

***IN SILICO* SELECTION AND VALIDATION OF  
DNA APTAMER AGAINST PROGESTERONE  
RECEPTOR DNA BINDING DOMAIN**

**NAVIEN A/L THOLASI NADHAN**

**UNIVERSITI SAINS MALAYSIA**

**2022**

***IN SILICO* SELECTION AND VALIDATION OF  
DNA APTAMER AGAINST PROGESTERONE  
RECEPTOR DNA BINDING DOMAIN**

by

**NAVIEN A/L THOLASI NADHAN**

**Thesis submitted in fulfilment of the requirements  
for the degree of  
Master of Science**

**August 2022**

## **ACKNOWLEDGEMENT**

To begin with, thanks to God for being my strength from the very beginning till now. Besides that, I specially thank my supervisors, Prof. Dr. Tang Thean Hock and Dr. Citartan Marimuthu for all the guidance and assistance throughout the research period. I also would like to take this opportunity to express my deepest appreciation to my parents, Mr. Tholasi Nadhan and Mrs. Malini and family members Ms. Netangily and Ms. Rinita for the motivation, sacrifice and love that enable me to complete this postgraduate study successfully. In addition to that, I want to extend my heartfelt gratitude to my fellow friends Dr. Prabu, Dr. Presela, Mr. Tzi Shien Yeoh, Mrs. Anna Andrew, Dr. Alvin Paul, Mrs. Yik Wei Chew, Mrs. Sharmilla, Ms. Wen Siew, Mr. Hoong Kuan, Ms. Teh Hui Wen and Mr. Cheah Hong Leong for all the unwavering support and exceptional care. It was a wonderful journey with all of my colleagues in the laboratory. Thanks to Dr. Ng Siew Kit for sharing his knowledge of expertise and thanks to the RNA-Bio Group's members as well as the IPPT administration team for their assistance in multiple circumstances. My sincere gratitude to USM for the financial support provided through the USM Fellowship Award throughout my study. Last but not least, my deepest sense of thanks and gratitude to all my well-wishers towards the completion of Master of Science.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENT.....</b>	<b>ii</b>
<b>TABLE OF CONTENTS .....</b>	<b>iii</b>
<b>LIST OF TABLES .....</b>	<b>vi</b>
<b>LIST OF FIGURES .....</b>	<b>vii</b>
<b>LIST OF SYMBOLS AND ABBREVIATIONS.....</b>	<b>ix</b>
<b>ABSTRAK....</b>	<b>xii</b>
<b>ABSTRACT.....</b>	<b>xiv</b>
<b>CHAPTER 1 INTRODUCTION.....</b>	<b>1</b>
1.1 Human Progesterone Receptor (hPR) .....	1
1.2 PR as a Breast Cancer Biomarker .....	3
1.3 Methods for the Detection of PR.....	4
1.4 Aptamer .....	5
1.5 Advantages of Aptamers Versus Antibodies .....	6
1.6 SELEX .....	6
1.7 Disadvantages of SELEX.....	8
1.8 <i>In Silico</i> Selection of Aptamers.....	8
1.9 Problem Statement .....	9
1.10 Objectives.....	9
<b>CHAPTER 2 <i>IN SILICO</i> SELECTION OF DNA APTAMER AGAINST PR DBD PROTEIN .....</b>	<b>12</b>
2.1 Introduction.....	12
2.2 Materials and Methods.....	13
2.2.1 Designing of PRE-derived ssDNA Library .....	13
2.2.2 DNA Secondary Structure Prediction.....	13
2.2.3 DNA Tertiary Structure Prediction.....	14

2.2.4	Molecular Docking .....	15
2.3	Results and Discussion.....	16
2.3.1	Designing of PRE-derived ssDNA Library .....	16
2.3.2	DNA Secondary Structure Prediction.....	20
2.3.3	DNA Tertiary Structure Prediction.....	26
2.3.4	Molecular Docking Analysis of the ssDNA Candidates and PR DBD.....	31
2.3.5	Determination of the Interacting Region of Both the Aptamer and PR DBD .....	38
2.4	Conclusion.....	42
<b>CHAPTER 3 EXPRESSION AND PURIFICATION OF RECOMBINANT PROGESTERONE RECEPTOR DNA BINDING DOMAIN (PR DBD) FOR THE EXPERIMENTAL EVALUATION OF THE DNA APTAMER.....</b>		<b>43</b>
3.1	Introduction .....	43
3.2	Materials and Methods.....	43
3.2.1	Recombinant PR DBD Protein Expression via IPTG Induction.....	43
3.2.2	SDS-PAGE and Coomassie Blue Staining .....	44
3.2.3	Recombinant PR DBD Protein Purification.....	45
3.2.4	Protein Dialysis .....	45
3.2.5	Bradford Protein Quantification Assay.....	46
3.2.6	Western Blot Analysis .....	46
3.3	Results and Discussion.....	47
3.3.1	Recombinant PR DBD Protein Expression.....	47
3.3.2	Recombinant PR DBD Protein Purification.....	51
3.3.3	Confirmation of PR DBD Protein Identity by Western Blot Assay .	53
3.4	Conclusion.....	55
<b>CHAPTER 4 <i>IN VITRO</i> BINDING ANALYSIS OF DNA APTAMER .....</b>		<b>56</b>
4.1	Introduction .....	56

4.2	Materials and Methods.....	56
4.2.1	Polymerase Chain Reaction (PCR) for the Preparation of the DNA Template of the PRapt-3 RNA aptamer.....	56
4.2.2	Agarose Gel Electrophoresis.....	57
4.2.3	Ethanol Precipitation.....	58
4.2.4	<i>In Vitro</i> Transcription.....	58
4.2.5	Denaturing-Urea Polyacrylamide Gel-Electrophoresis .....	59
4.2.6	Rapid Crush and Soak-based RNA Purification.....	59
4.2.7	Functionalization of the Poly A-tailed PRapt-3 RNA Aptamer with Biotin at the 3'-end .....	60
4.2.8	Preparation of Biotinylated DNA Aptamer Candidates.....	60
4.2.9	Direct Enzyme-Linked Apta-Sorbent Assay (ELASA) .....	61
4.2.10	Statistical Analysis.....	62
4.2.11	Estimation of Dissociation Constant, $K_d$ .....	62
4.3	Results and Discussion.....	62
4.3.1	Functionalization of the Aptamer Candidates with Biotin prior to ELASA.....	62
4.3.2	Validation of the Aptamer Candidates using Direct ELASA Reveals PRDBDapt17 as the Most Potent DNA aptamer .....	66
4.3.3	Determination of Equilibrium Dissociation Constant of PRDBDapt17 DNA Aptamer Against Recombinant PR DBD.....	73
4.4	Conclusion.....	75
	<b>CHAPTER 5 CONCLUSION AND FUTURE PERSPECTIVES.....</b>	<b>76</b>
5.1	Conclusion.....	76
5.2	Future Perspectives .....	77
	<b>REFERENCES.....</b>	<b>78</b>
	<b>APPENDICES</b>	
	<b>LIST OF PUBLICATIONS</b>	

