Identification of novel breast cancer-related biomarkers through an aptamer-based proteomic approach

Beatriz Santos[1][0000−0002−5056−1629], Débora Ferreira[2][0000−0002−0937−4047], and Sérgio F. Sousa[3][0000−0002−6560−5284]

[1] Informatics Department, University of Minho, 4710-057 Braga, Portugal
[2] Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal
[3] LAQV/REQUIMTE, BioSIM - Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal

1 Introduction

1.1 Understanding Triple-Negative Breast Cancer

Breast cancer (BC) is a prevalent form of cancer that remains a prevalent global health challenge, constituting the most frequently diagnosed malignancy among women worldwide. In 2022, there were 2.3 million women diagnosed with breast cancer globally, resulting in 670,000 deaths (source `www.who.int/news-room/fact-sheets/detail/breast-cancer`; accessed on 13 March 2024 [1]).

Triple-negative breast cancer (TNBC) is a rare BC subtype known for its aggressive and metastatic nature. It is characterized by elevated rates of relapse, resulting in a poorer prognosis and a low response to current therapies. TNBC is highly heterogeneous, distinguished by the absence of the receptors commonly found in other BCs, such as estrogen (ER) and progesterone (PR) receptors, as well as the lack of amplification/overexpression of the human epidermal growth factor receptor 2 (HER2) [11,20,29].

Standard therapeutic strategies for TNBC primarily rely on chemotherapy, including anthracyclines and taxanes. More recently, some therapies have been exploited such as epidermal growth factor receptor (EGFR) targeted agents, androgen receptor-targeted agents, anti-antigenic agents, and poly-ADP ribose polymerase (PARP) inhibitors [2,5,7,14]. However, despite initial responses to these therapies, recurrence and high molecular heterogeneity inherent to TNBC further complicate treatment. Therefore, there is a growing need for novel therapeutic approaches tailored to the unique molecular landscape of this disease [31].

1.2 Aptamers in Targeted Therapies: A Revolution in Precision Medicine for TNBC

The emergence of biological drugs, such as monoclonal antibodies, has revolutionized the field of medicine by offering innovative therapies for diseases that

are traditionally challenging to address with small-molecule medications. While these advancements have expanded treatment options, they have also resulted in increased healthcare expenses and potential impacts on patient well-being. Thus, the pursuit of targeted therapies in TNBC demands the identification of specific biomarkers and molecular targets implicated in disease progression and therapeutic response [9,15,25,30,31].

Aptamers have emerged as highly effective biorecognition elements. They are artificial short single-stranded DNA (ssDNA) or RNA molecules, typically ranging from 25 to 50 nucleotides (nt) in length. Compared to other ligands, such as antibodies, aptamers have numerous benefits, including being non-toxic, easily manufactured, and chemically modified. Aptamers also exhibit a quick tissue penetration rate and low immunogenicity. Furthermore, aptamers demonstrate high specificity, selectivity, and binding affinity for a wide variety of targets, including proteins, small molecules, viruses, bacteria, and live cells [13,16]. Additionally, their ability to adopt several three-dimensional (3D) configurations enhances their ability to bind to specific protein targets.

1.3 Enhancing Aptamer Discovery: Leveraging *In silico* Tools and High Throughput Sequencing for Targeted Oncology Applications

In silico tools have garnered increasing interest in aptamer research as they provide insights into the binding affinity between a specific aptamer and its target, at a fraction of the time and with reduced costs while maintaining satisfactory accuracy [6]. Within the field of oncology, several investigations have utilized computational techniques, including structure prediction, docking, and molecular dynamics (MD). Subsequent experimental studies have confirmed these computational analyses, further enhancing the credibility of their findings. Indeed, numerous studies underscore the significance of *in silico* tools in advancing the development and refinement of novel nucleic acid-based molecules [3,4,21,27].

