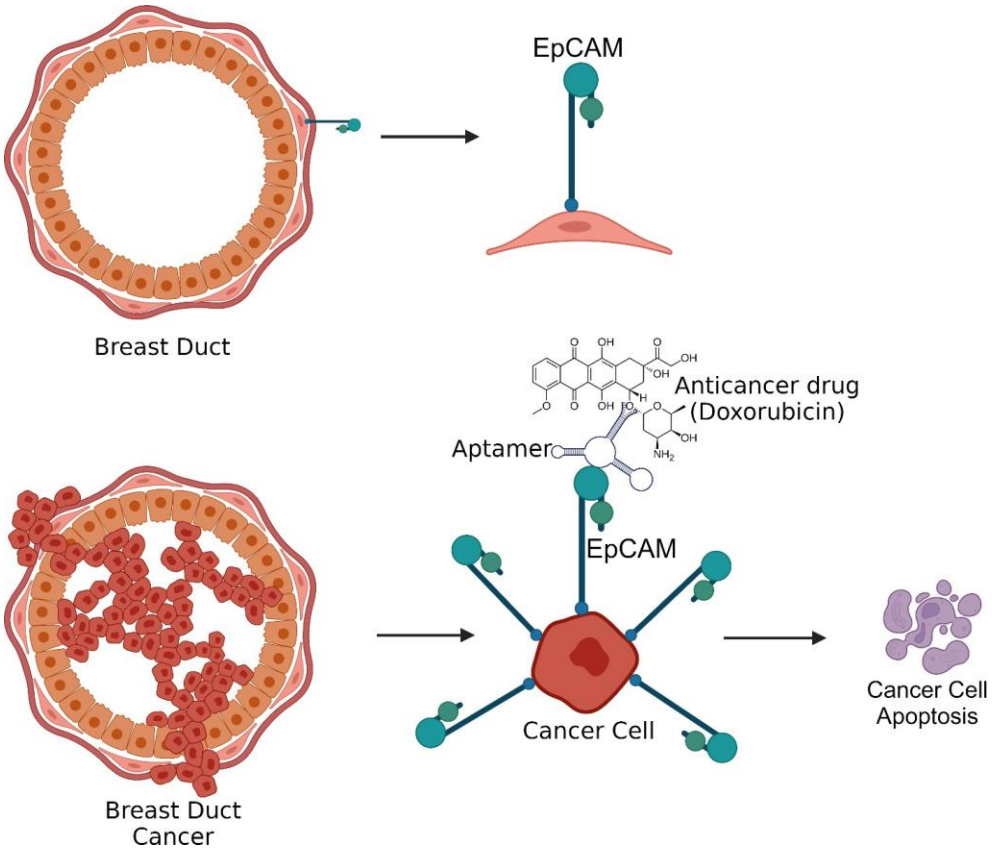

Fig. 2: Schematic of the current research work: EpCAM is less expressed on cells with healthy breast ducts (top). EpCAM is present in epithelial cells in these ducts. However, in the breast cancer duct (bottom), EpCAM is overexpressed. This targeted delivery leads to cancer cells' selective apoptosis (cell death), sparing healthy cells.

and dot-bracket notation, set the number of clusters to 1, and submit the job to get the aptamers 3D structure. To perform the molecular docking simulations, we have used the HDOCK2.0 software [14]. HDOCK is an online platform for predicting protein docking on other elements using advanced computer algorithms to model and predict the interactions as accurately as possible. P2Rank, on the other hand, focuses more on predicting an exact binding position between proteins and/or DNA [15]. The predictions are based on sequence data, and machine learning is used to make the predictions. PDA-PRED is a tool that predicts protein-DNA by calculating the binding energy between protein and DNA sequences [16]. The method focuses on the physical and chemical properties necessary for binding. The tool integrates features like amino acid composition, secondary structure, and DNA sequence-specific characteristics to make accurate predictions.

Summary of key aptamers evaluated in this study, including their nucleotide sequences, predicted secondary structures, root mean square deviation (RMSD) values from structural alignment (in Å), and molecular docking scores indicating binding affinity. Lower RMSD and docking scores suggest greater structural stability and stronger target binding, respectively.

