Journal of Associated Medical Sciences 2025; 58(3): 248-253

JAMS
Journal of Associated Medical Sciences

Scopus Indexed & Thai-Journal Citation Index Centre (TCI)

Journal of Associated Medical Sciences

Journal homepage: https://www.tci-thaijo.org/index.php/bulletinAMS/index

Deep neural network-based prediction of RNA aptamers targeting E6 protein of high-risk human papilloma virus

Bundit Promraksa[1]* Yingpinyapat Kittirat[1], Dujdao Boonyod[1], Chawisa Phetumpai[1], Malinee Thanee[2], Anchalee Techasen[3]

[1]Regional Medical Sciences Center 2 Phitsanulok, Department of Medical Sciences, Ministry of Public Health, Phitsanulok Province, Thailand.
[2]Department of Pathology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand.
[3]Centre for Research and Development of Medical Diagnostic Laboratories (CMDL), Faculty of Associated Medical Science, Khon Kaen University, Khon Kaen Province, Thailand.

ARTICLE INFO

Article history:
Received 30 March 2025
Accepted as revised 13 July 2025
Available online 15 July 2025

Keywords:
RNA aptamer, human papillomavirus, deep neural network, cervical cancer, E6 protein

ABSTRACT

Background: The Human papilloma virus is the primary cause of cervical cancer. The virus integrates with the human genome to produce the E6 oncoprotein. Therefore, the E6 oncoprotein is a crucial molecular target for cancer progression or treatment. The development of aptamers is beneficial for interacting with the target protein and serves as a new strategy for detection or delivery systems.

Objectives: We aim to explore the candidate aptamer sequence against E6 oncoprotein using a computational-based method.

Materials and methods: This study designed the candidate aptamer against the target protein based on computational approaches using the AptaTrans pipeline. After obtaining the candidate aptamer sequences, the minimum free energy was calculated using the RNAfold web server. The tertiary structure was then generated using RNAComposer. Next, the molecular docking score was acquired from the GRAMM web server.

Results: The aptamer sequences with the best stability, as indicated by minimum free energy (MFE), are Sq3_16E6, Sq3_Actn, and Sq3_18E6, respectively. The aptamer sequences of Sq3_16E6 and Sq2_18E6 showed potential interactions with 8GCR and 6SJV, respectively.

Conclusion: Sq3_16E6 and Sq2_18E6 are appropriate for the development of the detection of the E6 protein in cervical swabs. Further investigation should be performed.

Introduction

High risk human papilloma virus (HPV) is non enveloped DNA virus, can cause the abnormalities of cervical epithelial cells, which primarily cause of cervical cancer. The individuals with risky behaviors, such as frequently changing sexual partners, engaging in sexual activity at a young age, and direct contact with lesions of an infected person, are at significant risk of developing cervical cancer [1]. Currently, HPV DNA testing is promoted with regular screening of cervical cancer in Thailand. This policy is recommended in response to these second most common cancers so that the detection of HPV DNA collected by self-cervical swabs have been promoted to prevent the development of these cancers in early stages[2]. Furthermore, the vaccination program in age under 12-13 years old significantly reduced caused of cervical cancer[3]. Regional Medical Sciences Center 2 Phitsanulok offers cervical cancer screening services using the HPV DNA test, which detects 14 high-risk

* **Corresponding contributor.**
Author's Address: *Regional Medical Sciences Center 2 Phitsanulok, Department of Medical Sciences, Ministry of Public Health, Phitsanulok Province, Thailand.*
E-mail address: bundit.p@dmsc.mail.go.th
doi: 10.12982/JAMS.2025.094
E-ISSN: 2539-6056