Aptamers are derived from random initial oligonucleotide/peptide library and can be isolated by SELEX (Systematic Evolution of Ligands by EXponential enrichment), which involves repetitive rounds of partitioning and enrichment commonly performed with purified target proteins immobilized on a solid support [8,23,26]. One variant of SELEX, called cell-SELEX, enables the screening of cell-specific aptamers without requiring prior knowledge of the molecular signatures of the target cells [4,19,22,24,30].

The identification of new specific-target aptamers has posed challenges primarily due to difficulties associated with characterizing potential sequences from enriched libraries. Traditionally, the identification of enriched sequences relies on cloning and Sanger sequencing of the library resulting from multiple selection cycles [18,28]. Consequently, only sequenced clones are recognized as potential targeting ligands, thereby excluding candidate aptamers that perform well but have low copy numbers in the enriched library. To address this issue, specialized technologies have been integrated into the original SELEX process, such as

high-throughput sequencing (HTS) and bioinformatics analysis. HTS and bioinformatics, when combined with SELEX (referred to as HT-SELEX), facilitate the identification of numerous aptamer candidates, total reads, frequencies of each unique sequence, distribution of each nucleotide in the random sequence, and the rate of molecular enrichment [12]. Moreover, HT-SELEX enables the identification of aptamers with high affinity and specificity in earlier selection rounds, thereby potentially reducing over-selection, which is time-consuming, and avoiding potential PCR artifacts during the amplification steps [17].

In a previous study, novel single-stranded DNA (ssDNA) aptamer sequences (Apt1 and Apt2) that specifically recognize metastatic TNBC cell lines were identified through the combination of cell-SELEX with HTS [10]. An in-depth binding analysis not only showcased the high specificity and selectivity of the identified aptamers towards TNBC cell lines but also provided compelling evidence indicating a strong affinity for targeting specific cell surface receptors. This comprehensive examination under-scored the potential of Apt1 and Apt2 to precisely recognize and interact with TNBC cells, offering valuable insights into their therapeutic targeting capabilities at the molecular level [10]. Particularly, proteomic analysis was employed to identify the potential biomarker protein that Apt2 (Nomenclature mentioned in REF) binds to, and several relevant targets were found, including the following proteins: P09493, P13984, Q14141, Q9BTC0, Q9NSK0, Q8N8S7 (Primary protein accession IDs). Further studies need to be performed in order to assess their described information and to study how they can interact with Apt2 in the context of BC.

2 Objectives

The main goal of this project is to apply computational methodologies, such as structure prediction, aptamer-protein docking, and molecular dynamics, to corroborate and validate the experimental studies. To achieve this, molecular dynamic simulations of the docked complexes and free binding energy calculations by molecular mechanics-generalized born surface area (MM-GBSA) will be performed. These analyses will facilitate the identification of the potential biomarker(s) to which the mentioned aptamer binds. Ultimately, Apt2 could be a truly helpful tool for targeted delivery in TNBC tumors. The specific aims for this project include:

1. Use of bioinformatics tools and molecular modeling techniques to generate accurate 3D models of Apt2, ensuring its stability and functionality;
2. Investigating the molecular interactions between Apt2 and potential biomarker proteins expressed on the surface of TNBC cells by employing molecular docking simulations to predict the binding modes and affinities of Apt2-protein complexes;
3. Conducting molecular dynamics simulations to analyze the stability and dynamics of the docked complexes, providing insights into the strength and specificity of Apt2-protein interactions.

3 Research methodology

Task 1 - Generation of Apt2 aptamer secondary structure

The process of generating a 3D Apt2 DNA aptamer model from its nucleotide sequence involves three sequential steps first described by [13]. The workflow will be adapted and modified as outlined in Figure 1 and will facilitate the study of interactions with target molecules.

Fig. 1. Workflow employed to generate the three-dimensional structure of the DNA aptamer from its nucleotide sequence, which includes predicting their secondary structures and subsequently modeling them into three dimensions. UNAFold stands for Unified Nucleic Acid Folding and Hybridization Package. DNA - Deoxyribonucleic acid; 3D - Three dimensional; RNA – Ribonucleic acid. Adapted from Pereira et al, 2022 [21].