## LIST OF TABLES

	<b>Page</b>
Table 1.1	Detailed overview of PR isoforms ..... 2
Table 1.2	Major steps involved in SELEX ..... 7
Table 2.1	Sequence details of the PRE-derived ssDNA sequences ..... 17
Table 2.2	The $\Delta G$ value of the secondary structures of DNA aptamer candidates ..... 23
Table 2.3	The predicted binding scores of DNA aptamer candidates with PR DBD protein (highest to the lowest) ..... 35

## LIST OF FIGURES

	<b>Page</b>
Figure 1.1	Study flowchart..... 11
Figure 2.1	Secondary structures of DNA sequences predicted by mfold server. 21
Figure 2.2	Secondary structure prediction of the PRapt-3 RNA aptamer using mfold..... 22
Figure 2.3	Schematic workflow of the <i>in silico</i> selection of DNA aptamer against PR DBD protein..... 28
Figure 2.4	Tertiary structures of DNA sequences predicted with the aid of RNA Composer and VMD ..... 29
Figure 2.5	Tertiary structure of PRapt-3 RNA aptamer as predicted by RNA Composer ..... 30
Figure 2.6	Three-dimensional crystal structure of PR DBD protein..... 32
Figure 2.7	Aptamer-PR DBD complex predicted by PatchDock..... 33
Figure 2.8	PRapt-3-PR DBD complex predicted by PatchDock..... 34
Figure 2.9	Hydrogen bonding between PRDBDapt17 and PR DBD..... 39
Figure 2.10	Hydrogen bonding between PRDBDapt5 and PR DBD..... 40
Figure 2.11	Secondary structure of aptamer candidates showing the contact nucleotide regions with PR DBD (red): (a) PRDBDapt17 and (b) PRDBDapt5. .... 41
Figure 3.1	Schematic diagram of the recombinant PR DBD protein expression and purification workflow..... 49
Figure 3.2	SDS-PAGE analysis on the time-course expression profile of PR DBD protein with the Rosetta 2(DE3)pLysS..... 50

Figure 3.3	SDS-PAGE analysis of the dialyzed PR DBD protein eluates purified by native IMAC purification.....	52
Figure 3.4	Western blot analysis of the purified PR DBD protein using nickel-HRP .....	54
Figure 4.1	Poly A-tail extension of PRapt-3 RNA aptamer at the 3'-end.....	64
Figure 4.2	Secondary structure prediction of the poly A-tailed PRapt-3 RNA aptamer via mfold .....	65
Figure 4.3	Schematic diagram of the ELASA workflow .....	68
Figure 4.4	Direct ELASA with PRapt-3 RNA aptamer .....	69
Figure 4.5	Direct ELASA with PRDBDapt17 DNA aptamer candidate.....	70
Figure 4.6	Direct ELASA with PRDBDapt18 DNA aptamer candidate.....	71
Figure 4.7	Direct ELASA with PRDBDapt5 DNA aptamer candidate.....	72
Figure 4.8	Dissociation constant determination of the PRDBDapt17 DNA aptamer.....	74

## LIST OF SYMBOLS AND ABBREVIATIONS

A	Adenine
aa	Amino acid
APS	Ammonium persulfate
AMD	Age-related macular degeneration
ASCO	American Society of Clinical Oncology
Bis	N, N'-methylene bisacrylamide
bp	Base pair (s)
BSA	Bovine serum albumin
C	Cytosine
CAP	College of American Pathologists
ddH <sub>2</sub> O	Double-distilled water
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic Acid
ELASA	Enzyme-Linked Apta-Sorbent Assay
et al.	And others
ER	Estrogen Receptor
FDA	USA Food and Drug Administration
g	Gram
G	Guanine
HCl	Hydrochloric acid
HRP	Horseradish peroxidase
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
KCl	Potassium chloride
K <sub>d</sub>	Dissociation constant
kDa	Kilodalton
kHz	Kilohertz

LB	Luria Bertani medium
LiCl	Lithium Chloride
LINA	Lithium Chloride and Sodium Chloride
LOD	Limit of Detection
M	Molar, [(Mole)/(Litre)]
Min	Minute (s)
mL	Milliliter
mM	Millimolar
MRE	Molecular Recognition Element
MWCO	Molecular Weight Cut-Off
NaCl	Sodium chloride
NaOAc.3H <sub>2</sub> O	Sodium acetate trihydrate
NaOH	Sodium hydroxide
nM	Nanomolar
-OH	Hydroxyl
PAGE	Polyacrylamide gel electrophoresis
PBST	Phosphate buffered saline with Tween 20
PCR	Polymerase chain reaction
PDB	Protein Data Bank
pmol	Picomole
PR	Progesterone Receptor
PR DBD	Progesterone Receptor DNA binding domain
PRapt-3	Progesterone Receptor aptamer – 3
RCSB	The Research Collaboratory for Structural Bioinformatics
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Rotations per minute
RT	Room temperature
RT-PCR	Reverse transcription-PCR
s	Second (s)
SELEX	Systematic Evolution of Ligands via Exponential Enrichment
ssDNA	Single-stranded DNA

T	Thymine
TAE	Tris-Acetic Acid-EDTA
TBE	Tris-Boric Acid-EDTA
TEMED	N,N,N',N'-Tetramethylethylenediamine
TMB	3,3',5,5'-tetramethylethylenediamine
Tris	Tris-(Hydroxymethyl)-Aminomethane
tRNA	Transfer RNA
UV	Ultraviolet
V	Volt (s)
VEGF	Vascular Endothelial Growth Factor
v/v	Volume per volume
W	Watt
WHO	World Health Organization
w/v	Weight per volume
x g	Relative Centrifugal Force
µg	Microgram
µL	Microliter
µM	Micromolar
°C	Degrees Celcius
%	Percentage