strains in its responsible area within Health Region 2. If patients test positive for HPV types 16 and 18, confirmation with colposcopy is recommended. For other strains, likely types 52 and 58, patients are taken liquid-based cytology to diagnose the progression of cervical cancer[4]. Notably,HPV DNA encodes several oncoproteins, including the oncogenic E6 and E7 proteins, which promote cervical carcinogenesis and are overexpressed during cervical transformation[5]. Therefore, an oncoprotein-based diagnostic test holds promise for being especially specific in detecting precancerous lesions that have progressed to a high-grade CIN stage or to cancer. The E6 protein binds to a short LxxLL consensus sequence within the cellular ubiquitin ligase E6-associated protein (E6AP), subsequently leading to degrad pro-apoptotic tumour suppressor p53. The meta-analysis suggested that the detection of E6 oncoprotein may be useful for triaging HPV-positive women by predicting the risk of developing cervical pre-cancer and cancer[6]. In consistency with the study by Ferrera et al. (2019), the detection of E6 oncoprotein is highly sensitive and serves as a specific marker for HPV16/18-related High-grade Squamous Intraepithelial Lesion (HSIL) lesions. Moreover, the development of lateral flow assays for E6 oncoprotein is useful for direct triage to treatment in resource-limited settings[7]. In summary, detecting HPV DNA, along with E6 and E7 proteins, is beneficial for identifying high-risk HPV infections. This approach is particularly important for assessing cervical cancer progression before the patient undergoes colposcopy by an obstetrician. Moreover, E6 is also the molecular target for treatment of cervical cancer[8]. Aptamers are the short nucleotides of DNA and RNA molecules which can bind to specific targets such as ions, small molecules or specific proteins. Aptamers are generally developed by the conventional method known as systematic evolution of ligands by exponential enrichment (SELEX)[9]. The limitation of SELEX is labor- and time- consuming. Moreover, this method rarely yields the number of effective candidate aptamers for further performance evaluation and validation. The use of a computational based method is alternative precision to predict the candidate aptamer sequence at the monomer level. AptaTrans, a deep neural network (DNN) model, was developed to utilize Monte-Carlo tree search (Apta-MCTS) for the exploration of the recommending RNA aptamer candidates[10,11]. This pipeline pretrain structural representation pretrained encoders to generates the deep neural network model. Then, the binding capabilities of aptamer and protein targets can investigate by the molecular docking tools. The predicted 3D model results explore different poses of the aptamer-protein interaction and identify the complexes with the lowest binding energies.

In this study, we aim to investigate the candidate aptamer sequence using a computational-based method. Following the AptaTrans pipeline, we predict candidate RNA aptamers that interact with the protein. We are initially exploring these RNA aptamers using computational methods for HPV detection (Figure 1).

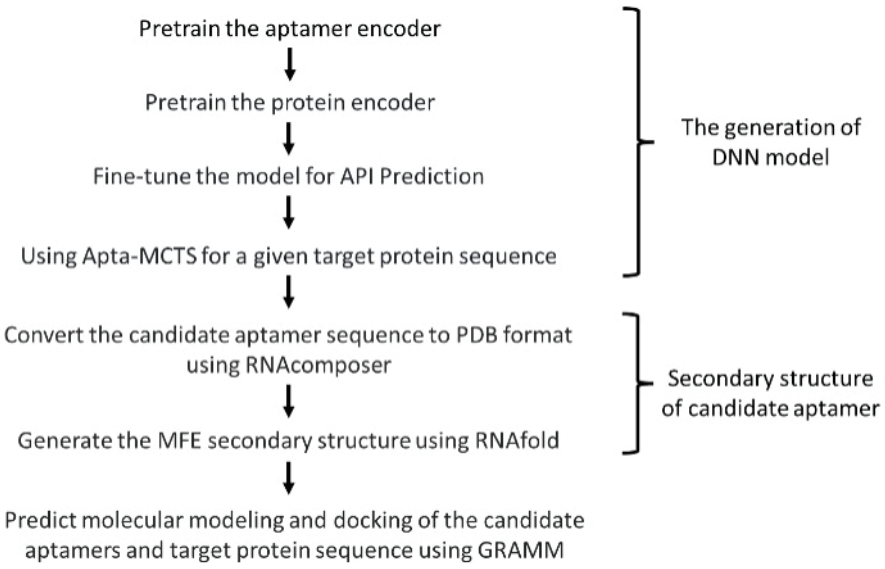


Figure 1. The workflow of this study to generate the candidate aptamer against E6 oncoprotein

