



**STRUCTURAL AND AFFINITY ANALYSIS OF SELEX-  
GENERATED APTAMERS AND OPTIMIZATION  
THROUGH *IN-SILICO* MUTAGENESIS**

**YAĞMUR YEŞİLYURT**

Thesis for the Master's Program in Bioengineering

Graduate School  
Izmir University of Economics

Izmir

2024

**STRUCTURAL AND AFFINITY ANALYSIS OF SELEX-  
GENERATED APTAMERS AND OPTIMIZATION  
THROUGH *IN-SILICO* MUTAGENESIS**

**YAĞMUR YEŞİLYURT**

THESIS ADVISOR: ASSOC. PROF. DR. OSMAN DOLUCA

A Master's Thesis  
Submitted to  
the Graduate School of Izmir University of Economics  
the Department of Bioengineering

Izmir

2024

## ETHICAL DECLARATION

I hereby declare that I am the sole author of this thesis and that I have conducted my work in accordance with academic rules and ethical behavior at every stage from the planning of the thesis to its defense. I confirm that I have cited all ideas, information and findings that are not specific to my study, as required by the code of ethical behavior, and that all statements not cited are my own.

Name, Surname:

Yağmur YEŞİLYURT

Date:

18.01.2024

Signature:

# ABSTRACT

## STRUCTURAL AND AFFINITY ANALYSIS OF SELEX-GENERATED APTAMERS AND OPTIMIZATION THROUGH *IN-SILICO* MUTAGENESIS

YEŞİLYURT, Yağmur

Master's Program in Bioengineering

Advisor: Assoc. Prof. Dr. Osman DOLUCA

January, 2024

The aim of this thesis is to perform a comprehensive structural analysis of aptamers produced by the SELEX method and design a workflow to generate an aptamer modified to increase its affinity. Firstly, the physical structures of aptamers were analyzed in detail by obtaining 2D and 3D structures from an aptamer sequence with bioinformatics software. Using molecular docking and dynamics, the interaction of the aptamer with the target molecule was analyzed. Subsequently, it was investigated whether the affinity of the target molecule with the existing aptamer could be improved through *in-silico* mutagenesis. Molecular interactions were analyzed to improve the performance of the aptamer against the target hormones, and mutations that could increase the binding affinity were proposed. In the study, the 2D and 3D structures of the alsager22 aptamer in the literature and the binding site with estradiol were obtained. Affinity analysis revealed unexpected interactions with non-target hormones such as progesterone, testosterone, and androstenedione, although it showed the best affinity with estradiol. Further analysis with the selected mutant alsager22 showed that

the mutation alters the overall structure of the aptamer and the hormone binding site. Accordingly, it cannot be generally concluded that better affinity would be achieved by mutation, as it is not a reliable and reproducible process. Nevertheless, the highest affinity was observed in the MD simulation with the selected mutant aptamer and estradiol. These findings suggest that the possibility of obtaining better affinity by mutation exists but requires extensive processing power as there are too many mutation possibilities.

Keywords: SELEX-Generated Aptamers, Structural Analysis, Affinity Analysis, *In-Silico* Mutagenesis, Aptamer Optimization.



# ÖZET

## SELEX İLE ÜRETİLEN APTAMERLERİN YAPISAL VE AFİNİTE ANALİZİ VE *İN-SİLİKO* MUTAJEN YOLUYLA OPTİMİZASYONU

YEŞİLYURT, Yağmur

Biyomühendislik Yüksek Lisans Programı

Tez Danışmanı: Doç. Dr. Osman DOLUCA

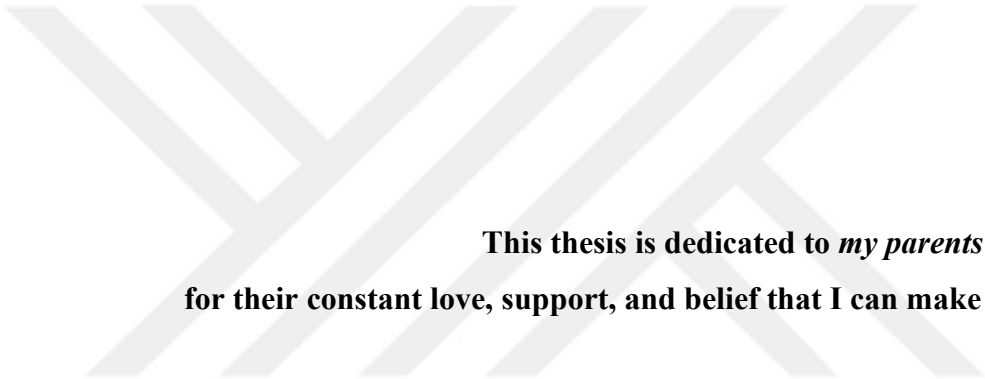
Ocak, 2024

Bu tezin amacı, SELEX metodu ile üretilen aptamerlerin kapsamlı bir yapısal analizini gerçekleştirmek ve bu analize dayanarak, afinitesini artırmak için modifiye edilmiş bir aptamer oluşturmak amacıyla bir iş akışı tasarlamaktır. İlk olarak, biyoformatik yazılımları ile bir aptamer dizisinden 2D ve 3D yapılar elde edilerek, aptamerlerin fiziksel yapıları detaylıca incelenmiştir. Moleküler yerleştirme yöntemi ve moleküler dinamik (MD) simülasyonları kullanılarak, aptamerin hedef moleküle etkileşimi analiz edilmiştir. Devamında, *in-silico* mutajen yoluyla elde edilen aptamerle hedef molekülün afinitesinin iyileştirilip iyileştirilemeyeceği araştırılmıştır. Hedef hormonuna karşı, aptamerin performansının iyileştirilmesi için moleküler etkileşimler analiz edilmiş ve bağlanma afinitesini artırabilecek mutasyonlar önerilmiştir. Çalışmada, literatürdeki alsager22 aptamerinin 2D ve 3D yapıları ve hedef hormonu olan östradiol ile bağlama bölgesi elde edilmiştir. Afinite analizi sonucunda, aptamerin östradiolle en iyi afinite göstermesine rağmen, hedef dışı hormonlar olan progesteron, testosteron, androstenedionla beklenmedik etkileşimleri saptanmıştır. 63 potansiyel

mutant aptamer arasından seçilen mutant alsager22 ile yapılan ileri analizler, mutasyonun aptamerin genel yapısını ve hormon bağlanma bölgesini değiştirdiğini göstermiştir. Bu doğrultuda, güvenilir ve tekrarlanabilir bir süreç olmadığı için mutasyon ile daha iyi afinite elde edileceği genel kanısına varılamamıştır. Yine de  $\Delta G$  değerleri karşılaştırıldığında çalışma boyunca en yüksek afinite, seçilen mutant aptamer ve östradiol ile gerçekleştirilen MD simülasyonunda gözlemlenmiştir. Bu bulgular, mutasyonla daha iyi afinite elde etme olasılığının varlığını, ancak çok fazla mutasyon olasılığı olduğu için kapsamlı bir işlem gücü gerektiğini ortaya koymaktadır.

Anahtar Kelimeler: SELEX ile Üretilen Aptamerler, Yapısal Analiz, Afinite Analizi, *In-Silico* Mutagenez, Aptamer Optimizasyonu.





**This thesis is dedicated to *my parents & my love,*  
for their constant love, support, and belief that I can make it possible.**



## ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Assoc. Prof. Dr. Osman DOLUCA, for their expert guidance, support, patience, and for providing me with the opportunity to undertake this project. His insights and advice have been fundamental in shaping this thesis and my master's degree journey.

Special thanks to my beloved parents, Canan and Burhanettin YEŞİLYURT, for their unconditional love, endless encouragement, and unwavering belief in my abilities. Their sacrifices and support have been the pillars of my strength and motivation.

I am immensely grateful to my love, Yağız TANŞU, for their understanding, love, and constant encouragement. His presence and support have been a source of comfort and inspiration during the most challenging phases of my academic pursuit. Whenever I felt like giving up, he went above and beyond to motivate me and keep me on track. He made efforts to comprehend aspects of my studies that were outside his area of expertise, simply to provide me with the support I needed. No matter what I say, it won't fully capture the extent of my gratitude and appreciation.

I would also like to extend my sincere appreciation to my colleague and mentor, Res. Asst. Hüseyin Saygın PORTAKAL, with whom I shared the journey of the TÜBİTAK 1001 project. His collaboration, knowledge, and dedication have significantly contributed to my personal academic development.

Additionally, I wish to extend my appreciation to the Scientific and Technological Research Council of Turkey (TÜBİTAK) for their financial support and resources (project no: 121Z972 & 2165854) that were crucial for the completion of this research.

This journey would not have been possible without the collective support of all those mentioned above. I am eternally grateful for their contributions to my academic and personal growth.

## PREFACE

This thesis is a journey of discovery in the discipline of "Structural Bioinformatics." It primarily delves into the structural examination of aptamers, crafted via the SELEX method, and their optimization using *in-silico* mutagenesis. At its core, structural bioinformatics serves as an insightful lens, granting us a clearer understanding of biological structures and their intricate interactions. My fascination with this field stems from my interest in computational biology and software.

Throughout my thesis journey, I have employed a suite of sophisticated bioinformatics tools and molecular modeling techniques. These were instrumental in revealing the interactions between aptamers and their target molecules. The hurdles encountered along the way were not just obstacles but valuable lessons in persistence and creative problem-solving. Each phase of my research not only enriched my understanding of the field but also honed my research skills.

IZMIR

18/01/2024

Yağmur YEŞİLYURT

## TABLE OF CONTENTS

ABSTRACT.....	iv
ÖZET.....	vi
ACKNOWLEDGEMENTS .....	ix
PREFACE .....	x
TABLE OF CONTENTS.....	xi
LIST OF TABLES .....	xiii
LIST OF FIGURES .....	xiv
LIST OF ABBREVIATIONS .....	xvii
CHAPTER 1: INTRODUCTION .....	1
1.1. Aptamers generated with SELEX .....	1
1.1.2. Aptamers .....	1
1.1.3. SELEX method .....	2
1.2. Secondary structure prediction.....	3
1.2.2. UNAFOLD.....	4
1.3. Tertiary structure prediction.....	4
1.3.2. XIAOLAB.....	5
1.4. Interaction analysis.....	6
1.4.1. Molecular docking .....	6
1.4.1.1. AUTODOCK tools & Vina .....	7
1.4.2. Molecular dynamics simulations .....	7
1.4.2.1. Force fields .....	8
1.4.2.2. GROMACS.....	9
1.4.3. Receptor-ligand interaction analysis .....	10
1.5. <i>In-silico</i> mutagenesis .....	11
1.6. Purpose and importance of the thesis .....	12
CHAPTER 2: METHODOLOGY .....	13
2.1. Obtaining 2D aptamer structures.....	14
2.2. Obtaining 3D aptamer structures.....	14
2.3. Molecular docking with the aptamer and target molecules.....	15
2.4. Molecular dynamics simulations.....	17

2.5.	$\Delta G$ calculation of target molecule and the aptamer .....	29
2.6.	RMSD and RMSF Calculations .....	30
2.6.1.	RMSD Analysis .....	30
2.6.2.	RMSF Analysis .....	30
2.7.	Visualization of target molecule and the aptamer interaction .....	31
2.8.	Determination of possible mutation sites .....	31
2.9.	<i>In-silico</i> mutagenesis of the three closest bases in PYMOL .....	34
2.10.	Docking with mutant aptamers .....	36
2.11.	MD simulation of mutant aptamer with the best dock.....	38
CHAPTER 3: RESULTS & DISSCUSSION .....		39
3.1.	Structural and affinity analysis.....	39
3.1.1.	Obtained 2D aptamer structures.....	40
3.1.2.	Obtained 3D aptamer structures.....	44
3.1.3.	Molecular docking with the aptamer and target molecules .....	48
3.1.4.	Molecular dynamics simulations .....	52
3.1.5.	$\Delta G$ calculations .....	62
3.1.6.	Visualization of the interactions.....	64
3.2.	Optimization through <i>in-silico</i> mutagenesis.....	70
3.2.1.	Proposing targets and mutant sequences.....	70
3.2.1.1.	Determination of possible mutation sites.....	70
3.2.1.2.	<i>In-silico</i> mutagenesis of the three closest bases in PYMOL .....	71
3.2.2.	Docking with mutant aptamers .....	71
3.2.3.	MD simulations of mutant aptamer with the best dock .....	74
3.2.4.	$\Delta G$ calculations .....	80
3.2.5.	Visualization of the interactions.....	81
CHAPTER 4: CONCLUSION.....		83
REFERENCES.....		85

## LIST OF TABLES

Table 1. Affinity and RMSD Analysis of Alsager22-Ligands Dockings. ....	50
Table 2. $\Delta G$ values of estradiol, progesterone, testosterone, and androstenedione molecules calculated using the C2 approach as a result of four MD simulation experiments with Alsager22.....	63
Table 3. $\Delta G$ values of estradiol, progesterone, testosterone, and androstenedione molecules calculated using the IE approach as a result of four MD simulation experiments with Alsager22.....	63
Table 4. Docking results of mutant aptamers and estradiol. The unit of higher affinity is kcal/mol. ....	72
Table 5. Table 4 (continued) .....	73
Table 6. $\Delta G$ values of estradiol calculated using the C2 approach as a result of four MD simulation experiments with mutant Alsager22 and original Alsager22.....	80
experiments with Alsager22.....	80
Table 7. $\Delta G$ values of estradiol calculated using the IE approach as a result of four MD simulation experiments with mutant Alsager22 and original Alsager22. ....	80

## LIST OF FIGURES

Figure 1. Methodology flow chart for the study. ....	13
Figure 2. The first of the predicted structural configurations. ....	41
Figure 3. The second of the predicted structural configurations.....	42
Figure 4. The third of the predicted structural configurations. ....	43
Figure 5. Visualization of 5 different structures obtained in XiaoLab 3dRNA/DNA structure prediction software.....	46
Figure 6. Superimposed visualization of 5 different structure predictions obtained from XiaoLab 3dRNA/DNA service. ....	47
Figure 7. Distance between N3 of Thymine 5 - O4 of Thymine 17 and O2 of Thymine 5 - N3 of Thymine 17. The unit is Armstrong. ....	47
Figure 8. The rotatable bond in the progesterone shown in yellow. ....	49
Figure 9. Visualization of the modes in which the hormones androstenedione, estradiol, progesterone, and testosterone dock with the aptamer with the highest affinity.....	51
Figure 10. Positions of Alsager22 aptamer and androstenedione hormone after 10 nanoseconds of four different MD simulations.....	53
Figure 11. Positions of Alsager22 aptamer and estradiol hormone after 10 nanoseconds of four different MD simulations. ....	54
Figure 12. Positions of Alsager22 aptamer and progesterone hormone after 10 nanoseconds of four different MD simulations.....	55
Figure 13. Positions of Alsager22 aptamer and testosterone hormone after 10 nanoseconds of four different MD simulations.....	56
Figure 14. First of the MD simulation trials of Estradiol and Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Alsager22 over time. (B) Calculation of the RMSD values of Alsager22 after least square fit to Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Alsager22. (E) Calculation of RMSF values for Estradiol.....	58
Figure 15. Second of the MD simulation trials of Estradiol and Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Alsager22 over time. (B) Calculation of the RMSD values of Alsager22 after least square fit to	

Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Alsager22. (E) Calculation of RMSF values for Estradiol.....	59
Figure 16. Third of the MD simulation trials of Estradiol and Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Alsager22 over time. (B) Calculation of the RMSD values of Alsager22 after least square fit to Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Alsager22. (E) Calculation of RMSF values for Estradiol.....	60
Figure 17. Fourth of the MD simulation trials of Estradiol and Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Alsager22 over time. (B) Calculation of the RMSD values of Alsager22 after least square fit to Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Alsager22. (E) Calculation of RMSF values for Estradiol.....	61
Figure 18. Interactions of estradiol hormone with Alsager22 aptamer at the end of 10 nanoseconds of four different MD simulations.....	65
Figure 19. Interactions of progesterone hormone with Alsager22 aptamer at the end of 10 nanoseconds of four different MD simulations.....	66
Figure 20. Interactions of testosterone hormone with Alsager22 aptamer at the end of 10 nanoseconds of four different MD simulations.....	67
Figure 21. Interactions of androstenedione hormone with Alsager22 aptamer at the end of 10 nanoseconds of four different MD simulations. ....	68
Figure 22. Possible mutation sites.....	70
Figure 24. First of the MD simulation trials of Estradiol and Mutant Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Mutant Alsager22 over time. (B) Calculation of the RMSD values of Mutant Alsager22 after least square fit to Mutant Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Mutant Alsager22. (E) Calculation of RMSF values for Estradiol. ....	76
Figure 25. Second of the MD simulation trials of Estradiol and Mutant Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Mutant Alsager22 over time. (B) Calculation of the RMSD values of Mutant Alsager22 after least square fit to Mutant Alsager22 over time. (C) Calculation of the RMSD values of Estradiol	

after least square fit to Estradiol over time. (D) Calculation of RMSF values for Mutant Alsager22. (E) Calculation of RMSF values for Estradiol. ....	77
Figure 26. Third of the MD simulation trials of Estradiol and Mutant Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Mutant Alsager22 over time. (B) Calculation of the RMSD values of Mutant Alsager22 after least square fit to Mutant Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Mutant Alsager22. (E) Calculation of RMSF values for Estradiol. ....	78
Figure 27. Fourth of the MD simulation trials of Estradiol and Mutant Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Mutant Alsager22 over time. (B) Calculation of the RMSD values of Mutant Alsager22 after least square fit to Mutant Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Mutant Alsager22. (E) Calculation of RMSF values for Estradiol. ....	79
Figure 28. Interactions of estradiol hormone with Alsager22 aptamer at the end of 10 nanoseconds of four different MD simulations.....	82



## LIST OF ABBREVIATIONS

2D: Two Dimensional

3D: Three Dimensional

A: Adenine

C: Cytosine

DBN: Dot Bracket Notation

G: Guanine

GROMACS: Groningen Machine for Chemical Simulations

MD: Molecular Dynamics

MFE: Minimum Free Energy

MMPBSA: Molecular Mechanics Poisson- Boltzmann Surface Area

N: Nitrogen

NMR: Nuclear Magnetic Resonance

NPT: Box volume constant, Particle number constant, Pressure constant

NVT: Box volume constant, Particle number constant, Temperature constant

O: Oxygen

RMSD l.b.: Root Mean Square Deviation of the Lower Bound

RMSD u.b.: Root Means Square Deviation of the Upper Bound

RMSF: Root Mean Square Fluctuation

SELEX: Systematic Evolution of Ligands by Exponential Enrichment

SSEs: Smallest Secondary Elements

T: Thymine

$\Delta G$ : Gibbs Free Energy

$\Delta S$ : Entropy

## CHAPTER 1: INTRODUCTION

### *1.1. Aptamers generated with SELEX*

#### *1.1.2. Aptamers*

Aptamers are small, single-stranded DNA or RNA molecules that have emerged as crucial tools in biomedical science. Their development is made possible through a sophisticated process known as SELEX. This method involves screening a vast library of random nucleic acid sequences to identify those that bind most effectively and specifically to a selected target. These targets can be quite diverse, ranging from small molecules and proteins to entire cells. The SELEX process is iterative, involving repeated cycles of binding the sequences to the target, partitioning the bound from the unbound sequences, and amplifying the bound sequences. This cycle is repeated multiple times, leading to the isolation of aptamers with superior binding characteristics. In biology and medicine, aptamers play a significant and varied role. In diagnostics, their high specificity for biomarkers renders them ideal for use in biosensors and diagnostic assays. As therapeutic agents, aptamers offer a non-immunogenic and highly specific alternative to traditional antibodies, making them particularly valuable in targeted drug delivery. This is especially crucial in cancer treatment, where their ability to selectively bind to tumor cells minimizes side effects and enhances treatment efficacy. Aptamers have numerous advantages over antibodies. They can be chemically synthesized, ensuring consistent quality, and allowing for modifications to enhance stability, binding affinity, and specificity. Moreover, aptamers are typically smaller than antibodies, which improves their tissue penetration, a critical factor for diagnostic and therapeutic applications. Aptamers' smaller size and ability to tailor their properties position them as a highly versatile and promising tool in modern biomedical science.

### ***1.1.3. SELEX method***

First introduced in 1990, SELEX is a sophisticated molecular technique designed for identifying and evolving ligands with high affinity and specificity for a given target molecule (Blackwell and Weintraub, 1990; Wright, Binder and Funk, 1991). This target can be a variety of biomolecules, such as proteins, nucleic acids, or even small molecules, making SELEX a versatile tool in both research and therapeutic development (Darmostuk et al., 2015). The process starts with a large, diverse library of oligonucleotides (short DNA or RNA sequences), often comprising millions of different sequences with random regions. These sequences represent a vast combinatorial space of potential ligands. The library is incubated with the target molecule, and through a process of binding and washing, sequences that do not bind or bind weakly to the target are removed. The remaining sequences, which have a higher affinity for the target, are then separated and amplified. This cycle of binding, separation, and amplification is repeated multiple times, with each iteration enriching the pool with sequences that have the highest affinity for the target. This repeated process is an important part of SELEX because it lets the most specific and strong ligands be chosen over time from a pool that is initially random. After several rounds, the sequences that bind most effectively to the target dominate the pool of oligonucleotides. These sequences are then cloned and sequenced to be identified. These identified oligonucleotides can be further characterized and modified for enhanced stability, binding affinity, or specificity, making them valuable for various applications such as therapeutic agents, diagnostic tools, or to study biomolecular interactions. In particular, four main post-SELEX optimization methods are in focus: truncation, chemical modification, construction of bivalent or multivalent aptamers, and mutagenesis (Gao et al., 2016).

## ***1.2. Secondary structure prediction***

Nucleic acid aptamers have gathered significant attention in various biotechnological and medical applications, primarily owing to their remarkable specificity and binding capability towards the intended molecule. The comprehension of aptamer structure aids in elucidating the functionality of these molecules and enhancing their design. Consequently, the prediction of aptamer secondary structure emerges as a pivotal step towards comprehending the operation of these specialized nucleic acid entities. The prediction of aptamer secondary structure assumes paramount importance when data acquired through experimental means is limited or costly. This secondary structure encompasses certain structural attributes, such as base pairs, loops, and terminal regions, which exist among diverse segments of the nucleotide sequence within an aptamer. Acquiring knowledge about this structure allows us to comprehend the ways in which aptamers bind to target molecules and their interaction mechanisms at a higher level. A multitude of computational and bioinformatics methodologies have been devised to forecast the secondary structure of aptamers, each employing distinct methods and applications. These methodologies generate secondary structure predictions by utilizing the thermodynamic and physical properties of the nucleotide sequence. RNAfold and UNAFold (Mfold) are popular software tools used to predict the secondary structure of nucleic acids (Zuker, 2003; Gruber et al., 2008; Markham and Zuker, 2008). These tools work based on the MFE method and can generate secondary structure predictions with different options . Accurate prediction of the secondary structure of aptamers constitutes a fundamental stride in modeling and enhancing the interactions between these molecules and target entities. Moreover, these predictions facilitate the comprehension of aptamer behavior under varying conditions and the optimization of design processes.

### **1.2.2. UNAFOLD**

UNAFold is a software tool for the prediction of secondary structures of RNA and DNA molecules. The basic working principle of this software is to analyze the possible structural configurations of nucleic acid chains to identify stable structures. UNAFold takes a single-stranded sequence of an RNA or DNA molecule and predicts the possible secondary structures of this sequence by performing thermodynamic calculations. By considering the base pairings between nucleotides, the software determines the possible secondary structures. The thermodynamic stability of these structures is calculated based on the principle of free energy minimization. That is, UNAFold calculates the  $\Delta G$  for each possible structure and identifies the structure with the lowest free energy as the most stable structures. These thermodynamic calculations include parameters such as experimentally determined base pairing energies and how the bases interact depending on environmental conditions. As a result, UNAFold is a powerful thermodynamics-based nucleic acid secondary structure prediction tool with a wide range of applications in molecular biology and genetic engineering. Thanks to this tool, the structural and functional properties of nucleic acids are better understood, and this knowledge contributes to the development of new biotechnological applications.

### **1.3. Tertiary structure prediction**

3D structural estimates of nucleic acids are a research area that has made major advances in molecular biology and biophysics in recent years. Nucleic acids, such as RNA and DNA, play a fundamental role in the storage and expression of genetic information, so understanding the 3D structure of these molecules is critical to understanding their biological functions in depth. Traditional structural biological methods, especially X-ray crystallography NMR spectroscopy, have been used as important tools in obtaining detailed structural information on nucleic acids at the atomic level. While these methods provide high-resolution structural information, they may sometimes be inadequate or quite time-consuming to solve large and complex nucleic acid structures. In this context, in recent years, computer-aided structural prediction methods, especially machine learning and AI-based algorithms, have great potential to predict the 3D structure of nucleic acids. These algorithms are trained on large data sets, learn the properties of molecular structures, and are used to predict new structures.

Alpha-Fold, developed by DeepMind, is acknowledged as innovative tools that have transformed the field of protein folding prediction (Wei, 2019). In case of the nucleic acid structural prediction, these technologies contribute significantly to scientific research as well, such as the 3DRNA/DNA service by the XiaoLab (Zhang, Wang and Xiao, 2022; Zhang, Xiong and Xiao, 2022). Progress in this area has enabled us to better understand the biological functions of nucleic acids, and it has also become a key tool for exploring the molecular foundations of diseases and developing new therapeutic strategies. Furthermore, advances in these technologies open new doors in areas such as synthetic biology. The design and manufacture of synthetic nucleic acid structures creates new possibilities for biotechnological applications. Briefly, 3D structural predictions of nucleic acids are a rapidly evolving field. Research in this area has the potential to provide better understanding of biological systems, but also to provide innovative solutions in the fields of medicine and biotechnology.

### **1.3.2. XIAOLAB**

The XiaoLab 3DRNA/DNA Service focuses on creating 3D models of nucleic acid structures using advanced algorithms and computer simulation techniques. The study of Xiao et al. in this field provides solutions to fundamental problems in the structural biology of nucleic acids, which find applications in fields such as molecular biology, genetics, and drug design (Zhang, Wang and Xiao, 2022; Zhang, Xiong and Xiao, 2022). This method generates 3D structures of RNA and DNA using sequence and secondary structure information. The first step is to decompose the given secondary structure into the SSEs. This involves an effort to find templates for SSEs from a library of templates created from crystal or NMR structures. If a template cannot be found, a distance-geometry-based looping method can be used to construct the SSE *ab initio*. Then, for each SSE, the structure built with the found templates forms the assembly module, ignoring pseudo-nodes at this stage. Finally, starting from the resulting assembly module or a given structure, it optimizes the structure using the Simulated Annealing Monte Carlo method. This optimization process can be performed using a coarse-grained energy function that includes constraints such as pseudo-nodes, residue-residue interactions, or distances. In this way, the resulting structure is optimized in accordance with the constraints specified by pseudo nodes, residue-residue interactions, and distances.