# PEMILIHAN *IN SILICO* DAN VALIDASI APTAMER DNA TERHADAP DOMAIN PENGIKAT DNA RESEPTOR PROGESTERON

## ABSTRAK

Reseptor progesteron memainkan peranan penting dalam perkembangan kanser payudara. Pada masa ini, 'Immunohistochemistry' berasaskan antibodi digunakan dalam penilaian patologi tahap PR untuk pengesanan kanser payudara. Kelemahan penggunaan antibodi membuka laluan untuk menggunakan aptamers sebagai alternatif. Aptamer ialah oligonukleotida DNA atau RNA untai tunggal yang dijana oleh SELEX yang mampu mengikat molekul sasaran serumpunya dengan pertalian dan kekhususan yang tinggi berdasarkan kapasiti lipatan struktur uniknya. Kelemahan langkah-langkah tertentu SELEX konvensional meningkatkan usaha untuk memilih aptamer DNA menggunakan dok *in silico*. Oleh itu, kami melaporkan pemilihan *in silico* dan validasi aptamer DNA terhadap domain pengikat DNA reseptor progesteron (PR DBD) menggunakan urutan ssDNA yang diperoleh daripada elemen reseptor progesteron manusia (PREs). Pertama, perpustakaan enam puluh empat analog ssDNA yang hampir sepadan dengan PRE telah direka bentuk dan tertakluk kepada penentuan struktur tertiar. Selepas itu, dok antara struktur tertiar ssDNA dengan PR DBD telah dijalankan menggunakan PatchDock. Urutan ssDNA dengan skor pengikatan tertinggi telah dipilih sebagai calon aptamer dan selanjutnya disahkan oleh ELISA langsung *in vitro*. Di antara calon, kami memilih jujukan ssDNA (PRDBDapt17; 5'- AGAACAGCGTGTCT -3'), yang menunjukkan skor dok tertinggi 11334 sebagai aptamer pengikat PR DBD yang berpotensi. Di samping itu, PRDBDapt17 mengesan PR DBD rekombinan dalam ELISA langsung dengan had pengesanan 3.91 nM. Pemalar pemisahan dianggarkan

pada 366.6 nM. Justeru, PRDBDapt17 merupakan aptamer yang berpotensi digunakan dalam diagnosis kanser payudara.

# ***IN SILICO* SELECTION AND VALIDATION OF DNA APTAMER AGAINST PROGESTERONE RECEPTOR DNA BINDING DOMAIN**

## **ABSTRACT**

Progesterone receptor plays an important role in the progression of breast cancer. Currently, antibody-based Immunohistochemistry is used in pathological assessment of PR levels for the detection of breast cancer. The shortcomings associated with antibodies pave the path to use aptamers as the alternatives. Aptamers are single-stranded DNA or RNA oligonucleotides generated by SELEX that are capable of binding to their cognate target molecules with high affinity and specificity based on their unique structural folding capacity. The tediousness and rigor associated with certain steps of the conventional SELEX intensify the efforts to select DNA aptamers using *in silico*-docking approach. That said, we report an *in silico* selection and validation of DNA aptamer to the progesterone receptor DNA binding domain (PR DBD) using ssDNA sequences derived from human progesterone response elements (PREs). Firstly, a library of sixty-four different near-native ssDNA analogs of the corresponding PRE sequences was designed and subjected to secondary and tertiary structural determination. After that, docking between the ssDNA tertiary structures with the PR DBD was carried out using PatchDock. The sequence with the highest docking score was chosen as the aptamer candidate and further validated by *in vitro* direct ELISA. Among the candidates, we selected the ssDNA sequence (PRDBDapt17; 5'- AGAACAGCGTGTCT -3'), which showed the highest docking scores of 11334 as a promising PR DBD binding aptamer. In addition, the PRDBDapt17 detected recombinant PR DBD in direct ELISA with a limit of detection of 3.91 nM. The dissociation constant was

estimated at 366.6 nM. Therefore, PRDBDapt17 is a potential aptamer that can be used in the diagnosis of breast cancer.

## CHAPTER 1

### INTRODUCTION

#### 1.1 Human Progesterone Receptor (hPR)

The progesterone receptor (PR) belongs to the type I nuclear receptors, which are also known as steroid hormone receptors. It is primarily expressed in the female genitalia tract, breasts, and brain and has a role in the development, differentiation, and maintenance of female reproductive tissues (Li et al., 2003). PR is made up of three key domains: an amino-terminal domain (NTD), a highly conserved and a centrally positioned DNA binding domain (DBD) and a C-terminal ligand-binding domain (LBD). Human PR is divided into three distinct isoforms: PR-A, PR-B and PR-C. PR-A and PR-B share similar DBD (552 - 632 aa) and LBD (687 - 933 aa) that are separated by a hinge while PR-C only has a LBD at the C-terminal. In addition, PR-B consists of the entire length of PR whereas PR-A does not contain the first 164 amino acids at the N-terminus and thus is shorter than PR-B. The phenomenon of different isoforms of PR is due to the expression from different transcriptional and translational start sites (Hill et al., 2012; Li et al., 2003). Further description concerning PR isoforms is summarized in Table 1.1.

**Table 1.1** Detailed overview of PR isoforms

<b>PR isoforms</b>	<b>Size</b>	<b>Functions</b>	<b>References</b>
PR-A	94 kDa	Uterine development and reproductive function	(Jacobsen et al., 2002; Richer et al., 2002)
PR-B	116 kDa	Normal mammary gland development	(Jacobsen et al., 2002; Mulac-Jericevic et al., 2003; Richer et al., 2002)
PR-C	60 kDa	Transcriptional activity, dominant inhibitor of uterine PR-B in the fundal myometrium during labor	(Condon et al., 2006; Jacobsen et al., 2002; Richer et al., 2002)

PR is mainly localized in the nucleus, which substantiates its involvement in transcriptional regulation (Arnett-Mansfield et al., 2004; Arnett-Mansfield et al., 2007; Scarpin et al., 2009). PR modulates transcriptional activity associated with mammary gland proliferation in women by binding directly to DNA through progesterone response elements (PREs), in a regulation, which is also known as the classical genomic signaling event. In the classical genomic signaling event, the regulation begins with binding of the ligand progesterone to the progesterone receptor ligand-binding domain (PR LBD). As a result, PR undergoes conformational changes and dimerization. Following that, the PR binds directly to the progesterone response element (PRE) on the target genes promoters via the DBD (Scarpin et al., 2009; Tata, 2002). PRE is made up of a 5'-AGAACA-3' inverted repeat hexanucleotide sequence with a 3 nucleotide (3N) space in the middle, which is distinctively specific for PR (Hill et al., 2012; Umesono et al., 1989). The natural binding affinity of the DBD of PR accounts for its attachment to the PRE (Hill et al., 2012; Roemer et al., 2006). Upon binding to PR, the transcription of the corresponding genes will be activated (Scarpin et al., 2009; Tata, 2002). The genes that are activated for transcription are *STAT5A* (transcription), *Bcl-X<sub>L</sub>* (cell growth and apoptosis), *NPC1* (lipid metabolism) and *ADARB1* (nucleic acid and protein processing) (Bertucci et al., 2013; Cerliani et al., 2011; Graham et al., 2005; Obr et al., 2013; Xie et al., 2006).