To start, the secondary structure of the DNA aptamer will be predicted based on its nucleotide sequence. The DNA sequence of Apt2 will be submitted to the 3dRNA/DNA web server, which is now integrated into the Xiao lab web server (`http://biophy.hust.edu.cn/new/3dRNA/create`). Utilizing free energy minimization algorithms, this web server predicts secondary structures of ssDNA, while considering a folding temperature of 37^{o}C, in 187 mM Na^{+} and 0.5 mM Mg^{2+} at pH 7.4. Only foldings within 5 % of the minimum free energy will be computed, with a maximum of 50 foldings and no restrictions on the maximum distance between paired bases. Subsequently, the resulting structure with the lowest free energy will be utilized in the subsequent steps. The secondary structure obtained from the server analysis will be employed to generate the initial 3D models.

Task 2 - Generation of three-dimensional models for Apt2

As there are no currently available tools for predicting 3D ssDNA structures, 3D RNA structures will be generated instead. Subsequently, we will download the resulting RNA structures in the protein data bank (PDB) file format to convert them back to DNA. To accomplish this, we will use an in-house bash script to remove the hydroxyl group at the C2 position of the ribose ring and the hydrogen atom at the C5 position of uracil. Next, we will make use of the LEaP program from the AmberTools software package to add all the missing atoms, including the methyl group required to convert uracil into thymine. The resulting structures will be suitable for the next steps of the computational pipeline required for this work.

Task 3 - Main conformations and molecular dynamics simulations of Apt2 aptamer

Firstly, in order to prepare the Apt2 aptamer for MD simulations, several steps must be undertaken, including solvating the aptamer with water molecules and adding a physiological concentration of NaCl in the system by the LeaP package of the AmberTools software. The number of Na^+ (to neutralize the DNA) and Cl^- ions will be added according to the method screening layer tally by container average potential (SLTCAP) method. Energy minimizations and MD simulations will be performed with bsc1, considering a force field developed for the atomistic simulation of nucleic acids. The AMBER18 software package, within the Sander program, will be used. The minimization of the system involves four steps: (1) water molecules; (2) hydrogen atoms; (3) all non-backbone atoms; and (4) all atoms. Following system equilibration using the NVT ensemble, there will be a production phase. Then, to constrain all bonds involving hydrogen atoms, it will be used the SHAKE38 algorithm. [13]

Task 4 - Prediction and generation of a three-dimensional model for protein structures.

Based on previous research conducted by our team, proteomics analysis has identified various targets of interest that may serve as potential therapeutic targets for TNBC. To predict the structure of these proteins, we will use the AlphaFold system, a state of the art deep learning-based method for protein structure prediction, developed by DeepMind. Input sequences for AlphaFold prediction will be prepared by extracting amino acid sequences of selected proteins from relevant databases. These sequences will be then submitted to the AlphaFold prediction server. Leveraging AlphaFold's advanced neural network architecture and extensive training data, we aim to predict the 3D structures of the selected proteins with high accuracy.

Task 5 - Aptamer-protein docking

To identify representative aptamer structures from the MD simulations suitable for molecular docking with the selected proteins, cluster analysis will be conducted. Using the K-means clustering method, clusters will be generated based on the root mean square deviation (RMSD) of all non-hydrogen atoms. The representative structure of each cluster will be selected for further analysis. Next, the complexes between proteins and Apt2 will be predicted using the HADDOCK 2.4 web server. Ambiguous interaction restraints (AIRs) will be randomly determined from residues. The ranking of the different complexes for each protein will be determined based on the HADDOCK score (HS), which is calculated as a linear weighted sum of energetic terms and buried surface area:

$$\mathrm{HS} = 1\mathrm{Evdw} + 0.2\mathrm{Ee1ec} + 1\mathrm{Edeso1} + 0.1\mathrm{EAIR} \quad (1)$$

Here, each term of equation 1 represents van der Waals (Evdw), Coulomb electrostatics (Eelec), desolvation (Edesol), and restraint energies (EAIR), respectively.