Materials and methods

Generating a deep neural network model

The deep neural network model was executed followed by AptaTrans pipeline https://github.com/PNUMLB/AptaTrans) in the Windows subsystem for Linux environment. The core of the model is based on transformer-based encoders, which effectively capture the complex interactions between RNA aptamers and their target proteins. To generate neural network model, we used a pre-trained encoders of aptamers[12] and proteins[13]. A batch size of 16 was used, and the model was trained for 20 epochs. The neural network model was then fine-tuned for to further enhance performance.

Predicting candidate RNA aptamer using computational methods in deep neural networks

To obtain the RNA aptamer at the monomer level, the candidate aptamer was encoded by specifying the targeting

protein sequence and the depth of the Monte-Carlo search tree. The E6 oncoprotein sequence of HPV type 16 and 18 was retrieved from Uniprot (UniProt IDs: P06463 for HPV16 and P03255 for HPV18). 8GCR and 6SJV response to the HPV16 E6-E6AP-p53 complex and E6AP-LXXLL motifs of HPV type 16 and 18, respectively. Whereas, 3BYH represent actin. The number of iterations for the MCTS was set to 50, ensuring a thorough search for potential aptamer candidates.

Table 1. Candidate aptamer sequence

Candidate aptamer			MFE (-kcal/mol)
HPV16 (8GCR)	Sq1_16E6	GUUUAGCGAAUGCCCUUCAGUCUCUAACAAGAUGA	-1.90
	Sq2_16E6	CGCGGACCCCCUACAUUCCGCCGGGAUAUUACUAAAAGCGCUUUAAUCGU	-7.10
	Sq3_16E6	GAGGCGCAAGGCCGAACUGUAGAUUUAUAGGGGUGAACCAAGGACAUGC CGCGCGACUCC	-14.70
HPV18 (6SJV)	Sq1_18E6	GGUGAGCGAGCCCAUAGGUGGCUUACAGAGUUUUG	-10.70
	Sq2_18E6	GCGAAUUCGCCUUGGCACGAGAUCCGUAGGCAGGAGACGAAUUCGCGAUU	-19.10
	Sq3_18E6	CGCCGCGAGGUUAGCCUUACCGCACCCCCCGUCGACGAGAGGACGGGGCGG UCUCACGAA	-23.40
Actin (3BYH)	Sq1_Actn	AGAACACAUCCUUUAGUGCCAGUCCGACUUCUCGUUAGGUUUGACUGGGU	-14.50
	Sq2_Actn	AACCGCGCGGCCAUGUAUGCACGGAGUGUAGCCCUACUGUAGUCAAACUGA AACCGCCGC	-11.70
	Sq3_Actn	UAGACGCAGUUCAUACGAGCAGUUCGAUGUAUACUGGAACCCAAGUGAGU GGUGCAGAUUGUACAGCAAAUGGCUCGCCGCCUCCACGUCCGAAUCAAAG	-20.90

