IN SILICO SELEX IN DESIGNING NOVEL DNA APTAMER HAIRPIN FOR IN VITRO DETECTION OF HE4: AN OVARIAN CANCER BIOMARKER

BY

NUR NADIAH ABDUL RASHID

A thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy in Pharmacy

Kulliyyah of Pharmacy
International Islamic University Malaysia

NOVEMBER 2024

ABSTRACT

Ovarian cancer (OC) poses a significant risk as it usually remains asymptomatic until advanced stages, resulting in delayed diagnosis and lowering the chances of survival. This fifth most common cancer among women globally lacks efficient screening approaches for early stages OC (stage I and II), increasing the threat of OC progression and mortality. In this study, the *in silico* method was applied to design DNA aptamer hairpins for the detection of human epididymis protein 4 (HE4), an OC biomarker. The *in silico* work outcome was supported by an *in vitro* assay. The work began with the HE4 protein modelling using AlphaFold, I-TASSER, and Robetta protein structure prediction servers. Subsequently, molecular dynamics (MD) simulation was conducted on each predicted model for 100 ns using the OPLS force field, for structure refinement. The tertiary structure quality was validated by PROCHECK and ERRAT, showing the refined model from AlphaFold, RF1, was the highest-quality HE4 tertiary conformation. All amino acids were located in the favoured regions of Ramachandran plot with ERRAT overall quality score of 97.701. Next, this HE4 structure was docked using AutoDock Vina with four HE4 aptamer candidates (A1, A2, A3, and A4). The HE4-A4 binding energy was -6.0 kcal/mol, and the complex formed 24 hydrogen bonds (H-bonds); 5 identified at the aptamer hairpin loop region. A4 was chosen as the most suitable candidate to be utilised in the designing of the DNA hairpin, as it exhibited good affinity with highest number of H-bonds at the hairpin loop. To initiate the *in silico* design of new hairpins, the 25-mer A4 aptamer was truncated at the hairpin region, 5'-CGCAAG-3' and the stem was extended, forming the 5'-GCGCAAGC-3' sequence. The loop nucleotides, -GCAA-, were substitutionally mutated, producing 256 sequences. These 256 hairpins were docked with HE4 individually, and H16, H101, and H256 have shown good binding affinities with binding energies ranging between -10.6 and -11.6 kcal/mol. Consequently, 100 ns MD simulation using CHARMM27 force field was applied to the HE4-H16, HE4-H101, and HE4-H256 complexes. Based on the RMSD, radius of gyration, number of H-bonds, H-bond occupancy, and the overall total energy, H256 was deduced as the most promising DNA hairpin against HE4 marker with great affinity and stability throughout the simulation. This H256 hairpin was synthesised and its binding with HE4 was analysed *via in vitro* electrophoretic mobility shift assay (EMSA). Based on the DNA band intensities, the designed H256 (3.27 %) bound four times better to HE4 than the A4 aptamer (0.84 %). Finally, a preliminary study for future diagnostic potential was carried out by conjugating gold nanoparticle (GNP) with H256. The FTIR and Raman spectra confirmed the presence of amide group, formed by the successful conjugation of the carboxylated GNP with the aminated-H256. The GNP solution changed from red to purple-red, indicating the size increment after conjugation, that was confirmed by particle size analyser. In conclusion, H256 is a promising DNA hairpin in HE4 screening and is recommended for future development of a fully functional OC diagnostic kit, suitable to be used in routine screening for all women, with or without symptoms. This potentially improves the detection among early stages (I and II) patients, enhancing patients' outcome.

ملخص البحث

يشكل سرطان المبيض (*Ovarian Cancer-OC*) خطرًا كبيرًا لأنه عادة ما يظل بدون أعراض حتى مراحل متقدمة، مما يؤدي إلى تأخير التشخيص، وتقليل فرص البقاء على قيد الحياة. يفتقر هذا النوع الخامس من السرطان الأكثر شيوعًا بين النساء على مستوى العالم إلى طرق فحص فعالة للمراحل المبكرة من سرطان المبيض (المرحلة الأولى والثانية). في هذه الدراسة، تم تطبيق طريقة السيليكو (*in silico*) لتصميم دبابيس الشعر *DNA aptamer* للكشف عن بروتين البربخ البشري *4(HE4)* ، وهو علامة بيولوجية *OC* . تم دعم نتائج العمل في السيليكو من خلال اختبار في المختبر. بدأ العمل بنمذجة بروتين *HE4* باستخدام خوادم التنبؤ ببنية البروتين *AlphaFold* و *I-TASSER* و *Robetta*. وفي وقت لاحق، تم إجراء محاكاة الديناميكيات الجزيئية (*Molecular Dynamics-MD*) على كل نموذج متوقع لمدة *100 ns* باستخدام مجال القوة *OPLS* لتحسين الهيكل. تم التحقق من صحة جودة الهيكل الثالث بواسطة *PROCHECK*، و*ERRAT* مما يوضح أن النموذج المكرر من *AlphaFold* ، و*RF1* كان أعلى جودة من حيث التشكل الثالثي *HE4* .تم العثور على جميع الأحماض الأمينية في المناطق المفضلة في مؤامرة راماشاندران *(Ramachandran plot)* مع درجة الجودة الإجمالية لـ *ERRAT* البالغة *97.701*. بعد ذلك، تم إرساء هيكل *HE4* هذا باستخدام *AutoDock Vina* مع أربعة مرشحين لـ*HE4 aptamer (A1*، و *A2*، و *A3*، و *A4)*. كانت طاقة الربط *HE4-A4 -6.0* كيلو كالوري/مول، وشكل المجمع 24 رابطة هيدروجينية *(H-bonds)*، تم تحديد 5 منها في منطقة حلقة دبوس الشعر الأبتمر *(aptamer hairpin loop region)*. تم اختيار *A4* باعتباره المرشح الأكثر ملاءمة لاستخدامه في تصميم دبوس الشعر *(DNA hairpin)*، حيث أظهر تقاربًا جيدًا مع أكبر عدد من روابط هيدروجينية في حلقة دبوس الشعر. لبدء تصميم السيليكو لدبابيس الشعر الجديدة، تم اقتطاع أبتمر *A4* ذو 25 مير في منطقة دبوس الشعر، تم تمديد *5′-CGCAAG-3′* والجذع، ليشكل تسلسل *5′-GCGCAAGC-3′*. تم تحور نيوكليوتيدات الحلقة *(nucleotides loop)*، *-GCAA-*،

بشكل استبدالي، منتجة 256 تسلسل. تم إرساء دبابيس الشعر ال 256 هذه مع *HE4* بشكل فردي، وأظهرت *H16* و *H101*و *H256* ارتباطات ارتباط جيدة مع طاقات ربط تتراوح بين -10.6 و-11.6 كيلو كالوري/مول. ونتيجة لذلك، تم تطبيق محاكاة *100 MD ns* باستخدام مجال القوة *CHARMM27* على مجمعات *HE4-H16* و *HE4-H101*و *HE4-H256*. استنادًا إلى *RMSD* ،ونصف قطر الدوران، وعدد الروابط الهيدروجينية ، وشغل السندات الهيدروجينية، والطاقة الإجمالية، تم استنتاج *H256* باعتباره دبوس الشعر الواحد للحمض النووي ضد علامة *HE4* مع تقارب واستقرار كبيرين طوال المحاكاة. تم تصنيع دبوس الشعر *H256* هذا وتم تحليل ارتباطه بـ *HE4* عبر اختبار التحول الحركي الكهربي (*Electrophoretic Mobility Shift Assay-EMSA*) في المختبر. استنادًا إلى شدة نطاق الحمض النووي *(DNA)*، يرتبط *H256* المصمم *(3.27%)* بـ *HE4* أفضل بأربع مرات من *A4 (0.84%)* من الدراسة السابقة. أخيرًا، تم إجراء دراسة أولية لإمكانات التشخيص المستقبلية عن طريق اقتران جسيمات الذهب النانوية *(Gold Nanoparticle-GNP)* مع *H256* . أكد *FTIR* وأطياف *Raman* وجود مجموعة الأميد *(amide)*، التي تكونت من اقتران الناتج القومي الإجمالي الكربوكسيلي (*carboxylated GNP*) مع الأمين *H256-*. تغير حل الناتج القومي الإجمالي من الأحمر إلى الأحمر الأرجواني، مما يشير إلى زيادة الحجم بعد الاقتران. وقد تمّ تأكيد ذلك من خلال جهاز تحليل حجم الجسيمات. في الختام، *H256* هو دبوس الشعر الواحد للحمض النووي في فحص*HE4* ، ويُقترح باستخدام هذه التقنية في التطوير المستقبلي لمجموعة تشخيصية كاملة الوظائف لسرطان المبيض، ومناسبة للاستخدام في الفحص الروتيني لجميع النساء، سواء مع أو بدون أعراض. وهذا من شأنه أن يحسن الكشف بين المرضى في المراحل المبكرة (الأولى والثانية)، مما يعزز نتائج المرضى.

APPROVAL PAGE

The thesis of Nur Nadiah Abdul Rashid has been approved by the following:

Izzat Fahimuddin Mohamed Suffian
Supervisor

Azzmer Azzar Abdul Hamid
Co-supervisor

Mohd Hamzah Mohd Nasir
Co-supervisor

Nurasyikinhamzah

Nurasyikin Hamzah
Co-supervisor

Muhammad Salahuddin Haris @ Harith
Internal Examiner

Saharuddin Mohamad
External Examiner

Khairani Idah Mokhtar @ Makhtar
Chairman

DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Nur Nadiah Abdul Rashid

Signature……………………………… Date 10/10/2024……………………………...

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

IN SILICO SELEX IN DESIGNING NOVEL DNA APTAMER HAIRPIN FOR *IN VITRO* DETECTION OF HE4: AN OVARIAN CANCER BIOMARKER

I declare that the copyright holder of this thesis is jointly owned by the student and IIUM.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.

2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.

3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledge that I have read and understand the IIUM Intellectual Property Right and Commercialisation policy.

Affirmed by Nur Nadiah Abdul Rashid

……………………………
Signature

10/10/2024
……………………….
Date

This thesis is dedicated to my family, whose unwavering support and understanding have been the cornerstone of my Ph.D. journey.

ACKNOWLEDGEMENT

All praise belongs to Allah, the Almighty, for His endless Grace and Mercy that have accompanied me throughout my academic journey. Despite the challenges faced, it is His Merciful blessings that have lightened the burdensome task of finalising this whole research.

To my esteemed supervisors, Dr. Izzat, Assoc. Prof. Dr. Azzmer, Dr. Hamzah, and Dr. Syikin, your mentorship has been nothing short of transformative. Your expert guidance, constructive criticism, and unwavering belief in my potential have not only shaped this thesis but also my growth as a scholar and individual. Your dedication to excellence and nurturing of my academic curiosity have left an indelible mark on my intellectual journey. I am eternally grateful for your generosity of knowledge, time, and encouragement at ensuring the best quality of our research. I am thankful for the moral support the supervisors have provided, serving as a boost, particularly during my low moments, and they have undoubtedly played a significant role in the development and completion of this work.

My heartfelt gratitude goes out to my husband, Hijaz Ridzwan, and my little treasures, Aiman Yusuf, Adam Mikail, and Auni Sofiyya. Your unconditional love, unwavering support, and endless patience have not only been the cornerstone but also the sanctuary that provided me solace during the most challenging phases of my journey. Lastly, my enduring gratitude goes to my loving parents, who have been the guiding light, illuminating the path even in the darkest of times. Your faith in me has been the wellspring of my strength and determination. And not to forget, the many individuals who have helped me in realising this dream, sincere thanks to each of you.

TABLE OF CONTENT

CHAPTER ONE: INTRODUCTION

CHAPTER TWO: LITERATURE REVIEW

CHAPTER THREE: THE MODELLING OF HE4 PROTEIN TERTIARY STRUCTURE USING PROTEIN STRUCTURE PREDICTION SERVERS AND ITS VALIDATIONS

CHAPTER FOUR: SCREENING OF THREE-DIMENSIONAL DNA APTAMERS AGAINST HE4 PROTEIN THROUGH MOLECULAR DOCKING AND ITS BINDING ANALYSIS

CHAPTER FIVE: THREE-DIMENSIONAL DNA APTAMER HAIRPIN DESIGN BY MUTATION, MOLECULAR DOCKING, AND MOLECULAR DYNAMICS SIMULATION

CHAPTER SIX: *IN VITRO* DETECTION OF HE4 BIOMARKER USING THE DESIGNED DNA APTAMER HAIRPIN, AND GOLD NANOPARTICLE-DNA CONJUGATES FOR FUTURE CANCER DIAGNOSTIC

6.2 Methodology

6.3 Results and Discussion

CHAPTER SEVEN: GENERAL CONCLUSION AND RECOMMENDATIONS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

3D	Three-dimensional
A	Adenine
AD4	AutoDock 4
ADT	AutoDock Tools
AFP	Alpha-fetoprotein
Apo	Apolipoprotein
C	Cytosine
CA125	Cancer antigen-125
CA15-3	Cancer antigen 15-3
CA19-9	Cancer antigen 19-9
CA72-4	Cancer antigen 72-4
CASP	Critical Assessment of Structure Prediction
CEA	Carcinoembryonic Antigen
CHARMM	Chemistry at Harvard Molecular Mechanics
ChIP	Chromatin immunoprecipitation
C-score	Confidence score
CT	Computed tomography
DNA	Deoxyribonucleic acid
DpAu	Deposited gold nanocrystals
EGFR	Epidermal growth factor receptor
EMSA	Electrophoretic mobility shift assay
ePAD	Electrochemical paper analytical device
FIGO	Federation of Gynaecology and Obstetrics
G	Guanine
GTD	Gestational trophoblastic disorders
H-bond	Hydrogen bond
hCG	Human chorionic gonadotropin
HDI	Human Development Index
HE4	Human epididymis protein 4

HGSC	High-grade serous carcinoma
I-TASSER	Iterative threading assembly refinement
LGSC	Low-grade serous carcinoma
LRET	Luminescence resonance energy transfer
MD	Molecular dynamics
MMGB/SA	Molecular Mechanics/Generalized Born Surface Area
MRI	Magnetic resonance imaging
NCRD	National Cancer Registry Department
OC	Ovarian cancer
OPLS	Optimised Potentials for Liquid Simulations
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDGF	Platelet-derived growth factor
PET/CT	Positron emission tomography/computed tomography
pLDDT	Per-residue local distance difference test
PLIP	Protein-Ligand Interaction Profiler
pTM-score	Predicted TM score
Rg	Radius of gyration
RMSF	Root-mean-square fluctuations
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
SELEX	Systematic evolution of ligands by exponential enrichment
SPR	Surface plasmon resonance
T	Thymine
TVUS	Transvaginal ultrasound
UCNP	Upconverting nanoparticles
UKCTOCS	United Kingdom Collaborative Trial of Ovarian Cancer Screening
Vina	AutoDock Vina
VMD	Visual Molecular Dynamics
WFDC2	Whey acidic four disulphide core 2 protein
ZEA	Zearalenone

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF STUDY

Ovarian cancer (OC) is a serious health concern with an estimation of 314,000 new cases and approximately 207,000 deaths worldwide in 2020 (Reid, 2023). The Global Cancer Observatory anticipated 359,000 new cases globally among women aged 20 to 84 in 2030, with Asia accounting more than 50% with 201,000 cases, as it is the most populated continent (World Health Organization, 2024). In Malaysia, there were more than 3500 cases reported between 2012 to 2016 (Manan et al., 2019). The disease is frequently diagnosed at an advanced stage because of vague clinical symptoms at early stages of the disease, which complicates effective screening and early diagnosis, resulting in 5-year survival rate of less than 30% (Lu et al., 2022; Lv et al., 2023; Yang et al., 2021). The subtle symptoms of early stages OC may include nausea, diarrhoea, and constipations, which leads to confusion with other gastrointestinal issues (Babaei et al., 2022). In late stages of OC, patients experience severe abdominal pain, fatigue and unexplained weight loss (Andreou et al., 2023). Early diagnosis and treatment of OC pose significant challenges as the current OC diagnostic approaches have been a subject of extensive research and clinical trials.

In the past decade, the biomarker cancer antigen-125 (CA125) and transvaginal ultrasound were commonly utilised for the diagnosis of OC (Doubeni et al., 2016; Elias et al., 2018; Stewart et al., 2019). However, studies have shown that CA125 is not sensitive enough to identify OC when it is limited to the ovary, while transvaginal ultrasound does not consistently distinguish between benign and malignant ovarian tumours (Sopik et al., 2015; Stirling et al., 2005). Additionally, the specificity of CA125 among 120 patients, including benign tumours and ovarian carcinomas patients, was reported to be 62.50 % (Ali et al., 2022).

Ongoing research on the molecular basis of the disease provides potential for improvements in the future. For instance, the combination measurement of CA125 biomarker with human epididymis protein 4 (HE4) have shown increased efficiency in OC diagnosis (Dochez et al., 2019; Liang et al., 2020). Multiple studies have proven that HE4 has the potential to be a specific biomarker for OC, distinguishing between malignant ovarian tumours and benign diseases such as ovarian endometriotic cysts (Anastasi et al., 2013; Huhtinen et al., 2009; Moore et al., 2012). This HE4 protein which is highly expressed in OC cells has demonstrated more specificity than the standard CA125 biomarker for detecting OC which suggests its potential to enhance the diagnostic precision (Aafreen et al., 2022; Kumarasamy et al., 2019; Nahar et al., 2019). Whether HE4 is used alone or in conjunction with other biomarkers, this protein improves the accuracy of the diagnoses (Kang et al., 2022).

Previously, high-affinity monoclonal antibodies were used in detecting HE4 in human serum samples (Shen et al., 2018; Zhou et al., 2016). However, recent years have seen the evolution of aptamer-integrated assays in the diagnostic field as aptamers offer various advantages over antibodies, including easy synthesis, non-immunogenicity, and greater cost-effectiveness (Röthlisberger & Hollenstein, 2018). Aptamers are single-stranded DNA or RNA molecules that exhibit great specificity, affinity, and stability, making them valuable tools for cancer detection and treatment (Ma et al., 2019; Takita et al., 2021). They are also known as chemical antibodies with unique secondary and tertiary structures that allow them to recognise the specific target, thus, functioning similarly to antibodies (Hori et al., 2018). Furthermore, unlike antibodies, these synthetic oligonucleotides can be chemically modified to precisely bind to various cancer biomarkers, including HE4 and CA125, and they hold great promise for cancer diagnostics (Elskens et al., 2020; Hanzek et al., 2021). The modifications of aptamers improve their durability in biological samples, making them more effective in various diagnostic platforms, from basic colourimetric assays to advanced biosensor approaches. Additionally, the chemical synthesis of aptamers guarantees uniform quality and availability, overcoming a typical issue linked to manufacturing biological reagents such as antibodies (Bauer et al., 2019; Shigdar et al., 2013).

Typically, the selection of oligonucleotides with high affinity to their target applies *in vitro* molecular biology approach called Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Byun, 2021; Zhu et al., 2024). A conventional SELEX process requires multiple cycles of bound aptamer elution and amplification of oligonucleotide library, which might take days or months to complete (Wang et al., 2019). Consequently, computational approach or *in silico* SELEX was introduced as it offers a ground-breaking alternative to conventional SELEX techniques for discovering and refining aptamers. This approach uses advanced algorithms and molecular modelling to replicate the iterative process of binding, partitioning, and amplification seen in physical SELEX, but in a virtual environment (Navien et al., 2021; Rasouli Jazi et al., 2023; Sabri et al., 2019). *In silico* SELEX offers several advantages over traditional SELEX such as improved speed, efficiency, cost reduction, and enhanced affinity and specificity (Douaki et al., 2022). The time can be shortened to few hours, significantly faster than the time-consuming traditional SELEX method, designing good affinity aptamers, investigated through molecular docking and MD simulations.

The *in silico*-designed DNA aptamer or DNA aptamer hairpin, expected to exhibit strong binding affinity and interactions with HE4 protein will be verified *via in vitro* Electrophoretic Mobility Shift Assay (EMSA). This confirms the conjugation of the designed hairpin with HE4 protein (Cozzolino et al., 2021). Following this, the conjugation of the DNA aptamer hairpin on the surface of gold nanoparticles (GNP) and its characterisation, including the analysis of particle size, functional group spectral, and formulation stability serves as the groundwork in developing reliable OC diagnostic technique. From this perspective, with the advancement of aptamer-based diagnostics in research and their anticipated significant contribution to personalised medicine for OC, investigating the design of DNA aptamers against the HE4 protein through an *in silico* approach becomes essential.

1.2 PROBLEM STATEMENT

Owing to its asymptomatic character in the early stages and a lack of efficient early detection approaches, OC remains one of the most fatal gynaecological cancers globally (Zhang et al., 2022). The HE4 protein has emerged as a promising OC biomarker, with greater specificity than the previously utilised CA125 biomarker, particularly in distinguishing ovarian malignancy among pre- and post-menopausal women (Teh et al., 2018; Zhang et al., 2019). Up-to-date, the crystallised structure of HE4 protein has not been discovered, making it essential to elucidate the structure where the potential binding affinity and binding interactions with other biomolecules can be investigated (Petsko & Ringe, 2004). This causes limited understanding of the HE4 structural complexities, leading to present challenges in developing a highly affinity DNA aptamer against this protein, that can be potentially used in the development of a diagnostic technique. In consequence, this current work aims to predict the tertiary structure of HE4 protein using various structure prediction servers; AlphaFold, Iterative threading assembly refinement (I-TASSER), and Robetta. It is imperative to enhance these structural models and verify their structural quality to precisely target HE4 with the DNA aptamers, potentially resulting in improved detection and diagnostic applications.

Despite HE4's potential to improve early detection rates, existing clinical OC diagnostic approaches have considerable hurdles, such as complicated laboratory processes and the requirement for highly specialised equipment and staff (Barr et al., 2022; Huang et al., 2017). These limitations highlight the critical need for developing novel techniques for detecting HE4 with high specificity and sensitivity. The presence of reliable detection tools will improve patient outcomes, leading to lower mortality rates. In this context, DNA aptamers, that can attach to specific targets with great affinity, represent a viable path for HE4 biomarker detection. Aptamers have various benefits over traditional antibodies, including simplicity of synthesis and high stability, to bind with HE4 protein (Shaban & Kim, 2021). However, traditional techniques of aptamer selection, such as SELEX, are time-consuming and labour-intensive, other than involving complicated process, which at times fail to produce aptamers with the requisite specificity and affinity for their targets (Liu & Yu, 2018).

The process in selecting the aptamers with high affinities towards its target in traditional SELEX consists of multiple cycles, typically 5 to 15 cycles, takes days to months to complete (Komarova & Kuznetsov, 2019). Additionally, the whole process consisting of binding, separation, amplification, and iteration steps involves significant consumption of materials, making it a resource-demanding process. This study addresses the gap in the development of aptamer-based diagnostics by employing a computational approach to design novel DNA aptamer hairpins, ensuring robust interactions with the HE4 protein for reliable detection.

The development of *in silico* techniques provides a game-changing approach to aptamer design, possibly overcoming the limits of the traditional SELEX. *In silico* design simulates aptamer-target interactions, allowing for faster screening and minimise the intensive labour for optimisation of aptamer sequences with high affinity for HE4 protein (Lee et al., 2023). This computational technique has the potential to drastically decrease the time and cost involved with experimental aptamer selection, paving the door for the fast development of a new generation of diagnostic techniques. This work uses advanced computational techniques, involving HE4 protein modelling, molecular docking, and molecular dynamics simulations to determine aptamer sequences with optimum HE4 binding properties. The application of substitutional mutations of the aptamer nucleotides is to enhance the binding affinity between the DNA aptamer hairpin with its target, HE4 protein.

Subsequently, *in vitro* EMSA assay takes place in verifying the binding of the designed DNA aptamer hairpin with HE4 protein, the OC biomarker in this study (Cozzolino et al., 2021). This step is essential to validate the efficiency of the hairpin, ensuring strong binding capability to its target. The successful design and validation of such aptamer hairpin has the potential to transform OC diagnostics by providing a quick, affordable, and non-invasive technique for detection of the illness. The potential impact goes beyond OC which may lead to the development of personalised diagnostic tools that focus on specific biomarkers, tackling the inherent limitations of existing OC diagnostic methods. The combination of *in silico* DNA aptamer design and *in vitro* validation has the ability to enhance the development of diagnostic tools for various diseases, influencing wider fields of molecular diagnostics and personalised medicine.

1.3 OBJECTIVES OF STUDY

The general objective of the study is to develop a detection method for HE4, an ovarian cancer biomarker, through the *in silico* design of potential DNA aptamer hairpin. To achieve this general objective, four specific objectives are involved:

1. To design the three-dimensional structure of HE4 protein with protein prediction servers, refine its structure using molecular dynamics (MD) simulation, and validate the protein structures.
2. To determine the most suitable DNA aptamer against HE4 protein through molecular docking and its binding analysis.
3. To enhance the DNA aptamer hairpin using substitutional mutation method, estimate the binding affinities through molecular docking, and study their dynamics by MD simulations.
4. To verify the bioconjugation of the designed DNA aptamer hairpin with HE4 protein using *in vitro* electrophoretic mobility shift assay (EMSA) and formulate

a gold nanoparticle-DNA aptamer hairpin approach for preliminary HE4 detection method studies.

1.4 RESEARCH QUESTIONS

The primary aim of this study is to design and validate a novel DNA aptamer hairpin targeting the HE4 biomarker by integrating *in silico* and *in vitro* approaches. The following research questions are designed to address the key aspects of aptamer design and its diagnostic development:

1. What are the structural differences of the HE4 protein conformations modelled by AlphaFold, I-TASSER, and Robetta protein prediction servers?
2. How do the binding energies and hydrogen bonding patterns of the DNA aptamer candidates differ when docked with the HE4 protein?
3. Which factors contribute to the enhanced binding affinity of the designed DNA aptamer hairpin towards HE4 during MD simulations?
4. How does the *in vitro* assay binding performance of the designed DNA aptamer hairpin compared to the positive control aptamer in detecting the HE4 protein; and what are the potential implications of the GNP-hairpin conjugate’s spectral and particle size changes for the future development of a diagnostic technique for OC?

These elements offer a systematic framework for assessing the effectiveness of the computationally-designed hairpin in OC diagnosis, establishing a foundation for future study and advancement in this area.

1.5 RESEARCH HYPOTHESES

Based on the above research questions, the relevant research hypotheses can be outlined as follows.

1. The conformations of the HE4 protein, as modelled by AlphaFold, I-TASSER, and Robetta protein prediction servers, display significant structural variations that influence their stability, compactness, and suitability for binding with DNA aptamer hairpins due to different algorithms of each server.
2. The binding energies and hydrogen bonding patterns of the DNA aptamer candidates display significant variance when docked with the HE4 protein, affecting their overall binding affinity and binding poses.

3. Several factors, including root-mean-square deviation (RMSD), radius of gyration (Rg), and hydrogen bond occupancy, improve the binding affinity of the designed DNA aptamer hairpin towards HE4 during MD simulations.
4. The designed DNA aptamer hairpin is expected to display enhanced *in vitro* binding efficacy compared to the positive control aptamer in detecting the HE4 protein, indicating its better diagnostic potential, while the observed spectral and particle size changes in the H256-GNP conjugate will support its improved detection abilities, indicating its potential for future advancement into a diagnostic kit for OC.

1.6 SIGNIFICANCE OF THE RESEARCH

The significance of this study lies primarily in its potential to improve the detection methods for OC, a disease frequently identified at advanced stages due to inadequate screening technologies. This work advances the OC screening by developing a new DNA aptamer hairpin that targets the HE4 biomarker, which is essential for improving survival rates. The integration of *in silico* and *in vitro* approaches enhances the development of diagnostic tools that are predicted to be more sensitive and specific, offering a substantial improvement over existing OC detection technique.

Furthermore, this study emphasises the innovative application of *in silico* methods, including protein modelling and MD simulations, providing a cost-effective and efficient method for designing potential diagnostic agents. This approach establishes a benchmark for similar applications in other diseases, broadening the significance of the study beyond OC. The promising findings of the computationally-designed DNA aptamer hairpin sets a foundation for the development of a commercial diagnostic kit, potentially revolutionise existing clinical practices and enhance the outcomes of the OC patients. The potential for commercialisation highlights the practical significance of the study, establishing it as a key contributor to future diagnostic innovations.

CHAPTER TWO

LITERATURE REVIEW

2.1 OVARIAN CANCER

OC is a substantial health challenge for women globally, originates in the ovaries which is the female reproductive glands. This disease has several subtypes, each with different molecular compositions, cells of origin, reactions to treatment, and prognosis (Matulonis et al., 2016). The main subtypes include epithelial (~90%), stromal tumours (<5%), germ cell (<5%), and mixed-cell (~1%) type. Epithelial OC originates on the outside of the ovary, whereas stromal and germ cell OCs arise from hormone-producing ovarian cells and the egg cells, respectively. Although most OC cases are categorised as epithelial OC, there are various histological subtypes within this type of cancer, including serous (~70%), endometrioid (~10%), clear cell (10%), mucinous (~3%), malignant Brenner tumours (~1%), and mixed histology (~6%) (Figure 2.1) (Rojas et al., 2016). The epithelial OC mainly affects the ovaries, which are important for hormone productions and egg development (Lheureux et al., 2019).

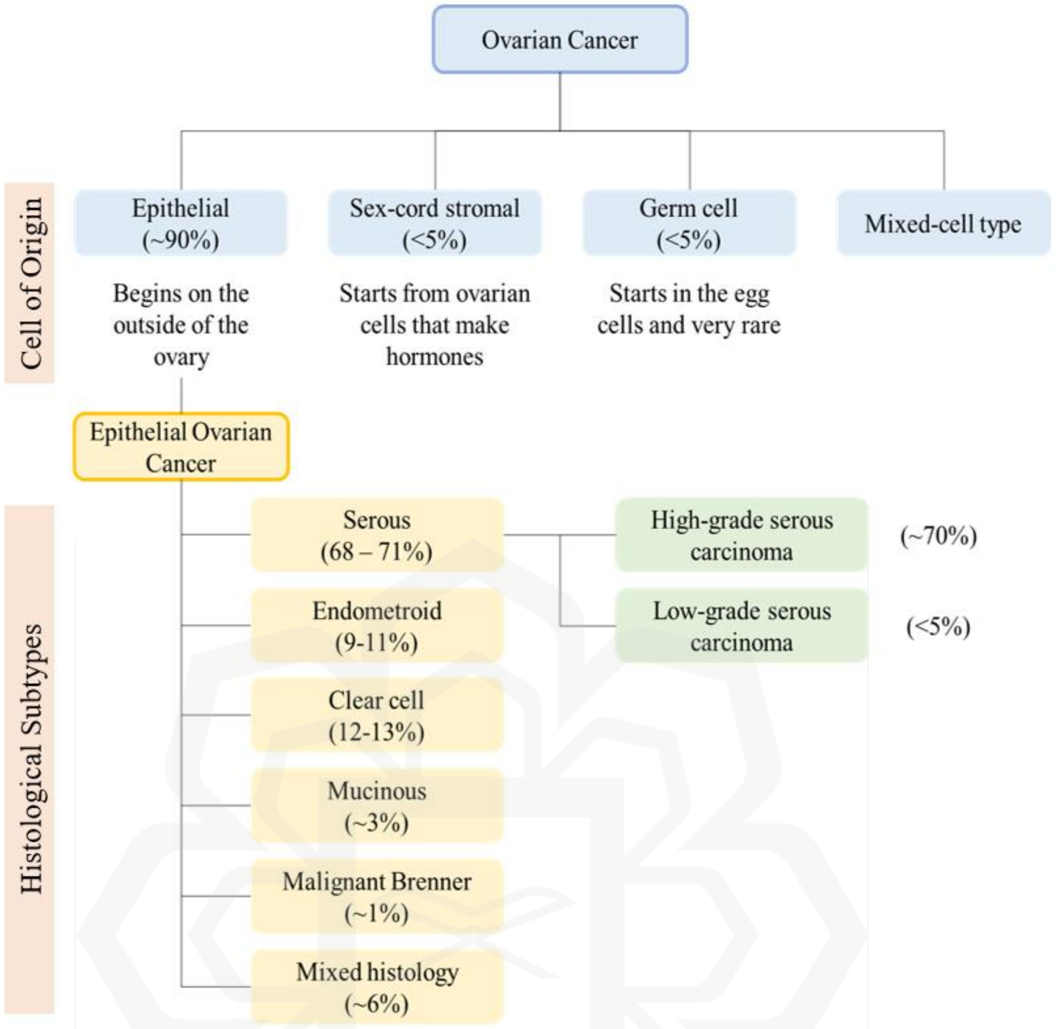


Figure 2.1 Types of OC based on the cell of origin, and the epithelial OC classes based on tumour histology. The occurrence percentages in all ovarian carcinomas are stated in brackets. The figure summarises the findings from Rojas et al. (2016).

Recent progress in the molecular characterisation of tumours, alongside clinicopathologic and molecular analyses of OCs, has led to the classification of epithelial OC into two main groups, Type I and Type II carcinomas (Kurman & Shih, 2016). Type I includes low-grade serous carcinoma (LGSC), endometroid, clear-cell, and mucinous carcinomas, and malignant Brenner tumour. The high-grade serous carcinoma (HGSC) and carcinosarcomas are classified as Type II. The clinical features of Type I OCs include slow growth, less aggressive clinical presentation, and highly possible to be diagnosed at stage I and II. Meanwhile, the Type II OCs are known for its rapid growth, aggressive clinical behaviour and mainly diagnosed at advanced stages, stage III and IV (Zhu et al., 2022).

According to the staging guideline proposed by the Federation of Gynaecology and Obstetrics (FIGO), stage III and IV are characterised by the spreading of cancer beyond the pelvis and to distant organs such as livers and lungs, respectively (Prat et al., 2015). It is at these advanced stages that patients experience symptoms such as abdominal or pelvic pain, bloating or increasing abdominal size, and unexplained weight loss (Gaona-Luviano et al., 2020). The disease's elusive characteristics, often manifesting non-specific symptoms or staying asymptomatic until reaching advanced stages (FIGO stage III and IV), have posed challenges in early diagnosis. The FIGO staging highlighted the confinement of cancer in one or both ovaries for stage I (Figure 2.2a), and the cancer spreads to other organs in the pelvic region when it reaches stage II (Figure 2.2b). Unfortunately, the cancer spreads even further, beyond the pelvis to the lining of the abdomen or to nearby lymph nodes for OC stage III (Figure 2.2c). And the last stage (stage IV), which is the worst stage have shown the spreading of cancer to distant organs as mentioned earlier (Figure 2.2d) (Fernandez-Garza et al., 2021).

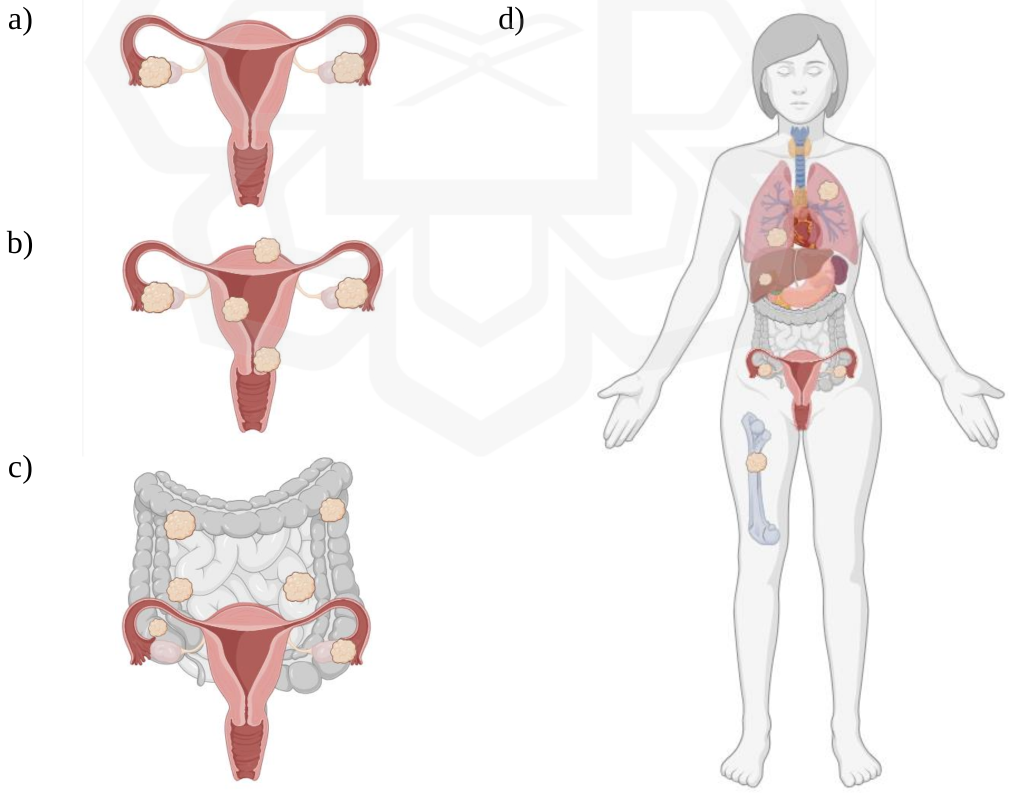


Figure 2.2 Cancer spread based on the FIGO staging of OC. The light orange-coloured lumps represent the spread of cancer or tumour. (a) Stage I, (b) Stage II, (c) Stage III, and (d) Stage IV. The diagram was illustrated using Biorender.com based on descriptions by Prat et al. (2015) and Fernandez-Garza et al. (2021).

Globally, from 1990 to 2017, the incidence of OC cases increased from 152,100 to approximately 286,100, while the fatalities rose from 95,500 to 176,000 (Zheng et al., 2020). These numbers of OC cases and fatalities increased to 314,000 and 207,000 for the year 2020 and expected to double by 2040, based on Human Development Index (HDI) and global assessments of 185 countries (Cabasag et al., 2022). OC ranks as the eighth most prevalent cancer worldwide and the fifth highest cause of cancer-related deaths in women, highlighting the serious threat to women's health and survival (Zhang et al., 2022). Although there have been improvements in therapy and diagnosis, the 5-year survival rate for OC remains poor, particularly for individuals diagnosed at late stages. The survival rates exhibit substantial variation depending on the stage of diagnosis, with early-stage identification providing a much more favourable prognosis in contrast to advanced-stage illness. More than 80% of ovarian carcinosarcomas, LGSCs and HGSCs were detected at stage III and IV, with 5- and 10-year survival rates of less than 20% and 40%, respectively (Hollis, 2023). Although the overall survival rate is low, it has a 90% 5-year survival rate for patients who were diagnosed at early stages (FIGO stage I and II) (Žilovič et al., 2021).

The common clinical presentation of OC includes the detection of an adnexal mass through pelvic examination or imaging methods, initiated by the pelvic pain or discovered incidentally during a routine medical check-up (Gaona-Luviano et al., 2020). This is one of the reasons the OC was considered as a silent disease, in which the adnexal masses are rarely noticed by the patients in early stages. Until today, detecting early-stage OCs has been challenging with the absence of public screening programmes. Currently, over 200 tumour markers are associated with OC but only cancer antigen-125 (CA125) and human epididymis protein 4 (HE4) markers have been utilised clinically (Savinova & Gataullin, 2022). Various studies have proposed the use of a combination of multiple biomarkers to increase the specificity and sensitivity of OC diagnosis, including cancer antigen 19-9 (CA19-9) and cancer antigen 15-3 (CA15-3) alongside HE4 and CA125 (Dochez et al., 2019; Matsas et al., 2023). The additional imaging modalities like transvaginal ultrasound and positron emission tomography/computed tomography (PET/CT) scans, enhance the confirmation of the disease (Kemppainen et al., 2019; Khiewvan et al., 2017). The various OC biomarkers and current diagnosis approaches are further discussed in the next subsections.

2.2 PREVALENCE OF OVARIAN CANCER IN MALAYSIA

In 2016, the National Cancer Registry Department (NCRD) of Ministry of Health Malaysia reported the Malaysian population was 31.60 million with 55.6% Malays, 23.4% Chinese, 13.0% Bumiputera, 7% Indians, and the rest belonged to other ethnic groups. There were 115,238 new cancer diagnoses between 2012 and 2016, with females accounting for 63,733 cases, representing 55.3% of the total. The most common cancers in females were breast cancer, followed by colorectal, cervix uteri, and ovary. The NCRD recorded 3,575 cases of OC between 2012 to 2016, 3% higher compared to the 2007 to 2011 report (3,472 cases). No latest report has been published by NCRD ever since. The highest number of OC was recorded in 2016 with 776 cases. Overall, the trend showed increment from 2013 to 2016 with 668, 696, 748, and 776 reported cases. Based on this NCRD report, OC is the tenth common cancer and fourth among females in Malaysia, which 56.3% were identified at stage III and IV. This was similar to the previous 2007 to 2011 report and the highest incidence occurred between age 65 and 69 for both reports. The incidence rate in 100,000 population of each ethnic was highest among the Chinese, followed by Indian and Malay between 2007-2011. However, the 2012 to 2016 report showed otherwise, Malay accounting the highest incidence rate (per 100,000 ethnic population) with a total of 1,984 cases (Table 2.1) (Manan et al., 2019).

Table 2.1 Ovarian cancer incidence summary by report year and ethnicity in Malaysia.

NCRD report year		2007 to 2011	2012 to 2016
Ethnicity	Malay	1,754	1,984
	Chinese	1,107	1,006
	Indian	277	272
	Others	334	313
TOTAL		3,472	3,575

In a case study reported in Sabah, a woman aged 58 was presented with symptoms such as constipation, lower abdominal pain and mass, and frequent urination. She was diagnosed with stage IV OC after various screening test were conducted, including ultrasound-guided biopsy, CT scan and tumour marker tests (Pei Yin, 2019). The CA125 level was 983.4 U/ml, significantly higher than their normal cut-off limit of 35 U/ml. Here, the CA125 level corresponds to the advanced-stage of the OC. This case highlights the reality that OC often undiagnosed until it has progressed significantly, emphasising the urgent need for more efficient OC screening techniques to enhance survival rates.

In a separate study, a survey was conducted by the Department of Community Health, Faculty of Medicine and Health Science, Universiti Putra Malaysia involving 366 female participants. The survey gathered information on the participant's awareness of OC, including its symptoms, risk factors, and screening approaches (Elmahdi et al., 2017). The study found a lack of awareness, influenced by certain socio-demographic traits of the participants. A comparable study was also carried out by School of Health Sciences, Universiti Sains Malaysia with 110 participants, yielding similar results (Hashim et al., 2021). Thus, to bridge this knowledge gap, these studies recommend the implementation of OC awareness and educational programmes, focusing on alerting women to the symptoms, clinical screening methods availability, while highlighting the risk factors. Lastly, the coping strategies among patients diagnosed with recurrent OC aged 52 to 84 were analysed, revealing their coping mechanisms involved maintaining a hopefulness mindset, seeking spiritual assistance, and relying on family support (Lee et al., 2021). Based on the limited studies mentioned, it is revealed that Malaysia is still lacking in the statistic and information concerning OC. Given that Malaysia is among the leading countries in OC occurrence, there is a pressing need for more comprehensive studies to be conducted and thoroughly documented.

In light of the worrying OC statistics and their fatality numbers, this study aims to address the urgent need to enhance the detection methods by designing DNA aptamer hairpin with high binding affinity towards OC biomarker, HE4. This research sets a groundwork of developing potential OC screening technique which hopefully can be utilised for routine screening in all women, regardless of whether the OC symptoms are present or not.

2.3 OVARIAN CANCER BIOMARKERS

Given that OC often progresses silently and is commonly diagnosed at an advanced stage, early detection and accurate diagnosis are essential for enhancing survival rates. Currently, CA125 and HE4 stand as the predominantly utilised biomarkers for OC (Wang et al., 2021). A thorough review highlights the significance of an array of biomarkers in the diagnostic process for OC, including cancer antigen 19-9 (CA19-9), cancer antigen 15-3 (CA15-3), and human chorionic gonadotropin (hCG), other than the two known biomarkers (Matsas et al., 2023).

2.3.1 Cancer Antigen-125 (CA125)

CA125 is a protein typically found on OC cells and it is the most thoroughly studied biomarker for OC, playing a crucial role in the diagnosis and treatment of the illness (Zhang et al., 2021). The serum CA125 assay showed limited sensitivity in the early stages and exhibited higher levels in certain conditions, including menstruation and endometriosis (Dochez et al., 2019). The upper limit of CA125 is 35 U/ml and has not changed since the first report in 1983 (Bast et al., 1983). The CA125 assay conducted in that study demonstrated only 1% of 888 healthy persons had CA125 level higher than 35 U/ml. In contrast, elevated antigen levels were observed in 83 out of 101 patients (82%) who had been confirmed to have ovarian carcinoma through surgeries. However, the CA125 levels are also elevated during menstruation, premenstrual phase, and at certain pregnancy phases (Charkhchi et al., 2020). In a recent study involving 131 healthy control and 131 patients, CA125 levels were not significant among stage I and II patients, with mean values of 39.8 and 56.2 U/ml, respectively. However, the mean values elevated drastically for patients with stage III and IV OC, recorded at 193.8 and 573.1 U/ml, respectively (Bilonia et al., 2022).

Although CA125 is the main tumour marker for OC, however its predictive accuracy is not ideal. CA125 testing was conducted on 50,780 women where 3.4% under the age of 50 and 15.2% aged over 50 had CA125 levels higher than the upper limit (35 U/ml) for healthy women with non-ovarian cancer (Funston et al., 2020). Among women with elevated CA125 levels, only 10.1 % were diagnosed with OC, while 12.3% were found to have different types of cancer. CA125's specificity is limited due to its potential to elevate in several common benign gynaecologic or nongynaecologic medical conditions, such as pelvic inflammatory disease and uterine myomas (Medeiros et al., 2009). CA125 does not possess adequate sensitivity and specificity to serve as a reliable tool for population screening to detect early-stage OC as they may not generate sufficient CA125 for detection (Cohen et al., 2014). Furthermore, as a standalone test performed on a sample size consisting of 2,305 women, this biomarker missed a significant number of ovarian malignancies, especially among premenopausal women and early-stage OCs (Dunton et al., 2021). Research shows that elevated CA125 levels are primarily correlated with tumour burden rather than the presence of early malignant changes, explaining the lack of sensitivity in detecting OC early stages cases (Jin et al., 2022). This suggests that CA125 is a useful OC marker to monitor the progress of the disease and track treatment response, and not a reliable detector for early stages OC (Zhang et al., 2021). Early detection can significantly improve the patients' outcomes and increases the survival rate for patients with OC.

Ongoing research aims to discover more biomarkers and create multi-marker screening tests to enhance the accuracy and effectiveness of detecting OC. Utilising CA125 with other biomarkers such as HE4 and cancer antigen 19-9 (CA19-9), combining with advanced imaging methods could enhance the early detection rates (Guo et al., 2019; Shin et al., 2020). Only 50% of early-stage patients have increased CA125 levels, which results in a poor sensitivity (50–62%) for diagnosing early-stage OC (Zhang et al, 2022). As a result, the accuracy of CA125 for detecting early-stage OC is limited. Furthermore, about 60% of women with elevated CA125 values do not have OC, and the specificity of CA125 is often poor (73–77%). Its use as a screening tool for asymptomatic women or those at average risk for OC is not advisable due to its low specificity and sensitivity.

2.3.2 Cancer Antigen 19-9 (CA19-9)

Cancer antigen 19-9 (CA19-9) is a glycoprotein antigen included in a panel of blood tests utilised for the diagnosis and treatment of specific cancer types. It has been recognised as a tumour marker that is elevated in biliary tract, gastrointestinal, and ovarian tumours (Ahmadzadeh et al., 2022; Cherchi et al., 1999; Scarà et al., 2015). The CA19-9 and CA125 levels were above the cut-off values of 39 U/ml and 35 U/ml, respectively, in marker tests performed on 314 individuals with mucinous ovarian tumours (Lertkhachonsuk et al., 2020). However, comparing the two markers in this previous study, CA125 marker exhibited a better diagnostic performance with a lower *P*-value ($P < 0.001$) than CA19-9 ($P = 0.002$). In another study, a 58-year-old woman experiencing dyspnoea and difficulty of walking with significant abdominal distention, was examined with multiple tumour markers. All tumour marker measurements for this postmenopausal woman were within normal range except for the CA19-9 level which was very high (953 U/ml), followed by a huge abdominal mass finding in a computed tomography (CT) scan conducted, suggesting mucinous tumour of the right ovary (Francesk et al., 2020). The sensitivity and specificity of CA19-9 in screening for ovarian cancer were reported as 50% and 97.6%, respectively (Santotoribio et al., 2016). However, the sensitivity increased by 16.7% (to 66.7%) when using both CA19-9 and CA125 markers simultaneously.

Based on the mentioned studies, CA19-9 marker alone is not effective as a screening tool for OC because it lacks the necessary sensitivity and specificity for this specific disease. The antigen can be elevated in several circumstances, including benign gastrointestinal disorders, liver diseases, and malignancies such colorectal, stomach, and hepatobiliary tumours (Scarà et al., 2015). Elevated CA19-9 levels at diagnosis in OC patients can serve as a valuable diagnostic to monitor the efficacy of treatment approaches, where a decline in CA19-9 levels may signal a favourable reaction to therapy and rising levels may indicate disease advancement or relapse. This was proven by (Francesk et al., 2020) where the CA19-9 levels of the patient had regressed to normal range, four months after surgery, and remained normal until the end of the study, approximately one-year post-surgery. The limited specificity and sensitivity of CA19-9 marker require its combination with other diagnostic approaches for a comprehensive assessment of the illness.

2.3.3 Cancer Antigen 15-3 (CA15-3)

In 1988 research, it was discovered that 41% of 180 cancer patients had high cancer antigen 15-3 (CA15-3) levels especially in late stages OC cases, which the upper-limit level of the marker was set at 30 U/ml (Scambia et al., 1988). CA15-3 is a transmembrane glycoprotein expressed on the mucosal surfaces of epithelial cells and is commonly overexpressed in various carcinomas, including breast, ovarian, lungs, and pancreatic cancers (Li et al., 2020). Although the elevated CA15-3 levels have been established for diagnosing breast cancer since 1980s, a study on the marker levels conducted on samples from 19,789 healthy individuals and patients with 30 different clinically defined diseases has shown increased CA15-3 levels, not only among breast cancer patients (Li et al., 2019). The diseases exhibiting the highest mean values were lymphoma, OC, lung cancer, and nephrotic syndrome, in descending order.

Its limited sensitivity and specificity for OC prevent it from being employed as a screening tool for the illness. Out of 38 OC patients evaluated for CA15-3 serum levels, only 42% (16 patients) had elevated CA15-3 level (above 35 U/ml) at diagnosis, indicating not all individuals with OC have increased levels of the marker (Maccaroni et al., 2017). CA15-3 had high specificity (88.27%) and low sensitivity (57.81%) in diagnosing OC malignancies, making it not suitable to be utilised as a single biomarker to diagnose OC (Li et al., 2020). However, the combination of CA15-3 with cancer antigen 72-4 (CA72-4) marker was discovered to increase the sensitivity to 73.44%. It is suggested to incorporate CA15-3 into a wider range of biomarkers, where it can improve its usefulness by offering a more thorough assessment of the disease status and treatment effectiveness. In addition to CA72-4, the incorporation of carcinoembryonic antigen (CEA) and CA125 into marker testing shows promise for early-stage OC diagnosis (Dolscheid-Pommerich et al., 2017). Continued study on biomarker expression patterns in OC and the creation of multi-biomarker panels might help elucidate the significance of CA15-3 and improve the accuracy of disease monitoring.

2.3.4 Human Chorionic Gonadotropin (hCG)

Human Chorionic Gonadotropin (hCG) is a hormone most recognised for its function in pregnancy, increasingly secreted by the placenta following implantation. Notably, the significant levels can be detected 10 days post-fertilisation in the maternal blood (Makrigiannakis et al., 2017). Aside from its known function in supporting foetal development, hCG has also been recognised as a biomarker for certain diseases, with the hCG and its components linked to gestational trophoblastic disorders (GTDs) and germ cell tumours in the ovary (Nwabuobi et al., 2017). Germ cell tumours such as choriocarcinomas, although it is less prevalent than epithelial ovarian malignancies, secretes higher level hCG than its normal range. The normal range for non-pregnant women aged 18 to 40 years old and 41 to 55 years old are less than 5 IU/L and 8 IU/L, respectively (Ostreni et al., 2022). Choriocarcinoma is a rapid-growing malignant tumour may develop in the ovaries. The hCG levels were also studied in epithelial OC where the hCG expression of the cancerous sample was significantly higher, compared to normal ovarian tissues (Zhong et al., 2019). Observably, 80 out of 139 (57.55%) epithelial OC samples had high hCG expressions and none (0/48 samples) of the normal ovarian samples were experiencing high levels of hCG.

The usefulness of hCG as OC biomarker is in its ability to be used for diagnosis, tracking therapy response, and identifying recurrence in germ cell tumours of the ovary. High hCG levels among non-pregnant women can assist in the initial diagnosis and illness stage, offering vital information that enhances imaging investigations and surgical results. A case report in 2018 involving a non-pregnant, 19-year-old female with irregular weight gain and abdominal bloating was tested for hCG, CA125, CA19-9, and CEA marker tests. The hCG level was 94 mIU/ml at initial testing and elevated to 227 mIU/ml after one week, revealed with stage III mucinous adenocarcinoma of the ovary after biopsy was conducted (Wagner et al., 2018). The hCG decreased to 69 mIU/ml and CA125 was reported to be normal at 16 U/ml after completion of surgery and chemotherapy. Serial assessments of hCG levels can be used to monitor the tumour's response to therapy during and after treatment. A decline in hCG levels signifies a positive reaction to the therapy, but consistently high or increasing levels may signal the existence of remaining illness or recurrence. Although hCG is a useful indicator for some ovarian malignancies, it is not universally relevant to all types of OC. Hence, hCG is utilised alongside other indicators and diagnostic approaches to offer a thorough assessment of ovarian tumours, such as CA125 and alpha-fetoprotein (AFP) marker tests, biopsy, and transvaginal ultrasound (Nenciu et al., 2020).

2.3.5 Human Epididymis Protein 4 (HE4)

Human epididymis protein 4 (HE4) is a notable biomarker that is receiving more attention in relation to OC. HE4 was first discovered in the epithelial cells of the distal epididymis and has since been observed to be highly expressed in several carcinomas, particularly in OCs (Granato et al., 2015). This biomarker produced by the whey acidic four disulphide core 2 protein (WFDC2) gene is a glycoprotein which belongs to a group of proteins recognised for their function in defending against proteases in the external environment, playing an important part in natural immunity. HE4 also regulates the tumour microenvironment and preserves tissue integrity, potentially having an impact on tumour growth and metastasis (Lin et al., 2021). Moore et al. (2008) reported the mean HE4 concentrations among benign disease patients and OC patients were 50 pM and 544 pM, respectively, whereas, it was 40.5 pM in healthy controls.

A study of 233 individuals with pelvic masses found that HE4 had a greater sensitivity of 72.9% compared to CA125's sensitivity of 43.3% at specificities of 67 patients with epithelial OC (Moore et al., 2009). Unlike CA125, the HE4 level was not affected by menopausal state, making it capable of differentiating between benign and malignant tumours in premenopausal individuals, and it was also suggested to be valuable for distinguishing OCs from ovarian endometriotic cysts (Anastasi et al., 2013; Huhtinen et al., 2009; Moore et al., 2012). In a comprehensive systematic review covering literature from 1952 to 2016, 49 studies comprising of 12,631 women, including 4549 OC patients, had an overall sensitivity and specificity of 71% and 86%, respectively. These studies were done for screening the borderline or malignant ovarian tumours (Scaletta et al., 2017). In a different study, the HE4 levels were measured for the samples from 149 ovarian carcinoma patients, aged 19 to 75. The sensitivity and specificity recorded were 85.3% and 91.5%, respectively, indicating similar sensitivity compared with CA125 but higher specificity (Lakshmanan et al., 2019).

A study reported in 2022 evaluated the effectiveness of HE4 marker test alone and in combination with CA125 for detecting OC in a primary care population. The combined test demonstrated the best specificity at 97.3% and lower specificity for HE4 alone (80.4%). However, the sensitivity for HE4 alone for 640 samples obtained from women aged under 50 was 100%. Overall, the study suggested that HE4 marker test performed better in women aged under 50 and exhibited higher sensitivity to detect early- and late-stage OC compared to CA125 marker test (Barr et al., 2022). This was supported by another study involving 965 women of whom 161 were diagnosed with OC while the others with benign tumours, suggesting HE4 as a valuable biomarker in diagnosing OC among premenopausal patients (Braicu et al., 2022). The finding of both studies and many others indicated that the combination of HE4 and CA125 increased the specificity and efficiency in OC diagnosis (Dochez et al., 2019; Y. Li et al., 2020; Matsas et al., 2023; Montagnana et al., 2017).

The Risk of Ovarian Malignancy Algorithm (ROMA) when combined with HE4 and CA125 enhances the accuracy of predicting OC, as it assesses the risk of OC in women with a pelvic mass by integrating the serum levels of CA125 and HE4, together with the menopausal state of the patient (Dochez et al., 2019; Wang et al., 2021; Zhang et al., 2019). High-risk scores necessitate referral to gynaecologic oncologists for further evaluation , enabling rapid identification and management of OC. The ROMA cut-off values, determining the high and low risks, is depending on the laboratory methods applied and differ for pre- and postmenopausal women (Chudecka-Głaz, 2015). In 2022, novel research modelled a panel of multiple biomarkers, comprising of HE4, CA125, CA15-3, apolipoprotein (Apo) A1, and ApoA2, to diagnose OC. This optimal combination displayed a sensitivity of 93.71% and high specificity (93.63%) (Kang et al., 2022). HE4's function extends beyond diagnosis to include monitoring illness progression and assessing response to therapy. The level of this biomarker corresponds to the efficacy of treatment, where it is decreased after effective therapy and increased in situations of recurrence (Behrouzi et al., 2021; Scaletta et al., 2017). In summary, HE4 biomarker serves as a valuable tool for initial diagnosis and the management of patients. It is highly effective as a standalone biomarker compared to CA125, CA19-9, CA15-3, and hCG, due to the ability in distinguishing malignant from benign gynaecologic conditions. However, the diagnostic accuracy can be enhanced by combining HE4 with other biomarkers and advanced-imaging approaches. Table 2.2 summarises the specificities and sensitivities of the biomarkers.

Table 2.2 The sensitivities and specificities of 5 OC biomarkers.

OC biomarker	Sensitivity (%)	Specificity (%)	Reference
CA125	50.0 – 62.0	73.0 – 77.0	Zhang et al. (2022)
CA19-9	50.0	97.6	Santotoribio et al. (2016)
CA15-3	57.8	88.3	M. Li et al. (2020)
hCG	57.6	-	Zhong et al. (2019)
HE4	71.0 – 72.9	86.0 – 91.5	Scaletta et al. (2017) Lakshmanan et al. (2019)

This in-depth investigation on the OC biomarkers emphasises the significance and reliability of HE4 biomarker in detecting the illness. Thus, this research focuses on the designing of DNA aptamer hairpins targeting HE4 with high binding affinity, improved interactions, and good stability.

2.4 OVARIAN CANCER DIAGNOSTICS

The diagnosis of OC poses a challenge with no reliable early detection tools available at present. The current gold standard of diagnosing OC involves surgical pathological investigation with possibility of delays in early detection and clinical management (Guo et al., 2019). This resulted in lowered survival rates among OC patients. A standard pelvic examination is frequently the initial procedure employed when there is suspicion of OC, including the increment of abdominal mass and discomfort or pain at pelvic area (U.S. Preventive Services Task Force, 2018). However, its ability to identify early-stage OC and differentiating benign from malignancy conditions is restricted due to the location of ovaries, deep inside the pelvis (Ebell et al., 2015). The existing clinical and under investigation diagnostics for OC are comprehensive, encompassing physical examinations, imaging analyses such as transvaginal ultrasound (TVUS) and positron emission tomography (PET) scan, tumour marker tests, and surgical treatments to provide a clear diagnosis (Kang et al., 2022; Khiewvan et al., 2017). These techniques were thoroughly investigated to improve the accuracy and rapidness of OC diagnosis.

2.4.1 Imaging Techniques

The transvaginal ultrasound (TVUS), positron emission tomography-computed tomography (PET/CT), and magnetic resonance imaging (MRI) scans are important tools in enhancing the diagnosis and prognosis of OCs. TVUS is frequently utilised as a primary imaging approach or as a secondary examination after CA125 marker test, in examining and characterising ovarian masses by offering detailed images of the ovaries (Campbell & Gentry-Maharaj, 2018). The United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) is an approach assessing approximately 50,000 women from the general population for the detection of ovarian or tubal cancers, involving yearly TVUS (Kalsi et al., 2021). 960 women were screened with imaging abnormalities, leading to surgeries of which 113 (11.8%) OC cases were recorded. Although this effort assisted the detection of OCs in general population, it has not been a practice in most countries due to lack of required facilities and professional staff.