In conclusion, the XiaoLab 3DRNA/DNA Service is a powerful tool for understanding the 3D structures of nucleic acids and exploring the effects of these structures on their biological functions. This service makes significant contributions to molecular biology and genetics research, drug discovery, and genetic engineering projects and is a valuable resource for ongoing research in these fields.

#### ***1.4. Interaction analysis***

##### ***1.4.1. Molecular docking***

Molecular Docking is a computational method utilized in numerous fields of science like biochemistry, bioinformatics, drug design, and biomedical research. This method is used to understand how a ligand binds and interacts with a receptor biomolecule. The ligands may be small molecules, peptides, proteins, or even nucleic acids. On the other hand, usually pre-folded protein or DNA molecule structures are used as the receptors. The Molecular Docking process involves many computational steps. In the first step, the molecular structures representing the receptor and ligand are defined in a computer environment. A sequence of computations is subsequently conducted to ascertain the way the ligand interacts with the receptor. These calculations involve appropriately transforming the ligand towards the receptor, which is called the docking, and calculating the energies of the interactions. Moreover, this technique facilitates enhanced comprehension of biological processes and mechanisms underlying the action of pharmaceutical agents. One of the application areas where Molecular Docking is used is aptamer - target molecule interaction analysis. Aptamer-target molecule docking is employed to comprehend the mechanism by which aptamers interact with target molecules and how this interaction transpires on a molecular scale. This method is important to evaluate and improve the selective binding capabilities of aptamers. There are different approaches to Molecular Docking. Blind docking explores potential binding sites across the entire receptor surface, while site-directed docking focuses on a specific active site. Flexible docking, which accounts for the flexible structures of both ligand and receptor, offers more realistic interactions. In contrast, rigid docking, with simpler modeling based on fixed ligand and receptor, is also used. These methods enhance the understanding of molecular interactions and facilitate the development of targeted therapies and diagnostics.

#### ***1.4.1.1. AUTODOCK tools & Vina***

AutoDock Tools and Vina are two important software tools for molecular docking, widely used in drug discovery and structural biology. AutoDock Tools provides a comprehensive interface for users to prepare and analyze receptor and ligand docking simulations. AutoDock Tools (Sanner, 1999; Sanner and Stoffler, 2001) allow setting various parameters in the molecular docking process and visualizing simulation results. AutoDock Vina software executes the docking process. AutoDock Vina software is specifically designed for macromolecule-ligand interactions, enabling high-performance docking simulations (Trott and Olson, 2010). One of the most important features of AutoDock Vina is the high speed and accuracy of its algorithm, which is especially advantageous in analyzing large data sets and drug discovery projects. To figure out how well a receptor and ligand bind, AutoDock Vina uses energy terms that consider van der Waals interactions, hydrogen bonds, and electrostatic interactions. Considering the flexibility of the ligand, it explores different conformations and uses Monte Carlo simulations and genetic algorithms to find the optimal binding posture. A scoring function evaluates the potential binding postures of the ligand in the user-specified grid box, predicting the binding affinity. In conclusion, the combination of AutoDock Tools and Vina offers a powerful and flexible set of tools in the fields of drug discovery and structural biology.

#### ***1.4.2. Molecular dynamics simulations***

MD is a computational method used to understand how systems at the atomic and molecular level evolve over time. These simulations model the motion of atoms and molecules based on Newton's laws of motion (Rognan, 1998). In general, these simulations calculate the interactions between atoms using a force field, and these interactions determine the structural and dynamic properties of the system at the atomic level, such as temperature and pressure. The force fields known in the literature are usually based on experimental data or detailed quantum mechanical calculations and describe the potential energy between atoms. One of the applications of MD simulations is the understanding of the dynamics of biomolecules. MD studies predict the movement of proteins or other large biomolecules at the atomic level over time, showing how these molecules undergo structural changes and the effects of these changes on their function. Several powerful tools are available for MD simulations.



CHARMM is a flexible and comprehensive MD software that is widely used, especially for simulations on biological systems and polymers (Brooks et al., 1983). LAMMPS is the versatile and extensible tool of choice for simulations in materials science and solid-state physics (Thompson et al., 2022). Each of these tools offers unique features and capabilities to address different needs and research areas. GROMACS is known for being high-performance and very flexible; it is specifically designed for modeling the dynamics of biomolecules including nucleic acids (Abraham et al., 2015). The accuracy of MD simulations depends on the force field used and the correct choice of computational parameters. Using force fields and other parameters that are not suitable for the system may lead to misleading results. Therefore, the success of these simulations depends not only on powerful computer resources but also on detailed theoretical and experimental knowledge.

#### ***1.4.2.1. Force fields***

MD simulations are carried out using force fields describing the interactions between atoms and molecules. A force field is a mathematical description of the potential energy between atoms, and how this energy depends on distance or angle. The physical and chemical properties of molecular systems are modeled using force fields. With these simulations we can study such things as the folding dynamics of proteins, drug-target interactions, or a material's mechanical properties. In the literature there are many popular force fields used in MD simulations. For this reason, CHARMM is tailored to suit biomolecules and commonly used for in-depth research on proteins, nucleic acids, or lipids. Originally designed for proteins and nucleic acids, AMBER has been built up to cover a broad chemical space. GROMOS is often used to simulate biological molecules and is well-suited for modelling the behavior of molecules in aqueous solutions. OPLS is a simulation method, both for organic molecules and biomolecules. It can also be used to compute the thermodynamic properties of liquids.

Such force fields contain many energy terms, including bond lengths, angles between bonds and dihedrals, van der Waals interactions and electrostatic forces. These force fields are used in MD simulations, by which the motion of atoms and molecules is traced step-by-step over time.

The choice of the force field depends on whether you're studying a small or large system, how much detail is needed for your calculations, the quality of its parametrization, compatibility with available software and finally on what others have done before. It is an important decision because the accuracy of results depends on it.

#### ***1.4.2.2. GROMACS***

GROMACS is an open-source, high-performance software tool for MD simulations (Abraham et al., 2015). First developed at the University of Groningen in the early 1990s, it was specifically designed to study the dynamics of biological molecules and polymers. GROMACS is especially optimized for atomic-level simulations of biomolecules such as proteins, nucleic acids, and lipids, and plays an important role in understanding the structure and function of biological systems (Van Der Spoel et al., 2005). One of the most notable features of GROMACS is its parallel computing capabilities and fast algorithm structure. Parallel computing is the simultaneous execution of a computational operation by multiple processors or computers. This approach allows larger and complex calculations to be completed faster. In addition, it is compatible with different force fields and integration algorithms. Thanks to these features, simulations of large biological systems and longtime scales can be performed efficiently. GROMACS includes basic simulation techniques such as energy minimization, temperature, and pressure control, as well as offers a wide range of tools for more complex processes. These include detailed analysis of processes such as protein folding, ligand binding, enzyme mechanisms and cell membrane dynamics. Another important aspect of GROMACS is its ease of use through the terminal and its extensive documentation. These features allow users to easily set up and run their simulations and analyze the results. Furthermore, GROMACS is continuously being developed and updated by a community of users, which makes the software constantly adaptable to state of art research needs. GROMACS has a wide range of applications in many fields such as molecular biology, biophysics, materials science, and drug design. The detailed simulation capabilities provided by GROMACS allow for a better understanding of the processes occurring at the molecular level and the use of this knowledge in the development of new therapeutics.

### ***1.4.3. Receptor-ligand interaction analysis***

Thermodynamic analysis methods are extremely important in biochemistry and drug design. They are employed to give us important information about the character of chemical reactions and their changes in energy. The  $\Delta G$ , which is included among the thermodynamic parameters, can be taken as a basic standard to judge both whether a chemical reaction will occur spontaneously and if its direction of change in energy. As chemical reactions occur, they undergo energy transformations from the initial state to their final state. Once we comprehend this transformation in energy, then the state of the reactions becomes clear. If  $\Delta G$  is negative, then the reaction may occur spontaneously. This means that the reaction is energetically favorable and does not require any net input of external energy. If  $\Delta G$  is greater than zero, the reaction does not tend to occur spontaneously and must be driven from without. If  $\Delta G$  is zero, the reaction has reached equilibrium.  $\Delta G$  is very important in biochemistry and drug design, especially when it comes to evaluating the interactions of ligand-receptor complexes. Ligands act to do their biological work by attaching themselves to target receptors. If  $\Delta G$  is negative, the ligand tends to bind with the receptor and this reaction spontaneously occurs. In other words, the ligand-receptor interaction is advantageous from an energy point of view. But the calculation of these thermodynamic parameters is complex and requires us to carefully examine molecular-level interactions. It is at this point those computational methods such as MMPBSA are needed. MMPBSA is a computational protocol for analyzing the results of MD simulations and calculating estimates of binding free energies in molecular interactions. It is also an important means of predicting the efficacy of ligands and aptamers. This calculation method is part of numerous MD simulation tools. One of them is the GMX\_MMPBSA package from Gromacs (Valdés-Tresanco et al., 2021). To sum up, methods of thermodynamic analysis have become widely used in all manner of fields such as drug discovery and development; aptamer interaction; biochemistry more generally; pharmacology. These methods are making important contributions toward a better understanding of intermolecular interactions and the creation of targeted strategies.

### ***1.5. In-silico mutagenesis***

*In-silico* mutagenesis of nucleic acids is the process of artificially altering genetic material in a computerized environment, which is of increasing importance in molecular biology and drug design. This method involves replacing, deleting, or adding specific bases in DNA or RNA sequences, usually using bioinformatics tools and molecular modeling software. Such simulations are used to study in detail the effects of genetic changes on protein structures, functions, and intracellular interactions. *In-silico* mutagenesis is favored in studies aiming to understand the molecular mechanisms of genetic diseases, identify potential therapeutic targets and predict the effects of drug candidates. *In-silico* mutagenesis offers a faster, less resource-intensive and more comprehensive alternative to *in vitro* mutagenesis experiments performed in the laboratory. *In-silico* mutagenesis is also used to understand the basic mechanisms of genetic diseases and plays an important role in the design and optimization of potentially effective molecules and drug candidates for their treatment. In cancer genetics research, *in-silico* mutagenesis is used to model the genetic alterations involved in tumor development and their effects on cellular processes.

### ***1.6. Purpose and importance of the thesis***

The aim of this thesis is to perform a comprehensive structural analysis of aptamers produced by the SELEX method and based on this analysis, to design a workflow to create a modified aptamer to increase affinity. The objective of this study is to conduct a comprehensive examination of aptamers, in addition to their affinitive characteristics. First, a 2D aptamer structure was obtained based on aptamer sequences. Subsequently, methods were adopted to produce the 3D configuration. At this point, it is aimed to reveal the physical structure of aptamers in detail. The analysis of the interaction between the target molecule and the 3D aptamer structure was conducted utilizing the molecular docking method. This analysis was performed to evaluate the binding potential of aptamers with the target molecule. At this stage, specific configurations and binding patterns of aptamers were studied. The MD simulations were carried out to analyze the interaction between the target molecule and the aptamer in more detail, following the results of molecular docking. These simulations were used to understand the binding dynamics of aptamers with the target molecule. In addition, the evaluation of thermodynamic stability of this complex was conducted through the utilization of  $\Delta G$  calculations. Consequently, it was examined whether the performance of existing aptamers can be improved. The molecular interactions of the target with the structural features of the existing aptamer were studied and potential mutations on the aptamer that may increase binding affinity were proposed. Potential mutants were then analyzed using molecular docking and dynamics to finalize proposed mutants. In conclusion, this thesis aims to establish a comprehensive workflow for the structural analysis of aptamers and to investigate whether existing aptamers can be further improved with *in-silico* mutagenesis.

## CHAPTER 2: METHODOLOGY

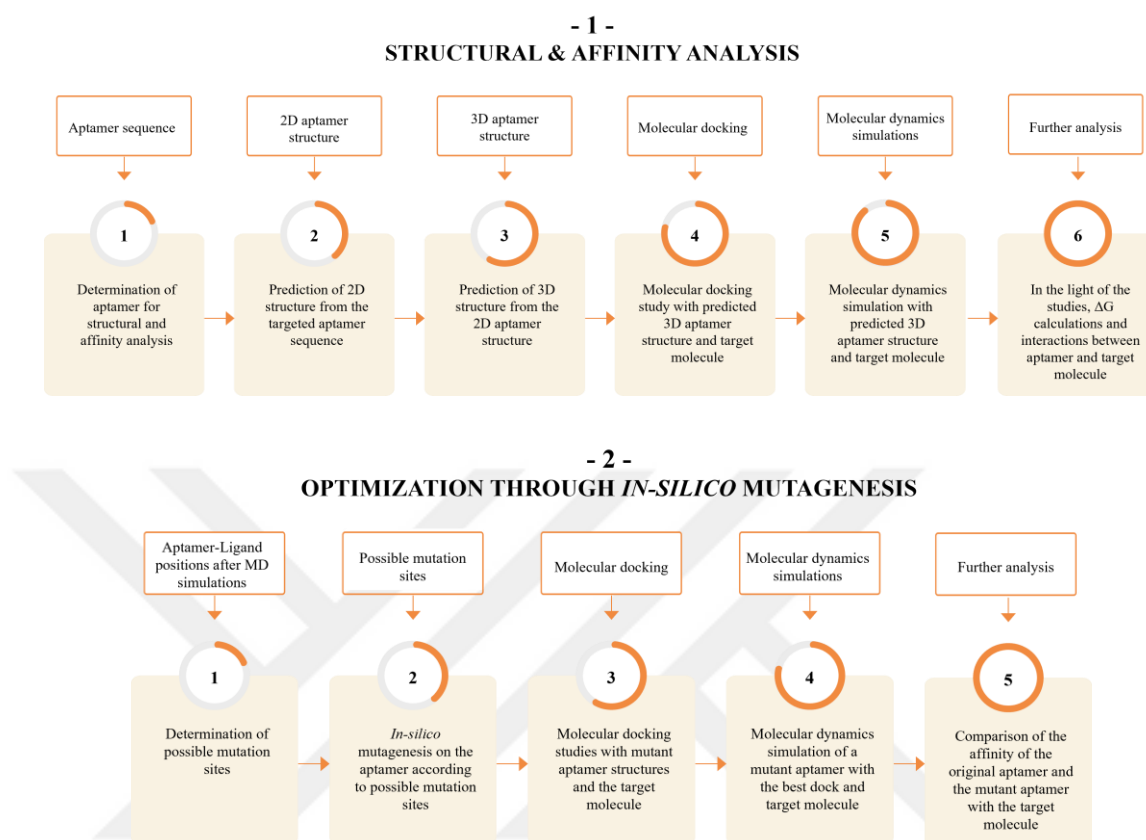


Figure 1. Methodology flow chart for the study.

### ***2.1. Obtaining 2D aptamer structures***

2D aptamers can be created by modifying the primary structure of DNA oligonucleotides. One of the bioinformatics tools, UNAFold, is used to predict secondary structures of DNA and RNA molecules based on thermodynamic stability. The first step of the structural bioinformatics part of this thesis is to use UNAFold to turn the target aptamer sequence into a 2D structure. The following steps are followed to generate the secondary structure of a DNA sequence using web tool at <http://www.unafold.org/mfold/applications/dna-folding-form.php>:

- a. The sequence of the target aptamer for which the 2D structure to be generated was entered to the relevant text box.
- b. In the "Advanced Options" section, the desired temperature and salt concentration parameters to be used for analysis were set.
- c. After clicking the "Submit" button, the secondary structure drawing(s), delta G values of each drawing, and DBN are displayed. DBN is a notation method to express the base pairing in secondary structures of nucleic acids.

### ***2.2. Obtaining 3D aptamer structures***

To determine the binding properties of aptamers, it is necessary to have information about their 3D structures. 3D structures were obtained using the DBN of aptamers obtained in the previous section. For this transformation, it was preferred to use a software tool named Xiao Lab. The Xiao Lab software tool is a web-based bioinformatics tool at <http://biophy.hust.edu.cn/new/3DRNA> that allows the calculation of 3D structures of nucleic acid using the secondary structure. The Xiao Lab software tool is used by following the steps below:

- a. "Optimize" approach was selected as the preferred calculation approach from the "Procedure" section.
- b. The sequence of the target aptamer was entered in the relevant text box.
- c. DBN notation was entered in the "2D Structure" section, as obtained in the previous 2D structure generation step.
- d. In the "# of Predictions" section, the number of structure predictions to be obtained from the tool at the end of the process was entered as "10".

- e. In the "Advanced Options" section, the default settings of certain parameters remained unchanged.
- f. The calculation was started using the "Submit" button. Processing time may vary depending on the size and complexity of the aptamer.
- g. After the process was completed, previews of the obtained 3D structures were displayed, and the PDB of the desired structure prediction were downloaded.

### ***2.3. Molecular docking with the aptamer and target molecules***

To predict correct binding poses of the ligand and the aptamer, molecular docking methods are used to analyze the 3D aptamer structures. In this context, specially developed bioinformatics software tools enable molecular docking using 3D structures of aptamers and target molecules. The binding position of the aptamer to the target molecule, obtained using molecular docking, provides a starting point for the prospective MD simulations. As a result of the studies, AutoDock Tools and AutoDock Vina bioinformatics software tools were preferred for DNA aptamer and small molecule docking. Molecular docking was performed in this thesis by following the steps:

- a. Created a docking folder containing the PDB files of the aptamer and the target molecule, as obtained from Xiao Lab and PubChem, respectively.
- b. aptamer.pdb in the docking folder was opened in the application with the "Read Molecule" option in AutoDock-Tools, (part of MGL tools).
- c. The all hydrogens in the structure were removed and readded with the "polar only" option in the "edit" section.
- d. To save the obtained structure in PDBQT format, "Macromolecule" option was clicked in the "grid" section, opening the window required to select the aptamer with the "choose" button. After selection of the aptamer using the "select molecule" button, the "aptamer.pdbqt" file is saved in the same directory.
- e. Next, the grid box was added to define the section where the docking program will perform the docking of the target molecule into the aptamer. In this thesis, the blind docking principle is preferred. Accordingly, the position of the grid box was determined to include the entire aptamer structure. "Grid Box" was clicked under the grid section and the grid options were set to this value.



- f. After the grid box is determined, the molecule.pdb in the placement file was loaded into the application with the "input" option under the "Ligand" section.
- g. After the target molecule is loaded, torsion was selected with the "Torsion Tree" option under the "Ligand" section to determine the rotatable bonds and completed by clicking the "done" button.
- h. The "molecule.pdbqt" file was saved in the same directory with the "output" option under the "Ligand" section.
- i. A configuration file in text format was created inside the docking file using Notepad. The file was modified to contain the Grid Box parameters, receptor, and ligand file names as in the example below. Since blind docking is preferred in this docking study, the exhaustiveness parameter was increased to 128 and specified in the configuration file since many repetitions are needed to find the correct binding. Example configuration file contents:

**conf.txt**

---

```
receptor = aptamer.pdbqt  
ligand = molecule.pdbqt  
center_x = 11  
center_y = 90.5  
center_z = 57.5  
size_x = 26  
size_y = 26  
size_z = 26  
exhaustiveness=128
```

---

- j. In the Windows operating system, command prompt is opened, and the directory where the docking folder is located was set as the current directory.
- k. To run the docking process, the following command is entered in the command prompt using proper full address of vina executable:

```
"C:\Program Files (x86)\The Scripps Research Institute\Vina\vina.exe" -  
-config conf.txt --log log.txt"
```

This generates the binding modes and outputs a PDBQT extension with the symbol molecule\_out.

#### **2.4. Molecular dynamics simulations**

The Swissparam tool was preferred for generating small-molecule topologies and force field parameters. The amber ff99SB-ILDN force field was used for nucleic acids in this study. MD simulations were performed using the GROMACS 2018.1 package. The environment in which the MD simulations were performed was solved in a cubic box containing the aptamer and target molecule complexes with the TIP3P water model. To ensure electrical neutrality, Na<sup>+</sup> or Cl<sup>-</sup> ions were added to neutralize the simulation system. Prior to the MD simulations, the steepest descent energy minimization algorithm was used to stabilize the system and obtain a stable starting point. During this minimization process, the system was allowed to rest for a maximum of 50,000 steps. The minimization process was set with a maximum force tolerance of 1000 kJ/mol and an energy step size of 0.01, and the minimization was stopped when the maximum force dropped below 10.0 kJ/mol. After energy minimization, the system was equilibrated for 0.1 ns by simulating NVT at 300 K and NPT at 1 bar pressure. Long-range Coulomb interactions were addressed using the Particle Mesh Ewald approximation, while Coulomb and van der Waals interactions were approximated using a cut-off distance of 1.2 nm. Finally, the equilibrated systems were simulated with a time step of 2 fs for a total duration of 10 ns. These parameters were entered into the simulation using the following steps:

- a.** Open the “molecule\_out.pdbqt” file obtained from the docking result in the Discovery Studio software tool, select the first position with the highest affinity, and delete the rest of the positions.
- b.** Add hydrogen to the structure in the "Chemistry" section. It is then saved in the name and format molecule.mol2.
- c.** The obtained "molecule.mol2" and "aptamer.pdb" files are added to a folder named "MD," and a terminal is opened in the Linux operating system at the target folder location.
- d.** With the pdb2gmx command, the aptamer file should be processed with the selected force field and water type information and converted to Gromacs structure format. To convert the "aptamer.pdb" file to the "aptamer\_processed.gro" file, the following command is entered into the terminal:

```
"gmx pdb2gmx -f aptamer.pdb -o aptamer_processed.gro"
```

- e. The appropriate forcefield is selected on the screen. In this project, "AMBER99SB-ILDN protein, nucleic AMBER94" was preferred.
- f. Then a water type selection is made. In this project, the TIP3P water type was preferred.
- g. 3D coordinate files for the "molecule.mol2" file are obtained from the Swissparam software tool, and the molecule topology files generated by the system are added to the MD folder.
- h. "molecule.pdb" file should be converted to Gromacs structure format with the editconf command. To convert the "molecule.pdb" file to the "molecule.gro" file, the following command is entered into the terminal:

```
"gmx editconf -f molecule.pdb -o molecule.gro"
```

- i. The resulting "molecule.gro" file is opened with a suitable application, and the molecule information is copied. Then, the "aptamer\_processed.gro" file is opened with a suitable application, and the copied molecule information is pasted below the last line of aptamer atoms and above the box vectors. The number of newly added molecules is added to the number of molecules at the beginning of the "aptamer\_processed.gro" file. aptamer\_processed.gro is renamed to "complex.gro".
- j. The resulting "topol.top" file is opened with an appropriate application, and the following information is added to the last line to indicate that it is a new molecule in the complex.gro file.

### Parameters

---

"#compound	#mols
MOLECULE	1"

---

- k. To include the parameters of the target molecule in the system topology, the molecule.itp file obtained with swissparam is added under the "include forcefield parameters" section of the "topol.top" file as follows:

## Parameters

---

```
"#include "molecule.itp""
```

---

- l.** The unit cell is where the simulation will be performed, and the initial structure must be prepared. In this step, the shape and size of the unit cell are determined by considering the complex structure and specified in the command. In this thesis, the shape of the unit cell is chosen to be cubic, and its dimensions are 2.0 nm x 2.0 nm x 2.0 nm. The command creates the unit cell, organizes the structure to fit inside the unit cell, and adjusts the initial coordinates. For these operations, the following command is entered into the terminal:

```
"gmx editconf -f complex.gro -o newbox.gro -bt cubic -d 2.0"
```

- m.** With the solvate command, the structure in the unit cell is filled with solution, and a solution system is created. For this operation, the following command is entered into the terminal:

```
"gmx solvate -cp newbox.gro -cs spc216.gro -p topol.top -o solv.gro"
```

- n.** In a solvated system, there is a charged complex. Since real life is not based on a specific charge, it is necessary to neutralize this system by adding ions. To do this, first a parameter file with the mdp extension is created. In this thesis, the content of the ions.mdp file is as follows:

### **ions.mdp** (Lemkul, 2019)

---

```
title = Minimization ; Title of run
; Parameters describing what to do, when to stop and what to save
integrator = steep ; Algorithm (steep = steepest descent minimization)
emtol = 1000.0 ; Stop minimization when the maximum force < 10.0 kJ/mol
emstep = 0.01 ; Energy step size
nsteps = 50000 ; Maximum number of steps to perform
; Parameters describing how to find the neighbors of each atom and how
to calculate the interactions
nstlist = 1 ; Frequency to update the neighbor list and long-range forces
cutoff-scheme = Verlet
ns_type = grid ; Method to determine neighbor list (simple, grid)
```

```
rlist = 1.0 ; Cut-off for making neighbor list (short range forces)
coulombtype = cutoff ; Treatment of long-range electrostatic interactions
rcoulomb = 1.0 ; long range electrostatic cut-off
rvdw = 1.0 ; long range Van der Waals cut-off
pbc = xyz ; Periodic Boundary Conditions
```

---

- o.** Next, it is necessary to create a tpr file by combining the parameter file (ions.mdp), the initial coordinate file (solv.gro), and the topology file (topol.top). For this, the following command is entered into the terminal:

```
"gmx grompp -f ions.mdp -c solv.gro -p topol.top -o ions.tpr"
```

- p.** The ions of the aptamer preferred for MD analysis are determined by examining the buffers used in the wetlab study. In this thesis, calcium was preferred as a positive ion and chloride as a negative ion. The following command is entered into the terminal to add ions to make the system neutral:

```
"gmx genion -s ions.tpr -o solv_ions.gro -p topol.top -pname CA -nname CL -neutral"
```

The "SOL" option is selected in the list that appears in the terminal with the command.

- q.** Before starting the MD simulation, the system, which became neutral in the previous stage, is brought to the minimum energy level. Energy minimization ensures that the atoms and bonds in the system approach the energy minimum so that the system can reach a more stable initial configuration. To minimize energy, first a parameter file with a mdp extension is created. In this thesis, the content of the em.mdp file is as follows:

### **em.mdp** (Lemkul, 2019)

---

```
title = Minimization ; Title of run
; Parameters describing what to do, when to stop and what to save
integrator = steep ; Algorithm (steep = steepest descent minimization)
emtol = 1000.0 ; Stop minimization when the maximum force < 10.0 kJ/mol
emstep = 0.01 ; Energy step size
nsteps = 50000 ; Maximum number of steps to perform
; Parameters describing how to find the neighbors of each atom and how
to calculate the interactions
nstlist = 1 ; Frequency to update the neighbor list and long-range forces
cutoff-scheme = Verlet
ns_type = grid ; Method to determine neighbor list (simple, grid)
rlist = 1.2 ; Cut-off for making neighbor list (short range forces)
coulombtype = PME ; Treatment of long-range electrostatic interactions
rcoulomb = 1.2 ; long range electrostatic cut-off
vdwtype = cutoff
vdw-modifier = force-switch
rvdw-switch = 1.0
rvdw = 1.2 ; long range Van der Waals cut-off
pbc = xyz ; Periodic Boundary Conditions
DispCorr = no
```

---

- r. Next, it is necessary to create a tpr file by combining the parameter file (em.mdp), the initial coordinate file with ions (solv\_ions.gro), and the topology file (topol.top). For this, the following command is entered into the terminal:

```
"gmx grompp -f em.mdp -c solv_ions.gro -p topol.top -o em.tpr"
```

- s. The following command starts the energy minimization simulation using em.tpr and other related input files.

```
"gmx mdrun -v -deffnm em"
```

- t. It is necessary to equilibrate the aptamer-target molecule complex in water. The first step in this equilibration is to restrict the ligand. For this, an index file is first created with the make\_ndx command to select various atomic groups or features from the Gromacs coordinate file named "molecule.gro".