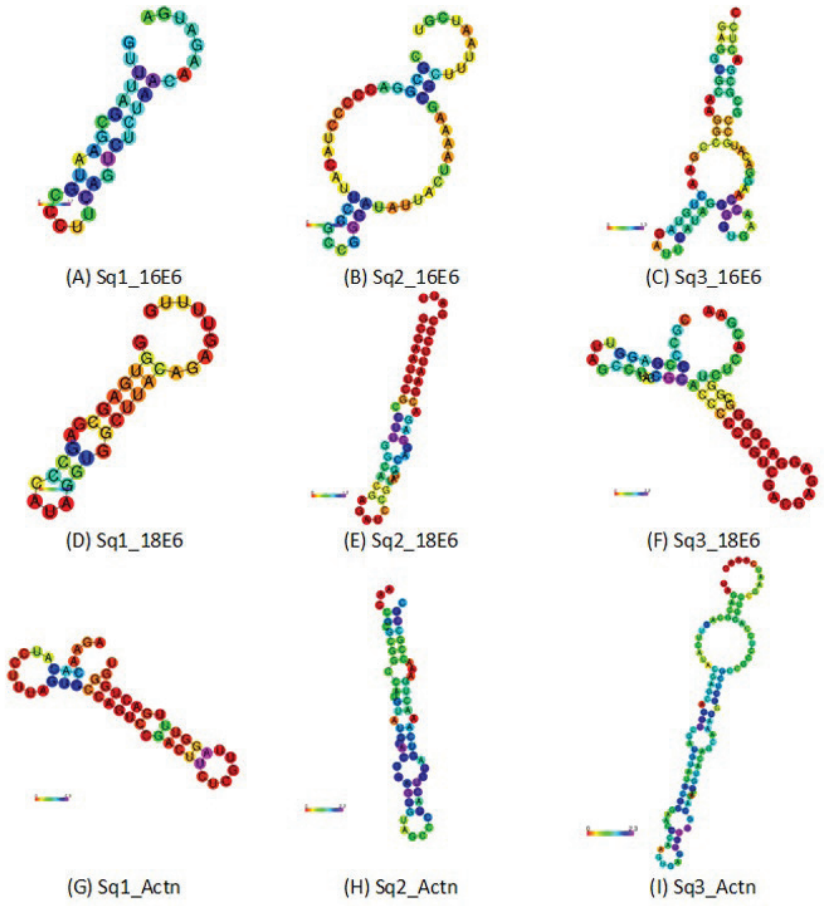


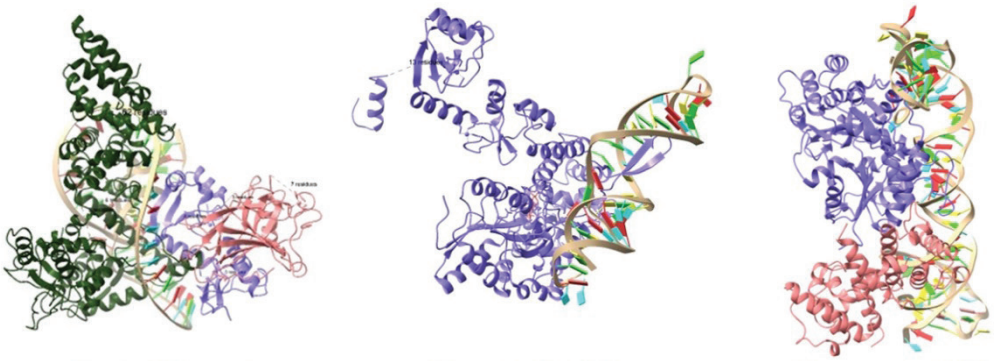
Figure 2. MFE secondary structure of the candidate aptamer sequence by RNAfold. Nucleotides are colored according to their positional entropy, as shown on the horizontal bar. Red colors indicate lower entropy, while blue colors indicate higher entropy

Molecular modeling and docking of the candidate aptamers

RNA aptamer sequences were converted to PDB format using RNAComposer web servers, to be used as ligands in subsequent analyses. The minimum free energy (MFE) secondary structure was obtained from RNAfold web server. The PDB format of target proteins, including 8GCR, 6SJV, and 3BYH, was downloaded from the Protein Data Bank (PDB). To predict the molecular interaction poses, RNA aptamer-protein dockings were performed using the GRAMM web server, which did not specify a binding site. The clustering threshold was set to 5 Å. Then, the 3D model structures were generated using the ChimeraX program (UCSF, USA).