3. RESULTS

To understand the EpCAM binding properties, we first explored this receptor's surface properties. First, the binding site on the surface of this receptor was obtained using the P2Rank software. A binding site is the site or region where the protein(s) bind with the aptamer. The sites can vary from aptamer to aptamer. Figure 2 shows the proposed binding site on the surface of the protein. Figure 2 also shows the three parts of the protein: i.e., ectodomain, which remains out of the cell; endo-domain, which stays inside the cell; and trans-domain, which stays in the plasma membrane.

I used molecular docking simulations to predict binding behaviors and interactions between a ligand and its target protein. [11] The results showed that EpCAM and Aptamers eventually become EpCAM-Aptamer complexes, which are displayed in Figure 3. [11] In the next step, we used the PLIP software to show the interactions between the EpCAM and aptamers, which are shown in Table 2. Finally, the binding energy was computed to get the strength of the EpCAM-aptamer interaction.

Table 1. Aptamer sequence and dot-bracket notations.

Aptamer Name	Sequence	Secondary Structure	RMSD (Å)	Docking Score
Sgc8c	AUCUAACUGCUGCGCCGCCGGGAAAAUACUGUACGGU-UAGA	.(((((((((.....)))))))))	2.1	125
AS1411	GGUGGUGGUGGUUGUGGUGGUGGUGG	......(.....).....	2.8	136
TCO1	ACCAAACACAGAUGCAACCUGACUUCUAACGUCAUU-UGGUG	((((((.....((((.....)))))))))).	1.9	117
Xq2d	ACUCAUAGGGUUAGGGGCUGCUGGCCAGAACUCA-GAUGGUAGGGUUACUAUGAGC	.(((((((.....(((.....)))..(((((.....))))))))))).	2.4	143
ApHER2	GCAGCGGUGUGGGGGCAGCGGUGUGGGGGCAGCG-GUGUGGGG	.((.((.(((....))).)).)).....	3.2	132
ApMUC1	GCAGUUGAUCCUUUGGAUACCCUGG	.....((((...))))....	1.7	119
EpCAM	CACUACAGAGGUUGCGUCUGUCCCACGUUGU-CAUGGGGGGUUGGCCUG	....(((((.....)))))((((.....)))).....2....5		138

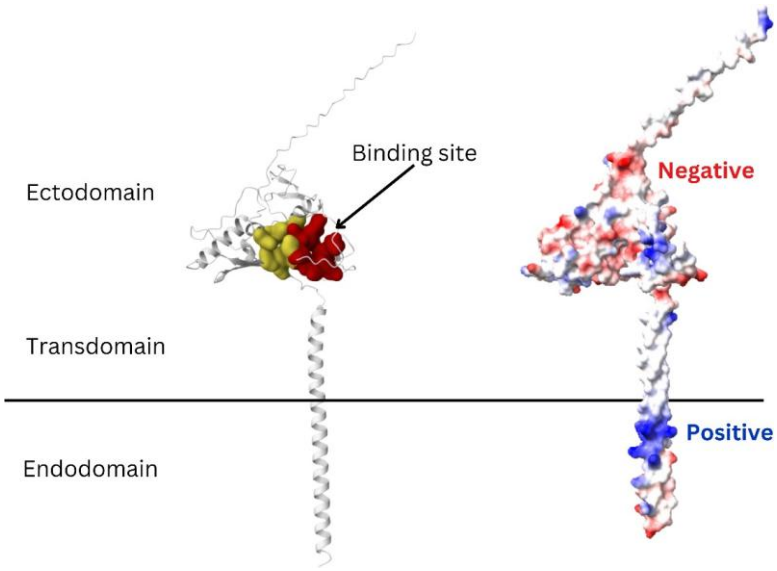

Fig. 3. Structural model of the EpCAM receptor, highlighting its ectodomain, transmembrane domain, and endodomain. The predicted binding site is indicated on the ectodomain (yellow and red regions), and the electrostatic potential model shows negative (red) and positive (blue) surface regions.

Table 2. Binding energy comparison of various aptamers when docked with the EpCAM receptor.

