The Study of SELEX Methodology in the Generation of Aptamers for Environmental Contaminant Nonylphenol

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Biotechnology

Sophia Isabel Mullan

VICTORIA UNIVERSITY OF WELLINGTON
TE HERENGA WAKA

School of Biological Sciences

Victoria University of Wellington

Te Herenga Waka

2022

Abstract

Nonylphenol is an emerging organic contaminant released into the environment by anthropogenic activity. Nonylphenol is a mix of 211 branched-chain isomers collectively referred to as nonylphenol technical equivalents (NPTE), with the main isomer being 4-nonylphenol (4-NP). These isomers are capable of interfering with developmental and physiological signalling in humans, as well as numerous other terrestrial and aquatic species. Its presence has been documented in sediment and water bodies worldwide, including Aotearoa New Zealand. Due to the current limitations of environmental monitoring, researchers are becoming increasingly motivated to develop alternative techniques for detecting contaminants. Aptamer-based biosensors are of interest as their low cost, ease of use and quick analyses times meets the requirements for frequent monitoring. Aptamers are synthetic DNA molecules which provide the molecular-recognition element to biosensors. This is derived from their ability to fold into complex three-dimensional structures and form non-covalent bonds with their target molecule. Aptamers are developed through an *in vitro* process called systematic evolution of ligands by exponential enrichment (SELEX).

The aim of this Masters study was to generate aptamers that bound single (4-NP) or multiple (NPTE) nonylphenol isomers. To achieve this, two SELEX experiments were performed against each target molecule(s) using a modified version of traditional SELEX methodology with additional counter- and negative-selection steps. Within each SELEX experiment, stringency was increased at each round in the form of Tween 20 detergent, starting at low (0.01% v/v) or high (0.1% v/v) concentrations to determine their effects on the libraries being enriched with sequences that bind either 4-NP or NPTE. Real-time PCR was used to monitor DNA quantity across selection rounds. Following this, four libraries were sequenced using Illumina sequencing and analysed using an aptamer-specific bioinformatics pipeline. The target-binding ability of ten candidates from each library were determined using an existing characterisation technique utilising gold nanoparticles (AuNPs). Prior to characterisation, due to the nature of the target molecule, preliminary studies were conducted to assess the effects of three organic solvents on the AuNP assay.

The results of this study demonstrated that the higher stringency conditions led to reduced DNA quantity and overall fewer unique sequences in the final library. Bioinformatics data

demonstrated that the top clusters in libraries with sequences enriched for 4-NP under high stringency conditions, compared to low stringency conditions, contained a greater range of unique sequences. Overall, more clusters were found in the NPTE library following high stringency selection compared to both 4-NP libraries, though it is unclear whether this was derived from non-specific binding. Results from the AuNP characterisation assay showed no aptamer candidates bound to either 4-NP or NPTE, although an additional common characterisation method should also be used. Given that no aptamers were identified to bind any of the nonylphenyl isomers, the effects of stringency conditions and multi- or single-target SELEX on aptamer characteristics could not be assessed. However, some insights from this work including the effect of detergent concentration on library populations and the use of solvents in the AuNP assay may be used to advise future experimental design for small molecule aptamer development.

Acknowledgements

As I am nearing the end of this 5-year chapter in my life I am able to look back at my experiences with great joy, admiration, and sadness. This would not have been possible without those around me.

First and foremost, to my primary supervisor Janet, I am so lucky to have met you. Your wisdom, kind nature and faith in me guided me through easily one of the most challenging periods of my life both personally and professionally. To my secondary supervisor Jen, you were a fundamental part of my journey, thank you for your patience, honesty, and support, it always kept me grounded. Although it has only been in the scheme of things a short period of time, you both have made an immeasurable difference in my life. I know I would not have been the easiest person to mentor as we both know I am incredibly prone to injury but you always put my health first which is a reflection of your character and the great relationship we have built. Thank you for everything you have both done.

To my unofficial mentors Hamish and Odey, thanks for showing me the ropes, you taught me everything I needed to know from heavy lab work to the ins and outs of excel. You've always been there when I needed to throw idea's around and prompted me to think bigger. You helped me laugh in the face of adversity and I can't thank you both enough. You are both brilliant scientists and can't wait to see what you do in the future. Specifically, Ham thanks for teaching me how to bowl. It's all in the wrist!

Dearest Bex, you are the greatest friend throughout my years at university I could ask for. It has been a pleasure defrosting the freezers, sharing breakfast with you and spending immense amounts of time together throughout this journey. I will never forget it.

To the rest of the Pitman lab group Zara, Mel, Sarah, Matire, Ellie, Niamh and Amey, your endless kindness and encouragement is something I will always look back on fondly. I hope we can meet again in the future and work on something great together.

To my friends and family, I am so grateful for the constant love and support you've shown and continue to show me. It has done more than you know! I hope I have made you proud.

Lastly to the reader, thank you for showing interest in my work, I hope this offers new insight into the world of aptamer technology.

Two years later, one international pandemic, two severely broken bones, two surgeries, here is one thesis.

Contents

List of Figures

List of Tables

List of Abbreviations

3D	3-dimensional
4-NP	4-nonylphenol
A	Adenine
ANZECC	Australian and New Zealand Environmental and Conservation Council
AuNP	Gold nanoparticles
BDDE	Butanediol diglycidyl ether
BWB	Binding wash buffer
C	Cytosine
Ct	Cycle threshold
ddH_2O	Distilled water
DNA	Deoxyribonucleic acid
ds	Double-stranded
E2	17β -oestradiol
EPA	Environmental Protection Agency
EtOH	Ethanol
EOC	Emerging organic contaminant
EU	European Union
FDA	Food and Drug Administration
G	Guanine
GC-MS	Gas-Chromatography Mass Spectrometry
HCl	Hydrochloric acid
HNO_3	Nitric acid
HTS	High-throughput sequencing
ITC	Isothermal Titration Calorimetry
Kd	Dissociation constant
MeOH	Methanol
MB	Methylene blue
MFE	Minimum free energy
$MgCl_2$	Magnesium Chloride
μM	Micromole per litre (or $\mu mol/L$)
$MnCl_2$	Manganese Chloride
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NGS	Next generation sequencing
nM	Nanomole per litre (or $nmol/L$)
NP	Nonylphenol
$NPEO_1$	Nonylphenol monoethoxyate

NPEO$_2$	Nonylphenol diethoxylate
NPTE	Nonylphenol technical equivalents
NZ	New Zealand
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pM	Picomole per litre (or pmol/L)
ROS	Reactive oxygen species
RPM	Revolutions per minute
SELEX	Systematic evolution of ligands by exponential enrichment
SEM	Standard error of the mean
ss	Single-stranded
T	Thymine

Chapter 1.   Introduction

1.1 Fundamental Importance of Water

Water is an essential natural resource for the survival of all living organisms. Humans require water for fundamental bodily functions, food production and economic development (Popkin et al., 2010). More than 70 % of the earth's surface is covered by water, but only 2.5 % of this is fresh water, and thus consumable. Moreover, of this small fraction, the majority of freshwater is held in glaciers, polar ice caps and snowfields, leaving only 0.5 % of freshwater accessible for human use (*Clean Water Crisis Facts and Information*, 2021)(Karpińska & Kotowska, 2019). In recent years, urbanization, climate change, industrialisation, increasing populations, and human activity has caused a progressive deterioration of water quality threatening the already limited and valuable necessity of freshwater (Ahamad et al., 2020). Thus, water scarcity is a major global issue that is forecast to worsen over the next few decades. The 2018 edition of the United Nations Water Development report stated that by 2050, nearly 6 billion people will suffer from clean water scarcity (Boretti & Rosa, 2019)*.*

1.2 Water Pollution and Emerging Organic Contaminants

Water pollution is the contamination of water bodies including lakes, oceans, groundwater, and rivers through the introduction of contaminants or toxic substances Contamination can come from various routes including sea-based sources such as contamination from mariculture, tourism, and shipping, as well as land-based sources such as wastewater discharge from hospitals, domestic sewage, agriculture and chemical manufacturing. The nature and quantity of contaminants in water can determine its suitability for use through direct consumption, agricultural, industrial or recreational use (*Tackling Global Water Pollution | UNEP - UN Environment Programme*, 2020). There is a natural capacity for the environment, particularly oceans and estuaries, to absorb pollutants. However, this is limited and once capacity is reached, it may overwhelm natural restorative processes and the ability to sustain human life and ecosystems is compromised (Ahamad et al., 2020). Greater amounts of pollution correspond with greater adverse effects on species and the balance of natural ecosystems. Globally, the deterioration of fresh-water ecosystems from water pollution has led to the loss of 30 % of total biodiversity (Boretti & Rosa, 2019). Currently, approximately

730 million tonnes of untreated sewage and effluent are discharged into the environment annually (Boretti & Rosa, 2019). The global population is ~7.8 billion but is projected to reach 9.9 billion by 2050 (United Nations Department of Economic and Social Affairs, 2017).

Recently, emerging organic contaminants (EOC's) are receiving increasing recognition due to their adverse human and ecological effects at low concentrations and the ease of their ability to enter the environment (Rosenfeld & Feng, 2011). EOCs are synthetic or naturally occurring chemicals consisting of pesticides, herbicides, pharmaceuticals and endocrine-disrupting hormones (Moreau et al., 2019). They encompass both newly synthesised compounds and those that may have been introduced into the environment a long time ago, however, only recently has their significance been realised earning them their name: contaminants of emerging concern. Historically, contaminants may have been present in such low concentrations that older techniques and equipment were not capable of detecting them or contaminants were not searched for as their effects were largely unknown. In the last few decades, advances in technology and analytical instruments have greatly aided the detection of an increasing number of EOCs in different and complex environmental matrices which were previously not possible. Specifically, the development of high-resolution mass spectrometry allows non-targeted screening of contaminants in trace amounts leading to their identification. Since then, considerable effort has been made to detect, characterise, and learn more about environmentally relevant contaminants. This research will focus on one EOC named nonylphenol (NP), which is classed as an alkylphenol ethoxylate.

1.3 Alkylphenol Ethoxylates

Alkylphenol ethoxylates are a category of synthetic organic chemicals that contain two compounds: octyl phenols and nonylphenol ethoxylates (NPEO) and are formed in the refinement of petroleum. The NPEO represent 80 % of all alkylphenols produced dominating the global commercial market (Priac et al., 2017). They are primarily used to manufacture surfactants for industrial applications including paper production, pesticide formulation, detergents, leather, and textile processing to commercial household cleaning products. The ability of surfactants to form aggregates called micelles help change the properties of a chemical mixture by decreasing the surface tension between liquids contributing to solution

stability (Ahamad et al., 2020). The NPEO enter the environment after use by either being discharged directly or following waste-water treatment (see Figure 1.1).
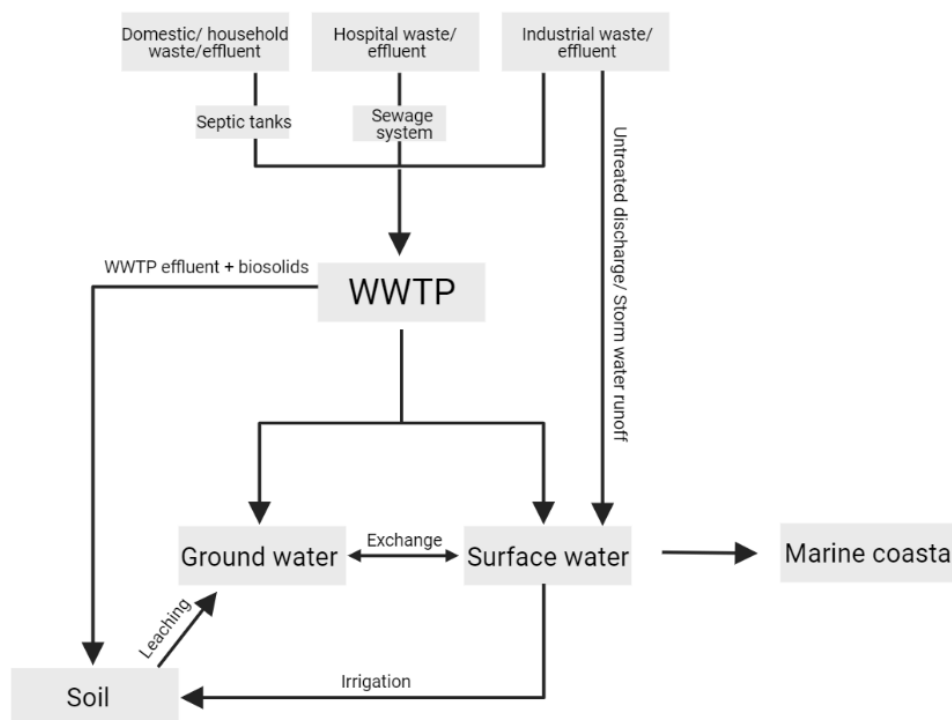
Figure 1.1: Nonylphenol ethoxylate sources, distribution routes and receiving environments. The majority of contamination comes from wastewater treatment plants and are distributed across the various water bodies.

Current technology employed in the treatment of wastewater cannot remove all contaminants or their metabolites. Depending on treatment efficiency, processes have been reported to remove between 9-94 % of alkylphenol ethoxylates in sewage. The remaining quantities of NPEOs break down in aerobic and anaerobic conditions to form nonylphenol (NP) metabolites (Chokwe et al., 2017).

1.4 Nonylphenol

Degradation of NPEO parent molecule forms shorter chain metabolites, namely nonylphenol monoethoxyate (NPEO$_1$), nonylphenol diethoxylate (NPEO$_2$) and nonylphenol (NP) as depicted in Figure 1.2 (Priac et al., 2017). The transformation from NPEO to NP is achieved

through a decarboxylation process. These are much more stable than their parent molecule, thus more persistent in the environment (Chokwe et al., 2017).
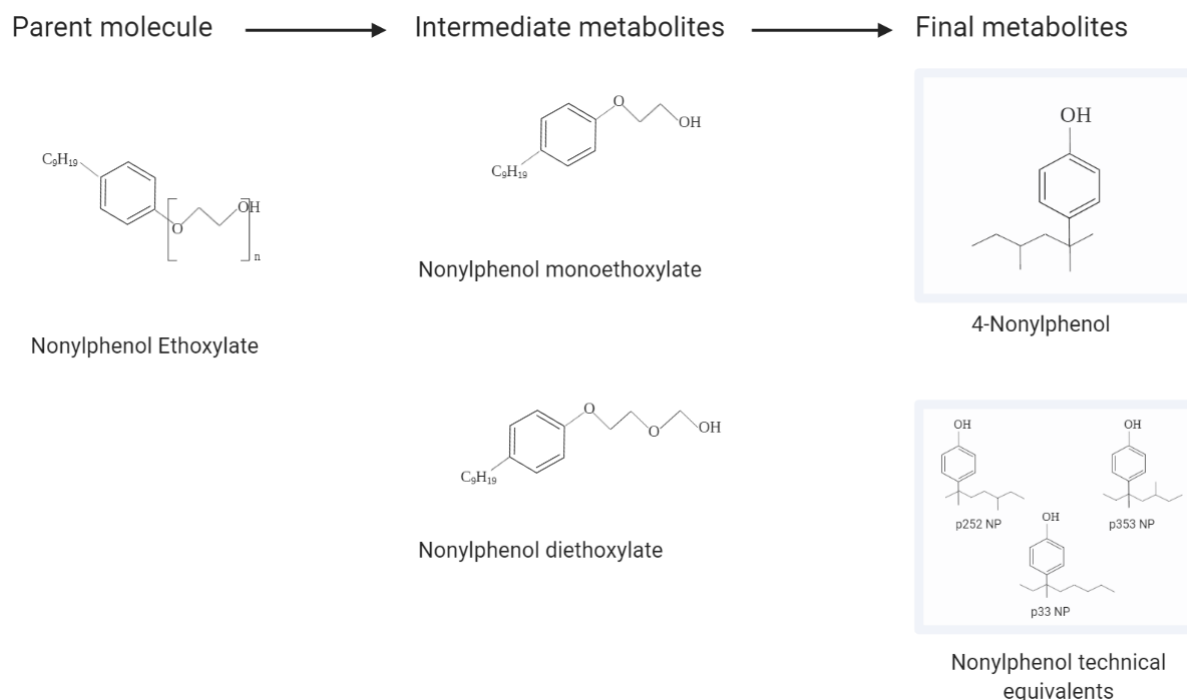
Figure 1.2: Nonylphenol ethoxylates and degradation products. *This figure was created with BioRender.com.*

The final degradation product, NP, has the chemical formula $C_{15}H_{24}O$ (*4-Nonylphenol | C15H24O - PubChem*, 2021) and contains a lipophilic linear nonylphenol chain and a phenolic group. It is not one chemical structure but a combination of 211 possible branched-chain isomers which collectively are referred to as technical nonylphenol equivalents (NPTE) (Priac et al., 2017). The most widely produced isomer is 4-n-nonylphenol which will be referred to as 4-nonylphenol or 4-NP. This group of NP compounds has characteristics of low solubility and high hydrophobicity which contributes to its bioaccumulation in humans, aquatic species and environments with a high organic content (Bai et al., 2017).

1.4.1 Nonylphenol Degradation

The main factors controlling NP degradation is initial concentration, environmental conditions, and soil parameters. Physicochemical properties of the receiving environment can influence the spatiotemporal fate and persistence of NP. Dissemination into terrestrial, aquatic, and atmospheric environments is determined by properties such as temperature, irradiation, water flow, and precipitation. Environments with reduced levels of ammonium

4

and sulphate seem to inhibit or delay the aerobic degradation of EOCs (Mao et al., 2012). Nonylphenol has an estimated half-life of 58 days within aquatic environments but this timeline can be accelerated in the presence of sunlight (Rougée et al., 2021). Although this may seem short, as NP readily binds organic matter, it can be absorbed into soil and sediment with an estimated half-life of over 60 years. Thus, soil and sediment act as natural reservoirs for this contaminant with the potential to bio-magnify (Raju et al., 2018).

1.4.2 Nonylphenol Toxicity

NP is classed as a xenobiotic compound meaning once it is absorbed by the body, it can cause direct or indirect modifications to the body's natural endocrine system causing undesirable effects (Ji et al., 2019). NP can mimic natural hormones produced by the body and disrupt normal developmental and physiological signalling. Specifically, its structural similarity to the sex hormone 17β -oestradiol (E2) facilitates its binding to natural hormone receptors to induce biological effects culminating in hormonal disorders (Kim et al., 2016). This has been demonstrated across numerous species, with evidence in humans revealing it increases the risk of autoimmune diseases, DNA damage, carcinogenesis, and cancer progression (Noorimotlagh et al., 2018). Ingestion of NP results in the production of reactive oxygen species (ROS), thus inducing oxidative stress. Free ROS radicals can cause oxidation of DNA molecules leading to DNA damage in areas including the lung, colon, kidney, liver, urinary bladder, and thyroid gland (Noorimotlagh et al., 2018). For example, studies have shown that NP may enhance prostate cancer through mechanisms of inhibiting apoptosis, accelerating cell cycle progression and facilitating the viability of LNCaP prostate cancer cells (Lee et al., 2017). It has also been associated with the development of human breast cancer and ovarian cancer (Noorimotlagh et al., 2020). Moreover, chronic exposure to NP can cause biochemical and pathological changes within the body, specifically in the reproductive tract by altering testis and epididymis size, sperm viability and quality causing infertility in males (Noorimotlagh et al., 2018).

Due to the lipophilic nature of NP, it can bioaccumulate readily in aquatic species. Extensive studies on catfish, carp-fish and zebrafish have demonstrated NP can induce DNA damage at sublethal levels and interfere with sexual differentiation, endocrine function and the growth and development of fish. A study conducted on the African catfish *Clarias gariepinus* found

exposure to sublethal concentrations of NP resulted in deleterious effects on fish physiology through increased levels of blood glucose, cholesterol, apoptosis and DNA damage (Sayed & Hamed, 2017). These environmentally-relevant concentrations of NP can inhibit growth through reducing growth hormone receptor 1 mRNA expression in rainbow trout (Hanson et al., 2014) as well as disruption of notochord morphogenesis, muscle function and the expression of neuroendocrine hormones in zebrafish (Sun et al., 2017). In terms of sexual differentiation, NP acts as a strong oestrogen agonist and can alter the transcriptional expression of sexual-differentiation genes leading to a change of sex ratio in fish (Sun et al., 2017). Laboratory studies assessing NP toxicity have also demonstrated its ability to cause harm to the endocrine system of divergent species including birds by disrupting normal signalling in the hypothalamic-pituitary-ovarian axis causing developmental abnormalities in ovaries and oviducts (Zhang et al., 2017).

Recent *in vitro* studies demonstrated that NP disrupts the innate and adaptive immune responses in several species which is facilitated by oestrogen receptors (e.g. ERα and ERβ) on immune cells (DeWitt & Patisaul, 2018). For example, the production of interleukins (e.g. IL-4, IL-6) is partially controlled by oestrogen receptors and play a key role in cellular signalling and regulation of immune and inflammatory responses (Nowak et al., 2019). Interruption of cellular signalling can lead to insufficient immune responses against foreign invading pathogens (e.g. bacteria, viruses, fungi) and the ability to prevent tumours and disease.

Overall, these studies demonstrate the ability of NP to mimic the action of natural hormones involved in homeostatic, immune and developmental processes in several species including humans, culminating in dysregulation and disease.