In this thesis, to constrain the target molecule, it is preferred to create an index group containing only non-hydrogen atoms when creating the position restriction topology. To do this, the following command is entered into the terminal:

```
"gmx make_ndx -f molecule.gro -o index_molecule.ndx"
```

Then the following information is entered into the input system on the terminal screen:

```
"0 & ! a H*"
```

- u. Next, an itp file called "posre\_molecule.itp" is created using information from the coordinate file (molecule.gro) and the generated index file (index\_molecule.ndx). This itp file contains the parameters required for position restriction and other molecule identification. For this, the genrestr command is used, and the -fc flag is used to specify the force constants of the position restriction. In this thesis, position restrictions were applied in the x, y, and z directions with force constants of 1000 kJ/mol/nm<sup>2</sup>. For this operation, the following command is entered into the terminal:

```
"gmx genrestr -f molecule.gro -n index_molecule.ndx -o posre_molecule.itp  
-fc 1000 1000 1000 1000"
```

In the selection displayed by the command, select the group labeled "System\_&!H\*".

- v. The information from the previous step needs to be added to the topology file. The topol.top file should be opened with a suitable application, and the following code block should be added just below the position restraint information:

#### **Code block**

---

```
"; Ligand position restraints  
#ifdef POSRES_molecule  
#include "posre_molecule.itp"  
#endif"
```

---

- w. Accurate temperature control is important. The target molecule is grouped together in the temperature equilibration to ensure perfect compatibility with the aptamer. Likewise, the added ions are considered an integral part of the solvent environment and are grouped together. In this context, a special index group is required that brings the aptamer and the target molecule together. For this, it creates an index file to select target groups from the coordinate file "em.gro". For this operation, the following command is entered into the terminal:

```
"gmx make_ndx -f em.gro -o index.ndx"
```

On the page that comes with the command, the numbers of the aptamer (DNA) and the target molecule must be entered as follows.

```
"1 | 2"
```

- x. NVT is a simulation condition in MD simulations with a constant number (N), a constant volume (V), and a constant temperature (T). This condition is used to monitor the behavior of systems. For this, an MDP file is prepared specifying the parameters of the NVT simulation. As mentioned in the previous step, attention should be paid to the line "tc-grps = DNA\_molecule Water\_and\_ions" when preparing this file. In this thesis, the following parameters were preferred:

**nvt.mdp** (Lemkul, 2019)

---

```
title = Aptamer-Target molecule complex NVT equilibration
define = -DPOSRES ; position restrain the Aptamer-Target molecule
; Run parameters
integrator = md ; leap-frog integrator
nsteps = 50000 ; 2 * 50000 = 100 ps
dt = 0.002 ; 2 fs
; Output control
nstenergy = 500 ; save energies every 1.0 ps
nstlog = 500 ; update log file every 1.0 ps
nstxout-compressed = 500 ; save coordinates every 1.0 ps
; Bond parameters
continuation = no ; first dynamics run
constraint_algorithm = lincs ; holonomic constraints
```



```

constraints = h-bonds ; bonds to H are constrained
lincs_iter  = 1      ; accuracy of LINCS
lincs_order = 4      ; also related to accuracy
; Neighbor searching and vdW
cutoff-scheme = Verlet
ns_type       = grid ; search neighboring grid cells
nstlist       = 20   ; largely irrelevant with Verlet
rlist         = 1.2
vdwtype       = cutoff
vdw-modifier  = force-switch
rvdw-switch   = 1.0
rvdw          = 1.2 ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype   = PME ; Particle Mesh Ewald for long-range
electrostatics
rcoulomb      = 1.2 ; short-range electrostatic cutoff (in nm)
pme_order     = 4   ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
; Temperature coupling
tcoupl        = V-rescale ; modified Berendsen thermostat
tc-grps       = DNA_molecule Water_and_ions ; two coupling groups - more
accurate
tau_t         = 0.1 0.1 ; time constant, in ps
ref_t         = 300 300 ; reference temperature, one for each group, in K
; Pressure coupling
pcoupl        = no ; no pressure coupling in NVT
; Periodic boundary conditions
pbc           = xyz ; 3-D PBC
DispCorr      = no
; Velocity generation
gen_vel       = yes ; assign velocities from Maxwell distribution
gen_temp      = 300 ; temperature for Maxwell distribution
gen_seed      = -1 ; generate a random seed

```

---

For this step, first a simulation tpr file (nvt.tpr) is created with the grompp command using the mdp file, initial coordinate file, reference coordinate file, topology file, and index file. For this operation, the following command is entered into the terminal:

```
"gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -n index.ndx -o nvt.tpr"
```

Then start the NVT simulation by calling mdrun.

```
"gmx mdrun -v -deffnm nvt"
```

- y. NPT is a simulation condition in MD simulations at a constant number (N), a constant pressure (P), and a constant temperature (T). This is another condition for monitoring the behavior of systems. For this, an MDP file is prepared specifying the parameters of the NPT simulation. As mentioned in the previous step, attention should be paid to the line "tc-grps = DNA\_molecule Water\_and\_ions" when preparing this file. In this thesis, the following parameters were preferred:

#### **npt.mdp** (Lemkul, 2019)

---

```
title = Aptamer-Target molecule complex NVT equilibration
define = -DPOSRES ; position restrain the Aptamer-Target molecule
; Run parameters
integrator = md ; leap-frog integrator
nsteps = 50000 ; 2 * 50000 = 100 ps
dt = 0.002 ; 2 fs
; Output control
nstenergy = 500 ; save energies every 1.0 ps
nstlog = 500 ; update log file every 1.0 ps
nstxout-compressed = 500 ; save coordinates every 1.0 ps
; Bond parameters
continuation = yes ; continuing from NVT
constraint_algorithm = lincs ; holonomic constraints
constraints = h-bonds ; bonds to H are constrained
lincs_iter = 1 ; accuracy of LINCS
lincs_order = 4 ; also related to accuracy
; Neighbor searching and vdW
cutoff-scheme = Verlet
```

```

ns_type      = grid      ; search neighboring grid cells
nstlist      = 20        ; largely irrelevant with Verlet
rlist        = 1.2
vdwtype      = cutoff
vdw-modifier  = force-switch
rvdw-switch  = 1.0
rvdw         = 1.2      ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype  = PME          ; Particle Mesh Ewald for long-range
electrostatics
rcoulomb     = 1.2
pme_order    = 4          ; cubic interpolation
fourierspacing = 0.16    ; grid spacing for FFT
; Temperature coupling
tcoupl       = V-rescale   ; modified Berendsen thermostat
tc-grps      = DNA_LIG Water_and_ions ; two coupling groups - more
accurate
tau_t        = 0.1 0.1    ; time constant, in ps
ref_t        = 300 300    ; reference temperature, one for each
group, in K
; Pressure coupling
pcoupl       = Berendsen   ; pressure coupling is on for NPT
pcoupltype   = isotropic   ; uniform scaling of box vectors
tau_p        = 2.0        ; time constant, in ps
ref_p        = 1.0        ; reference pressure, in bar
compressibility = 4.5e-5    ; isothermal compressibility of
water, bar^-1
refcoord_scaling = com
; Periodic boundary conditions
pbc          = xyz        ; 3-D PBC
DispCorr     = no
; Velocity generation
gen_vel      = no        ; velocity generation off after NVT

```

---

For this step, first a simulation tpr file (npt.tpr) is created with the grompp command using the mdp file, initial coordinate file, reference coordinate file, topology file, and index file.

For this operation, the following command is entered into the terminal:

```
"gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -r nvt.gro -p topol.top -n
index.ndx -o npt.tpr"
```

Then start the NPT simulation by calling mdrun.

```
"gmx mdrun -v -deffnm npt"
```

- z. At the end of the NVT and NPT equilibration stages, the system was brought to an equilibrium state at the specified temperature and pressure. The position constraints will be released at this stage, allowing the initiation of production MD simulations to collect data. In this thesis, MD simulations with a duration of 10 nanoseconds were performed. For this, an MDP file is prepared specifying the parameters of the MD simulation. As mentioned in the previous step, the line "tc-grps = DNA\_molecule Water\_and\_ions" should be considered when preparing this file. The following parameters are preferred in this thesis:

#### **md.mdp** (Lemkul, 2019)

---

```
title = Aptamer-Target molecule complex MD simulation
; Run Parameters
integrator      = md      ; Leap-frog integrator
nsteps         = 5000000  ; 2 * 5000000 = 10000 ps (10 ns)
dt             = 0.002    ; 2 fs
; Output Control
nstenergy      = 5000     ; Save energies every 10.0 ps
nstlog         = 5000     ; Update log file every 10.0 ps
nstxout-compressed = 5000  ; Save coordinates every 10.0 ps
; Bond Parameters
continuation   = yes      ; Continuing from NPT
constraint_algorithm = lincs ; Holonomic constraints
constraints    = h-bonds  ; Bonds to H are constrained
lincs_iter     = 1 ; Accuracy of LINCS
lincs_order    = 4      ; Also related to accuracy
; Neighbor Searching and vdW
cutoff-scheme  = Verlet
ns_type        = grid    ; Search neighboring grid cells
nstlist        = 20      ; Largely irrelevant with Verlet
rlist          = 1.2
```

```

vdwtype      = cutoff
vdw-modifier  = force-switch
rvdw-switch   = 1.0
rvdw = 1.2          ; Short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype   = PME          ; Particle Mesh Ewald for long-range
electrostatics
rcoulomb      = 1.2
pme_order     = 4            ; Cubic interpolation
fourierspacing = 0.16       ; Grid spacing for FFT
; Temperature Coupling
tcoupl       = V-rescale     ; Modified Berendsen thermostat
tc-grps      = DNA_LIG Water_and_ions ; Two coupling groups - more accurate
tau_t = 0.1 0.1           ; Time constant, in ps
ref_t       = 300 300       ; Reference temperature, one for each group, in
K
; Pressure Coupling
pcoupl       = Parrinello-Rahman ; Pressure coupling is on for NPT
pcoupltype   = isotropic     ; Uniform scaling of box vectors
tau_p        = 2.0          ; Time constant, in ps
ref_p        = 1.0          ; Reference pressure, in bar
compressibility = 4.5e-5; Isothermal compressibility of water, bar^-1
; Periodic Boundary Conditions
pbc          = xyz          ; 3-D PBC
DispCorr     = no
; Velocity Generation
gen_vel      = no          ; Continuing from NPT equilibration

```

---

For this step, first a simulation tpr file (md\_0\_10.tpr) is created with the grompp command using the mdp file, initial coordinate file, reference coordinate file, topology file, and index file. For this operation, the following command is entered into the terminal:

```
"gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -n index.ndx -o md_0_10.tpr"
```

Then start the MD simulation by calling mdrun.

```
"gmx mdrun -v -deffnm md_0_10"
```

## 2.5. $\Delta G$ calculation of target molecule and the aptamer

$\Delta G$  values are a good choice to obtain information about the thermodynamic stability of the aptamer-target molecule complex. In this project, we will calculate  $\Delta G$  values after MD simulation with Gromacs using GMX\_MMPBSA. GMX\_MMPBSA calculates the energy changes after simulation using the MMPBSA approach. For GMX\_MMPBSA to work, post-simulation coordinate file, topology file, energy file and MMPBSA parameter files are required. When calculating  $\Delta G$  using GMXPBSA, there are different approaches. Two of them are C2 and Interaction Entropy. In this thesis, results will be obtained with the Interaction Entropy approach. For this, a mmpbsa.in file is prepared with the following content:

### **mmpbsa.mdp**

---

```
&general sys_name="IE_entropy",
startframe=1,
endframe=1001,
forcefields="oldff/leaprc.ff99SB,leaprc.gaff",
interaction_entropy=1,
temperature=300,
/
&gb
igb=8, saltcon=0.150, intdiel=10,
/
```

---

After preparing the parameter file, enter the following command to the terminal:

```
"gmx_MMPBSA -O -i mmpbsa.in -cs md_0_10.tpr -ci index.ndx -cg 1 2 -ct
md_0_10.xtc -cp topol.top"
```

## **2.6. RMSD and RMSF Calculations**

Following the MD simulations performed with GROMACS, Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) analyses are important to evaluate structural changes and dynamics. These analyses allow a detailed examination of the conformational changes and stability of aptamer and hormone molecules during the simulation process. RMSD and RMSF plots provide valuable information about the overall stability of the simulation and the nature of the molecular interactions.

### **2.6.1. RMSD Analysis**

RMSD analysis is a metric that measures how far atoms move away from their initial conformation over time in the simulation process. This analysis is used to track structural changes in the aptamer, hormone, or aptamer-hormone complex. Using the trajectory file obtained at the end of the simulation, the RMSD calculation is performed with the "gmx rms" command. This command calculates the root of the mean square difference between the selected groups of atoms over time.

```
"gmx rms -s md_0_10.tpr -f md_0_10.xtc -o rmsd.xvg -tu ns"
```

This command produces an "xvg" file containing the RMSD values with time. The resulting file is used for graph plotting.

### **2.6.2. RMSF Analysis**

RMSF analysis measures how much the atoms in the aptamer chains or hormone molecules fluctuate over the course of the simulation. This analysis provides information about the flexibility and dynamics of specific regions of the aptamer, or hormone. The "gmx rmsf" command is used to calculate the mean square fluctuations of atoms over the trajectory. This command averages the fluctuation over time for each atom.

```
"gmx rmsf -s md_0_10.tpr -f md_0_10.xtc -o rmsf.xvg -res"
```

This results in an "xvg" file containing the RMSF values for each residue. This file is used to plot graphs showing which parts of the aptamer or hormone molecule are more flexible or more rigid.

### ***2.7. Visualization of target molecule and the aptamer interaction***

The result of the MD simulation is played back using the USCF Chimera bioinformatics tool. In the MD movie feature provided by this tool, the md\_0\_10.xtc and md\_0\_10.tpr files were generated by mdrun. This process allows us to view and analyze the results of the MD simulation in a more meaningful and detailed way. At the end of 10 nanoseconds, to visualize the interaction of the target molecule and aptamer and the bonds formed, the last frame of the MD simulation is first saved with the "save PDB" feature of the USCF Chimera tool. The saved PDB file (1001st-frame.pdb) is uploaded to the Discovery Studio bioinformatics tool. The receptor-ligand interactions section of this tool is opened, and the aptamer is first defined as a "receptor" and the target molecule as a "ligand" to determine the interactions of interest. After the identification process, the interactions of the aptamer and the target molecule in water at the end of 10 nanoseconds are visualized in 2D with the "display ligand interactions" feature.

### ***2.8. Determination of possible mutation sites***

It was important to use MD simulations to see how the aptamer and target molecule interacts over time in aqueous environment and to predict the  $\Delta G$  values of the complexes. Following this step, *in-silico* mutagenesis experiments were performed to try to increase the affinity of the aptamer. We aimed to propose candidate mutation sites for this purpose. After MD simulations, a Python script was written to determine the three closest bases of the aptamer to the center of the target molecule at its position at the end of 10 nanoseconds-long simulation. This script was developed to study the interactions between the target molecule and the aptamer and to determine which bases could be candidates for the *in-silico* mutagenesis.



## find\_close\_bases.py

---

```
from Bio import PDB
import numpy as np
import pandas as pd

# Loading the PDB file and getting its information
def get_pdb_info(pdb_file):
    parser = PDB.PDBParser(QUIET=True)
    structure = parser.get_structure('pdb', pdb_file)
    model = structure[0]
    ligand_atoms = []
    dna_atoms = []

    for chain in model:
        for residue in chain:
            if "UNK" in residue.resname or "LIG" in residue.resname: #
Checking the residue name to find the target molecule
                ligand_atoms.extend(residue.get_atoms())

            if chain.get_id() != 'UNK' and chain.get_id() != 'LIG':
                dna_atoms.extend(residue.get_atoms())

    return ligand_atoms, dna_atoms

# Function to find the center of the target molecule
def get_ligand_center(ligand_atoms):
    coords = [atom.get_coord() for atom in ligand_atoms]
    center = np.mean(coords, axis=0)
    return center

# Function to calculate the distance from the center of the DNA aptamer
def distance_to_center(atom_coords, center):
    return np.linalg.norm(atom_coords - center)

# To specify the target PDB file
pdb_file_path = "md_0_10_last_frame.pdb"

# Getting target molecule and DNA aptamer information from PDB file
ligand_atoms, dna_atoms = get_pdb_info(pdb_file_path)
```

```

# Finding the center of the target molecule
ligand_center = get_ligand_center(ligand_atoms)
# To rank the bases of the DNA aptamer according to the distance from the
center of the ligand
distances = [(atom, distance_to_center(atom.get_coord(), ligand_center)) for
atom in dna_atoms]
distances.sort(key=lambda x: x[1])

# Create a list to store the results
results = []

# Printing the bases of the 3 nearest DNA aptamers
print("3 nearest DNA aptamers closest to Ligand:")
seen_resseqs = set()
count = 0
for atom, dist in distances:
    resseq = atom.parent.id[1]
    if resseq not in seen_resseqs:
        seen_resseqs.add(resseq)
        result = {
            "Residue Number": resseq,
            "DNA Base": atom.parent.resname,
            "Distance": dist
        }
        results.append(result)
        count += 1
    if count >= 3:
        break

# Create a DataFrame from the results
df = pd.DataFrame(results)

# Save the DataFrame to an Excel file
output_excel_file = "md_0_10_last_frame.xlsx"
df.to_excel(output_excel_file, index=False)

print(f"Results saved to {output_excel_file}")

```

---

## 2.9. *In-silico* mutagenesis of the three closest bases in PYMOL

In this part of the thesis, different mutant aptamer structures are made by changing the bases in the aptamer sequence, which means adding new nucleotides in their place. The three aptamer bases from the previous step are used for *in-silico* mutagenesis. These are the ones that are closest to the center of the target molecule after MD. There are four different possible nucleotide options for each residue, making a total of  $4^3$  or 64 different combinations. One of these 64 combinations is the order in the original aptamer structure, which means that there are a total of 63 different mutant aptamer options. These combinations represent different structures of the aptamer sequence, and these structures are important to study because they can affect the interaction and affinity of the aptamer with the target molecule. To achieve this in this thesis, the nucleotide mutagenesis feature of the Pymol bioinformatics tool was used. A Python script was written for this process. This script converts the nucleotide at that position to the targeted base by means of the residue and new base information of the base to be mutagenized *in-silico* and saves this change in a new PDB file. This step results in a total of 63 different mutant aptamer PDB files.

### **In\_silico\_mutagenesis.py**

---

```
from enum import Enum
import pymol
from pymol import cmd
pymol.finish_launching()

# All combinations other than original (TTC) for mutation
class Base(Enum):
    Adenine = "Adenine"
    Cytosine = "Cytosine"
    Guanine = "Guanine"
    Thymine = "Thymine"

mutations = []

for i in Base:
    for j in Base:
        for k in Base:
```

```

        if i.value != "Thymine" or j.value != "Thymine" or k.value !=
"Cytosine":
            mutations.append((i.value, j.value, k.value))

# Create a PDB file for each mutation
for i, mut in enumerate(mutations):
    cmd.load("aptamer.pdb")
    cmd.wizard("nucmutagenesis")
    cmd.refresh_wizard()

    # Mutation for residue #
    cmd.get_wizard().do_select("resi #")
    cmd.get_wizard().set_mode(mut[0])
    cmd.get_wizard().apply()

    # Mutation for residue #
    cmd.get_wizard().do_select("resi #")
    cmd.get_wizard().set_mode(mut[1])
    cmd.get_wizard().apply()
    # Mutation for residue #
    cmd.get_wizard().do_select("resi #")
    cmd.get_wizard().set_mode(mut[2])
    cmd.get_wizard().apply()
    cmd.set_wizard()

    # Use mutation name as file name
    mutation_name = "_".join(mut)
    cmd.save(f"mutated_structure_{mutation_name}_{i+1}.pdb")

    # Clear current structure for next mutation
    cmd.delete("all")
pymol.cmd.quit()

```

---

## 2.10. Docking with mutant aptamers

Docking studies were carried out at this stage to evaluate the interaction of 63 different mutant aptamers for an aptamer obtained by making base changes in the genetic structure of an aptamer in the *in-silico* mutagenesis step with the target molecule of the original aptamer. The primary goal of this step was to identify the mutant aptamers that would have the strongest interaction with the target molecule and have the highest binding strength.

At this stage, the procedure in 2.3. was followed for the docking of mutant DNA aptamers and small molecules, and this step was automated with a Python script. Unlike the procedure in 2.3, site-specific docking was performed instead of blind docking. The coordinates of the C3 atom in the center of the ligand in the last frame of the MD simulation were taken with Discovery Studio software, determined as the x, y, and z coordinates of the grid box, and given as input to the script.

### Automated\_docking\_with\_Vina.py

---

```
import subprocess
import os

# AutoDock Vina's path
vina_path = r"C:\the_path\Vina.exe"

# Path to the Estradiol PDBQT file
estradiol_pdbqt = r"C:\the_path\estradiol.pdbqt"

# Folder with mutant aptamer PDB files
aptamer_folder = r"C:\the_path\aptamer_folder"

# Folder to save the results
output_folder = r"C:\the_path\output_folder"

# Docking parameters
center_x, center_y, center_z = 42.4, 47.49, 39.9
size_x, size_y, size_z = 50, 50, 50
```

```

exhaustiveness = 128

# Loop for all mutant aptamers
for aptamer_file in os.listdir(aptamer_folder):
    if aptamer_file.endswith(".pdb"):
        aptamer_path = os.path.join(aptamer_folder, aptamer_file)
        output_file_name = "out_" + os.path.splitext(aptamer_file)[0] +
        ".pdbqt"
        output_path = os.path.join(output_folder, output_file_name)
        log_file_name = "log_" + os.path.splitext(aptamer_file)[0] +
        ".txt"
        log_path = os.path.join(output_folder, log_file_name)

        # Creating the AutoDock Vina command
        vina_command = [
            vina_path, "--receptor", aptamer_path, "--ligand",
            estradiol_pdbqt,
            "--center_x", str(center_x), "--center_y", str(center_y), "--
            center_z", str(center_z),
            "--size_x", str(size_x), "--size_y", str(size_y), "--size_z",
            str(size_z),
            "--exhaustiveness", str(exhaustiveness), "--out", output_path,
            "--log", log_path
        ]

        # Start the docking process
        subprocess.run(vina_command)

print("Docking operations and log records are complete.")

```

---

### **2.11. MD simulation of mutant aptamer with the best dock**

To evaluate in detail, the interaction between the mutant aptamers, which have the strongest interaction with the target molecule and have the highest binding strength, and the target molecules and to make a comparison with the original aptamer, MD simulations are performed between the mutant aptamers selected at this stage and the target molecules. The results of MD simulations contain thermodynamic and energetic information. In this context, the energy profiles and thermodynamic parameters obtained from the simulation results are used for  $\Delta G$  analysis. The obtained MD simulation results are analyzed in comparison with the original aptamer, by determining the interaction strengths of the mutant aptamers and the target molecules through  $\Delta G$  analysis. This analysis helps us to understand the interaction mechanisms of mutant aptamer with target molecule and sheds light on whether the mutant aptamer provides higher affinity than the original aptamer.

At this stage, the procedure in 2.4. was followed for MD simulation of mutant DNA aptamer and target molecule and the procedure in 2.5. & 2.6. were followed for  $\Delta G$ , RMSD-RMSF calculations of mutant DNA aptamer and target molecule.

## CHAPTER 3: RESULTS & DISCUSSION

### 3.1. Structural and affinity analysis

In a 2014 study by Alsager et al., a specific 75-mer aptamer for 17 $\beta$ -estradiol was selected by SELEX and shown to be capable of low nanomolar detection of 17 $\beta$ -estradiol (Alsager et al., 2014). In a follow-up study published in 2015, truncated versions of this 75-mer aptamer (35-mer and 22-mer) were developed and these truncated aptamers showed 25-fold higher sensitivity for 17 $\beta$ -estradiol detection than the original (Alsager et al., 2015). This improvement is due to the enhancement of signal transduction by removing excess nucleotides. These results emphasize the importance of short aptamer sequences in biosensors. In a 2017 study by Svobodová et al. to develop a rapid and high-throughput technique for the analysis of aptamers interacting with small molecules, aptamers reported against the hormone 17 $\beta$ -estradiol were used as a model system. The study confirms that this aptamer interacts with estradiol, but not with androstenedione, progesterone, and testosterone (Svobodová et al., 2017).

In this thesis, Alsager22, which was truncated in the article, was chosen as the aptamer to be subjected to structural and affinity analysis followed by *in-silico* mutagenesis and optimization studies. Alsager22 aptamer has the sequence “GCCGTTTGGGCCCAAGTTCGGC” and shows high affinity especially for estradiol, but minimal affinity for other tested hormones (testosterone, progesterone, and androstenedione).



### **3.1.1. Obtained 2D aptamer structures**

In this part of the thesis, the results of the prediction of the 2D structure of an aptamer named “Alsager22” are discussed. The UNAFold web server was used for 2D structure prediction. UNAFold predicts the possible secondary structures of nucleic acid sequences based on energy minimization principles. UNAFold determined the structures with the lowest free energy from the nucleotide sequence of Alsager22. The results include several structural configurations indicating possible secondary structures. These structures clearly reveal the binding sites between bases and hairpins of the aptamer. The prediction of the 2D structure in UNAFold was performed at a temperature of 37°C and ion concentrations of  $[Na^+] = 10.5 \text{ mM}$  &  $[Mg^{++}] = 0.3 \text{ mM}$ . The first of the predicted structural configurations is labeled as “Structure 1”, and its  $\Delta G$  value was calculated as  $-1.62 \text{ kcal/mol}$ . (Figure 2) The 2D structure of Structure 1 is shown in DBN format as “((((.....))))”. The second structure configuration is labeled “Structure 2”, and the  $\Delta G$  value is calculated as  $-1.22 \text{ kcal/mol}$ . (Figure 3) The 2D structure of Structure 2 is shown in DBN format as “(((.....))).....”. The final structure configuration is labeled “Structure 3”, and the  $\Delta G$  value is calculated as  $-0.73 \text{ kcal/mol}$ . (Figure 4) The 2D structure of Structure 3 is shown in DBN format as “(((((((.....))))))”. These energy values help in understanding the relationship between the different structural configurations of Alsager22. While Structure 1 and 2 showed singular hairpin topology, Structure 3 had a hairpin and an internal loop at 5, 16, 17 and 18th residues. While Structure 1 showed lower  $\Delta G$ , the complexity of Structure 3 and lower  $\Delta S$  lead us to continue our studies using the former. Due to complexity of the structure, it is expected to have more rigid binding sites towards target molecules.

