Development and Application of Ion Pair Reverse Phase HPLC for the Analysis of DNA Aptamers, dsRNA Biocontrols and mRNA Therapeutics

Paul Edward Coombes

Submitted for the degree of Doctor of Philosophy

September 2024

Department of Chemical, Materials and Biological Engineering

University of Sheffield

Declaration of Originality

In accordance with university regulations, I hereby declare that:

1. This thesis has been produced solely by me.
2. It is entirely my own work.
3. It has not been submitted in part or whole for any other degree or personal qualification.

Based on work within this thesis, the following article has been published.

Chapter 3:

P.E. Coombes, M.J. Dickman, Optimisation of denaturing ion pair reversed phase HPLC for the purification of ssDNA in SELEX., J Chromatogr A (2024) 1–10. https://doi.org/10.1016/j.chroma.2024.464699.

Acknowledgements

First and foremost, I would like to thank Professor Mark Dickman for providing me with the opportunity to undertake this PhD; his constant guidance and mentorship ensured that the project could progress even during the most challenging stages. I want to thank Dr Caroline Evans, Dr Alison Nwokeoji, and Dr Phillip Jackson. Their years of experience have proved vital to me as an organic chemist entering the world of molecular biology.

I want to thank other members of the MJD group, including James Grinham for help with HPLC maintenance, Gareth Owen for his help with mass spectrometry data processing and Sebastion Ross for assistance with microbiology and cell culture. Additionally, I would like to thank the rest of the MJD group for their support, professionally and as friends, during my time there.

Finally, I want to thank my mum for her constant support over the last four years.

Abbreviations

AEX – Anion exchange

AGO – Argonaut Protein

ASOs – Antisense oligonucleotides

BP – Base pairs

BSA – Bovine serum albumin

CAS-9 – CRISPR associated protein-9

CE – Capillary Electrophoresis

CGE – Capillary gel electrophoresis

CRISPR – Clustered regularly interspaced short palindromic repeats

CRNA – CRISPR-RNA

CSP - Covid Spike protein

CZE - Capillary zone electrophoresis

DBAA – Di-butyl ammonium acetate

DNA – Deoxyribose nucleic acid

DNTPS – Deoxynucleotide triphosphates

Ds – double-stranded

E. coli - Escherichia coli

EMA – European medicines agency

EPA – Environmental protection agency

ESI – Electron spray ionisation

FDA – Food and Drug Administration

GalNAc- N-acetylgalactosamine

GFP – Green fluorescent protein

GMP – Good manufacturing practice

gRNA- Guide RNA

HFIP – Hexafluoro isopropanol

IMAC – Immobilised metal affinity chromatography

ITC – Isothermal titration calorimetry

IVT – In vitro transcription

LNP – Lipid nanoparticle

MDA-9 - Melanoma differentiation-associated gene-9

MHRA - Medicines and healthcare products regulatory agency

mRNA – Messenger ribose nucleic acid

MS – Mass spectrometry

NGS – Next generation sequencing

NS – Negative selection

NT - Nucleotides

NTPS – Nucleotide triphosphates

OGN – Oligonucleotides

PEG – Polyethylene glycol

PCR – Polymerase chain reaction

PSDVB – Poly(styrene divinylbenzene)

RNA – Ribose nucleic acids

RISC – RNA induced silencing complex

RP IP HPLC – Reverse phase ion-pair high-performance liquid chromatography

SDS – Sodium dodecyl sulphate

SELEX – Selective enhancement of ligands by exponential enrichment

siRNA – Small Interfering RNA

SPR – Surface plasmon resonance

Ss – single-stranded

TBAA – Tributyl ammonium acetate

TEA – Triethyl ammonia

TEAA – Triethyl ammonium acetate

USP – United states pharmacopeia

UTR – Untranslated region

UV – Ultra-violet

VEGF – Vascular endothelial growth factor

Abstract

Nucleic acids are emerging as a powerful new class of therapeutic and play an increasingly important role in biotechnology. The global mRNA therapeutic market alone is expected to reach $128.14 billion by 2030. mRNA vaccines significantly reduce vaccine development time due to their ease of customisation, and DNA/RNA aptamers are emerging as potential therapeutics and have broader applications in biotechnology. dsRNA biopesticides are also emerging as novel sustainable pesticides. However, each of these faces unique challenges; while dsRNA biocontrols such as Ledprona have been successfully deployed and mRNA vaccines proved vital during the COVID pandemic, they are subject to strict regulations and quality control. Robust analytical methods are required to fully characterise these biomolecules to support manufacturing and ensure their safety and efficacy. ssDNA aptamers are currently developed using in vitro selection (SELEX). However, high failure rates can impact the development and generation of new aptamers.

This study aimed to develop and utilise ion pair reverse phase HPLC (IP RP HPLC) to analyse and characterise mRNA therapeutics and dsRNA biocontrols. In addition, it was proposed that the SELEX procedure be optimised before generating ssDNA aptamers for the melanoma differentiation-associated gene-9 (MDA-9). IP RP HPLC was used under denaturing conditions to develop a novel method for the isolation of ssDNA in SELEX. The novel method enabled the rapid purification of ssDNA with an 80% yield, demonstrating significant advantages over existing methods. This method was subsequently utilised in SELEX to obtain ssDNA aptamers to Syntenin-1. Modifications to the SELEX approach resulted in the successful generation of ssDNA aptamers against Syntenin-1. Binding affinity studies demonstrated that the ssDNA aptamers had binding affinities in the 100-600 nM range, with the highest affinity aptamer at 25 nM.

IP RP HPLC was also developed to accurately size dsRNA biocontrols and mRNA therapeutics. A wide range of ion pair reagents were studied under denaturing and non-denaturing conditions. The results showed that accurate sizing of a range of dsRNA biocontrols resulted in less than 2.5% error using dibutyl ammonium acetate in the mobile phase. The IP RP HPLC methods developed in this thesis offer significant advantages over existing methods for generating ssDNA and accurately sizing dsRNA biocontrols. In addition, further work building on the ssDNA aptamers against Syntenin-1 could develop novel therapeutics as inhibitors to a protein important to cancer proliferation.

Table of Contents

List of Tables

List of Figures

1 Chapter 1 Nucleic acids, applications and analysis.

1.1 DNA, RNA, and the Genetic Code

The importance of nucleic acids and their central role in life's origin cannot be overstated. Allowing for the storage of genetic information, nucleic acids allowed early organisms to replicate themselves, driving the formation of the biosphere as we know it today. It is essential to contemplate how their structure facilitates their functional role within cells and how they can be exploited for other applications.

1.1.1 Isolation and Discovery

In one early account, the isolation of nucleic acids happened almost by accident; Friedrich Miescher was studying the physiology of human lymph cells when he precipitated a phosphorus-containing substance from cell nuclei, calling this substance nucleolin, he went on to find it in many cell types[1]. Richard Altman later gave nucleic acids their name, where it was later evident that they played a crucial role in cell division [1]. The first detailed molecular studies of nucleic acids came from the work of Levene and Jacobs at the Rockefeller Institute; alkaline hydrolysis revealed what made up these enigmatic species [2],[3].

1.1.2 Structure of Nucleic Acids

A nucleotide unit comprises a phosphate group, a pentose sugar, and a nitrogenous base. The primary structure of nucleic acids (DNA) is a linear polynucleotide in which 2’ deoxyribonucleoside is linked to the next through a 3’ to 5’ phosphodiester (see Figure 1.1).

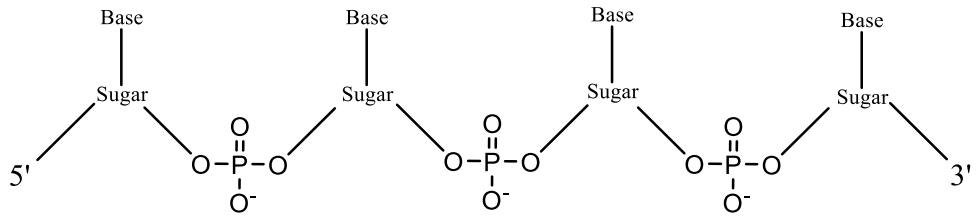


Figure 1.1: Phosphodiester backbone of oligonucleotides: *monomers consist of a sugar bonded to a base and a phosphate, which acts as the linker in the polymer chain.*

The two nucleic acids found commonly in nature are ribonucleic acids (RNA) and deoxyribonucleic acids (DNA). RNA is made up of ribonucleotides, whereas the monomers of DNA are 2'-deoxyribonucleotides. The corresponding ribose and 2' deoxyribose sugars are shown in Figure 1.2.

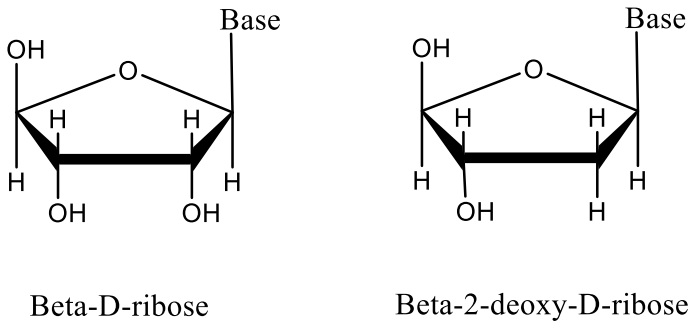


Figure 1.2: Structures of the sugar groups of RNA (left) and DNA (right): *RNA consists of three hydroxyl groups on the sugar, while DNA, being reduced, has two.*

The third major component of nucleotides is nitrogenous bases; there are five common bases in nature that occur in nucleic acids (see Figure 1.3**).** Adenine cytosine and guanine are found in RNA and DNA, whereas uracil is found primarily in RNA and thymine primarily in DNA. The bases themselves are conjugated heterocyclic compounds. The primary structure of DNA is shown in Figure 1.4.

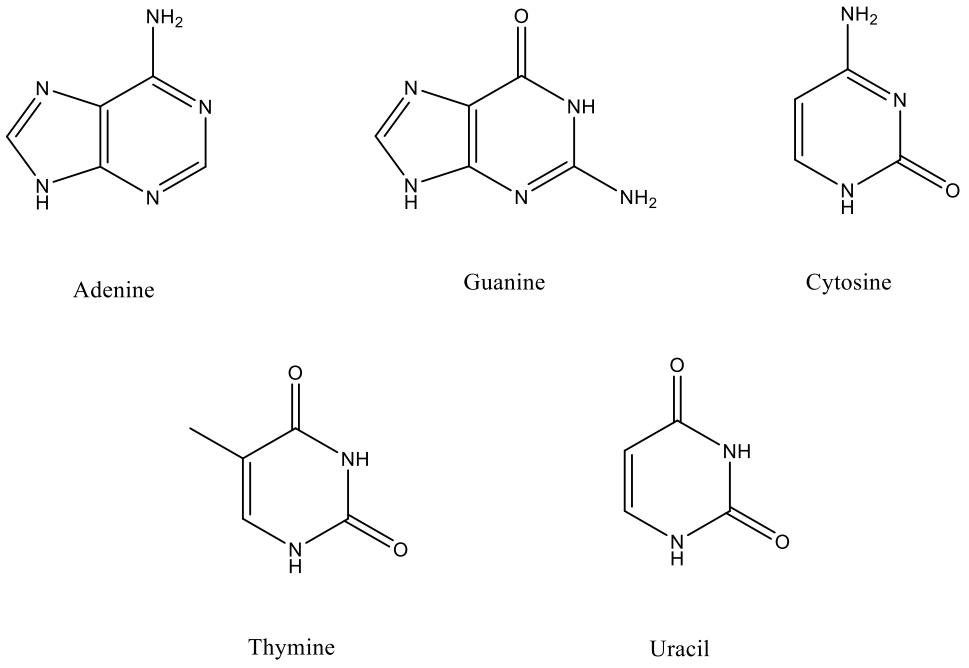


Figure 1.3: The five naturally occurring bases in nucleic acids: *Two purines, adenine, and guanine, and the three pyrimidines, cytosine, thymine, and uracil.*

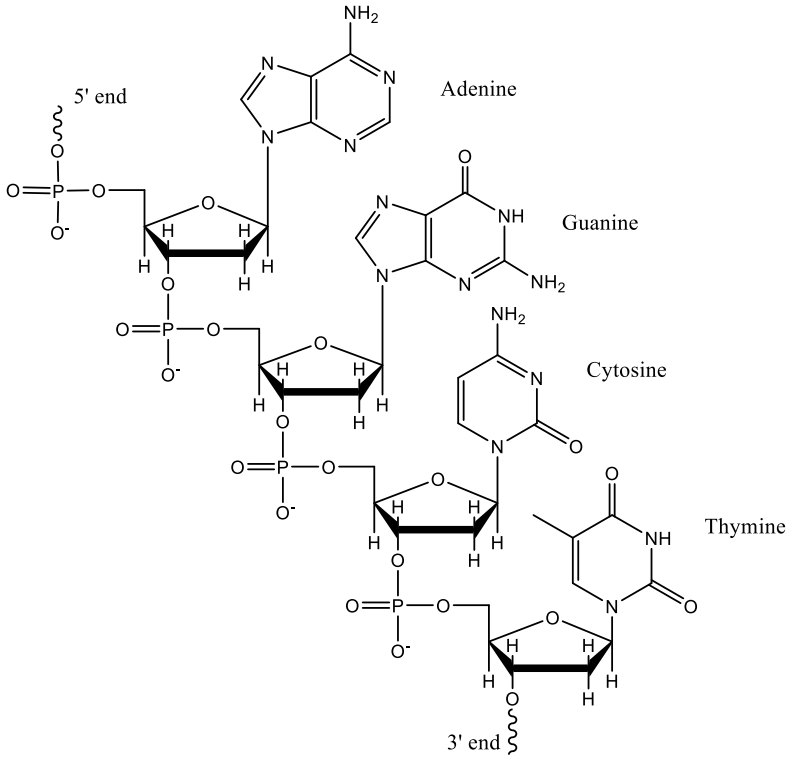


Figure 1.4: The primary structure of nucleic acids: *DNA represented in chain form containing its four constituent nucleotides, Adenine, Guanine, Cytosine and Thymine*

In 1953, the supramolecular structure of DNA was revealed to be a double-stranded right-handed helix (B form duplex DNA) by the work of Franklin and Gosling, Watson and Crick, and Wilkins Stokes and Wilson [4–6]. The helix formation is facilitated by hydrogen bonding between the nitrogenous bases. The nitrogenous bases can act as both proton donors and acceptors, the electronegative oxygen and nitrogen atoms act as hydrogen acceptors, and the protons on the amino groups are sufficiently positively charged (due to their proximity to electronegative nitrogen atoms) to interact with the acceptors forming hydrogen bonds. Some significant motifs involved in hydrogen bonding in the case of guanine and cytosine are shown in Figure 1.5**.** Initial studies into the helical structure of DNA showed that nucleotide bases preferentially form hydrogen bonds with specific complementary bases; the base pairs are adenine and thymine, guanine and cytosine and in RNA where thymine is not present, adenine bonds to uracil Figure 1.5.

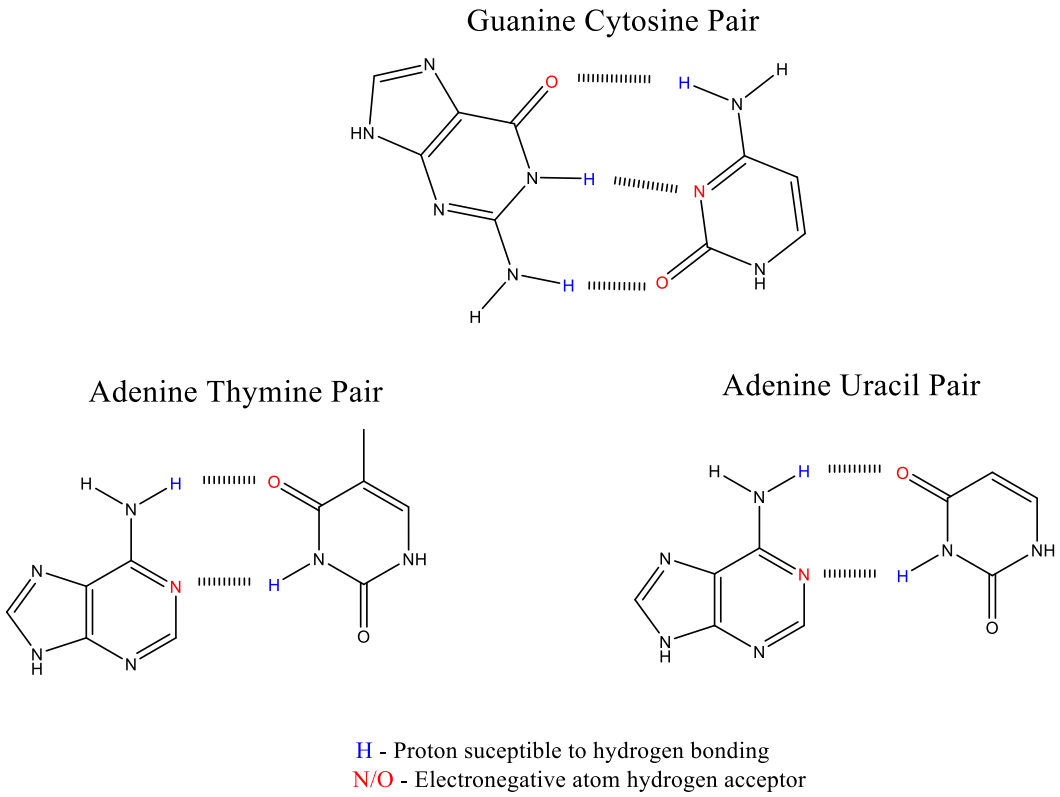


Figure 1.5: Complementary base pairing in DNA and RNA: *Adenine-Thymine, Guanine-Cytosine, Adenine-Uracil.*

The consistent specificity of the base pairing is what allows the DNA strands to form a helix once joined; it is important to note that the guanine-cytosine pair contains three hydrogen bonds, whereas the adenine-thymine/uracil pair has two, the guanine-cytosine linkage is stronger meaning DNA with a higher G-C content has more robust physical properties such as higher melting points. While the right-handed helix was the first secondary structure of nucleic acids to be discovered, it became abundantly clear with further research that DNA can adopt a variety of conformations depending on the DNA sequence and interaction with proteins [7]. This can occur when bases "break the rules" of the Watson-Crick model, forming mismatched pairs; these cause steric changes within the structure, leading to the formation of non-standard duplex and hairpin motifs [8]. Guanine-rich sequences in DNA have been demonstrated to form four-stranded complexes based on a central four Guanine bases forming a tetrad. This is notable as the G-tetrad structures are highly thermostable relative to their counterparts [9]. Studying these non-Watson crick base pairs and their resulting structures has led to the discovery of ten different DNA conformations, each capable of exciting interactions and chemistries. This opens the possibility of nucleic acids being capable of much more than just storing genetic information [10].

1.2 Applications of Nucleic Acids in Biotechnology and Medicine

1.2.1 Nucleic Acid Therapeutics

In 1972, Friedmann and Roblin hypothesised that disorders arising from a faulty gene variant could be treated by introducing a functional gene variant [11]. Therapy utilizing gene replacement was the first to be conceptualised, but prevention of transcription via blocking with an anti-sense fragment or introducing genetic material to express a deficient protein or enzyme was hypothesised [12]. Since this early conception, nucleic acid therapeutics have emerged as an important new class of drug, as demonstrated by the recent approval of several therapeutics by the United States Food and Drug Administration (FDA) and the European Medicines Agency (EMA) [12].

There are several alternative classes of nucleic acid therapeutics (NATs), including:

1. Antisense oligonucleotides (ASOs)
2. Small interfering RNA (siRNA)
3. Aptamers
4. Guide RNAs (gRNAs)

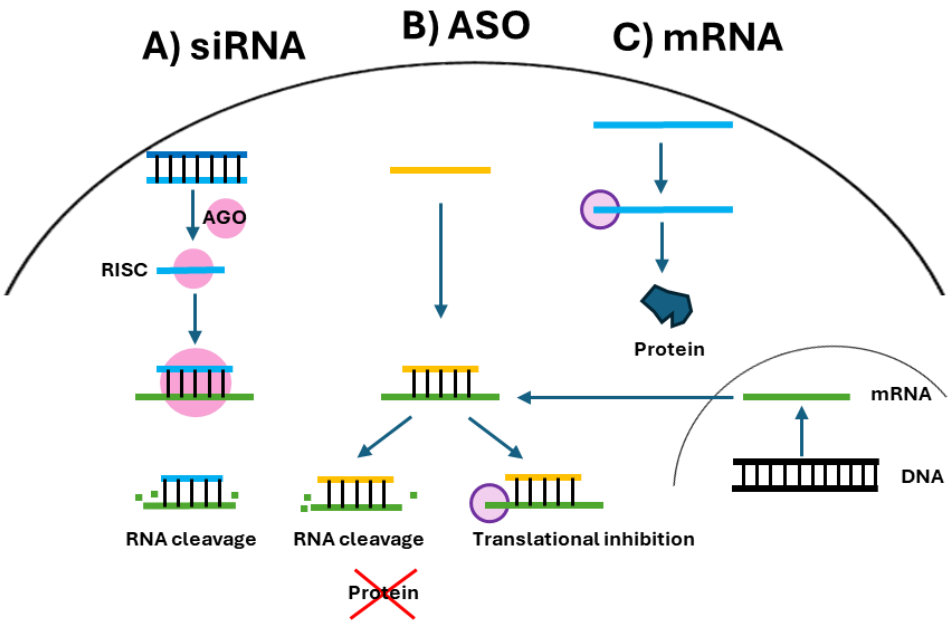
Antisense oligonucleotides (ASOs), small interfering RNAs (siRNAs)are involved in RNA interference (RNAi), aptamers, and guide RNAs (gRNAs), which are involved in CRISPR gene editing technology (Figure 1.6).

Currently, 15 such drugs have been clinically approved [13]. Most of those approved in the United States are ASOs or siRNAs [14]. A key challenge facing these technologies is the delivery of exogenous genetic material and its susceptibility to nuclease degradation. Several platforms have been developed to achieve delivery of genetic information in vivo; these include chemically modified antisense oligonucleotides (ASOs), N-acetyl galactosamine (GalNAc) ligand-modified short interfering RNA (siRNA) conjugates, lipid nanoparticles (LNPs), and adeno-associated virus (AAV) vectors [12]

1.2.2 mRNA therapeutics/vaccines

Prior to the COVID-19 pandemic, mRNA vaccines were already showing great promise as therapeutics combining molecular biology and immunology as an

alternative to protein or live virus vaccines [15]. Unlike traditional vaccines, mRNA carries genetic information, allowing the production of the desired antigen in vivo [16]. Injection of mRNA trans dermally affords the expression of a desired protein in the local area (Figure 1.6) [17]. mRNA offers numerous advantages over classical approaches to vaccination, including ease of large-scale production and customisation to new pathogens [18]. Years of prior research led to the mRNA vaccine system being ready for use during the COVID-19 pandemic; the most commonly used vaccines were Pfizer BionTech's BNT162b2 mRNA and Moderna's mRNA-1273 [19].



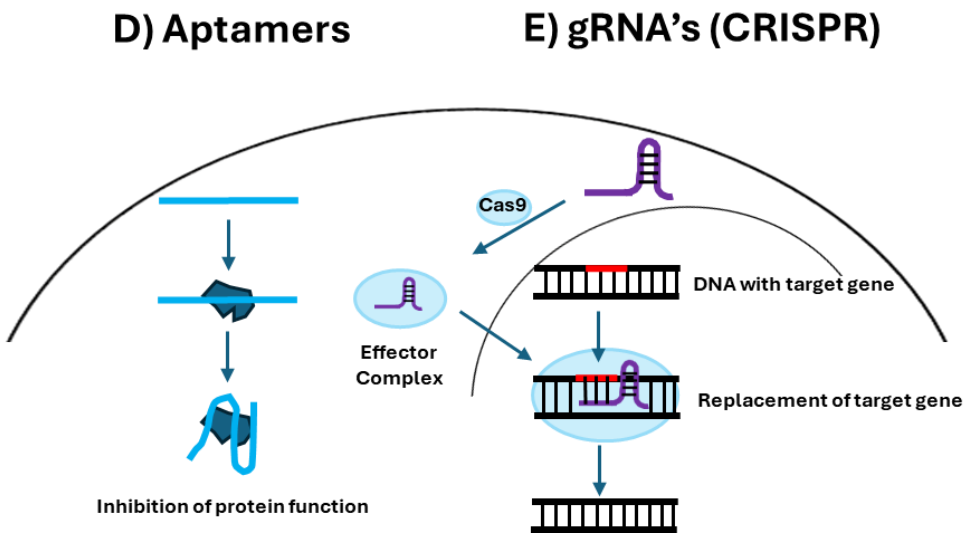


Figure 1.6: Mechanism of action or RNA therapeutics and mRNA vaccines. ***(A)*** *siRNAs are small double-stranded RNA molecules with exact complementarity to a target messenger RNA. Once associated with the Argonaute protein (AGO) and RNA induced silencing (RISC) complex, it binds to its target mRNA and induces gene silencing by preventing translation.* ***(B)*** *ASOs are small single-stranded RNA molecules with complementarity to a target mRNA. Once bound, they induce post-transcriptional gene silencing by preventing translation.* ***(C)*** *mRNA vaccination, exogenous mRNA is injected transdermally and transcribed into a protein antigen in local cells, triggering an immune response.* ***(D)*** *Aptamers are short RNA, DNA molecules that fold into secondary and tertiary structures then bind to a target molecule (often a protein), acting as an antagonist suppressing the pathway the target is involved in.* ***(E)*** *Guide RNA, consisting of tracer RNA linked CRISPR-RNA (cRNA) combines with CRISPR associated protein-9 (CAS-9) to form an effector complex which can cleave DNA where one strand can hybridise with the gRNA.[20]*

1.2.3 Production of mRNA for vaccines

mRNA vaccines are huge nucleic acids, typically over 3000 nucleotides in length. They are composed of the gene of interest, a poly-adenosine (poly[A]) tail > 80 bases in length at the 3´ end, which enhances stability and translation efficiency and a 7-methylguanosine cap structure at the 5´ end, which enables stable translation [21]. At the 3´ and/or 5´ ends, untranslated regions (UTRs) can be incorporated to enhance stability and translation efficiency [19,22].

In addition to the wide range of mRNA sizes used in vaccines, structural modifications on the individual nucleotides have been utilised within mRNA vaccines for increased therapeutic efficacy. As early as 2008, pseudo-uridine-modified mRNA showed a greater translation rate in mammalian cells. This demonstrated an increase in the stability of modified mRNAs [23]. Additionally, methylation has been shown to reduce the immunogenicity of mRNAs [24].

Production of mRNA for use in vaccinations/therapeutics is currently performed via in vivo transcription (IVT), an enzymatic process which involves the generation of mRNA from a DNA template using an RNA polymerase (T7, SP6 or T3), nucleotide triphosphates (NTP's), $MgCl_2$ (polymerase co-factor) and a suitable pH buffer [25]. Following IVT, mRNA must be isolated and purified from the reaction mixture, which contains residual enzyme, DNA template, NTPs and mRNA by-products; standard purification methods such as DNase treatment followed by column purification can remove most of these impurities, but IVT by-products remain [26].

1.3 RNAi-based crop protection

Plant pests and diseases currently account for a loss of 20-40% of global crop yields annually, with insects consuming 5-20% of major grain crops [27]. With the global population set to hit 9 billion within the coming decades, this will pose an unprecedented threat to global food security [28]. Alongside the increase in pest populations, pesticide use has steadily declined since 2007 due to increased regulatory pressure and negative public opinion [28]. This has primarily been driven by increasing pesticide resistance and public concern with industrial farming practices, notably off-target effects on vital pollinating insects [29,30]. Current pest control methods rely on insecticidal and fungicidal sprays and inducing plant genetic

resistance to pests via genetic modification. Therefore, there is increasing demand for novel, sustainable strategies for crop protection driven by population increase, regulatory demands, pest range expansion and increasing resistance to current methods [31].

dsRNA pesticides have been demonstrated as successful gene silencers via RNA interference. The post-transcriptional inhibition of gene expression, utilising short dsRNA to downregulate the expression of genes with complementary sequences, was first discovered in roundworms and subsequently seen in plants [32,33]. Fire et al. (1998) demonstrated that oral delivery of dsRNA induced the RNA interference pathway as early as 1998, and recently, in 2023, 'Ledprona', a dsRNA produced by Greenlight Biosciences, was approved by the US Environmental Protection Agency EPA for use on the Colorado potato beetle [28,34]. Ledprona targets the Potato Beetle via the RNA interference mechanism, silencing the PSMB5 gene via antisense hybridisation it achieves 90% mortality after six days of exposure [35].

dsRNA is produced in bulk either by using in vitro transcription (IVT) of the two strands and their subsequent annealing or in vivo via expression in a living system such as *E. coli*, the most prevalent production method [36,37]. Following expression of the dsRNA, it is extracted from microbial cells and purified via RNAase treatment to remove 5' and 3' ssRNA overhangs. The short half-life of dsRNA in the environment, combined with the efficiency of the interference mechanism, varies significantly from species to species, which means that dsRNA must be produced in large quantities with approximately 2-8 g required per hectare [38–40].

1.4 Aptamers: Nucleic Acids as Biochemical Agents

The critical roles nucleic acid plays in storing genetic information and templating for replication have been known for decades. In the 1960s, suggestions came that, in addition to this, DNA and RNA may have the potential to fold up into complex 3D structures and perform functional roles in the cell, like proteins [41]. In the 1980s, it was demonstrated that RNA could catalyse chemical reactions like enzymes, which opened the door to greater speculation about the role of RNA in the early origins of life on Earth, the theory of the RNA world [42][43].

With the notion that short nucleic acids are capable of introductory chemistry came the idea that they could act as nucleic acid-based counterparts to antibodies. In 1990,

Tuerk and Gold published a paper describing the driven evolution of 3D RNA oligonucleotide structures to act as high-affinity ligands for bacterial polymerase [44]. The work entailed subjecting a library of RNA oligonucleotides (eight bases in length) to successive rounds of selection against the desired target so that a ligand to the polymerase would be isolated through gradual enrichment [44]. They isolated two ligands from a pool of an estimated 65,000. Tuerk and Gold stated that the selection/amplification protocols could yield high-affinity ligands for any protein that binds nucleic acids as part of its function; high-affinity ligands could conceivably be developed for any target molecule [44].

This work was carried out at the same time as a group headed by Ellington and Szostak, who published a paper in the same year, “In vitro selection of RNA molecules that bind specific ligands [45]. The work entailed the progressive isolation of ligands via in vitro selection against various organic dyes from a library of $1x10^{10}$ sequences. Ellington deduced that one species in a library of $1x10^{10}$ would fold in such a way that would allow it to bind specifically to the desired target. This was exciting because he deduced that the same library could isolate ligands to multiple targets.

Aptamers, as this new class of affinity ligands would be called, come from the Latin “Aptus” to fit, and the Greek “meros”, meaning region, are short single-stranded oligonucleotides produced by an in vitro enrichment workflow which is driven by Darwinian evolution [46],[47]. Aptamers can bind to small molecules, metal ions, proteins, and whole cells. As a class of affinity ligands, their broad capabilities have led to their utilisation in various applications such as drug delivery, diagnosis and biosensing [48][49]. Before aptamers appeared, many biomolecules were available to act as affinity ligands for given targets, including antibodies, nanobodies, and affibodies, which were the most common [50]. Tuerk and Gold always believed nucleic acid aptamers would replace antibodies for in vitro and in vivo applications. A comparison of the different properties of antibodies and aptamers is shown in Table 1.1 [51].

Table 1.1 shows that aptamers offer many advantages as a therapeutic, especially in the long term; post-approval, a clinically approved aptamer would be cheaper to produce and keep a longer shelf life than an antibody developed for the same purpose.

Table 1.1: Comparison of Aptamers to Antibodies [52].

Property	Aptamer	Antibodies
Molecular weight	15-30 kDa	150-180 kDa
Generation time	Hours to months	Around six months
Immunogenicity	Low	High
Targets	Wide range of targets	Immunogenic molecules
Shelf life (4 ºC)	Years	Weeks to months
In vivo half-life	Tens of minutes	1 month
Nuclease degradation	Susceptible	Resistant
Stability	Stable	pH and thermosensitive

1.4.1 Aptamer generation via SELEX

After the early discoveries of Gold, Ellington and Szostak, further research was directed towards improving the aptamer generation procedure; this led to a standardised approach to in vitro selection, SELEX. The process of SELEX is a cycle that involves five main steps: (i) Design of the nucleic acid library (ii) placing the library in contact with the target to separate molecules in the library that bind from those that do not; (iii) amplifying “survivors” by the polymerase chain reaction (PCR), (iv) re-generation of ssDNA following PCR and (v) Sequencing of the aptamer pool (after an appropriate number of cycles) [53]. The first SELEX protocol devised by Tuerk and Gold was laborious. However, each process step has been closely scrutinised and improved in the following decades [54]. A schematic representation of the basic steps of SELEX is shown in Figure 1.7.

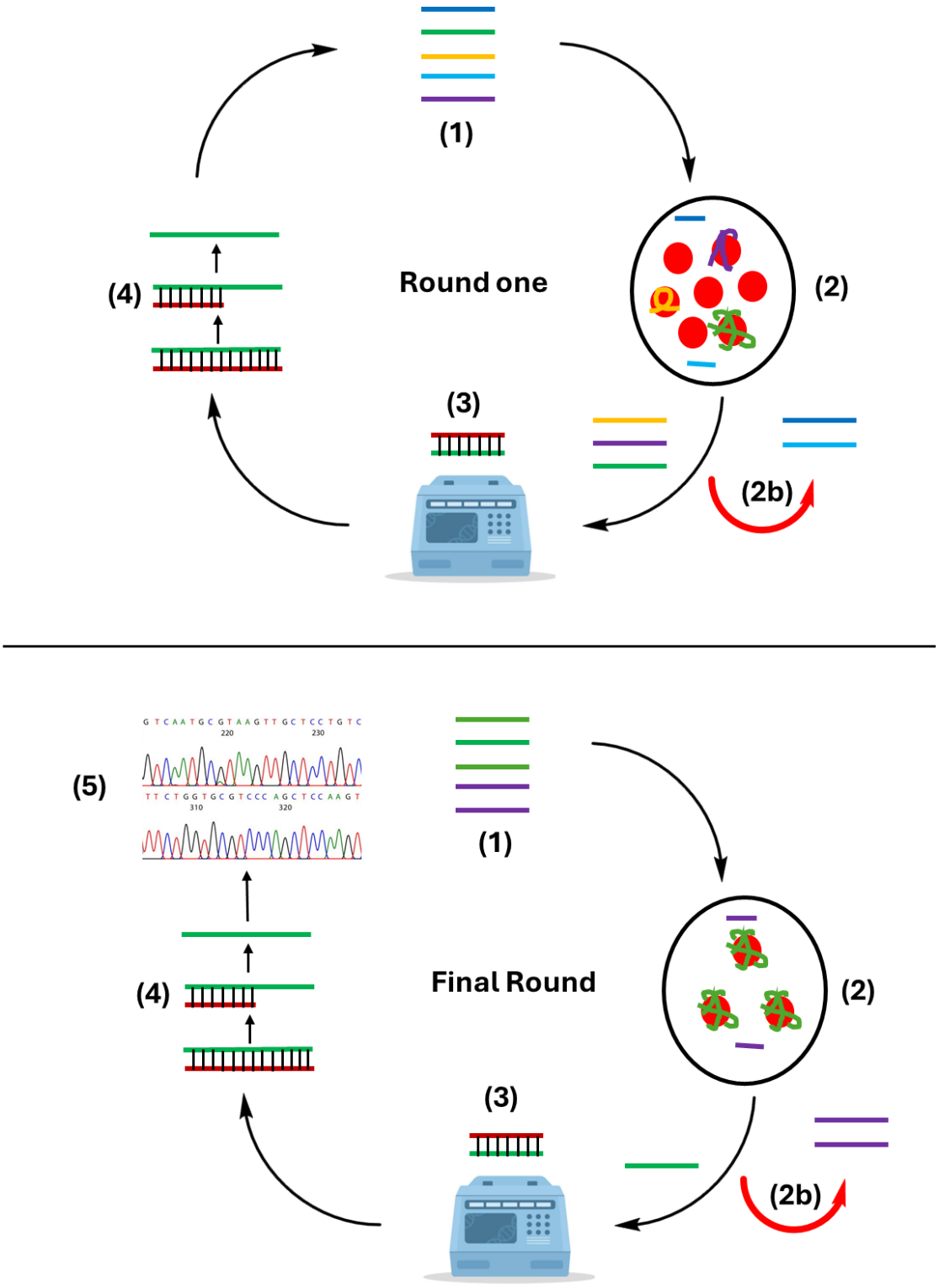


Figure 1.7: The conventional SELEX process, the first and final rounds: ***(1)*** *Initial ssDNA library,* ***(2)*** *Incubation of library with a target,* ***(2b)*** *Elution of low-affinity species,* ***(3)*** *Amplification of desired, high-affinity species (aptamers)* ***(4)*** *Re-generation of single-stranded DNA (ssDNA)* ***(5)*** *Sequencing of aptamers*

As described in Figure 1.7, the conventional SELEX process (as it was in the 1990s) can be broken down into five key stages, each of which has its key steps depending on the methods employed. All protocols begin at the same starting point, a library of single-stranded (ss) nucleic acids chemically synthesised to have random centralised regions, which can be anything from 8 to over 100 bases in lengths flanked on either side by priming regions for PCR amplification, which are typically ~20 bases in length [55]. While the initial library is extensive, containing approximately 10^{15} sequences, only a few are expected to have any natural affinity for the target. Therefore, a series of purifications or 'washes' must be performed to eliminate non-binding species. This is the key focus of the selection step [47]. Following the isolation of a small number of sequences with high affinity for the target molecule, these must be amplified before further rounds of selection; this generally occurs via PCR. Following amplification, the ssDNA must be re-generated from the dsDNA PCR products before additional selection rounds. Following the final rounds of selection, when the library has become enriched with sequences binding to the target, sequencing is performed to identify aptamer candidates, which are screened for their affinity to the target.

Aside from the initial two papers, single-stranded DNA aptamers were predominantly the focus of the early SELEX experiments in the 1990s. However, due to the potential more significant structural and conformational diversity that RNA sequences can adopt, RNA was used in SELEX experiments to generate higher affinity aptamers [56][57]. However, many studies generating ssDNA aptamers showed that ssDNA has a similar capacity to RNA in forming complex 3D structures and has potential advantages over RNA in SELEX. It is widely known that DNA is more chemically stable than RNA, making selection easier and offering aptamers with a longer shelf life. This amplification of DNA does not require the reverse transcription steps of RNA, making ssDNA SELEX protocols more time and cost-effective[58].

The minor differences in the structure of DNA and RNA result in significant differences in their reactivity and stability. The additional 2' hydroxyl group on RNA makes it more susceptible to nucleophile attack. The protocol for the generation of RNA aptamers is more laborious. Following elution from the target, ssRNA must be reverse transcribed and amplified as dsDNA via PCR prior to transcription to RNA and continuation of the protocol, of course, adds extra steps, increasing the potential for by-products, reducing the yield of RNA and the risk of nuclease degradation of the RNA.

1.5 Improving the chemical repertoire of nucleic acids in SELEX

The concern of Tuerk and Gold from the early days was the need for the selection phase of SELEX to be efficient; strong interactions lead to higher SELEX success rates [51]. Strategies have been developed to chemically modify the pyrimidine bases (uracil, thymine and adenine) at the 2 position with alkyl, hydrophobic and hydrophilic groups to increase the potential range of nucleic acid interactions. This allowed researchers to fine-tune the electrostatic and hydrophobic interactions of the nucleic acids and their targets [59].

1.5.1 Macugen

The in-vivo success of aptamer modifications led to the approval of Macugen by the FDA in 2004 for treating age-related macular degeneration. Macugen is a chemically modified RNA aptamer to vascular endothelial growth factor VEGF. VEGF is associated with ocular angiogenesis, the excessive sprouting of new blood vessels in the eyes, which leads to the degeneration of sight with age. Macugen has been clinically proven to be an effective treatment of age-related macular degeneration and was approved for use in the European Union in 2006 [60]. It took sixteen years for Aptamers to move from the laboratory bench to clinical approval, but it showed that the field had matured, SELEX had been shown as a viable method of drug discovery, and the resulting commercial exploitation led to remarkable progress in aptamer research [51,61].

1.6 An Overview of SELEX Methodology for generation of aptamers to in vivo protein targets

1.6.1 Library design and optimisation

In conventional methods, SELEX libraries contain a core random region (N flanked by two constant regions that act as primer binding sites for PCR [62]. The maximum theoretical size of the random region (N) is related to the number of randomised positions and the oligonucleotide diversity (y). A natural ssDNA library of 40 randomised positions has a maximum theoretical diversity of y^N and contains $4^{40} = 1.2 \times 10^{24}$ different molecules [63]. The diversity of the starting library holds the key to a successful SELEX protocol. The extensive array of sequences must, in turn, fold into

an extensive array of secondary and tertiary structures, where, in theory, a small number are proposed to bind to the target with high affinity,

A random region would not have to be particularly large to achieve a significant level of diversity (approximately 25 nucleotides). However, the aptamer must be long enough to fold into a complex 3D structure that can bind to a target, so most starting libraries have a core region between 30-50 nucleotides in length [63][64]. A summary of different length random regions used to obtain aptamers to different protein targets is shown in Table 1.2.

Table 1.2: Comparison of random region size in protein binding aptamers

Group/Year	Target	Size (nucleotides)	Citation
Jeong, Han, 2010	Kras Protein	40	[65]
Tan, 2006	Protein Kinase C	30	[66]
Parma, 2001	Tenascin C	40	[67]
Vater 2013	CXCL12	45	[68]
Xiang 2017	FOXM1	42	[69]

In the original in vitro evolution experiments, random pools of genomic nucleic acids were created by introducing random bases during in vitro transcription. However, the advent of more advanced synthetic techniques, including phosphoramidite chemistry and H- phosphate chemistry, allow for the synthesis of random regions up to 150 bases in length. Theoretically, mixing each phosphoramidite in a ratio of 25% (each base) should result in an equal ratio in the ssDNA library. However, due to known bias in the solid phase synthesis, this generally produces libraries with a few percentage points of drift from the desired 25% ratio [55].

While much research has been done into the starting library for SELEX, nucleic acids are simple molecules. Four bases dictate how the structure folds and forms motifs to bind to targets. The four bases do not interact electrostatically in the same fashion, so it stands to reason that different ratios of bases will have different potentials to form better aptamers. Computational studies have been carried out into what ratios of bases give rise to sequences with greater structural stability and selection potential. It was shown that pyrimidine-rich sequences (G and C) give these desired qualities; in

later studies, it was shown that a ratio of 25% U, 15% C, 20% A, and 40% G would give a high probability of finding aptamers with sufficient affinity [70][71].

1.6.2 Pattern SELEX

Aptamers require a secondary structure to bind to their targets; an approach was developed to incorporate design patterns into the initial library design to enrich the library into known structures that would act as stabilisers, allowing information-rich sequences to fold into complex 3D structures and bind to targets more effectively [48]. The first pattern to be exploited was the well-defined G-Quadruplex structure consisting of four guanine bases that folded on each other and formed a tetrad. These structures act as a scaffold for the various looping motifs of the aptamers, similar to the core structures of antibodies [48]. The core structure of the G-Quadruplex is shown in Figure 1.8.

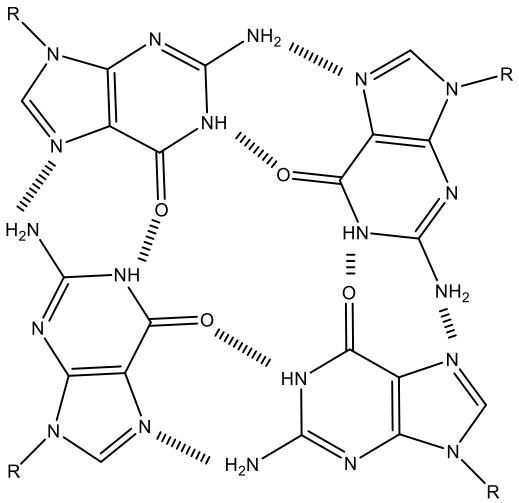


Figure 1.8: Secondary Structure of G-Quadruplex: *Four Guanine bases in a planar tetrad are linked by hydrogen bonding in a non-Watson-Crick manner*

G-Quadruplexes form naturally in guanine-rich oligonucleotides AS11 described in section 1.13 is an example of such a case, and its progress as the most advanced aptamer in cancer treatment shows the significance of G-Quadruplex structures [72]. In the case of protein targets, G-quadruplexes offer numerous advantages; the significant charge density of G-quadruplex DNA—twice the negative charge per unit length compared to duplex provides a high electrostatic potential for strong binding to

the positively charged surfaces of proteins, although hydrophobic interactions can also be critical for robust and specific binding to the target [73].

1.6.3 Fixed regions

One significant difficulty in generating high-affinity ssDNA aptamers in SELEX is the issue centred on the PCR amplification process that can generate high molecular weight (MW) by-products after repeated selection rounds [71,74]. Careful design of the primer binding regions has been used to help prevent such “size heterogeneity” issues during the PCR. Crameri and Stemmer demonstrated that introducing AT-rich sequences at the 5’ and 3’ ends improved the quality of PCR products, the principle being that A-T linkages are much weaker than G-C linkages, meaning primer chimaeras are less likely to form [75]. They also found that incorporating *E. coli* single-strand binding protein into the PCR reaction matrix destabilised large products [75][76].

1.6.4 Aptamer modification.

Nuclease resistance remains a limitation on aptamer performance in vivo. Nuclease enzymes readily cleave the phosphodiester backbone of DNA [77][78]. Chemical modifications of aptamers have been utilised to overcome these limitations while additionally providing greater structural diversity to pools, theoretically increasing SELEX success rate [79]. Modifications have been successfully utilised in SELEX on each of the three components of the structure: the phosphodiester backbone, the sugar and the base, as shown schematically in Figure 1.9 [80].

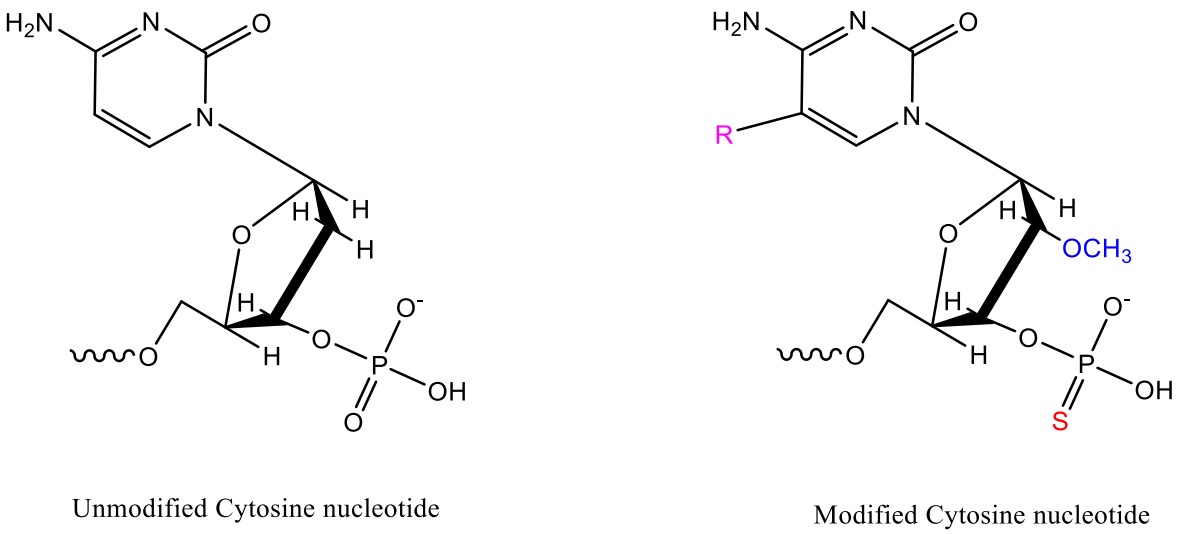


Figure 1.9: Comparison of common nucleotide modifications: *thiophosphate (Red), 2' substitution (Blue) and base modification (Pink) R = alkyl chain.*

Nuclease enzymes readily cleave the phosphodiester bonds which form the backbone of DNA; it was hypothesised that incorporating a sulphur atom onto the phosphate groups of DNA would make it less susceptible to the attack of the enzyme [78]. This should give the aptamer a longer half-life in vivo and a greater chance of finding and binding to the designated target. A thiophosphate-modified nucleotide is presented in Figure 1.9. The most common substitution on modified aptamers is the modification at the 2' position of the ribose sugar (see Figure 1.9). Various alternate substituents, including fluoro, amino, and hydroxymethyl groups, have been utilised, the main aim of which is increased resistance to nucleases [81]. The third class of nucleotide modifications is modifications to the base. The highly charged nature of oligonucleotides makes them hydrophilic, but many proteins show hydrophobic properties. To mimic this, slow off-rate Modified Aptamers (Somamers) introduced protein-like side chains into the structure of oligonucleotide aptamers; these alkyl chains increase the lipophilicity of the structure, allowing it to bind to lipophilic proteins [82].

Modified libraries can be synthesised by incorporating modified nucleotide precursors during the synthesis step; this is much more efficient than modification following synthesis [79]. Incorporation of modified deoxynucleotide triphosphates (dNTPs) into PCR is more challenging depending on the nature of the modification; enzymes may be more or less tolerant of its presence, for example, in the case of RNA aptamers.

While T7 RNA polymerase is tolerant of base modifications, Taq DNA polymerase is inefficient at incorporating modified dNTPs, this has led to screening for more efficient enzymes as well as attempts to engineer Taq polymerase to achieve this [79,83].

A comparison of in vivo longevity of thionucleotide and O-methyl substituted showed increased tissue penetration for both modifications, interestingly showing penetration of brain tissue, which was not seen in the control unmodified nucleotide aptamer. The thiophosphate aptamer also showed a greater half-life than the control aptamer [84].

1.7 Aptamer Selection

The concentration of the protein target in the selection step is a critical variable in the process. In contrast, the concentration needs to be high enough for aptamers with affinity to interact and bind to the target; weaker ligands may be retained following harsh selection pressures if they are in a matrix saturated with the target protein. Studies have shown that lower target concentrations give a higher probability of isolating high-affinity ligands, while inversely, higher target concentrations yield more low-affinity aptamers [54]. From a selection perspective, low-affinity aptamers are more damaging to the effectiveness of the process than negligible-affinity aptamers [54]. It is typical for the target-to-aptamer ratio to be adjusted throughout the protocol, starting with 1:100 and slowly approaching 1:1000 to increase competition between the ligands [54].

1.7.1 Negative Selection Steps

Conventionally, the target is bound to a solid support that enables affinity chromatography-based methods in the SELEX protocol. However, precautions must ensure no selection pressure towards aptamers that bind the solid support. Additional negative selections can be performed against components of the background matrix. Counter-SELEX or “negative selection” may reduce the possibility of evolving nonspecific binders or improve the odds of recovering aptamers specific for one target over another [85].

1.7.2 Washing and Elution

With the target successfully immobilised, the library is loaded onto the matrix in a loading buffer; a series of washes and elutions must be carried out to eliminate weakly

binding species. The stringency of the washing procedure is critical to the efficiency of the SELEX protocol; factors such as pH and salt concentration must be adjusted carefully to ensure the selection pressure is neither too weak nor too firm. The stringency of the washing conditions is typically increased over successive elutions and several rounds of SELEX; this typically involves an increase in the concentration of the monovalent salt [86]. The washing buffer usually consists of a salt such as NaCl, as its ability to interfere with electrostatic interactions (the primary binding method of aptamers) allows for the separation of unbound species [87].

Following the removal of unbound species via washing, the aptamers must now be separated from the target to be amplified; some common methods for elution are heating, adding salt to destabilise electrostatic interactions and the use of strong chemical denaturants such as urea [88][45][89].

Since the development of conventional SELEX in the 1990s, its time-consuming nature has led to the development of alternative approaches to the selection stage. These range from using magnetic approaches to separate low-affinity species to using cells as selectors [64]. For peptide/protein targets, immobilized metal affinity chromatography (IMAC) resins are frequently used [90]. In magnetic bead SELEX, the target protein is immobilised via chelation to magnetic beads; upon exposure to a library, various aptamer target complexes form; this allows for the separation of the beads (along with high-affinity aptamers) from non-binding species by magnetic separation [91]. In addition, Stoltenburg developed ‘Flu-mag SELEX’, during which the dissociation constants of the species can be determined by fluorescence [88]. Many new methods have been used to target proteins; a comparison of alternative selection methods utilised for protein SELEX is shown in Table 1.3.

Table 1.3: Comparison of Protein SELEX methods[64]

Selex method	Advantages	Limitations	Citation
Nitrocellulose filter binding	The relative ease of performance Aptamer isolation for multiple targets in parallel	Background binding A large number of rounds	[92]
Capillary Electrophoresis SELEX	Rapid partitioning of bound aptamers from unbound ones A small amount of analyte is required	Capillary electrophoresis instrument required Restricted to targets that cause a shift in aptamer electrophoretic mobility pattern	[93]
Magnetic Bead SELEX	Applicable to most targets Easy fine-tuning of selection pressure	Target immobilization is required.	[94]
Microcolumn Bead SELEX	Minimizing the amount of target and aptamer	Target immobilization required Need for automated devices.	[95]

1.8 Monitoring Selection Pressure

While SELEX is theoretically straightforward, the background binding level due to the amplification of unspecific binding sequences often results in unsuccessful SELEX protocols. Therefore, SELEX's progress can be monitored using simple yet reliable methods [96]. In Flu-mag SELEX, following the initial selection, the proto-aptamers are labelled using fluorescein-modified primers; this allows for the binding of aptamers to the beads to be determined by fluorescence emission. The enrichment progress can then be monitored easily through SELEX rounds, and the selection pressure can be adjusted accordingly. This offers a safer alternative to radiolabelling, available to most

molecular biology laboratories [88]. During selection, environmental factors such as pH, salt concentration, and temperature are highly influential, so the conditions for selection must mimic those of the in vivo environment to which the aptamer will be subjected [97]. For example, aptamers are against protein factors important to tumour metabolism, and it is known that tumour tissue tends to have a lower pH than healthy tissue [98].

1.9 PCR Amplification

The polymerase chain reaction (PCR) amplifies libraries for the following selection round. The PCR reaction has several 'key components': the DNA polymerase, two oligonucleotide primers, dNTPs, magnesium ions and reaction buffer [99]. The primers act as initiators binding to their corresponding sites on the template strand, facilitating extension via the DNA polymerase. Following amplification to generate dsDNA, further steps must be performed to re-generate the ssDNA [100].

The use of conventional PCR to amplify ssDNA libraries has been shown to produce artificial by-products that reduce the efficiency of the process. Studies have broken down by-products into two specific types: ladder and non-ladder [74]. The ladder-type by-product is most concerning as PCR can amplify it; ever-longer by-products can be formed and will eventually dominate the library composition. Therefore, the ladder-type by-product has the most severe implications for SELEX [74]. PCR cannot amplify the non-ladder type; thus, it will not be exponentially amplified [74].

While the selection step of SELEX is the primary driver of evolution in the SELEX process, selection pressure towards genotypes more easily amplified by the DNA polymerase enzyme is a factor to consider. PCR tends to show a bias towards shorter and less stable sequences, as well as those for which the polymerase enzyme has a preference [101]. Conventional PCR typically results in a loss of 50% of the library diversity within each round, which is a critical factor to the detriment of many SELEX protocols. The optimisation of PCR parameters is crucial to the development of new aptamers [102]. Paradoxically, while SELEX aims to isolate aptamers with diverse secondary structures that bind to the target, complex secondary structures such as stem-loops and G tetrads create challenges for the PCR stage, preventing efficient amplification due to their secondary structures [100].

Strategies to overcome these PCR issues address the parasitic buildup of byproducts; this is typically done by breaking the PCR reaction down into smaller compartments, typically done via digital droplet or emulsion PCR. In emulsion PCR, the PCR mixture is dispersed within an oil-based emulsion as droplets, whereas in digital droplets, the mixture is divided into tiny droplets within a specialised thermal cycler [101,103]. This compartmentalisation effectively limits the growth of byproducts by starving them of more material to accumulate.

1.9.1 Production of ssDNA following PCR.

All the PCR methods discussed amplify ssDNA to generate dsDNA; therefore, the ssDNA must be re-generated from the dsDNA prior to further rounds of SELEX [104]. A wide range of alternative methods have been employed for the purification of ssDNA from dsDNA in SELEX.

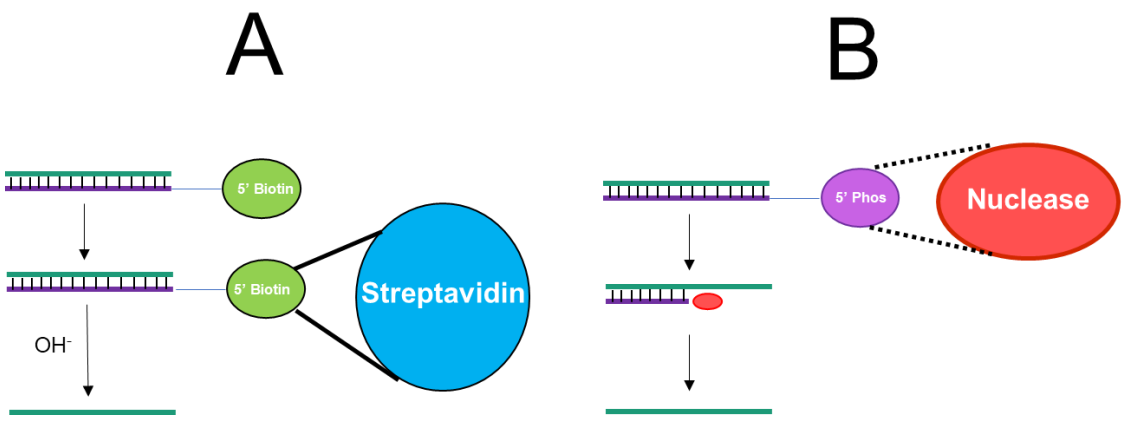


Figure 1.10: Regeneration of ssDNA following PCR: ***A)*** *Separation via streptavidin-biotin affinity chromatography, 5' biotinylated dsDNA binds to streptavidin and alkaline denaturation releases non-biotinylated strand.* ***B)*** *separation via exonuclease digestion, 5' phosphate on complimentary strand makes it susceptible to exonuclease digestion.*

Streptavidin-biotin affinity chromatography involves the use of a 5' biotin-labelled PCR reverse primer. This results in the biotin labelling of the complementary strand, and the resulting dsDNA is captured by binding to streptavidin-coated beads. The DNA strands are then separated by alkaline denaturation (NaOH), enabling elution of the ssDNA, whilst the biotinylated strand remains attached to the streptavidin-coated beads (see Figure 1.10 A).[105] A primary concern with this method is the potential for streptavidin to remain in the sample matrix after the separation; this would act as a secondary selection target in later rounds of SELEX. In addition, PCR by-products

containing biotin will also be captured, and ssDNA will be regenerated, facilitating the parasitic growth of PCR byproducts from round to round [106].

Introducing a 5’ phosphate group on the reverse primer in the PCR reaction generates a 5’ phosphate labelled complementary strand, susceptible to selective degradation by exonucleases (see Figure 1.10B). This selectivity allows for complete digestion of the complementary strand with minimal template loss. This approach offers advantages over the streptavidin/biotin method as no potential selection targets remain in the matrix [107,108]. Asymmetric PCR involves adjusting the primer ratios to amplify the desired strand many times more than the complementary strand. While asymmetric PCR offers an excellent solution to the double-stranded product problem, there are two significant drawbacks: asymmetric PCR offers an efficiency of 60-70% compared to regular PCR, and the reaction still generates double-stranded products, meaning an adequate purification method is still required [109,110].

1.9.2 Sequencing of DNA libraries

DNA sequencing approaches are utilised primarily at two points within a SELEX protocol, at the start to validate the diversity of the aptamer library and in the later rounds to examine the sequences of the aptamers being produced. Historically, sequencing was predominately performed using Sanger-based methods [111]. In the late 2000s, the advancement of high throughput sequencing technologies gave birth to next-generation sequencing (NGS) methods, which enabled faster and deeper insight into the aptamer library as SELEX progressed round to round [112,113]. In addition, the NGS methods have offered further potential insight into the flaws of SELEX protocols [114].

These techniques include the Illumina, 454 Roche, and IonTorrent sequencing platforms, the Illumina being the most widespread [115]. The basic steps of Illumina Sequencing involve the addition of adapters (known sequences) to the matrix and their ligation; this allows the sequences to attach to a synthetic manifold from where sequencing can take place. Sequences bound to slides and then amplified by bridge amplification. Fluorescently tagged nucleotides are added to the slide, which binds to the various amplified gene clusters. The sequence of fluorescent nucleotides will depend on the gene sequences, so their emission wavelengths can be used to determine the base sequence [116].

1.10 Measuring Aptamer Performance

The primary measure of an aptamer's ability to 'stick' to a target is its binding affinity, measured in terms of its equilibrium dissociation constant Kd, defined according to Equation 1.

$$Kd = \frac{[A][B]}{[AB]}$$

Equation 1: Dissociation Constant

Where:

K_d is the dissociation constant.

[A] is the concentration of the unbound aptamer.

[B] is the concentration of the unbound target.

[AB] is the concentration of the aptamer target complex.

The affinity of the aptamer for its target is inversely proportional to the magnitude of its dissociation constant as K_d is effectively the ratio of aptamers bound to unbound ones and thermodynamically related to the Gibbs free energy [117]. Gibbs free energy and its relation to K_d is shown in Equation 2.

$$\Delta G = -RT\ln(1/K_d) = \Delta H - (T\Delta S)$$

Equation 2: Relation of Gibbs free energy to the dissociation constant

Where: K_d is the equilibrium dissociation constant; R is the ideal gas constant; ΔG is the change in free energy; ΔH is the change in enthalpy; ΔS is the change in entropy; T is the temperature.

Equation 2 shows that the dissociation constant is subject to the influence of several environmental variables, notably T, temperature, ΔH enthalpy and ΔS entropy. In terms of in vivo performance, this means the dissociation constant of an aptamer is influenced by the nature of the tissue it is in; salt concentration, temperature and pH will all play a role. The second key factor of aptamer performance specificity is

effectively the binding affinity of the aptamer to its target versus background species and other proteins within the same family. For example, an aptamer where K_d to the target is << K_d to proteins similar in nature would be specific.

1.11 Methods for measuring aptamer protein binding affinity

Many methods are available for measuring aptamer binding affinity to protein targets; these can be split into label-based and label-free techniques. Two popular label-free methods are isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR) [118]. ITC involves measuring the thermodynamic properties of the interaction between the aptamer and target through titrations where one species (usually the protein) is in excess through molar equilibria; the Gibbs free energy is used to calculate K_d [119].

SPR involves the immobilisation of one of the two species involved in an interaction on a surface; a solution containing the other species (analyte) is passed over the immobilised species (the target); as binding occurs, the refractive index of the surface changes, monitoring this throughout the interaction allows binding strength and specificity to be determined [120]. Label-based techniques are often partition coefficient-based and can be done simply using fluorescently tagged aptamers and an immobilised target or in a more complicated fashion such as shift/separation-based assays, including gel shift assays capillary electrophoresis and Microscale thermophoresis [118]. Each technique has advantages and pitfalls, but ultimately, selecting a technique that suits the Kd range of the selected aptamers is essential. A summary of some common affinity screening techniques is presented in Table 1.4.

Table 1.4: Summary of aptamer protein affinity assays:

Method	Lower Kd limit	Principle	Advantages	Citations
Surface Plasmon resonance	1 pM	Refractive index changes	Highly sensitive, requiring low sample volumes	[121,122]
Isothermal Titration Calorimetry	1 nM	Heat of binding	Highly sensitive, provides stoichiometric information	[123–125]
Microscale Thermophoresis	~1 pM – 1 nM	Thermophoretic mobility	Highly sensitive, low sample volumes	[126,127]
Enzyme-linked immunosorbent Assay	100 pM	Binding-mediated colour change	Low-cost reagents, high throughput screening	[128,129]
Capillary Electrophoresis	1-10 nM	Size-dependent electrophoretic shift	Resolution of different complexes, low sample volumes	[124,130]

1.12 Aptamer anti-cancer Therapeutics

Aptamers against human proteins for use as anti-cancer therapeutics are the focus of this project. However, their uses in a non-medicinal and industrial capacity are broad, including the detection of cocaine in blood samples, the detection of norovirus in food products and their use in stationary phases for affinity-based purification chromatography [131–133].

Aptamers for cancer therapeutics can generally be divided into four categories [134]. Cancer cell adhesion and invasion inhibition are achieved by blocking involved molecules. E-selectin and P-selectin, mediated adhesion of tumour cells to vascular endothelium is a pivotal step of metastasis formation; SDA, a DNA aptamer to E/P-selectin, showed a dissociation constant of 100 nM inhibiting the interaction between selectin and its ligands [135].

Secondly, modulation of the immune system could be deactivated by tumour cells. Cytotoxic T cell antigen-4 (CTLA-4) is associated with immune system suppression, RNA aptamers to CTLA-4 inhibited function in vitro and enhanced tumour immunity in mice [136].

Thirdly, blockage of signalling pathways by inhibiting kinases, phosphatases, carboxypeptidases, etc., to stop downstream activation and signalling for tumour growth. The epidermal growth factor receptor (EGFR) has been implicated in the development of a wide range of human cancers; an RNA aptamer developed by Esposito et al. binds to EGFR with a binding constant of 10 nM and inhibits tumour growth in a mouse xenograft model of human non-small-cell lung cancer (NSCLC) [137].

Fourthly, binding to target proteins is closely connected with tumour development. For example, Aptamers targeting b-catenin can inhibit multiple oncogenic functions of target proteins in HCT116 colon cancer cells [138].

1.13 Aptamer cancer drugs in the clinic

Nucleolin, or C23, is one of the most abundant non-ribosomal phosphoproteins of the nucleolus. However, Nucleolin is highly expressed intracellularly in several cancers and on the cell surface [139]. AS1411, formerly AGRO100, is an aptamer to nucleolin, discovered by chance by Bates et al [72]. AS1411 is a 26-mer unmodified guanosine-rich oligonucleotide, which induces growth inhibition in vitro. It has shown activity against human tumour xenografts in vivo, and its progression to clinical trials has led to it being hailed as the most advanced aptamer in cancer therapy, both in vitro and in vivo [140]. However, AS1411 was later discontinued from phase II trials for renal cell carcinoma (RCC) due to limited activity; further evaluation against this type of cancer has been discontinued [141]. Despite these setbacks, the nuclease-resistant nature of AS1411's structure and its ability to enter cancer tissue means that research focusing on optimizing the structure of AS1411 remains popular [142]. Aside from AS1411, as of 2023, three anti-cancer aptamers were registered in clinical trials; these are typically administered in combination with another drug but are not conjugated with it. The progression of aptamers in clinical trials up to 2023 is summarised in Table 1.5.

Table 1.5: Anti-cancer aptamers progression in clinical trials up to 2023.[143]

Aptamer	Number of trials	Target	Latest Phase
AS1411	4	Nucleolin	2
EYE0001	1	VEGF	1
NOX-A12	5	CXCL12	2
NOX-H94	1	Hepcidin	2

Aside from direct anti-cancer action, one of the most promising outlooks for aptamers in cancer treatment is aptamer-conjugates; these broadly come in two forms: aptamer conjugation to an anti-cancer drug with the hope of increasing a drug specificity or aptamer conjugation to radiosensitisers to increase the efficiency of radiotherapy [144]. Aptamer therapeutics are typically administered in vivo via intravenous, subcutaneous or intravitreal injection; when in the bloodstream, they should migrate naturally through the body until they reach and bind to the desired target [145]. This is advantageous when targeting proteins that are naturally overexpressed in cancers compared to healthy tissue, as it means the aptamers will have a passive selectivity for tumour tissue due to the higher target concentration.

1.14 Challenges in the Aptamer Field

1.14.1 The Thrombin problem

The nature of the thrombin problem is that hundreds of researchers have focused on studying thrombin aptamers or developing assay methods instead of more clinically relevant targets [146]. Aptamers have suffered from their success since it is easy to manipulate DNA to develop interesting detection methods. Aptamers such as the thrombin aptamer are available cheaply as they can be manufactured on kilogram scales. It is also much faster to generate publishable research if you have modified a thrombin aptamer to suit the needs of an experiment than if you have built one from the ground up via SELEX [146].