EpCAM-Aptamer Complex	EpCAM	HER2	AS1411	Sgc8c1	TCO1	Xq2d
Binding Energy	-14.81	-14.43	-17.39	-14.28	-14.65	-9.4

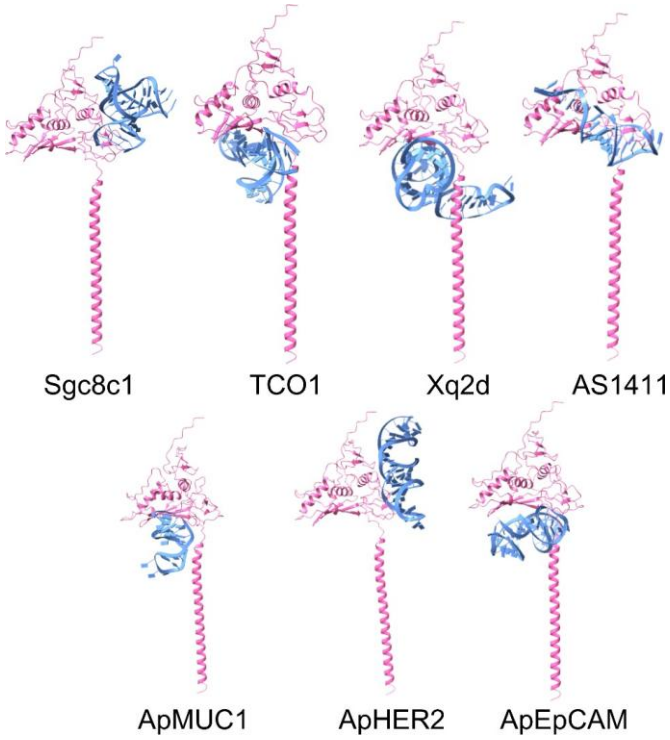

Fig. 4. Docked structure of EpCAM-aptamer complexes. The aptamers are shown interacting with the predicted binding site on the ectodomain of the EpCAM receptor.

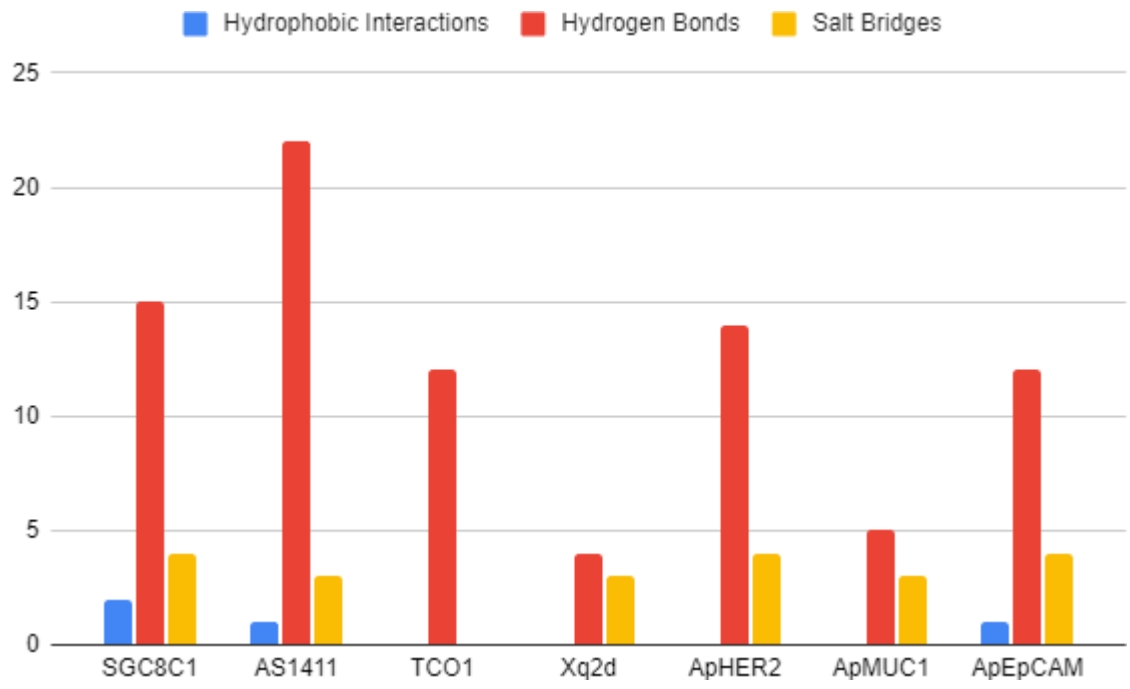

Fig. 5. Graphical representation of the number of interactions between EpCAM and different aptamers, as computed using the PLIP software. The interactions include hydrogen bonds, hydrophobic contacts, and electrostatic interactions.

