

Structure-Guided Modeling of Proteolysis-Targeting Chimera (PROTAC)-Aptamer Conjugates Targeting Nucleolin Protein in Glioblastoma

MEERA MISHRA^{1,2}, Gaurav Sharma²

¹Stratford Preparatory High School, Sunnyvale, CA

²Gaurav Sharma, Eigen Sciences, Apex, NC

Published November, 2025

Glioblastoma (GBM) is an aggressive and fast-growing brain cancer with no known cure, and current research focuses on developing new strategies to slow its progression and improve patient outcomes. Nucleolin is an intracellular protein frequently overexpressed in cancer cells, where it plays key roles in cell cycle regulation, protein quality control, apoptosis, and antiviral immunity. Its upregulation contributes to oncogenesis, making it a critical therapeutic target in GBM. Proteolysis Targeting Chimeras (PROTACs) are bifunctional molecules that induce targeted protein degradation by linking a target protein to an E3 ligase, leading to its ubiquitination and destruction by the proteasome. We hypothesize that an aptamer-based PROTAC can effectively bring nucleolin into proximity with an E3 ligase, thereby facilitating nucleolin's proteasomal degradation. This research involves the computational modeling of the nucleolin protein, an E3 ligase, and a PROTAC-aptamer conjugate using AlphaFold. The results demonstrate that the designed aptamer binds strongly to a predicted site on nucleolin, and that the PROTAC architecture enables proximity to the predicted binding site of an E3 ligase. This current research demonstrates the potential of aptamer-based PROTACs to selectively degrade oncogenic nucleolin proteins, offering a promising new approach to slow tumor growth and overcome therapeutic resistance. Future applications include the development of targeted treatments that improve patient outcomes.

1. INTRODUCTION

Glioblastoma multiforme (GBM) most often affects older adults, with the highest rates seen between the ages of 75 and 84 and a median age of diagnosis around 64.[1] Glioblastoma (GBM) occurs worldwide, but it is most commonly diagnosed in countries with large, well-developed healthcare systems because of better MRI/CT access and cancer reporting. Men are more likely to be diagnosed than women, with the incidence being roughly 1.6 times higher in males.[1] In terms of race and ethnicity, white individuals, specifically non-Hispanic European Americans, have significantly higher rates of GBM compared to African Americans, Hispanics, Asians, and American Indians. Geographically, the tumor most frequently appears in the frontal and temporal lobes of the brain and is rarely found in the cerebellum or spinal cord. Genetic factors play a role, as GBM can occur more often in individuals with certain hereditary tumor syndromes such as Li-Fraumeni or Turcot syndrome. While GBM typically arises without a clear genetic link, specific polymorphisms have been

associated with an increased risk in certain populations.[2]

GBM is an aggressive and fast-growing cancer that originates in astrocytes (cells that support the brain and spinal cord).[1] Common symptoms include vision changes, seizures, memory loss, confusion, and changes in mood or personality.[3] Current treatments typically involve a combination of surgery, radiation, and chemotherapy, with newer options like tumor-treating fields and targeted therapy sometimes added. Surgery is usually the first step, but complete removal is often impossible because the tumor spreads into healthy brain tissue. Radiation and chemotherapy can help slow the tumor's growth but usually come with serious side effects and limited long-term success. Unfortunately, despite aggressive treatment, GBM almost always returns, and the average survival time remains about 15 months, highlighting the urgent need for more effective therapies.[2]

Nucleolin [4] is a multifunctional protein primarily located in the nucleolus, where it plays a central role in ribosome biogenesis, including rRNA transcription, maturation, and assembly. Although nuclear principally, nucleolin can also be found on the cell surface, particularly in cancer cells, where its pres-

ence is associated with altered signaling and tumor progression. Functionally, nucleolin interacts with both proteins and RNA, regulating processes like chromatin remodeling, mRNA stability, and translation.[5] In GBM, nucleolin is upregulated and contributes to several cancer hallmarks, such as evading apoptosis, promoting angiogenesis, and supporting uncontrolled cell proliferation.[6] It enhances the translation of pro-survival and pro-growth proteins like Bcl-2, Akt1, and MMP9, while suppressing tumor suppressors like p53. Recent studies show that nucleolin also drives endothelial cell proliferation and angiogenesis in GBM, making it a potential therapeutic target for this aggressive brain tumor.