1.14.2 Clinical Trial issues

A large phase clinical trial that held great expectations using an anticoagulant aptamer (RADAR phase 2b clinical trial) showed toxicity in a small fraction of patients. Unfortunately, this outcome likely discouraged the launch of other clinical trials using this technology [97]. It is important to highlight that the toxicity observed in this clinical trial was associated with the pre-existence of antibodies against PEG (polyethene glycol), a modification included in the anti-coagulant aptamer to enhance its half-life in serum [97]. While the fact that it was the structural modification that caused the allergic response may allow this trial data to be dismissed, it is important to remember that one of the key advantages of aptamers is the ease by which their structure is modified, indeed in vivo studies of aptamers suggest that modification is necessary to overcome their shortcomings, what this trial suggests is that modification of aptamers potentially makes them susceptible to detection by the immune system reducing the numbers of possible modifications that can be made thus undermining one of the key advantages of aptamers in the first place.

1.14.3 In Vivo Targets: The Medicinal Chemistry Problem

The inhibition route of cancer treatment is simple: a drug must find and bind to the right target with sufficient affinity to inhibit its ability to aid cancer proliferation. However, there is a lack of drugs that can marry affinity with specificity. Since the late 19th century, mankind's ability to utilise organic chemistry for molecular medicine has improved drastically. However, scientists are left with a significant problem: how do they design a drug or therapy that selectively binds to specific targets in an aqueous solution or even better cellular targets such as membrane surface proteins [49]? Cis-platin was a revolutionary drug when approved in the 1970s; it is an effective killer of cancer cells but also causes a wide range of side effects, including neural toxicity and bone marrow suppression [147]. Despite these effects, cis-platin chemotherapy is a routine cancer treatment. The inability of chemists to marry efficacy with specificity is in stark contrast to nature, which delivers molecules that bind to targets with extraordinary affinity and specificity [49].

1.15 Development of analytical methods for the analysis of mRNA therapeutics and dsRNA biocontrols

The regulatory requirements facing clinically approved vaccines are consistent post-marketing surveillance, particularly quality control in production [148]. Like any other therapeutic, mRNA production is governed by good manufacturing practice (GMP), which, while varying somewhat between regulatory authorities (MHRA or FDA), lays out strict rules for quality control; GMP production of large quantities of mRNA remains challenging [15].

In 2022, the United States Pharmacopeia released draft guidelines for mRNA vaccines, focusing on critical quality attributes, including mRNA identity, purity, content, and safety [149]. Many issues arise during the IVT production of mRNA and its by-products when meeting these attributes. In addition to the leftover DNA template and reaction components, IVT by-products include truncated mRNA and small quantities of immunostimulatory dsRNA; for therapeutic mRNA to be released, these by-products must be eliminated [150].

Environmental protection agencies strictly govern the use of pesticides to ensure quality and minimise the risk of environmental damage and off-target effects on other organisms; as such, the quality of dsRNA pesticides is held to a stringent standard [36]. Particularly in the case of dsRNA expressed in bacterial systems, care must be taken to remove contaminants and ensure that GMO material is not released into the environment [151]. For any quality control/assurance process, size and purity are required to ensure that the dsRNA produced is the correct sequence.

Additionally. products can be sized and identified; this is commonly done by gel electrophoresis, which, while useful at a glance, lacks actual sized-based resolution for larger fragments. Additionally, dyes employed in gel electrophoresis are often more sensitive to double-stranded over single-stranded species, leading to issues with sensitivity when detecting single-stranded products. Nwokeoji et al. previously demonstrated that dsRNA could be separated from other RNA products left over from RNA extraction from *E. coli* via HPLC [152].

Functional RNAs in the form of vaccines and biocontrols offer medicinal and industrial solutions to global problems, and their production will only become more critical as we move forward. The strict regulation of these products requires robust analytical

methodologies to ensure consistent, high-quality production from batch to batch. A requirement of these methods must be integrity and impurities analysis, which largely depends on the size-based separation of a fragment from associated byproducts or synthetic leftovers.

1.16 Gel electrophoresis

Gel electrophoresis was one of the earliest methods developed for the analysis of nucleic acids and has been widely employed in molecular biology for the separation and analysis of DNA and RNA [153]. The nucleic acid fragment migration rate through the gel is determined by its size as the negatively charged phosphates are attracted to the positive electrode. Gel electrophoresis has demonstrated accurate sizing of RNA in the 200 to 6000 nt (nucleotides) in length using a linear relationship between the logarithm of relative mobility ($cm^2V^{-1}h$) and the square root of molecular weight, but this method is crude. It cannot provide complete purity and integrity analysis [154].

However, gel electrophoresis is somewhat limited, offering only low-resolution separations of DNA/RNA. For the analysis of RNA, denaturing conditions often need to be employed, including the use of hazardous chemical denaturants. Like anion exchange chromatography, Curved DNA fragments have been shown to possess reduced mobility during gel electrophoresis [155,156]. As migration depends on end-to-end length, it results in problems in accurately sizing nucleic acids using this approach [157].

1.17 Capillary Electrophoresis

Capillary electrophoresis (CE) has been demonstrated to be one of the most efficient separation methods for nucleic acids [158]. Being well suited to the analysis of charged solutes, it lends itself well to the analysis of highly soluble, multiply-charged nucleic acids [159]. CE separates fragments based on their migration rates through a capillary subjected to an electric field; in the case of nucleic acids, their size is directly proportional to the number of negative charges; therefore, the migration rate should theoretically be size-dependent [160]. Two CE modes are generally used to analyse nucleic acids: capillary gel electrophoresis (CGE) and capillary zone electrophoresis (CZE). The resolving power of CGE is superior to that of CZE and, as such, is more suited to a size-based assay [160]. A concern with this method is that it is trickier to couple CE with MS, which may limit its compatibility with techniques such as RNA

mass mapping, which is of increasing interest in the analysis of mRNA vaccines [160,161].

1.18 HPLC

The principles of HPLC are the same as those of paper chromatography, but the separation is much more efficient. A pump administers a constant mobile phase flow at pressure through a column containing a packed material of small particles (the stationary phase). The sample is introduced into HPLC via an injection valve just after the pump; as the mixture of analytes passes through the column, the affinity of each analyte for the stationary phase vs its affinity (solubility) for the mobile phase determines its retention time and thus separates it from other analytes [162]. Following elution from the column, analytes are detected using various methods. Ultraviolet (UV) detection is the most common for organic molecules, and this can be coupled with mass spectrometry to perform a detailed mass analysis of analytes [162].

The use of HPLC for nucleic acid analysis has emerged as a powerful alternative to conventional methods such as gel electrophoresis. HPLC's superior capabilities become apparent when identifying, separating, and quantifying nucleic acids. These advantages include sample automation, fraction collection, and data reporting functions, which promote higher sample analysis throughput and recovery rates [163]. Various HPLC approaches have been employed for the analysis and purification of nucleic acids, with ion-pair reverse-phase high-performance liquid chromatography (IP RP HPLC) and anion exchange chromatography (AEX) being widely used [164].

1.18.1 Anion-exchange chromatography

The difference between the isoelectric point of mRNA (2-2.5) and that of impurities generated in the IVT process and leftover DNA allows for using AEX chromatography to separate nucleic acids [154]. The separation mechanism occurs via the interaction of the positively charged stationary phase with negatively charged analytes; a gradient of increasing ionic strength causes competition for binding sites [155]. The strength of the interaction of mRNA molecules with the AEX resin is proportional to their length, specifically the number of phosphates in their backbone, so of course, mRNA’s 4000 bases in length will have a much stronger interaction than shorter dsRNA’s or leftover NTP’s from the IVT reaction, in real terms this means shorter RNA’s and NTP’s tend

to give sharp peaks. At the same time, significant band broadening occurs in larger molecules; this is disadvantageous for mRNA analysis in particular [156].

The separation principle of anion-exchange chromatography implies that elution order is determined only by size. However, anion-exchange separation of dsDNA fragments on various stationary phases has suffered from inversions in retention time as a function of chain length, preventing its use for accurate-size fragment identification [168]. Curved DNA fragments, including DNA fragments with poly-A runs, curved fragments exhibit high AT-rich regions and have been shown to over retain during anion exchange HPLC through the preferential attachment of the curved DNA to the ionic groups of the column [169].

1.18.2 Ion Pair Reversed Phase HPLC

IP RP HPLC has been widely employed to analyse nucleic acids [170,171]. The development of non-porous alkylated PS DVB by Bonn et al. and commercialised via Transgenomic, USA, allowed for high-resolution small and large nucleic acid separations in as little as 10 minutes [172,173]. They demonstrated a significant increase in resolution compared to conventional porous stationary phases.

The IP RP HPLC mechanism involves using an ion pair reagent, a salt, to mediate between the hydrophobic stationary phase and the phosphate groups on the backbone of the nucleic acid (polar molecule). This allows both single and double strands to interact similarly with the column. The ion pair reagent, typically an alkylammonium acetate salt such as triethylammonium acetate (TEAA), contains a positively charged ammonium ion which interacts electrostatically with the phosphate backbone of nucleic acids and an alkyl chain which binds to the stationary phase (Figure 1.11) [172].

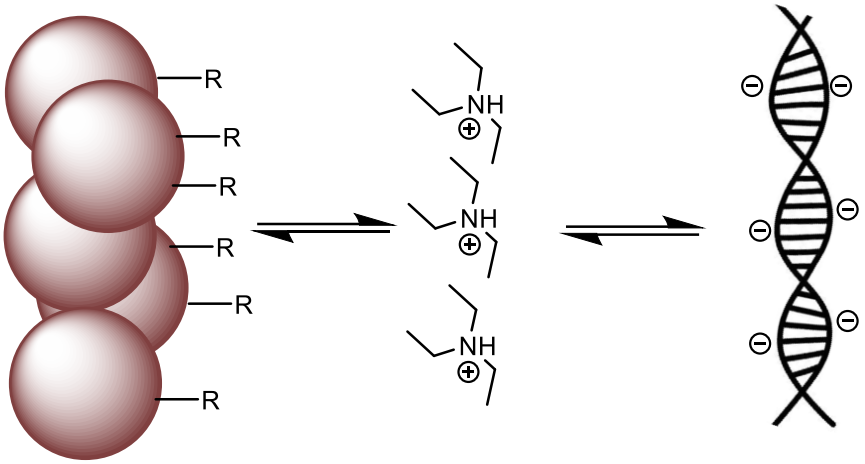


Figure 1.11: The ion-pairing mechanism in IP RP HPLC: *The amphiphilic nature of thetriethylammonium ion allows it to facilitate the interaction of negatively charged phosphate groups on nucleic acids with the hydrophobic stationary phase.*

Introducing an organic modifier such as acetonitrile in the secondary mobile or mobile phase B disrupts this interaction. An increase in the concentration of the organic results in the elution of oligonucleotides of increasing length to elute from the column. The primary ion-pairing reagent for nucleic acids is alkylamines, typically a tertiary alkylamine with a counter ion forming an ammonium salt. Many alkylamines, including triethyl, hexyl, dibutyl, and tributylamine, are suitable for this mechanism. All demonstrate different resolutions and retention strengths for nucleic acids of the same length, which can be attributed to chain alkyl length and water solubility [174].

One crucial advantage of IP RP chromatography over the other HPLC techniques is the ease with which this technique can be combined with mass spectrometry (MS) detection. IP RP chromatography with electrospray ionisation mass spectrometry (ESI-MS) has been used extensively to study single and double-stranded oligonucleotides and a wide range of ion-pairing reagents have been used for the interfacing of IP RP HPLC to ESI MS [171,175,176]. In recent studies using IP RP HPLC to analyse and purify RNA, results have indicated its capability in a wide range of RNA applications, from the purification of synthetic oligoribonucleotides to the analysis of more complex RNAs [177–181].

Dependence of chromatographic retention on fragment size and base sequence

Various biochemical and molecular biological techniques and the manufacturing of important nucleic acid therapeutics yield complex mixtures of nucleic acids, which subsequently must be separated and identified. Separation of nucleic acid mixtures according to size is critical in manufacturing quality control as it allows for the identification of the product and any by-products left over from production [168]. Other than size, properties such as shape, conformation, curvature, melting behaviour, and sequence variations are relevant to the production and purification process [168].

Predicting retention and size of nucleic acids using IP RP HPLC

A key advantage of HPLC is its ability to separate species based on slight differences in size and physical properties. The ability to accurately size nucleic acid and separate nucleic acids based on their length has previously been demonstrated in several studies where models have been established describing the relationship between size and retention time in ion-pairing systems. Initial models of the relationship between nucleic acid size and retention time were purely empirical, with statistical relationships between various independent variables such as organic content, alkyl chain length and concentration of ion pair reagent used to predict retention based on previous data alone.[182] In 1995, Huber described retention capacity factor (s), which is calculated as s k = b log bp + a where k is a capacity factor, bp is the length of dsDNA fragments (in base pairs), and a and b are constants calculated via least-squares linear regression analysis. It was found that there is an almost linear correlation between the capacity factor and logarithm of base pair up to 400-500 bp TEAA on the non-porous Poly(styrene-divinylbenzene) (PSDVB)-C_{18} column [183].

In 1998, Huber simplified this model, demonstrating that this separation is size-dependent to such a high degree by plotting the retention times of 57 DNA fragments in TEAA against their retention times; the high separation efficiency of this system did show signs of the influence of sequence composition and structure on retention; however, it was found that there is a linear relationship between retention time and the logarithm of length, in base pairs up to 500 bp [168]. The plot from this study is shown in Figure 1.12

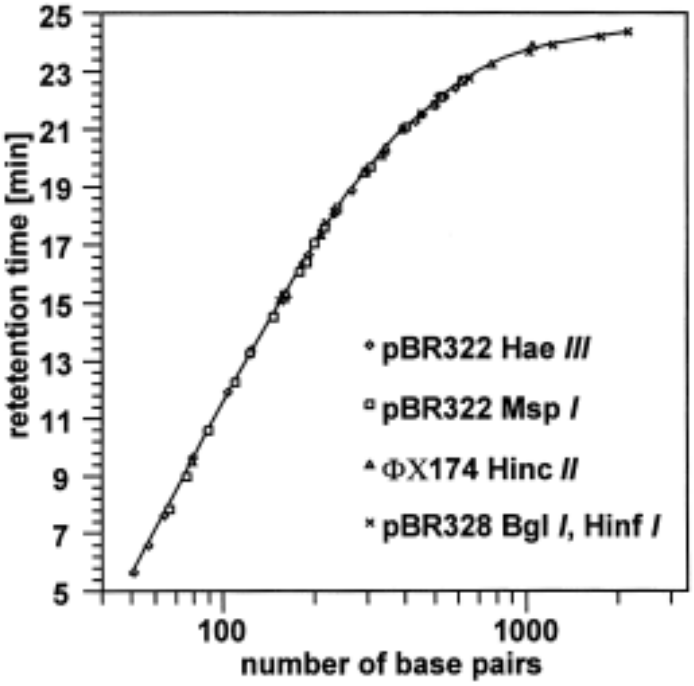


Figure 1.12: The linear relationship between retention time and the logarithm of length in base pairs: *Taken from Huber et al. 1998, Column, PS–DVB-C18 (2.1 mm, 5034.6-mm I.D.); mobile phase, (A) 0.1 M TEAA, pH 7.0, 8% acetonitrile, (B) 0.1 M TEAA, 20% acetonitrile, pH 7.0; linear-gradient, 0–100% B in 30 min; flow-rate, 1 ml/min; temperature, 50 ºC; sample, 57 restriction fragments ranging from 51 to 2176 bp in length.[168]*

In 2002, Gilar studied the influence of organic modifiers, establishing the relationship between the log of the base length of a fragment and the organic concentration required to elute it [184]. These capacity factor models provide accurate predictions for retention, but the regression fitting parameters are wholly dependent on the choice of mobile phase and must be recalculated for any change in the independent variable. In 2013, Zhang et al. went a step further, developing a model that included the impact of both column properties and experimental variables; they could independently calculate dsDNA size based on their retention times, thus improving and accelerating separations [185].

Understanding the interaction of analytes with a stationary phase is critical for all techniques discussed in this chapter, but as RP-IP HPLC is the focus of this body of work, it will only be discussed in this context. When researchers first tried to model the retention of analytes IP RP HPLC, the models fell into stoichiometric and electrostatic/nonstoichiometric [186]. Stoichiometric models focus on forming ion pairs

within the matrix between the analyte and the ion-pairing reagent, in this case, the nucleic acid phosphate and the alkylamine, thus enabling the retention of the column [187].

Electrostatic models focus on the binding of the ion pair reagent to the stationary phase and the subsequent formation of an electric double layer when the analyte interacts with the now-charged surface [188]. What became apparent is that both these scenarios seem to co-occur and that the failure of each model class to account for the other led to inaccurate retention models.

Cecchi et al. combine the two scenarios into one model using a modified Gibbs free energy equation and derive an equilibrium constant (K) for four scenarios.

- Interaction of the ligand binding site with the analyte
- Interaction of the ligand binding site with the ion pair reagent
- Interaction of the ion pair reagent with the analyte
- Simultaneous interaction of the ligand binding site with the ion pair reagent and the analyte:

Therefore, the analyte's capacity factor is a summary of all these parts [189].

1.19 Summary

Nucleic acids have emerged as effective platforms for solutions to industrial and therapeutic problems, but each comes with challenges. Rapid quality control methods are needed for mRNA vaccines and dsRNA biocontrols to meet the demand for batch production. Due to their similar chemistries, RP IP HPLC is a likely candidate to satisfy quality control requirements for both platforms, offering a size-based resolution that allows for the assessment of identity, integrity, and purity. There is already strong precedent that reverse phase ion pair chromatography can be used to size DNA in the 100-1000 bp range accurately. Can this mode of analysis be used to size functional RNAs in the form of dsRNA pesticides in the 100-500 bp range and mRNA vaccines in the 1000-5000 nt range?

Aptamers are incredibly versatile molecules and have been shown to bind to many targets with high affinity. However, a significant challenge is the lack of a universal, well-developed SELEX methodology. While the field has advanced much in the last three decades, its multi-disciplinary nature often means that each target is approached

differently without well-established protocols. The development of SELEX methods to protein targets will involve utilising the innovations of the last three decades where possible. There is ample evidence regarding protein SELEX that a well-designed SELEX protocol can generate aptamers that show in vitro and in vivo binding. To do this, the methods employed must take full advantage of recent innovations in SELEX, such as NGS.

1.20 Thesis Aims and Objectives

The experimental section of this thesis can be divided into two sections with discrete aims and objectives. The first section aimed to develop an in-house SELEX protocol to generate high-affinity aptamers to a protein target Melanoma differentiation-associated gene-9 (MDA-9)/Syntenin-1. Chapter 3 focuses on the development and optimisation of a SELEX protocol for the generation of ssDNA aptamers. Following optimisation of the ssDNA library design and PCR amplification conditions, novel IP RP HPLC methods were developed to purify ssDNA.

Chapter 4 utilised the developed IP RP HPLC method to generate ssDNA as part of a SELEX protocol to generate ssDNA aptamers against a target protein. Following development and changes to the SELEX protocol, SELEX was successfully used with next-generation sequencing methods to identify potential ssDNA aptamers, which were screened for affinity to the target using ITC.

The project's second phase involved developing and applying IP RP HPLC to analyse dsRNA biocontrols and mRNA therapeutics. Chapter 5 focused on two crucial emerging classes of nucleic acids: dsRNA biocontrols and mRNA therapeutics. Developing and applying novel analytical methods for characterising and analysing dsRNA biocontrols and mRNA therapeutics is a significant challenge. IP RP HPLC methods using various ion-pair reagents were assessed for the accurate sizing and analysis of dsRNA biocontrols and mRNA therapeutics. HPLC separations were performed on monolithic and macro-porous PS DVB stationary phases.

2 Chapter 2 Materials and Methods

2.1 Chemical Reagents

Unless specified, all chemical reagents and solvents were purchased from Sigma-Aldrich, Thermo Fisher, Pharmacia, Promega, Qiagen, Bio-Rad, Invitrogen and New England Biolabs.

2.2 PCR

PCR reactions were performed in a final volume of 50 µl using a Techne Prime thermal cycler. Standard PCR conditions used to amplify the DNA are shown in Tables 2.1 and 2.2. Variations of the standard PCR conditions are described in the relevant chapters. Taq polymerase was sourced from New England Biolabs.

Table 2.1: PCR conditions for initial PCR amplification.

Component	**Desired Concentration**	**Volume Required µl (25 µl reaction)**
dNTP mix	0.2 mM	0.5
PCR buffer	1x	5
Forward Primer	250 nM	0.625
Reverse Primer	250 nM	0.625
$MgCl_2$	1.5 mM	2
Taq polymerase	0.1 u/µl	0.5
Template DNA	1 ng/ µl	1
H_2O	N/a	14.75

Table 2.2: PCR cycles for initial amplification

Phase	**Temperature °C**	**Time (s)**
Denaturation	95	30
Annealing	47	30
Extension	72	30
Initial Denaturation	95	60
Final Extension	72	300

2.3 DNA purification

Ethanol precipitation

Following PCR, dsDNA was resuspended in a final volume of 300 µl HPLC grade water, and 300 µl of phenol-chloroform solution (Sigma Aldrich) was added. The solution was vortexed at 13000 rpm for 5 minutes, and the aqueous fraction was removed and mixed with two volumes of ice-cold ethanol, 1 µl of glycogen was added, and the solution was cooled at -20 °C for 4 hours. The solution was centrifuged at 13,000 rpm for 5 minutes, and the pellet was washed twice with 300 µl of a 70% ethanol solution before re-suspension in H_2O. For the precipitation of ssDNA from the HPLC, the above procedure was performed without adding phenol-chloroform.

Silica magnetic bead clean-up

DNA was mixed with 2X volumes of 5 M guanidinium chloride in 90% ethanol (v/v) and 50 µl of silica magnetic beads (Dynabeads, Thermo Scientific) for 15 minutes on a gentle rocker; beads were removed from the solution via magnetic separation and washed twice with 80% ethanol (v/v). The beads were then dried in a 37 °C incubator, and the DNA was eluted in HPLC-grade water.

Freeze drying

The eluted fractions were frozen at -80°C and freeze-dried overnight at -120 °C. Following lyophilisation, the sample was resuspended in 500 µl of HPLC-grade water. DNA fractions were dried using vacuum centrifugation at 60 °C until dry and resuspended in 500 µl of HPLC grade water.

2.4 ssDNA re-generation

2.4.1 Exonuclease digestion

One µg of ds PCR product was made up to 100 µl with H_2O, 10 µl Lambda exonuclease reaction buffer (10x) and 2 µl (5 units) of Lambda exonuclease (New England Biolabs). Digestion was carried out at 37 ºC in a Techne thermocycler for a

time defined by the experiment. Following digestion, heat activation occurred at 75 ºC for 10 minutes.

2.4.2 Purification of ssDNA using streptavidin magnetic beads

Table 2.3: Composition of purification buffers

Buffer	Composition
Binding/Wash buffer	10 mM Tris-HCl (pH 7.5), 1 mM EDTA 2 M NaCl
Elution Buffer	0.15 M NaOH

The desired volume of magnetic beads was added to a 1.5 ml Eppendorf tube and washed three times with 2X binding buffer, beads were then resuspended in 2X binding buffer, and an equal volume of the biotinylated PCR product beads were incubated for 15 min at room temperature using gentle rotation. Beads were separated from the supernatant with a magnet; the supernatant was discarded. Beads were washed once with 1X binding/wash buffer. Beads were resuspended in 50 µl of 0.15 M NaOH and incubated at room temperature for 10 minutes with gentle shaking. The magnetic particles were captured using a magnet, and the supernatant was removed and neutralised by adding 5 µl 1.5 M HCl; the elution was repeated twice.

2.5 HPLC

Samples were analysed by IP-RP-HPLC on a Vanquish HPLC (ThermoFisher), Agilent 1100 HPLC or a U3000 (ThermoFisher), consisting of a degasser, pump, autosampler, column oven, and UV detector. The chromatography software used was Chromeleon versions 6.8 and 7.3 Dionex Corporation.

2.5.1 HPLC columns

Table 2.4: HPLC columns for oligonucleotide analysis

Mode of Analysis	Manufacturer	Dimensions	Stationary phase	Particle size µm
Weak IP (HPLC flow)	Thermo	100 mm x 2.1 cm	DNApac	8
Weak IP (Capillary flow)	Thermo	0.1 mm x 5.0 cm	PSDVB (Polystyrene Divinylbenzene)	Monolith
Strong IP (Capillary flow	Thermo	0.1 mm x 5.0 cm	PSDVB (Polystyrene Divinylbenzene)	Monolith

2.5.2 HPLC Buffers

Triethylammonium acetate (TEAA, Sigma-Aldrich), triethylamine (TEA, ThermoFisher), Tributylamine (TBA, Acros Organics), acetonitrile and water (HPLC grade, Fisher Scientific), 1,1,1,3,3,3,-Hexafluoro-2-propanol (HFIP, Sigma-Aldrich). Glacial acetic acid, combined with tributylamine (TBA) to prepare tributylammonium acetate (TBAA), was sourced from VWR. All oligonucleotides were synthesised by Eurofins, GeneWiz or IDT. SELEX library was sourced from IDT. For MS analysis, acetonitrile and water (UHPLC MS grade, ThermoFisher) and 1,1,1,3,3,3,-Hexafluoro-2-propanol LC-MS grade (99.9% ThermoFisher) were used. TBAA stock (100 mM) was prepared by combining 6 ml tributylamine (Acros organics) with 1.5 ml glacial acetic acid (VWR) and made up to 250 ml with 225 ml ACN and 17.5 ml water.

2.5.3 HPLC Gradients

Gradient 1: Mobile phase A 0.1 M TEAA pH 7.0; Mobile phase B 0.1 M TEAA, pH 7.0 containing 25% acetonitrile. Gradient starting at 10% buffer B for 2 minutes, followed by a non-linear extension (curve 3) to 100% buffer B over 18 minutes at a flow rate of 0.6 ml/min at 40 °C.

Gradient 2: Mobile Phase A: TBAA 5 mM 10% acetonitrile. Mobile Phase B: TBAA 5 mM 80% acetonitrile. Gradient starting at 5% buffer B for 2 minutes, followed by a

non-linear extension (curve 3) of 40% buffer B over 15 minutes, then extended to 100% buffer B over 2.5 minutes at a flow rate of 2.2 µl/min at 50 °C.

Gradient 3: Mobile phase A 0.1 M TEAA pH 7.0; Mobile phase B 0.1 M TEAA, pH 7.0 containing 25% acetonitrile. Gradient starting at 35% buffer B for 2 minutes, followed by a non-linear extension (curve 2) to 80% buffer B over 18 minutes, then extended to 100% buffer B over 5 minutes at a flow rate of 0.6 ml/min at 40 °C.

Gradient 4: Mobile Phase A: 0.2% TEA, 50 mM HFIP. Mobile Phase B: 0.2% TEA, 50 mM HFIP, 20% acetonitrile. Gradient starting at 10% buffer B followed by a non-linear extension (curve 2) to 100% buffer B over 18 minutes, then extend to 100% Buffer B over 5 minutes at a flow rate of 0.2 ml/min at 30 or 95 °C.

Gradient 5: Mobile Phase A: 0.1 M TEAA 50 mM HFIP. Mobile Phase B: 0.1 M TEAA, 50 mM HFIP, 25% acetonitrile. Gradient starting at 35% buffer B for 2 minutes, followed by a linear extension to 80% buffer B over 10 minutes, then extended to 100% buffer B over 5 minutes at a flow rate of 0.6 ml/min at 80 °C.

Gradient 6: Mobile Phase A: 0.1 M TEAA 50 mM HFIP. Mobile Phase B: 0.1 M TEAA 50 mM HFIP 25% acetonitrile. Gradient starting at 25% buffer B for 2 minutes, followed by a linear extension to 80% buffer B over 22 minutes, then extended to 100% buffer B over 5 minutes at a flow rate of 0.3 ml/min at 80 °C.

Gradient 7: Mobile Phase A: TBAA 5 mM 10% acetonitrile. Mobile Phase B: TBAA 5 mM 80% acetonitrile. Gradient starting at 20% buffer B for 1 minute, followed by a linear extension to 30% buffer B over 1 minute, followed by a linear extension to 60% buffer B over 13 minutes, then extended to 90% buffer B over 2.5 minutes at a flow rate of 2.2 µl/min at 40 °C.

Gradient 8: Mobile phase A 0.1 M TEAA pH 7.0; Mobile phase B 0.1 M TEAA, pH 7.0 containing 25% acetonitrile. Gradient starting at 25% buffer B for one minute, followed by a non-linear extension (curve 3) to 72.5% buffer B over 16 minutes at a 2.2 µl/min flow rate.

Gradient 9: Mobile phase A 0.1 M TEAA pH 7.0; Mobile phase B 0.1 M TEAA, pH 7.0 containing 25% acetonitrile. Gradient starting at 30% buffer B for one minute, followed by a non-linear extension (curve 2) to 48% buffer B over 15 minutes, followed by a

non-linear extension (curve 2) to 65% buffer B over 5 minutes at a flow rate of 0.3 ml/min at 95 °C

Gradient 10: Mobile phase A: 15 mM DBAA; Mobile phase B: 15 mM DBAA containing 50% acetonitrile. Gradient starting at 45% buffer B for one minute, followed by a linear extension to 55% buffer B over one minute, followed by a linear extension to extension to 80% buffer B over 13 minutes at a flow rate of 2.2 µl/min at 50 °C

Gradient 11: Mobile phase A: 15 mM DBAA, mobile phase B: 15 mM DBAA containing 50% acetonitrile. Gradient starting at 40% buffer B for one minute, followed by a linear extension to 50% buffer B over one minute, followed by a linear extension to 75% buffer B over thirteen minutes, at a flow rate of 2.2 µl/min at 75 °C.

Gradient 12: Mobile phase A: 15 mM DBAA 50 mM HFIP, mobile phase B 15 mM DBAA 50 mM HFIP containing 50% acetonitrile. Gradient starting at 20% buffer B for 5 minutes, followed by a non-linear extension to 85% buffer B over 17 minutes at a flow rate of 2.2 µl/min at 50 °C

Gradient 13: Mobile phase A: TBAA 5 mM containing 10% acetonitrile. Mobile Phase B: TBAA 5 mM containing 80% acetonitrile. Gradient starting at 35% buffer B for 5 minutes, followed by a non-linear extension (curve 3) to 65% buffer B over 14 minutes at a flow rate of 2.2 µl/min at 50 °C.

Gradient 14: Mobile phase A: TBAA 5 mM containing 10% acetonitrile. Mobile Phase B: TBAA 5 mM containing 80% acetonitrile. Gradient starting at 15% buffer B for 5 minutes, followed by a non-linear extension (curve 2) to 60% buffer B over 14 minutes at a flow rate of 2.2 µl/min at 75 °C.

Gradient 15: Mobile phase A 0.1 M TEAA pH 7.0; Mobile phase B 0.1 M TEAA, pH 7.0 containing 25% acetonitrile. Gradient starting at 35% B for one minute following a non-linear extension (curve 2) to 57.5% B over 21 minutes at a flow rate of 0.25 ml/min 50 ºC.

Gradient 16: Mobile phase A 0.1 M TEAA pH 7.0; Mobile phase B 0.1 M TEAA, pH 7.0 containing 25% acetonitrile. Gradient starting at 35% B for one minute following a non-linear extension (curve 2) to 52.5% B over 23 minutes at a flow rate of 0.25 ml/min 95 ºC.

Gradient 17: Mobile phase A: 15 mM DBAA, mobile phase B: 15 mM DBAA containing 50% acetonitrile. Gradient starting at 40% B for five minutes, followed by a non-linear extension (curve 3) to 80% B over 15 minutes at a flow rate of 2.2 µl/min at 50 ºC.

Gradient 18: Mobile phase A: 15 mM DBAA, mobile phase B: 15 mM DBAA containing 50% acetonitrile. Gradient starting at 40% B for five minutes, followed by a non-linear extension (curve 3) to 80% B over 15 minutes at a flow rate of 2.2 µl/min at 75 ºC.

Gradient 19: Mobile phase A: TBAA 5 mM containing 10% acetonitrile. Mobile Phase B: TBAA 5 mM containing 80% acetonitrile. Gradient starting at 40% B for 1 minute, followed by a non-linear extension (curve 3) to 65% B over 17.5 minutes at a flow rate of 2.2 µl/min at 50 ºC.

Gradient 20: Mobile phase A: TBAA 5 mM containing 10% acetonitrile. Mobile Phase B: TBAA 5 mM containing 80% acetonitrile. Gradient starting at 40% B for 1 followed by a non-linear extension (curve 3) to 65% B over 21 minutes at a flow rate of 2.2 µl/min at 75 ºC.

2.6 Gel Electrophoresis Analysis

2.6.1 Agarose gel electrophoresis

Agarose gels were prepared within a concentration of 3-4% with Ultrapro Agarose (molecular biology grade, Appleton) for analysis of oligonucleotides. Gels were run in a Bio-Rad gel tank in 1 x TAE buffer (Sigma Aldrich) at 100 mv for 45 minutes. Midori green direct dye (Nippon Genetics) was used for staining. 1 µl of Midori green dye was added to 4 µl of sample and 1 µl of Midori green dye. The gels were visualised in a blue light illuminator.

2.6.2 SDS PAGE gel electrophoresis

Proteins were denatured at 95 ºC in mercaptoethanol and flash-cooled on ice before SDS PAGE gel electrophoresis, 6% stacking gel, and 12% resolving gel. Gels were stained with Coomassie Blue (Thermo Fisher).

Table 2.5: SDS PAGE gel mix composition

Component (7cm x 7cm x 0.75ml)	**6% stacking Gel**	**12% Resolving Gel**
MilliQ Water	1.64ml	2.44ml
40% Acrylamide/bis acrylamide (5% bis)	562 µl	1.25ml
0.5M Tris-HCl pH6.3, 0.4% SDS	375 µl	-
1.5M Tris-HCl pH8.8, 0.4% SDS	-	1.25ml
10% ammonium persulphate	37.5 µl	50µl
Tetramethyl ethylenediamine	7.5 µl	10µl

2.7 SELEX

2.7.1 SELEX buffers

Table 2.6: SELEX Binding/Wash Buffer Composition

Component	**Concentration**
Sodium Chloride	140 mM
Potassium Chloride	2 mM
Magnesium chloride	5 mM
Calcium Chloride	2 mM
Tris HCL	20 mM (pH 7.6)

2.7.2 Conventional SELEX protocol

Table 2.7: Conventional SELEX, bead volume, library mass and incubation time, round-by-round

Round number	**Bead volume µl**	**Library mass µg**	**Incubation time min**
1	240	74.0	90
2	20	12.0	75
3	24	15.0	75
4	8	4.9	60
5	8	5.0	60
6	5.3	5.0	60

NS	30	6.0	45
7	5.3	5.5	45
8	1.64	2.0	45

Table 2.8: Conventional SELEX, Library: MDA-9 molar ratios, round by round

Round Number	Library nmol (approximate)	MDA-9 nmol	Molar Ratio (Lib: MDA-9)
1	3	3	1
2	0.5	0.25	2
3	0.6	0.3	2
4	0.2	0.1	2
5	0.2	0.1	2
6	0.2	0.067	3
7	0.225	0.067	3
8	0.08	0.02	4

Table 2.9: Conventional SELEX, washes and PCR cycles, round-by-round

Round Number	Washes	Wash Time	PCR Cycles
1	1	5	8
2	1	5	18
3	1	7.5	15
4	1	10	12
5	1	15	12
6	1	20	10
7	1	25	8
8	1	30	6

2.7.3 Low amplification SELEX protocol

Table 2.10: Low amplification SELEX, bead volume, library mass and incubation time, round-by-round

Round number	Bead volume µl	Library mass µg	Incubation time
1	214	400	90
2	40	2.5	75
3	40	0.6	60
NS	32	3.8	30
4	12	3.6	60
5	7.2	5.5	45
6	4.0	6.2	45

Table 2.11: Low amplification SELEX, Library: MDA-9 molar ratios, round by round

Round Number	Library nmol	MDA-9 nmol	Ratio (Lib: MDA-9)
1	16	2.67	6
2	0.1	0.5	0.2
3	0.025	0.5	0.05
4	0.15	0.15	1
5	0.225	0.09	2.5
6	0.25	0.05	5

Table 2.12: Low amplification SELEX, washes and PCR cycles, round-by-round

Round Number	Washes	Wash Time	PCR Cycles
1	1	5	0
2	1	5	0
3	1	5	12
4	2	10	12
5	2	20	12
6	3	30	8

1.1 Quantification of Oligonucleotides

Quantification of nucleotides was carried out on a Thermo Scientific NanoDrop 2000/2000c Spectrophotometer at 260 nm.

2.8 List of Oligonucleotides

Table 2.13: List of oligonucleotides used in PCR

Name	Length (bp)	Sequence 5'-3'	Modification	Manufacturer
PC1	20	**TATATACTGAGCGCATATAT**	5' Phosphate	GENEWIZ
PC2	20	**TATATACTGAGCGCATATAT**	5' Biotin	GENEWIZ
PC3	20	**ATATATGACTCGGGTATATA**	N/a	GENEWIZ
PC4	80	**ATATATGACTCGGGTATATA-(N40)-ATAT ATGCGCTCAGTATATA**	N/a	IDT
PC5	20	**TATATACTGAGCGCATATAT**	5' 12 C Spacer	Eurofins
PC6	20	**TATATACTGAGCGCATATAT**	5' Texas red	Eurofins
PC7	20	**ATATATGACTCGGGTATATA**	5' Texas red	Eurofins
PC8	20	**TATATACTGAGCGCATATAT**	Phosphorthioate	Eurofins
PC9	80	**ATATATGACTCGGGT ATATAATGGTCCTAATAGCCCTATAGGTA CACTGACTATAGGCCGATATATGCGCTCA GTATATA**	N/a	Eurofins
PC10	80	**GTCTATATGATCTGTAACTCATGGTCCTAA TAGCCCTATAGGTACACTGACTAT AGGCCGCCAGCAGTGAGTCATCAGAT**	N/a	Eurofins
PC11	20	**GTCTATATGATCTGTAACTC**	N/a	Eurofins
PC12	20	**ATCTGATGACTCACTGCTGG**	N/a	Eurofins

2.9 Mass spectrometry

2.9.1 Instrument Methods and Data Acquisition: Oligonucleotides

Mass spectrometry analysis was performed using an Orbitrap Exploris 240 mass spectrometer (Thermo Fisher) in negative ion mode. Instrument operation was carried out using Thermo Scientific Tune and XCalibur software. Data acquisition was performed using data-dependent acquisition in full-scan negative mode, scanning from 450 to 3000 m/z, with an MS1 resolution of 240 000 and a normalised automatic gain control (AGC) target of 300%. Intact mass analysis and deconvolution were performed using the 'Xtract deconvolution" algorithm within Freestyle software (Thermo Fisher). Deconvolution was performed with the software considering negative charges and the nucleotide isotope table. Mass accuracy calibration is performed weekly on this instrument; additionally, mass accuracy was determined on the day using a 36-base oligo with a mass error of -13.569 ppm.

2.10 Microbial Cell Culture

2.10.1Media

LB Broth: 25 g LB broth, miller (Sigma-Aldrich) was dissolved in 1 L deionised water and sterilised via autoclave (121 ºC for 15 mins at 15 psi).

LB Plates (ampicillin): 40 g LB broth, miller, with agar (Sigma-Aldrich) was dissolved in 1 L deionised water and sterilised via autoclave (121 ºC for 15 mins at 15 psi), following cooling, but prior to the agar setting, ampicillin was added to a final concentration of 100 mg/L.

2.11 *E. coli* transformation

E. coli DH5α cells were used for plasmid cloning, and BL21 cells for MDA-9 expression. Competent cells were prepared using a Zymo ® Mix and Go kit. Briefly, 0.5 ml of *E. coli* liquid culture (grown overnight) was used to inoculate 50 ml of ZymoBroth™ in a 500 ml culture flask. The flask was shaken at 250 rpm until it reached an OD 0.6 (600 nm). The culture was transferred to ice and, following 10 minutes of cooling was pelleted via centrifugation at 2,500 xg for 10 minutes at 4 ºC. The supernatant was removed, and the cells were resuspended in 5 ml ice-cold Zymo

wash buffer. The cells were re-pelleted and resuspended in 5 ml ice-cold 1x competent buffer, divided into 0.2 ml fractions and stored at -80 ºC before transformation.

A 0.2 ml aliquot of competent cells prepared in section 2.11 was thawed on ice; 5 ng of plasmid DNA was made up to a 5 µl volume of deionised water and added to the cells on ice. Mixing was done by gentle tapping followed by incubation for 30 minutes. The cells were heat shocked for 30 seconds at 42 ºC on a hot plate and then immediately placed on ice for 5 minutes. 250 ml of SOC medium (ThermoFisher) was pre-warmed and added to the cells; the cells were then shaken at 225 rpm at 37 ºC in a shaking incubator. Ampicillin plates were inoculated with the cells with an additional negative control inoculated with SOC medium and incubated at 37 ºC overnight.

A single colony was taken from a freshly incubated plate and used to inoculate 5 ml of LB media containing 100 µg/ml ampicillin; the media was then incubated in a culture tube at 37 ºC for 12 hr with shaking at 250 rpm. Following outgrowth, 2 ml of this culture was used to inoculate 50 ml of LB media in a culture flask containing 100 µg/ml ampicillin (2% inoculum). The culture was incubated until it reached OD 0.6 (600 nm), and the plasmid was either extracted (DH5-α) or the culture was used for protein expression (BL21).

2.11.1 Plasmid DNA Extraction

Plasmid DNA extraction was carried out using a Qiagen miniprep kit; briefly, the cell culture was spun down and pelleted at 6800 xg in a benchtop centrifuge for 3 minutes at room temperature. The pelleted cells were resuspended in 250 µl buffer P1 and mixed with 250 µL buffer P2. Resuspend pelleted bacterial cells in 250 µl Buffer P1 and transfer to a microcentrifuge tube. 350 µl of buffer N3 was added and mixed via inverting six times prior to centrifugation at 17,900 xg for 10 minutes. The supernatant was applied to a QIAprep spin column and centrifuged for 60 seconds. The flow-through was discarded, and washes were performed with buffers PB and PE. To elute the plasmid, 50 µl of buffer EB was added and centrifuged for 1 min.

2.11.2 Restriction digests

Restriction digests were performed with either SphI, XbaI (New England Biolabs) or a combination of the two; 1 µg of plasmid was digested in each reaction in 50 µl of 1x Cut smart buffer (New England Biolabs) containing ten units of enzyme 37 ºC for 1 hour.

2.12 MDA-9 Expression

The pet151/MDA-9 plasmid was transformed into BL21 *E. coli* as described in 2.11 and grown on an ampicillin plate overnight at 37 ºC. A single colony was taken and used to inoculate 5 ml of LB media containing 100 µg/ml ampicillin; the media was then incubated in a culture tube at 37 ºC for 12 hr with shaking at 250 rpm. Following outgrowth, 2 ml of this culture was used to inoculate 50 ml of LB media in a culture flask containing 100 µg/ml ampicillin (2% inoculum).

Following overnight growth of a 50 ml BL21 culture, 1 ml of culture was used to inoculate 100 ml of LB media containing 100 µg/ml ampicillin (1% inoculum). The cells were grown to OD 0.6 (600 nm) and induced with IPTG (1 mM) before incubation at 37 ºC for four hours. Following induction, cells were centrifuged at 8000 xg for 10 minutes, and the supernatant was removed. Cell pellets were frozen at -80 ºC prior to further use.

2.12.1 Cell lysis

Cell pellets were made up to 50 ml in deionised water containing 10 % glycerol, protease inhibitors (Roche) and lysozyme to a concentration of 1 mg/ml. The cells were warmed to 37 ºC, frozen at -80 ºC and thawed at 37 ºC. The freeze-thaw was repeated once; the cells were centrifuged at 8000 xg for 10 minutes, and the supernatant was divided into 2 ml quantities and frozen at -80 ºC for further use.

2.12.2 Protein purification: HisPur™ Ni-NTA Resin

Table 2.14: Protein purification buffers

Buffer	Components
Equilibration	20 mM sodium phosphate 300 mM sodium chloride 10 mM imidazole
Wash	20 mM sodium phosphate 300 mM sodium chloride 25 mM imidazole
Elution	20 mM sodium phosphate 300 mM sodium chloride 250 mM imidazole

To prepare for binding, 1 ml of Ni-NTA resin (Thermo Fisher) was added to a 15 ml falcon tube. The tube was centrifuged for 2 minutes at 700 xg, and the supernatant was removed. Two resin bed volumes of equilibration buffer were added, and the resin was resuspended. Centrifugation was repeated, and the supernatant was removed. Following preparation, 5 ml of cell lysate was made up with an equal volume of equilibration buffer and incubated with the resin for 30 minutes with gentle rotation. Following incubation, the tube was centrifuged, and the supernatant was removed. Two 5ml washes were carried out with wash buffer for 5 minutes, followed by centrifugation and supernatant removal.

One volume of elution buffer was incubated with the resin for 2 minutes with gentle shaking to elute MDA-9. The tube was centrifuged, the supernatant was collected, and the elution was repeated twice. Prior to further use, the protein was dialysed against an appropriate buffer to remove imidazole using D-Tubes™ (Merck Millipore)

2.12.3 Protein-magnetic bead binding for SELEX

For in vitro selection, purified MDA-9 fractions were bound to HisPur™ Ni-NTA Magnetic Beads (Thermo Scientific). The buffers for each stage were the same as in Table 2.14. The binding proceeded per the manufacturers' instructions; 300 µl of beads were placed into an Eppendorf tube, followed by an initial wash with 500 µl of equilibration buffer. 1 ml of purified protein fraction, dialysed against the equilibration buffer, was incubated with the beads for 30 minutes. Following binding, the unbound MDA-9 was removed and the beads were stored in 300 µl of buffer (to maintain bead concentration). A small fraction of beads was taken (40 µl), and the protein was eluted in two aliquots of 50 µl elution buffer for quantification.

2.12.4 Protein quantification

Quantification of MDA-9 eluted from Ni-NTA beads was carried out using a NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific) at 280 nm prior to accurate quantification via Bradford assay, using a TEECAN plate reader at 595 nm. A range of concentrations of Bovine serum albumin (BSA, Sigma Aldrich, see Table 2.15) were used to generate a seven-point standard curve. The beads were calculated to have a binding capacity of 0.5 µg/µl of bead suspension for MDA-9. Bradford reagent was sourced from Bio-Rad.

Table 2.15: Bradford Assay standard solution concentrations

Solution Number	BSA Concentration (µg/ml)
Stock	1500
1	0
2	0.75
3	1.5
4	2.25
5	3
6	3.75
7	4.5
8	5.25

2.13 Proteomic analysis

2.13.1 Reduction, Alkylation and Tryptic digestion

Using a clean scalpel, MDA-9 bands were excised from a stained SDS PAGE gel, cut into 1 mm^2 cubes and stored in an Eppendorf tube. They were suspended in 750 µl of deionised water to destain the cubes before adding 750 µl 200 mM ammonium carbonate /40% v/v ACN. The tube was vortexed and incubated at 37 ºC for five minutes. The solution was replaced, and incubation was repeated until the cubes were destained. The gel pieces were dehydrated by adding 750 µl ACN. Reduction and alkylation was performed by adding 20 µl of 10 mM dithiothreitol (Sigma-Aldrich) in 100 mM ammonium carbonate and incubated for 45 minutes at 56 ºC. The dithiothreitol solution was discarded, and 20 µl of 55 mM iodoacetamide in 100 mM ammonium carbonate was added prior to incubation on ice in darkness for 20 minutes. Following this, 20 µl of 10 mM dithiothreitol in 100 mM ammonium carbonate was added to quench the alkylation reaction and the supernatant was removed. The gel pieces were then dehydrated with ACN and stored. The gel pieces were rehydrated with water before digestion with 50 µl of 5 ng/ µl trypsin solution (Sigma-Aldrich) in 50 mM ammonium carbonate and 1% Protease Max overnight at 37 ºC. Following tryptic digest, the peptide solution was removed before mass spectrometry analysis.

Mass Spectrometry- Peptide Analysis

LC-MS/MS analysis using nano-flow liquid chromatography (U3000 RSLCnano, Thermo Scientific) coupled to a hybrid quadrupole-orbitrap mass spectrometer (Q Exactive HF, Thermo Scientific). Peptides were separated on an Easy-Spray C18 column (75 µm x 50 cm) using a 2-step gradient from 3% solvent A (0.1% formic acid in water) to 50% solvent B (0.1% formic acid in 80% acetonitrile) over 30 minutes at 300 nL min-1, 40 °C. The mass spectrometer was programmed for data-dependent acquisition with 10 product ion scans (resolution 30,000, automatic gain control 1e5, maximum injection time 60 ms, isolation window 1.2 Th, normalised collision energy 27, intensity threshold 3.3e4) per full MS scan (resolution 120,000, automatic gain control 1e6, maximum injection time 60 ms) with a 20 second exclusion time.

Data was searched using MASCOT v 2.5.0 using *E. coli* database 03032022 (Uniprot), including the MDA-9 sequence. Search parameters “tryptic enzyme“ with a specificity of up to three one cleavages, variable modification of methionine residue oxidation and carbamidomethylation of cysteine residues. The databases were searched peptide tolerance = 0.1 Da, MS/MS tolerance = 0.1 ppm Da and peptide charge 2+, 3+ and 4+). A false discovery rate filter was not applied.

2.14 Isothermal Titration Calorimetry

2.14.1 Binding buffers

Table 2.16: BSA aptamer ITC buffer composition

Component	Concentration
Sodium Chloride	100 mM
Magnesium Chloride	2 mM
Potassium Chloride	5 mM
Tris HCl	20 mM (pH 7.6)
Tween 20	1% v/v
Glycerol	10% v/v

Table 2.17: Thrombin Aptamer ITC buffer composition

Component	Concentration

Sodium Chloride	140 mM
Potassium Chloride	5 mM
Calcium Chloride	1 mM
Tris-acetate	20 mM (pH 7.6)
Glycerol	10% v/v

Table 2.18: MDA-9 ITC buffer composition

Component	**Concentration**
Sodium Chloride	140 mM
Potassium Chloride	2 mM
Magnesium chloride	5 mM
Calcium Chloride	2 mM
Tris HCL	20 mM (pH 7.6)

2.14.2 ITC method

ITC experiments were run using TA instruments Nano ITC. All aptamers were heated to 95 °C in the appropriate buffer and then flash-cooled on ice for ten minutes; this ensures the denaturation of secondary structures. The samples were then degassed using the degasser associated with the Nano ITC. The desired volume of sample was injected into the ITC chamber; the titrant was loaded into the 50 µl syringe, which was screwed into the burette and installed above the sample chamber.

Each titration consisted of a five-minute equilibration followed by an initial purge injection of 0.2 µl and 19 successive injections of 1.82-2.39 µl (depending on titration) with a spacing of 180 sec between each injection. The sample chamber was heated to 25 ºC or 37 ºC, depending on the titration and stirred at 300 rpm. The raw data was first corrected for the dilution heat of the ligand (calculated as the mean injection heats of the last five injections) and then analysed with the Nano analyse software (TA instruments); the isotherms were fitted with a single-site binding model.

3 Chapter 3 Development and optimisation of critical steps in SELEX

3.1 Abstract

Critical steps in the production of DNA aptamers via SELEX include the amplification of ssDNA following in vitro selection and the generation of ssDNA from the resulting dsDNA. Several caveats are associated with current ssDNA amplification and re-generation methods, which can lower the success rates of SELEX experiments. They often result in low yields, thereby decreasing diversity or failing to eliminate parasitic PCR by-products, accumulating by-products from round to round. Both contribute to the failure of SELEX protocols.

In this chapter, the aim was to optimise each of the critical steps of the SELEX method before implementing the optimised SELEX method to generate ssDNA aptamers in future work. In part A of this chapter, library design was informed via extensive literature research, and PCR conditions were optimised to prevent the formation of PCR by-products. Following the optimization of the library design in conjunction with PCR conditions, a comparison was performed between an alternative library design and PCR conditions from the literature. The optimised PCR method from this study generated higher-quality PCR products without the presence of by-products observed in the PCR conditions from the literature.

In part B, 5’ chemical modifications on PCR primers were used to amplify PCR fragments before separating and purifying the DNA strands using denaturing IP RP HPLC. This novel method offers significant advantages over existing methods for the rapid generation and purification of ssDNA of the correct size and offers the opportunity to improve the development of new aptamers via SELEX. The results show that using a 5’ Texas Red modification on the reverse primer in the PCR stage enabled purification of the ssDNA from its complementary strand via IP RP HPLC under denaturing conditions. Following lyophilisation, ssDNA yields of up to 80% were obtained. In comparison, the streptavidin-biotin affinity chromatography generates ssDNA with a yield of only 55%.

3.2 Introduction

Aptamers have been isolated for various applications ranging from drug detection to affinity chromatography [5][6]. However, several caveats are associated with successful aptamer selection via selective enhancement of ligands via exponential enrichment (SELEX) [7]. A significant problem in SELEX protocols is their low success rate, partly due to complications relating to the generation of ssDNA following PCR. This step is crucial as the quality and quantity of ssDNA generated greatly influences the successful evolution of aptamers [8]. Current methods are low-yielding or fail to eliminate by-products, which can grow exponentially from round to round [9]. The two most common methods, exonuclease digestion and streptavidin/biotin affinity chromatography, particularly suffer from this as they cannot distinguish between strands of different sizes.

Exonuclease digestion involves labelling the reverse primer in PCR with a 5' phosphate substrate for phosphate-dependent exonucleases [107]. However, it must be noted that these nucleases are not completely specific, and so some of the non-phosphorylated strands can be digested. In addition, any PCR by-products that do not have a 5' phosphate group will primarily be undigested by the enzyme and taken through to the selection stage. Furthermore, any nucleases in the product require careful removal prior to downstream applications. The streptavidin-biotin affinity chromatography method involves the labelling of the reverse primer of the PCR reaction with a 5' biotin modification; this causes the complementary strand to interact strongly with streptavidin such that the DNA duplex can be denatured, and the ssDNA separated [190]. This method also suffers from the inability to distinguish between products and by-products, potentially contaminating the next selection stage with streptavidin if additional purification is not carried out [105].

Co-polymerisation within an agarose gel using a polymerisable 5' modification has also demonstrated effective separation of the two strands [4]. The gel mixture is loaded into the wells of a pre-set polyacrylamide gel. When set, high temperatures denature the duplex and release the strand not covalently linked to the acrylamide [191]. This separation is efficient, offering yields of over 80 %. Methods such as gel extraction via crush and soak and asymmetric PCR have yielded reasonable quantities of ssDNA. Asymmetric PCR involves biasing primer concentrations so that the concentration of the forward primer exceeds that of the reverse primer. This

leads to the generation of a higher quantity of the desired strand, making purification more efficient. When coupled with exonuclease digestion, asymmetric PCR yields over 80% [110]. However, these methods still do not address the issue of by-product formation, particularly those close in size to the DNA template, which are typically not removed using existing methods [101][71]. In addition, re-generation of ssDNA is essential for other molecular biology protocols, including sequencing, sample preparation for mass spectrometry and microarray technology [192][110].

3.3 Results and Discussion Part A: ssDNA Library design and optimisation of PCR amplification

3.3.1 ssDNA library design

Before implementing SELEX to generate high-affinity nucleic acid aptamers, each step of the SELEX protocol must be optimised to increase the likelihood of successful aptamer isolation. Essential parameters to optimise in SELEX methodology include the design of the aptamer library, PCR and ssDNA re-generation stages and ssDNA purification. The first part of any SELEX research is library design (see Chapter 1), which has been shown to impact the overall outcome of a SELEX experiment significantly. In particular, the design of the priming regions of the aptamer library is critical to the success of SELEX methods [48,54]. The library was designed based on previous literature utilising SELEX to successfully generate ssDNA aptamers to diverse protein targets (see Chapter 1, section 1.6.1). Based on previous studies, a ssDNA library (80 nts) was synthesised with two fixed 20 nt priming regions flanking a 40 nt random region.

The random region of the library (N40) was synthesised with a bias in the base composition, 25% T, 15% C, 20% A, and 40% G. A higher %G was used to favour the formation of G quadruplexes. G4 quadruplexes have been found to occur in many aptamers selected against biologically relevant protein targets, this should not be surprising as the G tetrad can act as a stable foundation for a vast array of secondary structures, as discussed in Chapter 1 [193,194]. Besides the benefits of G-quadruplex forming aptamers, some G-rich DNA sequences demonstrate cancer-selective antiproliferative activity. While favouring this latent anti-cancer activity in a SELEX

protocol is tricky, it could be a relevant, beneficial property for any aptamers generated with an abundance of guanine [195].

The primer regions were selected to be A-T-rich based on previous studies by Crameri and Stemmer, which showed that their A-T-rich primers significantly increased the quality of PCR products [75]. The 5' and 3' ends of primers can act as starting points for by-product formation, with A-T consisting of two hydrogen bonds compared to the three of G-C; as a result, any by-product form should be weaker, and the annealing temperature should be high enough to hinder their formation following denaturation. The ssDNA library sequence is shown below.

Library 5'-3'

5'-ATATATGACTCGGGTATATA-(N40)-ATATATGCGCTCAGTATATA-3'

3.3.2 Optimisation of ssDNA amplification using PCR

a. Optimising cycle number

The amplification of the ssDNA using PCR can be a significant challenge in SELEX. ssDNA aptamers often fold in stable secondary structures, presenting challenges for PCR amplification, resulting in the formation of PCR by-products. As discussed in Chapter 1, a poorly optimised PCR method can lead to many problems later in the SELEX process. Typical PCR by-products include "ladder type" by-products that behave parasitically and can thus grow accumulatively from round to round, becoming the major component of the aptamer library and detrimental to the SELEX protocol that, otherwise, may have succeeded [74].

Therefore, initial work focussed on optimising PCR conditions in SELEX protocols. The aim was to study the effect of a range of different PCR parameters on the amplification of the ssDNA library; the parameters optimised throughout the initial experiments are shown in Table 3.1. Details of the PCR components and experimental parameters are shown in Chapter 2, section 2.2.

Table 3.1: PCR variables for rection optimisation

Parameter	Range of variation
PCR cycles	10-35
$MgCl_2$ concentration	3.0-5.0 mM
Primer concentration	62.5-1000 nM
Annealing Temperature	47-62 ºC

In the initial experiment, the ssDNA random library was amplified from 10 to 35 cycles to determine the optimal number of amplification cycles needed to achieve maximum yield without generating significant by-products. Six replicates were run for each cycle number: 15, 20, 25, 30, and 35. A negative control was run simultaneously in the absence of DNA polymerase. Gel electrophoresis analysis of these samples is shown in Figure 3.1A. The results indicate that the initial amplification was successful, producing the expected 80 bp dsDNA product. However, the gel electrophoresis analysis revealed two major bands rather than the anticipated single dsDNA band. At cycle 10, these bands appear as a smear and become more defined at higher cycle numbers. PCR amplification in SELEX protocols typically runs between 6 and 18 cycles, with a decrease in amplification cycles as the protocol progresses.

b. Optimising MgCl2 concentrations

Magnesium chloride is commonly added to PCR reactions to increase the specificity of the polymerase, thereby enhancing yield and decreasing by-product formation [196]. However, concentrations exceeding five mM can cause non-specific interactions between the primers and non-priming template regions, forming by-products [197,198].

Five PCR reactions were performed with $MgCl_2$ concentrations ranging from 1.5 to 6.0 mM. (see Figure 3.1B). In typical SELEX protocols, the $MgCl_2$ concentration ranges from 0.5-5 mM, with the higher concentration theoretically providing higher yields. [135,199]. The results show that PCR product quality did not improve significantly

across the different $MgCl_2$ concentrations. However, there was a notable increase in yield when the concentration was increased from 1.5 mM to 3.0 mM. Therefore, it was decided to increase the $MgCl_2$ concentration. As there was no apparent visual improvement in yield from 3.0 to 6.0 mM, it was decided to use 3.0 mM going forward. This would increase yield without the risk of forming byproducts

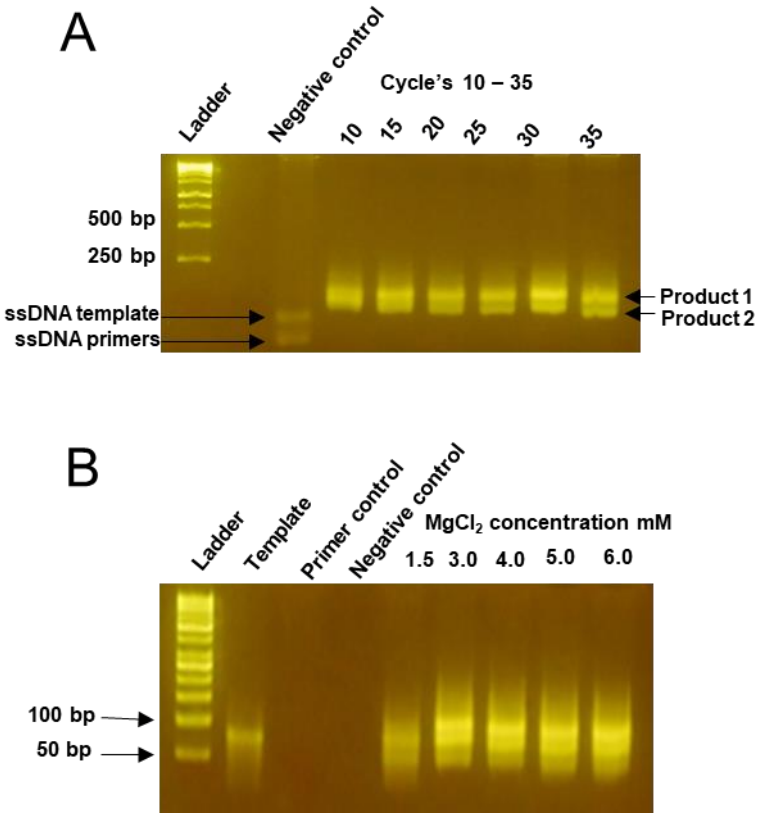


Figure 3.1: Optimisation of PCR amplification of the ssDNA SELEX library sections A-B. *A) 4% agarose gel electrophoretogram showing the effect of different cycle numbers in the PCR B) 4% agarose gel electrophoretogram showing the effect of varying $MgCl_2$ concentration in the PCR.*

c. Optimising Agarose Gel analysis

In addition to optimising the PCR conditions, additional work was performed to optimise the analysis of the PCR products using gel electrophoresis. Initial experiments used high percentage agarose gels, which required long run times for each analysis. Furthermore, it was noted that using high percentage agarose gels (4%) resulted in a significant temperature increase in the gel during the electrophoresis. It was hypothesised that this could be causing the denaturing of the dsDNA and the formation of multiple bands in the electrophoresis analysis of the SELEX PCR products. Therefore, the agarose concentration was lowered to 3%, the

double-banded product was no longer seen, and only the expected single band was observed. A comparison of the same PCR products run on a 3% and 4% constant voltage (100 mV) gel is shown in Figure 3.2C and D.The results show that multiple bands were observed in the 4% (Figure 3.2C) gel at 100 mV, consistent with previous results, and as previously noted, it resulted in a high gel temperature during electrophoresis. In contrast, the 3% gel at (Figure 3.2D) 100 mV resulted in a single band and no significant increase in temperature during the electrophoresis. These results confirm that using high % agarose resulted in partial denaturing of the dsDNA product and the appearance of multiple bands. Therefore, 3% agarose gels were used for subsequent analysis.

d. Optimising primer concentrations

Primer-primer and primer-product interactions are known to be significant contributors to by-product formation in PCR. However, primers must be present at sufficient concentration in the reaction to yield a high dsDNA yield. The objective was to determine if a lower primer concentration would reduce by-product formation, thereby establishing a correlation between by-product levels and primer concentration. PCR amplifications were performed over 10 or 20 cycles using primer concentration over a range (62.5 to 250 nM**)**. The results are shown in Figure 3.2E and show that in all reactions, the successful amplification of the DNA. However, there is no significant increase in the yield of PCR with varying primer concertation based on gel staining intensity. An excess of unused primers is evident in the PCR with the highest concentration of primers (500-1000 nM). Therefore, based on this data, a final primer concentration of 125 nM was selected, which provided a sufficient yield for the PCR product and minimised the potential formation of primer dimers.

e. Optimisation of annealing temperature

The annealing temperature in PCR is crucial as it needs to be sufficiently low for primers to anneal to the template and ensure a high amplification yield, yet not so low that non-specific binding occurs, which can lead to the formation of PCR parasites over successive amplification cycles. PCR reactions were conducted across an annealing temperature gradient from 47 to 62 °C. The goal was to determine if a higher annealing temperature could improve PCR product quality without negatively affecting yield. If successful, a higher temperature will be adopted for future experiments.

Additionally, a negative control was run without any DNA template but with DNA polymerase and primers present to identify if any primer-primer dimers or larger by-products were formed. The results of this annealing temperature optimisation experiment are shown in Figure 3.2F. The results show no significant change in product quality across the temperature range. Therefore, it was decided to retain the current annealing temperature of 47 °C, with the expectation that a lower temperature would increase the yield of the PCR product. At this temperature, there is no sign of by-product formation in the template control, indicating no significant interaction between the primers (facilitated by the polymerase) at this annealing temperature.

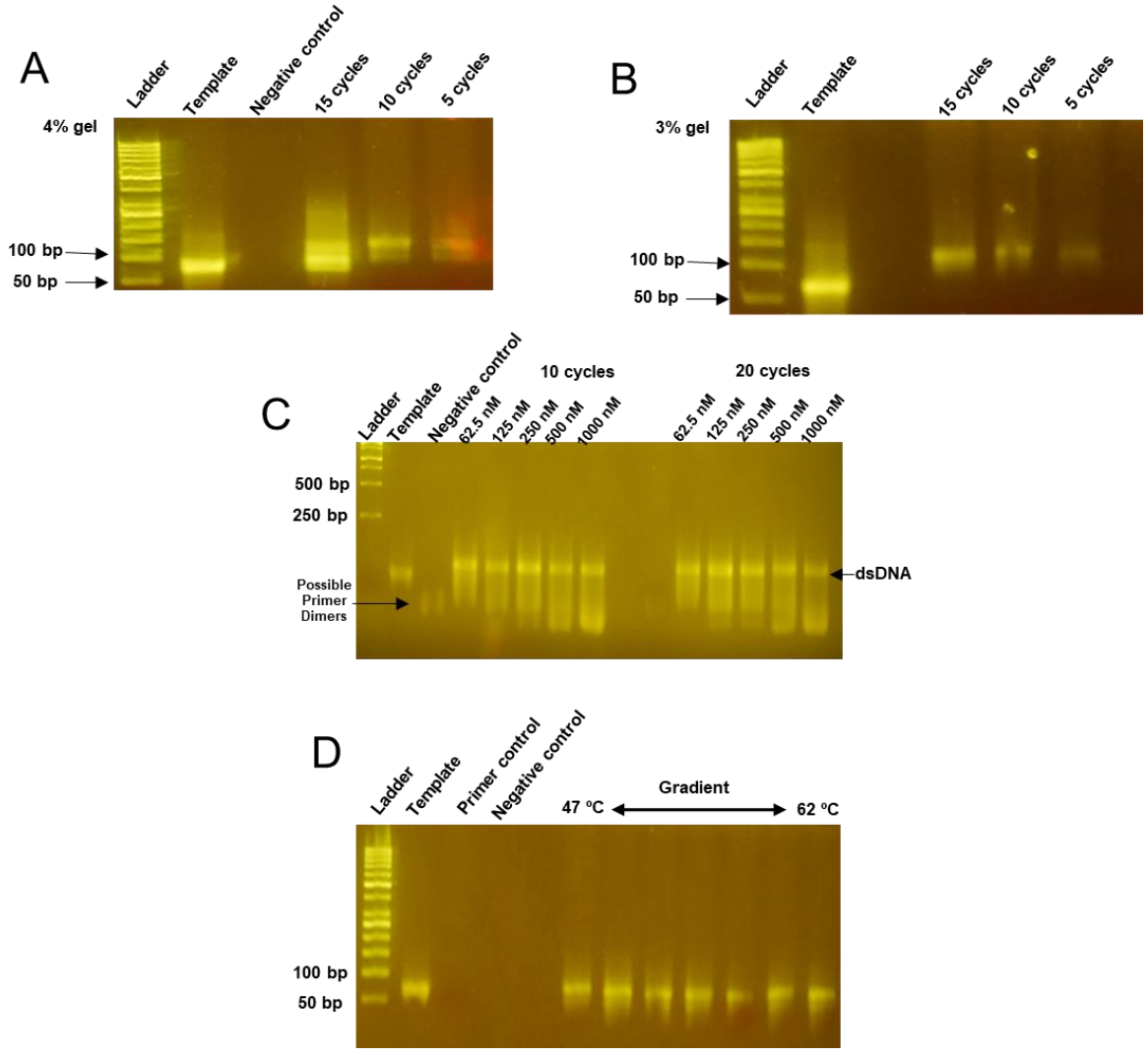


Figure 3.2: Optimisation of PCR amplification of the ssDNA SELEX library sections C-E. A) *4% Agarose gel showing PCR product after 5, 10 and 15 cycles of amplification. B) 3% Agarose gel showing PCR product after 5, 10 and 15 cycles of amplification. C) 3% Agarose gel comparing PCR product with varying primer concentrations after 10 and 20 amplification cycles. D) 3% Agarose gel showing the effect of annealing temperature on PCR product quality. 5 µl of PCR product was loaded in each lane; agarose gels were stained using Midori direct.*

3.3.3 Comparative analysis of optimised library design and PCR conditions against an alternative ssDNA library and PCR method.