4. DISCUSSION

The epithelial cell adhesion molecule (EPCAM) is a well-established biomarker in breast cancer diagnostics due to its frequent overexpression in epithelial-derived tumors. Panel (b) illustrates the protein-protein interaction (PPI) network of EPCAM, highlighting its associations with key cancer-related proteins such as TP53, IGFBPs, and CD74. These interactions suggest EPCAM's involvement in pathways regulating cell adhesion, proliferation, and immune response—all processes commonly dysregulated in breast cancer. Panel (c) provides a detailed topological representation of the EPCAM protein, showing its single transmembrane domain, extracellular accessibility, post-translational modifications (PTMs), and variants. This structural information is crucial for developing antibody- or aptamer-based diagnostic tools targeting the extracellular domain of EPCAM in tumor tissues or circulating tumor cells (CTCs). Although panel (a) focuses on TBKBP1 localization in the nucleoplasm, its inclusion offers a systems biology context, potentially connecting EPCAM's regulatory environment with nuclear signaling proteins that may influence breast cancer progression. Together, these data support the utility of EPCAM not only as a surface biomarker for detection and as a functional player in breast cancer pathophysiology.

The figure provides essential insights into the relevance of aptamer sequence relationships for breast cancer detection, explicitly focusing on EpCAM-targeting aptamers. EpCAM (Epithelial Cell Adhesion Molecule) is a well-established surface biomarker overexpressed in many epithelial cancers, including breast cancer. The aptamer EpCAMis designed to specifically bind to this protein, making it a valuable tool for early detection and targeted diagnostics. In panel (a), the multiple sequence alignment shows regions of sequence conservation and divergence among various aptamers, including EpCAM, ApHER2, AS1411, and others. The moderate conservation in EpCAM sequence suggests it retains unique binding motifs that are likely important for its specificity to EpCAM. Panel (b) illustrates a phylogenetic tree where EpCAM is shown to be evolutionarily distinct from aptamers such as ApHER2, which targets a different breast cancer biomarker (HER2). This separation underscores EpCAM potential for specific detection without cross-reactivity. Lastly, panel (c) presents a percent identity matrix showing that EpCAM shares only moderate sequence similarity (around 51–61%) with other aptamers. This degree of divergence supports the aptamer's specificity and reduces the risk of off-target interactions, a critical factor for developing reliable diagnostic tools. Collectively, the data validate EpCAM suitability as a selective probe for EpCAM-positive breast cancer cells and demonstrate its potential in clinical applications such as biosensing, targeted imaging, and precision diagnostics.

Application: Breast cancer is one of the most common types of cancer worldwide, and improving how we treat and detect it is very important. Targeted therapy is a newer, more efficient approach that focuses on the specific characteristics of cancer cells. This includes particular proteins or genes while avoiding harm to healthy cells, unlike traditional chemotherapy, which attacks all cells. For example, treatments like HER2 inhibitors block the growth of cancer cells with their proteins, which helps reduce damage to normal tissues and lowers the risk of countless side effects. Hormone-based therapies are another option for cancers that rely on hormones to grow. At the same time, advancements in diagnostic tools, like improved imaging and biomarker tests, make it so much easier for medical professionals to find breast cancer at an earlier stage– with much more accuracy. Early detection is crucial because it gives patients more treatment options and a better chance of recovery, especially when the tumor is still growing. By combining these targeted therapies with advanced diagnostic methods, doctors can provide more personalized care, helping patients live longer and healthier lives.

Current Treatments: Current breast cancer treatments typically include surgery, radiation therapy, chemotherapy, and hormone therapy, depending on the type and stage of the tumor. Surgery is often the first step (specifically for non-metastatic tumors), which gives the patient the option of simply removing the cancer or a full mastectomy. Radiation therapy has also proven to be effective by destroying the remaining cancer cells after surgery. Chemotherapy may be prescribed to kill cancer cells or shrink tumors drastically to prepare for removal. Hormone therapy specifically targets hormone-sensitive cancers, while a similar method, targeted therapy, focuses on the individual molecules involved in cancer growth. Newer techniques, such as Immunotherapy, boost the patient's immune system to fight cancer. Overall, treatment plans vary from patient to patient and are incredibly personalized based on the genetic and molecular profiling of the tumor.