PROTAC is a small, bifunctional molecule that targets specific proteins for degradation by the cell's own proteasome system.[7] It links a target protein to an E3 ligase, triggering the protein's breakdown. PROTACs are used in therapeutic research to treat diseases like cancer, neurodegenerative disorders, and autoimmune conditions. They are beneficial for targeting proteins that are hard to inhibit with traditional drugs. A PROTAC is made by chemically linking two ligands (one that binds the target protein and one that binds an E3 ligase) using a flexible linker. This forms a single molecule that brings the two proteins together to initiate degradation.[7]

2. METHOD

UniProt is a high-quality, comprehensive, and freely accessible resource of protein sequence information, including detailed annotations about their function, taxonomy, literature citations, and more.[8] We used UniProt to determine the amino acid sequence of the protein nucleolin. AlphaFold 3 is an AI model developed by Google DeepMind and Isomorphic Labs that predicts the 3D structures of proteins, DNA, RNA, ligands, and other biomolecules based on their amino acid sequences, as well as how these molecules interact with one another. We used AlphaFold 3 to model the protein nucleolin based on its amino acid sequence.[9] This allowed us to observe various features of the protein, including its 3D shape, coulombic charge distribution, and hydrogen bonding. PROTTER is a web-based tool that enables users to visualize and integrate both predicted and annotated protein sequence features.[10] It also maps experimental proteomic data, such as peptides and post-translational modifications (PTMs), onto the protein's transmembrane topology. We used PROTTER to identify the spatial location of the protein nucleolin- namely if it is an extracellular, transmembrane, or intracellular protein. From this topology visualization, we concluded that the protein nucleolin is an intracellular protein. The Human Protein Atlas (HPA) is an open-access database that provides detailed information on the expression of human proteins across various tissues, cells, and organs.[11] Its goal is to provide a comprehensive overview of the human proteome, supporting research in human biology, disease biomarker identification, and the development of new treatments. We used The Human Protein Atlas to identify the subcellular localization of the protein nucleolin. From this analysis, we concluded that nucleolin is localized to the nucleoplasm, nucleolus, and mitotic chromosomes. A predicted binding site is a region on a protein surface where it is likely to interact with another molecule, such as a ligand, another protein, or a drug. Computational tools predict these sites based on the protein's 3D structure, identifying pockets or surfaces with features that suggest potential binding activity. P2Rank is a machine learning-based tool that predicts ligand binding sites on proteins using their 3D structures. It analyzes

the geometry and properties of the protein surface to identify the most probable binding pockets.[12] We used P2Rank by inputting the structures of E3 ligase and nucleolin to determine their predicted binding sites. This helped us understand where interactions or modifications (like small molecule binding) are most likely to occur. ScanNet is a deep learning tool that predicts protein binding sites.[13] It uses neural networks trained on 3D structural data to generate heatmaps that visualize probable interaction regions. We used ScanNet by uploading our proteins to visualize and confirm potential binding interfaces. The resulting figures helped us better understand which regions of the proteins might be involved in protein-protein interactions. Aptamer modeling was conducted in three stages: transitioning from the 1D sequence to a 2D structure, and ultimately to a 3D conformation. First, the linear nucleotide sequence of the aptamer was uploaded to the UNAFold and RNAstructure Predict1 Server[14] (<https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>), which generated a connectivity table (ct file) and a predicted 2D structure image; the latter was saved for visualization, and the ct file was downloaded for further use. Next, the ct file was uploaded to the ct2dot server (<https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/ct2dot/ct2dot.html>), where the sequence's secondary structure was converted into dot-bracket notation, which was recorded in the manuscript for reference. Finally, both the sequence and its corresponding dot-bracket notation were submitted to the FARFAR2 server on ROSIE (<https://rosie.graylab.jhu.edu/farf2>) to predict the aptamer's 3D structure.[15] Molecular docking simulations are computational methods used to predict how two molecules, such as a protein and an RNA aptamer, interact and bind at the atomic level. They help visualize the binding orientation and estimate the strength and stability of the interaction. We used molecular docking to evaluate how well each designed aptamer could bind to the nucleolin protein, helping us identify the most promising candidates for therapeutic targeting. The PROTAC was modelled using the YASARA software.[16] The aptamers and PROTACs were docked on the respective protein using the HDock software.[17]