Further analysis compared the in-house primer design and optimised PCR conditions against an alternative set of primer sequences and PCR conditions previously used for SELEX from the literature [200]. An alternative ssDNA template with alternative primer binding sites was taken from a SELEX protocol optimised for selection against a protein target [200]. This protocol has been used by independent researchers from the original study to generate aptamers to COVID-19 spike protein for lateral flow assays [201]. Independent, successful use of this protocol should indicate that the primer PCR reaction is robust for SELEX applications and will make a valuable comparison to the in-house primer/library design and PCR optimisation. An 80 mer ssDNA oligo and primers were ordered with priming regions that matched the primers from the previous study and amplified according to the method set out in their publication for 20 cycles (see Figure 3.2A right panel).

Forward primer- GTCTATATGATCTGTAACTC

Reverse primer- ATCTGATGACTCACTGCTGG

Oligo sequence-

GTCTATATGATCTGTAACTCATGGTCCTAATAGCCCTATAGGTACACTGACTAT
AGGCCGCCAGCAGTGAGTCATCAGAT

PCR products generated from the primer design and previously optimised PCR conditions were compared to those produced using the alternative ssDNA template and primer set from the literature after 20 cycles of amplification prior to analysis using both gel electrophoresis and IP RP HPLC (see Figure 3.2B/C). The agarose gel electrophoresis analysis reveals that the template amplified with literature-based primers produced PCR by-products, with two distinct bands between 100 and 150 bp. In addition, the IP RP HPLC analysis indicates four peaks within this size range, attributable to by-products. No PCR by-products were observed in this study's optimised PCR and library design. These findings suggest that the in-house PCR primers and reaction parameters yield superior quality PCR products compared to the primers and PCR conditions taken from the literature, which were optimised for a SELEX protocol against a protein target. Consequently, there is confidence that the in-house primers and PCR reaction conditions are well-suited for this project. It should be noted, however, that as defined sequences were used for this experiment the

quality of PCR product may be poorer in SELEX itself, due to binding of the primers to complimentary sequences in the random regions within the library as well as problems arising from sequences with complex secondary structures.

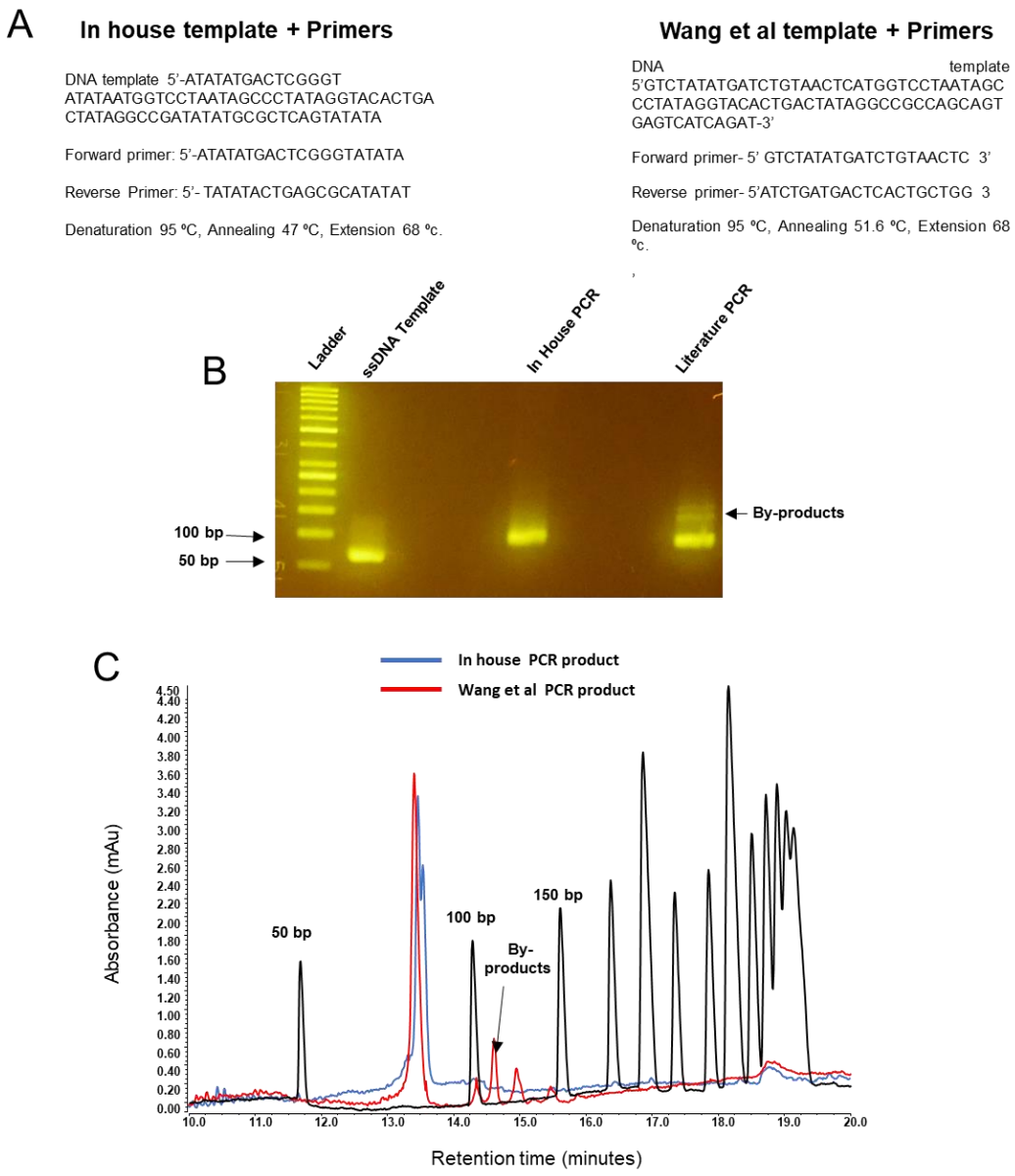


Figure 3.3: Comparative analysis of ssDNA amplification using optimised PCR conditions. *A) Schematic illustration of the PCR workflows used in this study. B) 3% agarose gel electrophoretogram of the PCR products generated from the two different PCR conditions. 5 µl of PCR sample was loaded and gel stained with Midori direct. The PCR by-products (higher molecular weight) DNA fragments are highlighted. C) IP RP chromatogram of the PCR products generated from the two different PCR conditions. 1 µl of PCR products were analysed using mobile phase A, 5 mM TBAA 10% ACN; mobile phase B, five mM TBAA 80% ACN, gradient 7 flow rate 2.2 µl/min at 50 ºC with UV detection at 260 nm. The PCR by-products (larger-sized DNA fragments) are highlighted.*

3.4 Results and discussion Part B: Comparison of alternative methods for the generation of ssDNA in SELEX

3.4.1 Introduction

Following PCR amplification, it is essential to regenerate ssDNA from the duplex PCR products before proceeding to the following selection round. Various methods have been reported in the literature for this purpose (see Chapter 1, section 1.9.1). Before proceeding to the next selection round, the ideal method should provide high yields, high purity, and minimal additional steps, such as desalting. Exonuclease digestion and streptavidin/biotin affinity chromatography have been widely used to generate ssDNA in SELEX and were assessed for their suitability. Initial work focussed on utilising these methods for the generation of ssDNA.

3.4.2 Generation of ssDNA using exonuclease digestion

The exonuclease digestion method for ssDNA isolation requires the production of dsDNA via PCR in which one of the strands has a 5' phosphate group. This labelling renders the complementary strand of the PCR product susceptible to phosphate-dependent exonuclease digestion [107]. PCR products were generated from ssDNA using a 5' phosphate labelled primer before exonuclease digestion (see Figure 3.4A). The results of the initial digestion showed that no obvious ssDNA was generated, which was consistent with the migration of the ssDNA control template. A potential third band below the main product, which could be ssDNA, was observed. From this data, it was concluded that the digestion time should be increased. Therefore, additional exonuclease digests were carried out over 20-40 minutes (see Figure 3.4B). The results show that like the previous experiment, no obvious ssDNA was generated, consistent with the migration of the ssDNA control template. To further investigate, digestion times were extended to two and eight hours to assess whether the enzyme had a lower activity rate. The results of these extended digestions are shown in Figure 3.4C. The results show that the two- and eight-hour digests show signs of digestion, but no ssDNA is generated. It was hypothesised that the enzyme might not have been as specific to 5' phosphate-labelled DNA as initially thought and could have some activity against 5' OH ssDNA. Given the uncertainty surrounding the feasibility of the exonuclease method, it was decided to explore alternative methods for this stage of SELEX.

3.4.3 Generation of ssDNA using streptavidin affinity chromatography

Streptavidin affinity chromatography is commonly utilised for ssDNA generation during SELEX [190]. A 5' biotinylated primer is used in the PCR to generate 5' biotinylated dsDNA, which is subsequently bound to streptavidin magnetic beads. Due to the nature and strength of the biotin-streptavidin interaction, this enables elution of the complementary ssDNA strand via alkaline denaturation. For the initial experiment, PCR amplification using the previously developed optimised PCR conditions was employed with a 5' biotinylated primer. A large-scale binding experiment was conducted to ensure method viability prior to optimisation. Fifteen PCR reactions (total volume 750 µl) were pooled and incubated with 500 µl of streptavidin beads (5 mg beads); the binding, washing, and elution steps were performed according to the protocol described (see Chapter 2, section 2.4.2). Agarose gel electrophoresis analysis of the eluted ssDNA is shown in Figure 3.4D. While electrophoresis demonstrated that the ssDNA generated was pure, it would not show whether some complementary strand was eluted with the desired ssDNA, which is a common problem with this method [105]. Additional, LC/MS analysis of the purified ssDNA is shown in Figure 3.4F. The IP RP HPLC and mass spectrometry data verify the identification of a single significant mass corresponding to the expected mass of the ssDNA. Therefore, these results confirm that the ssDNA generated from the streptavidin/biotin affinity purification is pure and viable for use in SELEX. However, salt removal would be necessary before proceeding with a selection round following the use of this method.

3.4.4 Determination of ssDNA yield using streptavidin affinity chromatography

In the initial purification, large quantities of streptavidin beads and PCR product were used to ensure the generation of a significant quantity of ssDNA. However, using 0.5 ml of beads in each round of a SELEX protocol, which could potentially exceed ten rounds, would not be economical as the beads cost ~£200 per ml. Based on the manufacturer's claim of binding capacity for biotinylated OGNs (Oligonucleotides) to the beads, 6.75 µg PCR product should bind to 70 µl of beads. To ensure complete binding of the dsDNA, 140 µl of streptavidin beads was used. Following streptavidin affinity chromatography and elution of the ssDNA, agarose gel electrophoresis analysis (see Figure 3.4E) was performed in conjunction with nanodrop UV

spectrophotometry to determine the amount of ssDNA eluted (see Table 3.2). The results show that a yield of 50-60% was obtained using this approach based on 6.75 µg of input dsDNA and elution of 1857.3 ng ssDNA from a theoretical yield of 3.377 µg, resulting in a percentage yield of 54.9%. These results are consistent with previous observations using streptavidin affinity chromatography to generate ssDNA [110].

Table 3.2: Determination of ssDNA yield

Fraction	**Volume (µl)**	**A260 nm**	**Concentration (ng/µl)**	**Mass (ng)**
NaOH elution one	50	1.23	35.27	1,763.3
NaOH elution two	50	0.063	1.88	94
NaOH elution three	50	blank	n/a	n/a

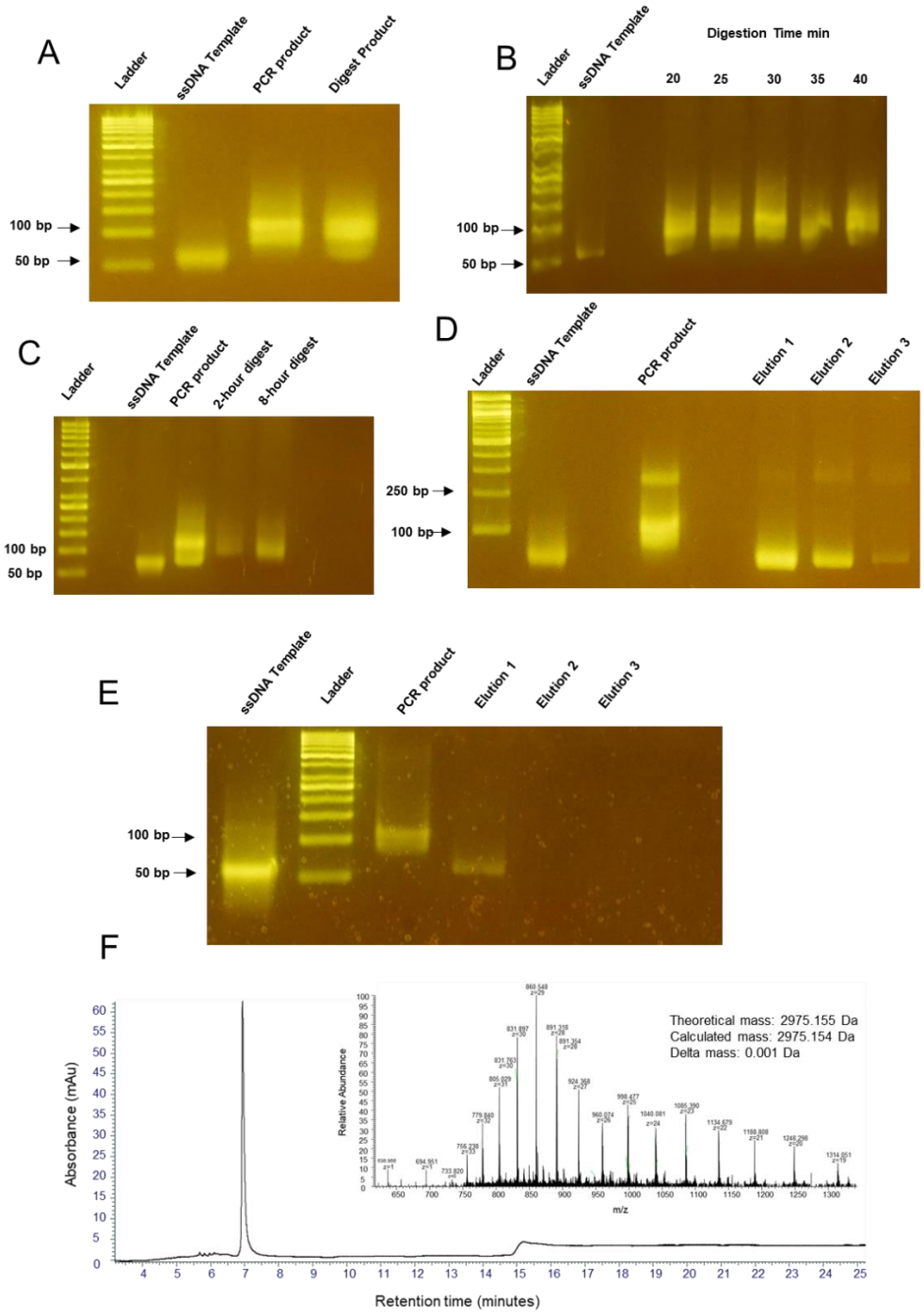


Figure 3.4: Analysis of ssDNA purified using exonuclease digestion and streptavidin/biotin affinity chromatography. *Gel electrophoresis analysis of exonuclease digestion of the dsDNA PCR products from A) 20-minute digestion, B) A range of extended digestion times, and C) 2-8 hr exonuclease digestion time. Gel electrophoresis analysis of ssDNA eluted from streptavidin beads from D) Large scale purification E) Small scale purification. All gel electrophoresis analysis was performed using 4% agarose gels. 10 µl of PCR product or DNA digest was loaded and stained with Midori direct. F) IP RP HPLC UV analysis of purified ssDNA, with ESI MS spectra shown inset. Mobile phase A 0.2% TEA 50 mM HFIP in H_2O, mobile phase B 0.2% TEA 50 mM HFIP in 80:20 H_2O: ACN, Gradient 4, flow rate 0.2 ml/min. The theoretical and calculated monoisotopic masses are highlighted. It should be noted that the exonuclease and streptavidin methods were being worked on before it was discovered that 4% gels denatured the PCR product, so the product bands are smudged*

3.5 Optimisation of denaturing ion pair reversed phase HPLC for the purification of ssDNA in SELEX.

Previous results showed that the generation of ssDNA was inefficient using exonuclease digestion approach in this study. Streptavidin/biotin affinity chromatography enabled the efficient generation of ssDNA, however this method suffers from the inability to distinguish between products and by-products as well as potentially contaminating the next stage of selection with streptavidin if additional purification is not carried out [105]. Therefore, the aim was to develop a novel method for generating ssDNA using IP RP HPLC. Previous work had demonstrated that 5' chemical modifications have a noticeable effect on the retention time of OGNs of the same length [192]. It was hypothesised that use of various hydrophobic 5' modifications on one strand of a DNA duplex would allow for the two stands to be resolved in a denaturing HPLC system enabling purification of the ssDNA.

3.5.1 Retention comparison of 5' modified oligonucleotides using IP RP HPLC

In the initial experiments I aimed to assess the effect of a variety of 5' chemical modifications of oligonucleotides on their retention time when analysed using IP RP HPLC. It was hypothesised that increasing the hydrophobicity of the 5' modification would result in the ability to separate the two strands of the DNA duplex when analysed under denaturing conditions using IP RP HPLC and therefore enable the separation of the two single strands from an 80 bp dsDNA fragment generated using PCR products via the same principle. Initial work focused on the analysis of a range of 5' modifications on the PCR primers including C18, Biotin, Texas Red and a 20-base primer with a phosphorothioate backbone which were selected due to the increased hydrophobicity of phosphorothioate OGNs [202–204].

The modified primers were analysed via both weak IP RP HPLC (triethylammonium acetate) and strong IP RP HPLC (tributyl ammonium acetate). The strength of an ion-pair reagent is defined by its overall hydrophobicity. The alkylamine, TEA (short alkyl chains) is termed a "weak" ion-pair reagent and the more hydrophobic TBA (longer alkyl chains) is termed a "strong" ion-pair reagent, which has a stronger interaction with the stationary phase [205]. Comparison of the retention times of the various 5' chemical modifications on the ssDNA primers are shown in Figure 3.5A/B and demonstrate that the phosphorothioate and 5' Texas Red result in the most

significant difference in retention compared to the unmodified OGN. The results show significant shifts in retention time, reflecting the different OGN hydrophobicity in the order unmodified <C18 < biotin < phosphorothioate < Texas Red. These results are consistent with previous IP RP HPLC analysis of oligonucleotides with hydrophobic fluorophores [206–209]. The phosphorothioate OGN is notably broader when analysed under weak IP RP HPLC due to the large number of diastereoisomers present [202,204]. This may have implications for its use in the denaturing HPLC method if fractionating and collecting the phosphorothioate ssDNA for analysis. Collecting much broader peaks may involve splitting the fractions over multiple tubes and potentially decreasing yield.

To further analyse the effects of the 5’ modifications, analysis was also performed using strong IP RP HPLC utilising the larger alkylamine (tributylamine) (see Figure 3.4B). Under these conditions size-based separations largely dominate and the influence of the 5’ modifications is expected to be reduced [202–204]. The results show a notably reduced difference in retention time of the various 5’ modifications compared to weak IP RP HPLC (Figure 3.5A). However, the results show that a difference in overall hydrophobicity of the OGNs is still observed under strong IP RP HPLC conditions, highlighting that fully size based separations are not achieved. The results also show that peak broadening of the phosphorothioate OGN is not seen in the strong ion pair analysis as the strength of interaction between the analyte column is much stronger hence structural features such as base sequence or structural isomerism have less impact on retention time. Consistent with weak IP RP HPLC analysis the phosphorothioate and 5’ Texas Red modifications result in the most considerable difference in retention compared to the unmodified OGNs.

The advantage of comparing strong and weak IP RP HPLC is that strong IP RP HPLC affords much greater resolution and would allow for a more accurate assessment of the hydrophobicity of each modification, however strong ion pair reagents such as TBA are less volatile and therefore less suitable for the purification of ssDNA.

3.5.2 IP RP HPLC analysis of PCR products amplified using 5’ modified PCR primers

Analysis of the 5’ modified PCR primers via strong and weak IP RP HPLC showed that the phosphorthioate and 5’ Texas Red modifications had the most significant influence on the hydrophobicity of the oligonucleotides and subsequently lead to the greatest shift in retention time relative to the unmodified primer. Further experiments were performed to generate double stranded DNA (dsDNA) to demonstrate that using the 5’ chemically modified primer (reverse) did not decrease the yield of the PCR reaction or generate by-products, as this would prevent it from use in a SELEX protocol. PCR products were generated from each primer and analysed via agarose gel electrophoresis (see Figure 3.5C). In addition, the effect of the 5’ modification on the overall hydrophobicity of the dsDNA PCR product was also investigated using IP RP HPLC (see Figure 3.5D).The results in Figure 3.4 C/D show the successful PCR amplification of the dsDNA using different 5’ modified primers. No significant difference in the yield or purity of the PCR products was observed in comparison to the unmodified control, demonstrating that in principle they are suitable for the PCR amplification stage of SELEX. It was expected that the contribution of the 5’ modification on the overall hydrophobicity would be reduced in larger PCR products, compared to shorter primer sequences due to the increased overall hydrophobicity of the longer dsDNA. The results show that consistent with IP RP HPLC analysis of the modified primers (see Figure 3.5A/B) the 5’ phosphorothioate and Texas Red modifications cause the greatest shift in retention time due to their hydrophobicity. A broader peak for the 5’- phosphorothioate dsDNA was also observed due to the presence of the diastereoisomers consistent with previous observations.

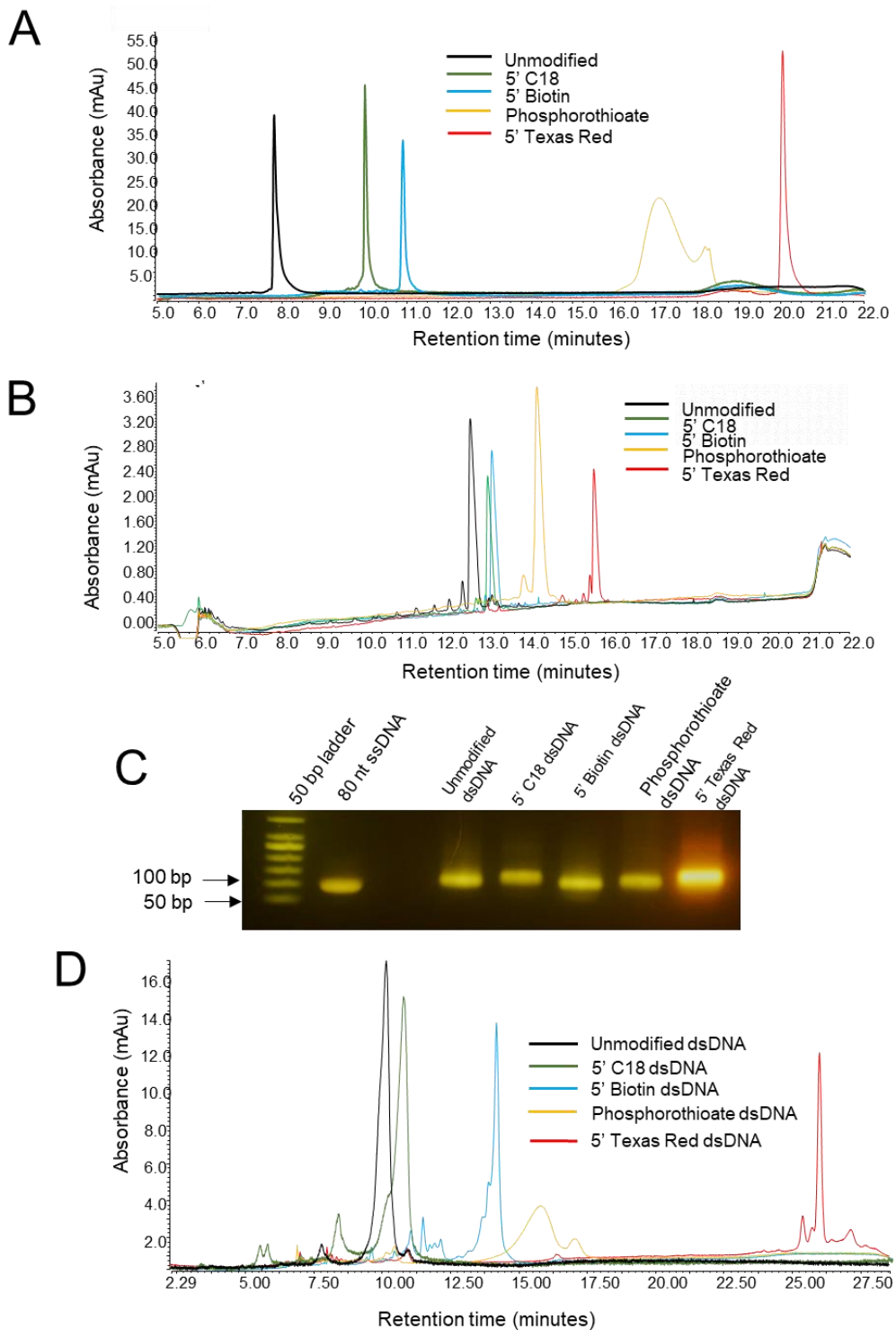


Figure 3.5: IP RP HPLC analysis of 5' modified primers and PCR products. *A) Weak IP RP HPLC analysis. Mobile phase A 0.1 M TEAA; mobile phase B 0.1M TEAA, 25% ACN. Gradient 1, flow rate 0.3 ml/min at 40 °C. 5 pmol primer analysed on a DNAPac RP column. B) Strong IP RP HPLC analysis. Mobile phase A 5 mM TBAA; mobile phase B 5 mM TBAA, 80% ACN. Gradient 2, flow rate 2.2 µl/min at 50 °C, 3 pmol primer injected, separation performed on a monolithic PS DVB capillary column. C) Agarose gel electrophoresis analysis, 50 ng PCR product was analysed for each dsDNA PCR product. Staining was performed using Midori Green Direct. D) IP RP HPLC chromatogram of dsDNA PCR products. Mobile phase A 0.1 M TEAA; mobile phase B 0.1M TEAA in 25% ACN. Gradient 3, flow rate 0.3 ml/min at 40 °C UV detection at 260 nm Chapter 2 section 2.5.3.*

3.5.3 Denaturing IP RP HPLC analysis of 5’ modified dsDNA

Further studies were performed using IP RP HPLC under denaturing conditions to denature the duplex dsDNA with the aim to resolve the single stranded DNA based on differences in their hydrophobicity due to the 5’ chemical modification. Initial work focussed on the analysis using denaturing IP RP HPLC in conjunction with TEAA as the mobile phase (see Figure 3.6A and Appendix section 7.1). The results for the 5’ Texas Red modified dsDNA analysed under denaturing conditions (95 °C) show the separation of the two single strands, with the 5’ Texas Red labelled strand eluting later due to the increased hydrophobicity.

To confirm the identity of the peaks and corresponding ssDNA, the IP RP HPLC was interfaced with high resolution accurate mass spectrometry (see Figure 3.6B/C). The results in Figure 3.6 confirm the 2 peaks are the single stranded DNA and the difference in mass was used to confirm that the later eluting species was the 5’ Texas Red modified ssDNA. Initial mass deconvolution of the two fragments generated accurate masses that did not match the expected theoretical mass of the DNA due to the addition of adenine to the 3’ end of each fragment via *Taq* DNA polymerase in the PCR reaction as described by McCarthy *et al* [210]. Taking into account the addition of 3’ adenine to the DNA products led to parity between the theoretical and experimental mass for each ssDNA fragment. LC-MS analysis of all the corresponding 5’ modified dsDNA under non-denaturing and denaturing conditions is shown in Appendix section 7.2.

Following separation and validation of the ssDNA using denaturing IP RP HPLC, further work was performed to isolate and purify the ssDNA. 2 µg of 5’ modified Texas Red dsDNA was separated, and the unmodified ssDNA was fraction collected from the HPLC prior to freeze drying. Following purification of the ssDNA, further analysis was performed using IP RP HPLC and agarose gel electrophoresis (see Figure 3.7). The results in Figure 3.7 show the successful isolation and purification of ssDNA from dsDNA using IP RP HPLC under denaturing conditions which was validated using LC MS.

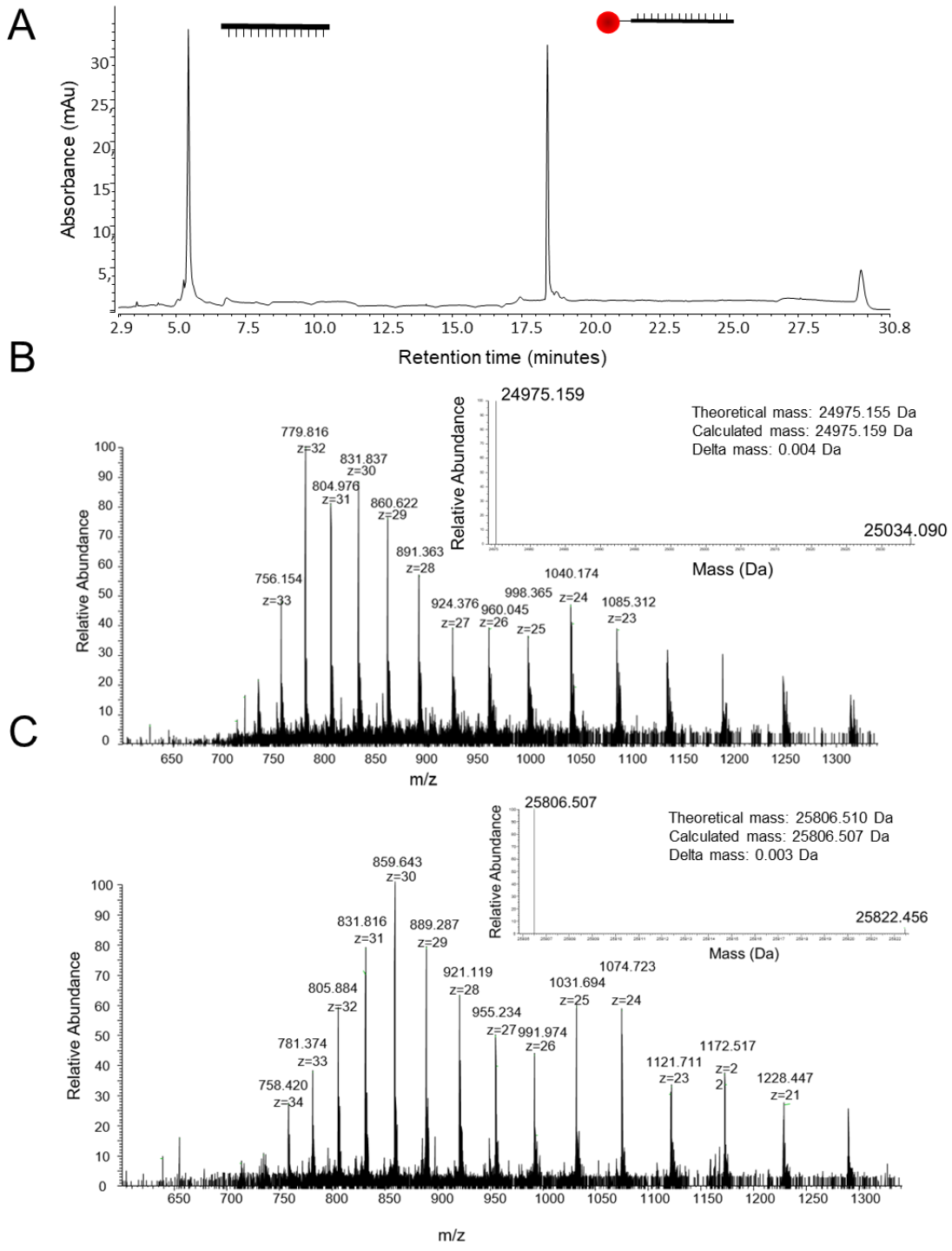


Figure 3.6: Denaturing IP RP HPLC analysis. *A) Texas Red labelled dsDNA was analysed using weak IP RP HPLC at 95 °C, UV detection 260 nm. Mobile phase A 0.2% TEA, 50 mM HFIP; mobile phase B 0.2% TEA 50 mM HFIP in 25% ACN analysed on a DNAPac RP column. Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. B) ESI MS spectra unlabelled ssDNA. The theoretical and calculated monoisotopic masses are highlighted. C) ESI MS spectra of Texas Red labelled ssDNA. The theoretical and calculated monoisotopic masses are highlighted. Deconvoluted masses are highlighted within the ESI MS spectra shown in the inset.*

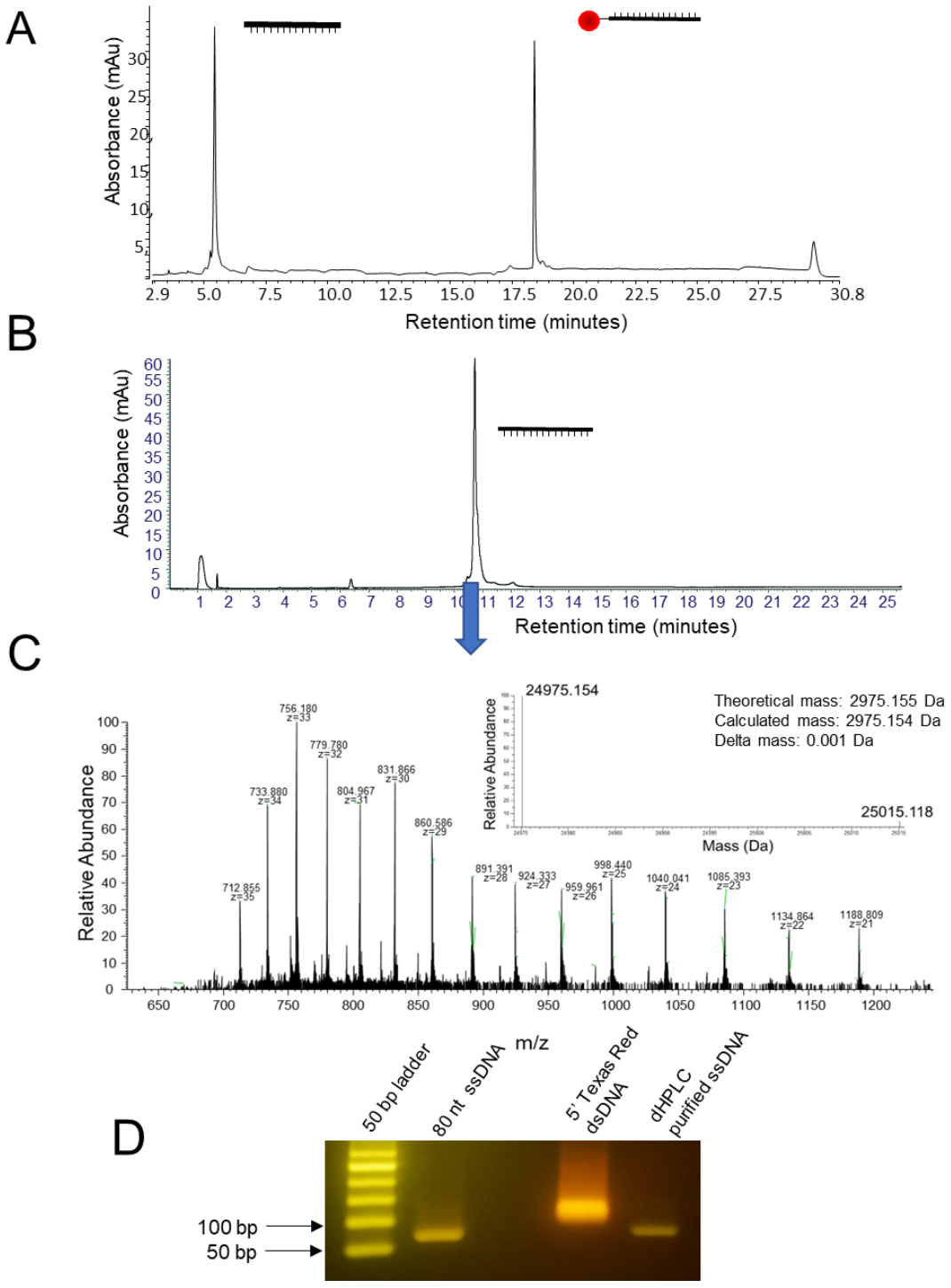


Figure 3.7: Denaturing IP RP HPLC purification of ssDNA. *A) IP RP chromatogram of the Texas Red labelled dsDNA at 95 °C, UV detection 260 nm. Mobile phase A 0.2% TEA, 50 mM HFIP; mobile phase B 0.2% TEA 50 mM HFIP in 25% ACN. Gradient 4 flow rate 0.2 ml/min, DNAPac RP column Chapter 2 section 2.5.3. B) LC UV chromatogram of the purified unlabelled ssDNA at 40 °C, UV detection 260 nm. Mobile phase A 0.1 M TEAA, 50 mM HFIP; mobile phase B 0.1M TEAA 50 mM HFIP in 25% ACN. Gradient 4 flow rate 0.2 ml/min. C) ESI MS spectra of the purified unlabelled ssDNA. The theoretical and calculated monoisotopic masses are highlighted. Deconvoluted masses are highlighted within the ESI MS spectra shown in the inset. D) Agarose gel electrophoresis analysis of the dsDNA and purified ssDNA. Staining was performed using Midori Green Direct.*

3.5.4 Optimisation of denaturing IP RP HPLC conditions for the separation of ssDNA

Previous work had demonstrated that using TEAA as the mobile phase required the use of high temperatures (95 °C) to completely denature the dsDNA PCR product. Furthermore, it was noted that at lower temperatures it was not possible to denature the two DNA strands (see Figure 3.7A). G rich sequences are often identified or utilised in aptamer libraries, due to their ability to form G-quadruplexes which are desirable secondary structure elements of aptamers [211]. Therefore, to ensure that efficient denaturation was possible even for potential G rich sequences an alternative mobile phase system was investigated to enable more efficient denaturation of the dsDNA at lower temperatures. Initial work focused on the addition of 1,1,1,3,3,3,-Hexafluoro-2-propanol (HFIP) to the TEAA mobile phases. The addition of HFIP to the mobile phases has been shown to have a denaturing effect on duplexes, while also reducing the effect of oligonucleotide hydrophobicity upon retention [210]. The results show that under these conditions effective denaturing was achieved at lower temperatures (80 °C) compared to previous high temperatures required using TEAA mobile phase (see Figure 3.8B). In addition, an alternative mobile phase system was employed using TEA/HFIP in conjunction with denaturing IP RP HPLC. Under these mobile phase conditions it was noted that the PCR products denature at much lower temperatures in comparison to TEAA. This lower melting temperature is due to the chaotropic nature of HFIP, disrupting the internal hydrogen bonds of the duplex.

The results show that the duplex dsDNA remains intact only at lower temperatures typically 30 °C and elevated temperature above this effectively denature the dsDNA. Therefore, effective denaturation using this mobile phase was observed at 40 °C (see Figure 3.8D). These results demonstrate that employing TEA/HFIP mobile phase or simply the addition of HFIP to the TEAA mobile phase enables effective denaturing of the dsDNA without the requirement for very high temperature and the requirement of specialised columns ovens that can operate at such temperatures. Furthermore, the ability to denature the dsDNA more effectively at lower temperatures is beneficial for the isolation of ssDNA from PCR products with high melting temperatures and high GC content.

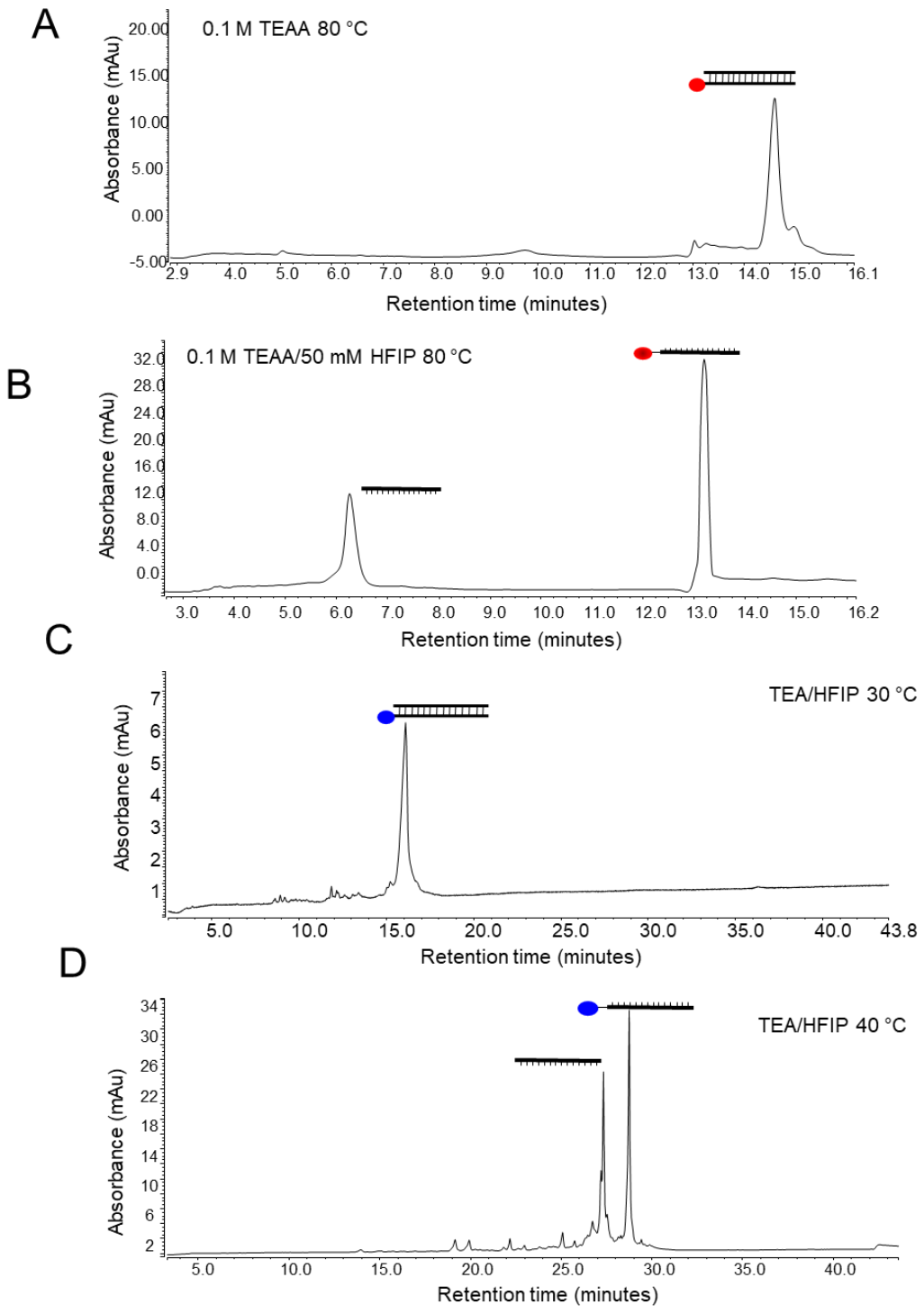


Figure 3.8: Optimisation of mobile phases for the denaturation of dsDNA. *A) IP RP chromatogram of the Texas Red labelled dsDNA at 80 °C mobile phase A 0.1 M TEAA, mobile phase B 0.1M TEAA 25% ACN. B) IP RP chromatogram of the Texas Red labelled dsDNA at 80 °C mobile phase A 0.1 M TEAA 50 mM HFIP, mobile phase B 0.1M TEAA 50 mM HFIP 25% ACN. Gradient 7, flow rate 0.6 ml/min. The resulting ds and ssDNA species are highlighted. C) Denaturing IP RP HPLC analysis of C18 labelled PCR product mobile phase A 0.2% TEA, 50 mM HFIP, mobile phase B 0.2% TEA, 50 mM HFIP, 0.2 ml/min, 25% ACN at C) 30°C Gradient 4 D) 40°C Gradient 8 Chapter 2 section 2.5.3. The resulting ds and ssDNA species are highlighted. Analysis was performed on a DNAPac RP column.*

3.5.5 ssDNA library elution via denaturing HPLC

For application of the method in a SELEX protocol it must be successfully applied to the isolation of ssDNA from the PCR products of a SELEX library containing a random region of DNA. This presents a greater challenge compared to the non-randomised DNA template previously used, as the DNA peaks observed on the IP RP HPLC will likely be broader because of the random region and further complicated by the presence of PCR by-products which are often generated in the PCR of DNA during SELEX.

Following the successful purification of ssDNA using denaturing IP RP HPLC, the method was adopted for the analysis and purification of ssDNA generated from a random ssDNA library (N40) used in SELEX protocols. Texas Red labelled PCR product was generated from the random ssDNA library (N40) and analysed via denaturing HPLC (see Figure 3.9). The results show the separation of the two ssDNA strands as previously observed for the non-random DNA. A broader peak for the non-labelled ssDNA was observed as expected due to the presence of the random region (N40) and a large number of different sequences present. Furthermore, a potential PCR impurity (ssDNA of a different size) is also observed in the chromatogram. The non-labelled ssDNA was fraction collected and freeze-dried before analysis using agarose gel electrophoresis and IP RP HPLC (see Figure 3.8 B/C). The results show the successful purification of the ssDNA (N40) and furthermore demonstrates the purification of the ssDNA from potential larger length impurities that were generated in the PCR, which often hamper SELEX methods. In this case mass spectrometry analysis cannot be used to verify the sequences of the two ssDNA following denaturation of and separation of the ssDNA due to the random sequence nature. However, the presence of the 5’ Texas red group allows for fluorescence analysis of both the denatured PCR product and the purified ssDNA (see Figure 3.10). The results in Figure 3.10 show that as expected the later eluting ssDNA gave the corresponding fluorescence, consistent with previous analysis. In addition, no fluorescence signal was observed for the purified ssDNA which lacks the Texas Red label.

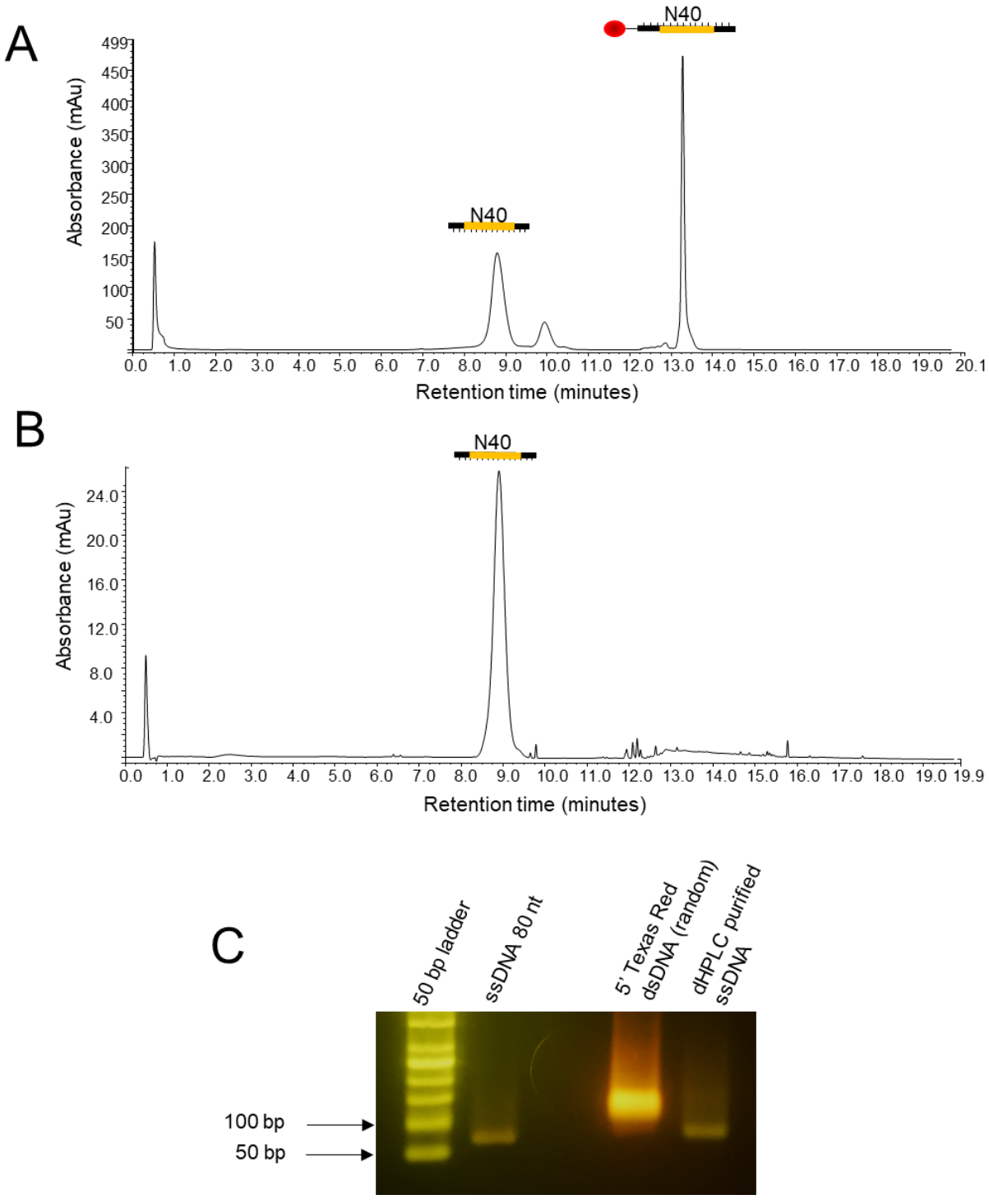


Figure 3.9: Purification of ssDNA generated from a SELEX library. *A) IP RP HPLC analysis of dsDNA generated from a SELEX library (random region N40) under denaturing conditions (80 °C). B) IP RP HPLC chromatogram of the purified ssDNA (N40). Mobile phase A 0.1 M TEAA 50 mM HFIP, mobile phase B 0.1M TEAA 50 mM HFIP 25% ACN Gradient 5 flow rate of 0.6 ml/min UV detection 260 nm Chapter 2 section 2.5.3. 100 ng of dsDNA and 100 ng ssDNA were injected. C) Agarose gel electrophoresis analysis of the dsDNA and purified ssDNA. Staining was performed using Midori Green Direct. IP RP HPLC analysis was performed using a DNAPac column.*

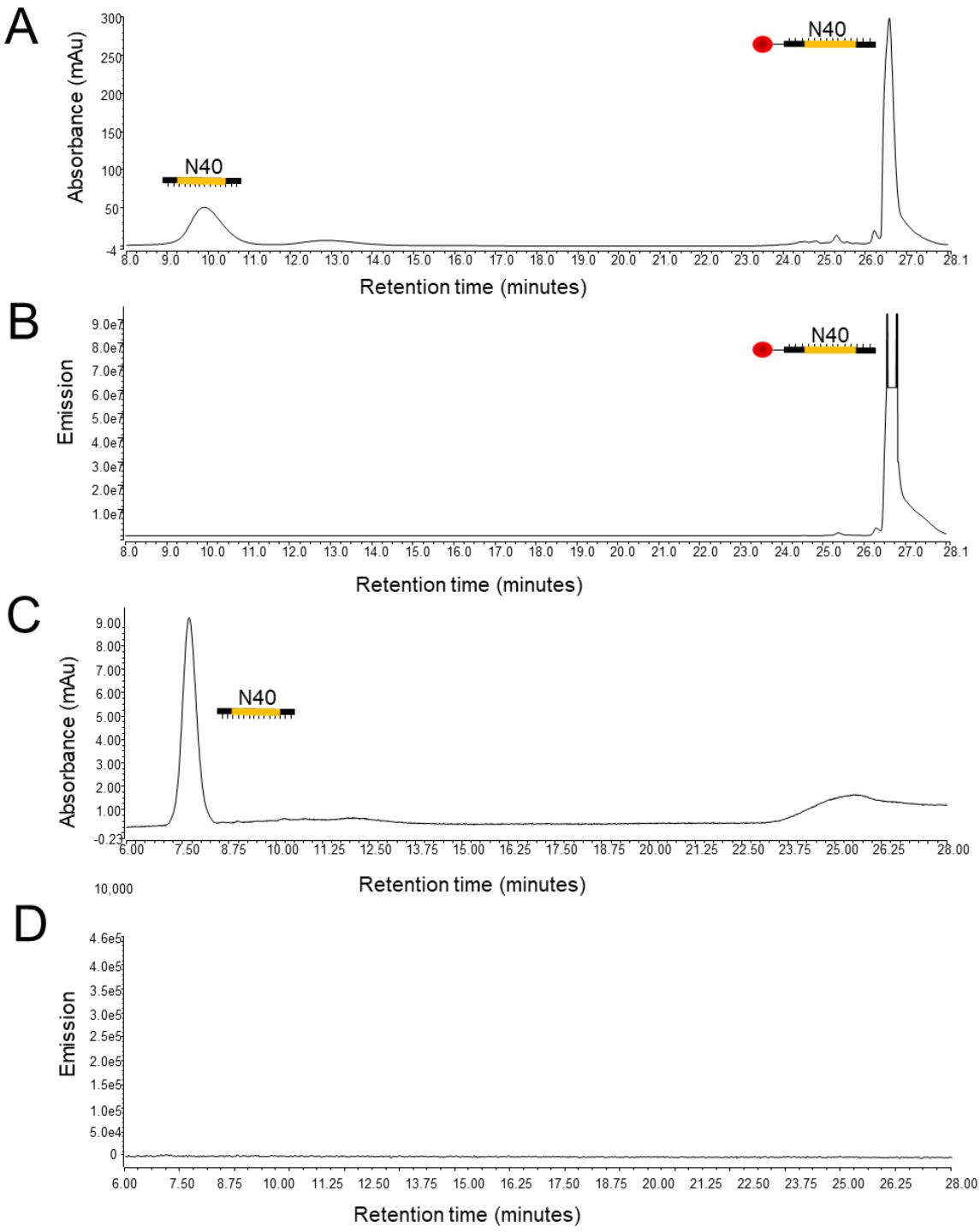


Figure 3.10: IP RP HPLC analysis in conjunction with fluorescence detection.
Denaturing IP RP chromatogram of the Texas red labelled dsDNA (N40). Gradient 6, 0.3 ml/min Chapter 2 section 2.5.3. A) UV detection 260 nm. B) Fluorescence detection: excitation 565 nm, emission 615 nm. Denaturing IP RP HPLC chromatogram of the purified ssDNA (N40) C) UV detection 260 nm D) Fluorescence detection: excitation 565 nm, emission 615 nm. IP RP HPLC analysis was performed using a DNAPac RP column.

3.5.6 Purification and fractionation of ssDNA using IP RP HPLC

Following purification of the ssDNA from SELEX PCR products using denaturing IP RP HPLC it is crucial to ensure that there is minimal sample loss during the removal of the mobile phase prior to re-suspension and folding of the ssDNA in the SELEX binding buffer. The criteria for the desalting method is that it must effectively remove salts that may interfere with the selection stage, must not be too laborious and must not significantly impact yield. Four alternative methods were assessed and the % recovery of the ssDNA were evaluated, including ethanol precipitation, capture on silica coated magnetic beads and removal of the volatile mobile phases using either rotary evaporation or freeze drying [212] [213].

In each case, an equal mass of the ssDNA was collected via HPLC, and the % recovery of the ssDNA was determined for each of the different methods. The results in Figure 3.11 show that the highest yield recovered was obtained using freeze-drying (80.8 ± 10.6%). The freeze-drying method was taken through to the higher scale, comparison of the yield of the 2 µg and 8 µg scales resulted in similar yields of the ssDNA. Therefore, these results demonstrate that high recovery and yields of ssDNA can be obtained using denaturing IP RP HPLC in conjunction with freeze drying.

Furthermore, such approaches enable purification of the correct size single stranded DNA from potential PCR artefacts commonly generated in SELEX procedures. Verification of the ssDNA obtained was confirmed by both mass spectrometry and fluorescence analysis. Therefore, these results demonstrate that higher yield of ssDNA is obtained using denaturing IP RP HPLC optimised in this workflow in comparison to streptavidin-biotin affinity chromatography.

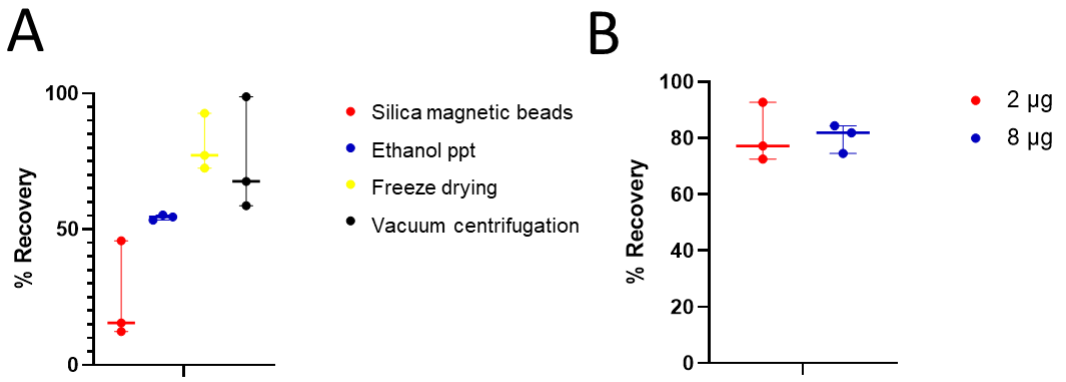


Figure 3.11: Comparison of DNA purification methods. *A) % recovery was determined using 2 µg of ssDNA purified using silica coated magnetic beads, ethanol precipitation, freeze drying, and vacuum centrifugation (n=3). B): % recovery of 2 µg and 8 µg ssDNA using freeze-drying (n=3).*

3.6 Conclusions

One of the main challenges when performing successful SELEX procedures to generate ssDNA aptamers is the ability to generate high-quality ssDNA in high yields from the dsDNA PCR products. In this chapter, I have developed and optimised critical aspects of the SELEX PCR amplification. Optimisation of the library design in conjunction with PCR conditions was performed. Following optimisation, a comparison against an alternative ssDNA library and PCR conditions from the literature was performed. The results showed an improvement in the dsDNA quality from the PCR conditions I optimised, and the primer sequences used in the design compared to those taken from a successful SELEX application from the literature. In addition, further comparative analysis of alternative methods of ssDNA generation was performed, including exonuclease digestion and streptavidin affinity chromatography. Several caveats were identified in the above methods, including low yields and the purification of ssDNA by-products resulting from the PCR. Therefore, I developed and optimised a novel denaturing IP RP HPLC for the purification of ssDNA, demonstrating significant advantages over existing affinity chromatography and nuclease digestion methods. PCR amplification was performed utilising a PCR primer containing a variety of hydrophobic tags (5' Texas Red, C18, Biotin and phosphorothioate modifications), which enables the separation of the ssDNA based on differences in their hydrophobicity. Optimal separations were

achieved using a 5’ Texas Red labelled PCR primer to amplify DNA containing a random (N40) region. Furthermore, optimisation of the mobile phase conditions was also performed using the addition of HFIP to the mobile phases, which enables complete denaturation of the dsDNA at moderate temperatures and circumvents the high temperatures often required using TEAA. Validation of the ssDNA obtained was performed using high-resolution, accurate mass spectrometry analysis and fluorescence-based HPLC detection. High yields and recovery (80%) of the ssDNA were obtained using freeze-drying to remove the volatile mobile phase. Furthermore, this approach enables purification of the desired ssDNA of the correct size, removing potential DNA artefacts often generated in the PCR amplification during SELEX procedures, therefore demonstrating advantages over current affinity chromatography and nuclease digestion methods.

4 Chapter 4 *In vitro* Selection of ssDNA Aptamers against Syntenin-1/MDA-9

4.1 **Abstract**

According to the World Health Organization, the ageing of the global population is set to lead to an increasing cancer burden on global societies. Current cancer treatments are associated with poor outcomes, including off-target effects and cancer resistance. Aptamers present an alternative class of affinity ligands, with their ability to interfere with the metabolic pathways of their targets, a notable example of this being AS1411, an aptamer to nucleolin, demonstrating cancer cell cytotoxicity both *in vitro* and *in vivo.* This chapter aimed to generate ssDNA aptamers against the target Syntenin-1 (MDA-9). MDA-9 was overexpressed in *E. coli* and purified before performing SELEX to select for ssDNA aptamers binding to MDA-9. A random ssDNA library (N40) was used in the SELEX protocol, utilising a novel denaturing HPLC method to generate ssDNA at each SELEX round. Initial attempts at the SELEX method resulted in low-quality PCR products and no significant increase in the % of ssDNA binding to MDA-9. Therefore, alternative SELEX methods were developed, with three rounds of selection of ssDNA aptamers without PCR amplification and three rounds of selection with amplification. The resulting pool was sequenced using NGS, and sequences of interest were identified using Galaxy tools and FASTAptamer. Using this optimised SELEX protocol, no significant decline in the quality of PCR product was observed. However, PCR performed during the sequencing workflow generated poor-quality products, resulting in lower-quality sequencing data.

Twelve ssDNA sequences identified from the sequencing analysis were synthesised before determining their binding affinity to MDA-9 using ITC. The results showed that three different aptamer sequences had binding affinities in the 25-600 nM range. These results demonstrate the successful isolation of ssDNA aptamers against MDA-9. Future studies will include enhancing the affinity of these aptamers for MDA-9 by removing the primer regions, further truncation of the sequence and incorporation of modified nucleotides to enhance *in vivo* stability. Following the enhancement of affinity, a study demonstrating the inhibition of MDA-9's function in vitro would provide insight into any potential in vivo application.

4.2 Introduction

Aptamers are incredibly versatile molecules and have been shown to bind to a plethora of targets with high affinity; the issue, however, is the lack of a universal, well-developed SELEX methodology. While the field has advanced much in the last three decades, its multi-disciplinary nature often means each target is approached differently without well-established protocols. The development of SELEX methods for protein targets will involve the utilisation of the innovations of the last three decades where possible. There is plenty of evidence regarding protein SELEX that a well-designed SELEX protocol can generate aptamers that show in vitro and in vivo binding.

4.2.1 Cancer in the 21st Century

The World Health Organisation (WHO) predicts that cancer is now a greater cause of mortality than coronary heart disease, with the cancer burden on populations expected to increase over the coming decades; in 2012, cancer accounted for 8.2 million deaths worldwide [214]. Common cancer treatments include radiotherapy and chemotherapy, and while they have shown some success in reducing mortality rates, they are plagued by side effects that often cause chronic pain in patients [215]. Difficulties in treating cancer come from insufficiently selective treatments, being unable to distinguish tumours from healthy tissues, as well as drug resistance and cancer metastasis [216].

The cellular changes that lead to cancer development occur within the genome and involve the aberrant overexpression of genes in the fledgling tumour tissue compared to the healthy tissue that it originated from [217]. However, this overexpression only occurs for a fraction of the genes in the genome, with the majority not being overexpressed relative to the healthy cells [218]. The implication of this is that cancer does not randomly express genes in the hope of proliferation, as this would be a waste of vital resources. The proteins these selectively expressed genes code for must be critical to many physiological processes that are key to cancer survival.

4.2.2 Syntenin (MDA-9)

Originally identified in the late 1990s as a syndecan-binding protein, Syntenin is a PDZ domain-containing protein that binds the cytoplasmic C-terminal FYA motif of the syndecans; it is a widely distributed cytosolic protein that performs several essential

functions, including recruiting other proteins to form complexes [219][220][221]. The multi-domain protein consists of 298 amino acids and four domains- N-terminal, PDZ1, PDZ2 and a C-terminal domain (Figure 4.1)[222]. PDZ domains bind proteins to the terminus of other proteins, allowing for protein-protein complex formation and enabling the control of complex cellular functions [222]. Its role in cancer metastasis and angiogenesis has led to its selection as a target in this project.

Figure 4.1: MDA-9/Syntenin-1: *Four domain structures, two PDZ domains (red) flanked by N-terminal and C-terminal domains https://www.rcsb.org/structure/1N99.*

MDA-9 and Cancer

Cancer metastasis, the ability of cancer cells to migrate from their native tissues and spread to other parts of the body, is the leading cause of mortality in patients with the disease [223]. In fact, in the case of melanoma, the one-year survival rate of chemotherapy patients whose cancer has metastasised is less than 30% [224]. MDA-9 is upregulated in several skin, head and neck cancers [225]. In the case of melanoma, MDA-9 acts as an essential scaffold that enhances the assembly of protein complexes, which actively increases metastatic competence in human melanoma cells [226].

Angiogenesis is a crucial stage in tumour development; in its early stages of growth, a tumour is supported by diffusion from surrounding tissues. However, when it reaches a specific size that is no longer sufficient to allow further growth, the tumour releases chemical signals such as VEGF vascular endothelial growth factor, which promotes the sprouting of blood vessels from nearby vasculature, essentially enabling the tumour to develop its own vasculature system, driving its growth [227]. In vitro and in vivo studies confirm the role of MDA-9 in the expression of various angiogenic factors, thus promoting angiogenesis [228].

Aptamers to protein targets have been generated with binding affinities in the picomolar to micromolar range [229]. However, what makes an aptamer useful is more than just considering binding affinity. In the case of an enzyme, this is relatively easy to define: something that inhibits it is function. An aptamer with a Kd (600 nM) effectively inhibited the function of methyl transferase 1 [230]. This counters the idea that a very low Kd strong binding aptamer is a 'better' aptamer, both the suitability for its operation environment (in vitro/in vivo) and the effect on the target of interest that defines it is suitability for an intended purpose. In the context of aptamers as anti-cancer therapeutics, they must be sufficiently large to conceal their target protein, thereby acting as an antagonist that blocks their interaction with protein counterparts or substrates [229]. In the case of the target of this project, Syntenin-1, it is a complex builder; an aptamer must impede this through steric hindrance of the binding sites (PDZ domains) to its syndecan targets, and it may not need such a low Kd value (sub 100 nM) to do so.

Previous studies have demonstrated that peptide inhbitors of Syntenin-1 bind to and 'block' the PDZ domains, leading to the limitation of cancer cell in vitro, with Kd's ranging from 0.8-98 nM when used to simultaneously inhibit PDZ domains via blocking the domain pocket which dominates the interaction of MDA-9 with ligands. Any aptamers generated would ideally bind to these domains to achieve the same outcome [231].Following the development and optimisation of the SELEX methodology outlined in Chapter 3, this chapter aimed to generate ssDNA aptamers against MDA-9. The MDA-9 protein was expressed in a suitable host; *E. coli* was selected due to its ease of growth and suitability for recombinant protein expression [232]. The gene for MDA-9 was cloned into a plasmid vector with a histidine tag at the C terminus to allow for

affinity purification following expression. In addition, this will enable the protein to be bound to Ni-NTA magnetic beads for the selection stage.

The methods developed in Chapter Three were utilised in the SELEX method to generate aptamers against MDA-9. NGS sequencing was performed on the resulting ssDNA library against MDA-9, and bioinformatic analysis was used to identify potential aptamer sequences. Finally, analysis of several potential ssDNA aptamers against MDA-9 were synthesised before further analysis using ITC to determine their binding affinity for MDA-9.

4.3 Results and Discussion

4.3.1 Expression and purification of MDA-9 in *E. coli.*

To overexpress the MDA-9 protein in *E. coli,* the MDA-9 gene sequence (Human, isoform 1, Uniprot) was codon optimised, synthesised and cloned into the plasmid pET151 D-TOPO vector by ThermoFisher to generate the plasmid pMDA-9. The design of the pET151 vector includes a His tag at the N terminus of the inserted MDA-9 gene to enable purification of the MDA-9 using Ni-NTA affinity chromatography (see Figure 4.2A [233]. The complete gene sequence used and cloned is shown in Appendix 12.