Task 6 - Molecular dynamics simulations of the docked aptamer-protein complexes

The MD simulations of the top-ranked docked aptamer-protein complexes will follow the same protocol as described earlier for the initial simulation of the Apt2 aptamer (Task 3). All parameters will remain consistent, with the exception of employing the ff14SB force field, specifically designed for protein simulations, in addition to the force field used for the aptamer. These analyses will facilitate the identification of the potential biomarker(s) to which Apt2 binds, and later be validated experimentally. Ultimately, this novel aptamer could be a helpful tool for targeted delivery in TNBC tumors.

4 Timeline

The project timeline is shown in Figure 2.

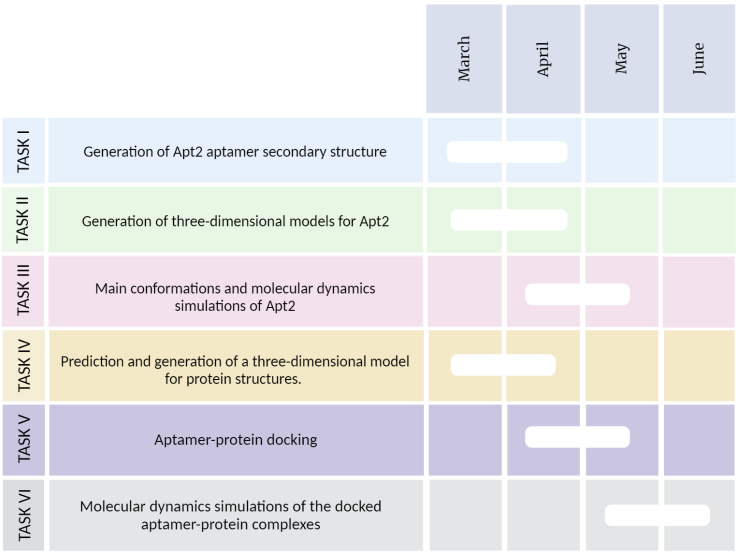

Fig. 2. Proposed timeline for the project. The time dedicated to each task considers the optimizations required for the different procedures.