3. RESULTS

The structure of the E3 ligase was analyzed to identify a potential binding site, predicted using P2Rank (Fig. 2a). This pocket is highlighted in red and appears suitable for small molecule interaction, potentially being able to serve as the E3-binding region of a PROTAC. A surface model (Fig. 2b), created using ScanNet, shows that this site is easily accessible, while electrostatic mapping using UCSF Chimera reveals a mix of positive and negative charges, which influence the binding of the ligand. The overall fold of the protein is shown in the ribbon diagram (Fig. 2c), allowing us to visualize how the binding site is structurally supported. Similarly, in Figure 2d-f, the predicted binding pocket on nucleolin is marked in red (Fig. 2e). ScanNet predictions (Fig. 2d) indicate two likely protein-protein interaction regions, and the electrostatic surface analysis (Fig. 2e) shows these areas have polarized charges, which would support binding. The ribbon diagram (Fig. 2f) outlines nucleolin tertiary structure and how the binding pockets are integrated into the protein's fold.

4. DISCUSSION

The blood-brain barrier (BBB) is a specialized, selectively permeable interface that separates the circulating blood from the central nervous system (CNS).[18] It is primarily composed of

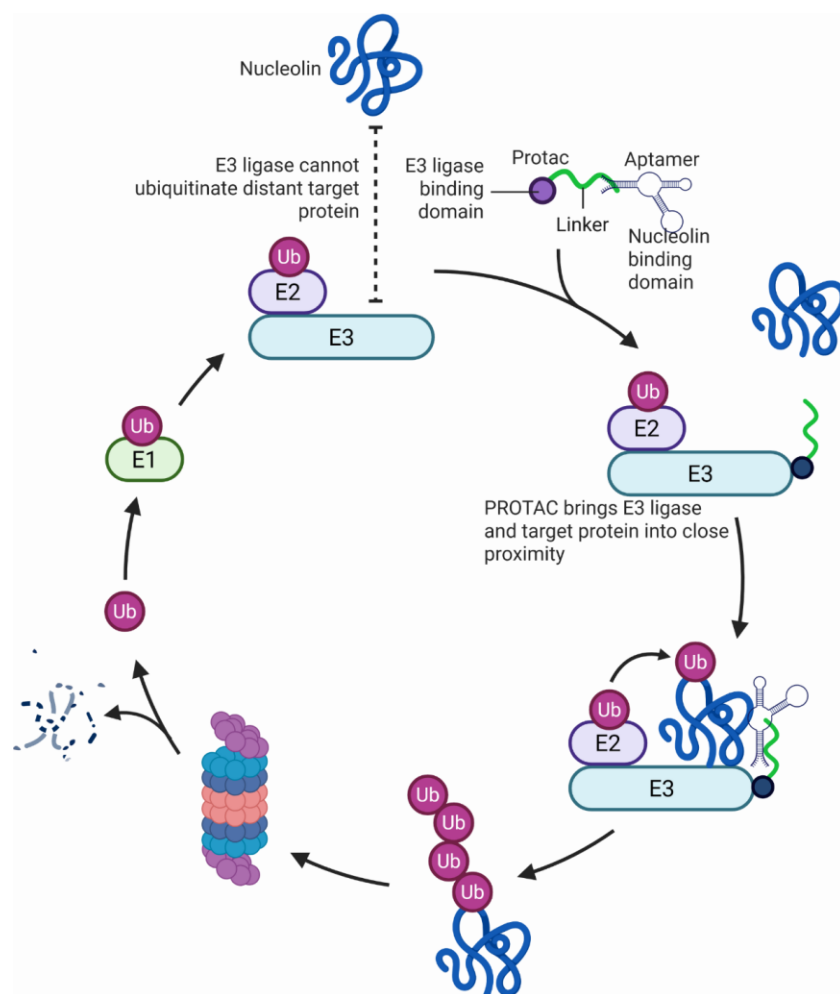


Fig. 1. The figure shows how an aptamer-based PROTAC recruits nucleolin to an E3 ligase, enabling its ubiquitination and subsequent degradation by the proteasome. Without the PROTAC, the E3 ligase cannot reach nucleolin, but the aptamer brings them into proximity to trigger targeted degradation. The figure was generated by using BioRender.com

endothelial cells joined by tight junctions, supported by pericytes and astrocytic end-feet. The BBB plays a vital role in maintaining CNS homeostasis by restricting the passive diffusion of most compounds, particularly polar and large molecules, while allowing the regulated transport of essential nutrients.[19] Due to this restrictive nature, the ability of a drug to penetrate the BBB is a key determinant in the development of effective CNS-targeted therapeutics.