The plasmid was transformed into competent DH5-α *E. coli* cells (see Chapter 2, section 2.11). The cells were streaked onto an ampicillin agar plate and incubated overnight at 37 °C. A single colony was selected and grown in a LB media overnight before extraction and purification of the plasmid DNA and analysis using gel electrophoresis (see Figure 4.2B). The results confirm the successful cloning of the plasmid DNA into DH5-α *E. coli* cells. In addition, restriction digestion in conjunction with agarose gel electrophoresis was used to confirm the identity of the plasmid DNA (see Figure 4.2C). Linearization of the pMDA-9 with SphI (single cut) generates the corresponding correct size DNA (plasmid size 6654 bp), and the digest with XbaI and SphI generates the expected size of DNA fragments (6399 and 255 bp, respectively).

4.3.2 Overexpression and purification of MDA-9

After successfully cloning the plasmid pMDA-9, it was transformed into competent BL21 *E. coli* cells prepared as previously described. Following overnight growth, a 1% inoculum was used in 100 ml LB media. Cells were grown to OD 0.6 and induced with IPTG prior to a further four hours of incubation. The cell culture methodology and protein expression details can be found in Chapter 2, section 2.12. Cells were harvested before cell lysis, and protein purification was performed via Ni-NTA affinity chromatography using HisPur™ Ni-NTA Resin (ThermoFisher).

Analysis during protein expression and purification was performed using SDS PAGE (see Figure 4.2 D/E). The SDS PAGE results show the purification of a single protein to a high purity in the expected size range (36 kDa). To confirm the successful expression and purification of MDA-9, gel trypsin digestion in conjunction with LC MS/MS was performed (see Chapter 2, section 2.13). Data analysis was performed using the Mascot search engine for protein identification, using the reference protein sequence database for *Escherichia coli* (strain K12) (UniProt), 4448 sequences, downloaded 03032022). The recombinant His tagged human Syntenin (MDA-9) protein sequence was included as an additional entry to this reference proteome. The results from the mascot software analysis are shown in Figure 4.2C, which shows the identification of Syntenin with high (89%) sequence coverage, validating the successful expression and purification of MDA-9 in *E. coli*.

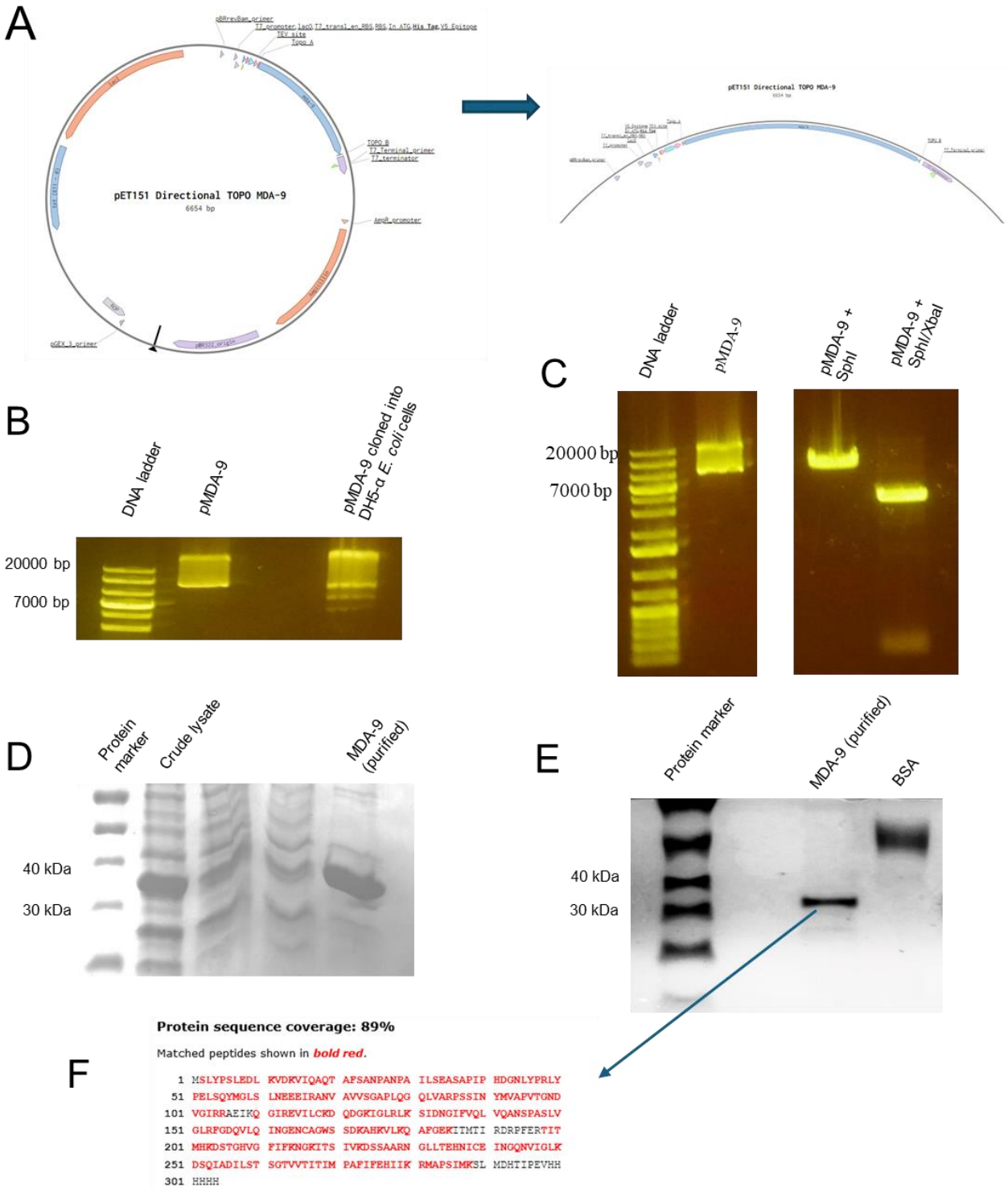


Figure 4.2: Cloning and overexpression of MDA-9 into E. coli. DH5-α cells. *A) Plasmid map (Genohub) of pET151 vector with MDA-9 gene including N-terminal 6x histidine tag. B) Gel electrophoresis analysis of pMDA-9 from the manufacturer stock and following cloning into DH5-α cells. C) Gel electrophoresis analysis of pMDA-9 following restriction digestion with XbaI, SphI and XbaI + SphI. D/E) SDS PAGE analysis following overexpression of MDA-9 in E. coli BL21 cells and purification using Ni-NTA affinity chromatography, 5 µl protein ladder was loaded MW of the 40 kDa and 30 kDa proteins are highlighted. E) Mass Spectrometry verification of MDA-9 following in gel tryptic digestion. A sequence coverage map for MDA-9 is shown with identified regions/peptides in red.*

4.3.3 Generating ssDNA aptamers against MDA-9

SELEX method -1

As discussed in Chapter 1, a protein target can be immobilised in many ways to facilitate selection. This study used MDA-9 bound to Ni-NTA magnetic beads as this approach enables rapid isolation, has low non-specific binding and has successful application in previously published SELEX methods [234,235]. Fellows et al. (2020) demonstrated that an optimised magnetic bead-based SELEX method generated aptamers in as little as three rounds; this minimises the risk of PCR parasite formation, which was a significant concern if high numbers of rounds are required to enrich the library [236]. For the SELEX protocol, the library designed in Chapter 3 consisted of 80 nts in length, with a central random region of 40 nts in length flanked by two 20 nt random regions.

The initial protocol adopted a conventional approach to selection, as summarised in Figure 4.3A. Briefly, 74 µg (3 nmol) of the library was heated to 95 ºC in binding buffer (Chapter 2 section 2.7.1), then flash cooled on ice and immediately incubated with 40 µl of Ni-NTA beads containing no protein as an initial negative selection step to filter out sequences with an affinity for Ni-NTA beads. The unbound ssDNA sequences were subsequently incubated with 214 µl of Ni-NTA beads with 96 µg (3 nmol) of MDA-9 bound for 90 minutes at 25 ºC with gentle shaking. In the initial round, the protein was in an equal ratio to ssDNA to capture every sequence with affinity for the target.

Following incubation, stringent wash procedures were introduced to effectively remove the non-specific or low-affinity ssDNA binding sequences. However, initial results demonstrated that it was impossible to progress beyond the early selection rounds. When three washes were carried out following the initial binding, only a minimal quantity of ssDNA was eluted from the beads. Two phases of PCR were required to amplify the DNA to a quantity sufficient for ssDNA re-generation and further rounds of selection. Gels electrophoresis analysis of the DNA products of these PCR steps is shown in figure 4.3B/C. The initial PCR step was carried out for only six cycles of amplification (Figure 4.3B) prior to fractions undergoing an additional 12 amplification cycles (Figure 4.3C). The results show little or no PCR product from the first six cycles of amplification; a further 12 cycles of amplification results in a visible PCR product;

however, the presence of significant by-products is visible. These results demonstrate that due to the low recovery of ssDNA, it was impossible to amplify it enough without by-product formation. Therefore, it was decided to modify the SELEX procedure to increase the ssDNA eluted at each stage.

SELEX method -2

An alternative selection method was developed utilising a “lower selection pressure” approach. Following the binding of the ssDNA pool with the Ni-NTA-MDA9 beads, a single 750 µl wash was used before eluting the ssDNA in water at 95 ºC as previously utilised to remove ssDNA bound to target effectively. This elution method is common in literature and does not require the addition of reagents such as urea, which will interfere with the PCR step.

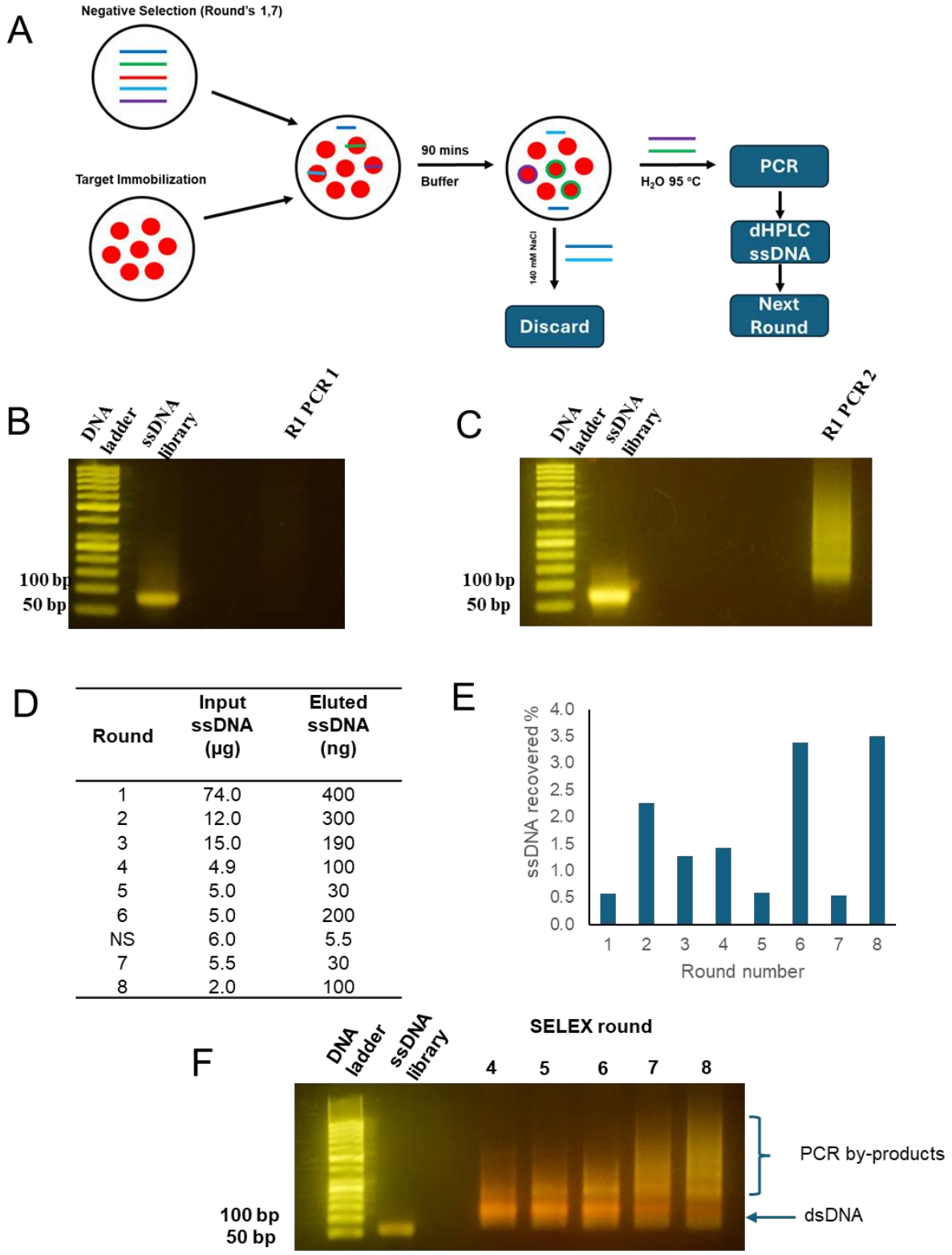
Following each selection round, an initial PCR step was performed to gauge PCR product quality; decreasing the number of PCR rounds became necessary as the protocol progressed. The number of PCR cycles by round is shown in Table 4.1. The only exception was the first round, in which the library yield was high enough to warrant only a small number of PCR amplification cycles. In addition, in rounds where the amount of eluted ssDNA was less than 30 ng, an initial six rounds of PCR were carried out to amplify the ssDNA to enough dsDNA for further amplification.

Following PCR amplification with a 5’ Texas Red labelled reverse primer, denaturing HPLC was used to isolate the ssDNA as previously developed and described in Chapter 3. Negative selection “NS” was performed between rounds 6 and 7. Despite using only a single wash, the selection pressure was increased throughout the protocol, from the second round onwards, by decreasing the volume of Ni-NTA-MDA-9 beads to increase the ratio of ssDNA library to MDA-9, increasing competition between binders. This ratio started at 1:1 (Library: protein) in round two and finished at 4:1 by round eight. In addition, the incubation time was reduced from 90 minutes in round one to 45 minutes in round 8. A breakdown of library quantities and wash/incubation times for each round is presented in Chapter 2, section 2.7.2

Table 4.1: PCR cycle numbers by round in the conventional SELEX protocol

Round number	Number of PCR cycles
1	8
2	18
3	15
4	12
5	12
6	10
7	8
8	6

Following the SELEX protocol outlined in Figure 4.3A and the modifications described above, a summary of the results based on the quality of PCR product generated during the SELEX rounds and the yield of ssDNA recovered based on quantification of ssDNA yield (post elution) from round to round is shown in Figure 4.3 D/E. Following multiple attempts at selection, the same problems repeatedly arose: a decrease in the quality of PCR product post-selection (see Figure 4.3F) and a lack of enrichment of ssDNA bound to MDA-9 (see Figure 4.3E). Typically, in a successful SELEX method, the % of ssDNA recovered (eluted ssDNA/input ssDNA) is small and slowly increases as rounds increase[237]. Thereby, the selection pressure is increased, removing weaker binding species. However, numerous attempts at the SELEX method failed to enrich for ssDNA binding to MDA-9 (see Figure 4.3 D/E).



Round	Input ssDNA (µg)	Eluted ssDNA (ng)
1	74.0	400
2	12.0	300
3	15.0	190
4	4.9	100
5	5.0	30
6	5.0	200
NS	6.0	5.5
7	5.5	30
8	2.0	100

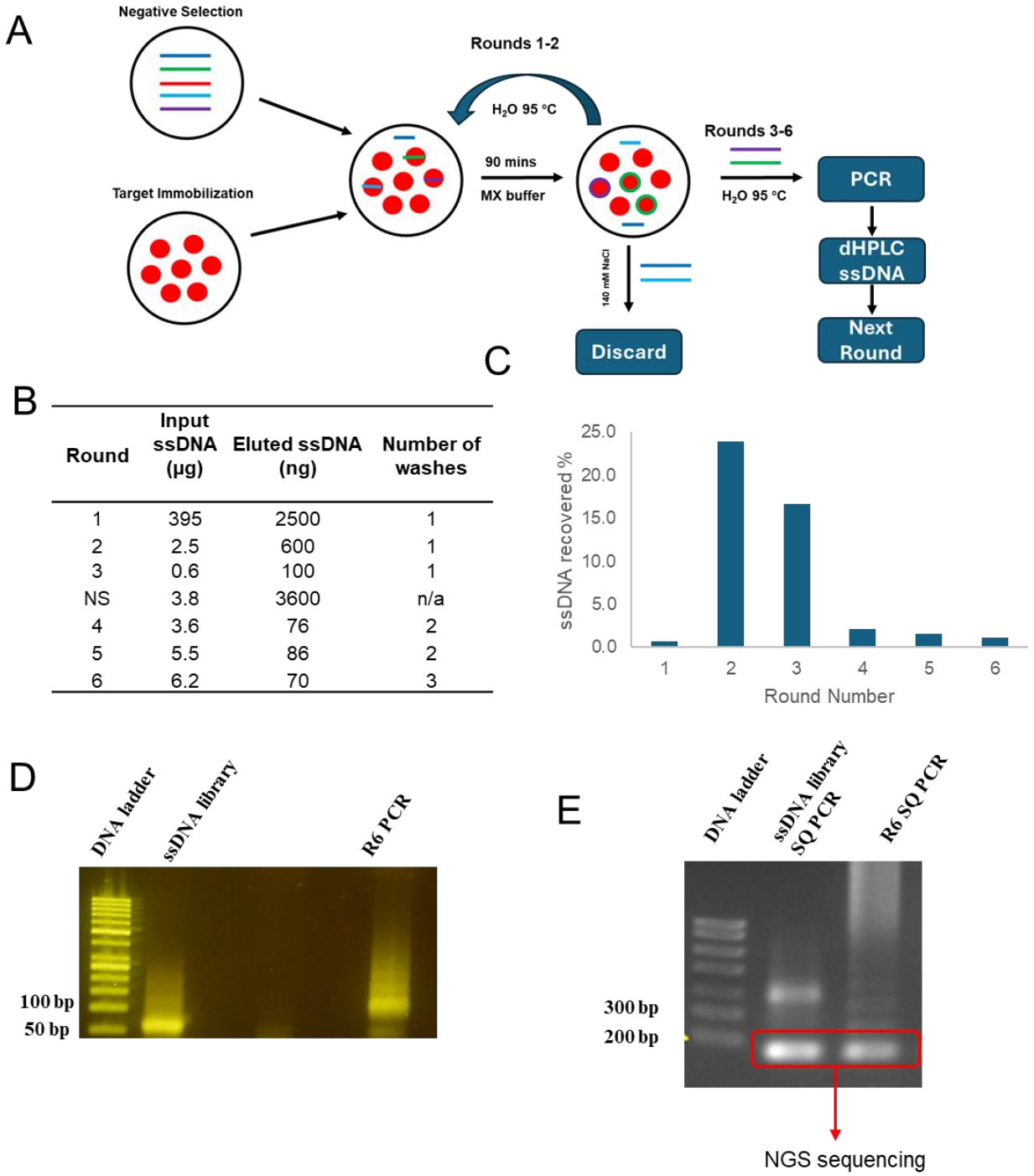
Figure 4.3: SELEX against MDA-9, method one. *A) Schematic illustration of conventional SELEX protocol utilized to generate ssDNA aptamers against MDA-9.B/C) Gel electrophoretograms showing PCR (six cycles of amplification) of ssDNA isolated after round one of SELEX. C) Second stage PCR of dsDNA from the initial PCR (R1). D) Table summarizing input ssDNA vs output ssDNA for SELEX rounds 1-8. E) Bar chart showing the % ssDNA recovered for rounds 1-8.F) Gel electrophoretogram of the PCR products generated from SELEX rounds 4-8, ssDNA control is highlighted.*

SELEX method -3

With no success at enriching the ssDNA library against MDA-9, alternative SELEX approaches were explored. More recent SELEX approaches with magnetic bead-based selections demonstrated that the required selection rounds can be drastically reduced (as low as three rounds) compared to more traditional SELEX methods [238]. The main problem of the previous SELEX methods was associated with the low quality of PCR product in the later rounds (see Figure 4.3F), caused by excessive PCR amplification. PCR by-products are a known problem in SELEX protocols, mainly the "ladder" type by-products described by Tolle et al. (2014), which accumulate parasitically from round to round [74]. Aiming to reduce the impact of PCR on the protocol, SELEX can be carried out without amplification or with fewer amplification steps, where a significantly larger quantity of ssDNA library and target have been previously used; eluted ssDNA from this larger scale selection is directly used for the next round of selection(s); finally, PCR is carried out on the few ssDNA species that remain, some of which should be high-affinity aptamers [239–241].

Therefore, an alternative SELEX strategy was adopted (summarised in Fig 4.4A). A larger quantity of ssDNA library was initially used (395 µg, 16 nmol), and the eluted ssDNA was directly used in two additional selection rounds. This was done without PCR amplification to ensure a significant selection pressure at the start of the protocol consistent with other successful SELEX approaches[241,242]. A negative selection step occurred between rounds three and four following the first PCR amplification/ssDNA re-generation. In addition to the early selection pressure, the number of wash steps increased, and the volume of beads decreased as the protocol progressed, see Chapter 2 section 2.7.3.

A summary of the results of the % ssDNA eluted and PCR product quality is shown in Figure 4.4B-D. Figure 4.4C shows that in the later rounds (4-6), there was still a relatively low % recovery of ssDNA. However, the number of washes was successfully increased to remove lower affinity binders. Figure 4.4D shows the PCR product following the final round of selection. The results show a higher quality PCR product than was seen previously due to the reduction in the number of rounds of SELEX requiring PCR amplification. After six rounds of selection, it was decided to proceed with sequencing, although no enrichment of % ssDNA bound to MDA-9 was observed.



B

Round	Input ssDNA (µg)	Eluted ssDNA (ng)	Number of washes
1	395	2500	1
2	2.5	600	1
3	0.6	100	1
NS	3.8	3600	n/a
4	3.6	76	2
5	5.5	86	2
6	6.2	70	3

Figure 4.4: Modified SELEX against MDA-9. *A) Schematic illustration of a modified SELEX protocol utilized to generate ssDNA aptamers against MDA-9.B) Table summarizing input ssDNA vs output ssDNA for SELEX rounds 1-8 with the number of washes shown. C) Bar chart showing the % ssDNA recovered for rounds 1-6.D) Agarose gel electrophoretogram of the PCR products generated from SELEX round 6, ssDNA control is highlighted. E) Gel electrophoretogram of the dsDNA prepared for next-generation sequencing (NGS) following PCR generated from SELEX round 6 and the original ssDNA library. Following library preparation, the DNA of the correct size was cut and eluted from the gel (highlighted in red box).*

Next Generation Sequencing of ssDNA

Following the final round of selection, the ssDNA round 6 library and the initial ssDNA library were amplified with a 5' Texas Red reverse primer and 1.2 µg of ssDNA was re-generated using denaturing HPLC. The resulting ssDNA was sent for NGS sequencing (CD genomics). Before sequencing, the manufacturer had to complete a sequencing preparation workflow on the ssDNA library. The library must have binding sites for sequencing adaptors for the Illumina NGS platform. The manufacturer used the 'TrueSeq' method. This involves phosphorylation of the 5' ends, A-tailing of the 3' ends to enable ligation of sequencing adaptors, ligation of adaptors, and further amplification by PCR [243]. The adaptors contain a unique identifier for that sequence (barcode) and the adaptor sequence itself, which enables the hybridisation of the adaptor region of the sequence with DNA oligos bound to the surface of the Illumina flow cell. In Illumina systems, sequencing occurs via the 'clonal' method, whereby the sequences bound to the surface are amplified by further PCR reactions before the primary sequencing process.

These additional PCR steps in the library preparation and sequencing stages increase the risk of PCR by-product formation, especially as the sequencing provider did not have enough time to optimise the PCR variables. Gel electrophoresis analysis of the PCR products of the resulting library (now containing barcodes and adaptors with sequencing primers attached) is shown in Figure 4.4E. The PCR product for both the random library and the post-SELEX library contained additional PCR by-products, with high MW by-products present in the products of both reactions. Therefore, before sequencing, the correct-sized DNA was cut and eluted from the gel before NGS (see Figure 4.4E).

4.3.4 Sequencing Data Processing Workflow

Illumina NGS produces data in the form of FASTQ files, which contain reads of 150 bases in length, 5' (80 bp library)(barcode)(adaptor) 3'. NGS resulted in 12.2 million reads for each library with sequencing artefacts removed. Additional rounds of NGS data processing were carried out in Galaxy Tools before examining the library in FASTAptamer, an R studio program developed for analysing SELEX libraries [244].

A summary of the workflow used to analyse the NGS sequencing is shown in Figure 4.5. The layout of the Galaxy Tools workflow is based on the work of Thiel et al. (2016) but with a modification at the end of the workflow to remove primers from the library [245].

1. **Upload**: The FastQ files are uploaded to Galaxy Tools.
2. **Trim Sequences**: The random region (bases 21-60 of 150) is isolated from the rest of the sequence
3. **Prinseq**: Using the 'Prinseq' tool, reads with a mean quality score (Q) below 30 were discarded; a Q score of 30 represents an error rate of 0.1%.[246]
4. **FASTQ to FASTA and Merge**: The FASTQ file is converted to a much smaller FASTA file, which contains only the sequence and a unique identifier; the two files corresponding to the forward and reverse end reads are then merged
5. **FASTA to CSV:** The FASTA file is converted to a tabular then CSV format, which allows it to be opened in Microsoft Excel
6. **Cut Primers:** The find and replace tool replaces all primer sequences (presumed to be present in the random region due to a PCR error or other sequencing error) with the character 'A'.
7. **CSV to FASTA and Filter:** the CSV file is uploaded to Galaxy Tools and converted into FASTA format. Any sequences with the primer or priming regions swapped for an 'A' character are now shorter than 40 bases and can be filtered out by length.
8. **FASTAptamer:** The FASTA file is uploaded to Fast Aptamer, which ranks the sequences by the number of reads and gives a value for the number of unique sequences within the library.[244]

4.3.5 FASTAptamer

FASTAptamer is capable of a broad number of data analysis and processing applications, including counting, normalizing, ranking and sorting the abundance of each unique sequence in a population, comparing sequence distributions for two populations, clustering sequences into sequence families based on Levenshtein edit distance, calculating fold-enrichment for all the sequences present across populations, and searching degenerately for nucleotide sequence motifs [244].

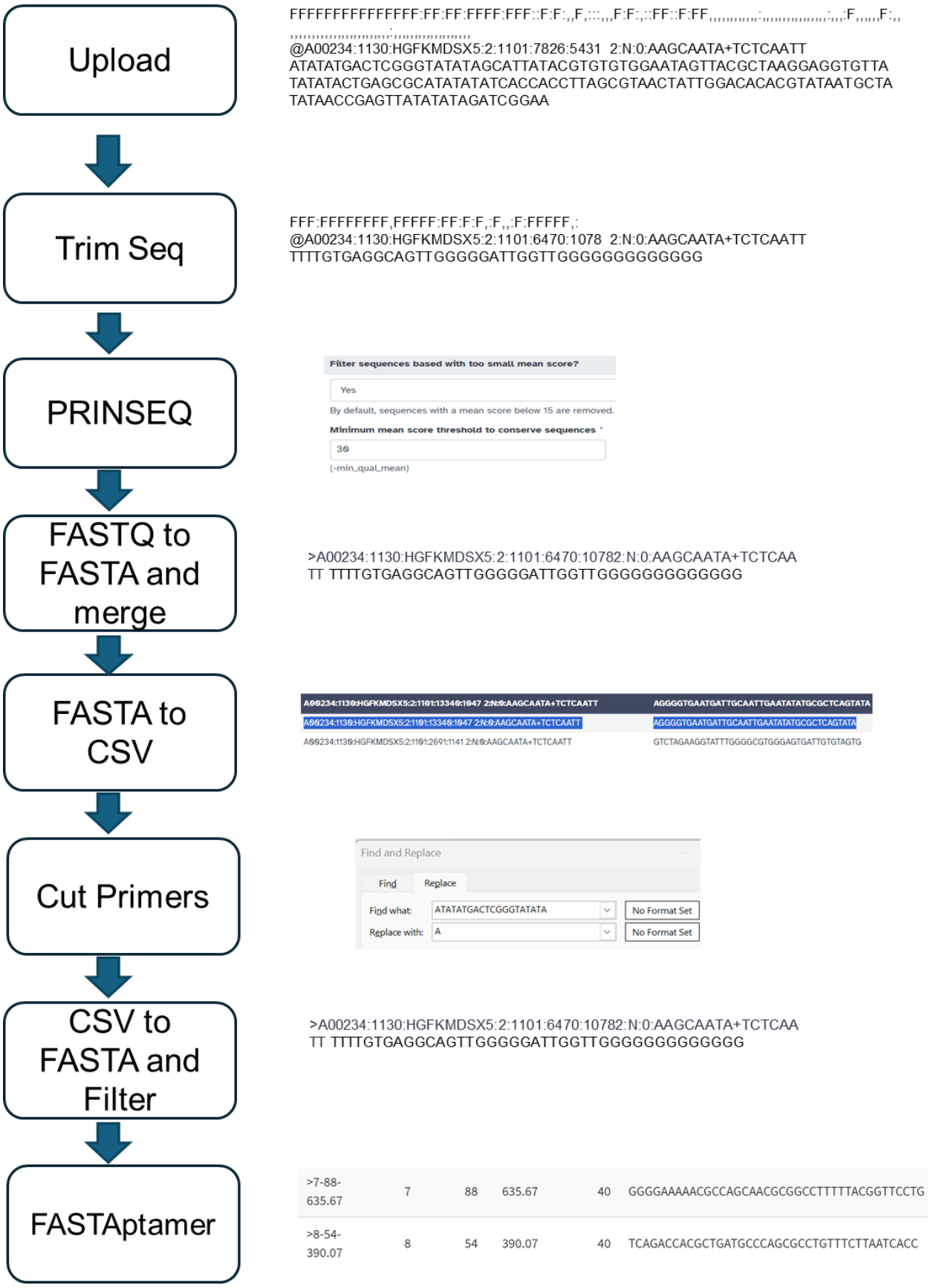


Figure 4.5: *Workflow used to analyse NGS data and identify potential ssDNA aptamer sequences.*

4.3.6 Random/pre-SELEX Library Sequencing Analysis

The FASTQ files for the random library were processed according to the workflow outlined in Figure 4.5. Ideally, the primer removal step would be omitted, but as PCR amplified the library in both the library prep and sequencing stages, PCR by-products containing the primer sequences could be present. A breakdown of how many reads passed each stage of filtering is shown in Figure 4.6. Sequencing artefact removal was performed by the sequence provider (19% of the initial reads), leaving 9.9 million (Figure 4.6A left panel). The following filtering stage removed low-quality reads with a mean Q score below 30 million (Figure 4.6B left panel).

Many reads failed this filter, removing 86.3% from the library, leaving 1.3 million reads. Finally, any N40 regions which contained sequences matching either the primers or priming regions were removed from the dataset million (Figure 4.6C left panel). The impact of the PCR amplification is noticeable as 52.5% of reads contained the primer sequences and were potential PCR by-products million (Figure 4.6D left panel). Following filtering, FASTAptamer was used to assess the number of duplicate reads. 94% of the initial library consisted of unique sequences, as expected for the initial random DNA library.

4.3.7 Post SELEX Library Sequencing Analysis and Identification of Aptamer Candidates

Following the analysis of the pre-SELEX, the post-SELEX library data was processed using the same workflow. The supplier's artefact filtering workflow removed 97.9% of reads from the library, leaving only 256,000 reads for further filtering (see Figure 4.6A right panel). This significantly reduces the number of reads available for processing and may lead to fewer potential aptamers being identified. The quality of the sequencing data is significantly lower for the post-SELEX library than for the pre-SELEX library. A potential reason for this could be an increased number of PCR amplifications performed, leading to parasitic by-products, which caused problems during the NGS sequencing. Alternatively, it is proposed that the post-SELEX library contains some enriched sequences with a high degree of secondary structure formation, which makes PCR and other steps within the sequencing workflow more challenging, leading to lower-quality data. Further filtering of the sequences by Q score

eliminated 18.1% of reads; a much higher proportion of the library at this stage had a mean Q score greater than 30 compared to the pre-SELEX library (Figure 4.6B right panel). This is most likely due to many low-quality reads being removed during the removal of sequencing artefacts.

During the removal of sequences containing PCR primers, 40.7% of reads contained the primer or sequences complementary to the primers. This is a significant amount, which indicates that the quality of the PCR product was poor during the sequencing workflow and has led to additional sequencing errors/ PCR by-products (Figure 4.6C). Following filtering, FASTAptamer was used to rank the sequences by number of reads and identify the number of unique and enriched sequences within the library. Figure 4.6D right shows the degree of enrichment within the library; 82.4% of sequences are unique compared to the libraries' 96%. It is important at this stage to consider the literature to infer what the abundance of enriched vs unique sequences can reveal about the degree of enrichment in a SELEX protocol. One way of measuring enrichment is the proportion of unique sequences within the library as the protocol progresses; each sequence in the library should be unique (depending on the length of random regions); as selection progresses and amplification of binders occurs, the percentage of unique sequences decreases. However, even in later rounds, they make up most of the sequences in the library; NGS studies of SELEX libraries have shown that enrichment typically plateaus when the number of unique reads gets to 80% of the library [112].

In the post-SELEX library, the value for the number of unique reads or background binders is 82.4%; this is not far from the theoretical plateau of enrichment described in the literature. Potential enrichment was achieved in the post-SELEX library, and the remaining reads were screened for aptamer candidates. The base composition of the pre and post-SELEX libraries was compared to the base composition specified in the library design, shown in Table 4.2. The pre-SELEX library, while having a slight abundance of guanine, differs significantly in composition from the desired ratio; the post-SELEX library differs even more so, implying there was a selection pressure towards a natural base composition and no evidence of guanine-rich sequences becoming enriched, it can be concluded then, from this experiment, there was no evidence of evolutionary favourability for G rich sequences in this case despite their known ability to form stable secondary structures.

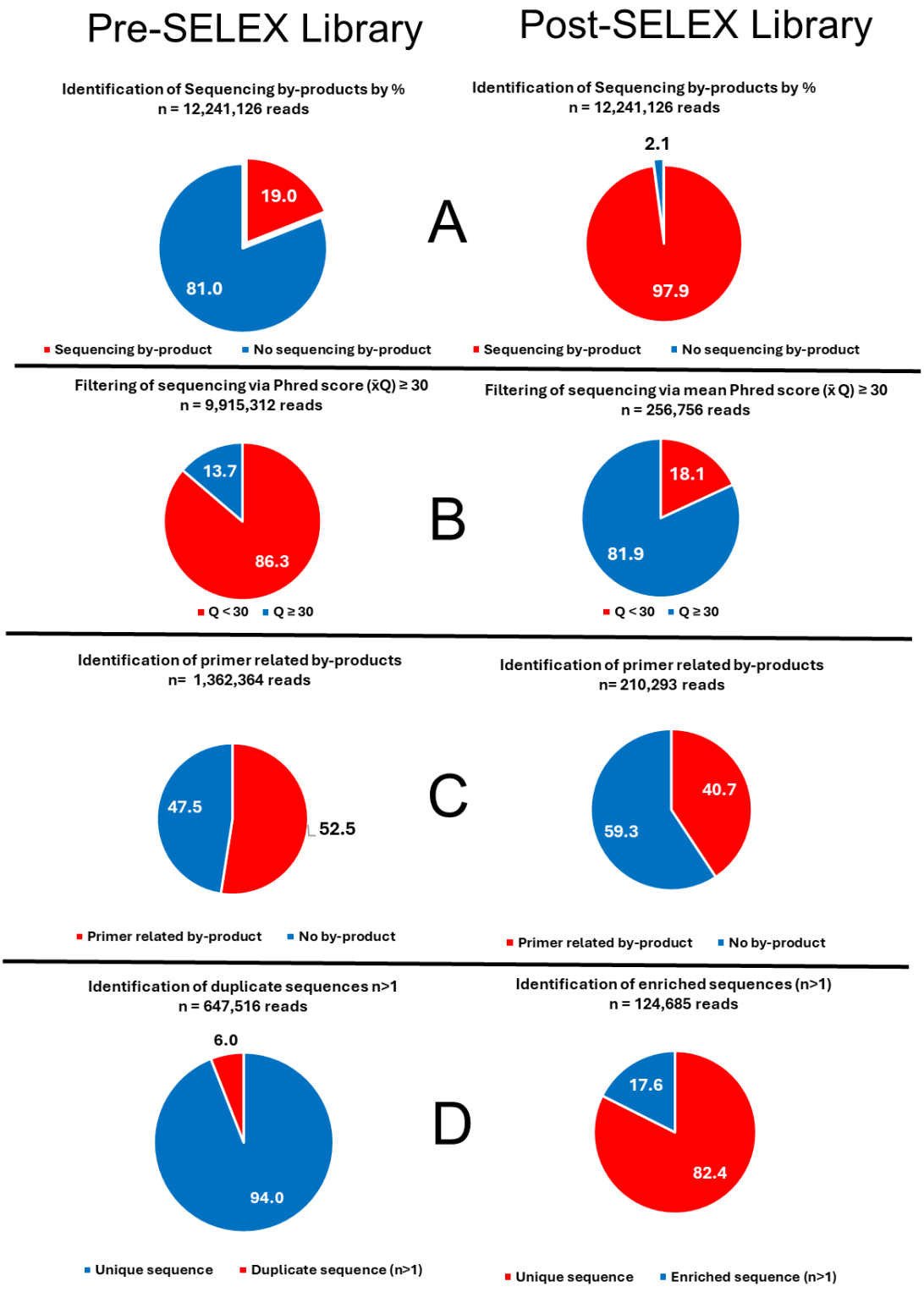


Figure 4.6: NGS outputs summarizing the sequencing data and processing of the pre and post-SELEX aptamer libraries. *A) Identification of sequencing by-products (done by manufacturer). B) Filtering sequences by mean Phred score (lower than 30) using the PRINSEQ tool on Galaxy tools. C) Identification of primer-related by-products within the library, defined as sequences containing eight or more bases in a sequence matching one of the primers. D) Identifying enriched sequences, defined as sequences with more than one read, uses a fast aptamer application within R Studio.*

Table 4.2: Base composition of pre- and post-SELEX libraries: Composition is *calculated using the "Base Composition" tool in Galaxy Tools.*

Base	Specification	Pre SELEX library	Post SELEX library
T	25%	26.8%	26.4%
C	15%	18%	21.2%
A	20%	24.6%	23.4%
G	40%	30.6%	29.0%

Twelve sequences with a range of read numbers were selected from the library. Previous sequencing workflows select the top 10-25 sequences by read count, but many of these in this dataset are similar. For an initial study, I wanted to select reads with diverse sequences. Then, the data could be revisited using the ability of FASTAptamer to find sequences within a family for higher read numbers if candidates baring similar sequences were found to have high affinity; the top 20 reads for the post and pre-SELEX libraries can be found in Appendix section 7.4. The reads selected for further characterisation are shown in Table 4.3.

Table 4.3: Sequences identified from sequencing data selected from FastAptamer

Sequence number	Sequence	Number of reads
1	GGTCGGAAGAGCGTCGTGTAGGGAAAGAGTGTTTGCGGAC	70
2	TCGGCCTATAGTCAGTGTACCTATAGGGCTATTAGGACCA	225
3	TATATCGGCCTATAGTCAGTGTACCTATAGGGCTATTAGG	180
4	GATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAA	93
5	TCAGACCACGCTGATGCCCAGCGCCTGTTTCTTAATCACC	54
6	GGGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTG	88
7	GGGAAAGAGTGTTTGCGGACGTGTAGATCTCGGTGGTCGC	22
8	TTAACGTGAAGTACCGTTATGAGCTGACGGACAGTGTGGG	12
9	TGGTCCTAATAGCCCTATAGGTACACTGATTATAGGCCGA	10
10	GAAGCTTCTCACTGTTTTATAATAAAACGCCCGTTCCCGG	9
11	TTCTCCCGTATTGTTGACATGCCAGCGGGTCGGGGAAACG	52
12	CCCAGAGGCAAAGGGCTCCCAGGATCCCTCAGAGGTCTTT	31

4.3.8 Characterisation of ssDNA aptamers to MDA-9 using ITC.

After identifying several potential ssDNA aptamer sequences (see Table 4.3), these 12 sequences were synthesised for further analysis to measure their respective binding affinities to MDA-9. ITC has been widely used to determine aptamer binding affinity. Several protein-binding aptamers have been characterised using ITC, including aptamers for thrombin [123,247,248].

ITC was used to determine the binding affinity of selected aptamers. Furthermore, a previously identified ssDNA aptamer for thrombin was synthesised as a positive control. Several biophysical techniques, including ITC, have studied the thrombin aptamer (TBA) with a Kd of ~50 nM. Although previous studies also report a much lower Kd of ~8 nM [248,249]. A BSA aptamer with a reported Kd of ~70 nM was also synthesised as a suitable control [250]. The ITC method is described in detail in Chapter 2, section 2.14.2.

The binding buffers used in each experiment were the same as described in the literature for previous studies of these aptamers; in the case of the thrombin aptamer, many different buffers have been used for studying the affinity of this aptamer, but for this study, the binding buffer from the original SELEX protocol in which this aptamer was isolated was used. Each experiment takes the mean heat of the last five injections as the subtract constant.

Measuring binding affinity of ssDNA aptamers

Three titrations were performed to assess the suitability of the nano ITC for protein-aptamer binding studies: a water: water titration, a buffer: buffer titration (SELEX buffer) and an EDTA: $CaCl_2$ titration as directed by the instrument manual. The raw data for these titrations is shown in Figure 4.7A-C and the binding isotherm, for the EDTA: $CaCl_2$ titration. The water and buffer titrations show a clear negative result, injection heat, but no binding occurs. The EDTA calcium titration was processed into a binding curve; Kd and stoichiometry calculations are within range of the expected values; these are established in ITC test kit documents as a stoichiometry (n) between 0.75 and 1.25, a Kd nM between 500 and 850 (nM). Therefore, it can be concluded that ITC is suitable for use in further binding studies.

Following initial validation of the ITC equipment, two positive control aptamers were analysed using ITC. A thrombin aptamer (TBA) is one of the earliest high-affinity protein binding aptamers to be isolated, and its mechanism of binding is well understood and has been analysed via ITC previously.[251] An aptamer to BSA has previously been identified by Wang et al. (2019); the Kd was determined via a partition-based method and was also selected as a secondary positive control.[250] For the thrombin assay, five µM thrombin (Cambridge Biosciences) was titrated with 100 µM thrombin binding aptamer; for the BSA assay, 5 µM BSA was titrated with 100 µM BSA binding aptamer, and the ITC data is shown in Figure 4.7E-H. Figure 4.7F and H titration determined a Kd of 16.6 nM ± 15.6 nM for the Thrombin-TBA interaction. These results are consistent with literature values for this aptamer of ~10-50 nM.

The shape of the curve is slightly below the ideal value, with the stoichiometry slightly lower than 0.75, which is considered the “good” threshold for ITC curves. However, literature studies of this aptamer also show curves with low stoichiometries [248,249,252]. The ITC analysis of the BSA aptamer (Figure 4.7 E/G) results in a Kd for the BSA titration (190 nM), which is similar (although higher) to the reported value of 70 nM. It should be noted that the quality of the ITC baseline is not ideal in this experiment; visual inspection of the raw data in Figure 4.8E reveals a broadening in the peaks at the start and end of the titration, indicating possible protein aggregation.

It should also be noted that this is not a pure fraction of BSA but was bought as a powder from Sigma Aldrich. Additionally, this glycerol and tween-20 were added to the binding buffer out of necessity due to the poor ‘behaviour’ of BSA in solution; a titration without these detergents added is shown in appendix section 7.5. This and the fact that the BSA is not entirely pure may be the reason for the deviation of Kd from the literature value. To critique the suitability of this titration as a positive control for ITC, it must be considered that BSA in this slightly impure form is available for very cheap gram quantities and can be used regularly compared to more expensive recombinant proteins.

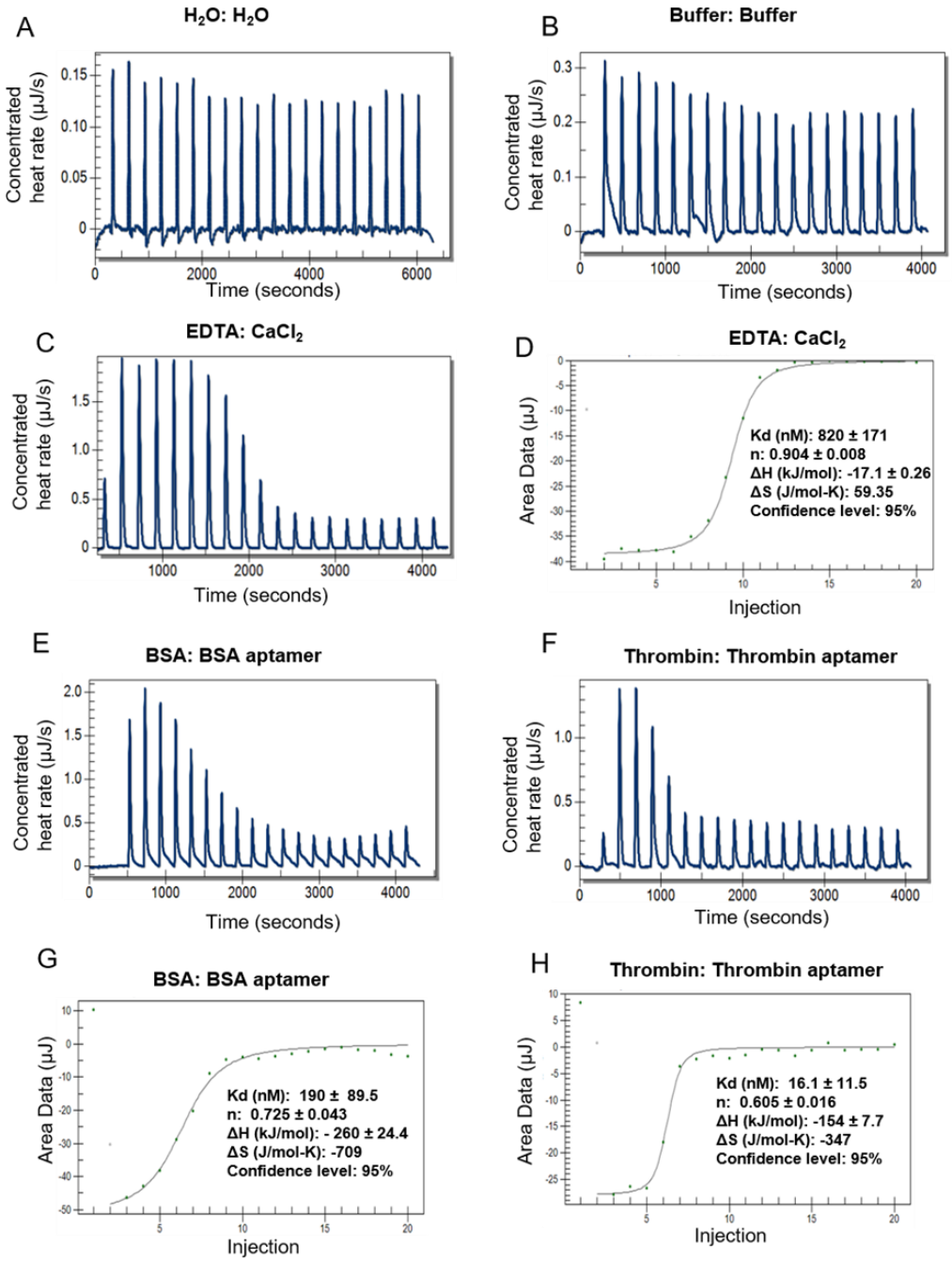


Figure 4.7: . Analysis of blanks and positive controls using ITC*. A-B) ITC analysis of water on water, buffer on buffer titration, heat vs time curve 25 °C, 20 injections, injection time 300 s, cell volume 300 µl, injection volume 2 µl. C) ITC analysis of EDTA: $CaCl_2$ titration, 19 injections, injection time 300 s, syringe concentration 0.95 mM $CaCl_2$, cell concentration 0.15 mM EDTA, cell volume 160 µl, injection volume 2.4 µl, 25 ºC. D) ITC analysis of heat vs time curve for EDTA: $CaCl_2$ titration plotted against injection number, the data was analyzed using a single site binding model and the best fit is shown as a solid line in (D). E) ITC analysis of BSA: aptamer titration 20 injections, injection time 300 s, syringe concentration 100 µM aptamer, cell concentration 5 µM BSA, cell volume 300 µl, injection volume 2 µl, 25 ºC. F) ITC analysis of Thrombin: aptamer titration 20 injections, injection time 300 s, syringe concentration 100 µM aptamer, cell concentration 5 µM thrombin, cell volume 300 µl, injection volume 1.8 µl, 37 ºC G) Heat vs time curve for BSA: aptamer titration plotted against injection number, the data was analyzed using a single site binding model and the best fit is shown as a solid line in (G). H) Heat vs time curve for Thrombin: Aptamer titration plotted against injection number. The data was analyzed using a single site binding model, and the best fit was shown as a solid line in (H).*

MDA-9 Aptamer binding

The ITC titration and integrated isotherms or ssDNA aptamer sequences 1-4 are shown in Figure 4.8. A comparison of the baseline data (panels A-D) shows a good baseline in all cases except sequence 4. This is due to sequence 4 giving a low enthalpy interaction with the protein; as a result, any baseline anomalies have a more significant impact on the curve fitting. The kinetic data (panels E-H) show suitable stoichiometry in all cases (typically defined as an n value between 0.750 and 1.250). The Kd values are in the nanomolar (163 nM) to low micromolar range (1.2 µM), demonstrating the successful isolation of several ssDNA aptamer candidates to MDA-9. Despite multiple attempts at repeating the titration, sequence four remains difficult to fit an isotherm to, with integration being challenging due to high signal-to-noise due to it being a low enthalpy interaction. Therefore, ITC may not be a suitable method for measuring the Kd of this aptamer.

The titration and curves for ssDNA sequences 5-8 are shown in Figure 4.9. The ITC baselines panels A-D are of high quality (except for sequence 7) but the kinetic data (panels E-H) show that only one of the sequences has a Kd indicating a specific interaction. Sequence 5 is a non-specific binder with an isotherm that has low stoichiometry. Sequence 6 is likely a nonspecific binder (approximately 1 µM Kd); the curve shows acceptable stoichiometry. Sequence 7 appears to have no interaction with MDA-9, hence why no isotherm can be generated. The baseline is poor due to the low enthalpy nature of this interaction. The results show that sequence 8 has a moderate affinity for MDA-9. However, the independent (one-site) binding model, typically used for aptamer studies, does not fit this data well, as the injection heat does not plateau as expected for a high-affinity binder. Other techniques may be more suitable for studying sequence eight as this shows a moderate affinity for the target. The ITC analysis for sequences 9-12 is shown in Figure 4.10. The injection data (panels A-D) is of high quality with minimal baseline issues. Examination of the kinetic data (panels E-H) reveals that sequence 9 has a Kd in the 600 nM range but with a low stoichiometry. Sequence 10 displays interesting data with the last few injections dipping below the baseline (see Figure 4.10B); other techniques may be necessary to understand this interaction, as seen in previous studies of protein binding aptamers [248]. However, due to the low affinity, this aptamer is only relevant for further study with the priming regions removed, as this may show a greater affinity for the target;

the anomalous baseline behaviour contributes to an extensive error range of the Kd. Sequence 11 demonstrates a high affinity for MDA-9 with a Kd 25 nM ±10.8 nM. The stoichiometry (0.75) is at the lower end of the "good" range. Sequence 12 is a non-specific binder and does not warrant further study.

The ITC analysis has identified several potential ssDNA aptamers that bind with Kd's in the 10-100s nM range. The ITC analysis also showed various trends, with several of the ssDNA-protein interactions characterised by high enthalpies. This highlights a potential caveat of ITC as a method for studying aptamer-protein interactions, as an aptamer with a low enthalpy but high-affinity interaction would suffer from signal-to-noise issues due to low injection heats, thus making it challenging to generate and fit an isotherm to the data. Some ssDNA aptamer candidates are promising, particularly sequence 11, which has a very low nanomolar Kd. Additional replicates and a secondary method would be required to confirm. In the initial binding study for the aptamer candidates, the sequences had the priming regions attached, intending that the following study would use sequences without the priming region. The aim was to compare binding affinities with and without primers, which has been done in previous aptamer studies; a significant difference between the two would demonstrate that the priming regions were interfering with the binding and that the priming regions would need to be redesigned [253]. Future work must cover the analysis of the sequences without the priming regions.

Negative controls

In addition to the ITC analysis of the MDA-9 aptamer sequences, further negative control studies were performed using the thrombin and BSA aptamers titrated against MDA-9 in their respective binding buffers. In addition, ssDNA sequence 1 was titrated against BSA in the SELEX binding buffer used for selection against MDA-9. The ITC analysis is shown in Figure 4.11, showing that the thrombin aptamer has no interaction with MDA-9 as expected. Furthermore, the BSA aptamer (panels B and E) shows an anomalous interaction with MDA-9. Sequence 1 shows a low affinity and non-specific interaction with BSA (panels C and F), demonstrating a much lower Kd compared to MDA-9, as expected.

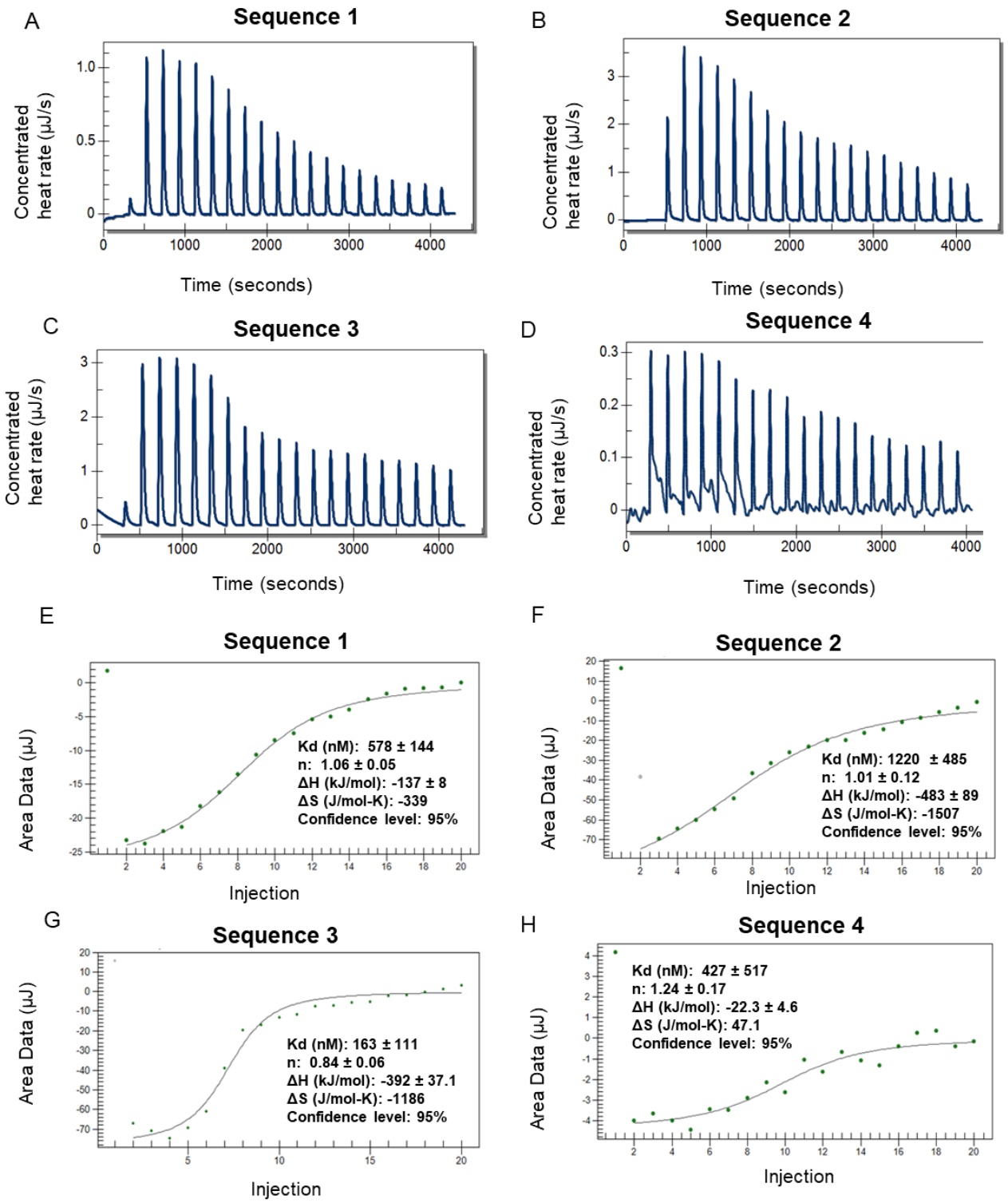


Figure 4.8: Analysis of ssDNA aptamers against MDA-9 using ITC. *ITC data for titration of MDA-9 with ssDNA sequences 1-4 (S1-4), 19 injections, injection time 300 s, syringe concentration 100 µM ssDNA aptamer, cell concentration 5 µM MDA-9, cell volume 300 µl, injection volume 2.0 µl, 25 ºC. A-D) Raw calorimetric data plotted as heat vs time curves. E-H) Corresponding heats of reaction plotted against injection number, the data was analyzed using a single site binding model and the best fit is shown as a solid line in (E-H).*

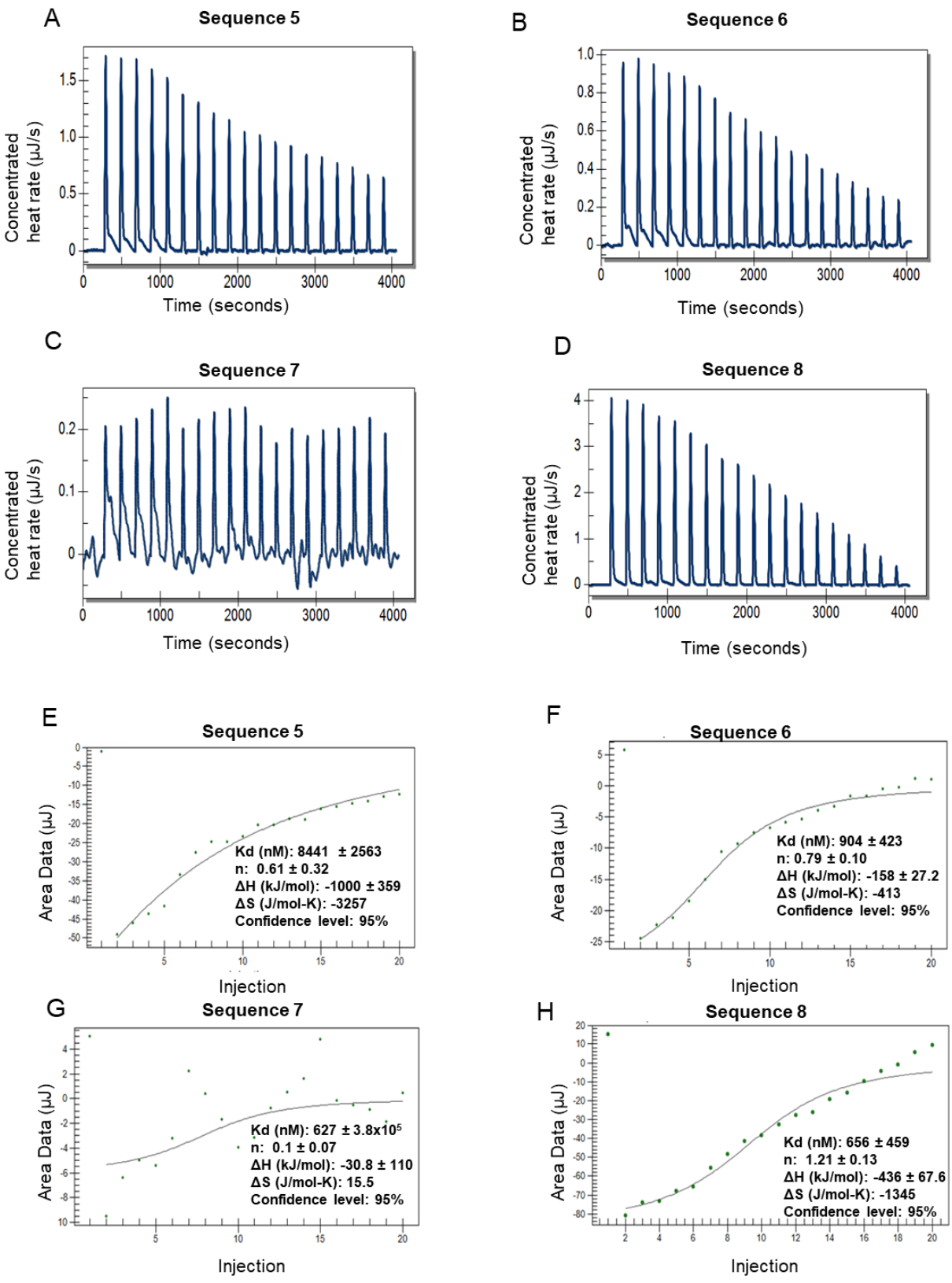


Figure 4.9: Analysis of ssDNA aptamers against MDA-9 using ITC. *ITC data for titration of MDA-9 with sequences 5-8 (S5-8), 19 injections, injection time 300 s, syringe concentration 100 µM aptamer, cell concentration 5 µM MDA-9, cell volume 300 µl, injection volume 2.0 µl, 25 ºC. A-D) Raw calorimetric data plotted as heat vs time curves. E-H) Corresponding heats of reaction plotted against injection number, the data was analyzed using a single site binding model and the best fit is shown as a solid line in (E-H).*

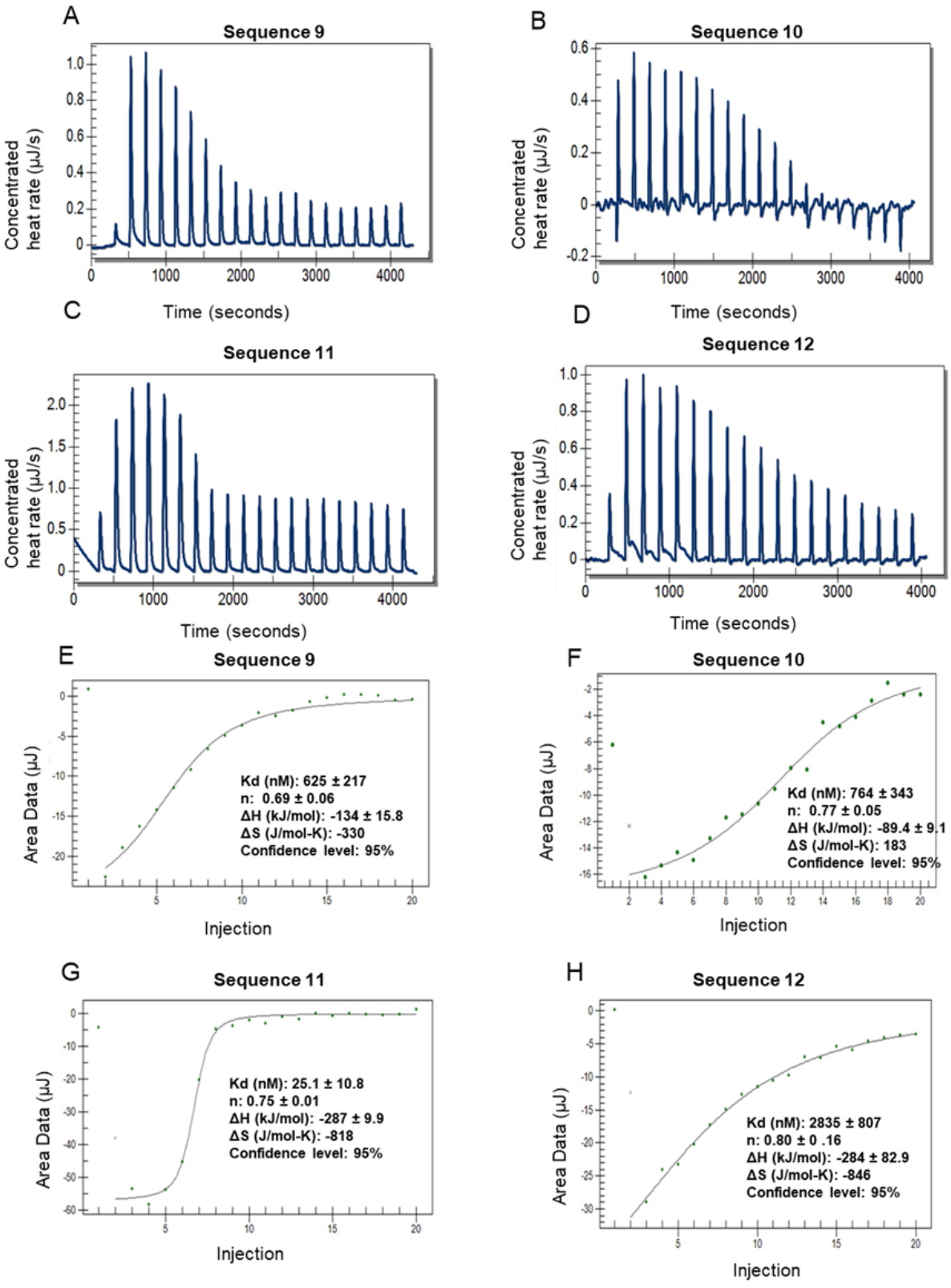


Figure 4.10: Analysis of ssDNA aptamers against MDA-9 using ITC. *ITC data for titration of MDA-9 with sequences 9-12 (S9-12), 19 injections, injection time 300 s, syringe concentration 100 µM aptamer, cell concentration 5 µM MDA-9, cell volume 300 µl, injection volume 2.0 µl, 25 ºC. A-D) Raw calorimetric data plotted as heat vs time curves. E-H) Corresponding heats of reaction plotted against injection number, the data was analyzed using a single site binding model and the best fit is shown as a solid line in (E-H).*

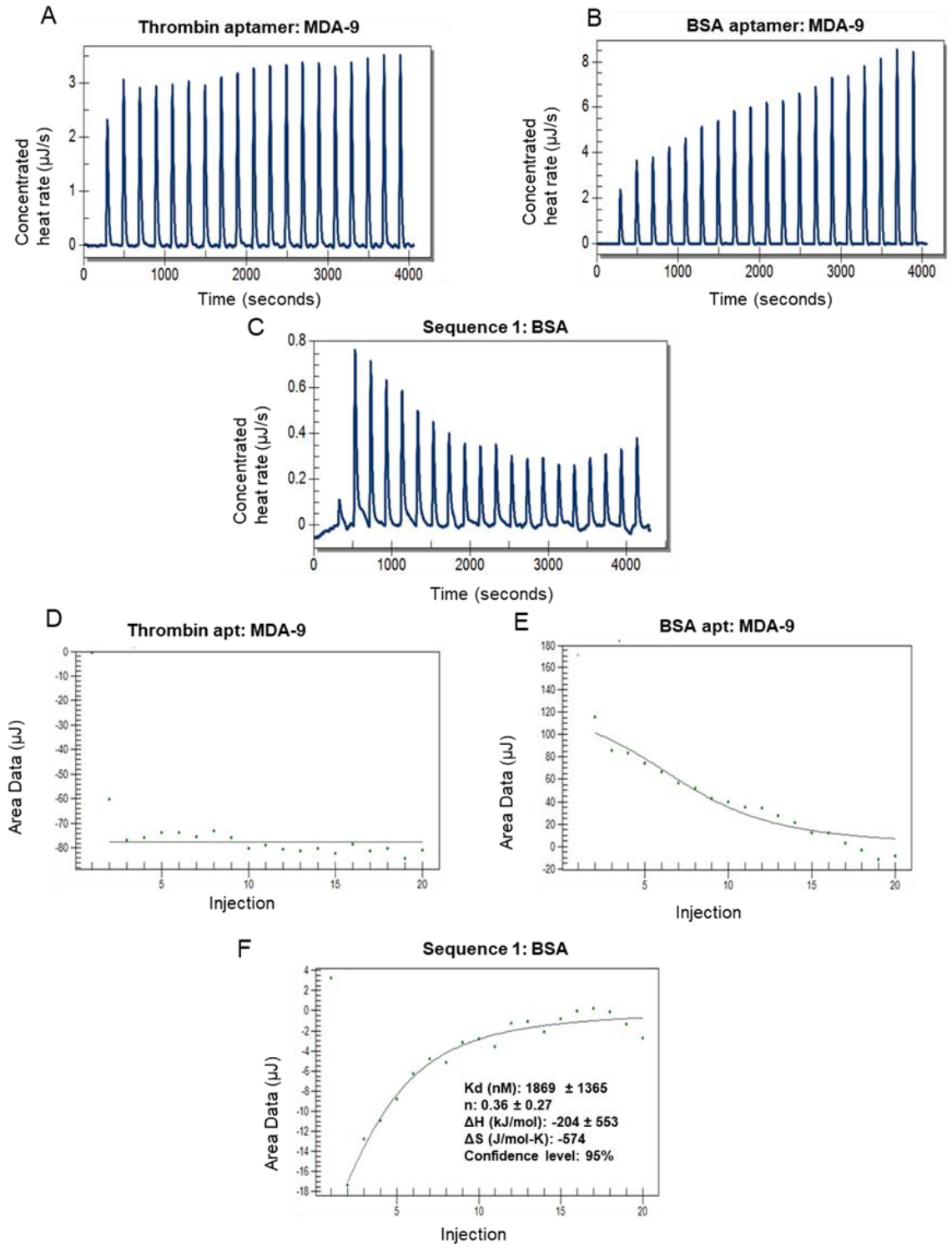


Figure 4.11: Analysis of ssDNA aptamers against MDA-9 using ITC. *ITC data for titration of MDA-9 with a BSA aptamer, thrombin aptamer (TBA 100 nm Kd) and BSA with ssDNA sequence 1. A-C) Raw calorimetric data plotted as heat vs time curves. BSA aptamer: 19 injections, injection time 300 s, syringe concentration 50 µM aptamer, cell concentration 5 µM MDA-9, cell volume 300 µl, injection volume 2 µl, 25 ºC. Thrombin aptamer: 20 injections, injection time 300 s, syringe concentration 100 µM aptamer, cell concentration 5 µM MDA-9, cell volume 300 µl, injection volume 1.8 µl, 37 ºC. Sequence 1: 20 injections, injection time 300 s, syringe concentration 100 µM aptamer, cell concentration 5 µM BSA, cell volume 300 µl, injection volume 2.0 µl, 25 ºC. D-F) Corresponding heats of reaction plotted against injection number, the data was analyzed using a single site binding model and the best fit is shown as a solid line in (D-F).*

4.3.9 ITC Data summary:

A summary of the ITC data for the sequences identified from the sequencing data is shown in Table 4.4.