## **1.2 PR as a Breast Cancer Biomarker**

Despite the fact that PR is necessary for the growth and development of the female reproductive system, however, its overexpression of PR causes malignancies such as breast cancer, ovarian cancer, cervical cancer and endometrial cancer (Daniel et al., 2011; Hong et al., 2017; Kleine et al., 1990; Lee et al., 2005). Studies have

proven that ligand-activated PR is generally associated with the occurrence of breast cancer, the foremost reason of death among women worldwide (Daniel et al., 2011; Torre et al., 2015).

PR expression is a potential prognostic biomarker in breast cancer, according to one of the largest population-based cohort studies encompassing 1074 patients (Purdie et al., 2014). Collectively, it is undeniable that PR expression is a valuable biomarker for the pathological diagnosis of breast cancer.

### **1.3 Methods for the Detection of PR**

The 'gold standard' approach for pathological assessment of PR levels in breast tumor biopsies was first a bio-chemical steroid-binding assay of cytosol extracts of tumor samples using dextran-coated charcoal (DCC). The DCC assay was used to distinguish between receptor-bound and unbound steroids. Nonetheless, considering the need to diagnose breast cancer at an early stage, the amount of tissue samples available for DCC is often insufficient. Thus, the biochemical DCC assay has been replaced by the advent of monoclonal antibody-based immunohistochemistry (IHC) procedures (Elashry-Stowers et al., 1988).

IHC turns out to be a better technique as it is relatively cheaper, more efficient, uses fewer tumor samples, able to distinguish between tumor and normal cell expression of receptors, detects tumor cell heterogeneity, and can be performed with paraffin-embedded tissues (Allred, 1993; Mohsin et al., 2004; Press et al., 2002). Nonetheless, antibody selection could jeopardize the accuracy and reproducibility of the IHC results (Nadji et al., 2005; Wludarski et al., 2011). However, antibodies as the diagnostic agent have several disadvantages such as high cost of production and batch-to-batch variation. These issues associated with

antibodies can be compensated by another class of molecular recognition element (MRE) known as aptamers.

#### **1.4 Aptamer**

The term “aptamer” comes from the combination of the Latin word “aptus”, which means “to fit” and the Greek word “meros”, which means “particle” (Ellington et al., 1990). Aptamers are single-stranded DNA or RNA sequences that, because of their unique structural folding capacity, bind to their cognate target molecules with high affinity and specificity (Savory et al., 2013). Van der Waals forces, hydrogen bonds, electrostatic interactions, aromatic ring stacking, and shape complementarity all influence how the aptamer interacts with its target (Hermann et al., 2000b; Keefe et al., 2010; Nimjee et al., 2005). Thousands of aptamers have been selected against a broad spectrum of targets such as small inorganic metal ions (Raducanu et al., 2020), drugs (Tran et al., 2020), peptides (Kutovy et al., 2020), hormones (Chergui et al., 2020), antibiotics (Yue et al., 2021), vitamins (Wadhwa et al., 2020), toxins (Song et al., 2020), proteins (Chen et al., 2020), cells (Gao et al., 2020), tissues (Liu et al., 2021), bacteria (Zhang et al., 2021) and viruses (Kumar et al., 2021). The term “aptamer” was used in over 15000 published papers in the PubMed database as of December 2021. Pegaptanib (Macugen; Pfizer/Eyetech), an RNA aptamer selective against vascular endothelial growth factor (VEGF), is the only aptamer-based medication licenced by the US Food and Drug Administration (FDA) to treat wet Age-related Macular Degeneration (AMD) (Ng et al., 2006). This was a watershed moment in aptamer technology’s application.

## **1.5 Advantages of Aptamers Versus Antibodies**

Aptamers have received more traction than antibodies due to numerous advantages such as smaller size, lower molecular weight, better structural flexibility, low degree of immunogenicity, prolonged shelf life, reusability, no batch-to-batch variation, lower production cost and easier to be synthesized (Breaker, 1997; Kruspe et al., 2008; Wang et al., 2015; Zhang et al., 2018; Zhou et al., 2017). In addition, aptamers selectively discriminate between very similarly structured biomolecules like caffeine and theophylline, which differs only by a methyl group, which is absent in the latter (Hermann et al., 2000a; Jenison et al., 1994). Aptamers have gained a lot of attention due to their exceptional qualities in applications spanning from diagnostics to therapeutics, targeted drug delivery, gene therapy (Zhao et al., 2017) and bio-sensing (Alshamaileh et al., 2017; Citartan et al., 2019; Khan et al., 2018; Shigdar et al., 2013). Due to the aforementioned factors, aptamers are being considered as an alternative to antibodies.

## **1.6 SELEX**

Aptamers are generally chosen *in vitro* from a random oligonucleotide library using an iterative technique known as Systematic Evolution of Ligands by Exponential enrichment (SELEX). This process consists of four major steps, which include preparation of initial oligonucleotide library, incubation, partitioning and amplification (Sun et al., 2015). Table 1.2 elaborates the steps in detail (Darmostuk et al., 2015; Sun et al., 2015).

**Table 1.2** Major steps involved in SELEX

<b>SELEX steps</b>	<b>Description</b>
(i) Initial random oligonucleotide combinatorial library preparation	A single-stranded DNA oligonucleotide pool comprised of $10^{14}$ - $10^{15}$ random sequences flanked by predetermined sequences is designed and chemically synthesized
(ii) Incubation	The randomized sequences in the initial library will fold into their unique structural conformation, which is then incubated with immobilized or free targets under optimal conditions to form aptamer-target complexes
(iii) Partitioning	Separation of the unbound sequences from the target-bound sequences through methods such as membrane filtration, affinity columns, magnetic beads, or capillary electrophoresis
(iv) Amplification	Target-bound sequences are amplified by PCR (DNA aptamers) or RT-PCR (RNA aptamers)

Following PCR amplification, the resulting double-stranded DNA is converted to ssDNA. These steps are repeated for at least 8 to 15 cycles before a pool of nucleic acids having a high binding affinity towards the target molecule is attained. The enriched pool of sequences is cloned and studied by Sanger sequencing or high-throughput sequencing methods to reveal the putative aptamer (Ellington et al., 1990; Tuerk et al., 1990).

### **1.7 Disadvantages of SELEX**

Although traditional SELEX techniques have undergone various improvisations to develop aptamers against hundreds of targets, so far, there is no one ideal SELEX protocol that exists for aptamer selection. As a result, the researchers must choose the one that suits their need. One of the limitations of the SELEX process is the requirement to synthesize a large amount of the ssDNA library to initiate the selection process, which is expensive and tedious (Xu et al., 2021). In addition, conventional SELEX process generally takes from weeks to months to generate aptamer candidates and the possibility of amplification bias occurrence is undeniable (Darmostuk et al., 2015). As a result, obtaining high-quality aptamers for targets continues to be a challenge.