[18] Molecules that can cross the BBB typically possess specific physicochemical properties that enable passive diffusion through the lipophilic membranes of the barrier or are substrates for active transport mechanisms.[20] In the broader context of drug discovery, the Lipinski Rule of Five provides a set of empirical guidelines used to predict the oral bioavailability of small-molecule compounds.[20] According to this rule, a drug-like compound is more likely to be orally active if it has a molecular weight of less than or equal to 500 Daltons, a logP value (the logarithm of the octanol-water partition coefficient) of 5 or less, no more than five hydrogen bond donors, and no more than ten hydrogen bond acceptors. These parameters reflect the balance between hydrophilicity and lipophilicity required for adequate membrane permeability and solubility.[20]

While these rules do not guarantee BBB penetration, they serve as a valuable baseline for assessing general drug-likeness.

Beyond Lipinski's criteria, several additional physicochemical properties make up the pharmacokinetic profile of a compound, including human intestinal absorption, oral bioavailability, aqueous solubility, and lipophilicity. Human intestinal absorption (HIA) refers to the proportion of an orally administered compound that is absorbed through the intestinal epithelium into the systemic circulation.[21] High HIA is needed for effective oral bioavailability and is often correlated with moderate lipophilicity and low molecular polarity.[21] Oral bioavailability, more broadly, is defined as the fraction of an administered dose that reaches systemic circulation in its active form.[22] This parameter is determined by multiple factors including solubility, stability in the gastrointestinal tract, permeability across intestinal membranes, and the extent of first-pass metabolism. Aqueous solubility is the concentration at which a compound dissolves in water and is a critical factor for drug absorption. Poor solubility often limits the dissolution of the compound in the gastrointestinal tract, thereby reducing its absorption, even if membrane permeability is high. Lipophilicity, typically measured by