Table 4.4: Comparison of binding affinities for ssDNA aptamer sequences screened against MDA-9

Sequence number	Number of reads	Kd (nM)	Stoichiometry (n)
BSA Apt	N/a	190 ± 90	0.73
Thrombin Apt	N/a	16.1 ± 11.5	0.61
1	70	579 ± 140	1.06
2	225	1,222 ± 467	1.01
3	180	163 ± 110	0.84
4	93	427 ± 486	1.24
5	54	8,441 ± 2563	0.61
6	88	904 ± 423	0.79
7	22	627 ± >10,000	0.1
8	12	1,133 ± 478	1.21
9	10	625 ± 217	0.69
10	9	764 ± 335	0.77
11	52	25.1 ± 10.7	0.75
12	31	2,835 ± 800	0.80

To summarize, many of the aptamers identified are in the 100-1000 nM Kd range, which implies they are specific binders; in addition, there is one aptamer in the <50 nM Kd range, which is high affinity. Further work, including studies with alternative biophysical techniques such as SPR or ELONA, would further validate the analysis of the ssDNA aptamers for MDA-9 and Kds obtained in this study.

4.4 Conclusions

This chapter's primary aim was to isolate high-affinity ssDNA aptamers against MDA-9. MDA-9 is a protein associated with cancer metastasis and proliferation; therefore, aptamers acting as antagonists to this protein offer therapeutic potential, as has been demonstrated by aptamers to other cancer-associated proteins of interest, such as AS1411. In addition to therapeutic potential, strong binders with poor antagonistic performance offer diagnostic potential due to the relative ease of aptamer labelling and the abundance of this protein in cancerous tissue.

A secondary aim was to develop an in-house SELEX protocol that could be used for future studies. MDA-9 was cloned into *E. coli* prior to overexpression and purification of the recombinant protein. Ni-NTA affinity chromatography was used to purify the protein prior to verification using mass spectrometry. Using SELEX methods optimised in Chapter 3, a conventional SELEX protocol was implemented. However, applying selection pressures in conjunction with stringent washing to remove non-specific ssDNA binders to MDA-9 resulted in low recovery of ssDNA, and the SELEX procedure either failed in the early rounds due to complete loss of library or the PCR product quality deteriorated to such an extent that further selection was not possible.

To counter this problem, a low amplification SELEX protocol was designed to minimize the impact of PCR by-products on the process. A larger amount of ssDNA starting library was used, and three section rounds were performed with no amplification steps, followed by three rounds of conventional SELEX. Following the six rounds of selection, there was no evidence of significant enrichment of the ssDNA pool to MDA-9. NGS sequencing was performed at this stage. The sequencing data revealed that the PCR product quality had an impact on the NGS workflow; following the ligation of the sequencing adaptors to the library, the following amplification yielded poor quality PCR product, while the desired band could be excised from the gel, PCR occurs during the sequencing process itself and the impact of by-products in this step cannot be directly mitigated.

As a result, the quality of sequencing data with sequencing-related artefacts, filtering by Q score and primer-related artefacts reduced the number of reads from 12 million to 600 thousand in the pre-SELEX library and just 120 thousand in the post-SELEX

library. This loss of information is critical as previous studies have shown that high-affinity aptamers often do not have exceptionally high read counts, and so many could have been lost due to the poor quality of sequencing data.

Despite the potential low quality of the NGS of the SELEX round 6 library against MDA-9, twelve potential aptamer candidates were identified for further study using ITC. ITC was established as a suitable method for analysing aptamer-protein interaction by analysing well-characterised thrombin and BSA aptamer to their respective proteins. Kd values were obtained that were consistent with previous studies. The ITC analysis of the potential MDA-9 aptamers demonstrated that three selected ssDNA sequences demonstrate high affinity binding to MDA-9. Sequences 1 and 3 are in the sub 600 nM Kd range, and sequence 11 was the binder with the highest affinity, with a Kd of approximately 25 nM. Further studies are proposed to examine a wider number of ssDNA sequences and analysis on selected sequences with the primer regions removed (truncated sequences) using only the central 40 nucleotides, which may yield higher affinity aptamers to MDA-9.

5 Chapter 5: Accurate sizing of dsRNA biocontrols and mRNA therapeutics using ion pair reversed phase chromatography

5.1 Abstract

RNA biomolecules present a new frontier of solutions for therapeutic and industrial challenges. dsRNA biocontrols are emerging as a new sustainable approach for pest management strategies as alternatives to chemical pesticides. In addition, mRNA has emerged as a powerful new class of therapeutic, as demonstrated by the recent mRNA vaccines developed during the Coronavirus pandemic. The analysis of large RNAs is challenging, and there is significant demand for improved analytical methods to characterise functional RNAs, including methods for their accurate sizing. In this chapter, a wide range of ion pair reagents and HPLC conditions were optimised for the accurate sizing of dsRNA biocontrols and mRNA therapeutics. Under different IP

RP HPLC conditions, a range of different commercial dsDNA, dsRNA and ssRNA nucleic acid ladders were used to determine the extent of the linear relationship between retention time and the logarithm of nucleic acid size (nts or bps). Based on their retention times, this relationship was used to predict the lengths of various nucleic acid species, enabling the determination of the most effective chromatography conditions for sizing dsRNA biocontrols (100-500 bp) and therapeutic mRNAs (800-4500 nts). The results show that dsRNAs can be sized accurately (less than 5% error) against a dsRNA ladder with weak ion pair reagents (TEAA) under non-denaturing conditions (50 ºC).

Furthermore, using stronger ion pair reagents (DBAA) under denaturing conditions resulted in less than 2.5% error in sizing the dsRNA. The analysis of mRNA using IP RP HPLC for accurate sizing was less accurate, resulting in a 15% error against a high-range RNA ladder under denaturing conditions in DBAA. Furthermore, the linear relationship between retention time and length breaks down above ~1000 nucleotides due to a loss of resolution, leading to poor sizing accuracy. The results show that IP RP chromatography effectively determines single and double-stranded RNA size < 1000 nts/bps. However, resolving therapeutic mRNAs, typically in the 4000-6000 base range, is challenging using IP RP HPLC.

5.2 Introduction

RNA is utilised in the single and double-stranded form for several purposes, both therapeutically and industrially. DsRNA biocontrols have demonstrated success as biopesticides, with Fire et al. first demonstrating successful gene silencing with oral delivery of dsRNA to insects, leading to death, in 1998 [37]. Decades of research eventually led to the first approval of a dsRNA pesticide via the US EPA for use on the Colorado potato beetle [36,37]. During and after the COVID-19 pandemic, the utility of mRNA vaccines became abundantly clear. Their direct transdermal injection affords in vivo expression of a desired antigen in a small local area, effectively inducing immune response and immunity. The nature of mRNA means that only a change in sequence is required to produce a different antigen with minimal changes to the rest of the formulation [16,18,254].

Despite the promise of these platforms, with their success comes greater demand for large-scale batch production and associated regulatory challenges. The use of pesticides is governed strictly by environmental legislation to mitigate the risk of environmental damage and (in the case of biocontrols) releasing GMO material into ecosystems [36,151]. Following approval, mRNA vaccines, like any therapeutic, require post-marketing surveillance, focusing on quality control in batch production. The USP guidelines for vaccine quality control include critical quality attributes: identity, purity, content, and safety, all of which must be satisfied [148,149].

The analytical challenges facing both these platforms are significant. However, while their functions may differ, their fundamental chemistries are very similar, in theory allowing for one analytical technique to provide a robust analysis of both these species.

Several techniques have been demonstrated for nucleic acid analysis, including gel electrophoresis, capillary electrophoresis and HPLC via anion exchange or ion pair see Chapter 1). However, IP RP HPLC offers rapid separation of nucleic acids and has previously been shown to separate dsDNA size-dependently [168,255]. This is key as it allows for qualitative analysis via size-based separation of products from by-products and insight into product integrity [256]. This potentially allows for the satisfaction of all critical quality attributes in a rapid quality control assay.

This chapter will focus on IP RP HPLC to accurately size two specific RNA products, mRNA for use in vaccines and dsRNA for use as gene silencing pesticides. In addition, the retention behaviour of modified and unmodified mRNAs will be compared to assess the sequence dependence of each ion pair system. This study will investigate the retention behaviour of double and single-stranded species, following previous work but on a much larger scale, comparing weak and strong ion pair reagents based on size and sequence-dependent separation.

This chapter aimed to study the ability of alternative ion pair regents of varying strength (hydrophobicity) under denaturing and non-denaturing conditions to accurately size both dsRNA biocontrols and mRNA therapeutics. Ultimately, the aim was to identify the ideal ion pair reagent and nucleic acid ladder to size each species for a size-dependent quality assurance assay following production. A secondary aim was to establish an upper limit for the linear relationship between retention time and the

logarithm of length of a nucleic acid, which was established in previous studies. Once this was established, methods to improve this technique could be explored.

5.3 Results and Discussion

Initial work compared five mobile phase compositions using different ion pair reagents, including TEAA, DBAA, DBAA HFIP and TBAA. In each mobile phase composition, three nucleic acid ladders, dsRNA (New England Biolabs), dsDNA (Generuler, Thermo Fisher), and ssRNA (Riboruler, Thermo Fisher), were analysed under non-denaturing conditions at 50 °C (see Figure 5.1) and denaturing conditions at 75/95 °C (see Figure 5.2). To determine the relationship between retention time and nucleic acid length, a plot of retention times against nucleic acid length (log) was performed (see Figures 5.1B and 5.2B). For completely size-dependent separations where the sequence, chemistry and structure of the nucleic acid do not affect the hydrophobicity of the nucleic acids and subsequent retention time, this would mean all the different nucleic acid markers of the same size would co-elute. In the initial experiments, the aim was to compare the three commercial nucleic acid ladders to each other in each ion pair reagent to assess whether it was possible to form linear relationships between retention time and the logarithm of the length in each case (R^2 >0.9) this would allow for them to be used to estimate the length of the corresponding mRNA or dsRNA. Ladders that did not correlate (R^2<0.9) would likely not be suitable for sizing.

This would indicate a lack of size or greater sequence dependence when comparing ladders of different species, i.e. ssRNA to dsRNA. Using this linear relationship, the length of each RNA fragment could be estimated using their retention time. The difference between the predicted value and the actual length would determine which ladder/mobile phase system was most accurate at sizing the RNA. The analysis was run in triplicate for each sample in this study, with the mean retention time of the three replicates used for sizing calculations and correlation plots. In all cases, unless specified injections were highly reproducible (< 0.1-minute retention time drift), error bars are not plotted on any correlation plots.

Figure 5.1A (left panel) shows that under non-denaturing conditions using weak IP RP HPLC (TEAA), there are apparent differences in RT of the corresponding ds and ss species of the same length (nt vs bp), confirming that under these non-denaturing conditions, significant differences in overall hydrophobicity are observed between ss

and ds nucleic acids. These results are consistent with previous work demonstrating that in TEAA, the separation of double-stranded fragments is typically size-dependent, while the retention of single-stranded fragments is greatly influenced by base composition [257].

Analysis using the stronger ion pair reagents (DBAA/TBAA) under non-denaturing conditions (Figure 5.1B) resulted in improved correlations between the ds and ss nucleic acids. Therefore, a significant reduction in the difference in RT was observed between ds and ss species using DBAA or TBAA compared to TEAA. However, the results show a difference in retention time for shorter single and double-stranded fragments of the same length under these conditions in strong ion pair reagents.

Consistent with the results obtained for TEAA, only minor differences in RT are observed when comparing dsRNA and dsDNA fragments of the same length under non-denaturing conditions in stronger ion pair reagents (see Figure 5.1B). These results demonstrate that predominantly size-dependent separation of the ds nucleic acids is obtained using either weak or strong ion pair mobile phases for analysis. The results also show that the most significant difference in RT between ssRNA and dsDNA fragments of the same length occurs in the lower size range in TBAA (see Figure 5.1A right panel). This is expected as the effects of the individual sequence (AT/GC) content on overall hydrophobicity will play a more significant role where resolution is superior.

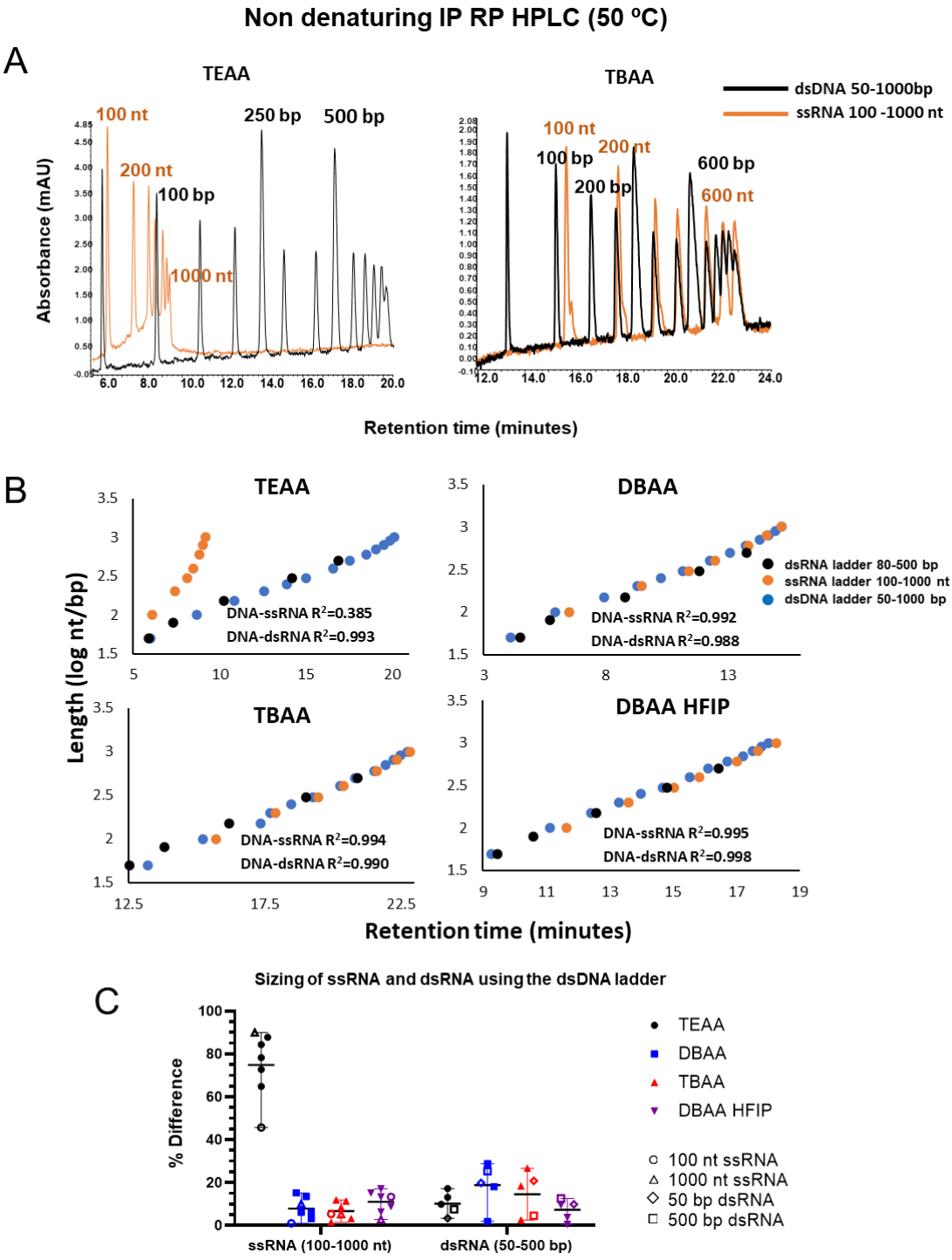


Figure 5.1: IP RP HPLC analysis of dsDNA, ssRNA and dsRNA ladders at 50 °C. *A) IP RP chromatogram of dsDNA ladder compared against ssRNA ladder in TEAA and TBAA at 50 °C. B) Graph showing the correlation between log nucleic acid length for ss and dsRNA against retention time for different IP reagents at 50 °C. Correlations were calculated by combining the series of two ladders; for example, the series containing the DNA ladder has 12 points; this was combined with the RNA ladder series to make a series with 19 points for which an R^2 was calculated. C) Accuracy of sizing of ssRNA and dsRNA ladders using the dsDNA ladder. The % difference of the calculated length vs theoretical length of the RNA is shown. IP RP HPLC separations were performed on a monolithic PS DVB capillary column at a 2.2 µl/min flow rate at 50 °C with UV analysis at 260 nm. Separations were performed using gradient 8 (TEAA), 10 (DBAA), 12 (DBAA HFIP), and 14 (TBAA) see Chapter 2 section 2.5.3.*

Figure 5.1C shows the results of estimating the size of each fragment in the ss and dsRNA ladders based on their retention times using the linear regression of the dsDNA ladder. The results show that under native conditions, it is feasible to effectively size dsRNA against the dsDNA ladder in both weak and strong IP mobile phases. This is consistent with previous observations that show the separation of ds nucleic acids largely depends on its length because the bases are within the helix structure and do not interact with the stationary phase [257]. However, it should be noted that sizing of the dsRNA using the dsDNA ladder resulted in typical errors ~10-20%, therefore demonstrating the differences in hydrophobicity between RNA and DNA and potential structural differences that result in differences in RT of dsRNA and dsDNA of the same length.

Further work was performed to analyse the nucleic acid ladders under denaturing conditions (75/95 °C, see Figure 5.2). Under these conditions, the dsRNA and dsDNA are expected to denature their corresponding ssRNA and ssDNA species, enabling comparison of only the ss species. The effects of different sequences (base composition), the effects of 2’ OH vs 2’ H on the ribose sugar and overall structural differences can impact the overall hydrophobicity and retention time. The results demonstrate that, as expected under denaturing weak IP RP HPLC (see Figure 5.2 A left panel), the significant difference in RT of ss vs ds nucleic acids observed under non-denaturing conditions has been eliminated. The only exception is that dsDNA fragments above 500 bp do not fully denature in TEAA (Figure 5.2 left panel). However, differences in RT of the same-sized fragments (ssRNA vs ssDNA) are observed, demonstrating the influence of different chemistry/sequences on the overall hydrophobicity and RT.

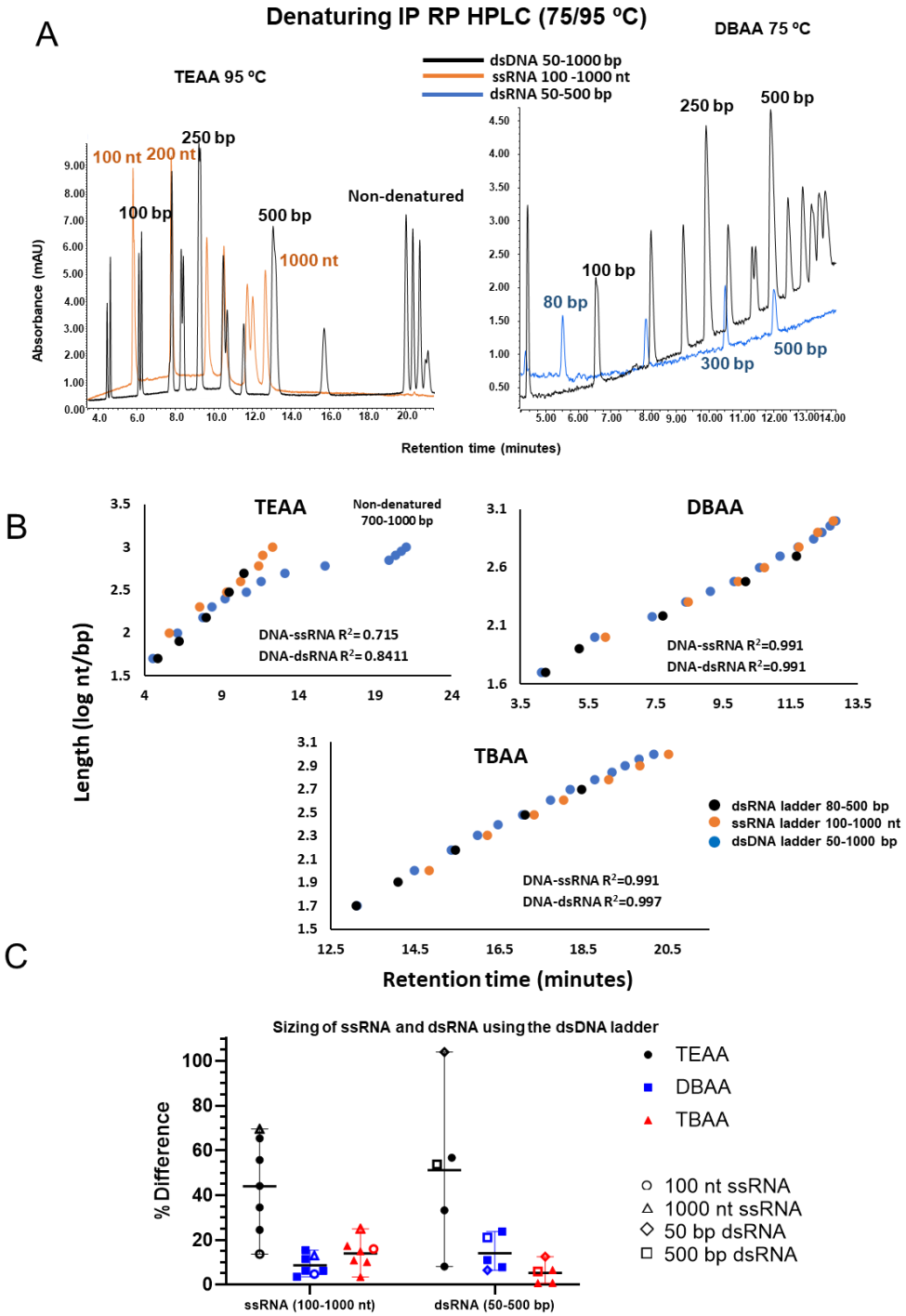


Figure 5.2: IP RP HPLC analysis of dsDNA, ssRNA and dsRNA ladders at 75 °C. *A) IP RP chromatogram of dsDNA ladder compared against ssRNA ladder in TEAA and TBAA at 75 °C. B) Graph showing the correlation between log nucleic acid length for ss and dsRNA against retention time for different IP reagents at 75 °C. Correlations were calculated by combining the series of two ladders; for example, the DNA ladder series has 12 points, which was combined with the RNA ladder series to make a series with 19 points for which an R^2 was calculated. C) The sizing of ssRNA and dsRNA was accurate using the dsDNA ladder. The % difference of the calculated length vs theoretical length of the RNA is shown. IP RP HPLC separations were performed on a monolithic PS DVB capillary column) at a flow rate of 2.2 µl/min at 75 °C with UV analysis at 260 nm, except for TEAA separations, which were performed on a DNAPac column at a flow rate of 0.3 ml/min at 95 ºC. Separations were performed using gradient 9 (TEAA), 11 (DBAA), and 15 (TBAA) see Chapter 2 section 2.5.3.*

The results shown in Figure 5.2A (left panel) demonstrate the separation of the two strands from the corresponding dsDNA (same length), highlighting this effect. This effect is more pronounced in shorter fragments in TEAA, where resolution is greater and differences in base sequence impact overall hydrophobicity.

Further analysis of the nucleic acids using stronger IP RP HPLC (DBAA/TBAA) under denaturing conditions (see Figure 5.2 right panel) shows a reduction in the influence of the sequence and chemistry (ribose vs deoxyribose) on the retention time. Each of the ss species from the corresponding dsDNA and dsRNA co-elutes. The now denatured DNA strands notably elute later than the corresponding denatured dsRNA stands; this is likely due to the increased hydrophobicity of thymine compared to uracil, which implies sequence effects are playing a role. Under these conditions, where size-dependent separations dominate, typical % differences in the calculated vs theoretical size of the ssRNA and dsRNA fragments for DBAA/TBAA (denatured to ssRNA) are <20%, as shown in Figure 5.2C. Due to its partial denaturation, the dsDNA ladder is unreliable for sizing in TEAA at 95 ºC; higher temperatures may resolve this, but few column ovens can reach temperatures over 100 ºC.

5.3.1 Accurate sizing of dsRNA biocontrols using non-denaturing IP RP HPLC

Following a comparison of nucleic acid standards under different IP RP HPLC conditions, further work was performed to analyse various dsRNA biocontrols and mRNA therapeutics. Different dsRNA biocontrols were expressed in microbial cells and purified with RNase T1 treatment to generate a range of dsRNAs (see Ross et al. 2024).[258] Analysis of the dsRNA biocontrols under non-denaturing conditions on both weak (TEAA) and strong (TBAA) is shown in Figure 5.3A.

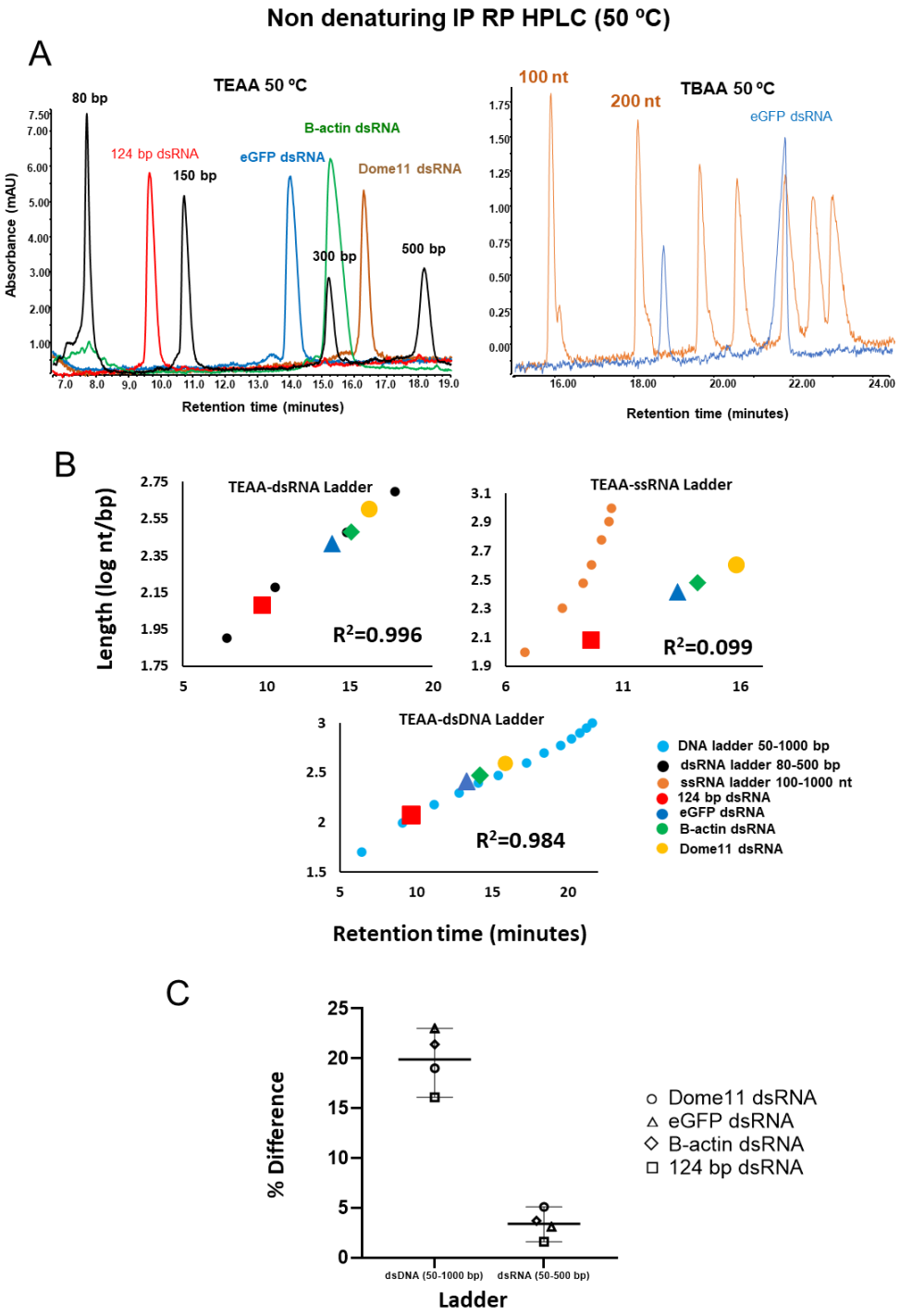


Figure 5.3: Comparison of sizing accuracy of dsRNA biocontrols against dsDNA, ssRNA and dsRNA ladders at *50 °C A) IP RP chromatogram of dsRNA biocontrols compared against dsRNA ladder in TEAA and a ssRNA ladder in TBAA at 50 °C. B) Graphs comparing the correlation between the dsRNA biocontrols to the ss and dsRNA ladders in TEAA at 50 °C. Correlations were calculated by adding the data points for the dsRNA biocontrols to each ladder series and calculating an R^2.C) Accuracy of sizing of dsRNA biocontrols compared to dsDNA and dsRNA ladders at 50 °C, the % difference of the calculated length vs theoretical length of the RNA is shown. IP RP HPLC separations were performed on a monolithic PS DVB capillary column) at a flow rate of 2.2 µl/min at 50 °C with UV analysis at 260 nm, mobile phase A 0.1 M TEAA, mobile phase B 0.1M TEAA 25% ACN. Separations were performed using gradient 8 (TEAA), 10 (DBAA), 12 (DBAA HFIP), and 14 (TBAA) see Chapter 2 section 2.5.3.*

The results show that the dsRNA elute near their expected retention times relative to the dsRNA ladder. Notably, the 124 bp fragment elutes between the 80 bp and 150 bp fragments, and the B-actin dsRNA (306 bp) elutes slightly later than the 300 bp ladder fragment. In the presence of strong ion pair reagents (DBAA/TBAA), all dsRNA biocontrols analysed, except for the smaller 120 bp fragment, consistently exhibit two peaks in the IP RP chromatogram (see Figure 5.3A right panel). Interestingly, this phenomenon was not observed in the dsRNA ladder, although small shoulders/peaks to the right-hand side of the peak are observed. The earlier of the two peaks has the expected retention time for the expected length of the dsRNA fragments. The cause of this discrepancy between the dsRNA ladder and dsRNA biocontrols remains unclear, as both the samples of dsRNA and the dsRNA ladder were purified using Monarch RNA cleanup kits prior to analysis. This split peak pattern has negative implications for using strong ion pair reagents to size these fragments under non-denaturing conditions and may reflect potential multimers or aggregates.

Further analysis showing the correlation between the dsRNA biocontrols and dsRNA ladder and the accuracy of sizing using TEAA under non-denaturing conditions is shown in Figure 5.3B. The results show a strong correlation coefficient (R^2>0.995) and high-resolution separation under non-denaturing conditions. Conversely, in TEAA, the fragments size poorly against the ssRNA ladder, closely matching the data in Figure 5.1 and previous studies [257]. Further analysis was performed to determine the length of the dsRNA biocontrols using linear regression of the nucleic acid ladder. The percentage difference between each fragment's size and the predicted size, derived from its retention time, is shown in Figure 5.3C and summarised in Table 5.1. The results show that under native conditions, TEAA is the most suitable ion pair reagent for sizing dsRNA biocontrols using the dsRNA ladder. Based on the differences in retention between single and double-stranded species, it is unsurprising that the sizing error associated with using the single-stranded ladder for sizing dsRNA is out by two orders of magnitudes and so has not been plotted. The difference in accuracy between sizing with the dsDNA and dsRNA ladders is significant; this is likely due to the difference in hydrophobicity between the two species due to differences in structure (A form vs B form helices) and the effects of the 2’ OH group present in the structure of DNA, causing it to be less suited for sizing the dsRNA fragments.

There was an attempt to accurately size the later eluting of the two peaks in TBAA for each fragment. As they elute within the gradient, the later peaks in DBAA and DBAA HFIP typically elute in the wash phase and cannot be incorporated into the sizing model. This data is shown in appendix section 7.6, Table 7.1. It was found that these peaks may be the result of dimers as they size approximately 1.8 to 2.2 times the expected length. Further analysis under denaturing conditions may reveal more about the nature of these later eluting peaks.

5.3.2 Accurate sizing of dsRNA biocontrols using denaturing IP RP HPLC

Following IP RP HPLC analysis under non-denaturing conditions, the dsRNA biocontrols were further analysed under denaturing conditions. Figure 5.4A shows the overlaid chromatograms for the dsRNA biocontrols in weak (TEAA left panel) and strong ion pair (DBAA right). In TEAA, as expected under denaturing conditions, the corresponding ss species are separated for the dsRNA ladder and dsRNA biocontrols. These results are consistent with data obtained from denatured DNA and dsRNA ladders (see Figure 5.2), highlighting the sequence effects on the overall hydrophobicity of the RNA, resulting in different RT for the same length ssRNA using TEAA. Analysis in the stronger IP reagent DBAA (see Figure 5.4A right panel) shows that under denaturing conditions, only a single peak is now observed in contrast with the two peaks observed under non-denaturing conditions (see Figure 5.3A). This observation suggests that the previously observed double peaks are likely attributable to aggregation, dimerisation, or some other weak secondary structure, which can be effectively denatured at elevated temperatures.

Further analysis was performed by plotting the retention time of each fragment in the nucleic acid ladder and the dsRNA to examine their correlation(see Figure 5.4B). In the case of TEAA, where some of the ladder peaks and all the dsRNA peaks are split, the mean of the two retention times was plotted prior to size calculation, the caveat to this being a loss of sizing accuracy due to the sequence dependence. The results show that using TEAA under denaturing conditions is markedly less accurate for accurate sizing of dsRNA compared to the previous data under non-denaturing conditions at 50 ºC. This is attributed to the impact of sequence-dependent separations on retention in weak IP RP HPLC.

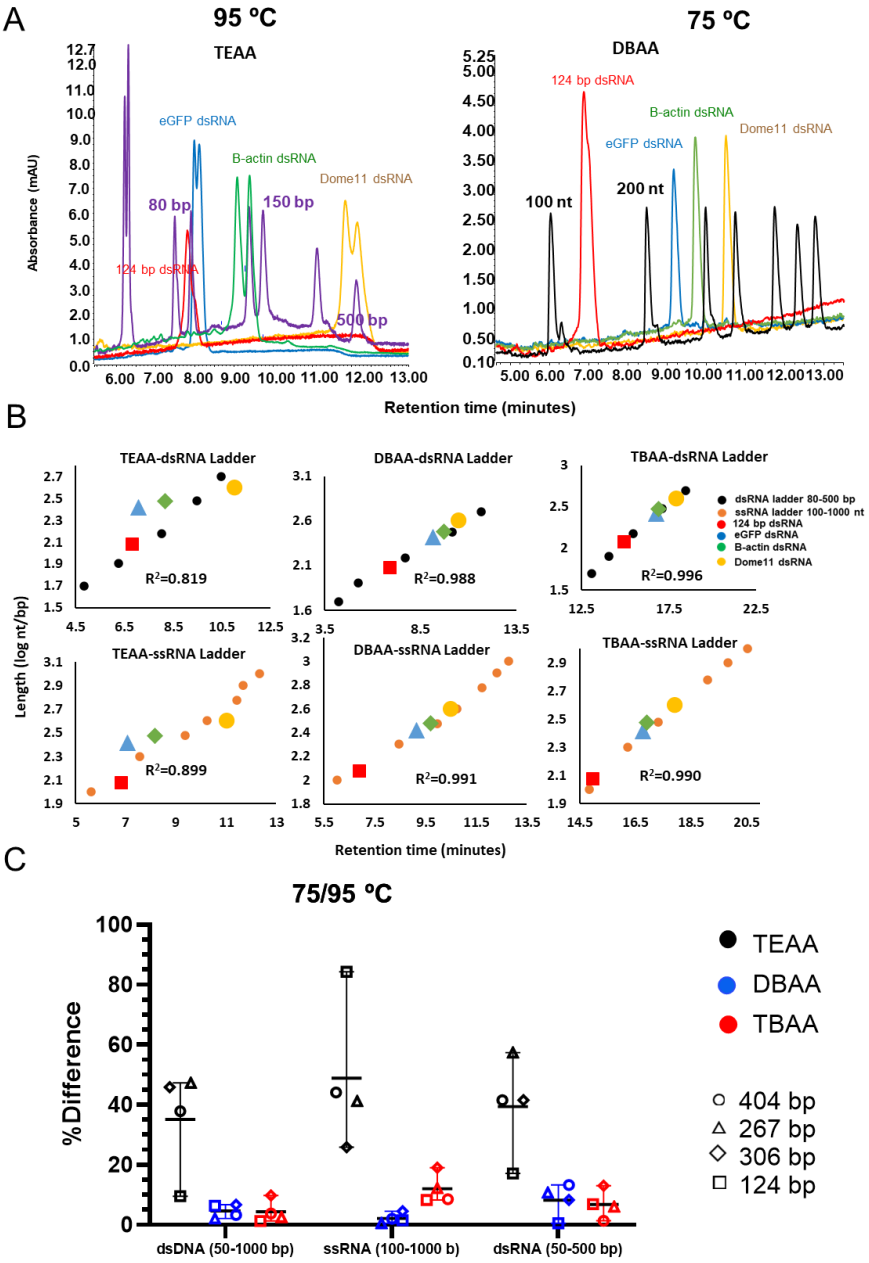


Figure 5.4: Comparison of sizing accuracy of dsRNA biocontrols using dsDNA, ssRNA and dsRNA ladders under denaturing conditions. *A) IP RP chromatogram of dsRNA biocontrols compared against a dsRNA ladder in TEAA and a ssRNA ladder in TBAA at 75/95 °C. B) Graphs comparing the correlation between the dsRNA biocontrols to the ss and dsRNA ladders in TEAA, DBAA and TBAA at 75/95 °C. Correlations were calculated by adding the data points for the dsRNA biocontrols to each ladder series and calculating an R2.C) Sizing accuracy of dsRNA biocontrols using dsDNA, ssRNA and dsRNA ladders at 75/95 °C all ion pairs, the % difference of the calculated length vs theoretical length of the RNA is shown. IP RP HPLC separations were performed on a monolithic PS DVB capillary column at a flow rate of 2.2 µl/min at 50 °C with UV analysis at 260 nm, except for TEAA separations, which were performed on a nonporous DNAPacRP column at a flow rate of 0.3 ml/min at 95 ºC. Separations were performed using gradient 9 (TEAA), 11 (DBAA), and 15 (TBAA) see Chapter 2 section 2.5.3.*

The results of the correlation analysis using stronger IP reagents (TBAA/DBAA) typically show R2 values >0.985, demonstrating that using stronger IP reagents reduces the influence of the RNA sequence on RT when analysing ssRNA under denaturing conditions. With the correlations between the dsRNA(corresponding ssRNA) ssRNA ladders determined, the sizing accuracy for each fragment was calculated against each ladder in each mobile phase composition (see Figure 5.4 C).

The results show that using TBAA/DBAA, an average of less than 10% deviation in predicted length from actual length was obtained for sizing the dsRNA biocontrols using either the dsDNA, ssRNA or dsRNA ladders. Interestingly, using the dsDNA ladder enables accurate sizing of the dsRNA biocontrol using strong IP RP HPLC under denaturing conditions. Therefore, despite the strands being denatured, the difference in the sequence/chemistry of the ssDNA and ssRNA does not significantly impact retention time; this is consistent with the findings of the ladder analysis study in Figure 5.2.

In summary, the results show that the analysis of dsRNA biocontrols using weak IP RP HPLC in conjunction with TEAA mobile phases under non-denaturing conditions and DBAA/TBAA mobile phases under denaturing temperatures can both be utilised for the accurate sizing and analysis of dsRNA biocontrols. A summary of the data obtained for each dsRNA in the three most accurate phase/temperature combinations for sizing is shown in Table 5.1.

Table 5.1: Comparison of sizing accuracy of dsRNA biocontrols using different ion pair reagent/ladder combinations.

Ion Pair	Temp ºC	Ladder	dsRNA	Predicted	Actual	%Error	Mean %Error
TEAA	50	dsRNA	124 bp	122	124	1.6	**3.4**
			eGFP	257	267	3.7	
			Beta-actin	315	306	3.1	
			Dome 11	383	404	5.1	
DBAA	75	ssRNA	124 bp	122	124	1.7	**2.0**
			eGFP	260	267	2.7	
			Beta-actin	314	306	2.6	
			Dome 11	408	404	1.0	
TBAA	75	dsDNA	124 bp	122	124	1.3	**4.4**
			eGFP	260	267	2.5	
			Beta-actin	276	306	9.8	
			Dome 11	419	404	3.8	

Due to the peak splitting phenomena, TEAA is the only suitable ion pair reagent for analysis at 50 °C. Under denaturing conditions, the performance of TBAA and DBAA is very similar when sizing using all three different nucleic acid ladders (dsRNA, dsDNA and ssRNA). However, visual inspection of the chromatography data for the ladders shows DBAA affords better resolution for larger fragments, so it may be more suitable for the analysis of larger dsRNA in the 500-1000 bp region, with the ssRNA ladder being the best for predicting the lengths of dsRNA fragments in the size range of this study.

5.3.3 Analysis of long mRNA using IP RP HPLC

Following the development and application of IP RP HPLC for sizing dsRNA fragments (100-500 bp), further analysis was performed to assess the feasibility of accurate sizing of larger mRNA therapeutics/vaccines. These mRNA typically range from 800-4500 nts in length and provide a significant analytical challenge to achieve high-

resolution separations and accurate sizing of such large RNAs. Initial work focussed on optimising the analysis of a commercial high-range RNA ladder (200-6000 nt) in various IP mobile phases under native and denaturing conditions, with particular attention to the resolution of larger fragments. The results of this gradient optimisation are presented in Figure 5.5

The results show that for the non-denaturing analysis using TEAA, most of the ssRNA species are separated (see Figure 5.5A). However, the 1500/2000 and 3000/4000 nt peaks are not fully baseline resolved. Under strong IP conditions, the loss of resolution of the larger RNA is evident where DBAA and TBAA exhibit complete loss of baseline resolution above 1000 nt. The analysis under denaturing conditions (see Figure 5.5B) shows an improvement in resolution using TEAA, where most RNAs are now baseline resolved. However, no significant improvement in resolution has been achieved for the larger RNAs in DBAA/TBAA. Such losses in resolution will impact the sizing accuracy of mRNAs in downstream studies. Although high-resolution separations were achieved using TEAA, previous results demonstrate that sequence-specific effects dominate these separations, and therefore, large RNA fragments may not elute in the correct order of size, which may need to be considered.

Further analysis was performed using four mRNAs of varying sizes, summarised in Table 5.2. mRNA was generated via IVT and provided by Dr Zoltan Kis' laboratory (University of Sheffield).

Table 5.2: mRNAs used for IP RP HPLC analysis

mRNA	Gene	Source	Size nt
Nluc	Nano luciferase	IVT	870
eGFP	GFP	IVT	960
Fluc	Firefly luciferase	Trilink	1929
CSP	Covid spike protein	IVT	4283

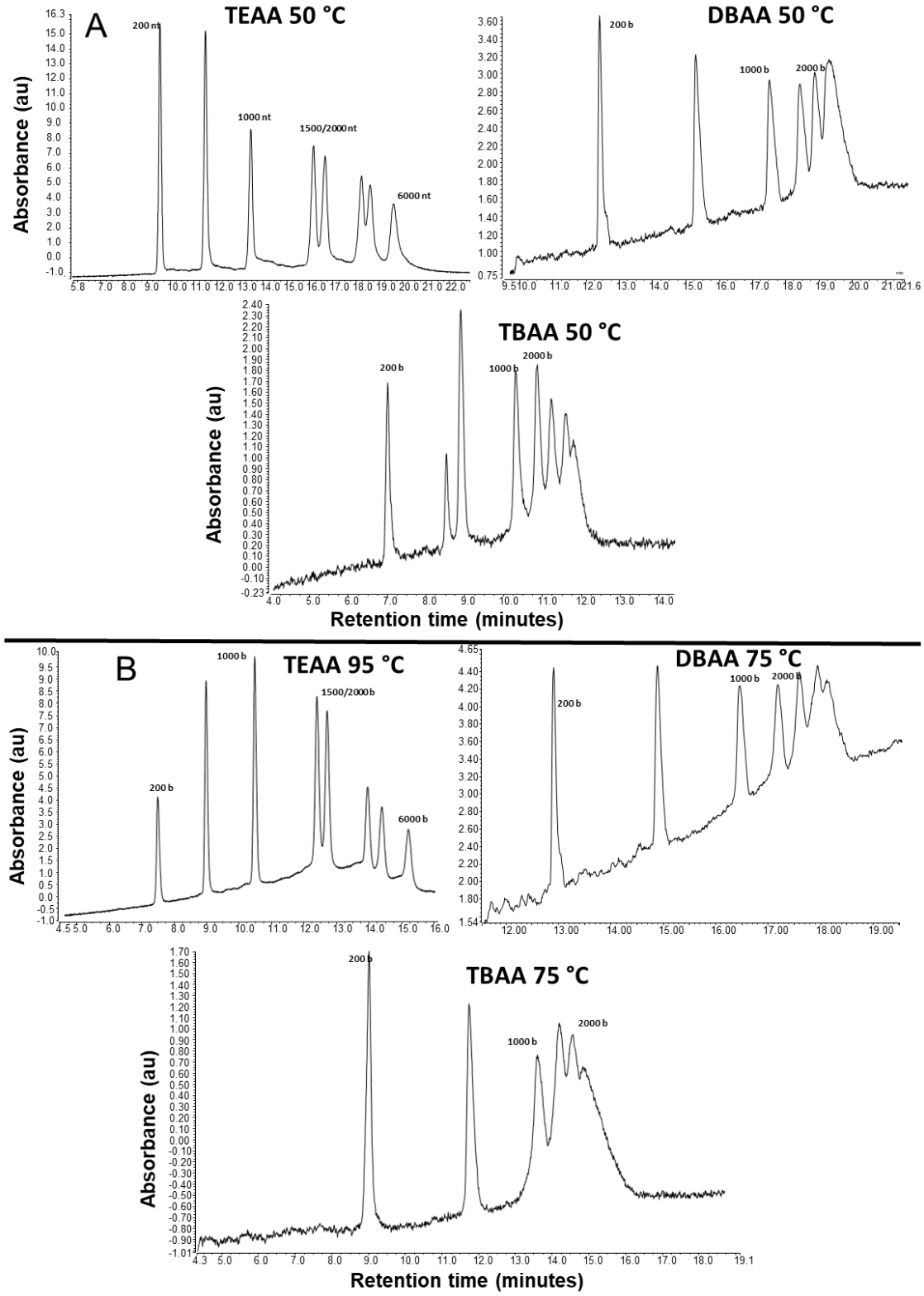


Figure 5.5: IP RP HPLC analysis of a high range ssRNA ladder (200-6000 bp, Riboruler HR, Thermo Fisher) in weak and strong ion-pair reagents *A) HPLC analysis at 50 °C B) HPLC analysis at 50/75 °C. IP RP HPLC separations were performed on a monolithic PS DVB capillary column at a flow rate of 2.2 µl/min with UV analysis at 260 nm, except for denaturing TEAA separations, which were performed on a DNAPac RP column at a flow rate of 0.3 ml/min at 95 ºC. Separations were performed using gradient 15/16 (TEAA), 17/18 (DBAA), and 19/20(TBAA) see Chapter 2, section 2.5.3.*

The different mRNA was analysed under non-denaturing conditions using TEAA, DBAA and TBAA (see Figure 5.6). Figure 5.6A left panel shows an overlay of the mRNA and the high-range ssRNA ladder. The results show that the elution order is not size-dependent for analysing large RNA and mRNA using TEAA at 50 °C. This is clearly shown in the elution of Nluc mRNA (870 nts), which elutes later than eGFP (Green fluorescent protein) mRNA (960 nts). Examining the base composition of each mRNA (see Table 5.3) shows that the NLuc mRNA has a significantly higher % of AU (the most hydrophobic bases) [259].

Table 5.3: Base composition comparison of the four mRNA

Fragment	%AU
Nluc mRNA 870 nt	64.3
eGFP mRNA 960 nt	43.9
Fluc mRNA 1929 nt	34.6
CSP mRNA 4183 nt	44.5

A higher composition of these bases in the sequence increases the overall hydrophobicity, resulting in increased retention compared to mRNA with a lower overall %AU composition but of a larger size. Further effects, such as overall structure, may also play a role in the overall hydrophobicity and differences seen between the mRNAs

Analysis of the mRNAs using a stronger IP (DBAA) mobile phase is shown in Figure 5.6 (right panel). The results show that under these conditions, the NLuc mRNA now elutes prior to the eGFP as expected, according to its size. However, interestingly, the results also show that the Covid Spike mRNA (4284 nt) is eluting earlier than the FLuc mRNA (1929 nt), which is less than half its length. The base composition does not explain this anomaly, as the %AU composition of the larger mRNA is more significant (see Table 5.3). At this stage, there is no explanation for this anomalous data; it is only seen in the DBAA data, the alternative strong IP reagent (TBAA), and the fragments elute in size order, except for Nluc which elutes slightly before GFP. A closer examination of the retention times reveals that Nluc is eluting in the right time frame relative to the ladder and that GFP is eluting slightly earlier; these off-trend

effects would indicate that sequence effects play a role even in this strong ion pair system.

Further analysis was performed to establish the extent of the correlation between the retention of the mRNA fragments and the high range ssRNA ladder (see Figures 5.6B/C). The data shows that only TBAA shows a strong linear correlation (>0.9) between retention time and the logarithm of length, with high range linearity observed in the later eluting larger RNAs. As expected, a lower correlation is observed in TEAA, where sequence-dependent separations dominate and have a considerable influence on the RT, leading to inaccurate sizing of the mRNA. Therefore, TBAA emerges as the only promising candidate for accurately sizing these larger mRNA species under non-denaturing conditions.

The sizing of the mRNA was performed differently from the previous dsRNA analysis. In calculating the equation of the line of best fit, peaks presumed to be co-eluting were excluded. For example, in the DBAA data, only the first six fragments, which were at least partially resolved, were included in the trendline. Although this approach meant attempting to size fragments outside of the established linear range, the calculations were done to assess the sizing accuracy of the larger fragments.

Figure 5.6C shows the sizing data for the four mRNAs. The sizing accuracy improves with the order of hydrophobicity of the three ion pairing reagents (TEAA<DBAA<TBAA). However, even the TBAA data shows inferior sizing accuracy compared to the dsRNA study, with an average deviation of more than 15% between predicted and actual lengths. It should be noted that the trendline is skewed by the inclusion of larger non-baseline resolved fragments in the case of TBAA. For instance, the 870 nt fragment is the least accurate prediction of the model. If only the first three fragments are included, the prediction accuracy for the 870 nt fragment improves.

In conclusion, the non-denaturing analysis and sizing indicate that both TEAA and DBAA are unsuitable due to the non-size-based elution order of mRNA. While TBAA offers better sizing accuracy, its resolution deteriorates significantly for RNA fragments over 2000 nt. Therefore, with the current methodologies, achieving simultaneous accurate sizing for fragments under 1000 nt and those over 2000-3000 nt may not be feasible.

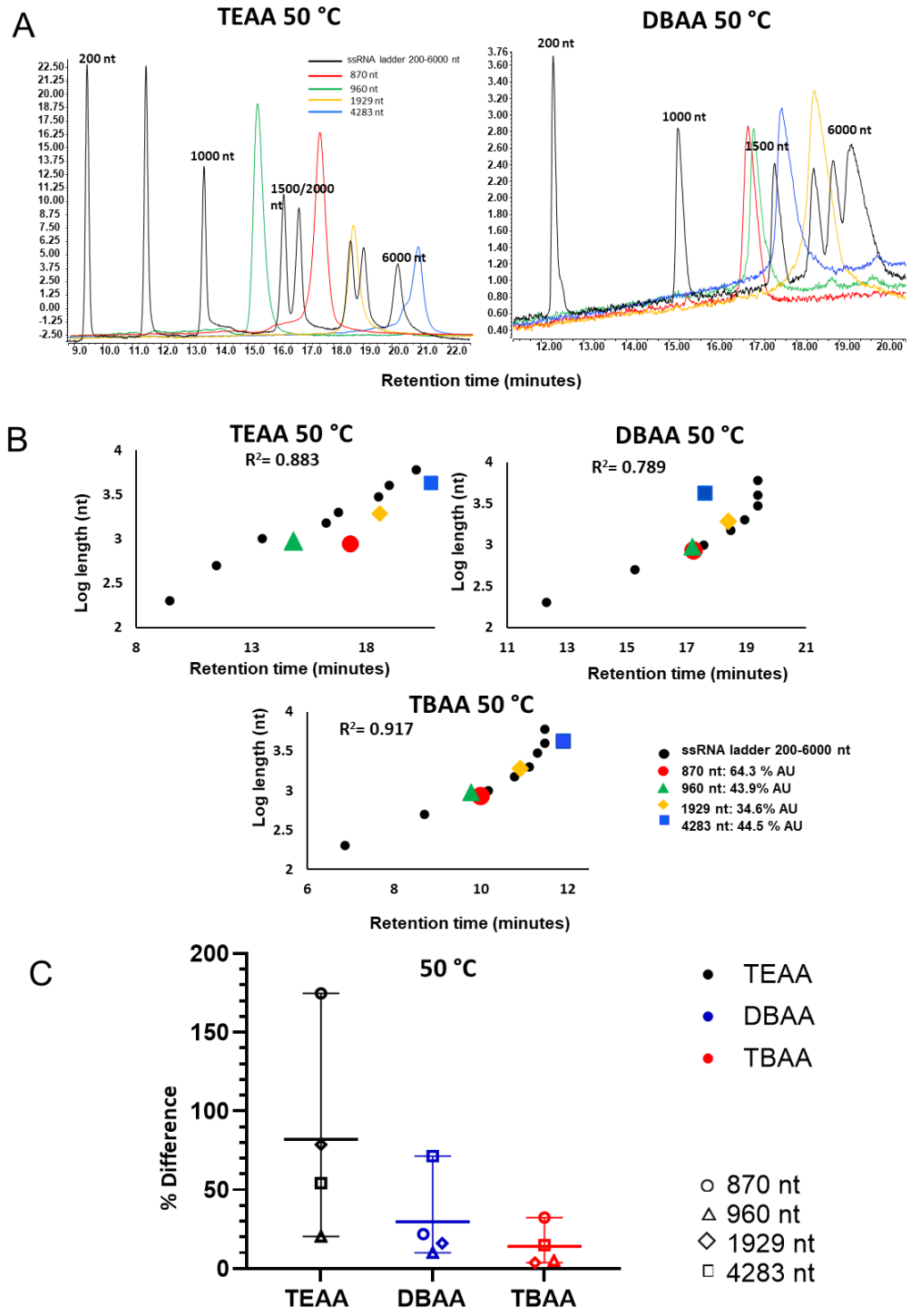


Figure 5.6 Comparison of sizing accuracy at 50 °C of mRNAs against a high range ssRNA ladder (200-6000 nt). *A) IP RP chromatogram of mRNAs overlaid with a high range ssRNA ladder in TEAA and DBAA at 50 °C. B) Graphs comparing the correlation between mRNAs and high-range ssRNA ladder in TEAA, DBAA and TBAA at 50 °C. Correlations were calculated by adding the data points for each mRNA species to the ladder series and calculating an R^2.C) Sizing accuracy of mRNAs using a high range ssRNA ladder at 50 °C all ion pairs, the % difference of the calculated length vs theoretical length of the RNA is shown. IP RP HPLC separations were performed on a monolithic PS DVB capillary column at a 2.2 µl/min flow rate at 50 °C with UV analysis at 260 nm. Separations were performed using gradient 15 (TEAA), 17 (DBAA), and 19(TBAA) see Chapter 2 section 2.5.3.*

5.3.4 Accurate sizing of mRNA using denaturing IP RP HPLC

The mRNA analysis was repeated under denaturing conditions to assess whether the resolution could be improved or whether anomalous data observed in the DBAA analysis could be linked to mRNA structure/conformation. Since all fragments analysed were single-stranded, significant changes in retention behaviour or sizing accuracy were not anticipated. The results of this analysis are shown in Figure 5.7, and the order of elution of the mRNA in TEAA under denaturing conditions is the same as previously observed for non-denaturing conditions. Figure 5.7A (right) illustrates the corresponding stronger ion pair analysis in DBAA, where the elution order is the same as previous non-denaturing conditions. The underlying cause of the early elution of the 4283 nt mRNA remains unexplained; if this phenomenon is conformation-based, it appears to be thermostable under these conditions.

The logarithm of length vs time plots was created for the mRNA fragments against the high-range ssRNA ladder in each ion pair reagent (see Figure 5.7B). Consistent with the non-denaturing data, only TBAA shows a significant linear correlation between the mRNA fragments and the ssRNA ladder, although GFP is still eluting slightly early relative to Nluc and the ladder. The sizing data for the denaturing analysis is presented in Figure 5.7C. For the TBAA data, the 4283 nt mRNA was excluded from the calculations due to poor resolution and inconsistent peak shape across multiple replicates. The sizing data indicates that TBAA is the best performer for sizing, like the non-denaturing conditions. However, the resolution of the larger RNA fragments in the strong ion pair conditions remains suboptimal.

While a curved fit could have been applied to estimate the sizes outside the linear range, the HPLC traces indicate a significant loss of resolution beyond 2000 nt, likely due to inefficient ion pairing. Although previous models have employed non-linear functions for size estimation, these approaches depend on consistent retention behaviour, which is not observed for fragments within this size range. No prior studies have successfully demonstrated accurate sizing in this range, likely due to the resolution loss and inconsistent separation between injections. Identifying the upper limit of the linear relationship between size and retention time has been a key objective of this chapter, and this limit has now been established.

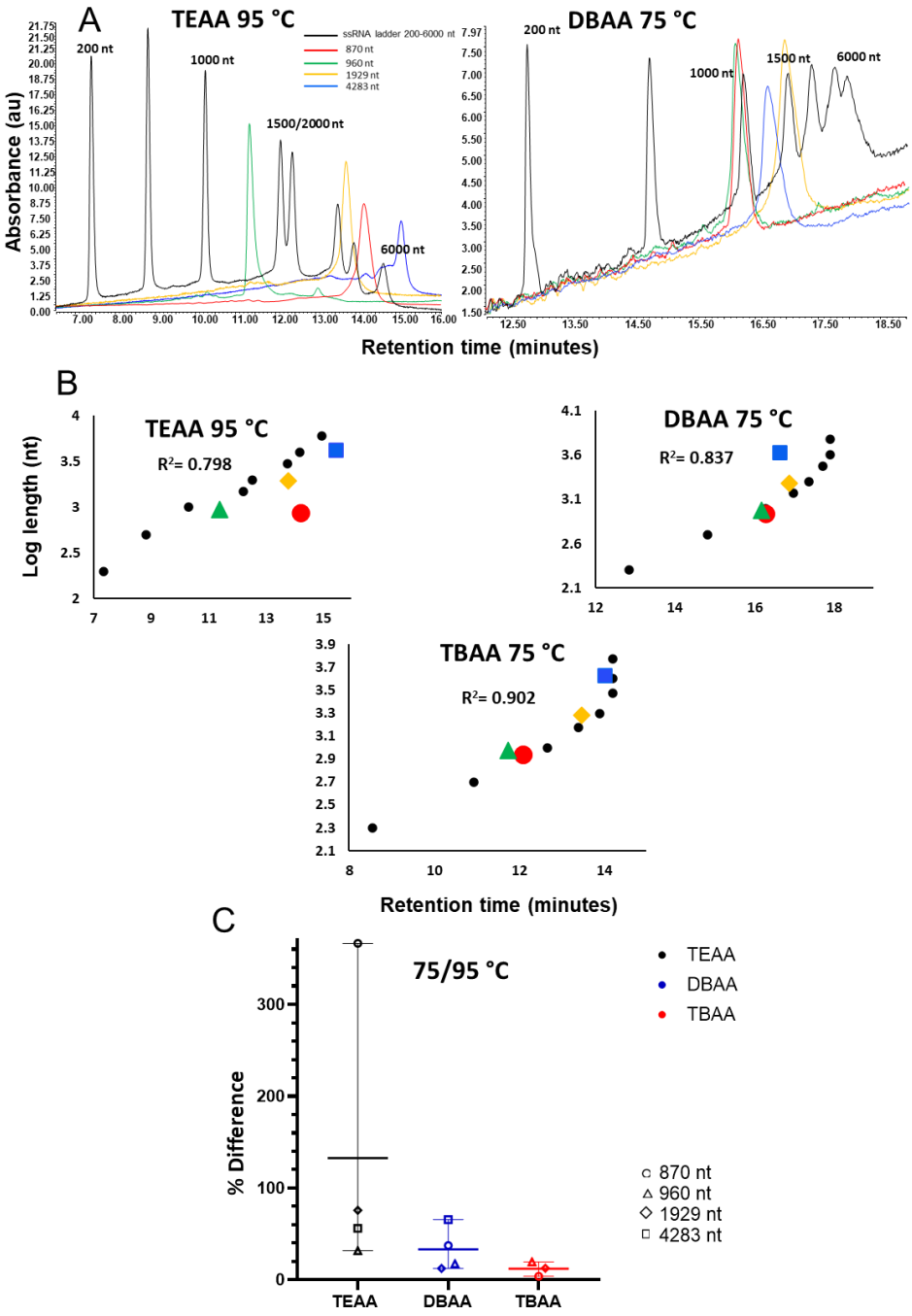


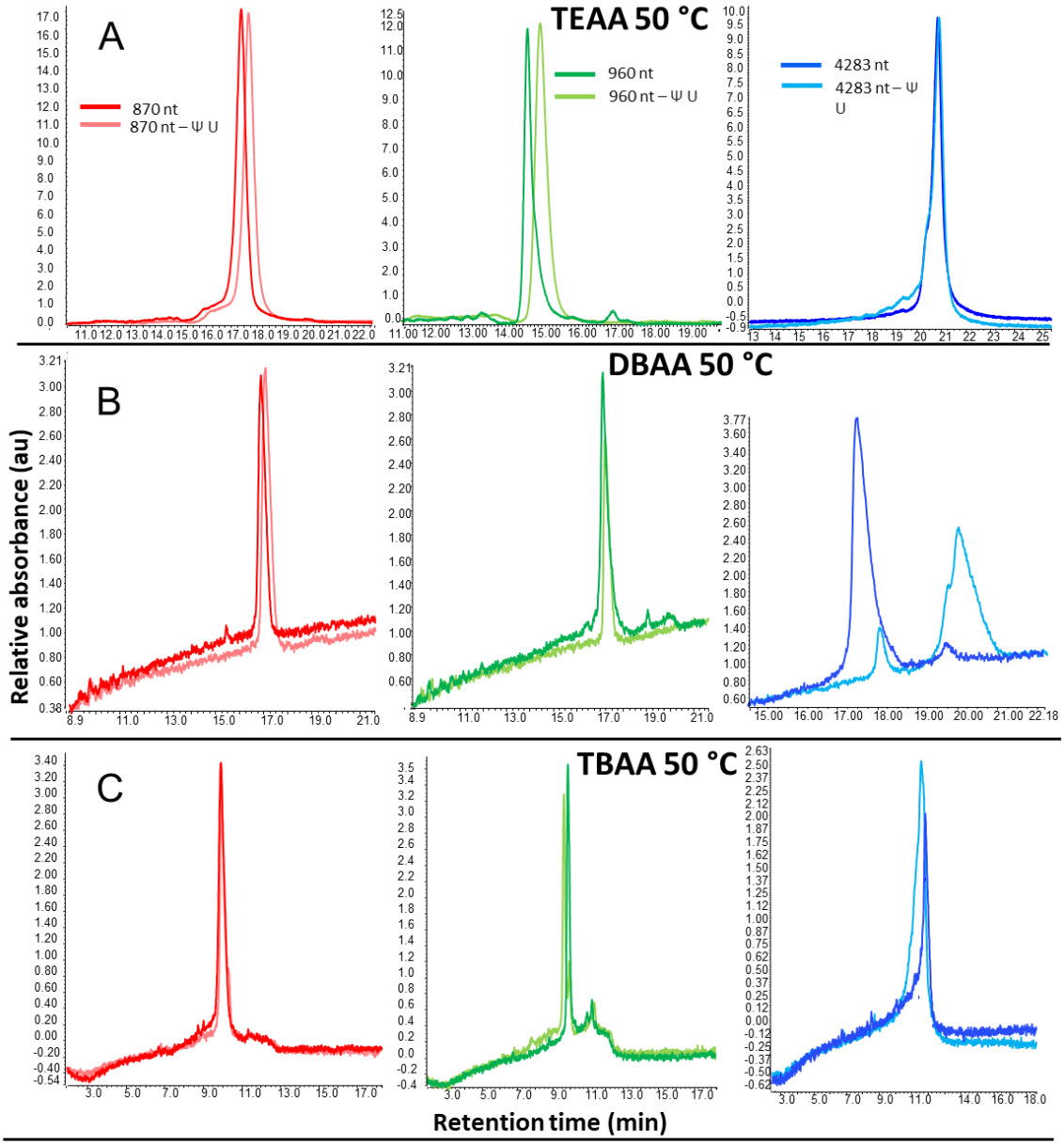
Figure 5.7: Comparison of sizing accuracy at 75/95 °C of mRNAs against a high-range ssRNA ladder (200-6000 nt). *A) IP RP chromatogram of mRNA overlayed with a high range ssRNA ladder in TEAA and DBAA at 75/95 °C. B) Graphs comparing the correlation between mRNAs to the high-range ssRNA ladder in TEAA, DBAA and TBAA at 75/95 °C. Correlations were calculated by combining the series of two ladders; for example, the DNA ladder series has 12 points, which was combined with the RNA ladder series to make a series with 19 points for which an R^2 was calculated. C) Sizing accuracy of mRNAs using a high range ssRNA ladder at 75/95 °C all ion pairs, the % difference of the calculated length vs theoretical length of the RNA is shown. IP RP HPLC separations were performed on a monolithic PS DVB capillary column at a flow rate of 2.2 µl/min at 50 °C with UV analysis at 260 nm, except for TEAA separations, which were performed on a DNAPac RP column at a flow rate of 0.3 ml/min at 95 ºC. Separations were performed using gradient 16 (TEAA), 18 (DBAA), and 20(TBAA) see Chapter 2 section 2.5.3.*

Given the lack of sizing accuracy achieved with the larger mRNA fragments, it is crucial to critically evaluate whether IP RP HPLC using TBAA can fulfil the objectives of this study and serve as a reliable method for accurate sizing of mRNA, particularly considering the resolution limitations for larger fragments. Therapeutic mRNAs, such as the mRNA for the COVID spike protein, typically span several thousand nucleotides. The fact that the best-performing assay fails for only mRNA within the therapeutic size range suggests that size-based analysis using IP RP HPLC is not accurate nor adequate for mRNA species of this size.