1.5 Water Contamination of Nonylphenol in Global Studies

Despite clear evidence of the harmful repercussions of NP in the environment, this chemical continues to be produced. In fact, approximately 154,200, 16,000 and 16,500 tons of NP is produced annually in the USA, China, and Japan, respectively. This is alarming considering 60% of NP and its metabolites disseminate into surrounding environments (Bhandari et al., 2021). Its presence in the environment has been reported in countries including Japan, Taiwan, USA, Germany and Spain (Cherniaev et al., 2016). Worryingly, the study conducted in

Spain revealed alkylphenols including NP were ubiquitously present in sediment samples across the Spanish south Atlantic coast, the Mediterranean continental shelf and subregions with concentrations ranging from 31-253 ng/g indicating medium-high risk (León et al., 2020). In Taiwan, NP was detected in marine sediment in concentrations ranging between 18 and 27,882 ng/g, which is higher than contamination found in marine sediment throughout Germany (10-153 ng/g), Japan (2.2-4,560 ng/g), and China (3.4-2,900 ng/g) (Dong et al., 2015). Furthermore, NP has been detected at concentrations ranging from 0.015-179.6 µg/L in surface water, groundwater and drinking water, and up to 1350 µg/L in untreated wastewater (Mao et al., 2012;Bhandari et al., 2021). It has also been demonstrated in air samples across the USA indicating vaporisation and atmospheric transport (Rudel et al., 2003;Ferrey et al., 2018). The results of these studies highlight the widespread distribution of concerning levels of NP in diverse environments across the world.

1.5.1 Global-Regulation & Mitigation Strategies

The harmful effects resulting from the exposure to NP has highlighted the need for mitigation strategies to monitor and reduce its dissemination into the environment. Several major regulatory bodies have recognised this and implemented restrictions on the use of this chemical and the maximum concentrations permitted in the water. The European Union was the first to place restrictions on the concentration of nonylphenol permitted in the water, with strict monitoring in place in several other countries including Canada and Japan (de Araujo et al., 2018a). Recently, the production and use of NP along with its precursors have also been banned in European Union countries. Similar to the EU, Russian legislation has prohibited the discharge of NP into marine environments but this is exclusive to the economic zone of the Russian Federation (Cherniaev et al., 2016). In the USA, the Environmental Protection Agency (EPA) recommended the maximum concentration of NP in water to be 6.6 mg/L in reflection of the 1977 Clean Water Act. In 2014, the EPA additionally proposed a significant new use rule which required the evaluation of the intended use of fifteen NP and NP polyethoxylates in manufacturing to determine its necessity and/or regulation (EPA, 2005). Japan, Europe and Canada took action to replace alkylphenols with alcohol ethoxylates, a more environmentally conscious alternative. Responses from Government and regulatory bodies reflects the severity of consequences induced by NP contamination.

However, in many countries including India, China and South America, steps to regulate, eliminate or replace NP have not been taken and it continues to be produced (Bhandari et al., 2021).

Regulation strategies that have been implemented across countries are a significant step in reducing the risk of exposure to emerging environmental contaminants, however, this is not the case across all countries. There is a lack of information regarding contaminant emission and distribution in some regions of the world, particularly in developing countries. One example is Latin America where the few studies conducted have revealed a lack of knowledge about water pollution caused by contaminants due to two major factors: a lack of legislation and the high costs of analysis. Latin America is home to many wetlands that sustain biodiversity and ecosystems, therefore, these are most at risk due to the lack of regulation of contaminants (Vargas-Berrones et al., 2020). To fully grasp the scope of contamination in the environment, analytical techniques which are sensitive, relatively easy to use and cost-effective are integral. Effective analytical techniques are the first step in comprehending the sources, distribution routes and extent of environmental contamination and determining the risk they present. This will help prioritise which contaminants should be regularly monitored and whether mitigation strategies to reduce their emission are necessary.

1.6 Water Contamination In New Zealand

New Zealand (NZ) is home to many unique and diverse species, habitats, and ecosystems. The majority (92 %) of freshwater species present in NZ are endemic to NZ. The growth, survival and lifestyles of native species such as plants, fish, birds and invertebrates, their habitats, and wider ecosystems depend on water resources.

Māori people have a deep physical and spiritual relationship with wai (water) such as rivers, streams, lakes, and springs. Wai is the essence of all life, it represents the blood of Papatūāuuku (Earth mother) and Ranginui (sky father) which share a whakapapa (genealogy) with Maori people and connects them to the natural land (*Māori Values: Protecting NZ's Rivers*, 2021). Wai is also used for cleansing and ceremonial purposes, so it is central to the sense of identity of tangata whenua (people of the land), tribal identity and the survival of native Taonga species. Water bodies have mauri meaning they carry life force and vitality. The

Māori community is becoming increasingly concerned regarding the decline in water quality in NZ as it is a valuable life source that must be protected; therefore, actions must be made to restore and increase the mauri of water (Harmsworth et al., 2016; *Wai Māori-Māori Values in Water*, 2010).

Water is a resource that is used in a range of industries such as farming, forestry, manufacturing and a range of other industrial and commercial operations. New Zealand's primary industries such as farming and agriculture require water for crop growth, pasture, and forestry (*Why Freshwater Matters | Ministry for the Environment*, 2021). Irrigation is a technique that is becoming increasingly popular in agriculture; therefore, natural water quality can directly influence crop health. New Zealand's clean and green natural landscape is also an attraction for tourism. Before the Covid-19 pandemic, the thousands of tourists every year contributed to the economy, promoting revenue and employment. In 2009, tourism contributed $21 billion to the NZ economy. Of these tourists, a survey found over 121,000, 165,000, and 600,000 people participated in water sports, swimming, and boating, respectively (*Freshwater Values • Environment Guide*, 2021). But tourists are not the only ones enjoying water activities. The lifestyle of many New Zealanders involves water-related activities such as fishing, jet boating, swimming, and kayaking. Many settlements across the country are centred around water bodies such as rivers and lakes, therefore, many of us grew up around these and are part of our identity and sense of place. For socio-economic and cultural well-being, clean water is imperative.

Relatively few assessments have been conducted in NZ regarding EOC's, but those studies that have demonstrated that the sources, concentrations, and accumulation of EOC's in the environment and biota are similar to that observed overseas. A study conducted in 2018 found 227 types of EOCs in groundwater sites across NZ (Close et al., 2021). Many of the sites sampled also contained a mixture of EOCs which raises concerns about cumulative or multiplicative effects. Another study found in 91 % of sites tested, EOC's were ranged from 0.1-11,000 ng/L demonstrating EOC's are ubiquitously present in the environment and some at environmentally concerning concentrations (Moreau et al., 2019). NP has been detected in surface water and sediment across NZ. Sediment samples from estuaries across Auckland contained NP at 145-32,000 ng/g with a median concentration of 153 ng/g (Stewart et al.,

2014). Additionally, NP equivalents were found at a mean concentration of 58 ng/L in surface water, with a high of 167 ng/L, exceeding the Australian and New Zealand Environmental and Conservation Council (ANZECC) national water quality guidelines of 100 ng/L indicating risk to species (Murray & Zak, 2020).

Our waters provide vital ecosystem services and are important for social, cultural, and environmental reasons, but also biodiscovery, therefore, we have domestic and international obligations to protect these resources. However, despite international actions, there is no existing legislation to limit EOC's including NP in the environment and they are not included in routine monitoring systems within NZ. This is largely due to the limited information concerning EOCs in NZ and the analytical capabilities of regulatory bodies. In reflection of international efforts, more studies have been undertaken but these are still done on an ad-hoc basis and there is no central coordination effort. Tracing contaminants will provide a greater depth of understanding of their sources, distribution and receiving environments. Since the abundance of EOCs are affected by spatiotemporal factors and physiochemical properties in the environment, frequent monitoring is required. This can help inform a multi-barrier approach to mitigate pollution at all steps as recommended by World Health Organisation.

1.7 Current Detection Techniques

Detection techniques are tools used to identify and quantify a molecule of interest present in a sample. Extraction and quantification methods vary depending on the type of sample being tested. Alkylphenols such as NP can distribute in a range of environments as their presence has been reported in wastewater, surface water, sediment and air (Chokwe et al., 2017). Current detection methods for environmental contaminants generally require initial isolation, pre-concentration, and cleaning steps. Due to the complexity of environmental samples, these steps attempt to remove any interferences which may impede the accurate detection of the desired target molecule (de Araujo et al., 2018).

Extraction techniques for NP in water samples include liquid-liquid extraction, solid-phase extraction, solid-phase micro-extraction and dispersive liquid-liquid microextraction (de Araujo et al., 2018). These techniques are employed to separate compounds by their physical

and chemical properties. Each technique carries advantages and disadvantages; therefore, the selection of a technique should be based on specific criteria such as cost, time, and reagent requirements. For example, solid-phase extraction is the most used technique attributable to its simultaneous pre-concentration and purification step. It is also robust and cheap, however, is more time consuming than other techniques (Chokwe et al., 2017). Additionally, solid-phase and liquid-phase extraction both require large amounts of organic solvents making them environmentally unfriendly and a less popular technique (Chokwe et al., 2017). In comparison, dispersive liquid-liquid microextraction is rapid, requires a low sample volume and low consumption of organic solvent, however, the choice of extraction solvent is highly limited (Belay, 2016). For the isolation of alkylphenols in solid samples, extraction techniques employed include microwave-assisted extraction, Soxhlet extraction, ultrasonic-assisted extraction and pressurized liquid extraction. Soxhlet extraction is the most common technique due to the complexity and cost of other techniques (Chokwe et al., 2017).

Due to the chemical properties of NP, the most common technique for detection in environmental samples is gas-chromatography coupled with mass spectrometry (GC-MS) (de Araujo et al., 2018). GC-MS is powerful in its ability to separate complex mixtures in liquid, gaseous and solid samples where NP has been found (*Gas Chromatography Mass Spectrometry (GC-MS) Information | Thermo Fisher Scientific - NZ*, 2021). Other analytical methods used for the detection of NP include high-performance liquid chromatography coupled with mass spectrometry, ultraviolet or fluorescence detection (Inoue et al., 2006).

The increasing presence of contaminants in the environment with potential adverse effects on human, animal, and plant health, has stressed the need for high-throughput analytical systems that can rapidly and cheaply detect target molecules in complex samples. Collectively these analytical methods discussed present advantages of low detection limits and the ability to test a range of environmental samples. However, overall, these lab-based techniques are costly, time-consuming and require specialist training, thus, impractical for frequent use. To monitor water quality and ensure it complies with water quality standards, detection techniques which require minimal equipment and are quick and simple to use would be far more desirable in terms of efficiency and sustainability. One potential option in the rapid and

reliable assessment of environmental contaminants over these commonly practised methods are aptamers.

1.8 Introduction to Aptamers

Aptamers are small synthetic RNA or DNA strands usually 20-80 nucleotides long, that are capable of folding into unique three-dimensional (3D) structures and act like molecular recognition elements (Zhuo et al., 2017). They bind to a variety of target molecules through hydrogen bonding, pie-stacking, ionic bonding, van der Waals force, electrostatic and hydrophobic interactions or a combination of these, with dissociation constants in the pico- to nano-molar range (Sun & Zu, 2015). Target-binding RNA was first discovered in 1990 by two independent studies. In their investigation of translational regulation by T4 bacteriophages, Tuerk and Gold found two hairpin-shaped RNA ligands from a pool of 65,536 possible combinations that were able to bind T4 polymerase (Tuerk & Gold, 1990). Within the same year, Ellington and Szostak discovered a selection of RNA ligands from a population of random sequence RNA molecules that bound organic dyes (Ellington & Szostak, 1990). These experiments demonstrated that RNA molecules can fold and create specific binding sites for molecules. Since then, target-binding RNA and DNA structures and their potential applications have been heavily researched. Ellington and Szostak termed the resultant ligands *aptamers* derived from the Latin word "Aptus" meaning to fit and the Greek word "mers" meaning particle.

Aptamers are generated *in vitro* through a highly iterative process called systematic evolution of ligands by exponential enrichment or SELEX. Since its introduction, modifications to SELEX have been made to attain aptamer products more efficiently and for divergent purposes (Sun & Zu, 2015). At present, thousands of aptamers have been generated to molecules such as proteins, amino acids, bacteria, viruses, organic molecules, and environmental contaminants (Darmostuk et al., 2014).

Currently, antibodies dominate the market for target-detecting ligands and are used in several diagnostic and therapeutic applications (Thiviyanathan & Gorenstein, 2012). Aptamers are like antibodies in the sense that they function as affinity reagents that bind target molecules to identify, pinpoint or influence their activity (Zhu & Chen, 2018). However,

unlike antibodies, aptamer production is a chemical process compared to a biological process. This means that aptamers can be reproducibly manufactured with low batch-batch variability and cost. Additionally, due to their synthetic nature, aptamers are stable and can resist changes in the environment and degradation from enzymes and toxic substances (Zhu & Chen, 2018). Aptamers are also small molecules, relative to antibodies, increasing their bioavailability and exhibiting low immunogenicity which decreases the chance of prompting an immune response if used in living systems. This makes aptamers attractive candidates for diagnostic and therapeutic purposes.

Significant progress has been made in aptamer research since its introduction and some have reached the commercial market. Pegaptanib is an anti-vascular endothelial growth factor (VEGF) aptamer that binds and blocks VEGF inhibiting the stimulation of neovascularization due to age-related macular degeneration. Antibodies have been used to target VEGF but pegaptanib was approved by the Food and Drug Administration (FDA) as a treatment for age-related macular degeneration due to its low immunogenicity (Darmostuk et al., 2014). One of the key benefits of aptamers is their flexibility to be engineered into sensors and devices in emerging technologies. This has attracted attention for several potential applications including biosensing, imaging, therapeutics, diagnostics and detection of contamination in food and environmental samples.

1.9 SELEX Parameters

To generate an aptamer to a specific target molecule, a SELEX process must be selected and modified to fit the aptamers end use. Ideal aptamer characteristics are high affinity, specificity, and selectivity. High affinity refers to strong interactions with the target(s), therefore, aptamers that exhibit high affinity are sensitive reagents and can bind targets in low abundance. A highly specific aptamer will only bind the desired target and no other molecules in a sample mixture, thus avoiding non-target interactions. However, generating an aptamer with a high degree of affinity and specificity is not always desirable. Aptamers that are selective, meaning they bind maximally to the desired target but may bind other similar molecules such as metabolites or family members of the target molecule at several-fold less affinity, may be a desirable alternative. Modifications to SELEX methodologies encourage the development of these and other useful characteristics (Kalra et al., 2018).

Conventional SELEX involves challenging a large pool of random nucleic acid sequences to bind the desired target under certain conditions such as pH, salt concentration and temperature (Tombelli et al., 2007). Sequences bind to the immobilization target and those that are unbound are eluted and discarded. Following this, bound sequences are chemically eluted from the immobilization matrix and amplified via polymerase chain reaction (PCR). Products are then purified and strand-separated to re-establish the single-stranded (ss) DNA library containing sequences with an affinity for the target required in the next selection round. This cycle is repeated (i.e. rounds) until an aptamer with a high affinity for the target is generated (Figure 1.3). This conventional method is time-consuming and has had a low success rate so each step of the SELEX process may be optimized to improve SELEX efficiency and aptamer binding properties under specific binding conditions.
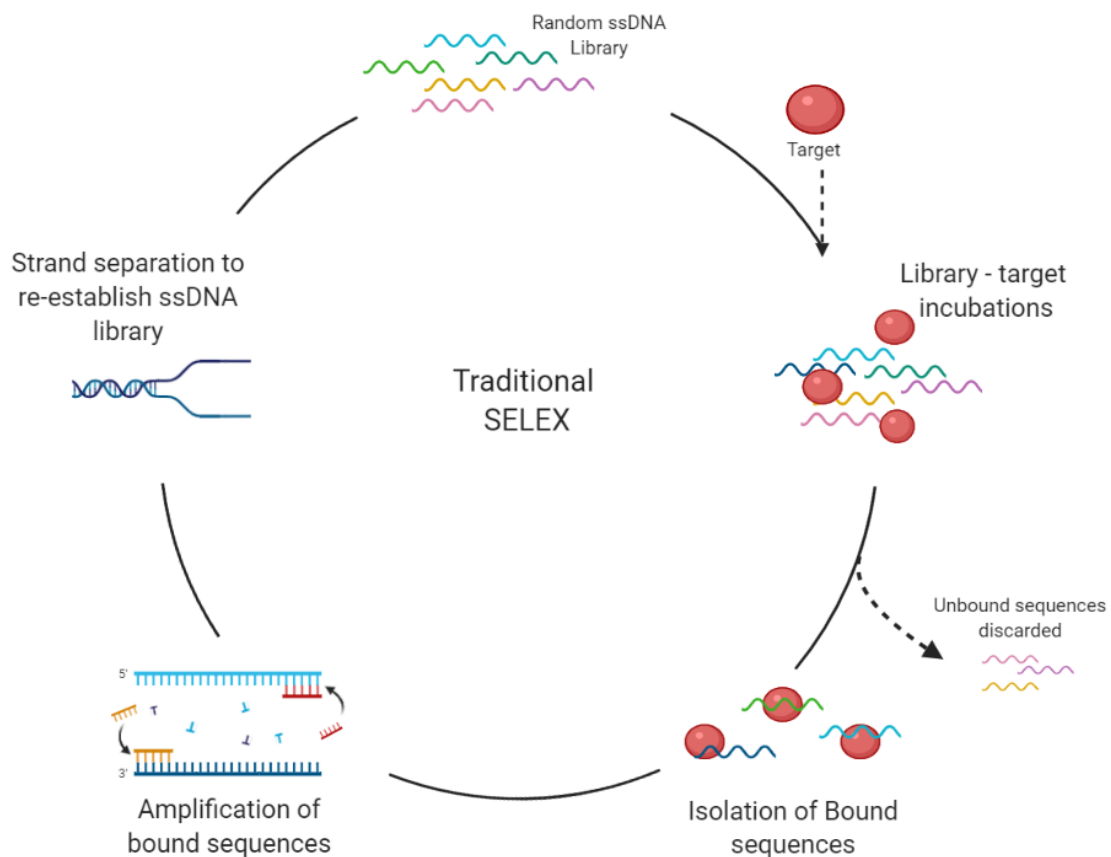

Figure 1.3: The basic protocol of SELEX: Repeated cycles of random nucleotide library incubations with target, elution of unbound sequences and retainment of bound sequences for exponential amplification, followed by strand separation to re-establish an ssDNA library for further SELEX rounds. *This figure was created with BioRender.com*.

1.9.1 Library Design

Firstly, the selection process requires a diverse library of random oligonucleotide sequences. Both DNA or RNA sequences can be used in SELEX, however, DNA sequences are favoured given their stability owing to the lack of a 2' hydroxyl group on the DNA sugar, and thus can be used for a greater number of applications (McKeague et al., 2015). The DNA SELEX process, unlike for RNA, does not require extra steps such as reverse transcription and therefore is also more cost and time-efficient (Wang et al., 2019). Alternatively, chemically modified nucleotide libraries can be used for improved stability. Modifications are made to nucleic acid components such as the backbone, base or more commonly pentose sugar (Wang et al., 2019). Important aspects of library design to consider is the core region and conserved regions of the sequences. The core region typically ranges from 20-60 random nucleotides and is flanked by two conserved regions approximately 18-21 nucleotides in length (Komarova & Kuznetsov, 2019; Blind & Blank, 2015; Sampson, 2003). These conserved regions are used as binding sites for short segments of DNA called primers to bind to during DNA amplification. The design of primers are important and should be optimised to minimize the formation of primer-dimers and hairpin structures that may lead to PCR by-products (Kohlberger & Gadermaier, 2021).

The core region determines the maximum possible sequence diversity in the initial nucleic acid library. Longer core regions can form more complex structures and binding sites which can accommodate more complex targets. Additionally, the nature of the nucleic acid sequences in the core region determines the formation of two-dimensional and three-dimensional structures (Sun & Zu, 2015). Theoretically, a longer core region would correlate with higher sequence diversity. However, the interaction between the conserved and core regions can affect the formation of structures, shift the evolution of the aptamer pool and block amplification required for further selection rounds. Shorter core regions are beneficial as there is less opportunity for primer regions to interact and they are preferable for practical applications (Komarova & Kuznetsov, 2019). Therefore, a nucleotide pool consisting of a core region 40 nucleotides in length allows for greater sequence diversity while avoiding the limitation associated with longer core regions. During the synthesis of the core region, the four nucleotide bases adenine (A), thymine (T), guanine (G) and cytosine (C) were distributed

using a skewed molar ratio of bases to remove the natural bias of G-C incorporation (Wang et al., 2019). This initial nucleotide pool will then be subjected to iterative rounds of binding, separation, and amplification until the resultant pool is enriched with sequences that exhibit high target-binding affinity (Komarova & Kuznetsov, 2019). Depending on the parameters used, this is generally achieved after 5-15 rounds.

1.9.2 Incubation of Library with Target

The first step in SELEX is to incubate the target molecule with the previously designed diverse oligonucleotide pool to promote binding. During library-target incubations, the target(s) should be pure and present in sufficient concentration to promote specific binding. In solution, sequences will form three-dimensional structures with binding pockets, and those which have an affinity to the target will bind while those that don't will remain free in solution (Sun & Zu, 2015). Conditions during library-target incubations including temperature, pH, and target concentration can be manipulated to create conditions that imitate the environment in which the aptamer is intended for use. For example, if an aptamer is desired for *in vivo* biological purposes, selections can be performed at 37 °C to mimic normal physiological conditions (McKeague et al., 2015). Similarly, if an aptamer is engineered to measure river water samples, the pH may be modified which can influence the stability, charge and structure of nucleic acids and target molecules, and therefore, the type of interaction that occurs (McKeague et al., 2015).