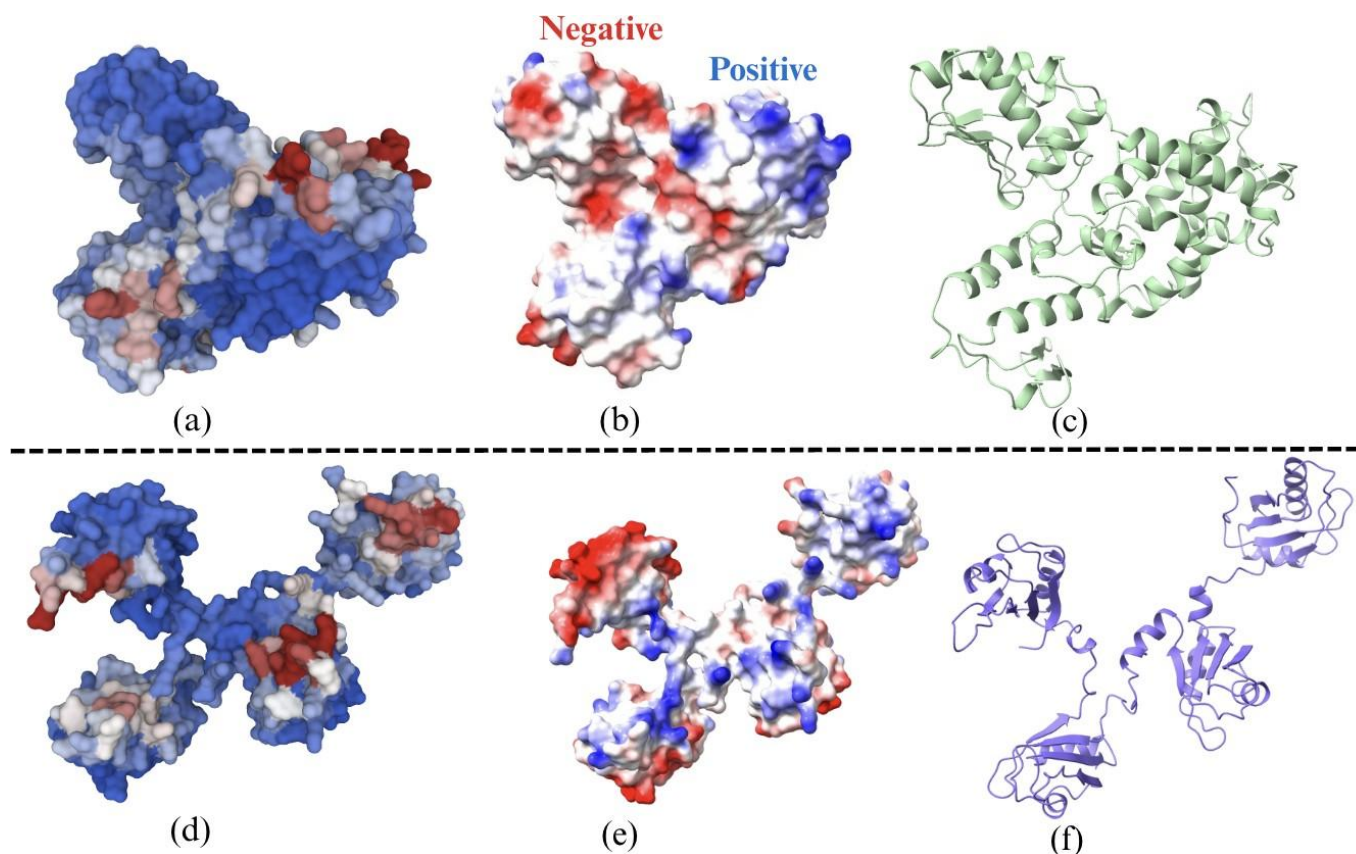


Fig. 2. Binding site and electrostatic surface predictions for E3 ligase and nucleolin. (a–c) Binding site prediction and surface analysis of E3 ligase; (a) P2Rank-predicted binding pocket (red) shown on the protein structure; (b) Electrostatic surface potential: red = negative, blue = positive; (c) Ribbon diagram showing overall fold; (d) P2Rank binding site (red/yellow) overlaid on the protein; (e) ScanNet-identified protein-protein interaction sites boxed on the surface; (g) Electrostatic potential map with interaction regions highlighted; and (f) Ribbon structure of nucleolin.