### **1.8 *In Silico* Selection of Aptamers**

*In silico*-selection is a strategy that incorporates the usage of computer-based simulations to model nucleic acid-target interactions prior to the selection of the best aptamer candidates. The reliance on *in silico* approach could compensate for the drawbacks of *in vitro* SELEX procedures (Navien et al., 2021). *In silico* docking comprises two major steps. The first step involves simulation of binding between the target and the oligonucleotides, completion of which is usually accompanied by

scoring that reflects on the strength of the interaction. In the second step, the binding site of the interaction will be determined. The scoring function depends on the interaction between a receptor (target) and the corresponding ligand (aptamer), which is governed by a number of intermolecular forces such as hydrogen bonding, electrostatic interactions, base stacking interactions, hydrophobic and Van der Waals interactions. In most of the studies, the *in silico* docking approach is performed to determine the binding strength of the nucleic acid sequences to their cognate target after several cycles of *in vitro* SELEX (Navien et al., 2021). However, if any potential sequences that could harbor binding properties towards a target are pre-determined, *in silico* docking of these potential sequences with the target is able to eventually unveil the most potent aptamer candidate (Ahirwar et al., 2016; Kandasamy et al., 2020).

## **1.9 Problem Statement**

The drawbacks associated with the conventional SELEX bolster the efforts to generate aptamers using *in silico*-selection methods. That said, utilizing DNA sequences derived from human PREs, we report on *in silico* selection of DNA aptamers to the PR DBD (Figure 1.1).

## **1.10 Objectives**

Therefore, the objectives of this study are as follow:

**i) To select PRE-derived DNA aptamer candidate against PR DBD using *in silico*-docking approach**

- a) Design of PRE-derived DNA library
- b) Determination of the secondary structures of the generated DNA library

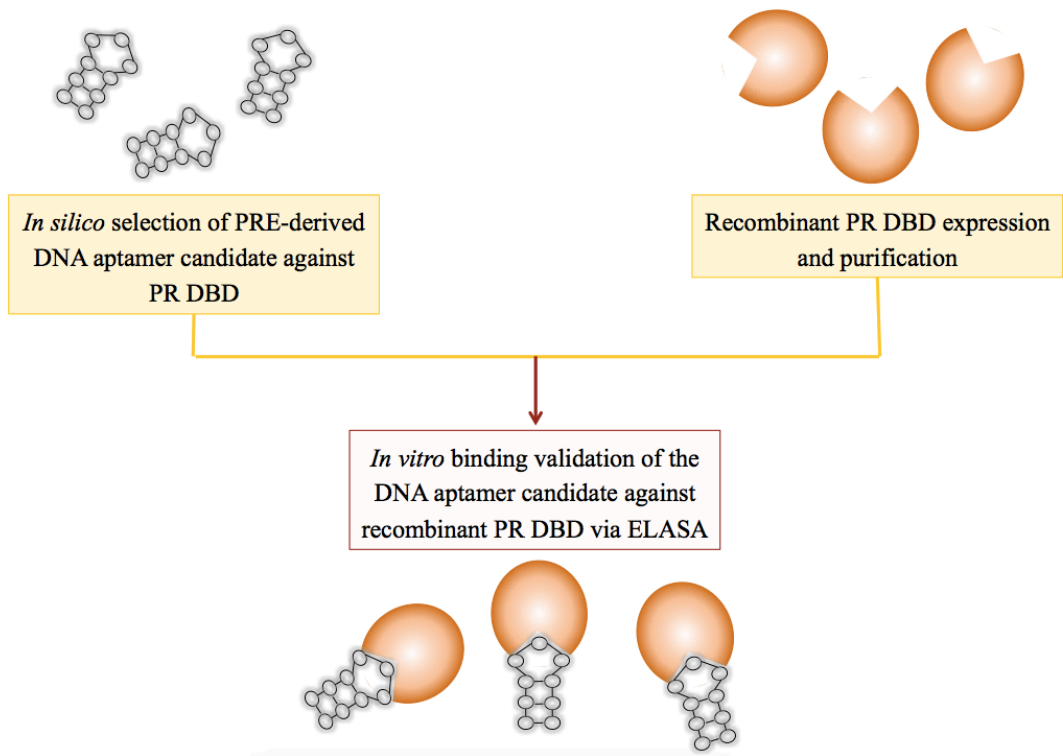
- c) Determination of the tertiary structures of the generated DNA library
- d) Molecular docking analysis of the DNA library-PR DBD protein complexes
- e) Prediction of the binding sites of the DNA aptamer candidate-PR DBD protein complex

**ii) To express and purify recombinant PR DBD**

- a) Expression and purification of recombinant PR DBD
- b) Protein identity confirmation using western blot assay probed with nickel-HRP

**iii) To validate the binding of the DNA aptamer candidate against recombinant PR DBD via Enzyme-Linked Apta-Sorbent Assay (ELASA)**

- a) Determination of the limit of detection (LOD) of the selected DNA aptamer candidate against recombinant PR DBD
- b) Determination of the equilibrium dissociation constant ( $K_d$ ) of the selected DNA aptamer candidate against recombinant PR DBD



**Figure 1.1** Study flowchart

## CHAPTER 2

### *IN SILICO* SELECTION OF DNA APTAMER AGAINST PR DBD PROTEIN

#### 2.1 Introduction

Aptamers are generally selected through conventional SELEX. The time-consuming, costly nature associated with conventional SELEX and issue of amplification bias that erroneously results in the thriving of non-binding sequences raises the need to use *in silico* molecular docking approach in aptamer selection. PatchDock server, which uses local feature matching, analyses the ligand-receptor interaction more efficiently than other docking tools. Furthermore, PatchDock server employs high-tech data structures and spatial pattern identification algorithms, such as geometric hashing and poses clustering, to reduce computational processing time and enable large-scale docking experiments (Schneidman-duhovny et al., 2005). Recently, PatchDock was applied to simulate the development of a sandwich DNA aptamer assay for the recognition of human brain natriuretic peptide (BNP). Bruno used the PatchDock docking algorithm to dock BNP to the 25cF capture aptamer. Following that, the capture aptamer-BNP complex was docked with another 2F reporter aptamer based on shape complementarity, facilitating the development of a sandwich DNA aptamer complex. Both the 25cF and 2F DNA aptamers were found to recognize different epitopes of BNP target, which comprises of 32-amino acids sequence (Bruno, 2019).