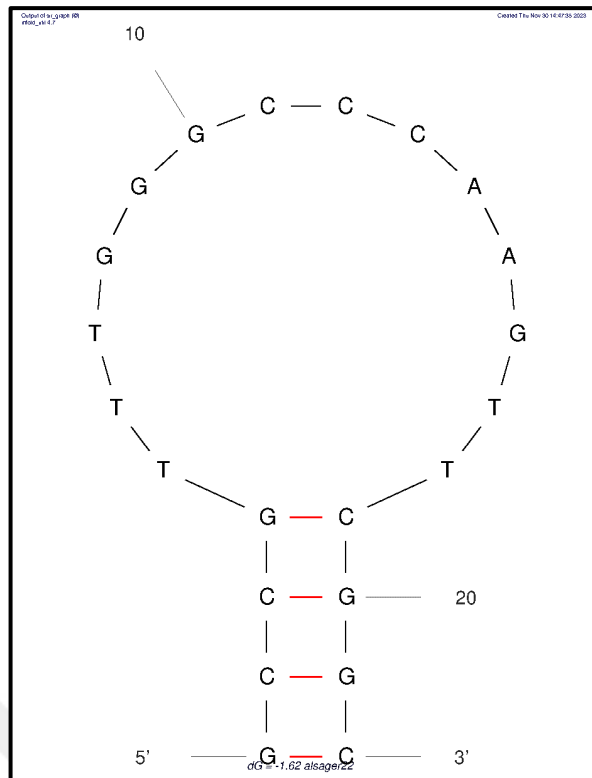


Figure 2. The first of the predicted structural configurations.

$\Delta G = -1.62$  kcal/mol at  $37^\circ\text{C}$ ,  $\Delta H = -38.90$  kcal/mol,  $\Delta S = -120.2$  cal/(K·mol),  $T_m = 50.4^\circ\text{C}$  assuming a 2-state model, linear DNA folding, Ionic conditions:  $[\text{Na}^+] = 0.0105$  M,  $[\text{Mg}^{++}] = 0.0003$  M

(Standard errors are roughly  $\pm 5\%$ ,  $\pm 10\%$ ,  $\pm 11\%$  and  $2\text{-}4^\circ\text{C}$  for free energy, enthalpy, entropy, and  $T_m$ , respectively.)

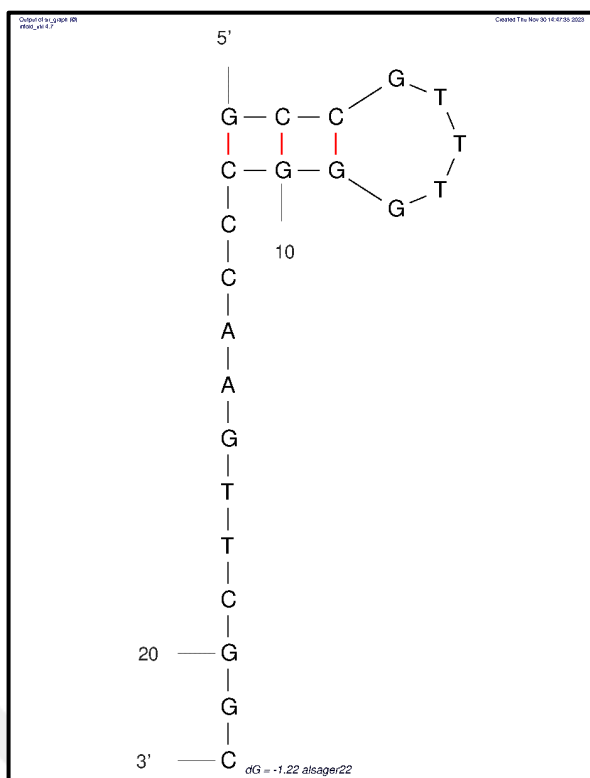


Figure 3. The second of the predicted structural configurations.

$\Delta G = -1.22$  kcal/mol at 37 °C,  $\Delta H = -38.00$  kcal/mol,  $\Delta S = -118.5$  cal/(K·mol),  $T_m = 47.2$  °C assuming a 2-state model, linear DNA folding, Ionic conditions:  $[Na^+] = 0.0105$  M,  $[Mg^{++}] = 0.0003$

(Standard errors are roughly  $\pm 5\%$ ,  $\pm 10\%$ ,  $\pm 11\%$  and 2-4 °C for free energy, enthalpy, entropy, and  $T_m$ , respectively.)

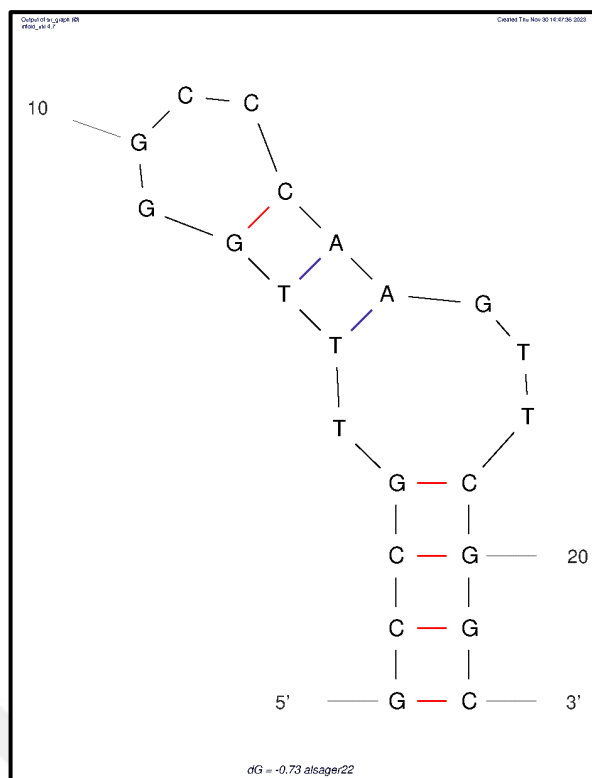


Figure 4. The third of the predicted structural configurations.

$\Delta G = -0.73$  kcal/mol at 37 °C,  $\Delta H = -51.10$  kcal/mol,  $\Delta S = -162.4$  cal/(K·mol),  $T_m = 41.4$  °C assuming a 2-state model, linear DNA folding, Ionic conditions:  $[Na^+] = 0.0105$  M,  $[Mg^{++}] = 0.0003$  M

(Standard errors are roughly  $\pm 5\%$ ,  $\pm 10\%$ ,  $\pm 11\%$  and 2-4 °C for free energy, enthalpy, entropy, and  $T_m$ , respectively.)

The choice of Structure 3 among the three different structural configurations, based on the prediction results for its 2D structure, is a critical decision point. The main reasons for this choice are structural complexity, low entropy, and topological properties. The structure of Structure 3 is complex, incorporating a hairpin and an inner loop. This complexity may allow the aptamer to bind more specifically and strongly to target molecules. Among others, the existence of an internal loop in Structure 3 could provide additional sites to bind with the target molecule that are considerably more stable and selective. A smaller  $\Delta S$  value means that Structure 3 is more ordered and rigid. This may make the interactions with target molecules more stable. High stability means that the aptamer will only need to undergo fewer conformational changes in its encounter with, or reaction with, target molecules, which may make binding more efficient.

In addition to all this, it was surprising that the  $\Delta G$  value calculated by UNAFold of the 3rd structure, which was expected to be more stable, was higher compared to the other two structures. Although the  $\Delta G$  values for Structure 1 and Structure 2 are lower than those for Structure 3, the unique topological characteristics of Structure 3 make it advantageous. These topological features may be conducive to more effective interaction, especially when considering the structure and binding requirements of a target molecule. In addition, model 3 was consistent with the 2D structure favored in the research article (Svobodová et al., 2017).

### **3.1.2. *Obtained 3D aptamer structures***

This part of the thesis details how the 3D structure of the Alsager22 aptamer was predicted and the details of this process. As described in the previous chapter, three different 2D structures of the Alsager22 aptamer were obtained using Mfold software. For the proceeding studies, we have chosen the “Structure 3” since this structure showed lowest  $\Delta S$  and highest  $T_m$ . Consequently, this structure was also predicted in the original literature. The DBN representation of this structure, “(((((((.....))))))..))”, was used for the prediction.

XiaoLab’s 3DRNA/DNA service was the preferred choice for creating the 3D structure. This service was developed to predict the 3D modeling of nucleic acid structures. The prediction process started by determining the molecule type of the Alsager22 aptamer as DNA and selecting the optimization procedure as the structure generation procedure. Then, the nucleotide sequence of the aptamer and its 2D structure, represented in DBN format, were input into the system. This information formed the basis of the 3D modeling process. The system predicted five different 3D structures based on the input information. Each prediction represents the possible 3D conformations of the Alsager22 aptamer. The system performed a minimization process for each of the structures obtained. Minimization is a process to optimize the energy levels of 3D structures and find the most stable conformation without changing overall structure. This step will minimize the final structures predicted by AMBER using the ff14SB force field, and after this process, the atom conflict will be eliminated.

Five models were obtained from XiaoLab's 3DRNA/DNA service. (Figure 5) Unfortunately, these models were provided with no score, or energy parameters associated with the models. When superimposed these structures showed minimal variation. (Figure 6) The 18th Thymine bulge was noticeable in five models. All models retained the hairpin topology, however, the Thymine at position 5 and Thymine at position 17 pose in a paired manner. According to all models, N3 of Thymine 5 forms a H-bond with O connected to C4 of Thymine 17, and N3 of Thymine 17 forms a H-bond with O connected to C2 of Thymine 5. (Figure 7) It is important to note this new pairing is previously undocumented to our knowledge, and its energetic favorability may be revealed during MD studies.



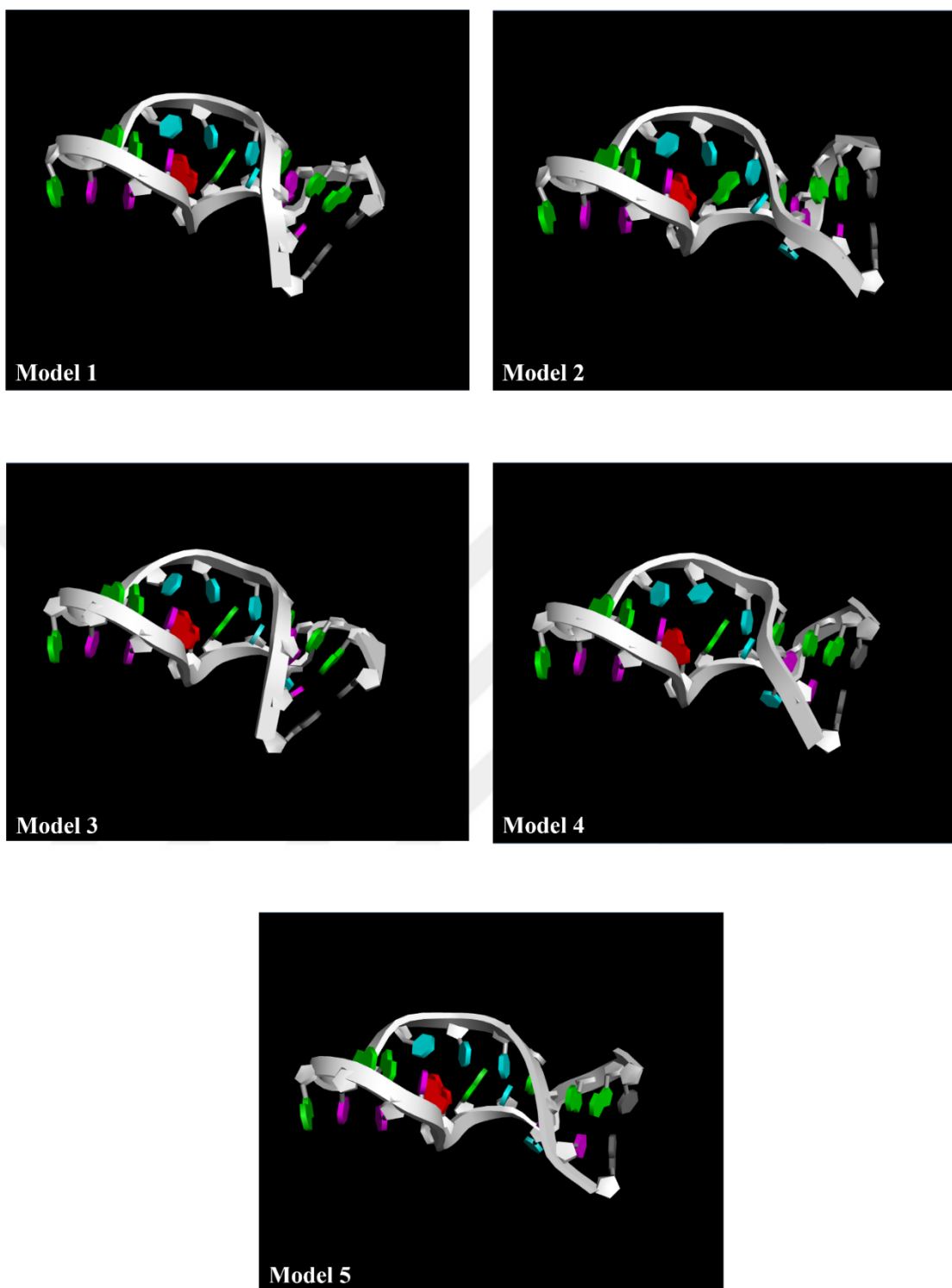


Figure 5. Visualization of 5 different structures obtained in XiaoLab 3dRNA/DNA structure prediction software.

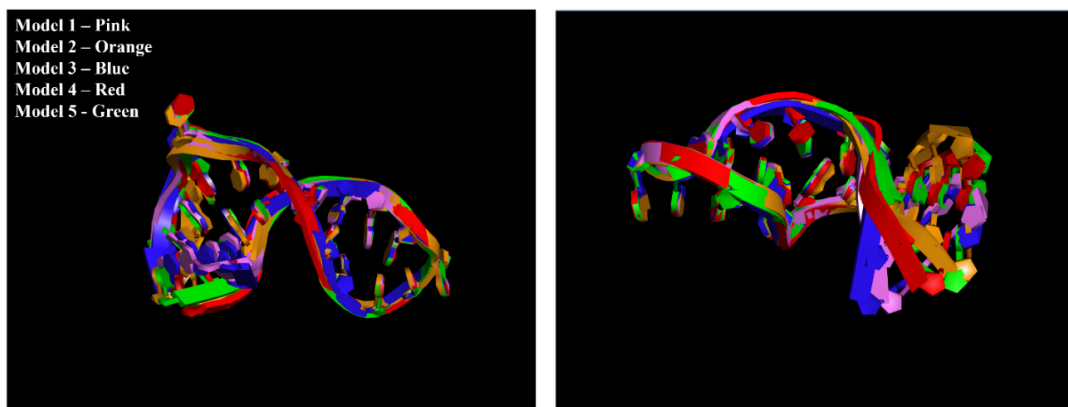


Figure 6. Superimposed visualization of 5 different structure predictions obtained from XiaoLab 3dRNA/DNA service.

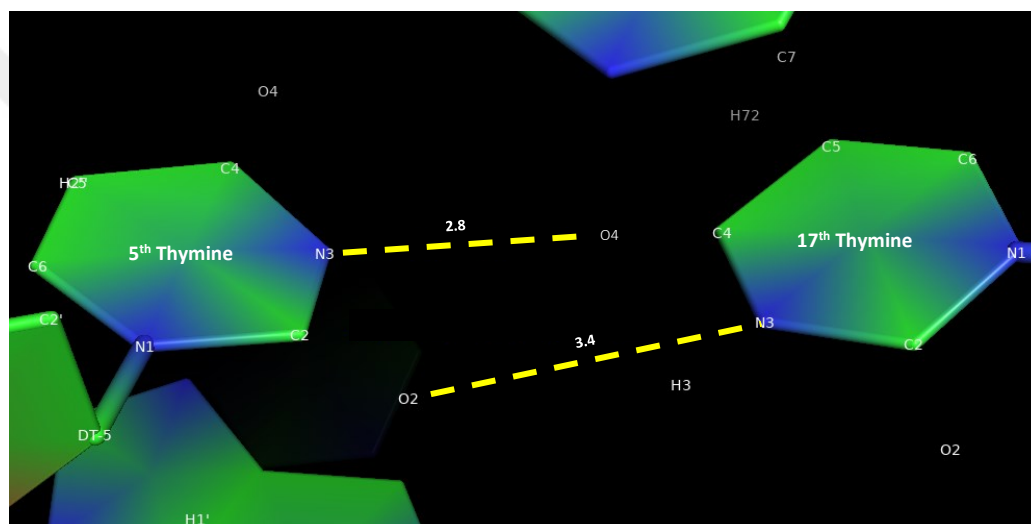


Figure 7. Distance between N3 of Thymine 5 - O4 of Thymine 17 and O2 of Thymine 5 - N3 of Thymine 17. The unit is Armstrong.



The five different models obtained in the process of predicting the 3D structure of the Alsager22 aptamer resulted in an important observation: In general, the 5<sup>th</sup> and 17<sup>th</sup> T bases in all models face each other and could potentially form an H-bond. This structural feature may have a profound effect on the 3D conformation and function of the aptamer. The small discrepancy among these models may suggest that the 3D structural integrity and stability of the Alsager22 aptamer can be maintained under various circumstances. This suggests that the aptamer in fact has a stable structure within living organisms, and such structural characteristics can be quite critical, not only for the recognition of targets but also for binding to them. This novel and thus far undocumented T-T pairing might be more understandable during MD simulation studies of the aptamer. Such an energetically favored pairing might reveal a neglected or misleadingly understood kind of interaction. The H-bond formed between the N atom (5T), and the O atom (17T) may also have significant effects on structure stability, or it could be very important to the interaction of the aptamer with its target molecule. The 18<sup>th</sup> Thymine bulge could be important to the interaction of the aptamer with its target molecule.

### ***3.1.3. Molecular docking with the aptamer and target molecules***

This part of the thesis presents the details of the molecular docking process and the obtained results. Firstly, the docking study on Model 1, selected from the 3D structure predictions, with the ligand estradiol, which is shown in the literature as the target of the Alsager22 aptamer, is examined. This study represents a fundamental step in enhancing the understanding of the interaction between the aptamer and its target molecule. Docking was performed using the AutoDock Vina. In this process, preparations were first made for the aptamer acting as a receptor and the estradiol accepted as a ligand. The grid box was set to encapsulate the whole structure to perform a blind docking. Torsion settings were made to determine the rotatable bonds of the ligand. No rotatable bonds were detected in 3 of the 4 hormones, with only one bond detected in progesterone. (Figure 8) Then only polar hydrogens were added to all molecules. Furthermore, non-specific targets, progesterone, testosterone, and androstenedione, were also docked as the off-target controls, i.e., four different studies in total. These additional studies are important to evaluate the potential interactions of the aptamer with off-target molecules. For each docking study, a configuration file containing information about the receptor, ligand, and grid box was prepared.

Finally, 4 different docking runs were performed using AutoDock Vina, with the exhaustiveness value set to 128 for each docking. These procedures were designed to allow a comprehensive study of the interactions with the target molecule. AutoDock Vina returns a log file with the results as output. The log file contains various binding energies under different “modes” and their distance from the “best mode”. Each mode represents a different conformation at the binding site. Of particular interest are the affinity values. These values are given in kcal/mol. Lower affinity values generally imply stronger binding.

The log file also contains “RMSD l.b.” and “RMSD u.b.” values, which express how far away each mode is from the 'best mode'. These values indicate the posing similarities between the different binding modes.

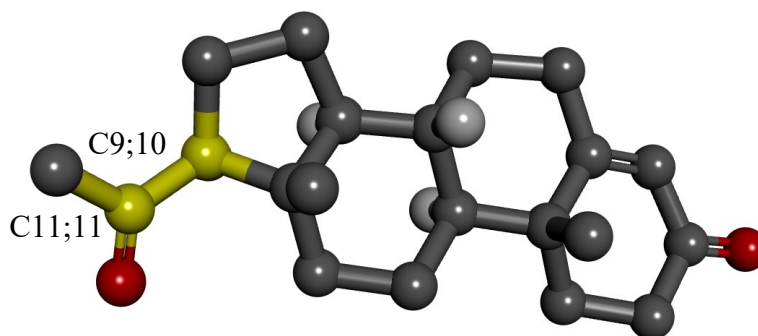


Figure 8. The rotatable bond in the progesterone shown in yellow.

Table 1. Affinity and RMSD Analysis of Alsager22-Ligands Dockings.

RECEPTOR	LIGAND	DOCKING MODE	AFFINITY (kcal/mol)	RMSD LOWER BOUND	RMSD UPPER BOUND
<b>alsager22</b>	<b>estradiol</b>	<b>1</b>	<b>-8.1</b>	<b>*</b>	<b>*</b>
alsager22	estradiol	2	-8	1.149	2.705
alsager22	estradiol	3	-8	1.529	6.072
alsager22	estradiol	4	-7	1.32	2.009
alsager22	estradiol	5	-6.9	2.039	6.469
alsager22	estradiol	6	-6.8	1.258	6.634
alsager22	estradiol	7	-6.8	1.912	3.529
alsager22	estradiol	8	-6.7	1.716	6.829
alsager22	estradiol	9	-6.6	2.715	8.219
<b>alsager22</b>	<b>progesterone</b>	<b>1</b>	<b>-8.1</b>	<b>*</b>	<b>*</b>
alsager22	progesterone	2	-8	1.68	7.073
alsager22	progesterone	3	-7.6	1.751	6.801
alsager22	progesterone	4	-7.3	1.442	7.171
alsager22	progesterone	5	-7.2	1.556	2.852
alsager22	progesterone	6	-7	1.995	3.067
alsager22	progesterone	7	-7	1.876	7.62
alsager22	progesterone	8	-6.6	11.902	13.184
alsager22	progesterone	9	-6.5	11.547	12.683
<b>alsager22</b>	<b>testosterone</b>	<b>1</b>	<b>-8.5</b>	<b>*</b>	<b>*</b>
alsager22	testosterone	2	-7.7	1.338	6.49
alsager22	testosterone	3	-7.6	1.236	2.706
alsager22	testosterone	4	-7.6	1.933	6.595
alsager22	testosterone	5	-7.1	2.018	6.323
alsager22	testosterone	6	-7	2.791	4.544
alsager22	testosterone	7	-6.7	12.141	14.394
alsager22	testosterone	8	-6.5	3.609	7.611
alsager22	testosterone	9	-6.4	2.869	6.633
<b>alsager22</b>	<b>androstenedione</b>	<b>1</b>	<b>-8.3</b>	<b>*</b>	<b>*</b>
alsager22	androstenedione	2	-8	1.089	1.223
alsager22	androstenedione	3	-7.8	1.423	3.149
alsager22	androstenedione	4	-7.7	1.259	6.495
alsager22	androstenedione	5	-7.5	1.93	6.617
alsager22	androstenedione	6	-7.5	1.34	2.815
alsager22	androstenedione	7	-7.2	1.615	6.157
alsager22	androstenedione	8	-7.1	2.336	6.515
alsager22	androstenedione	9	-7.1	2.693	4.436

A negative energy value indicates thermodynamic favorability of binding, with lower energy values generally corresponding to stronger bond formation between the receptor and ligand. At the molecular level, androstenedione exhibited a high binding affinity with a determined best binding value of -8.3 kcal/mol. Similarly, estradiol and progesterone exhibited a calculated value of -8.1 kcal/mol, indicating their strong

binding affinity to the receptors. Testosterone exhibited the highest binding affinity with a calculated value of -8.5 kcal/mol. (Table 1)

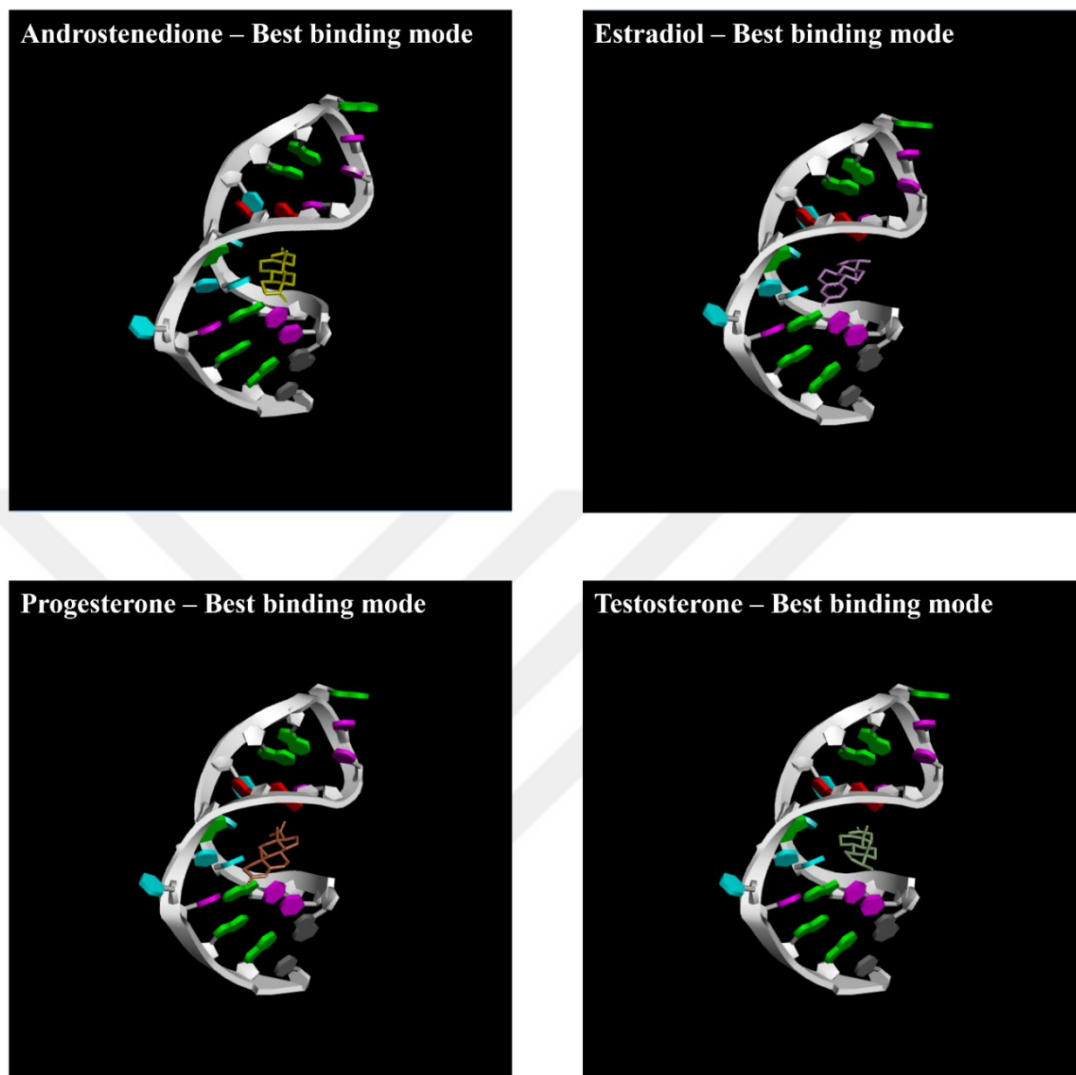


Figure 9. Visualization of the modes in which the hormones androstenedione, estradiol, progesterone, and testosterone dock with the aptamer with the highest affinity.