The combination of PET/CT scan and 2-deoxy-2[fluorine-18]fluoro-D-glucose (^{18}F-FDG) helps differentiate between benign and malignant ovarian tumours by identifying elevated glucose metabolism linked to cancer cells (Kosinska et al., 2023; Mbakaza & Vangu, 2022). A PET/CT scan integrates functional data from a PET scan with anatomical information and precise cross-sectional images from a CT scan, to offer a thorough evaluation of tumour metabolism and location which is commonly use in suspected recurrent OCs examination (Reyners et al., 2016; Sultana et al., 2021). 45 out of 52 patients with elevated CA125 showed signs of recurring OC on PET/CT imaging, including three false negative cases and one false positive case. The sensitivity and specificity of ^{18}F-FDG PET/CT were determined as 94% and 75%, respectively (Cengiz et al., 2019). This technique provides whole-body imaging, allowing for the detection of distant metastases as well as abdominal and pelvic lesions, which can aid in patient care. In a similar study, the PET scan was combined with MRI in the presence of ^{18}F-FDG to evaluate the staging in OC patients where the lesion-based sensitivity and specificity were estimated at 78.6% and 95.7%, respectively, promoting 93.9% accuracy (Tsuyoshi et al., 2020). MRI scans provide images of the pelvis and abdomen, assisting the identification of ovarian masses, evaluating tumour size, and detecting metastases (Kasper et al., 2017).

Each imaging technique has a distinct role, advantages, and limitations. The selection of imaging modality is based on aspects such as clinical symptoms, the necessity for in-depth analysis of masses, staging criteria, and the demand for monitoring therapy efficacy or disease recurrence. Combining these imaging methods with clinical examination and laboratory testing, such as CA-125 and HE4 marker tests, offers a thorough approach to diagnosing, staging, and treating ovarian cancer.

2.4.2 Ovarian Cancer Marker Tests

OC marker tests are diagnostic tools used in screening biomarkers associated with OC, including HE4, CA125, and hCG. Even though these markers are not exclusive to OC only, they might suggest the existence of the illness when their levels are elevated.

Matsas et al. (2023) summarised the diagnostic importance of CA125, HE4, CA15-3, CA19-9, hCG, and AFP markers. HE4, the glycoprotein by WFDC2 gene, can be screened in the blood and urine of OC patients, whereas the common practice for CA125 test sampling is through blood samples.

Historically, CEA marker was labelled as OC biomarker in 1976 but was first discovered in 1965 as mucinous colon cancer biomarker (Ueland, 2017). Moving forward, the CA125 marker was explored in 1981 and announced as a cancer biomarker specifically for OCs soon after. However, an increased CA125 level above its cut-off limit (35 U/ml) can also arise from non-cancerous diseases and other conditions such as pregnancy and menstruation (Neogi & Srivastava, 2014).

In 2008, the Food and Drug Administration (FDA) in the United States declared HE4 as a marker to monitor patients already diagnosed with OC, detecting the recurrence of epithelial OC and overcoming the traditional role of CA125 (Capriglione et al., 2017; Chudecka-Głaz, 2015; Hellstrom et al., 2010). However, various recent reports have explored the combination of HE4 test with other marker tests in early detection of OC, resulting in high specificity and sensitivity. Research suggested that the performance of HE4 marker test was better in women aged less than 50, whereas, CA125 marker test worked more ideal in older women (Barr et al., 2022). Thus, the combination of these two markers has been tested numerous times and outperformed a single biomarker test in early stages of the disease while precisely distinguishing between benign and malignancy cases (Han et al., 2018; Y. Li et al., 2020). Moreover, combining CA125 with CA19-9, epidermal growth factor receptor (EGFR), and few other markers increased sensitivity to 98.2% and specificity to 98.7% in the OC early detection (Muinao et al., 2019). Nevertheless, the usual clinical practice in confirming the OC diagnosis after marker tests is by including imaging modalities and/or surgical pathologies (Grayson et al., 2019).

2.4.3 Surgeries

Surgical procedures are commonly more invasive than the primary and secondary diagnostic steps, which are the marker tests and imaging scans. These surgeries serve two main purposes; diagnosing the illness and identifying the illness stage, in which involving laparoscopy or laparotomy (Nezhat et al., 2013). The diagnostic laparoscopy allows medical personnel to observe the ovaries and nearby organs using a camera attached at the end of a thin tube called laparoscope. Small incisions are made when suspicious and abnormal regions are detected to remove tissue samples for biopsy. Histological analysis of tissue acquired by biopsy remains the most reliable method for definitively diagnosing OC. In contrast, laparotomy is a more invasive surgery procedure, involving an open and larger incision in the abdomen with risk of complications (Sincavage et al., 2021). This procedure allows for a more comprehensive examination of the abdominal cavity, possibly determining the staging of the illness and identifying the extent of the spread of the cancer to the neighbouring organs.

Traditional surgical staging for OC typically involves an exploratory laparotomy along with peritoneal washings, removal of the uterus (hysterectomy), removal of both the fallopian tubes and ovaries (salpingo-oophorectomy), removal of the omentum (omentectomy), taking several biopsies from different peritoneal locations, and potentially removing lymph nodes in the pelvic and para-aortic areas (Nezhat et al., 2013). However, at present, laparoscopy has been the common approach to surgical staging of early-stage OC, minimising the complications experienced by post-laparotomy patients, such as bleeding and post-surgery hernia (Falcetta et al., 2016). This is supported by a study involving 778 staging laparoscopy cases, which discovered that laparoscopy after initial diagnostic (marker tests and/or imaging scans) was effective in identifying the staging of OC (Zeff, 2018). A comparative study on OC surgical staging involving 60 patients concluded that the laparoscopy group experienced longer surgeries but benefited from shorter hospital stays and received earlier treatment compared to the laparotomy group who had a longer post-surgery recovery time (Baiomy et al., 2020).

In summary, the diagnosis of OC requires a wide range of clinical diagnostic tests, each of which plays a vital role in identifying, determining the stage, and treating the illness. Despite the existence of notable obstacles in the field of early OC detection, continuous research and technological progress offer promising screening techniques.

2.5 APTAMERS

2.5.1 Introduction and Aptamer Selection

Aptamer, also known as synthetic antibody is one of the main components in this study. It is a short, single-stranded oligonucleotide made up of nucleic acids, and it is typically 20 to 100 nucleotide long (Röthlisberger & Hollenstein, 2018). Nucleic acids are essential for all living forms, carrying genetic information and essentially important for protein production. These macromolecules are made of nucleotide chains, consisting of five-carbon sugars, phosphate groups, and nitrogenous bases (Alberts et al., 2002). There are two main forms of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA contains the genetic information necessary for the growth, functioning, and reproduction of living organisms and certain viruses, while RNA serves as the intermediary that conveys instructions from DNA to regulate protein production (Minchin & Lodge, 2019). RNA also performs catalytic and regulatory activities inside cells.

This study emphasises the utilisation of DNA aptamers, instead of RNA aptamers. The advantages of DNA aptamers over RNA aptamers include their thermal and chemical stability, alongside their compatibility with various chemical modifications (Ni et al., 2011). Furthermore, comparing to the conventional and widely utilised antibodies, aptamers offer various advantages, offering smaller molecule size, greater structure flexibility, higher biocompatibility, and lower immunogenicity (Byun, 2021). RNA and DNA aptamers were discovered in the early 1990s using a ground-breaking selection process known as Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Ellington & Szostak, 1990, 1992). SELEX enables the *in vitro* evolution of aptamers by successive rounds of selection, amplification, and enrichment (Figure 2.3), leading to the identification of aptamer sequences that bind strongly to the desired target (Blind & Blank, 2015). The process begins with a large library containing random aptamer sequences, which the pool was subjected to iterative rounds of selection, enriching the sequences that bind to the target during the incubation period. The sequences that did not bind to target are washed away, leaving only those sequences that bound to be eluted and collected. The collected sequences are then amplified and gone through additional rounds of SELEX process. Lastly, the sequencing of the final pool was done, identifying the aptamers bound to target with high affinity.

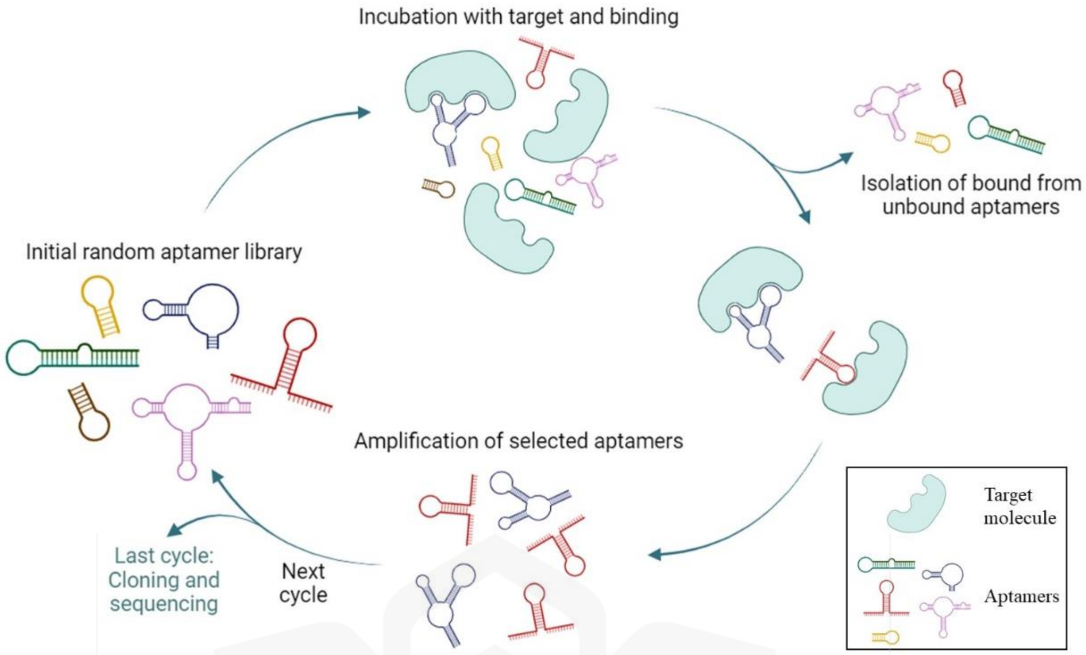


Figure 2.3 Schematic representation of SELEX in aptamer selection.

The past three decades have witnessed SELEX as the primary technique in aptamer selection (Sampson, 2003), but in recent years, researchers have begun to investigate the potential of integrating computational or *in silico* approaches with the SELEX selection method, which is typically followed by aptamer design enhancement. A few studies have named this computational selection method as ‘*in silico* SELEX', a revolutionary strategy to designing and evaluating aptamers (Poustforoosh et al., 2022; Rasouli Jazi et al., 2023). The *in silico* SELEX method differs from one study to another due to different computational methodologies applied, including various software, tools, and work designs (Buglak et al., 2020), which the work mostly includes molecular docking and molecular dynamics (MD) simulations (Figure 2.4). Molecular docking has become an important tool in computer-assisted drug design and analyte screening, estimating binding affinity and analysing the interaction mode (Fan et al., 2019). Whereas, MD simulations depict the behaviour of proteins and other biomolecules in atomic detail and at extremely fine temporal resolution (Hollingsworth & Dror, 2018). This significantly improves efficiency and lower research costs.

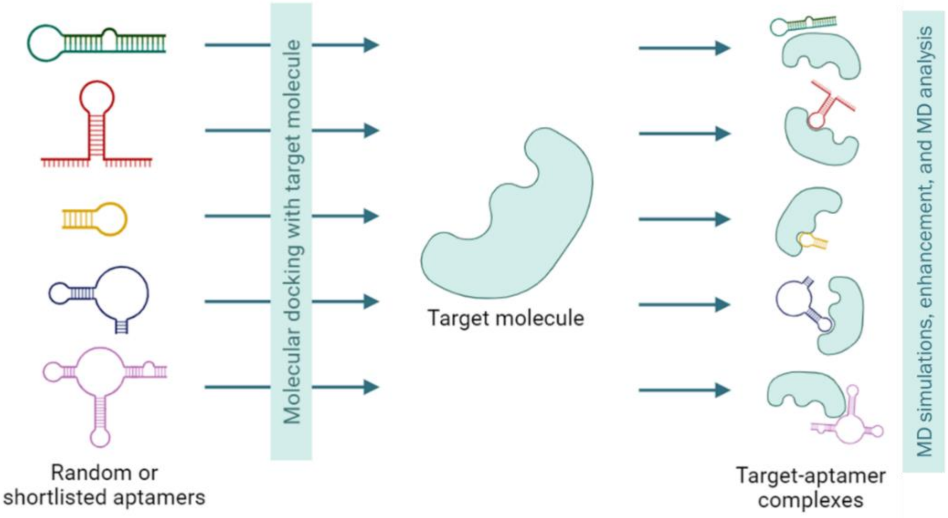


Figure 2.4 Schematic representation of a typical workflow of *in silico* SELEX.

In silico SELEX emerges as an effective approach for aptamer selection, providing numerous advantages. For instance, enhanced exploration of sequences, minimised the time in experimental rounds, and increased efficiency in high-affinity aptamers identification (Thevendran et al., 2023). By integrating computational simulations with experimental evaluations, *in silico* SELEX represents an effective method in the development of aptamers for various applications, including diagnostic purposes. This approach facilitates the swift evaluation of extensive libraries, removing the necessity for time-consuming experimental iterations while forecasting off-target interactions, thereby enhancing aptamer specificity at an early stage in the selection process (Zarandi et al., 2020).

Furthermore, *in silico* SELEX benefits from advancement in molecular docking and MD simulations, offering atomic-level insights into protein interactions and binding affinities (Almazar et al., 2023). The application of artificial intelligence and machine learning in this method significantly improves the optimisation of aptamer structures, thereby enhancing their binding affinity and stability in physiological conditions (Zhao et al., 2022). The combination of computational and experimental methods streamlines the selection process and speeds up the development of aptamers for therapeutic and diagnostic applications. In summary, in silico SELEX stands out as a highly effective approach for creating superior aptamers, providing advantages such as lower research expenses and expedited development schedules.

2.5.2 The Fundamental of DNA Aptamer Structure

DNA aptamers are composed of the same four nitrogenous bases found in a typical DNA double helix, which are adenine (A), thymine (T), guanine (G), and cytosine (C). The thymine found in DNA is replaced with uracil for RNA composition. The process of DNA aptamer synthesis involves the formation of a deoxyribose sugar-phosphate backbone and nucleotide base pairing, where adenine pairs with thymine *via* two hydrogen bonds (H-bonds), and guanine pairs with cytosine through three H-bonds (Figure 2.5a). This specific hydrogen bonding pattern among the bases generates the double helix structure based on Watson-Crick base pairing, maintaining the stability of the DNA (Takahashi & Sugimoto, 2021; Watson & Crick, 1953).

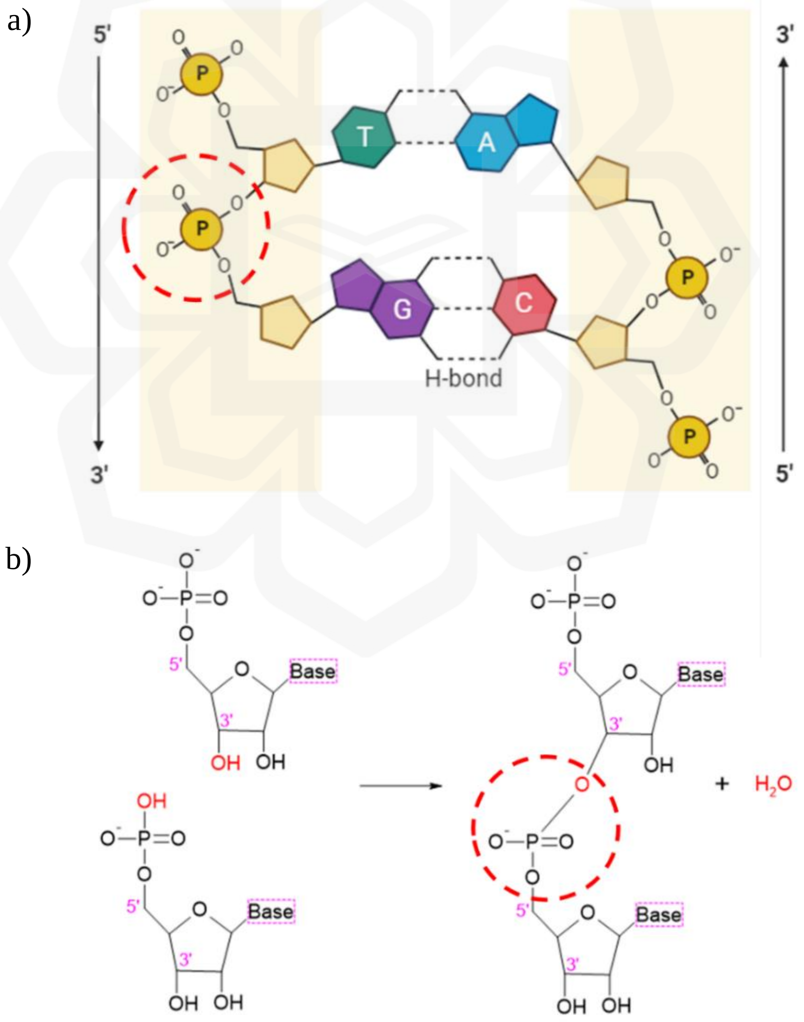


Figure 2.5 The formation of phosphodiester bond (in red circles).

The nucleic acid chains are formed by polymerisation, which occur through the establishment of phosphodiester bonds linking the 3’ hydroxyl group of one nucleotide's sugar to the 5’ hydroxyl group of the following nucleotide, resulting in chain expansion from the 3’ to the 5’ direction (Figure 2.5b) (Mikkola et al., 2018). This phosphodiester bond is essential in linking nucleotides in the DNA chain and due to the presence of phosphate groups throughout the molecules, DNA aptamers are negatively charged molecules. This negative charge is important in a variety of biological activities, such as DNA aptamer interactions with positively charged molecules. The nucleobases of aptamer serve as the foundational elements, and by arranging the bases in certain order, they fold into unique three-dimensional (3D) structures, allowing them to bind with high affinity and specificity to their target molecules (Odeh et al., 2019).

2.5.3 DNA Aptamers in Ovarian Cancer Diagnostics

DNA aptamers have been recognised as reliable diagnostic tools, offering innovative and diverse approaches to identify and measure a wide range of biomolecules and analytes (Hu et al., 2022). The capacity to modify these aptamers results in great binding specificity and affinity with the target molecules, making them viable candidates for recognition elements and revolutionising the field of diagnosis. In comparison to commonly utilised antibodies in diagnostic applications, aptamers are manufactured using cell-free chemical synthesis, resulting in lower large-scale manufacturing costs (Bauer et al., 2019). When these aptamers interact with their target, they use a lock-and-key model, which is made possible by the various conformations. Furthermore, because aptamers are smaller than antibodies, they can penetrate tumour cores more efficiently, making them useful for drug delivery (Sett, 2020). DNA aptamers have been used in a variety of diagnostic platforms during the last decade, including electrochemical biosensors, surface plasmon resonance (SPR)-based sensors, lateral flow assays, and polymerase chain reaction (PCR) assays (Kim et al., 2016; Mehmood et al., 2019).

The utilisation of DNA aptamers onto biosensor platforms is a critical step towards developing potential diagnostic tools. These biosensors use the specific binding interactions between aptamers and their target analytes to create measurable signals (Wan et al., 2021). Depending on the type of biosensor, the aptamer-target contact causes a variety of signal transduction pathways. For example, SPR sensors rely on changes in the refractive index of the sensor surface, electrochemical biosensors control electrical currents, and fluorescence-based sensors use conformational changes in aptamers to adjust fluorescence emissions. This signal transduction technique enables sensitive detection and quantification of diagnostic biomarkers, opening the path for early illness detection (Futane et al., 2023). Antibody-based platforms are commonly developed in the past decade for the screening of OC markers, including a work by Kovarova et al. (2023). The study presented an electrochemical sandwich-type magneto-immunosensor using three specific antibodies to screen CA125, HE4, and AFP marker levels that fit the criteria as an early-stage OC detection kit (Kovarova et al., 2023). However, the expensive nature of antibodies is making researchers to seek for more cost-effective and easier to synthesise alternatives like aptamers. Table 2.3 shows recent developed aptamer-based diagnostic approach for OC.

Table 2.3 Aptamer-based approaches in recent ovarian cancer diagnostics.

Aptamer-based diagnostic approach	Target	Limit of detection (LOD)	Reference
Lanthanide-doped upconverting nanoparticles (UCNPs) luminescence resonance energy transfer (LRET) nanoprobe containing HE4 aptamer	HE4	0.021 ng/ml	(Ma et al., 2021)
Rapid intraoperative imaging technique by staining primary serous OC (pSOC) cells with fluorescein amidite-labelled mApoc46 aptamer	pSOC cells	-	(J. Wang et al., 2021)
CA125 aptamer immobilisation on silicon-on-insulator (SOI) nanoribbon (NR) biosensor	CA125	0.015 fM	(Ivanov et al., 2021)
LRET-based aptasensor using aptamer-modified near-infrared (NIR)-excitable UCNPs	CA125	0.009 U/ml	(X. Zhang et al., 2021)
Colourimetric biosensor using citrate-gold nanoparticles with PDGF-specific aptamer	PDGF	0.01 µg/ml	(Hasan et al., 2022)
Electrochemical aptasensor with nanocomposite as signal amplifier and deposited gold nanocrystals (DpAu) as sensing platform	HE4	0.41 fM	(R. Hu et al., 2023)
Magnetic $Fe_3O_4/\alpha\text{-}Fe_2O_3$ hetero-nanorods electrochemical bioplatform with DNA aptamer immobilisation	HE4	27.5 fg/ml	(Y. Zhang et al., 2023)
Electrochemical aptasensor combining molybdenum disulphide (MoS_2), CA125 aptamer, and gold nanoflowers (AuNFs)	CA125	0.0001 U/ml	(C. Hu et al., 2023)
Streptavidin-coated gold chip immobilised with biotinylated CA125 aptamer as SPR biosensor	CA125	0.01 U/ml	(Shahbazlou et al., 2023)
Aptamer-modified AgNP electrochemical paper analytical device (ePAD)	PDGF	0.01 µg/ml	(Sharma et al., 2023)

A recent study developed an electrochemical aptasensor with detection limit of 0.41 fM (fmol/L) using abundant amino groups in metal-organic framework/ketjen black (NH_2-MIL-53(Al)/KB) composite, where the target HE4 was bound to the aptamer (R. Hu et al., 2023). It is an ultrasensitive sensor as the cut-off level for pre-menopause and menopause OC patients are reported as 70 pmol/L and 140 pmol/L, respectively (Dochez et al., 2019). Luminescence-based nanoprobe were also developed earlier in HE4 marker screening, showing excellent sensitivity and selectivity (Ma et al., 2021). They constructed LRET-based probes where the lanthanide-doped UCNP and HHQ-1 dye were tagged with molecular beacon consisting of HE4 aptamer sequence. The aptamer hairpin structure played a significant role in the strong interaction between HE4 and its aptamer, leading to the LRET inhibition.

Recent years have also seen the rapid development of aptasensors in screening other OC markers, including CA125 and platelet-derived growth factor (PDGF). A rapid, low-cost, and disposable electrochemical paper analytical device (ePAD) were created by depositing silver nanoparticles on the paper analytical device and immobilising PDGF aptamer (Sharma et al., 2023). Another sensor was developed to screen PDGF by applying a different approach, which was the colourimetric assay for naked-eye detection, where these two sensors had similar sensitivities with LOD of 0.01 µg/ml (Hasan et al., 2022). A very low LOD (0.0001 U/ml) was achieved in the screening of CA125 marker level by using a gold nanoparticle-based sensor with a stable layered substrate of molybdenum disulphide (MoS_2) in the presence of CA125 aptamer, making it a sensitive detection method where the upper limit of CA125 was reported at 35 U/ml (Hu et al., 2023).

Highly sensitive and rapid cancer biomarkers detection is essential for early diagnosis and prognosis of the illness. Therefore, the recent studies highlighted in literature search have shown advancements and are aiming to simplify, accelerate, and improving the current diagnostic approaches.

2.6 CHOICE OF METHODS

Advancements in computational technology have significantly transformed the medical industry, especially diagnostics, by improving the precision, speed, and effectiveness of illness detection. Advanced technologies like bioinformatics and artificial intelligence

(AI) are crucial for analysing intricate biological data and detecting patterns that are too complicated, beyond human capability to screen and recognise manually. This enhances the successful rate of a therapy and greatly decreases the time required for diagnosis, allowing healthcare practitioners to respond to patient needs more quickly.

In this thesis, computational methodologies are integrated into most of the work, to achieve its main objective of designing potential DNA aptamer hairpin in the detection of OC biomarker, HE4. Additionally, the findings obtained from the *in silico* work are subsequently validated by *in vitro* assay, ensuring a precise outcome of the whole study. The choice of methods for the computational methods of protein modelling, DNA aptamer hairpin design, and MD simulations of protein-DNA complex, including experimental method to analyse the protein-DNA binding are further discussed.

2.6.1 Molecular Modelling of HE4 Protein

At the beginning of this study, the searching of the HE4 protein tertiary structure yielded no results, indicating that it was not available at the time of the research. The searching included the well-established Protein Data Bank (PDB) database and other previous studies involving this OC biomarker having a sequence length of 124 amino acids. Therefore, the potential of modelling the protein structure through computational approaches is highly aimed for, after thorough literature search. Initially, the secondary structure of HE4 protein will be predicted by PSIPRED Protein Analysis Workbench, utilising neural networks in analysing the sequence which generates its secondary structural motifs, including the coils, alpha-helices, and beta-strands (Buchan & Jones, 2019).

Next, the tertiary protein structure prediction tools included Phyre2 (Kelley et al., 2015), SWISS-MODEL (Waterhouse et al., 2018), MODELLER (Webb & Sali, 2016), AlphaFold (Jumper et al., 2021), Iterative Threading Assembly Refinement (I-TASSER) (Yang & Zhang, 2015), Robetta (Kim et al., 2004), HHPred (Söding et al., 2005), and RaptorX (Källberg et al., 2014). However, while each protein prediction tool is powerful and unique, they come with limitations. For instance, Phyre2 relies heavily on known template structures available in databases, making it difficult to model the proteins with novel folds. This common limitation applies to SWISS-MODEL, MODELLER, and HHPred prediction tools as well. For MODELLER, initial template of known structure is required to undergo homology or comparative modelling, making it nearly impossible to predict structures of proteins with no similar template (Bitencourt-Ferreira & de Azevedo, 2019). Furthermore, the MODELLER prediction process is computationally intensive, which makes it less favourable.

In contrast, AlphaFold, I-TASSER, and Robetta have brought significant progress to protein structure prediction, surpassing conventional techniques such as SWISS-MODEL and MODELLER. AlphaFold, the free online prediction server developed by DeepMind, has established a higher level of precision in predicting protein structures compared to other approaches in the Critical Assessment of Structure Prediction (CASP) competitions (Elofsson, 2023). The deep learning method forecasts protein structures as accurate as the experimental findings. This is due to its ability to predict novel proteins without known homologues as it does not rely solely on template-based approach. On the other hand, the combination of threading, *ab initio*, and homology modelling makes I-TASSER a unique approach in predicting the structure of a wide range of proteins (Yang & Zhang, 2016). Robetta adopts both homology and *ab initio* modelling, identifying the structure even without close homologues (Park et al., 2018). One distinct advantage of AlphaFold, I-TASSER, and Robetta is their capability to determine the full tertiary structure of the protein, and not only fragments of the full provided sequence. Current study utilised these three protein prediction servers to model the tertiary structure of HE4 protein and validated the structure quality using Ramachandran plot and ERRAT. A Ramachandran plot serves as a graphical tool to assess the dihedral angles (phi and psi) of amino acids within protein structures (Park et al., 2023), whereas ERRAT assesses the overall integrity of protein models through the examination of non-bonded atomic interactions. This analysis evaluates the interactions in relation to high-resolution crystallographic data to identify regions where the structure might be unreliable (Mohammed et al., 2022).

2.6.2 Design of 3D DNA Aptamer Hairpins

This study combines SELEX and *in silico* SELEX in the aptamer selection that binds strongly to HE4 protein. The screening of the DNA aptamer candidates against HE4 was previously done by Eaton et al. (2015) through SELEX selection, where four promising aptamer sequences were shortlisted by the work. These sequences were selected due to their enrichment throughout the selection process and they represented large clusters (Eaton et al., 2015). Further literature search did not lead to any other study with different aptamer candidates against HE4, thus, the four candidates mentioned earlier were chosen for the subsequent computational methods. Based on previous studies, molecular docking between protein and aptamers predicts the affinity and poses of the protein-aptamer complexes, subsequently identifying the most promising aptamer against its protein target (Azri et al., 2021; Navien et al., 2021; Sabri et al., 2019). To ensure that single-stranded DNA forms proper functional 3D DNA aptamer structures, it is important to initially construct the secondary DNA with correct base pairings before predicting its tertiary conformation (Zhang et al., 2019). This is done by utilising Mfold and RNAComposer servers. Sabri et al. (2020) then converted the predicted tertiary RNA structure to DNA composition by replacing uracil to thymine and changing the ribose sugar in the backbone to deoxyribose. This work adopts similar workflow as previous studies in predicting the secondary and tertiary structures, subsequently utilising AutoDock Vina in the molecular docking simulation between the four aptamer candidates with HE4 protein. The most potential aptamer candidate against HE4 is selected based on the greatest binding affinity.

Next, the truncation of the hairpin region of the most potential DNA aptamer candidate takes place, to initiate the designing process of multiple DNA aptamer hairpins with different sequences. The main purpose of this truncation to a shorter sequence length is to expedite the molecular docking of HE4 with hundreds of DNA sequences. The DNA aptamer hairpins undergo substitutional mutation of the nucleotides located at the loop region. This region plays a significant role in exposing the side chains of the loop nucleotides, enhancing the binding with its target (Ma et al., 2021). The substitutional mutation involves the replacement of a nucleobase with another nucleobase, such as thymine replaces by guanine and adenine replaces by cytosine, resulting in hundreds of different hairpin sequences. These hairpins are docked with HE4 and the HE4-hairpin complexes with lowest binding energies are further analysed, emphasising their binding poses. The selected, high affinity complexes are then subjected to MD simulations to characterise their molecular dynamics.

2.6.3 MD Simulations using GROMACS and Force Field of Choice

The *in silico* approach in this study continues with MD simulations of the HE4-hairpin complexes. MD simulation is a computer method that models the physical motion of atoms and molecules over time. The simulation comprises numerically solving Newton's equations of motion to study the atomic-level behaviour of nucleic acids (X. Wang et al., 2014). The key purposes of MD simulation include understanding molecular interactions, protein folding processes, and evaluating the dynamic behaviour of molecular systems under different conditions including temperature and pressure. GROMACS (Groningen Machine for Chemical Simulations) is one of the most widely utilised MD simulation tools. It supports various force fields, including OPLS (Optimised Potentials for Liquid Simulations), AMBER (Assisted Model Building with Energy Refinement), and CHARMM (Chemistry at Harvard Molecular Mechanics) (Chen et al., 2015). The software package covers a wide range of settings for customisable simulations and have been successfully study the molecular dynamics in different biomolecular systems, such as protein-ligand, membrane-protein, and protein-nucleic acids (Lehrer & Rheinstein, 2023; Smith et al., 2022; van Heesch et al., 2023).

The precision of MD simulations is largely dependent on the force fields employed, which depict the interactions among atoms in the system (Tucker et al., 2022). Historically, MD simulations of proteins have been more precise, but recent progress has improved force fields for MD simulation of nucleic acids. In this study, OPLS force field was applied to the MD simulations of HE4 protein, while the simulations for HE4-DNA aptamer hairpin complexes utilised CHARMM force field. The OPLS force field was developed by William L. Jorgensen and it is suitable to predict the dynamic properties of organic and biological molecules in liquid and gaseous states (Jorgensen et al., 1996). OPLS force field is recognised for its precise representation of torsional energetics, essential for simulating protein folding and dynamics. Research confirms the capability of this force field in accurately capture the protein and also dipeptide dynamics (Lee et al., 2019).

The CHARMM force field has been used previously in the studies of biomolecular systems such as proteins, nucleic acids, and carbohydrates. This force field provided great flexibility to the DNA structures in protein-nucleic acids simulation (Gallardo et al., 2022). Furthermore, it offers detailed parameterisation of nucleic acid, as well as protein, allowing accurate prediction of the structural and energetic aspects. Due to these criteria, CHARMM force field was selected to be applied in the MD simulations of HE4-DNA aptamer hairpin complexes.

2.6.4 Nucleic Acids-Protein Interactions and GNP-DNA Conjugation

Protein-nucleic acid interactions are important in biological systems and essential for replication, transcription, translation, repair, and gene expression control (Valuchova et al., 2016). These interactions occur when proteins attach to DNA or RNA molecules, allowing the proteins to carry out their regulatory roles, contribute to the structural arrangement of genetic material. They are very specific and determined by the sequence and molecular structure of the nucleic acid, as well as the structure and chemistry of the protein. The interactions between those two molecules may arise from the formation of hydrogen bonds, hydrophobic interactions, salt bridges, and electrostatic interactions (Ferreira de Freitas & Schapira, 2017; Kurczab et al., 2018).

The analysis and confirmation of protein-nucleic acid binding can be done through electrophoretic mobility shift assay (EMSA), chromatin immunoprecipitation (ChIP), and footprinting assays. EMSA, also known as gel shift assay uses non-denaturing polyacrylamide or agarose gel in electrophoresis (Cozzolino et al., 2021). The nucleic acid changes in mobility as it binds to the protein. The nucleic acid bands are usually visualised under ultraviolet light with the presence of gel stain that binds to the DNA molecules. ChIP is utilised to study the interaction between proteins and DNA within chromatin. This method includes attaching proteins to DNA, fragmenting the DNA, and subsequently isolating the DNA-protein complex by immunoprecipitation using an antibody that targets the particular protein (Ferraz et al., 2021). EMSA is selected to analyse the binding between HE4 protein and DNA aptamer hairpin in this study due to its rapid and simple technique. Additionally, the process doesn’t require the usage of expensive antibodies, as is necessary in ChIP assay.

In recent years, there has been a significant trend towards integrating DNA with gold nanoparticles (GNPs) in the development of diagnostic methods. GNPs have been a popular choice for their diagnostic potential, attributed to their ease of functionalisation, biocompatibility, and unique optical properties (Bai et al., 2020). Study have been conducted to screen superoxidase dismutase 4 (SOD 4) markers by conjugating GNPs with DNA aptamer *via* two approaches, carbodiimide chemistry method and a complexation approach (Dekhili et al., 2020). This thesis adopts the carbodiimide chemistry protocol, in which the conjugation process between GNP and DNA aptamer involves the activation of the carboxyl group on the GNP using 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), followed by the formation of amide bonds with N-Hydroxysuccinimide (NHS). The resulting GNP-DNA conjugate is analysed for formulation characteristics such as particle size and zeta potential.

CHAPTER THREE

THE MODELLING OF HE4 PROTEIN TERTIARY STRUCTURE USING PROTEIN STRUCTURE PREDICTION SERVERS AND ITS VALIDATIONS

3.1 INTRODUCTION

Protein modelling by prediction servers is a pivotal tool in the advancement of diagnostic research, fundamentally improving our understanding of illnesses. Protein modelling primarily entails the prediction of proteins' three-dimensional structures, which is essential in understanding their functions and interactions in biological systems (Al-Lazikani et al., 2001; Petsko & Ringe, 2004). Protein modelling plays a crucial role in pioneering diagnostic methods as precise models facilitate the detection of putative diagnostic biomarkers, revealing the intricate structural details of the proteins linked to specific illnesses. As computational protein modelling gains popularity, various protein prediction servers have been widely utilised such as AlphaFold, Rosetta, RoseTTAFold, OpenFold, I-TASSER, UniFold, and ESMFold (Avery et al., 2022; Elofsson, 2023; J. Yang et al., 2020). A recent study used the IntFOLD7 and MultiFOLD servers to predict the tertiary structure of protein T1170, which is a homomer with six identical monomers. Subsequently, the best IntFOLD7 model was superimposed on the native structure, PDB ID 7pbr, showing high similarity with a score of 82.04 (McGuffin et al., 2023).

This chapter focuses on the HE4 protein tertiary structure prediction using three different prediction servers: AlphaFold, I-TASSER, and Robetta, which were followed by molecular dynamics (MD) simulations and structure validations of the predicted models. One significant advantage of the 3 selected prediction servers is their ability to predict complete protein tertiary structures using the primary sequences provided. This contrast with servers such as Phyre2 which may not always model the full conformation and instead only generate specific fragments of the protein (Kelley et al., 2015). The AlphaFold server, recognised for its ground-breaking advancements in predicting protein structures, provides unmatched precision, enabling researchers to unravel complex protein architectures implicated in various diseases (Jumper et al., 2021; Read et al., 2023). This significantly reduces the gap between computational predictions and existing experimental protein structures with rapid prediction, which leads to quicker insights into the protein structure. AlphaFold utilises a deep learning neural network to predict protein structures, and the process involves the input of amino acid sequence, feature extraction, distance prediction, and finally, 3D structure prediction, which includes model refinement to improve accuracy. The method heavily depends on multi-sequence analysis, utilising data on conserved peptide structures (Perrakis & Sixma, 2021). In the challenging 14th Critical Assessment of Protein Structure Prediction (CASP14), the AlphaFold predicted model of CASP14 target T1049 closely resembled the experimental structure (PDB ID 6Y4F) with a high TM-score of 0.93, indicating high similarity (Figure 3.1) (Jumper et al., 2021).

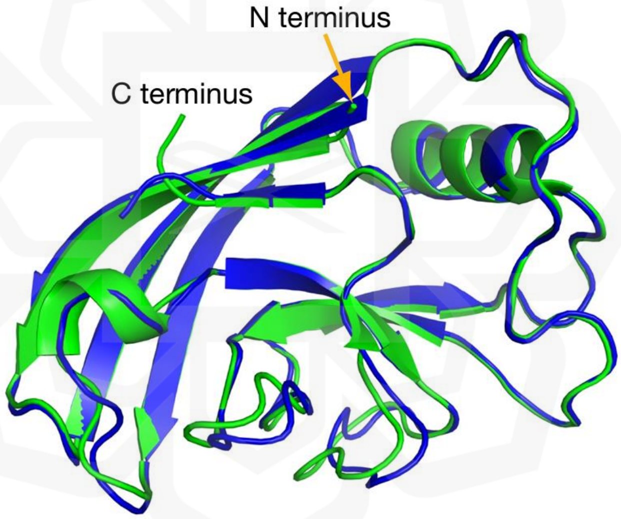


Figure 3.1 AlphaFold prediction of target T1049 (in blue) compared to the experimental structure of the target (in green), retrieved from the Protein Data Bank (Jumper et al., 2021).

Iterative Threading Assembly Refinement (I-TASSER), known for its expertise in comparative modelling, assists in elucidating structures when direct templates are not accessible. It employs a multiple-step approach where the threading and iterative assembly refinement steps are combined to predict protein structures accurately (Yang & Zhang, 2016; Zhang, 2008). The prediction process begins with identifying similar structural templates of known protein structures using threading algorithms and subsequently generates models with full length-sequence by integrating the identified templates (J. Yang & Zhang, 2023). The system finalises the modelling by performing iterative simulations for fragment assembly and refinement. Transferrin receptor protein 1 (TfR1), the main receptor that controls the process of absorbing iron from transferrin, was modelled using I-TASSER and showed an average quality model after validation through the Ramachandran plot, where 6.9% of the residues were located in the disallowed regions (outliers) (Al-Refaei et al., 2020). I-TASSER's strength rests in its ability to deal with diverse protein structures, especially in situations where similar homologues exist, allowing reliable tertiary protein model prediction.

Meanwhile, Robetta is a prominent protein structure prediction server that relies on the algorithms of the Rosetta software package. It emphasises *ab initio* modelling, which involves predicting protein structures from scratch based on the fundamental laws of chemistry and physics (Hardin et al., 2002; Lee et al., 2017). This method utilises a wide range of realistic potentials and energy functions to investigate the different shapes that proteins may take, predicting models from the beginning without heavily depending on pre-existing templates. Robetta utilises a fragment-based assembly methodology, which involves dividing protein sequences into smaller pieces before generating the complete models (Kim et al., 2004). While it primarily focuses on *ab initio* modelling, it also incorporates homology-based modelling, where it uses the information from available templates to generate the model before refinement procedures take place. These improve the precision of the predictions.

Table 3.1 The similarities and differences between AlphaFold, I-TASSER, and Robetta protein prediction servers.

		AlphaFold	I-TASSER	Robetta
Similarities	Purpose	Computationally predicting protein structures, facilitating the understanding of protein interactions		
	Utility	Utilised extensively in bioinformatics for numerous applications, including drug discovery and diagnostic development		
Differences	Developer	DeepMind Technologies (Alphabet Inc.)	Zhang Lab (University of Michigan)	Baker Lab (University of Washington)
	Methodology	Uses deep learning and neural networks to predict high accuracy structure	Adapts threading and iterative assembly simulations, applying comparative modelling technique	Combines *ab initio* and comparative modelling, not relying solely on existing templates
	Accuracy and performance	High accuracy, high precision	Moderate to high accuracy, depending on the availability of close homologues for comparative modelling	Moderate to high performance but might be less accurate for larger or more complex protein structures
	Number of AA allowed	1000	1500	1201
References		(Jumper et al., 2021)	(J. Yang et al., 2014; Zhang, 2008)	(Baek et al., 2021; Baker Lab, 2023; Kim et al., 2004)

Structure refinement by MD simulations (Fan & Mark, 2004; Feig, 2016; Kurniasih et al., 2021; R. Yang et al., 2022) and structural validations are necessary following structure prediction. This approach is important to optimise the modelled HE4 structures, as the simulation minimises energy, hence stabilising the overall conformation of the HE4 protein. This simulation was applied to the best predicted model of each server for 100 ns with a selection of Optimised Potentials for Liquid Simulations (OPLS) force field, as this force field has been widely used for numerous biological systems, including proteins and nucleic acids (Robertson et al., 2015). MD simulation is the most in-demand computational method for analysing the equilibrium structures and dynamic interactions of biological systems, which in this study focuses on the HE4 protein system and HE4-DNA complex systems. Finally, the PROCHECK and ERRAT tools were used to validate the structure of the predicted and refined models. These tools produced Ramachandran plots and overall quality factor scores, respectively (Mohammed et al., 2022; Nene et al., 2022). The Ramachandran plot shows the dihedral angles of amino acid residues in the protein backbone, which is important for checking the accuracy of protein structures. ERRAT, on the other hand, looks at non-bonded atomic interactions to check the overall quality of protein structures.

3.2 METHODOLOGY

3.2.1 HE4 Protein Modelling by Alphafold, I-TASSER and Robetta Protein Prediction Servers

Prior to HE4 tertiary structure modelling, the primary sequence of the protein is required, which was retrieved from UniProt (https://www.uniprot.org/) by searching for the keyword ‘HE4’. The full amino acid sequence for the human type of HE4 protein was selected and retrieved. Based on the protein sequence, the ProtParam tool (https://web.expasy.org/protparam/) was used to determine many parameters, including the molecular weight, amino acid compositions, and atomic components. The HE4 protein sequence, which consists of 124 amino acids, was applied to three different protein structure prediction web servers: AlphaFold (https://alphafold.ebi.ac.uk/), I-TASSER (https://zhanggroup.org/I-TASSER/), and Robetta (https://robetta.bakerlab.org/). The FASTA format of the protein sequence was downloaded from UniProt and used as the sequence input for each programme. The workflow was adopted from Sawal et al. (2023). The settings and parameters remained as default, followed by clicking the ‘submit’ button. Each predicted structure was visualised using the PyMOL molecular viewer, and the best model predicted by each server was downloaded as pdb files for further study.

3.2.2 HE4 Structure Refinement by Molecular Dynamics (MD) Simulations

Prior to the optimisation of the HE4 tertiary structures through MD simulations, the best model predicted by AlphaFold, I-TASSER, and Robetta was downloaded in pdb files. The MD simulation of each model was conducted using the GROMACS 5.1 software package (Royal Institute of Technology and Uppsala University, Stockholm, Sweden) separately. The proteins were simulated within a virtual cubic box with a 1.0 nm distance between the protein and the box faces at constant pressure and temperature of 1 atm and 310 K, respectively. The water molecules were added in the solvation step using simple point charge (SPC216), followed by the neutralisation of the protein by using sodium ions, Na^+ and chloride ions, Cl^-. The simulations were conducted using the OPLS forcefield for 100 ns. All MD simulations were analysed for their root-mean-square deviation (RMSD) and radius of gyration (Rg). The most stable HE4 conformation, which was the middle structure of the top cluster after the clustering step of each simulation was extracted as pdb files: RF1, RF2, and RF3, from the MD simulations of AlphaFold, I-TASSER, and Robetta, respectively. The 3D structures of RF1, RF2, and RF3 were viewed on PyMOL molecular viewer, and their secondary structure elements were analysed by PDBsum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/). This methodology was altered based on the design pipeline by Sabri et al. (2021).

3.2.3 HE4 Protein Structure Validations

Two protein structure validation programmes, PROCHECK and ERRAT, available on the SAVES 6.0 website (https://saves.mbi.ucla.edu/) (Sawal et al., 2023) were used in evaluating the quality of the best predicted models by AlphaFold, I-TASSER, Robetta, and the middle structures extracted from the top cluster after the MD simulations of AlphaFold, I-TASSER, and Robetta systems: RF1, RF2, and RF3. The pdb files were uploaded to the validation programme server and evaluated. The information on the backbone conformation and residue interactions was presented by Ramachandran plots, ERRAT plots, and overall quality factor values. Additionally, the pathway or tunnel and accessibility of ligands and other molecules to the potential binding sites within the best HE4 model were identified by CAVER 3.0 tool.

3.3 RESULTS AND DISCUSSION

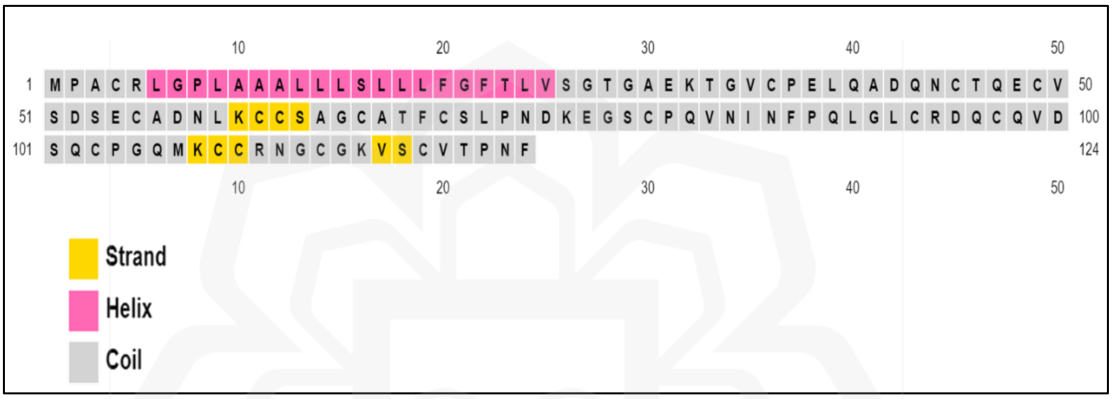
3.3.1 HE4 Predicted Models by Alphafold, I-TASSER, and Robetta

The search conducted on the UniProt website revealed the protein sequence associated with UniProt ID Q14508. This protein sequence has a length of 124 amino acids and a mass of 12993 Da. The chemical formula is $C_{541}H_{874}N_{154}O_{178}S_{19}$, which involves 1766 atoms and has a primary sequence of:

MPACRLGPLAAALLLSLLLFGFTLVSGTGAEKTGVCPELQADQNCTQECVSD
SECADNLKCCSAGCATFCSLPNDKEGSCPQVNINFPQLGLCRDQCQVDSQCP
GQMKCCRNGCGKVSCVTPNF

According to the ProtParam analysis, the predominant amino acid in the sequence is cysteine, accounting for 13.7% of the total. It is closely followed by leucine, which makes up 11.3% of the sequence. The proportion of the whole sequence that consists of amino acids with electrically charged side chains, namely arginine, lysine, aspartic acid, and glutamic acid, is 15.3%. Subsequently, the prediction of HE4 secondary structure was done by PSIPRED server based on the amino acid sequence above. PSIPRED categorises each residue in the HE4 protein into helix, strand, or coil, where this approach yields precise predictions through the examination of sequence homology and the integration of information from established protein structures (Buchan & Jones, 2019).

The output of PSIPRED features graphical representations, enabling the visualisation of the locations of helices, beta-strands, and random coils within the protein, as presented in Figure 3.2. The regions highlighted in pink indicate the presence of alpha-helices, particularly within the N-terminal segment, covering residues 5 to 25, which implies that this section establishes a stable helical configuration. The yellow-coloured segments indicate beta-strands, especially near the cysteine residues (-KCCSand -KCC-), suggesting their role in beta-sheet formation, which may play a part in the protein's structural framework. The grey-coloured regions represent random coils, highlighting areas of flexibility. The primary sequence from UniProt was then used to make accurate predictions of the HE4 protein's tertiary structure by the AlphaFold, I-TASSER, and Robetta servers. Each protein structure prediction server generated five models (Table 3.2).



1 M P A C R L G P L A A A L L L S L L L F G F T L V S G T G A E K T G V C P E L Q A D Q N C T Q E C V 50

51 S D S E C A D N L K C C S A G C A T F C S L P N D K E G S C P Q V N I N F P Q L G L C R D Q C Q V D 100

101 S Q C P G Q M K C C R N G C G K V S C V T P N F 124

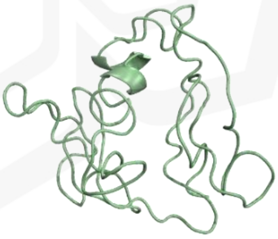
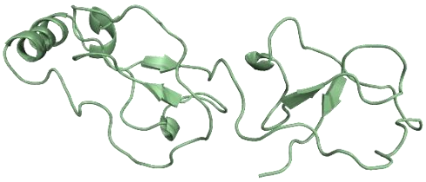
🟨 Strand

🟪 Helix

⬜ Coil

Figure 3.2 The secondary elements (strand, helix, and coil) represented by the PSIPRED server.

Table 3.2 The predicted models retrieved from AlphaFold, I-TASSER, and Robetta, and their scores based on their scoring systems

Server	Scoring system	Predicted model	Scores
AlphaFold	a) pLDDT	Model 1	pLDDT: 80.7 pTM-score: 0.614
	b) pTM-score	Model 2	pLDDT: 80.7 pTM-score: 0.611
		Model 3	pLDDT: 82.1 pTM-score: 0.635
		Model 4	pLDDT: 80.3 pTM-score: 0.567
		Model 5	pLDDT: 73.0 pTM-score: 0.450
		 Best model: Model 3	
I-TASSER	a) C-score	Model 1	C-score: -3.31 pTM-score: 0.35±0.12
	b) pTM-score (provided only for model 1)	Model 2	C-score: -3.62
		Model 3	C-score: -3.87
		Model 4	C-score: -4.07
		Model 5	C-score: -3.62
		 Best model: Model 1	
Robetta	C-score	Model 1	0.72
		Model 2-5	No C-score was provided. Only the 3D models were given.
		 Best model: Model 1	

As introduced earlier in this chapter, the I-TASSER server constructed the model using multiple threading alignments derived from the PDB library. In this study, the I-TASSER models were built using the 1zlgA, 1udkA, 7mn5B, 1udk, and 7fdeP templates available. Subsequently, the whole structure was constructed by integrating fragments derived from the existing templates, and the overall accuracy of the topology is assessed based on the predicted TM score (pTM-score) and confidence score (C-score). Models with a pTM-score greater than 0.5 are regarded as having a closely comparable fold to related proteins (Yin et al., 2022). The highest pTM-score achieved by the top-performing model from I-TASSER was 0.35 ± 0.12, accompanied by a C-score of -3.31. The C-score assesses the quality of the structure by analysing the alignments of the threading templates. It is measured on a scale from -5 to 2 (Zhang, 2008). A higher C-score indicates a model with greater confidence. The per-residue local distance difference test (pLDDT) scores range from 0 to 100, with a score of 100 indicating the highest level of confidence in the projected model's resemblance to the genuine protein structure (Guo et al., 2022). AlphaFold, which employs a template-free prediction method, uses the pLDDT score to assess the confidence and reliability of predicted structures on top of the pTM-score.

Out of the five models generated by the AlphaFold server, the model that performed the best was model 3, which achieved pLDDT and pTM-scores of 82.1 and 0.635, respectively. Robetta utilises a combination of template-based and template-free approaches for protein folding prediction. The C-score for the Robetta-generated structure that was the most optimal was 0.72, indicating that it was a more accurate model than the one that I-TASSER had created. The most accurate model for each server was examined using the PyMOL molecular visualisation software, allowing for detailed observations of the folded HE4 protein.

3.3.2 Molecular Dynamics Simulations Analysis

MD simulations were used to optimise the projected HE4 tertiary structures by analysing the motions of molecules and atoms over a specific time frame. The aim is to refine and adapt the HE4 models in order to get them closer to their native state. The root mean square deviation (RMSD) is used to assess the conformational changes in protein backbones throughout the MD simulation, comparing them to their initial structural conformation (Aier et al., 2016). The RMSD graph of the AlphaFold protein exhibited initial fluctuations during the course and eventually stabilised at 12 Å after 8 ns (Figure 3.3). The highly fluctuated RMSD values at the beginning of MD simulation indicate the initial adjustment or relaxation of the protein structures from its initial conformations, which were the predicted models by AlphaFold, I-TASSER, and Robetta. This phenomenon occurs especially when the initial structure is significantly far from equilibrium, where the conformational changes will reach a more stabilised and energetically favoured state (Harris et al., 2014, Sarker et al., 2023). Evidently, the RMSD for both I-TASSER and Robetta stabilised at approximately 15 ns, exhibiting RMSD values between 5 and 6 Å respectively, lower than the AlphaFold. Low RMSD values, ideally around 2 Å throughout the simulation, imply a better level of conformational stability (Bahaman et al., 2021), however, certain molecules are anticipated to exhibit greater and more substantial fluctuations in RMSD values, indicating a higher degree of trajectory deviation where the conformations deviate considerably from the original conformation.

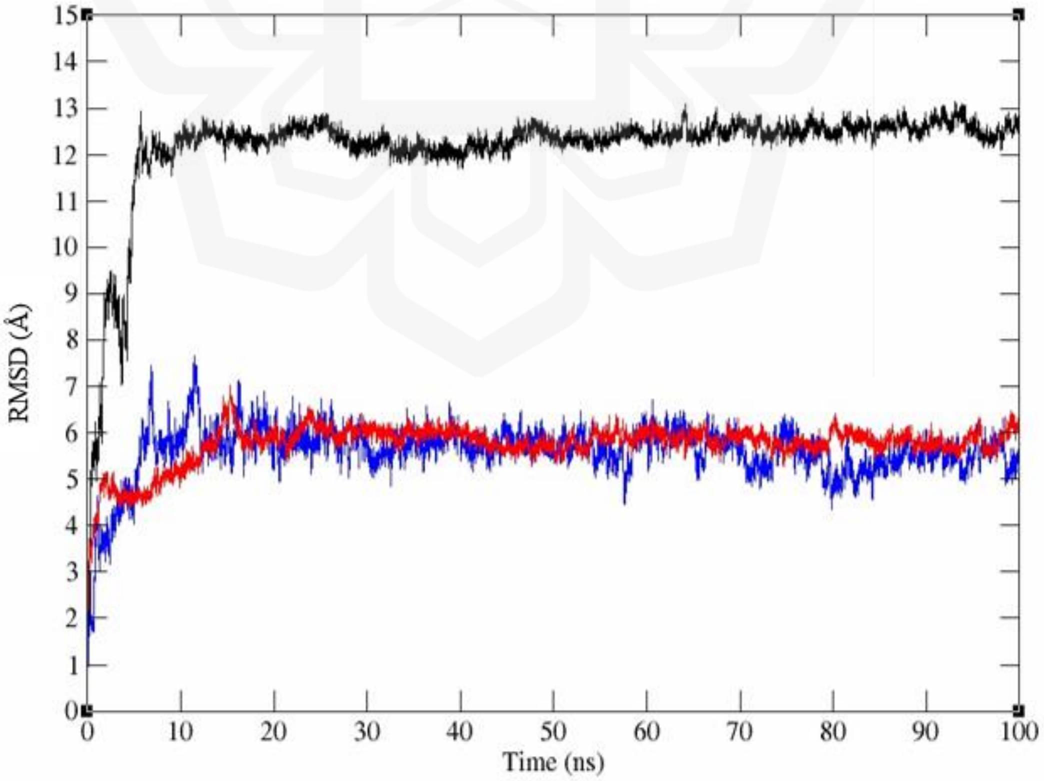


Figure 3.3 The RMSD of AlphaFold (black), I-TASSER (red), and Robetta (blue) 3D models during 100 ns of MD simulations.

Equilibration is a crucial phase in MD simulations, where the protein attains a stable and consistent state. In this phase, the system adapts to the simulation circumstances, reaches its lowest energy state, and attains an equilibrium distribution of velocities and positions for its constituent components, such as atoms (Frenkel & Smit, 1996). Most molecules undergo significant fluctuations at the beginning of the MD simulations before the protein adopts a more stable conformation (Harris et al., 2014).

The RMSD exhibited fluctuations at the start of the simulations for all three systems. However, the AlphaFold system had a greater degree of deviation in its trajectory, with the RMSD value fluctuating significantly between 5 ns and 8 ns. This is influenced by significant structural alterations between its initial conformation and its stable conformation once equilibrium has been achieved (Sarker et al., 2023). The flattening of the RMSD graph, which suggests the attainment of stable conformations, indicates that the equilibration phase was complete at 8 ns. After each MD simulation, a clustering step was performed to identify the representative structure and simplify the analysis of the simulation results (Phillips et al., 2011). The clustering stage identifies the predominant conformations in the simulated ensemble while emphasising the crucial structural patterns, as it groups similar conformations together.

Table 3.3 Clustering analysis after MD simulations of the AlphaFold, I-TASSER, and Robetta predicted HE4 protein models.

Clustering analysis	Protein prediction server		
	AlphaFold	I-TASSER	Robetta
Total number of structures	9001	9001	9001
Number of clusters	353	150	542
Number of structures in top cluster	429	1062	150
Middle structure of top cluster	RF1 At 83770 (83.77 ns)	RF2 At 84040 (84.04 ns)	RF3 At 48930 (48.93 ns)

According to the clustering analysis (Table 3.3), which utilised the Gromos clustering approach with an RMSD cut-off of 0.15 nm, it was determined that there were 429 structures that exhibited a significant degree of similarity in their conformations during the MD simulation of the AlphaFold projected model. Out of the 9001 I-TASSER conformations, around 11.80% (1062) exhibit a significant level of similarity. The refined HE4 model for the AlphaFold, I-TASSER, and Robetta systems was obtained from the middle structure of the top cluster labelled as RF1, RF2, and RF3, respectively. It is important to mention that the refined model reflects the most dominant conformation. In addition to the RMSD curve explained earlier, the Rg is a useful parameter for assessing the stability and compactness of a protein structure. The protein structure exhibits a high level of compactness as it maintains stable values throughout the simulation (Kumar D et al., 2017).

Figure 3.4 illustrates the fluctuation of the Rg graph of the AlphaFold system, which eventually stabilised after 8 ns at approximately 19 Å. Similar to the RMSD, this is a result of significant structural differences between its initial and stable conformation, RF1. The initial conformation was cylindrical in shape before it was folded into a compact form. No significant changes in the Rg were seen in the I-TASSER and Robetta systems. Their Rg values were roughly 19 Å and 14.5 Å, respectively. Based on the Rg graph, it can be inferred that all three systems exhibited relatively steady values after 8 ns, indicating that they maintained a compact and stable folded shape for the majority of the simulation duration. For Robetta, the Rg remained constant since 0 ns, denoting that the structure did not significantly fold or unfold which implies that the conformation has achieved its most compact and stable conformation since the beginning of the MD simulation.

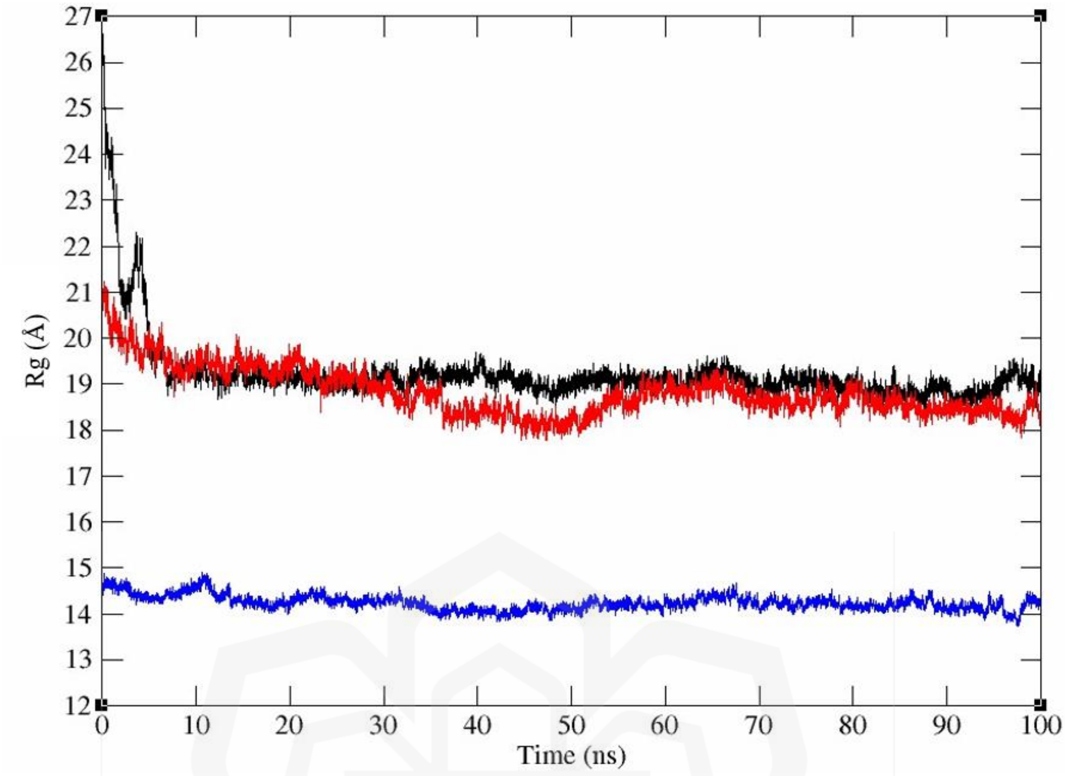


Figure 3.4 The radius of gyration of AlphaFold (black), I-TASSER (red), and Robetta (blue) 3D models during 100 ns of MD simulations.