Table 2. GRAMM's docking scores of candidate aptamers to target protein

HPV types	Candidate Aptamer sequences	Docking scores		
		8GCR	6SJV	3BYH
HPV16	Sq1_16E6	-630	-523	-634
	Sq2_16E6	-686	-654	-649
	Sq3_16E6	-802	-719	-712
HPV18	Sq1_18E6	-558	-613	-573
	Sq2_18E6	-661	-682	-659
	Sq3_18E6	-670	-670	-635
Actin	Sq1_Actn	-616	-640	-726
	Sq2_Actn	-670	-802	-818
	Sq3_Actn	-709	-664	-780



(A) Sq3_16E6-8GCR (B) Sq2_18E6-6SJV (C) Sq3_Actn-3BYH

Figure 3. The representative structure of the interaction illustrates the binding interfaces between the aptamer and target protein

Results

Candidate aptamers generation from DNN model

In this study, we performed deep neural networks to predict the sequences of aptamers targeting the proteins 8GCR, 6SJV, and 3BYH, which represent the E6 protein of HPV types 16 and 18, and actin, respectively. The candidate aptamer sequences are shown in Table 1. Based on the MFE results, the folding stability of each aptamer depends on sequence length. Additionally, the use of actin has been iterated as an internal control in the study. For this purpose, the RNA aptamer sequence most specific for this protein is Sq3_Actn, as it shows the highest binding affinity, indicated by the lowest docking scores and demonstrates lack of cross-reactivity with other proteins studied. The secondary structure of aptamer was shown in Figure 2.

The molecular docking of candidate aptamer to protein

The binding affinities of the candidate aptamers across all three proteins were evaluated using GRAMM web server (Table 2). Between candidate RNA aptamer against 8GCR protein, Sq3_16E6 exhibited the highest binding affinity based on its docking score. The docking scores of Sq2_18E6 and Sq1_18E6 against 6SJV protein were -682 and -613, respectively. Sq3_Actn showed a specific interaction with 3BYH protein, and demonstrated lower cross reactivity with other studied proteins. The interaction between representative candidate aptamer and the target protein has binding interfaces as shown Figure 3.

Discussion

HPV infection is a major risk factor for cervical cancer. An early region of HPV E6 and E7 genes integrate into the human genome, encode the oncoproteins playing a significant role during cervical cancer progression. The detection of the E6 oncoprotein serves as a specific marker for HPV-induced cancer progression. However, the development of detection methods may be problematic due to issues with the storage of cervical swab samples and the freeze-thaw cycles. Aptamers not only improve detection methods but also play a crucial role in the

development of next-generation drugs[14,15]. The process of obtaining specific recognition sequences, such as antibodies or high-throughput SELEX (HT-SELEX) aptamers, is still time consuming. Therefore, computational approaches are an alternative choice for improving the successful identification of aptamer sequence.

We obtained the aptamer sequence after inputting the target protein sequence into a DNN model, which predicted the candidate aptamer sequence based on Monte-Carlo Tree Search (MCTS)[11]. The general size of aptamer is between 30-100 nucleotides, so the length of candidate aptamer set up in these intervals. The selection and production of aptamers have been mentioned in which be challenges. Muhammad et al., 2022 have been generate the novel aptamer against NT-3 growth factor receptor based on computational approach with a relatively stable structure[16]. Moreover, the use of RNA aptamer, which was designed in silico, also demonstrated the ability to adhere on the surface of MCF-7 and MDA-MB-231 cells, which have been beneficial for cancer cell imaging[17]. In our study, Sq3_16E6 and Sq3_Actn expressed the low MFE, which relate to their stable positional entropy in the secondary structure. The tertiary structure was obtained from RNAcomposer webserver. This is the first study to apply the DNN predictive abilities using Aptatrans for aptamer sequence generation on cervical oncoprotein.