1.9.3 Partitioning

The success of SELEX depends on the ability to accurately separate sequences that bind to the target molecule from those that do not, to ensure the evolution of exclusively target-binding sequences (Wang et al., 2019). Following library-target incubations, sequences that have an affinity to the target remain bound and non-interacting sequences can be eluted and discarded through different physical or chemical methods (McKeague et al., 2015). The method employed for partitioning is dependent on whether the target molecule is immobilised or free in solution. To address the latter, strategies for sequence separation include nitrocellulose membrane filtration, electrophoretic mobility shift assay, and capillary electrophoresis, however, these techniques generally rely on the difference in molecular

weight and charge between nucleic acid sequences and target molecule for separation (Wang et al., 2019). This may prove difficult for small-molecule targets as there is little difference between bound and unbound oligonucleotide sequences. Thus, before library-target incubations, target molecules can be immobilised on a matrix including nitrocellulose membrane, magnetic particles, agarose-bead resin, microfluidic chips or Sepharose beads for efficient separation of bound and unbound sequences using well-designed wash buffers (Wang et al., 2019). Greater partitioning efficiency can result in fewer selection cycles for the isolation of high-affinity aptamers, which is commercially advantageous (Komarova & Kuznetsov, 2019).

One limitation with immobilisation matrices is although different conjunction chemistries exist, immobilisation onto solid supports generally requires the use of the targets functional groups and therefore, makes these groups inaccessible to the nucleotide library pool (McKeague et al., 2015). The inaccessibility of the functional group(s), especially for small target molecules, to the library may lead to a few specific aptamer sequences (Ruscito et al., 2017). This is also confounded by the fact that sequences may interact non-specifically with these matrices. This can be avoided by introducing an additional negative control step in the SELEX process which involves incubating the library with the immobilisation matrix alone. The sequences which do not have an affinity to the matrix are retained and used for further SELEX rounds (Wang et al., 2019). A study conducted by Ellington & Szostak found this negative selection technique increased the aptamer affinity by over 10-fold (Aquino-Jarquin & Toscano-Garibay, 2011). An alternative option is to perform library-target incubations in a matrix-free condition. This is beneficial as some aptamer applications require the aptamer to be free in solution (McKeague & Derosa, 2012). However, depending on other selection factors, this may not be practicable.

1.9.4 Stringency in SELEX

The number of cycles to obtain aptamer candidates within SELEX can be directly influenced by the selection strategy and the stringency applied. Increasing stringency refers to a condition or conditions that make binding interactions more difficult and results in the removal of low-affinity binders. This serves to isolate the oligonucleotide sequences that can bind under highly adverse conditions (Bhardwaj et al., 2020). Decreasing the concentration of

target available for sequences to bind, decreasing target incubation time, and increasing the number of wash steps and detergent concentrations can all result in higher stringency selection. Greater stringency conditions are generally introduced in later cycles to preserve initial library diversity since many sequences are in low frequency and can be easily lost (Ruscito & DeRosa, 2016). With each cycle, stringency is increased to raise the competition between aptamers so only high-affinity binders persist. In theory, the removal of unspecific or low-affinity binders during selection steps will decrease the competition for high-affinity binders leading to their enrichment. One study introduced stringency in a three-stage process, starting with a flexible phase, followed by a medium and then high stringency phase to boost the ssDNA library and remove low-medium affinity binders through decreasing target concentration, incubation time and increasing wash steps. The target was Lucentis, a monoclonal antibody fragment used as treatment for age-related macular degeneration. Lucentis-binding aptamers were desired to improve process analytical techniques to increase the yield of Lucentis production. A highly specific and selective aptamer for Lucentis was obtained in only ten SELEX cycles (Bhardwaj et al., 2020).

The selection and stringency method chosen is dependent on experiment specific metrics such as the target molecule. With target molecules immobilised onto a solid support, detergent can be introduced in wash steps to increase selection pressures and isolate high-affinity aptamers. Detergents such as Tween 20 can interfere with hydrogen bonds thus lowering levels of binding (Esaka et al., 1997).

To enhance the specificity of the sequence library and isolate highly specific aptamers, counter-selection steps can be performed. Counter-selection is used to remove sequences that cross-react with non-target molecules. The library is incubated with molecules that are structurally similar to the target, and those that bind are removed from the sequence library and further SELEX rounds (Kohlberger & Gadermaier, 2021). This step is particularly relevant for target molecules found in complex matrices such as environmental samples where similarly structured molecules are likely to be present.

1.9.5 Selection Pool Amplification

Following partitioning steps, sequences which exhibit an affinity for the target molecule are isolated and exponentially amplified through PCR for use in the next selection round. PCR is a revolutionary method developed by Kary Mullis in the 1980s which allows the amplification of DNA more than a billion-fold (Ozer et al., 2014). PCR consists of several cycles of denaturing, annealing and elongation. First, double-stranded (ds) DNA is heated at high temperatures to denature the hydrogen bonds holding the dsDNA together, producing ssDNA. Then the DNA is cooled which supports the stabilisation of the newly formed ssDNA allowing primers to anneal to the primer binding sites. Following this, heat-resistant enzyme *Taq polymerase* initiates elongation by adding free single nucleotide bases along the 3' end of the primer, synthesising a new and complete complementary strand (Valasek & Repa, 2005). This cycle is then repeated until the DNA has amplified to a sufficient concentration (Komarova & Kuznetsov, 2019).

There are three stages to PCR amplification: the initial lag phase, the exponential phase and the plateau phase. The exponential phase is where PCR products are exactly doubling, and the DNA is proportional to the amount of starting template. As the reaction continues there is a natural change in reaction efficiency when reagents become depleted causing the reaction to slow down and level off, this is referred to as the plateau phase (Valasek & Repa, 2005). To avoid undesirable by-products being formed during later cycles of PCR, real-time PCR can be used to monitor the amplification products after each amplification cycle. Real-time PCR works by exploiting a fluorescent probe that intercalates in dsDNA emitting fluorescence and producing a signal read-out (Valasek & Repa, 2005). This means it is possible to track the amplification results precisely and terminate the reaction at the exponential phase of amplification, where the PCR products accurately reflect the starting template.

The amplification of undesirable and non-specific sequences can arise from the diverse nature of the initial sequence pool. Because the starting pool has different structural properties (e.g. secondary and tertiary structures, sequence, and GC content), and a natural PCR bias exists, sequences that better adapt to PCR will amplify more than those less well adapted (Wang et al., 2019; Sun & Zu, 2015). This may result in a loss of unique sequences and potential aptamer candidates. The complexity of the pool can also result in inefficient amplification and the

production of by-products. Reaction conditions such as primer and reagent concentrations, the number of amplification cycles, and the temperature can be optimised to both improve amplification performance and circumvent by-products that may accumulate during the later cycles of PCR (Wang et al., 2019). To promote evolution of the sequence pool during SELEX, modifications to the PCR method can be made. The addition of chemical compounds such as magnesium chloride (MgCl$_2$,) and manganese chloride (MnCl$_2$) may also be added to enhance the existing error rate of *Taq polymerase.* The increased mutation rate can improve sequence properties (e.g. binding, stability) and result in greater library diversity and the evolution of sequences (Arnold et al., 2003).

1.9.6 Single-strand DNA Regeneration

The products that arise from PCR are in double-stranded form. However, it is ssDNA that is incubated with the target molecule for target-binding in subsequent selection rounds, and thus it is this form that is essential for the identification of potential aptamer sequences (Tolle & Mayer, 2016). Four methods are commonly used for the regeneration of ssDNA from dsDNA including enzymatic digestion of the undesired strand, asymmetric PCR, separation using streptavidin-coated magnetic beads and size separation of strands by denaturing Polyacrylamide gel electrophoresis (PAGE).

Enzymatic digestion requires the modification of the reverse primer before PCR amplification (Gao et al., 2003). The reverse primer is phosphorylated, and the *lambda exonuclease* enzyme hydrolyses and digests the phosphorylated strand in a 5' to 3' direction, generating ssDNA. The unmodified strand is protected from hydrolytic activity (Avci-Adali et al., 2009; Svobodová et al., 2012). Once the phosphorylated strands are digested, *lambda exonuclease* is deactivated via heat treatment or alkaline denaturation to prevent further digestion of unphosphorylated DNA (Gao et al., 2003).

Asymmetric PCR involves the preferential amplification of one strand over the other. This is achieved by using an unequal molar concentration of forward and reverse primers (Tan et al., 2012). The higher concentration primer is used to synthesise an excess of the desired ssDNA while the lower concentration primer becomes exhausted. Asymmetric PCR and PCR have similar reaction conditions excluding the ratio of primers used (Tan et al., 2012). Following

asymmetric PCR, the PCR products will contain both ssDNA and dsDNA, thus separation and purification of products through PAGE is necessary to obtain the desired ssDNA sequences (Svobodová et al., 2012).

Separation using streptavidin-coated magnetic beads involves two types of primers: namely 5' biotin-labelled primers and 3' unmodified primers. These primers are added to the amplification reaction resulting in one strand of the dsDNA being biotinylated (Gao et al., 2003). Following PCR, dsDNA is incubated with streptavidin-coated magnetic beads whereby the biotinylated strand is immobilised onto the beads. The complementary strand may then be mechanically or chemically eluted off (e.g. by alkaline denaturation through the addition of NaOH) (Gao et al., 2003). The undesired biotinylated strands remain bead-bound, and the eluted strands are used to create a pool of ssDNA ready for target incubation in sequential SELEX rounds.

Lastly, ssDNA can be formed using denaturing PAGE gel. The PCR is performed with a primer that contains a spacer e.g. hexaethylene glycol that acts as a terminator, followed by 20 adenine nucleotides commonly known as a polyA extension. This long primer, when incorporated into the PCR, produces products with different strand lengths which may then be separated through denaturing gel electrophoresis and the desired strands may be excised from the gel.

The ability to extract ssDNA from dsDNA products can significantly impact the success of aptamer discovery, therefore, the appropriate technique must be employed(Tolle & Mayer, 2016).

1.10 Sequencing Libraries for Candidate Aptamer Identification

Once all the SELEX rounds are complete, the pool of enriched target-binding oligonucleotides are isolated and sequenced. Sequence information is essential for the synthetic generation of potential aptamers for characterisation and *in silico* modelling using the sequence information can help deduce its 3D conformation to guide further applications.

Previously, the traditional approach to identify aptamer candidates involved cloning sequences into bacterial vectors such as plasmids, which are then transformed into

competent bacterial cells and grown (Alberts et al., 2002). Subsequently, a small number of colonies were picked, and Sanger sequencing was used to determine the nucleotide sequence present in those cells. Sanger sequencing works using chain-termination PCR whereby fluorescent chain-terminating nucleotides are incorporated into DNA replication resulting in several copies of the sequence of interest terminated at random lengths. These products are then separated by gel electrophoresis and either manually or automatically analysed to determine the DNA sequence (*Sanger Sequencing Steps & Method*, 2021). One limitation to this method is that as sequences are isolated from a small number of colonies, this method provides only minimal information. Therefore, there is a bias towards the most frequent clones, but not necessarily the unique binding sequences (Blind & Blank, 2015). Sanger-sequencing is low-throughput and generally, less than 100 clones are sequenced at a time. Additionally, some sequences may be lost during bacterial vector cloning, therefore, this method may not provide a thorough representation of possible aptamer candidates following SELEX.

Since its introduction in the 1990's, high-throughput sequencing (HTS) using next-generation sequencing (NGS) techniques has revolutionised many disciplines in biology (Wuyts & Segata, 2019). NGS is a technique that produces millions of sequence reads per run. This allows comprehensive sequence coverage for greater read accuracy, higher sensitivity for low-frequency sequences and the ability to sequence hundreds of gene regions in parallel. Recently HTS has been used in conjunction with aptamer science as the data obtained from this method provides an informative basis for accurate aptamer identification. HTS offers more information that can't be accessed from low-throughput sequencing which is useful when investigating SELEX methodology (Komarova et al., 2020). There are several different strategies for NGS including Illumina, 454 Roche and Ion Torrent platforms. All strategies employ a sequencing by synthesis technique where target sequences act as a template for the synthesis of a complementary sequence. Prior to HTS, sequence libraries are transformed into dsDNA with flanking adaptor regions that contains sequence-specific data. During complementary strand synthesis, each base that is incorporated produces a signal which is identified by a sequencing device and output for further analysis.

A considerable limitation to NGS is that because it provides such a large raw data set, it can prove difficult to organise this large quantity of information, and computational analyses and bioinformatics are required to deduce sequence reads (Behjati & Tarpey, 2013). Thankfully, new bioinformatics approaches have been developed in response to this. Analyses of aptamer specific HTS data generally includes data pre-processing, library clustering, enrichment analysis and motif search. Aptamer specific programs including The Galaxy Project, Aptaplex, FASTAptamer, AptaSUITE, APTANI and MEMESUITE specifically address one or all HTS data processing requirements for the identification of aptamers (Blind & Blank, 2015 ; Komarova et al., 2020). However, many of these programs still require a foundational understanding of the command line.

1.11 Aptamer Characterisation Methods

Following the completion of SELEX, characterisation methods are employed to identify target-binding oligonucleotides and elucidate the strength and type of bonding that occurs between this binding complex.

Isothermal Titration Calorimetry (ITC) is among the number of technologies used as a standard practice to characterise candidate aptamers. ITC is a powerful, robust, sensitive and label-free method to study biochemical or molecular interactions (Daems et al., 2021). Aptamer-target molecule interactions depend on the nature of the aptamer and the characteristics of the target molecule. Aptamers can interact and bind to a target molecule via hydrogen bonding, van der Waals, electrostatic interactions, pie stacking or a combination of multiple forces (Cai et al., 2018). ITC is extremely sensitive and can measure these binding interactions with a K_d in the micromolar to nanomolar range. A typical ITC instrument contains two cells kept at a constant temperature; one is a reference cell and the other is loaded with the sample. An aptamer or target molecule is incrementally titrated by injection at short intervals into the cell containing the target molecule or aptamer, respectively. The calorimeter measures to one-millionth of a degree Celsius, the heat exchange (released or absorbed) by the target molecule-aptamer interactions, in comparison to the reference cell (Sakamoto et al., 2018). Thermodynamic parameters (enthalpy change, entropy change, Gibbs free energy change, dissociation constants and stoichiometry) associated with molecular interactions may then be calculated. Understanding these parameters can aid in

determining target molecule-aptamer binding affinities and provide information to inform potential end use applications.

Functionalised nanoparticles that utilise aptamers can be used as label-free colorimetric sensors to detect target analytes. Since its introduction in 1996 (Zhang & Liu, 2021) this optical assay has been modified to detect a range of target molecules including, but not limited to, small molecules, proteins, and metal ions (Saha et al., 2012). Generally, nanoparticles are of a size that span between 1 – 100 nm and their physical properties can be modified to control their shape, composition, and surface chemistry (Wang & Yu, 2013). Surface-modified nanoparticles such as gold nanoparticles (AuNPs) can exhibit specific, and sensitive binding properties. Their unique optical properties are a result of the surface plasmon resonance effect which causes the nanoparticles to exhibit a deep red colour. This means low concentrations of nanoparticles in the nm to pm range can be observed by the naked eye. Oligonucleotides, such as an aptamer, that are pre-absorbed onto the AuNP surface creates a protective effect against AuNP aggregation. Aggregation of nanoparticles causes a coupling of surface plasmon resonance resulting in a visual colour change from red to blue and is promoted by the addition of salt. Thus, when AuNPs that have an aptamer absorbed onto their surface and are placed in the presence of the aptamers target molecule, the aptamer undergoes structural change desorbing from the AuNP surface to bind the target molecule. Following the addition of salt, the now unprotected AuNP undergo salt-induced aggregation as illustrated in Figure 1.4. Therefore, interactions between the aptamer and its target molecule(s) can be quantified using colour absorbance from non-aggregated (red) to aggregated (blue) AuNPs. This allows a sensitive but simple and quantitative method for determining aptamer-target binding (Zhang & Liu, 2021).

1.12 Suitable Biosensing Platforms for Aptamers

When an aptamer binds to its target molecule, it can undergo significant structural rearrangements which can be transduced into a detectable signal. Due to the synthetic nature of aptamers, they can be coupled to carriers, surfaces, and nanoparticles (Song et al., 2008). The design of these sensors relies heavily on the type of recognition mode between each aptamer and target molecule pair. There are several suitable sensing platforms in which aptamer-target molecule interactions can be transduced into a signal which may be based on

electrochemical, optical, or mass-sensitive properties. These platforms have broad applications including targeted therapeutics, biomedical diagnostics, environmental monitoring, food safety analysis, personalised healthcare, and point of care detection of biomarkers for chronic and emergent diseases (Song et al., 2008; Villalonga et al., 2020).
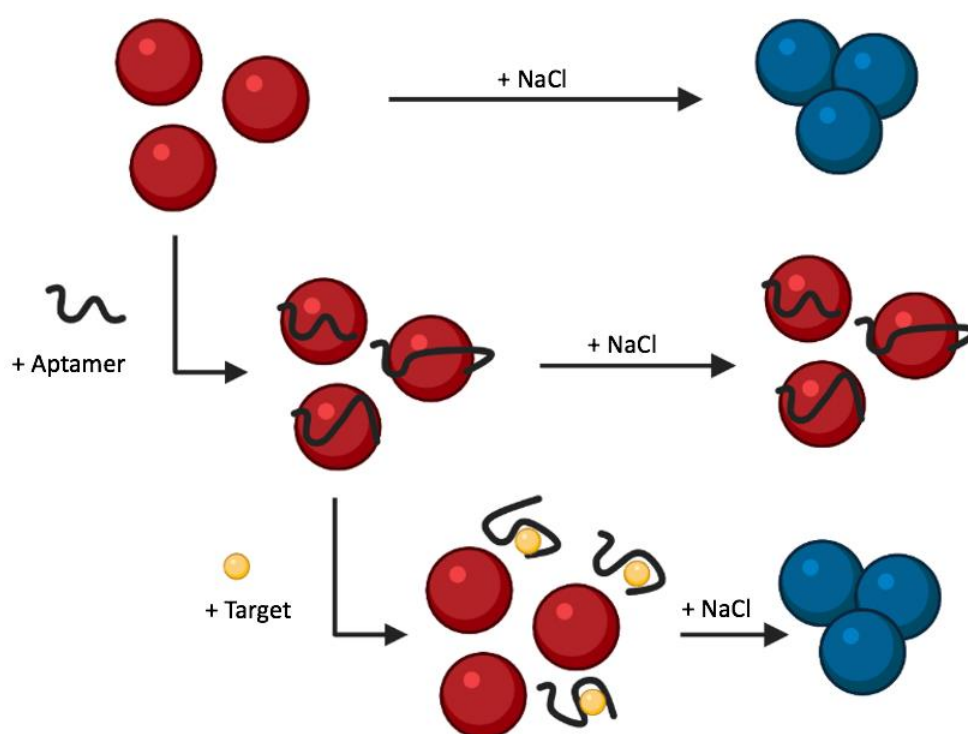

Figure 1.4: AuNP aggregation scheme. A) AuNP in the presence of salt aggregate accompanied by a colour change from red to blue. B) Aptamer-coated AuNPs are protected from salt induced aggregation. C) In the presence of target, target-binding aptamers dissociate exposing AuNP surface to salt leading to coupling of surface plasmon resonance observed by a visual colour change from red to blue. Modified figure from William Odey, VUW, unpublished. *Figure was created with BioRender.com*.

One application of aptamers is optical bioassays. Of these, fluorescence and colourimetric assays are commonly used techniques (Liu et al., 2020). Aptamers that form hairpin structures can be end-labelled with a fluorescent dye that is in close proximity to a quenching molecule. The binding of a target molecule to the aptamer will cause a conformational change separating the fluorophore and the quencher, resulting in a fluorescent signal (Song et al., 2008). Other frequently used formats include a fluorophore-labelled aptamer coupled to a complementary sequence, labelled with a quencher. The presence of the target molecule, to which the aptamer has a higher binding affinity to, will cause the departure of the aptamer

from the complementary sequence and hence the quencher, resulting in a measurable fluorescence change.

Electrochemical aptasensors work by immobilising an aptamer onto a conduction support or electrode (Song et al., 2008). Upon target molecule-binding, the aptamer will undergo a conformational transition resulting in a change in electric potential or current (Liu et al., 2020). This target molecule-induced conformational change will either lead to an increase or decrease in electrochemical signal due to the change in electron-transfer ability between an electro-active label (e.g. methylene blue MB) and the electrode surface (Villalonga et al., 2020). Recently, label-free electrochemical impedance spectroscopy has received attention due to its high sensitivity and simplicity and it is being investigated for the detection of low-molecular-weight targets (Nguyen et al., 2017; Li et al., 2007). Label-free electrochemical impedance spectroscopy is where the target molecule interacts with an immobilised probe and changes the electrical properties of the surface for impedance sensing. This removes the requirement for labelling the aptamer with a probe which could change the molecule's binding properties (Villalonga et al., 2020; Daniels & Pourmand, 2007). Several biosensors have been developed for environmental contaminants including heavy metals, water-borne bacterial pathogens, and small-molecule agricultural toxins (McConnell et al., 2020; Mishra et al., 2018). Electrochemical and optic aptasensors have been developed for similarly-structured molecules to NP including bisphenol-A (BPA) and E2 and used in their detection in water samples (Yang et al., 2019; Yu et al., 2019). Recently an aptamer to 4-nonylphenol has been developed using reduced-graphene oxide (rGO) SELEX and tested in a gold nanoparticle-based sensing system (Kim et al., 2019). However, 4-NP was the sole target and next-generation sequencing was not used to analyse sequence libraries. This work aims to investigate alternative strategies to improve the selection and development of an aptamer to both 4-NP and NPTE.