In this study, the fact that PR binds directly to PRE has inspired us to directly derive DNA sequences from PRE. Thus, a library of sixty-four different ssDNA sequences was successfully derived from PRE. These sequences were further subjected to secondary structure determination followed by tertiary structure prediction. A rigid body docking analysis was conducted with the tertiary structures

of the PRE-derived DNA sequences and the PR DBD protein using PatchDock. The potential aptamer candidate against PR DBD was chosen based on the DNA sequence with the greatest docking scores. The regions responsible for the binding interaction between the selected aptamer candidate and the PR DBD were further predicted.

## **2.2 Materials and Methods**

### **2.2.1 Designing of PRE-derived ssDNA Library**

The pseudo palindromic sequence, PRE (5'-AGAACANNNTGTTCT-3'), specifically interacts to the progesterone receptor DNA binding domain (Li et al., 2003). Based on PRE sequence, a library of sixty-four possible different ssDNA sequences were designed based on two criteria. i) The inclusion of the palindromic sequence (AGAACA) derived from PRE both at the 5' and 3'. ii) The preparation of a total of 64 different sequences as result of permutation derived from the 3-nt randomized region ( $3^4=64$ ) in the middle sandwiched by the palindromic sequences at both ends. On the other hand, PRapt-3 RNA aptamer (5'-GGAGCUCAGCCUUCACUGCUUGCUUUUGGGUGCCGUACGCAUUGCGGCAGGGGGAAGAGGAGGGUAGCGACCAGUCGAAGGCACCACGGUCGGAUCAC-3') isolated by SELEX in our previous study was used as a positive control in this test.

### **2.2.2 DNA Secondary Structure Prediction**

DNA secondary structure prediction is a prerequisite before the determination of DNA tertiary structures. The sixty-four DNA sequences were subjected to secondary structure prediction with the aid of multiple folds (mfold)

(<http://unafold.rna.albany.edu>), using default conditions (Zuker, 2003). In parallel, the secondary structure of PRapt-3 RNA aptamer was also predicted. The most thermodynamically stable structure with the lowest free energy,  $\Delta G$  value, was chosen and downloaded in dot-bracket notation (Vienna) file format, which makes it suitable for tertiary structure prediction.

### **2.2.3 DNA Tertiary Structure Prediction**

The DNA and PRapt-3 secondary structures were converted to their corresponding RNA tertiary structures via the tertiary structure RNA modeling software, RNA Composer. The RNA Composer service (<http://rnacomposer.ibch.poznan.pl/Home>) is a publicly accessible tool for predicting RNA tertiary structures (Heiat et al., 2016; Yarizadeh et al., 2019). The RNA Composer was fed secondary structures of DNA sequences represented in dot-bracket notation. Upon feeding the input, RNA Composer automatically recognized and replaced thymine with uracil residues in the input sequence. The resulting RNA's tertiary structure, which contains uracil and an extra hydroxyl group at the 2'-carbon atom of ribose, was downloaded in PDB file format. The RNA tertiary structures were then transformed into their equivalent DNA tertiary structures using the AutoPSF Visual Molecular Dynamics (VMD) plugin (<http://www.ks.uiuc.edu/Research/vmd/plugins/autopsf/>) by replacing the functional group at the 2'-position of the sugar residue (2-OH to 2-H) and the bases (uracil to thymine). Finally, the resulting tertiary structures of the DNA aptamers were subjected to energy minimization by AutoIMD VMD plugin (<http://www.ks.uiuc.edu/Research/vmd/imd/>) to evaluate the molecules' optimum local energy. The resulting files were saved in PDB format and visualized using the UCSF Chimera version 1.15.

#### 2.2.4 Molecular Docking

The crystal structure of the PR DBD was obtained from the PDB repository of the Research Collaboratory for Structural Bioinformatics (RCSB) (PDB ID: 2AA5) and was prepared for docking by eliminating unwanted protein chains and ligands. Molecular docking was performed using the PatchDock online software. PatchDock is a free programme that operates by geometry-based rigid-body docking algorithm (<http://bioinfo3d.cs.tau.ac.il/PatchDock/>) (Schneidman-duhovny et al., 2005). The PDB files of all the DNA aptamer candidates and PRapt-3 were treated as ligands while that of PR DBD were regarded as a receptor. All these files were directly submitted to PatchDock server. The Clustering Root-Mean-Square Deviation (RMSD) value was set at 4 Å. The Connolly dot surface representations of the molecules were established based on various patches such as concave, convex, and flat. After that, the complementary patches were matched to develop aptamer-target complexes. Subsequently, a scoring formula was used to evaluate the aptamer-target complex, which takes into account both geometric fit and atomic desolvation energy. Finally, the highest-scoring aptamer-target complexes were identified using RMSD clustering. At the end of the process, a link to a web page displaying the docking findings was sent via email. The geometric score, desolvation energy, interface area size, and actual rigid transformation of the top-scoring aptamer-target complexes were all accessible through the portal. This procedure was carried out independently for each of the sixty-four DNA sequences including PRapt-3. The PatchDock binding scores of each of the DNA sequences were analyzed. As a DNA aptamer candidate against PR DBD, the DNA sequence with the greatest docking score was chosen. Finally, the UCSF Chimera version 1.15 was utilized to analyze the binding

interactions between the chosen aptamer candidate and PR DBD. The interface of binding between DNA aptamer candidate and PR DBD was discovered.

## 2.3 Results and Discussion

### 2.3.1 Design of PRE-derived ssDNA Library

Dimerized progesterone receptors bind to the double-stranded PRE to stimulate the transcription of the target genes. The affinity that PRE harbours against PR inspired us to derive a library of ssDNA sequences from PRE (Table 2.1). Because of their inverted repeat nature, the resulting ssDNA can adopt stable hairpins and interact with PR, eventually allowing us to identify the most potent binders. The central 3-nt randomized region stems from the unknown 3-nt region present in PRE. Inclusion of a randomized region at the central, ssDNA sequences should form a secondary hairpin loop structure. Hairpin confirmation creates an optimum conformation to stabilize the aptamer (Sabri et al., 2019). All 64 different sequences were subjected to secondary structure prediction. The details of the generated sixty-four DNA sequences are shown in Table 2.1. Parallel to this, PRapt-3 RNA aptamer (5'-GGAGCUCAGCCUUCACUGCUUGCUUUUGGGUGCCGUACGCAUUGCGGCAGGGGGAAGAGGAGGGUAGCGACCAGUCGAAGGCACCACGGUCGGAUCAC-3') was also subjected to secondary structure formation.