The molecular docking has been used as bioinformatics tool to predict the RNA aptamer-protein interactions[18]. Recently, the development of apta-sensors performed the molecular docking to predict the interaction between calcium/calmodulin-dependent serine protein kinase (CASK) protein, which propose to be breast cancer screening methods, against candidate aptamers[19]. Therefore, the molecular docking is the powerful tools to identify how the aptamer interacts with the target before further verification with other molecular techniques. In this study, we use GRAMM web server for protein docking to predict docking poses based on the Fourier transformation of the possible interaction sites of macromolecules and proteins[20]. The binding interfaces of the Sq3_16E6 groove can interact with the 8GRC protein, which represents a high docking score compared to other studied proteins. This interaction of Aptamers against proteins involves hydrogen bonding, electrostatic interactions, hydrophobic interactions and van der Waals forces[21]. However, the low binding affinity of RNA aptamers for HPV18E6 might result from poor shape complementarity, which is due to rigid and geometric-based scoring. RNA aptamers are highly flexible, and this limitation can cause inaccurate or low docking scores. Our result noticed that candidate aptamers, which expressed low MFE, had high potential interact with E6 oncoprotein.

Limitation

This study relies on computational approaches using the existing capabilities of the DNN model, which may not always accurately predict the complex structures and behaviors of aptamers. Next, the experimental validation is necessary to confirm these predictions.

Conclusion

The computational approach is a powerful tool for generating candidate aptamer. The molecular docking revealed the potential on the candidate aptamer and protein interactions. This study suggests that Sq3_16E6 and Sq2_18E6 are appropriate for further evaluation of their ability on E6 HPV type 16 and 18 oncoproteins, respectively.

Ethical approval

Not applicable

Funding

This work received no specific grant from any funding agency.

Conflict of interest

The authors declare no conflict of interest.

CReDIT authorship contribution statement

Bundit Promraksa: conceptualization, methodology, visualization, investigation, data curation, validation, writing original draft preparation; **Yingpinyapat Kittirat:** conceptualization, data curation; **Dujdao Boonyod:** conceptualization, data curation, writing-reviewing and editing; **Chawisa Phetumpai:** investigation, data curation; **Malinee Thanee:** data curation, writing-reviewing and editing; **Anchalee Techasen:** data curation, writing-reviewing and editing.

Acknowledgements

I would also like to thank Regional Medical Sciences Center 2 Phitsanulok for providing the resources and facilities needed to conduct my research. I would like to express my deepest gratitude to Ms. Amornrat Tassanakit, Director of Regional Medical Sciences Center 2 Phitsanulok for support and encouragement.

References

[1] Woodman CBJ, Collins SI, Young LS. The natural history of cervical HPV infection: unresolved issues. Nat Rev Cancer. 2007; 7(1): 11-22. doi: 10.1038/nrc2050.

[2] Ploysawang P, Rojanamatin J, Prapakorn S, Jamsri P, Pangmuang P, Seeda K, et al. National cervical cancer screening in Thailand. Asian Pac J Cancer Prev. 2021; 22(1): 25-30. doi: 10.31557/APJCP.2021.22.1.25.

[3] Falcaro M, Castañon A, Ndlela B, Checchi M, Soldan K, Lopez-Bernal J, et al. The effects of the national HPV vaccination programme in England, UK, on cervical cancer and grade 3 cervical intraepithelial neoplasia incidence: a register-based observational study. The Lancet. 2021; 398(10316): 2084-92. doi: 10.1016/S0140-6736(21)02178-4.

[4] Huh WK, Ault KA, Chelmow D, Davey DD, Goulart RA, Garcia FAR, et al. Use of primary high-risk human papillomavirus testing for cervical cancer screening: Interim clinical guidance. Obstet Gynecol. 2015; 125(2): 330-7. doi: 10.1097/AOG.0000000000000669.

[5] Derbie A, Mekonnen D, Woldeamanuel Y, Van Ostade X, Abebe T. HPV E6/E7 mRNA test for the detection of

high grade cervical intraepithelial neoplasia (CIN2+): a systematic review. Infect Agent Cancer. 2020; 15(1): 9. doi: 10.1186/s13027-020-0278-x.

[6] Downham L, Jaafar I, Rol ML, Nyawira Nyaga V, Valls J, Baena A, et al. Accuracy of HPV E6/E7 oncoprotein tests to detect high-grade cervical lesions: a systematic literature review and meta-analysis. Br J Cancer. 2024; 130(4): 517-25. doi: 10.1038/s41416-023-02490-w.