The advantages of aptamer-based biosensors over traditional analytical methods are great. Aptamers have characteristics of high stability and reproducibility, are amenable to simple chemical modification, have enhanced target affinity and specificity, and have limitless selection capacity for a diverse range of targets (Villalonga et al., 2020; McConnell et al., 2020). Due to their synthetic nature, they can be generated to detect toxic molecules, a key

26

advantage they have over antibodies. They can be applied to several sensing platforms which are cost-effective, portable, and user-friendly, thus making them ideal for environmental monitoring (Moreno, 2015). Aptamer-based detection methods can overcome current monitoring limitations and allow for regular surveillance of environmental contaminants. Understanding the sources, distribution routes and fate of contaminants in various environmental matrices will inform mitigation strategies required to reduce pollution at all stages to protect our precious natural recourses and enhance socio-economic, and cultural well-being in Aotearoa New Zealand and internationally.

1.13 Masters Objectives

The overall objective for my Master's Part 2 project was to generate an aptamer(s) that successfully binds to NP. A future aspiration is that this candidate aptamer may then be engineered into a biosensing platform that can overcome the limitations of current analytical techniques. This would aid in the regular monitoring of environmental samples to gauge levels of contamination in Aotearoa New Zealand, owing to the lack of information that exists on the topic.

Hypothesis

This work will address two hypotheses. My first hypothesis is that performing SELEX using a single target (4-NP), compared to multiple targets (nonylphenol technical equivalents, NPTE), will be more effective for the generation of an aptamer more specific to 4-NP. My second hypothesis is that the addition of increasing detergent concentrations in the wash steps of SELEX will increase stringency conditions resulting in fewer nucleotide sequences but of these, they will exhibit a higher affinity towards 4-NP.

Aims

My hypotheses will be addressed by achieving three specific aims:

1) Compare the resultant enriched libraries from two independent SELEX experiments using a singular target molecule (4-NP) and a solution containing multiple branched-chain variants (NPTE), and;

2) Compare the resultant enriched libraries from two independent SELEX strategies of low and high stringency conditions through the addition of detergent in the wash steps, and;

3) Identify and characterise potential aptamer sequences using a AuNP colorimetric assay to determine their binding affinity and suitability for adaptation to a biosensing platform.

Chapter 2. Materials and Methods

2.1 Target Molecules

Two molecules were used as the targets in independent SELEX experiments; namely 4-nonylphenol (4-NP; Sigma Aldrich), and a technical grade solution containing a mixture of ring and chain nonylphenol isomers named nonylphenol equivalents (NPTE; Sigma Aldrich). The certificate of analysis for each are given in Appendices A.4 and 4.5.

2.2 Buffer Preparation

A 10x binding wash buffer (BWB) stock was made by combining 1000 mM NaCl, 50 mM KCl, 20 mM MgCl$_2$, 10 mM CaCl$_2$ and 200 mM Tris-HCl in 1 litre of double distilled water (ddH$_2$O) and the solution was adjusted to a pH of 7.5 . The buffer was then sterilised by autoclaving at 121˚C for 20 minutes and stored at room temperature. This stock was then diluted 1:10 in ddH$_2$O to create a 1x working concentration for use in various SELEX processes.

2.3 Preparation of Tween 20 Concentrations

Solutions containing different concentrations of Tween 20 (Thermo Fisher Scientific) were prepared in 1 x BWB stock such that the final concentrations of Tween 20 were 0.01, 0.02, 0.03, 0.04, 0.05 % (v/v) and 0.1, 0.2, 0.3, 0.4, 0.5 % (v/v). These combinations represent the low stringency and high stringency conditions, respectively.

2.4 Synthesising Target-conjugated Sepharose Beads

To isolate the oligonucleotides that have an affinity to the desired target molecule, the target molecule was first conjugated to an affinity matrix. The conjugation method is displayed schematically in Figure 2.1. All bead preparations were performed in a fume hood.
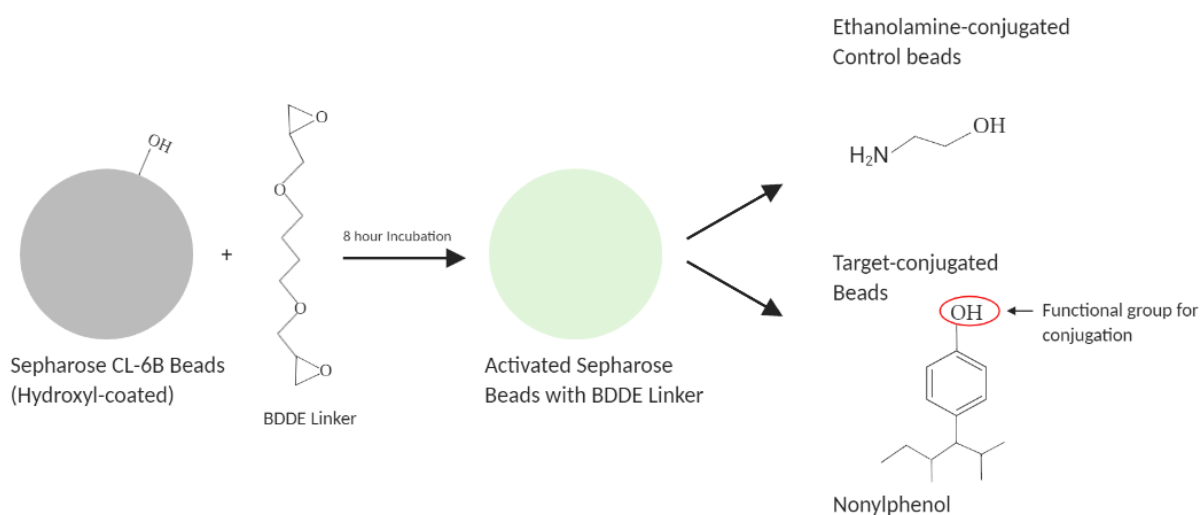
Figure 2.1: Scheme of CL-6B Sepharose bead conjugation to target(s) or control. *This figure was created with BioRender.com.*

To achieve this, 8 mL of CL-6B Sepharose beads (Sigma Aldrich) were pipetted into a sintered glass funnel. Sepharose beads were washed first with 500 mL of 0.5 M NaCl, followed by 1000 mL of ddH$_2$O and a final wash of 200 mL of 1 M NaOH. Each solution was removed before the addition of the next. The beads were then dried for 10 minutes in a vacuum and their weight was measured and recorded. To conjugate a target molecule onto the CL-6B Sepharose bead affinity matrix, a linker molecule must first be attached to the Sepharose beads. This was achieved through incubating the dried beads with 15 mL of an activating mixture comprised of 8 mL of 1 M NaOH, 2.4 mL of 1,4-butanediol diglycidyl ether (BDDE) (Sigma Aldrich) and 4.6 mL of ddH$_2$O, and then incubated for 8 hours on a rotating wheel at 300 rpm.

Succeeding the incubation, the bead and activating mixture were washed in a sintered glass funnel with 500 mL of 1 M NaOH followed by 2000 mL of ddH$_2$O. The beads were then dried again for 10 minutes in a vacuum and divided into two parts labelled control beads and target beads. For the target beads, a 1.33 mM solution of NPTE (Sigma Aldrich) suspended in 15 mL of 1 M NaOH was added to the dried beads. The solution was then incubated at room temperature on a rotating wheel at 300 rpm for 8 hours. The purpose of this was to conjugate the target to the BDDE linker. For the control beads, a 15 mL solution of 1 M ethanolamine was added and incubated in the same manner as the target beads to block any reactive groups on the beads and decrease non-specific binding.

30

Following incubation, the control beads were washed with 2000 mL of ddH$_2$O in a sintered glass funnel and dried for 10 minutes in a vacuum. Bead weight was recorded and finally, the beads were stored at 4°C until further use. The target beads were washed with 500 mL of 1 M NaOH then 1000 mL of ddH$_2$O. They were then subjected to a blocking step consisting of an 8-hour incubation with a 15 mL solution of 1 M ethanolamine on a rotating wheel at 300rpm. Lastly, the target beads were washed with 2000 mL of ddH$_2$O and dried for 10 minutes under a vacuum and stored at 4˚C until further use. This protocol was repeated with 1.33 mM of 4-NP. A 1.33 mM solution was made by dissolving 4-NP in 1 mL of 100 % acetone, then 14 mL of NaOH to make a total volume of 15 mL, before adding to the linker conjugated beads.

2.5 Target-bead Conjugation Assessment

To determine whether the target molecules successfully conjugated to the Sepharose beads via the linker molecule, a conjugation assessment was performed using a Spectrostar Nano microplate reader (BMG LABTECH). A measure of 200 mg of target-conjugated and control beads were resuspended in 1 mL of 20 % (v/v) ethanol each and 5 µL of both bead suspensions were loaded onto an L-vis plate in duplicate. The samples were measured at 220-500 wavelength at 5 nm resolution. The presence or absence of an absorption peak at 225 and 280 nm indicates the presence or absence of 4-NP and NPTE conjugated to the target and control beads.

2.6 Oligonucleotide Library Incubation with Target-conjugated Sepharose Beads

To isolate sequences that bind our target molecules, first a random diverse nucleotide library was incubated with the target-conjugated Sepharose beads and the sequences which had an affinity to the target molecule(s) bound, while those that do not remained free in solution. Two grams of target conjugated beads were resuspended in 60 µL of 1 X BWB and incubated with 8 µL of 100 mM N40 library (Integrated DNA Technologies) at 4˚C on a rotating wheel at 1.5 x speed for 1 hr. This was repeated for each SELEX round where following the previous round, the evolved library was incubated with 2 g of target-conjugated beads resuspended in 60 µL of 1 x BWB for consistency.

2.7 Washing of Sepharose Beads

The sequences that had an affinity for the target molecule(s) were isolated via wash steps from those that did not exhibit affinity. In SELEX Rounds 1 and 2, there was one pool of sequences to ensure conditions were equal preceding the introduction of different stringency conditions. Directly preceding partitioning steps in SELEX Round 3, each evolved library was split evenly into two separate elution columns. Libraries were then washed with increasing concentrations of Tween 20 detergent in each SELEX round representing either a low (0.01, 0.02, 0.03 and 0.04 %) or high (0.1, 0.2 0.3 and 0.4 %) stringency wash. Solutions were added and passed through the column by gravity. See Table 2.1 for the details of the wash conditions for each SELEX round.

Table 2.1: Volume and detergent concentration in wash steps across SELEX rounds

SELEX ROUND	LOW STRINGENCY CONDITION	HIGH STRINGENCY CONDITION
1	15 mL 1 x BWB	
2	15 mL 1 x BWB	
3	5 mL 1 x BWB 5 mL 0.01 % (v/v) Tween 20 5 mL 1 x BWB	5 mL 1 x BWB 5 mL 0.1 % (v/v) Tween 20 5 mL 1 x BWB
4	5 mL 1 x BWB 5 mL 0.02 % (v/v) Tween 20 5 mL 1 x BWB	5 mL 1 x BWB 5 mL 0.2 % (v/v) Tween 20 5 mL 1 x BWB
5	5 mL 1 x BWB 5 mL 0.03 % (v/v) Tween 20 5 mL 1 x BWB	5 mL 1 x BWB 5 mL 0.3 % (v/v) Tween 20 5 mL 1 x BWB
6	5 mL 1 x BWB 5 mL 0.04 % (v/v) Tween 20 5 mL 1 x BWB	5 mL 1 x BWB 5 mL 0.4 % (v/v) Tween 20 5 mL 1 x BWB
7	5 mL 1 x BWB 5 mL 0.05 % (v/v) Tween 20 5 mL 1 x BWB	5 mL 1 x BWB 5 mL 0.5 % (v/v) Tween 20 5 mL 1 x BWB

Following the wash steps, the Sepharose beads in the columns were resuspended in 100 μL of 1 x BWB and pipetted off the top of the column into separate 1.5 mL Eppendorf tubes. This step was repeated two more times to ensure all beads were removed from the column making a total volume of 300 μl. The library-bound beads were then pelleted by centrifugation at 13,000 g for 2 minutes and the supernatant was removed and discarded. The pelleted beads were then resuspended in 100 μL of 1 x BWB and subjected to 95˚C on a heat block for 5 minutes to dissociate the nucleotide sequences bound to the target. Subsequently, naked bead and library suspension were centrifuged at 13,000 g for 3 minutes and the supernatant containing the library was collected and stored at -20˚C until further use.

2.8 Polymerase Chain Reaction

The supernatant collected from the wash steps contained oligonucleotides that exhibited an affinity for the target molecule(s). These were then amplified exponentially through Polymerase Chain Reaction (PCR). The forward and reverse primers used in this procedure are complementary to the conserved regions of the oligonucleotide library and are shown in Table 2.2. The preparation of the 20 μL PCR master mix for each oligonucleotide library template consisted of 15.65 μL of Ultra-Pure (U/P) H_2O (Thermo Fisher Scientific), 2.5 μL of 10 x PCR buffer (Qiagen Taq Polymerase Kit), 0.5 μL of 10 mM dNTPs (Thermo Fisher Scientific), 0.55 μL of 10 μM forward primer, 0.55 μL of 10 μM biotin-labelled reverse primer (Integrated DNA Technologies), 0.1 μL of 30 x SYBR green (Life Technologies) and 0.15 μL of 0.75 U/μL Taq DNA polymerase (Biolabs, New England). Following this, 5 μL of the oligonucleotide library template was added making a total reaction volume of 25 μL. Ten 25 μL reaction tubes were set up for each low stringency and high stringency library strategy. Positive and negative controls were included using the same method as above, however, substituting 5 μL of template with 5 μL of 10 nm N40 library or U/P H_2O, respectively. SYBR green was used in conjugation with Real-Time PCR to visualise the production of the amplification DNA products following each amplification round. This enabled the avoidance of over-amplification and the formation of spurious products, characteristic of later PCR cycles when reagents become scarce.

To introduce random mutations into the nucleotide library which may be beneficial, mutagenesis was performed in SELEX Rounds 3 and 6. This was achieved by increasing the

natural error rate of Taq polymerase through the addition of 2.65 μL of Mg^{2+}. A Corbett RotorGene 6000 instrument (Qiagen) was used to perform the amplification cycles which went as follows: an initial denaturation step at 95˚C for 5 minutes, then repeated cycles (between 5-25 depending on the DNA amplicon production) of denaturation for 40 seconds at 94˚C, annealing for 30 seconds at 53˚C and lastly an extension step for 15 seconds at 72˚C. The PCR products were then subjected to a final extension step for 60 seconds at 72˚C, prior to a melt curve analysis. All PCR products were stored at -20˚C until required.

Table 2.2: Nucleotide sequences of the starting library, forward and reverse primers. Note the primer regions in the library span a 40mer random nucleotide region

Name	Sequence
N40 Starting Library	TAACCACATAACCGCAAGANNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNNNNNNNN NNNNTATTGTGCTACTCTCCTCGT
Forward primer	5'TAA CCA CAT AAC CGC AAG A-3'
Reverse primer	5'-ACG AGG AGA GTA GCA CAA TA-3'
Biotin-labelled reverse primer	5'/5Biosg/ ACG AGG AGA GTA GCA CAA TA-3'

The N40 starting library and primers listed in Table 2.2. (Integrated DNA Technologies) were provided as a lyophilised powder. Reagents were spun down at 14,100 *g* for 1 minute, then hydrated to 100 μM with U/P H$_2$O and vortexed. The library and primers were aliquoted and stored at -20˚C until further use.

2.9 Concentration of Amplified PCR Products

PCR products were concentrated to reduce the volume of liquid present whilst retain the DNA content. Amplification products in the 25 μL reaction volume were pipetted into 1.5 mL Eppendorf tubes. Then, x mL (5 x volume of gel weight; e.g. 5 mL for 1 g of excised gel) of Qiagen phosphate buffer (PB Buffer) was added to the amplified products and vortexed. This

helps to precipitate the dsDNA onto the filter membrane. Following this, an aliquot of 700 µL was then pipetted into a Qiagen MinElute spin column and eluted by centrifugation at 9,000 *g* for 60 seconds. The flow-through was put through the column again and centrifuged for a further 60 seconds after which, the flow-through was discarded. These steps were repeated until the total volume was used. To ensure the removal of PCR reagents, a final 700 µL of Qiagen PE Buffer was pipetted into the MinElute and centrifuged at 13,300 *g* for 60 seconds. The flow-through was put through the column and centrifuged once more. Following this, the empty MinElute spin column was centrifuged at 13,300 *g* for 60 seconds to remove all liquid and the column was transferred to a fresh 1.5 mL Eppendorf tube. Then, 30 µL of U/P H_2O was loaded directly onto the column membrane and incubated for 1 minute, then spun down at 13,300 *g* for 60 seconds. This step was repeated with 20 µL of U/P H_2O to make up a final volume of 50 µL. The concentrated dsDNA was then stored at -20°C until further use.

2.10 Agarose Gel Electrophoresis

Amplified PCR products were subjected to electrophoresis to enable their purification by separation by size and to estimate their amount by comparing band densities. To achieve this, 25 µL of concentrated dsDNA from each condition were combined with 2 µL of 100 x SYBR green and 6 µL of Thermo Scientific 6 x Tritrack DNA loading dye and then were incubated for approximately 8 minutes at room temperature. A DNA ladder was prepared by incubating 1 µL of Gene Ruler low range DNA ladder (Thermo Fisher Scientific) with 6 µL U/P H_2O, 2 µL of 100 x SYBR green, 6 µL of 6 x Tritrack DNA loading dye (Thermo Fisher Scientific) at room temperature for 8 minutes. Meanwhile, a 4 % (w/v) agarose gel was prepared by melting 4 g of GTG agarose (LONZA Bioscience) in 100 mL of 1 x TAE buffer (TAE buffer recipe; 0.4 M tris acetate: 0.01 M EDTA: pH of 8.3) in a 250 mL Erlenmeyer flask and microwaved for 30-second intervals, until the mixture did not have visible bubbles and the agarose granules had melted. The mixture was then poured into a gel tray containing a 15 well comb insert. Once the gel was set, the comb was removed, and the gel was submerged in an electrophoresis tank containing 1 x TAE buffer. The dsDNA sample preparations were then loaded into the wells and the gel underwent electrophoresis at 90 volts for 90 minutes. An image of the gel was recorded using an Omega Lum G imaging system (Aplegen) and the gel was placed on an Ultraviolet Transilluminator (Bio-Rad) to excise the desired bands of DNA

using a sterile razor blade. The excised bands were then weighed and stored in a 15 mL falcon tube at -20˚C until further use.

2.11 DNA Extraction from Agarose Gel

Extraction techniques were used to isolate dsDNA from the agarose gel. A 3 x volume per weight of gel of Qiagen QG solubilisation buffer was added to the falcon tube containing the excised gel bands and placed in a 55˚C water bath for approximately 30 minutes. Once fully dissolved, a 1 x volume of the original gel weight of 100 % isopropanol was added and vortexed. Following this, a 700 μL aliquot of the solution was added to a Qiagen MinElute column and centrifuged at 5,400 g for 60 seconds. The flow-through solution was passed through the column again and centrifuged two more times for 60 seconds and then the flow-through was discarded. This step was repeated until all of the dissolved gel solution was had been transferred through the MinElute column. Subsequently, 500 μL of QG solubilisation buffer was added to each column and centrifuged for a further 60 seconds at 5,400 g. Once the flow-through was discarded, 700 μL of Qiagen PE wash buffer was added to each column and centrifuged at 11,300 g for 60 seconds. The dsDNA was eluted from the column membrane using a 1:50 dilution of Elution buffer in ddH$_2$0. This was performed by adding 30 μL of the diluted Elution buffer directly onto the column membrane, incubating for 1 minute and centrifuging at 11,300 g for 60 seconds. This step was repeated with a further 20 μL of diluted elution buffer, making a total volume of 50 μL. The purified dsDNA was then stored at -20˚C until further use.

2.12 Streptavidin Magnetic Bead Capture and Alkaline Denaturation Strand Separation of Biotin Labelled DsDNA

For placement into further SELEX rounds, the dsDNA must be strand separated to obtain ssDNA. This was achieved using streptavidin-coated magnetic beads and alkaline denaturation. During PCR, biotin-labelled reverse primers were used which prime the synthesis of the complementary strands of DNA. The biotin label present on the reverse primer binds the streptavidin-conjugated magnetic beads through a protein-ligand interaction. Single-stranded sequences are then isolated through alkaline denaturation. The details of this method follow; 50 μL of Streptavidin magnetic beads (BioLabs, New England)

were resuspended and pipetted into a 1.5 mL Eppendorf tube along with 50 μL of 2 x sodium citrate buffer (300 mM NaCl: 40 mM citrate: pH 7.0). The beads were then washed in 50 μL of 1 x sodium citrate buffer and then added to 50 μL of dsDNA, along with 50 μL 2 x sodium citrate buffer and incubated at room temperature for 45 minutes with periodic mixing every 20 minutes. This step allows the biotin and streptavidin interaction to occur. Following this, beads were collected and washed with 50 μL of 1 x sodium citrate buffer, and then twice with U/P H$_2$O. Subsequently, beads were transferred to a 1.5 mL Eppendorf tube containing 50 μL of 0.3 M of NaOH and incubated for 2 minutes and 15 seconds. The beads were removed and discarded and the NaOH solution containing the ssDNA was neutralised using 300 μL of QX1 solubilisation buffer (Qiagen). An aliquot of 20 μL of 3 M sodium acetate was added followed by 10 μL of agarose beads from Qiagen II gel extraction kit and vortexed gently. The solution was then incubated on a heat block at 50˚C for 10 minutes with periodic mixing in 2-minute intervals. The agarose beads were centrifuged at 5,400 *g* for 30 seconds and the supernatant was removed. Beads were resuspended in 500 μL of QX1 buffer (Qiagen) and centrifuged for a further 30 seconds and the supernatant was removed. Beads were then resuspended in 500 μL of Qiagen PE buffer and centrifuged in the same manner. The supernatant was then removed, and the agarose beads were dried in a Biohazard Cabinet for 30 minutes. Following incubation, agarose beads were then resuspended in the volume required for the next SELEX round and incubated in a heat block at 95˚C for 10 minutes with periodic mixing. Finally, the bead pellets were centrifuged at 11,300 *g* for 90 seconds and the supernatant containing the ssDNA was collected and stored at -20˚C, pending further use.