**Table 2.1** Sequence details of the PRE-derived ssDNA sequences

<b>Number</b>	<b>Sequence Name</b>	<b>Sequence (5'-3')</b>
1	PRDBDapt1	AGAACAGGGTGTCT
2	PRDBDapt2	AGAACAGGATGTCT
3	PRDBDapt3	AGAACAGGTTGTCT
4	PRDBDapt4	AGAACAGGCTGTCT
5	PRDBDapt5	AGAACAAGGTGTCT
6	PRDBDapt6	AGAACAAGATGTCT
7	PRDBDapt7	AGAACAAGTTGTCT
8	PRDBDapt8	AGAACAAGCTGTCT
9	PRDBDapt9	AGAACATGGTGTCT
10	PRDBDapt10	AGAACATGATGTCT
11	PRDBDapt11	AGAACATGTTGTCT
12	PRDBDapt12	AGAACATGCTGTCT
13	PRDBDapt13	AGAACACGGTGTCT
14	PRDBDapt14	AGAACACGATGTCT
15	PRDBDapt15	AGAACACGTTGTCT
16	PRDBDapt16	AGAACACGCTGTCT
17	PRDBDapt17	AGAACAGCGTGTCT
18	PRDBDapt18	AGAACAGCATGTCT
19	PRDBDapt19	AGAACAGCTGTCT
20	PRDBDapt20	AGAACAGCCTGTCT
21	PRDBDapt21	AGAACAACCTGTCT
22	PRDBDapt22	AGAACAACATGTCT
23	PRDBDapt23	AGAACAACCTGTCT

24	PRDBDapt24	AGAACAACCTGTTCT
25	PRDBDapt25	AGAACATCGTGTTCT
26	PRDBDapt26	AGAACATCATGTTCT
27	PRDBDapt27	AGAACATCTTGTTCT
28	PRDBDapt28	AGAACATCCTGTTCT
29	PRDBDapt29	AGAACACCGTGTTCT
30	PRDBDapt30	AGAACACCATGTTCT
31	PRDBDapt31	AGAACACCTTGTTCT
32	PRDBDapt32	AGAACACCCTGTTCT
33	PRDBDapt33	AGAACAGTGTGTTCT
34	PRDBDapt34	AGAACAGTATGTTCT
35	PRDBDapt35	AGAACAGTTTGTTCT
36	PRDBDapt36	AGAACAGTCTGTTCT
37	PRDBDapt37	AGAACAATCTGTTCT
38	PRDBDapt38	AGAACAATATGTTCT
39	PRDBDapt39	AGAACAATTTGTTCT
40	PRDBDapt40	AGAACAATCTGTTCT
41	PRDBDapt41	AGAACATTGTGTTCT
42	PRDBDapt42	AGAACATTATGTTCT
43	PRDBDapt43	AGAACATTTTGTTCT
44	PRDBDapt44	AGAACATTCTGTTCT
45	PRDBDapt45	AGAACACTGTGTTCT
46	PRDBDapt46	AGAACACTATGTTCT
47	PRDBDapt47	AGAACACTTTGTTCT

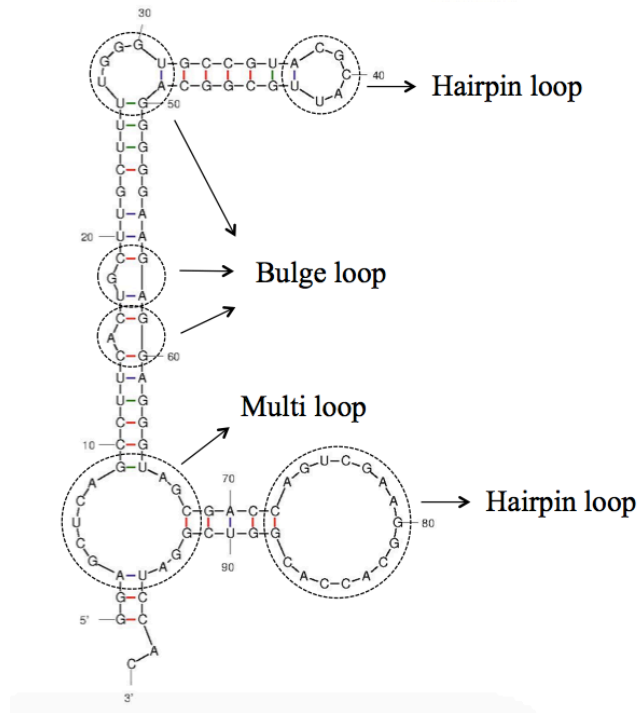
48	PRDBDapt48	AGAACACTCTGTTCT
49	PRDBDapt49	AGAACAGAGTGTTCT
50	PRDBDapt50	AGAACAGAATGTTCT
51	PRDBDapt51	AGAACAGATTGTTCT
52	PRDBDapt52	AGAACAGACTGTTCT
53	PRDBDapt53	AGAACAAACTGTTCT
54	PRDBDapt54	AGAACAAAATGTTCT
55	PRDBDapt55	AGAACAAATTGTTCT
56	PRDBDapt56	AGAACAAACTGTTCT
57	PRDBDapt57	AGAACATAGTGTTCT
58	PRDBDapt58	AGAACATAATGTTCT
59	PRDBDapt59	AGAACATATTGTTCT
60	PRDBDapt60	AGAACATACTGTTCT
61	PRDBDapt61	AGAACACAGTGTTCT
62	PRDBDapt62	AGAACACAATGTTCT
63	PRDBDapt63	AGAACACATTGTTCT
64	PRDBDapt64	AGAACACACTGTTCT

### 2.3.2 DNA Secondary Structure Prediction

The secondary structures of the DNA sequences were determined using the mfold online software based on the free energy minimization algorithm. The mfold server was chosen because of its feasibility and widespread usage for DNA secondary structure determination in many studies (Antunes et al., 2018; Baig et al., 2015; Yarizadeh et al., 2019). Moreover, DNA stacking, single mismatch, hanging end, terminal stacking, and loop parameters are also available on the site (Xia et al., 1998; Zuker, 2003). The server revealed that all the sixty-four ssDNA sequences were able to assume a simple secondary structure, forming one hairpin loop at the region A6-T10 (Figure 2.1). The secondary structure of PRapt-3 RNA aptamer also comprises hairpin loop structure (Figure 2.2). The secondary structure of the PRapt-3 has a  $\Delta G$  value of -30.05 kcal/mol whereas the  $\Delta G$  values of the aptamer candidates were either -2.87 or -3.87 kcal/mol (Table 2.2).