[7] Schweizer J, Lu PS, Mahoney CW, Berard-Bergery M, Ho M, Ramasamy V, et al. Feasibility Study of a Human Papillomavirus E6 Oncoprotein Test for Diagnosis of Cervical Precancer and Cancer. J Clin Microbiol. 2010; 48(12): 4646-8. doi: 10.1128/JCM.01315-10.

[8] Tan S, G.E. De Vries E, G.J. Van Der Zee A, De Jong S. Anticancer drugs aimed at E6 and E7 activity in HPV-positive cervical cancer. Curr Cancer Drug Targets. 2012; 12(2): 170-84. doi: 10.2174/156800912799095135.

[9] Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. Nature. 1990; 346(6287): 818-22. doi: 10.1038/346818a0.

[10] Shin I, Kang K, Kim J, Sel S, Choi J, Lee JW, et al. AptaTrans: a deep neural network for predicting aptamer-protein interaction using pretrained encoders. BMC Bioinformatics. 2023; 24(1): 447. doi: 10.1186/s12859-023-05577-6.

[11] Lee G, Jang GH, Kang HY, Song G. Predicting aptamer sequences that interact with target proteins using an aptamer-protein interaction classifier and a Monte Carlo tree search approach. Nebel JC, editor. PLOS ONE. 2021; 16(6): e0253760. doi: 10.1371/journal.pone.0253760.

[12] Danaee P, Rouches M, Wiley M, Deng D, Huang L, Hendrix D. bpRNA: large-scale automated annotation and analysis of RNA secondary structure. Nucleic Acids Res. 2018; 46(11): 5381-94. doi: 10.1093/nar/gky285.

[13] Berman HM. The protein data bank. Nucleic Acids Res. 2000; 28(1): 235-42. doi: 10.1093/nar/28.1.235.

[14] Khabbazian M, Jabbari H. AI-powered aptamer generation. Nat Comput Sci. 2022; 2(6): 356-7. doi: 10.1038/s43588-022-00253-w.

[15] Keefe AD, Pai S, Ellington A. Aptamers as therapeutics. Nat Rev Drug Discov. 2010; 9(7): 537-50. doi: 10.1038/nrd3141.

[16] Muhammad AM, Zari A, Alsubhi NH, Al-Zahrani MH, Alghamdi RA, Labib MM. Novel design of RNA aptamers as cancer inhibitors and diagnosis targeting the tyrosine kinase domain of the NT-3 growth factor receptor using a computational sequence-based approach. Molecules. 2022; 27(14): 4518. doi: 10.3390/molecules27144518.

[17] Bavi R, Liu Z, Han Z, Zhang H, Gu Y. In silico designed RNA aptamer against epithelial cell adhesion molecule for cancer cell imaging. Biochem Biophys Res Commun. 2019; 509(4): 937-42. doi: 10.1016/j.bbrc.2019.01.028.

[18] Navien TN, Thevendran R, Hamdani HY, Tang TH, Citartan M. In silico molecular docking in DNA aptamer development. Biochimie. 2021; 180:5 4-67. doi: 10.1016/j.biochi.2020.10.005.

[19] De Araújo NS, Moreira ADS, Abreu RDS, Junior VV, Antunes D, Mendonça JB, et al. Aptamer-based recognition of breast tumor cells: A new rra for breast cancer diagnosis. Int J Mol Sci. 2024; 25(2): 840. doi: 10.3390/ijms25020840.

[20] Singh A, Copeland MM, Kundrotas PJ, Vakser IA. GRAMM web server for protein docking. In: Gore M, Jagtap UB, Editors. Computational rug Discovery and Design (Methods in Molecular Biology; Vol. 2714). New York, NY: Springer US, pp 101-12. doi: 10.1007/978-1-0716-3441-7_5.

[21] Ji C, Wei J, Zhang L, Hou X, Tan J, Yuan Q, et al. Aptamer–protein interactions: From regulation to biomolecular detection. Chem Rev. 2023; 123(22): 12471-506. doi: 10.1021/acs.chemrev.3c00377.