2.13 Negative Selection

A negative selection step was employed to remove sequences that may non-specifically bind the affinity matrix and remove these from the enriched library pool. The negative selection step was performed prior to SELEX Rounds 3 and 6 as follows. An aliquot of 2 mg of dried control beads were resuspended in 60 μL of 1 x BWB in a 1.5 mL Eppendorf tube and incubated with the evolved library pool for 1hr on a rotating wheel at 1.5 speed in 4˚C, imitating the conditions of the library-target incubation. This was repeated before Round 6 with 10 mg of dried control beads to increase stringency in negative selection. Following incubation, the Eppendorf tube was centrifuged at 1,000 *g* for 60 seconds and the

supernatant containing the sequences that didn't bind the control matrix were isolated and used in the following SELEX round.

2.14 Counter Selection

The purpose of incorporating counter selection steps within SELEX is to increase the specificity of oligonucleotide sequences in the evolved library and reduce non-specific binding. This was achieved by incubating the library with structurally similar molecules. Sequences that bound these similar molecules were removed from the library such that only sequences that could distinguish between structural analogues remained. Three molecules were chosen for counter selection due to their structural similarity to the target molecule NP, and they are all synthetically produced and present as contaminants in similar environments. These molecules were bisphenol A, dibutyl phthalate and 2-phenylphenol (Sigma Aldrich), shown in Figure 2.4. All three molecules were solubilised in 100 % EtOH, then diluted in 1 x BWB to a final concentration of 1 µM in 1 % EtOH (v/v). Following library-target molecule incubations and prior to the washing steps, the incubation solution was loaded into an elution column and washed with 1 mL of 1 x BWB. Following this, the beads were resuspended with 100 µL of 1 µM of each counter molecule and incubated for 15 minutes, and then regular wash steps were applied (see Section 2.7 for details).
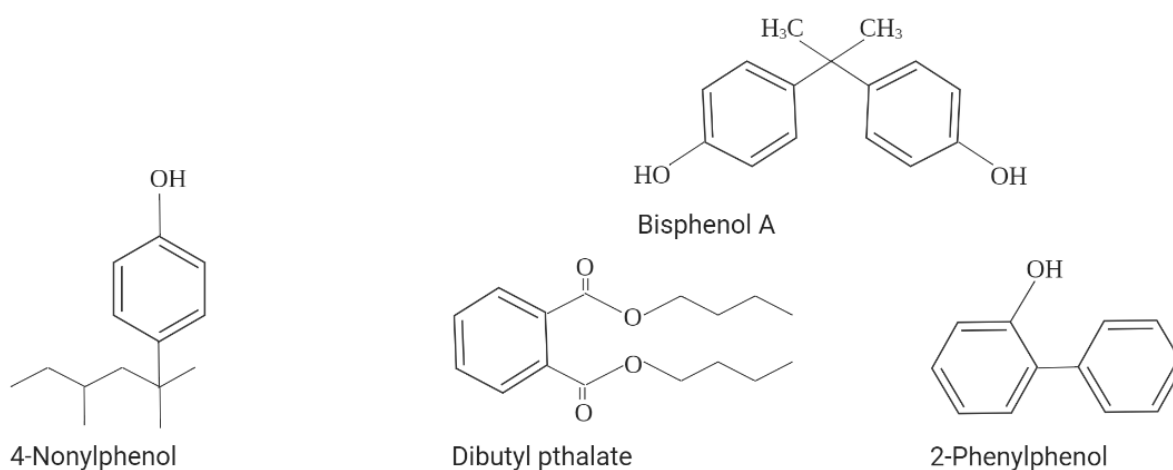
Figure 2.2: Molecules used in counter selection steps: Bisphenol A, Dibutyl phthalate and 2-Phenylphenol. Molecules are structurally similar to target Nonylphenol.

2.15 Next Generation Sequencing Preparation

In the final round of SELEX, non-biotin labelled reverse primers were used in PCR in preparation for next-generation sequencing. Following PCR, PCR amplicons were purified and concentrated using the Monarch PCR and DNA clean-up kit following the manufacturer's instructions (Biolabs, New England). This method was used to reduce the loss of DNA amplicons that occurs through extraction from agarose gel and to concentrate the dsDNA. Each centrifugation step was performed at 16,000 x g. In brief, the PCR amplicons from each condition were pipetted into separate 1.5 mL Eppendorf tubes and a 5:1 ratio (v/v) of binding buffer: sample was added and vortexed. A column was inserted into a collection tube and 500 µL of the sample was loaded onto the column and centrifuged for 1 minute. This step was repeated until the sample was used. Subsequently, 200 µL of DNA wash buffer (Biolabs, New England) was loaded onto the column and spun down. The column was then transferred to a clean 1.5 mL microfuge tube and 20 µL of DNA elution buffer (Biolabs, New England) was loaded into the column, incubated for 1 minute and centrifuged at 16,000 g. The flow-through solution was collected and stored.

Aliquots of the starting ssDNA N40 library needed to be converted to dsDNA for sequencing. This was achieved by diluting the 100 mM library to 10 nm using U/P H_2O, and then amplifying by PCR. The PCR products were then purified in the same manner as the final SELEX libraries.

The concentrations of each enriched and control library sample were then measured using a Qubit high-sensitivity dsDNA kit on a Qubit3 (Life Technologies, United States). Samples were labelled and sent to Gen X Pro (Frankfurt, Germany) for sequencing. The preparation of sequences for Illumina Nextseq Platform was carried out by Gen X Pro. Adaptor sequences were ligated onto the sequences which contained integrated nucleotide barcodes and molecular identifiers. Libraries were then sequenced using Illumina NextSeq500 applying a sequencing depth of one million reads.

Following sequencing, adaptor sequences were removed by Gen X Pro and sequencing data was returned in a FASTQ format and was accessed through the Gen X Pro server for further bioinformatics analyses.

2.16 Description of Bioinformatics Pipeline

The bioinformatics pipeline used in this study is illustrated in Figure 2.5 is described in detail below.
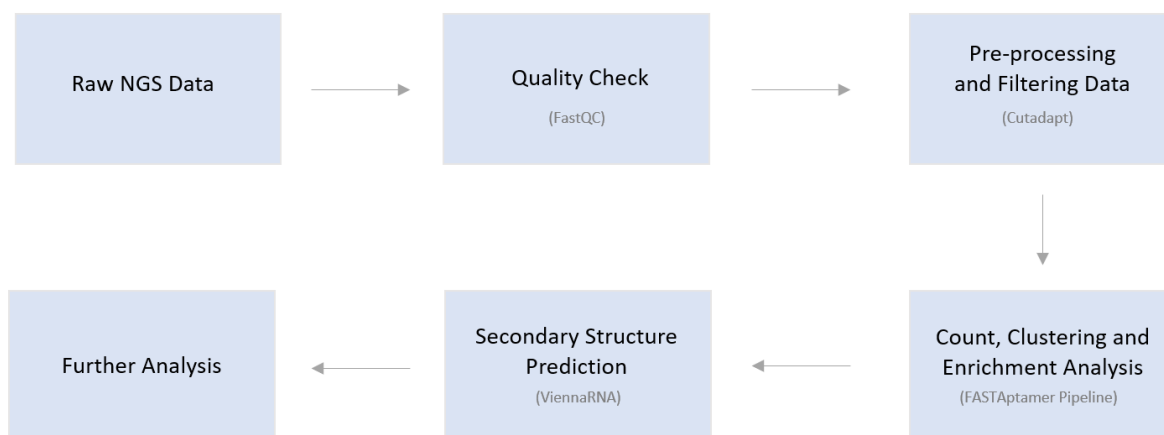
Figure 2.3.: Outline of bioinformatics pipeline.

2.16.1 Quality Check of HTS Data

Software available online (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to assess the quality of the raw sequencing data (Andrews, 2010; Leggett et al., 2013). All five FASTQ files containing data from the N40, low and high stringencies for 4-NP, and low and high stringencies for NPTE libraries were analysed through FASTQC Version 0.11.9. The sequence quality, nucleotide content, length distribution and sequence duplication levels of the library were assessed. Phred scores, which indicate individual nucleotide base quality, of above 30 indicates a 99.9 % accuracy in the identification of each nucleotide. Primer regions were excluded from this analysis. Phred scores for the N40 region of sequences in the low and high stringency libraries to 4-NP were >32. Phred scores for sequences in low and high stringency libraries to NPTE were >27 and >32, respectively.

2.16.2 Pre-processing and Filtering of Sequence Data

A Bash script that utilised Cutadapt Version 2.8 was used to perform the following pre-processing and filtering of raw high throughput sequence libraries (Martin, 2011). The script can be found in Appendix C.2.

Throughout SELEX, only the forward strand of DNA was used for selection in each round. Prior to sequencing, ssDNA from each library was converted into dsDNA through PCR. Therefore, complementary sequences of the forward strands were present in the final sequenced libraries. Forward strands in the library contained a forward primer sequence of 19 nucleotides long, a 40-mer random region and a complementary region of the reverse primer of 20 nucleotides in length. Complimentary sequences contained a reverse primer sequence of 20 nucleotides long, a 40-mer random region and the reverse complement of the forward primer of 19 nucleotides in length. Sequences that did not contain the forward primer or reverse complement of the reverse primer were removed from the sequence pool using Cutadapt.

Pre-processing of the forward strands also involved the removal of flanking primer sequences to isolate the random nucleotide region. Cutadapt was used to remove forward primers and the reverse complement of the reverse primers. Specifications of 14 minimum overlap and 0.3 maximum error rate were made. Primer sequences which flanked the random nucleotide regions were present within each *in vitro* selection round, however, prior to bioinformatic analyses, primers were removed. The rationale for primer removal prior to analyses was that as primer regions were consistent across all sequences in each population, they do not influence the analyses of the variable regions being investigated. In contrast, the presence of the primers would affect any clustering analyses (as margin of error was allowed for primer regions so any substitutions would have been included in the levenshein edit distance for clustering; see Section 2.16.3 for more detail) and conserved motif searches. Thus, in the interest of scope, time and clarity they were excluded in the following bioinformatics pipeline.

The original selection library was designed to have a random region consisting of 40 nucleotides long. Sequence reads were filtered through Cutadapt to only include sequences between 37 and 41 nucleotides long to allow for minimal insertions and deletions. All remaining sequences were outputted as a FASTQ file then analysed through the aptamer-specific bioinformatics toolkit called FASTAptamer (Version 1.0.14) (Alam et al., 2015).

2.16.3 FASTAptamer Pipeline

FASTAptamer is an open-source toolkit that is comprised of several modular Perl scripts forming a pipeline including FASTAptamer-count, -cluster, -compare and -enrich. A description of the pipeline is included below and the command-line information for each script is available in Appendix C.

FASTAptamer-count takes sequences in FASTQ files and assigns tags with unique identifying features required for further analysis while preserving sequence metrics. It then outputs the data into FASTA format compatible with future bioinformatics scripts. FASTAptamer-count also counts the abundance of each sequence (reads), ranks sequences on their abundance in the population and normalises abundance to reads per million (RPM) to ensure different sequence populations are comparable.

FASTAptamer-compare is a script used to compare the sequence distribution of two input populations and indicates conserved sequences present in these populations. Information is input in FASTA format and outputted via a plain text file.

FASTAptamer-cluster is a tool that groups closely related sequences into "clusters" or families. Clustering is based on a Levenshtein edit distance defined by the user. An edit distance of 7 was selected to allow for substitutions, insertions and deletions. To generate a cluster, a parent or seed sequence was identified which was based on sequence abundance. Each sequence in the population was then compared to the seed sequence and clustered together if the edit distance was less or equal to the Levenshtein edit distance defined. For example, the most frequent sequence in the population will be the parent sequence of cluster Number 1. Any sequences which require 7 or fewer nucleotide changes to transform into the parent sequence will be included in Cluster 1. Then the second most abundant sequence in the population will be the parent sequence of Cluster 2 and all unclustered sequences will be compared. Clustering is performed to identify primary and mutant sequences and allows for the selection of oligonucleotides with different structures and features for investigation and characterisation, while also reducing the large oligonucleotide pool for analysis. The output of clustering is in FASTA format and includes identity information provided by FASTAptamer-

count, the cluster each sequence was congregated and the rank of each sequence within its cluster.

Lastly, FASTAptamer-enrich calculates the fold-enrichment ratio of sequences that are present in two populations. The output of this script is directed into a tab-separated plain text file containing all information provided from FASTAptamer-count, -cluster and -enrich including sequence length, reads, rank, RPM, cluster and enrichment information. Relevant information was extracted and analysed in the results.

Following the FASTAptamer pipeline, sequences from the top 100 clusters across the four specific libraries (low and high stringency libraries for both 4-NP and NPTE) were investigated further. This number was chosen to incorporate a sufficient quantity of data to make a meaningful analysis, whilst also considering time, clarity and scope. In each condition, the top 100 clusters were ranked based on their total reads. The total reads included the number of reads attributed by the parent sequence and the reads contributed from all the unique sequences within that cluster to determine the most abundant cluster. From this, the top 10 parent sequences of each cluster, across all four libraries were investigated further for possible secondary structures, presence of sequence motifs and target affinity through characterisation.

2.16.4 Secondary Structure Prediction and Motif Search

The top 10 candidates from each of the four libraries were folded with the presence of flanking primer regions using ViennaRNA software package (Version 2.4.18). ViennaRNA is a computational method that contains command-line programs that use algorithms to predict RNA or DNA structures. There are several functions available in the ViennaRNA package including RNAfold, Alifold and inverse fold that predict structures using different parameters such as using a consensus structure based on aligned sequences or finding sequences based on pre-defined structures. RNAfold is a basic function and was used to obtain secondary structure predictions and minimum free energy (MFE) of all 40 aptamer candidates. Parameters specified were folding DNA at 21°C with maximum expected accuracy, partition function and base pairing probability matrix (Lorenz et al., 2011). Further information can be found in Appendix C.7.

2.17 Gold Nanoparticle Characterisation

Ten aptamer candidates from each of the four libraries were selected based on total cluster frequency and screened using a medium-throughput gold nanoparticle-based assay to detect any aptamer-target binding events.

Prior to characterisation, candidate aptamers were synthesized that included the primer regions for medium-throughput gold nanoparticle (AuNP) characterisation. Primer regions are not considered to contribute to the formation of aptamer structures (Cowperthwaite & Ellington, 2008), however, there is evidence of primer regions contributing to the aptamer binding pocket (Famulok, 2002), therefore, removal of these regions may interfere with its binding to the target molecule. To replicate *in vitro* selection conditions and maximise the chance of observing aptamer-target binding events, primer regions were included for the characterisation of oligonucleotides. In addition, the gold nanoparticle assay for characterising aptamer candidates was optimised using an 81-mer aptamer. Thus, the libraries investigated herein were similar in length (i.e.79-80-mer). Synthetic oligonucleotides were synthesised and provided by Sangon Biotech in a lyophilised state. The powder was centrifuged for 1 minute at 14,100 g before being hydrated to 10 µM in U/P H_2O and vortexed.

2.17.1 Gold Nanoparticle Synthesis

2.17.2 Preparation for AuNP Synthesis

Prior to AuNP synthesis, an Aqua Regia solution of 1:3 ratio of concentrated nitric acid (HNO_3) : concentrated hydrochloric acid (HCL) was prepared to wash all glassware. All procedures were conducted in a fume hood using protective eyewear. First, a 1 L Schott bottle was washed with Aqua Regia then rinsed thoroughly with double distilled (dd)H_2O. This bottle was used to carry ddH_2O to rinse all remaining glassware that has undergone acid washing. A 500 mL flat bottom flask, magnetic stir bar, and condenser were washed with Aqua Regia for 15 minutes then rinsed thoroughly with ddH_2O to remove any remaining traces of the solution. Once finished, all Aqua Regia solutions were disposed of in a neutralisation tub containing an aqueous basic solution of sodium bicarbonate.

2.17.3 Synthesis of AuNP's

Gold nanoparticles were prepared through a gold III chloride trihydrate (HAuCl$_4$.H$_3$O) citrate reduction. To achieve this, a 95 mL aqueous HAuCl$_4$.H$_3$O solution was added to a flat-bottomed reaction flask, placed on a hot plate set at 400°C and 370 rpm, and refluxed for 30 minutes. Simultaneously, a 70.54 mM of sodium citrate was prepared and filtered into a 100 mL Schott bottle using a 0.45 μm syringe filter onto a 50 mL syringe and heated in a water bath set to 95°C. After 30 minutes of boiling, 5 mL of the sodium citrate solution was added to the boiling HAuCl$_4$.H$_3$O solution and stirred for 15 minutes until the golden colour of the solution turned from grey to purple and then to a final deep red colour, symbolic of 10-15 nm AuNP's. Once cooled at room temperature, the AuNP solution was stored at 4°C until further use.

UV absorption was performed using a nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific) to determine the theoretical concentration of the newly synthesised AuNPs. A 10 μL aliquot of the AuNPs was loaded onto the nanodrop and absorbance was measured at wavelengths of 520 and 625 nm. Thereafter, the Beers Lambert Law A = Ɛ C l, was applied to determine the AuNP concentration for following experiments.

2.18 Preliminary Experiments determining Solvent Effects in the AuNP Assay

Due to the chemical properties of the target molecule(s), a solvent was required to solubilise a sufficient concentration of the target molecules for AuNP characterisation. No previous literature describing the effects of various solvents on the AuNP assay was found, thus, preliminary experiments were conducted.

The AuNPs were resuspended in varying concentrations of common organic solvents to determine the solvents at differing concentrations that were sufficient to destabilise the AuNPs. An aliquot of AuNPs were suspended in ddH$_2$O or 5 %, 10 %, 15 %, 20 %, 30 %, 40 % and 50 % of ethanol, methanol, and acetone to achieve a final concentration of 12 nM and then vortexed gently. A 100 μL aliquot of each AuNP solution was pipetted into a 96-well plate and left to react at room temperate for 30 minutes. Modification to the AuNP surface charge was determined by UV vis spectroscopy using a Spectro star nano at wavelengths of 625 and 520 nm, and the 625:520 nm absorbance ratio was determined.

To determine whether solvent concentrations interfered with salt-induced aggregation after an aptamer-target binding event, a positive control assay was run using a validated 81mer aptamer-target pair. For this, two test rows of AuNPs were required. In the first row of 12 wells, 100 μL of 12 nM of AuNP resuspended in various concentrations (5-40 %) of organic solvents (ethanol, methanol, acetone) was reacted with 36 nM of OM5-C6 aptamer (William Odey, VUW, unpublished results) in a 96-well plate. In the second row of 12 wells, 100 μL of 12 nM AuNP resuspended in ddH$_2$O was reacted with 36 nM of OM5-C6 aptamer in a 96-well plate. Each line was incubated for 30 minutes before 12 μL of the target molecule methamphetamine (BDG Synthesis) resuspended in differing amounts of U/P H$_2$O were added to both test lines to make a final concentration range of 0, 0.15, 0.625, 2.5 and 10 μM in each well of each row. The reactions were then incubated for a further 10 minutes. Subsequently, 8 μL of 0.5 M of NaCl was added to make a final reaction volume of 120 μL per well. Following a 10-minute incubation, AuNP aggregation was measured as previously described and the 625:520 nm absorbance ratio was determined.

According to the results of the positive control assay (Figure 3.23), 5 % Ethanol and Acetone appeared to have minimal effect on target-induced AuNP aggregation compared to the control line. Concentrations of Ethanol and Acetone >5 % and all concentrations of methanol demonstrated a difference in the sensitivity of the assay in comparison to the control line and therefore, were not considered for further use. The presence of the target in solution was determined from 225 280 nm UV absorbance using a nanodrop ND-1000 spectrophotometer. Readings of target suspended in Acetone showed variable UV absorbance peaks and as such, Ethanol was selected to use in the following experiments.

2.19 Characterisation of Candidate Aptamers using AuNP's

To find a suitable aptamer, the binding affinity of 40 aptamer candidates selected from bioinformatics analyses were assessed using the AuNP assay. A 100 μL aliquot of 12 nM of AuNP were incubated with 36 nM of each aptamer candidate in 2.5 % (v/v) EtOH for 30 minutes on a 96-well plate. Then, 12 μL of 0, 0.016, 0.08, 0.4, 2, 10, 50 and 250 μM of 4-NP or NPTE resuspended in 25 % (v/v) EtOH were added to wells to obtain a final concentration of 0, 0.0016, 0.008, 0.2, 1, 5, 25 μM in 5% (v/v) EtOH in each well. This mixture was incubated for 10 minutes. Subsequently, to aggregate the AuNPs, 8 μL of 0.7 M NaCl was added to each

well to make a final volume of 120 µL. The AuNP aggregation and the 625:520 nm absorbance ratio was determined as previously described in Section 2.18. Two negative controls were implemented, and these were prepared in the same method as detailed above (see Section 2.19). Candidate aptamers were substituted with either ddH$_2$O or a 75mer aptamer not specific for NP to represent the no aptamer and non-specific aptamer controls, respectively.

Oligonucleotides were visually and spectroscopically assessed for target-binding and candidates which appeared to be removed from the AuNP (as evidenced by salt-induced aggregation) after target molecule addition were further validated using technical repeats.