The refined models (RF1, RF2, and RF3) were analysed using PyMOL molecular viewer, and the structural changes were compared with their initial conformation, which was the best model predicted by AlphaFold, I-TASSER, and Robetta, respectively (Figure 3.5). The RF2 and RF3 models exhibit minor changes in their structural conformations compared to the I-TASSER and Robetta projected models, respectively. On the other hand, the RF1 model displays significant deviations from its original conformation, which is based on the AlphaFold predicted model. The observed differences between the AlphaFold and RF1 models corresponded to the fluctuations in RMSD that occurred during the simulations. Specifically, the AlphaFold system exhibited a significant level of deviation. The modifications in the configurations of RF2 and RF3 from their original conformations are very discreet, corresponding to the slight variations observed in the RMSD curve.

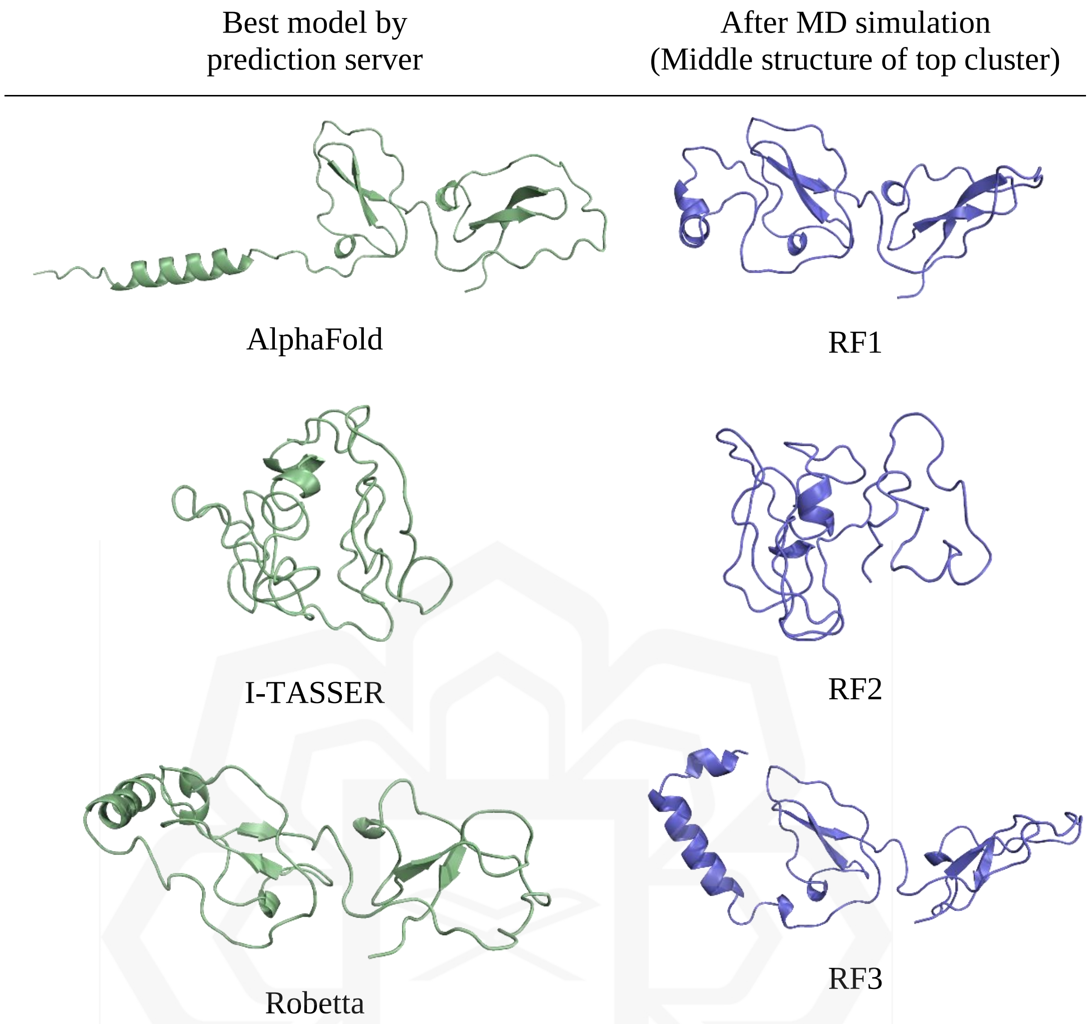


Figure 3.5 The HE4 protein models. The best model predicted by AlphaFold, I-TASSER, and Robetta (in green) and the dominant conformations, taken from the middle structure of the top cluster after the MD simulations of the predicted model: RF1, RF2 and RF3 (in indigo).

In terms of structural conformations, the RF2 and RF3 models differ slightly from the I-TASSER and Robetta models, respectively, however the RF1 model differs significantly from its initial conformation, which is the AlphaFold model. These significant differences between the AlphaFold and RF1 models (Figure 3.6) corresponded to the RMSD fluctuations seen at the beginning of the simulation (Figure 3.3), where the AlphaFold system experienced a substantial degree of deviation from 0 ns to 8 ns. The changes in the structures of RF2 and RF3 from their original conformations are less visible, but the rearrangement of the structures still exists, corresponding to slight fluctuations in the RMSD curve.

Initial conformation (at 0 ns)
(AlphaFold best model)

Conformation at 2 ns

Conformation at 4 ns

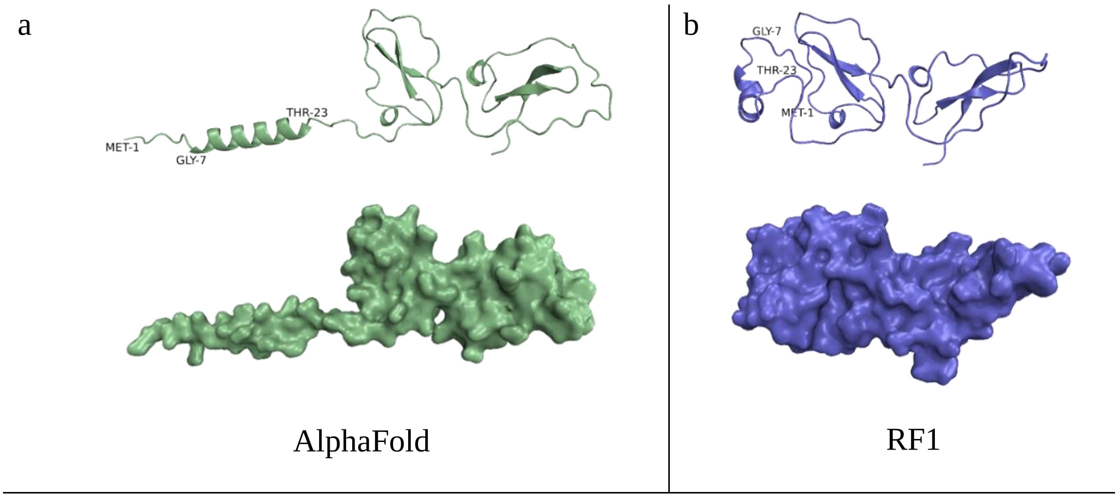
Conformation at 6 ns

Conformation at 8 ns

Middle structure of top cluster
(RF1)

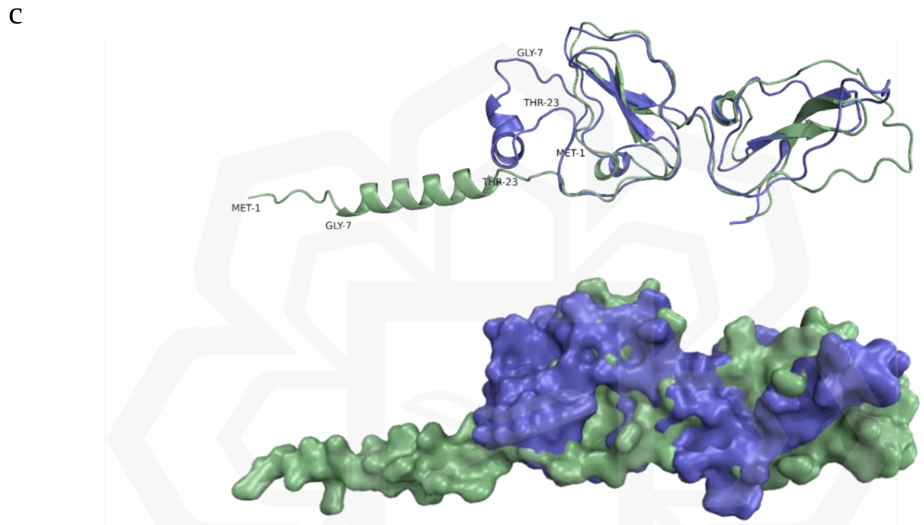
Figure 3.6 The progress of HE4 predicted conformation modelled by AlphaFold during MD simulation until it reaches equilibrium. The protein undergoes a conformational change towards a state of greater stability, forming (f) dominant conformation, RF1.

The notable changes in structure between the AlphaFold and RF1 models are obvious in the helix motif spanning from glycine-7 (GLY-7) to threonine-23 (THR-23) (Figure 3.7). Upon attaining stable conformations, this area underwent inward folding from 0 ns until 8 ns, resulting in increased atom-atom interactions with adjacent residues, hence contributing to the overall stability of the three-dimensional structure. These interactions can take place *via* hydrogen bonds, disulphide bonds, Van der Waals forces, ionic contacts, and hydrophobic interactions (Petsko & Ringe, 2004). These modes of interaction are essential for stabilising the secondary and tertiary structures of the protein. Proper protein folding is crucial for ensuring optimal activity and stability, as proteins are less susceptible to denaturation.



AlphaFold

RF1



AlphaFold and RF1 aligned

Figure 3.7 The tertiary structures of HE4 by AlphaFold, RF1, and the alignment of both. (a) HE4 modelled by AlphaFold, (b) RF1, and (c) the alignment of (a) and (b), showing the structural differences between the two HE4 models. The 3D structures are shown in cartoon and surface viewing modes.

The RF1, RF2, and RF3 models were analysed further with PBDsum, which led to the creation of wiring diagrams for the secondary structure (Figure 3.8). This wiring diagram serves as a useful tool for comprehending the interactions between the components of the protein. Similar to the PSIPRED server prediction, PDBsum visually represents the overall secondary structure of the protein by illustrating the helices, strands, and numerous motifs such as the beta-turn and beta-hairpin (Laskowski et al., 2018). The RF1, RF2, and RF3 models exhibited similar structural patterns found in native proteins, including alpha-helices and beta-strands. The RF1 model has three helices (H1, H2, and H3) and multiple strands derived from various beta sheets (A, B, and C). The RF3 model demonstrates a greater number of helices spanning amino acids 1 to 124, with fewer strands. The yellow-linking bars in the three wiring schematics exhibited disulphide bonds. For RF2, which is the refined model from the MD simulation of the I-TASSER model, the protein structure exhibited a lack of helices and strands, rendering it a less desirable HE4 model in comparison to RF1 and RF3, as it did not resemble the structural motifs of native proteins.

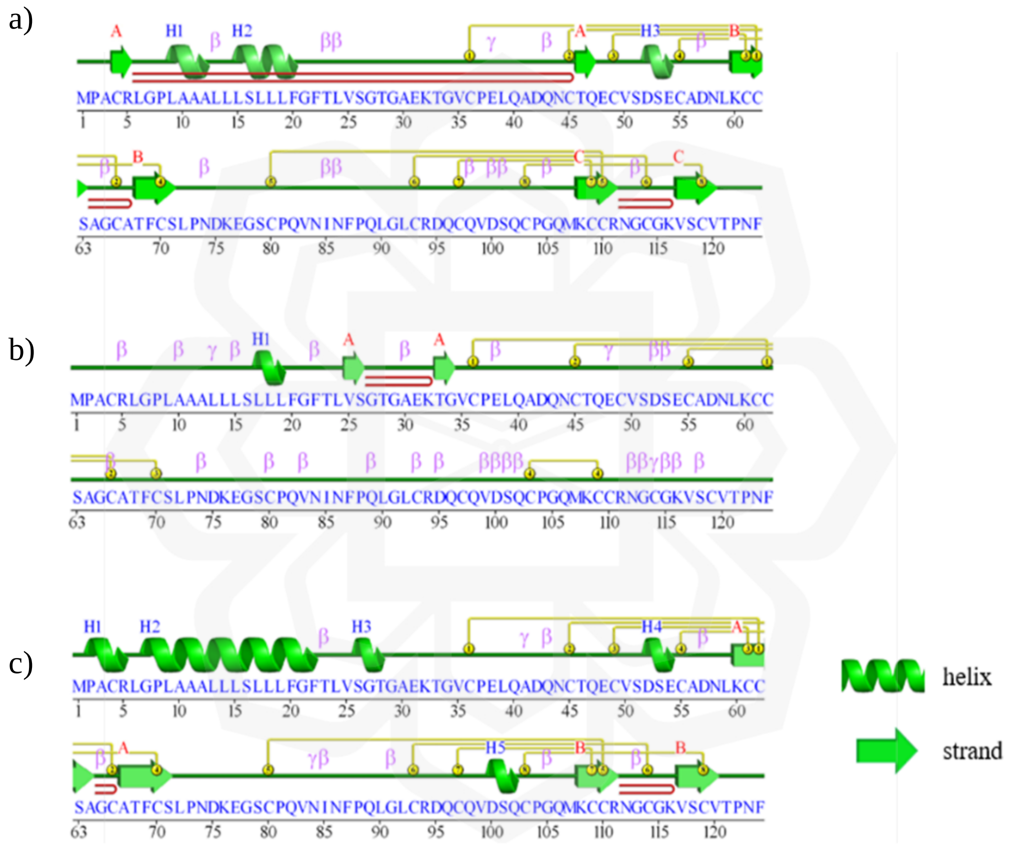


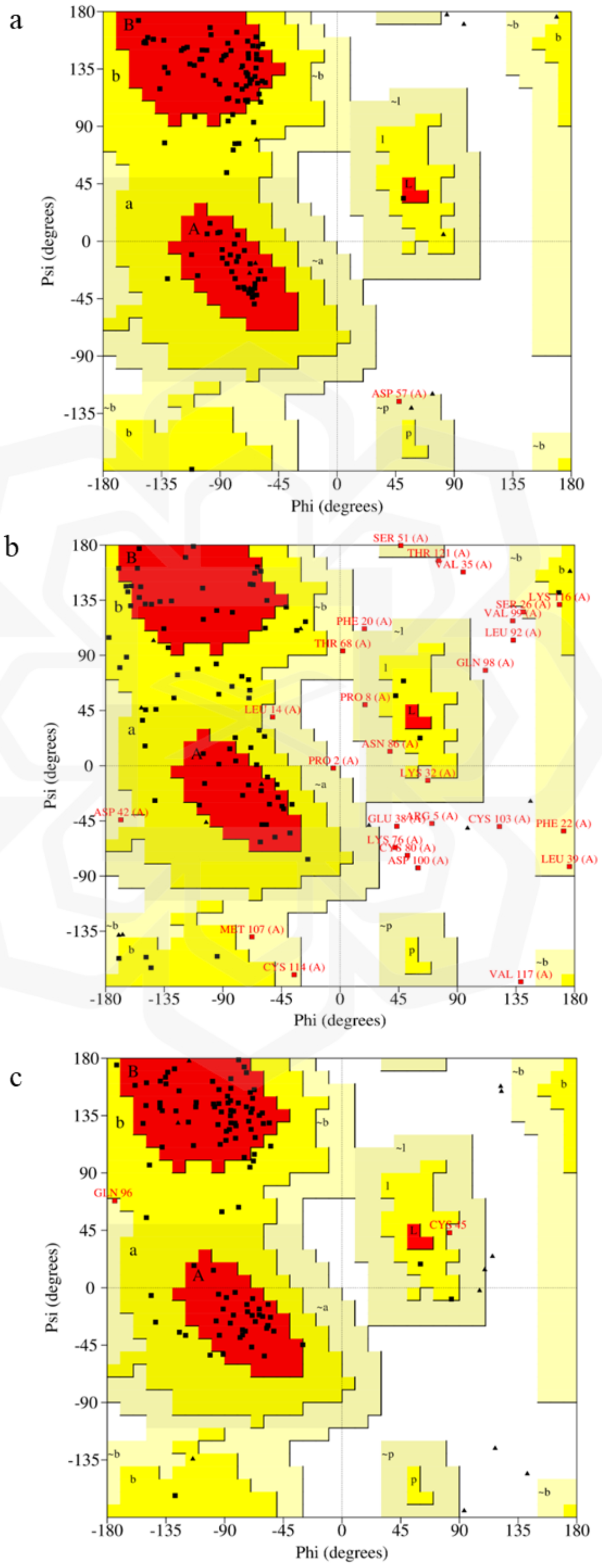
Figure 3.8 The schematic wiring diagram of the HE4 representing the secondary structure elements: alpha-helices and beta-sheets, obtained from the PDBsum analysis; (a) RF1; (b) RF2; (c) RF3. The helices are marked as H1, H2, H3, and H4, and the strands are labelled according to which beta sheet they belong to (A, B, C).

Alpha-helices play a crucial role in proteins by providing structural stability and maintaining the protein's structure (Popot & Engelman, 2000). They facilitate the efficient arrangement of the polypeptide chain, promoting a more compact protein structure that optimises interactions among the amino acids. Beta-strands, on the other hand, play an important role in upholding the structural stability of proteins through the establishment of hydrogen bonds with neighbouring strands, leading to the creation of beta sheets. Based on these essential roles of the motifs, the protein structures of RF1 and RF3 exhibited a strong resemblance to the typical secondary structure components found in native proteins, as proven by the presence of numerous helices and strands.

3.3.3 HE4 Structure Validations

The predicted models (AlphaFold, I-TASSER, and Robetta) and the refined models (RF1, RF2, and RF3) were validated using PROCHECK and ERRAT on the SAVeS server. The PROCHECK programme constructed a Ramachandran plot illustrating the φ (phi) and ψ (psi) torsional angles of the polypeptide backbone of the amino acids (Hollingsworth & Karplus, 2010). The quality of the HE4 protein structure can be classified according to the distribution of residues within the favoured, allowed, and disallowed regions. A high-quality model should exhibit more than 90% of its residues within the favoured regions, with merely 0-5% in the allowed regions and under 0.5% in the disallowed regions, indicating strong structural reliability (Elalouf, 2023). A good-quality model generally exhibits 80 to 90% of residues in the favoured areas, approximately 10% in the allowed areas, and less than 1.5% in the disallowed areas. Models of lower quality show fewer than 80% of residues in the favoured regions, exceed 10% in the allowed regions, and surpass 1.5% in disallowed regions, suggesting possible errors or instability in the protein structure.

All residues in the AlphaFold and Robetta models, including their refined models RF1 and RF3, were located within the allowed regions. On the other hand, the I-TASSER model and its refined model RF2 had 13 and five residues (glutamic acid-38, valine-35, cysteine-66, valine-99, and aspartic acid-100), respectively, that were found in the disallowed region (Figure 3.9). The disallowed regions are characterised as areas where there is significant steric hindrance between the torsions (Gunasekaran et al., 1996; Ho & Brasseur, 2005). This region corresponds to the combinations of φ and ψ angles that are not often observed in well-structured native protein conformations, as they would cause strain in the protein's backbone. Glycine lacks a side chain, allowing it to freely adopt the torsional angles inside any quadrant of the Ramachandran plot. According to Ho and Brasseur (2005), the first quadrant (top-left) represents the beta-sheet area, the bottom-left quadrant represents the right-handed alpha-helix, and the top-right quadrant represents the left-handed alpha-helix region.



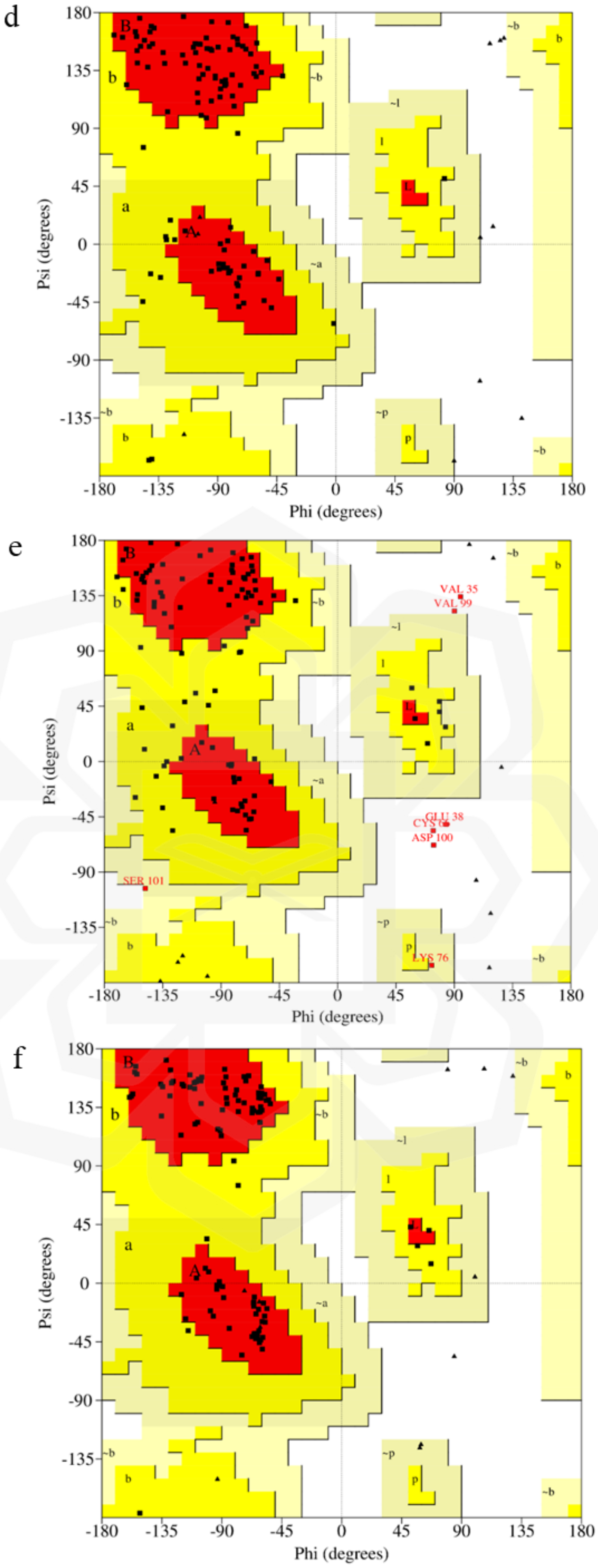


Figure 3.9 The Ramachandran plots of the HE4 protein models: (a) AlphaFold, (b) I-TASSER, (c) Robetta, (d) RF1, (e) RF2, and (f) RF3. The residues were plotted in different regions: most favoured (red), additional allowed (yellow), generously allowed (pale yellow), and disallowed region (white).

The RF1, RF2, and RF3 models exhibited greater stability and accuracy, compared to AlphaFold, I-TASSER, and Robetta, respectively, based on the distribution of the φ and ψ dihedral angles of the amino acids. In Table 3.4, the number of amino acids found in the most favoured regions increased for the refined models, where the residues located within these regions exhibit stable backbone geometries without steric strains (Zhou et al., 2011). On the other hand, the residues in the additional and generously allowed regions are less favourable but are still permitted, where steric strains can possibly occur owing to local structural constraints or specific interactions. Based on the Ramachandran plot analysis, the RF1 model exhibits the highest quality protein structure, as 87.4% of the residues are situated inside the most favoured regions. Additionally, none of the residues were found in the disallowed regions.

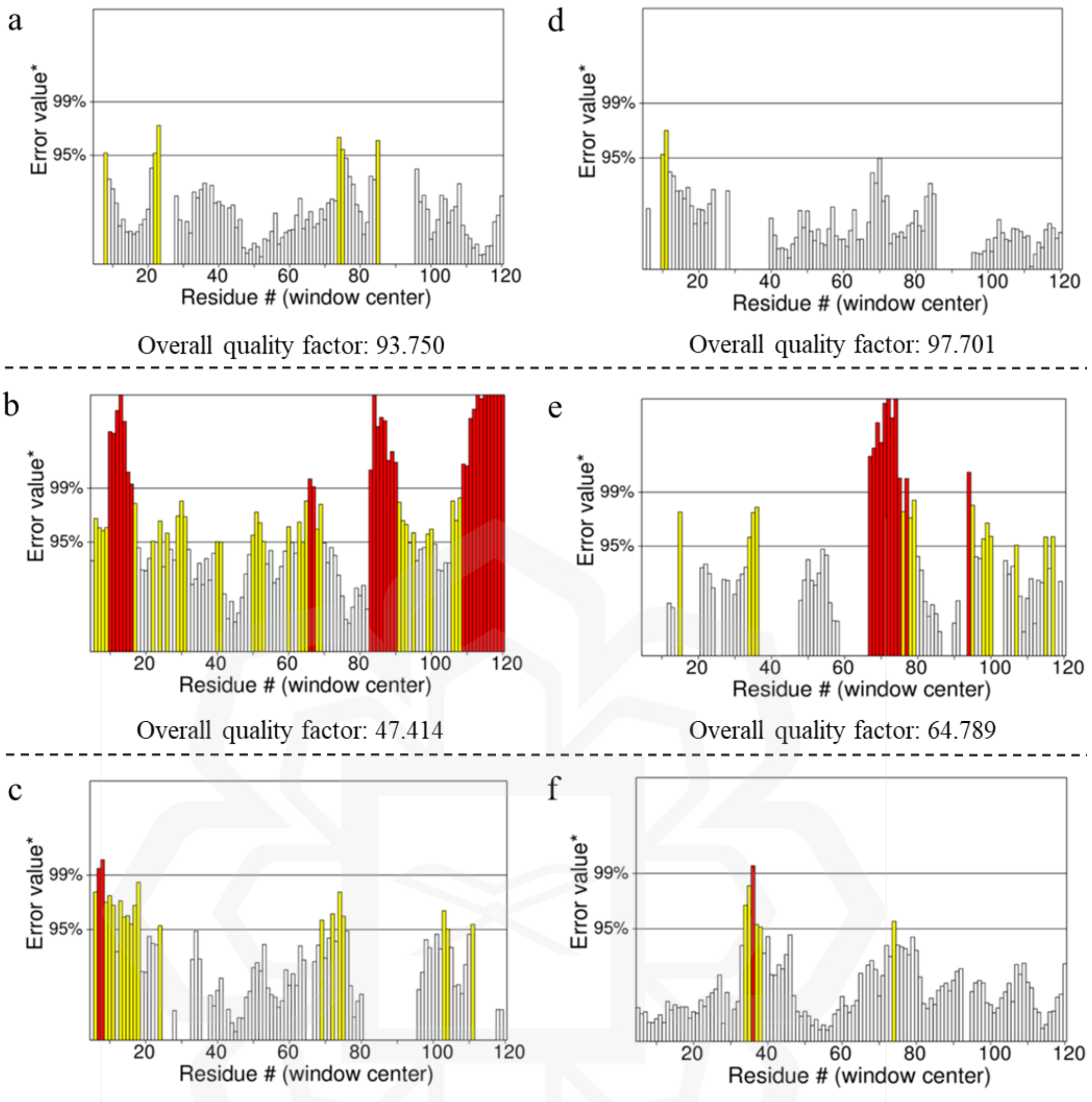
Table 3.4 The percentage of HE4 residues in regions of the Ramachandran plot

Region	Region colour	AlphaFold	RF1	I-TASSER	RF2	Robetta	RF3
Residues in most favoured regions (A, B, L)	Red	85.4%	87.4%	35.0%	65.0%	80.6%	82.5%
Residues in additional allowed regions (a, b, l, p)	Yellow	13.6%	12.6%	40.8%	28.2%	17.5%	17.5%
Residues in generously allowed regions (~a, ~b, ~l, ~p)	Pale yellow	1.0%	-	11.7%	1.9%	1.9%	-
Residues in disallowed regions	White	-	-	12.6%	4.9 %	-	-

In addition to the Ramachandran plot validation, the ERRAT tool is used to assess the protein structures by calculating overall-quality-factor scores, relying on the interactions of the non-bonded atoms between the residues (Pradeepkiran et al., 2021). This study observes the distribution of atomic interactions inside the protein and compares it to statistical data obtained from high-resolution crystallographic structures. The approach entails comparing the observed non-bonded interactions between residues with the statistical database, enabling the identification of potential errors or anomalies in the protein model and validating its structural accuracy in conjunction with other tools such as the Ramachandran plot.

ERRAT is represented by a numerical number ranging from 0 to 100, where 100 indicates the model with the best quality. According to the evaluation (Figure 3.10), RF1 demonstrates high-quality HE4 protein structure, achieving an overall quality factor score of 97.701, which is an improvement compared to the AlphaFold model that scored 93.750. The I-TASSER model has the lowest quality, with an overall quality factor score of just 47.414. An error value greater than 99% indicates a residue region that has been poorly modelled (Al-Khayyat & Al-Dabbagh, 2016), where 29 residues of the I-TASSER HE4 model are in this region, resulting in the low score.

Despite performing MD simulation on the I-TASSER system, the overall quality factor score for the refined conformation, RF2, improved but remained low at 64.789, with 11 outliers. The outliers (in red bars) suggest inaccuracies or lower-quality protein conformation, which might affect its functionality. According to the Ramachandran and ERRAT plots, both the RF1 and RF3 models exhibit high-quality tertiary structures for HE4, highlighting that the RF1 model performed slightly better since it did not have any outliers in either plot.



Overall quality factor: 93.750

Overall quality factor: 97.701

Overall quality factor: 47.414

Overall quality factor: 64.789

Overall quality factor: 75.000

Overall quality factor: 94.683

Figure 3.10 The ERRAT plots of the HE4 protein models: (a) AlphaFold, (b) I-TASSER, (c) Robetta, (d) RF1, (e) RF2, and (f) RF3. The red bars represent outlier residues where the region significantly deviates from typical, well-folded, native protein structures.

In a separate analysis, the CAVER 3.0 tool was utilised to analyse protein structures, emphasising the identification of the protein tunnels and cavities within these structures (Chovancova et al., 2012; Stourac et al., 2019). The CAVER offers a user-friendly interface and generates 3D visualisations of the predicted tunnels and cavities. It calculates the paths and cavities using an algorithm based on Voronoi diagrams and MD simulations, where it predicts dynamic changes in tunnels and channels by combining geometric analysis with the MD simulations. In this study, the best-modelled structure RF1, was analysed, and it was revealed that a single tunnel was identified (Figure 3.11), with a length of 1.4 Å, a bottleneck radius of 2.0 Å, and a distance-to-surface of 1.1 Å. The bottleneck and anticipated cavity involved 15 amino acid residues: LEU24, VAL25, SER26, ALA30, GLU31, LYS32, CYS49, VAL50, SER51, ASP52, SER53, LYS60, CYS61, CYS62, and SER63 (Table 3.5). The tunnel facilitates the transportation of tiny molecules, including water, ions, and substrates, into and out of proteins, therefore directing the selective binding of a ligand to its active site (Brezovsky et al., 2013).

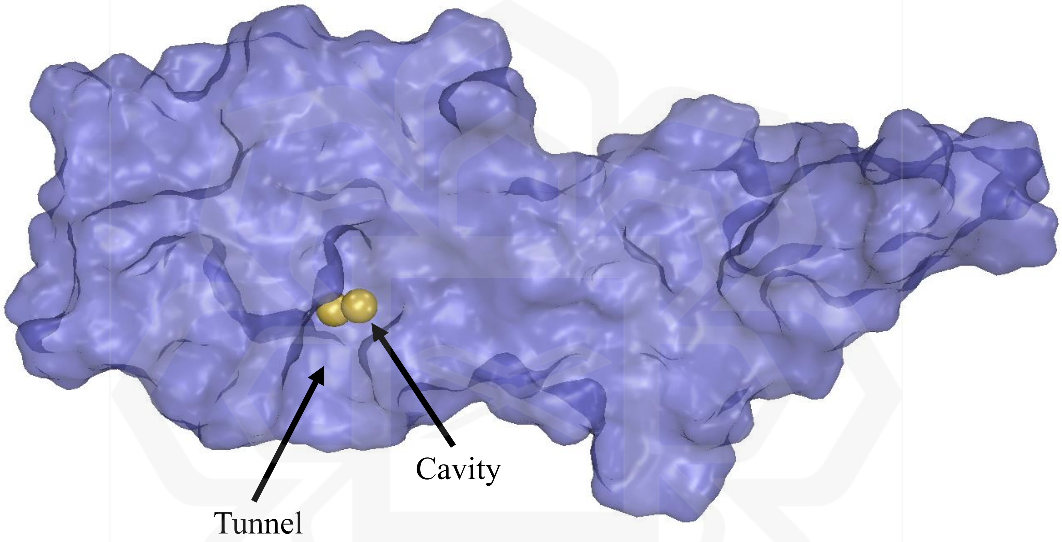


Figure 3.11 The tunnel and binding site identified by CAVER 3.0. This tunnel visualises and analyses the potential pathway or binding site for ligands to move through them and bind strongly with the surrounding residues.

Table 3.5 Tunnel details and the amino acids in the region

Number of bottlenecks	1	
Bottleneck radius	2.0 Å	
Tunnel length	1.4 Å	
Distance to surface	1.1 Å	
Number of residues	15	LEU24, VAL25, SER26, ALA30, GLU31, LYS32, CYS49, VAL50, SER51, ASP52, SER53, LYS60, CYS61, CYS62, SER63

This discovery enhances our understanding of the protein interactions with other biomolecules, thus increasing the possibility for its utilisation in other computational studies, particularly in this current study where this modelled HE4 protein structure was further used in investigating its binding with DNA aptamers through molecular docking and MD simulations.

3.4 CONCLUSION

The tertiary structure of HE4 was accurately predicted using various computational tools, and the results were validated satisfactorily. The molecular dynamics simulations demonstrated that the conformation acquired stability within 8 ns and maintained it until the end of the simulations for all three systems (AlphaFold, I-TASSER, and Robetta). The secondary structural components shown in the wiring diagram show that the RF1 model share few similar features as native protein structures, such as having multiple helices and strands. According to the structure evaluation tools PROCHECK and ERRAT, this RF1 model has the highest quality HE4 protein tertiary structure. All amino acids were located in the most favoured regions of the Ramachandran plot, and the model achieved the highest overall quality factor score of 97.701 from the ERRAT analysis. Finally, the CAVER 3.0 tool identified a small tunnel that might potentially enable the movement of molecules towards the active site. Undoubtedly, the findings of this computational study can be extensively employed for future investigations, including the advancement of diagnostic methods and drug delivery. The most accurately represented tertiary structure of HE4 (RF1) derived from this research can be utilised for subsequent computational investigations, such as molecular docking and molecular dynamic (MD) simulations, using different ligands.

CHAPTER FOUR

SCREENING OF THREE-DIMENSIONAL DNA APTAMERS AGAINST HE4 PROTEIN THROUGH MOLECULAR DOCKING AND ITS BINDING ANALYSIS

4.1 INTRODUCTION

DNA aptamers' tertiary (or 3D) structure determines their functionality, specificity, stability, and application in a variety of biotechnological and biological disciplines, including the applications in diagnosis and therapy (Chen et al., 2017). Based on algorithms and experimental validations, *in silico* or computational method can design the secondary and tertiary conformations of aptamers with high binding affinities (Buglak et al., 2020). The foundation of the aptamer folding that leads to secondary and tertiary structures is the primary sequence, which is the linear arrangement of nucleotides. The sequence determines the potential folding pathways and the interactions that take place during aptamer folding. Based on complementary base pairing, or known as Watson-Crick base pairing, and the stacking interactions between nucleotides, aptamers may form a variety of secondary structural motifs such as hairpins, G-quadruplex, pseudoknots, and bulges (Duan et al., 2020; Sullivan et al., 2019; Yu et al., 2021).

In 1953, Watson and Crick scientifically reported a right-handed double helix structure composed of two DNA strands (Watson & Crick, 1953). This structure was deduced by considering the relative quantity of each DNA base and the necessary conditions for forming a double helix with opposite orientations. The helix structure is formed by hydrogen bonding between the nucleobases adenine (A) and thymine (T, or uracil; U for RNA), as well as guanine (G) and cytosine (C), resulting in the formation of Watson−Crick base pairs (Figure 4.1). The stability of the double helix in DNA primarily relies on the hydrogen bonds (H-bonds) formed between A-T and G-C base pairs through non-covalent contacts, the stacking interactions between neighbouring bases throughout the helix, and the cross-interactions between base pairs (Beiranvand et al., 2021; Poater et al., 2014). Adenine and thymine consist of one donor and one acceptor, whereas each cytosine has one donor and two acceptors. Guanine, on the other hand, contains two H-bond donors and one acceptor.

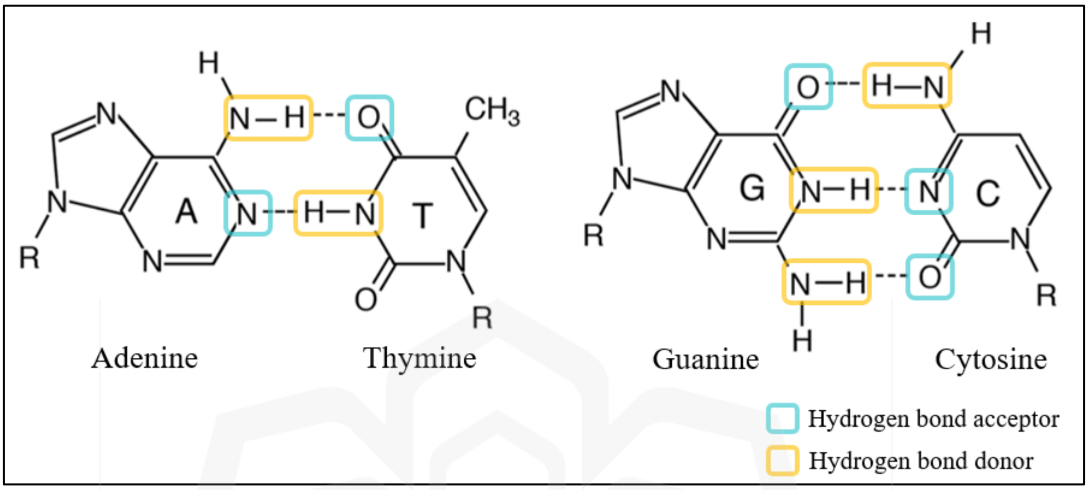


Figure 4.1 The hydrogen bonds (dashes) between adenine (A) and thymine (T), and guanine (G) and cytosine (C) based on Watson-Crick base pairs (Takahashi & Sugimoto, 2021).

The secondary structures of aptamers contribute to the determination of the overall tertiary structure conformations. The hairpin (stem-loop) structure is a common and significant structural motif in aptamers, where the stem is a double-stranded sequence region formed by complementary base pairing that provides structural stability to the aptamer while the loop is a single-stranded length that connects the stem's two arms and is generally unpaired and more flexible (Ilgu et al., 2019). The loop typically includes the precise binding location of the target molecule, proven in an earlier study where 15-mer and 28-mer fluorescein amidite-labelled oligonucleotides used in the docking with peptides and amino acid (Fadeev et al., 2022). The binding occurred at T5 residue located at the hairpin loop, alongside with multiple interactions at the open strands of the 3’ and 5’ end terminals. The stem of the hairpin is usually lacked binding occurrence with target molecule due to the rigidity of atoms that are held by strong H-bonds in the Watson-Crick base-pairs (Long et al., 2023).

There are various RNA and DNA secondary structure prediction tools available, including RNAfold, CentroidFold, Mfold, RNAstructure and Vfold (Afanasyeva et al., 2019; Lee et al., 2023; Sato et al., 2009; Xu et al., 2014). The Mfold web server is a popular tool for predicting the secondary structure of single-stranded nucleic acids. It determines the minimum free energy secondary structures and is currently used in a variety of research to predict nucleic acid folding and hybridisation (Zuker, 2003). In a recent work, the secondary structure of nucleic acid mimics (NAM), RNA and DNA aptamers with PDB ID of 5D3G, 5HRT, 2NP9, 6U82, and 4PDB were successfully designed by Mfold, where the Gibbs Free Energy, ΔG ranged between -33.50 and 4.58 kcal/mol for the selected structures (Oliveira et al., 2022). The 3D structures of the aptamers were then generated by a fully automatic web server, 3dRNA and refined by QRNAS software where it utilises existing database structures as templates to create a comparative model. For aptamer tertiary structures, they can be generated by RNAComposer (Sarzynska et al., 2023), trRosettaRNA (W. Wang et al., 2023), 3dRNA (Zhang et al., 2020), and many other servers. These servers are generally used to predict RNA structures; however, due to the non-existent DNA aptamer tertiary structure prediction servers, they were also utilised to generate DNA aptamer 3D structures. This can be achieved with additional steps which include manual conversions of the uracil to thymine and the ribose sugar to deoxyribose as applied by several previous studies (Oliveira et al., 2022; Sabri et al., 2020).

The aptamer tertiary structure, despite being relatively short in comparison to proteins, it is critical for several reasons, including assuring their ability to bind to target molecules with high specificity and affinity, determining aptamer functional activity, and promoting stability and durability (Zhou & Rossi, 2017). Following the aptamer secondary and tertiary conformations prediction, molecular docking with target, HE4, take place. Molecular docking is a process that utilises computational simulations to predict the interaction, including binding poses and affinities, between a small molecule and a protein, and this underlying concept was adapted to study the interactions between aptamers and proteins. This docking can be achieved by HDOCK (Yan & Huang, 2020), AutoDock (Seeliger & De Groot, 2010), GOLD (Verdonk et al., 2003), and Patchdock (Schneidman-Duhovny et al., 2005) web servers. AutoDock software package includes two applications, the AutoDock4 (Morris et al., 2009) and AutoDock Vina (Trott & Olson, 2010). Due to improved search algorithm and enhanced scoring function, AutoDock Vina is a more popular docking tool compared to AutoDock4, resulting in more precise and faster predictions (Nguyen et al., 2020). The selection and suitability of a docking server is based on the specific interaction being analysed and the available structural data pertaining to the molecules studied. Although numerous docking models were generated for most docking tools, frequently, the model with the highest binding affinity for each aptamer was extracted for further study.

This chapter emphasises on the *in silico* DNA aptamer structural design, based on available primary sequence of aptamer candidates against HE4 biomarker, obtained from a study that applied the *in vitro* approach, SELEX (Eaton et al., 2015). The designing process was firstly initiated by generating the secondary aptamer conformations, by the Mfold nucleic acids folding programme, followed with 3D structure prediction using RNAComposer. As the RNAComposer converted the DNA sequence to RNA 3D conformation, manual editing of the atoms was needed, to translate the conformation to the desired 3D DNA aptamer structure (Sabri et al., 2019; Wang et al., 2019). However, there’s no validation programme available to verify the accuracy and quality of the DNA 3D structure, unlike verification of protein tertiary structures. Hence, comparing computational predictions to experimental data helps to establish the model's accuracy. This was done by comparing the aptamer's three-dimensional foldability using the sequential procedure above with its aptamer structure that was experimentally produced and deposited in the PDB (Sabri et al., 2020). This similar technique was used to predict the three-dimensional folding of proteins in Critical Assessment of Structure Prediction (CASP) (Kryshtafovych et al., 2021). This research is also significant since it is hypothesised that the folded hairpin structure, consisting of the stem and loop regions, offers a functional 3D structure which are capable to bind strongly to the target. It is exhibited in this study that the stem and loop regions of the aptamers were discovered, and the binding affinities were analysed.

4.2 METHODOLOGY

Determining the most potential DNA aptamer against HE4 biomarker involved the preparation of the aptamer candidates, followed by the molecular docking of the tertiary structure of the aptamers with HE4 structure that was modelled in the previous chapter. The binding affinities and binding regions were analysed through molecular docking, prior to choosing the most potential DNA aptamer to be further utilised in this study.

4.2.1 Preparation of DNA Aptamer Candidates

HE4 aptamer candidates were obtained from Eaton et al. (2015) where the study elucidates the use of capillary electrophoresis (CE)-SELEX as the separation technique in the procedure of identifying DNA aptamers that exhibit affinity towards HE4 biomarker. The four aptamer candidates were renamed as A1, A2, A3, and A4. Sequences A1 and A4 were the most enriched sequence in round 5 and round 2 of the CE-SELEX, respectively, and they did not appear in any of the negative selection rounds. A2 ranked third, while A3 ranked second most enriched sequence in rounds 5 and round 4, respectively (Eaton et al., 2015).

Table 4.1 Four aptamer candidates against HE4 protein (Eaton et al., 2015)

Aptamer candidate	Sequence	Length (base)
A1	TTATCGTACGACAGTCATCCTACAC	25
A2	CACAGTGCGTCACATTTAGGGCATT	25
A3	CAGTGCGTGCTTATTGGCGTAGCGTC	26
A4	ATGGTCGCAAGAACTGAGAATTTAC	25

The secondary structure of each candidate was predicted using the Mfold server (http://www.unafold.org/mfold/applications/dna-folding-form.php), a web programme that predicts the secondary structure of single strand nucleic acids. The parameters were set as default with temperature of 37°C, and ionic conditions of 1.0 M Na^+ and 0.0 M Mg^{2+} ions. The most thermodynamically stable structure with the lowest folding energy (in kcal/mol) were selected and the dot-bracket conformation, or known as Vienna format, was downloaded to be utilised in determining its tertiary structure using RNAComposer web server. The Vienna format of the DNA aptamer candidate was used as the input for the RNAComposer server and equivalent RNA tertiary structure was assembled. The pdb file of this RNA structure was downloaded and opened in PyMOL molecular viewer, which was then manually converted to DNA aptamer. This was done by converting the ribose sugar to deoxyribose, by changing the hydroxyl group (2'-OH) to –H on the backbone. All uracil residues were converted to thymine by adding methyl group ($-CH_3$) to C-5 in the aromatic ring. All manual editing was done using the builder tool in PyMOL.

4.2.2 Validation of Aptamer Designing Method

In previous chapter, the modelled HE4 protein was validated according to established validation approach, including Ramachandran plot and ERRAT. Unfortunately, for aptamers, no validation tools are available, up to date. Thus, this study utilised the following approach to validate the aptamer designing method applied in subchapter 4.2.1 above. The sequence of six random DNA aptamers, where the crystallised structures have been discovered and deposited in the PDB, were selected and tested. The tertiary structures for the aptamers were designed by Mfold and RNAComposer, before the atoms were converted from RNA to DNA aptamers. These steps followed the design pipeline in previous section. The six aptamers are listed in Table 4.2.

Table 4.2 Sequence and length of six DNA aptamers retrieved from PDB

PDB ID	Sequence	Length
1ZHU	CAATGCAATG	10
1BJH	GTACAAAGTAC	11
2M8Y	CGCGAAGCATTCGCG	15
1ECU	GCGCGAAACTGTTTCGCGC	19
2L5K	CAGTTGATCCTTTGGATACCCTG	23
1JVE	CCTAATTATAACGAAGTTATAATTAGG	27

The pdb file of each crystallised aptamer structure were downloaded and viewed in PyMOL molecular viewer, together with the 3D structure of the designed aptamer. These two 3D structures were superimposed and the RMSD value between the two was recorded. In this analysis, the calculation of RMSD provides the information on the structural similarity between the designed and crystallised structures.

4.2.3 Molecular docking of HE4 and DNA aptamer candidates

Molecular docking was applied in predicting and analysing the interactions occurred between the HE4 protein and DNA aptamer candidates at an atomic level. The molecular docking approach between the two molecules was adopted from a work done by Sabri et al (2015). To initiate molecular docking, preparation of the input, ligand, and receptor, is required. This was done by AutoDock Tools (ADT) where the water molecules were removed and the grid box of the receptor (HE4 protein) was created to ensure the whole protein structure was covered, as the grid defines the region where docking will be conducted, exploring possible binding orientations. The ADT assigned polar hydrogens and calculated the Kollman and Gasteiger charges. These charges were applied to assist the docking step to estimate the interactions between the ligand (DNA aptamer) and HE4 protein. For the preparation of the ligand, the root of torsion tree was determined by ADT before the torsions was chosen. Torsions represent the rotations around the single bonds of the molecule which the rigidity and flexibility allow them to adopt different orientations to fit into the binding site (Elokely & Doerksen, 2013). This increases the chances of identifying the most favourable binding interaction between the DNA aptamer and HE4 protein. Both aptamer and protein prepared structures were downloaded as pdbqt files to undergo molecular docking.

The aptamer candidates were docked with HE4 protein, individually, using AutoDock Vina. This docking tool uses a hybrid scoring function, integrating empirical and knowledge-based approach which provides 10 lowest binding energy (kcal/mol) of the DNA aptamer-protein complexes, arranged from the most negative to the least negative energy. The grid box size was set to 82 × 44 × 84 of x, y, and z points with grid centre at the dimensions of 64.672, 65.933 and 65.56 for centre x, centre y and centre z, respectively. This grid box fully covered the protein structure, enabling blind docking to be performed. These parameters were set in the configuration file, along with the receptor and ligand pdbqt file names. Molecular docking was performed and the pose with the lowest energy of binding energy was extracted and aligned with the HE4 protein structure in PyMOL for further analysis. The analysis includes the binding energies, and the residues involved in the interactions between the DNA aptamer candidate and the HE4 protein, for all four DNA aptamer-HE4 complexes. The receptor-ligand interactions were analysed by Protein-Ligand Interaction Profiler (PLIP), available at https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index.

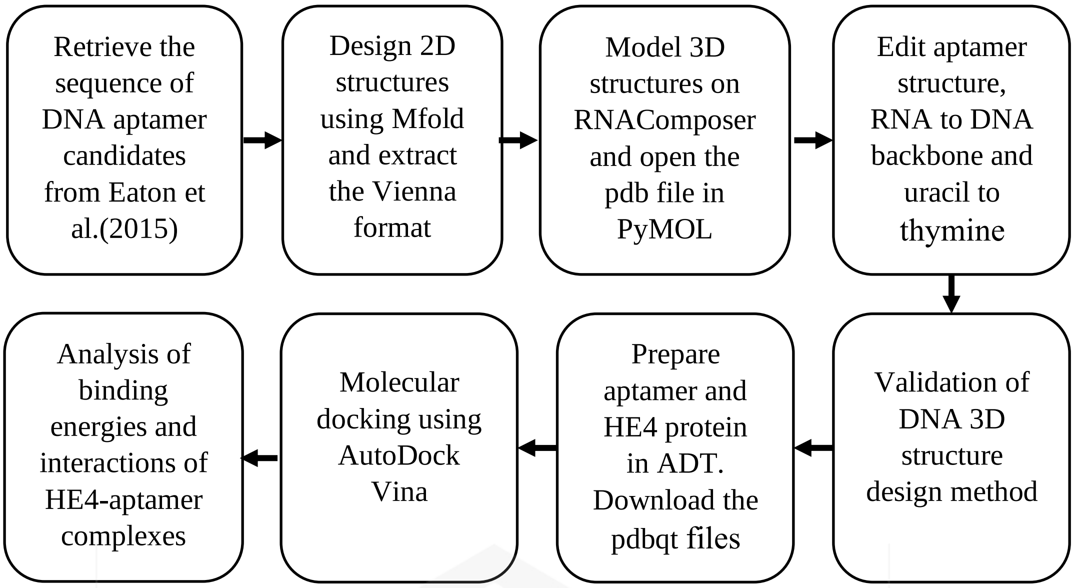


Figure 4.2 Pipeline for the DNA aptamer structural design, its validation and HE4-aptamer molecular docking.

4.3 RESULTS AND DISCUSSION

4.3.1 Secondary and tertiary structures of aptamer candidates

The prediction of aptamer secondary structures began with the nucleotide sequences of four aptamers, taken from an experimental study by (Eaton et al., 2015). Table 4.1 displays the aptamer sequences (A1 to A4) that were selected from the CE-SELEX process conducted in that study, where its main objective is to select potential DNA aptamer candidates, detecting ovarian cancer biomarker, HE4 (Eaton et al., 2015). A1, A2, A3, and A4 were the four potential aptamer sequences where they were enriched through the selection course and were indicative of big clusters. This current study focuses on the aptamer secondary and tertiary structure predictions by Mfold and RNAComposer servers. The Mfold server determines the secondary structure with the minimum free energy, which corresponds to the most stable arrangement of the RNA (Ojha et al., 2021) which is then provided as the output of the Mfold prediction. It often offers a graphical depiction of the anticipated secondary structure, showcasing the pairing of nucleotide bases, stem-loop formations, internal loops, bulges, and other characteristics (Duan et al., 2020). The Mfold calculated and generated the free binding energies and dot-bracket conformations (Vienna format), displaying the regions of the hairpin (stem and loop), and bulges, where the brackets representing the residues involved in the Watson-Crick base-pairings (Table 4.3).

Table 4.3 The sequences of aptamer candidates and their dot-bracket conformations, representing secondary structures, retrieved from Mfold server.

Aptamer	Number of nucleotides and sequence		Vienna format	Free energy, kcal/mol	Number of Watson-Crick base pairs		
					G-C	A-T	Total
A1	25	TTATCGT ACGACAG TCATCCT ACAC	.......((....))..........	0.21	1	1	2
A2	25	CACAGTG CGTCACA TTTAGGG CATT	....(((...)))..........	-0.79	2	1	3
A3	26	CAGTGCG TGCTTAT TGGCGTA GCGTC	................((...))...	-1.31	2	0	2
A4	25	ATGGTCG CAAGAAC TGAGAAT TTAC	.....(....)..............	-0.60	1	0	1

The core algorithm of Mfold relies on the dynamic programming methodology, notably the Zuker algorithm which it systematically examines all potential secondary structures and effectively computes the corresponding free energy for each potential structure (Zuker, 2003). The method utilises a recursive strategy to assess the stability of RNA structures by considering the energy contributions of each individual base pair. Table 4.3 describes the secondary structure of the four aptamers exhibited a range of minimal free energy values, spanning from 0.21 to -1.31 kcal/mol. The length of the base-pairing stem, as measured by the number of hydrogen bonds produced by Watson-Crick base pair, especially the G-C pairs, was shown to be correlated with the aptamer's free binding energy. This relates to the conformation stability, as a lower free binding energy (more negative) resulted in a more stable structure, demonstrated by previous study that conducted SELEX process to select DNA aptamers for histatin 3 (Ojha et al., 2021). The aptamers were more stable and formed more spontaneously, with lower

Gibbs free energy value. Referring to Table 4.3, the A3 aptamer has the longest sequence with 26 nucleotides with two G-C base pairs, producing the lowest minimum free energy of -1.31 kcal/mol. Although the A2 aptamer has more base pairs (three), but due to the presence of A-T base pair, it resulted in higher minimum free energy (-0.79 kcal/mol). This was proven in an earlier study on three aptamers with similar length; 1DGO (18 nucleotides), 1ECU (19 nucleotides), and 2VAI (18 nucleotides) with 7 base pairs each. The minimum free energies were -4.73, -7.16, and -4.73 kcal/mol, respectively (Sabri, 2021). Out of the 7 base pairs, 1ECU had five G-C base pairs, while the other two only had four, and the higher number of G-C base pair led to much lower free energy (-7.16 kcal/mol).

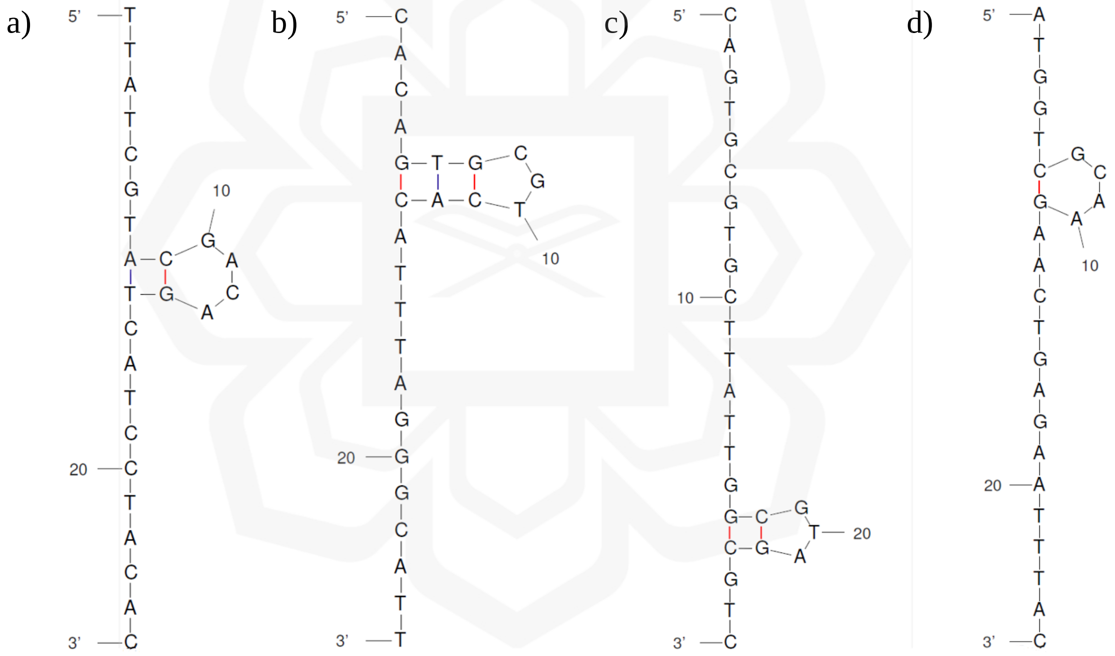


Figure 4.3 Secondary structures of the aptamer candidates, generated by Mfold. (a) A1, (b) A2, (c) A3, and (d) A4. All aptamers consist one hairpin structural motif. The stem, made up of Watson-Crick base-pairs, shown in red (A-T) and blue lines (C-G).

In the context of aptamers, a hairpin region denotes a structural motif present in the aptamer sequence, wherein a stretch of single-stranded RNA or DNA folds back into itself, resulting in the formation of a stem-loop structure. The stem is constructed by the base-pairing between nucleotides that are complementary to each other, whereas the loop is a segment that is single-stranded and connects the stems that are base-paired (Yu et al., 2021). At times, the loop region experiences conformational changes when it binds to the target molecule, as the bonds are rotatable and flexible. This leads to stronger interactions between aptamer and the protein. The aptamer's function relies on these crucial structural rearrangements, which can either inhibit a biological activity or initiate a signalling event. Following the secondary structure prediction by Mfold, the dot-bracket conformations were extracted and utilised in RNAComposer software, to generate tertiary structures of the aptamers. Based on its software name, it is widely used to predict RNA 3D structures, but due to the absence of DNA 3D structure prediction tool, the RNAComposer is integrated with other procedures to construct the DNA 3D structures (Oliveira et al., 2022). This includes the alterations of the backbone and other atoms (Figure 4.4) in the RNA predicted conformation.