logP or logD values, describes a molecule's affinity for lipid environments relative to aqueous environments.[23] Compounds intended to cross the BBB generally require moderate to high lipophilicity to diffuse across the lipid-rich endothelial membranes of the brain vasculature. Taken together, these factors determine whether a compound is likely to be orally bioavailable, and thus capable of penetrating the BBB. Understanding these properties allows us to properly design CNS-active drugs with favorable pharmacokinetic and pharmacodynamic profiles. A PROTAC–aptamer conjugation strategy integrates the selectivity of nucleic acid aptamers with the targeted protein degradation capability of PROTACs to achieve precise, cell-specific therapeutics. In this design, the aptamer—engineered to bind a specific cell surface receptor such as EGFR, PSMA, or CD133—is chemically linked to a PROTAC molecule via a flexible and biocompatible linker. The conjugation can be achieved using bioorthogonal chemistries such as azide–alkyne “click” reactions, amide coupling, or maleimide–thiol conjugation, typically through modifications at the 5' or 3' terminus of the aptamer. The linker, often composed of PEG or a cleavable disulfide bond, preserves aptamer folding while allowing intracellular release of the PROTAC once internalized. Upon receptor-mediated endocytosis, the reducing environment of the cytosol cleaves the linker, releasing the active PROTAC. The PROTAC then induces ubiqu-

itination and proteasomal degradation of the target protein by bridging it to an E3 ligase such as VHL or CRBN. This modular approach enhances therapeutic precision by combining aptamer-guided targeting with PROTAC-driven degradation, minimizing off-target effects and improving intracellular delivery efficiency. The experimental validation of a PROTAC–aptamer conjugate involves confirming its synthesis, binding specificity, cellular uptake, and degradation efficiency. First, MALDI-TOF, HPLC, and NMR are used to verify successful conjugation and purity, while CD spectroscopy ensures the aptamer retains its proper folding. Next, SPR, MST, or flow cytometry confirm specific receptor binding. Fluorescent microscopy then demonstrates receptor-mediated internalization and intracellular localization. Functionally, Western blotting and immunoprecipitation confirm target protein degradation and ubiquitination following PROTAC release. Finally, cell viability assays (MTT/CCK-8) and apoptosis analyses (Annexin V/PI) evaluate therapeutic efficacy and cytotoxicity, ensuring the conjugate maintains both aptamer targeting and PROTAC degradation capabilities.

While our computational modeling successfully identified potential binding sites on both nucleolin and the E3 ligase, and demonstrated favorable aptamer–nucleolin interactions, it also presented significant pharmacokinetic challenges for clinical application. The modeled PROTAC–aptamer conjugate, being

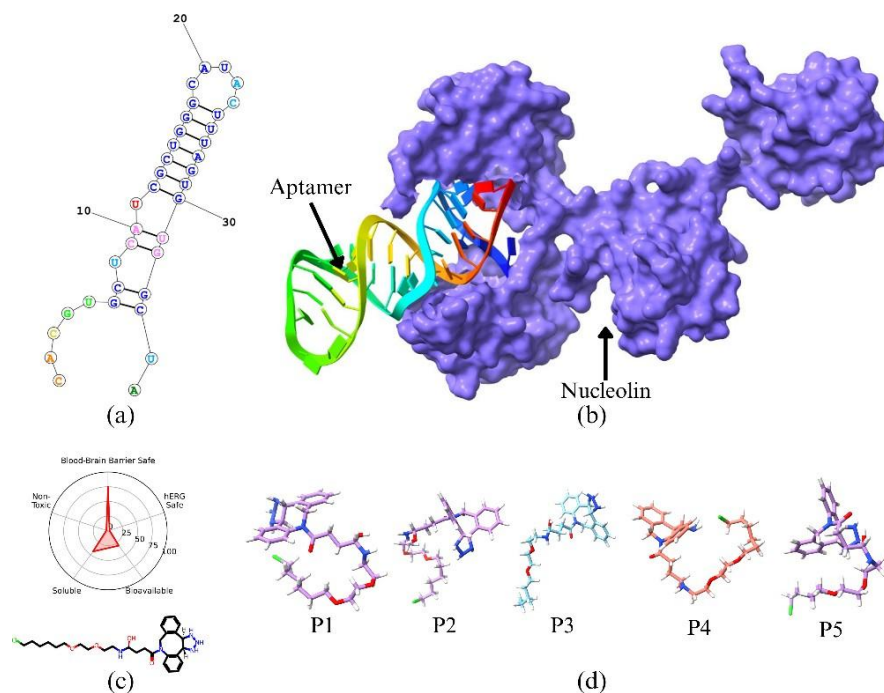


Fig. 3. (a) The secondary structure of the selected aptamer; (b) the nucleolin-bound aptamers obtained from molecular docking simulations; (c) radar graph of nucleolin; and (d) different conformers of the selected PROTAC.