The molecular docking of the Alsager22 aptamer with estradiol and other hormones produced unexpected results. Noteworthy is the difference in affinity between estradiol and other non-specific targets such as progesterone, testosterone, and androstenedione. The Alsager22 aptamer in the literature is a target molecule of estradiol. The docking results show that estradiol has a lower affinity than other hormones. Although it is mentioned in the study that it is not the target of the aptamer, testosterone has the highest binding affinity value (-8.5 kcal/mol). Furthermore, the fact that all these hormones had the highest binding affinity to the same particular groove provides crucial insights into how the aptamer's structural properties effect these binding interactions. (Figure 9) The internal loop creates this groove, which is what causes the structure to bend. These results show that the way in which aptamers recognize their targets and determine how they interact with them cannot be assessed purely based on affinity values. It is important to note, docking results indicate the best possible mode of binding, but not whether such binding may occur in physiological conditions. Therefore, MD simulations are required for more reliable and comprehensive analysis.

#### ***3.1.4. Molecular dynamics simulations***

In this part of the thesis, MD simulations of aptamer with bound hormones are discussed. The MD simulations were performed starting from the docking results obtained in the previous section. The simulations were performed for 10 ns with 2 fs time step at 300 K temperature target. After completion, each simulation was visualized using Chimera software.

In case of Estradiol-Alsager22 simulations, Estradiol was found to be bound to the aptamer at the middle of the stem, except one simulation, where the compound was found unbound. (Figure 11) In case of Androstenedione-Alsager22 simulations, Androstenedione was found to be bound to the aptamer at positions closer to the middle of the stem, as well, except one simulation, where the compound was found unbound. (Figure 10) At the end of the first MD simulation, progesterone and aptamer are quite far from each other in terms of position. At the end of the second, third, and fourth simulations, they are close enough to form bonds, but at different points of the aptamer in all three. (Figure 12) At the end of the first and third MD simulations, testosterone and aptamer are quite far from each other in terms of position.

At the end of the second and fourth simulations, they are close enough to form a bond, but both are at a different point of the aptamer. (Figure 13)

In all MDs, the hairpin structure remained after 10 ns of simulation with marginal structural changes. 5<sup>th</sup> and 17<sup>th</sup> Thymine bases remained in H-bonding in most structures with 18<sup>th</sup> Thymine bulging. This unique conformation has been the main target for the Estradiol and Androstenedione. Despite that this formation remained in simulations with other ligands, they did not pose close to it at the end.

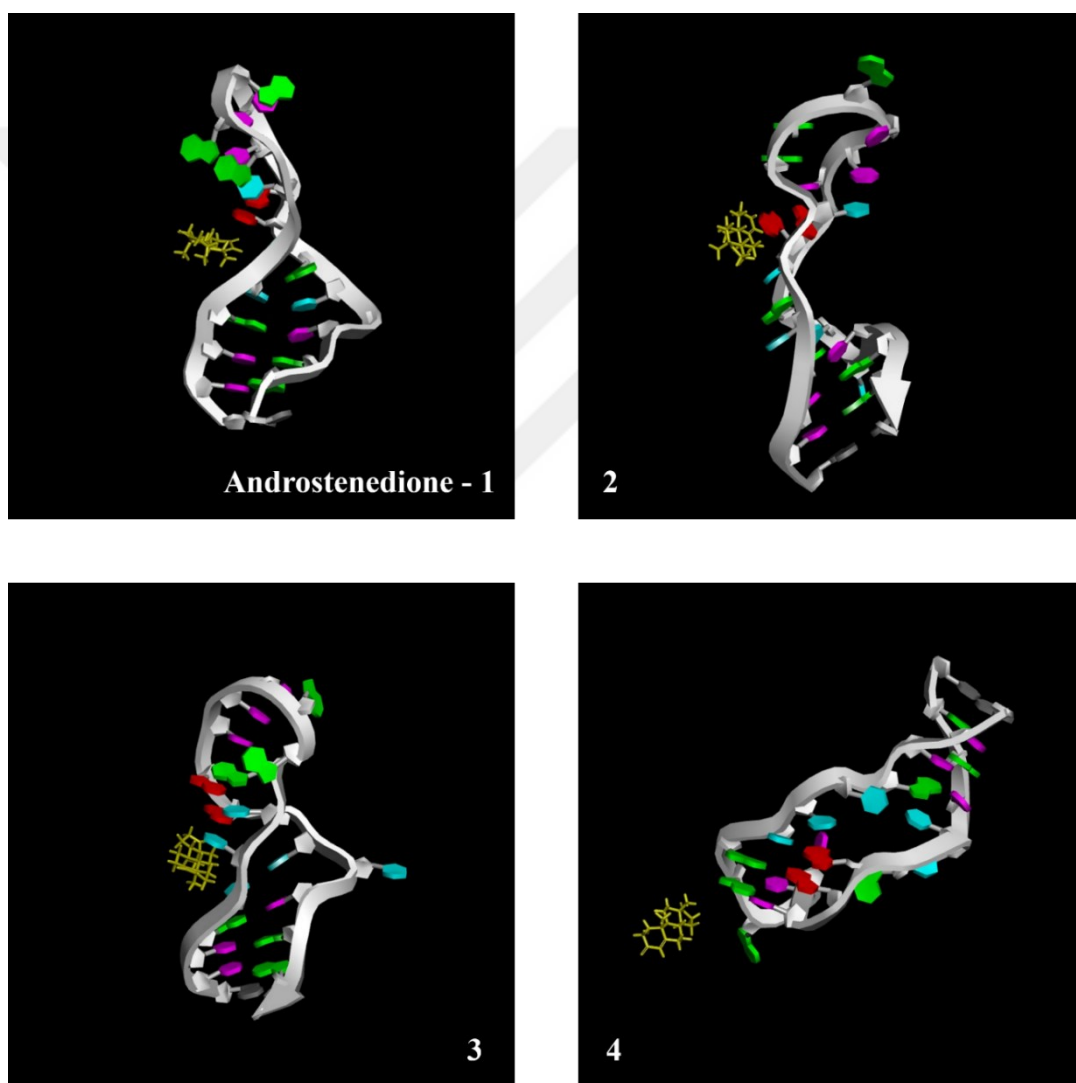


Figure 10. Positions of Alsager22 aptamer and androstenedione hormone after 10 nanoseconds of four different MD simulations.

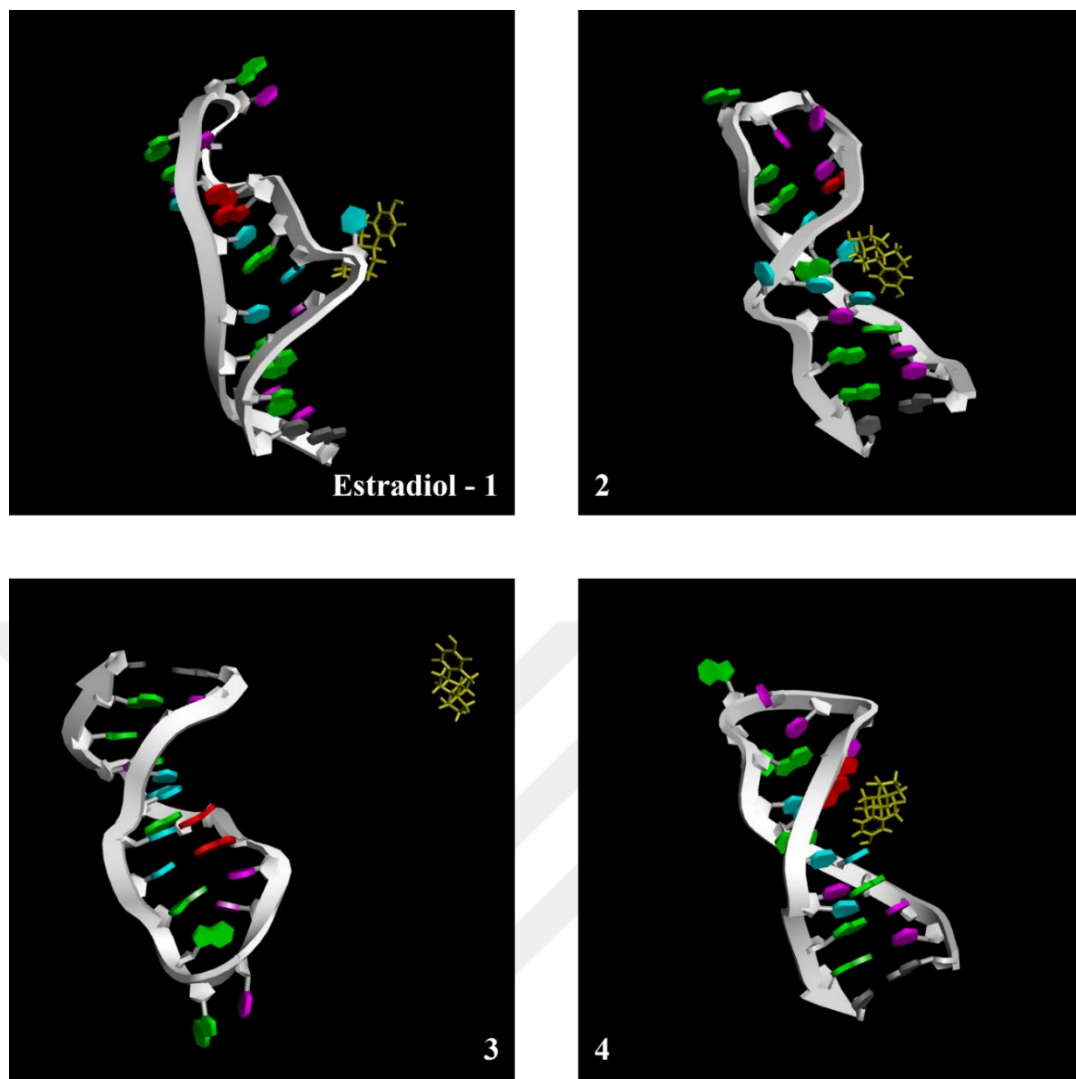


Figure 11. Positions of Alsager22 aptamer and estradiol hormone after 10 nanoseconds of four different MD simulations.

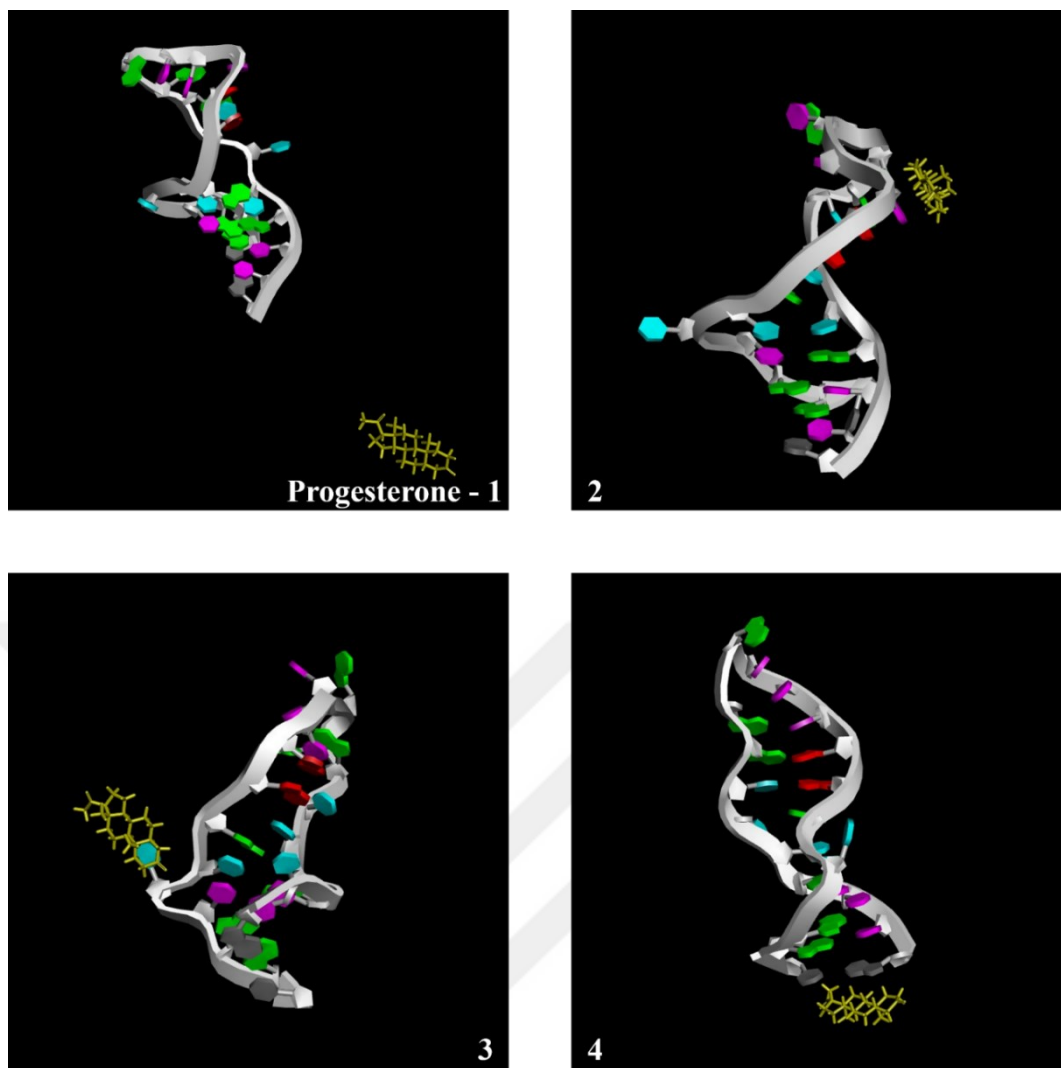


Figure 12. Positions of Alsager22 aptamer and progesterone hormone after 10 nanoseconds of four different MD simulations.



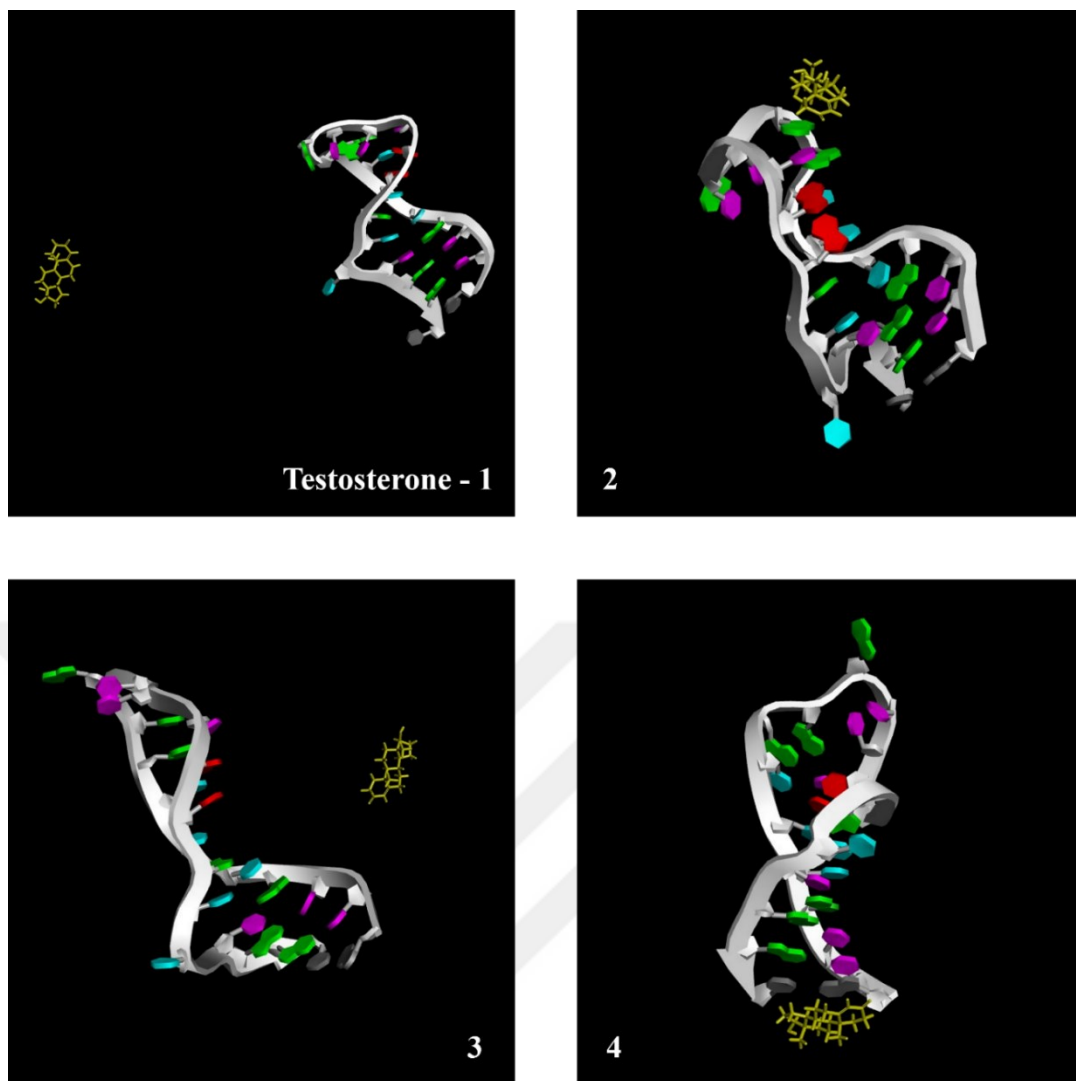


Figure 13. Positions of Alsager22 aptamer and testosterone hormone after 10 nanoseconds of four different MD simulations.

RMSD analysis after least square fit measurements of Alsager22 to itself over time has shown that within 2 ns, the DNA structure undergoes significant change and remains relatively unchanged after that. This indicates that predicted 3D structure has changed during early stages of the simulation. Past 2 ns, the structures have been relatively stable. (Figure 14B, 15B, 16B, 17B)

Fluctuation analysis showed increased fluctuations for DNA at the 3' and 5' ends. This is an expected result since these moieties are at the ends and more flexible. Dramatic increase of fluctuations at atoms around 300 and 550, which correspond to G10 and T18. This is expected for both the bulging Thymine at 18th position and the Guanine at position 10 which corresponds to the tip of the hairpin loop. (Figure 14D, 15D, 16D, 17D) In case of the ligand, fluctuations occur only at hydrogens that are connected through a methyl (Hydrogens 1, 2, 3) or hydroxyl group (Hydrogens 12 and 24). (Figure 14E, 15E, 16E, 17E) This too is expected since their carbon and oxygens are connected through rotatable bonds.

RMSD analysis for the ligand versus the DNA has been more informative regarding the interactions between the ligand and the DNA. For the MD simulations 1, 2 and 4, the ligand became stable on the DNA past the 4th ns. (Figure 14A and 15A) It is expected since the first 2 ns were the period for the DNA structure to fold into a stable state. (Figure 14B and 15B) In the case of the third simulation where the ligand was disconnected at the final pose, after the 3rd ns, the complex had been stable until 6th ns, after which the ligand was removed from the stable pose yet remained close to DNA until 9<sup>th</sup> ns. For this reason, the delta G measurements were done using the frames between 3rd and 6th ns for this simulation.

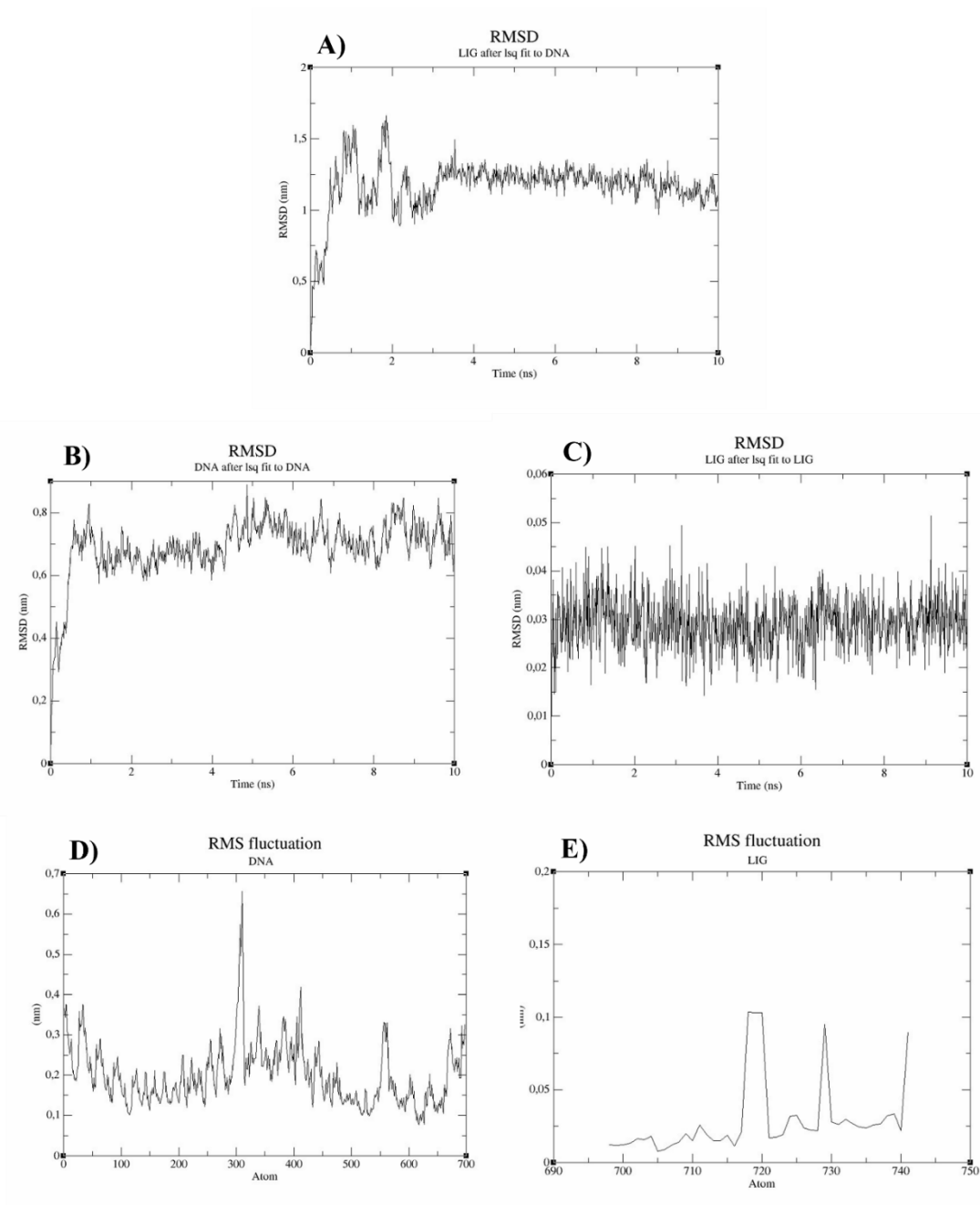


Figure 14. First of the MD simulation trials of Estradiol and Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Alsager22 over time. (B) Calculation of the RMSD values of Alsager22 after least square fit to Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Alsager22. (E) Calculation of RMSF values for Estradiol.

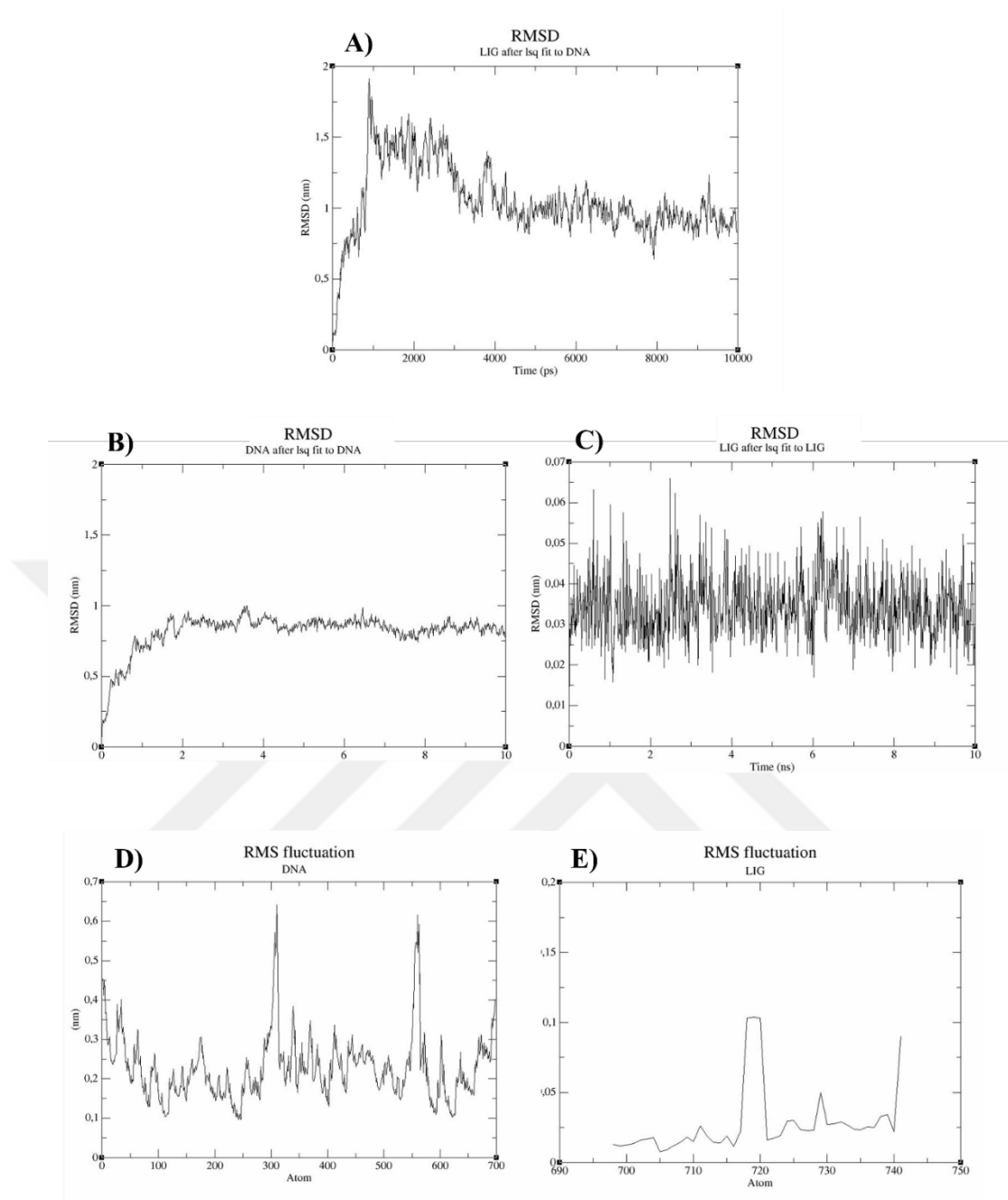


Figure 15. Second of the MD simulation trials of Estradiol and Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Alsager22 over time. (B) Calculation of the RMSD values of Alsager22 after least square fit to Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Alsager22. (E) Calculation of RMSF values for Estradiol.