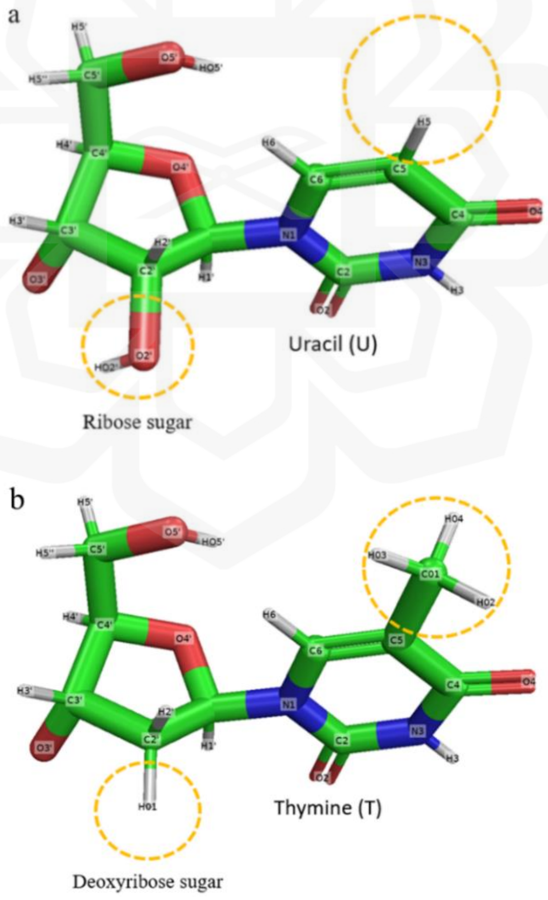
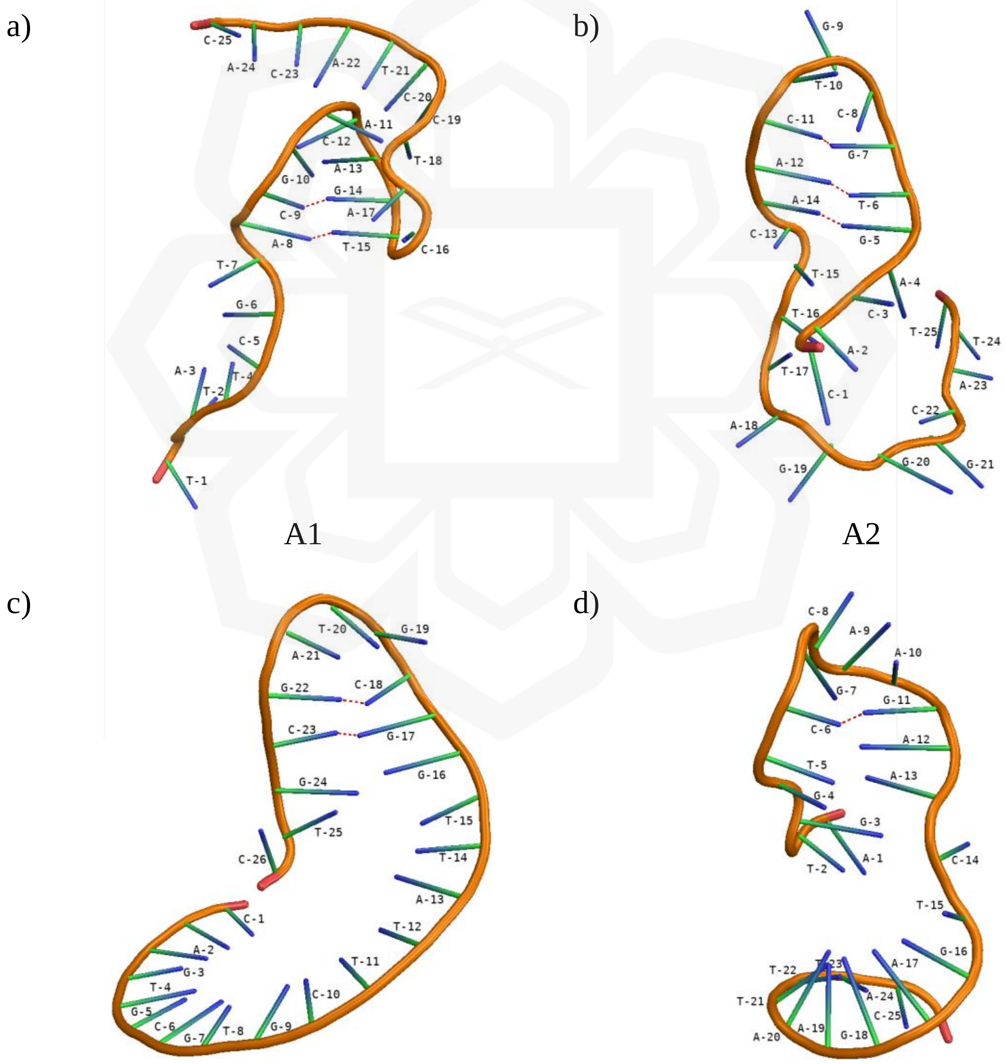


Figure 4.4 The conversions of ribose to deoxyribose sugar, and uracil to thymine. (a) The ribose sugar and uracil base of an RNA nucleotide, (b) The deoxyribose sugar and thymine base of a DNA nucleotide. The 2’-OH of the ribose was replaced

The main differences between DNA and RNA are their sugar that is part of the backbone, and the nucleotides. The sugar in RNA aptamer is ribose, where a hydroxyl (-OH) group can be found at the 2'-position (Egli, 2022). In this study, the hydroxyl was replaced with -H, converting the sugar to deoxyribose, which is the sugar of DNA. The conversion of RNA to DNA aptamer were achieved when the uracil (U) nucleobase of the RNA is altered to thymine (T) by substituting the H5 atom with a methyl ($-CH_3$) group. Figure 4.5 displays the 3D structures of the aptamer candidates after the required conversions were conducted in PyMOL molecular viewer. The red dashes represent the H-bonds formed by the Watson-Crick base pairing theory.



A1

A2

A3

A4

Figure 4.5 The tertiary structures of aptamer candidates, modelled by RNAComposer and converted to DNA aptamer. The H-bonds are shown in red dashes. These bonds corresponded to the number of Watson-Crick base-pairs in Table 4.2 and Figure 4.3.

In Figure 4.5, the stem and loop regions of the hairpin can be clearly observed where the stems are made up of the Watson-Crick base pairs, C-G and A-T. In example, the stem of A1 aptamer consists of A8-T15 and C9-A17 base pairs (Figure 4.5a), while the three base pairs that made up the single stem in A2 aptamer are G5-A14, T6-A12, and G7-C11 (Figure 4.5b). They have three and four nucleotides, respectively, forming the loop regions. A4 aptamer has only one base pair, formed by C6 and G11 (Figure 4.5c). Looking closely to the interactions between the Watson-Crick base pairs, Figure 4.6 illustrates three H-bonds in C-G base pair and two H-bonds in A-T base pair. This tertiary structure was truncated from the A1 aptamer, focusing on the hairpin region of the aptamer.

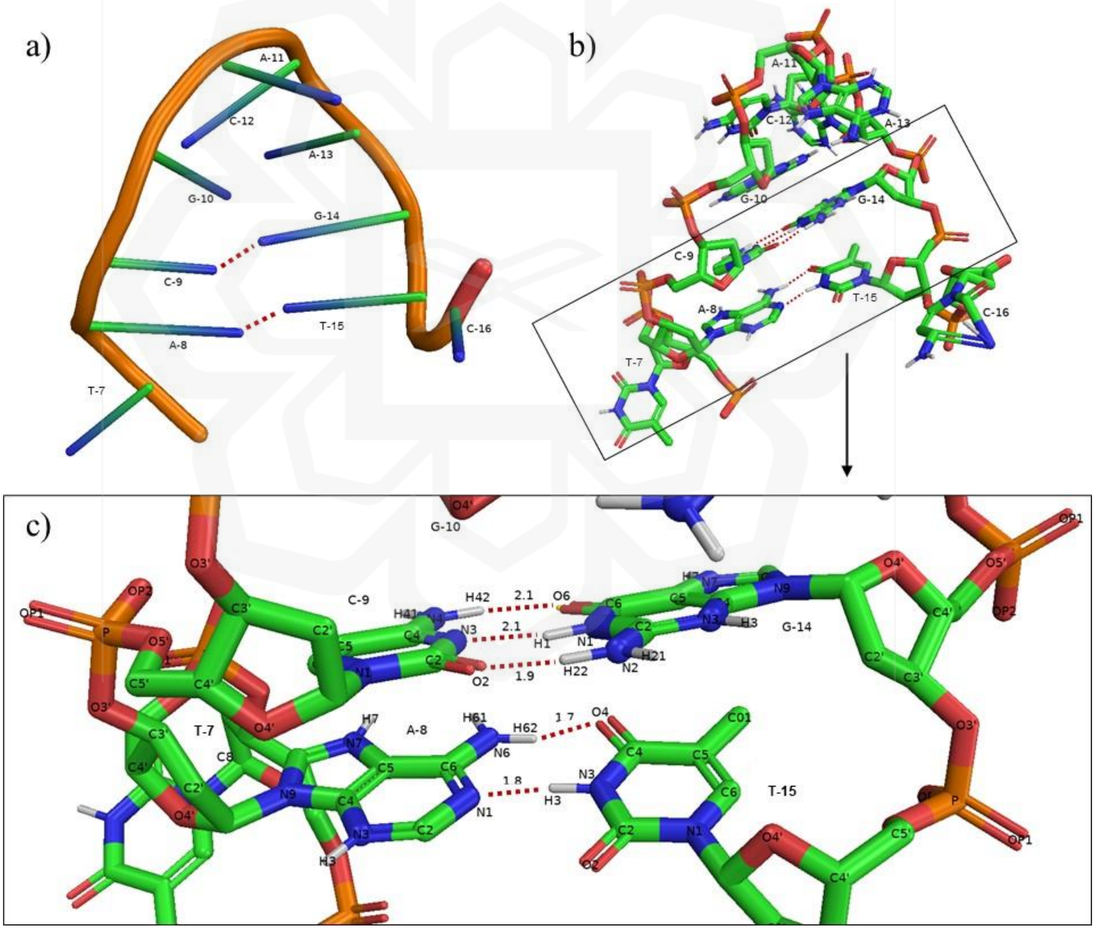


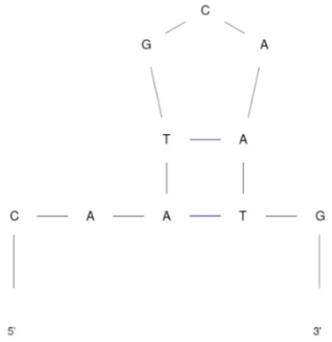
Figure 4.6 The 3D structure of truncated A1 aptamer, from T7 to C16. The H-bonds formed in the stem are shown in red dashes. (a) Cartoon view mode, showing the stem (base pairing between C9-G14 and A8-T15) and loop (G10 to A13), (b) Stick view mode, and (c) The H-bonds and their lengths (in angstrom, Å) of the C9-G14 and A8-T15 base pairs.

The hairpin consists of 8 nucleotides, four at the rigid stem (A8, C9, G14, and T15) and four at the flexible loop region (G10, A11, C12, and A13). The H-bonds lengths ranged from 1.7 Å to 2.1 Å (Figure 4.6c). According to Protein Structure and Function textbook, the lengths of H-bonds should be less than 3.5 Å, which the interactions occurred when a hydrogen atom bonded to electronegative atom such as oxygen and hydrogen atom is attracted to a highly electronegative atom of another molecule (Petsko & Ringe, 2004). As discussed earlier, G-C base pair causes stronger interaction compared to A-T, proven by higher number of H-bonds formed between guanine and cytosine residues.

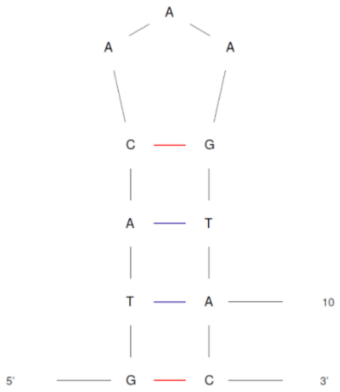
4.3.2 Aptamer Design Validation

This study validated the proposed pipeline for designing 3D aptamers where it involved the utilisation of web server applications, Mfold and RNAComposer, followed by manual atom alterations in PyMOL. The validation was conducted by comparing the all-atom RMSD score of the DNA aptamer crystallised structure available in the PDB database and the designed DNA aptamers after they were aligned (Sabri et al., 2021).

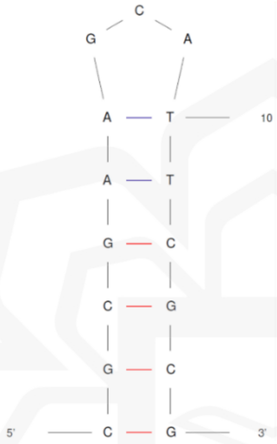
Six DNA aptamers with crystallised structures available in PDB selected for this validation study ranged from 10 to 27 nucleotides in length, while the actual HE4 aptamer candidates obtained from Eaton et al (2015) being studied in Subsection 4.3.1 had 25 to 26 nucleotides. This is to prove that the length of the actual aptamer candidates can apply the same pipeline in designing the 2D and 3D conformations. The sequences of the 6 selected DNA aptamers were utilised to predict the secondary structures by Mfold, followed by tertiary structures prediction by RNAComposer. The secondary structures of all 6 DNA aptamers have shown the presence of only one hairpin for each, and 1ECU aptamer consists of the greatest number of nucleotides at the loop of the hairpin with five nucleotides (Figure 4.7). The other aptamers, 1ZHU, 1BJH, 2M8Y, 2L5K, and 1JVE, has three nucleotides each at their loop with various numbers of base pairs making up their stem. Additionally, only aptamer 2L5K has an internal bulge formed by T4, T5, G6, A18, C19, and C20 residues, other than the hairpin loop made up of T11, T12, and T13.



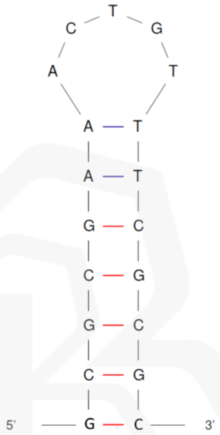
1ZHU (0.71)



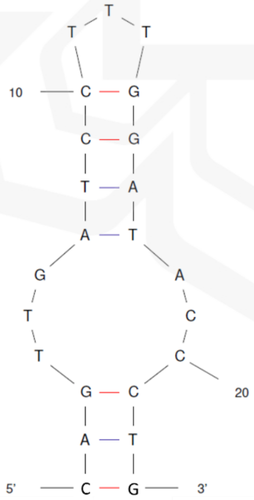
1BJH (0.04)



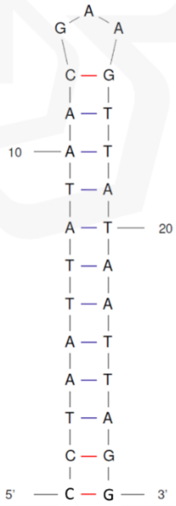
2M8Y (-6.33)



1ECU (-8.72)



2L5K (-1.57)



1JVE (-9.56)

Figure 4.7 Secondary structures of six selected DNA aptamers, modelled by Mfold server. The free energy (kcal/mol) of each conformation is shown in brackets.

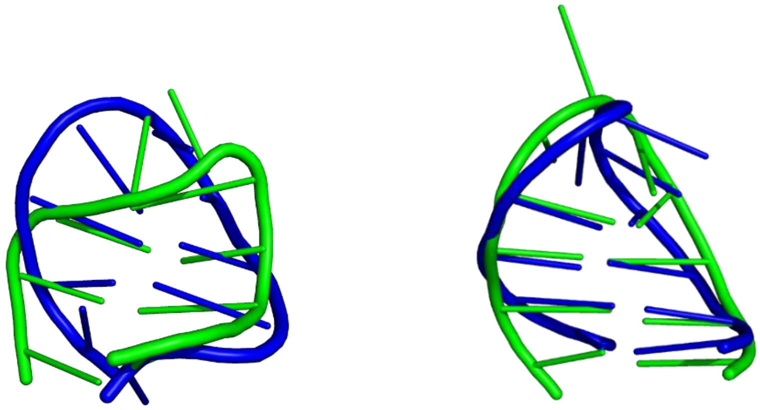
As Mfold predicted multiple secondary conformations for each DNA, only the conformation with the lowest free energy was taken as it is the most stable conformation. The folding of the six aptamers resulted in minimum free energy ranged from 0.71 to -9.56 kcal/mol. The study demonstrated a positive relationship between the length of the stem, measured by the number of H-bonds formed, and the aptamer's free energy. A longer stem region corresponded to a lower free energy and increased structural stability. It was proven in the 1JVE aptamer, which consisted of 27 nucleotides that formed 12 base pairs and had a minimum free energy of -9.56 kcal/mol. The 1JVE and 1ECU aptamers exhibit high stability in this study with the lowest free energies, while 1ZHU has the lowest stability among the DNA structures, with a positive free energy value of 0.71 kcal/mol. Despite the formation of many Watson-Crick base pairs in the 2D conformation of 2L5K, the existence of an internal bulge results in a less stable conformation, with a minimum free energy of -1.57 kcal/mol. The nucleotides comprising the internal bulge are free-moving and lack rigidity, resulting in reduced stability. The 2D and 3D conformations of the six aptamers were also designed by Sabri (2021) and Table 4.4 displays the differences and similarities between these two studies.

Table 4.4 Comparisons of the free energies and Watson-Crick base pairs between the 2D conformations of six aptamers designed by this study and the study by Sabri (2021)

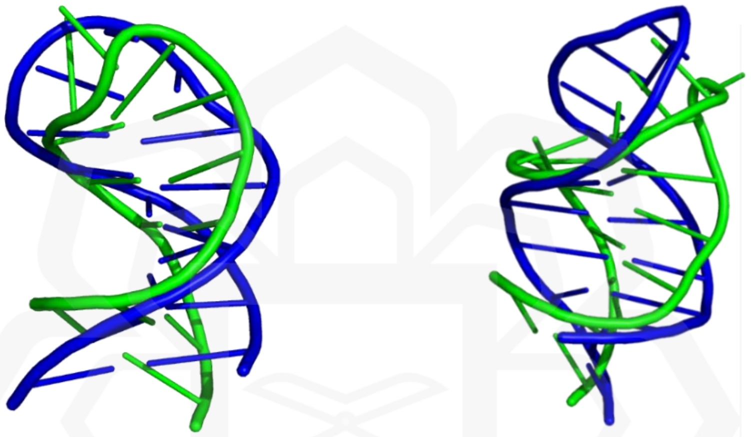
PDB ID	Length	Free energy (kcal/mol)		Watson-Crick base pair G-C		Watson-Crick base pair A-T	
		This study	Sabri (2021)	This study	Sabri (2021)	This study	Sabri (2021)
1ZHU	10	0.71	1.23	-	-	2	2
1BJH	11	0.04	0.80	2	1	2	2
2M8Y	15	-6.33	-5.05	4	4	2	2
1ECU	19	-8.72	-7.16	5	5	2	2
2L5K	23	-1.57	-0.53	4	4	3	3
1JVE	27	-9.56	-6.70	3	3	9	9

Focusing on the minimum free energies, the 2D conformations predicted in this current study have shown lower values for all six aptamers, where the most significant can be found for 1JVE with -9.56 kcal/mol, while it was only -6.70 for the previous study. This is possibly due to the time when the studies were conducted, as the algorithm and efficiency of the Mfold programme might have been upgraded recently by the developer. In the context of Watson-Crick base pairs, they were highly identical. Only 1BJH showed a difference in the number of G-C pairs. The pattern of free energies in the previous study is similar to this current study, as longer stem led to more negative free energy. The 2L5K designed by Sabri (2021) formed an internal bulge too, making the conformation less stable with free energy of -0.53 kcal/mol.

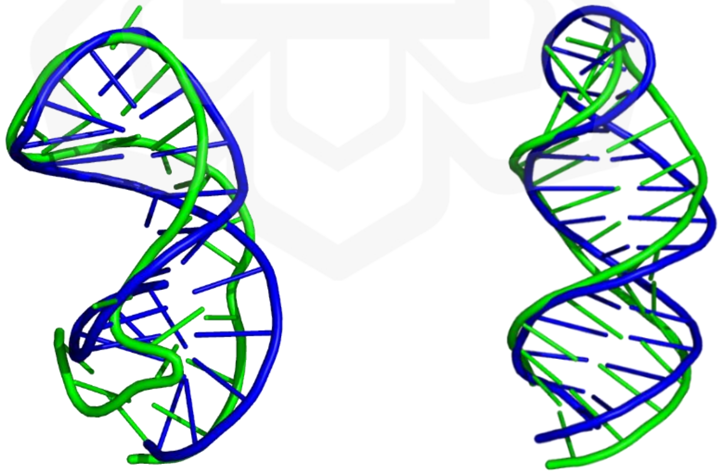
This previous study utilised RNAComposer to generate 3D structure of the DNA aptamers too, but all the 3D models underwent MD simulations for 20 ns to optimise the models, where energy minimisation steps were applied. However, this current study did not apply the same procedure, as the Mfold applied energy minimisation step during the modelling process (Markham & Zuker, 2008; Zuker, 2003). After manual alterations of the atoms, converting the RNA to DNA aptamer, the designed 3D structure was superimposed with the original crystallised structure retrieved from PDB database. The main purpose is to evaluate the structural and RMSD values differences of the conformations. Figure 4.8 illustrates the superimposed structures and the RMSD values.



1ZHU (RMSD = 4.445) 1BJH (RMSD = 2.798)



2M8Y (RMSD = 3.330) 1ECU (RMSD = 6.931)



2L5K (RMSD = 6.232) 1JVE (RMSD = 5.190)

Figure 4.8 Superimposed DNA structures of PDB crystallised structure (blue) and modelled by RNAComposer (green). The RMSD values (in Å) are shown in brackets.

A published work in 2023 retrieved the sequences of 20 crystallised structures of single-stranded DNA (ssDNA) aptamers from PDB and computationally designed the 3D conformations, which was then followed by superimposing the designed models to the crystallised structures (Mu et al., 2023). It was reported that the mean RMSD for the superimposed aptamers ranged from 2.0 Å to 6.0 Å with overall mean RMSD of 3.4 Å. In an earlier study, 24 ssDNA hairpin structures were designed based on their sequences deposited in PDB. The designed models were also superimposed to the crystallised structures and the calculated RMSD values varied from 1.87 Å to 8.59 Å (Jeddi & Saiz, 2017). The lowest RMSD came from 1PQT, consists of 7 nucleotides, while the highest RMSD belongs to 1NGU with 27 nucleotides. From this previous study, it can be observed that the length and the structural motifs of the ssDNA affected the RMSD values. For an instance, the RMSD values of 1EN1 (18 nucleotides) and 1NGO (27 nucleotides) are 6.89 Å and 6.42 Å, respectively, while the 1KR8 (7 nucleotides) has low RMSD value, 1.99 Å. The designed short ssDNAs are highly similar to their crystallised structures, as lower RMSD corresponds to high similarity of the two conformations. As for 1EN1, even though the length is shorter than 1NGO, due to the presence of an open strand at the 3’ end terminal, the region rotated freely and was not modelled precisely by the server, thus, generated high RMSD value.

Referring to 1ZHU in this study, the structure consists of open strands at both 3’ and 5’ end terminals. As a result, it has moderate RMSD value of 4.445 even though it is a short aptamer, made up of only 10 nucleotides. The 1BJH aptamer generated the lowest RMSD value (2.798 Å), indicating good structural similarity with only slight differences. According to the RMSD calculations in PyMOL programme, out of 330 atoms in 1BJH, 42 atoms were rejected during the alignment process. For 1JVE, only 3.1 % of the atoms (25 out of 796) were rejected during alignment, showing high similarity between the designed model and the crystallised structure.

Pursuant to the procedures applied and the output generated for this validation section, it can be deduced that the suggested pipeline to model the 2D and 3D DNA structures using Mfold and RNAComposer servers with additional steps of atom alterations are acceptable. By comparing the designed models with the actual ssDNA aptamers from PDB, it is proven that this suggested pipeline generates accurate conformations.

4.3.3 HE4-DNA Aptamer Complexes Binding Analysis

Four promising DNA aptamer sequences were chosen through *in vitro* CE-SELEX technique against HE4 biomarker from a study mentioned earlier (Eaton et al., 2015). The exact four aptamer sequences were retrieved, and their binding affinities towards the modelled HE4 tertiary structure (Chapter 2) were analysed computationally. This involved the utilisation of AutoDockTools (ADT), in preparing the aptamer (ligand) and HE4 protein (receptor) input files (Morris et al., 2008), and AutoDock Vina to conduct the docking between the two molecules (Trott & Olson, 2010). The tertiary conformations of the DNA aptamers were shown earlier in Figure 4.5 where each hairpin can be seen. During the input preparation in ADT, the backbone and the stem region were set as rigid with non-rotatable bonds between atoms, conserving the helical structure of the stem, while the bonds in the nucleobases found at the loop region were set to be rotatable. This ensures better interactions between the aptamer and HE4 protein as the bonds between the atoms are adaptable and get closer to the atoms of the receptor, thus, increasing the binding affinity. Additionally, the HE4 protein was also set as rigid, as this approach simplifies the complexity of the computational simulations and enables rapid screening of potential binding poses (Meng et al., 2011).

AutoDock 4 (AD4) and AutoDock Vina (Vina) are prominent molecular docking tools widely used in computational biology and drug discovery (Lee et al., 2023; Nguyen et al., 2020). AD4 employs a Lamarckian Genetic Algorithm in conjunction with local search techniques to navigate the conformational space of ligands, providing important predictions of binding modes. The scoring mechanism involves empirical force-field parameters and affinity terms, which allows it to effectively handle various docking conditions (Bitencourt-Ferreira et al., 2019). However, Vina is an improved version that offers advances in both speed and accuracy when compared to AD4. The Vina methodology utilises an iterative stochastic global optimisation method, along with local optimisation, to enable more rapid ligand docking computations. It is distinguished by its utilisation of a sophisticated scoring mechanism that encompasses factors like steric interactions, hydrogen bonding, and other energy components. This enhances its precision in predicting ligand binding modes, rendering it especially suitable for high-throughput virtual screening programmes (Vieira & Sousa, 2019). The tool's focus on efficiency and user-friendliness is apparent through its graphical user interface (GUI), which allows researchers with different levels of competence to easily utilise it.

A published work in 2019 utilised AD4 and Vina to evaluate the docking performance of a huge dataset consisted of 102 protein receptors, 22 432 active compounds, and over 1.3 million decoy molecules (Vieira & Sousa, 2019). Overall, the results indicated that Vina and AD4 performed similarly in distinguishing between active compounds and decoys. Nevertheless, the outcomes exhibited substantial variations depending on the specific target. The AD4 exhibited superior discriminatory ability between ligands and decoys in hydrophobic, weakly polar, and poorly charged pockets, whereas Vina shown higher performance in polar and charged binding pockets. These results are supported by the findings of a recent study that operated the AD4 and Vina to analyse the binding poses and affinities of 800 protein-ligand complexes where the PDB structures of the complexes and their experimental binding affinities are known (Nguyen et al., 2020). Both docking tools were also assessed for their ability to accurately replicate experimental binding poses and it was discovered that Vina outperformed AD4 in predicting the native poses, with a success rate of 81% compared to 77% for AD4. These two studies have shown the accuracy of using Vina for molecular docking involving polar and charged binding pockets, thus, it is a suitable docking tool for this current study. For the output, Vina docking listed 10 lowest energy conformations of the docked aptamer, but only the lowest (most negative) binding energy for each system was extracted and analysed further.

Table 4.5 The binding energy and the number of hydrogen bonds formed between HE4 protein and aptamers.

Protein target	Aptamer	Binding energy (kcal/mol)	No. of H-bonds (total)	No. of H-bonds at hairpin region	
				Stem	Loop
HE4	A1	-3.7	15	1	2
	A2	-4.6	19	2	1
	A3	-6.5	26	3	0
	A4	-6.0	24	0	5

Table 4.5 shows that HE4-A3 has the lowest binding energy whereas HE4-A1 has the highest. However, the lowest binding energy does not always indicate the best binding pose, as the binding site and interactions between the two molecules should also be considered. Based on the results obtained, as the number of H-bonds increases, the binding energy decreases, indicating stronger interactions. The H-bonds produced in the hairpin region were highlighted, other than the overall number of H-bonds formed between the aptamer and HE4. The significance lies in the flexibility of the atoms of the loop residues, as the bonds connected them are rotatable, thus, they have greater ability to bind to the atoms of the receptor (HE4) by bond rotations. Previous studies have shown the variations of the loop sequence, including the alterations within the loop regions improved the stability and enhanced the binding properties (Macdonald et al., 2017, Lakshmipriya et al., 2018). Comparing the two lowest energy complexes, HE4-A3 (-6.5 kcal/mol) and HE4-A4 (-6.0 kcal/mol), it is clear that A4 formed the highest number of H-bonds at the hairpin region and all of them are located at the loop region of the aptamer, unlike A3 which has H-bonds only at the stem of the hairpin, excluding the other H-bonds at the open strands of the 3’ or 5’ end terminals. Figure 4.9 shows the binding poses of the HE4-aptamer complexes with the lowest binding energy, generated by Vina. These poses were viewed in PyMOL molecular viewer, displayed in mesh (for HE4 protein) and surface (for the DNA aptamers) view modes.

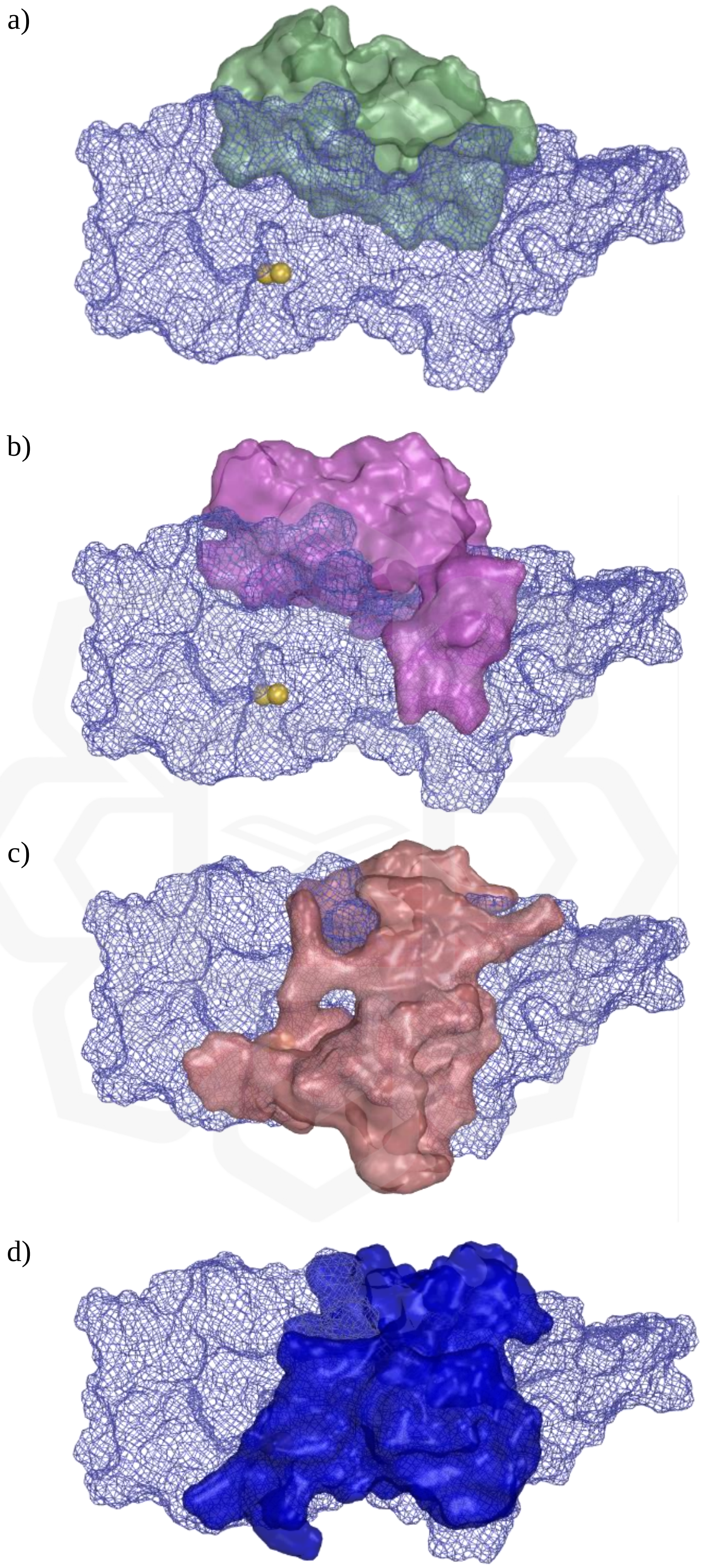


Figure 4.9 DNA aptamer-HE4 complexes. (a) HE4-A1, (b) HE4-A2, (c) HE4-A3, and (d) HE4-A4. The gold dots represent the cavity determined by CAVER 3.0 tool.

Based on these binding poses, A1 and A2 aptamers did not bind to any residue of the potential binding site (gold dots) or the small tunnel leading to the binding site that was predicted by CAVER tool in the previous chapter (Figure 3.11). However, looking at the HE4-A3 and HE4-A4 binding poses, the aptamers might form interactions with the HE4 protein at the tunnel and binding site. The pdb file of the HE4-aptamer complex was downloaded and the HE4-aptamer interactions were identified using PLIP server, developed by Michael Schroeder Group at the Biotechnology Center TU Dresden (Adasme et al., 2021). This server has successfully characterised docking experiments and evaluate ligand-protein complexes, including RNA- and DNA-protein complex. The output file of PLIP assessment was retrieved and viewed in PyMOL.

The conformational study of docked HE4-A3 and HE4-A4 structures revealed that both A3 and A4 aptamers bind specifically to the residues available at or near the potential binding site of HE4 protein *via* H-bonds and hydrophobic interactions. The strength of hydrogen bonds arises from the presence of a partial positive charge on a hydrogen atom and a partial negative charge on the electronegative atom, participating in the bond. In contrast, hydrophobic interactions mostly arise from the tendency of nonpolar molecules to reduce their interaction with water molecules. Even though hydrophobic interactions are significant in stabilising protein structures and aiding in ligand binding, they are often weaker compared to hydrogen bonds (Ferenczy & Kellermayer, 2022; Y. Wang et al., 2020). There were 9 and 10 hydrophobic interactions formed between the HE4 protein and the aptamer of HE4-A3 and HE4-A4 complexes, respectively (Table 4.6). Some of them were contributed by PRO37, LEU72, and GLN98 with distance ranged between 3.57 Å to 3.87 Å.

Table 4.6 The hydrophobic interactions between HE4 protein and aptamer in HE4-A3 and HE4-A4 complexes.

Complex	Residues involved	Distance (Å)
HE4-A3	GLU31-T15	3.76
	PRO37-T8	3.86
	GLU38-T12	3.62
	LEU39-G7	3.54
	LEU72-T8	3.76
	ASP75-T11	3.95
	GLU77-C23	3.88
	GLN98-T4	3.85
	VAL117-C10	3.82
HE4-A4	PHE22-T2	3.58
	VAL35-C6	3.61
	PRO37-A19	3.57
	GLU48-T21	3.83
	PHE69-T21	3.51
	LEU72-T23 (1)	3.87
	LEU72-T23 (2)	3.81
	GLN98-G18	3.41
	GLN98-C25	3.82
	VAL99-G18	3.51

Meanwhile, Figures 4.10 and 4.11 display the details of conformational analysis of HE4-A3 and HE4-A4 complexes, as they both exhibited good interactions with low binding energies and accurate binding region. The binding regions were investigated further, focusing on the residues of the aptamer hairpin regions and the predicted binding site. In Table 4.5, HE4-A3 complex formed higher number of H-bonds (26 H-bonds) compared to HE4-A4 (24 H-bonds). However, HE4-A4 has five H-bonds formed at the aptamer hairpin region, while HE4-A3 has only three (Figure 4.10c and 4.11c). While higher number of H-bonds indicate stronger binding, it is important to locate the region where the interactions occurred, specifically the loop region. Only HE4-A4 complex formed H-bonds at the loop region, while all hairpin H-bonds of HE4-A3 are only available at the stem region. As discussed earlier, the flexibility of the loop atoms allows for a better ability to bind to HE4 rigid atoms. Figures 4.10d and 4.11d highlighted the hairpin residues involved in forming H-bonds with HE4 protein. There were three H-bonds formed by G22 and C23 of A3 which are located at the rigid stem, whereas, the G7, C8 and A10 residues of A4 hairpin loop formed five H-bonds. Both A3 and A4 hairpins bound to the GLU77 and GLY78 protein residues with H-bond lengths between 2.1 Å to 3.5 Å. Additionally, all residues involved in this HE4-hairpin H-bonds are listed in Table 4.7, alongside the residues formed the H-bonds at the HE4 predicted tunnel and active binding site. The H-bonds available at the tunnel-binding site area and their bond length are displayed in Figure 4.10b and 4.11b. A4 aptamer formed stronger interaction around this region with five H-bonds, compared to A3 with only two. The HE4 protein residues contributed to the binding in both complexes were identical; SER63 and GLU31, with HE4-A4 has a slight addition of SER51 residue.

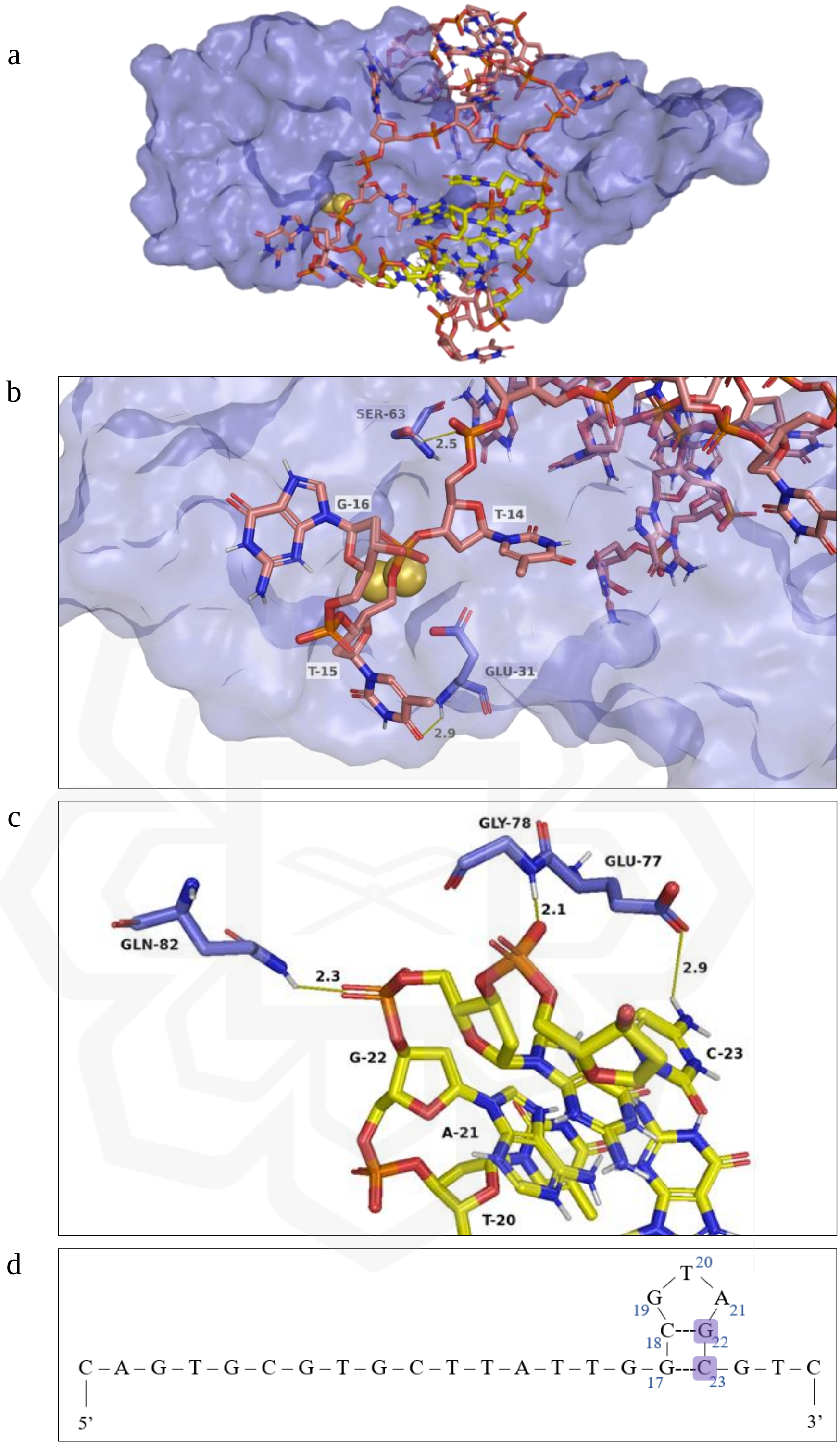


Figure 4.10 The binding pose of HE4-A3 complex. (a) The overall binding (yellow stick represents the hairpin), (b) The residues involved in H-bonds at the CAVER predicted tunnel and binding site (gold dots), (c) The hairpin residues that made up the H-bonds with HE4 residues, and (d) The 2D structure, illustrating the hairpin residues contribute to the H-bonds formation. Yellow lines represent the H-bonds, together with the bond length (in Å).

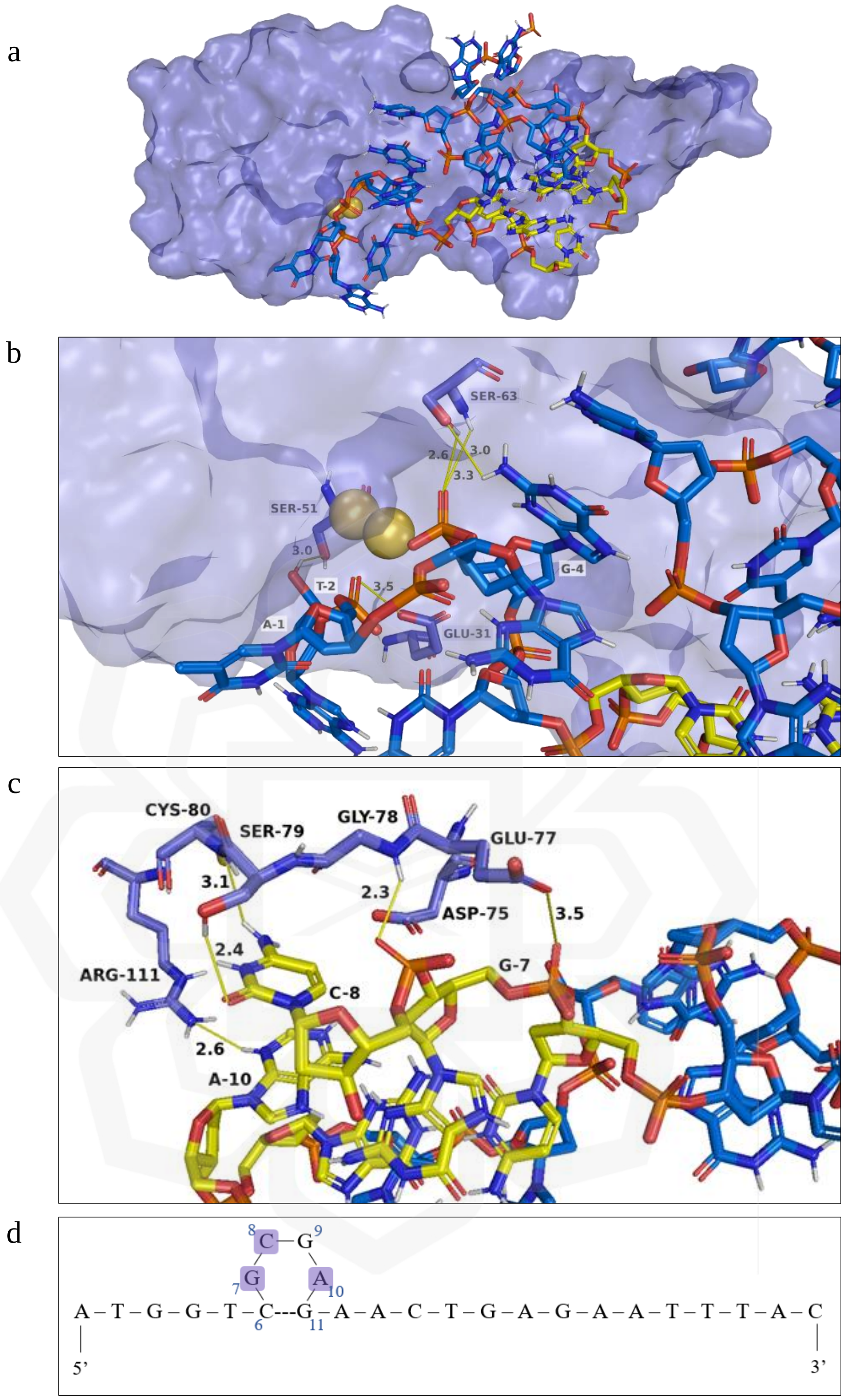


Figure 4.11 The binding pose of HE4-A4 complex. (a) The overall binding (yellow stick represents the hairpin), (b) The residues involved in H-bonds at the CAVER predicted tunnel and binding site (gold dots), (c) The hairpin residues that made up the H-bonds with HE4 residues, and (d) The 2D structure, illustrating the hairpin residues contribute to the H-bonds formation. Yellow lines represent the H-bonds, together with the bond length (in Å).

Table 4.7 The binding energy and the number of hydrogen bonds formed between HE4 protein and aptamers.

Complex	Residues involved in hydrogen bonds		
	Tunnel-binding site region	Hairpin region	
		Stem	Loop
HE4-A3	SER63-T14 GLU31-T15	GLN82-G22 GLY78-C23 GLU77-C23	none
HE4-A4	GLU31-A1 SER51-T2 SER63-G4 (1) SER63-G4 (2) SER63-G4 (3)	none	GLU77-G7 GLY78-C8 SER79-C8 CYS80-C8 ARG111-A10

Both A3 and A4 aptamers had good interactions with the HE4 protein model. However, in agreement with the data and results presented, A4 aptamer had better binding after thorough evaluations including the binding energy, binding region, and quantity of interactions (H-bonds and hydrophobic).

4.4 CONCLUSION

This chapter provides an in-depth description of the aptamer modelling of the secondary and tertiary conformations, followed by the validation of the choice of method. Consequential to this, the molecular docking of HE4 protein with four aptamer candidates were conducted using AutoDock Vina, with the objective of identifying the most potential DNA aptamer to bind to the ovarian cancer biomarker, HE4. It was discovered that the hairpin (stem-loop) of aptamer is an important structural motif, as it was presented in every aptamer designed by Mfold and RNAComposer servers. This was also proven by the experimentally-obtained aptamer structures retrieved from PDB database, which were then utilised to assist the validation of the aptamer design method. The designed 3D models and the PDB structures were highly similar after the conformations were aligned and RMSD values were calculated by PyMOL, ranging from 2.798 Å to 6.931 Å. The aptamer length and type of structural motifs influenced the RMSD value, generally, the longer the aptamer, the higher the RMSD value. The computational technique to designing the 2D and 3D structures of the DNA aptamers was deemed reliable although manual inspections and alterations of atoms are still required.

Following the validation of the aptamer design method, four aptamer candidates chosen from a previous CE-SELEX study against HE4 biomarker were modelled according to the suggested pipeline. The 3D conformations were prepared in ADT as ligand, where the bonds at the loop nucleobases were set as flexible, whereas the 3D HE4 structure modelled in previous chapter was set as the receptor with rigid bonds. The Vina docking resulted in binding energies ranged from -3.7 kcal/mol to -6.5 kcal/mol for the HE4-aptamer complexes, having HE4-A3 as the lowest (-6.5 kcal/mol), followed by HE4-A4 (-6.0 kcal/mol). After thorough analysis on the interactions between HE4 and both aptamers, A3 and A4, it is deduced that the A4 is the most suitable aptamer to be further utilised in this study where the hairpin will be truncated and altered, improving the hairpin structures and enhancing the interactions with HE4 protein.

CHAPTER FIVE

THREE-DIMENSIONAL DNA APTAMER HAIRPIN DESIGN BY MUTATION, MOLECULAR DOCKING, AND MOLECULAR DYNAMICS SIMULATION

5.1 INTRODUCTION

The hairpin of aptamer with loop and stem motifs is essential for the design and functionality of aptamers (Achar & Sætrom, 2015). This hairpin structure typically consists of a binding region that complements the target molecule, enabling it to selectively attach to the desired target within a complex mixture of molecules (Hoinka et al., 2012). Moreover, hairpins have the ability to undergo conformational alterations when they attach to their target molecules as they are single-stranded, making them structurally flexible, specifically at the loop region, enabling the hairpin to adopt several conformations and selectively interact with the target molecule (Dubey et al., 2005; Z. Zhang & Liu, 2018). The recognition and the binding process rely on this crucial dynamic behaviour. The presence of the hairpin structure enhances the stability and facilitates the right folding of the aptamer. In a previous study, the significance of hairpin loops in the functionality of aptamers was discussed where 13 aptamer candidates were incubated with *Vibrio fisheri* cells under the same conditions (Shin et al., 2018). The secondary structures of the highest efficiency aptamers, VFCA-02 and VFCA-03 were predicted by Mfold server and it was discovered that both had identical hairpin loops, unlike the other 11 aptamers, indicating the importance of hairpin motif in the functionality of aptamer. Another study in 2018 reported great signalling properties with higher signal intensity when utilising an aptamer consists of hairpin motif, leading to greater binding of adenosine on graphene oxide (Zhang & Liu, 2018).

Nucleotide, the building block of aptamers, consists of a nitrogenous base, a phosphate group, and a sugar molecule, and may undergo alteration which is known as mutation, involving a sequence change of the nitrogenous bases. The types of nucleotide mutation include substitution, insertion, and deletion which may occur naturally or synthetically (Gunter, 2024). For example, a substitution involves the replacement of one nucleotide to another while insertion occurs when one or more nucleotides are incorporated into a sequence. Mutation of nucleotides can happen spontaneously, induced by errors in DNA replication, radiation, or exposure to certain chemicals (Cooper et al., 2011; Ma et al., 2013). These mutations may affect its binding ability towards the target molecule and trigger conformational changes which leads to changing the overall structural motifs. The advantages of nucleotide mutation in aptamers include the enhancement of binding affinity and increasing structural stability (Chen et al., 2019). In this study, the DNA aptamer candidate identified in the Chapter 4, which is the A4 aptamer, was truncated at the stem region for further analysis against the target molecule, HE4. Mutations to the nucleotides of the loop were applied to enhance the binding affinity with HE4 protein. A shorter sequence after truncation led to faster molecular docking and molecular dynamics simulation, allowing hundreds of different sequences to be studied, to determine the most potential hairpin aptamer to screen this ovarian cancer biomarker.

It has been proven that truncated aptamers from a long aptamer resulted in enhancement of biosensor performance (Azri et al., 2021). Long aptamers have potential drawbacks that need to be considered when conducting *in silico* and *in vitro* studies which include difficulties in synthesis, decreased binding strength, increased complexity, hindrance due to spatial arrangement, reduced sensitivity, inaccuracy in determining their secondary and tertiary structures, and higher costs of production (Chen et al., 2021; Le et al., 2014; Mu et al., 2023; Ni et al., 2021). The intricate secondary and tertiary conformations of long aptamers might complicate the process of designing and optimising them, hence posing a challenge in accurately predicting their folding patterns (Jeddi & Saiz, 2017; Sabri, 2021). It was proven in the Chapter 4 (Subsection 4.3.2) that shorter aptamers generated lower free energies, indicating higher stabilities. Whereas, in long aptamers, more atoms are present which increases the occurrence of steric hindrance, restricting the ability to reach binding sites, especially when engaging with large molecules (Mahshid et al., 2015).

Nucleic acids and proteins are important biological macromolecules in cells where their interactions play a vital role in several biological processes, including signal transmission, cell control, protein synthesis, DNA replication and repair, and RNA transcription (Minchin & Lodge, 2019; Tagami, 2023). Hence, the identification of their 3D organisation of protein-DNA complex is crucial for comprehending the atomic-level mechanisms underlying biological processes (Emamjomeh et al., 2019) and this can be achieved by molecular docking. As stated in the previous chapter, various molecular docking programmes have been established including AutoDock4 (Forli et al., 2016), AutoDock Vina (Trott & Olson, 2010), GOLD (Verdonk et al., 2003), and HDOCK (Yan & Huang, 2020). Protein-DNA docking, utilising AutoDock Vina, is an *in silico* approach used to predict the spatial configuration and binding modes of proteins and DNA molecules. Vina is widely recognised for its exceptional efficiency, rapidity, reproducibility, and intuitive user interface (Trott & Olson, 2010; Vieira & Sousa, 2019). Due to the advantages, Vina was used in numerous aptamer studies against different targets such as wild-type EGFR, EGFRvIII, and zearalenone (Zavyalova et al., 2020; Y. Zhang et al., 2018). The affinity of the protein-aptamer interaction is dependent on the structural characteristics of the target molecules and aptamers. Aptamers are able to form many structural motifs, including hairpins, bulges, pseudoknots, and G-quadruplexes (Duan et al., 2020; Sullivan et al., 2019; H. Yu et al., 2021), and they are able to attach to targets by various forces, including hydrogen bonding, electrostatic interactions, hydrophobic interaction, van der Waals forces, or a combination of these forces (Tan et al., 2016). Following the completion of the docking simulation, the binding modes and interaction energies are analysed and MD simulations were introduced to enhance the protein-ligand computational studies.

Understanding the dynamic behaviour of protein-ligand complexes over time possibly improves their affinity experimentally (Salmaso & Moro, 2018). MD simulations with GROMACS for protein-DNA complexes entail a meticulous analysis of the complex's dynamic behaviour and interactions across a certain period. GROMACS, a commonly utilised simulation programme, uses force-field-based molecular mechanics to mimic the motions of atoms inside the system which in this study, is the Chemistry at Harvard Macromolecular Mechanics 27 (CHARMM27) force field. This force field includes parameters for amino acids and nucleic acids with parameters specifically designed for nucleotide building blocks, making it a suitable force field to be used in this current study (Huang et al., 2017; Vanommeslaeghe et al., 2010). It was utilised in earlier DNA studies, including ssDNA-hepatitis B surface antigen study (Sabri et al., 2019) and research related to B-form (BI) and A-form (BII) in DNA for dynamics representation improvement (Hart et al., 2012). CHARMM27 was deduced as the best force field compared to Amber99sb, Amber14sb, and Gromos43a1 force fields in a study performing MD simulations using CutA1 protein obtained from *Pyrococcus horikoshii* (Matsuura et al., 2019). Post MD simulations analysis such as RMSD, Rg, and Root Mean Square Fluctuation (RMSF), offer valuable information on the structural stability, compactness, and the fluctuations it experiences during the simulation (Kamaraj & Bogaerts, 2015; Rampogu et al., 2022).

In this chapter, various computational methods were integrated in the designing of DNA aptamer hairpin against HE4 protein. This includes the design of hairpin 3D structures using Mfold and RNAComposer, the molecular docking by Vina, and MD simulations of complexes using GROMACS (Figure 5.1).

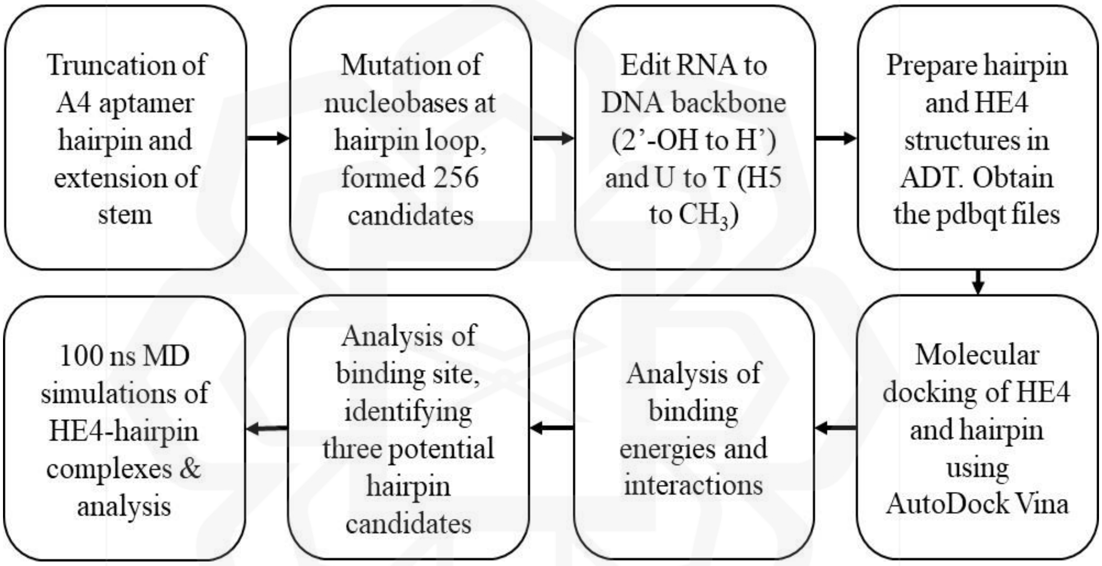


Figure 5.1 Pipeline for the mutation of aptamer hairpin, molecular docking of hairpin-HE4 protein, and MD simulation of HE4-hairpin complexes.

5.2 METHODOLOGY

5.2.1 3D DNA Aptamer Hairpins Design *via* Nucleotide Mutation

The promising HE4 aptamer candidate identified in Chapter 4 was further utilised, to enhance the interactions of the aptamer hairpin with HE4 protein. The workflow consisted of truncation of the A4 aptamer at the hairpin region, the extension of the hairpin stem, the nucleotide mutation at the hairpin loop, and finally the modelling of hairpin tertiary structures which involved the utilisation of Mfold (Zuker, 2003) and RNAComposer (Sarzynska et al, 2023) servers together with the atom's alteration in converting the RNA to DNA hairpins. Firstly, the A4 aptamer was truncated at the hairpin region, -CGCAAG- and the length of the stem was extended at the 5’ and 3’ ends with one guanine and one cytosine, respectively. Truncations on the hairpin regions of aptamers have significantly improved the binding affinity towards the targets, highlighting the importance of the region (Kaur & Yung, 2012). The overall length of the hairpin was 8-mer, with a sequence of 5’–GCGCAAGC-3’, where the loop region (-GCAA-) was mutated by substituting the nucleotides. This resulted in the formation of 256 8-mer oligonucleotides (refer to supplementary data in Appendix A) as each position of the nucleotide at the loop has four possible combinations of thymine, adenine, guanine, and cytosine bases. The stems of all 256 aptamer hairpins were kept constant, 5’-GC-(loop)-GC-3’.

Table 5.1 Sequences of A4 aptamer, truncated and extended hairpins

Oligonucleotide	Sequence (5’ to 3’)	Length (base)
Full A4 aptamer	ATGGTCGCAAGAACTGAGAATTTAC	25
A4 hairpin only	CGCAAG	6
Extended A4 hairpin	GCGCAAGC	8

*Highlighted sequences represent the hairpin loop region

Next, the Mfold server (http://www.unafold.org/mfold/applications/dna-folding-form.php) was employed to predict the secondary structure of each 8-mer oligonucleotide. Default parameters were utilised, including a temperature of 37°C and ionic conditions of 1.0 M Na^+ and 0.0 M Mg^{2+} ions. Among the predicted structures, the one exhibiting the most favourable thermodynamic profile, determined by the lowest folding energy value (in kcal/mol), was chosen. This selected structure, presented in Vienna format, was obtained for further analysis of its tertiary structure using the RNAComposer web server (https://rnacomposer.cs.put.poznan.pl/), where the server generated an equivalent RNA structure. The resultant pdb file representing this RNA structure was downloaded and accessed through the PyMOL molecular viewer programme. Subsequently, a manual conversion was performed to transform the RNA hairpin structure into a DNA aptamer hairpin, by the conversion of the ribose sugar to deoxyribose by replacing the hydroxyl group (-OH) of ribose with –H on the backbone. Additionally, all uracil residues were replaced by thymine, accomplished by substituting the H5 atom with a methyl group (-CH_3). All these modifications were executed by PyMOL's builder tool. These steps were applied to all 256 8-mer nucleotides, forming 256 DNA aptamer hairpins.

5.2.2 HE4-DNA Aptamer Hairpin Molecular Docking

The 256 hairpins underwent individual docking with HE4 protein using AutoDock Vina. Similar to subsection 4.2.3, this section analyses the binding interactions between HE4 protein and the 256 computationally-designed hairpins obtained from subsection 5.2.1 through molecular docking. This approach was adopted from a previous study by Nguyen et al (2020). The parameters and configures were set according to the docking in the previous chapter 4. The preparation of the input, receptor (HE4) and ligand (DNA aptamer hairpin) was conducted using AutoDock Tools (ADT). After the removal of water molecules, the polar hydrogens were assigned, followed by Kollman and Gasteiger calculations by the programme. The grid box dimensions were set to 82 × 44 × 84 for x, y, and z points, enclosing the protein structure completely to run a blind docking. Then, the box was centred at coordinates 64.672, 65.933, and 65.56 for x, y, and z respectively. The molecular docking was initiated. Subsequently, the docking output provided 256 HE4-hairpin complexes where they were sorted according to their binding energies, from the most negative to the least. The 10 complexes with the most negative binding energies were extracted and aligned with the HE4 protein structure using PyMOL for further assessment. The presence of H-bonds was recorded, followed by the identification of three complexes with a high number of H-bonds exhibited at the hairpin loop region. The hairpins formed most H-bonds with the residues of HE4 tunnel and cavity identified by the CAVER 3.0 tool in the previous chapter were prioritised in the selection of the three most potential hairpins.

5.2.3 Molecular Dynamics Simulation of HE4-Hairpin Complexes

Following molecular docking analysis, three potential complexes were chosen out of the 256 HE4-hairpin complexes and MD simulations were carried out. This work assesses the stability and binding interactions over time at atomic level, providing in-depth understanding of the dynamic behaviour of the complexes. The MD simulations of these three complexes were performed using GROMACS software, version 5.1.1 (Royal Institute of Technology and Uppsala University, Stockholm, Sweden). The software's primary focus lies in biochemical molecules such as proteins, lipids, and nucleic acids, characterised by complex bonded interactions. The simulations of the complexes were conducted using CHARMM27 force field for 100 ns. The complex structure was solvated within a cubic box, maintaining a distance of 1 nm between the structure and each side of the box. To neutralise the system's charge, sodium, and chloride ions (Na^+ and Cl^-) were introduced. The energy minimisation of the solvated systems was then conducted, aiming to diminish unfavourable van der Waals interactions which were completed within 1622 steps until the system reached convergence. As the system consisted of protein and DNA, both molecules were merged as a new index, named 'DNA_protein'. Position-restrained equilibration was performed by applying a force constant of 1000 kJ mol^{-1} nm^{-1}. Following equilibration, unrestricted MD simulation was executed to the system for 100 ns and repeated twice for all systems, resulting in triplicates.

The suite within the GROMACS software (Abraham et al., 2023) was employed for MD analysis, emphasising the backbone RMSD, the Rg, root-mean-square fluctuations (RMSF), the minimum distance between the protein and DNA aptamer hairpin in the complex, and number of H-bonds. Additionally, the H-bond occupancy analysis was done, exhibiting the frequencies of H-bonds within the complexes using Visual Molecular Dynamics (VMD) (Ranade & Ramalingam, 2020; Zikri et al., 2021). This measures the amino acids and nucleic acids interactions, specifically the H-bonds between HE4 protein and DNA aptamer hairpin. Finally, the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method was executed, estimating the free energy of binding in molecular complexes, especially protein-ligand interactions. This approach determines the binding free energy by assessing the gas phase energy, alongside the electrostatic solvation energy and non-electrostatic contribution to the solvation energy (Ylilauri & Pentikäinen, 2013).