References

1. Breast cancer fact sheet. https://www.who.int/news-room/fact-sheets/detail/breast-cancer, accessed: April 13, 2024
2. André, F., Zielinski, C.: Optimal strategies for the treatment of metastatic triple-negative breast cancer with currently approved agents. Annals of Oncology **23**, vi46–vi51 (8 2012). https://doi.org/10.1093/annonc/mds195
3. Bavi, R., Liu, Z., Han, Z., Zhang, H., Gu, Y.: In silico designed rna aptamer against epithelial cell adhesion molecule for cancer cell imaging. Biochemical and Biophysical Research Communications **509**, 937–942 (2 2019). https://doi.org/10.1016/j.bbrc.2019.01.028
4. Bell, D.R., Weber, J.K., Yin, W., Huynh, T., Duan, W., Zhou, R.: In silico design and validation of high-affinity rna aptamers targeting epithelial cellular adhesion molecule dimers. Proceedings of the National Academy of Sciences **117**, 8486–8493 (4 2020). https://doi.org/10.1073/pnas.1913242117
5. Beniey, M., Haque, T., Hassan, S.: Translating the role of parp inhibitors in triple-negative breast cancer. Oncoscience **6**, 287–288 (1 2019). https://doi.org/10.18632/oncoscience.474
6. Buglak, A.A., Samokhvalov, A.V., Zherdev, A.V., Dzantiev, B.B.: Methods and applications of in silico aptamer design and modeling. International Journal of Molecular Sciences **21**, 8420 (11 2020). https://doi.org/10.3390/ijms21228420
7. Cortés, J., André, F., Gonçalves, A., Kümmel, S., Martín, M., Schmid, P., Schuetz, F., Swain, S.M., Easton, V., Pollex, E., Deurloo, R., Dent, R.: Impassion132 phase iii trial: atezolizumab and chemotherapy in early relapsing metastatic triple-negative breast cancer. Future Oncology **15**, 1951–1961 (6 2019). https://doi.org/10.2217/fon-2019-0059
8. Ellington, A.D., Szostak, J.W.: In vitro selection of rna molecules that bind specific ligands. Nature **346**, 818–822 (8 1990). https://doi.org/10.1038/346818a0
9. Fattal, E., Hillaireau, H., Ismail, S.I.: Aptamers in therapeutics and drug delivery. Advanced Drug Delivery Reviews **134**, 1–2 (9 2018). https://doi.org/10.1016/j.addr.2018.11.001
10. Ferreira, D., Barbosa, J., Sousa, D.A., Silva, C., Melo, L.D.R., Avci-Adali, M., Wendel, H.P., Rodrigues, L.R.: Selection of aptamers against triple negative breast cancer cells using high throughput sequencing. Scientific Reports **11**, 8614 (4 2021). https://doi.org/10.1038/s41598-021-87998-y
11. Gucalp, A., Traina, T.A.: Triple-negative breast cancer: Adjuvant therapeutic options. Chemotherapy Research and Practice **2011**, 1–13 (6 2011). https://doi.org/10.1155/2011/696208
12. Hoinka, J., Berezhnoy, A., Dao, P., Sauna, Z.E., Gilboa, E., Przytycka, T.M.: Large scale analysis of the mutational landscape in ht-selex improves aptamer discovery. Nucleic Acids Research **43**, 5699–5707 (7 2015). https://doi.org/10.1093/nar/gkv308
13. Jeddi, I., Saiz, L.: Three-dimensional modeling of single stranded dna hairpins for aptamer-based biosensors. Scientific Reports **7**, 1178 (4 2017). https://doi.org/10.1038/s41598-017-01348-5
14. Keung, M.Y., Wu, Y., Badar, F., Vadgama, J.V.: Response of breast cancer cells to parp inhibitors is independent of brca status. Journal of Clinical Medicine **9**, 940 (3 2020). https://doi.org/10.3390/jcm9040940
15. Khot, V.M., Salunkhe, A.B., Pricl, S., Bauer, J., Thorat, N.D., Townley, H.: Nanomedicine-driven molecular targeting, drug delivery, and therapeutic ap-