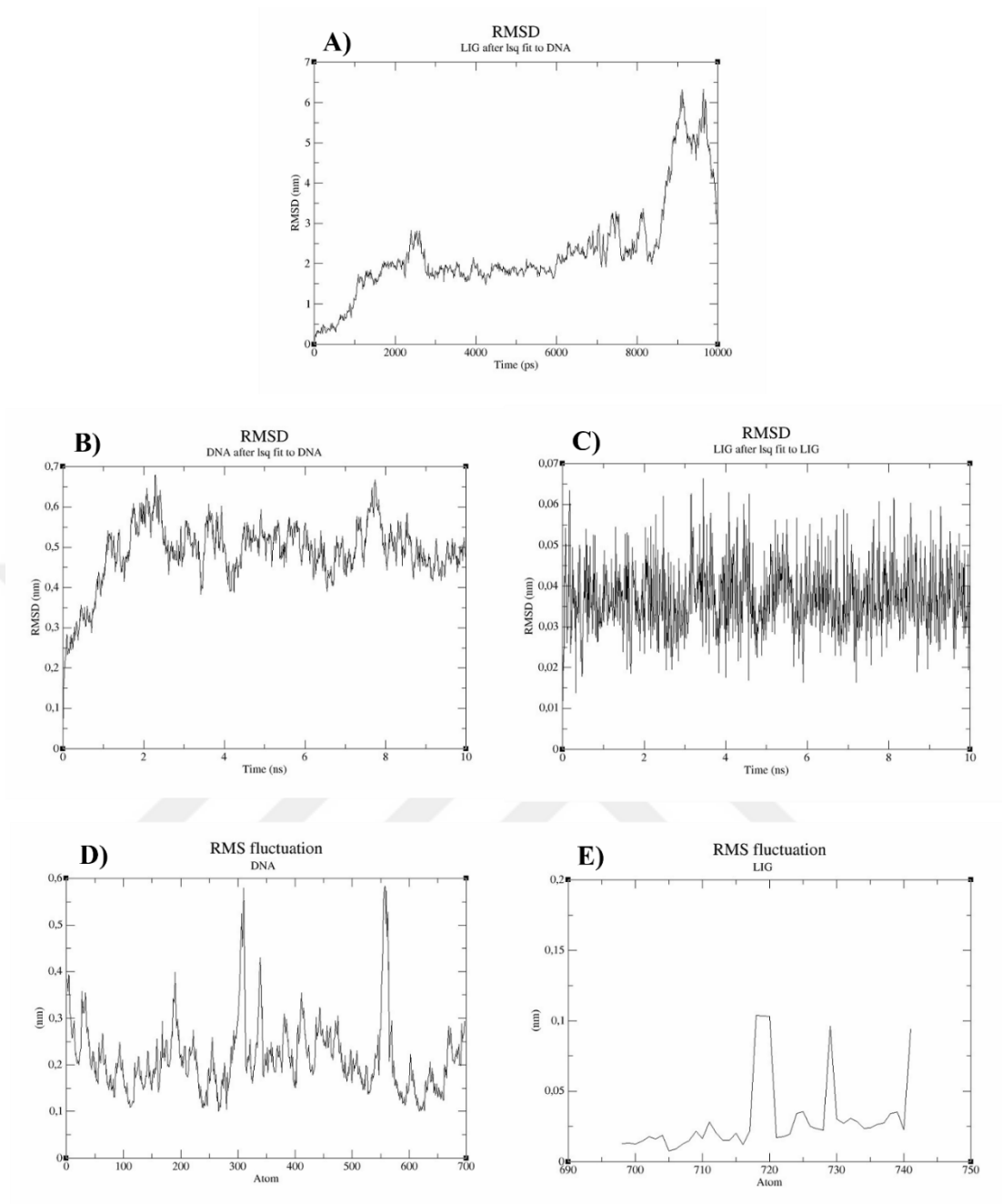


Figure 16. Third of the MD simulation trials of Estradiol and Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Alsager22 over time. (B) Calculation of the RMSD values of Alsager22 after least square fit to Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Alsager22. (E) Calculation of RMSF values for Estradiol.

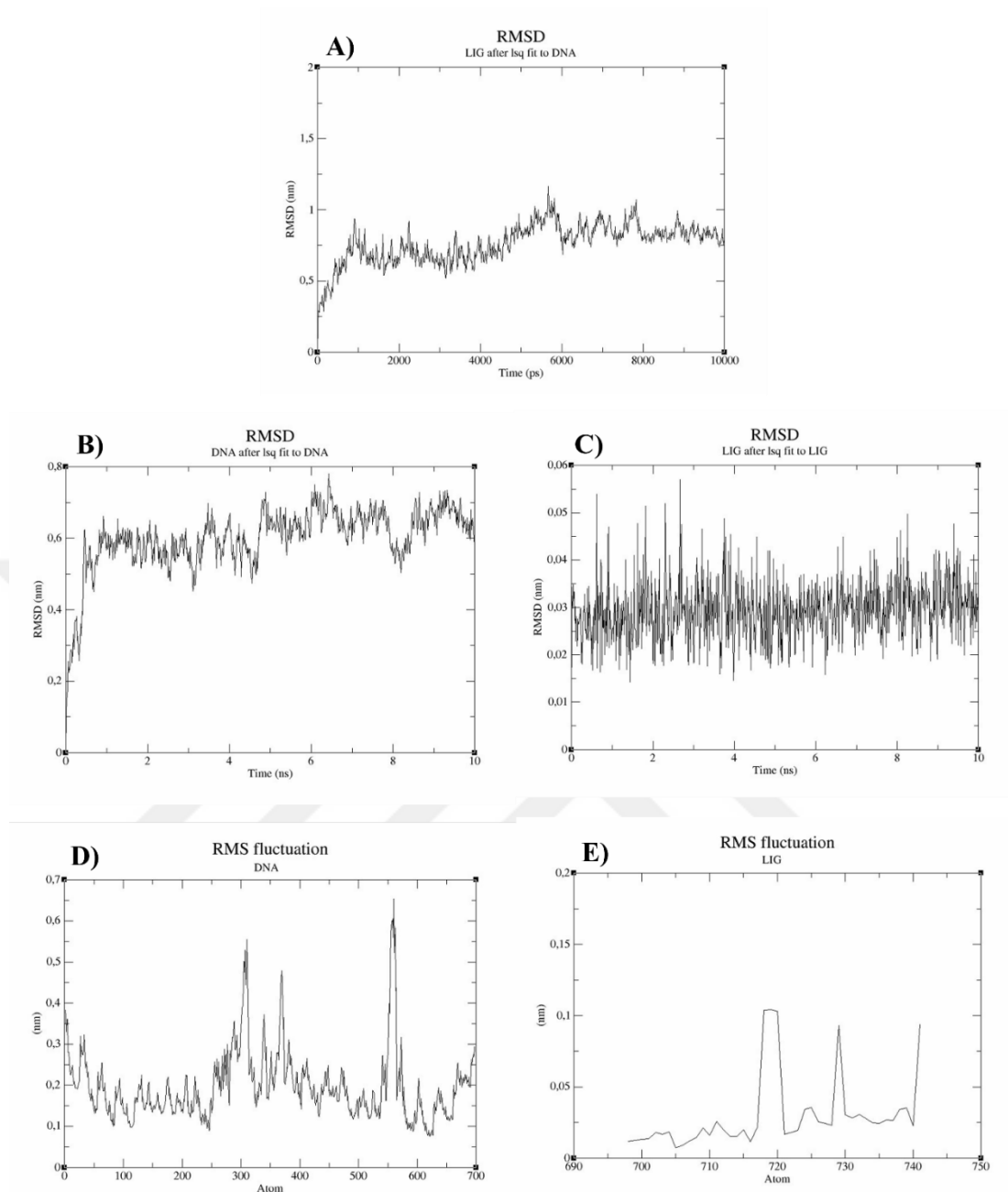


Figure 17. Fourth of the MD simulation trials of Estradiol and Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Alsager22 over time. (B) Calculation of the RMSD values of Alsager22 after least square fit to Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Alsager22. (E) Calculation of RMSF values for Estradiol.

### **3.1.5. $\Delta G$ calculations**

This section focuses on calculating the  $\Delta G$  of the target molecule and aptamer. In this thesis, a total of 16 MD simulations were performed using four different hormones, four times for each of them. These simulations were run under the same conditions and parameters, and the results were evaluated using gmxMMPBSA.  $\Delta G$  calculations were performed using two different approaches. The first method is the C2-entropy approach which measures the effect of disorder and random motions on energy. The second method is the interaction-entropy approach which involves calculating the intermolecular interactions and their contribution to the energy. Using these two approaches, a total of 32  $\Delta G$  results were obtained for each MD simulation. Each of the MD simulations performed consists of 1001 frames. These frames show the time-dependent motions of atoms and water molecules in detail. In the analysis process, the last 250 frames of the simulations were used as the basis for the  $\Delta G$  change calculations. These frames are generally thought to be closer to the equilibrium state than the beginning of the system and provide more statistically valid results. These results contribute to a deeper understanding of the thermodynamic stability of the target molecule and aptamer and their interactions. The lowest  $\Delta G$  scores indicate that Estradiol was a preferred target for Alsager22, while other compounds also found to be capable of forming stable complexes with the aptamer.

Table 2.  $\Delta G$  values of estradiol, progesterone, testosterone, and androstenedione molecules calculated using the C2 approach as a result of four MD simulation experiments with Alsager22.

<b>C2 - <math>\Delta G</math> (kcal/mol)</b>	<b>Estradiol</b>	<b>Progesterone</b>	<b>Testosterone</b>	<b>Androstenedione</b>
1st try	-17.37	0.03	0.01	5.66
2nd try	-21.49	-17.42	-9.26	-11.25
3rd try	0.03	3.2	0.02	-12.75
4th try	-12.13	-19.06	-15.38	0.06
<b><i>Average (all results)</i></b>	<b>-12.74</b>	<b>-8.3125</b>	<b>-6.1525</b>	<b>-4.57</b>
<b><i>Average (except for positive results)</i></b>	<b>-16.997</b>	<b>-18.24</b>	<b>-12.32</b>	<b>-12</b>
<b><i>Variance (except for positive results)</i></b>	<b>22.007</b>	<b>1.345</b>	<b>18.727</b>	<b>1.125</b>

Table 3.  $\Delta G$  values of estradiol, progesterone, testosterone, and androstenedione molecules calculated using the IE approach as a result of four MD simulation experiments with Alsager22.

<b>IE - <math>\Delta G</math> (kcal/mol)</b>	<b>Estradiol</b>	<b>Progesterone</b>	<b>Testosterone</b>	<b>Androstenedione</b>
1st try	-16.06	11.15	0.02	-6.61
2nd try	-23.14	-8.79	-10.88	-7.38
3rd try	1.89	-2.29	5.63	-13.96
4th try	-8.31	-7.05	-7.69	10.04
<b><i>Average (all results)</i></b>	<b>-11.405</b>	<b>-1.745</b>	<b>-3.23</b>	<b>-4.478</b>
<b><i>Average (except for positive results)</i></b>	<b>-15.837</b>	<b>-6.043</b>	<b>-9.285</b>	<b>-9.317</b>
<b><i>Variance (except for positive results)</i></b>	<b>55.020</b>	<b>11.323</b>	<b>5.088</b>	<b>16.319</b>

The calculations were done in two different ways, using the C2 entropy and IE entropy methods. The data indicate that estradiol maintains the lowest  $\Delta G$  values among all ligands in both methods, which may mean it has the strongest and most stable interaction with the aptamer. Overall, there are differences, but all the hormones tested proved capable of forming stable complexes with the aptamer. The negative and positive values found in the results of various experiments for a particular ligand may be caused by the randomness intrinsic to MD simulations. Most commonly, the GROMACS software used to run MD simulations employs random seeds to initialize simulation conditions. This results in different trajectories and outcomes each time a program is run. Comparing the results between the C2 and IE methods, it is noticeable that there are small differences in the  $\Delta G$  scores; these differences become more pronounced at higher values. Such differences may be the result of different approaches utilized by the two methods. These differences highlight the need for using more than one method to analyze the stability and strength of molecular interactions.



### ***3.1.6. Visualization of the interactions***

This part of the thesis focuses on the nature of the interactions of the Alsager22 aptamer with four different steroid hormones: progesterone, androstenedione, testosterone, and estradiol, after MD simulations. A total of 16 simulations were conducted, with four separate MD simulations performed for each steroid hormone under the same conditions and parameters. These simulations are long processes, each lasting 10 nanoseconds and containing 1001 frames. The final moment of each simulation, i.e., 1001st frame obtained at the end of 10 nanoseconds, was converted to PDB format using Chimera software, and then these PDB files were analyzed in Discovery Studio. The ligand interactions module of Discovery Studio visualized the interactions between the aptamer and each hormone. This visualization process reveals in detail the binding sites and types of interactions of the hormones with the aptamer.

At the end of the first MD simulation of Alsager22-estradiol, pi-alkyl interactions exist with the hormone and the thymine with residue numbers 17 and 18. There is also a pi-pi-stacked interaction between the hormone and the thymine with residue number 18. Furthermore, the hormone and the cytosine with residue number 19 formed a conventional hydrogen bond. At the end of the second MD simulation, pi-alkyl interactions exist with the hormone and the guanine with residue number 16, thymine with residue number 6, and thymine with residue number 5. The hormone and the guanine with residue number 16, and the thymine with residue number 5 formed a conventional hydrogen bond. Furthermore, the hormone and the adenine with residue number 15 formed a carbon-hydrogen bond. There is also a pi-pi-stacked interaction between the hormone and the thymine with residue number 5. No interaction between the hormone and the aptamer was observed at the end of the third MD simulation. Finally, at the end of the fourth MD simulation, pi-alkyl interactions exist with the hormone and the adenine with residue number 15. Furthermore, the hormone and the thymine with residue number 5 formed a conventional hydrogen bond. (Figure 18)

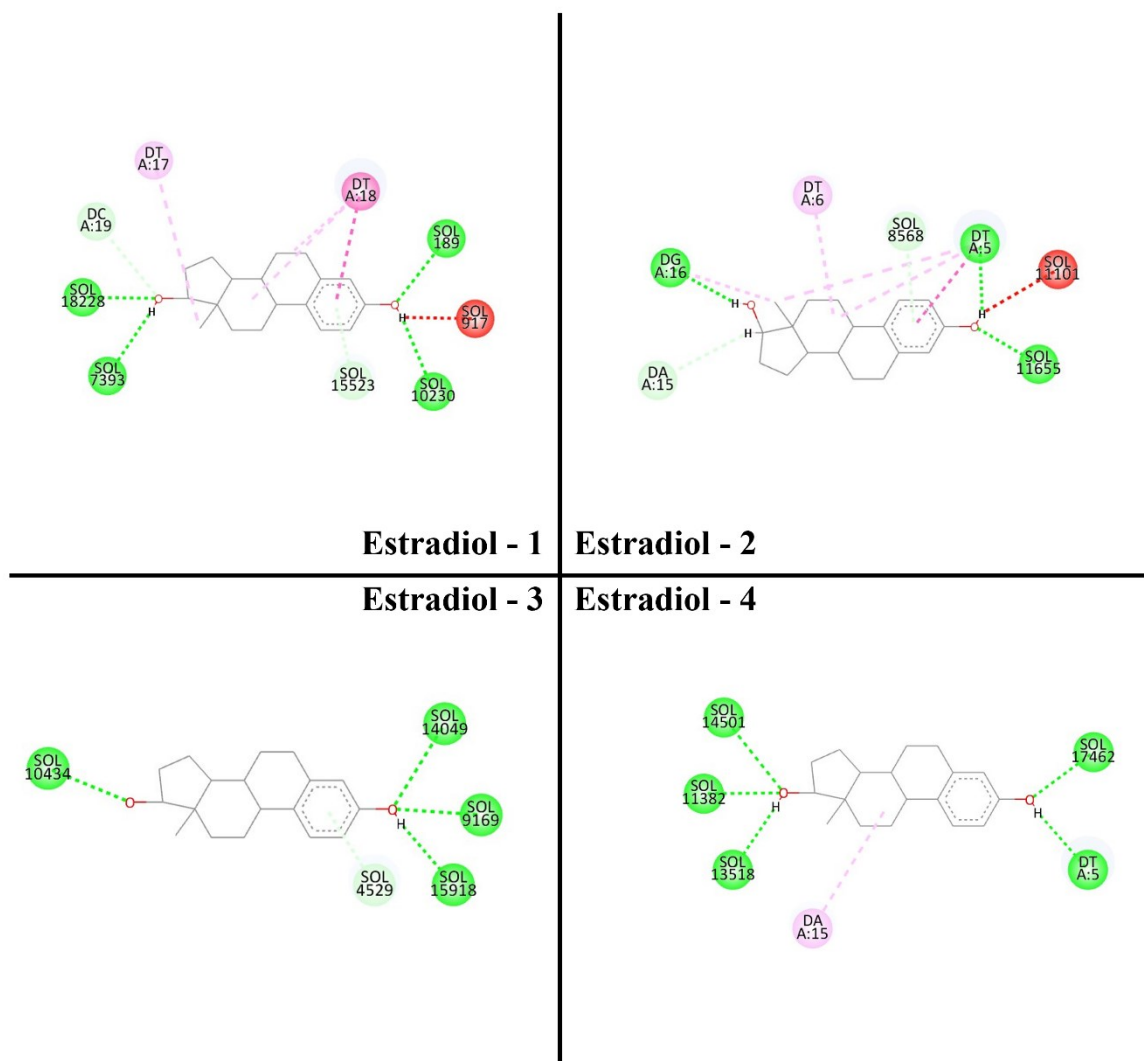


Figure 18. Interactions of estradiol hormone with Alsager22 aptamer at the end of 10 nanoseconds of four different MD simulations.

At the end of the first MD simulation of Alsager22-progesterone, no interaction between the hormone and the aptamer was observed. At the end of the second MD simulation, pi-alkyl interactions exist with the hormone and the guanine with residue number 9, and the cytosine with residue number 13. Furthermore, pi-sigma interactions exist with the hormone and the cytosine with residue number 13. At the end of the third MD simulation, pi-alkyl interactions exist with the hormone and the thymine with residue number 18. Finally, at the end of the third MD simulation, pi-alkyl interactions exist with the hormone and the guanine with residue number 1 and the cytosine with residue number 22. (Figure 19)

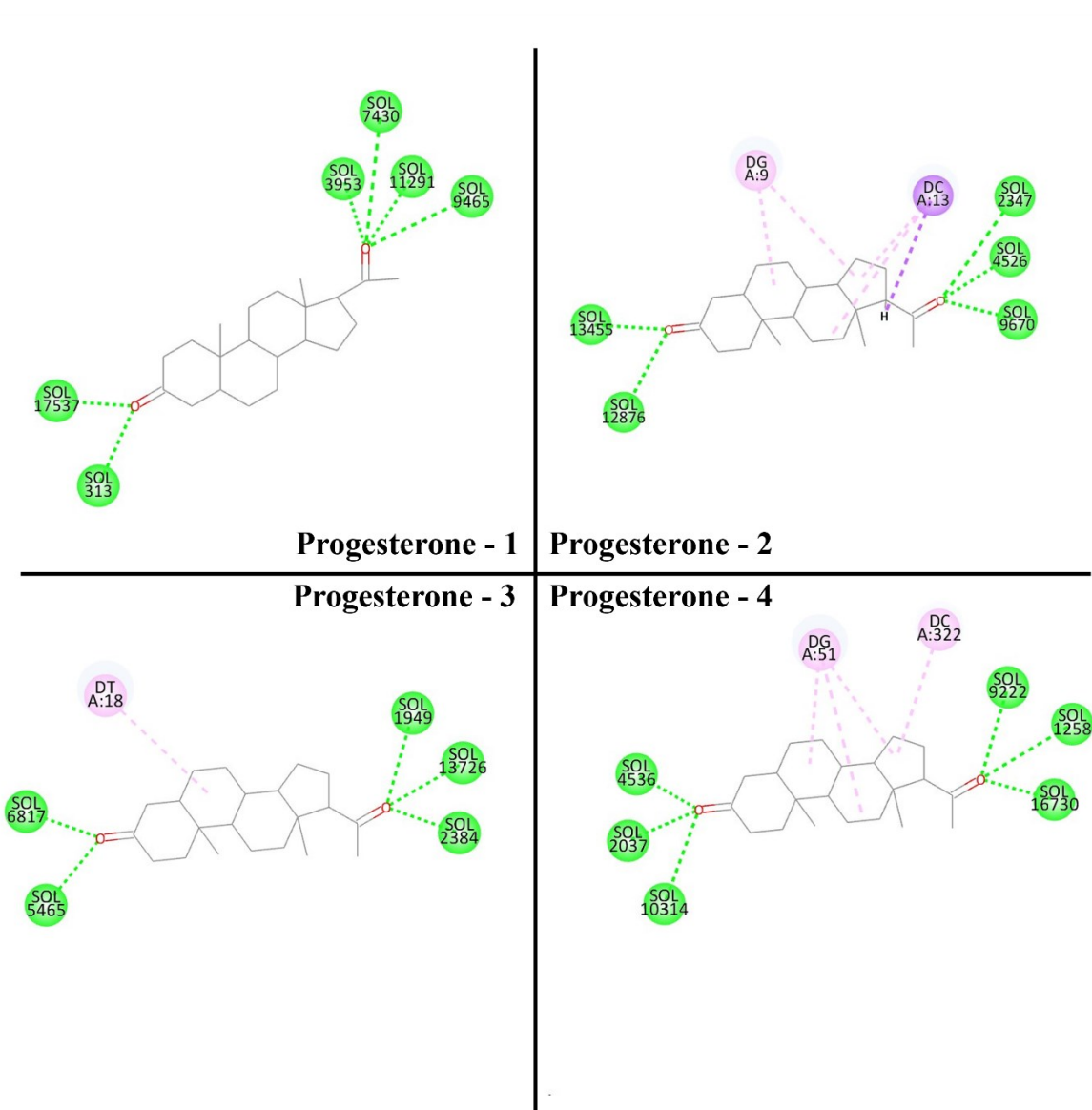


Figure 19. Interactions of progesterone hormone with Alsager22 aptamer at the end of 10 nanoseconds of four different MD simulations.

At the end of the first MD simulation of Alsager22-testosterone, no interaction between the hormone and the aptamer was observed. At the end of the second MD simulation, pi-alkyl interactions exist with the hormone and the guanine with residue number 9. At the end of the third MD simulation, no interaction between the hormone and the aptamer was observed. Finally, at the end of the third MD simulation, pi-alkyl interactions exist with the hormone, the guanine with residue number 1, and the cytosine with residue number 22. Furthermore, pi-sigma interactions exist with the hormone and the guanine with residue number 1. (Figure 20)

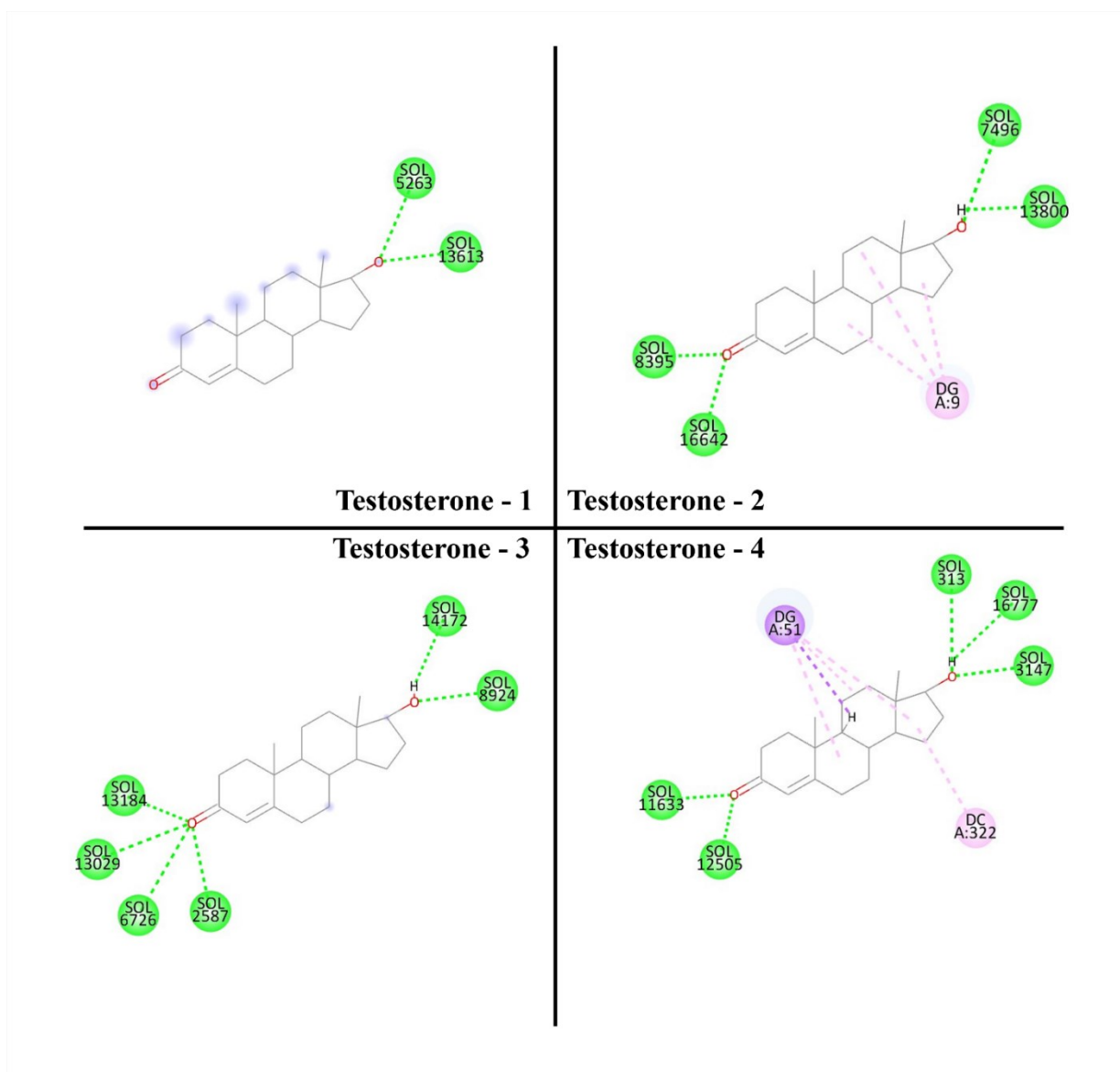


Figure 20. Interactions of testosterone hormone with Alsager22 aptamer at the end of 10 nanoseconds of four different MD simulations.