5.3 RESULTS AND DISCUSSION

5.3.1 DNA Aptamer Hairpin Tertiary Structure

Aptamers are able to interact with their target proteins and form complexes, which allows them to carry out a wide range of biological tasks. Given that the functionalities of biological molecules are determined by their tertiary structures, it is crucial to have accurate secondary and tertiary modelling for aptamers (Chen et al., 2021). The modelling and validation of aptamer tertiary modelling have been discussed in Chapter 4, followed by molecular docking using Vina. Based on the docking results in Section 4.3.3, A4 aptamer was selected as the most suitable aptamer against the protein target, HE4. As this aptamer consists of 25 nucleotides, the efficiency and feasibility of computational studies that involve docking and MD simulations may be affected, due to the larger conformational space to be explored in long aptamer. Various studies have obtained better accuracies in 2D and 3D modelling, and more specific binding to targets with shorter RNA or DNA aptamer length (Azri et al., 2021; Mu et al., 2023; Sabri et al., 2019). Truncation of long aptamer structure shortens the nucleotide length, making it simpler to work with, without compromising the main component of an aptamer which is the hairpin region. The study by Azri et al. (2021) truncated an 80-mer zearalenone (ZEA) aptamer, leading to 15 shorter sequences, ranging from 12-mer to 45-mer which consisted of one to two hairpins. It was confirmed that the binding affinities of the truncated aptamers against ZEA protein increased significantly, compared to the initial 80-mer ZEA aptamer, mostly with binding energies lower than -6.5 kcal/mol. A study designing DNA aptamer hairpins against hepatitis B surface antigen (HBsAg) truncated three long aptamers (77-mer to 85-mer) into five short hairpins (8-mer and 9-mer) which resulted in good Vina scores (binding energies) between -22.175 kJ/mol (-5.3 kcal/mol) and -31.798 kJ/mol (-7.6 kcal/mol) (Sabri et al., 2019).

In this study, the 25-mer A4 aptamer was truncated at the hairpin stem, C6 and G11, shortening the length to 6-mer with a sequence of 5’-CGCAAG-3’ (Figure 5.2a). As there was only one base pair (-CG-) forming the stem, the extension of the stem was done by adding another Watson-Crick base pair (-GC-) at the 5’ and 3’ ends as shown in Figure 5.2b. This increased the interactions between the base-pairings in the stem, making the conformation more stable as more guanine-cytosine pairs are present in this new 8-mer hairpin, making a greater number of H-bonds (Watson & Crick, 1953). The free energy of A4 aptamer predicted by Mfold in the previous chapter was -0.60 kcal/mol, however, it was improved for the 8-mer hairpin with -1.34 kcal/mol. Figure 5.2c illustrates the Watson-Crick base pairs formation at the stem for the truncated 6-mer hairpin (Figure 5.2c) and extended 8-mer hairpin (Figure 5.2d).

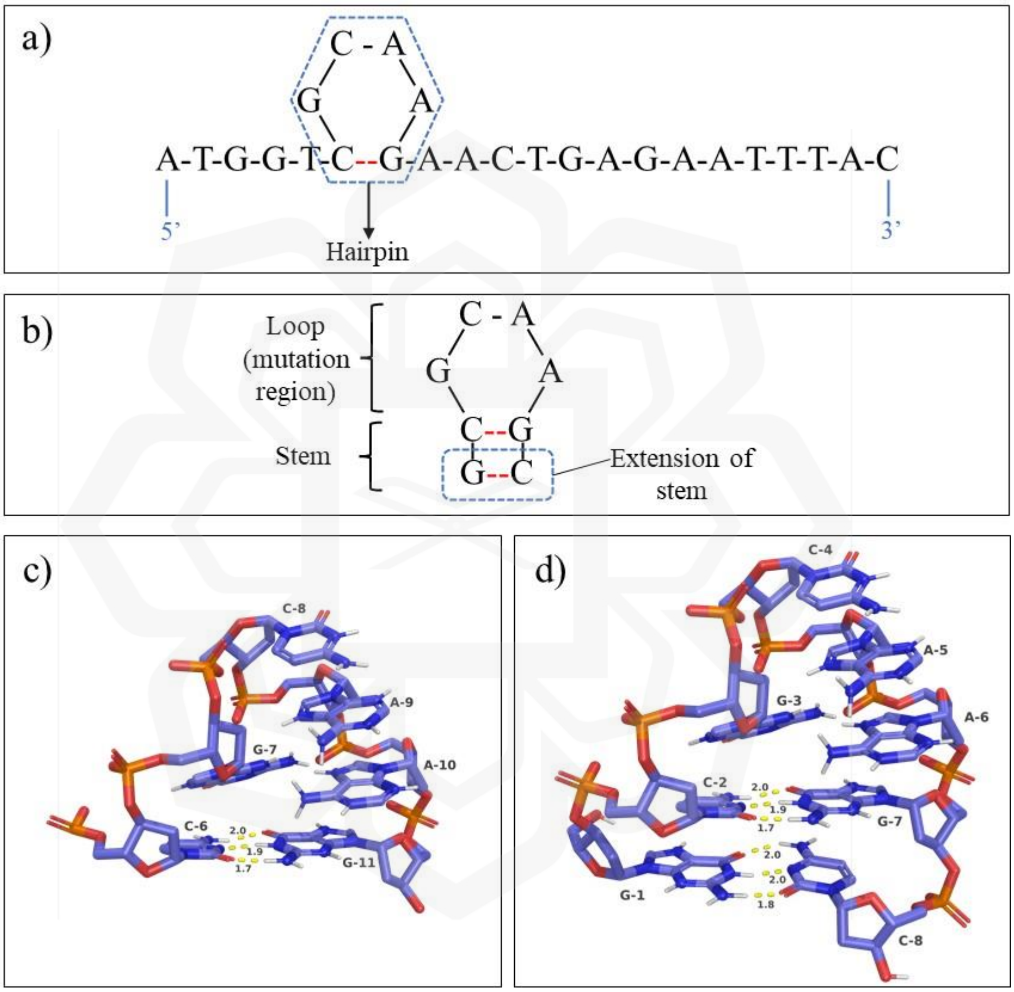


Figure 5.2 The A4 aptamer and truncated hairpin. (a) The 2D structure of 25-mer A4 aptamer, highlighting the hairpin region, (b) the mutation region and stem extension, forming 8-mer oligonucleotide, (c) the 3D structure of the truncated 6-mer hairpin, and (d) the 3D structure of the extended hairpin.

The nucleotides in the truncated hairpin (Figure 5.2c) were renumbered after stem extension, starting at nucleotide number one, G1 until C8 (Figure 5.2d) as only this 8-mer hairpin was utilised as the scaffold structure to design the DNA aptamer hairpin candidates against OC biomarker, HE4. Subsequently, the loop nucleotides (G3 to A6) of this scaffold hairpin were mutated, where each of the four nucleotides was substituted with thymine, adenine, guanine, and cytosine, one at a time, leading to the formation of 256 8-mer hairpins. Some of the hairpins are shown in Table 5.2, while the rest are available in the Supplementary Table (Appendix B).

Table 5.2 Examples of mutated hairpins from the mutation of loop nucleotides

No.	Hairpin	Sequence 5’-GC(loop)GC-3’
1	Scaffold	GC(GCAA)GC
2	H1	GC(ATCG)GC
3	H2	GC(ATCC)GC
4	H3	GC(ATCA)GC
5	H4	GC(ATCT)GC
6	H5	GC(ATGG)GC
7	H6	GC(ATGC)GC
8	H7	GC(ATGA)GC
9	H8	GC(ATGT)GC
10	H9	GC(ATAG)GC

Subsequently, the secondary and tertiary conformations of the 256 hairpins were modelled by Mfold (Markham & Zuker, 2008; Zuker, 2003) and RNAComposer (Oliveira et al., 2022; Sarzynska et al., 2023), following similar procedures in the Chapter 4 (Section 4.2.1). The secondary conformations of the 256 hairpins had very similar free energy values of approximately -1.34 kcal/mol, and the dot-bracket conformations (Vienna format) generated by the Mfold server were utilised in predicting the 3D structures. As RNA hairpins were modelled by the RNAComposer, they were converted to DNA hairpins, according to the manual designed earlier.

Molecular docking of these 256 DNA hairpins was conducted and analysed in the next section.

5.3.2 HE4 Protein and DNA Aptamer Hairpin Docking

AutoDock Vina was chosen to conduct the molecular docking between the target protein, HE4 with the 256 hairpins, individually, mainly due to its outstanding efficiency and easy-to-use tool (Nguyen et al., 2020; Vieira & Sousa, 2019). This includes the ability to continuously dock different hairpins to the same target, one after another, with simple commands that saved time and effort in completing the 256 dockings. Unfortunately, the ligand (hairpin) preparations were time-consuming as the prediction by Mfold and RNAComposer servers needed to be conducted one-by-one, followed by manual editing of the atoms, converting the ribose to deoxyribose sugar and replacing all uracil to thymine. Subsequently, each hairpin was assessed in ADT software where the water molecules were removed, polar hydrogens were assigned, and Kollman and Gasteiger charges were estimated before the torsion tree was determined. The Kollman and Gasteiger partial atomic charges represent the electron density distribution in the ligand which assists in the prediction of electrostatic interactions with the protein target (Lozano-Aponte et al., 2019).

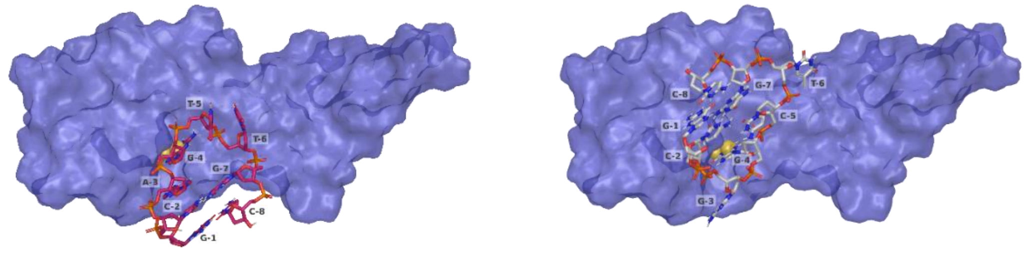
Torsions are the rotations around the single bonds of the molecule, either rigid or flexible, adopting various orientations and angles to fit into the binding site (Elokely & Doerksen, 2013). The presence of torsional degrees of freedom in conformational sampling facilitates a thorough investigation of the aptamer's possible binding configurations. The flexibility offered by the torsional angles increases the binding interaction of the aptamer against its protein (Kothandan et al., 2021). In this study, the torsions of the hairpin backbone (deoxyribose and phosphate groups) and the stem nucleotides were set as rigid, non-rotatable single bonds, as this conserved the main feature of aptamers which is the stem-loop motif. Only the torsions of the nucleobases at the hairpin loop were set as rotatable, enabling the hairpin to form stronger binding with the rigid HE4 protein structure, as the bonds engaged different orientations, and this was reported to be more accurate than rigid ligand-rigid receptor docking (Fan et al., 2019). Furthermore, this approach decreased the docking time and prevented the interactions within the base pairs of the stem from dissociating which may result in a linear ssDNA structure that occurred in earlier studies (Sabri et al., 2019). In common computational docking studies, ligands are freely docked onto the rigid receptor where the flexibility of the ligand side chain allows the ligand to alter its orientation according to the binding site of the receptor (Pagadala et al., 2017).

Vina molecular docking predicted the 10 best binding poses with the lowest Vina scores (binding energies) in kcal/mol and for this study, 10 complexes with the lowest binding energies out of 256 HE4-hairpin complexes were selected for thorough assessments. Table 5.3 shows the 10 complexes with the lowest binding energies ranging from -11.6 to -10.6 kcal/mol. The lowest energy is not necessarily the best binding pose as multiple analyses are required, including the binding site assessment, ligand and receptor residues identification, followed by MD simulations analysis which will be discussed comprehensively in the next subsection.

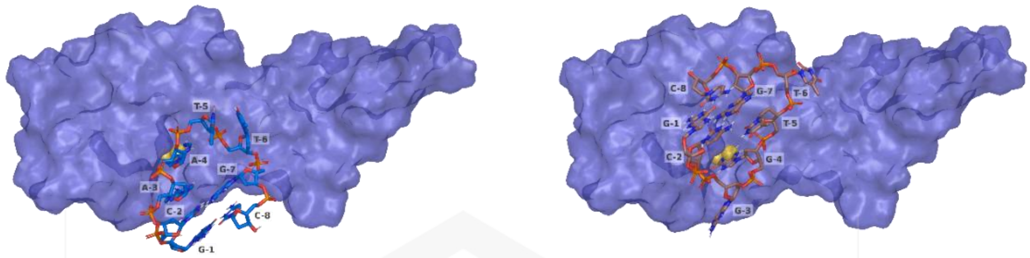
Table 5.3 Ten HE4-hairpin complexes with the lowest binding energies, in ascending order

HE4-hairpin complex	Sequence	Binding energy (kcal/mol)
HE4-H101	GCAGTTGC	-11.6
HE4-H197	GCGGCTGC	-11.5
HE4-H256	GCAATTGC	-11.2
HE4-H98	GCGGTTGC	-11.2
HE4-H143	GCGGATGC	-11.1
HE4-H252	GCAGTCGC	-11.1
HE4-H40	GCGGGTGC	-11.0
HE4-H44	GCAGGTGC	-11.0
HE4-H124	GCTTGGGC	-10.8
HE4-H16	GCATTTGC	-10.6

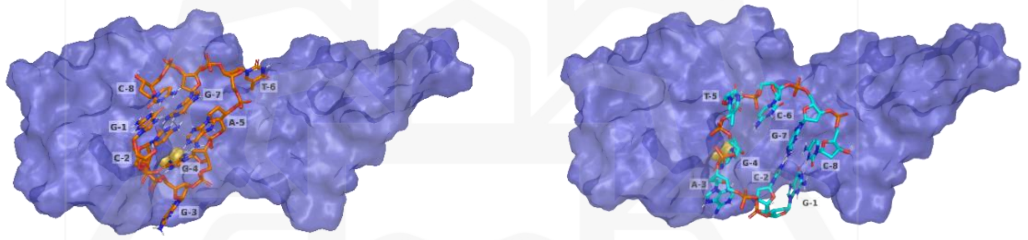
The binding energy reported in the previous chapter (subsection 4.3.3) for docked HE4-A4 aptamer was -6.0 kcal/mol. Notably, the current chapter displays significant improvements in the binding interactions, as the molecular dockings of the designed DNA aptamer hairpins with HE4 revealed substantially lower binding energies, ranging from -11.6 to -10.6 kcal/mol for 10 lowest binding energy complexes (Table 5.3). A study proved that lower binding energies corresponded to greater protein-ligand binding affinities after analysing the conformations of multiple chemical compounds with L99A T4 lysozyme receptor (Gallicchio et al., 2010). The 10 docked HE4-hairpin complexes were viewed in PyMOL molecular viewer, assessing their binding conformations and residues involved in forming the binding interactions, where the interactions were identified by the Protein-Ligand Interaction Profiler (PLIP) server. Figure 5.3 illustrates the docked conformations of 10 HE4-hairpin complexes before the three most potential hairpins were chosen to undergo MD simulations.



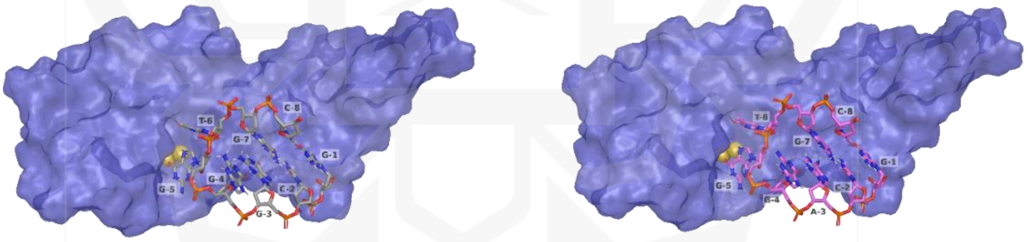
HE4-H101 (-11.6) HE4-H197 (-11.5)



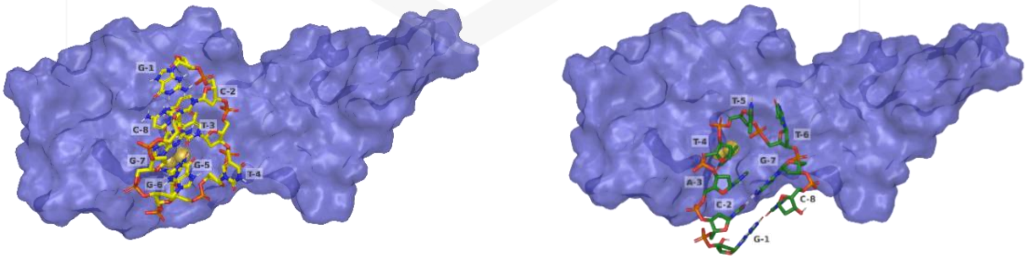
HE4-H256 (-11.2) HE4-H98 (-11.2)



HE4-H143 (-11.1) HE4-H252 (-11.1)



HE4-H40 (-11.0) HE4-H44 (-11.0)

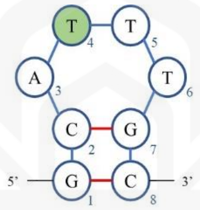
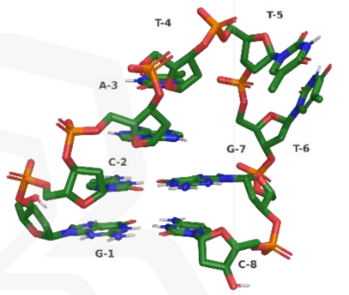
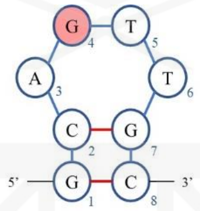
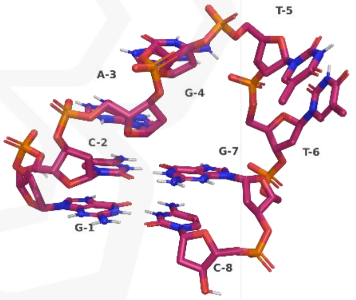
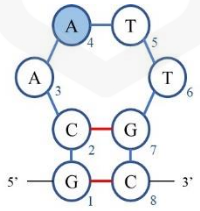
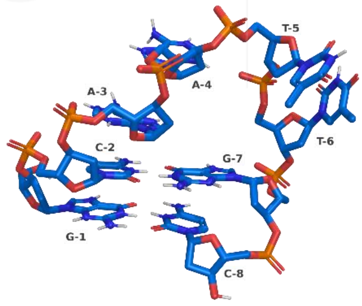


HE4-H124 (-10.8) HE4-H16 (-10.6)

Figure 5.3 The 3D conformations of 10 docked HE4-hairpin complexes with the lowest binding energies. The binding energies (in kcal/mol) are shown in brackets.

Out of 10 docked complexes, HE4-H101, HE4-H256, and HE4-H16 were selected for further analysis as they formed multiple hydrogen bonds between hairpin loop regions and the cavity residues predicted by CAVER 3.0 tool earlier, including the tunnel residues. The 15 HE4 tunnel-cavity residues determined were LEU24, VAL25, SER26, ALA30, GLU31, LYS32, CYS49, VAL50, SER51, ASP52, SER53, LYS60, CYS61, CYS62, and SER63.

Table 5.4 Primary sequence, secondary, and tertiary structures of the selected hairpins.

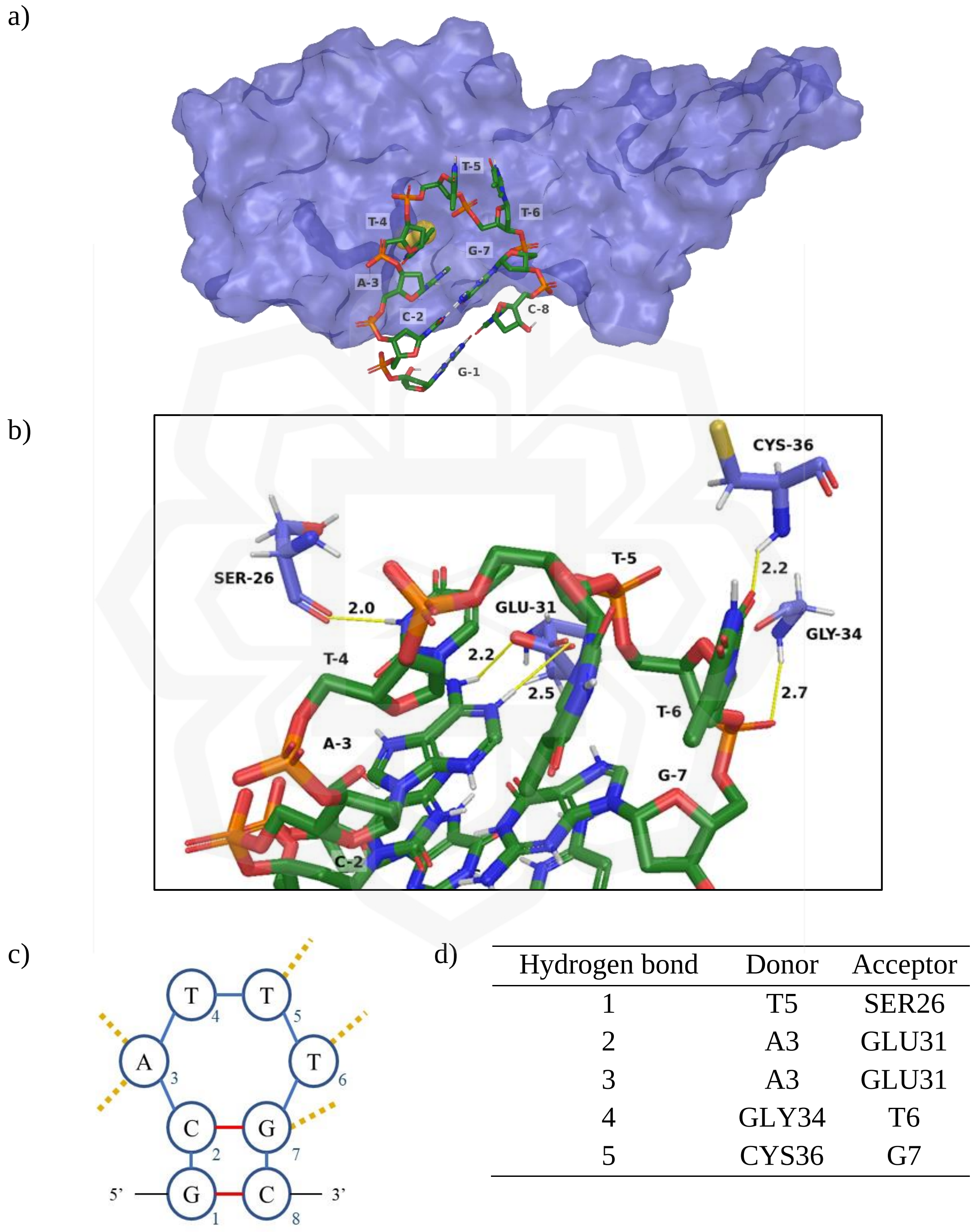
Hairpin	Sequence	Structure	
		Secondary	Tertiary
H16	GCATTTGC		
H101	GCAGTTGC		
H256	GCAATTGC		

*The different colours in secondary structures (green, red, and blue) represent the nucleotides that vary from one hairpin to another.

Figures 5.4, 5.5, and 5.6 display the intermolecular H-bonds within the complex receptor and ligand molecules. A hydrogen bond occurs when a highly electronegative atom (oxygen and nitrogen) is strongly attracted to a hydrogen atom that is bonded to another highly electronegative atom, and the typical distance is between 1.8 Å to 3.5 Å (Petsko & Ringe, 2004; Grabowski, 2006). It was reported that hydrogen bonding frequently occurs between hydroxyl (-OH) groups on protein side chains and phosphate groups on DNA in protein-DNA complexes, where they are crucial for maintaining the stability and specificity of these complexes (Lin & Guo, 2019; B. Yu et al., 2020). H101, H256, and H16 hairpins formed nine, eight, and four H-bonds, respectively, with HE4 tunnel-cavity residues. Notably, 80% of the H-bonds formed between HE4 and H16 occurred at the protein's tunnel-cavity site. Coincidentally, the sequences of these three selected hairpins are highly similar and they only differ at nucleotide number 4, as highlighted in Table 5.4 where the tertiary conformations are also displayed.

The conformational analysis of docked complexes in Figures 5.4a, 5.5a, and 5.6a, displayed that H16, H101, and H256 hairpins bound to similar sites of the HE4 protein where the hairpin residues formed multiple H-bonds at identical protein residues, SER26, GLU31, GLY34, and CYS36 with lengths between 2.0 Å and 3.3 Å. Figures 5.4b, 5.5b, and 5.6b show most of the H-bonds (yellow dashes) formed between the HE4 and hairpin residues, while Figures 5.4c, 5.5c, and 5.6c simplify the Figures 5.4b, 5.5b, and 5.6b to show the hairpin residues involved in the hydrogen bonding (yellow dashes). For H101 and H256 hairpins, it was obvious that the nucleotide G4 (in H101) and A4 (in H256) contributed significantly to the binding affinity of the HE4-hairpin complexes as they made up 6 and 5 H-bonds, respectively. The G4 and A4 residues are both located at the loop regions and the importance of the hairpin loop was reported in an earlier study, where anti-MUC1 ssDNA aptamer hairpin were immobilised on biosensor substrate, and it was predicted that the upright orientation of the aptamer where it has the most exposure of the hairpin loop region exhibited the most enhanced biosensor performance (Jeddi & Saiz, 2021). This single-stranded loop consists of the sequence responsible for target binding as it exposes the flexible atoms of the side chains, making it accessible for interactions with the target molecule (H. Yu et al., 2021; Z. Zhang & Liu, 2018). In a recent report, a study successfully developed an improved and highly sensitive electrochemical sensor targeting tobramycin, where the hairpin loop region was incorporated with a poly-thymine base sequence (Meng et al., 2023). The re-engineered loop region experienced huge conformational change upon the detection of tobramycin in real samples with a signal change percentage increased to

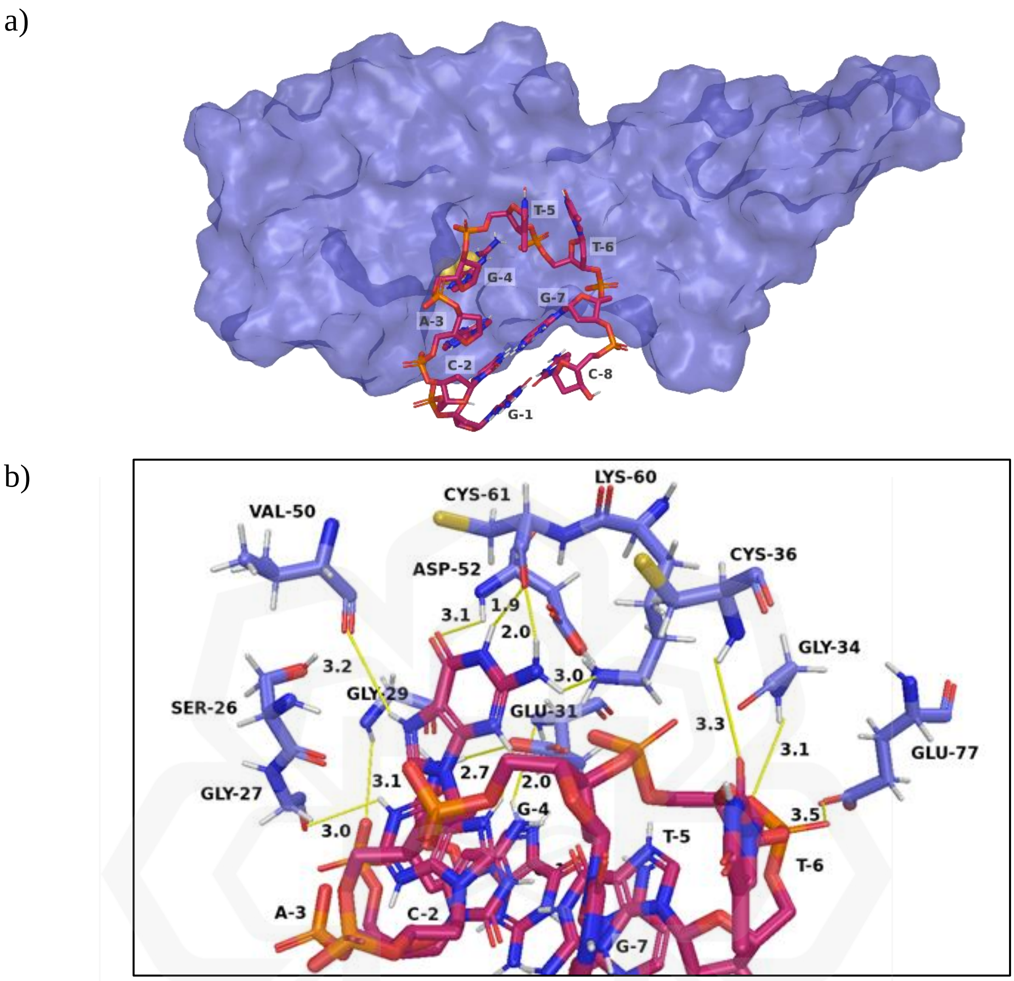
767.1%. Lastly, Figures 5.4d, 5.5d, and 5.6d summarised the total number of H-bonds that existed between HE4 and hairpin molecules, and their residues responsible as the H-bond donors and acceptors.



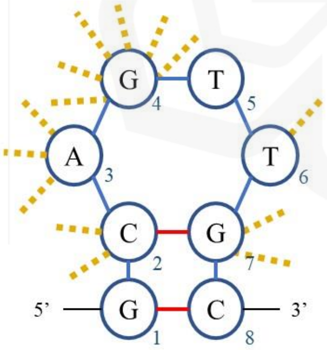
Hydrogen bond	Donor	Acceptor
1	T5	SER26
2	A3	GLU31
3	A3	GLU31
4	GLY34	T6
5	CYS36	G7

Figure 5.4 Binding pose and hydrogen bond formation in HE4-H16 complex. (a) Whole tertiary conformation of the complex, (b) a closer view of the interactions, (c) 2D structure, showing the residues involved in the H-bonds, and (d) The H-bond donors and acceptors.

a)



c)

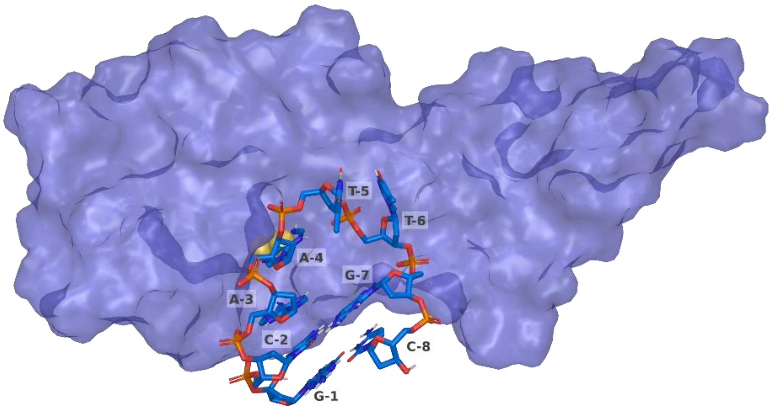


d)

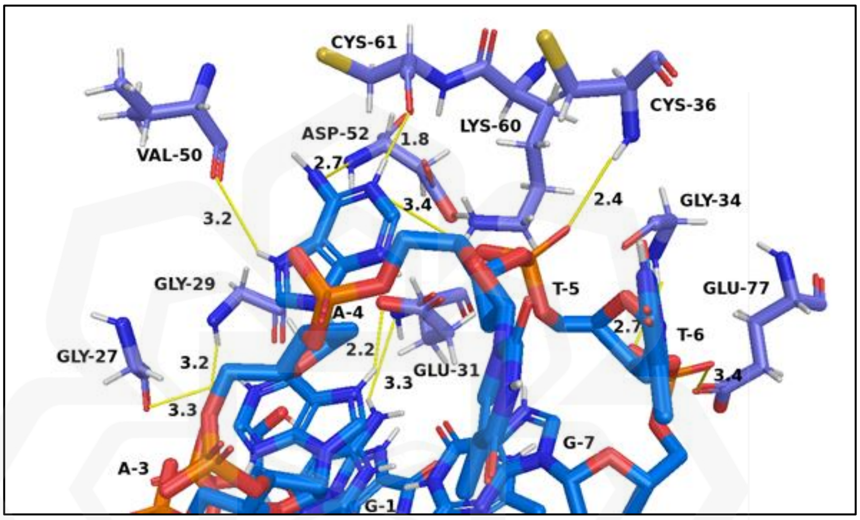
Hydrogen bond	Donor	Acceptor
1	SER26	G4
2	A3	GLY27
3	GLY29	C2
4	A3	GLU31
5	GLU31	C2
6	A3	GLU31
7	GLY34	G7
8	CYS36	T6
9	G4	VAL50
10	ASP52	G4
11	LYS60	G4
12	G4	CYS61
13	G4	CYS61
14	GLU77	G7

Figure 5.5 Binding pose and hydrogen bond formation in HE4-H101 complex. (a) Whole tertiary conformation of the complex, (b) a closer view of the interactions, (c) 2D structure, showing the residues involved in the H-bonds, and (d) The H-bond donors and acceptors.

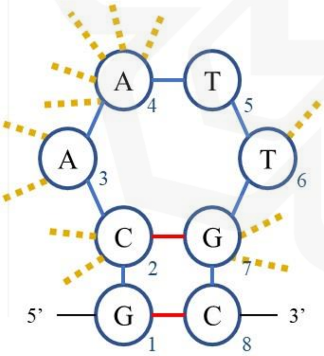
a)



b)



c)



d)

Hydrogen bond	Donor	Acceptor
1	A3	GLY27
2	GLY29	C2
3	A3	GLU31
4	GLU31	C2
5	GLY34	G7
6	CYS36	T6
7	A4	VAL50
8	ASP52	A4
9	A4	ASP52
10	LYS60	A4
11	A4	CYS61
12	GLU77	G7

Figure 5.6 Binding pose and hydrogen bond formation in HE4-H256 complex. (a) Whole tertiary conformation of the complex, (b) a closer view of the interactions, (c) 2D structure, showing the residues involved in the H-bonds, and (d) The H-bond donors and acceptors.

The adenine and guanine bases formed multiple H-bonds with the protein residues. G4 base in H101 formed 6 H-bonds and they are all bound to the active binding site (tunnel and cavity) residues, SER26, VAL50, ASP52, LYS60, and CYS61, that were predicted by the CAVER 3.0 tool. In the HE4-H256 complex, there were 7 H-bonds formed by A3 and A4 residues, where five of them bound to the protein active binding site residues. To explain this, the structures of the nitrogenous bases, adenine (A), thymine (T), cytosine (C), and guanine (G) are illustrated in Figure 5.7. A and G are purine bases with five-membered and six-membered ring structures which made them larger and more complex than C and T, the single six-membered ring pyrimidine bases (Minchin & Lodge, 2019). Both A and G consist of two H-bond donors and one H-bond acceptor, unlike T, which has only one donor and one acceptor. In the G base, the amino (-NH_2) group and nitrogen atom act as the donors, having the oxygen atom as the H-bond acceptor. Due to the presence of the H-bond donors and acceptor, a single A or G can form multiple H-bonds with the receptor residue at one time. Referring to Figures 5.4 to 5.6, the T bases located at H16, H101, and H256 loop formed one or no H-bond with the protein residue.

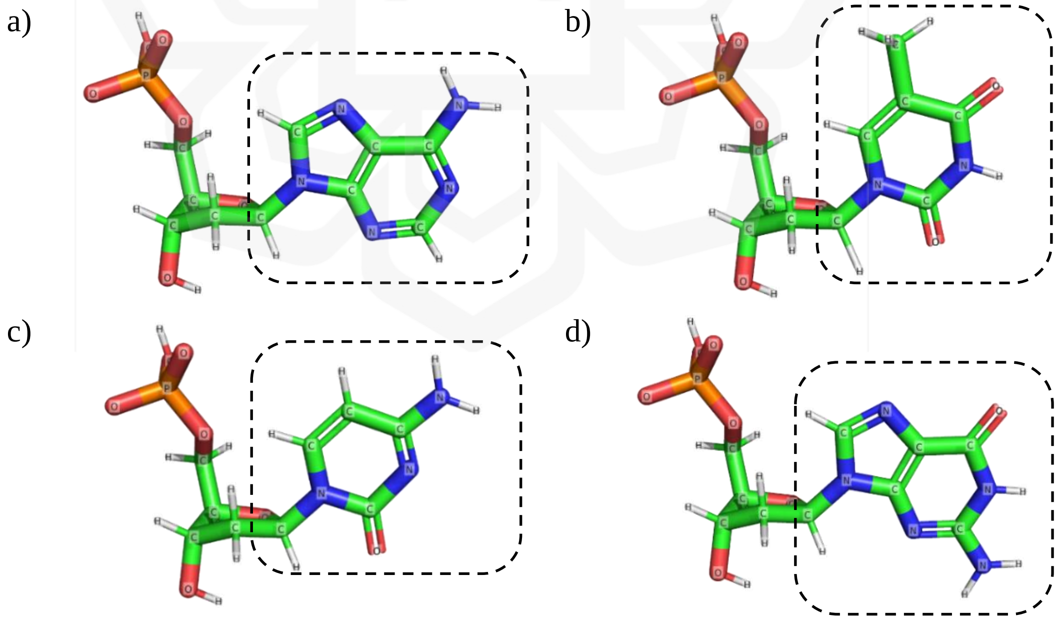


Figure 5.7 The nitrogenous bases of DNA. (a) adenine, (b) thymine, (c) cytosine, and (d) guanine, in the dotted boxes.

Apart from H-bonds, the intermolecular interactions within HE4-hairpin complexes may be contributed by salt bridges and hydrophobic interactions. Their presence was identified by the PLIP tool, together with the H-bonds earlier. Hydrophobic interactions were not significant in this study as only one hydrophobic interaction formed between VAL25 and T4 in HE4-H16 complex with a distance of 3.6 Å, while none existed between the HE4 and hairpin molecules in HE4-H101 and HE4-H256 complexes. This interaction occurs between nonpolar molecules or the side chains of the amino acids and ligand residues (Ferreira de Freitas & Schapira, 2017). The other interaction, salt bridges, were formed within all complexes and they are listed in Table 5.5. Salt bridges arise from the electrostatic interactions contributed by oppositely charged groups (Kurczab et al., 2018).

Table 5.5 The salt bridges occurred within HE4-hairpin complexes.

Complex	Residues/atoms			Distance (Å)
	Protein	DNA aptamer hairpin		
HE4-H16	LYS60	T5	O3’	4.01
		T6	P	
		T6	O5’	
		T6	OP1	
		T6	OP2	
HE4-H101	GLU31	A3	N9	5.46
	GLU31	G4	N1	3.72
		G4	N2	
		G4	N3	
	ASP52	G4	N1	4.37
		G4	N2	
		G4	N3	
	LYS60	T5	O3’	4.44
		T6	P	
		T6	O5’	
		T6	OP1	
		T6	OP2	
HE4-H256	GLU31	A4	N9	3.73
	LYS60	T5	O3’	4.75
		T6	P	
		T6	O5’	
		T6	OP1	
		T6	OP2	

The salt bridge formations involved oxygen, nitrogen, and phosphorus atoms located at hairpin loops. Nucleotides at positions 3 to 6 were located at these flexible loop regions where they are exposed to form greater interactions with the receptor residues. Identically, the LYS60 residue in all three complexes formed salt bridges with T5 and T6 hairpin residues, with a distance ranging from 4.01 Å to 4.75 Å. A salt bridge helps neutralising the negative charge of the DNA backbone, enabling the protein to closely interact with the DNA (Vo et al., 2021). This interaction is essential for the stability of the complex and is frequently implicated in the initial recognition and binding phase.

Even though the HE4-H101 complex generated the lowest binding energy of -11.6 kcal/mol and formed the highest number of H-bonds, further analysis was still needed before the most potential DNA aptamer hairpin against HE4 protein could be determined. Upon the completion of molecular docking and docked complexes binding analysis, 100 ns MD simulations using CHARMM27 force field were subsequently conducted on these HE4-H16, HE4-H101, and HE4-H256 complexes, individually. The stability of the complex conformations and binding interactions in long simulations, mimicking the actual environment are comprehensively discussed in the molecular dynamics study in the following section.

5.3.3 Molecular Dynamics (MD) Simulations of HE4-Hairpin Complexes

MD simulations using GROMACS are an effective computational method for investigating the dynamic behaviour of biomolecular systems, such as protein-protein and protein-DNA complexes. GROMACS, a widely used software programme in computational studies, uses force fields to simulate atom interactions and describe molecule motions within a particular timeframe (Lemkul, 2018). The simulation setup consists of identifying the molecular structure, followed by solvation, ions addition, and force field assignment. Subsequently, energy minimisation was applied, before MD simulation. GROMACS aids the investigation of these dynamic events by numerically integrating Newton's equations of motion, allowing for the prediction of structural changes and energetics at the atomic level, with many improvements, up-to-date (Pronk et al., 2013; Rakhshani et al., 2019). With various force fields available today, CHARMM force field is widely applied to the MD simulations of proteins, nucleic acids, and other molecules like lipids and carbohydrates, and it is supported by many MD simulations programme, including GROMACS (Guterres et al., 2022; Huang et al., 2017; Vanommeslaeghe et al., 2010). Two recent studies applied CHARMM force field in the MD simulations of heparin, an anticoagulant that binds to multiple proteins in the human body; and a protein-DNA/RNA system consisting of elongation factor SPT5 (Gallardo et al., 2022; Janke et al., 2022). Protein-DNA complexes serve important roles in cellular activities, and knowing their dynamics is essential for unravelling biological systems. Here, the stability of binding interfaces, the conformational changes in both protein and DNA components, and the atomic position changes are evaluated. Before conducting MD simulation analyses such as RMSD and Rg evaluations, the clustering step was applied, right after MD simulations. This step analyses and organises the MD trajectories, enabling the identification of distinct conformational structures as a vast amount of data was generated during the MD simulations and various clustering algorithms have been adopted by previous studies (Hao et al., 2015; Shao et al., 2007). This simplifies the complex binding analysis.

5.3.3.1 RMSD, radius of gyration and RMSF

Root-mean-square-deviation (RMSD) quantifies the mean deviation of atoms in a molecular structure from a reference structure during a simulation. It offers a measure of the extent to which the simulated structure differs from the initial or other conformations (Kufareva & Abagyan, 2012; Ramírez & Caballero, 2018). It is determined by aligning the atomic coordinates of the simulated structure with the corresponding atoms in the reference structure and subsequently calculating the root mean square of the positional differences.

Figure 5.8 shows the RMSD and radius of gyration (Rg) graphs of HE4-H16, HE4-H101, and HE4-H256 complexes over 100 ns of simulations using the CHARMM27 force field. Similar to RMSD, Rg provides information on structural stability, together with the overall compactness of the structure folding (Rampogu et al., 2022). Changes in the Rg during an MD simulation can be taken as an indication of conformational changes in the molecule. A decreasing Rg may be due to folding or compaction, while an increase in Rg might indicate the unfolding or expansion of the molecule, and this is essential for interpreting the dynamic behaviour of the simulated system.

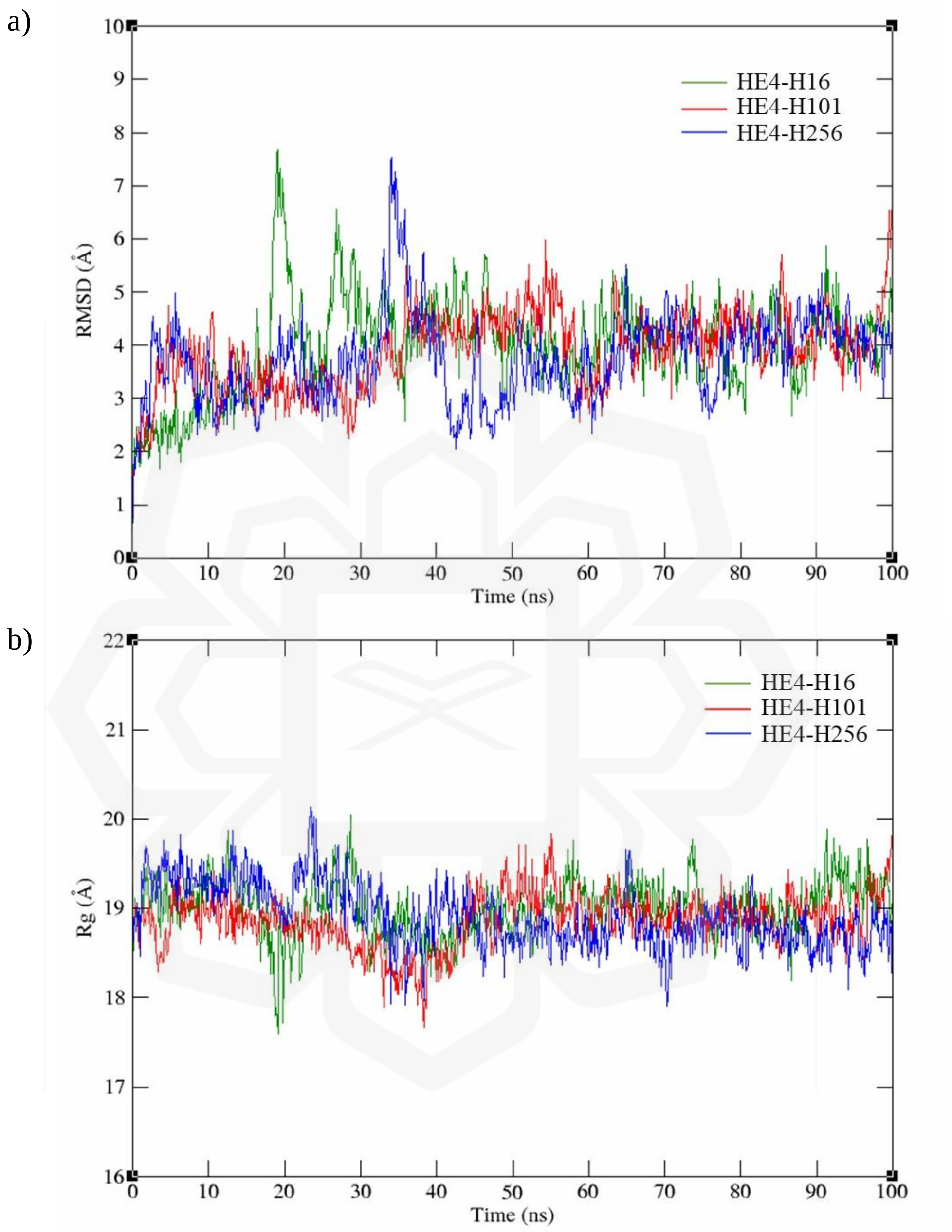


Figure 5.8 (a) RMSD and (b) Rg plots of protein backbone over 100 ns of MD simulations of HE4-hairpin complexes. HE4-H16 (green), HE4-H101 (red), and HE4-H256 (blue).

The backbone RMSD of the docked HE4 protein in HE4-hairpin complexes reached equilibrium after 50 ns for all systems (Figure 5.8a). In other protein-DNA MD simulations studies, the RMSD values of the protein backbone achieved equilibrium after 7ns (out of 20) in HBsAg protein-DNA complexes (Sabri, 2021), after 100 ns (out of 300 ns) in five different systems (Xu et al., 2023), and 40 ns (out of 100 ns) in p53 protein-DNA complexes (Kamaraj & Bogaerts, 2015). These various outcomes arise from the different structures of each system. The RMSD plot displays different deviation patterns for HE4-H16, HE4-H101, and HE4-H256 systems from 0 ns to 50 ns. HE4-H16 system experienced significant fluctuations between 15 ns to 25 ns and these fluctuations are reflected in the Rg graph. As discussed in the Section 5.3.2, the docked conformation of HE4 protein and H16 hairpin simulated by Vina exhibited the least number of intermolecular H-bonds (five H-bonds). This shows weak interactions between the protein and ligand throughout the MD simulations course, in which the interactions might dissociate, resulting in high fluctuations of RMSD value. Notably, after 50 ns, the RMSD values deviated only within ±2 Å, ranging between 3 Å to 5 Å for all systems, indicating that the protein was highly stable. The backbone RMSD of a study on DNA-tumour protein p53 achieved equilibrium after 40ns, out of 100 ns simulation course, and had an average Rg value of 1.69 nm (16.9 Å) (Balasundaram & Doss, 2022).

Referring to the Rg plot of this study, all systems maintained relatively steady values after 50ns at approximately 19 Å, indicating compact and folded protein backbone structures, closely resembling the native structure (Kamaraj & Bogaerts, 2015). In the protein structures compactness analysis of leukaemia mutations in protein-DNA complexes, the wildtype backbone reached steady values after 65 ns out of 150 ns simulation time, while the mutants experienced slight fluctuations, and only achieved low and steady values slightly later (Kamaraj et al., 2021). Based on RMSD and Rg graphs in Figure 5.7, the most potential DNA aptamer hairpin against HE4 protein was not able to be determined as they exhibited very similar outcomes, thus, more evaluations were conducted, including the analysis of root-mean-square-fluctuation (RMSF), number of hydrogen bonds, and the minimum distance between the protein and hairpin molecules.

Root-mean-square-fluctuation (RMSF) quantifies the fluctuation of individual residues within a molecule throughout the simulation (Sharma et al., 2021). It offers valuable information on regions of the structure that undergo substantial conformational changes. It represents the variety in the locations of atoms or residues within the molecule. Greater RMSF values imply increased flexibility or movement in certain regions, whilst lower RMSF values indicate more stable regions. Both RMSD and RMSF are evaluated for the accuracy of simulations, and different complex trajectories while highlighting the areas of interest that experience conformational changes. Figure 5.9 shows the RMSF values of 124 amino acids in HE4 protein in the complex system.

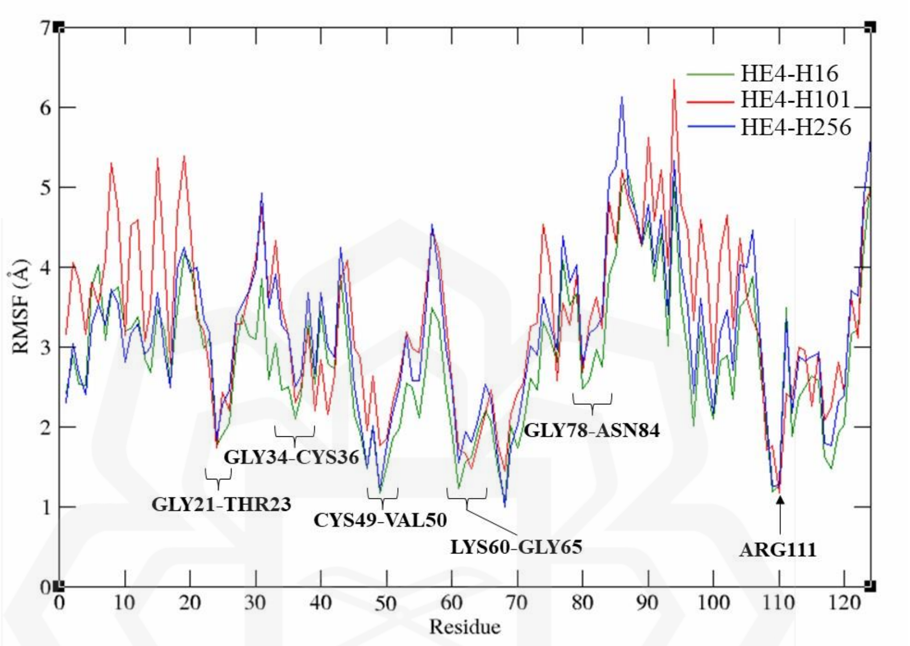


Figure 5.9 RMSF plot of HE4-hairpin complexes. HE4-H16 (green), HE4-H101 (red) and HE4-H256 (blue). The labelled residues are some HE4 residues that formed hydrogen bonds with DNA aptamer hairpin

In the plot above, the x and y axis represent the HE4 protein residues and RMSF values, respectively. The residues ranging from 21 to 23, 34 to 36, 49 to 50, 60 to 65, and 78 to 84 showed low RMSF values, suggesting more stable regions. According to the results obtained from Vina molecular docking (Figures 5.4 to 5.6), these are the common residues involved in forming the H-bonds between protein and hairpin molecules. For example, the HE4-H256 complex had 12 H-bonds, 8 of which involved the residues mentioned here. All three hairpins (H16, H101, and H256) identically formed H-bonds with GLY34 and CYS36 protein residues with lengths ranging from 2.1 Å to 3.1 Å. High fluctuations shown between residues 5 to 18 and 86 to 100 indicated less stable protein regions, which was proven by the non-existent interactions or H-bonds between the protein and DNA molecules, based on the docked complex analysis in the previous section. Figure 5.9 displays HE4 tunnel-cavity residues predicted by the CAVER tool earlier which are presented in green, and the neighbouring residues (in pink) that were commonly involved in making the intermolecular H-bonds in HE4-H16, HE4-H101, and HE4-H256 complexes. These residues corresponded to low RMSF values they generated and labelled in Figure 5.9. The H-bonds formed (listed in Figure 5.4d, 5.5d, and 5.6d) in all three complexes were centralised at this region highlighted in Figure 5.10.

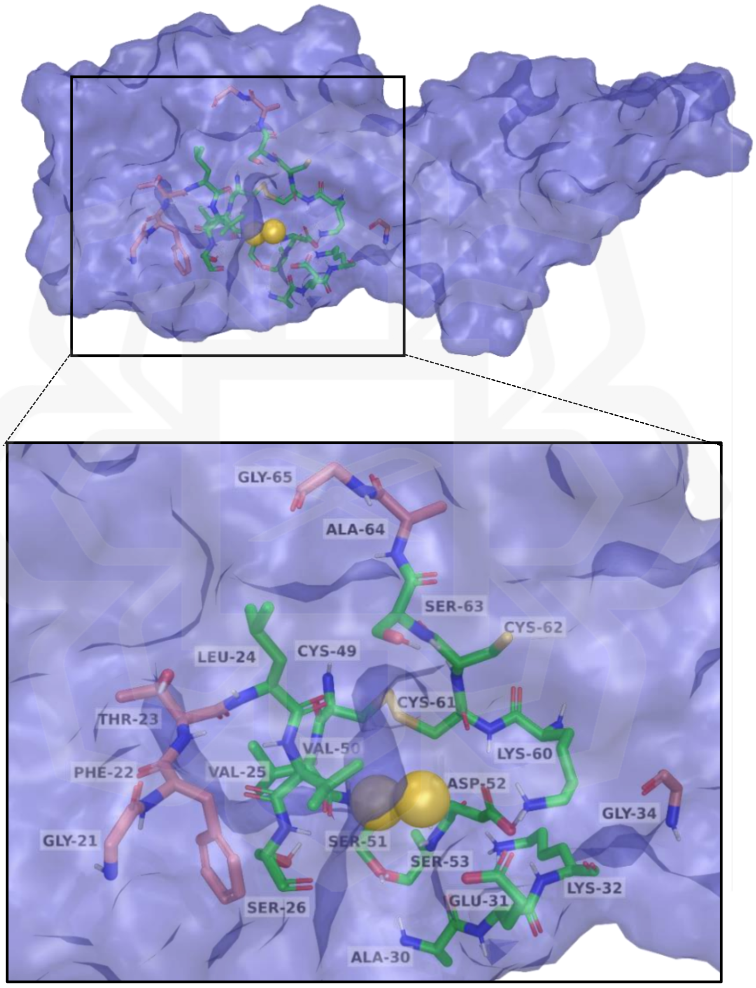


Figure 5.10 HE4 protein residues. The tunnel-cavity residues determined by CAVER 3.0 tool (in green) and the common residues involved in forming intermolecular H-bonds, in HE4-H16, HE4-H101, and HE4-H256 complexes. The yellow dots represent the predicted binding site by CAVER 3.0.

5.3.3.2 Hydrogen bonds and minimum distance

The number of hydrogen bonds formed between HE4 protein and H16, H101, and H256 hairpins correlate with the minimum distance between protein and DNA molecules

(Figure 5.11). According to the Protein Structure and Function textbook, the length of a hydrogen bond is up to 3.5 Å (Petsko & Ringe, 2004). Based on the molecular docking analysis earlier, there were 6, 14, and 12 intermolecular H-bonds formed within HE4-H16, HE4-H101, and HE4-H256 complexes, respectively. It is worth highlighting that the docking analysis was done before MD simulations were conducted, thus, the dynamic behaviour including the stability of H-bond formation was not certain in a simulated system of 100 ns.

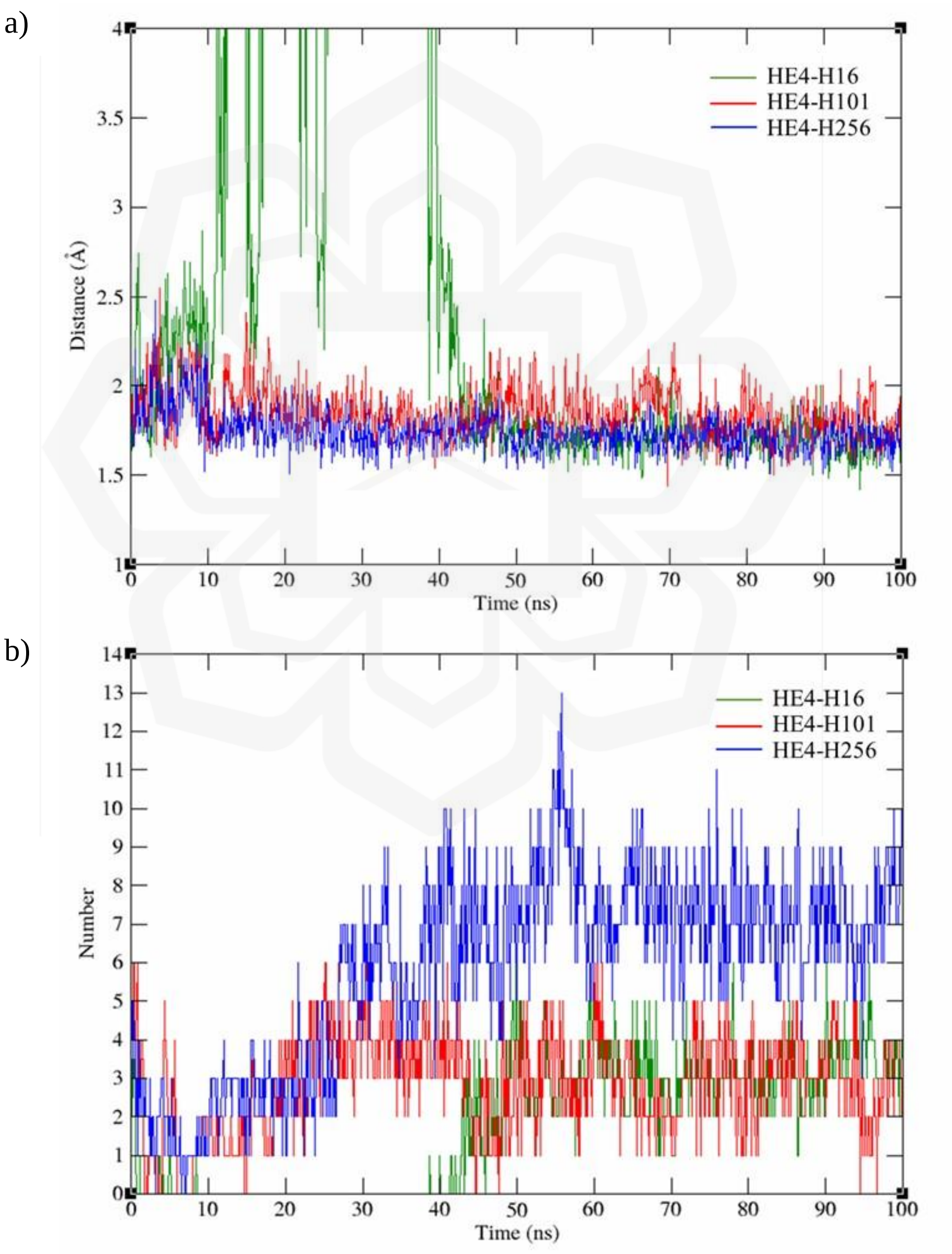


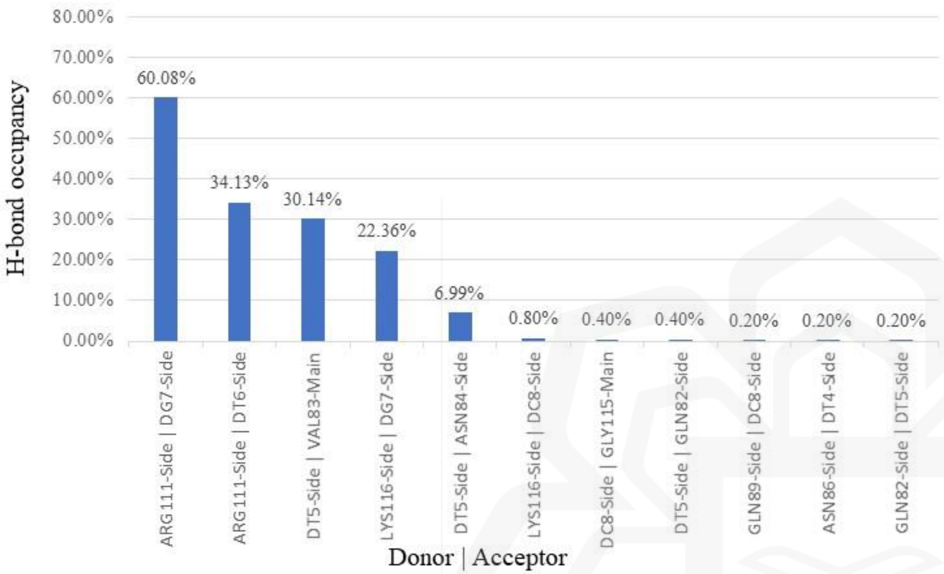
Figure 5.11 The (a) minimum distance and (b) number of hydrogen bonds between HE4 protein and H16, H101, and H256 hairpins.