<pre> G G G A T C G A T A T G C A T </pre>	<pre> G G A A T C G A T A T G C A T </pre>	<pre> G G T A T C G A T A T G C A T </pre>	<pre> G G C A T C G A T A T G C A T </pre>	<pre> A G G A T C G A T A T G C A T </pre>	<pre> A G A A T C G A T A T G C A T </pre>	<pre> A G T A T C G A T A T G C A T </pre>	<pre> A G C A T C G A T A T G C A T </pre>
PRDBDapt1	PRDBDapt2	PRDBDapt3	PRDBDapt4	PRDBDapt5	PRDBDapt6	PRDBDapt7	PRDBDapt8
<pre> T G A T C G A T A T G C A T </pre>	<pre> T A A T C G A T A T G C A T </pre>	<pre> T T A T C G A T A T G C A T </pre>	<pre> T C A T C G A T A T G C A T </pre>	<pre> C G A T C G A T A T G C A T </pre>	<pre> C A A T C G A T A T G C A T </pre>	<pre> C T A T C G A T A T G C A T </pre>	<pre> C C A T C G A T A T G C A T </pre>
PRDBDapt9	PRDBDapt10	PRDBDapt11	PRDBDapt12	PRDBDapt13	PRDBDapt14	PRDBDapt15	PRDBDapt16
<pre> C G G A T C G A T A T G C A T </pre>	<pre> C G A A T C G A T A T G C A T </pre>	<pre> C G T A T C G A T A T G C A T </pre>	<pre> C G C A T C G A T A T G C A T </pre>	<pre> A C A T C G A T A T G C A T </pre>	<pre> A C A A A T C G A T A T G C A T </pre>	<pre> C A T A T C G A T A T G C A T </pre>	<pre> A C A C A T C G A T A T G C A T </pre>
PRDBDapt17	PRDBDapt18	PRDBDapt19	PRDBDapt20	PRDBDapt21	PRDBDapt22	PRDBDapt23	PRDBDapt24
<pre> T G A T C G A T A T G C A T </pre>	<pre> T A A T C G A T A T G C A T </pre>	<pre> T T A T C G A T A T G C A T </pre>	<pre> T C A T C G A T A T G C A T </pre>	<pre> C G A T C G A T A T G C A T </pre>	<pre> C A A T C G A T A T G C A T </pre>	<pre> C T A T C G A T A T G C A T </pre>	<pre> C C A T C G A T A T G C A T </pre>
PRDBDapt25	PRDBDapt26	PRDBDapt27	PRDBDapt28	PRDBDapt29	PRDBDapt30	PRDBDapt31	PRDBDapt32
<pre> T G A T C G A T A T G C A T </pre>	<pre> T A A T C G A T A T G C A T </pre>	<pre> T T A T C G A T A T G C A T </pre>	<pre> T C A T C G A T A T G C A T </pre>	<pre> C G A T C G A T A T G C A T </pre>	<pre> C A A T C G A T A T G C A T </pre>	<pre> C T A T C G A T A T G C A T </pre>	<pre> C C A T C G A T A T G C A T </pre>
PRDBDapt33	PRDBDapt34	PRDBDapt35	PRDBDapt36	PRDBDapt37	PRDBDapt38	PRDBDapt39	PRDBDapt40
<pre> T G A T C G A T A T G C A T </pre>	<pre> T A A T C G A T A T G C A T </pre>	<pre> T T A T C G A T A T G C A T </pre>	<pre> T C A T C G A T A T G C A T </pre>	<pre> C G A T C G A T A T G C A T </pre>	<pre> C A A T C G A T A T G C A T </pre>	<pre> C T A T C G A T A T G C A T </pre>	<pre> C C A T C G A T A T G C A T </pre>
PRDBDapt41	PRDBDapt42	PRDBDapt43	PRDBDapt44	PRDBDapt45	PRDBDapt46	PRDBDapt47	PRDBDapt48
<pre> G A G A T C G A T A T G C A T </pre>	<pre> G A A A T C G A T A T G C A T </pre>	<pre> G A T A T C G A T A T G C A T </pre>	<pre> G A C A T C G A T A T G C A T </pre>	<pre> A A C A T C G A T A T G C A T </pre>	<pre> A A A A T C G A T A T G C A T </pre>	<pre> A A T A T C G A T A T G C A T </pre>	<pre> A A C A T C G A T A T G C A T </pre>
PRDBDapt49	PRDBDapt50	PRDBDapt51	PRDBDapt52	PRDBDapt53	PRDBDapt54	PRDBDapt55	PRDBDapt56
<pre> T A A T C G A T A T G C A T </pre>	<pre> T A A T C G A T A T G C A T </pre>	<pre> T A T A T C G A T A T G C A T </pre>	<pre> T A C A T C G A T A T G C A T </pre>	<pre> C A G A T C G A T A T G C A T </pre>	<pre> C A A A T C G A T A T G C A T </pre>	<pre> C A T A T C G A T A T G C A T </pre>	<pre> C A C A T C G A T A T G C A T </pre>
PRDBDapt57	PRDBDapt58	PRDBDapt59	PRDBDapt60	PRDBDapt61	PRDBDapt62	PRDBDapt63	PRDBDapt64

Figure 2.1 Secondary structures of DNA sequences predicted by mfold server



**Figure 2.2** Secondary structure prediction of the PRapt-3 RNA aptamer using mfold

**Table 2.2** The  $\Delta G$  value of the secondary structures of DNA aptamer candidates

Number	Sequence name	$\Delta G$ value (kcal/mol)
1	PRDBDapt1	-2.87
2	PRDBDapt2	-3.87
3	PRDBDapt3	-2.87
4	PRDBDapt4	-2.87
5	PRDBDapt5	-2.87
6	PRDBDapt6	-2.87
7	PRDBDapt7	-2.87
8	PRDBDapt8	-2.87
9	PRDBDapt9	-2.87
10	PRDBDapt10	-2.87
11	PRDBDapt11	-2.87
12	PRDBDapt12	-2.87
13	PRDBDapt13	-2.87
14	PRDBDapt14	-2.87
15	PRDBDapt15	-2.87
16	PRDBDapt16	-2.87
17	PRDBDapt17	-2.87
18	PRDBDapt18	-3.87
19	PRDBDapt19	-2.87
20	PRDBDapt20	-2.87
21	PRDBDapt21	-2.87
22	PRDBDapt22	-2.87
23	PRDBDapt23	-2.87

24	PRDBDapt24	-2.87
25	PRDBDapt25	-2.87
26	PRDBDapt26	-2.87
27	PRDBDapt27	-2.87
28	PRDBDapt28	-2.87
29	PRDBDapt29	-2.87
30	PRDBDapt30	-2.87
31	PRDBDapt31	-2.87
32	PRDBDapt32	-2.87
33	PRDBDapt33	-2.87
34	PRDBDapt34	-3.87
35	PRDBDapt35	-2.87
36	PRDBDapt36	-2.87
37	PRDBDapt37	-2.87
38	PRDBDapt38	-2.87
39	PRDBDapt39	-2.87
40	PRDBDapt40	-2.87
41	PRDBDapt41	-2.87
42	PRDBDapt42	-2.87
43	PRDBDapt43	-2.87
44	PRDBDapt44	-2.87
45	PRDBDapt45	-2.87
46	PRDBDapt46	-2.87
47	PRDBDapt47	-2.87