At the end of the first MD simulation of Alsager22-androstenedione, pi-alkyl interactions exist with the hormone and the adenine with residue number 15. At the end of the second MD simulation, pi-sigma and pi-alkyl interactions exist with the hormone and the adenine with residue number 15. At the end of the third MD simulation, pi-alkyl interactions exist with the hormone and the adenine with residue number 15, the thymine with residue number 5, and the thymine with residue number 6. Furthermore, pi-sigma interactions exist with the hormone and thymine with residue number 6. Finally, no interaction between the hormone and the aptamer was observed at the end of the fourth MD simulation. Adenine with residue number 15 seems to be a common site for interaction with the hormone. All interactions were pi-alkyl. (Figure 21)

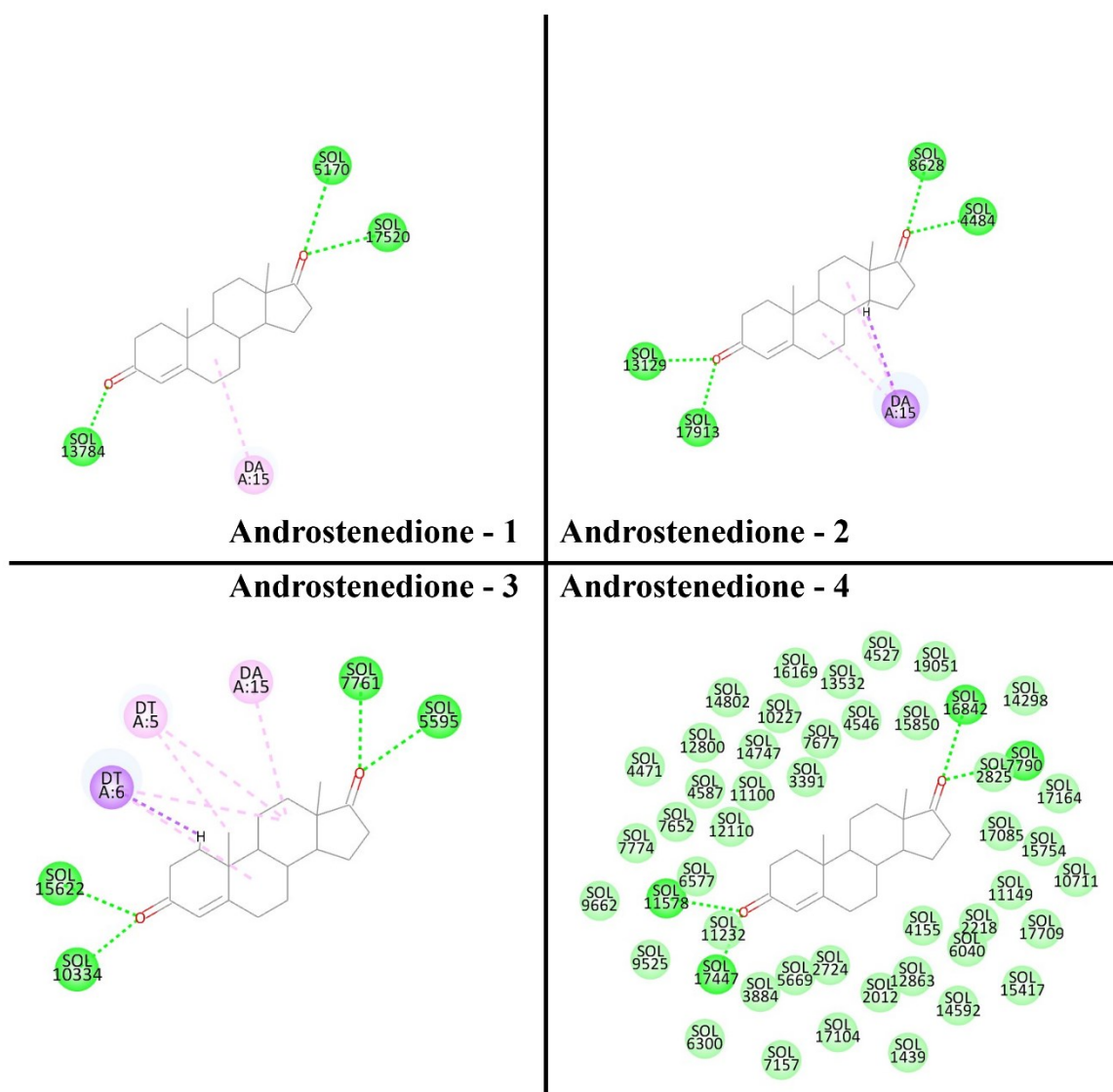


Figure 21. Interactions of androstenedione hormone with Alsager22 aptamer at the end of 10 nanoseconds of four different MD simulations.

In case of Estradiol, the aromatic moiety of the compound is found to be a favorable target for pi-pi stacking with Thymine bases, 5 or 18, at the end of the first simulation. There are several pi-alkyl simulations that contribute to the complex, however, the lowest  $\Delta G$  were observed for the two simulations where pi-pi stacking was observed. The lack of this aromatic moiety in all other hormones is probably what promotes the higher affinity of the aptamer structure to Estradiol. Besides the pi-pi interaction that formed between Estradiol and the aptamer, the majority of the bonds were pi-alkyl in nature. Most of these pi-alkyl bonds were formed with Adenine 15, Thymine 18, Thymine 5, Thymine 6, Guanine 1, and Cytosine 22. The former four nucleotides lay at the groove in the middle of the internal loop, while the latter two are present at the root of the stem. These positions also correspond to the binding sites of the compounds as discovered from the docking results. As these findings demonstrate, the interaction between Estradiol and the aptamer has significant affinity, particularly due to the pi-pi stacking ability of the aromatic structure. Such interactions significantly increase the binding strength and stability of the molecule with nucleotides. Moreover, the prevalence of pi-alkyl interactions indicates the structural flexibility of the aptamer and its ability to interact with the target molecule.

## 3.2. Optimization through *in-silico* mutagenesis

### 3.2.1. Proposing targets and mutant sequences

#### 3.2.1.1. Determination of possible mutation sites

In this thesis, *in-silico* mutagenesis was preferred to increase the affinity of the aptamer for estradiol, the target hormone of Alsager22. Before this, candidate residue proposals to be mutated are needed. In this direction, the second trial with the highest  $\Delta G$  among the four different MD simulation results of estradiol and Alsager22 was taken as the basis, and the Python script written to determine the three bases of the aptamer located closest to the center of estradiol in the last frame of this MD simulation was used. Based on the output shown in Figure 24, residues 5<sup>th</sup> thymine, 6<sup>th</sup> thymine, and 16<sup>th</sup> guanine were selected as candidates for *in-silico* mutagenesis.

```
*****
*****
3 different bases of DNA aptamer closest to Ligand in Alsager22-estradiol-2nd-MD-1001st-frame.pdb
<Residue DT het= resseq=5 icode= > Distance: 3.7373633
<Residue DT het= resseq=6 icode= > Distance: 4.816036
<Residue DG het= resseq=16 icode= > Distance: 5.372099
*****
*****
```

Figure 22. Possible mutation sites.

These three bases, which were found to be closest, were found to be compatible with the bases that interact with estradiol at the end of the second MD simulation. This selection highlights these nucleotides as potential candidates for targeted mutations. By further strengthening the interaction that the aptamer already establishes with estradiol, the proposed mutations can increase the specificity and binding strength in this molecular recognition process.

### **3.2.1.2. *In-silico* mutagenesis of the three closest bases in PYMOL**

The three residues of the aptamer identified in the previous step and closest to the center of estradiol after MD simulation were mutagenized *in-silico* to generate different mutant aptamer structures. Since there are four different possible nucleotides for each residue, this means 64 aptamer sequences, one of which is the order in the original aptamer structure. As a result, there are a total of 63 different mutant aptamers. For the *in-silico* mutagenesis step, a Python script written by integrating the nucleotide mutagenesis feature of the Pymol bioinformatics tool was used. In this script, *in-silico* mutagenesis was performed by providing the bases to be mutagenized, namely residues 5, 6, 16, and 63 different base arrangements. As the output of the script, a total of 63 different mutant aptamer PDB files with mutations at bases 5, 6, and 16 were obtained.

### **3.2.2. *Docking with mutant aptamers***

In this step, molecular docking was performed to examine the interaction of 63 mutant Alsager22 aptamers obtained in the previous step with estradiol. These mutants were based on the modification of the closest 3 bases to the ligand at the end of the second simulation with estradiol. The aim of the study was to investigate possible mutant aptamers that might show higher affinity to estradiol. The docking process was automated using a Python script and site-specific docking was performed. For site-specific docking, the coordinates of the C3 atom at the center of the estradiol hormone were determined using Discovery Studio software and these values were used as the x, y, and z coordinates of the grid box. For each docking, 63 different docking runs were performed with the exhaustiveness value set to 128. Higher affinities are presented in Table 4 and 5.

The docking results of the mutants were relatively similar as all of them ranged between -6.9 and -7.3 kcal/mol while the unchanged structure showed the affinity of -7.2 kcal/mol. Yet once sorted, a pattern among the highest affinities were revealed. According to this, having Guanine at the 5th position and a Guanine or Adenine at 16<sup>th</sup> position proved to be necessary to get the highest affinity. This pattern can be shortened as “GNR” pattern. We have chosen the GTG pattern as a mutant for further MD analysis.



Table 4. Docking results of mutant aptamers and estradiol. The unit of higher affinity is kcal/mol.

<b>Mutant Name</b>	<b>Best Affinity</b>
log_mutated_structure_Guanine_Adenine_Adenine_33.txt	-7.3
log_mutated_structure_Guanine_Adenine_Guanine_35.txt	-7.3
log_mutated_structure_Guanine_Cytosine_Adenine_37.txt	-7.3
log_mutated_structure_Guanine_Cytosine_Guanine_39.txt	-7.3
log_mutated_structure_Guanine_Guanine_Adenine_41.txt	-7.3
log_mutated_structure_Guanine_Guanine_Guanine_43.txt	-7.3
log_mutated_structure_Guanine_Thymine_Adenine_45.txt	-7.3
log_mutated_structure_Guanine_Thymine_Guanine_47.txt	-7.3
log_mutated_structure_Adenine_Adenine_Adenine_1.txt	-7.2
log_mutated_structure_Adenine_Adenine_Guanine_3.txt	-7.2
log_mutated_structure_Adenine_Cytosine_Adenine_5.txt	-7.2
log_mutated_structure_Adenine_Cytosine_Guanine_7.txt	-7.2
log_mutated_structure_Adenine_Guanine_Adenine_9.txt	-7.2
log_mutated_structure_Adenine_Guanine_Guanine_11.txt	-7.2
log_mutated_structure_Adenine_Thymine_Adenine_13.txt	-7.2
log_mutated_structure_Adenine_Thymine_Guanine_15.txt	-7.2
log_mutated_structure_Cytosine_Adenine_Adenine_17.txt	-7.2
log_mutated_structure_Cytosine_Adenine_Guanine_19.txt	-7.2
log_mutated_structure_Cytosine_Cytosine_Adenine_21.txt	-7.2
log_mutated_structure_Cytosine_Cytosine_Guanine_23.txt	-7.2
log_mutated_structure_Cytosine_Guanine_Adenine_25.txt	-7.2
log_mutated_structure_Cytosine_Guanine_Guanine_27.txt	-7.2
log_mutated_structure_Cytosine_Thymine_Adenine_29.txt	-7.2
log_mutated_structure_Cytosine_Thymine_Guanine_31.txt	-7.2
log_mutated_structure_Thymine_Adenine_Adenine_49.txt	-7.2
log_mutated_structure_Thymine_Adenine_Guanine_51.txt	-7.2
log_mutated_structure_Thymine_Cytosine_Adenine_53.txt	-7.2
log_mutated_structure_Thymine_Cytosine_Guanine_55.txt	-7.2
log_mutated_structure_Thymine_Guanine_Adenine_57.txt	-7.2
log_mutated_structure_Thymine_Guanine_Guanine_59.txt	-7.2
log_mutated_structure_Thymine_Thymine_Adenine_61.txt	-7.2

Table 5. Table 4 (continued)

<b>Mutant Name</b>	<b>Best Affinity</b>
log_mutated_structure_Adenine_Adenine_Thymine_4.txt	-7.1
log_mutated_structure_Adenine_Cytosine_Cytosine_6.txt	-7.1
log_mutated_structure_Adenine_Guanine_Cytosine_10.txt	-7.1
log_mutated_structure_Adenine_Guanine_Thymine_12.txt	-7.1
log_mutated_structure_Adenine_Thymine_Cytosine_14.txt	-7.1
log_mutated_structure_Guanine_Adenine_Cytosine_34.txt	-7.1
log_mutated_structure_Guanine_Adenine_Thymine_36.txt	-7.1
log_mutated_structure_Guanine_Cytosine_Cytosine_38.txt	-7.1
log_mutated_structure_Guanine_Cytosine_Thymine_40.txt	-7.1
log_mutated_structure_Guanine_Guanine_Cytosine_42.txt	-7.1
log_mutated_structure_Guanine_Guanine_Thymine_44.txt	-7.1
log_mutated_structure_Guanine_Thymine_Cytosine_46.txt	-7.1
log_mutated_structure_Guanine_Thymine_Thymine_48.txt	-7.1
log_mutated_structure_Adenine_Adenine_Cytosine_2.txt	-7
log_mutated_structure_Adenine_Cytosine_Thymine_8.txt	-7
log_mutated_structure_Adenine_Thymine_Thymine_16.txt	-7
log_mutated_structure_Cytosine_Adenine_Thymine_20.txt	-7
log_mutated_structure_Cytosine_Cytosine_Thymine_24.txt	-7
log_mutated_structure_Cytosine_Guanine_Cytosine_26.txt	-7
log_mutated_structure_Cytosine_Guanine_Thymine_28.txt	-7
log_mutated_structure_Cytosine_Thymine_Thymine_32.txt	-7
log_mutated_structure_Thymine_Adenine_Cytosine_50.txt	-7
log_mutated_structure_Thymine_Adenine_Thymine_52.txt	-7
log_mutated_structure_Thymine_Cytosine_Cytosine_54.txt	-7
log_mutated_structure_Thymine_Cytosine_Thymine_56.txt	-7
log_mutated_structure_Thymine_Guanine_Cytosine_58.txt	-7
log_mutated_structure_Thymine_Guanine_Thymine_60.txt	-7
log_mutated_structure_Thymine_Thymine_Cytosine_62.txt	-7
log_mutated_structure_Thymine_Thymine_Thymine_63.txt	-7
log_mutated_structure_Cytosine_Adenine_Cytosine_18.txt	-6.9
log_mutated_structure_Cytosine_Cytosine_Cytosine_22.txt	-6.9
log_mutated_structure_Cytosine_Thymine_Cytosine_30.txt	-6.9

### 3.2.3. MD simulations of mutant aptamer with the best dock

In this part of the thesis, the MD simulation of estradiol with the mutant aptamer selected in the previous step was performed. The simulations were run four times under the same conditions for 10 ns at a temperature target of 300 K with a time step of 2 fs. After the MD simulation, the final frame of each experiment was visualized. In the Estradiol-Mutant Alsager22 simulations, it was observed that Estradiol bound to the mutant aptamer in similar positions for three simulations, except for the first simulation in which the compound was not bound. (Figure 23) In the three MD simulations in which aptamer and ligand interacted, the salient point was that the interaction site of the ligand with the aptamer in the original alsager22-estradiol MD experiment selected for further analysis (2<sup>nd</sup> simulation of alsager22-estradiol) was change. The ligand did not interact with the mutated region, which was performed to increase the affinity at the interaction site. This suggests that the mutation did not have the expected effect. Nevertheless, in the three MD simulation experiments where the aptamer and ligand interact, the location of the ligand at the end of 10 ns is quite similar. The mutation did not increase the affinity of the target site, in fact it caused a loss of interaction, but it may have provided another site with a better affinity. The next step for this was to perform  $\Delta G$  analyses.

For further analysis, RMSD and RMSF plots were drawn as a result of all MD simulations. RMSD analysis of the ligand versus DNA allowed for better interpretation of *in-silico* mutagenesis. For MD simulation 1, the ligand was stable from 1<sup>st</sup> ns to 3<sup>rd</sup> ns, after which the ligand exited the stable pose. (Figure 24A) For MD simulation 2, the ligand became stable on DNA after ns 3. (Figure 25A) Unlike the others, for MD simulation 3, the ligand was stable at one location from 1<sup>st</sup> ns to 5<sup>th</sup> ns, followed by 5<sup>th</sup> ns to 8<sup>th</sup> ns, then stabilized at a different location from 5<sup>th</sup> ns to 8<sup>th</sup> ns, and changed its stable location again after 8<sup>th</sup> ns. (Figure 26A) For MD simulation 4, the ligand was stable from 1<sup>st</sup> ns to 6<sup>th</sup> ns, after which the ligand left the stable position. (Figure 27A)

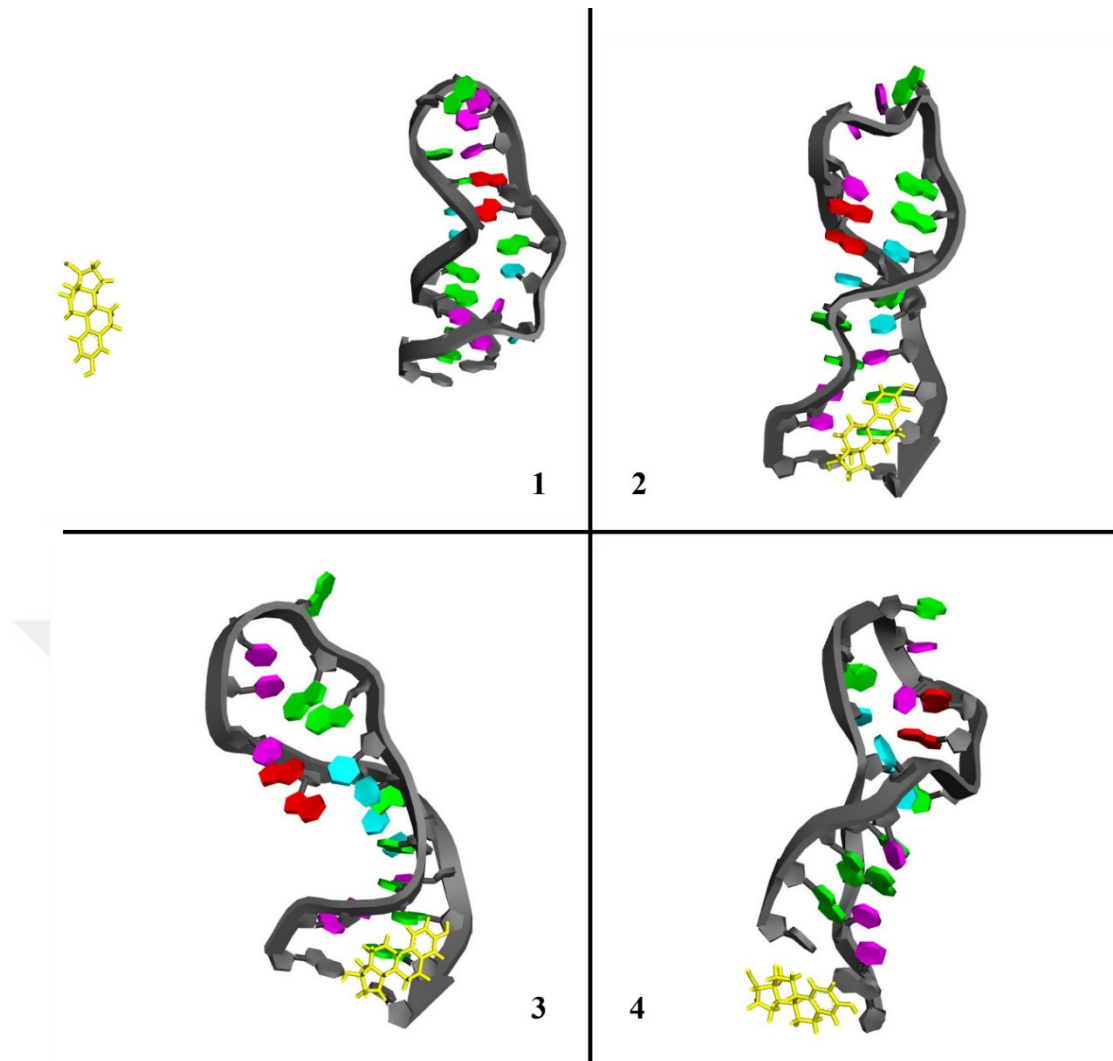


Figure 23. Positions of Mutant Alsager22 and estradiol hormone after 10 nanoseconds of four different MD simulations.

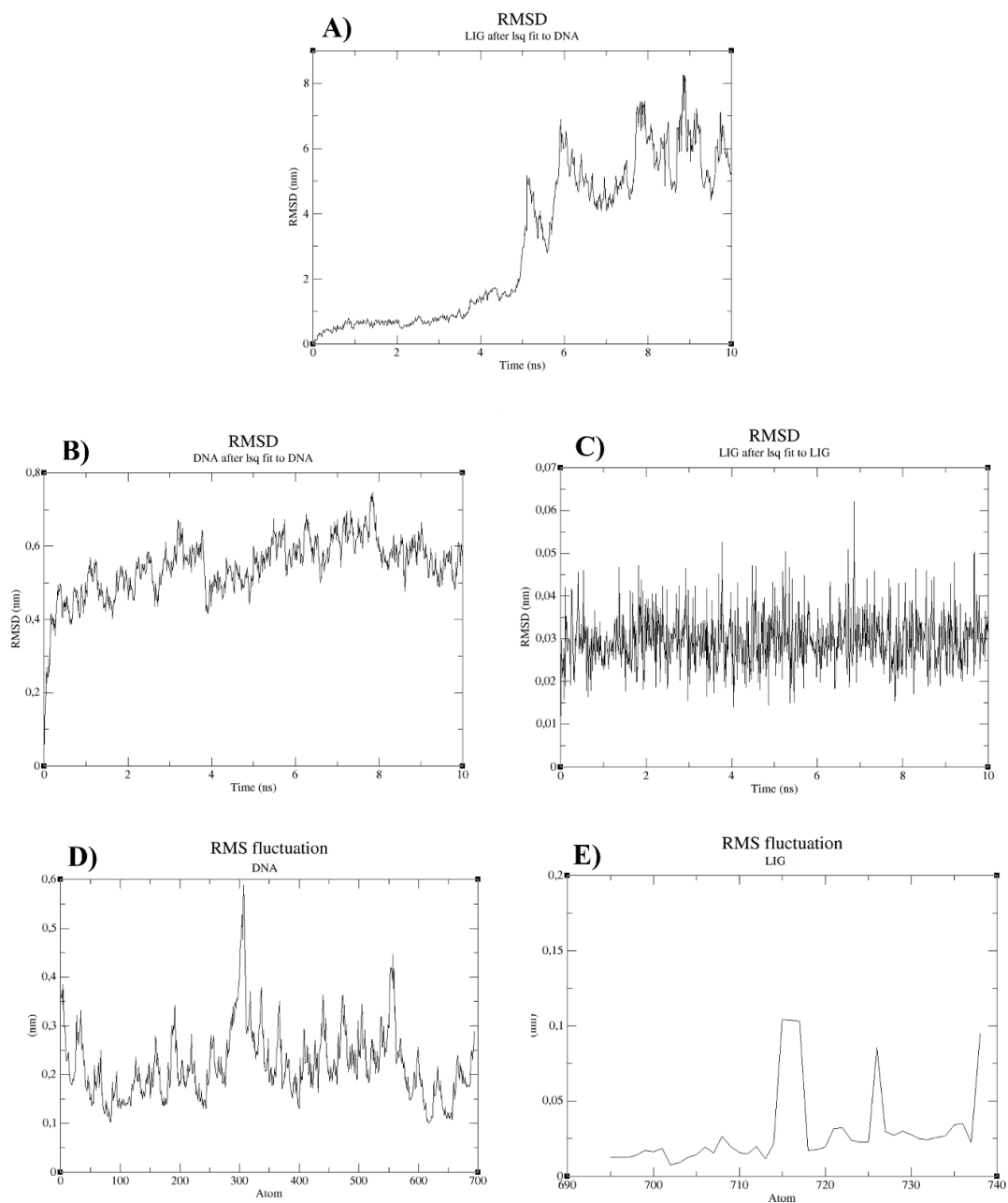


Figure 24. First of the MD simulation trials of Estradiol and Mutant Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Mutant Alsager22 over time. (B) Calculation of the RMSD values of Mutant Alsager22 after least square fit to Mutant Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Mutant Alsager22. (E) Calculation of RMSF values for Estradiol.