In Figure 5.11a, the minimum distance between the protein and DNA molecules for HE4-H101 and HE4-H256 complexes maintained below 2.0 Å for most of the simulation time, showing the presence of multiple H-bonds (Figure 5.11b). Meanwhile, for HE4-H16 complex, there was no formation of H-bonds between 10 to 40 ns, as the distance fluctuated to higher than 3.5 Å. This occurred due to the conformational changes where the HE4 protein and hairpin H16 experiences structural shifts, leading to the H-bonds donor and acceptor to shift away from each other (Battistel et al., 2017). This reduces the overall binding interaction within the HE4-H16 system. Based on the RMSD, higher deviation elucidates the effect of fewer H-bonds formed between protein and DNA (Balasundaram & Doss, 2022). This can be seen in the HE4-H16 system as it had multiple high deviations within this timeframe (Figure 5.8a). As for HE4-H101 and HE4-H256 complexes, they maintained a minimum distance at approximately 2 Å, showing how closely intact the protein and DNA molecules were. However, the HE4-H256 complex has shown remarkable output where it maintained a high number of H-bonds, between 6 to 13 H-bonds for nearly 70% of the 100 ns simulation course. The complex also maintained the minimum distance at approximately 1.7 Å, maintaining the close interactions between the molecules, from 10 ns to 100 ns. Based on these minimum distances and the number of H-bonds analyses, the HE4-H256 complex is likely to be selected as the most potential or promising hairpin against HE4 protein, but additional analyses such as hydrogen bond occupancy and the total energy using MM/GBSA were conducted to strengthen the selection.

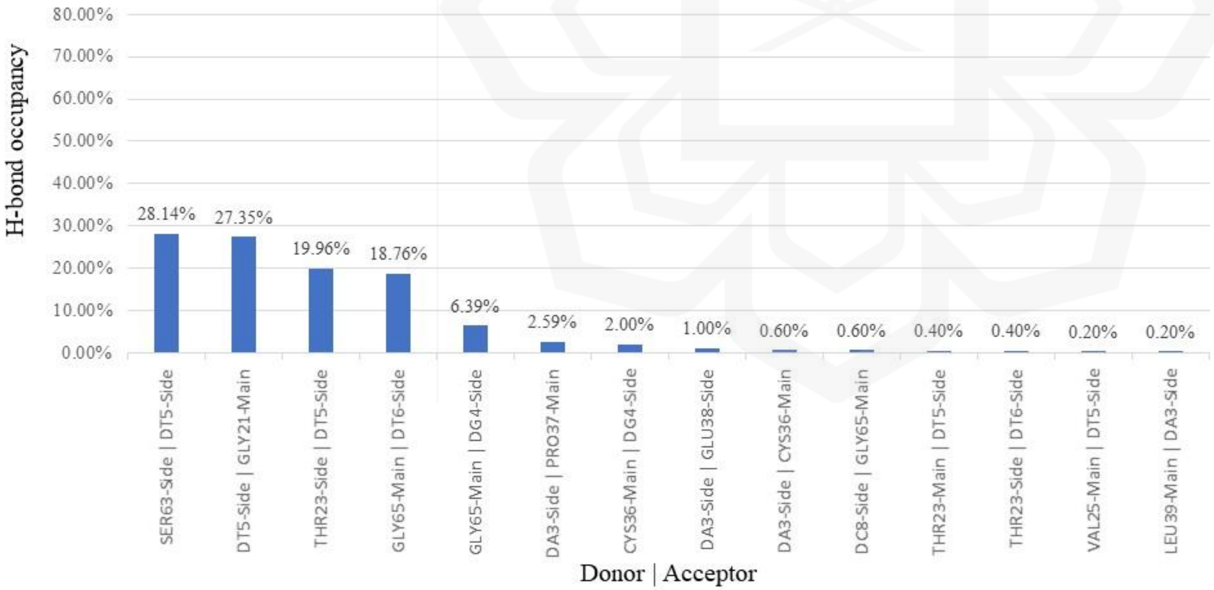
5.3.3.3 Hydrogen bond occupancy and MM/GBSA analysis

In this study, the HE4 protein tertiary structure was modelled computationally, while the active binding residues were predicted by CAVER 3.0 tool, as discussed in the earlier chapter. By analysing the hydrogen bond occupancy of the HE4-hairpin complexes during the MD simulations, the significant interactions contributed by the HE4 active binding residues, and the hairpin residues were determined. Here, the VMD software was utilised to visually represent the development of dynamic hydrogen bonds between the complex molecules using the ‘hydrogen bonds’ plugin (Sabri, 2021). As mentioned earlier, an H-bond is established when a donor, a hydrogen atom, and an acceptor (oxygen, nitrogen, or fluorine) are present, and a typical H-bond distance is less than 3.5 Å. Figure 5.12 depicts all the HE4 and hairpin residues involved in the H-bonds where the residues are stated as a hydrogen donor or acceptor. The y-axis shows the H-bond occupancy (in %) for 50 ns, starting at 50 ns until 100 ns of MD simulations. This timeframe was chosen based on the RMSD and Rg plots where they achieved stability and compactness after 50 ns until the end of the simulation course.

a)


H-bond occupancy
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
60.08%
34.13%
30.14%
22.36%
6.99%
0.80%
0.40%
0.40%
0.20%
0.20%
0.20%
ARG111-Side | DG7-Side
ARG111-Side | DT6-Side
DT5-Side | VAL83-Main
LYS116-Side | DG7-Side
DT5-Side | ASN84-Side
LYS116-Side | DC8-Side
DC8-Side | GLY115-Main
DT5-Side | GLN82-Side
GLN89-Side | DC8-Side
ASN86-Side | DT4-Side
GLN82-Side | DT5-Side
Donor | Acceptor

b)


H-bond occupancy
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
28.14%
27.35%
19.96%
18.76%
6.39%
2.59%
2.00%
1.00%
0.60%
0.60%
0.40%
0.40%
0.20%
0.20%
SER63-Side | DT5-Side
DT5-Side | GLY21-Main
THR23-Side | DT5-Side
GLY65-Main | DT6-Side
GLY65-Main | DG4-Side
DA3-Side | PRO37-Main
CYS36-Main | DG4-Side
DA3-Side | GLU38-Side
DA3-Side | CYS36-Main
DC8-Side | GLY65-Main
THR23-Main | DT5-Side
THR23-Side | DT6-Side
VAL25-Main | DT5-Side
LEU39-Main | DA3-Side
Donor | Acceptor

c)

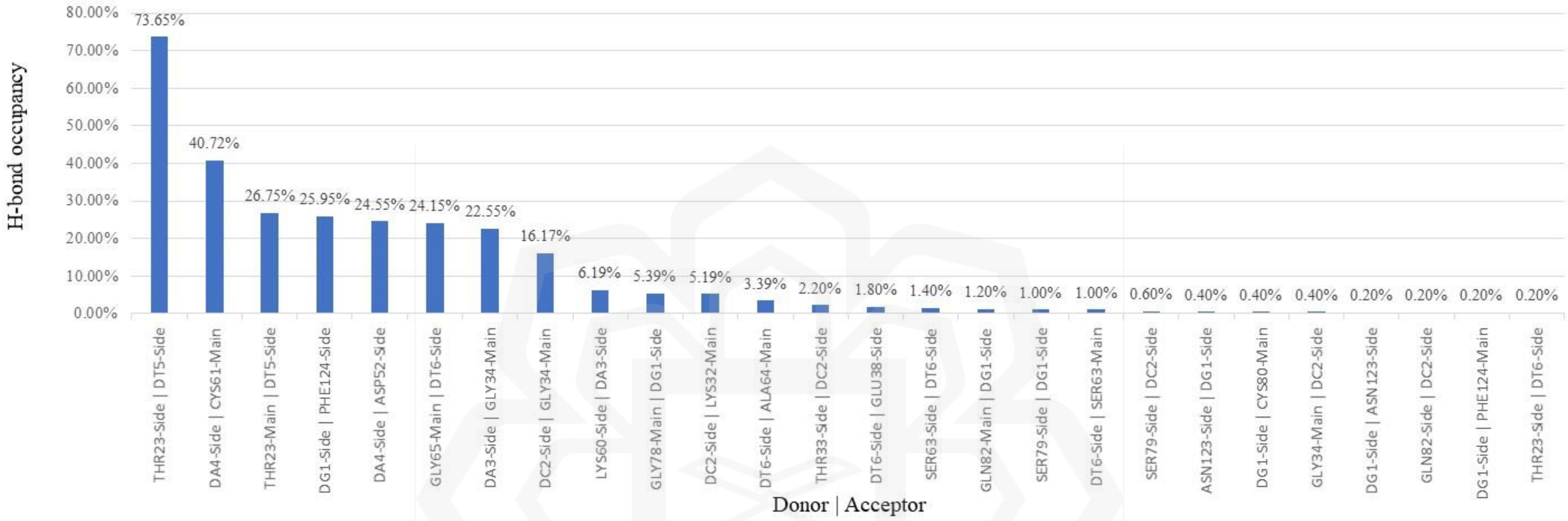


Figure 5.12 Hydrogen bond occupancy between HE4 and hairpin residues within the complexes, (a) HE4-H16, (b) HE4-H101, and (c) HE4-H256.

Figure 5.12c shows the highest number of H-bonds and it has the highest occupancy that was formed between the THR23 residue side chain of HE4 protein and the side chain of T5 residue of H256 hairpin. The recorded occupancy was at 73.65 %, showing that this particular H-bond was intact for more than 2/3 of the simulation time, indicating strong and stable interaction. T5 residue in this H256 hairpin was located at the loop region, where the side chain atoms were flexible, exposing a larger contact area with the HE4 protein (Jeddi & Saiz, 2021). The 10 highest H-bond occupancies (out of 26 total H-bonds) within this HE4-H256 complex were mainly formed by the nucleotides located at the hairpin loop, including A3, A4, and T6 residues, specifically at the flexible side chain. The occupancies ranged between 5.39 % to 73.65 %. 11 intermolecular H-bonds existed within the HE4-H16 complex with occupancies ranging from 0.20 % to 60.08 %, with only five having occupancies of more than 5 %. The highest H-bond occupancy generated within HE4-H101 was low compared to the other two complexes, at 28.14 %, formed by the side chain of SER63 residue and the side chain of T5 residue located at the hairpin loop region.

Table 5.6 lists the hydrogen donors and acceptors involved in the formation of the H-bonds with occupancies of more than 5 %. The H-bonds highlighted in yellow involved the active binding site residues while the H-bonds in green were formed by the neighbouring residues of the active binding site. Only five H-bonds with the said occupancies were present within HE4-H16 and HE4-H101 complexes, and none of the H-bonds in the HE4-H16 complex involved the protein tunnel and cavity residues, or its neighbouring residues as highlighted earlier in Figure 5.10. Furthermore, the highest occupancy of this complex, 60.08 %, was formed far from the protein’s active binding site, at ARG111 residue, and G7 residue of the H16 hairpin stem. Focusing on the hydrogen donors and acceptors of the intermolecular H-bonds within the HE4-H101 complex, even though the SER63, GLY21, THR23, and GLY65 were among the active binding site and its neighbouring residues, the occupancies were not convincing, as the interactions existed for only 28.14 % (and below) of the final 50 ns simulation time. The findings indicate that while the overall quantity of H-bond interactions is significant in determining the affinity of the complex, the percentage of H-bond occupancy has a greater influence on the stability of the protein-aptamer complex (Zikri et al., 2021). Based on Table 5.6 and other analyses conducted earlier, including RMSD, Rg, RMSF, and the number of H-bonds, the H256 hairpin has shown remarkable results to be entitled as the most promising DNA aptamer hairpin against ovarian cancer biomarker, HE4. This was further supported by the calculation of total binding energy by the MM/GBSA approach.

Table 5.6 List of hydrogen bond occupancies, revealing the hydrogen donor, acceptor and specific hairpin region involved in the interaction.

Complex	Hydrogen bond				Hairpin	
	Total no.	Donor	Acceptor	% Occupancy	Stem	Loop
HE4-H16	11	ARG111-Side	DG7-Side	60.08	●	
		ARG111-Side	DT6-Side	34.13		●
		DT5-Side	VAL83-Main	30.14		●
		LYS116-Side	DG7-Side	22.36	●	
		DT5-Side	ASN84-Side	6.99		●
HE4-H101	14	SER63-Side	DT5-Side	28.14		●
		DT5-Side	GLY21-Main	27.35		●
		THR23-Side	DT5-Side	19.96		●
		GLY65-Main	DT6-Side	18.76		●
		GLY65-Main	DG4-Side	6.39		●
HE4-H256	26	THR23-Side	DT5-Side	73.65		●
		DA4-Side	CYS61-Main	40.72		●
		THR23-Main	DT5-Side	26.75		●
		DG1-Side	PHE124-Side	25.95	●	
		DA4-Side	ASP52-Side	24.55		●
		GLY65-Main	DT6-Side	24.15		●
		DA3-Side	GLY34-Main	22.55		●
		DC2-Side	GLY34-Main	16.17	●	
		LYS60-Side	DA3-Side	6.19		●
		GLY78-Main	DG1-Side	5.39	●	
		DC2-Side	LYS32-Main	5.19	●	

*This table comprises H-bonds with occupancy of at least 5% over the last 50 ns simulation time.

The total binding energy of a complex can be computed with Molecular Mechanics Generalised Born Surface Area (MM/GBSA), aiming to calculate the binding free energy between a ligand and a receptor, such as a protein (Genheden &

Ryde, 2015). This method utilises the Generalised Born model to calculate solvation energies, specifically the electrostatic component (Mishra & Koča, 2018). This model applies quicker and easier approximation in comparison to the Poisson-Boltzmann model employed in MM/PBSA, making it more ideal for high-throughput investigations or bigger systems while the MM/PBSA approach is more suitable for individual-target binding energy calculation (Sun et al., 2014; X. Zhang et al., 2017). The calculations of binding free energy between the protein and ligand used the following equation:

$$\Delta \mathrm{G}_{binding} = \mathrm{G}_{complex} - G_{protein} - G_{ligand} \qquad \text{(Eq. 5.1)}$$

where the free energy of each component is calculated by Equation 4.2 below:

$$G = E_{int} + E_{ele} + E_{vdw} + G_{pol} + G_{np} - TS \qquad \text{(Eq. 5.2)}$$

E_{int}, E_{ele}, and E_{vdw} represent the molecular-mechanics of internal, electrostatic, and van der Waals energies, respectively, while the polar and non-polar solvation energies are denoted by G_{pol} and G_{np}. Lastly, T and S are the absolute temperature and entropy (Godschalk et al., 2013; Kumari et al., 2014).

In this study, only the overall binding free energies were evaluated, and the lowest binding free energy estimated by MM/GBSA method shown in Figure 5.13 belongs to the HE4-H256 system after the method was employed to the final 50 ns of the MD simulation course where the RMSD and Rg plots were stabilised. This system maintained negative binding free energies between -10 kcal/mol to -40 kcal/mol, indicating that the H256 hairpin bound tightly to HE4 protein, compared to the other two hairpins, H16 and H101. Negative and lower binding free energies indicate the overall stability of the system (Ganjali Koli et al., 2023; Swain et al., 2018). The binding of the molecules within HE4-H16 and HE4-H101 complexes were unstable and not as strong as HE4-H256 due to the binding free energies over 50 ns in Figure 5.13a and 5.13b experiencing fluctuations between positive and negative energy values. This last analysis of the binding energies strongly supported H256 hairpin as the most promising DNA aptamer hairpin with great binding affinity against HE4 protein, alongside the outcomes of other analyses that have been comprehensively discussed in this current chapter. The confirmation of the strong binding of H256 and HE4 was tested *via in vitro* assay in the next chapter.

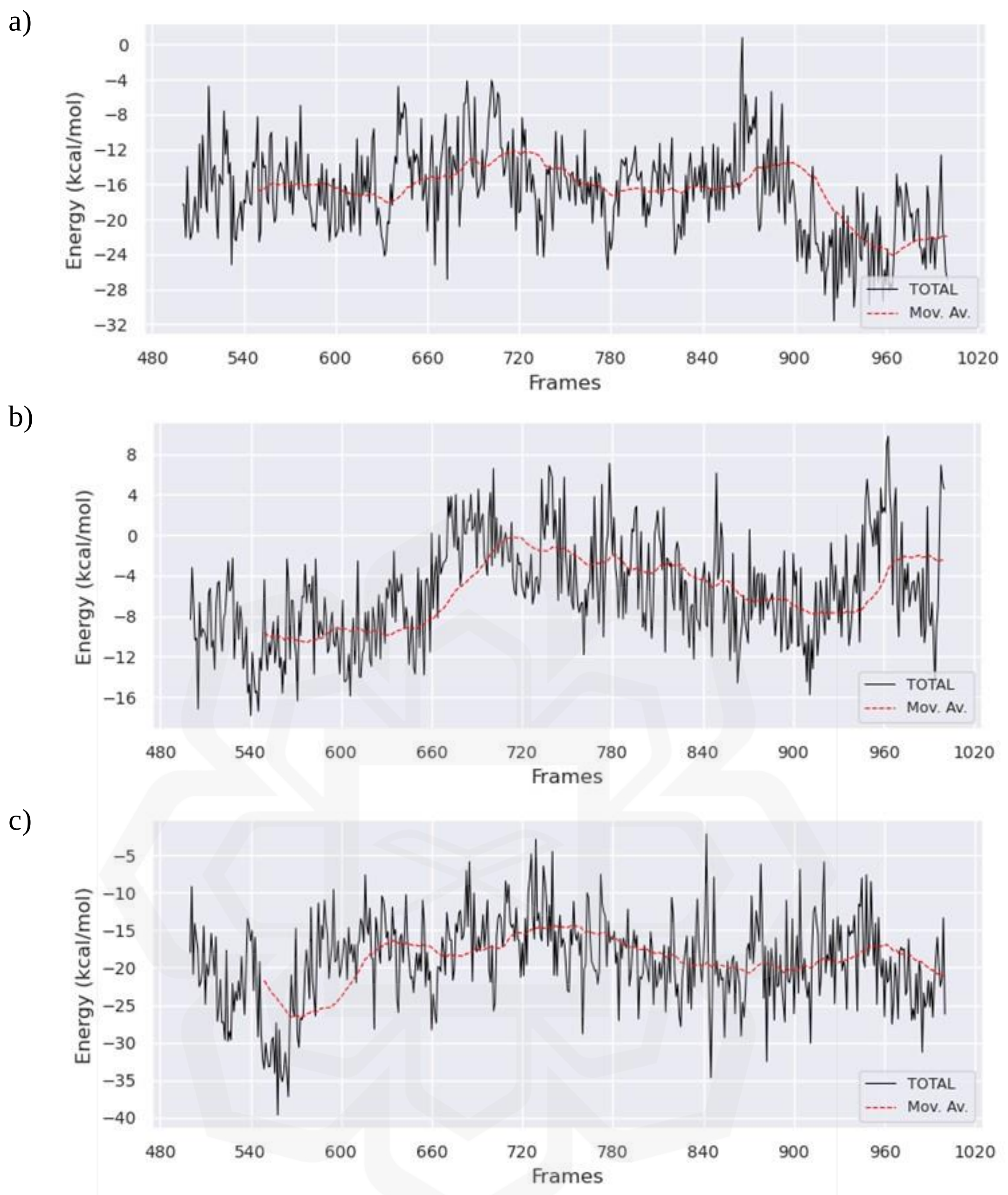


Figure 5.13 The binding free energies of the protein-hairpin complexes between 50 to 100 ns of MD simulations; (a) HE4-H16, (b) HE4-H101, and (c) HE4-H256.

5.4 CONCLUSION

AutoDock Vina molecular docking and GROMACS molecular dynamics simulations were employed to elucidate the interaction mechanisms between the HE4 protein and DNA aptamer hairpin structures. The work began with the truncation of a long 25-mer A4 aptamer into a 6-mer hairpin, followed by stem extension, forming an 8-mer hairpin.

Next, the nucleotides at the single-stranded loop of this 5’-GCGCAAGC-3’ hairpin were mutated, resulting in 256 hairpins where each of them was docked with HE4 protein using Vina. The binding poses of the 10 lowest binding energies out of 256 HE4-hairpin complexes were thoroughly examined and the three most potential complexes were selected: HE4-H16, HE4-H101, and HE4-H256. Subsequently, 100 ns MD simulations using GROMACS were performed to investigate the stability and dynamics of each complex, analysing different parameters including RMSD, Rg, RMSF, minimum distance, number of hydrogen bonds, hydrogen bond occupancy, and the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) free binding energy. The RMSD analysis was similar for all complexes, exception for HE4-H16 which experienced significant fluctuations before 40 ns. The RMSD values showed an initial increase in all complexes, indicating conformational adjustments at the beginning, but reached a plateau phase, suggesting that the complex achieved a stable conformation.

The Rg results for HE4-H16, HE4-H101, and HE4-H256 complexes were in agreement with the RMSD findings, showing a slight initial fluctuation before maintaining relatively steady values. The constant Rg values after 50 ns of simulation indicated that the overall folding and compactness of the complex remained stable. Next, the RMSF analysis provided insights into the dynamic behaviour of individual residues in the complex, where the regions with higher RMSF values indicated more flexibility, which is essential for accommodating structural adjustments during HE4-hairpin binding. The residues, specifically CYS49, VAL50, and LYS60 to GLY65, played a crucial role in facilitating the binding which was proven by the H-bonds formed by HE4-H101 and HE4-H256 complexes. The minimum distance between HE4 and the hairpin molecules was monitored where the consistently low minimum distance values proposed close and stable interactions. Both HE4-H101 and HE4-H256 maintained low values at less than 2 Å. But the HE4-H256 complex showed remarkably lowest distance over the simulation time with the highest number of H-bonds, compared to the other two complexes. The number of H-bonds and their occupancy rates were key indicators of the interaction strength, where a high number of H-bonds, coupled with high H-bond occupancy, demonstrated strong and specific interactions. The residues involved were mostly located at the HE4 active binding site predicted by the CAVER 3.0 tool, alongside the neighbouring residues. Based on the H-bond occupancy, H256 maintained strong binding with HE4 for 73 % of the simulation time. Finally, the MM/GBSA binding energy calculation provided a quantitative measure of the binding affinity between HE4 and the DNA aptamer hairpin, where HE4-H256 was the only system that maintained negative binding energies, and strong and spontaneous interactions.

In conclusion, the comprehensive analysis of the HE4-H16, HE4-101, and HE4-H256 complexes through molecular docking and MD simulations provided valuable insights into the interaction dynamics between the HE4 protein and the DNA aptamer hairpin. Following all the extensive analysis, the H256 hairpin is deduced as the most promising DNA aptamer hairpin against its target HE4 protein, the ovarian cancer biomarker. This H256 DNA aptamer hairpin was synthesised and an *in vitro* assay was conducted using an actual HE4 protein sample, which is discussed in the next chapter.

CHAPTER SIX

IN VITRO DETECTION OF HE4 BIOMARKER USING THE DESIGNED DNA APTAMER HAIRPIN, AND GOLD NANOPARTICLE-DNA CONJUGATES FOR FUTURE CANCER DIAGNOSTICS

6.1 INTRODUCTION

The interactions between proteins and nucleic acids play a crucial role in the DNA aptamer-based diagnostic sector, offering important insights into the molecular mechanisms that underlie different illnesses and facilitating the creation of precise diagnostic instruments (Zhang et al., 2019). These interactions, involving the binding of proteins to DNA or RNA, have a pivotal function in cellular activities such as gene expression, replication, and RNA processing. The specificity and affinity of these interactions are crucial for their effectiveness in diagnostics. Methods such as Electrophoretic Mobility Shift Assays (EMSAs), Chromatin Immunoprecipitation (ChIP), DNA Footprinting, and Fluorescence Resonance Energy Transfer (FRET) are employed to investigate these interactions, aiding in the identification of precise binding patterns (Chen & Goller, 2020; Ferraz et al., 2021; Sui & Imamichi, 2023; Yan & Huang, 2023).

The EMSA, commonly known as the gel shift assay, is a flexible and extensively employed tool in molecular biology that investigates protein-nucleic acid interactions (Seo et al., 2019). This approach is highly valuable for the identification and characterisation of protein binding to specific DNA sequences. The underlying concept of EMSA is straightforward where the phenomenon relies on the fact that the interaction between a protein and a DNA fragment causes a change in the fragment's mobility during electrophoresis (Holden & Tacon, 2011). Usually, the DNA is subjected to incubation with the target protein which is the HE4 in this study. The presence of the protein binding to the DNA alters the characteristics of the DNA-protein complex, such as its size and charge, which in turn affects its mobility when subjected to non-denaturing gel electrophoresis, commonly using polyacrylamide or agarose gels (Ream et al., 2016). The protein in the protein-DNA complex limits the mobility of its bound

DNA, while the free unbound DNA migrates faster through the gel. EMSA offers a notable benefit in terms of its high sensitivity and specificity. EMSA is highly versatile as this technique is applicable for investigating a broad spectrum of protein-DNA interactions, encompassing the binding of transcription factors to promoter regions as well as the interaction between nucleosomes and genomic DNA (Chen & Goller, 2020; Daras et al., 2019). A study demonstrated using agarose gel EMSA yielded high band resolution in a short running time in the association of the siRNA-binding protein p19 with the nucleic acids of its target RNA (Ream et al., 2016).

The capacity to recognise and measure protein-aptamer interactions has significant implications for the early diagnosis of diseases, monitoring the progress of diseases, and assessing the effectiveness of treatments, establishing it as a fundamental aspect of diagnostic approaches (N. Zhang et al., 2021). For instance, aptamers have been successfully applied to the screening of tumour-associated biomarkers such as human epidermal growth factor receptor 2 (HER2), prostate-specific antigen (PSA), and carcinoembryonic antigen (CEA) (Yang et al., 2015; F. Zhang et al., 2015). Additionally, DNA aptamers were utilised to identify specific targets, including polychlorinated biphenyls (PCBs) (Mehta et al., 2012) and gastrointestinal stromal tumours (GIST) (Banerjee et al., 2020). One recent study identified aptamer candidates using Graphene Oxide-Systematic Evolution of Ligands by Exponential Enrichment (GO-Cell-SELEX) to successfully detect various strains of SARS-CoV-2 virus (Moshref et al., 2023). Moreover, DNA aptamer-based diagnostic approaches that have been developed in the past 10 years include the production of sensors (electrochemical, surface plasmon resonance (SPR)-based, and fluorescent-colourimetry sensors), lateral flow assays (LFA), and nanoparticle assays. They successfully detected various proteins, related to different types of illnesses, such as small cell lung cancer (SCLC) biomarker (Sun et al., 2021), HER2 biomarker (Ranganathan et al., 2020), penicillin-binding protein (PBP2a) (Fan et al., 2020), and vascular endothelial growth factor (VEGF) protein (Stawicki et al., 2021).

Nanoparticle-based assays have been utilised extensively in clinical diagnostics, in example, the detection of superoxidase dismutase (SOD) cancer biomarker by using gold nanoparticles (GNPs) coated with SOD aptamer 4 (Dekhili et al., 2020). The use of aptamer and GNP-based detection techniques is an advanced and innovative strategy in the realm of disease diagnostics as these approaches exploit the distinctive optical characteristics of GNPs and the selectivity of aptamers. DNA aptamers, which are artificial short sequences of nucleotides, are commonly designed to specifically attach to particular molecules, such as proteins, peptides, or even cells that indicate the presence of diseases (Kaur et al., 2018), while GNPs are recognised for their unique optical characteristics, including their SPR, which resulted in a noticeable alteration in colour under specific conditions (Jazayeri et al., 2018; Khan et al., 2019; Nooranian et al., 2022). GNPs offer several advantages over other nanoparticles such as silver in biomedical fields like diagnostics. The chemical stability of GNPs is better than silver NPs where it shows good resistance to oxidation, enhances the durability in biological environments (Stein et al., 2021). GNPs also exhibit good biocompatibility, providing lower risks of toxicity, and their surface chemistry is easily modifiable, facilitating the functionalisation with other biomolecules (Sztandera, et al., 2018).

In diagnostic, the detection approach involves the conjugation of aptamers to GNPs, followed by the aptamers directing the GNPs towards their target molecules, which may be found in blood, saliva, or other bodily fluids. When the target molecule is not present, the aptamers prevent the GNPs from aggregating together due to the presence of salt, resulting in a red solution. Nevertheless, in the presence of the target molecules, the aptamers bind to them, resulting in the GNPs losing their protective layer and clustering together in high ionic strength solutions and resulting in a noticeable colour shift from red to blue or purple (Akshaya et al., 2020; Dekhili et al., 2020). This change serves as an indication that the disease marker is present. This method is extremely responsive and capable of detecting different levels of biomarkers. Furthermore, it enables quick and easily comprehensible outcomes without the requirement of advanced equipment, rendering it appropriate for on-site testing and application in the field, making it a suitable candidate for the development of point-of-care-testing (POCT) kit (Larsson et al., 2015).

In this study, the affinity of the designed hairpin, H256 to HE4 protein was characterised and compared to the affinity of HE4-A4 in which the A4 aptamer is a positive control, obtained from the SELEX selection (Eaton et al., 2015) and evaluated thoroughly in the Chapter 4. These complexes were screened through an *in vitro* study using EMSA where the samples of equal amounts were loaded into the wells on the gel and the intensity of the DNA bands (bound and unbound) were relatively measured. According to the computational studies conducted earlier, the binding energies calculated by AutoDock Vina for HE4-A4 and HE4-H256 complexes were -6.0 kcal/mol and -11.2 kcal/mol, respectively. The lower energy indicated that the affinity of the designed hairpin H256 to HE4 protein was higher, making the intensity of the bound H256 band would have been expected to be higher in the agarose gel. Additionally, a preliminary study to conjugate GNP to the designed hairpin H256 was conducted, involving the formulation and characterisations (particle size, zeta potential, and functional group analyses) of the GNP-H256 solution.

6.2 METHODOLOGY

6.2.1 Electrophoretic Mobility Shift Assay (EMSA)

The implementation of EMSA offers a cost-effective, expeditious, and dependable approach for quantifying apparent dissociation constants concerning protein-ligand binding affinity. This study utilised agarose-based EMSA to verify the bioconjugation between the HE4 protein and the DNA aptamer hairpin designed computationally in the Chapter 5. The 1.5% agarose gel solution was firstly prepared by dissolving 0.75 g agarose (molecular biology grade) powder (Vivantis Technologies Sdn. Bhd.) in 50 mL of pH7, 1x Tris-acetate-EDTA (TAE) buffer, heating in microwave at low heat until all particles are completely dissolved. The solution was cooled down until safe to touch and 2x SYBR-safe DNA stain (Invitrogen, USA) was added. Then, the agarose solution was poured into the gel cast, where the comb was placed earlier. The cast was put aside until the gel sets.

Next, the lyophilised oligonucleotides of the designed hairpin, H256, and positive control aptamer, A4 (IDT Integrated Pte. Ltd.) were dissolved in pH7, 1x Tris-EDTA (TE) buffer to 500 µM. The oligonucleotides were heated to 94°C for 2 minutes, to denature or separate the double-stranded DNA into single strands, before they reformed the double-stranded structures and formed proper DNA folding, as they cooled down to 25 °C. Once they cooled down, both H256 and A4 were incubated with recombinant human HE4 protein (Elabscience, USA) in pH 7, 1x phosphate buffer saline (PBS) buffer for 24 hours at 37°C on a rotational shaker. Loading dye was added to each sample. The samples were loaded into the wells of the prepared agarose gel that was submerged in pH 7, 1x TAE buffer, and electrophoresis was conducted at 100 V for 30 minutes. The amount of each component is described in Figure 6.1. Lastly, the gel bands were visualised under ultraviolet (UV) light using a VILBER E-box DNA gel stain imager, and the band intensities were measured by GelAnalyzer software. The experiment was repeated only once, due to limitation of materials.

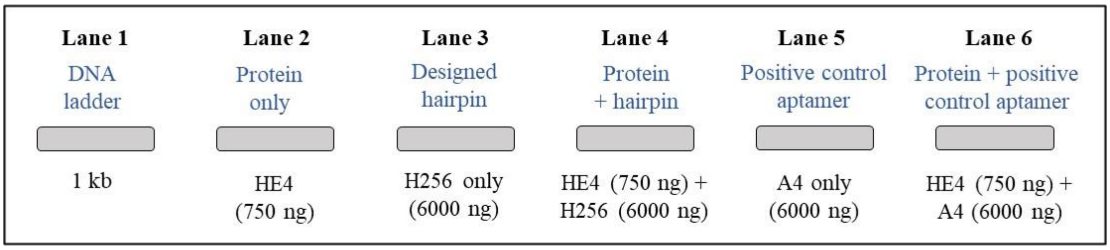


Figure 6.1 The samples loaded to each well of the agarose gel for EMSA

6.2.2 Preliminary Studies of Detection Method Development

The conjugation of GNP-DNA hairpin follows the work flow established by Dekhili *et al.* (2020). The formulation of GNP-H256 solution began with the modification of H256 hairpin with an amino group ($-NH_2$) that was added to the 5'-end terminal, forming an aminated-H256 hairpin (AH256). This was modified and supplied by IDT Integrated Pte. Ltd., Singapore. 40 nm carboxylated-gold nanospheres (GNP-COOH) with optical density (OD) of 20 in water were purchased from NanoComposix Inc., USA while 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and N-hydroxysuccinimide (NHS) were obtained from Tokyo Chemical Inc. and Nacalai Tesque Inc., Japan, respectively. Firstly, the OD 20 GNP-COOH, known as Raw GNP in this study, was diluted to OD 5 with deionised water, followed by the preparation of AH256 aptamer. The AH256 was dissolved to 100 µM according to the specification sheet provided by the supplier with TE buffer (pH 7.4) that was prepared earlier by mixing tris(hydroxymethyl)aminomethane (Tris) and ethylenediaminetetraacetic acid (EDTA). TE buffer was used to protect the DNA from degradation. This 100 µM AH256 was diluted to 10 µM in 1X, pH 7.4 phosphate-buffered saline (PBS) buffer. The diluted DNA was heated to 94 °C for 2 minutes to denature any double-stranded formations. The solution was then left to cool to 25 °C, allowing the DNA to refold to its native structure, following a previous study (Alnaimi et al., 2022).

In a 1.5 mL microcentrifuge tube, 1.2 mL of OD5 Raw GNP solution was added, then immediately followed by 10 µL of 10 mg/ml EDC and 10 µL of 10 mg/ml NHS that was prepared right before their additions. Next, the mixture was vortexed and incubated at 25 °C for 30 minutes to react with the EDC/NHS with the GNPs. After incubation, the solution was centrifuged at 8000 rpm for 10 minutes, and the supernatant consisting of excess EDC and NHS was removed. Resuspended in 1.2 mL of PBS, the pellet was vortexed and sonicated in a bath for 10 minutes at 25 °C. To initiate aptamer binding, 80 µL of the 10 µM AH256 was added to this GNP solution and incubated on a rocker at 25 °C for 1 hour. Upon completion, 20 µL of 10% bovine serum albumin (BSA) was added as a blocking agent and incubated for 10 minutes at 25 °C. Another 10-minute, 8000-rpm centrifugation followed. The pellet was resuspended in 1.2 mL PBS after discarding the supernatant. The resuspension procedure required extensive vortexing and 10-minute sonication. Next, the solution was examined for the colour change, which would indicate effective GNP-AH256 conjugation, followed by the analyses of its particle size and zeta potential using Zetasizer Nano Series Model 1600 (Malvern, United Kingdom) and Zetasizer Nano Series Model 2600 (Malvern, United Kingdom). Finally, the identification of functional groups was achieved by Spectrum Two Fourier Transform Infrared (FTIR) spectrometer (PerkinElmer, United States) and handheld Raman spectrometer (Bruker, Germany).

6.3 RESULTS AND DISCUSSION

6.3.1 Electrophoretic Mobility Shift Assay (EMSA)

The verification of the binding of HE4-H256 and HE4-A4 was conducted using an *in vitro* assay, agarose gel-based EMSA, where it separates the protein-DNA complex from the free DNAs by electrophoresis (Ream et al., 2019). The agarose gel was chosen over polyacrylamide gel as the agarose gel required the usage of standard molecular biology reagents, equipment, and a buffer system with low conductivity, making the whole process simpler and faster than the traditional polyacrylamide procedures (Ream et al., 2016). The polyacrylamide gel electrophoresis involves a buffer system which maintains the pH to ensure the unfolding and separation of proteins are correct. Whereas, the milder conditions of agarose gel electrophoresis are less likely to adjust the pH and disrupts the protein-DNA complexes. The staining of the DNA bands on electrophoresis gels using ethidium bromide (EtBr) has been widely utilised for many decades, but it is reportedly to be toxic and mutagenic in humans (Singh & Singh, 2018). Due to this, the SYBR Safe DNA gel stain was selected in this study as it is suitable and safe to be used as a pre-casting dye where the dye was introduced to the gel solution before it was poured into the cast (Bawane et al., 2023).

As shown in Figure 6.2a, the unbound DNA bands were seen near the middle region of the gel, for all loaded samples. However, the intensities of the unbound DNA bands for HE4-H256 (lane 4) and HE4-A4 (lane 6) were lower compared to their controls, the H256 (lane 3) and A4 aptamer (lane 5), respectively. As all wells were initially loaded with the same amount of DNA (H256 hairpin or A4 aptamer), these lower intensities indicated that some of the DNA molecules were successfully bound to its target protein, HE4. Unfortunately, the bands of the bound DNA were not observed in Figure 6.2a which was expected to be at the top part of the gel, just below the wells. This was solved by adjusting the contrast of the gel image (Figure 6.2b) and inverting the image colour (Figure 6.2c), where the bound DNA bands are seen as labelled in the figures. The electrophoretic mobility of the HE4-DNA complexes was less due to the large size of complex molecules, making them migrate slower than the small, free, and unbound DNA molecules (Hellman & Fried, 2007; Seo et al.,2019). Through these gel images, both H256 hairpin and A4 aptamer were proven to have a good affinity towards HE4 protein, however, this *in vitro* study does not provide quantification analysis. Nevertheless, the relative binding affinity of HE4-H256 and HE4-A4 were evaluated through the band intensities which are demonstrated and listed in Figure 6.3 and Table 6.1.

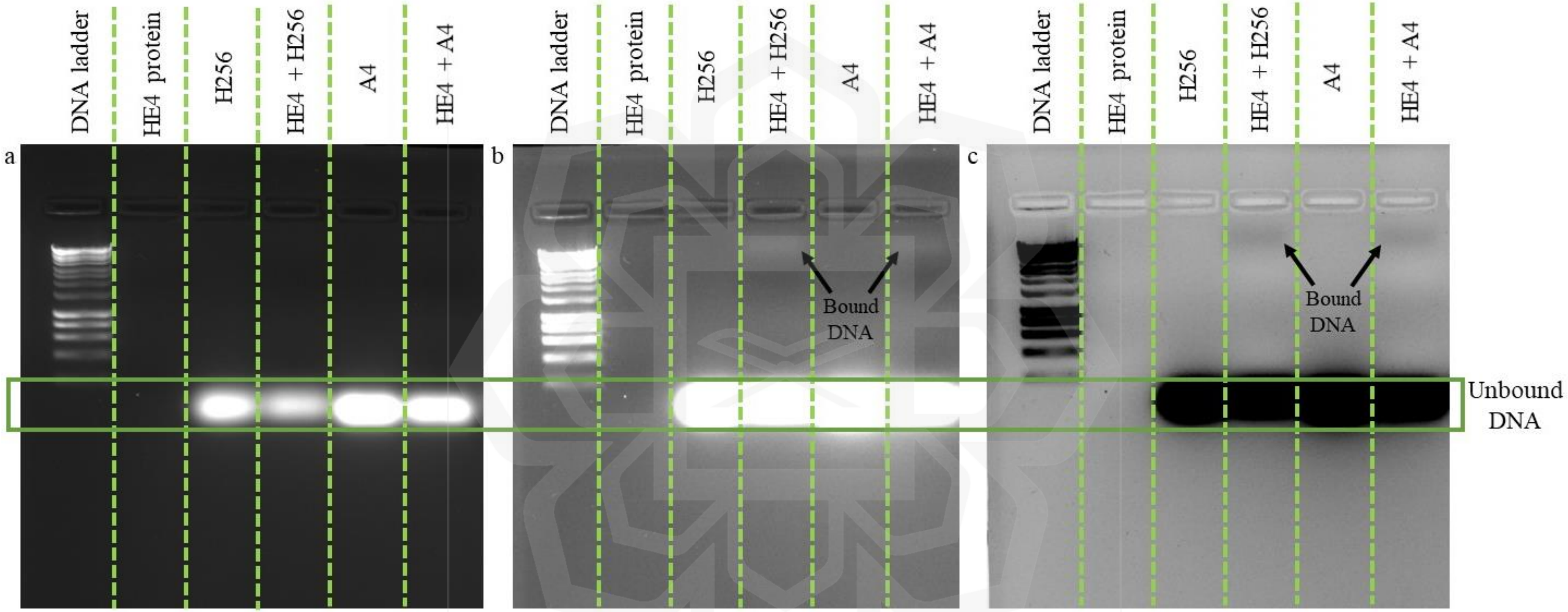


Figure 6.2 EMSA of the designed DNA hairpin, H256, and the positive control, A4 aptamer against HE4 protein. (a) Raw gel image showing the DNA bands, (b) Gel image after adjusting the contrast, and (c) Gel image after inverting the image and adjusting the contrast. The bands of the bound DNA were observed in both adjusted gel images.

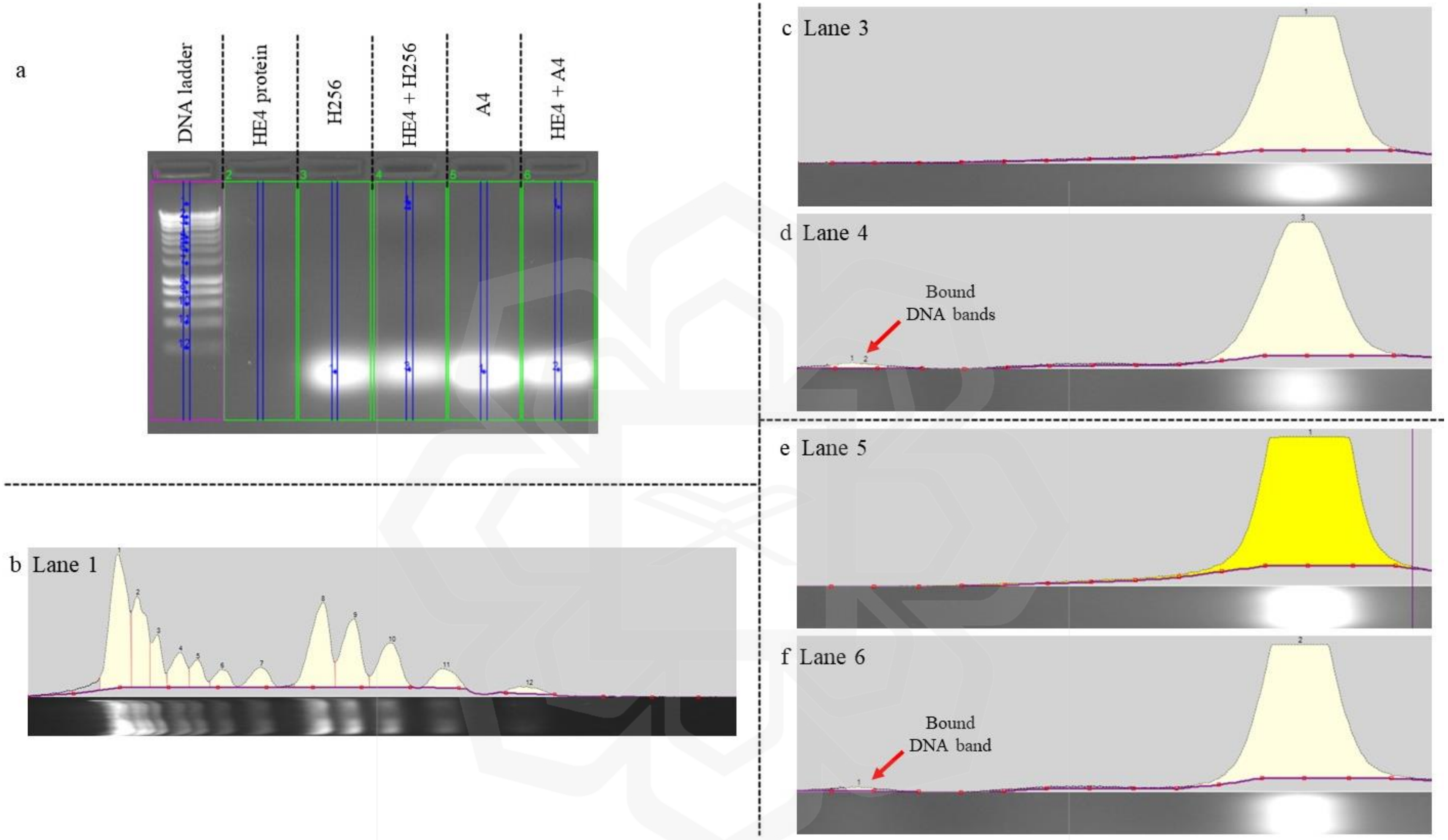


Figure 6.3 The band peaks of each lane, observed after 30 minutes of electrophoresis at 100 V. (a) The agarose gel, illustrating the lanes and band peaks, where different band intensities were obtained, (b) 12 peaks detected for the lane loaded with 1 kb DNA ladder, (c) Only one band seen for H256 (control), (d) three bands for sample where H256 was incubated with HE4 for 24 hours, (e) Only one band displayed by A4 aptamer (control), and (f) two bands for sample where A4 aptamer was incubated with HE4 for 24 hours.

Based on Figure 6.3a, the blue dots within the blue lines represent the DNA bands identified by the GelAnalyzer software. This analysis revealed that Lane 1 which consisted of the 1kb DNA ladder had 12 DNA bands. In contrast, Lane 2 which contained only the HE4 protein, showed no bands, indicating the absence of nucleic acids. The H256 control (Lane 3), HE4-H256 (Lane 4), A4 aptamer (Lane 5), and HE4-A4 (Lane 6) displayed one, three, one, and two bands, respectively. The single band observed in Lane 3 exclusively represented the unbound H256 hairpin, as depicted in Figure 6.3c. Similarly, the band in Lane 5, shown in Figure 6.3e, corresponded solely to the A4 aptamer. These two lanes acted as the controls for the HE4-H256 and HE4-A4 complexes as only the H256 and A4 raw samples were loaded. Nevertheless, Figures 6.3d and 6.3f exhibit additional bands at the upper part of the agarose gel, representative of the DNA bound to the target protein in the HE4-H256 and HE4-A4 complexes, respectively. Specifically, the HE4-H256 sample (Lane 4) features an extra band, with a total of three bands, in comparison to Lane 6, which contains the HE4-A4 sample. The area of each DNA band was measured by GelAnalyzer software and recorded in Table 6.1.

Table 6.1 The width, height, and area of each gel band for all samples. A larger area corresponds to a higher intensity and higher amount of DNA.

Lane	Sample	Peak	Band type	Width	Height	Area	Percentage (%)
2	HE4 protein	-	-	-	-	-	-
3	H256 hairpin	1	Unbound DNA	124	207	12728	100
4	HE4 + H256	1	Bound DNA	21	8	133	1.25
		2		36	7	218	2.02
		3	Unbound DNA	123	207	10714	96.73
5	A4 aptamer	1	Unbound DNA	377	195	13825	100
6	HE4 + A4 (positive control)	1	Bound DNA	19	6	97	0.84
		2	Unbound DNA	134	206	12153	99.16

In Lane 4, consisting of the HE4-H256 sample, the band representing bound H256 accounted for 3.27% of the total, with the remainder appearing towards the bottom part of the gel, indicating it was an unbound H256 hairpin. The band intensity of the DNA that bound to HE4 was four times more intense compared to the band displayed by the HE4-A4 sample (0.84%). Here, the H256 hairpin demonstrated a greater binding affinity towards HE4 protein, an improvement from the A4 aptamer, which was the aptamer candidate selected *via* SELEX from the previous study by Eaton et al. (2015). The H256 hairpin was a mutation from the A4 aptamer truncated hairpin that was computationally designed in the Chapter 5. This analysis substantiates the reliability of the computational studies conducted earlier, as discussed in previous chapters which included structure predictions using Mfold and RNAComposer servers, molecular docking by AutoDock Vina, and molecular dynamics simulations employing GROMACS.

6.3.2 Characterisation of GNP-Aptamer Formulation

The successful binding of the designed hairpin, H256, with the HE4 protein opens numerous possibilities for its development as a detection technique for this ovarian cancer biomarker. Therefore, this study undertook an initial exploration into formulating a potential detection solution using gold nanoparticles (GNPs), given their strong suitability for colourimetric assays (Jans & Huo, 2012). The unique properties, ease of functionalisation, and biocompatibility have made them a valuable tool for the development of sensitive and selective diagnostic assays, including the diagnosis of head and neck upper aerodigestive tract cancer (Andrade & Costa, 2023), detection of carbendazim in agriproducts (Mao et al., 2022), and breast cancer biomarker HER2 (Ranganathan et al., 2020). In a previous study mentioned earlier, the SOD marker that can be found in body fluids was detected using GNPs coated with SOD aptamer 4, showing colour changes of the formulations (Dekhili et al., 2020). The colour of the GNP formulations varied, ranging from red to blue-purple, depending on the methods used and the agglomeration of the GNP particles (Aldewachi et al., 2018; Alizadeh & Nazari, 2020). In this study, the 40 nm raw GNP solution appeared to be red, while it turned purple after the AH256 DNA aptamer was incorporated into the GNP solution (Figure 6.4). According to a work reported in 2018, the colour of the reaction changed from red to purple after the DNA aptamer was introduced to the GNP and incubated for 10 minutes (He & Yang, 2018). This colour change was caused by the DNA-induced agglomeration of the gold particles which was proven by a shift in the surface plasmon band from 524 to 556 nm.

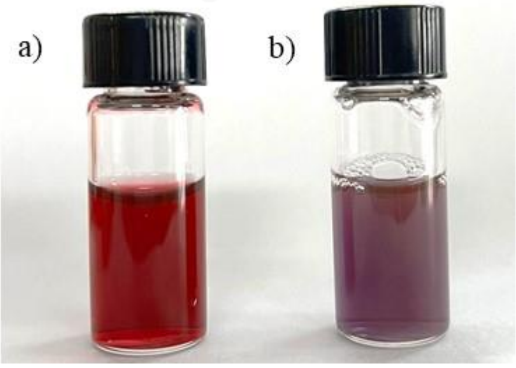


Figure 6.4 The (a) 40 nm raw gold nanoparticle and (b) GNP-AH256 solution.

The particle size of GNPs influences the colour of the solutions, and this phenomenon is deeply based on the principles of nanoscale physics and SPR. SPR is the resonant oscillation of conduction electrons at the surface of the nanoparticles stimulated by incident light, which depends on the particles' size, shape, and the dielectric environment around them (Amendola et al., 2017). Larger nanoparticles have broader SPR peaks, indicating a wider range of absorbed wavelengths, which alters the solution colour. The specific interplay of particle size, SPR, and the optical properties of gold nanoparticles makes them valuable in numerous applications, including sensing, imaging, and medicine, where colour changes can be used as indicators of nanoparticle behaviour or interactions (Elahi et al., 2018; Wu et al., 2019). It was reported earlier that a GNP-aptamer-based aggregation assay for the detection of interleukin-6 (IL6) marker changed the red colour of the GNP-aptamer to purple-blue hue after incubation of the formulation with samples containing IL6 at room temperature (Giorgi-Coll et al., 2019). The shift to a purple colour is due to an increase in particle size, as previously reported where GNPs with a particle size ranging from 30 to 80 nm display a red colour, while larger particles, such as 90 nm, resulting in a reddish-purple hue (Alex & Tiwari, 2015). Thus, the emergence of the purple colour in the GNP solution of this current study is possibly caused by the effective conjugation of the AH256 aptamer to the surface of the gold particles, resulting from the interaction between the carboxyl (-COOH) group of the GNP and the amino group of the AH256, forming a molecule covalently bonded by an amide bond that causes the particle size increment (Figure 6.5).

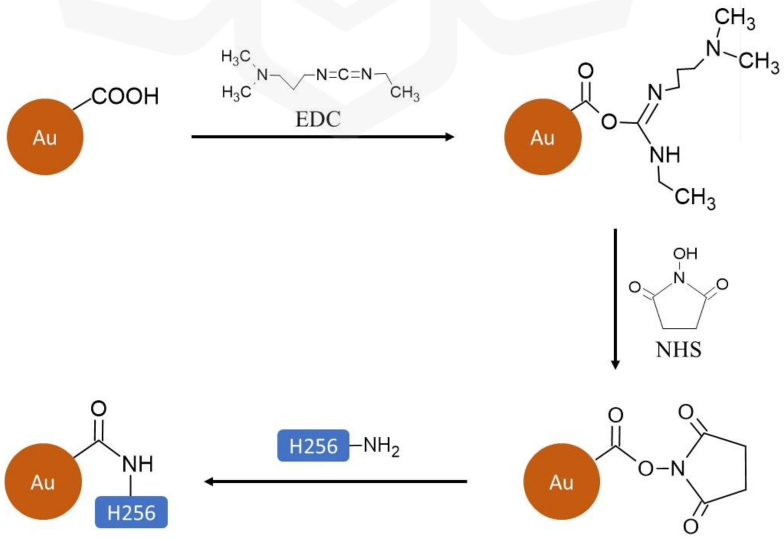


Figure 6.5 EDC (carbodiimide) crosslinking reaction. The carboxylated GNP formed an amide bond with the designed hairpin, aminated-H256 ($H256-NH_2$).

The EDC/NHS coupling reaction is commonly employed in bioconjugation chemistry to covalently link biomolecules to surfaces like gold nanoparticles as demonstrated in many previous studies (Busch et al., 2019; Jazayeri et al., 2016; Zhang et al., 2017). For the process, the EDC activates GNP carboxyl (-COOH) groups to generate reactive O-acylisourea intermediates. This intermediate hydrolyses easily, reducing amine coupling efficiency, hence, it was stabilised by adding NHS, which interacts with O-acylisourea to generate a stable NHS ester. This activated ester reacts with primary amine groups on the aminated-H256 to produce a stable amide bond and covalently bond to GNP. Functionalised GNPs with aptamers may precisely attach to target molecules, making them useful in nanotechnology and molecular diagnostics. Figure 6.6 shows the infrared spectrum of the conjugation between GNP and AH256 that was obtained by the Fourier transform infrared (FTIR) instrumentation.

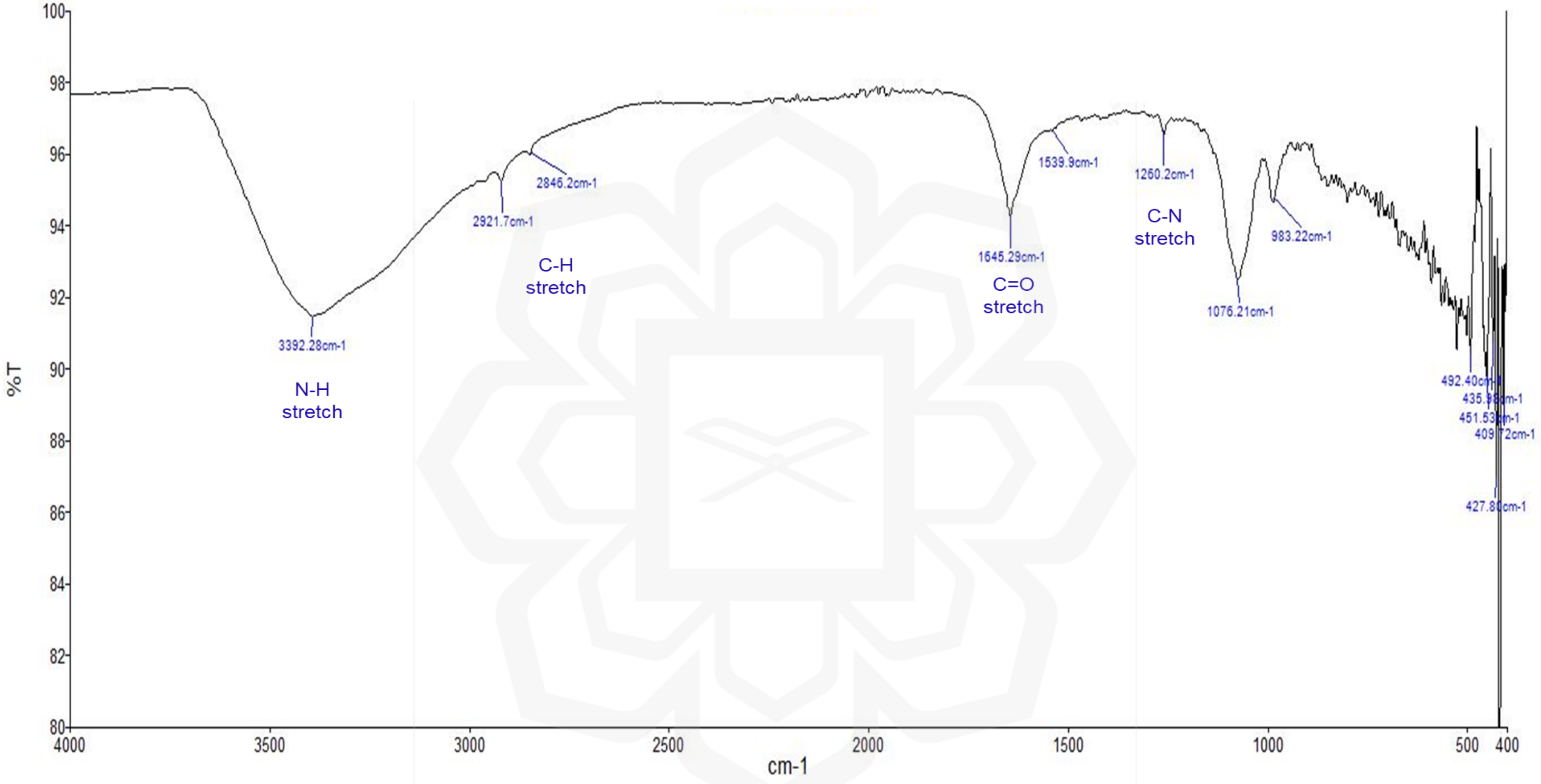


Figure 6.6 Infrared spectrum of GNP-AH256 conjugation.

Unfortunately, the quantity of the dried sample yielded from the freeze-drying process of the GNP-AH256 solution was extremely small, making the amount for FTIR sampling limited in this study. The typical pattern of peaks for amides is the N-H stretching peaks and a C=O stretch available at 3370 to 3170 and 1680 to 1630 cm^{-1}, respectively (Smith, 2020). In the spectrum (Figure 6.6), the C=O stretch was shown at 1645 cm$^{-1,}$ and the N-H stretch of the amide was at 3392 cm^{-1}. Here, the formation of an amide bond was detected, however, more analysis should be done to confirm this conjugation as the spectrum shows low signal strength in its peaks and several bands or stretches are missing. The deformation of samples and insufficient samples can impede the precise and exhaustive examination of samples by FTIR.

FTIR and Raman spectroscopy are complementary techniques to analyse the molecules vibrational modes, offering important insights into their molecular composition and structure, including the identification of functional groups (Barnas et al., 2020). Both techniques assess the interactions of light with the vibrations of the molecules. The utilisation of Raman spectroscopy in this study validated the presence of C=O stretch of amide at 1753 cm^{-1} for the conjugated GNP-AH256 sample (Figure 6.7b), alongside the C=O stretch of the carboxyl group in the raw GNP sample at 1642 cm^{-1} (Figure 6.7a). Metals are generally considered Raman-inactive because they do not have the vibrational modes like organic molecules. Specific chemical groups, like metal-carbonyl (metal-C=O), exhibit distinct peaks between 2000 and 2700 cm^{-1}, which is the Raman silent region, suggesting that metals typically do not exhibit strong signals in these regions (Zhao et al., 2003). The raw GNP sample, which was readily carboxyl-functionalised, displays a weak to moderate peak at 1995 cm^{-1}. Whereas, the peak of conjugated GNP-AH256 at the same region is more intense with the presence of DNA. These gold particles have a crucial role in amplifying the Raman signals of the associated DNA *via* the surface-enhanced Raman scattering (SERS) phenomenon, forming signals at approximately 2000 cm^{-1} (Barhoumi et al., 2008). Other than the GNP-DNA peak, the DNA presented peaks at 803 and 1156 cm^{-1}, similar to various other nucleic acids studies with two peaks at the same regions (Chandra et al., 2015; Vaverkova et al., 2014; X. Zhang et al., 2012). Furthermore, the purines (adenine and guanine) and pyrimidines (cytosine and thymine) of the H256 hairpin consist of five- and/or six- membered aromatic rings (Takahashi & Sumimoto, 2021), contributing to the formation of peak at 1481 cm^{-1} in Figure 6.7b.