an RNA aptamer, inherently exceeds Lipinski's Rule of Five criteria for small molecules, with a molecular weight of 500.14 Daltons and a Topological Polar Surface Area (TPSA) of 107.12 Å². Our DrugBank analysis further predicts a human intestinal absorption of 0.97 (33.85th percentile), low oral bioavailability of 0.06 (30.83rd percentile), and limited aqueous solubility of -3.38 log(mol/L) (43.20th percentile). Critically, the predicted BBB penetration is only 0.42 (25.51th percentile), strongly indicating that passive diffusion will be insufficient for CNS targeting. Furthermore, preliminary toxicity predictions show a clinical toxicity of 0.70 (97.90th percentile) and an acute toxicity (LD₅₀) of 3.05 (81.16th percentile). All of these results emphasize the necessity of comprehensive experimental validation of both efficacy and safety. These computational limitations also reveal that while our findings provide a promising initial structural basis, extensive experimental work, particularly addressing BBB penetration and potential toxicity, will be extremely critical for the development of these aptamer-based PROTACs as viable therapeutics for GBM. Finally, to overcome pharmacokinetic challenges and enhance blood-brain barrier (BBB) penetration of RNA-based PROTACs, advanced delivery systems such as liposomal encapsulation, polymeric nanoparticles, and aptamer-decorated nanocarriers can be employed. These carriers protect the RNA from enzymatic degradation, prolong circulation time, and facilitate receptor-mediated transcytosis across the BBB. Incorporating PEGylation or targeted ligands (e.g., transferrin or insulin receptor peptides) further improves stability and brain-specific delivery, thereby maximizing the therapeutic potential of RNA-based PROTACs for neurological disorders.

5. CONCLUSION

In our research, we successfully modeled a PROTAC-aptamer conjugate targeting the nucleolin protein, an oncogenic driver in

GBM, using various computational tools, including AlphaFold, P2Rank, ScanNet, and molecular docking with HDOCK. We identified predicted binding pockets on both the E3 ligase and nucleolin protein, constructed the aptamer's 3D structure from sequence data, and simulated both aptamer-nucleolin and E3 ligase-PROTAC binding, which demonstrated favorable surface complementarity and interaction potential. These findings support the feasibility of designing PROTAC systems that selectively degrade nucleolin by facilitating its proximity to the E3 ligase protein. While our approach is computational, further research can be done experimentally and therapeutically. Future research should focus on in vitro and in vivo testing of the designed conjugates to confirm nucleolin degradation and efficacy, while exploring advanced drug delivery strategies to overcome BBB limitations and improve the safety profile of these aptamer-based drug treatments for GBM.

6. REFERENCES

- Alifieris, C., and Trafalis, D.T.: 'Glioblastoma multiforme: Pathogenesis and treatment', *Pharmacology & Therapeutics*, 2015, 152, pp. 63-82
- Grochans, S., Cybulska, A.M., Simińska, D., Korbecki, J., Kojder, K., Chlubek, D., and Baranowska-Bosiacka, I.: 'Epidemiology of Glioblastoma Multiforme-Literature Review', *Cancers (Basel)*, 2022, 14, (10)
- Davis, M.E.: 'Glioblastoma: Overview of Disease and Treatment', *Clin J Oncol Nurs*, 2016, 20, (5 Suppl), pp. S2-8
- Tonello, F., Massimino, M.L., and Peggion, C.: 'Nucleolin: a cell portal for viruses, bacteria, and toxins', *Cellular and Molecular Life Sciences*, 2022, 79, (5), pp. 271
- Ginisty, H., Sicard, H., Roger, B., and Bouvet, P.: 'Structure and functions of nucleolin', *Journal of Cell Science*, 1999, 112, (6), pp. 761-772
- Pearson, J.R.D., and Regad, T.: 'Targeting cellular pathways in glioblastoma multiforme', *Signal Transduction and Targeted Therapy*, 2017, 2, (1), pp. 17040
- Gao, H., Sun, X., and Rao, Y.: 'PROTAC Technology: Opportunities and Challenges', *ACS Medicinal Chemistry Letters*, 2020, 11, (3), pp. 237-240