5.3.5 Comparison of the retention behaviour of modified and unmodified mRNAs using IP RP HPLC

As discussed in Chapter 1, Section 1.2.3, there is significant interest in utilising mRNAs with modified nucleotides to improve therapeutic outcomes. This necessitates adjustments in analytical methods to accommodate the chemical variations introduced by these structural modifications. Although achieving sizing accuracy for larger mRNAs proved challenging, this study sought to explore the effect of chemical modifications on the accuracy of mRNA sizing using IP RP HPLC. Specifically, the investigation focused on whether modified and unmodified mRNAs of the same size would co-elute, thereby indicating size-based separation. Differences in retention time would indicate the predominantly sequence-based mechanism of separation. Three mRNAs, with lengths of 870, 960, and 4283 nts, with either uridine or N1-methyl pseudo uridine, were provided by the research group of Dr Zoltan Kis (University of Sheffield) and analysed using IP RP HPLC in TEAA, DBAA, and TBAA under both native and denaturing conditions. These mRNAs were sized against the ssRNA ladder, following the same methodology as the previous mRNA study.

The data from the analysis under non-denaturing conditions is presented in Figure 5.8. Panels A-C show the three mRNAs overlaid with N1-methyl pseudo-uridine-containing counterparts across the three ion pair reagents. The results show that for eGFP and NLuc mRNA, a difference in retention time is observed between the modified and unmodified mRNA in TEAA but not in the two stronger IP reagents. These results are consistent with previous data where IP RP HPLC analysis using TEAA results in more sequence-dependent separations of single-stranded RNA.



D

mRNA	Mod vs unmod predictions (nucleotides)								
	TEAA			DBAA			TBAA		
	M	U	D	M	U	D	M	U	D
870 nt	2363	2390	27	999	1060	61	1194	1152	42
960 nt	1324	1156	168	1077	1056	21	920	1009	89
4283 nt	6719	6603	116	2761	1225	201	3055	3645	590

Figure 5.8: Comparison of RP IP HPLC analysis and size prediction of modified and unmodified mRNAs using a high range ssRNA at 50 °C *A) HPLC analysis in TEAA. B) HPLC analysis in DBAA.C) HPLC analysis in TBAA. D) Table comparing the difference in predicted size in nucleotides between modified (M) and unmodified (U) mRNA in TEAA, DBAA and TBAA using an ssRNA ladder. IP RP HPLC separations were performed on a monolithic PS DVB capillary column at a 2.2 µl/min flow rate at 50 °C with UV analysis at 260 nm. Separations were performed using gradient 15 (TEAA), 17 (DBAA), and 19(TBAA) see Chapter 2 section 2.5.3.*

Interestingly, the larger CSP (Covid Spike Protein)mRNA does not exhibit the same behaviour as its smaller counterparts across any ion pair reagents. In TEAA, the retention difference between the modified and unmodified fragments is minimal. When examining the sizing estimates shown in Figure 5.8D, the difference (D) between the predicted lengths of the modified (M) and unmodified (U) CSP mRNA is only 116 nt. This data does not suggest size-dependent separation; instead, it implies that at this size, the overall hydrophobicity of the mRNA is so high that the hydrophobicity difference between uridine and N1-methyl pseudouridine becomes negligible. However, it should be noted that the lower resolution of the larger RNA will also play a role.

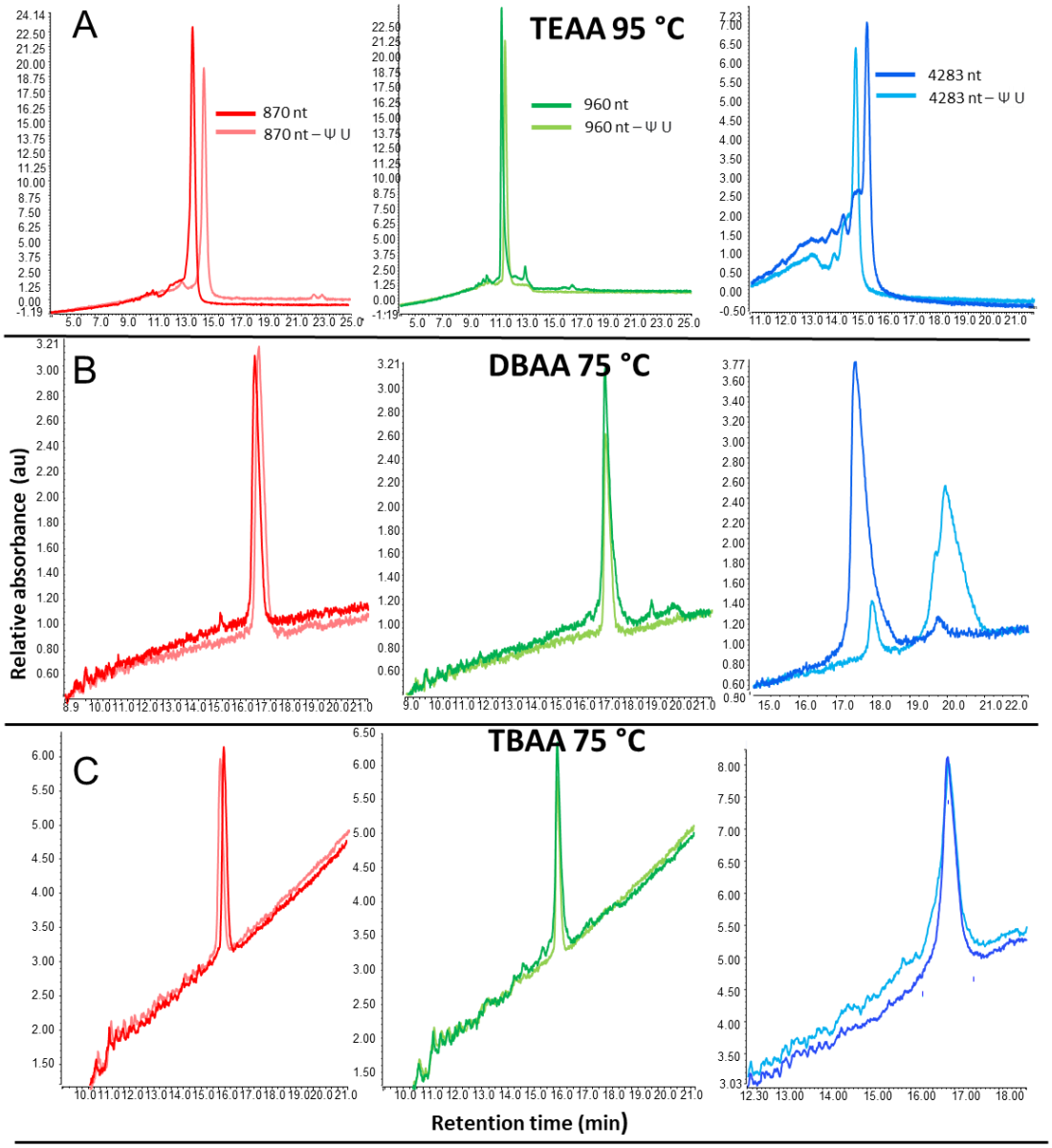
Examining the DBAA data in Figure 5.8B reveals that the largest CSP mRNA again behaves as an outlier, exhibiting a substantial difference in retention time between the modified and unmodified strands. Despite this anomalous behaviour, the analysis has provided insights that may explain the unusual behaviour observed for this mRNA in the sizing study (section 5.3.3). The chromatogram for the modified 4283 nt mRNA reveals two distinct peaks: a major peak, sized at 2761 nt, and a minor peak, sized at 1300 nt. Interestingly, this minor peak elutes within a retention window similar to the unmodified mRNA. Additionally, close inspection of the baseline for the unmodified strand reveals an additional, later-eluting peak, which appears within a time window like the larger peak in the trace of the modified strand. It is proposed that the anomalous behaviour could be due to different conformations of the mRNA, the abundance of which is influenced by the absence or presence of chemical modifications. Further studies are required to investigate this hypothesis.

The analysis of mRNA using TBAA (see Figure 5.8C) shows that the modified and unmodified eGFP, NLuc and CSP mRNA co-elute and, therefore, does not affect the sizing or influence the retention time of these mRNAs. However, considerable variation in retention time and chromatographic quality (e.g., peak widths and intensity) across different replicates of CSP mRNA was observed in TBAA; therefore, it can be concluded that TBAA is unsuitable for analysing these larger mRNAs due to the inconsistency and lack of reproducibility in the chromatography data.

5.3.6 Comparison of the retention behaviour of modified and unmodified mRNAs in denaturing IP RP HPLC

Despite the mRNA sizing study indicating minimal differences between denaturing and non-denaturing conditions for these species, the comparison was repeated at denaturing temperatures (see Figure 5.9). A comparison of the non-denaturing (Figure 5.8) with denaturing analysis (Figure 5.9) reveals minimal differences between the two conditions. The most notable differences are the later retention time of the unmodified mRNA compared to the modified mRNA in TEAA and the similar retention times of the Nluc and eGFP mRNA in DBAA and TBAA. These findings suggest that denaturing conditions do not significantly alter the retention characteristics of these mRNA fragments.

In conclusion, the IP RP HPLC analysis demonstrates that for mRNAs (under 1000 nt), the N1-methyl pseudo uridine modification increases the retention time of the mRNA using TEAA under both native and denaturing conditions. Therefore, this results in the prediction of different sizes of the mRNA determined using a high-range ssRNA ladder. In contrast, analyses conducted with stronger ion pair reagents under denaturing conditions result in similar retention times for both modified and unmodified. Therefore, the chemical modification does not affect the accuracy of mRNA sizing in conjunction with the high-range ssRNA ladder using strong IP RP HPLC (Figure 5.8D and Figure 5.9D). These results are consistent with previous data that show that strong ion pair reagents result in more accurate sizing of mRNAs. However, the application for analysis of larger mRNA >1000 nt is limited due to the low-resolution separations where close or co-elution of these mRNAs are observed.



D

mRNA	Mod vs unmod predictions (nucleotides)								
	TEAA			DBAA			TBAA		
	M	U	D	M	U	D	M	U	D
870 nt	5345	4060	1285	1039	1196	157	716	905	189
960 nt	1407	1264	143	1081	1129	38	803	772	31
4283 nt	6202	6687	485	3199	1471	1728	2200	2150	50

Figure 5.9: Comparison of RP IP HPLC analysis and size prediction of modified and unmodified mRNAs using a high range ssRNA at 75.95 °C *A) HPLC analysis in TEAA. B) HPLC analysis in DBAA.C) HPLC analysis in TBAA. D) Table comparing the difference in predicted size (D) in nucleotides between modified (M) and unmodified (U) mRNA in TEAA, DBAA and TBAA using an ssRNA ladder. IP RP HPLC separations were performed on a monolithic PS DVB capillary column at a flow rate of 2.2 µl/min at 75 °C with UV analysis at 260 nm, except for TEAA separations, which were performed on a nonporous DNAPac RP column at a flow rate of 0.3 ml/min at 95 ºC. Separations were performed using gradient 16 (TEAA), 18 (DBAA), and 20(TBAA) see Chapter 2 section 2.5.3.*

It is hypothesised that this issue may be related to column capacity. In addition to their intrinsic strength, the ion pair reagents are used at varying concentrations: TEAA at 100 mM, DBAA at 15 mM, and TBAA at 5 mM. As outlined in previous studies, the concentration of the ion pair reagent influences the capacity factor. It is plausible that larger fragments require a higher concentration of ion pair reagent to adequately bind the nucleic acids to the column and maintain resolution. However, low concentrations of the strong IP reagent are required in conjunction with a high % organic solvent to elute these large nucleic acids from the stationary phase. Further research is necessary to test this hypothesis and determine whether increasing the concentration of the ion pair reagent in conjunction with using alternative organic solvents could enhance the resolution of larger mRNA fragments.

5.4 Conclusions

One of the most significant challenges facing RNA platforms is rapid, robust quality control methods which identify the product, determine its integrity, and identify by-products or leftover materials from batch production. The primary aim of this study was to evaluate the suitability of IP RP HPLC chromatography for the accurate sizing of dsRNA biocontrols and mRNA. The ability to accurately size the mRNA assists with the identity testing of the manufactured RNA and identification (and sizing) of potential by-products.

The results demonstrate that dsRNA biocontrols can be accurately sized using IP RP HPLC under non-denaturing and denaturing conditions, with a sizing error of less than 5% under native and denaturing conditions using alternative nucleic acid ladders. It can be concluded that the most accurate sizing method was mobile phase DBAA under denaturing conditions in conjunction with a ssRNA ladder. This method is precise and efficient, with the entire process completed in under twenty minutes. Moreover, as evidenced by previous HPLC studies, it can be seamlessly integrated into existing dsRNA production platforms, making it a practical tool for ensuring product quality in biocontrol applications.

The analysis and accurate sizing of mRNA using IP RP HPLC is more challenging and was shown to be unsuitable for the analysis of large mRNA fragments exceeding 1000

nts in length. Given that therapeutic mRNAs used in vaccines, such as the COVID-19 spike protein mRNA, are typically much larger than 1000 nts, IP RP HPLC is not currently viable as a size-dependent quality control assay for these mRNAs. Although IP RP HPLC using TEAA (non-denaturing and denaturing conditions) enabled the separation of the high-range ssRNA ladder, the predominant sequence dependence in this separation mode results in inaccurate mRNA sizing. This was clearly demonstrated by the later elution of Nluc mRNA (870 nt) compared to eGFP mRNA (960 nt) and the separation of chemically modified mRNA from unmodified mRNA of the same size.

In contrast, the stronger IP reagents TBAA and DBAA result in predominantly size-dependent separations of mRNA, with slight variation in retention between modified and unmodified mRNAs; however, there is a loss of resolution of larger mRNAs (>1000nt), therefore limiting applications to only smaller mRNA for accurate sizing using this approach.

6 Chapter 6 Conclusions and recommendations for future work

6.1 **Conclusions**

Initial research presented in Chapters 3 and 4 aimed to develop and optimise critical aspects of the SELEX workflow prior to the generation of ssDNA aptamers against MDA-9. A significant caveat associated with SELEX experiments is their failure due to issues associated with PCR amplification. As the rounds progress through SELEX, PCR by-products can dominate. In addition, current methods used to isolate ssDNA can result in low yields and the purification of ssDNA from these PCR by-products, hampering the SELEX process. Initial work, therefore, focussed on optimising the ssDNA library design and PCR conditions to minimise PCR by-products. Using these optimised conditions, successful amplification of the ssDNA library was achieved, resulting in a higher quality dsDNA (no by-products) compared to an alternative ssDNA library design and PCR amplification conditions used in previous SELEX methods in the literature.

Re-generation of single-stranded DNA is a key and often fatal step in SELEX protocols, with current methods of exonuclease digestion and streptavidin affinity chromatography either delivering poor yields (52-69% for the former and 50-55% for the latter) or failing to eliminate PCR by-products.[231] This creates significant issues early in SELEX protocols when the copy number of species is low, and inefficient ssDNA re-generation can lead to a loss of information from the pool. Additionally, the propensity of these methods to carry through by-products through to later rounds of selection facilitates their growth.

IP RP HPLC is a powerful method for the analysis of nucleic acids. In this Thesis, I have utilised and developed IP RP HPLC in several studies. IP RP HPLC was used to optimise the SELEX methodology and develop a novel method for ssDNA regeneration, separating the two strands of a PCR product where the reverse primer contained a hydrophobic 5’ tag, allowing for the isolation of ssDNA with an 80% yield, superior to current methods. The method was validated using ESI mass spectrometry and fluorescence detection of the Texas red 5’ tag. Following the development of this approach to purify ssDNA, this was subsequently employed to generate ssDNA aptamers to MDA-9. This novel method can be used on antiquated

HPLC systems, which are ubiquitous in research institutions, allowing for easy implementation into SELEX workflows, improving success rates, and generating more aptamers.

6.1.1 Generation of ssDNA aptamers against Syntenin (MDA-9)

MDA-9 is a differentially regulated prometastatic gene with elevated expression in advanced stages of Melanoma. MDA-9 is a PDZ protein that binds the cytoplasmic C-terminal FYA motif of the syndecans, a widely distributed cytosolic protein. PDZ domains interact with several proteins, many critical regulators of signalling cascades in cancer. MDA-9 role in cancer metastasis and angiogenesis has led to its selection as a target in this project. Recombinant MDA-9 was successfully cloned into *E. coli* cells prior to overexpression and purification.

Initial studies using SELEX to obtain ssDNA aptamers against MDA-9 were unsuccessful due to the poor-quality PCR product obtained in later selection rounds. These results demonstrate that although initial optimisation of the PCR and library design was performed, this did not prevent the poor-quality PCR product from being obtained in later selection rounds. Therefore, alterations to the SELEX protocol were developed to minimise the number of stages where PCR amplification is used, relying on the ability of next-generation sequencing methods to sequence the final aptamer pool effectively. This modified approach (three rounds of selection of ssDNA aptamers without PCR amplification and three rounds of selection with amplification) was performed prior to next-generation sequencing and bioinformatics approaches to identify potential ssDNA aptamers.

Twelve ssDNA sequences identified from the sequencing analysis were synthesised before determining their binding affinity to MDA-9 using isothermal titration calorimetry (ITC). The results showed that three different aptamer sequences had binding affinities in the 25-600 nM range, with one high-affinity aptamer sequence identified. These results demonstrate the successful isolation of ssDNA aptamers against recombinant MDA-9. A wide range of further work is proposed to build upon identifying these initial ssDNA aptamers, including screening alternative sequences identified from NGS for binding to MDA-9 and using truncation to enhance the binding affinity of identified aptamers and determine the required minimum OGN sequence. Furthermore, it should

be noted that in this study, recombinant MDA-9 expressed in *E. coli* was used rather than human endogenous MDA-9. Therefore, further studies on MDA-9 expressed in mammalian cells should be performed to compare the binding affinity to recombinant MDA-9 expressed in *E. coli*.

6.1.2 Analysis of mRNA therapeutics and dsRNA biocontrols using IP RP HPLC

RNA therapeutics and RNA biocontrols are incredibly promising alternatives to small molecule-based platforms. Their success has already been proven both during the COVID pandemic and by the approval of Ledprona as a dsRNA biocontrol. With success comes greater demand for analytical characterisation of the RNA for batch release, the measurement of critical quality attributes and final commercial release. Therefore, there is a need for rapid, reliable methods for qualitative analysis of these molecules.

Current methods for qualitatively analysing these molecules include gel electrophoresis and anion exchange chromatography; however, they do not afford efficient size-based resolution due to structure and sequence effects. This study aimed to develop and optimise IP RP HPLC to analyse and accurately size mRNA therapeutics and dsRNA biocontrols. A wide range of alternative IP reagents were analysed, and IP RP HPLC was performed under denaturing and non-denaturing conditions. Analysis of a range of nucleic acid markers, including dsRNA, dsDNA and ssRNA, was used to assess the upper limit of the linear relationship between retention time and the logarithm of length (in base pairs/nucleotides) and compared to previous studies. The ladders were used to estimate the size of four dsRNA biocontrols based on their retention times using linear regression: a 124 bp dsRNA, eGFP dsRNA (267 bp), β-actin dsRNA (306 bp) and Dome11 dsRNA (404 bp). The major conclusions are that dsRNA can be accurately sized to within 5% error in the 124-404 bp region in TEAA under non-denaturing conditions and within 2% in DBAA under denaturing conditions.

These methods offer robust sizing of dsRNAs and will facilitate rapid QC methods, increasing the efficiency of production platforms. This will benefit the industry significantly, increasing the speed at which new dsRNA biocontrols are rolled out.

Analysis of mRNA therapeutics and accurate sizing using IP RP HPLC was more challenging than their smaller double-stranded counterparts. Large mRNA (1000-6000 nts) was separated using TEAA at elevated temperatures. However, under these conditions, although separation of the RNAs was achieved, the separation mechanism was predominantly sequence-based and, therefore, unsuitable for accurate mRNA sizing. This was demonstrated with the analysis of a range of mRNAs of different sizes and sequences where elution was not in the order of size (nts) of the mRNAs. Analysis of large RNAs using alternative stronger IP reagents showed that more size-dependent separations were achieved.

In DBAA, it was found that larger RNAs could be sized within 15% accuracy at 75 ºC up to 2000 nt, with the caveat that typically, under these conditions, a loss of resolution above 1500-2000 nt prevents accurate sizing of larger therapeutic mRNAs as the mRNAs either elute with very similar retention times or co-elute. Therefore, further work is required, focussing on increasing the resolution under stronger IP RP HPLC conditions to enable the separation of larger mRNA molecules, which is discussed below.

6.2 **Future Work**

The results presented in Chapter 4 demonstrated the successful isolation of ssDNA aptamers against recombinant MDA-9. However, a wide range of further work is proposed to build upon identifying these initial ssDNA aptamers. Previous work has shown that removing the primer sites in ssDNA aptamers increases the binding affinity of identified aptamers [253]. Therefore, further work should focus on ITC analysis of the ssDNA aptamer sequences identified from NGS for binding to MDA-9, without the primer regions, to assess their effect on MDA-9 binding. Furthermore, such studies can determine the minimal sequence required for binding to MDA-9. Truncation can boost affinity in libraries where the priming regions are not playing a significant role in binding; if they interact with the core region, affinity can be dramatically reduced [260,261].

Further binding studies should be carried out with ITC and at least one other biophysical technique to validate the binding affinity data from the ITC analysis. SPR would be an ideal alternative technique as it is commonly used for affinity screening and is orthogonal, using light scattering to determine binding as opposed to enthalpy

in ITC. MST, relying on the mobility of bound vs unbound aptamers would also be appropriate. Following the priming region, species with an affinity for the target can be re-screened using Fast Aptamer to find sequences within the same family in the library for affinity screening. It may be that species from the same cluster as one of the sequences screened with lower read counts have greater affinity for the target. When several sequences with affinity have been identified, screening for specificity against the target must be the next stage. Counter-screening candidates include the magnetic bead substrate and abundant cellular proteins. Candidates that emerge from the counter screening must then be optimised for any in vivo application, considering that this section of the Thesis's main aim was to isolate aptamers as antagonists to MDA-9, thereby acting as an anti-cancer therapeutic.

One of the simplest ways to increase nuclease resistance is the thiolation of the non-bridging oxygen atoms in the phosphate backbone. However, as discussed in Chapter 1, the major caveat is that if this is performed after selection, this may change how the aptamer folds and could completely nullify its affinity for the target. However, complete thiolation of the phosphate backbone is not, in fact, necessary to increase nuclease resistance; previous studies have demonstrated that partial thiolation yields significant exonuclease resistance [262].

With this in mind, and the fact that exonucleases cleave from the 5’ and 3’ ends inwards, a study involving both affinity screening and nuclease digestion of an aptamer with an increasing number of phosphorthioate groups would assess how much nuclease resistance could be achieved without a significant impact on affinity. A similar approach could be taken with other modifications, but phosphorthioate modifications are very economical.

6.2.1 **Optimisation of SELEX methodology**

The results of the SELEX study demonstrated that although several aptamer candidates were identified, the protocol requires further optimisation. There are two reasons for this conclusion; the first is the poor quality of PCR products during SELEX and sequencing preparation, which contributed to the lower quality of the sequencing data compared to the random ssDNA library. The first step in future work is to assess what role the primers played in forming PCR by-products; the truncation study will

achieve this. If it is found that the primers are playing a significant role in binding, further optimisation of the PCR steps may be required; increasing the annealing temperature may destabilise semi-stable by-products.

Further optimisation of next-generation sequencing could be performed, where it is proposed that designing the fixed regions of the ssDNA to match the adaptors of the sequencing provider would eliminate the need for a post-ligation amplification, one of the critical drivers of by-product formation in this study. However, a caveat of this approach is that unique barcodes cannot be used for each sequence. Therefore, only one library can be sequenced on one chip, making sequencing multiple SELEX rounds more expensive [263].

The second major flaw of all SELEX methodologies employed in this study was the lack of enrichment after multiple selection rounds. During SELEX protocols, the infamous “black box” problem was repeatedly encountered; it was unknown whether the lack of enrichment was due to a genuine lack of species with affinity for the target or some other problem within the workflow. There are, in fact, SELEX methods that do offer some insight into the binding stage; these include labelling strategies such as “flu-mag” SELEX. However, when attaching a fluorescent label to the 5’ end of a library, how this will affect folding must be considered[238]. Capillary electrophoresis SELEX may also offer an insight into the proportion of binders in the library at each round as this approach separates the bound from unbound ssDNA during the electrophoresis. The ability to directly measure and quantify the amount of ssDNA binding to the protein target will provide further insight during the SELEX procedure.

6.2.2 **Analysis of mRNA therapeutics using IP HPLC**

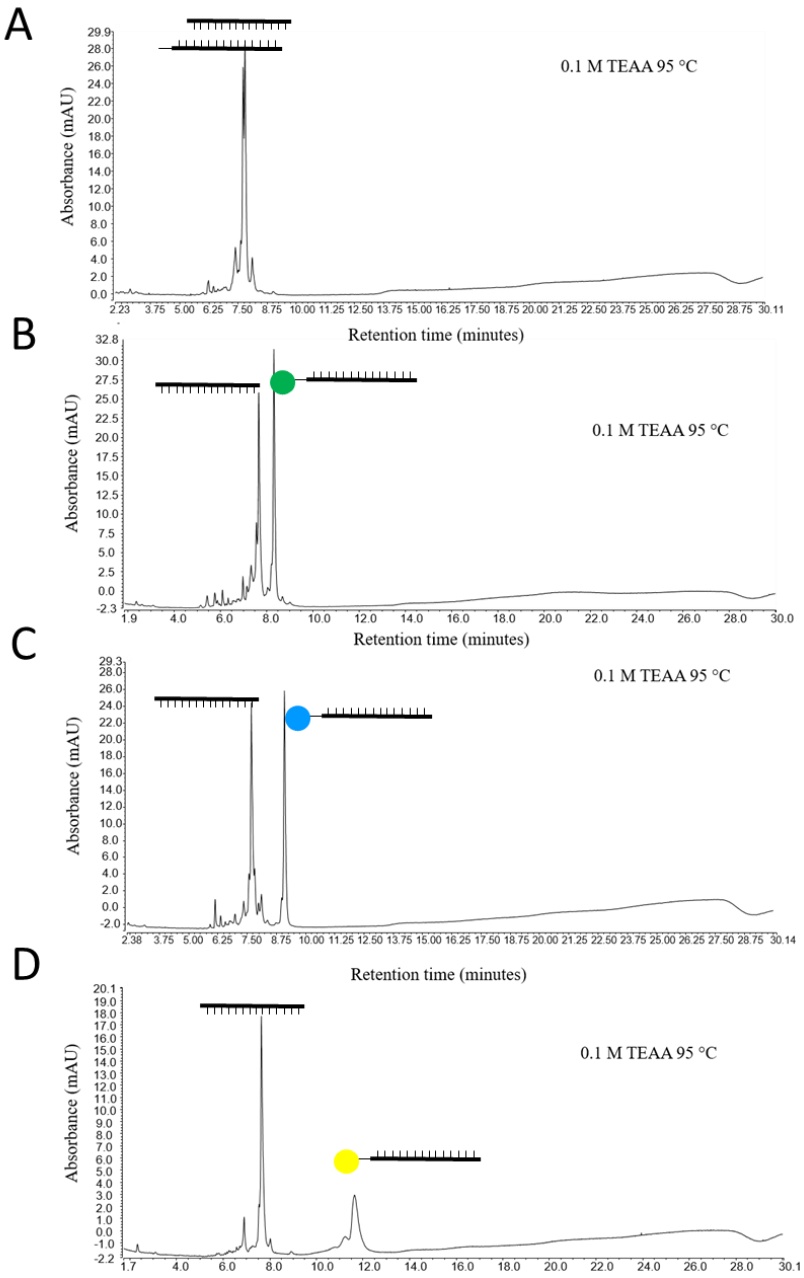
RNA platforms require robust quality control methods; the dsRNA biocontrol study satisfies this criterion, offering rapid, accurate sizing. However, further work is required to achieve the same outcome in mRNA analysis. IP RP HPLC analysis of mRNA therapeutics has been established for the routine analysis of mRNA integrity, typically using TEAA. However, the ability to accurately size mRNA presents significant challenges. The results in Chapter 5 demonstrate that although separation of large-sized mRNA fragments can be achieved using TEAA (200-6000 nts), loss of resolution

for mRNAs >1000-1500 was observed. This loss of resolution was particularly evident using the stronger ion pair reagents, DBAA and TBAA. The notable difference other than hydrophobicity between these and TEAA is the concentration at which they are used. TBAA at 5 mM and DBAA at 15 mM are considerably lower concentrations than TEAA at 100 mM. In the work of Huber and later Gilar, peak capacity is associated with ion pair concentration [183,264].

It is proposed that the efficiency of the ion-pairing mechanism decreases with fragment length. Therefore, larger mRNAs require more saturation (a higher ratio of ion-pairing molecules to phosphates) to interact efficiently with the stationary phase and maintain resolution. Unfortunately, increasing the concentration of TBAA and DBAA is not feasible, as it is likely to cause over-retention without any improvement in resolution. A potential solution is the use of combined IP reagents in the same mobile phase (weaker and stronger ion pair reagents). Further studies using TEAA and DBAA mixed at varying concentrations and with varying percentages of organic modifiers could be performed to extend the range at which accurate sizing of mRNA is feasible. To fine-tune accurate sizing, it would be more appropriate to use sizing standards with poly(A) tails the same length as the mRNA fragments being analysed due to the significant difference in base composition of RNAs with a poly(A) tail and the effect this may have on the overall hydrophobicity of the RNA compared to RNA fragments of the same length without a poly(A) tail.

7 Appendix

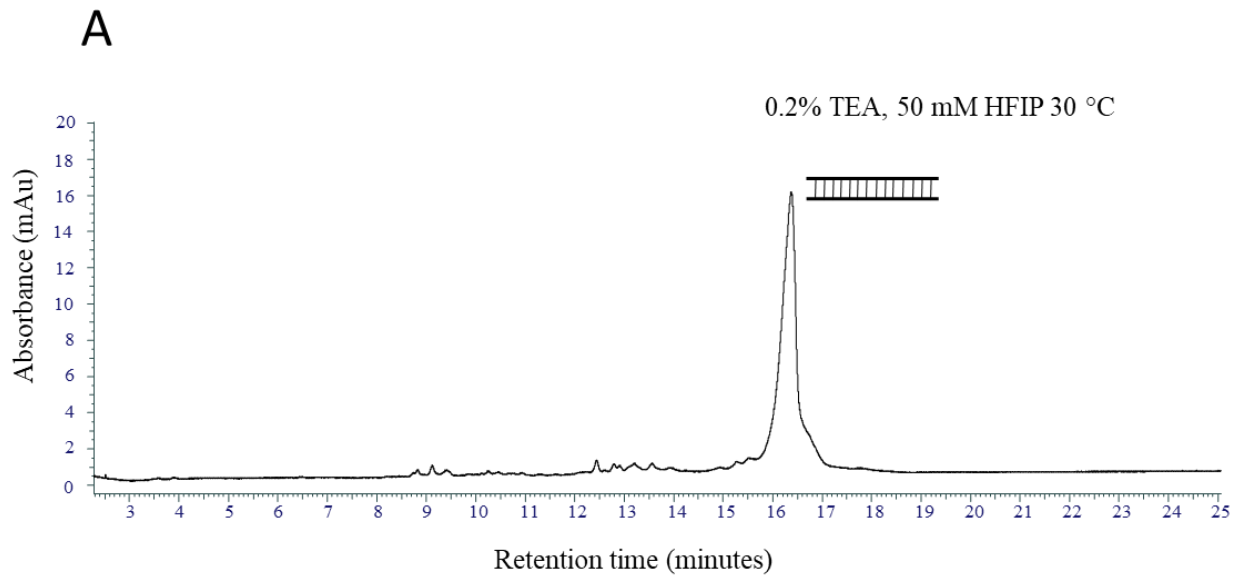
7.1 Denaturing RP IP HPLC separation of strands with various 5' modifications

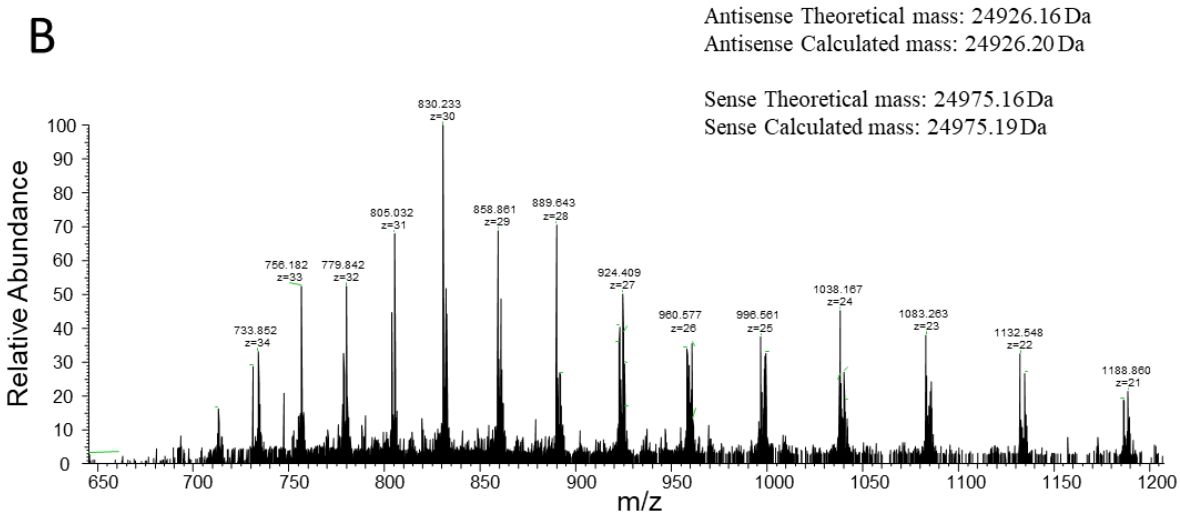


Appendix 1: **Denaturing IP RP HPLC analysis of labelled PCR products** A) IP RP chromatogram of unlabeled dsDNA at 95 °C. B) IP RP chromatogram of the 5' C18 labelled dsDNA at 95 °C. C) IP RP chromatogram of 5' Biotin labeled dsDNA at 95 °C. IP RP chromatogram of phosphorothioate dsDNA at 95 °C. Mobile phase A 0.1 M TEAA, mobile phase B 0.1M TEAA 25%

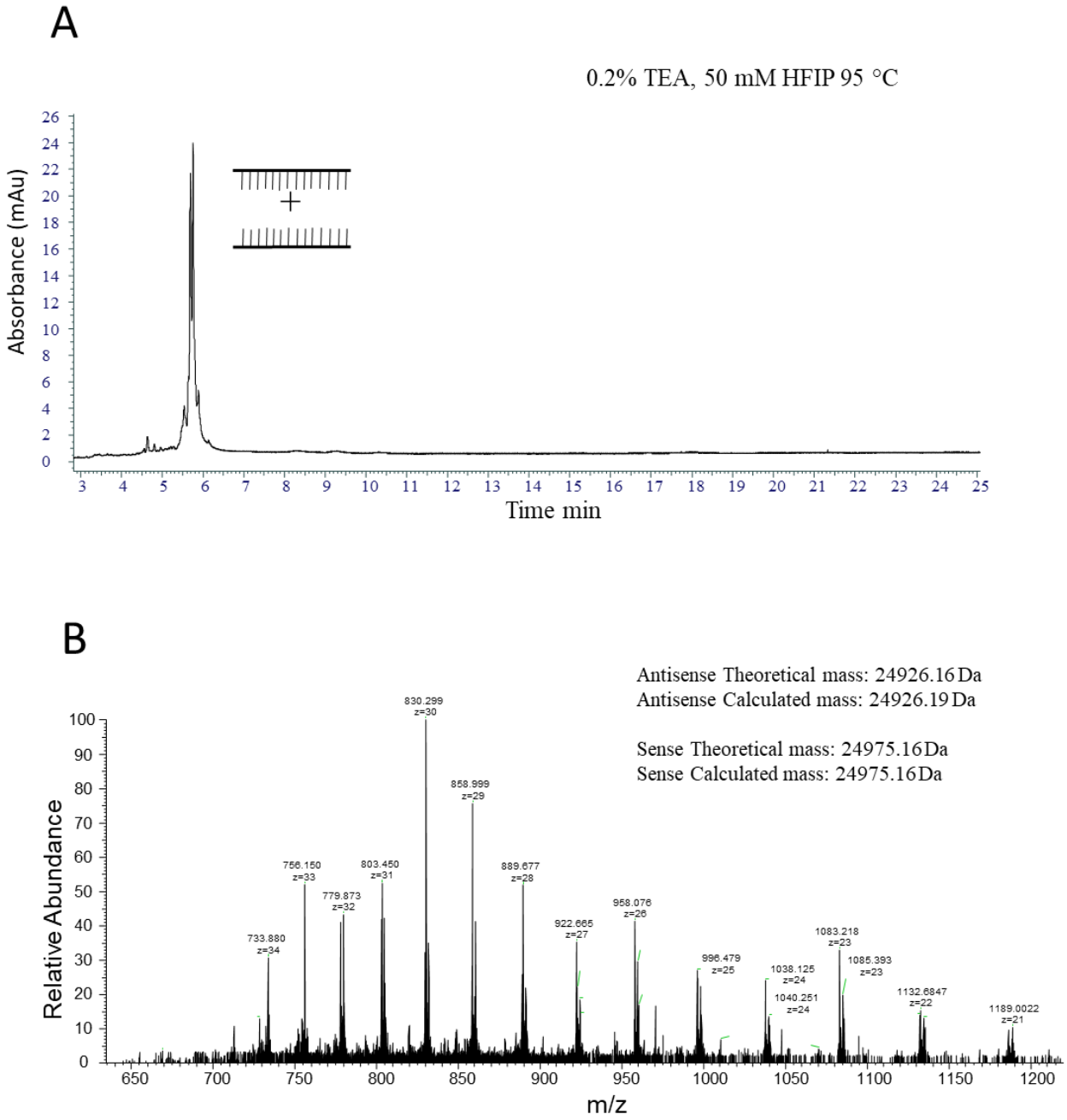
ACN. Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The ssDNA is highlighted. *Nonporous DNApac column (Thermofisher) used for separation*

7.2 LCMS analysis of all DHPLC strands

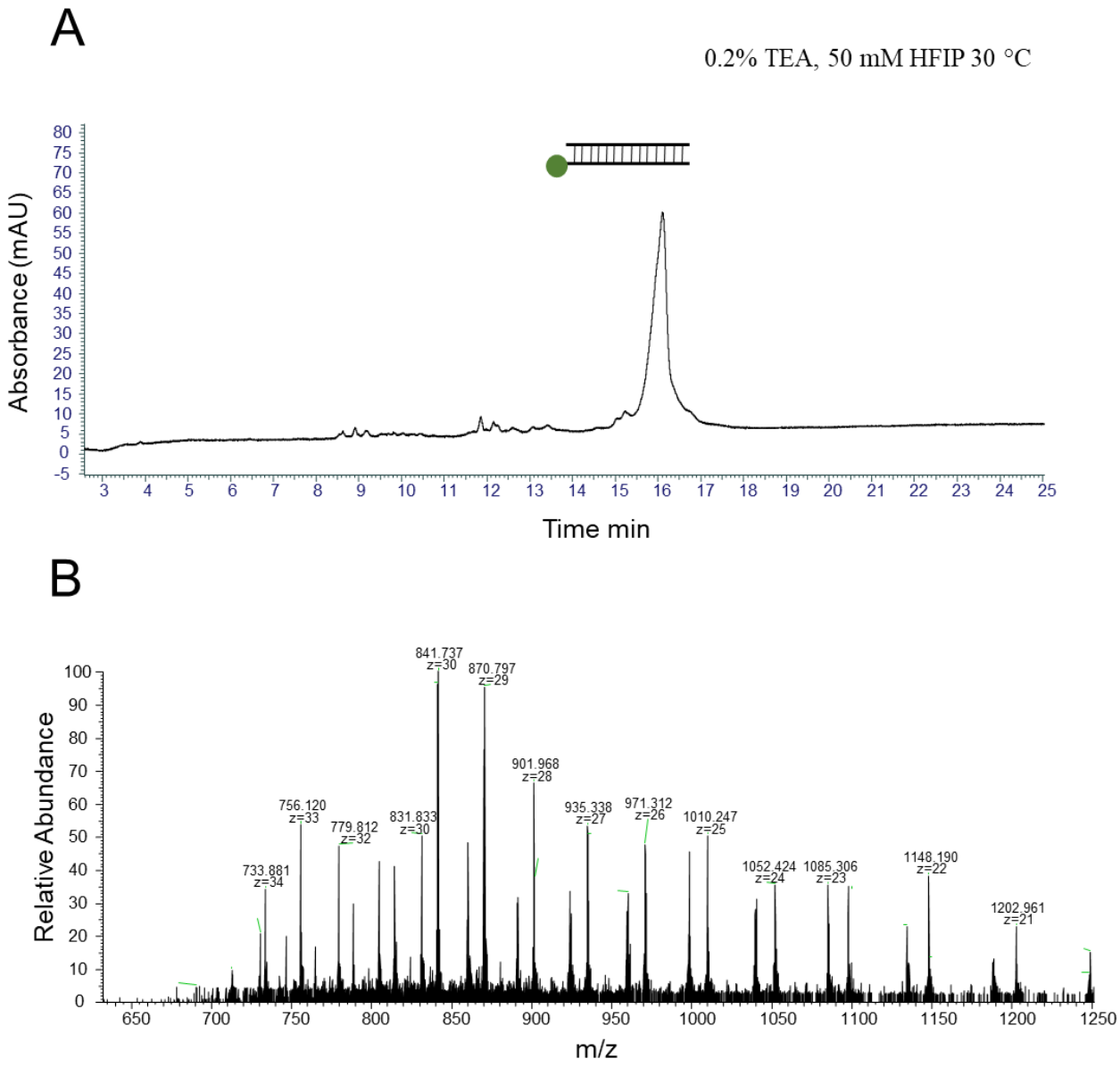




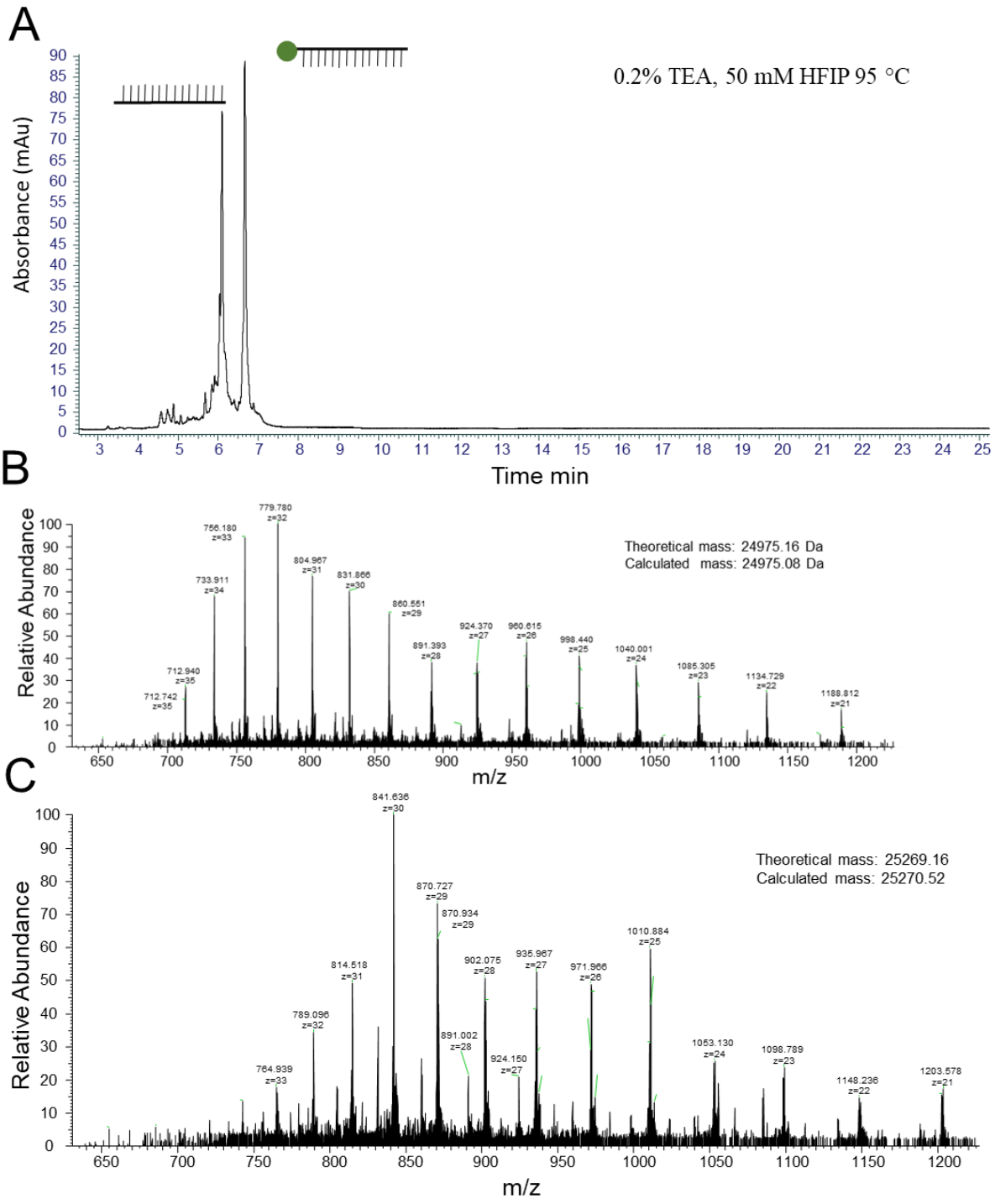
Appendix 2: Analysis of unmodified dsDNA using non-denaturing IP RP HPLC/MS. *A) HPLC UV analysis of unmodified dsDNA. B) ESI MS spectra of unmodified dsDNA. Mobile phase A 0.2% TEA 50 mM HFIP in H_2O, mobile phase B 0.2% TEA 50 mM HFIP in 80:20 H_2O: ACN, Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The theoretical and calculated monoisotopic masses are highlighted. Nonporous DNApac column (Thermofisher) used for separation*



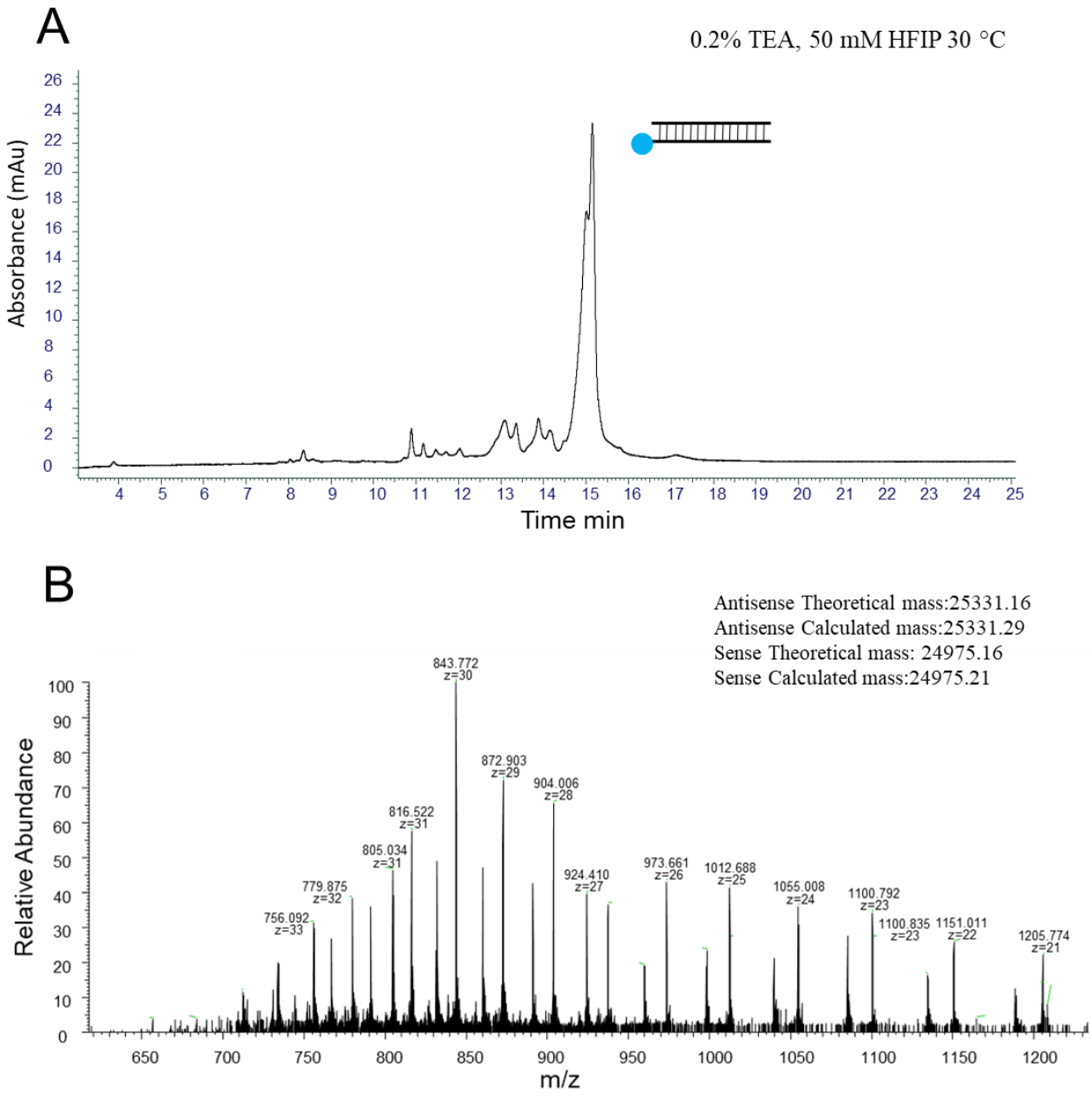
Appendix 3: Analysis of 5' C18 modified dsDNA using non-denaturing IP RP HPLC/MS. *A) HPLC UV analysis of unmodified dsDNA. B) ESI MS spectra of unmodified dsDNA. Mobile phase A 0.2% TEA 50 mM HFIP in H_2O, mobile phase B 0.2% TEA 50 mM HFIP in 80:20 H_2O: ACN, Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The theoretical and calculated monoisotopic masses are highlighted. Nonporous DNApac column (Thermofisher) used for separation*



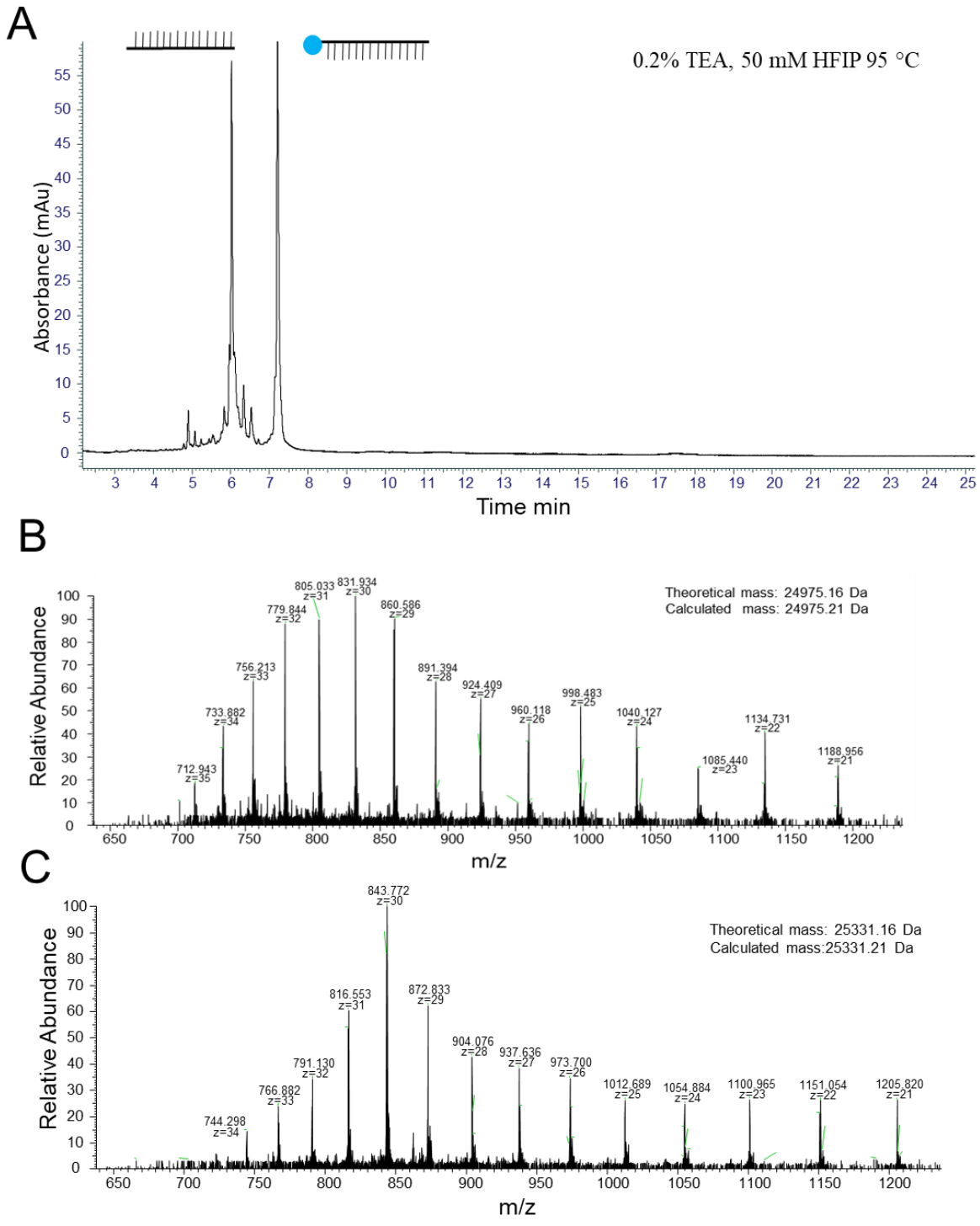
Appendix 4: Analysis of 5' C18 modified dsDNA using non-denaturing IP RP HPLC/MS. *A) HPLC UV analysis of unmodified dsDNA. B) ESI MS spectra of unmodified dsDNA. Mobile phase A 0.2% TEA 50 mM HFIP in H_2O, mobile phase B 0.2% TEA 50 mM HFIP in 80:20 H_2O: ACN, Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The theoretical and calculated monoisotopic masses are highlighted. Nonporous DNApac column (Thermofisher) used for separation*



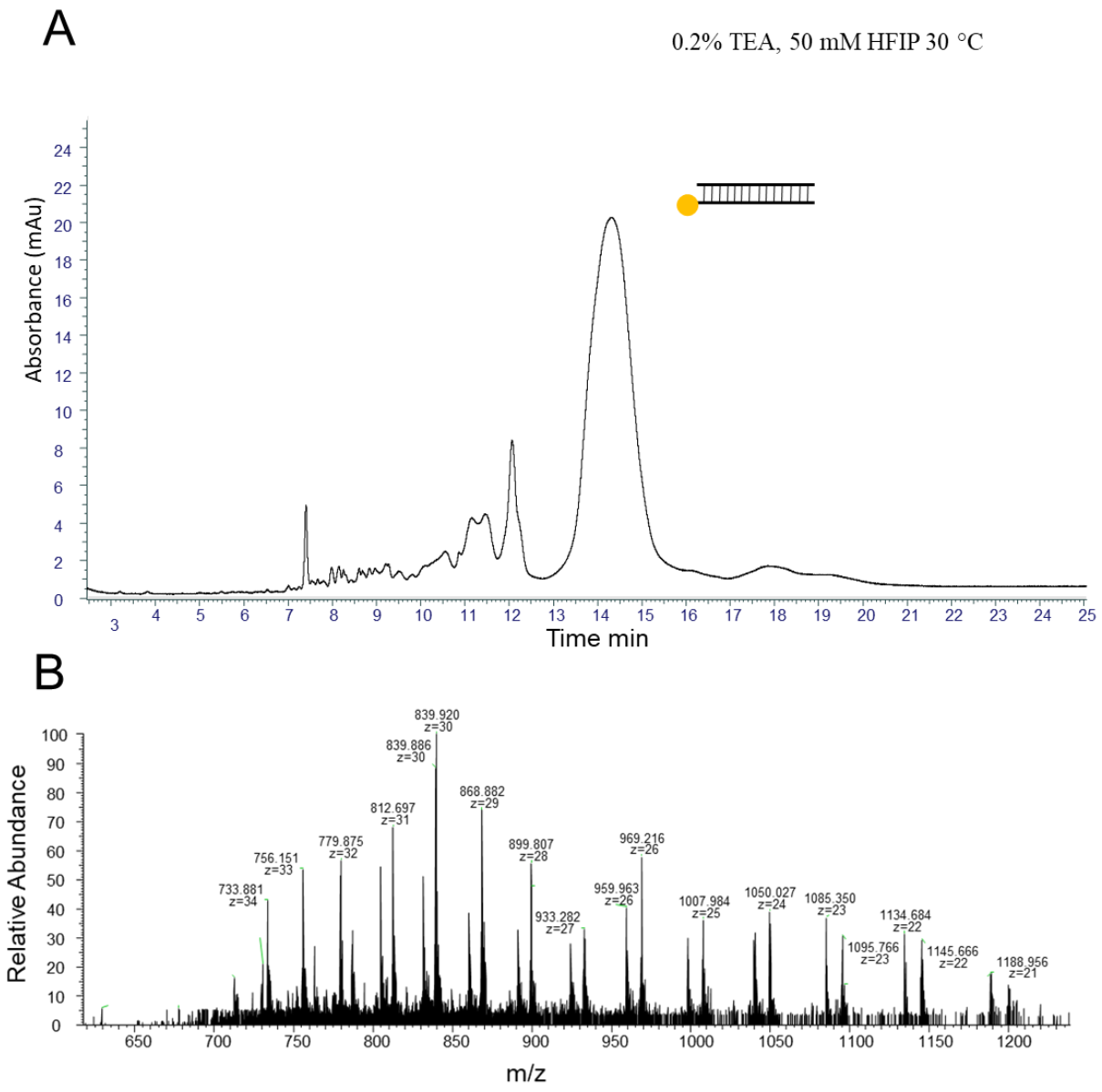
Appendix 5: Analysis of 5'C18 modified dsDNA using denaturing IP RP HPLC/MS. *A) HPLC UV analysis of unmodified dsDNA. B) ESI MS spectra of unmodified dsDNA. Mobile phase A 0.2% TEA 50 mM HFIP in H2O, mobile phase B 0.2% TEA 50 mM HFIP in 80:20 H2O: ACN, Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The theoretical and calculated monoisotopic masses are highlighted. Nonporous DNApac column (Thermofisher) used for separation*



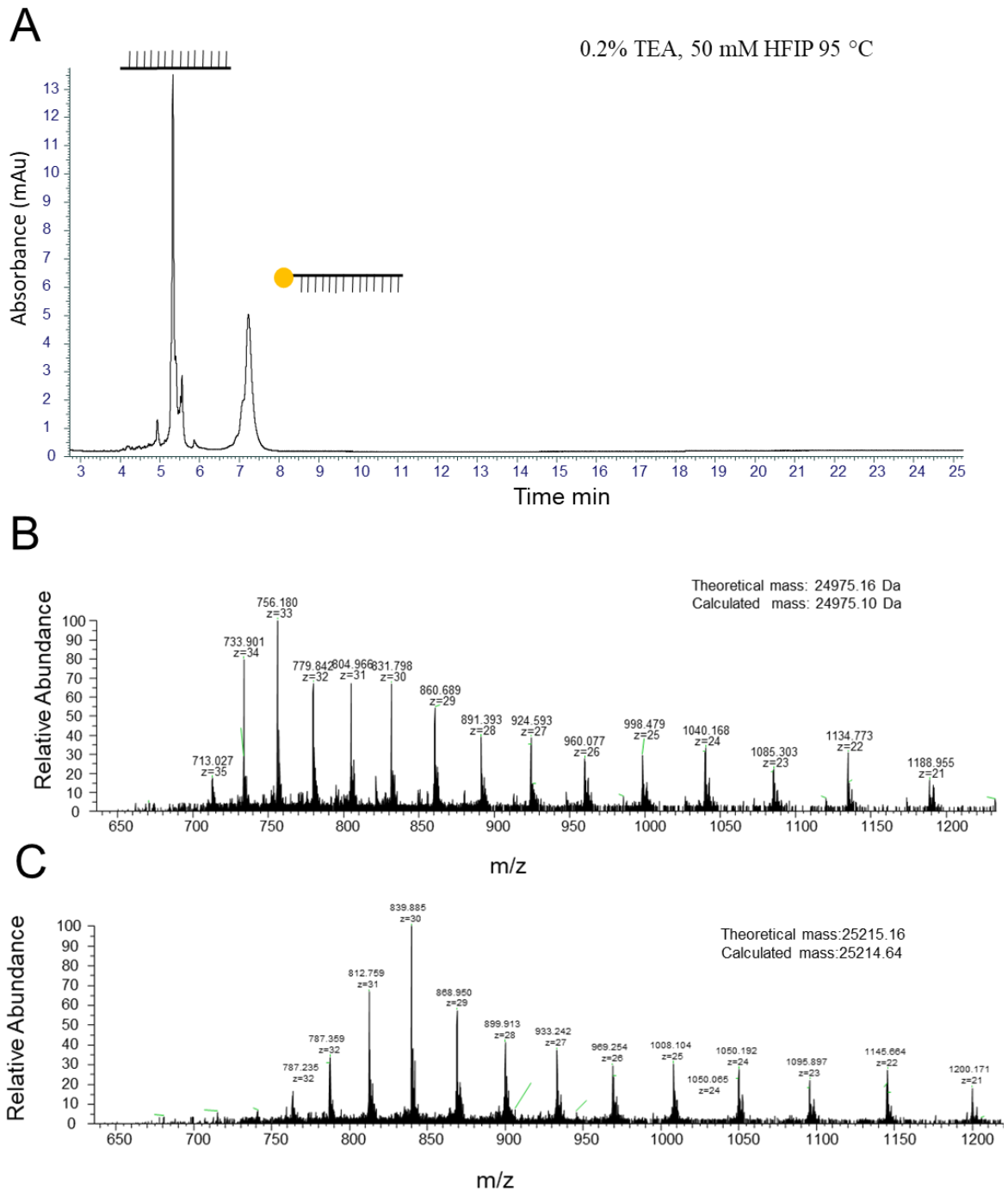
Appendix 6: Analysis of 5' Biotin modified dsDNA using non-denaturing IP RP HPLC/MS. *A) HPLC UV analysis of unmodified dsDNA. B) ESI MS spectra of unmodified dsDNA. Mobile phase A 0.2% TEA 50 mM HFIP in H_2O, mobile phase B 0.2% TEA 50 mM HFIP in 80:20 H_2O: ACN, Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The theoretical and calculated monoisotopic masses are highlighted. Nonporous DNApac column (Thermofisher) used for separation*



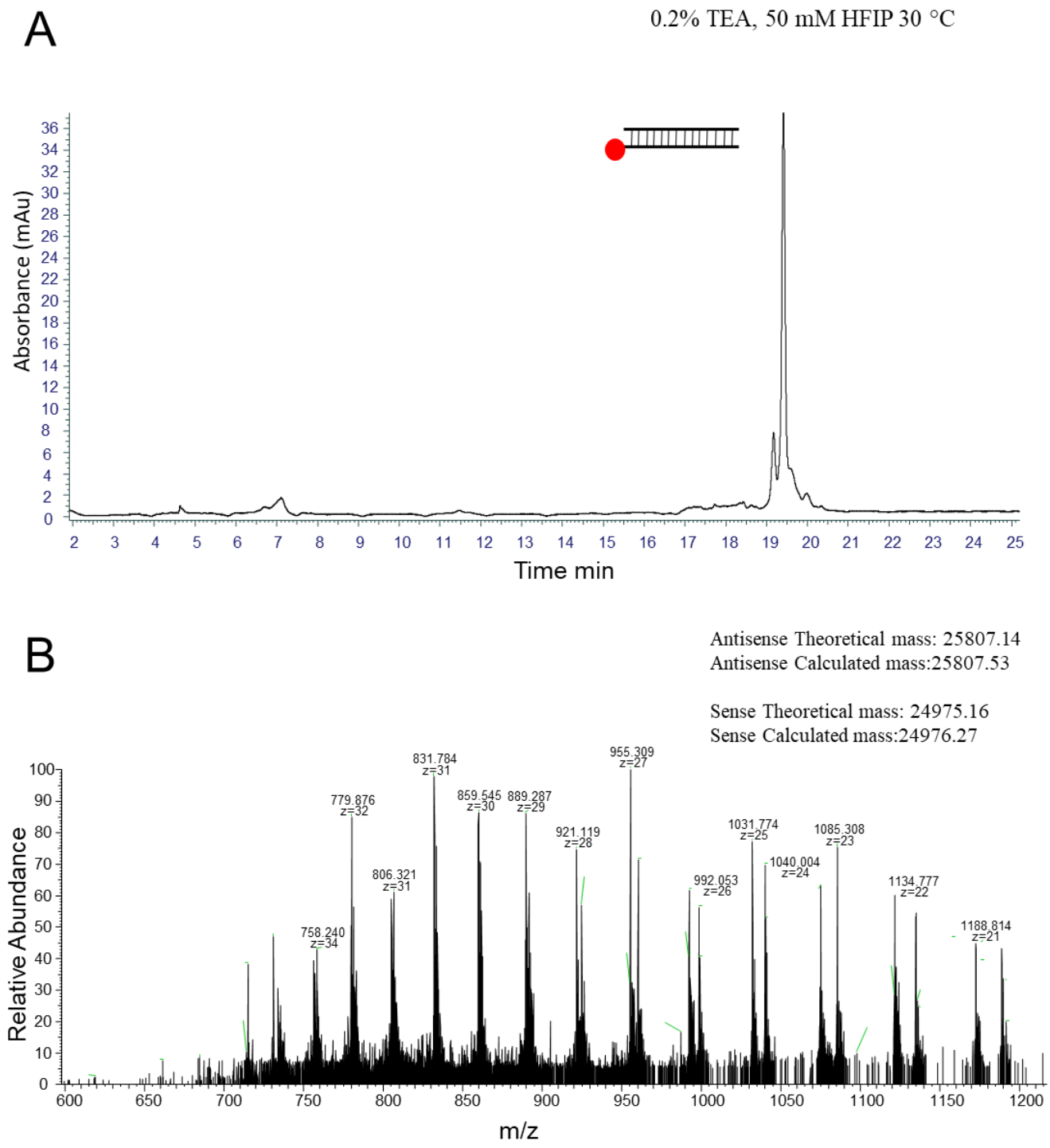
Appendix 7: Analysis of 5'Biotin modified dsDNA using denaturing IP RP HPLC/MS. *A) HPLC UV analysis of unmodified dsDNA. B) ESI MS spectra of unmodified dsDNA. Mobile phase A 0.2% TEA 50 mM HFIP in H_2O, mobile phase B 0.2% TEA 50 mM HFIP in 80:20 H_2O: ACN. Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The theoretical and calculated monoisotopic masses are highlighted. Nonporous DNApac column (Thermofisher) used for separation*