Chapter 3. Results

3.1 Target Preparation

The success of SELEX depends on the ability to separate unbound oligonucleotides from those that bind the target molecule. As 4-NP is a small molecule, it conjugation to a solid support enabled the isolation of target-binding oligonucleotides for further selection rounds. The conjugation of NPTE and 4-NP to Sepharose beads was performed. For both, controls represented Sepharose beads conjugated to a BDDE linker molecule and no target molecule. In Figure 3.1, a clear difference in optical density for control compared to NPTE-conjugated beads was apparent. There was an increase of optical density at 225nm wavelength for NPTE-conjugated beads that was absent for the control beads. The NPTE contains compounds with phenol rings which absorbs at 280nm (Puteh et al., 2015). However, no distinguishable peaks were observed at 280nm indicating the lack of conjugated NPTE. However, the lack of visible peaks may also be owing to a low concentration of conjugated NPTE due to poor solubility. It is also possible that the beads masked the small target molecules, therefore blocking the absorption spectra. Thus, the protocol was modified for 4-NP conjugation whereby an alternative solvent was used to dissolve the 4-NP and two conjugation assessments were performed.
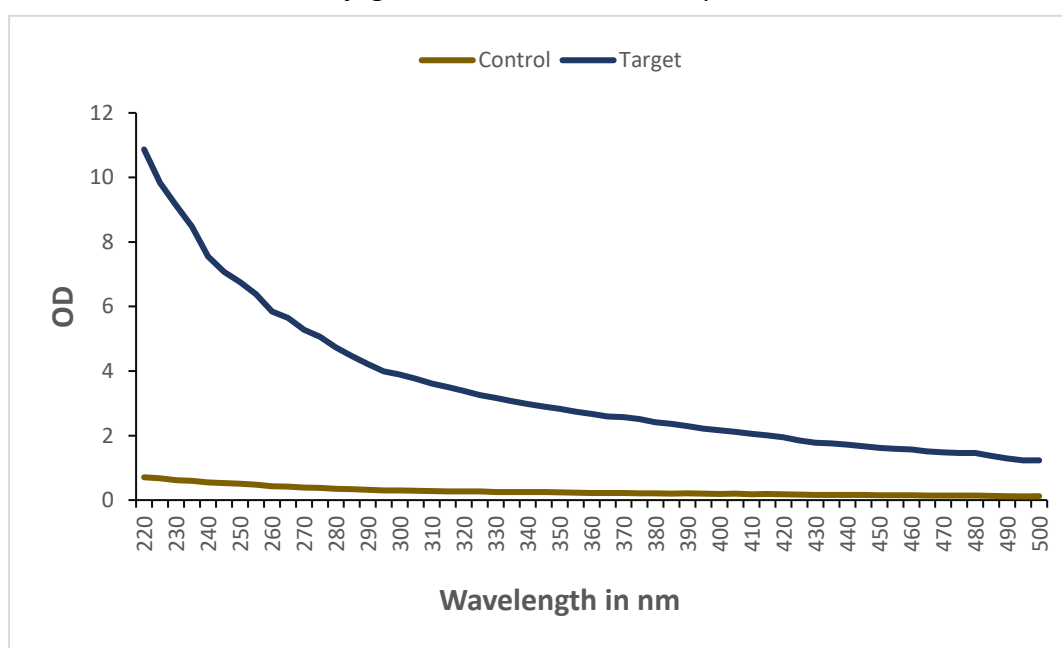
Figure 3.1: Assessment of nonylphenol technical equivalents (NPTE) conjugation onto Sepharose beads. Optical density of presumptive NPTE-conjugated and non-conjugated (control) beads measured at 225nm wavelength.

The single target molecule, 4-NP, was also conjugated onto Sepharose beads and the result was assessed by the absorbance measurements at 225 and 280 nm wavelengths. Figure 3.2 depicts the normalised data from two assessments of one conjugation process. The presence of absorption peaks at 225 and 280nm on the presumptive 4-NP-conjugated beads and their absence on the control beads, indicates the successful conjugation of 4-NP onto the Sepharose beads.
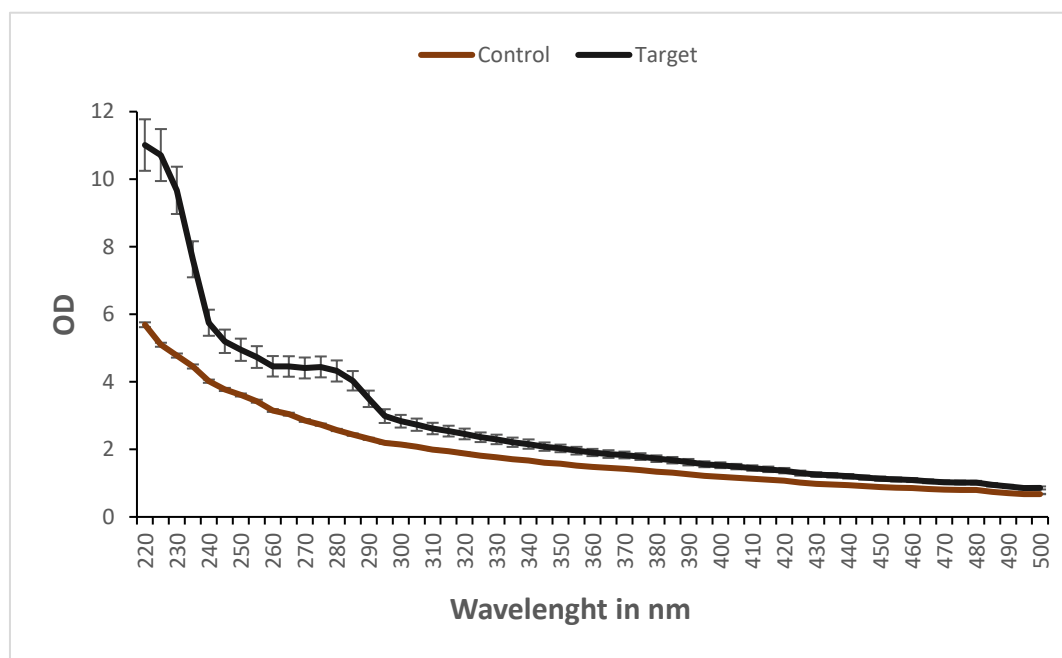
Figure 3.2: Bead conjugation assessment of 4-nonylphenol to Sepharose beads. Control beads show no absorption peaks while target-conjugated beads show a peak at 225 and 280nm indicating the presence of 4-nonylphenol on the beads. Error bars represent the range of data points from the conjugation assessment done in duplicate.

3.2 Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

SELEX was performed consecutively on 4-NP and NPTE binding enriched libraries subjected to either low or high stringency. Each library went through seven rounds of selection before the final libraries were prepared and sent for next-generation sequencing. Within each round, Real-Time qPCR was used on the isolated enriched libraries after each target molecule-incubation to re-amplify the library for the next round and to assess the effect of the low and high stringency measures on the quantity of DNA amplicons generated.
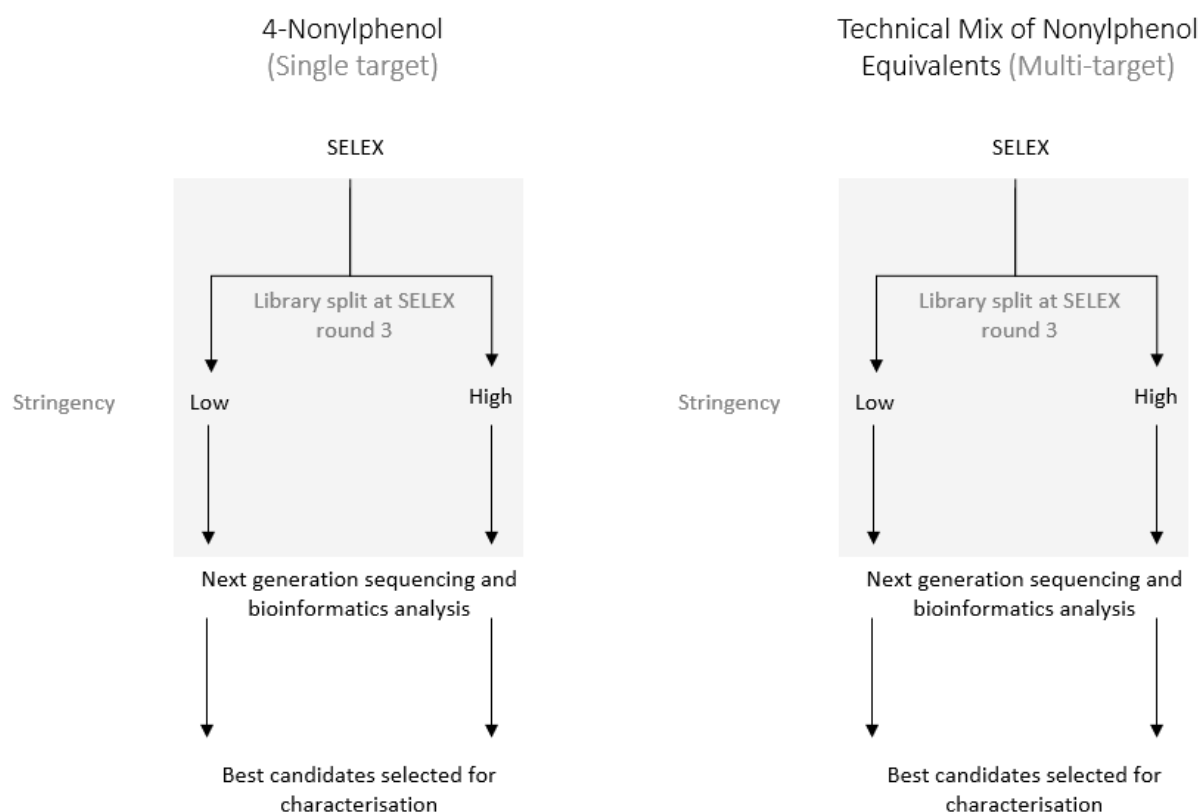
Figure 3.3: Scheme for objectives pipeline for single (4-NP) and multi-target (NPTE) SELEX. The pipeline was identical for both target molecules such that their results could be compared under both low and high stringency conditions.

3.2.1 Real-time PCR Results for Nonylphenol Technical Equivalents

Following each incubation of the ssDNA library with NPTE-conjugated beads, beads were washed with increasing concentrations of Tween 20 detergent in each SELEX round and subsequently sequences bound to NPTE were isolated for amplification via Real-Time PCR.

Figures 3.4 and 3.5 are representative images of the Real-Time PCR raw data from the NPTE SELEX. Figure 3.4 depicts the increasing fluorescence across amplification cycles, as the amount of dsDNA increases there is a corresponding increase in fluorescence signal from SYBR green dye intercalation with the dsDNA.

The cycle threshold (Ct) is the number of cycles required for the fluorescence signal to exceed the background fluorescence noise created during initial cycles of PCR. The Ct value is set at the exponential phase of amplification where the products are directly proportionate to the

50

starting template concentration. Once the cycle number of each reaction surpasses the set threshold value, all reactions can be compared. The Ct values are inversely proportionate where a more concentrated starting template results in lower Ct values.
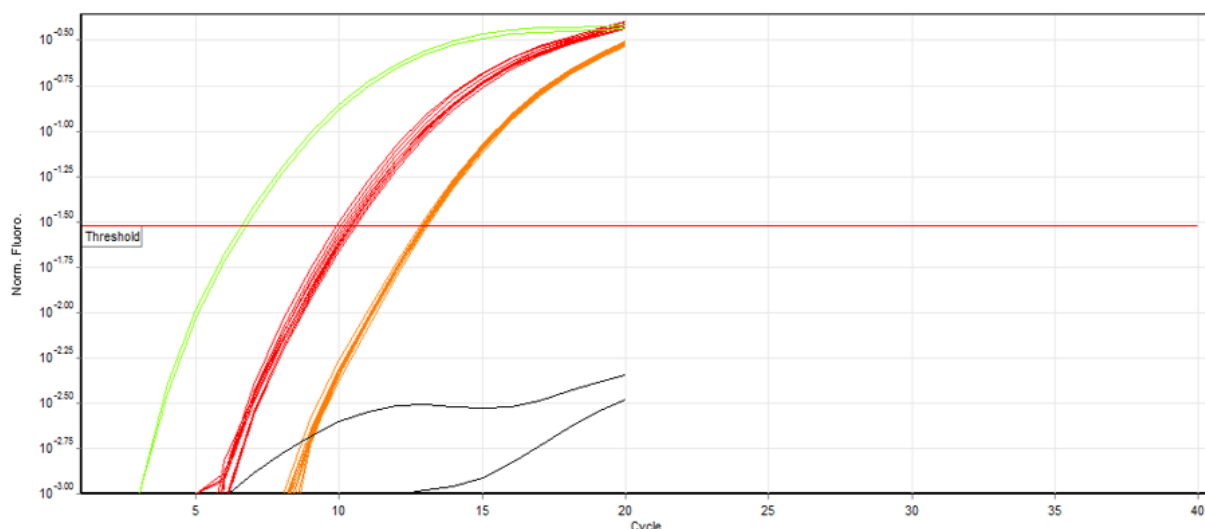
Figure 3.4: A representative image of raw quantification data using Real-Time PCR of an oligonucleotide library enriched for sequences that bind nonylphenol technical equivalents. The lines in green, black, red and orange represent the positive (N40 library) and negative (U/P H$_2$O) controls, and amplicons from the enriched libraries subjected to low and high stringency conditions, respectively. The horizontal red line depicts the cycle threshold.

As seen in Figure 3.4, the threshold depicted as a red horizontal line was set at the halfway point of amplification (when in the logarithmic scale) to ensure the fluorescence observed is sufficiently above background noise and that it remains in the exponential phase of the amplification profile. The positive control reached the fluorescence threshold earlier at a lower Ct value (6.65 Ct value) compared to the enriched libraries subjected to low and high stringency conditions (6.65, 10.30 and 12.95, respectively). As expected, the negative control did not cross the threshold indicating there was no contamination within the PCR reactions.

Following Real-Time PCR, the amplicons were subjected to a melt curve analysis to assess the reaction specificity and homogeneity of products. Thus, the thermal cycler measured the fluorescence at a low starting temperature, and then incrementally increased the temperature while continuing to measure fluorescence. A peak of fluorescence can be seen when the temperature is sufficient to denature dsDNA into ssDNA, causing dissociation of the SYBR green dye. The melting temperature of a PCR product is based on factors including sequence length and guanine-cytosine (GC) content. If multiple distinct melt temperatures were observed, it would indicate products of distinct size or composition. As SELEX contains

51

a diverse pool of oligonucleotides with different sequence characteristics, a broad melt temperature peak range would expected to be observed.

Figure 3.5 shows the raw melt curve of the DNA products following Real-Time PCR of the libraries enriched for NPTE binding sequences and subjected to low and high stringency conditions. A single broad peak at approximately 84˚C indicated that amplicons contained similar sequences and that the reaction contained negligible primer-dimer artefacts.
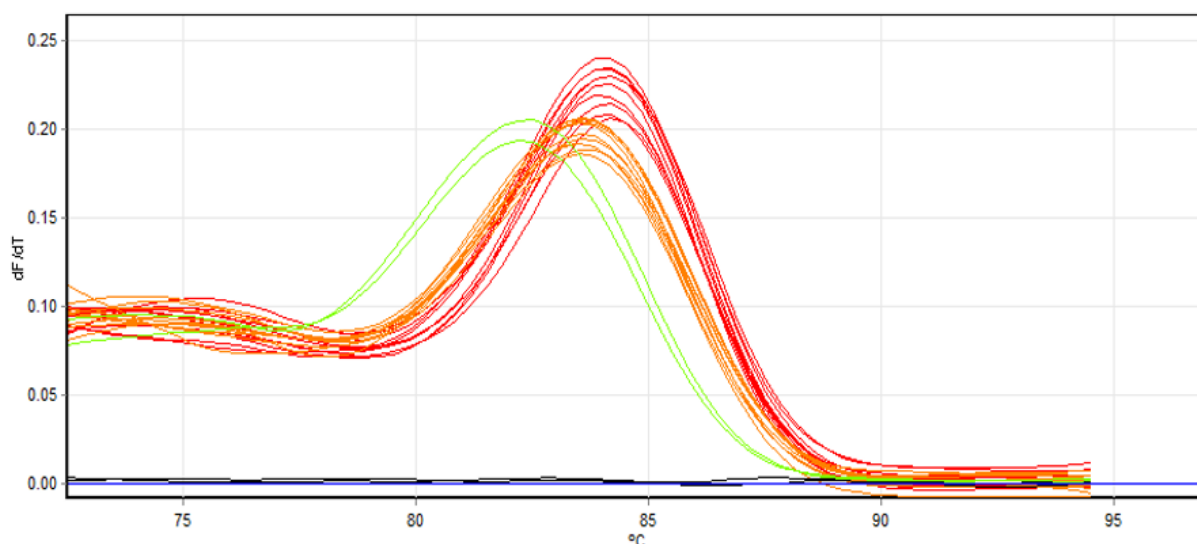
Figure 3.5: A representative image of a melt curve for PCR products generated from an oligonucleotide library enriched with sequences that bind to nonylphenol technical equivalents (NPTE) as measured by Real-Time PCR. The lines in green, black, red and orange represent the positive (N40 library) and negative (U/P H_2O) controls, and amplicons from the enriched libraries subjected to low and high stringency conditions, respectively.

The differing Ct values for the enriched libraries over the successive SELEX rounds infers relative fold changes and evolution of the library. A relative fold change denotes the difference in PCR product quantity from an original to a subsequent measure. In this case, comparing the quantity of DNA recovered following low and high stringency washes in each SELEX round, to the quantity of DNA recovered from the first low stringency wash in SELEX Round 3 when detergent was introduced. Relative fold change is calculated using the equation $2^{-(\Delta Ct)}$ (Livak & Schmittgen, 2001). The delta Ct value was calculated by subtracting the average Ct value from the low stringency condition in SELEX Round 3 from the average Ct

values for both low and high stringency conditions for each subsequent round inputted into the relative fold change equation, as illustrated in Figure 3.6.
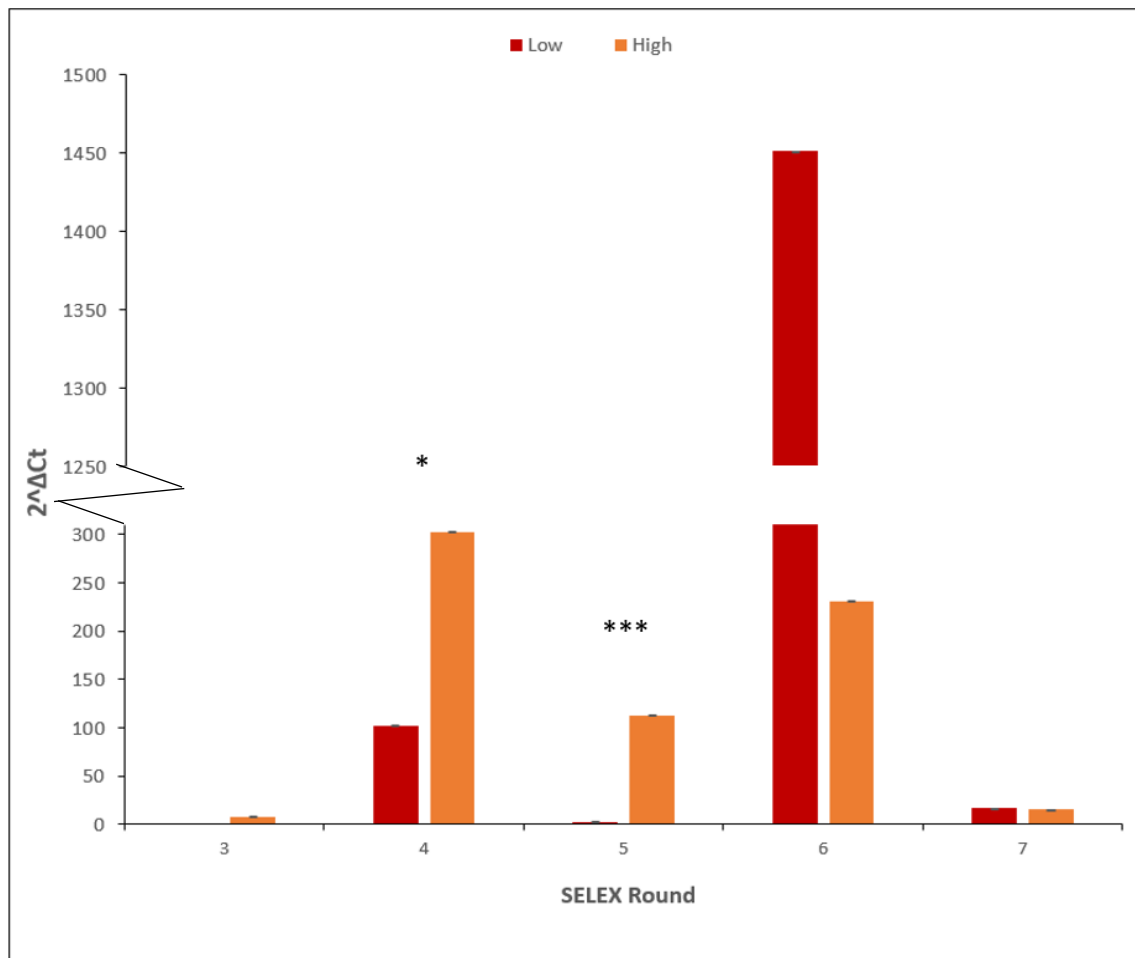
Figure 3.6: Relative mean fold change ($2^{-\Delta Ct}$) of PCR products recovered from low and high stringency washes across SELEX rounds for libraries being enriched for sequences that bind nonylphenol equivalents (NPTE). The bars and error bars represent the SEM from 10 replicates.