- 8 Consortium, T.U.: 'UniProt: a hub for protein information', *Nucleic Acids Research*, 2014, 43, (D1), pp. D204-D212
- 9 Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.L., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Židek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., and Jumper, J.M.: 'Accurate structure prediction of biomolecular interactions with AlphaFold 3', *Nature*, 2024, 630, (8016), pp. 493-500
- 10 Omasits, U., Ahrens, C.H., Müller, S., and Wollscheid, B.: 'Protter: interactive protein feature visualization and integration with experimental proteomic data', *Bioinformatics*, 2013, 30, (6), pp. 884-886
- 11 Pontén, F., Jirström, K., and Uhlen, M.: 'The Human Protein Atlas—a tool for pathology', *The Journal of Pathology*, 2008, 216, (4), pp. 387-393
- 12 Krivák, R., and Hoksza, D.: 'P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure', *Journal of Cheminformatics*, 2018, 10, (1), pp. 39
- 13 Tubiana, J., Schneidman-Duhovny, D., and Wolfson, H.J.: 'ScanNet: an interpretable geometric deep learning model for structure-based protein binding site prediction', *Nature Methods*, 2022, 19, (6), pp. 730-739
- 14 Reuter, J.S., and Mathews, D.H.: 'RNAstructure: software for RNA secondary structure prediction and analysis', *BMC Bioinformatics*, 2010, 11, (1), pp. 129
- 15 Watkins, A.M., and Das, R.: 'RNA 3D Modeling with FARFAR2, Online', in Kawaguchi, R.K., and Iwakiri, J. (Eds.): 'RNA Structure Prediction' (Springer US, 2023), pp. 233-249
- 16 Ozvoldik, K., Stockner, T., and Krieger, E.: 'YASARA Model-Interactive Molecular Modeling from Two Dimensions to Virtual Realities', *Journal of Chemical Information and Modeling*, 2023, 63, (20), pp. 6177-6182
- 17 Yan, Y., Zhang, D., Zhou, P., Li, B., and Huang, S.Y.: 'HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy', *Nucleic Acids Res*, 2017, 45, (W1), pp. W365-w373
- 18 Abbott, N.J., Patabendige, A.A.K., Dolman, D.E.M., Yusof, S.R., and Begley, D.J.: 'Structure and function of the blood-brain barrier', *Neurobiology of Disease*, 2010, 37, (1), pp. 13-25
- 19 Risau, W., and Wolburg, H.: 'Development of the blood-brain barrier', *Trends in Neurosciences*, 1990, 13, (5), pp. 174-178
- 20 Chen, X., Li, H., Tian, L., Li, Q., Luo, J., and Zhang, Y.: 'Analysis of the Physicochemical Properties of Acaricides Based on Lipinski's Rule of Five', *Journal of Computational Biology*, 2020, 27, (9), pp. 1397-1406
- 21 Wessel, M.D., Jurs, P.C., Tolan, J.W., and Muskal, S.M.: 'Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure', *Journal of Chemical Information and Computer Sciences*, 1998, 38, (4), pp. 726-735
- 22 Veber, D.F., Johnson, S.R., Cheng, H.-Y., Smith, B.R., Ward, K.W., and Kopple, K.D.: 'Molecular Properties That Influence the Oral Bioavailability of Drug Candidates', *Journal of Medicinal Chemistry*, 2002, 45, (12), pp. 2615-2623
- 23 Kah, M., and Brown, C.D.: 'LogD: Lipophilicity for ionisable compounds', *Chemosphere*, 2008, 72, (10), pp. 1401-1408