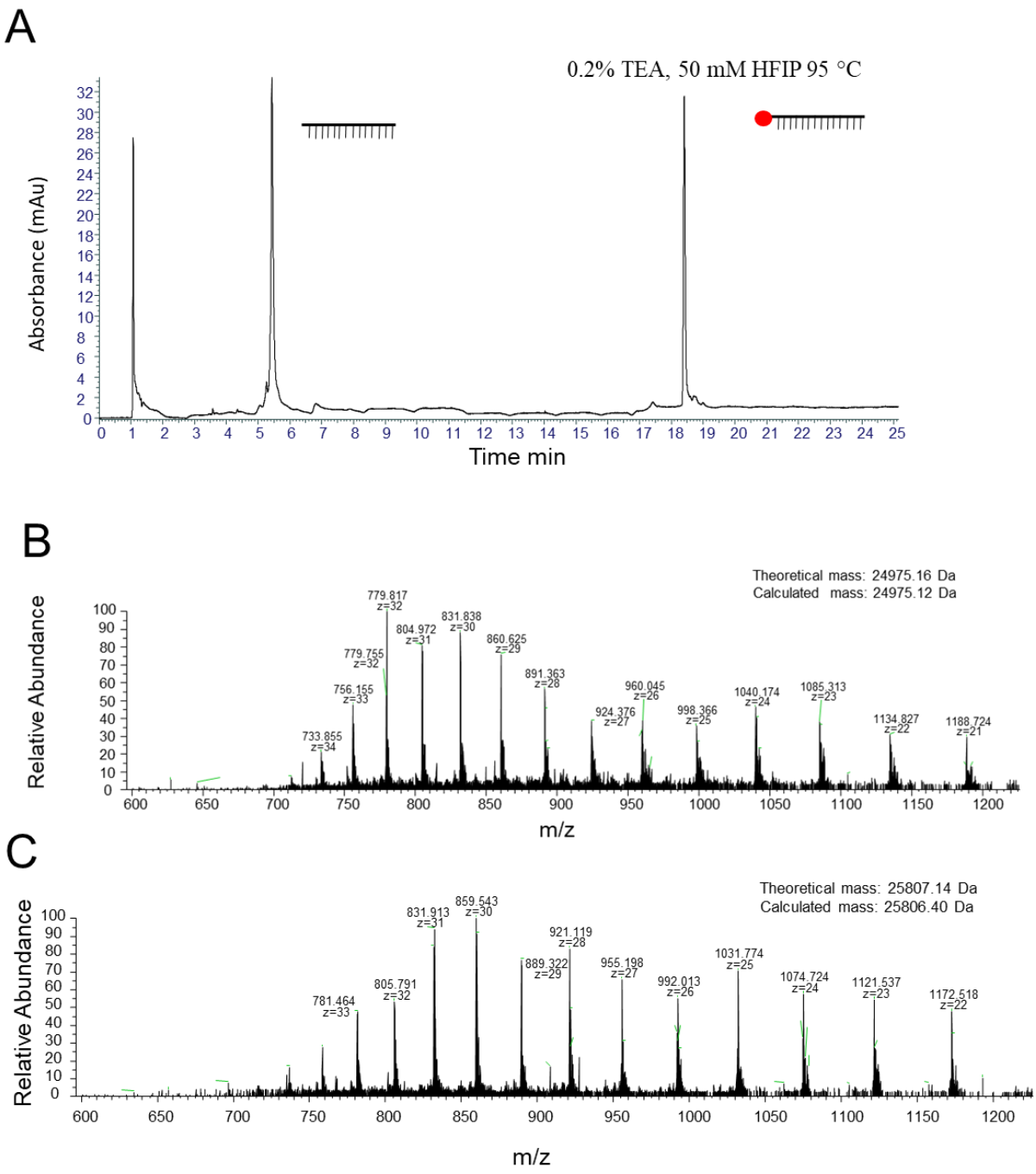
Appendix 8: Analysis of phosphorothioate dsDNA using non-denaturing IP RP HPLC/MS. *A) HPLC UV analysis of unmodified dsDNA. B) ESI MS spectra of unmodified dsDNA. Mobile phase A 0.2% TEA 50 mM HFIP in H_2O, mobile phase B 0.2% TEA 50 mM HFIP in 80:20 H_2O: ACN. Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The theoretical and calculated monoisotopic masses are highlighted. Nonporous DNApac column (Thermofisher) used for separation*



Appendix 9: Analysis of phosphorothioate dsDNA using denaturing IP RP HPLC/MS. *A) HPLC UV analysis of unmodified dsDNA. B) ESI MS spectra of unmodified dsDNA. Mobile phase A 0.1% TEA 50 mM HFIP in H_2O, mobile phase B 0.1% TEA 50 mM HFIP in 80:20 H_2O: ACN, Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The theoretical and calculated monoisotopic masses are highlighted. Nonporous DNApac column (Thermofisher) used for separation*



Appendix 10: Analysis of 5' Texas Red dsDNA using non-denaturing IP RP HPLC/MS. *A) HPLC UV analysis of unmodified dsDNA. B) ESI MS spectra of unmodified dsDNA. Mobile phase A 0.1% TEA 50 mM HFIP in H_2O, mobile phase B 0.1% TEA 50 mM HFIP in 80:20 H_2O: ACN, Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The theoretical and calculated monoisotopic masses are highlighted. Nonporous DNApac column (Thermofisher) used for separation*



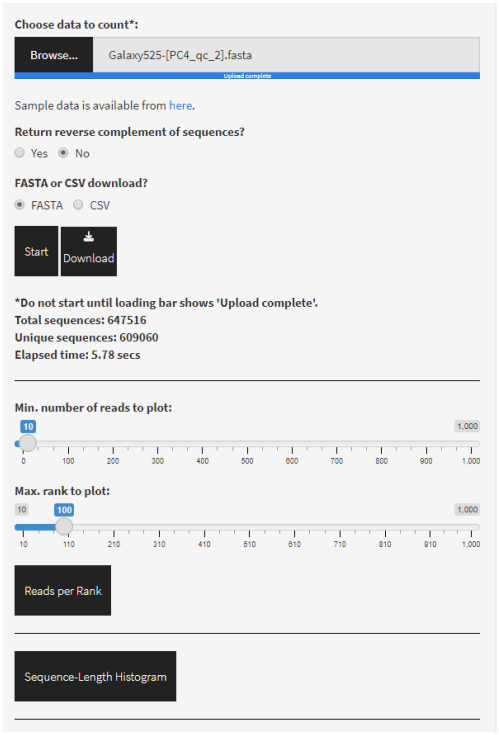
Appendix 11: Analysis of 5' Texas Red dsDNA using denaturing IP RP HPLC/MS. *A) HPLC UV analysis of unmodified dsDNA. B) ESI MS spectra of unmodified dsDNA. Mobile phase A 0.1% TEA 50 mM HFIP in H_2O, mobile phase B 0.1% TEA 50 mM HFIP in 80:20 H_2O: ACN, Gradient 4, flow rate 0.2 ml/min Chapter 2 section 2.5.3. The theoretical and calculated monoisotopic masses are highlighted. Nonporous DNApac column (Thermofisher) used for separation*

7.3 MDA-9 Gene sequence

ATGTCTCTGTACCCGTCTCTGGAAGACCTGAAAGTTGACAAAGTTATCCAGGCTCAGAC
CGCTTTCTCTGCTAACCCGGCTAACCCGGCTATCCTGTCTGAAGCTTCTGCTCCGATCC
CGCACGACGGTAACCTGTACCCGCGTCTGTACCCGGAACTGTCTCAGTACATGGGTCT
GTCTCTGAACGAAGAAGAAATCCGTGCTAACGTTGCTGTTGTTTCTGGTGCTCCGCTGC
AGGGTCAGCTGGTTGCTCGTCCGTCTTCTATCAACTACATGGTTGCTCCGGTTACCGGT
AACGACGTTGGTATCCGTCGTGCTGAAATCAAACAGGGTATCCGTGAAGTTATCCTGTG
CAAAGACCAGGACGGTAAAATCGGTCTGCGTCTGAAATCTATCGACAACGGTATCTTCG
TTCAGCTGGTTCAGGCTAACTCTCCGGCTTCTCTGGTTGGTCTGCGTTTCGGTGACCAG
GTTCTGCAGATCAACGGTGAAAACTGCGCTGGTTGGTCTTCTGACAAAGCTCACAAAGT
TCTGAAACAGGCTTTCGGTGAAAAAATCACCATGACCATCCGTGACCGTCCGTTCGAAC
GTACCATCACCATGCACAAAGACTCTACCGGTCACGTTGGTTTCATCTTCAAAAACGGTA
AAATCACCTCTATCGTTAAAGACTCTTCTGCTGCTCGTAACGGTCTGCTGACCGAACACA
ACATCTGCGAAATCAACGGTCAGAACGTTATCGGTCTGAAAGACTCTCAGATCGCTGAC
ATCCTGTCTACCTCTGGTACCGTTGTTACCATCACCATCATGCCGGCTTTCATCTTCGAA
CACATCATCAAACGTATGGCTCCGTCTATCATGAAATCTCTGATGGACCACACCATCCC
GGAAGTT

Appendix 12: MDA-9 sequence cloned into DH5-α: *Human, isoform 1, Uniprot*

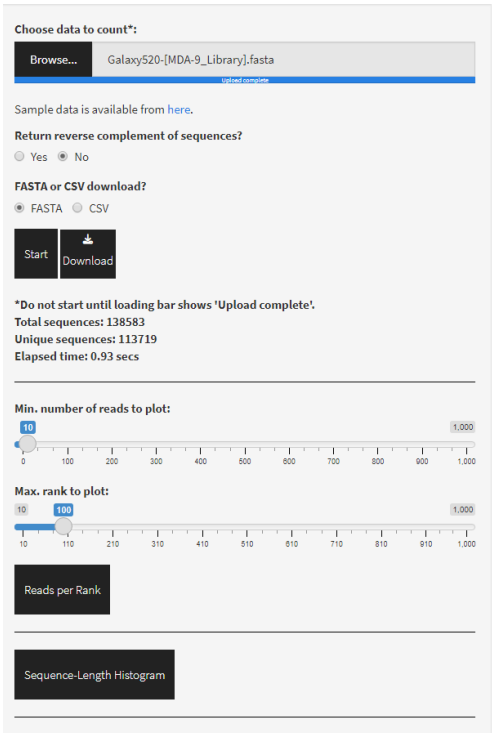
7.4 Fast Aptamer library analysis, pre and post-SELEX



Show 25 entries Search:

id	Rank	Reads	RPM	Length	seqs
>1-30-46.33	1	30	46.33	40	TATTGTTGACATGCCAGCGGGTCGGGGAAACGTGATCCTG
>2-20-30.89	2	20	30.89	40	ATTATTGCTTGTGACTGGAGTTCAGACGTGTGCTCTTCCG
>3-19-29.34	3	19	29.34	40	GCCGTGTGGACGTAAGCGTGAACGTCAGGATCACGTTTCC
>4-16-24.71	4	16	24.71	40	GCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCAG
>5-15-23.17	5	15	23.17	40	ACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATA
>6-15-23.17	6	15	23.17	40	CAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTT
>7-13-20.08	7	13	20.08	40	CTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACC
>8-13-20.08	8	13	20.08	40	GCACACCGGTATCCCCGCCCGACAACGACGGCAATGTACC
>9-12-18.53	9	12	18.53	40	AGCCCCGGGAGGAACCCGGGGCCGCAAGTGCGTTCGAAGT
>10-12-18.53	10	12	18.53	40	TAAGCGACGCTCAGGCAGGCGTAGCCCCGGGAGGAACCCG
>11-9-13.9	11	9	13.9	40	GATGACAGAAATTCTGCTTAAGCAGGCAATGGTGGGGATT
>12-9-13.9	12	9	13.9	40	GTAGCCCCGGGAGGAACCCGGGGCCGCAAGTGCGTTCGAA
>13-7-10.81	13	7	10.81	40	CAGCCACGCAGACCTTTGATGGTATTGCACAGAATATGGC
>14-7-10.81	14	7	10.81	40	CCCCAACGCCATAGCCCCTAATAGCCTCTTACCTCCTTCA
>15-6-9.27	15	6	9.27	40	ACTGGATGTTGAGGCCTTGTGGCGGGAATTGCGGTCGGGT
>16-6-9.27	16	6	9.27	40	AGGGGTTGTTTCGTGGATGAAAAGGGATTGGATATATGCG
>17-6-9.27	17	6	9.27	40	CATCCAATCCAAATGCGGCATCTTCAAACCTCCATGATGC
>18-6-9.27	18	6	9.27	40	CCCGCGTTCTCACCCTCCCGTCTCACTATATACCCGAGTC
>19-6-9.27	19	6	9.27	40	CCTAAAGCTGTATTTCCCACCCCTTCGCTATCTAATGACC
>20-6-9.27	20	6	9.27	40	GCGTAATTCGTGCAGGGTGGAACATGGTCTTGCGTTCGAG

Appendix 13: Top 20 reads (by read count) in pre SELEX libray, screenshot taken from FastAptamer

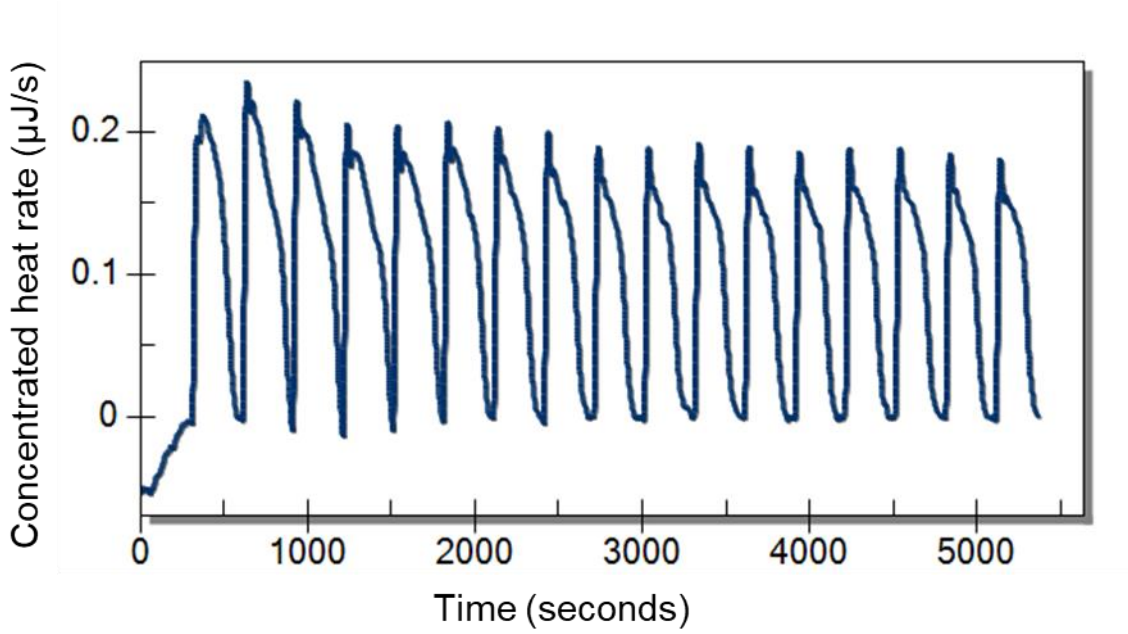


Show 25 entries Search:

id	Rank	Reads	RPM	Length	seqs
>1-225-1623.58	1	225	1623.58	40	TCGGCCTATAGTCAGTGTACCTATAGGGCTATTAGGACCA
>2-180-1298.86	2	180	1298.86	40	TATATCGGCCTATAGTCAGTGTACCTATAGGGCTATTAGG
>3-169-1219.49	3	169	1219.49	40	GGCCTATAGTCAGTGTACCTATAGGGCTATTAGGACCATT
>4-164-1183.41	4	164	1183.41	40	CCTATAGTCAGTGTACCTATAGGGCTATTAGGACCATTAT
>5-93-671.08	5	93	671.08	40	GATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAA
>6-88-635	6	88	635	40	GGGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTG
>7-80-577.27	7	80	577.27	40	GCCTATAGTCAGTGTACCTATAGGGCTATTAGGACCATTA
>8-74-533.98	8	74	533.98	40	TATAGTCAGTGTACCTATAGGGCTATTAGGACCATTATAT
>9-70-505.11	9	70	505.11	40	GGTCGGAAGAGCGTCGTGTAGGGAAAGAGTGTTTGCGGAC
>10-54-389.66	10	54	389.66	40	TCAGACCACGCTGATGCCCAGCGCCTGTTTCTTAATCACC
>11-52-375.23	11	52	375.23	40	TTCTCCCGTATTGTTGACATGCCAGCGGGTCGGGGAAACG
>12-43-310.28	12	43	310.28	40	GAGTCCTATGGACAAGTGGCCACAAACCACCAGAGTGCCC
>13-33-238.12	13	33	238.12	40	ATATCGGCCTATAGTCAGTGTACCTATAGGGCTATTAGGA
>14-32-230.91	14	32	230.91	40	GCACACCGGTATCCCCGCCCGACAACGACGGCAATGTACC
>15-31-223.69	15	31	223.69	40	CCCAGAGGCAAAGGGCTCCCAGGATCCCTCAGAGGTCTTT
>16-29-209.26	16	29	209.26	40	GAACGATGCGTAATGTGTGTATTGCCGTTGCTGTCTTTGC
>17-22-158.75	17	22	158.75	40	GGGAAAGAGTGTTTGCGGACGTGTAGATCTCGGTGGTCGC
>18-21-151.53	18	21	151.53	40	CTATAGTCAGTGTACCTATAGGGCTATTAGGACCATTATA
>19-19-137.1	19	19	137.1	40	CAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTT
>20-16-115.45	20	16	115.45	40	GATCGGAAGAGCACAGTCTGAACTCCAGTCACACGGCAGA

Appendix 14: Top 20 reads (by read count) in post SELEX libray, screenshot taken from FastAptamer

7.5 ITC Titration of BSA with BSA aptamer without detergents



Appendix 15: BSA-BSA aptamer titration: *ITC analysis of BSA: aptamer titration 20 injections, injection time 300 s, syringe concentration 100 µM aptamer, cell concentration 5 µM BSA.*

7.6 Sizing of later eluting peaks in analysis of dsRNAs in TBAA under non-denaturing conditions

Table 7.1: Size estimation of later eluting dsRNA peaks in TBAA at 50 ℃

Fragment	Size (bp)	Later peak size (bp)	Ratio
eGFP	267	574	2.15
B-actin	306	679	2.22
Dome11	404	983	2.43

7.7 Published Article, Coombes et al., 2024

Journal of Chromatography A 1719 (2024) 464699

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Optimisation of denaturing ion pair reversed phase HPLC for the purification of ssDNA in SELEX

Paul E. Coombes, Mark J. Dickman *

Department of Chemical & Biological Engineering, University of Sheffield, Sheffield, S1 3JD, UK

ARTICLE INFO

Keywords:
Ion Pair Reversed Phase HPLC
SELEX
ssDNA
Aptamers
Purification

ABSTRACT

Aptamers have shown great promise as oligonucleotide-based affinity ligands for various medicinal and industrial applications. A critical step in the production of DNA aptamers via selective enhancement of ligands by exponential enrichment (SELEX) is the generation of ssDNA from dsDNA. There are a number of caveats associated with current methods for ssDNA generation, which can lower success rates of SELEX experiments. They often result in low yields thereby decreasing diversity or fail to eliminate parasitic PCR by-products leading to accumulation of by-products from round to round. Both contribute to the failure of SELEX protocols and therefore potentially limit the impact of aptamers compared to their peptide-based antibody counterparts.

We have developed a novel method using ion pair reversed phase HPLC (IP RP HPLC) employed under denaturing conditions for the ssDNA re-generation stage of SELEX following PCR. We have utilised a range of 5′ chemical modifications on PCR primers to amplify PCR fragments prior to separation and purification of the DNA strands using denaturing IP RP HPLC. We have optimised mobile phases to enable complete denaturation of the dsDNA at moderate temperatures that circumvents the requirement of high temperatures and results in separation of the ssDNA based on differences in their hydrophobicity. Validation of the ssDNA isolation and purity assessment was performed by interfacing the IP RP HPLC with mass spectrometry and fluorescence-based detection.

The results show that using a 5′ Texas Red modification on the reverse primer in the PCR stage enabled purification of the ssDNA from its complimentary strand via IP RP HPLC under denaturing conditions. Additionally, we have confirmed the purity of the ssDNA generated as well as the complete denaturation of the PCR product via the use of mass-spectrometry and fluorescence analysis therefore proving the selective elimination of PCR by-products and the unwanted complementary strand. Following lyophilisation, ssDNA yields of up to 80% were obtained. In comparison the streptavidin biotin affinity chromatography also generates pure ssDNA with a yield of 55%. The application of this method to rapidly generate and purify ssDNA of the correct size, offers the opportunity to improve the development of new aptamers via SELEX.

1. Introduction

In just 30 years, nucleic acid aptamers have gone from a theoretical concept, to the lab bench, to the clinic and beyond [1]. Aptamers can bind to small molecules, metal ions, proteins and even whole cells, as a class of affinity ligands their broad capabilities has led to their utilisation in a range of applications such as drug delivery, diagnosis and biosensing [2]. Their rise as a feasible oligonucleotide counterpart to protein-based antibodies was seen as a new chapter in biomedicine, an inexpensive, non-immunogenic alternative to peptide based antibodies [3]. However, few aptamers have ever been selected for clinical trials. The sparsity of clinically approved aptamers can largely be attributed to in vivo issues, their susceptibility to nuclease degradation causing short half-lives in human serum [4].

Aside from their in vivo problems their wide range of potential targets have led to a plethora of diagnostic and industrial applications. Aptamers have been isolated for a variety of applications ranging from drug detection to affinity chromatography [5,6]. However, there are a number of caveats associated with successful aptamer selection via selective enhancement of ligands via exponential enrichment (SELEX) [7].

A significant problem in SELEX protocols is their low success rate, this is, in part due to complications relating to the generation of ssDNA

* Corresponding author.
E-mail address: m.dickman@sheffield.ac.uk (M.J. Dickman).

https://doi.org/10.1016/j.chroma.2024.464699
Received 27 September 2023; Received in revised form 29 January 2024; Accepted 30 January 2024
Available online 2 February 2024

following PCR. This step is crucial as the quality and quantity of ssDNA generated greatly influences the successful evolution of aptamers [8]. Current methods are low yielding or fail to eliminate by-products which can grow exponentially from round to round [9]. The two most common methods, exonuclease digestion and streptavidin/biotin affinity chromatography particularly suffer from this, due to their lack of distinguishing between DNA of different sizes.

Exonuclease digestion involves labelling of the reverse primer in PCR with a 5′ phosphate, which is a substrate for phosphate dependant exonucleases [10]. However, it must be noted that these nucleases are not completely specific and so some of the non-phosphorylated strand can be digested. In addition, any PCR by-products which do not have a 5′ phosphate group will largely be undigested by the enzyme and will be taken through to the selection stage. Furthermore, the presence of any nucleases in the product requires careful removal prior to downstream applications The streptavidin biotin affinity chromatography method involves the labelling of the reverse primer of the PCR reaction with a 5′ biotin modification, this causes the complimentary strand to interact strongly with streptavidin such that the DNA duplex can be denatured, and the ssDNA separated [11]. This method also suffers from the inability to distinguish between products and by-products as well as potentially contaminating the next stage of selection with streptavidin if additional purification is not carried out [12].

Co-polymerisation within an agarose gel using a polymerizable 5′ modification has demonstrated effective separation of the two strands. The gel mixture is loaded into the wells of a pre-set polyacrylamide gel. When it has set, high temperatures are used to denature the duplex and release the strand which is not covalently linked to the acrylamide [13]. This separation is efficient offering yields of over 80%.

Methods such as gel extraction via crush and soak and asymmetric PCR have been demonstrated to yield reasonable quantities of ssDNA. Asymmetric PCR involves biasing primer concentrations so that the concentration of the forward primer exceeds that of the reverse primer. This leads to the generation of a higher quantity of the desired strand making purification more efficient. When coupled with exonuclease digestion, asymmetric PCR affords yields of over 80% [14] However, these methods still do not address the issue of by-product formation, particularly those close in size to the DNA template which are typically not removed using existing methods [15,16]. In addition, re-generation of ssDNA is important for other molecular biology protocols, including sequencing, sample preparation for mass spectrometry and microarray technology [17,14].

In this study we have utilised a range of 5′ chemical modifications on PCR primers to amplify PCR fragments prior to separation and purification of the DNA strands using denaturing IP RP HPLC which offers significant advantages over existing methods. The application of this method to rapidly generate and purify ssDNA of the correct size, offers the opportunity to improve the development of new aptamers via SELEX.

2. Experimental

2.1. Materials

Triethylammonium acetate (TEAA, Sigma-Aldrich), triethylamine (TEA, ThermoFisher), Tributylamine (TBA, Acros Organics), acetonitrile and water (HPLC grade, Fisher Scientific), 1,1,1,3,3,3,-Hexafluoro-2-propanol (HFIP, Sigma-Aldrich). Glacial acetic acid, combined with tributylamine for the preparation of tributylammonium acetate (TBAA) was sourced from VWR. All oligonucleotides were synthesised by Eurofins, SELEX library was sourced from IDT. For MS analysis acetonitrile and water (UHPLC MS grade, ThermoFisher) and 1,1,1,3,3,3,-Hexafluoro-2-propanol LC MS grade (99.9% ThermoFisher) were used.

2.2. Ion pair-reverse phase high performance liquid chromatography (IP-RP HPLC)

Samples were analysed by IP-RP-HPLC on a Vanquish UHPLC (ThermoFisher) or Agilent 1100 HPLC using a DNAPacRP column (100 mm x 2.1 mm I.D. ThermoFisher). Chromatograms were generated using UV detection at a wavelength of 260 nm. Weak IP RP HPLC analysis was performed using the following conditions: Mobile phase A 0.1 M TEAA pH 7.0; Mobile phase B 0.1 M TEAA, pH 7.0 containing 25% acetonitrile or Mobile Phase A: 0.2% TEA, 50 mM HFIP. Mobile Phase B: 0.2% TEA, 50 mM HFIP, 20% acetonitrile. Mobile Phase A: 0.1 M TEAA, 50 mM HFIP. Mobile Phase B: 0.1 M TEAA, 50 mM HFIP, 25% acetonitrile. Strong IP RP HPLC was performed using Mobile Phase A: 5 mM TBAA, 10% acetonitrile. Mobile Phase B: 5 mM TBAA, 80% acetonitrile.

HPLC was performed using the following gradients:

Gradient 1. Mobile phase A 0.1 M TEAA pH 7.0; Mobile phase B 0.1 M TEAA, pH 7.0 containing 25% acetonitrile. Gradient starting at 10% buffer B for 2 min, followed by a non-linear extension (curve 3) to 100% buffer B over 18 min at a flow rate of 0.6 ml/min at 40 °C.

Gradient 2. Mobile Phase A: TBAA 5 mM, 10% acetonitrile. Mobile Phase B: TBAA 5 mM 80% acetonitrile. Gradient starting at 5% buffer B for 2 min, followed by a non-linear extension (curve 3) 40% buffer B over 15 min, then extended to 100% buffer B over 2.5 min at a flow rate of 2.2 µl/min at 50 °C..

Gradient 3. Mobile phase A 0.1 M TEAA pH 7.0; Mobile phase B 0.1 M TEAA, pH 7.0 containing 25% acetonitrile. Gradient starting at 35% buffer B for 2 min, followed by a non-linear extension (curve 2) to 80% buffer B over 18 min, then extended to 100% buffer B over 5 min at a flow rate of 0.6 ml/min at 40 °C.

Gradient 4. Mobile Phase A: 0.2% TEA, 50 mM HFIP. Mobile Phase B: 0.2% TEA, 50 mM HFIP, 20% acetonitrile. Gradient starting at 10% buffer B followed by a non-linear extension (curve 2) to 100% buffer B over 18minutes then extend to 100% Buffer B over 5 min at a flow rate of 0.2 ml/min at 30 or 95 °C.

Gradient 5. Mobile Phase A: 0.1 M TEAA 50 mM HFIP. Mobile Phase B: 0.1 M TEAA, 50 mM HFIP, 25% acetonitrile. Gradient starting at 35% buffer B for 2 min, followed by a linear extension to 80% buffer B over 10 min, then extended to 100% buffer B over 5 min at a flow rate of 0.6 ml/min at 80 °C..

Gradient 6. Mobile Phase A: 0.1 M TEAA 50 mM HFIP. Mobile Phase B: 0.1 M TEAA 50 mM HFIP 25% acetonitrile. Gradient starting at 25% buffer B for 2 min, followed by a linear extension to 80% buffer B over 22 min, then extended to 100% buffer B over 5 min at a flow rate of 0.3 ml/min at 80 °C.

2.3. Library and primers

The library (5′ ATATATGACTCGGGTATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATAT ATGCGCTCAGTATATA) was sourced from Integrated DNA Technologies. The non-randomised template (5′-ATATATGACTCGGGTATATAATGGTCCTAATAGCCCTATAGGTACACTGACTATAGGCCGATATATGCGCTCAGTATATA) along with the forward (5′-ATATATGACTCGGGTATATA) and reverse (5′- TATATACTGAGCGCATATAT) primers (unmodified and modified) were sourced from Eurofins.

2.4. PCR

PCR amplification was carried out by amplifying either the template or the library (50 ng per reaction) with a reverse primer (with the 5′

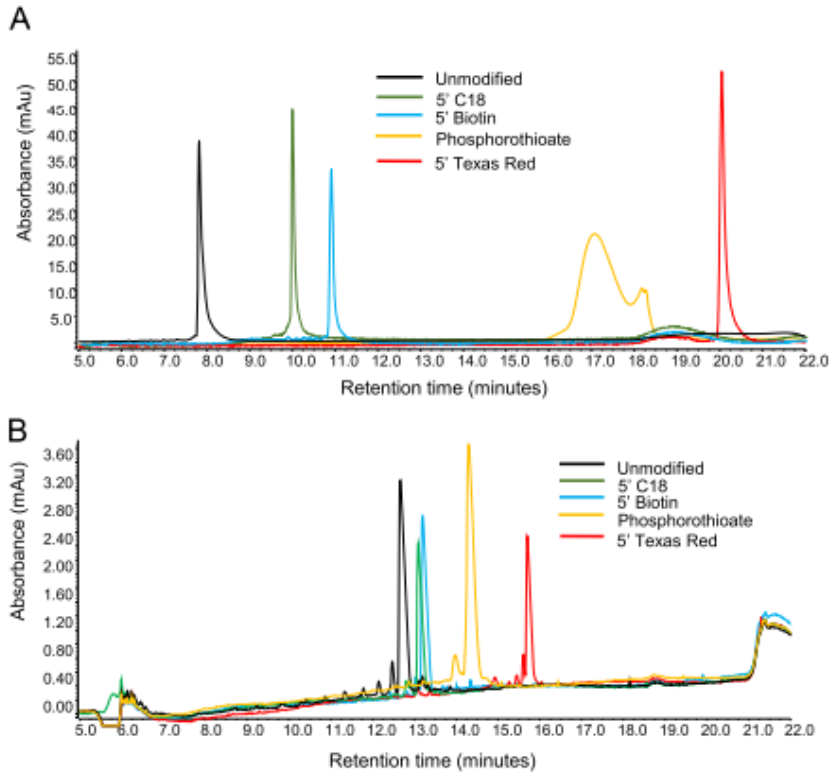


Fig. 1. IP RP HPLC analysis of 5′ modified and phosphorothioate modified primers. A) Weak IP RP HPLC analysis. Mobile phase A 0.1 M TEAA; mobile phase B 0.1 M TEAA, 25% ACN. Gradient 1, flow rate 0.3 ml/min at 40 °C. 5 pmol primer injected. B) Strong IP RP HPLC analysis Mobile phase A 5 mM TBAA; mobile phase B, 5 mM TBAA 80% ACN. Gradient 2, flow rate 2.2 µl/min at 50 °C, 3 pmol primer injected. UV detection at 260 nm.

chemical modification) and an unmodified forward primer. PCRs were performed in 50 µl final volume using 50 ng of a ssDNA template, 1 µl of 10 µM primer, 1 µl of 10 mM dNTPs (New England Biolabs), and 5 units Taq polymerase (New England Biolabs). The PCRs were carried out at 95 °C for 30 s, followed by 47 °C for 30 s and 68 °C for 30 s for 15 cycles. An initial denaturation step was carried out at 95 °C at the start of the PCR for 5 min and a final extension step carried out at the end at 68 °C for 5 min. Following PCR DNA was purified via ethanol precipitation or use of silica magnetic beads (Section 2.5).

2.5. DNA purification and desalting

2.5.1. Ethanol precipitation

Following PCR, dsDNA was re-suspended in a final volume of 300 µl HPLC grade water and 300 µl of phenol-chloroform solution added. The solution was vortexed at 13,000 rpm for 5 min and the aqueous fraction removed and mixed with two volumes of ice-cold ethanol, 1 µl of glycogen was added and the solution was cooled at −20 °C for four hours. The solution was centrifuged at 13,000 rpm for 5 min, and the pellet was washed twice with 300 µl of a 70% ethanol solution prior to re-suspension in H_2O. For the precipitation of ssDNA from the HPLC, the above procedure was performed without the addition phenol-chloroform step.

2.5.2. Magnetic bead clean-up

DNA was mixed with 2X volumes 5 M guanidinium chloride in 90% ethanol (v/v) and 50 µl of silica magnetic beads (Dynabeads, Thermo Scientific) for 15 min on a gentle rocker, beads were removed from solution via magnetic separation and washed twice with 80% ethanol (v/v). The beads were then dried in a 37 °C incubator and the DNA eluted in HPLC grade water.

2.5.3. Freeze drying

The eluted fractions were frozen at −80 °C and freeze-dried overnight at −120 °C. Following lyophilisation the sample was re-suspended in 500 µl of HPLC grade water. DNA fractions were dried in using a vacuum centrifugation at 60 °C until dry and re-suspended in 500 µl of HPLC grade water.

2.5.4. Separation of ssDNA via streptavidin agarose

Biotinylated PCR product were re-suspended in 1x PBS and incubated with 100 µl streptavidin agarose (Thermo Fisher Scientific) for 30 min. Following incubation, the gel was washed with 3x volumes 1x PBS. The ssDNA was eluted with 50 µl 150 mM NaOH (repeated twice for a total of three elutions) and neutralised with 5 µl 1.5 M HCl.

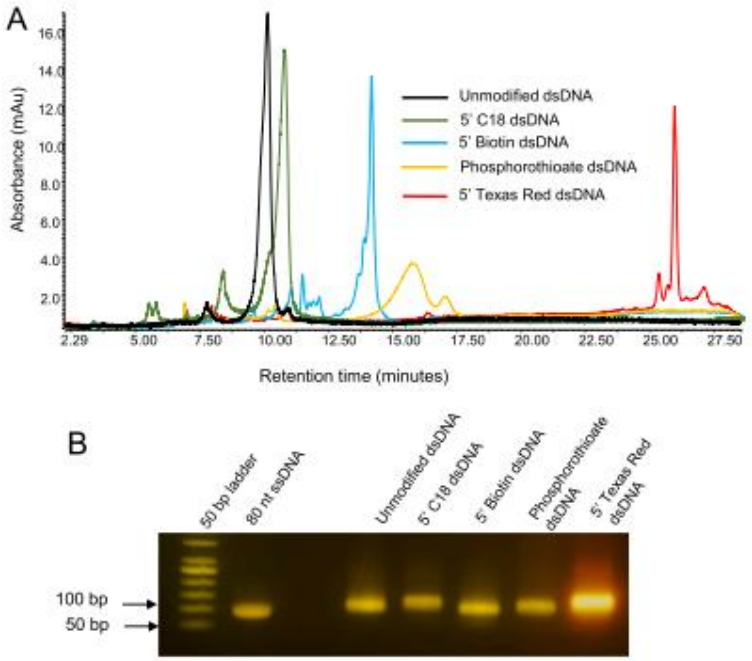


Fig. 2. Analysis of dsDNA PCR products generated using a 5′ modified reverse primer. A) IP RP HPLC chromatogram. Mobile phase A 0.1 M TEAA; mobile phase B 0.1 M TEAA in 25% ACN. Gradient 3, flow rate 0.6 ml/min at 40 °C UV detection at 260 nm. B) Agarose gel electrophoresis analysis, 50 ng PCR product was analysed for each of the dsDNA PCR products. Staining was performed using Midori green direct. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.6. Mass spectrometry

2.6.1. Instrument methods and data acquisition

Mass spectrometry analysis was performed using an Orbitrap Exploris 240 mass spectrometer (ThermoFisher) in negative ion mode. Instrument operation was carried out with Thermo Scientific Tune and XCalibur software. Data acquisition was performed using data-dependant acquisition (DDA) in full-scan negative mode, scanning from 450 to 3000 *m/z*, with an MS1 resolution of 240 000 and a normalized automatic gain control (AGC) target of 300%. Intact mass analysis and deconvolution was performed using the 'Xtract deconvolution" algorithm within FreeStyle software (ThermoFisher). Deconvolution was performed selecting the nucleotide isotope table, negative charge, charge range 5–50 prior to determining the monoisotopic mass. All mass spectra are shown in the supplementary section (**S2–12**).

3. Results and discussion

3.1. Retention comparison of 5′ modified oligonucleotides using IP RP HPLC

In our initial experiments we aimed to assess the effect of a variety of 5′ chemical modifications of oligonucleotides on their retention time when analysed using IP RP HPLC. It was hypothesised that increasing the hydrophobicity of the 5′ modification would result in the ability to separate the two strands of the DNA duplex when analysed under denaturing conditions using IP RP HPLC and therefore enable the separation of the two single strands from an 80 bp dsDNA fragment generated using PCR products via the same principle. Initial work focused on the analysis of a range of 5′ modifications on the PCR primers including C18, Biotin, Texas Red and a 20-base primer with a phosphorothioate backbone which were selected due to the increased hydrophobicity of phosphorothioate oligonucleotides (OGNs) [18–20].

The modified primers were analysed via both weak IP RP HPLC (triethylammonium acetate) and strong IP RP HPLC (tributylammonium acetate) see Fig. 1. The strength of an ion-pair reagent is defined by its overall hydrophobicity. The alkylamine, TEA (short alkyl chains) is termed a "weak" ion-pair reagent and the more hydrophobic TBA (longer alkyl chains) is termed a "strong" ion-pair reagent, which has a stronger interaction with the stationary phase [21]. Comparison of the retention times of the various 5′ chemical modifications demonstrates that the phosphorothioate and 5′ Texas Red result in the largest difference in retention compared to the unmodified OGN. The results show significant shifts in retention time, reflecting the different OGN hydrophobicity in the order unmodified <C18 < biotin < phosphorothioate < Texas Red. These results are consistent with previous IP RP HPLC analysis of oligonucleotides with hydrophobic fluorophores [22–25]. The phosphorothioate OGN is notably broader when analysed under weak IP RP HPLC due to the large number of diastereoisomers present [18,20]. This may have implications for its use in the denaturing HPLC method if fractionating and collecting the phosphorothioate ssDNA for analysis. Collecting much broader peaks may involve splitting the fractions over multiple tubes and potentially decreasing yield.

To further analyse the effects of the 5′ modifications, analysis was also performed using strong IP RP HPLC utilising the larger alkylamine (tributylamine). Under these conditions size-based separations largely dominate and the influence of the 5′ modifications is expected to be reduced [18–20] The results are shown in Fig. 1B and show a reduced difference in retention time of the various 5′ modification compared to weak IP RP HPLC (Fig. 1A). However, the results show that a difference

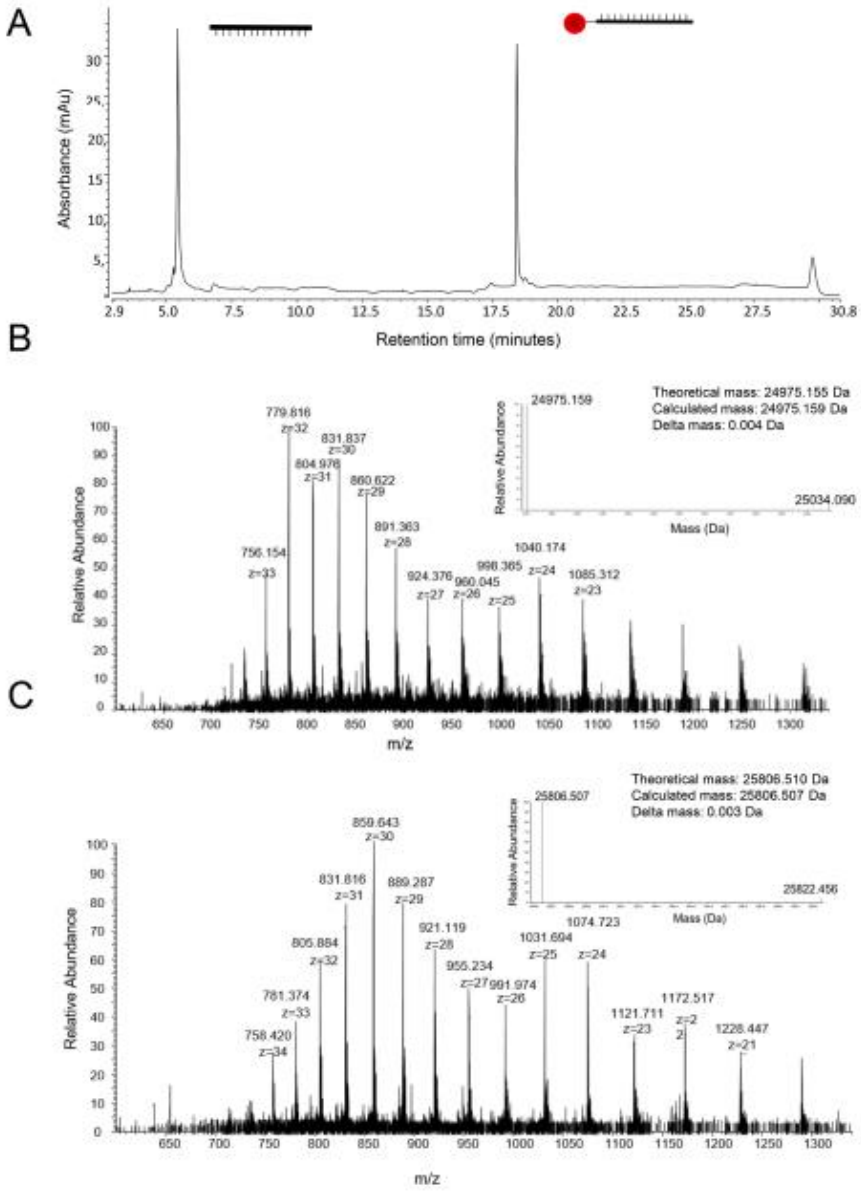


Fig. 3. Denaturing IP RP HPLC analysis. A) Texas Red labelled dsDNA was analysed using weak IP RP HPLC at 95 °C, UV detection 260 nm. Mobile phase A 0.2% TEA, 50 mM HFIP; mobile phase B 0.2% TEA, 50 mM HFIP in 20% ACN. Gradient 4, flow rate 0.2 ml/min. B) ESI MS spectra unlabelled ssDNA. C) ESI MS spectra Texas Red labelled ssDNA. The theoretical and calculated monoisotopic masses are highlighted. Deconvoluted masses are highlighted within the ESI MS spectra shown inset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in overall hydrophobicity of the OGNs is still observed under strong IP RP HPLC conditions, highlighting that fully size based separations are not achieved. The results also show that peak broadening of the phosphorothioate OGN is not seen in the strong ion pair analysis as the strength of interaction between the analyte column is much stronger hence structural features such as base sequence or structural isomerism have less impact on retention time. Consistent with weak IP RP HPLC analysis the phosphorothioate and 5′ Texas Red modifications result in the largest difference in retention compared to the unmodified OGNs. However, strong ion pair reagents are less volatile and therefore less suitable for the purification of ssDNA.

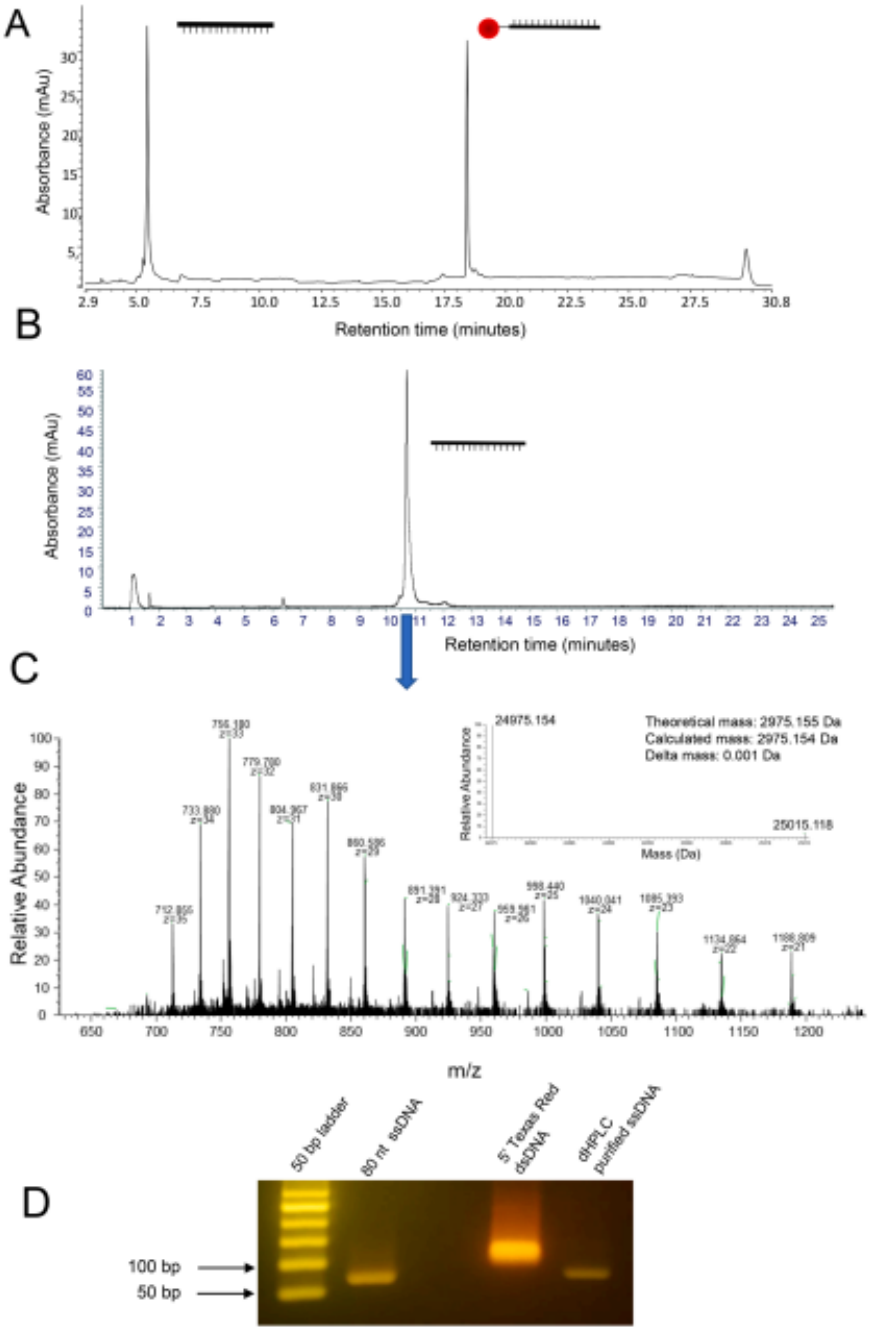


Fig. 4. Denaturing IP RP HPLC purification of ssDNA. A) IP RP chromatogram of the Texas Red labelled dsDNA at 95 °C, UV detection 260 nm. Mobile phase A 0.2% TEA, 50 mM HFIP; mobile phase B 0.2% TEA, 50 mM HFIP in 20% ACN. Gradient 4 flow rate 0.2 ml/min. B) LC UV chromatogram of the purified unlabelled ssDNA at 40 °C, UV detection 260 nm. Gradient 4 flow rate 0.2 ml/min. C) ESI MS spectra of the purified unlabelled ssDNA. The theoretical and calculated monoisotopic masses are highlighted. Deconvoluted masses are highlighted within the ESI MS spectra shown inset. D) Agarose gel electrophoresis analysis of the dsDNA and purified ssDNA. Staining was performed using Midori green direct. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

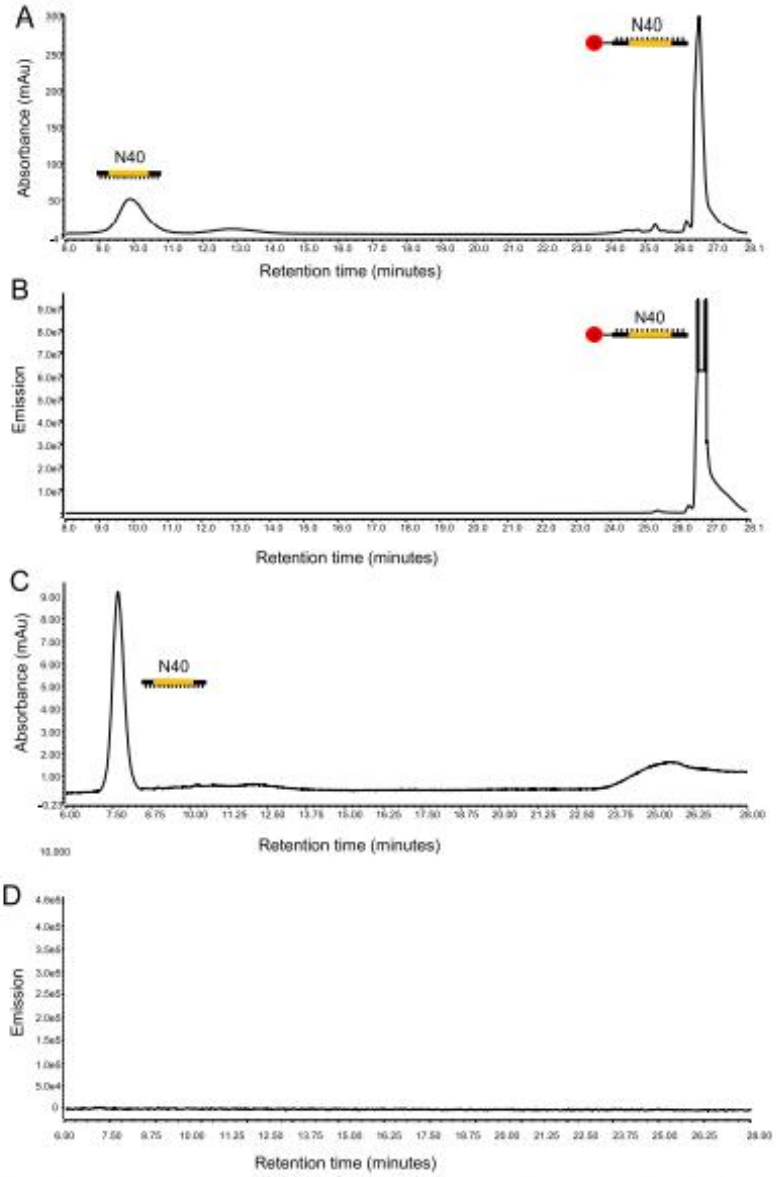


Fig. 6. IP RP HPLC analysis in conjunction with fluorescence detection. Denaturing IP RP chromatogram of the Texas red labelled dsDNA (N40) Gradient 6, 0.3 ml/min A) UV detection 260 nm B) Fluorescence detection: excitation 565 nm, emission 615 nm. Denaturing IP RP HPLC chromatogram of the purified ssDNA (N40). IP RP HPLC chromatogram of the purified ssDNA (N40). A) UV detection 260 nm B) Fluorescence detection: excitation 565 nm, emission 615 nm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

phase (see Supplementary Figure S13). In addition, an alternative mobile phase system was employed using TEA/HFIP in conjunction with denaturing IP RP HPLC. Under these mobile phase conditions it was noted that the PCR products denature at much lower temperatures in comparison to TEAA (see Supplementary Figure S14).

The results show that the duplex dsDNA remains intact only at lower temperatures typically 30 °C and elevated temperature above this effectively denature the dsDNA. Therefore, effective denaturation using this mobile phase was observed at 40 °C (see Supplementary Figure S14). These results demonstrate that employing TEA/HFIP mobile phase or simply the addition of HFIP to the TEAA mobile phase enables effective denaturing of the dsDNA without the requirement for very high temperature and the requirement of specialised columns ovens that can operate at such temperatures. Furthermore, the ability to

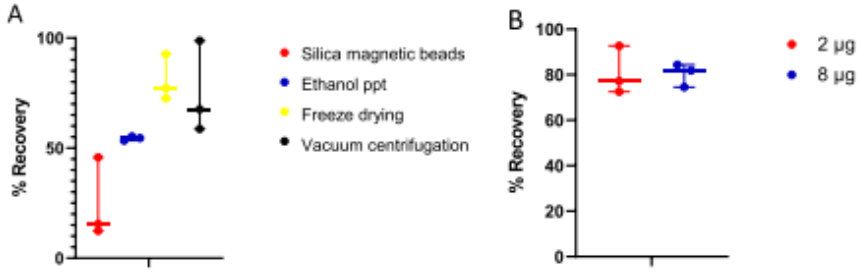


Fig. 7. Comparison of DNA purification methods.% recovery was determined using 2 µg of ssDNA purified using silica coated magnetic beads, ethanol precipitation, freeze drying, and vacuum centrifugation ($n = 3$). B):% recovery of 2 µg and 8 µg ssDNA using freeze-drying ($n = 3$).

denature the dsDNA more effectively at lower temperatures is beneficial for the isolation of ssDNA from PCR products with high melting temperatures and high GC content.

3.5. ssDNA library elution via denaturing HPLC

For application of the method in a SELEX protocol it must be successfully applied to the isolation of ssDNA from the PCR products of a SELEX library containing a random region of DNA. This presents a greater challenge compared to the non-randomised DNA template previously used as the DNA peaks observed on the IP RP HPLC will likely be broader due to the effects of the random region and further complicated by the presence of PCR by-products which are often generated in the PCR of DNA during SELEX.

Following the successful purification of ssDNA, the method was adopted for the analysis and purification of ssDNA generated from a random ssDNA library (N40) used in SELEX protocols. Texas Red labelled PCR product was generated from the random ssDNA library (N40) and analysed via denaturing HPLC (see Fig. 5A). The results show the separation of the two ssDNA strands as previously observed for the non-random DNA. A broader peak for the non-labelled ssDNA was observed as expected due to the presence of the random region (N40) and a large number of different sequences present. Furthermore, a potential PCR impurity (ssDNA of a different size) is also observed in the chromatogram. The non-labelled ssDNA was fraction collected, freeze dried prior to analysis using agarose gel electrophoresis and IP RP HPLC (see Figure 5B/C). The results show the successful purification of the ssDNA (N40) and furthermore demonstrates the purification of the ssDNA from potential larger length impurities that were generated in the PCR, which often hamper SELEX methods. In this case mass spectrometry analysis cannot be used to verify the sequences of the two ssDNA following denaturation of and separation of the ssDNA due to the random sequence nature. However, the presence of the 5′ Texas red group allows for fluorescence analysis of both the denatured PCR product and the purified ssDNA (see Fig. 6). The results show that as expected the later eluting ssDNA gave the corresponding fluorescence consistent with previous analysis. In addition, no fluorescence signal was observed for the purified ssDNA which lacks the Texas Red label.

3.6. Purification and fractionation of ssDNA using IP RP HPLC

Following purification of the ssDNA from SELEX PCR products using denaturing IP RP HPLC it is important to ensure that there is minimal sample loss during the removal of the mobile phase prior to re-suspension and folding of the ssDNA in the SELEX binding buffer. The criteria for the desalting method is that it must effectively remove salts that may interfere with the selection stage, must not be too laborious and must not significantly impact yield. Four alternative methods were assessed and the% recovery of the ssDNA were evaluated, including ethanol precipitation, capture on silica coated magnetic beads [28,29] and removal of the volatile mobile phases using either rotary evaporation or freeze drying. In each case equal mass of the ssDNA was collected via HPLC and the% recovery of the ssDNA determined for each of the different methods (see Fig. 7). The results show that the highest yield recovered was obtained using freeze-drying (80.8 ± 10.6%). The freeze-drying method was taken through to the higher scale, comparison of the yield of the 2 µg and 8 µg scales resulted in similar % yields of the ssDNA. Therefore, these results demonstrate that high recovery and yields of ssDNA can be obtained using denaturing IP RP HPLC in conjunction with freeze drying. Furthermore, such approaches enable purification of the correct size single stranded DNA from potential PCR artefacts commonly generated in SELEX procedures. Verification of the ssDNA obtained was confirmed by both mass spectrometry and fluorescence analysis. A comparison was also performed using streptavidin-biotin affinity chromatography to purify ssDNA. The yield was determined to be 55% consistent with previous reports [30]. Verification of the ssDNA isolated using streptavidin-biotin affinity chromatography was performed using LC MS. Therefore, these results demonstrate that a higher yield of ssDNA is obtained using denaturing IP RP HPLC optimised in this workflow in comparison to streptavidin-biotin affinity chromatography.

4. Conclusions

One of the main challenges when performing successful SELEX procedures to generate ssDNA aptamers, is the ability to generate high-quality ssDNA in high yields from the dsDNA PCR products. In this study, we have developed and optimised denaturing IP RP HPLC for the purification of ssDNA demonstrating significant advantages over existing affinity chromatography and nuclease digestion methods. We have optimised the PCR amplification utilising a PCR primer containing a variety of hydrophobic tags (5′ Texas Red, C18, Biotin and phosphorothioate modifications) which enables the separation of the ssDNA based on differences in their hydrophobicity. Optimal separations were achieved using a 5′ Texas Red labelled PCR primer to amplify DNA containing a random (N40) region. Furthermore, optimisation of the mobile phase conditions was also performed using with the addition of HFIP to the mobile phases which enables complete denaturation of the dsDNA at moderate temperatures and circumvents the requirement of high temperatures often required using TEAA. Validation of the ssDNA obtained was performed using high resolution accurate mass spectrometry analysis and fluorescence based HPLC detection. High yields and recovery (80%) of the ssDNA were obtained using freeze-drying to remove the volatile mobile phase. Furthermore this approach enables purification of the desired ssDNA of the correct size removing potential DNA artefacts often generated in the PCR amplification during SELEX procedures, therefore demonstrating advantages over current affinity chromatography and nuclease digestion methods.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

PEC is an Engineering and Physical Sciences Research Council iCASE funded student in collaboration with Porton Biopharma Limited. MJD acknowledges further support from the Biotechnology and Biological Science Research Council (BB/M012166/1).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2024.464699.

References

[1] L. Gold, Epilogue: a Personal Perspective: aptamers after 15 Years. The Aptamer Handbook: Functional Oligonucleotides and Their Applications, 2006, pp. 461–469, https://doi.org/10.1002/3527608192.epil.
[2] T. Wang, C. Chen, L.M. Larcher, R.A. Barrero, R.N. Veedu, Three decades of nucleic acid aptamer technologies: lessons learned, progress and opportunities on aptamer development, Biotechnol. Adv. 37 (2019) 28–50, https://doi.org/10.1016/j.biotechadv.2018.11.001.
[3] S.P. Chaudhari, P.U. Shinde, Aptamer: a new class of oligonucleotide for therapeutic and diagnostic use - review, J. Sci. Technol. 5 (2020) 19–24, https://doi.org/10.46243/jst.2020.v5.i3.pp19-24.
[4] C. Kratschmer, M. Levy, Effect of chemical modifications on Aptamer stability in serum, Nucleic Acid Ther. 27 (2017) 335–344, https://doi.org/10.1089/nat.2017.0680.
[5] M.N. Stojanovic, P. de Prada, D.W. Landry, Aptamer-based folding fluorescent sensor for cocaine, J. Am. Chem. Soc. 123 (2001) 4928–4931, https://doi.org/10.1021/ja0038171.
[6] M. Michaud, E. Jourdan, A. Villet, A. Ravel, C. Grosset, E. Peyrin, A DNA aptamer as a new target-specific chiral selector for HPLC, J. Am. Chem. Soc. 125 (2003) 8672–8679, https://doi.org/10.1021/ja034483t.
[7] H. Kuai, Z. Zhao, L. Mo, H. Liu, X. Hu, T. Fu, X. Zhang, W. Tan, Circular bivalent Aptamers enable in vivo stability and recognition, J. Am. Chem. Soc. 139 (2017) 9128–9131, https://doi.org/10.1021/jacs.7b04547.
[8] N. Komarova, A. Kuznetsov, Inside the black box: what makes Selex better? Molecules 24 (2019) https://doi.org/10.3390/molecules24193598.
[9] F. Tolle, J. Wilke, J. Wengel, G. Mayer, By-product formation in repetitive PCR amplification of DNA libraries during SELEX, PLoS ONE 9 (2014) 1–12, https://doi.org/10.1371/journal.pone.0114693.
[10] M. Avci-Adali, A. Paul, N. Wilhelm, G. Ziemer, H.P. Wendel, Upgrading SELEX technology by using lambda exonuclease digestion for single-stranded DNA generation, Molecules 15 (2010) 1–11, https://doi.org/10.3390/molecules15010001.
[11] R. Wilson, Preparation of single-stranded DNA from PCR products with streptavidin magnetic beads, Nucleic Acid Ther. 21 (2011) 437–440, https://doi.org/10.1089/nat.2011.0322.
[12] A. Paul, M. Avci-Adali, G. Ziemer, H.P. Wendel, Streptavidin-coated magnetic beads for DNA strand separation implicate a multitude of problems during cell-SELEX, Oligonucleotides 19 (2009) 243–254, https://doi.org/10.1089/oli.2009.0194.
[13] T.R. Damase, A.D. Ellington, P.B. Allen, Purification of single-stranded DNA by co-polymerization with acrylamide and electrophoresis, BioTechniques 62 (2017) 275–282, https://doi.org/10.2144/000114557.
[14] M. Svobodová, A. Pinto, P. Nadal, C.K. O' Sullivan, Comparison of different methods for generation of single-stranded DNA for SELEX processes, Anal. Bioanal Chem 404 (2012) 835–842, https://doi.org/10.1007/s00216-012-6183-4.
[15] K. Shao, W. Ding, F. Wang, H. Li, D. Ma, H. Wang, Emulsion PCR: a high efficient way of PCR amplification of random DNA libraries in aptamer selection, PLoS ONE 6 (2011) 1–7, https://doi.org/10.1371/journal.pone.0024910.
[16] M. Takahashi, X. Wu, M. Ho, P. Chomchan, J.J. Rossi, J.C. Burnett, J. Zhou, High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency, Sci. Rep. 6 (2016) 1–14, https://doi.org/10.1038/srep33697.
[17] M. Dickman, D.P. Hornby, Isolation of single-stranded DNA using denaturing DNA chromatography, Anal. Biochem. 284 (2000) 164–167, https://doi.org/10.1006/abio.2000.4669.
[18] Z. Kadlecová, K. Kalíková, E. Tesařová, M. Gilar, Phosphorothioate oligonucleotides separation in ion-pairing reversed-phase liquid chromatography: effect of ion-pairing system, J. Chromatogr. A 1676 (2022), https://doi.org/10.1016/j.chroma.2022.463201.
[19] E.D. Close, A.O. Nwokeoji, D. Milton, K. Cook, D.M. Hindocha, E.C. Hook, H. Wood, M.J. Dickman, Nucleic acid separations using superficially porous silica particles, J. Chromatogr. A 1440 (2016) 135–144, https://doi.org/10.1016/j.chroma.2016.02.057.
[20] C. Vanhinsbergh, E.C. Hook, N. Oxby, M.J. Dickman, Optimization of orthogonal separations for the analysis of oligonucleotides using 2D-LC, J. Chromatogr. B (2023) 123812, https://doi.org/10.1016/j.jchromb.2023.123812.
[21] M. Donegan, J.M. Nguyen, M. Gilar, Effect of ion-pairing reagent hydrophobicity on liquid chromatography and mass spectrometry analysis of oligonucleotides, J. Chromatogr. A 1666 (2022), https://doi.org/10.1016/j.chroma.2022.462860.
[22] L. Peng, W. Jing, G. Yun-Hua, W. Li-Qing, S. Ling-Hui, F. Bo-Qiang, Analysis of fluorescent dye-labeled oligonucleotides by ion-pair reversed-phase high-performance liquid chromatography, Fenxi Huaxue/Chinese J. Anal. Chem. 37 (2009) 1722–1726, https://doi.org/10.1016/S1872-2040(08)60149-8.
[23] P.J. Oefner, C.G. Huber, F. Umlauft, G.-N. Berti, E. Stimpfl, G.K. Bons, High-resolution liquid chromatography of fluorescent dye-labeled nucleic acid, Anal. Biochem. 223 (1994) 39–46, https://www.sciencedirect.com/science/article/pii/S0003269784715430, accessed December 20, 2023.
[24] K.J. Fountain, M. Gilar, Y. Budman, J.C. Gebler, Purification of dye-labeled oligonucleotides by ion-pair reversed-phase high-performance liquid chromatography, 2003. www.elsevier.com/locate/chromb.
[25] C. Anacleto, R. Ouye, N. Schoenbrunner, Orthogonal ion pairing reversed phase liquid chromatography purification of oligonucleotides with bulky fluorophores, J. Chromatogr. A 1329 (2014) 78–82.
[26] S.M. McCarthy, M. Gilar, J. Gebler, Reversed-phase ion-pair liquid chromatography analysis and purification of small interfering RNA, Anal. Biochem. 390 (2009) 181–188, https://doi.org/10.1016/j.ab.2009.03.042.
[27] E.W. Choi, L.V. Nayak, P.J. Bates, Cancer-selective antiproliferative activity is a general property of some G-rich oligodeoxynucleotides, Nucleic. Acids. Res. 38 (2009) 1623–1635, https://doi.org/10.1093/nar/gkp1088.
[28] R. Fregel, A. González, V.M. Cabrera, Improved ethanol precipitation of DNA, Electrophoresis 31 (2010) 1350–1352, https://doi.org/10.1002/elps.200900721.
[29] P. Oberacker, P. Stepper, D.M. Bond, S. Höhn, J. Focken, V. Meyer, L. Schelle, V. J. Sugrue, G.J. Jeunen, T. Moser, S.R. Hore, F. von Meyenn, K. Hipp, T.A. Hore, T. P. Jurkowski, Bio-On-Magnetic-Beads (BOMB): open platform for high-throughput nucleic acid extraction and manipulation, PLoS Biol. 17 (2019) 1–16, https://doi.org/10.1371/journal.pbio.3000107.
[30] S.C. Huang, M.D. Stump, R. Weiss, K.D. Caldwell, Binding of biotinylated DNA to streptavidin-coated polystyrene latex: effects of chain length and particle size, Anal. Biochem. 237 (1996) 115–122, https://doi.org/10.1006/abio.1996.0208.

8 References

[1] G.M. Blackburn, M.J. Gait, D. Loakes, Williams D.M, Nucleic Acids in Chemistry and Biology, Royal Society of Chemistry, Cambridge, 2007. https://doi.org/10.1039/9781847555380.

[2] W.A. Jacobs, P.A. Levene, Further studies on the constitution of inosinic acid, Exp Biol Med 6 (1909) 90–90. https://doi.org/10.3181/00379727-6-42.

[3] Z. Huang, H. Chen, H. Ye, Z. Chen, N. Jaffrezic-Renault, Z. Guo, An ultrasensitive aptamer-antibody sandwich cortisol sensor for the noninvasive monitoring of stress state, Biosens Bioelectron 190 (2021) 113451. https://doi.org/10.1016/j.bios.2021.113451.

[4] R.E. Franklin, R.G. Gosling, Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate, Nature 172 (1953) 156–157. https://doi.org/10.1038/172156a0.

[5] J.D. Watson, F.H.C. Crick, Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid, Nature 171 (1953) 737–738. https://doi.org/10.1038/171737a0.

[6] R. Langridge, W.E. Seeds, H.R. Wilson, C.W. Hooper, M.H.F. Wilkins, L.D. Hamilton, Molecular Structure of Ribose Nucleic Acid (DNA), J Biophys Biochem Cytol 3 (1957) 767–778. https://doi.org/10.1083/jcb.3.5.767.

[7] M.L. Bochman, K. Paeschke, V.A. Zakian, DNA secondary structures: Stability and function of G-quadruplex structures, Nat Rev Genet 13 (2012) 770–780. https://doi.org/10.1038/nrg3296.

[8] S.H. Chou, K.H. Chin, A.H.J. Wang, Unusual DNA duplex and hairpin motifs, Nucleic Acids Res 31 (2003) 2461–2474. https://doi.org/10.1093/nar/gkg367.

[9] D. Sen, W. Gilbert, Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, Letters to Nature 334 (1988) 364–366.

[10] D. Svozil, J. Kalina, M. Omelka, B. Schneider, DNA conformations and their sequence preferences, Nucleic Acids Res 36 (2008) 3690–3706. https://doi.org/10.1093/nar/gkn260.

[11] T. Friedmann, R. Roblin, Gene Therapy for Human Genetic Disease?, Science (1979) 175 (1972) 949–955. https://www.science.org.

[12] J.A. Kulkarni, D. Witzigmann, S.B. Thomson, S. Chen, B.R. Leavitt, P.R. Cullis, R. van der Meel, The current landscape of nucleic acid therapeutics, Nat Nanotechnol 16 (2021) 630–643. https://doi.org/10.1038/s41565-021-00898-0.

[13] M. Egli, M. Manoharan, Chemistry, structure and function of approved oligonucleotide therapeutics, Nucleic Acids Res 51 (2023) 2529–2573. https://doi.org/10.1093/nar/gkad067.

[14] A. Curreri, D. Sankholkar, S. Mitragotri, Z. Zhao, RNA therapeutics in the clinic, Bioeng Transl Med 8 (2023). https://doi.org/10.1002/btm2.10374.

[15] N. Pardi, M.J. Hogan, F.W. Porter, D. Weissman, mRNA vaccines-a new era in vaccinology, Nat Rev Drug Discov 17 (2018) 261–279. https://doi.org/10.1038/nrd.2017.243.

[16] T. Kramps, J. Probst, Messenger RNA-based vaccines: Progress, challenges, applications, Wiley Interdiscip Rev RNA 4 (2013) 737–749. https://doi.org/10.1002/wrna.1189.