The relative fold change refers to the quantity of DNA recovered from all selection rounds normalised to the quantity of DNA recovered from selection Round 3. It was expected that increasing stringency conditions would result in fewer sequences recovered following each SELEX round, culminating in a decrease in fold change as the SELEX rounds progressed. Additionally, it would be expected that a lower quantity of DNA is recovered from the libraries subjected to the high, compared to low, stringency selection pressures in each round. However, Figure 3.6 shows disparate results where the greatest fold change was observed in Round 6 where increased concentrations of dsDNA were amplified despite the harsh negative selection step. It was expected that the negative selection step would reduce the library size. The fold change from each condition fluctuates greatly between rounds. This may be

indicative of a library that includes a large number of sequences that exhibit non-specific binding. Additionally, this data depicts that fewer sequences were recovered from low versus high stringency conditions in Selection Rounds 3, 4 and 5. Technical factors during the strand separation in SELEX Rounds 3, 4, and 5 of the aptamer library subjected to the low stringency conditions may have caused a loss of sequences contributing to the results depicted in Figure 3.6.

3.2.2 Real-time PCR Results for 4-Nonylphenol

In Figure 3.7, the red horizontal line indicates the threshold set at the halfway point of amplification at which the fluorescence levels rise above background noise. The positive control (shown in green) amplified indicating the PCR reaction was set up correctly and ran as expected. Amplicons from the libraries enriched for sequences that bind 4-NP and were subjected to low and high stringency conditions required an average of 19.35 and 18.90 cycles to reach the threshold, respectively. The negative control crosses the threshold at 26 cycles revealing a small amount of contamination in the reaction, however, there was sufficient separation in Ct value of the negative control and the Ct values of the enriched libraries, to have minimal impact.
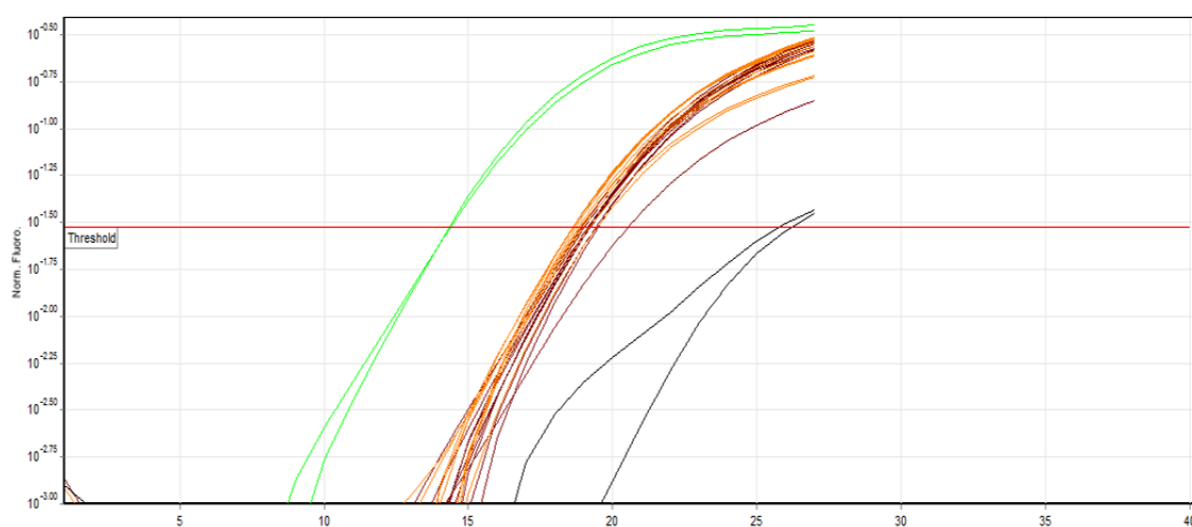
Figure 3.7: A representative image of raw quantification data using Real-Time PCR of an oligonucleotide library enriched for sequences that bind 4-nonylphenol (4-NP). The lines in green, black, brown and orange represent the positive (N40 library) and negative (U/P H$_2$O) controls and amplicons from the enriched libraries subjected to low and high stringency conditions, respectively. The horizontal red line depicts cycle threshold.

Sequence-specific characteristics such as high GC content of some PCR products can cause multistage melting transitions. Regions on dsDNA rich in GC content are more stable than others and do not melt immediately, thus remain as dsDNA until the temperature is high enough to cause melting into ssDNA. This results in a melt curve with multiple peaks. Figure 3.8 demonstrates such a curve for amplicons in the positive control and in both enriched libraries that were produced under low and high stringency conditions, indicating the amplicons in these libraries have a similar range of sequences to each other. Additionally, multiple peaks in the melt curve can also indicate different sequence products, however, these will be separated during the subsequent gel electrophoresis isolation method and only gel bands equivalent to the positive control were extracted. The multiple peaks in the melt curve of the positive control may have been due to the formation of parasitic sequences that are readily amplified under less favourable conditions such as when reagents become scarce in later cycles or may be the result of spurious product formation due to over-amplification of the diverse library. We could speculate that if the PCR reaction was stopped earlier, the products that form these additional peaks in the melt curve would not present. Nevertheless, undesirable DNA was removed during gel extraction and purification.
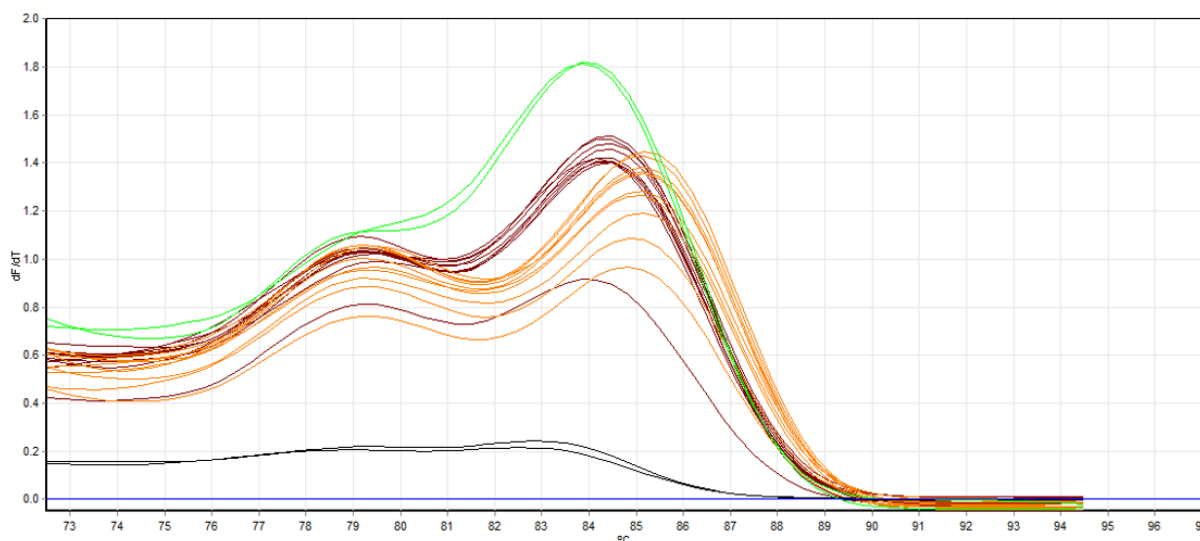
Figure 3.8: A representative image of a melt curve for PCR products generated from an oligonucleotide library enriched with sequences that bind to 4-nonylphenol (4-NP) as measured by Real-Time PCR. The lines in green, black, brown and orange represent the positive (N40 library) and negative (U/P H_2O) controls, and amplicons from the enriched libraries subjected to low and high stringency conditions, respectively.

The highest concentration of DNA recovered is following SELEX Round 3 for both enriched libraries (*i.e.* low and high stringency). This was followed by a decreased fold change for both conditions across SELEX Rounds 4,5 and 6. The overall fold change decreases as stringency increases across rounds. This was indicated by a lesser amount of DNA recovered (excluding SELEX round 7) as the detergent concentrations increased in each selection round in the enriched libraries subjected to both stringency conditions.

Moreover, there is a larger fold reduction in the library under the high, compared to the low, stringency conditions within each selection round up until Round 7. This indicated that fewer sequences were recovered from the elution steps following the high stringency washes. Selection Round 6 depicts minimal fold differences between low and high conditions suggesting the effects of the different stringencies are minimised. SELEX Round 7 shows an increase in fold change compared to Round 6 in both low and high stringency libraries. As library enrichment is occurring at Round 7, this could suggest that within each condition, the library of sequences was converging, resulting in the selection of specific sequences. If an eighth selection round was performed, we could speculate that a greater fold change and increased enrichment within each library would be observed, indicative of the selection of specific sequences. Before Selection Round 6, a rigorous negative control step was employed which may have contributed to these observations that minimal fold difference between low here and high stringency conditions and a reduced quantity of DNA was found compared to previous selection rounds.
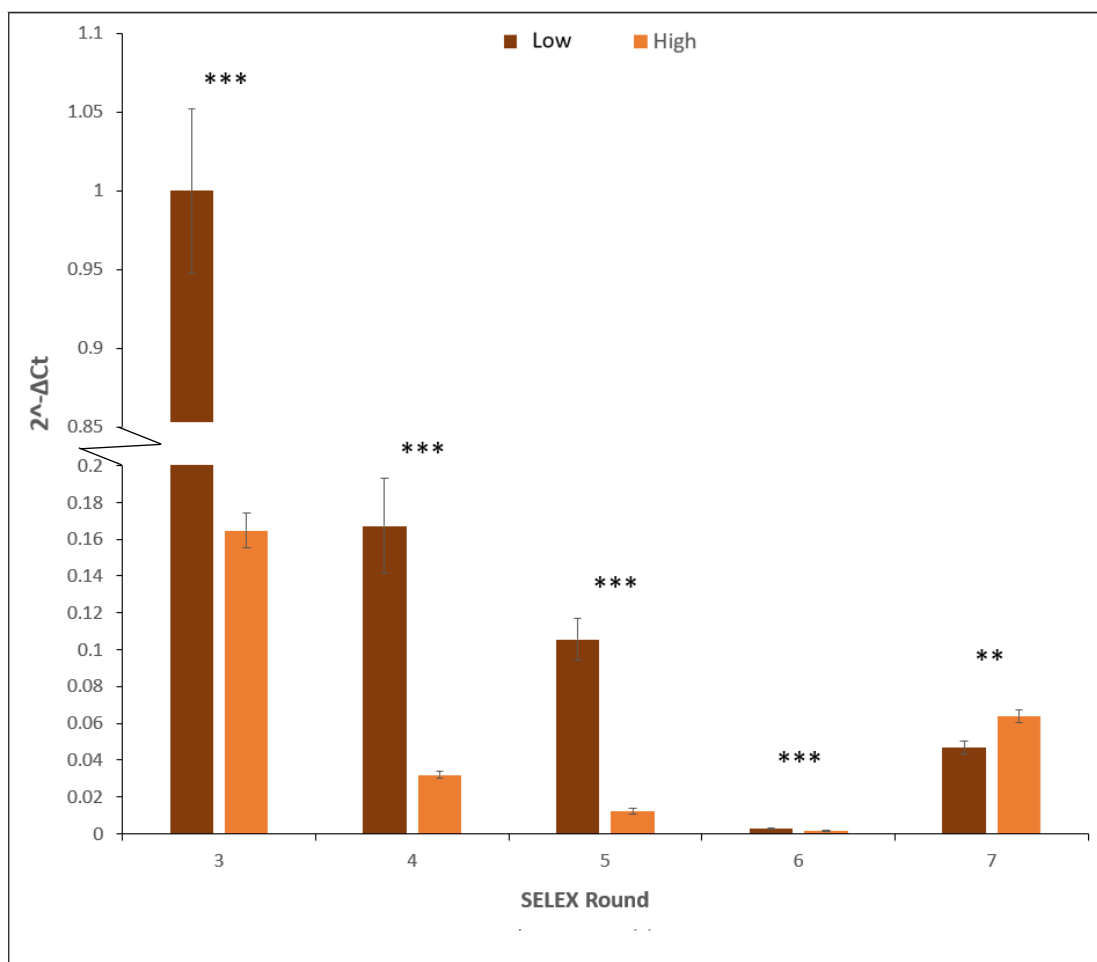
Figure 3.9: Relative mean fold change (2^{-DCt})of PCR products recovered from low or high stringency washes across SELEX rounds for libraries being enriched for sequences that bind 4-nonylphenol (4-NP). The bars and error bars represent the SEM from 10 replicates.

3.2.3 Agarose Gel Electrophoresis and DNA product Extraction

Following Real-Time PCR amplification of each selection round, the PCR products were separated, visualised and purified on an agarose gel. The following figures are representative images of PCR amplicons preceding (Figure 3.10) and following (Figure 3.11) the split of the sequence pool and introduction of different stringency conditions in libraries being enriched for sequences that bind 4-NP and NPTE.

In SELEX Rounds 1 and 2, a single library was amplified to increase the frequency of sequences prior to the division of the library and the introduction of stringency conditions. The library was amplified to the exponential phase of PCR each time to ensure sufficient DNA concentration for the following selection rounds, whilst reducing the risk of amplifying large amounts of parasitic DNA. In Figure 3.10, Lanes 3-7 contained amplicons from the same library that had not been subjected to any stringency measures. The DNA samples in Lanes 3-7 contain two distinct bands which may be due to the result of overamplification of parasitic DNA in the PCR or some spurious product formation. The yellow arrow signals the bands that are the same size as that in the positive control (Lane 11). It was these bands that were excised for further SELEX.
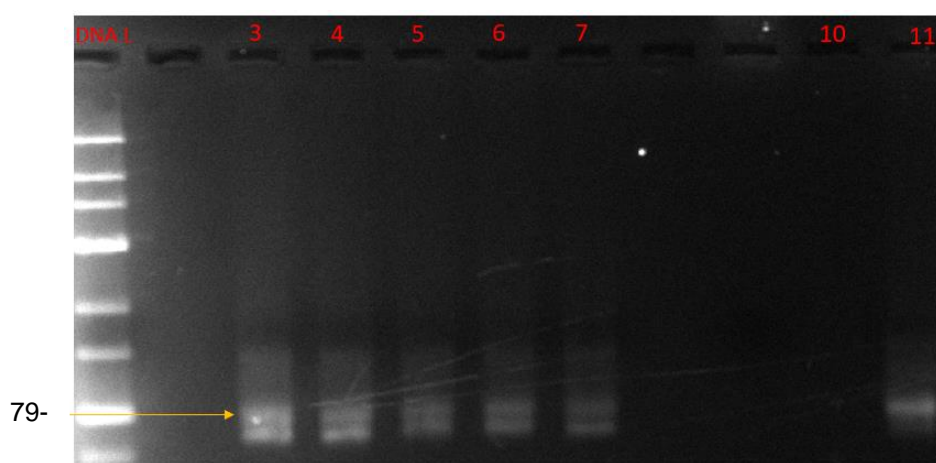
Figure 3.10: A representative image of amplicons visualised on an agarose gel from SELEX Rounds 1 and 2 for the libraries being enriched for sequences that bind 4-nonylphenol (4-NP) and nonylphenol technical equivalents (NPTE). Lane 1 contained the DNA ladder. Lanes 3-7 contained the enriched library for sequences that bind 4-NP and NPTE following non-stringent conditions. Lanes 10 and 11 contained the negative (U/P H_2O) and positive (N40 library) control amplicons, respectively. The yellow arrow represents the bands extracted that were approximately 79 bp in length.

In SELEX Rounds 3,4,5 and 6, the libraries from both low and high stringency conditions, were amplified to the end of the exponential phase of PCR in order to retain enough DNA for the following selection rounds. As illustrated in Figure 3.11, the PCR products across all conditions are of similar band intensity. No bands were observed in the negative control. The desired gel bands were excised then the dsDNA was extracted, purified then strand separated using methods described in Sections 2.11 and 2.12. The ssDNA was then used as the enriched library for the next round of selection.
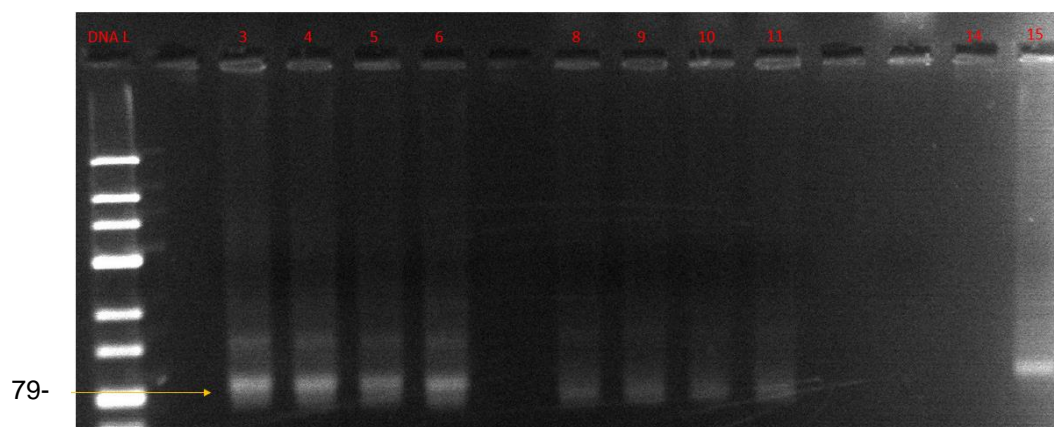
Figure 3.11: A representative image of amplicons visualised on an agarose gel from SELEX Rounds 3,4, 5 and 6 for the libraries being enriched for sequences that bind 4-nonylphenol (4-NP) and nonylphenol technical equivalents (NPTE). Lane 1 contained the DNA Ladder. Lanes 3-6 contained PCR products from low stringency library for Rounds 3, 4, 5 and 6. Lanes 8-11 contained PCR products from high stringency library for Rounds 3, 4, 5 and 6. Lanes 14 and 15 contained negative (U/P H$_2$O) and positive (N40 library) control amplicons, respectively. The yellow arrow represents the bands extracted that were approximately 79 bp in length.

3.3 Bioinformatics results

3.3.1 Sequence Reads

Succeeding Real-Time PCR at the seventh round of SELEX for both libraries enriched for sequences the bind 4-NP and NPTE, dsDNA from each condition was prepared for next-generation sequencing (see Section 2.15). Sequencing was performed by Gen X Pro using Illumina Nextseq500 with a sequencing depth of 1 million reads. The sequencing data was uploaded in a Fastq format then accessed on the Gen X Pro server for further analysis.

Raw sequencing data provided by Gen X Pro was first processed and analysed through FastQC, which indicated the quality of the sequencing data. Sequentially, the data was analysed through the FASTAptamer pipeline which provided the sequences present in each sample pool, the number of reads per sequence, reads per million (RPM), and cluster and enrichment information (see Section 2.16 for details for each stage of the pipeline). Data from these analyses are presented below.

Converting the number of reads to reads per million (RPM) is a form of data normalisation to remove technical biases in sequencing data and allow meaningful comparisons across

different population samples. The RPM values were used in the following graphs to observe the difference in abundance of conserved sequences under different stringency conditions. Larger RPM values correspond with greater sequence frequency and abundance within a population.

Figure 3.12 depicts the RPM of sequences present in both low and high stringency libraries following SELEX towards oligonucleotides that bind NPTE. The top two most abundant sequences were conserved across conditions, however, were excluded from Figure 3.12 as it skewed the data. The most abundant sequence had an RPM of 192,911 and 24,110 and the second most abundant sequence 46,091 and 7,195 in the low and high stringency conditions, respectively. These RPM values are demonstrative of selective amplification where certain sequence characteristics result in higher amplification efficiency due to natural PCR bias.
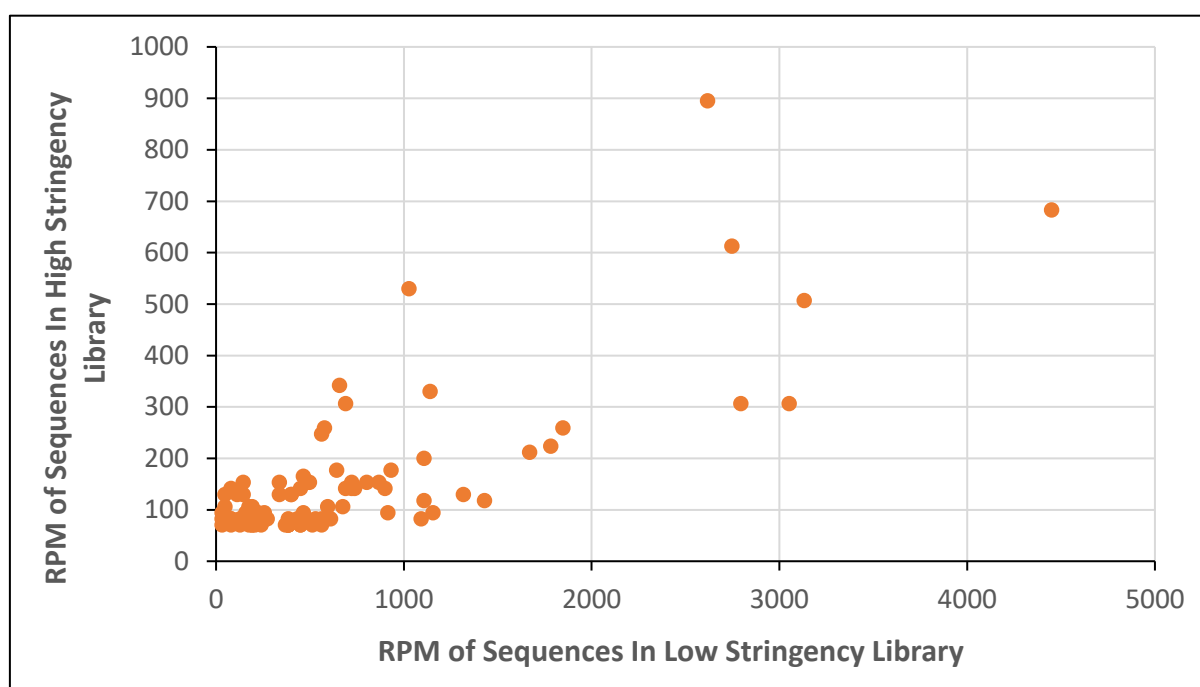
Figure 3.12: The reads per million (RPM) values for conserved sequences after low and high stringency selection in libraries enriched for sequences that bind nonylphenol technical equivalents (NPTE). Ninety-two sequences were conserved between the high stringency library and the top 100 most abundant sequences in the low stringency library.