proaches to cancer chemoresistance. Drug Discovery Today **26**, 724–739 (3 2021). https://doi.org/10.1016/j.drudis.2020.12.016
16. Ku, T.H., Zhang, T., Luo, H., Yen, T., Chen, P.W., Han, Y., Lo, Y.H.: Nucleic acid aptamers: An emerging tool for biotechnology and biomedical sensing. Sensors **15**, 16281–16313 (7 2015). https://doi.org/10.3390/s150716281
17. Levay, A., Brenneman, R., Hoinka, J., Sant, D., Cardone, M., Trinchieri, G., Przytycka, T.M., Berezhnoy, A.: Identifying high-affinity aptamer ligands with defined cross-reactivity using high-throughput guided systematic evolution of ligands by exponential enrichment. Nucleic Acids Research **43**, e82–e82 (7 2015). https://doi.org/10.1093/nar/gkv534
18. Li, X., Zhang, W., Liu, L., Zhu, Z., Ouyang, G., An, Y., Zhao, C., Yang, C.J.: In vitro selection of dna aptamers for metastatic breast cancer cell recognition and tissue imaging. Analytical Chemistry **86**, 6596–6603 (7 2014). https://doi.org/10.1021/ac501205q
19. Liu, M., Yang, T., Chen, Z., Wang, Z., He, N.: Differentiating breast cancer molecular subtypes using a dna aptamer selected against mcf-7 cells. Biomaterials Science **6**, 3152–3159 (2018). https://doi.org/10.1039/C8BM00787J
20. Mendes, T.F.S., Kluskens, L.D., Rodrigues, L.R.: Triple negative breast cancer: Nanosolutions for a big challenge. Advanced Science **2** (11 2015). https://doi.org/10.1002/advs.201500053
21. Pereira, A.C., Pina, A.F., Sousa, D., Ferreira, D., Santos-Pereira, C., Rodrigues, J.L., Melo, L.D.R., Sales, G., Sousa, S.F., Rodrigues, L.R.: Identification of novel aptamers targeting cathepsin b-overexpressing prostate cancer cells. Molecular Systems Design Engineering **7**, 637–650 (2022). https://doi.org/10.1039/D2ME00022A
22. Pleiko, K., Saulite, L., Parfejevs, V., Miculis, K., Vjaters, E., Riekstina, U.: Differential binding cell-selex method to identify cell-specific aptamers using high-throughput sequencing. Scientific Reports **9**, 8142 (5 2019). https://doi.org/10.1038/s41598-019-44654-w
23. Presnell, K.V., Alper, H.S.: Thermodynamic and first-principles biomolecular simulations applied to synthetic biology: promoter and aptamer designs. Molecular Systems Design Engineering **3**, 19–37 (2018). https://doi.org/10.1039/C7ME00083A
24. Sefah, K., Shangguan, D., Xiong, X., O'Donoghue, M.B., Tan, W.: Development of dna aptamers using cell-selex. Nature Protocols **5**, 1169–1185 (6 2010). https://doi.org/10.1038/nprot.2010.66
25. Stoltenburg, R., Reinemann, C., Strehlitz, B.: Selex—a (r)evolutionary method to generate high-affinity nucleic acid ligands. Biomolecular Engineering **24**, 381–403 (10 2007). https://doi.org/10.1016/j.bioeng.2007.06.001
26. Tuerk, C., Gold, L.: Systematic evolution of ligands by exponential enrichment: Rna ligands to bacteriophage t4 dna polymerase. Science **249**, 505–510 (8 1990). https://doi.org/10.1126/science.2200121
27. Wang, Y., Luo, J., Liu, J., Sun, S., Xiong, Y., Ma, Y., Yan, S., Yang, Y., Yin, H., Cai, X.: Label-free microfluidic paper-based electrochemical aptasensor for ultrasensitive and simultaneous multiplexed detection of cancer biomarkers. Biosensors and Bioelectronics **136**, 84–90 (7 2019). https://doi.org/10.1016/j.bios.2019.04.032
28. Wu, Q., Lin, N., Tian, T., Zhu, Z., Wu, L., Wang, H., Wang, D., Kang, D., Tian, R., Yang, C.: Evolution of nucleic acid aptamers capable of specifically targeting glioma stem cells via cell-selex. Analytical Chemistry **91**, 8070–8077 (7 2019). https://doi.org/10.1021/acs.analchem.8b05941

29. Zhang, J.F., Liu, J., Wang, Y., Zhang, B.: Novel therapeutic strategies for patients with triple-negative breast cancer. OncoTargets and Therapy **Volume 9**, 6519–6528 (10 2016). https://doi.org/10.2147/OTT.S105716
30. Zhong, Y., Zhao, J., Li, J., Liao, X., Chen, F.: Advances of aptamers screened by cell-selex in selection procedure, cancer diagnostics and therapeutics. Analytical Biochemistry **598**, 113620 (6 2020). https://doi.org/10.1016/j.ab.2020.113620
31. Zhou, J., Rossi, J.: Aptamers as targeted therapeutics: current potential and challenges. Nature Reviews Drug Discovery **16**, 181–202 (3 2017). https://doi.org/10.1038/nrd.2016.199