Limitations: The complete 3D structure of the occludin-EpCAM has yet to be available, so we have to use computational tools to model or predict its binding. While these models offer insights and are helpful visually, their validity must be scientifically checked. Future lab studies will be necessary to confirm the relationship between occludin and EpCAM.

5. CONCLUSION

In the current research, we have computationally simulated EpCAM receptors present on the surface of breast cancer. These EpCAM receptors become overexpressed, and RNA-based aptamers have been designed using computational techniques to identify the most appropriate aptamer that can bind to the receptor. Based on this research, aptamer AS1411 was selected as the most suitable candidate for detecting and treating breast cancer cells.

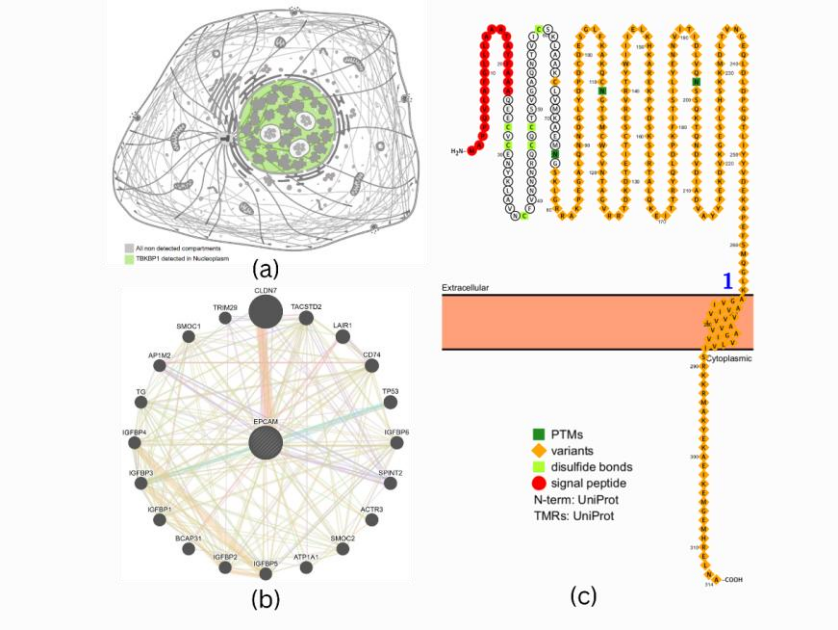

Fig. 6. (a) Subcellular Localization of TBKBP1 Protein: Diagrammatic representation of the human cell showing the subcellular localization of TBKBP1 (TANK-binding kinase 1-binding protein 1), predominantly detected in the nucleoplasm (highlighted in green). Other cellular compartments where TBKBP1 was not detected are shown in grey; (b) Protein-Protein Interaction (PPI) Network of EPCAM: A STRING-based interaction network centered around EPCAM (Epithelial Cell Adhesion Molecule), showing predicted and known protein-protein associations. Nodes represent proteins, while edges depict interactions supported by various evidence such as co-expression, experimental data, and text mining; and (c) Topological and Functional Annotation of EPCAM Protein: A schematic diagram showing the EPCAM protein topology. The protein contains a signal peptide (red), disulfide bonds (yellow lines), post-translational modification (PTM) sites (green), and known sequence variants (orange diamonds). The protein spans the membrane once, with extracellular and cytoplasmic regions indicated.