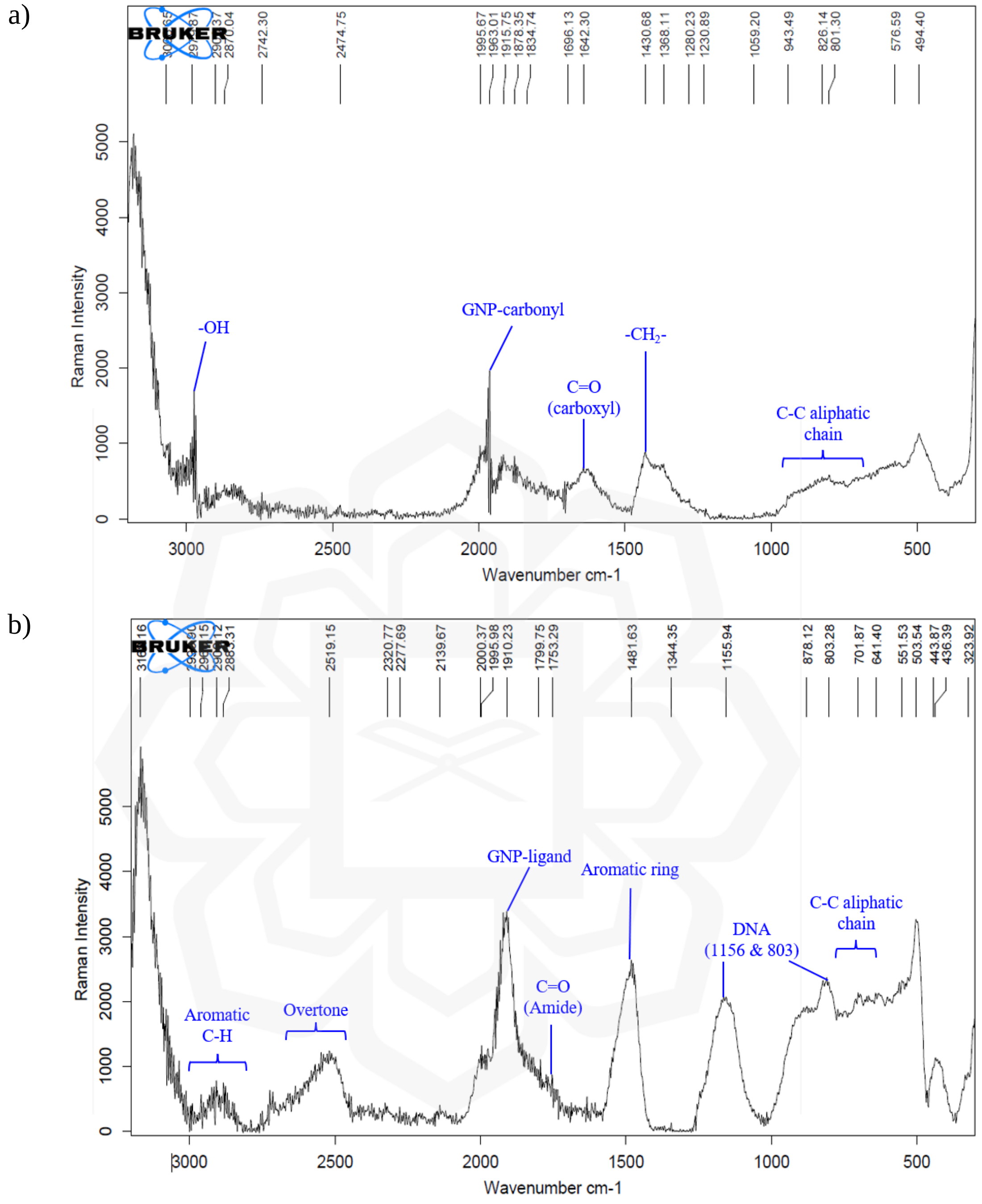


Figure 6.7 Raman spectra of (a) Raw GNP and (b) GNP-AH256.

Next, the raw GNP and conjugated GNP-AH256 samples were analysed for their solution colour, size, and surface charge measurement. Dynamic light scattering (DLS) technique was utilised in the instrumentation measuring the size of particles and the zeta potential which measures the surface charge that typically influences the stability of colloidal systems (Shah et al., 2014). This technique analyses the Brownian motion of particles, using laser light scattering and correlating the fluctuations in intensity to determine their hydrodynamic diameter. In a colourimetric assay detecting arboviruses, the formulated aptamer-GNP conjugates successfully bound to the mosquito salivary protein D7 and Zika virus envelope protein which visibly changed the solution colour from red to purple due to the aggregation of the particles while maintaining the negative zeta potential between -29.7 mV to -42.9 mV, caused by the electrostatic stability (Bosak et al., 2019). The initial particle size of the bare-GNP was approximately 30 nm and increased to 109 nm after conjugation with DNA aptamers. Similarly, the size of gold particles in this study increased from an average of 49 nm (raw GNP) to approximately 130 nm after the conjugation of aptamer through EDC/NHS coupling reaction, producing GNP-AH256 conjugate (Table 6.2). The observed increment in GNP particle size signifies the successful conjugation of the H256 aptamer onto the GNP surface. Furthermore, the GNP-AH256 solution was stable, following the zeta potential measurements where the particles were negatively charged, ranging from -33.8 mV to -38.6 mV (Carone et al., 2023; Vogel et al., 2017). Zeta potential measures the electrostatic magnitude between particles in a fluid, indicating the stability of colloidal dispersions where high zeta potential, either positive or negative, is electrically stable (Hunter, 1981).

Table 6.2 Average particle size, polydispersity index, and zeta potential of the raw GNP and GNP-AH256 solution.

Sample	Raw GNP			GNP-AH256		
Formulation	1	2	3	1	2	3
Particle size (nm)	48.75	48.75	49.78	129.8	126.8	133.0
Polydispersity index (PdI)	0.176	0.143	0.167	0.323	0.399	0.340
Zeta potential (mV)	-38.3	-38.9	-33.3	-33.8	-38.6	-34.8

The Polydispersity Index (PdI) serves as a quantitative indicator for analysing the size distribution homogeneity inside a colloidal or polymeric dispersion, delivering vital insights into the dispersion's uniformity and stability (Danaei et al., 2018). It is a dimensionless variable that ranges from 0 to 1 where the calculation involves the division of the weight-average molecular weight (Mw) with the number-average molecular weight (Mn). A PdI approaching zero signifies a monodisperse distribution, indicative of a sample wherein the constituent particles or molecules exhibit negligible variance in size. In contrast, a PdI close to a value of 1 suggests a polydisperse distribution, suggesting high size variation among the scattered particles. The PdI measurements in Table 6.2 display highly uniform particle distribution for raw GNP with values of less than 0.200. However, the value significantly increased to 0.399 after the conjugation with H256 aptamer, pointing to a wide range of particle size distribution. Solutions and mixtures with PDI value of less than 0.5 are recognised as homogenous Bennie et al., 2021). In an earlier report, colloidal citrate GNPs were conjugated with a 40-mer aptamer for the detection of fumonisin B1, a type of mycotoxin, measured with a PdI value of 0.199 ± 0.017 alongside a colour change from red to blue (Mirón-Mérida et al., 2021).

6.4 CONCLUSION

The main objective of this current chapter was to verify the binding affinity between HE4 protein with the most potential DNA aptamer hairpin, H256, which was comprehensively designed *via* computational approaches in earlier chapters. This H256 hairpin was derived from the substitutional mutation of the nucleotides of the A4 aptamer truncated hairpin, where this promising A4 aptamer was taken from the study conducted by Eaton et al. (2015) that consistently bound to the HE4 protein as it was the sequence most enriched in round 2 of the capillary electrophoresis SELEX. In comparison, the designed H256 hairpin bound to the HE4 protein better than the A4 aptamer as proven by the gel images obtained from the electrophoretic mobility shift assay. The relative binding affinity, referring to the bound DNA band intensities, between HE4 and the H256 hairpin (3.27%) was fourfold higher than that between HE4 and the A4 aptamer, which was only 0.84%. Furthermore, a preliminary study was undertaken that may serve as an initial framework for the future development of an ovarian cancer biomarker HE4 detection kit. The increment of particle size after the conjugation of H256 aptamer to the surface of GNPs was expected and aligned with many past studies. Moreover, the zeta potential maintained between -33 mV to -39 mV, signifying a good stability of the HE4-H256 formulation. The FTIR and Raman spectra confirmed the conjugation through the presence of amide group, arising from the reaction between the carboxyl group of the GNP and amine group of the aminated-H256 hairpin. While this investigation was limited to the analysis of particle size, polydispersity index, zeta potential, and functional group identification because of the material limitations, these parameters could be optimised in alignment with the requirements of future research endeavours, especially in diagnostic applications.

CHAPTER SEVEN

CONCLUSION AND RECOMMENDATIONS

7.1 GENERAL CONCLUSION

Diagnosing an early-stage ovarian cancer is crucial, as it improves survival rates, lowers risk of complications, and offers more treatment options. Moreover, the development of a promising, cost-effective alternative diagnostic approach will be an advantage over the current approaches available in the healthcare system. This thesis provides an in-depth analysis of designing potential DNA aptamer hairpin against OC biomarker, HE4. The study introduces a novel approach for the *in silico* design of DNA hairpin, applying multiple computational method to enhance the affinity towards the target.

Chapter Three modelled the 3D structure of HE4 protein using three web servers: AlphaFold, I-TASSER, and Robetta. Based on the validation approaches (PROCHECK and ERRAT) applied to the post-MD simulation models, RF1 displayed outstanding validation results and was deduced as the highest quality HE4 tertiary structure. This structure was selected to be utilised in the *in silico* studies of Chapter Four and Chapter Five. Furthermore, the tunnel and binding site of the protein was identified by CAVER 3.0 tools, showing the cavity size and the residues that made up the region.

Chapter Four focused on the process of designing secondary and tertiary structures of DNA aptamer using Mfold and RNAComposer, followed by the alterations of atoms, fulfilling the criteria of a DNA composition. Subsequently, the molecular docking of HE4 model with four HE4 aptamer candidates obtained from a previous study was conducted using AutoDock Vina. Following thorough examination of the interactions between HE4 protein and the DNA aptamer in the complexes, it is concluded that the A4 aptamer was the most suitable aptamer candidate against the OC biomarker.

The hairpin of A4 aptamer was truncated and modified through substitutional mutation of the loop nucleotides in Chapter 5, to achieve the objective of enhancing the DNA hairpin for better affinity against its target, HE4. This *in silico* modification of the

8-mer hairpin produced 256 different DNA sequences which were individually docked with HE4 using AutoDock Vina. H16, H101, and H256 hairpins were chosen as the three most suitable hairpins against HE4 with binding energies between -10.6 to -11.6 kcal/mol. Subsequently, MD simulations using GROMACS were conducted on HE4-H16, HE4-101, and HE4-H256 complexes and it is deduced that H256 hairpin was the most promising DNA hairpin to be synthesised and tested in *in vitro* assay. Computationally, it formed strong interactions, exhibiting highest number of hydrogen bonds between the loop nucleotides with the active binding site of HE4.

Finally, Chapter Six verified the binding between the designed H256 hairpin with HE4 protein through electrophoretic mobility shift assay, utilising agarose gel to capture the shifts of the DNA molecules. The computationally designed hairpin in this study have shown greater affinity towards HE4 protein, compared to the A4 aptamer that was selected through the SELEX study in previous research. This proves the potential of the novel *in silico* method developed in this thesis to replace or complement the existing SELEX method in screening the target marker. Ultimately, the main aim of this research has been successfully achieved, by fulfilling all the objectives stated in Chapter One. Additionally, a preliminary study was conducted, proposing a groundwork for future development of ovarian cancer marker detection tool.

7.2 LIMITATION OF STUDY

The development of detection method for HE4 protein in this thesis was not without limitation, as several of them were identified along the way. The challenges encountered during the computational work included the lack of HE4 aptamer information in literature search, constraints related to the computer workstation setup, and limited knowledge of the research team regarding aptamers, as this was the team’s first experience working with aptamers. Many studies have utilised the same HE4 aptamer candidate sequences obtained from the study by Eaton *et al.* as it was the most reliable and comprehensive study in the selection of HE4 aptamer. However, this study was done in 2015, hence, a newer discovery of aptamer against HE4 is needed as it could address the limitations of the original research.

Within the computational strategy, several elements demand attention, including the suitability of the methodologies used for modelling both the HE4 protein and particularly the DNA aptamer. The Ramachandran plot derived for the validation of protein structure provides a general overview regarding the sterically allowed regions for the backbone. However, it does not account for specific interactions, such as hydrogen bonding which at times allow residues to adopt conformations that are energetically unfavourable. Thus, employing a variety of validation methods is beneficial to assess the overall quality of predicted model.

Owing to the length of the computational research employed in this thesis, comprehensive experimental methods involving the use of real patient serum samples remain unexplored. This was another limitation in this study where the access to actual human samples was lacked for *in vitro* testing with the designed DNA hairpin, which would have made the findings more compelling. Nevertheless, the standard HE4 protein sample used in this study was adequate to verify the interaction between the DNA hairpin with the target protein, as the main objective of the research was to develop an *in silico* SELEX method for designing the DNA hairpin. Consequently, this demands the formulation of a new experimental framework for developing a ready-to-use ovarian cancer diagnostic technique for actual human samples.

7.3 FUTURE RECOMMENDATIONS

The work presented in this thesis is far from the creation of a fully developed ovarian cancer marker detection kit. Nevertheless, the thesis discusses the essential guidelines for designing the HE4 detection mechanism using computational approaches. For future studies, it is recommended to design aptamers with longer sequences with multiple hairpins as this can increase the selectivity and specificity of the aptamer towards its target. The presence of multiple hairpins in an aptamer increases the complexity of the structure, potentially allowing a more precise binding with the target molecule. This will to higher stability of the aptamer structure, as the hairpin stem regions consist of multiple complementary pairs of nucleotides. The helical structure formed by these pairing is more resistant to denaturation which explains the stability. Noteworthy, the molecular docking of the short DNA sequences in this thesis utilised AutoDock Vina, which this will not always be an optimal choice for every aptamer. The aptamer sequence length can significantly influence the suitability of molecular docking method selection. Thus, it is suggested to further investigate other docking alternatives, leading to more accurate and reliable docking results.

Based on the findings discussed in this work, the future work should emphasise on the development of the HE4 detection kit. As previously mentioned in the thesis, HE4 level is closely related to the diagnosis of OC. However, this detection kit should be paired with another diagnostic method, such as imaging scans to further confirm the presence of the illness. The kit can either be in the form of colourimetric assay, or biosensor such as electrochemical sensor. Partnering with an institution or organisation that has access to real human samples would be extremely advantageous, as this is a fundamental requirement for a fully developed OC detection kit.

REFERENCES

Aafreen, A., Pankaj, A., & Pankaj, A. (2022). A Study to Evaluate Risk of Ovarian Malignancy Algorithm (ROMA) in Patients with Ovarian Masses malignancy algorithm. *Asian Journal of Oncology*, 14–17.

Abraham, M., Alekseenko, A., Bergh, C., Blau, C., Briand, E., Doijade, M., Fleischmann, S., Gapsys, V., Garg, G., Gorelov, S., Gouaillardet, G., Gray, A., Irrgang, M. E., Jalalypour, F., Jordan, J., Junghans, C., Kanduri, P., Keller, S., Kutzner, C., Lindahl, E. (2023). GROMACS 2023 Manual.

Achar, A., & Sætrom, P. (2015). RNA motif discovery: a computational overview. *Biology direct*, *10*, 1-22.

Adasme, M. F., Linnemann, K. L., Bolz, S. N., Kaiser, F., Salentin, S., Haupt, V. J., & Schroeder, M. (2021). PLIP 2021: Expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *Nucleic acids research*, *49*(W1), W530-W534.

Afanasyeva, A., Nagao, C., & Mizuguchi, K. (2019). Prediction of the secondary structure of short DNA aptamers. *Biophysics and physicobiology*, *16*, 287-294.

Ahmadzadeh, A., Shahbazian, N., & Arkavazi, B. (2022). Examining CA 19-9 Tumor Marker in Ovarian Dermoid Cysts; A Retrospective Study. *Journal of Obstetrics, Gynecology and Cancer Research*, *3*(1), 19-22.

Aier, I., Varadwaj, P. K., & Raj, U. (2016). Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. *Scientific reports*, *6*(1), 34984.

Akshaya, K., Arthi, C., Pavithra, A. J., Poovizhi, P., Antinate, S. S., Hikku, G. S., ... & Murugesan, R. (2020). Bioconjugated gold nanoparticles as an efficient colorimetric sensor for cancer diagnostics. *Photodiagnosis and Photodynamic Therapy*, *30*, 101699.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). The structure and function of DNA. In *Molecular Biology of the Cell*. 4th edition. Garland Science.

Aldewachi, H., Chalati, T., Woodroofe, M. N., Bricklebank, N., Sharrack, B., & Gardiner, P. (2018). Gold nanoparticle-based colorimetric biosensors. *Nanoscale*, *10*(1), 18-33.

Alex, S., & Tiwari, A. (2015). Functionalized gold nanoparticles: synthesis, properties and applications—a review. *Journal of nanoscience and nanotechnology*, *15*(3), 1869-1894.

Alizadeh, S., & Nazari, Z. (2020). A review on gold nanoparticles aggregation and its applications. *J. Chem. Rev*, *2*(4), 228-242.

Al-Khayyat, M. Z. S., & Al-Dabbagh, A. G. A. (2016). In silico prediction and docking of tertiary structure of LuxI, an inducer synthase of Vibrio fischeri. *Reports of biochemistry & molecular biology*, *4*(2), 66.

Al-Lazikani, B., Jung, J., Xiang, Z., & Honig, B. (2001). Protein structure prediction. *Current Opinion in Chemical Biology, 5*(1), 51–56.

Almazar, C. A., Mendoza, M. V., & Rivera, W. L. (2023). In Silico Approaches for the Identification of Aptamer Binding Interactions to Leptospira spp. Cell Surface Proteins. *Tropical Medicine and Infectious Disease, 8*(2), 125.

Alnaimi, A., Al-Hamry, A., Makableh, Y., Adiraju, A., & Kanoun, O. (2022). Gold nanoparticles-MWCNT based aptasensor for early diagnosis of prostate cancer. *Biosensors*, *12*(12), 1130.

Al-Refaei, M. A., Makki, R. M., & Ali, H. M. (2020). Structure prediction of transferrin receptor protein 1 (TfR1) by homology modelling, docking, and molecular dynamics simulation studies. *Heliyon*, *6*(1).

Amendola, V., Pilot, R., Frasconi, M., Maragò, O. M., & Iatì, M. A. (2017). Surface plasmon resonance in gold nanoparticles: a review. *Journal of physics: Condensed matter*, *29*(20), 203002.

Anastasi, E., Granato, T., Falzarano, R., Storelli, P., Ticino, A., Frati, L., Panici, P. B., & Porpora, M. G. (2013). The use of HE4, CA125 and CA72-4 biomarkers for differential diagnosis between ovarian endometrioma and epithelial ovarian cancer. *Journal of Ovarian Research*, *6*(44), 1-8.

Andrade, L. M., & Costa, G. M. (2023). Insights into gold nanoparticles possibilities for diagnosis and treatment of the head and neck upper aerodigestive tract cancers. *Cancers*, *15*(7), 2080.

Andreou, M., Kyprianidou, M., Cortas, C., Polycarpou, I., Papamichael, D., Kountourakis, P., & Giannakou, K. (2023). Prognostic factors influencing survival in ovarian cancer patients: A 10-year retrospective study. *Cancers*, *15*(24), 5710.

Avery, C., Patterson, J., Grear, T., Frater, T., & Jacobs, D. J. (2022). Protein Function Analysis through Machine Learning. *Biomolecules, 12*(9), 1246.

Azri, F. A., Selamat, J., Sukor, R., Yusof, N. A., Raston, N. H. A., Eissa, S., Zourob, M., & Chinnappan, R. (2021). Determination of minimal sequence for zearalenone aptamer by computational docking and application on an indirect competitive electrochemical aptasensor. *Analytical and Bioanalytical Chemistry, 413*(15), 3861–3872.

Babaei, M., Salmani, F., Kariman, N., Nasiri, S., & Ozgoli, G. (2022). Women's awareness of ovarian cancer risk factors and symptoms in Western Iran in 2020-2021. *BMC Women's Health, 22*(1), 192.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., ... & Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science, 373*(6557), 871-876.

Bahaman, A. H., Wahab, R. A., Abdul Hamid, A. A., Abd Halim, K. B., & Kaya, Y. (2021). Molecular docking and molecular dynamics simulations studies on β-glucosidase and xylanase Trichoderma asperellum to predict degradation order of cellulosic components in oil palm leaves for nanocellulose preparation. *Journal of Biomolecular Structure & Dynamics, 39*(7), 2628–2641.

Bai, X., Wang, Y., Song, Z., Feng, Y., Chen, Y., Zhang, D., & Feng, L. (2020). The basic properties of gold nanoparticles and their applications in tumor diagnosis and treatment. *International journal of molecular sciences*, *21*(7), 2480.

Baiomy, T. A., Khalil, O. H., Abdallah, W. M., Ramadan, M. S. H., Abdou, A. M., Elshafaay, B. S., & Harb, O. A. (2020). Ovarian cancer surgical staging, laparoscopy versus laparotomy: a comparative study. *Journal of Gynecologic Surgery, 36*(4), 179–183.

Baker Lab. (2023). *Robetta*. https://robetta.bakerlab.org/

Balasundaram, A., & Doss, C. G. P. (2022). unraveling the structural changes in the DNA-binding region of tumor protein p53 (TP53) upon hotspot mutation p53 Arg248 by comparative computational approach. *International Journal of Molecular Sciences*, *23*(24), 15499.

Banerjee, S., Yoon, H., Yebra, M., Tang, C.-M., Gilardi, M., Shankara Narayanan, J. S., White, R. R., Sicklick, J. K., & Ray, P. (2020). Anti-KIT DNA Aptamer for Targeted Labeling of Gastrointestinal Stromal Tumor. *Molecular Cancer Therapeutics, 19*(5), 1173–1182.

Barhoumi, A., Zhang, D., Tam, F., & Halas, N. J. (2008). Surface-Enhanced Raman Spectroscopy of DNA. *Journal of the American Chemical Society*, *130*(16), 5523–5529.

Barnas, E., Skret-Magierlo, J., Skret, A., Kaznowska, E., Depciuch, J., Szmuc, K., & Cebulski, J. (2020). Simultaneous FTIR and Raman spectroscopy in endometrial atypical hyperplasia and cancer. *International Journal of Molecular Sciences*, *21*(14), 4828.

Barr, C. E., Funston, G., Jeevan, D., Sundar, S., Mounce, L. T. A., & Crosbie, E. J. (2022). The Performance of HE4 Alone and in Combination with CA125 for the Detection of Ovarian Cancer in an Enriched Primary Care Population. *Cancers, 14*(9), 1–18.

Bast, R. C. J., Klug, T. L., St John, E., Jenison, E., Niloff, J. M., Lazarus, H., Berkowitz, R. S., Leavitt, T., Griffiths, C. T., Parker, L., Zurawski, V. R. J., & Knapp, R. C. (1983). A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *The New England Journal of Medicine, 309*(15), 883–887.

Battistel, M. D., Azurmendi, H. F., & Freedberg, D. I. (2017). Glycan OH Exchange Rate Determination in Aqueous Solution: Seeking Evidence for Transient Hydrogen Bonds. *The Journal of Physical Chemistry B, 121*(4), 683–695

Bauer, M., Strom, M., Hammond, D. S., & Shigdar, S. (2019). Anything You Can Do, I Can Do Better: Can Aptamers Replace Antibodies in Clinical Diagnostic Applications? *In Molecules, 24*(23), 1-13.

Bawane, H., Kadam, K., Mahale, V., & Kulkarni, R. (2024). Comprehensive assessment of 12 commercial DNA-binding dyes as alternatives to ethidium bromide for agarose gel electrophoresis. *Electrophoresis*, *45*(5-6), 442-450.

Behrouzi, R., Barr, C. E., & Crosbie, E. J. (2021). HE4 as a Biomarker for Endometrial Cancer. *Cancers*, *13*(19), 4764.

Bennie, L. A., Feng, J., Emmerson, C., Hyland, W. B., Matchett, K. B., McCarthy, H. O., & Coulter, J. A. (2021). Formulating RALA/Au nanocomplexes to enhance nanoparticle internalisation efficiency, sensitising prostate tumour models to radiation treatment. *Journal of Nanobiotechnology, 19,* 1-13.

Beiranvand, N., Freindorf, M., & Kraka, E. (2021). Hydrogen Bonding in Natural and Unnatural Base Pairs—A Local Vibrational Mode Study. *Molecules, 26*(8), 2268.

Bilonia, G., Mathur, R., & Suresh, S. (2022). Comparative Study of Significance of Serum Human Epididymis Protein 4 (HE4) and Carbohydrate Antigen 125 (CA125) as an Early Tumour Marker in Female Patients with Ovarian Cancer. *International Journal of Clinical and Diagnostic Pathology, 5*(13), 12-15.

Bitencourt-Ferreira, G., & de Azevedo, W. F. (2019). Homology modeling of protein targets with MODELLER. *Docking Screens for Drug Discovery,* 231-249.

Blind, M., & Blank, M. (2015). Aptamer selection technology and recent advances. *Molecular Therapy-Nucleic Acids, 4*.

Bosak, A., Saraf, N., Willenberg, A., Kwan, M. W. C., Alto, B. W., Jackson, G. W., Batchelor, R. H., Nguyen-Huu, T. D., Sankarapani, V., Parks, G. D., Seal, S., & Willenberg, B. J. (2019). Aptamer-gold nanoparticle conjugates for the colorimetric detection of arboviruses and vector mosquito species. *RSC Advances, 9*(41), 23752–23763.

Braicu, E. I., Krause, C. L., Torsten, U., Mecke, H., Richter, R., Hellmeyer, L., Lanowska, M., Müller, B., Koch, E., & Boenneß-Zaloum, J. (2022). HE4 as a serum biomarker for the diagnosis of pelvic masses: a prospective, multicenter study in 965 patients. *BMC Cancer, 22*(1), 831.

Brezovsky, J., Chovancova, E., Gora, A., Pavelka, A., Biedermannova, L., & Damborsky, J. (2013). Software tools for identification, visualization and analysis of protein tunnels and channels. *Biotechnology Advances, 31*(1), 38–49.

Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research, 47*(W1), 402–407.

Buglak, A. A., Samokhvalov, A. V, Zherdev, A. V, & Dzantiev, B. B. (2020). Methods and Applications of In Silico Aptamer Design and Modeling. *International Journal of Molecular Sciences, 21*(22), 8420.

Busch, R. T., Karim, F., Weis, J., Sun, Y., Zhao, C., & Vasquez, E. S. (2019). Optimization and Structural Stability of Gold Nanoparticle-Antibody Bioconjugates. *ACS Omega, 4*(12), 15269–15279.

Byun, J. (2021). Recent Progress and Opportunities for Nucleic Acid Aptamers. *Life, 11*(3), 193.

Cabasag, C. J., Fagan, P. J., Ferlay, J., Vignat, J., Laversanne, M., Liu, L., van der Aa, M. A., Bray, F., & Soerjomataram, I. (2022). Ovarian cancer today and tomorrow: A global assessment by world region and Human Development Index using GLOBOCAN 2020. *International Journal of Cancer, 151*(9), 1535–1541.

Campbell, S., & Gentry-Maharaj, A. (2018). The role of transvaginal ultrasound in screening for ovarian cancer. *Climacteric, 21*(3), 221–226.

Caprigliione, S., Luvero, D., Plotti, F., Terranova, C., Montera, R., Scaletta, G., Schirò, T., Rossini, G., Benedetti Panici, P., & Angioli, R. (2017). Ovarian cancer recurrence and early detection: may HE4 play a key role in this open challenge? A systematic review of literature. *Medical Oncology, 34*(9), 164.

Carone, A., Emilsson, S., Mariani, P., Désert, A., & Parola, S. (2023). Gold nanoparticle shape dependence of colloidal stability domains. *Nanoscale Advances, 5*(7), 2017–2026.

Cengiz, A., Koç, Z. P., Özcan Kara, P., & Yürekli, Y. (2019). The Role of (18)F-FDG PET/CT in Detecting Ovarian Cancer Recurrence in Patients with Elevated CA-125 Levels. *Molecular Imaging and Radionuclide Therapy, 28*(1), 8–14.

Chandra, G. K., Eklouh-Molinier, C., Fere, M., Angiboust, J.-F., Gobinet, C., Van-Gulick, L., Jeannesson, P., & Piot, O. (2015). Probing in Vitro Ribose Induced DNA-Glycation Using Raman Microspectroscopy. *Analytical Chemistry, 87*(5), 2655–2664.

Charkhchi, P., Cybulski, C., Gronwald, J., Wong, F. O., Narod, S. A., & Akbari, M. R. (2020). CA125 and Ovarian Cancer: A Comprehensive Review. *Cancers, 12*(12), 3730.

Chen, C., Zhou, S., Cai, Y., & Tang, F. (2017). Nucleic acid aptamer application in diagnosis and therapy of colorectal cancer based on cell-SELEX technology. *Npj Precision Oncology, 1(*1), 37.

Chen, S. H., & Goller, C. C. (2020). Harnessing single-stranded DNA binding protein to explore protein–protein and protein–DNA interactions. *Biochemistry and Molecular Biology Education, 48*(2), 181–190.

Chen, W., Shi, C., MacKerell Jr, A. D., & Shen, J. (2015). Conformational dynamics of two natively unfolded fragment peptides: comparison of the AMBER and CHARMM force fields. *The Journal of Physical Chemistry B, 119*(25), 7902–7910.

Chen, Z., Hu, L., Zhang, B.-T., Lu, A., Wang, Y., Yu, Y., & Zhang, G. (2021). Artificial Intelligence in Aptamer–Target Binding Prediction. *International Journal of Molecular Sciences, 22*(7), 3605.

Chen, Z., Wang, X., Pang, L., Zhang, J. Z. H., & Zhu, T. (2019). Effect of mutations on binding of ligands to guanine riboswitch probed by free energy perturbation and molecular dynamics simulations. *Nucleic Acid Research, 47*(13), 6618-6631.

Cherchi, P. L., Dessole, S., Ruiu, G. A., Ambrosini, G., Farina, M., Capobianco, G., & Ambrosini, A. (1999). The value of serum CA 125 and association CA 125/CA 19-9 in endometrial carcinoma. *European Journal of Gynaecological Oncology, 20*(4), 315–317.

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., Biedermannova, L., Sochor, J., & Damborsky, J. (2012). CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLOS Computational Biology*, *8*(10), e1002708, 1-12.

Chudecka-Głaz, A. M. (2015). ROMA, an algorithm for ovarian cancer. *Clinica Chimica Acta*, *440*, 143-151.

Cohen, J. G., White, M., Cruz, A., & Farias-Eisner, R. (2014). In 2014, can we do better than CA125 in the early detection of ovarian cancer? *World Journal of Biological Chemistry*, *5*(3), 286–300.

Cooper, D. N., Bacolla, A., Férec, C., Vasquez, K. M., Kehrer-Sawatzki, H., & Chen, J.-M. (2011). On the sequence-directed nature of human gene mutation: The role of genomic architecture and the local DNA sequence environment in mediating gene mutations underlying human inherited disease. *Human Mutation*, *32*(10), 1075–1099.

Cozzolino, F., Iacobucci, I., Monaco, V., & Monti, M. (2021). Protein–DNA/RNA Interactions: An Overview of Investigation Methods in the -Omics Era. *Journal of Proteome Research*, *20*(6), 3018–3030.

Danaei, M., Dehghankhold, M., Ataei, S., Hasanzadeh Davarani, F., Javanmard, R., Dokhani, A., Khorasani, S., & Mozafari, M. R. (2018). Impact of Particle Size and Polydispersity Index on the Clinical Applications of Lipidic Nanocarrier Systems. *Pharmaceutics*, *10*(2), 57, 1-17.

Daras, G., Alatzas, A., Tsitsekian, D., Templalexis, D., Rigas, S., & Hatzopoulos, P. (2019). Detection of RNA‐protein interactions using a highly sensitive non‐radioactive electrophoretic mobility shift assay. *Electrophoresis*, *40*(9), 1365-1371.

Dekhili, R., Cherni, K., Liu, H., Li, X., Djaker, N., & Spadavecchia, J. (2020). Aptamer–Gold(III) Complex Nanoparticles: A New Way to Detect Cu, Zn SOD Glycoprotein. *ACS Omega*, *5*(23), 13851–13859.

Dochez, V., Caillon, H., Vaucel, E., Dimet, J., Winer, N., & Ducarme, G. (2019). Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. *Journal of Ovarian Research*, *12*(1), 1–9.

Dolscheid-Pommerich, R. C., Keyver-Paik, M., Hecking, T., Kuhn, W., Hartmann, G., Stoffel-Wagner, B., & Holdenrieder, S. (2017). Clinical performance of LOCITM-based tumor marker assays for tumor markers CA 15-3, CA 125, CEA, CA 19-9 and AFP in gynecological cancers. *Tumor Biology*, *39*(10),1-11.

Douaki, A., Garoli, D., Inam, A. K. M. S., Angeli, M. A., Cantarella, G., Rocchia, W., Wang, J., Petti, L., & Lugli, P. (2022). Smart Approach for the Design of Highly Selective Aptamer-Based Biosensors. *Biosensors*, *12*(8), 574.

Doubeni, C. A., Doubeni, A. R. B., & Myers, A. E. (2016). Diagnosis and management of ovarian cancer. *American Family Physician*, *93*(11), 937–944.

Duan, J., Wang, X., & Kizer, M. E. (2020). Biotechnological and Therapeutic Applications of Natural Nucleic Acid Structural Motifs. *Topics in Current Chemistry*, *378*(2), 26.

Dubey, A. K., Baker, C. S., Romeo, T., & Babitzke, P. (2005). RNA sequence and secondary structure participate in high-affinity CsrA–RNA interaction. *Rna*, *11*(10), 1579-1587.

Dunton, C. J., Hutchcraft, M. L., Bullock, R. G., Northrop, L. E., & Ueland, F. R. (2021). Salvaging Detection of Early-Stage Ovarian Malignancies When CA125 Is Not Informative. *Diagnostics, 11*(8), 1440.

Eaton, R. M., Shallcross, J. A., Mael, L. E., Mears, K. S., Minkoff, L., Scoville, D. J., & Whelan, R. J. (2015). Selection of DNA aptamers for ovarian cancer biomarker HE4 using CE-SELEX and high-throughput sequencing. A*nalytical and Bioanalytical Chemistry, 407*(23), 6965–6973.

Ebell, M. H., Culp, M., Lastinger, K., & Dasigi, T. (2015). A systematic review of the bimanual examination as a test for ovarian cancer. *American Journal of Preventive Medicine, 48*(3), 350–356.

Egli, M. (2022). DNA and RNA Structure. In G. M. Blackburn, M. Egli, M. J. Gait, & J. K. Watts (Eds.), *Nucleic Acids in Chemistry and Biology* (pp. 20-95). The Royal Society of Chemistry.

Elahi, N., Kamali, M., & Baghersad, M. H. (2018). Recent biomedical applications of gold nanoparticles: A review. *Talanta, 184*, 537–556.

Elalouf, A. (2023). In-silico Structural Modeling of Human Immunodeficiency Virus Proteins. *Biomedical Engineering and Computational Biology, 14*, 11795972231154402.

Elias, K. M., Guo, J., & Bast, R. C. (2018). Early detection of ovarian cancer. *Hematology/Oncology Clinics, 32*(6), 903–914.

Ellington, A. D., & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature, 346*(6287), 818–822.

Ellington, A. D., & Szostak, J. W. (1992). Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature, 355*(6363), 850–852.

Elmahdi, N. M., Latiff, L. A., Akhtari-Zavare, M., Ismail, M., Manaf, R. A., & Abu Bakar, A. S. Bin. (2017). A Cross-sectional Study of Ovarian Cancer Knowledge and Its Determinants among Female Employees in a Public University, Malaysia. *Malaysian Journal of Medicine & Health Sciences, 13*(3), 71-79.

Elofsson, A. (2023). Progress at protein structure prediction, as seen in CASP15. *Current Opinion in Structural Biology, 80*, 1-6.

Elokely, K. M., & Doerksen, R. J. (2013). Docking Challenge: Protein Sampling and Molecular Docking Performance. *Journal of Chemical Information and Modeling, 53*(8), 1934–1945.

Elskens, J. P., Elskens, J. M., & Madder, A. (2020). Chemical Modification of Aptamers for Increased Binding Affinity in Diagnostic Applications: Current Status and Future Prospects. *International Journal of Molecular Sciences, 21*(12), 4522.

Emamjomeh, A., Choobineh, D., Hajieghrari, B., MahdiNezhad, N., & Khodavirdipour, A. (2019). DNA–protein interaction: identification, prediction and data analysis. *Molecular Biology Reports, 46*(3), 3571–3596.

Fadeev, M., O’Hagan, M. P., Biniuri, Y., & Willner, I. (2022). Aptamer–Protein Structures Guide In Silico and Experimental Discovery of Aptamer–Short Peptide Recognition Complexes or Aptamer–Amino Acid Cluster Complexes. *The Journal of Physical Chemistry B, 126*(44), 8931–8939.

Falcetta, F. S., Lawrie, T. A., Medeiros, L. R. F., da Rosa, M. I., Edelweiss, M. I., Stein, A. T., Zelmanowicz, A., Moraes, A. B., Zanini, R. R., & Rosa, D. D. (2016). Laparoscopy versus laparotomy for FIGO stage I ovarian cancer. *Cochrane Database of Systematic Reviews, 10*, CD005344.

Fan, J., Fu, A., & Zhang, L. (2019). Progress in molecular docking. *Quantitative Biology, 7*(2), 83–89.

Fan, H., & Mark, A. E. (2004). Refinement of homology-based protein structures by molecular dynamics simulation techniques. *Protein Science, 13*(1), 211–220.

Fan, Y., Cui, M., Liu, Y., Jin, M., & Zhao, H. (2020). Selection and characterization of DNA aptamers for constructing colorimetric biosensor for detection of PBP2a. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 228,* 117735.

Feig, M. (2016). Local Protein Structure Refinement via Molecular Dynamics Simulations with locPREFMD. *Journal of Chemical Information and Modeling, 56*(7), 1304–1312.

Ferenczy, G. G., & Kellermayer, M. (2022). Contribution of hydrophobic interactions to protein mechanical stability. *Computational and Structural Biotechnology Journal*, *20*, 1946–1956.

Fernandez-Garza, L. E., Dominguez-Vigil, I. G., Garza-Martinez, J., Valdez-Aparicio, E. A., Barrera-Barrera, S. A., & Barrera-Saldana, H. A. (2021). Personalized Medicine in Ovarian Cancer: A Perspective from Mexico. *World Journal of Oncology*, *12*(4), 85–92.

Ferraz, R. A. C., Lopes, A. L. G., da Silva, J. A. F., Moreira, D. F. V., Ferreira, M. J. N., & de Almeida Coimbra, S. V. (2021). DNA–protein interaction studies: a historical and comparative analysis. *Plant Methods*, *17*(1), 82.

Ferreira de Freitas, R., & Schapira, M. (2017). A systematic analysis of atomic protein-ligand interactions in the PDB. *MedChemComm, 8*(10), 1970–1981.

Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J. (2016). Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols, 11*(5), 905–919.

Francesk, M., Nikoleta, O., Levan, T., Elias, L., & Ioannis, K. (2020). A giant ovarian mucinous tumor in a 58-year-old postmenopausal patient with persistent abdominal pain and high serum levels of CA 19-9. *PAMJ, 37*(76), 1-4.

Frenkel, D., & Smit, B. (2023). *Understanding molecular simulation: from algorithms to applications*. 2nd Ed. Elsevier.

Funston, G., Hamilton, W., Abel, G., Crosbie, E. J., Rous, B., & Walter, F. M. (2020). The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study. *PLoS medicine*, *17*(10), e1003295.

Futane, A., Narayanamurthy, V., Jadhav, P., & Srinivasan, A. (2023). Aptamer-based rapid diagnosis for point-of-care application. *Microfluidics and Nanofluidics, 27*(2), 15.

Gallardo, A., Bogart, B. M., & Dutagaci, B. (2022). Protein–Nucleic Acid Interactions for RNA Polymerase II Elongation Factors by Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling, 62*(12), 3079–3089.

Gallicchio, E., Lapelosa, M., & Levy, R. M. (2010). Binding Energy Distribution Analysis Method (BEDAM) for Estimation of Protein−Ligand Binding Affinities. *Journal of Chemical Theory and Computation, 6*(9), 2961–2977.

Ganjali Koli, M., Eshaghi Malekshah, R., & Hajiabadi, H. (2023). Insights from molecular dynamics and DFT calculations into the interaction of 1,4-benzodiazepines with 2-hydroxypropyl-βCD in a theoretical study. *Scientific Reports, 13*(1), 9866.

Gaona-Luviano, P., Medina-Gaona, L. A., & Magaña-Pérez, K. (2020). Epidemiology of ovarian cancer. *Chinese clinical oncology*, *9*(4), 47-47.

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*(5), 449–461.

Giorgi-Coll, S., Marín, M. J., Sule, O., Hutchinson, P. J., & Carpenter, K. L. H. (2019). Aptamer-modified gold nanoparticles for rapid aggregation-based detection of inflammation: an optical assay for interleukin-6. Microchimica *Acta, 187*(1), 13.

Godschalk, F., Genheden, S., Söderhjelm, P., & Ryde, U. (2013). Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations. *Physical Chemistry Chemical Physics, 15*(20), 7731–7739.

Grabowski, S. J. (2006). Hydrogen bonding: new insights (Vol. 3). Springer.

Granato, T., Porpora, M. G., Longo, F., Angeloni, A., Manganaro, L., & Anastasi, E. (2015). HE4 in the differential diagnosis of ovarian masses. *Clinica Chimica Acta, 446*, 147–155.

Grayson, K., Gregory, E., Khan, G., & Guinn, B.-A. (2019). Urine Biomarkers for the Early Detection of Ovarian Cancer - Are We There Yet? *Biomarkers in Cancer, 11*, 1-8.

Guo, B., Lian, W., Liu, S., Cao, Y., & Liu, J. (2019). Comparison of diagnostic values between CA125 combined with CA199 and ultrasound combined with CT in ovarian cancer. *Oncology letters*, *17*(6), 5523-5528.

Gunasekaran, K., Ramakrishnan, C., & Balaram, P. (1996). Disallowed Ramachandran Conformations of Amino Acid Residues in Protein Structures. *Journal of Molecular Biology, 264*(1), 191–198.

Gunter, C. (2024). *Point Mutation*. National Human Genome Research Institute. https://www.genome.gov/genetics-glossary/Point-Mutation

Guo, H.-B., Perminov, A., Bekele, S., Kedziora, G., Farajollahi, S., Varaljay, V., Hinkle, K., Molinero, V., Meister, K., Hung, C., Dennis, P., Kelley-Loughnane, N., & Berry, R. (2022). AlphaFold2 models indicate that protein sequence determines both structure and dynamics. *Scientific Reports, 12*(1), 10696.

Guterres, H., Park, S.-J., Zhang, H., Perone, T., Kim, J., & Im, W. (2022). CHARMM-GUI high-throughput simulator for efficient evaluation of protein–ligand interactions with different force fields. Protein Science, 31(9), e4413.

Han, C., Bellone, S., Siegel, E. R., Altwerger, G., Menderes, G., Bonazzoli, E., Egawa-Takata, T., Pettinella, F., Bianchi, A., Riccio, F., Zammataro, L., Yadav, G., Marto, J. A., Penet, M. F., Levine, D. A., Drapkin, R., Patel, A., Litkouhi, B., Ratner, E., Santin, A. D. (2018). A novel multiple biomarker panel for the early detection of high-grade serous ovarian carcinoma. *Gynecologic Oncology, 149*(3), 585–591.

Hao, G.-F., Xu, W.-F., Yang, S.-G., & Yang, G.-F. (2015). Multiple Simulated Annealing-Molecular Dynamics (MSA-MD) for Conformational Space Search of Peptide and Miniprotein. *Scientific Reports, 5*(1), 15568.

Hanzek, A., Siatka, C., & Duc, A.-C. (2021). High-specificity nucleic acid aptamers for detection of ovarian cancer protein biomarkers: Application in diagnostics. *Aptamers, 5*, 7–14.

Hardin, C., Pogorelov, T. V, & Luthey-Schulten, Z. (2002). Ab initio protein structure prediction. *Current Opinion in Structural Biology, 12*(2), 176–181.

Harris, B. J., Cheng, X., & Frymier, P. (2014). All-atom molecular dynamics simulation of a photosystem i/detergent complex. *The Journal of Physical Chemistry B*, *118*(40), 11633-11645.

Hart, K., Foloppe, N., Baker, C. M., Denning, E. J., Nilsson, L., & Mackerell, A. D. J. (2012). Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *Journal of Chemical Theory and Computation*, *8*(1), 348–362.

Hasan, M. R., Sharma, P., Pilloton, R., Khanuja, M., & Narang, J. (2022). Colorimetric biosensor for the naked-eye detection of ovarian cancer biomarker PDGF using citrate modified gold nanoparticles. *Biosensors and Bioelectronics: X, 11*, 1-6.

Hashim, N., Shet, D., & Zain, N. M. (2021). Women's Knowledge of Ovarian Cancer And Its Determinants Factors. *Quantum Journal of Medical and Health Sciences*, *1*(5), 1-10.

Hellstrom, I., Heagerty, P. J., Swisher, E. M., Liu, P., Jaffar, J., Agnew, K., & Hellstrom, K. E. (2010). Detection of the HE4 protein in urine as a biomarker for ovarian neoplasms. *Cancer Letters, 296*(1), 43–48.

He, Z., & Yang, H. (2018). Colourimetric detection of swine-specific DNA for halal authentication using gold nanoparticles. Food Control, 88, 9–14. Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature Protocols, 2*(8), 1849–1861.

Ho, B. K., & Brasseur, R. (2005). The Ramachandran plots of glycine and pre-proline. BMC Structural Biology, 5(1), 14. Hollingsworth, S. A., & Karplus, P. A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular Concepts, 1*(3–4), 271–283.

Hoinka, J., Zotenko, E., Friedman, A., Sauna, Z. E., & Przytycka, T. M. (2012). Identification of sequence–structure RNA binding motifs for SELEX-derived aptamers. *Bioinformatics, 28*(12), i215–i223.

Holden, N. S., & Tacon, C. E. (2011). Principles and problems of the electrophoretic mobility shift assay. *Journal of Pharmacological and Toxicological Methods, 63*(1), 7–14.

Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron, 99*(6), 1129–1143.

Hollis, R. L. (2023). Molecular characteristics and clinical behaviour of epithelial ovarian cancers. *Cancer Letters, 555*, 216057.

Hori, S., Herrera, A., Rossi, J. J., & Zhou, J. (2018). Current Advances in Aptamers for Cancer Diagnosis and Therapy. *Cancers,10*(1), 9.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., & MacKerell, A. D. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods, 14*(1), 71–73.

Huang, W., Wu, S., Lin, Z., Chen, P., & Wu, G. (2017). Evaluation of HE4 in the diagnosis and follow up of non-small cell lung cancers. *Clin Lab, 63*(3), 461-467.

Huhtinen, K., Suvitie, P., Hiissa, J., Junnila, J., Huvila, J., Kujari, H., Setälä, M., Härkki, P., Jalkanen, J., Fraser, J., Mäkinen, J., Auranen, A., Poutanen, M., & Perheentupa, A. (2009). Serum HE4 concentration differentiates malignant ovarian tumours from ovarian endometriotic cysts. *British Journal of Cancer, 100*(8), 1315–1319.

Hu, C., Qin, Z., Fu, J., Gao, Q., Chen, C., Tan, C. S., & Li, S. (2023). Aptamer-based carbohydrate antigen 125 sensor with molybdenum disulfide functional hybrid materials. *Analytical Biochemistry*, 115213.

Hunter, R. J. (2013). *Zeta potential in colloid science: principles and applications* (Vol. 2). Academic press.

Hu, R., Mu, Z., Gong, F., Qing, M., Yuan, Y., & Bai, L. (2023). A signal-on electrochemical aptasensor for sensitive detection of human epididymis protein 4 based on functionalized metal–organic framework/ketjen black nanocomposite. *Journal of Materials Science*, 1–13.

Ilgu, M., Yan, S., Khounlo, R. M., Lamm, M. H., & Nilsen-Hamilton, M. (2019). Common Secondary and Tertiary Structural Features of Aptamer–Ligand Interaction Shared by RNA Aptamers with Different Primary Sequences. *Molecules, 24*(24).

Ivanov, Y. D., Malsagova, K. A., Pleshakova, T. O., Galiullin, R. A., Kozlov, A. F., Shumov, I. D., Popov, V. P., Kapustina, S. I., Ivanova, I. A., & Isaeva, A. I. (2021). Aptamer-Sensitized Nanoribbon Biosensor for Ovarian Cancer Marker Detection in Plasma. *Chemosensors, 9*(8), 222.

Janke, J. J., Yu, Y., Pomin, V. H., Zhao, J., Wang, C., Linhardt, R. J., & García, A. E. (2022). Characterization of Heparin's Conformational Ensemble by Molecular Dynamics Simulations and Nuclear Magnetic Resonance Spectroscopy. *Journal of Chemical Theory and Computation, 18*(3), 1894–1904.

Jans, H., & Huo, Q. (2012). Gold nanoparticle-enabled biological and chemical detection and analysis. *Chemical Society Reviews, 41*(7), 2849–2866.

Jazayeri, M. H., Aghaie, T., Avan, A., Vatankhah, A., & Ghaffari, M. R. S. (2018). Colorimetric detection based on gold nano particles (GNPs): An easy, fast, inexpensive, low-cost and short time method in detection of analytes (protein, DNA, and ion). *Sensing and Bio-Sensing Research, 20*(2017), 1–8.

Jazayeri, M. H., Amani, H., Pourfatollah, A. A., Pazoki-Toroudi, H., & Sedighimoghaddam, B. (2016). Various methods of gold nanoparticles (GNPs) conjugation to antibodies. *Sensing and Bio-Sensing Research, 9*, 17–22.

Jeddi, I., & Saiz, L. (2017). Three-dimensional modeling of single stranded DNA hairpins for aptamer-based biosensors. *Scientific Reports, 7*(1), 1178.

Jeddi, I., & Saiz, L. (2021). Computational design of single-stranded DNA hairpin aptamers immobilized on a biosensor substrate. *Scientific Reports, 11*(1), 10984.

Jin, X., Du, M., Wang, Y., Wang, Y., Lu, Y., Xu, C., & Zhang, X. (2022). Evaluation of serum CA125-Tn glycoform in peritoneal dissemination and surgical completeness of high-grade serous ovarian cancer. *Journal of Ovarian Research, 15*(1), 134.

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society, 118*(45), 11225–11236.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583–589.

Källberg, M., Margaryan, G., Wang, S., Ma, J., & Xu, J. (2014). RaptorX server: a resource for template-based protein structure modeling. *Protein structure prediction*, 17-27.

Kalsi, J., Gentry-Maharaj, A., Ryan, A., Singh, N., Burnell, M., Massingham, S., Apostolidou, S., Sharma, A., Williamson, K., & Seif, M. (2021). Performance characteristics of the ultrasound strategy during incidence screening in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Cancers, 13*(4), 858.

Kamaraj, B., Al-Subaie, A. M., Ahmad, F., Surapaneni, K. M., & Alsamman, K. (2021). Effect of novel leukemia mutations (K75E & E222K) on interferon regulatory factor 1 and its interaction with DNA: insights from molecular dynamics simulations and docking studies. *Journal of Biomolecular Structure & Dynamics, 39*(14), 5235–5247

Kamaraj, B., & Bogaerts, A. (2015). Structure and Function of p53-DNA Complexes with Inactivation and Rescue Mutations: A Molecular Dynamics Simulation Study. PloS One, 10(8), e0134638, 1-16. Kaur H, & Yung LY. (2012). Probing high affinity sequences of DNA aptamer against VEGF165. *PLoS One 7*(2), e31196, 1-9.

Kang, K. N., Koh, E. Y., Jang, J. Y., & Kim, C. W. (2022). Multiple biomarkers are more accurate than a combination of carbohydrate antigen 125 and human epididymis protein 4 for ovarian cancer screening. *Obstetrics & Gynecology Science*, *65*(4), 346-354.

Kasper, S. M., Dueholm, M., Marinovskij, E., & Blaakær, J. (2017). Imaging diagnostics in ovarian cancer: magnetic resonance imaging and a scoring system guiding choice of primary treatment. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 210*, 83–89.

Kaur, H., Bruno, J. G., Kumar, A., & Sharma, T. K. (2018). Aptamers in the Therapeutics and Diagnostics Pipelines. Theranostics, 8(15), 4016–4032. Khan, I., Saeed, K., & Khan, I. (2019). Nanoparticles: Properties, applications and toxicities. *Arabian Journal of Chemistry*, *12*(7), 908–931.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols, 10*(6), 845–858.

Kemppainen, J., Hynninen, J., Virtanen, J., & Seppänen, M. (2019). PET/CT for evaluation of ovarian cancer. *Seminars in Nuclear Medicine, 49*(6), 484–492.

Khiewvan, B., Torigian, D. A., Emamzadehfard, S., Paydary, K., Salavati, A., Houshmand, S., Werner, T. J., & Alavi, A. (2017). An update on the role of PET/CT and PET/MRI in ovarian cancer. *European Journal of Nuclear Medicine and Molecular Imaging, 44*, 1079–1091.

Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic acids research*, *32*(suppl_2), W526-W531.

Kim, Y. S., Raston, N. H. A., & Gu, M. B. (2016). Aptamer-based nanobiosensors. *Biosensors and Bioelectronics, 76*, 2–19.

Komarova, N., & Kuznetsov, A. (2019). Inside the black box: what makes SELEX better?. *Molecules*, *24*(19), 3598.

Kosinska, M., Misiewicz, P., Kalita, K., Fijuth, J., Foks, M., Kuncman, L., & Gottwald, L. (2023). The value of [18F]FDG PET/CT examination in the detection and differentiation of recurrent ovarian cancer. *Nuclear Medicine Review, 26*(0), 98–105.

Kothandan, R., Uthayasooriyan, P., & Vairamani, S. (2021). Search for RNA aptamers against non-structural protein of SARS-CoV-2: Design using molecular dynamics approach. *Beni-Suef University journal of basic and applied sciences, 10,* 1-12.

Kovarova, A., Kastrati, G., Pekarkova, J., Metelka, R., Drbohlavova, J., Bilkova, Z., Selesovska, R., & Korecka, L. (2023). Biosensor with electrochemically active nanocomposites for signal amplification and simultaneous detection of three ovarian cancer biomarkers. *Electrochimica Acta, 469*, 143213.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., & Moult, J. (2021). Critical assessment of methods of protein structure prediction (CASP)-Round XIV. *Proteins, 89*(12), 1607–1617.

Kufareva, I., & Abagyan, R. (2012). Methods of protein structure comparison. Methods in Molecular Biology (Clifton, N.J.), 857, 231–257. Kumari, R., Kumar, R., & Lynn, A. (2014). g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. *Journal of Chemical Information and Modeling, 54*(7), 1951–1962.

Kumar D, D., Pandian, L., C, G. P. D., Tayubi, I., D. R., N. K., Inbanathan, F. Y., Siva, R., & Veeraraghavan, B. (2017). A Molecular Docking and Dynamics Approach to Screen Potent Inhibitors Against Fosfomycin Resistant Enzyme in Clinical Klebsiella pneumoniae. *Journal of Cellular Biochemistry, 118*, 4088-4094.

Kumarasamy, C., Madhav, M. R., Sabarimurugan, S., Lakhotiya, K., Pandey, V., Priyadharshini, T., Baxi, S., Gothandam, K. M., & Jayaraj, R. (2019). Diagnostic and prognostic role of HE4 expression in multiple carcinomas: A protocol for systematic review and meta-analysis. *Medicine, 98*(28), e15336.

Kurczab, R., Śliwa, P., Rataj, K., Kafel, R., & Bojarski, A. J. (2018). Salt Bridge in Ligand–Protein Complexes—Systematic Theoretical and Statistical Investigations. *Journal of Chemical Information and Modeling, 58*(11), 2224–2238.

Kurman, R. J., & Shih, I.-M. (2016). The Dualistic Model of Ovarian Carcinogenesis: Revisited, Revised, and Expanded. *The American Journal of Pathology, 186*(4), 733–747.

Kurniasih, R., Ambarsari, L., & Wahyudi, S. T. (2021). Structure Identification and Quality Assessment of Laccase (Lac InaCC) from Neurospora crassa by Using a Structure Prediction. *HAYATI Journal of Biosciences*, *28*(1), 1-1.

Lakshmanan, M., Kumar, V., Chaturvedi, A., Misra, S., Gupta, S., Akhtar, N., Rajan, S., Jain, K., & Garg, S. (2019). Role of serum HE4 as a prognostic marker in carcinoma of the ovary. *Indian Journal of Cancer, 56*(3), 216–221.

Lakshmipriya, T., Gopinath, S. C. B., Hashim, U., Chinni, S. V., & Thean-Hock, T. (2018). Computational base substitution analyses on secondary structure of aptamer: Conformational changes diminish complex formation. *Walailak Journal of Science and Technology (WJST), 15*(9), 645-657.

Larsson, A., Greig-Pylypczuk, R., & Huisman, A. (2015). The state of point-of-care testing: a european perspective. *Upsala Journal of Medical Sciences, 120*(1), 1–10.

Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Science, 27*(1), 129–134.

Lee, J., Freddolino, P. L., & Zhang, Y. (2017). Ab initio protein structure prediction. *From protein structure to function with bioinformatics*, 3-35.

Lee, J. H., Pollert, K., & Konermann, L. (2019). Testing the robustness of solution force fields for MD simulations on gaseous protein ions. *The Journal of Physical Chemistry B*, *123*(31), 6705-6715.

Lee, S. J., Cho, J., Lee, B.-H., Hwang, D., & Park, J.-W. (2023). Design and Prediction of Aptamers Assisted by In Silico Methods. *Biomedicines, 11*(2), 356.

Lee, Y. K., Praveena, K. A., Woo, Y. L., & Ng, C. J. (2021). Coping strategies among Malaysian women with recurrent ovarian cancer: A qualitative study. *Asia-Pacific Journal of Oncology Nursing, 8*(1), 40–45.

Lemkul, J. A. (2018). 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, 1*(1), 5068.

Lertkhachonsuk, A., Buranawongtrakoon, S., Lekskul, N., Rermluk, N., Wee-Stekly, W.-W., & Charakorn, C. (2020). Serum CA19-9, CA-125 and CEA as tumor markers for mucinous ovarian tumors. *Journal of Obstetrics and Gynaecology Research, 46*(11), 2287–2291.

Le, T. T., Chumphukam, O., & Cass, A. E. G. (2014). Determination of minimal sequence for binding of an aptamer. A comparison of truncation and hybridization inhibition methods. *RSC Advances, 4*(88), 47227–47233.

Lheureux, S., Gourley, C., Vergote, I., & Oza, A. M. (2019). Epithelial ovarian cancer. *The Lancet, 393*(10177), 1240–1253.

Li, J., Liu, L., Feng, Z., Wang, X., Huang, Y., Dai, H., Zhang, L., Song, F., Wang, D., & Zhang, P. (2020). Tumor markers CA15-3, CA125, CEA and breast cancer survival by molecular subtype: a cohort study. *Breast Cancer, 27*, 621–630.

Li, M., Men, X., & Zhang, X. (2020). Diagnostic value of carbohydrate antigen 72-4 combined with carbohydrate antigen 15.3 in ovarian cancer, cervical cancer and endometrial cancer. *J BUON, 25*(4), 1918–1927.

Li, X., Xu, Y., & Zhang, L. (2019). Serum CA153 as biomarker for cancer and noncancer diseases. *Progress in Molecular Biology and Translational Science, 162*, 265–276.

Li, Y., Wang, Z. C., Luo, L., Mu, C. Y., Xu, J., Feng, Q., ... & Lan, T. (2020). The clinical value of the combined detection of sEGFR, CA125 and HE4 for epithelial ovarian cancer diagnosis. *European Review for Medical & Pharmacological Sciences, 24*(2).

Liang, J., Yang, X., Liu, L., Qiao, L., Peng, P., & Zhou, J. (2020). Combined measurement of miRNA-183, HE4, and CA-125 increases diagnostic efficiency for ovarian cancer. *European Journal of Gynaecological Oncology, 41*(1), 30–35.

Lin, M., & Guo, J. (2019). New insights into protein–DNA binding specificity from hydrogen bond based comparative study. *Nucleic Acids Research, 47*(21), 11103–11113.

Lin, T., Xu, S., Wang, Y., Nian, X., Shan, X., Jiang, T., & Qiu, M. (2022). Human epididymis protein 4 as a new diagnostic biomarker for rheumatoid arthritis-associated interstitial lung disease. *Clinical and Experimental Rheumatology, 40,* 2167–2174.

Liu, H., & Yu, J. (2018). Challenges of SELEX and demerits of aptamer-based methods. *Aptamers for Analytical Applications*, 345-364.