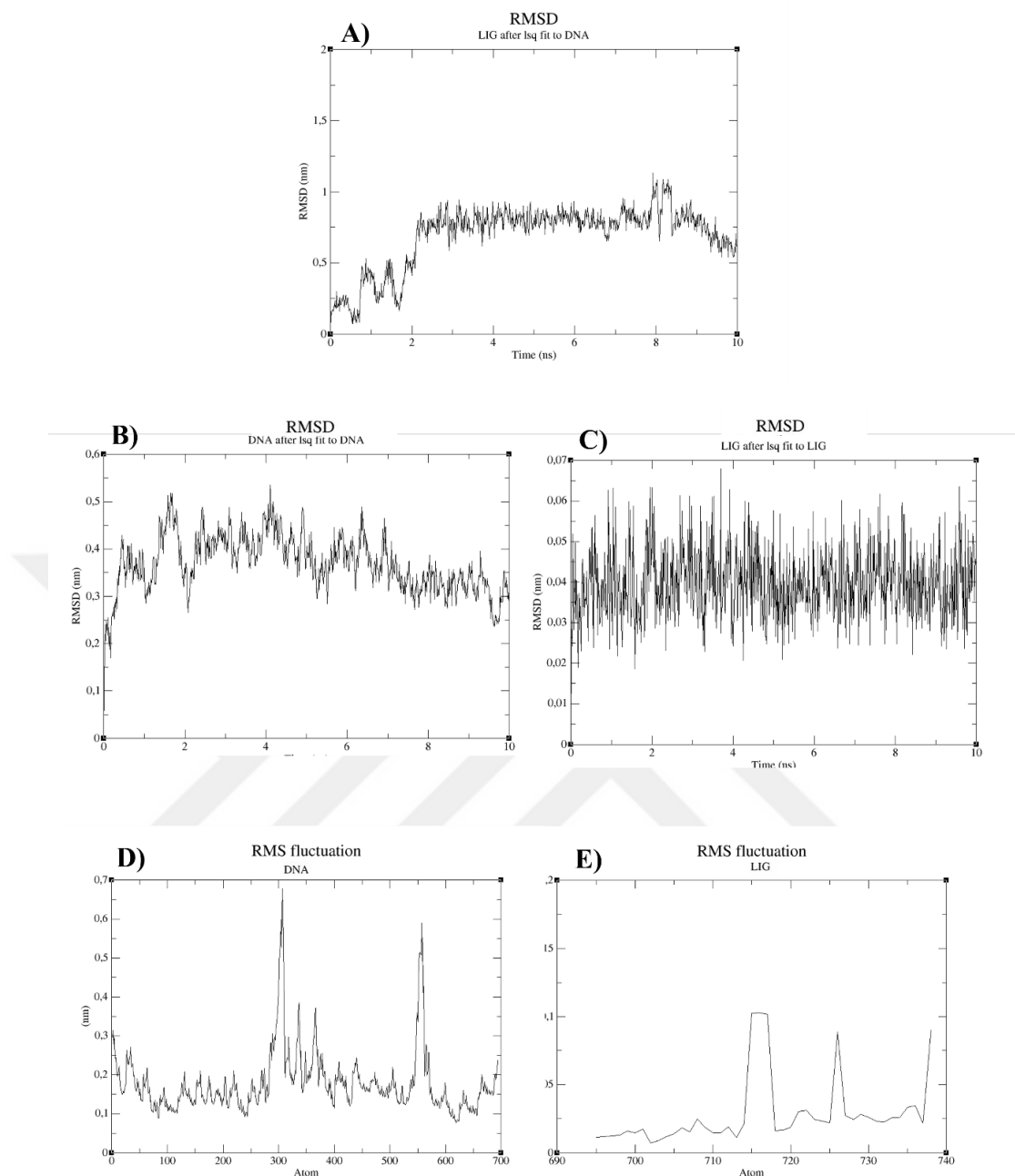


Figure 25. Second of the MD simulation trials of Estradiol and Mutant Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Mutant Alsager22 over time. (B) Calculation of the RMSD values of Mutant Alsager22 after least square fit to Mutant Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Mutant Alsager22. (E) Calculation of RMSF values for Estradiol.

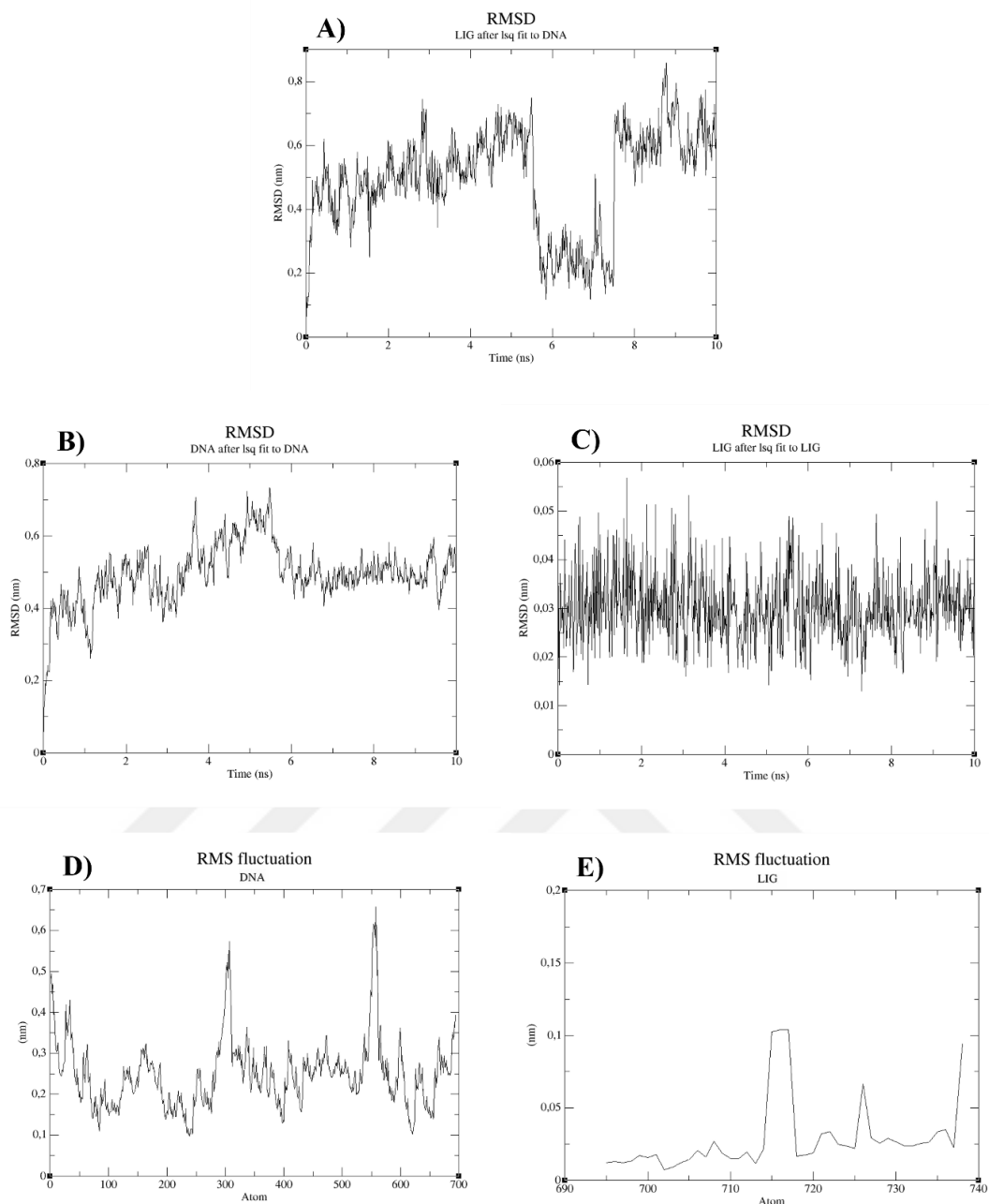


Figure 26. Third of the MD simulation trials of Estradiol and Mutant Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Mutant Alsager22 over time. (B) Calculation of the RMSD values of Mutant Alsager22 after least square fit to Mutant Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Mutant Alsager22. (E) Calculation of RMSF values for Estradiol.

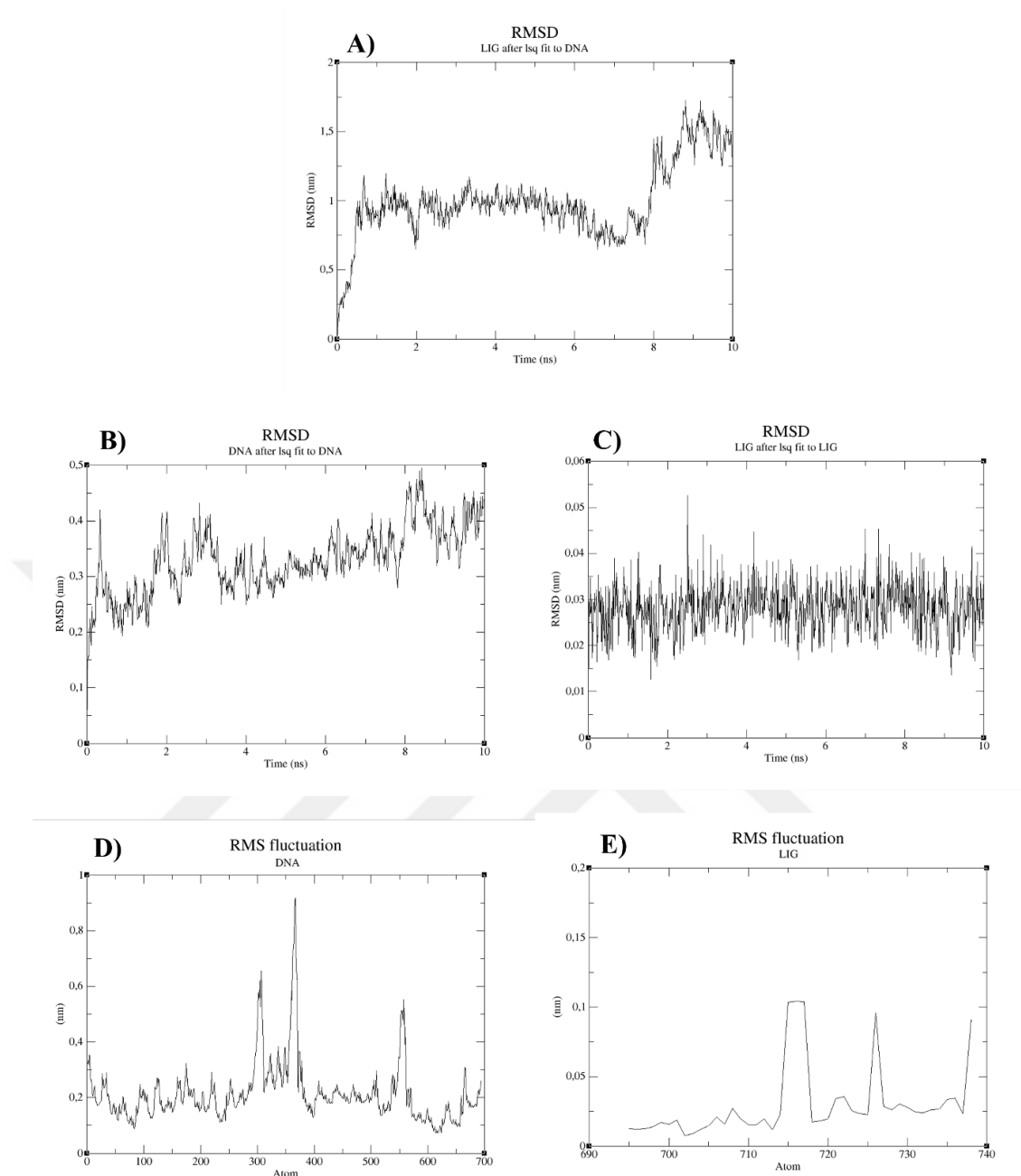


Figure 27. Fourth of the MD simulation trials of Estradiol and Mutant Alsager22. (A) Calculation of the RMSD values of Estradiol after least square fit to Mutant Alsager22 over time. (B) Calculation of the RMSD values of Mutant Alsager22 after least square fit to Mutant Alsager22 over time. (C) Calculation of the RMSD values of Estradiol after least square fit to Estradiol over time. (D) Calculation of RMSF values for Mutant Alsager22. (E) Calculation of RMSF values for Estradiol.



### 3.2.4. $\Delta G$ calculations

This section focuses on the calculation of the  $\Delta G$  of the target molecule and the mutant aptamer. In this step, a total of 4 MD simulations between estradiol and mutant aptamer under the same conditions and parameters were performed.  $\Delta G$  values were calculated using gmxMMPBSA. The  $\Delta G$  calculations were performed as in the other steps, using two different approaches, C2 and IE, and focusing on the last 250 frames of the simulations. Surprisingly, when comparing the mutant aptamer and the original aptamer, the lowest  $\Delta G$  score belongs to the 2nd MD simulation between Estradiol and mutant Alsager22 (-25.59 kcal/mol). (Table 6 and 7)

Table 6.  $\Delta G$  values of estradiol calculated using the C2 approach as a result of four MD simulation experiments with mutant Alsager22 and original Alsager22.

C2 - $\Delta G$ (kcal/mol)	After <i>in-silico</i> mutagenesis (5G - 6T - 16G)	Before <i>in-silico</i> mutagenesis (5T - 6T - 16G)
1st try	0	-17.37
2nd try	<b>-25.59</b>	-21.49
3rd try	-23.07	0.03
4th try	-2.44	-12.13
<i>Average (all results)</i>	<b>-12.775</b>	<b>-12.74</b>
<i>Average (except for positive results)</i>	<b>-17.033</b>	<b>-16.997</b>
<i>Variance (except for positive results)</i>	<b>161.312</b>	<b>22.007</b>

Table 7.  $\Delta G$  values of estradiol calculated using the IE approach as a result of four MD simulation experiments with mutant Alsager22 and original Alsager22.

IE - $\Delta G$ (kcal/mol)	After <i>in-silico</i> mutagenesis (5G - 6T - 16G)	Before <i>in-silico</i> mutagenesis (5T - 6T - 16G)
1st try	9.71	-16.06
2nd try	-20.9	-23.14
3rd try	-21.67	1.89
4th try	-6.24	-8.31
<i>Average (all results)</i>	<b>-9.775</b>	<b>-11.405</b>
<i>Average (except for positive results)</i>	<b>-16.270</b>	<b>-15.837</b>
<i>Variance (except for positive results)</i>	<b>75.599</b>	<b>55.020</b>

Estradiol had changed its interaction site after mutation and did not bind to the site targeted to increase its affinity by mutation. This is understandable as the overall structure of the aptamer was observed to change after mutation. The highest  $\Delta G$  value of mutant alsager22 and estradiol was calculated as -25.59 kcal/mol in the second MD simulation experiment. This is the highest  $\Delta G$  value calculated throughout the entire study. The highest  $\Delta G$  value of the original alsager22 and estradiol was calculated to be -23.14 kcal/mol. Although the affinity could not be increased at the mutation site, a higher potential was observed at the new site. However, the change in the overall structure of the aptamer and the interaction site of the hormone after mutation suggests that this workflow is not a reliable and repeatable process. Accordingly, it could not be concluded that better affinity would be achieved by mutation. Nevertheless, since the highest affinity was observed in the MD simulation with the mutant aptamer throughout the study, the possibility of obtaining better affinity with mutation is obvious. It should also be noted that only one of 63 different mutant aptamers could be tested in this study. In this case, extensive computational power is required as there are too many mutation possibilities.

### ***3.2.5. Visualization of the interactions***

This part of the thesis focuses on the visualization of the binding sites and types of interactions of the mutant Alsager22 aptamer with estradiol after MD simulations. The last frame of 4 simulations, obtained at the end of 10 nanoseconds, was converted to PDB format using Chimera software and then these PDB files were analyzed with the ligand interactions module in Discovery Studio.

At the end of the first MD simulation of mutant Alsager22-estradiol, no interaction between the hormone and the mutant aptamer was observed. At the end of the second MD simulation, pi-alkyl interactions are present with the hormone and guanine with residue number 1, guanine with residue number 21 and cytosine with residue number 22. At the end of the third MD simulation, a pi-donor hydrogen bond exists with the hormone and guanine with residue number 20. Finally, at the end of the fourth MD simulation, pi-alkyl interactions exist with the hormone and guanine with residue number 1 and cytosine with residue number 22. Also, there is a pi-pi stacked interaction with guanine with residue number 1.

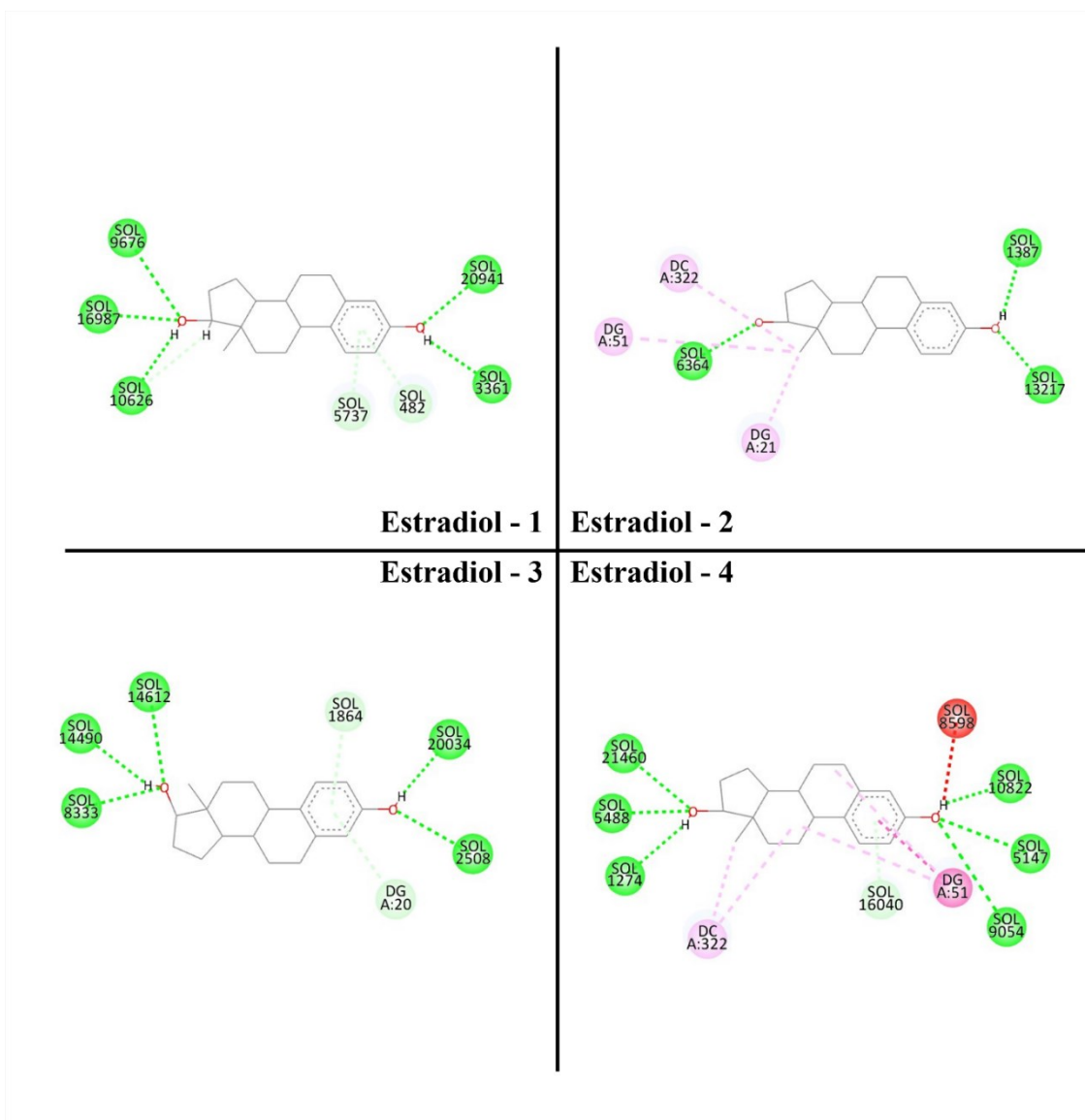


Figure 28. Interactions of estradiol hormone with Alsager22 aptamer at the end of 10 nanoseconds of four different MD simulations.

## CHAPTER 4: CONCLUSION

The aim of the thesis is to perform a comprehensive structural analysis of aptamers produced by the SELEX method and based on this analysis, to design a workflow to generate an aptamer modified to increase its binding affinity. First, a 2D aptamer structure was obtained based on the aptamer sequences, and a 3D structure was obtained from it. At this stage, we aimed to reveal the physical structure of aptamers in detail. Analysis of the interaction of the 3D aptamer structure with the target molecule was performed using the molecular docking method. Following the molecular docking results, MD simulations were performed to analyze the interaction between the target molecule and the aptamer in more detail. As a result, the binding site of estradiol was revealed. Furthermore, the evaluation of the thermodynamic stability of the complex was performed using  $\Delta G$  calculations. Structural analysis revealed that alsager22 interacts with off-target hormones, although the experimental article states that no binding with off-target hormones (progesterone, testosterone, and androstenedione) is observed, but the higher affinity is still for estradiol. Next, we examined whether the performance of the current aptamer could be improved against its target hormone (estradiol). Molecular interactions between the target molecule and the structural features of the current aptamer were analyzed, and possible mutations that could increase the binding affinity were proposed. After obtaining 63 potential mutant aptamers, the closest mutant aptamer to the original aptamer was selected, and molecular docking and molecular dynamics analyses were repeated. When the  $\Delta G$  values were compared with the original aptamer, the highest affinity was observed in the MD simulation performed with the mutant aptamer throughout the study. However, since the mutation changed the overall structure of the aptamer, the binding site of the hormone was also changed. In this case, we cannot conclude that better affinity will be obtained by mutation since it is not a reliable and repeatable process. Nevertheless, it became clear that there is a possibility to get better affinity with mutations, but it requires high processing power because there are so many possibilities to work with. In conclusion, this study is a step towards studying aptamers at the molecular level and increasing their therapeutic potential through their modifications.

This research not only advances our understanding of alsager22 aptamer structure but also sets a precedent for future structural analyses in this field. Studying on a workflow to modify aptamers for enhanced binding affinity is a significant step forward. This workflow has potential applications in improving the efficacy of aptamers as therapeutic agents, thereby broadening their utility in medical science. Additionally, the investigation into potential mutations for increasing binding affinity opens new avenues for aptamer optimization. While the results indicate that mutation can lead to changes in binding sites and is not a guaranteed route to increased affinity, the exploration of this possibility is a valuable contribution. It lays the groundwork for future research that might find more reliable methods for enhancing aptamer-target interactions through mutation.

Moreover, our findings challenge existing assumptions about the specificity of aptamers. The discovery that alsager22 interacts with off-target hormones, despite previous claims of specificity, underscores the need for more rigorous testing of aptamer-target interactions.

In summary, this thesis not only advances our understanding of aptamers and their interactions with target molecules but also opens up new questions and areas for future research, thereby significantly contributing to the field of molecular biology and therapeutic agent development.

## REFERENCES

Abraham, M.J., Murtola, T., Schulz, R., Pall, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) '*GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*', *SoftwareX*, Vol. 1, pp. 19–25.

Alsager, O.A., Kumar, S., Willmott, G.R., McNatty, K.P. and Hodgkiss, J.M. (2014) '*Small molecule detection in solution via the size contraction response of aptamer functionalized nanoparticles*', *Biosensors and Bioelectronics*, Vol. 57, pp. 262–268.

Alsager, O.A., Kumar, S., Zhu, B., Travas-Sejdic, J., McNatty, K.P. and Hodgkiss, J.M. (2015) '*Ultrasensitive Colorimetric Detection of 17 $\beta$ -Estradiol: The Effect of Shortening DNA Aptamer Sequences*', *Analytical Chemistry*, Vol. 87(8), pp. 4201–4209.

Blackwell, T.K. and Weintraub, H. (1990) '*Differences and Similarities in DNA-Binding Preferences of MyoD and E2A Protein Complexes Revealed by Binding Site Selection*', *Science*, Vol. 250(4984), pp. 1104–1110.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) '*CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*', *Journal of Computational Chemistry*, Vol. 4(2), pp. 187–217.

Darmostuk, M., Rimpelova, S., Gbelcova, H. and Ruml, T. (2015) '*Current approaches in SELEX: An update to aptamer selection technology*', *Biotechnology Advances*, Vol. 33(6 Pt 2), pp. 1141–1161.

Gao, S., Zheng, X., Jiao, B. and Wang, L. (2016) '*Post-SELEX optimization of aptamers*', *Analytical and Bioanalytical Chemistry*, Vol. 408(17), pp. 4567–4573.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R. and Hofacker, I.L. (2008) '*The Vienna RNA Websuite*', *Nucleic Acids Research*, Vol. 36(suppl\_2), pp. W70–W74.

Lemkul, J. (2019) *'From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]'*, Living Journal of Computational Molecular Science, Vol. 1(1).

Markham, N.R. and Zuker, M. (2008) *'UNAFold'*, in J.M. Keith (ed.) *Bioinformatics: Structure, Function and Applications*. Totowa, NJ: Humana Press (Methods in Molecular Biology™), pp. 3–31.

Rognan, D. (1998) *'Molecular dynamics simulations: a tool for drug design'*, Perspectives in Drug Discovery and Design, Vol. 9(0), pp. 181–209.

Svobodová, M., Skouridou, V., Botero, M.L., Jauset-Rubio, M., Schubert, T., Bashammakh, A.S., El-Shahawi, M.S., Alyoubi, A.O. and O'Sullivan, C.K. (2017) *'The characterization and validation of 17 $\beta$ -estradiol binding aptamers'*, The Journal of Steroid Biochemistry and Molecular Biology, Vol. 167, pp. 14–22.

Thompson, A.P., Aktulga, H.M., Berger, R., Bolintineanu, D.S., Brown, W.M., Crozier, P.S., in 't Veld, P.J., Kohlmeyer, A., Moore, S.G., Nguyen, T.D., Shan, R., Stevens, M.J., Tranchida, J., Trott, C. and Plimpton, S.J. (2022) *'LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales'*, Computer Physics Communications, Vol. 271, pp. 108171.

Trott, O. and Olson, A.J. (2010) *'AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading'*, Journal of Computational Chemistry, Vol. 31(2), pp. 455–461.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J.C. (2005) *'GROMACS: fast, flexible, and free'*, Journal of Computational Chemistry, Vol. 26(16), pp. 1701–1718.

Wei, G.-W. (2019) *'Protein structure prediction beyond AlphaFold'*, Nature Machine Intelligence, Vol. 1(8), pp. 336–337.

Wright, W.E., Binder, M. and Funk, W. (1991) '*Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site.*', Molecular and Cellular Biology, Vol. 11(8), pp. 4104–4110.

Zhang, Y., Wang, J. and Xiao, Y. (2022) '*3dRNA: 3D Structure Prediction from Linear to Circular RNAs*', Journal of Molecular Biology, Vol. 434(11), pp. 167452.

Zhang, Y., Xiong, Y. and Xiao, Y. (2022) '*3dDNA: A Computational Method of Building DNA 3D Structures*', Molecules, Vol. 27(18), pp. 5936.

Zuker, M. (2003) '*Mfold web server for nucleic acid folding and hybridization prediction*', Nucleic Acids Research, Vol. 31(13), pp. 3406–3415.