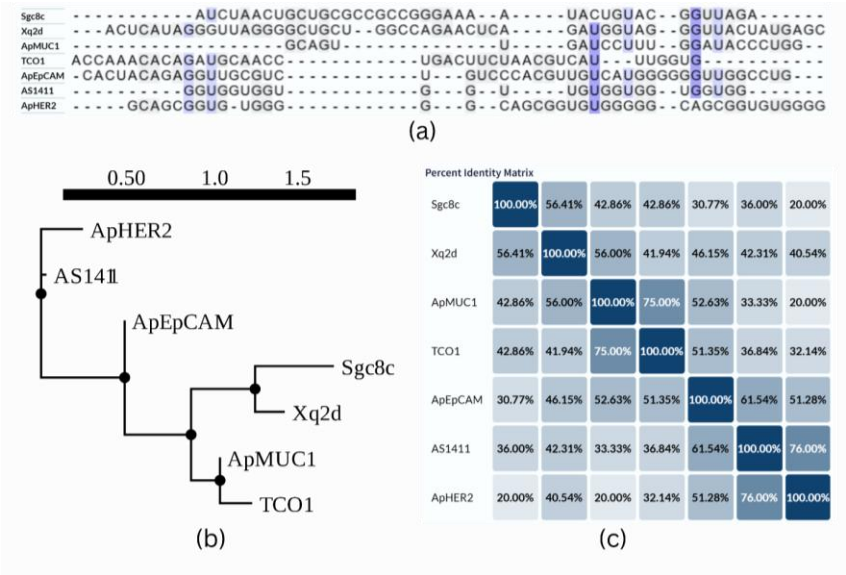

Fig. 7. (a) Multiple sequence alignment of various aptamer sequences, highlighting conserved regions (in purple) among Sgc8c, Xq2d, ApMUC1, TCO1, EpCAM, AS1411, and ApHER2; (b) Phylogenetic tree illustrating the evolutionary relationship between the aptamers based on sequence similarity. ApHER2 and AS1411 cluster closely, indicating higher similarity; and (c) Percent identity matrix showing pairwise sequence identity percentages among the aptamers. Darker shades represent higher sequence identity, with maximum identity (100%) along the diagonal.

REFERENCES

1. J. R. Benson and I. Jatoi, “The Global Breast Cancer Burden,” Future Oncology 8, 697–702.
2. A. M. Arozullah, “The financial burden of cancer: estimates from a study of insured women with breast cancer,” J Support Oncol 2, 271– 278 (2004).
3. Y. Yarden, “Biology of HER2 and its importance in breast cancer,” Oncology 61, 1–13 (2001).
4. G. D. Osborn, M. Hodin, P. J. Drew, H. Fielder, E. Vaughan-Williams, and H. M. Sweetland, “Patient demographics and treatment for early breast cancer: An observational study,” The Breast 15, 377–381 (2006).
5. P. A. Baeuerle and O. Gires, “EpCAM (CD326) finding its role in cancer,” British Journal of Cancer 96, 417–423 (2007).
6. U. Schnell, V. Cirulli, and B. N. G. Giepmans, “EpCAM: Structure and function in health and disease,” Biochimica et Biophysica Acta (BBA) - Biomembranes 1828 (1989).
7. A. D. Keefe, S. Pai, and A. Ellington, “Aptamers as therapeutics,” Nature Reviews Drug Discovery 9, 537–550 (2010).
8. G. T. Rozenblum, V. G. Lopez, A. D. Vitullo, and M. Radrizzani, “Ap- tamers: current challenges and future prospects,” Expert Opin Drug Discov 11, 127–162 (2016).
9. Y. Kim, N. H. A. Raston, M. Bock, and Gu, “Aptamer-based nanobiosen- sors,” Biosensors and Bioelectronics 76, 2–19 (2016).
10. M. R. Dunn, R. M. Jimenez, and J. C. Chaput, “Analysis of aptamer discovery and technology,” Nature Reviews Chemistry 1, 2017–2017 (2017).
11. J. Fan, A. Fu, and L. Zhang, “Progress in molecular docking,” Quantita- tive Biology 7, 83–89 (2019).
12. N. R. Markham and M. Zuker, “UNAFold,” in “Bioinformatics: Structure, Function and Applications,” J. M. K. Ed, ed. (Humana Press, 2008), pp. 3–31.
13. J. S. Reuter and D. H. Mathews, “RNAstructure: software for RNA secondary structure prediction and analysis,” BMC Bioinformatics 11, 2010–2010 (2010).
14. Y. Yan, D. Zhang, P. Zhou, B. Li, and S. Y. Huang, “HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy,” Nucleic Acids Research 45, 365–373 (2017).
15. R. Krivák and D. Hoksza, “P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure,” Journal of Cheminformatics 10, 39–39 (2018).
16. K. Harini, D. Kihara, and M. M. Gromiha, “PDA-Pred: Predicting the binding affinity of protein-DNA complexes using machine learning tech- niques and structural features,” Methods 213, 10–17 (2023).