Long, Z., Bing, T., Zhang, X., Sheng, J., Zu, S., Li, W., Liu, X., Zhang, N., & Shangguan, D. (2023). Structural Optimization and Interaction Study of a DNA Aptamer to L1 Cell Adhesion Molecule. *International Journal of Molecular Sciences, 24*(10), 8612.

Lozano-Aponte, J., Scior, T., Ambrosio, F. N. M., González-Melchor, M., & Alexander, C. (2019). Exploring electrostatic patterns of human, murine, equine and canine TLR4/MD-2 receptors. *Innate Immunity, 26*(5), 364–380.

Lu, Y., Mao, J., Xu, Y., Pan, H., Wang, Y., & Li, W. (2022). Ropivacaine represses the ovarian cancer cell stemness and facilitates cell ferroptosis through inactivating the PI3K/AKT signaling pathway. *Human & Experimental Toxicology, 41*, 1-11.

Lv, W., Guo, H., Wang, J., Ma, R., Niu, L., & Shang, Y. (2023). PDLIM2 can inactivate the TGF-β/Smad pathway to inhibit the malignant behavior of ovarian cancer cells. *Cell Biochemistry and Function, 41*(5), 542–552.

Ma, X., Huang, G., Ye, M., Zhang, X., Wang, Y., Liang, T., Deng, H., & Li, C. (2021). A homogeneous biosensor for Human Epididymis Protein 4 based on upconversion luminescence resonance energy transfer. *Microchemical Journal, 164*, 106083.

Ma, X., Lakshmipriya, T., & Gopinath, S. C. B. (2019). Recent Advances in Identifying Biomarkers and High-Affinity Aptamers for Gynecologic Cancers Diagnosis and Therapy. J*ournal of Analytical Methods in Chemistry, 2019*, 1-10.

Ma, Y., Hao, C., & Chang, H. (2013). Nucleotide Mismatches of Foot-and-Mouth Disease Virus During Replication. *Genetics and Molecular Research, 12*(2), 1022-1027.

Maccaroni, E., Giampieri, R., Del Prete, M., Principi, G., Bracci, R., Calcinari, A., Della Mora, A., Bianchi, F., Belvederesi, L., & Brugiati, C. (2017). Prognostic impact of CA15-3 pre-treatment levels in ovarian cancer patients. *Annals of Oncology, 28*, vi71–vi72.

Macdonald, J., Henri, J., Goodman, L., Xiang, D., Duan, W., & Shigdar, S. (2017). Development of a bifunctional aptamer targeting the transferrin receptor and epithelial cell adhesion molecule (EpCAM) for the treatment of brain cancer metastases. *ACS Chemical Neuroscience*, *8*(4), 777-784.

Mahshid, S. S., Camiré, S., Ricci, F., & Vallée-Bélisle, A. (2015). A highly Selective Electrochimical DNA-Based Sensor that Employs Steric Hindrance Effects to Detect Proteins Directly in Whole Blood. *Journal of the American Chemical Society, 137*(50), 15596-15599.

Makrigiannakis, A., Vrekoussis, T., Zoumakis, E., Kalantaridou, S. N., & Jeschke, U. (2017). The Role of HCG in Implantation: A Mini-Review of Molecular and Clinical Evidence. *International Journal of Molecular Sciences, 18*(6), 1-8.

Manan, A. A., Basri, H., Kaur, N., Rahman, S. Z. A., Amir, P. M., Ali, N., Raman, S., Bahtiar, B., Ramdzuan, N. S. M., Soffian, S. S. S., Othman, R., Othman, N. A., & Aziz, A. A. (2019). Malaysia National Cancer Registry Report 2012-2016. *Ministry of Health Malaysia.*

Mao, X., Wang, Y., Jiang, L., Zhang, H., Zhao, Y., Liu, P., Liu, J., Hammock, B. D., & Zhang, C. (2022). A Polydopamine-Coated Gold Nanoparticles Quenching Quantum Dots-Based Dual-Readout Lateral Flow Immunoassay for Sensitive Detection of Carbendazim in Agriproducts. *Biosensors, 12*(2), 83, 1-13.

Markham, N. R., & Zuker, M. (2008). UNAFold: Software for nucleic acid folding and hybridization. *Methods in Molecular Biology, 453*, 3–31.

Matsas, A., Stefanoudakis, D., Troupis, T., Kontzoglou, K., Eleftheriades, M., Christopoulos, P., Panoskaltsis, T., Stamoula, E., & Iliopoulos, D. C. (2023). Tumor Markers and Their Diagnostic Significance in Ovarian Cancer. *Life, 13*(8), 1689.

Matsuura, Y., Joti, Y., Bagautdinov, B., & Yutani, K. (2019). Evaluating the strengths of salt bridges in the CutA1 protein using molecular dynamic simulations: a comparison of different force fields. *FEBS Open Bio, 9*(11), 1939–1956.

Matulonis, U. A., Sood, A. K., Fallowfield, L., Howitt, B. E., Sehouli, J., & Karlan, B. Y. (2016). Ovarian cancer. *Nature Reviews Disease Primers, 2*, 1–22.

Mbakaza, O., & Vangu, M.-D.-T. W. (2022). 18F-FDG PET/CT Imaging: Normal Variants, Pitfalls, and Artifacts Musculoskeletal, Infection, and Inflammation. *Frontiers in Nuclear Medicine, 2*, 1-11.

McGuffin, L. J., Edmunds, N. S., Genc, A. G., Alharbi, S. M. A., Salehe, B. R., & Adiyaman, R. (2023). Prediction of protein structures, functions and interactions using the IntFOLD7, MultiFOLD and ModFOLDdock servers. *Nucleic Acids Research, 51*(W1), W274–W280.

Medeiros, L. R., Rosa, D. D., da Rosa, M. I., & Bozzetti, M. C. (2009). Accuracy of CA 125 in the diagnosis of ovarian tumors: A quantitative systematic review. European *Journal of Obstetrics & Gynecology and Reproductive Biology, 142*(2), 99–105.

Mehmood, S., Khan, A. Z., Bilal, M., Sohail, A., & Iqbal, H. M. N. (2019). Aptamer-based biosensors: a novel toolkit for early diagnosis of cancer. *Materials Today Chemistry, 12*, 353–360.

Mehta, J., Rouah-Martin, E., Van Dorst, B., Maes, B., Herrebout, W., Scippo, M.-L., Dardenne, F., Blust, R., & Robbens, J. (2012). Selection and Characterization of PCB-Binding DNA Aptamers. *Analytical Chemistry, 84*(3), 1669–1676.

Meng, X., Li, J., Wu, Y., Cao, X., & Zhang, Z. (2023). Rational design of hairpin aptamer using intrinsic disorder mechanism to enhance sensitivity of aptamer folding-based electrochemical sensor for tobramycin. *Sensors and Actuators B: Chemical, 394*, 134354.

Meng, X.-Y., Zhang, H.-X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design, 7*(2), 146–157.

Mikkola, S., Lönnberg, T., & Lönnberg, H. (2018). Phosphodiester models for cleavage of nucleic acids. *Beilstein Journal of Organic Chemistry, 14*, 803–837.

Minchin, S., & Lodge, J. (2019). Understanding biochemistry: structure and function of nucleic acids. *Essays in Biochemistry, 63*(4), 433–456.

Mirón-Mérida, V. A., González-Espinosa, Y., Collado-González, M., Gong, Y. Y., Guo, Y., & Goycoolea, F. M. (2021). Aptamer–Target–Gold Nanoparticle Conjugates for the Quantification of Fumonisin B1. *Biosensors, 11*(1), 18, 1-18.

Mishra, S. K., & Koča, J. (2018). Assessing the Performance of MM/PBSA, MM/GBSA, and QM–MM/GBSA Approaches on Protein/Carbohydrate Complexes: Effect of Implicit Solvent Models, QM Methods, and Entropic Contributions. *The Journal of Physical Chemistry B, 122*(34), 8113–8121.

Mohammed, S., Sa'idu, H., Manzo, J. O., Hussein, A. S., Alhassan, J. A., Muhammad, U. Z., & Aliyu, A. H. (2022). Prediction and Validation of 3-Dimensional Structure of Rice OsTHIC Abiotic Stress Responsive Gene. *Asian Journal of Plant Biology, 4*(1), 1–4.

Montagnana, M., Benati, M., & Danese, E. (2017). Circulating biomarkers in epithelial ovarian cancer diagnosis: From present to future perspective. *Annals of Translational Medicine, 5*(13), 1–11.

Moore, R. G., Brown, A. K., Miller, M. C., Skates, S., Allard, W. J., Verch, T., Steinhoff, M., Messerlian, G., DiSilvestro, P., Granai, C. O., & Bast Jr., R. C. (2008). The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. *Gynecologic Oncology, 108*(2), 402–408.

Moore, R. G., McMeekin, D. S., Brown, A. K., DiSilvestro, P., Miller, M. C., Allard, W. J., Gajewski, W., Kurman, R., Bast, R. C., & Skates, S. J. (2009). A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecologic Oncology*, *112*(1), 40–46.

Moore, R. G., Miller, M. C., Steinhoff, M. M., Skates, S. J., Lu, K. H., Lambert-Messerlian, G., & Bast Jr, R. C. (2012). Serum HE4 levels are less frequently elevated than CA125 in women with benign gynecologic disorders. *American Journal of Obstetrics & Gynecology*, *206*(4), 351.e1-351.e8.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry, 30*(16), 2785–2791.

Morris, G. M., Huey, R., & Olson, A. J. (2008). Using autodock for ligand-receptor docking. *Current protocols in bioinformatics*, *24*(1), 8-14.

Moshref, Z. S., Jalali, T., Rezaei Adriani, R., Soltati, E., & Mousavi Gargari, S. L. (2023). Aptamer-based diagnosis of various SARS-CoV2 strains isolated from clinical specimens. *Heliyon, 9*(6), e16458.

Mu, Z.-C., Tan, Y.-L., Liu, J., Zhang, B.-G., & Shi, Y.-Z. (2023). Computational Modeling of DNA 3D Structures: From Dynamics and Mechanics to Folding. *Molecules, 28*(12), 4833.

Muinao, T., Deka Boruah, H. P., & Pal, M. (2019). Multi-biomarker panel signature as the key to diagnosis of ovarian cancer. *Heliyon, 5*(12), e02826.

Mu, Z. C., Tan, Y. L., Liu, J., Zhang, B. G., & Shi, Y. Z. (2023). Computational modeling of DNA 3D structures: From dynamics and mechanics to folding. *Molecules*, *28*(12), 4833.

Nahar, B. A., Saha, R., & Das, C. (2019). Diagnostic utility of Immunohistochemical expression of HE4 in Epithelial Ovarian Neoplasm. *Annals of Pathology and Laboratory Medicine, 6*(11), 567–571.

Navien, T. N., Thevendran, R., Hamdani, H. Y., Tang, T.-H., & Citartan, M. (2021). In silico molecular docking in DNA aptamer development. *Biochimie, 180*, 54–67.

Nenciu, A.-E., Nenciu, C. G., Fodoroiu, R., Scalețchi, O., Șandru, F., & Dumitrașcu, M. C. (2020). Tumor Markers in Ovarian Cancer. *Journal of Surgical Sciences, 7*(2), 79.

Nene, T., Yadav, M., & Yadav, H. S. (2022). Plant catalase in silico characterization and phylogenetic analysis with structural modeling. *Journal of Genetic Engineering and Biotechnology, 20*(1), 125.

Neogi, S. S., & Srivastava, L. M. (2014). Elevated tumour marker CA125: Interpretations in clinical practice. *Current Medicine Research and Practice, 4*(5), 214–218.

Nezhat, F. R., Pejovic, T., Finger, T. N., & Khalil, S. S. (2013). Role of minimally invasive surgery in ovarian cancer. *Journal of Minimally Invasive Gynecology, 20*(6), 754–765.

Nguyen, N. T., Nguyen, T. H., Pham, T. N. H., Huy, N. T., Bay, M. Van, Pham, M. Q., Nam, P. C., Vu, V. V, & Ngo, S. T. (2020). Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. *Journal of Chemical Information and Modeling, 60*(1), 204–211.

Ni, S., Zhuo, Z., Pan, Y., Yu, Y., Li, F., Liu, J., Wang, L., Wu, X., Li, D., Wan, Y., Zhang, L., Yang, Z., Zhang, B.-T., Lu, A., & Zhang, G. (2021). Recent Progress in Aptamer Discoveries and Modifications for Therapeutic Applications. *ACS Applied Materials & Interfaces, 13*(8), 9500–9519.

Ni, X., Castanares, M., Mukherjee, A., & Lupold, S. E. (2011). Nucleic acid aptamers: clinical applications and promising new horizons. *Current Medicinal Chemistry, 18*(27), 4206–4214.

Nimjee, S. M., White, R. R., Becker, R. C., & Sullenger, B. A. (2017). Aptamers as Therapeutics. *Annual Review of Pharmacology and Toxicology, 57*(1), 61–79.

Nooranian, S., Mohammadinejad, A., Mohajeri, T., Aleyaghoob, G., & Kazemi Oskuee, R. (2022). Biosensors based on aptamer-conjugated gold nanoparticles: A review. *Biotechnology and Applied Biochemistry, 69*(4), 1517–1534.

Nwabuobi, C., Arlier, S., Schatz, F., Guzeloglu-Kayisli, O., Lockwood, C. J., & Kayisli, U. A. (2017). hCG: Biological Functions and Clinical Applications. *International Journal of Molecular Sciences*, *18*(10), 1-15.

Odeh, F., Nsairat, H., Alshaer, W., Ismail, M. A., Esawi, E., Qaqish, B., Bawab, A. Al, & Ismail, S. I. (2019). Aptamers Chemistry: Chemical Modifications and Conjugation Strategies. *Molecules, 25*(1), 1-51.

Ojha, Y. R., Giovannucci, D. R., & Cameron, B. D. (2021). Selection and characterization of structure-switching DNA aptamers for the salivary peptide histatin 3. *Journal of Biotechnology, 327*, 9–17.

Oliveira, R., Pinho, E., Sousa, A. L., Dias, Ó., Azevedo, N. F., & Almeida, C. (2022). Modelling aptamers with nucleic acid mimics (NAM): From sequence to three-dimensional docking. *PLoS ONE, 17*(3), e0264701 1-14.

Ostreni, I., Colatosti, A., Basile, E. J., & Rafa, O. (2022). Elevated Beta-Human Chorionic Gonadotropin in a Non-pregnant Female with Altered Kidney Function. *Cureus, 14*(4), e23747.

Owoloye, A. J., Ligali, F. C., Enejoh, O. A., Musa, A. Z., Aina, O., Idowu, E. T., & Oyebola, K. M. (2022). Molecular docking, simulation and binding free energy analysis of small molecules as PfHT1 inhibitors. *PLOS ONE, 17*(8), e0268269, 1-18.

Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: a review. *Biophysical Reviews, 9*(2), 91–102.

Park, H., Kim, D. E., Ovchinnikov, S., Baker, D., & DiMaio, F. (2018). Automatic structure prediction of oligomeric assemblies using Robetta in CASP12. *Proteins: Structure, Function, and Bioinformatics, 86*(S1), 283–291.

Park, S. W., Lee, B. H., Song, S. H., & Kim, M. K. (2023). Revisiting the Ramachandran plot based on statistical analysis of static and dynamic characteristics of protein structures. *Journal of Structural Biology*, *215*(1), 107939.

Pei Yin, K. (2019). Diagnosis and Management of Ovarian Cancer, Current Status and Future Potential. *Borneo Journal of Medical Sciences (BJMS), 13*(2 SE-Case report), 49.

Perrakis, A., & Sixma, T. K. (2021). AI revolutions in biology: The joys and perils of AlphaFold. *EMBO reports*, *22*(11), e54046.

Petsko, G. A., & Ringe, D. (2004). *Protein structure and function*. New Science.

Phillips, J. L., Colvin, M. E., & Newsam, S. (2011). Validating clustering of molecular dynamics simulations using polymer models. *BMC Bioinformatics, 12*(1), 445.

Poater, J., Swart, M., Bickelhaupt, F. M., & Fonseca Guerra, C. (2014). B-DNA structure and stability: the role of hydrogen bonding, π–π stacking interactions, twist-angle, and solvation. *Organic & Biomolecular Chemistry, 12*(26), 4691–4700.

Popot, J. L., & Engelman, D. M. (2000). Helical Membrane Protein Folding, Stability, and Evolution. *Annual Review of Biochemistry, 69*(1), 881–922.

Poustforoosh, A., Faramarz, S., Nematollahi, M. H., Hashemipour, H., Negahdaripour, M., & Pardakhty, A. (2022). In silico SELEX screening and statistical analysis of newly designed 5mer peptide-aptamers as Bcl-xl inhibitors using the Taguchi method. *Computers in Biology and Medicine, 146*, 105632.

Prat, J., Belhadj, H., Berek, J., Bermudez, A., Bhatla, N., Cain, J., Denny, L., Fujiwara, K., Hacker, N., Åvall-Lundqvist, E., Mutch, D., Odicino, F., Pecorelli, S., Prat, J., Quinn, M., Seoud, M. A. F., & Shrivastava, S. K. (2015). Figo’s staging classification for cancer of the ovary, fallopian tube, and peritoneum: Abridged republication. *Journal of Gynecologic Oncology, 26*(2), 87–89.

Pradeepkiran, J. A., Sainath, S. B., Balne, P. K., & Bhaskar, M. (2021). Computational modeling and evaluation of best potential drug targets through comparative modeling. In *Brucella Melitensis* (pp. 39-78). Academic Press.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open-source molecular simulation toolkit. *Bioinformatics, 29*(7), 845–854.

Rakhshani, H., Dehghanian, E., & Rahati, A. (2019). Enhanced GROMACS: toward a better numerical simulation framework. *Journal of Molecular Modeling, 25*(12), 355.

Ramírez, D., & Caballero, J. (2018). Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules, 23*(5), 1038, 1-17.

Rampogu, S., Lee, G., Park, J. S., Lee, K. W., & Kim, M. O. (2022). Molecular Docking and Molecular Dynamics Simulations Discover Curcumin Analogue as a Plausible Dual Inhibitor for SARS-CoV-2. *International Journal of Molecular Sciences, 23*(3), 1771, 1-20.

Ranade, S. S., & Ramalingam, R. (2020). Hydrogen bonds in anoplin peptides aid in identification of a structurally stable therapeutic drug scaffold. *Journal of Molecular Modeling, 26*(6), 155.

Ranganathan, V., Srinivasan, S., Singh, A., & DeRosa, M. C. (2020). An aptamer-based colorimetric lateral flow assay for the detection of human epidermal growth factor receptor 2 (HER2). *Analytical Biochemistry, 588*, 113471.

Rasouli Jazi, H. R., Zeinoddini, M., & Arab, S. S. (2023). A Novel in silico SELEX Method to Screen and Identify Aptamers against Vibrio cholerae. *Current Computer-Aided Drug Design, 19*(6), 416–424.

Read, R. J., Baker, E. N., Bond, C. S., Garman, E. F., & Van Raaij, M. J. (2023). AlphaFold and the future of structural biology. *Acta Crystallographica Section F: Structural Biology Communications*, *79*(7), 166-168.

Ream, J. A., Lewis, L. K., & Lewis, K. A. (2016). Rapid agarose gel electrophoretic mobility shift assay for quantitating protein: RNA interactions. *Analytical Biochemistry, 511*, 36–41.

Ream, J. A., Lewis, L. K., & Lewis, K. A. (2019). Horizontal Agarose Gel Mobility Shift Assay for Protein-RNA Complexes. *Electrophoretic Separation of Proteins: Methods and Protocols*, 363-370.

Reid, F. (2020). World Ovarian Cancer Coalition Atlas. 2020. *Global Trends in Incidence, Mortality, and Survival*, 1-42.

Reyners, A. K. L., Broekman, K. E., Glaudemans, A. W. J. M., Brouwers, A. H., Arts, H. J. G., van der Zee, A. G. J., ... & Jalving, M. (2016). Molecular imaging in ovarian cancer. *Annals of Oncology, 27*, i23-i29.

Robertson, M. J., Tirado-Rives, J., & Jorgensen, W. L. (2015). Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. Journal of Chemical Theory and Computation, 11(7), 3499–3509.

Rojas, V., Hirshfield, K. M., Ganesan, S., & Rodriguez-Rodriguez, L. (2016). Molecular Characterization of Epithelial Ovarian Cancer: Implications for Diagnosis and Treatment. *International Journal of Molecular Sciences, 17*(12), 1-23.

Röthlisberger, P., & Hollenstein, M. (2018). Aptamer chemistry. *Advanced Drug Delivery Reviews, 134,* 3–21.

Sabri, M. Z. (2021). In silico Assisted Design of Aptamers for Hepatitis B Antigen Detection. Universiti Kuala Lumpur.

Sabri, M. Z., Abdul Hamid, A. A., Sayed Hitam, S. M., & Abdul Rahim, M. Z. (2019). In silico screening of aptamers configuration against hepatitis B surface antigen. *Advances in bioinformatics*, *2019*(1), 6912914.

Sabri, M. Z., Hamid, A. A. A., Hitam, S. M. S., & Rahim, Mohd. Z. A. (2020). The assessment of three-dimensional modelling design for single strand DNA aptamers for computational chemistry application. *Biophysical Chemistry, 267*, 106492.

Salmaso, V., & Moro, S. (2018). Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. In *Frontiers in Pharmacology, 9*, 923.

Sampson, T. (2003). Aptamers and SELEX: the technology. World Patent Information, 25(2), 123–129.

Santotoribio, J. D., Garcia-de la Torre, A., Cañavate-Solano, C., Arce-Matute, F., Sanchez-del Pino, M. J., & Perez-Ramos, S. (2016). Cancer antigens 19.9 and 125 as tumor markers in patients with mucinous ovarian tumors. *European Journal of Gynaecological Oncology, 37*(1), 26–29.

Sarker, D. K., Ray, P., Rouf, R., Shilpi, J. A., & Uddin, S. J. (2023). In Silico Molecular Docking and Dynamic Investigations of Bioactive Phytoconstituents from Fenugreek Seeds as a Potent Drug against DPP-IV Enzyme. *ACS Food Science & Technology*, *3*(9), 1423-1439.

Sarzynska, J., Popenda, M., Antczak, M., & Szachniuk, M. (2023). RNA tertiary structure prediction using RNAComposer in CASP15. *Proteins: Structure, Function, and Bioinformatics*, *91*(12), 1790–1799.

Sato, K., Hamada, M., Asai, K., & Mituyama, T. (2009). CENTROIDFOLD: a web server for RNA secondary structure prediction. *Nucleic acids research*, *37*(suppl_2), W277-W280.

Savinova, A.R., & Gataullin, I.G. (2022). Early diagnostics and screening for ovarian cancer. *Kazan medical journal, 103*(3), 476–483

Sawal, H. A., Nighat, S., Safdar, T., & Anees, L. (2023). Comparative In Silico Analysis and Functional Characterization of TANK-Binding Kinase 1–Binding Protein 1. *Bioinformatics and Biology Insights*, *17*, 11779322231164828.

Scaletta, G., Plotti, F., Luvero, D., Capriglione, S., Montera, R., Miranda, A., Lopez, S., Terranova, C., De Cicco Nardone, C., & Angioli, R. (2017). The role of novel biomarker HE4 in the diagnosis, prognosis and follow-up of ovarian cancer: a systematic review. *Expert Review of Anticancer Therapy, 17*(9), 827–839.

Scambia, G., Panici, P. B., Baiocchi, G., Perrone, L., Greggi, S., & Mancuso, S. (1988). CA 15-3 as a tumor marker in gynecological malignancies. *Gynecologic Oncology, 30*(2), 265–273.

Scarà, S., Bottoni, P., & Scatena, R. (2015). CA 19-9: biochemical and clinical aspects. *Advances in Cancer Biomarkers: From biochemistry to clinic for a critical revision*, 247-260.

Seeliger, D., & De Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of Computer-Aided Molecular Design, 24*(5), 417–422.

Seo, M., Lei, L., & Egli, M. (2019). Label-free electrophoretic mobility shift assay (EMSA) for measuring dissociation constants of protein-RNA complexes. *Current Protocols in Nucleic Acid Chemistry, 76*, e70, 1-12.

Sett, A. (2020). Aptamers: Magic bullet for Theranostic applications. In *Theranostics-An Old Concept in New Clothing*. IntechOpen.

Shaban, S. M., & Kim, D.-H. (2021). Recent advances in aptamer sensors. *Sensors, 21*(3), 979.

Shah, R., Eldridge, D., Palombo, E., & Harding, I. (2014). Optimisation and stability assessment of solid lipid nanoparticles using particle size and zeta potential. *Journal of Physical Science, 25*(1), 59-75.

Shahbazlou, S. V., Vandghanooni, S., Dabirmanesh, B., Eskandani, M., & Hasannia, S. (2023). Biotinylated aptamer-based SPR biosensor for detection of CA125 antigen. *Microchemical Journal, 194*, 109276.

Shao, J., Tanner, S. W., Thompson, N., & Cheatham, T. E. (2007). Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *Journal of Chemical Theory and Computation, 3*(6), 2312–2334.

Sharma, J., Kumar Bhardwaj, V., Singh, R., Rajendran, V., Purohit, R., & Kumar, S. (2021). An in-silico evaluation of different bioactive molecules of tea for their inhibition potency against non-structural protein-15 of SARS-CoV-2. *Food Chemistry, 346*, 128933.

Sharma, P., Hasan, M. R., Khanuja, M., Rawal, R., Pilloton, R., & Narang, J. (2023). Aptamer-based silver nanoparticle decorated paper platform for electrochemical detection ovarian cancer biomarker PDGF. *Materials Chemistry and Physics, 306*, 128114.

Shen, Y., Wang, Y., Jiang, X., Lu, L., Wang, C., Luo, W., Zhang, Y., Li, P., Du, Z., & Dai, T. (2018). Preparation and characterization of a high-affinity monoclonal antibody against human epididymis protein-4. *Protein Expression and Purification, 141*, 44–51.

Shigdar, S., Macdonald, J., O'Connor, M., Wang, T., Xiang, D., Al.Shamaileh, H., Qiao, L., Wei, M., Zhou, S.-F., Zhu, Y., Kong, L., Bhattacharya, S., Li, C., & Duan, W. (2013). Aptamers as Theranostic Agents: Modifications, Serum Stability and Functionalisation. *Sensors, 13*(10), 13624–13637.

Shin, K. H., Kim, H. H., Kwon, B. S., Suh, D. S., Joo, J. K., & Kim, K. H. (2020). Clinical Usefulness of Cancer Antigen (CA) 125, Human Epididymis 4, and CA72-4 Levels and Risk of Ovarian Malignancy Algorithm Values for Diagnosing Ovarian Tumors in Korean Patients with and Without Endometriosis. *Annals of Laboratory Medicine, 40*(1), 40–47.

Shin, W. R., Sekhon, S. S., Rhee, S.-K., Ko, J. H., Ahn, J.-Y., Min, J., & Kim, Y.-H. (2018). Aptamer-Based Paper Strip Sensor for Detecting Vibrio fischeri. *ACS Combinatorial Science, 20*(5), 261–268.

Sincavage, J., Msosa, V. J., Katete, C., Purcell, L. N., & Charles, A. (2021). Postoperative Complications and Risk of Mortality after Laparotomy in a Resource-Limited Setting. *The Journal of Surgical Research, 260*, 428–435.

Singh, S., & Singh, A. N. (2018). Ethidium bromide: Is a stain turning into a pollutant? A synthesis on its status, waste management, monitoring challenges and ecological risks to the environment. *Int. J. Res. Anal. Rev*, *4*, 226-233.

Smith, A., Dong, X., & Raghavan, V. (2022). An Overview of Molecular Dynamics Simulation for Food Products and Processes. *Processes, 10*(1), 119.

Smith, B. C. (2020). Organic Nitrogen Compounds, VII: Amides - The rest of the story, 35(1), 10-15.

Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research, 33*, W244–W248.

Sopik, V., Akbari, M. R., & Narod, S. A. (2015). Genetic testing for RAD51C mutations: in the clinic and community. *Clinical Genetics, 88*(4), 303–312.

Stawicki, C. M., Rinker, T. E., Burns, M., Tonapi, S. S., Galimidi, R. P., Anumala, D., Robinson, J. K., Klein, J. S., & Mallick, P. (2021). Modular fluorescent nanoparticle DNA probes for detection of peptides and proteins. *Scientific Reports, 11*(1), 19921.

Stewart, C., Ralyea, C., & Lockwood, S. (2019). Ovarian Cancer: An Integrated Review. *Seminars in Oncology Nursing, 35*(2), 151–156.

Stirling, D., Evans, D. G. R., Pichert, G., Shenton, A., Kirk, E. N., Rimmer, S., Steel, C. M., Lawson, S., Busby-earle, R. M. C., Walker, J., Lalloo, F. I., Eccles, D. M., Lucassen, A. M., & Porteous, M. E. (2005). Screening for Familial Ovarian Cancer: Failure of Current Protocols to Detect Ovarian Cancer at an Early Stage According to the International Federation of Gynecology and Obstetrics System. *Journal of Clinical Oncology, 23*(24), 5588–5596.

Stourac, J., Vavra, O., Kokkonen, P., Filipovic, J., Pinto, G., Brezovsky, J., Damborsky, J., & Bednar, D. (2019). Caver Web 1.0: identification of tunnels and channels in proteins and analysis of ligand transport. *Nucleic Acids Research, 47*(W1), W414–W422.

Sui, H., & Imamichi, T. (2022). A DNA Pull-Down Assay with Diversity Forms of Competitor for Detecting or Evaluating Protein–DNA Interactions. In *DNA-Protein Interactions: Methods and Protocols* (pp. 1-10). New York, NY: Springer US.

Sullivan, R., Adams, M. C., Naik, R. R., & Milam, V. T. (2019). Analyzing Secondary Structure Patterns in DNA Aptamers Identified via CompELS. *Molecules, 24*(8), 1572.

Sultana, N., Shoukat, S., Nausheen, S., & Memon, B. (2021). Diagnostic Accuracy of PET/CT Scan in Evaluation of Clinically Suspected Recurrent Ovarian Cancer. *Journal of Fatima Jinnah Medical University,14*, 156-160.

Sun, H., Li, Y., Tian, S., Xu, L., & Hou, T. (2014). Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Physical Chemistry Chemical Physics, 16*(31), 16719–16729.

Sun, L., Shen, K., Zhang, J., Wan, W., Cao, W., Wang, Z., & Guo, C. (2021). Aptamer based surface plasma resonance assay for direct detection of neuron specific enolase and progastrin-releasing peptide (31-98). *RSC Advances, 11*(51), 32135–32142.

Swain, S. S., Paidesetty, S. K., Dehury, B., Sahoo, J., Vedithi, S. C., Mahapatra, N., Hussain, T., & Padhy, R. N. (2018). Molecular docking and simulation study for synthesis of alternative dapsone derivative as a newer antileprosy drug in multidrug therapy. *Journal of Cellular Biochemistry, 119*(12), 9838–9852.

Sztandera, K., Gorzkiewicz, M., & Klajnert-Maculewicz, B. (2018). Gold nanoparticles in cancer treatment. *Molecular pharmaceutics, 16*(1), 1-23.

Tagami, S. (2023). Why we are made of proteins and nucleic acids: Structural biology views on extraterrestrial life. *Biophysics and Physicobiology, 20*(2), e200026.

Takahashi, S., & Sugimoto, N. (2021). Watson–Crick versus Hoogsteen Base Pairs: Chemical Strategy to Encode and Express Genetic Information in Life. *Accounts of Chemical Research*, *54*(9), 2110–2120.

Takita, S., Nabok, A., Smith, D., & Lishchuk, A. (2021). Spectroscopic Ellipsometry Detection of Prostate Cancer Bio-Marker PCA3 Using Specific Non-Labeled Aptamer: Comparison with Electrochemical Detection. *Chemistry Proceedings,5*(1), 65.

Tan, S. Y., Acquah, C., Sidhu, A., Ongkudon, C. M., Yon, L. S., & Danquah, M. K. (2016). SELEX Modifications and Bioanalytical Techniques for Aptamer-Target Binding Characterization. *Critical Reviews in Analytical Chemistry, 46*(6), 521–537.

Teh, B. H., Yong, S. L., Sim, W. W., Lau, K. B., & Suharjono, H. N. (2018). Evaluation in the predictive value of serum human epididymal protein 4 (HE4), cancer antigen 125 (CA 125) and a combination of both in detecting ovarian malignancy. *Hormone Molecular Biology and Clinical Investigation, 35*(1), 20180029.

Thevendran, R., Tang, T.-H., & Citartan, M. (2023). In-silico selection employing rigid docking and molecular dynamic simulation in selecting DNA aptamers against androgen receptor. *Biotechnology Journal*, *18*(4), 2200092.

Trott, O., & 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, 31*(2), 455–461.

Tsuyoshi, H., Tsujikawa, T., Yamada, S., Okazawa, H., & Yoshida, Y. (2020). Diagnostic value of [18F]FDG PET/MRI for staging in patients with ovarian cancer. *EJNMMI Research, 10*(1), 117.

Tucker, M. R., Piana, S., Tan, D., Levine, M. V, & Shaw, D. E. (2022). Development of Force Field Parameters for the Simulation of Single- And Double-Stranded DNA Molecules and DNA–Protein Complexes. *The Journal of Physical Chemistry B, 126*(24), 4442-4457.

Ueland, F. (2017). A Perspective on Ovarian Cancer Biomarkers: Past, Present and Yet-To-Come. *Diagnostics, 7*(1), 14.

U.S. Preventive Services Task Force. (2018). Screening for ovarian cancer: Recommendation statement. *American Family Physician, 97*(12), 814A-814C.

Valuchova, S., Fulnecek, J., Petrov, A. P., Tripsianes, K., & Riha, K. (2016). A rapid method for detecting protein-nucleic acid interactions by protein induced fluorescence enhancement. *Scientific Reports, 6*(1), 39653.

van Heesch, T., Bolhuis, P. G., & Vreede, J. (2023). Decoding dissociation of sequence-specific protein–DNA complexes with non-equilibrium simulations. *Nucleic Acids Research, 51*(22), 12150–12160.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., & Mackerell, A. D. J. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry, 31*(4), 671–690.

Vaverkova, V., Vrana, O., Adam, V., Pekarek, T., Jampilek, J., & Babula, P. (2014). The study of naphthoquinones and their complexes with DNA by using Raman spectroscopy and surface enhanced Raman spectroscopy: new insight into interactions of DNA with plant secondary metabolites. *BioMed Research International, 2014*, 461393.

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics, 52*(4), 609–623.

Vieira, T. F., & Sousa, S. F. (2019). Comparing AutoDock and Vina in Ligand/Decoy Discrimination for Virtual Screening. *Applied Sciences, 9*(21), 4538.

Vo, T. D., Schneider, A. L., Wilson, W. D., & Poon, G. M. K. (2021). Salt bridge dynamics in protein/DNA recognition: a comparative analysis of Elk1 and ETV6. *Physical Chemistry Chemical Physics: PCCP, 23*(24), 13490–13502.

Vogel, R., Pal, A. K., Jambhrunkar, S., Patel, P., Thakur, S. S., Reátegui, E., Parekh, H. S., Saá, P., Stassinopoulos, A., & Broom, M. F. (2017). High-Resolution Single Particle Zeta Potential Characterisation of Biological Nanoparticles using Tunable Resistive Pulse Sensing. *Scientific Reports, 7*(1), 17479.

Wagner, V., Winn, H., Newtson, A., Bender, D., & McDonald, M. (2018). hCG production by mucinous adenocarcinoma of the ovary in a reproductive aged woman. *Gynecologic Oncology Reports, 26*, 102–104.

Wan, Q., Liu, X., & Zu, Y. (2021). Theranostics Oligonucleotide aptamers for pathogen detection and infectious disease control. *Theranostics, 11*(18), 9133–9161.

Wang, H., Liu, P., Xu, H., & Dai, H. (2021). Early diagonosis of ovarian cancer: serum HE4, CA125 and ROMA model. *American Journal of Translational Research, 13*(12), 14141–14148.

Wang, J., Fang, X., Zhang, C., Ji, H., Pang, Q., Li, X., Luo, Z., Wu, Q., & Zhang, L. (2021). Development of Aptamer-Based Molecular Tools for Rapid Intraoperative Diagnosis and In Vivo Imaging of Serous Ovarian Cancer. *ACS Applied Materials & Interfaces, 13*(14), 16118–16126.

Wang, Q. L., Cui, H.-F., Du, J.-F., Lv, Q.-Y., & Song, X. (2019). In silico post-SELEX screening and experimental characterizations for acquisition of high affinity DNA aptamers against carcinoembryonic antigen. *RSC Advances, 9*(11), 6328–6334.

Wang, T., Chen, C., Larcher, L. M., Barrero, R. A., & Veedu, R. N. (2019). Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. *Biotechnology Advances, 37*(1), 28–50.

Wang, W., Feng, C., Han, R., Wang, Z., Ye, L., Du, Z., Wei, H., Zhang, F., Peng, Z., & Yang, J. (2023). trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nature Communications, 14*(1), 7266.

Wang, X., Wang, Y., Zheng, L., & Chen, J. (2014). Molecular Dynamics Simulation in RNA Interference. *Current Medicinal Chemistry, 21*(17), 1968-1975.

Wang, Y., Liu, M., & Gao, J. (2020). Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. *Proceedings of the National Academy of Sciences, 117*(25), 13967–13974.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., & Bordoli, L. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research, 46*(W1), W296–W303.

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature, 171*(4356), 737–738.

Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics, 54*(1), 5–6.

World Health Organization. (2024). *Estimated Number of New Cases from 2022 to 2030, Females: Ovary.* https://gco.iarc.fr/tomorrow/en/dataviz/isotype?sexes=2&group_populations=0&mode=population&types=0&multiple_populations=0&single_unit=10000&cancers=25&years=2030&age_start=4&age_end=16&populations=935

Wu, X., Shaikh, A. B., Yu, Y., Li, Y., Ni, S., Lu, A., & Zhang, G. (2017). Potential Diagnostic and Therapeutic Applications of Oligonucleotide Aptamers in Breast Cancer. *International Journal of Molecular Sciences, 18*(9), 1851.

Wu, Y., Ali, M. R. K., Chen, K., Fang, N., & El-Sayed, M. A. (2019). Gold nanoparticles in biological optical imaging. Nano Today, 24, 120–140. Yan, Y., & Huang, T. (2023). The Interactome of Protein, DNA, and RNA. In T. Huang, J. Yang, & G. Tian (Eds.), *Liquid Biopsies: Methods and Protocols, 2695*, 89–110. Springer US.

Xu, S., Kondal, M. D., Ahmad, A., Zhu, R., Fan, L., Zaborniak, P., Madden, K. S., de Souza, J. V, & Bronowska, A. K. (2023). Mechanistic Investigation of the Androgen Receptor DNA-Binding Domain and Modulation via Direct Interactions with DNA Abasic Sites: Understanding the Mechanisms Involved in Castration-Resistant Prostate Cancer. *International Journal of Molecular Sciences, 24*(2), 1270, 1-16.

Xu, X., Zhao, P., & Chen, S.-J. (2014). Vfold: A Web Server for RNA Structure and Folding Thermodynamics Prediction. *PLoS ONE, 9*(9), e107504, 1-7.

Yan, Y., & Huang, S.-Y. (2020). Modeling Protein–Protein or Protein–DNA/RNA Complexes Using the HDOCK Webserver BT- Protein Structure Prediction. 217–229). Springer US.

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., & Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences of the United States of America, 117*(3), 1496–1503.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2014). The I-TASSER suite: Protein structure and function prediction. *Nature Methods, 12*(1), 7–8.

Yang, J., & Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research, 43*(W1), W174–W181.

Yang, J., & Zhang, Y. (2016). Protein Structure and Function Prediction Using I-TASSER. *Current Protocols in Bioinformatics, 52,* 5.8.1-5.815.

Yang, J., & Zhang, Y. (2023). *I-TASSER Protein Structure & Function Predictions*. https://zhanggroup.org/I-TASSER/

Yang, R., Wu, S., Wang, S., Rubino, G., Nickels, J. D., & Cheng, X. (2022). Refinement of SARS-CoV-2 envelope protein structure in a native-like environment by molecular dynamics simulations. *Frontiers in Molecular Biosciences, 9,* 1-15.

Yang, X., Wu, G., Yang, F., He, L., Xie, X., Li, L., Yang, L., Ma, Y., Zhang, Q., Chen, J., Zou, S., Han, Q., Wang, Y., Liu, S., Li, J., Han, B., & Liu, K. (2021). Elevated LINC00909 Promotes Tumor Progression of Ovarian Cancer via Regulating the miR-23b-3p/MRC2 Axis. *Oxidative Medicine and Cellular Longevity, 2021*, 1-23.

Yang, X., Zhuo, Y., Zhu, S., Luo, Y., Feng, Y., & Xu, Y. (2015). Selectively assaying CEA based on a creative strategy of gold nanoparticles enhancing silver nanoclusters' fluorescence. *Biosensors & Bioelectronics, 64,* 345–351.

Yin, R., Feng, B. Y., Varshney, A., & Pierce, B. G. (2022). Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Science, 31*(8), e4379.

Ylilauri, M., & Pentikäinen, O. (2013). MMGBSA as a Tool to Understand the Binding Affinities of Filamin-Peptide Interactions. *Journal of Chemical Information and Modeling, 53*(10), 2626-2633.

Yu, B., Pletka, C. C., & Iwahara, J. (2020). NMR Observation of Intermolecular Hydrogen Bonds between Protein Tyrosine Side-Chain OH and DNA Phosphate Groups. *The Journal of Physical Chemistry B, 124*(6), 1065–1070.

Yu, H., Alkhamis, O., Canoura, J., Liu, Y., & Xiao, Y. (2021). Advances and Challenges in Small-Molecule DNA Aptamer Isolation, Characterization, and Sensor Development. *Angewandte Chemie*, *133*(31), 16938–16961.

Zarandi, S. H., Behbahani, M., & Mohabatkar, H. (2020). In Silico Selection of Gp120 ssDNA Aptamer to HIV-1. *SLAS Discovery*, *25*(9), 1087–1093.

Zavyalova, E., Turashev, A., Novoseltseva, A., Legatova, V., Antipova, O., Savchenko, E., Balk, S., Golovin, A., Pavlova, G., & Kopylov, A. (2020). Pyrene-Modified DNA Aptamers with High Affinity to Wild-Type EGFR and EGFRvIII. *Nucleic Acid Therapeutics*, *30*(3), 175–187.

Zeff, N. (2018). Role of laparoscopy in initial tumour staging in advanced epithelial ovarian cancer: a systematic review. *Pleura and Peritoneum, 3*(1), 1-9.

Zhang, F., Li, S., Cao, K., Wang, P., Su, Y., Zhu, X., & Wan, Y. (2015). A microfluidic love-wave biosensing device for PSA detection based on an aptamer beacon probe. *Sensors*, *15*(6), 13839-13850.

Zhang, L., Chen, Y., & Wang, K. (2019). Comparison of CA125, HE4, and ROMA index for ovarian cancer diagnosis. Current Problems in Cancer, 43(2), 135–144.

Zhang, M., Cheng, S., Jin, Y., Zhao, Y., & Wang, Y. (2021). Roles of CA125 in diagnosis, prediction, and oncogenesis of ovarian cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 1875*(2), 188503.

Zhang, N., Chen, Z., Liu, D., Jiang, H., Zhang, Z.-K., Lu, A., Zhang, B.-T., Yu, Y., & Zhang, G. (2021). Structural Biology for the Molecular Insight between Aptamers and Target Proteins. *International Journal of Molecular Sciences, 22*(8), 4093, 1-27.

Zhang, Q., Rui-Xue, L. I., Xin, C., Xing-Xing, H. E., Ai-Ling, H. A. N., Guo-Zhen, F., Ji-Feng, L. I. U., & Shuo, W. (2017). Study of efficiency of coupling peptides with gold nanoparticles. *Chinese Journal of Analytical Chemistry, 45*(5), 662–667.

Zhang, R., Siu, M. K. Y., Ngan, H. Y. S., & Chan, K. K. L. (2022). Molecular Biomarkers for the Early Detection of Ovarian Cancer. *International Journal of Molecular Sciences, 23*(19), 1-16.

Zhang, W., Yang, F., Ou, D., Lin, G., Huang, A., Liu, N., & Li, P. (2019). Prediction, docking study and molecular simulation of 3D DNA aptamers to their targets of endocrine disrupting chemicals. *Journal of Biomolecular Structure & Dynamics, 37*(16), 4274–4282.

Zhang, X., Perez-Sanchez, H., & Lightstone, F. C. (2017). A Comprehensive Docking and MM/GBSA Rescoring Study of Ligand Recognition upon Binding Antithrombin. *Current Topics in Medicinal Chemistry, 17*(14), 1631-1639.

Zhang, X., Roeffaers, M. B. J., Basu, S., Daniele, J. R., Fu, D., Freudiger, C. W., Holtom, G. R., & Xie, X. S. (2012). Label-free live-cell imaging of nucleic acids using stimulated Raman scattering microscopy. Chemphyschem : A European *Journal of Chemical Physics and Physical Chemistry, 13*(4), 1054–1059.

Zhang, X., Wang, Y., Deng, H., Xiong, X., Zhang, H., Liang, T., & Li, C. (2021). An aptamer biosensor for CA125 quantification in human serum based on upconversion luminescence resonance energy transfer. *Microchemical Journal, 161,* 105761.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics, 9*(1), 40.

Zhang, Y., Lai, B. S., & Juhas, M. (2019). Recent Advances in Aptamer Discovery and Applications. *Molecules, 24*(5), 941, 1-22.

Zhang, Y., Lu, T., Wang, Y., Diao, C., Zhou, Y., Zhao, L., & Chen, H. (2018). Selection of a DNA Aptamer against Zearalenone and Docking Analysis for Highly Sensitive Rapid Visual Detection with Label-Free Aptasensor. *Journal of Agricultural and Food Chemistry, 66*(45), 12102–12110.

Zhang, Y., Ouyang, H., Zhang, S., Ni, Y., Zhu, Z., Ling, C., He, A., & Liu, R. (2023). Label-free electrochemical bioplatform based on Au-modified magnetic Fe3O4/α-Fe2O3 hetero-nanorods for sensitive quantification of ovarian cancer tumor marker. *Microchemical Journal*, *189*, 108546.

Zhang, Y., Wang, J., & Xiao, Y. (2020). 3dRNA: Building RNA 3D structure with improved template library. *Computational and Structural Biotechnology Journal*, *18*, 2416–2423.

Zhang, Z., & Liu, J. (2018). An engineered one-site aptamer with higher sensitivity for label-free detection of adenosine on graphene oxide. *Canadian Journal of Chemistry*, *96*(11), 957–963.

Zhao, L., Wang, Q., Yin, Y., Yang, Y., Cui, H., & Dong, Y. (2022). Evolution of Interferon-Gamma Aptamer with Good Affinity and Analytical Utility by a Rational In Silico Base Mutagenesis Post-SELEX Strategy. *Molecules 27*(17).

Zhao, X., Ando, Y., Liu, Y., Jinno, M., & Suzuki, T. (2003). Carbon Nanowire Made of a Long Linear Carbon Chain Inserted Inside a Multiwalled Carbon Nanotube. *Physical Review Letters, 90*(18), 187401.

Zheng, L., Cui, C., Shi, O., Lu, X., Li, Y., Wang, W., Li, Y., & Wang, Q. (2020). Incidence and mortality of ovarian cancer at the global, regional, and national levels, 1990–2017. *Gynecologic Oncology, 159*(1), 239–247.

Zhong, Y., Wang, Y., Huang, J., Xu, X., Pan, W., Gao, S., Zhang, Y., & Su, M. (2019). Association of hCG and LHCGR expression patterns with clinicopathological parameters in ovarian cancer. *Pathology-Research and Practice, 215*(4), 748–754.

Zhou, A. Q., O'Hern, C. S., & Regan, L. (2011). Revisiting the Ramachandran plot from a new angle. *Protein Science, 20*(7), 1166–1171.

Zhou, J., & Rossi, J. (2017). Aptamers as targeted therapeutics: current potential and challenges. *Nature Reviews. Drug Discovery, 16*(3), 181–202.

Zhou, L., Lv, Z., Shao, J., Xu, Y., Luo, X., Zhang, Y., Hu, Y., Zhang, W., Luo, S., & Fang, J. (2016). Detection of human epididymis protein 4 (HE4) in human serum samples using a specific monoclonal antibody-based sandwich enzyme-linked immunosorbent assay (ELISA). *Journal of Clinical Laboratory Analysis, 30*(5), 581–589.

Zhu, C., Feng, Z., Qin, H., Chen, L., Yan, M., Li, L., & Qu, F. (2024). Recent progress of SELEX methods for screening nucleic acid aptamers. *Talanta, 266*, 124998.

Zhu, J. W., Charkhchi, P., & Akbari, M. (2022). Potential clinical utility of liquid biopsies in ovarian cancer. *Molecular Cancer, 21*(114), 1-24.

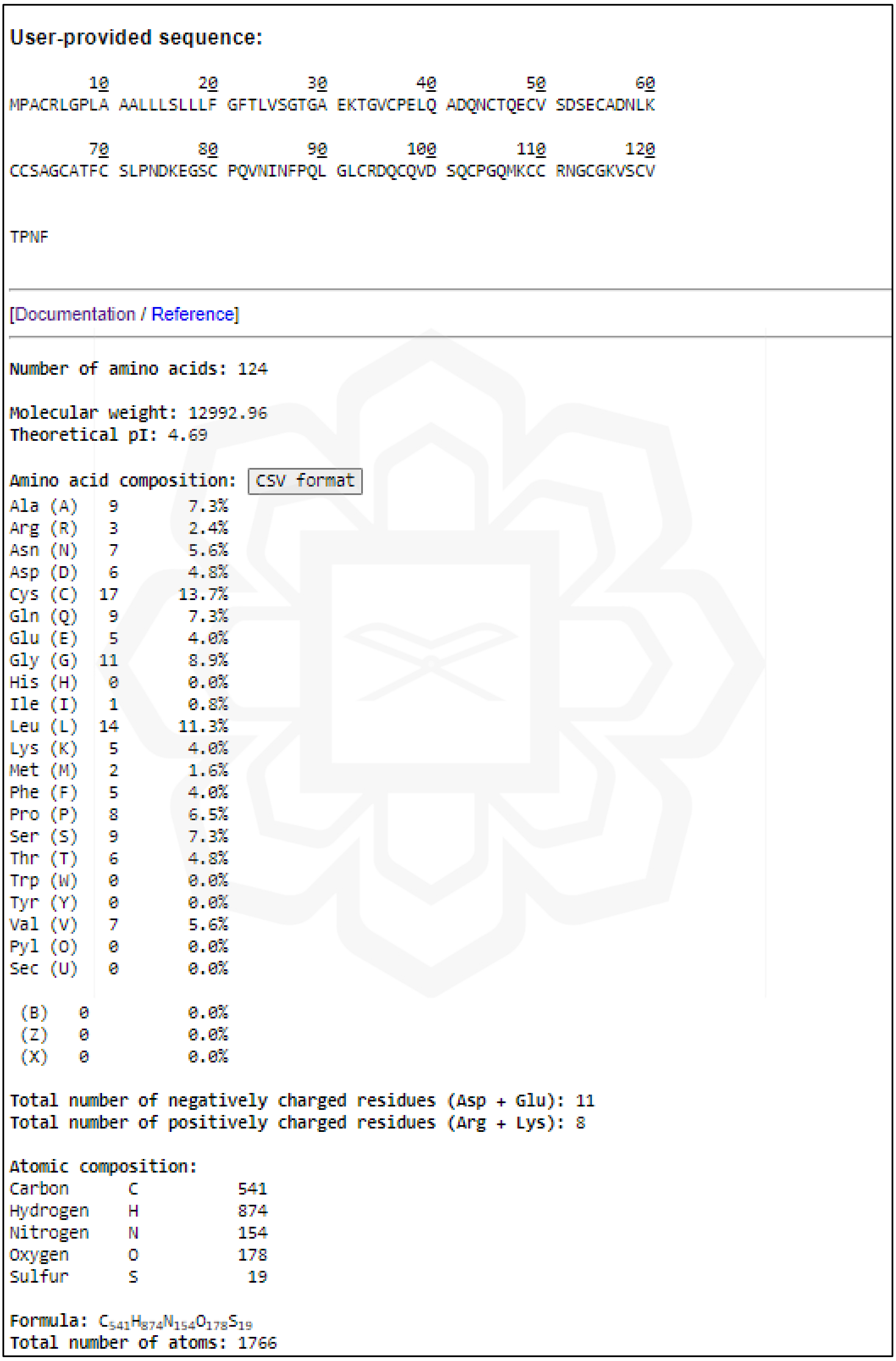
Zikri, A. T., Pranowo, H. D., & Haryadi, W. (2021). Stability, Hydrogen Bond Occupancy Analysis and Binding Free Energy Calculation from Flavonol Docked in DAPK1 Active Site Using Molecular Dynamic Simulation Approaches. *Indonesian Journal of Chemistry, 21*(2), 383–390.

Žilovič, D., Čiurlienė, R., Sabaliauskaitė, R., & Jarmalaitė, S. (2021). Future screening prospects for ovarian cancer. *Cancers, 13*(15), 3840.

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

APPENDIX A

The raw data of ProtParam analysis



```
User-provided sequence:

        10         20         30         40         50         60
MPACRLGPLA AALLLSLLLF GFTLVSGTGA EKTGVCPELQ ADQNCTQECV SDSECADNLK

        70         80         90        100        110        120
CCSAGCATFC SLPNDKEGSC PQVNINFPQL GLCRDQCQVD SQCPGQMKCC RNGCGKVSCV


TPNF

[Documentation / Reference]

Number of amino acids: 124

Molecular weight: 12992.96
Theoretical pI: 4.69

Amino acid composition: CSV format
Ala (A)   9         7.3%
Arg (R)   3         2.4%
Asn (N)   7         5.6%
Asp (D)   6         4.8%
Cys (C)  17        13.7%
Gln (Q)   9         7.3%
Glu (E)   5         4.0%
Gly (G)  11         8.9%
His (H)   0         0.0%
Ile (I)   1         0.8%
Leu (L)  14        11.3%
Lys (K)   5         4.0%
Met (M)   2         1.6%
Phe (F)   5         4.0%
Pro (P)   8         6.5%
Ser (S)   9         7.3%
Thr (T)   6         4.8%
Trp (W)   0         0.0%
Tyr (Y)   0         0.0%
Val (V)   7         5.6%
Pyl (O)   0         0.0%
Sec (U)   0         0.0%

 (B)   0            0.0%
 (Z)   0            0.0%
 (X)   0            0.0%

Total number of negatively charged residues (Asp + Glu): 11
Total number of positively charged residues (Arg + Lys): 8

Atomic composition:
Carbon      C          541
Hydrogen    H          874
Nitrogen    N          154
Oxygen      O          178
Sulfur      S           19
```

Formula: $C_{541}H_{874}N_{154}O_{178}S_{19}$

```
Total number of atoms: 1766
```

APPENDIX B

The sequences of 256 Mutated Hairpins

1	GCATCGGC	35	GCCGGAGC	69	GCGTTCGC	103	GCTGAAGC
2	GCATCCGC	36	GCCGGTGC	70	GCGTTGGC	104	GCTGTTGC
3	GCATCAGC	37	GCGGGCGC	71	GCGTTAGC	105	GCCACCGC
4	GCATCTGC	38	GCGGGGGC	72	GCGTTTGC	106	GCCAGGGC
5	GCATGGGC	39	GCGGGAGC	73	GCTTTCGC	107	GCCATAGC
6	GCATGCGC	40	GCGGGTGC	74	GCTTTGGC	108	GCGACCGC
7	GCATGAGC	41	GCAGGCGC	75	GCTTTAGC	109	GCGAGGGC
8	GCATGTGC	42	GCAGGGGC	76	GCTTTTGC	110	GCGATAGC
9	GCATAGGC	43	GCAGGAGC	77	GCCCGCGC	111	GCAACCGC
10	GCATACGC	44	GCAGGTGC	78	GCCCGGGC	112	GCAAGGGC
11	GCATAAGC	45	GCTGGCGC	79	GCCCGAGC	113	GCAATAGC
12	GCATATGC	46	GCTGGGGC	80	GCCCGTGC	114	GCTACCGC
13	GCATTGGC	47	GCTGGAGC	81	GCGCGCGC	115	GCTAGGGC
14	GCATTCGC	48	GCTGGTGC	82	GCGCGGGC	116	GCTATAGC
15	GCATTAGC	49	GCCAACGC	83	GCGCGAGC	117	GCCTCCGC
16	GCATTTGC	50	GCCAAGGC	84	GCGCGTGC	118	GCCTGGGC
17	GCCCCCGC	51	GCCAAAGC	85	GCACGCGC	119	GCCTAAGC
18	GCCCCGGC	52	GCCAATGC	86	GCACGGGC	120	GCGTCCGC
19	GCCCCAGC	53	GCGAACGC	87	GCACGAGC	121	GCGTGGGC
20	GCCCCTGC	54	GCGAAGGC	88	GCACGTGC	122	GCGTAAGC
21	GCGCCCGC	55	GCGAAAGC	89	GCTCGCGC	123	GCTTCCGC
22	GCGCCGGC	56	GCGAATGC	90	GCTCGGGC	124	GCTTGGGC
23	GCGCCAGC	57	GCAAACGC	91	GCTCGAGC	125	GCTTAAGC
24	GCGCCTGC	58	GCAAAGGC	92	GCTCGTGC	126	GCCCACGC
25	GCACCCGC	59	GCAAAAGC	93	GCCGCCGC	127	GCCCAGGC
26	GCACCGGC	60	GCAAATGC	94	GCCGAAGC	128	GCCCAAGC
27	GCACCAGC	61	GCTAACGC	95	GCCGTTGC	129	GCCCATGC
28	GCACCTGC	62	GCTAAGGC	96	GCGGCCGC	130	GCGCACGC
29	GCTCCCGC	63	GCTAAAGC	97	GCGGAAGC	131	GCGCAGGC
30	GCTCCGGC	64	GCTAATGC	98	GCGGTTGC	132	GCGCAAGC
31	GCTCCAGC	65	GCCTTCGC	99	GCAGCCGC	133	GCGCATGC
32	GCTCCTGC	66	GCCTTGGC	100	GCAGAAGC	134	GCTCACGC
33	GCCGGCGC	67	GCCTTAGC	101	GCAGTTGC	135	GCTCAGGC
34	GCCGGGGC	68	GCCTTTGC	102	GCTGCCGC	136	GCTCAAGC

137	GCTCATGC	171	GCGGTCGC	205	GCAGCAGC	239	GCTTGTGC
138	GCCGACGC	172	GCGGTGGC	206	GCAGCTGC	240	GCACAGGC
139	GCCGAGGC	173	GCGGTAGC	207	GCAACGGC	241	GCACACGC
140	GCCGATGC	174	GCTGTCGC	208	GCAACAGC	242	GCACAAGC
141	GCGGACGC	175	GCTGTGGC	209	GCAACTGC	243	GCACATGC
142	GCGGAGGC	176	GCTGTAGC	210	GCTGCGGC	244	GCAGAGGC
143	GCGGATGC	177	GCCATCGC	211	GCTGCAGC	245	GCAGACGC
144	GCTGACGC	178	GCCATGGC	212	GCTGCTGC	246	GCAGATGC
145	GCTGAGGC	179	GCCATTGC	213	GCTACGGC	247	GCACTGGC
146	GCTGATGC	180	GCGATCGC	214	GCTACAGC	248	GCACTCGC
147	GCCTACGC	181	GCGATGGC	215	GCTACTGC	249	GCACTAGC
148	GCCTAGGC	182	GCGATTGC	216	GCTTCGGC	250	GCACTTGC
149	GCCTATGC	183	GCTATCGC	217	GCTTCAGC	251	GCAGTGGC
150	GCGTACGC	184	GCTATGGC	218	GCTTCTGC	252	GCAGTCGC
151	GCGTAGGC	185	GCTATTGC	219	GCCAGCGC	253	GCAGTAGC
152	GCGTATGC	186	GCCGCGGC	220	GCCAGAGC	254	GCAATGGC
153	GCTTACGC	187	GCCGCAGC	221	GCCAGTGC	255	GCAATCGC
154	GCTTAGGC	188	GCCGCTGC	222	GCCTGCGC	256	GCAATTGC
155	GCTTATGC	189	GCCACGGC	223	GCCTGAGC		
156	GCCCTCGC	190	GCCACAGC	224	GCCTGTGC		
157	GCCCTGGC	191	GCCACTGC	225	GCGAGCGC		
158	GCCCTAGC	192	GCCTCGGC	226	GCGAGAGC		
159	GCCCTTGC	193	GCCTCAGC	227	GCGAGTGC		
160	GCGCTCGC	194	GCCTCTGC	228	GCGTGCGC		
161	GCGCTGGC	195	GCGGCGGC	229	GCGTGAGC		
162	GCGCTAGC	196	GCGGCAGC	230	GCGTGTGC		
163	GCGCTTGC	197	GCGGCTGC	231	GCAAGCGC		
164	GCTCTCGC	198	GCGACGGC	232	GCAAGAGC		
165	GCTCTGGC	199	GCGACAGC	233	GCAAGTGC		
166	GCTCTAGC	200	GCGACTGC	234	GCTAGCGC		
167	GCTCTTGC	201	GCGTCGGC	235	GCTAGAGC		
168	GCCGTCGC	202	GCGTCAGC	236	GCTAGTGC		
169	GCCGTGGC	203	GCGTCTGC	237	GCTTGCGC		
170	GCCGTAGC	204	GCAGCGGC	238	GCTTGAGC		