[17] J. Probst, B. Weide, B. Scheel, B.J. Pichler, I. Hoerr, H.G. Rammensee, S. Pascolo, Spontaneous cellular uptake of exogenous messenger RNA in vivo is nucleic acid-specific, saturable and ion dependent, Gene Ther 14 (2007) 1175–1180. https://doi.org/10.1038/sj.gt.3302964.

[18] P.A. Krieg, D.A. Melton, Functional messenger RNAs are produced by SP6 in vitro transcription of cloned cDNAs, Nucleic Acids Res 12 (1984) 7057–7070. https://doi.org/https://doi.org/10.1093/nar/12.18.7057.

[19] X. Xia, Detailed dissection and critical evaluation of the pfizer/biontech and moderna mrna vaccines, Vaccines (Basel) 9 (2021). https://doi.org/10.3390/vaccines9070734.

[20] H. Zogg, R. Singh, S. Ro, Current Advances in RNA Therapeutics for Human Diseases, Int J Mol Sci 23 (2022) 1–23. https://doi.org/10.3390/ijms23052736.

[21] V.R. Litvinova, A.P. Rudometov, L.I. Karpenko, A.A. Ilyichev, mRNA Vaccine Platform: mRNA Production and Delivery, Russ J Bioorg Chem 49 (2023) 220–235. https://doi.org/10.1134/S1068162023020152.

[22] R. Weiss, S. Scheiblhofer, E. Roesler, E. Weinberger, J. Thalhamer, MRNA vaccination as a safe approach for specific protection from type i allergy, Expert Rev Vaccines 11 (2012) 55–67. https://doi.org/10.1586/erv.11.168.

[23] K. Karikó, H. Muramatsu, F.A. Welsh, J. Ludwig, H. Kato, S. Akira, D. Weissman, Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability, Molecular Therapy 16 (2008) 1833–1840. https://doi.org/10.1038/mt.2008.200.

[24] K. Karikó, D. Weissman, Naturally occurring nucleoside modifications suppress the immunostimulatory activity of RNA: Implication for therapeutic RNA development, Curr Opin Drug Discov Devel 10 (2007) 523–532. https://www.researchgate.net/publication/6053392.

[25] A.L. Fuchs, A. Neu, R. Sprangers, A general method for rapid and cost-efficient large-scale production of 5’ capped RNA, RNA 22 (2016) 1454–1466. https://doi.org/10.1261/rna.056614.116.

[26] S.S. Rosa, D.M.F. Prazeres, A.M. Azevedo, M.P.C. Marques, mRNA vaccines manufacturing: Challenges and bottlenecks, Vaccine 39 (2021) 2190–2200. https://doi.org/10.1016/j.vaccine.2021.03.038.

[27] C.A. Deutsch, J.J. Tewksbury, M. Tigchelaar, D.S. Battisti, S.C. Merrill, R.B. Huey, R.L. Naylor, CLIMATE CHANGE Increase in crop losses to insect pests in a warming climate, 2018. https://www.science.org.

[28] J. Hough, J.D. Howard, S. Brown, D.E. Portwood, P.M. Kilby, M.J. Dickman, Strategies for the production of dsRNA biocontrols as alternatives to chemical pesticides, Front Bioeng Biotechnol 10 (2022). https://doi.org/10.3389/fbioe.2022.980592.

[29] V.A. Hilder, D. Boulter, Genetic engineering of crop plants for insect resistance - a critical review, Crop Protection 18 (1999) 177–191. https://doi.org/https://doi.org/10.1016/S0261-2194(99)00028-9.

[30] W. Zhang, Global pesticide use: Profile, trend, cost / benefit and more, Proceedings of the International Academy of Ecology and Environmental Sciences 8 (2018) 1–27. https://www.researchgate.net/publication/323302056_Global_pesticide_use_Profile_trend_cost_benefit_and_more (accessed September 22, 2024).

[31] S. Savary, A. Ficke, J.N. Aubertot, C. Hollier, Crop losses due to diseases and their implications for global food production losses and food security, Food Secur 4 (2012) 519–537. https://doi.org/10.1007/s12571-012-0200-5.

[32] A. Fire, S.Q. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, C.C. Mello, Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans, Letters to Nature 391 (1998) 806–811.

[33] A.J. Hamilton, D.C. Baulcombe, A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants, Proc. Natl. Acad. Sci. U.S.A 124 (1997) 2024. https://doi.org/https://doi.org/10.1126/science.286.5441.950.

[34] L. He, Y. Huang, X. Tang, RNAi-based pest control: Production, application and the fate of dsRNA, Front Bioeng Biotechnol 10 (2022). https://doi.org/10.3389/fbioe.2022.1080576.

[35] T.B. Rodrigues, S.K. Mishra, K. Sridharan, E.R. Barnes, A. Alyokhin, R. Tuttle, W. Kokulapalan, D. Garby, N.J. Skizim, Y.W. Tang, B. Manley, L. Aulisa, R.D. Flannagan, C. Cobb, K.E. Narva, First Sprayable Double-Stranded RNA-Based Biopesticide Product Targets Proteasome Subunit Beta Type-5 in Colorado Potato Beetle (Leptinotarsa decemlineata), Front Plant Sci 12 (2021). https://doi.org/10.3389/fpls.2021.728652.

[36] A. Dalakouras, V. Koidou, K. Papadopoulou, DsRNA-based pesticides: Considerations for efficiency and risk assessment, Chemosphere 352 (2024). https://doi.org/10.1016/j.chemosphere.2024.141530.

[37] L. Timmons, A. Fire, Specific interference by ingested dsRNA, Nature (1998).

[38] X.S. Wan, M.R. Shi, J. Xu, J.H. Liu, H. Ye, Interference Efficiency and Effects of Bacterium-mediated RNAi in the Fall Armyworm (Lepidoptera: Noctuidae), Journal of Insect Science 21 (2021). https://doi.org/10.1093/jisesa/ieab073.

[39] P. Bachman, J. Fischer, Z. Song, E. Urbanczyk-Wochniak, G. Watson, Environmental Fate and Dissipation of Applied dsRNA in Soil, Aquatic Systems, and Plants, Front Plant Sci 11 (2020). https://doi.org/10.3389/fpls.2020.00021.

[40] M. Zotti, E.A. dos Santos, D. Cagliari, O. Christiaens, C.N.T. Taning, G. Smagghe, RNA interference technology in crop protection against arthropod pests, pathogens and nematodes, Pest Manag Sci 74 (2018) 1239–1250. https://doi.org/10.1002/ps.4813.

[41] M.J. Lutz, D.W. Will, G. Breipohl, S.A. Benner, E. Uhlmann, Synthesis of a monocharged peptide nucleic acid (PNA) analog and its recognition as substrate by DNA polymerases, Nucleosides Nucleotides 18 (1999) 393–401. https://doi.org/10.1080/15257779908043084.

[42] K. Kruger, P.J. Grabowski, A.J. Zaug, J. Sands, D.E. Gottschling, T.R. Cech, Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena, Cell 31 (1982) 147–157. https://doi.org/10.1016/0092-8674(82)90414-7.

[43] W. Gilbert, The RNA world Superlattices point ahead, Nature 319 (1986) 618.

[44] C. Tuerk, L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, Science (1979) 249 (1990) 505–510. https://doi.org/10.1126/science.2200121.

[45] A.D. Ellington, J.W. Szostak, In vitro selection of RNA molecules that bind specific ligands, Nature 346 (1990) 818–822. https://doi.org/10.1038/346818a0.

[46] A. Galán, L. Comor, A. Horvatić, J. Kuleš, N. Guillemin, V. Mrljak, M. Bhide, Library-based display technologies: Where do we stand?, Mol Biosyst 12 (2016) 2342–2358. https://doi.org/10.1039/c6mb00219f.

[47] S.J. Klug, M. Famulok, All you wanted to know about SELEX, Mol Biol Rep 20 (1994) 97–107. https://doi.org/10.1007/BF00996358.

[48] T. Wang, C. Chen, L.M. Larcher, R.A. Barrero, R.N. Veedu, Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development, Biotechnol Adv 37 (2019) 28–50. https://doi.org/10.1016/j.biotechadv.2018.11.001.

[49] D. Han, C. Wu, W. Tan, Aptamers Selected by Cell-SELEX for Theranostics: Molecular engineering to enhance aptamer functionality, 2015. https://doi.org/10.1007/978-3-662-46226-3.

[50] C. Forier, E. Boschetti, M. Ouhammouch, A. Cibiel, F. Ducongé, M. Nogré, M. Tellier, D. Bataille, N. Bihoreau, P. Santambien, S. Chtourou, G. Perret, DNA aptamer affinity ligands for highly selective purification of human plasma-related proteins from multiple sources, J Chromatogr A 1489 (2017) 39–50. https://doi.org/10.1016/j.chroma.2017.01.031.

[51] L. Gold, Epilogue: A Personal Perspective: Aptamers after 15 Years, The Aptamer Handbook: Functional Oligonucleotides and Their Applications (2006) 461–469. https://doi.org/10.1002/3527608192.epil.

[52] S.P. Chaudhari, P.U. Shinde, Aptamer: A New Class of Oligonucleotide for Therapeutic and Diagnostic Use - Review, Journal of Science and Technology 5 (2020) 19–24. https://doi.org/10.46243/jst.2020.v5.i3.pp19-24.

[53] K. Sefah, Z. Yang, K.M. Bradley, S. Hoshika, E. Jiménez, L. Zhang, G. Zhu, S. Shanker, F. Yu, D. Turek, W. Tan, S.A. Benner, In vitro selection with artificial expanded genetic information systems, Proc Natl Acad Sci U S A 111 (2014) 1449–1454. https://doi.org/10.1073/pnas.1311778111.

[54] N. Komarova, A. Kuznetsov, Inside the black box: What makes Selex better?, Molecules 24 (2019) 1–23. https://doi.org/10.3390/molecules24193598.

[55] B. Hall, J.M. Micheletti, P. Satya, K. Ogle, J. Pollard, A.D. Ellington, Design, Synthesis, and Amplification of DNA Pools for In Vitro Selection, Curr Protoc Nucleic Acid Chem 39 (2009). https://doi.org/10.1002/0471142700.nc0902s39.

[56] J.W. Ellington, Andrew D. Szostak, Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures, Nature 355 (1992) 850–852.

[57] A.C. Cheng, V. Calabro, A.D. Frankel, Design of RNA-binding proteins and ligands, Curr Opin Struct Biol 11 (2001) 478–484. https://doi.org/10.1016/S0959-440X(00)00236-0.

[58] A. V. Lakhin, V.Z. Tarantul, L. V. Gening, Aptamers: Problems, solutions and prospects, Acta Naturae 5 (2013) 34–43. https://doi.org/10.32607/20758251-2013-5-4-34-43.

[59] B.E. Eaton, The joys of in vitro selection: Chemically dressing oligonucleotides to satiate protein targets, Curr Opin Chem Biol 1 (1997) 10–16. https://doi.org/10.1016/s1367-5931(97)80103-2.

[60] B. Katz, M. Goldbaum, Macugen (Pegaptanib Sodium), a Novel Ocular Therapeutic That Targets Vascular Endothelial Growth Factor (VEGF) Barrett Katz , MD , MBA Mauro Goldbaum , MD, Growth (Lakeland) (n.d.) 141–154.

[61] D.H.J. Bunka, P.G. Stockley, Aptamers come of age - At last, Nat Rev Microbiol 4 (2006) 588–596. https://doi.org/10.1038/nrmicro1458.

[62] Q. Zhu, G. Liu, M. Kai, A.O.A. Miller, DNA aptamers in the diagnosis and treatment of human diseases, Molecules 20 (2015) 20979–90997. https://doi.org/10.3390/molecules201219739.

[63] R. Stevenson, Specific Detection of 14-3-3 Proteins by ssDNA Aptamers Ross Stevenson The University of Edinburgh, University of Edinburgh, 2006.

[64] P. Bayat, R. Nosrati, M. Alibolandi, H. Rafatpanah, K. Abnous, M. Khedri, M. Ramezani, SELEX methods on the road to protein targeting with nucleic acid aptamers, Biochimie 154 (2018) 132–155. https://doi.org/10.1016/j.biochi.2018.09.001.

[65] S. Jeong, S.R. Han, Y.J. Lee, J.H. Kim, S.W. Lee, Identification of RNA aptamer specific to mutant KRAS protein, Oligonucleotides 20 (2010) 155–161. https://doi.org/10.1089/oli.2010.0231.

[66] R. Conrad, L.M. Keranen, A.D. Ellington, A.C. Newton, Isozyme-specific inhibition of protein kinase C by RNA aptamers, Journal of Biological Chemistry 269 (1994) 32051–32054. https://doi.org/https://doi.org/10.1016/S0021-9258(18)31598-9.

[67] B.J. Hicke, C. Marion, Y.F. Chang, T. Gould, C.K. Lynott, D. Parma, P.G. Schmidt, S. Warren, Tenascin-C Aptamers Are Generated Using Tumor Cells and Purified Protein, Journal of Biological Chemistry 276 (2001) 48644–48654. https://doi.org/10.1074/jbc.M104651200.

[68] S.I. Hori, A. Herrera, J.J. Rossi, J. Zhou, Current advances in aptamers for cancer diagnosis and therapy, Cancers (Basel) 10 (2018) 1–33. https://doi.org/10.3390/cancers10010009.

[69] Q. Xiang, G. Tan, X. Jiang, K. Wu, W. Tan, Y. Tan, Suppression of FOXM1 Transcriptional Activities via a Single-Stranded DNA Aptamer Generated by SELEX, Sci Rep 7 (2017) 1–12. https://doi.org/10.1038/srep45377.

[70] W.H. Thiel, T. Bair, K. Wyatt Thiel, J.P. Dassie, W.M. Rockey, C.A. Howell, X.Y. Liu, A.J. Dupuy, L. Huang, R. Owczarzy, M.A. Behlke, J.O. McNamara, P.H. Giangrande, Nucleotide bias observed with a short SELEX RNA aptamer library, Nucleic Acid Ther 21 (2011) 253–263. https://doi.org/10.1089/nat.2011.0288.

[71] M. Takahashi, X. Wu, M. Ho, P. Chomchan, J.J. Rossi, J.C. Burnett, J. Zhou, High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency, Sci Rep 6 (2016) 1–14. https://doi.org/10.1038/srep33697.

[72] P.J. Bates, D.A. Laber, D.M. Miller, S.D. Thomas, J.O. Trent, Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer, Exp Mol Pathol 86 (2009) 151–164. https://doi.org/10.1016/j.yexmp.2009.01.004.

[73] C. Platella, C. Riccardi, D. Montesarchio, G.N. Roviello, D. Musumeci, G-quadruplex-based aptamers against protein targets in therapy and diagnostics, Biochim Biophys Acta Gen Subj 1861 (2017) 1429–1447. https://doi.org/10.1016/j.bbagen.2016.11.027.

[74] F. Tolle, J. Wilke, J. Wengel, G. Mayer, By-product formation in repetitive PCR amplification of DNA libraries during SELEX, PLoS One 9 (2014) 1–12. https://doi.org/10.1371/journal.pone.0114693.

[75] A. Crameri, W.P.C. Stemmer, 1020Fold Aptamer Library Amplification Without Gel Purification, Nucleic Acids Res 21 (1993) 4410. https://doi.org/10.1093/nar/21.18.4410.

[76] J. Pollard, S.D. Bell, A.D. Ellington, Design, Synthesis, and Amplification of DNA Pools for Construction of Combinatorial Pools and Libraries , Curr Protoc Mol Biol 52 (2000) 1–24. https://doi.org/10.1002/0471142727.mb2402s52.

[77] K.D. Kovacevic, J.C. Gilbert, B. Jilma, Pharmacokinetics, pharmacodynamics and safety of aptamers, Adv Drug Deliv Rev 134 (2018) 36–50. https://doi.org/10.1016/j.addr.2018.10.008.

[78] M.L. Andreola, C. Calmels, J. Michel, J.J. Toulmé, S. Litvak, Towards the selection of phosphorothioate aptamers: Optimizing in vitro selection steps with phosphorothioate nucleotides, Eur J Biochem 267 (2000) 5032–5040. https://doi.org/10.1046/j.1432-1327.2000.01557.x.

[79] M.A. Dellafiore, J.M. Montserrat, A.M. Iribarren, Modified nucleoside triphosphates for in-vitro selection techniques, Front Chem 4 (2016) 1–13. https://doi.org/10.3389/fchem.2016.00018.

[80] W. Kusser, Chemically modified nucleic acid aptamers for in vitro selections: evolving evolution, Reviews in Molecular Biotechnology 74 (2000) 27–38. https://doi.org/https://doi.org/10.1016/S1389-0352(99)00002-1.

[81] R. E. Wang, H. Wu, Y. Niu, J. Cai, Improving the Stability of Aptamers by Chemical Modification, Curr Med Chem 18 (2011) 4126–4138. https://doi.org/10.2174/092986711797189565.

[82] J.C. Rohloff, A.D. Gelinas, T.C. Jarvis, U.A. Ochsner, D.J. Schneider, L. Gold, N. Janjic, Nucleic acid ligands with protein-like side chains: Modified aptamers and their use as diagnostic and therapeutic agents, Mol Ther Nucleic Acids 3 (2014) 201–214. https://doi.org/10.1038/mtna.2014.49.

[83] A.D. Keefe, S.T. Cload, SELEX with modified nucleotides, Curr Opin Chem Biol 12 (2008) 448–456. https://doi.org/10.1016/j.cbpa.2008.06.028.

[84] L. Cerchia, J. Hamm, D. Libri, B. Tavitian, V. De Franciscis, Nucleic acid aptamers in cancer medicine, FEBS Lett 528 (2002) 12–16. https://doi.org/10.1016/S0014-5793(02)03275-1.

[85] S.M. Shamah, J.M. Healy, S.T. Cload, Complex target SELEX, Acc Chem Res 41 (2008) 130–138. https://doi.org/10.1021/ar700142z.

[86] J.C. Cox, A. Hayhurst, J. Hesselberth, T.S. Bayer, G. Georgiou, A.D. Ellington, Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer., Nucleic Acids Res 30 (2002) 1–14. https://doi.org/10.1093/nar/gnf107.

[87] L.I. Hernandez, K.S. Flenker, F.J. Hernandez, A.J. Klingelhutz, J.O. McNamara, P.H. Giangrande, Methods for evaluating cell-specific, cell-internalizing RNA aptamers, Pharmaceuticals 6 (2013) 295–319. https://doi.org/10.3390/ph6030295.

[88] R. Stoltenburg, C. Reinemann, B. Strehlitz, FluMag-SELEX as an advantageous method for DNA aptamer selection, Anal Bioanal Chem 383 (2005) 83–91. https://doi.org/10.1007/s00216-005-3388-9.

[89] M. Bianchini, M. Radrizzani, M.G. Brocardo, G.B. Reyes, C. Gonzalez Solveyra, T.A. Santa-Coloma, Specific oligobodies against ERK-2 that recognize both the native and the denatured state of the protein, J Immunol Methods 252 (2001) 191–197. https://doi.org/10.1016/S0022-1759(01)00350-7.

[90] E. Kowalska, F. Bartnicki, K. Pels, W. Strzalka, The impact of immobilized metal affinity chromatography (IMAC) resins on DNA aptamer selection, Anal Bioanal Chem 406 (2014) 5495–5499. https://doi.org/10.1007/s00216-014-7937-y.

[91] M. Darmostuk, S. Rimpelova, H. Gbelcova, T. Ruml, Current approaches in SELEX: An update to aptamer selection technology, Biotechnol Adv 33 (2014) 1141–1161. https://doi.org/10.1016/j.biotechadv.2015.02.008.

[92] K.T. Urak, S. Shore, W.M. Rockey, S.J. Chen, A.P. McCaffrey, P.H. Giangrande, In vitro RNA SELEX for the generation of chemically-optimized therapeutic RNA drugs, Methods 103 (2016) 167–174. https://doi.org/10.1016/j.ymeth.2016.03.003.

[93] R.K. Mosing, S.D. Mendonsa, M.T. Bowser, Capillary electrophoresis-SELEX selection of aptamers with affinity for HIV-1 reverse transcriptase, Anal Chem 77 (2005) 6107–6112. https://doi.org/10.1021/ac050836q.

[94] J.G. Bruno, In vitro selection of DNA to chloroaromatics using magnetic microbead-based affinity separation and fluorescence detection, Biochem Biophys Res Commun 234 (1997) 117–120. https://doi.org/10.1006/bbrc.1997.6517.

[95] D.R. Latulippe, K. Szeto, A. Ozer, F.M. Duarte, C. V. Kelly, J.M. Pagano, B.S. White, D. Shalloway, J.T. Lis, H.G. Craighead, Multiplexed microcolumn-based process for efficient selection of RNA aptamers, Anal Chem 85 (2013) 3417–3424. https://doi.org/10.1021/ac400105e.

[96] N. Mencin, T. Šmuc, M. Vraničar, J. Mavri, M. Hren, K. Galeša, P. Krkoč, H. Ulrich, B. Šolar, Optimization of SELEX: Comparison of different methods for monitoring the progress of in vitro selection of aptamers, J Pharm Biomed Anal 91 (2014) 151–159. https://doi.org/10.1016/j.jpba.2013.12.031.

[97] M. Sola, A.P. Menon, B. Moreno, D. Meraviglia-Crivelli, M.M. Soldevilla, F. Cartón-García, F. Pastor, Aptamers Against Live Targets: Is In Vivo SELEX Finally Coming to the Edge?, Mol Ther Nucleic Acids 21 (2020) 192–204. https://doi.org/10.1016/j.omtn.2020.05.025.

[98] I.F. Tannock, D. Rotin, Acid pH in Tumors and Its Potential for Therapeutic Exploitation, Cancer Res 49 (1989) 4373–4384.

[99] R.K. Saiki, D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, H.A. Erlich, Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science (1979) 239 (1988) 487–491. https://doi.org/10.1126/science.2448875.

[100] V.K. Viswanathan, K. Krcmarik, N.P. Cianciotto, Template secondary structure promotes polymerase jumping during PCR amplification, Biotechniques 27 (1999) 508–511. https://doi.org/10.2144/99273st04.

[101] K. Shao, W. Ding, F. Wang, H. Li, D. Ma, H. Wang, Emulsion PCR: A high efficient way of PCR amplification of random DNA libraries in aptamer selection, PLoS One 6 (2011) 1–7. https://doi.org/10.1371/journal.pone.0024910.

[102] A. Levay, R. Brenneman, J. Hoinka, D. Sant, M. Cardone, G. Trinchieri, T.M. Przytycka, A. Berezhnoy, Identifying high-affinity aptamer ligands with defined cross-reactivity using high-throughput guided systematic evolution of ligands by exponential enrichment, Nucleic Acids Res 43 (2015) 1–10. https://doi.org/10.1093/nar/gkv534.

[103] E. Ouellet, J.H. Foley, E.M. Conway, C. Haynes, Hi-Fi SELEX: A high-fidelity digital-PCR based therapeutic aptamer discovery platform, Biotechnol Bioeng 112 (2015) 1506–1522. https://doi.org/10.1002/bit.25581.

[104] K. Kanokphandharangkul, Aptamer biosensors, Imperial College, 2016.

[105] A. Paul, M. Avci-Adali, G. Ziemer, H.P. Wendel, Streptavidin-coated magnetic beads for DNA strand separation implicate a multitude of problems during cell-SELEX, Oligonucleotides 19 (2009) 243–254. https://doi.org/10.1089/oli.2009.0194.

[106] P.E. Coombes, M.J. Dickman, Optimisation of denaturing ion pair reversed phase HPLC for the purification of ssDNA in SELEX., J Chromatogr A (2024) 1–10. https://doi.org/10.1016/j.chroma.2024.464699.

[107] M. Avci-Adali, A. Paul, N. Wilhelm, G. Ziemer, H.P. Wendel, Upgrading SELEX technology by using lambda exonuclease digestion for single-stranded DNA generation, Molecules 15 (2010) 1–11. https://doi.org/10.3390/molecules15010001.

[108] C. Marimuthu, T.H. Tang, J. Tominaga, S.C. Tan, S.C.B. Gopinath, Single-stranded DNA (ssDNA) production in DNA aptamer generation, Analyst 137 (2012) 1307–1315. https://doi.org/10.1039/c2an15905h.

[109] K. Ji, W.S. Lim, S.F.Y. Li, K. Bhakoo, A two-step stimulus-response cell-SELEX method to generate a DNA aptamer to recognize inflamed human aortic endothelial cells as a potential in vivo molecular probe for atherosclerosis plaque

detection, Anal Bioanal Chem 405 (2013) 6853–6861. https://doi.org/10.1007/s00216-013-7155-z.

[110] M. Svobodová, A. Pinto, P. Nadal, C.K. O' Sullivan, Comparison of different methods for generation of single-stranded DNA for SELEX processes., Anal Bioanal Chem 404 (2012) 835–842. https://doi.org/10.1007/s00216-012-6183-4.

[111] M. Takahashi, X. Wu, M. Ho, P. Chomchan, J.J. Rossi, J.C. Burnett, J. Zhou, High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency, Sci Rep 6 (2016) 1–14. https://doi.org/10.1038/srep33697.

[112] T. Schütze, B. Wilhelm, N. Greiner, H. Braun, F. Peter, M. Mörl, V.A. Erdmann, H. Lehrach, Z. Konthur, M. Menger, P.F. Arndt, J. Glökler, Probing the SELEX process with next-generation sequencing, PLoS One 6 (2011) 1–10. https://doi.org/10.1371/journal.pone.0029604.

[113] T. Yoshitomi, F. Wayama, K. Kimura, K. Wakui, H. Furusho, K. Yoshimoto, Screening of DNA signaling aptamer from multiple candidates obtained from SELEX with next-generation sequencing, Analytical Sciences 35 (2019) 113–116. https://doi.org/10.2116/analsci.18SDN05.

[114] A.B. Kinghorn, L.A. Fraser, S. Lang, S.C.C. Shiu, J.A. Tanner, Aptamer bioinformatics, Int J Mol Sci 18 (2017) 1–22. https://doi.org/10.3390/ijms18122516.

[115] N. Komarova, D. Barkova, A. Kuznetsov, Implementation of high-throughput sequencing (Hts) in aptamer selection technology, Int J Mol Sci 21 (2020) 1–22. https://doi.org/10.3390/ijms21228774.

[116] M. Kircher, P. Heyn, J. Kelso, Addressing challenges in the production and analysis of illumina sequencing data, BMC Genomics 12 (2011) 1–14. https://doi.org/10.1186/1471-2164-12-382.

[117] H. Hasegawa, N. Savory, K. Abe, K. Ikebukuro, Methods for improving aptamer binding affinity, Molecules 21 (2016) 1–15. https://doi.org/10.3390/molecules21040421.

[118] R. Thevendran, M. Citartan, Assays to Estimate the Binding Affinity of Aptamers, Talanta 238 (2022) 1–17. https://doi.org/10.1016/j.talanta.2021.122971.

[119] T. Sakamoto, E. Ennifar, Y. Nakamura, Thermodynamic study of aptamers binding to their target proteins, Biochimie 145 (2018) 91–97. https://doi.org/10.1016/j.biochi.2017.10.010.

[120] Z. Wang, T. Wilkop, D. Xu, Y. Dong, G. Ma, Q. Cheng, Surface plasmon resonance imaging for affinity analysis of aptamer-protein interactions with PDMS microfluidic chips, Anal Bioanal Chem 389 (2007) 819–825. https://doi.org/10.1007/s00216-007-1510-x.

[121] A.L. Chang, M. McKeague, J.C. Liang, C.D. Smolke, Kinetic and equilibrium binding characterization of aptamers to small molecules using a label-free, sensitive, and scalable platform, Anal Chem 86 (2014) 3273–3278. https://doi.org/10.1021/ac5001527.

[122] Y. Fa, M. Guan, H. Zhao, F. Li, H. Liu, Affinity analysis between trypsin and aptamers using surface plasmon resonance competition experiments in a steady state, Analytical Methods 11 (2019) 3061–3065. https://doi.org/10.1039/c9ay00861f.

[123] P.H. Lin, R.H. Chen, C.H. Lee, Y. Chang, C.S. Chen, W.Y. Chen, Studies of the binding mechanism between aptamers and thrombin by circular dichroism, surface plasmon resonance and isothermal titration calorimetry, Colloids Surf B Biointerfaces 88 (2011) 552–558. https://doi.org/10.1016/j.colsurfb.2011.07.032.

[124] M. Jing, M.T. Bowser, Methods for measuring aptamer-protein equilibria: A review, Anal Chim Acta 686 (2011) 9–18. https://doi.org/10.1016/j.aca.2010.10.032.

[125] S. Ratanabunyong, N. Aeksiri, S. Yanaka, M. Yagi-Utsumi, K. Kato, K. Choowongkomon, S. Hannongbua, Characterization of New DNA Aptamers for Anti-HIV-1 Reverse Transcriptase, ChemBioChem 22 (2021) 915–923. https://doi.org/10.1002/cbic.202000633.

[126] M. Jerabek-Willemsen, T. André, R. Wanner, H.M. Roth, S. Duhr, P. Baaske, D. Breitsprecher, MicroScale Thermophoresis: Interaction analysis and beyond, J Mol Struct 1077 (2014) 101–113. https://doi.org/10.1016/j.molstruc.2014.03.009.

[127] M. Jauset Rubio, M. Svobodová, T. Mairal, T. Schubert, S. Künne, G. Mayer, C.K. O'Sullivan, β-Conglutin dual aptamers binding distinct aptatopes, Anal Bioanal Chem 408 (2016) 875–884. https://doi.org/10.1007/s00216-015-9179-z.

[128] A. Miranda, T. Santos, J. Carvalho, D. Alexandre, A. Jardim, C.R.F. Caneira, V. Vaz, B. Pereira, R. Godinho, D. Brito, V. Chu, J.P. Conde, C. Cruz, Aptamer-based approaches to detect nucleolin in prostate cancer, Talanta 226 (2021) 1–9. https://doi.org/10.1016/j.talanta.2020.122037.

[129] Z.W. Wang, H.B. Wu, Z.F. Mao, X.P. Hu, H. Zhang, Z.P. Hu, Z.L. Ren, In vitro selection and identification of ssDNA aptamers recognizing the Ras protein, Mol Med Rep 10 (2014) 1481–1488. https://doi.org/10.3892/mmr.2014.2337.

[130] M. Berezovski, R. Nutiu, Y. Li, S.N. Krylov, Affinity analysis of a protein-aptamer complex using nonequilibrium capillary electrophoresis of equilibrium mixtures, Anal Chem 75 (2003) 1382–1386. https://doi.org/10.1021/ac026214b.

[131] M.N. Stojanovic, P. de Prada, D.W. Landry, Aptamer-based folding fluorescent sensor for cocaine, J Am Chem Soc 123 (2001) 4928–4931. https://doi.org/10.1021/ja0038171.

[132] P. Weerathunge, R. Ramanathan, V.A. Torok, K. Hodgson, Y. Xu, R. Goodacre, B.K. Behera, V. Bansal, Ultrasensitive Colorimetric Detection of Murine Norovirus Using NanoZyme Aptasensor, Anal Chem 91 (2019) 3270–3276. https://doi.org/10.1021/acs.analchem.8b03300.

[133] Q. Zhao, M. Wu, X. Chris Le, X.F. Li, Applications of aptamer affinity chromatography, TrAC - Trends in Analytical Chemistry 41 (2012) 46–57. https://doi.org/10.1016/j.trac.2012.08.005.

[134] H. Ma, J. Liu, M.M. Ali, M.A.I. Mahmood, L. Labanieh, M. Lu, S.M. Iqbal, Q. Zhang, W. Zhao, Y. Wan, Nucleic acid aptamers in cancer research, diagnosis

and therapy, Chem Soc Rev 44 (2015) 1240–1256. https://doi.org/10.1039/c4cs00357h.

[135] R. Faryammanesh, T. Lange, E. Magbanua, S. Haas, C. Meyer, D. Wicklein, U. Schumacher, U. Hahn, SDA, a DNA aptamer inhibiting E- And P-Selectin mediated adhesion of cancer and leukemia cells, the first and pivotal step in transendothelial migration during metastasis formation, PLoS One 9 (2014) 1–8. https://doi.org/10.1371/journal.pone.0093173.

[136] S. Santulli-Marotto, S.K. Nair, C. Rusconi, B. Sullenger, E. Gilboa, Multivalent RNA Aptamers That Inhibit CTLA-4 and Enhance Tumor Immunity, Cancer Res 63 (2003) 7483–7489.

[137] C.L. Esposito, D. Passaro, I. Longobardo, G. Condorelli, P. Marotta, A. Affuso, V. de Franciscis, L. Cerchia, A neutralizing rna aptamer against egfr causes selective apoptotic cell death, PLoS One 6 (2011) 1–12. https://doi.org/10.1371/journal.pone.0024071.

[138] M. Famulok, M. Blind, G. Mayer, Intramers as promising new tools in functional proteomics, Chem Biol 8 (2001) 931–939. https://doi.org/10.1016/S1074-5521(01)00070-9.

[139] R. Yazdian-Robati, P. Bayat, F. Oroojalian, M. Zargari, M. Ramezani, S.M. Taghdisi, K. Abnous, Therapeutic applications of AS1411 aptamer, an update review, Int J Biol Macromol 155 (2020) 1420–1431. https://doi.org/10.1016/j.ijbiomac.2019.11.118.

[140] C.R. Ireson, L.R. Kelland, Discovery and development of anticancer aptamers, Mol Cancer Ther 5 (2006) 2957–2962. https://doi.org/10.1158/1535-7163.MCT-06-0172.

[141] S. Thongchot, K. Aksonnam, P. Thuwajit, P.T. Yenchitsomanus, C. Thuwajit, Nucleolin-based targeting strategies in cancer treatment: Focus on cancer immunotherapy (Review), Int J Mol Med 52 (2023) 1–14. https://doi.org/10.3892/ijmm.2023.5284.

[142] X. Tong, L. Ga, J. Ai, Y. Wang, Progress in cancer drug delivery based on AS1411 oriented nanomaterials, J Nanobiotechnology 20 (2022) 1–36. https://doi.org/10.1186/s12951-022-01240-z.

[143] F. Mahmoudian, A. Ahmari, S. Shabani, B. Sadeghi, S. Fahimirad, F. Fattahi, Aptamers as an approach to targeted cancer therapy, Cancer Cell Int 24 (2024) 1–22. https://doi.org/10.1186/s12935-024-03295-4.

[144] S. Venkatesan, K. Chanda, M.M. Balamurali, Recent Advancements of Aptamers in Cancer Therapy, ACS Omega 8 (2023) 32231–32243. https://doi.org/10.1021/acsomega.3c04345.

[145] P.R. Bouchard, R.M. Hutabarat, K.M. Thompson, Discovery and development of therapeutic aptamers, Annu Rev Pharmacol Toxicol 50 (2010) 237–257. https://doi.org/10.1146/annurev.pharmtox.010909.105547.

[146] G.S. Baird, Where are all the aptamers?, Am J Clin Pathol 134 (2010) 529–531. https://doi.org/10.1309/AJCPFU4CG2WGJJKS.

[147] A.M. Florea, D. Büsselberg, Cisplatin as an anti-tumor drug: Cellular mechanisms of activity, drug resistance and induced side effects, Cancers (Basel) 3 (2011) 1351–1371. https://doi.org/10.3390/cancers3011351.

[148] I. Knezevic, M.A. Liu, K. Peden, T. Zhou, H.N. Kang, Development of mRNA vaccines: Scientific and regulatory issues, Vaccines (Basel) 9 (2021) 1–11. https://doi.org/10.3390/vaccines9020081.

[149] USP mRNA Vaccine Chapter ANALYTICAL PROCEDURES FOR mRNA VACCINES, (n.d.). https://www.uspnf.com/sites/default/files/usp_pdf/EN/USPNF/usp-nf-notices/gc-xxx-analytical-procedures-mRNA-vaccines.pdf.

[150] A. Dousis, K. Ravichandran, E.M. Hobert, M.J. Moore, A.E. Rabideau, An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts, Nat Biotechnol 41 (2023) 560–568. https://doi.org/10.1038/s41587-022-01525-6.

[151] A. Dietz-Pfeilstetter, M. Mendelsohn, A. Gathmann, D. Klinkenbuß, Considerations and Regulatory Approaches in the USA and in the EU for

dsRNA-Based Externally Applied Pesticides for Plant Protection, Front Plant Sci 12 (2021) 1–7. https://doi.org/10.3389/fpls.2021.682387.

[152] A.O. Nwokeoji, E.A. Nwokeoji, T. Chou, A. Togola, A novel sustainable platform for scaled manufacturing of double-stranded RNA biopesticides, Bioresour Bioprocess 9 (2022) 1–16. https://doi.org/10.1186/s40643-022-00596-2.

[153] J.R. Oann, Recent advances in the theory and practice of electrophoresis, Immunochemistry 5 (1968) 107–134. https://doi.org/https://doi.org/10.1016/0019-2791(68)90229-2.

[154] T. Masek, V. Vopalensky, P. Suchomelova, M. Pospisek, Denaturing RNA electrophoresis in TAE agarose gels, Anal Biochem 336 (2005) 46–50. https://doi.org/10.1016/j.ab.2004.09.010.

[155] J.C. Marini, S.D. Levenet, D.M. Crotherst, P.T. Englund, Bent helical structure in kinetoplast DNA (DNA conformation/sequence periodicity/trypanosomatid/mitochondrial DNA), Biochemistry 79 (1982) 7664–7668. https://doi.org/https://doi.org/10.1073/pnas.79.24.7664.

[156] H.-M. Wu, D.M. Crothers, The Locus of Sequence-directed and protein-induced DNA bending, Nature 5 (1984) 509–513.

[157] O.J. Lumpkin, P. Dlkjardin, B.H. Zimm, Theory of Gel Electrophoresis of DNA, Biopolymers (1985) 1–21.

[158] T. Lu, L.J. Klein, S. Ha, R.R. Rustandi, High-Resolution capillary electrophoresis separation of large RNA under non-aqueous conditions, J Chromatogr A 1618 (2020) 1–17. https://doi.org/10.1016/j.chroma.2020.460875.

[159] A.G. Ewing, R.A. Wallingford, T.M. Olefirowicz, Capillary Electrophoresis, Anal Chem 61 (1989) 292–303. https://pubs.acs.org/sharingguidelines.

[160] B. Wei, A. Goyon, K. Zhang, Analysis of therapeutic nucleic acids by capillary electrophoresis, J Pharm Biomed Anal 219 (2022) 1–11. https://doi.org/10.1016/j.jpba.2022.114928.

[161] C.J. Vanhinsbergh, A. Criscuolo, J.N. Sutton, K. Murphy, A.J.K. Williamson, K. Cook, M.J. Dickman, Characterization and Sequence Mapping of Large RNA

and mRNA Therapeutics Using Mass Spectrometry, Anal Chem 94 (2022) 7339–7349. https://doi.org/10.1021/acs.analchem.2c00765.

[162] S. Honoré Hansen, General Chromatographic Theory and Principles, in: Bioanalysis of Pharmaceuticals: Sample Preparation, Separation Techniques, and Mass Spectrometry, 2015: pp. 31–59.

[163] C.G. Huber, Micropellicular stationary phases for high-performance liquid chromatography of double-stranded DNA, J Chromatogr A 806 (1998) 3–30. https://doi.org/10.1016/S0021-9673(97)01124-2.

[164] A.O. Nwokeoji, A.W. Kung, P.M. Kilby, D.E. Portwood, M.J. Dickman, Purification and characterisation of dsRNA using ion pair reverse phase chromatography and mass spectrometry, J Chromatogr A 1484 (2017) 14–25. https://doi.org/10.1016/j.chroma.2016.12.062.

[165] X. Feng, Z. Su, Y. Cheng, G. Ma, S. Zhang, Messenger RNA chromatographic purification: advances and challenges, J Chromatogr A 1707 (2023) 1–14. https://doi.org/10.1016/j.chroma.2023.464321.

[166] J.R. Thayer, V. Barreto, S. Rao, C. Pohl, Control of oligonucleotide retention on a pH-stabilized strong anion exchange column, Anal Biochem 338 (2005) 39–47. https://doi.org/10.1016/j.ab.2004.11.013.

[167] S. Fekete, C. Doneanu, B. Addepalli, M. Gaye, J. Nguyen, B. Alden, R. Birdsall, D. Han, G. Isaac, M. Lauber, Challenges and emerging trends in liquid chromatography-based analyses of mRNA pharmaceuticals, J Pharm Biomed Anal 224 (2023) 1–13. https://doi.org/10.1016/j.jpba.2022.115174.

[168] C.G. Huber, Micropellicular stationary phases for high-performance liquid chromatography of double-stranded DNA, J Chromatogr A 806 (1998) 3–30. https://doi.org/https://doi.org/10.1016/S0021-9673(97)01124-2.

[169] F. Fack, V. Sarantoglou, Curved DNA fragments display retarded elution upon anion exchange HPLC, Nucleic Acids Res 19 (1991) 4181–4188. https://doi.org/https://doi.org/10.1093/nar/19.15.4181.

[170] P.J. Oefner, C.G. Huber, A decade of high-resolution liquid chromatography of nucleic acids on styrene-divinylbenzene copolymers, Journal of

Chromatography B 782 (2002) 27–55. https://doi.org/https://doi.org/10.1016/S1570-0232(02)00700-6.

[171] I.C. Santos, J.S. Brodbelt, Recent developments in the characterization of nucleic acids by liquid chromatography, capillary electrophoresis, ion mobility, and mass spectrometry (2010–2020), J Sep Sci 44 (2021) 340–372. https://doi.org/10.1002/jssc.202000833.

[172] M. Dickman, Ion pair reverse-phase chromatography Ion Pair Reverse-Phase Chromatography: A Versatile Platform for the Analysis of RNA, 2011. https://www.chromatographytoday.com/article/preparative/33/unassigned-independent-article/ion-pair-reverse-phase-chromatography-a-versatile-platform-for-the-analysis-of-rna/984.

[173] C.G. Huber, P.J. Oefner, E. Preuss, G.K. Bonn, High-resolution liquid chromatography of DNA fragments on non-porous poly(styrene-divinylbenzene) particles, Nucleic Acids Res 21 (1993) 1061–1066. https://academic.oup.com/nar/article/21/5/1061/1027180.

[174] N. Li, N.M. El Zahar, J.G. Saad, E.R.E. van der Hage, M.G. Bartlett, Alkylamine ion-pairing reagents and the chromatographic separation of oligonucleotides, J Chromatogr A 1580 (2018) 110–119. https://doi.org/10.1016/j.chroma.2018.10.040.

[175] C.G. Huber, H. Oberacher, Analysis of nucleic acids by on-line liquid chromatography-mass spectrometry, Mass Spectrom Rev 20 (2001) 310–343. https://doi.org/10.1002/mas.10011.

[176] A. Liu, M. Cheng, Y. Zhou, P. Deng, Bioanalysis of Oligonucleotide by LC–MS: Effects of Ion Pairing Regents and Recent Advances in Ion-Pairing-Free Analytical Strategies, Int J Mol Sci 23 (2022). https://doi.org/10.3390/ijms232415474.

[177] J. Farand, F. Gosselin, De novo sequence determination of modified oligonucleotides, Anal Chem 81 (2009) 3723–3730. https://doi.org/10.1021/ac802452p.

[178] A.O. Nwokeoji, P.M. Kilby, D.E. Portwood, M.J. Dickman, RNASwift: A rapid, versatile RNA extraction method free from phenol and chloroform, Anal Biochem 512 (2016) 36–46. https://doi.org/10.1016/j.ab.2016.08.001.

[179] H. Firczuk, S. Kannambath, J. Pahle, A. Claydon, R. Beynon, J. Duncan, H. Westerhoff, P. Mendes, J.E. McCarthy, An in vivo control map for the eukaryotic mRNA translation machinery, Mol Syst Biol 9 (2013) 1–13. https://doi.org/10.1038/msb.2012.73.

[180] G. Hölzl, H. Oberacher, S. Pitsch, A. Stutz, C.G. Huber, Analysis of biological and synthetic ribonucleic acids by liquid chromatography-mass spectrometry using monolithic capillary columns, Anal Chem 77 (2005) 673–680. https://doi.org/10.1021/ac0487395.

[181] J.B. Murray, A.K. Collier, J.R.P. Arnold, A general Purification Procedure for Chemically Synthesised Oligoribonucleotides, Anal Biochem 218 (1994) 177–184. https://doi.org/https://doi.org/10.1006/abio.1994.1157.

[182] C. Horvath, W. Melander, I. Molnar, P. Molnar, H. Knauer, H. Weg, Enhancement of Retention by Ion-Pair Formation in Liquid Chromatography with Nonpolar Stationary Phases In most cases 1 Present address, Anal Chem 49 (1977) 2295–2305. https://pubs.acs.org/sharingguidelines.

[183] C.G. Huber, P.J. Oefner, G.K. Bonn, Rapid and Accurate Sizing of DNA Fragments by Ion-Pair Chromatography on Alkylated Nonporous Poly(styrene-divinylbenzene) Particles, Anal. Chem 67 (1995) 578–585. https://pubs.acs.org/sharingguidelines.

[184] M. Gilar, Analysis and purification of synthetic oligonucleotides by reversed-phase high-performance liquid chromatography with photodiode array and mass spectrometry detection, Anal Biochem 298 (2001) 196–206. https://doi.org/10.1006/abio.2001.5386.

[185] L. Zhang, B. Majeed, L. Lagae, P. Peumans, C. Van Hoof, W. De Malsche, Ion-pair reversed-phase chromatography of short double-stranded deoxyribonucleic acid in silicon micro-pillar array columns: Retention model and applications, J Chromatogr A 1294 (2013) 1–9. https://doi.org/10.1016/j.chroma.2013.04.002.

[186] J. Stahlberg¨¨, S. Stahlberg¨¨, Retention models for ions in chromatography, J Chromatogr A 855 (1999) 3–55. https://doi.org/https://doi.org/10.1016/S0021-9673(99)00176-4.

[187] B.A. Bidlingmeyer, S.N. Deming, W.P. Price, B. Sachok, M. Petruseec, Retention mechanism for reversed-phase ion-pair liquid chromatography, J Chromatogr A 186 (1979) 419–434. https://doi.org/https://doi.org/10.1016/S0021-9673(00)95264-6.

[188] A. Bartha, J. Stihlberg, " Astra, Electrostatic retention model of reversed-phase ion-pair chromatography, J Chromatogr A 668 (1994) 255–284. https://doi.org/https://doi.org/10.1016/0021-9673(94)80116-9.

[189] T. Cecchi, Ion pairing chromatography, Crit Rev Anal Chem 38 (2008) 161–213. https://doi.org/10.1080/10408340802038882.

[190] R. Wilson, Preparation of single-stranded DNA from PCR products with streptavidin magnetic beads, Nucleic Acid Ther 21 (2011) 437–440. https://doi.org/10.1089/nat.2011.0322.

[191] T.R. Damase, A.D. Ellington, P.B. Allen, Purification of single-stranded DNA by co-polymerization with acrylamide and electrophoresis, Biotechniques 62 (2017) 275–282. https://doi.org/10.2144/000114557.

[192] M. Dickman, D.P. Hornby, Isolation of single-stranded DNA using denaturing DNA chromatography, Anal Biochem 284 (2000) 164–167. https://doi.org/10.1006/abio.2000.4669.

[193] C. Platella, C. Riccardi, D. Montesarchio, G.N. Roviello, D. Musumeci, G-quadruplex-based aptamers against protein targets in therapy and diagnostics, Biochim Biophys Acta Gen Subj 1861 (2017) 1429–1447. https://doi.org/10.1016/j.bbagen.2016.11.027.

[194] F. Pfeiffer, M. Rosenthal, J. Siegl, J. Ewers, G. Mayer, Customised nucleic acid libraries for enhanced aptamer selection and performance, Curr Opin Biotechnol 48 (2017) 111–118. https://doi.org/10.1016/j.copbio.2017.03.026.

[195] E.W. Choi, L. V. Nayak, P.J. Bates, Cancer-selective antiproliferative activity is a general property of some G-rich oligodeoxynucleotides, Nucleic Acids Res 38 (2009) 1623–1635. https://doi.org/10.1093/nar/gkp1088.

[196] O. Henegariu, N.A. Heerema, S.R. Dlouhy, G.H. Vance, P.H. Vogt, Multiplex PCR: Critical Parameters and Step-by-Step Protocol, Biotechniques 23 (1997) 504–511. https://doi.org/10.2144/97233rr01.

[197] P. Markoulatos, N. Siafakas, M. Moncany, Multiplex polymerase chain reaction: A practical approach, J Clin Lab Anal 16 (2002) 47–51. https://doi.org/10.1002/jcla.2058.

[198] J.M.S. Bartlett, D. Stirling, Methods in Molecular Biology, n.d.

[199] H. Liu, Y. Zhou, Q. Xu, S.M. Wong, Selection of DNA Aptamers for Subcellular Localization of RBSDV P10 Protein in the Midgut of Small Brown Planthoppers by Emulsion PCR-Based SELEX, Viruses 12 (2020) 1–19. https://doi.org/10.3390/v12111239.

[200] T. Wang, W. Yin, H. AlShamaileh, Y. Zhang, P.H.L. Tran, T.N.G. Nguyen, Y. Li, K. Chen, M. Sun, Y. Hou, W. Zhang, Q. Zhao, C. Chen, P.Z. Zhang, W. Duan, A Detailed Protein-SELEX Protocol Allowing Visual Assessments of Individual Steps for a High Success Rate, Hum Gene Ther Methods 30 (2019) 1–16. https://doi.org/10.1089/hgtb.2018.237.

[201] L.F. Yang, N. Kacherovsky, N. Panpradist, R. Wan, J. Liang, B. Zhang, S.J. Salipante, B.R. Lutz, S.H. Pun, Aptamer Sandwich Lateral Flow Assay (AptaFlow) for Antibody-Free SARS-CoV-2 Detection, Anal Chem 94 (2022) 7278–7285. https://doi.org/10.1021/acs.analchem.2c00554.

[202] Z. Kadlecová, K. Kalíková, E. Tesařová, M. Gilar, Phosphorothioate oligonucleotides separation in ion-pairing reversed-phase liquid chromatography: Effect of ion-pairing system, J Chromatogr A 1676 (2022) 1–8. https://doi.org/10.1016/j.chroma.2022.463201.

[203] E.D. Close, A.O. Nwokeoji, D. Milton, K. Cook, D.M. Hindocha, E.C. Hook, H. Wood, M.J. Dickman, Nucleic acid separations using superficially porous silica

particles, J Chromatogr A 1440 (2016) 135–144. https://doi.org/10.1016/j.chroma.2016.02.057.

[204] C. Vanhinsbergh, E.C. Hook, N. Oxby, M.J. Dickman, Optimization of orthogonal separations for the analysis of oligonucleotides using 2D-LC, Journal of Chromatography B (2023) 1–11. https://doi.org/10.1016/j.jchromb.2023.123812.

[205] M. Donegan, J.M. Nguyen, M. Gilar, Effect of ion-pairing reagent hydrophobicity on liquid chromatography and mass spectrometry analysis of oligonucleotides, J Chromatogr A 1666 (2022) 1–12. https://doi.org/10.1016/j.chroma.2022.462860.

[206] L. Peng, W. Jing, G. Yun-Hua, W. Li-Qing, S. Ling-Hui, F. Bo-Qiang, Analysis of fluorescent dye-labeled oligonucleotides by ion-pair reversed-phase high-performance liquid chromatography, Fenxi Huaxue/ Chinese Journal of Analytical Chemistry 37 (2009) 1722–1726. https://doi.org/10.1016/S1872-2040(08)60149-8.

[207] P.J. Oefner, C.G. Huber, F. Umlauft, G.-N. Berti, E. Stimpfl, G.K. Bonn, High-resolution liquid chromatography of fluorescent dye-labeled nucleic acids, Analytical Biochemistry 223 (1994) 39–46. https://doi.org/https://doi.org/10.1006/abio.1994.1543.

[208] K.J. Fountain, M. Gilar, Y. Budman, J.C. Gebler, Purification of dye-labeled oligonucleotides by ion-pair reversed-phase high-performance liquid chromatography, Journal of Chromatography B 783 (2003) 61–72. https://doi.org/https://doi.org/10.1016/S1570-0232(02)00490-7.

[209] C. Anacleto, R. Ouye, N. Schoenbrunner, Orthogonal ion pairing reversed phase liquid chromatography purification of oligonucleotides with bulky fluorophores, J Chromatogr A 1329 (2014) 78–82. https://doi.org/https://doi.org/10.1016/j.chroma.2013.12.072.

[210] S.M. McCarthy, M. Gilar, J. Gebler, Reversed-phase ion-pair liquid chromatography analysis and purification of small interfering RNA, Anal Biochem 390 (2009) 181–188. https://doi.org/10.1016/j.ab.2009.03.042.

[211] E.W. Choi, L. V. Nayak, P.J. Bates, Cancer-selective antiproliferative activity is a general property of some G-rich oligodeoxynucleotides, Nucleic Acids Res 38 (2009) 1623–1635. https://doi.org/10.1093/nar/gkp1088.

[212] R. Fregel, A. González, V.M. Cabrera, Improved ethanol precipitation of DNA, Electrophoresis 31 (2010) 1350–1352. https://doi.org/10.1002/elps.200900721.

[213] P. Oberacker, P. Stepper, D.M. Bond, S. Höhn, J. Focken, V. Meyer, L. Schelle, V.J. Sugrue, G.J. Jeunen, T. Moser, S.R. Hore, F. von Meyenn, K. Hipp, T.A. Hore, T.P. Jurkowski, Bio-On-Magnetic-Beads (BOMB): Open platform for high-throughput nucleic acid extraction and manipulation, PLoS Biol 17 (2019) 1–16. https://doi.org/10.1371/journal.pbio.3000107.

[214] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012, Int J Cancer 136 (2015) E359–E386. https://doi.org/10.1002/ijc.29210.

[215] P.W. Mantyh, Cancer pain and its impact on diagnosis, survival and quality of life, Nat Rev Neurosci 7 (2006) 797–809. https://doi.org/10.1038/nrn1914.

[216] S. Chakraborty, T. Rahman, The difficulties in cancer treatment., Ecancer 6 (2012) 1–5. https://doi.org/10.3332/ecancer.2012.ed16.

[217] J. Lotem, L. Sachs, Epigenetics and the plasticity of differentiation in normal and cancer stem cells, Oncogene 25 (2006) 7663–7672. https://doi.org/10.1038/sj.onc.1209816.

[218] J.B. Axelsen, J. Lotem, L. Sachs, E. Domany, Genes overexpressed in different human solid cancers exhibit different tissue-specific expression profiles, Proc Natl Acad Sci U S A 104 (2007) 13122–13127. https://doi.org/10.1073/pnas.0705824104.

[219] P. Zimmermann, D. Tomatis, M. Rosas, J. Grootjans, I. Leenaerts, G. Degeest, G. Reekmans, C. Coomans, G. David, Characterization of syntenin, a syndecan-binding PDZ protein, as a component of cell adhesion sites and microfilaments, Mol Biol Cell 12 (2001) 339–350. https://doi.org/10.1091/mbc.12.2.339.

[220] H. Boukerche, Z.Z. Su, L. Emdad, P. Baril, B. Balme, L. Thomas, A. Randolph, K. Valerie, D. Sarkar, P.B. Fisher, mda-9/Syntenin: A positive regulator of melanoma metastasis, Cancer Res 65 (2005) 10901–10911. https://doi.org/10.1158/0008-5472.CAN-05-1614.

[221] D. Sarkar, H. Boukerche, Z.Z. Su, P.B. Fisher, mda-9/syntenin: Recent insights into a novel cell signaling and metastasis-associated gene, Pharmacol Ther 104 (2004) 101–115. https://doi.org/10.1016/j.pharmthera.2004.08.004.

[222] J.J. Grootjans, P. Zimmermann, G. Reekmans, A. Smets, G. Degeest, J. Dürr, G. David, Syntenin, a PDZ protein that binds syndecan cytoplasmic domains, Proc Natl Acad Sci U S A 94 (1997) 13683–13688. https://doi.org/10.1073/pnas.94.25.13683.

[223] N. Makrilia, A. Kollias, L. Manolopoulos, K. Syrigos, Cell adhesion molecules: Role and clinical significance in cancer, Cancer Invest 27 (2009) 1023–1037. https://doi.org/10.3109/07357900902769749.

[224] A.S. Yang, P.B. Chapman, The History and Future of Chemotherapy for Melanoma, Hematol Oncol Clin North Am 23 (2009) 583–597. https://doi.org/10.1016/j.hoc.2009.03.006.

[225] L. Cui, S. Cheng, X. Liu, D. Messadi, Y. Yang, S. Hu, Syntenin-1 is a promoter and prognostic marker of head and neck squamous cell carcinoma invasion and metastasis, Oncotarget 7 (2016) 82634–82647. https://doi.org/10.18632/oncotarget.13020.

[226] H. Boukerche, Z.Z. Su, C. Prévot, D. Sarkar, P.B. Fisher, mda-9/Syntenin promotes metastasis in human melanoma cells by activating c-Src, Proc Natl Acad Sci U S A 105 (2008) 15914–15919. https://doi.org/10.1073/pnas.0808171105.

[227] A.M. Schor, S.L. Schor, Tumour angiogenesis, J Pathol 141 (1983) 385–413. https://doi.org/10.1002/path.1711410315.

[228] S.K. Das, S.K. Bhutia, B. Azab, T.P. Kegelman, L. Peachy, P.K. Santhekadur, S. Dasgupta, R. Dash, P. Dent, S. Grant, L. Emdad, M. Pellecchia, D. Sarkar, P.B. Fisher, MDA-9/Syntenin and IGFBP-2 promote angiogenesis in human

melanoma, Cancer Res 73 (2013) 844–854. https://doi.org/10.1158/0008-5472.CAN-12-1681.

[229] N. Zhang, Z. Chen, D. Liu, H. Jiang, Z.K. Zhang, A. Lu, B.T. Zhang, Y. Yu, G. Zhang, Structural biology for the molecular insight between aptamers and target proteins, Int J Mol Sci 22 (2021) 1–27. https://doi.org/10.3390/ijms22084093.

[230] L. Wang, J.Y. Lee, L. Gao, J. Yin, Y. Duan, L.A. Jimenez, G.B. Adkins, W. Ren, L. Li, J. Fang, Y. Wang, J. Song, W. Zhong, A DNA aptamer for binding and inhibition of DNA methyltransferase 1, Nucleic Acids Res 47 (2019) 11527–11537. https://doi.org/10.1093/nar/gkz1083.

[231] J. Liu, J. Qu, W. Zhou, Y. Huang, L. Jia, X. Huang, Z. Qian, J. Xia, Y. Yu, Syntenin-targeted peptide blocker inhibits progression of cancer cells, Eur J Med Chem 154 (2018) 354–366. https://doi.org/10.1016/j.ejmech.2018.05.015.

[232] R. Chen, Bacterial expression systems for recombinant protein production: E. coli and beyond, Biotechnol Adv 30 (2012) 1102–1107. https://doi.org/10.1016/j.biotechadv.2011.09.013.

[233] W.J. Park, S.H. You, H.A. Choi, Y.J. Chu, G.J. Kim, Over-expression of recombinant proteins with N-terminal His-tag via subcellular uneven distribution in Escherichia coli, Acta Biochim Biophys Sin (Shanghai) 47 (2015) 488–495. https://doi.org/10.1093/abbs/gmv036.

[234] I. Manea, M. Casian, O. Hosu-Stancioiu, N. de-los-Santos-Álvarez, M.J. Lobo-Castañón, C. Cristea, A review on magnetic beads-based SELEX technologies: Applications from small to large target molecules, Anal Chim Acta 1297 (2024) 1–25. https://doi.org/10.1016/j.aca.2024.342325.

[235] J. Lee, M. Ryu, D. Bae, H.M. Kim, S. il Eyun, J. Bae, K. Lee, Development of DNA aptamers specific for small therapeutic peptides using a modified SELEX method, Journal of Microbiology 60 (2022) 659–667. https://doi.org/10.1007/s12275-022-2235-4.

[236] T. Fellows, L. Ho, S. Flanagan, R. Fogel, D. Ojo, J. Limson, Gold nanoparticle-streptavidin conjugates for rapid and efficient screening of aptamer function in lateral flow sensors using novel CD4-binding aptamers identified through

Crossover-SELEX, Analyst 145 (2020) 5180–5193. https://doi.org/10.1039/d0an00634c.

[237] A.K. Kissmann, G. Bolotnikov, R. Li, F. Müller, H. Xing, M. Krämer, K.E. Gottschalk, J. Andersson, T. Weil, F. Rosenau, Impatient-qPCR: monitoring SELEX success during in vitro aptamer evolution, Appl Microbiol Biotechnol 108 (2024) 1–9. https://doi.org/10.1007/s00253-024-13085-7.

[238] C. Zhu, Z. Feng, H. Qin, L. Chen, M. Yan, L. Li, F. Qu, Recent progress of SELEX methods for screening nucleic acid aptamers, Talanta 266 (2024) 1–21. https://doi.org/10.1016/j.talanta.2023.124998.

[239] T. Hünniger, H. Wessels, C. Fischer, A. Paschke-Kratzin, M. Fischer, Just in time -selection: A rapid semiautomated SELEX of DNA aptamers using magnetic separation and BEAMing, Anal Chem 86 (2014) 10940–10947. https://doi.org/10.1021/ac503261b.

[240] K. Szeto, D.R. Latulippe, A. Ozer, J.M. Pagano, B.S. White, D. Shalloway, J.T. Lis, H.G. Craighead, RAPID-SELEX for RNA aptamers, PLoS One 8 (2013) 1–11. https://doi.org/10.1371/journal.pone.0082667.

[241] M. Berezovski, M. Musheev, A. Drabovich, S.N. Krylov, Non-SELEX selection of aptamers, J Am Chem Soc 128 (2006) 1410–1411. https://doi.org/10.1021/ja056943j.

[242] K. Szeto, D.R. Latulippe, A. Ozer, J.M. Pagano, B.S. White, D. Shalloway, J.T. Lis, H.G. Craighead, RAPID-SELEX for RNA aptamers, PLoS One 8 (2013) 1–11. https://doi.org/10.1371/journal.pone.0082667.

[243] S.R. Head, H. Kiyomi Komori, S.A. LaMere, T. Whisenant, F. Van Nieuwerburgh, D.R. Salomon, P. Ordoukhanian, Library construction for next-generation sequencing: Overviews and challenges, Biotechniques 56 (2014) 61–77. https://doi.org/10.2144/000114133.

[244] S.T. Kramer, P.R. Gruenke, K.K. Alam, D. Xu, D.H. Burke, FASTAptameR 2.0: A web tool for combinatorial sequence selections, Mol Ther Nucleic Acids 29 (2022) 862–870. https://doi.org/10.1016/j.omtn.2022.08.030.

[245] W.H. Thiel, P.H. Giangrande, Analyzing HT-SELEX data with the Galaxy Project tools - A web based bioinformatics platform for biomedical research, Methods 97 (2016) 3–10. https://doi.org/10.1016/j.ymeth.2015.10.008.

[246] H. Li, J. Ruan, R. Durbin, Mapping short DNA sequencing reads and calling variants using mapping quality scores, Genome Res 18 (2008) 1851–1858. https://doi.org/10.1101/gr.078212.108.

[247] T. Sakamoto, E. Ennifar, Y. Nakamura, Thermodynamic study of aptamers binding to their target proteins, Biochimie 145 (2018) 91–97. https://doi.org/10.1016/j.biochi.2017.10.010.

[248] T.B. Jensen, J.R. Henriksen, B.E. Rasmussen, L.M. Rasmussen, T.L. Andresen, J. Wengel, A. Pasternak, Thermodynamic and biological evaluation of a thrombin binding aptamer modified with several unlocked nucleic acid (UNA) monomers and a 2′-C-piperazino-UNA monomer, Bioorg Med Chem 19 (2011) 4739–4745. https://doi.org/10.1016/j.bmc.2011.06.087.

[249] T. Amato, A. Virgilio, L. Pirone, V. Vellecco, M. Bucci, E. Pedone, V. Esposito, A. Galeone, Investigating the properties of TBA variants with twin thrombin binding domains, Sci Rep 9 (2019) 1–8. https://doi.org/10.1038/s41598-019-45526-z.

[250] C. Wang, X. Du, T. Xie, H. Li, Label- and modification-free-based in situ selection of bovine serum albumin specific aptamer, J Sep Sci 42 (2019) 3571–3578. https://doi.org/10.1002/jssc.201900620.

[251] T.B. Jensen, J.R. Henriksen, B.E. Rasmussen, L.M. Rasmussen, T.L. Andresen, J. Wengel, A. Pasternak, Thermodynamic and biological evaluation of a thrombin binding aptamer modified with several unlocked nucleic acid (UNA) monomers and a 2′-C-piperazino-UNA monomer, Bioorg Med Chem 19 (2011) 4739–4745. https://doi.org/10.1016/j.bmc.2011.06.087.

[252] B. Pagano, C.A. Mattia, C. Giancola, Applications of isothermal titration calorimetry in biophysical studies of G-quadruplexes, Int J Mol Sci 10 (2009) 2935–2957. https://doi.org/10.3390/ijms10072935.

[253] N. Mohamad, A.M. Hashim, N.F. Khairil Mokhtar, M.H. Yuswan, S. Mustafa, Discovery of porcine proteins-binding DNA aptamer through SELEX and proteomics for pork authentication, Microchemical Journal 196 (2024) 1–12. https://doi.org/10.1016/j.microc.2023.109650.

[254] J. Probst, B. Weide, B. Scheel, B.J. Pichler, I. Hoerr, H.G. Rammensee, S. Pascolo, Spontaneous cellular uptake of exogenous messenger RNA in vivo is nucleic acid-specific, saturable and ion dependent, Gene Ther 14 (2007) 1175–1180. https://doi.org/10.1038/sj.gt.3302964.

[255] M. Gilar, K.J. Fountain, Y. Budman, U.D. Neue, K.R. Yardley, P.D. Rainville, R.J.R. Ii, J.C. Gebler, Ion-pair reversed-phase high-performance liquid chromatography analysis of oligonucleotides: Retention prediction, J Chromatogr A (2002) 167–182. https://doi.org/https://doi.org/10.1016/S0021-9673(02)00306-0.

[256] J. Talap, J. Zhao, M. Shen, Z. Song, H. Zhou, Y. Kang, L. Sun, L. Yu, S. Zeng, S. Cai, Recent advances in therapeutic nucleic acids and their analytical methods, J Pharm Biomed Anal 206 (2021) 1–14. https://doi.org/10.1016/j.jpba.2021.114368.

[257] M.J. Dickman, Effects of sequence and structure in the separation of nucleic acids using ion pair reverse phase liquid chromatography, J Chromatogr A 1076 (2005) 83–89. https://doi.org/10.1016/j.chroma.2005.04.018.

[258] S.J. Ross, G.R. Owen, J. Hough, A. Philips, W. Maddelein, J. Ray, P.M. Kilby, M.J. Dickman, Optimizing the production of dsRNA biocontrols in microbial systems using multiple transcriptional terminators, Biotechnol Bioeng (2024) 1–18. https://doi.org/10.1002/bit.28805.

[259] F. Ranjbarian, S. Sharma, G. Falappa, W. Taruschio, A. Chabes, A. Hofer, Isocratic HPLC analysis for the simultaneous determination of dNTPs, rNTPs and ADP in biological samples, Nucleic Acids Res 50 (2022) 1–13. https://doi.org/10.1093/nar/gkab1117.

[260] S. Gao, X. Zheng, B. Jiao, L. Wang, Post-SELEX optimization of aptamers, Anal Bioanal Chem 408 (2016) 4567–4573. https://doi.org/10.1007/s00216-016-9556-2.

[261] E. Ouellet, E.T. Lagally, K.C. Cheung, C.A. Haynes, A simple method for eliminating fixed-region interference of aptamer binding during SELEX, Biotechnol Bioeng 111 (2014) 2265–2279. https://doi.org/10.1002/bit.25294.

[262] T. Li, F. Yao, Y. An, X. Li, J. Duan, X. Da Yang, Novel complex of PD-l1 aptamer and holliday junction enhances antitumor efficacy in vivo, Molecules 26 (2021) 1–12. https://doi.org/10.3390/molecules26041067.

[263] N. Komarova, D. Barkova, A. Kuznetsov, Implementation of high-throughput sequencing (Hts) in aptamer selection technology, Int J Mol Sci 21 (2020) 1–22. https://doi.org/10.3390/ijms21228774.

[264] M. Gilar, K.J. Fountain, Y. Budman, U.D. Neue, K.R. Yardley, P.D. Rainville, R.J.R. Ii, J.C. Gebler, I on-pair reversed-phase high-performance liquid chromatography analysis of oligonucleotides: Retention prediction, 2002. www.elsevier.com/locate/chroma.