A large portion of the most abundant sequences were present in both stringency conditions. Furthermore, the majority of conserved sequences had greater RPM values following low stringency versus high stringency conditions indicating that sequences with weaker binding

to the target molecule were successfully removed under the higher stringent conditions resulting in a lower DNA abundance.

Figure 3.13A depicts the RPM of sequences present in both low and high stringency libraries based on the top 100 most abundant sequences in the low stringency library following SELEX towards oligonucleotides that bind 4-NP. Half of the sequences (50/100) from the low stringency library were also found in the high stringency library, however, mostly at a lower RPM (< 100RPM) indicating these sequences did not perform as well under high stringency conditions. Interesting to note was that two sequences were present at a higher RPM in the high versus the low stringency library indicating that these sequences were more favoured under high stringency conditions. Additionally, this data indicates that there is specific binding occurring between oligonucleotides and targets as there were conserved oligonucleotide sequences across both stringency conditions, despite libraries being separated by five rounds of selection.

Figure 3.13B depicts the RPM of sequences present in both low and high stringency libraries based on the top 100 sequences in the high stringency library following SELEX towards oligonucleotides that bind 4-NP. The most abundant sequence from the high stringency library had an RPM of 3158. Higher RPM values may indicate selective amplification. As demonstrated from the most abundant sequences in the high stringency library, there were few sequences conserved in the low stringency library and were present in low abundance inferred from low RPM values (<300 RPM). This indicates that these sequences in the library subjected to the high stringency conditions had a higher affinity for 4-NP, as they remained bound under the adverse conditions, however, were not favourably amplified under low stringency conditions in comparison.
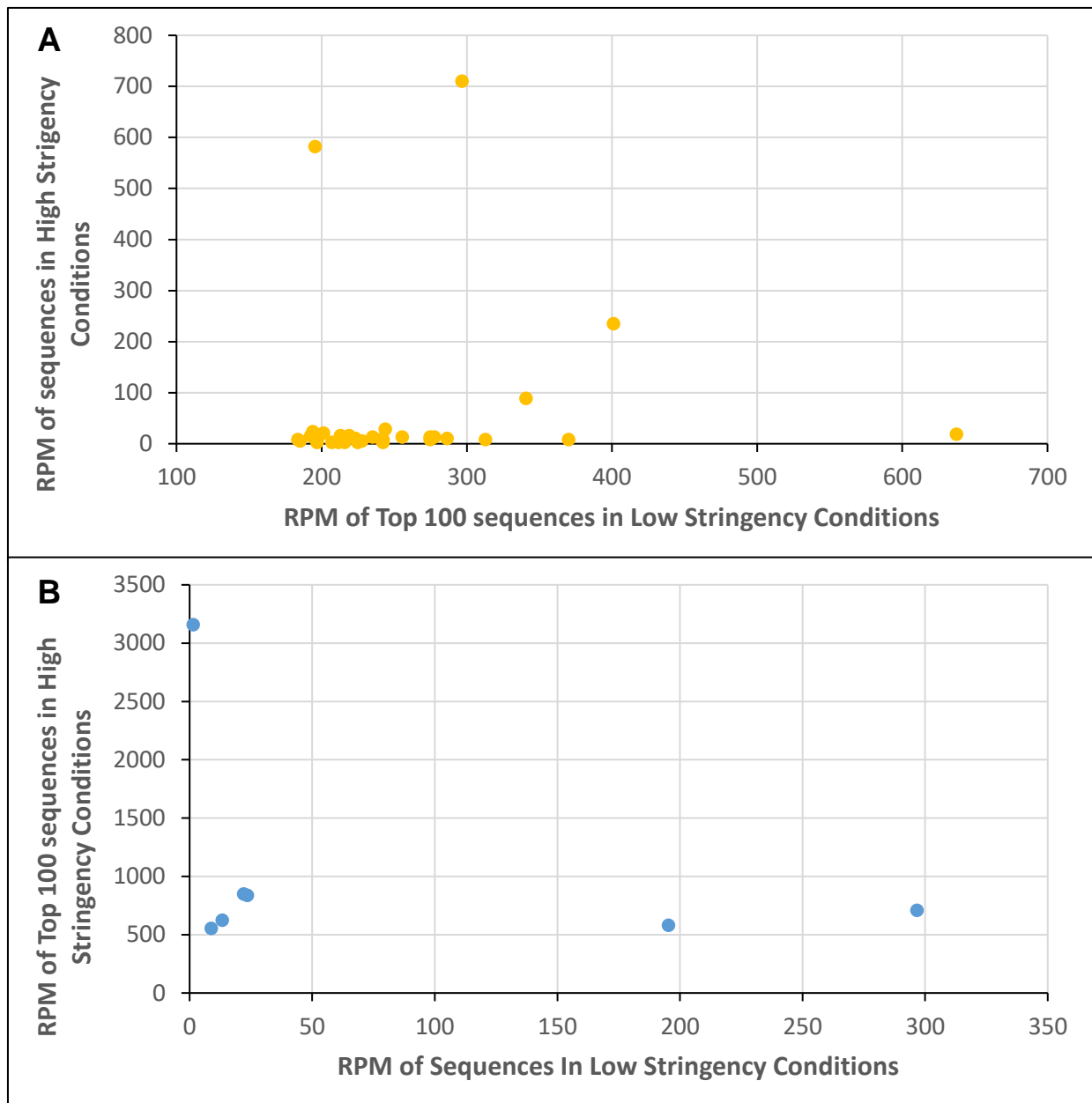
Figure 3.13: The reads per million (RPM) values of conserved sequences after low and high stringency selection in libraries enriched for sequences that bind to 4-NP. A) Fifty sequences were conserved between the high stringency library and the top 100 most abundant sequences in the low stringency library. B) Seven sequences were conserved between the low stringency library and the top 100 most abundant sequences in the high stringency library.

3.3.2 Cluster Data

Clustering refers to the grouping of oligonucleotides based on sequence similarity and allows groupings of oligonucleotide sequences that exhibit different features to each of the other groupings to be identified and investigated. Clustering can provide information such as the

number of unique sequences within a population and help identify primary (parent) and mutant sequences which can infer sequence evolution.

Figure 3.14 shows the numbers of clusters generated from the starting library (grey bar) and from the libraries enriched for sequences that bind to NPTE following low (red bar) or high (orange bar) stringency conditions. The N40 library had the greatest number of clusters, and therefore, unique sequences. This was not unexpected as the N40 library is a random library that is generated to be diverse to increase the likelihood of including sequences that bind well to any target molecule. This was followed by the enriched library from high stringency, then low stringency, conditions. This indicates that more unique sequences were found following high stringency selection. During Selection Rounds 3, 4, and 5 for NPTE-binders, technical faults resulting from a tear in the magnetic pen covering during dsDNA strand separation caused a loss of sequences from the low stringency library. This is likely to have also resulted in fewer unique sequences.
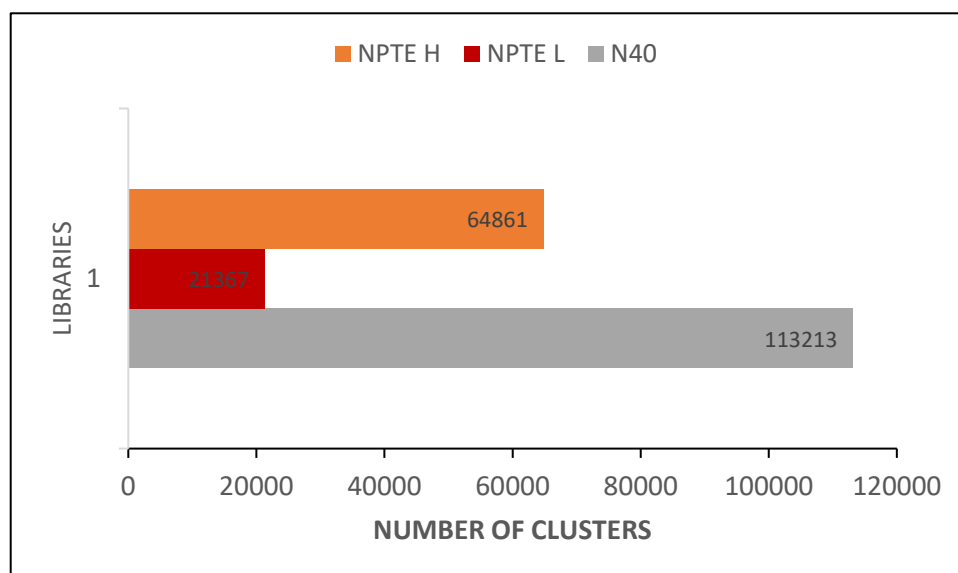
Figure 3.14: The number of clusters in the N40 starting library (grey), and libraries enriched for sequences that bind to nonylphenol equivalents (NPTE) that were selected under low (red) and high (orange) stringency conditions. Clustering was based off a Levenshtien distance of 7.

Figure 3.15 shows the numbers of clusters generated from the N40 starting library and from the libraries enriched for sequences that bind to 4-NP following low (red bar) or high (orange bar) stringency conditions. Overall, the cluster number, therefore, the number of unique sequences decreased as stringency increased. Overall, more clusters were generated in the

library generated under the high stringency conditions for binding to NPTE, compared to clusters generated in the library generated for 4-NP binders (see Figure 3.15).
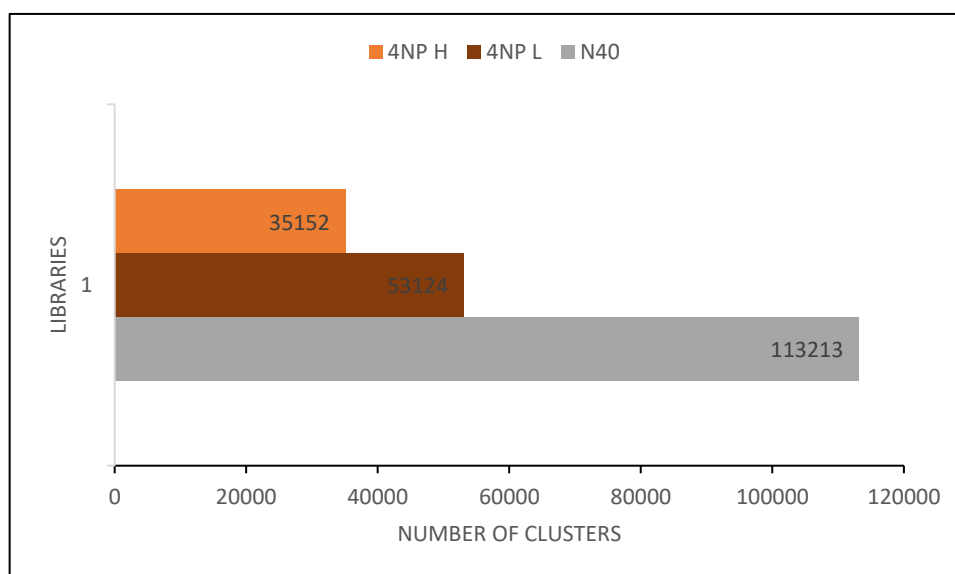
Figure 3.15: The number of clusters in the N40 starting library (grey), and libraries enriched for sequences that bind to 4-nonylphenol (4-NP) that were selected under low (brown) and high (orange) stringency conditions. Clustering was based off a Levenshtien distance of 7.

Following clustering, the top 100 clusters from each sequence library were analysed to incorporate as much information as possible to make meaningful analyses but also accounting for time and clarity due to the magnitude of data that next-generation sequencing provides. The number of unique sequences belonging to each cluster and the total number of reads within the cluster attributed by both the parent and unique (mutated) sequences was determined.

During SELEX, the generation of unique sequences was encouraged through error-prone PCR. The addition of magnesium during PCR causes the error rate of polymerase to rise to result in sequence variants containing random mutations in bases that may influence oligonucleotide sequence binding affinity or specificity.

Both libraries contained the same two sequence outliers that had a high number of total reads, however, were not represented in this graph as it would skew the data. Extremely high read counts comparative to other sequences within the libraries are indicative of selective amplification or high-affinity target-binding. As depicted in Figure 3.16, most sequence clusters in the low stringency library contained fewer than 10 unique sequences and a fewer

number of unique sequences per cluster were present in the high stringency library in comparison. The few unique sequences suggest non-specific binding or sequence variants may have been removed from the selection pool by the stringency measures applied, due to lack of affinity to target.

Figure 3.16 shows lower cluster numbers contain a greater number of unique sequences. As clustering is based on the frequency of the parent sequence, it is possible a highly abundant parent sequence subjected to error-prone PCR results in more variants and unique sequences within each cluster, however, target affinity will influence its persistence across selection rounds.
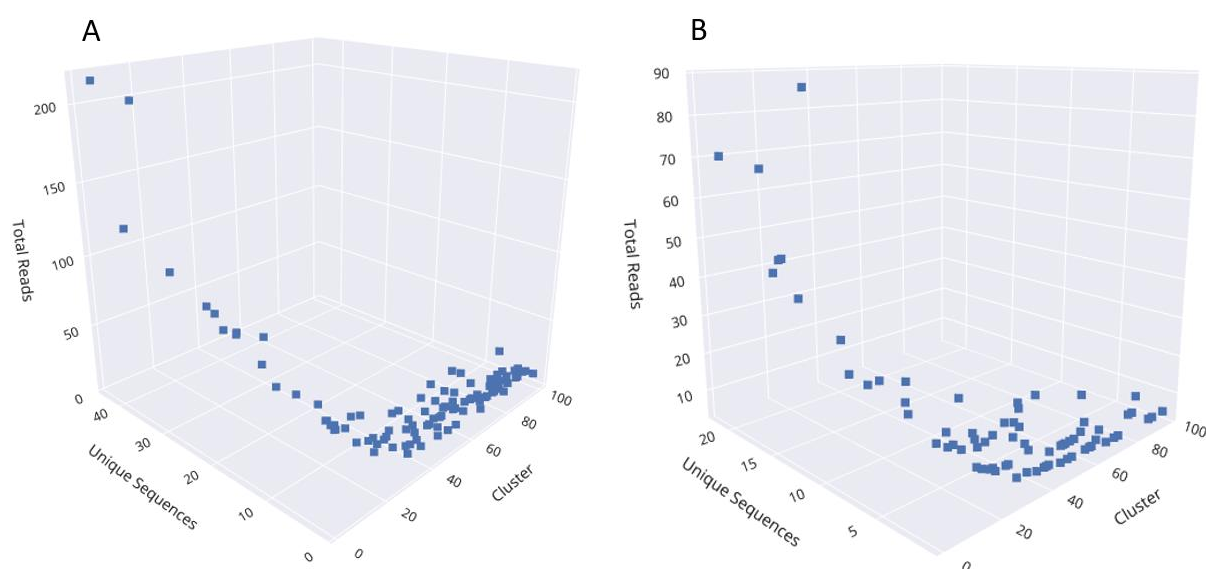
Figure 3.16: The number of unique sequences and total reads for the top 100 clusters in the libraries enriched for sequences that bind nonylphenol technical equivalents (NPTE) under low (A) and high (B) stringency conditions.

Figure 3.17B shows the range of unique sequences was greater in the library enriched for 4-NP binders under high stringency conditions. It is interesting to note that the most frequent sequence in the high stringency library, represented as Cluster 1 (circled in green), only contained two unique sequences. This may suggest that any further mutations of this sequence may not exhibit affinity to the target and therefore, under adverse conditions were removed from the selection pool.

When comparing Figures 3.16 and 3.17, more unique sequences were generated per cluster in both low and high stringency groups in libraries generated towards 4-NP compared NPTE.
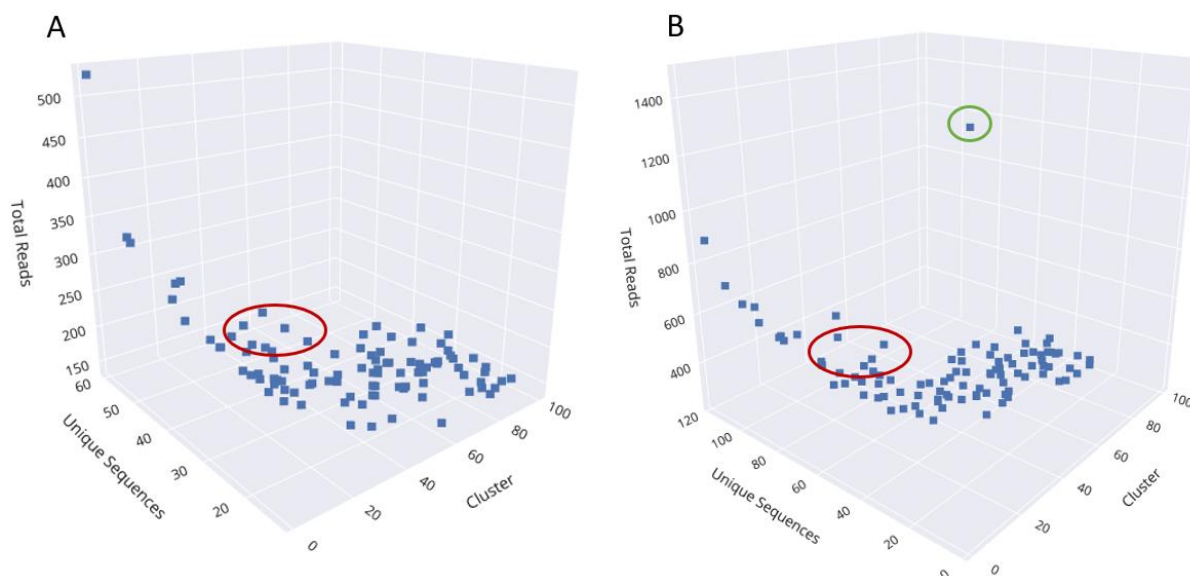
Figure 3.17: The number of unique sequences and total reads for the top 100 clusters in the libraries enriched for sequences that bind 4-nonylphenol (4-NP) under low (A) and high (B) stringency conditions.

The number of sequence reads is influenced by natural PCR selection bias, therefore, sequences with higher affinity to the target may not be represented as the most abundant sequence. Thus, it would be interesting to investigate sequences based on the relationship between cluster number, unique sequences, and total reads. Symbols circled in red in libraries under both low and high stringency conditions are representative of sequences that were present at a lower total read count, however, contained a greater number of unique sequences. This is indicative of sequences that persisted and then mutated, but still exhibited affinity to the target.

3.3.3 FASTAptamer Compare Results

In 4-NP SELEX, 1,949 sequences were conserved in both low and high stringency libraries. Similarly, 1,309 sequences were conserved between low and high stringency libraries following nonylphenol equivalent SELEX. Within each SELEX, sequence libraries were

consistently evolved towards the same target under near-identical conditions excluding detergent concentration and, therefore, a degree of sequence similarity is expected.

Table 3.1: Number of sequences conserved across the N40 library and libraries after low and high stringency selection in single or multi-target SELEX

	4NP L	4NP H	NP-eq L	NP-eq H
4NP L		1,949	30	31
4NP H	1,949		58	41
NP-eq L	30	58		1,309
NP-eq H	31	41	1,309	

Table 3.1 depicts sequences that were conserved in libraries across single- (4-NP) and multi-target (NPTE) SELEX under different stringency conditions. The mix of NPTE would also include 4-NP. Sequences conserved in both the 4-NP and NPTE libraries may bind this target isomer, however, some may also be non-specific binders to the control matrix.

Although it was unclear whether the NPTE conjugated to the beads at the concentration specified, or in fact, if they were even present throughout the selection, evidence from Real-Time PCR and bioinformatics data indicates this may not be the case. Therefore, accurate conclusions based on the assumption of target presence cannot be made.

3.4 Preliminary Experiments for Candidate Aptamer Characterisation using Gold Nanoparticles (AuNPs)

Following the completion of SELEX, characterisation methods were employed to identify target-binding oligonucleotides and assess their binding properties. Surface-modified nanoparticles or AuNPs can be a sensitive method due to their optical properties which allow the assessment of aptamer-target binding through visual or spectroscopic data.

The target molecules, 4-NP and NPTE, require a solvent for solubilisation. No previous literature was found on the effects of organic solvent on the stability of AuNPs or their influence on the sensitivity of aptamer target-binding assays. Therefore, preliminary tests were conducted.

Alterations to the AuNP surface plasmon resonance can be determined visually (through a colour change from red to blue) or spectroscopically using ultraviolet-visible (UV-vis) absorption spectra. Increases in the ratio of absorbance at 625 and 520 nm wavelengths (A625:A520) are indicative of greater amounts of AuNP aggregation. To determine whether solvent interferes with AuNPs in their native state, AuNPs were incubated with increasing concentrations of the organic solvent's ethanol, methanol, and acetone.

Figure 3.18 demonstrates a minimal difference in the A625/A520 ratio in AuNPs resuspended in 5-40 % (v/v) ethanol and methanol. The absorbance ratio was slightly greater in AuNPs suspended in acetone which can indicate AuNPs were more susceptible to aggregation when resuspended with acetone.
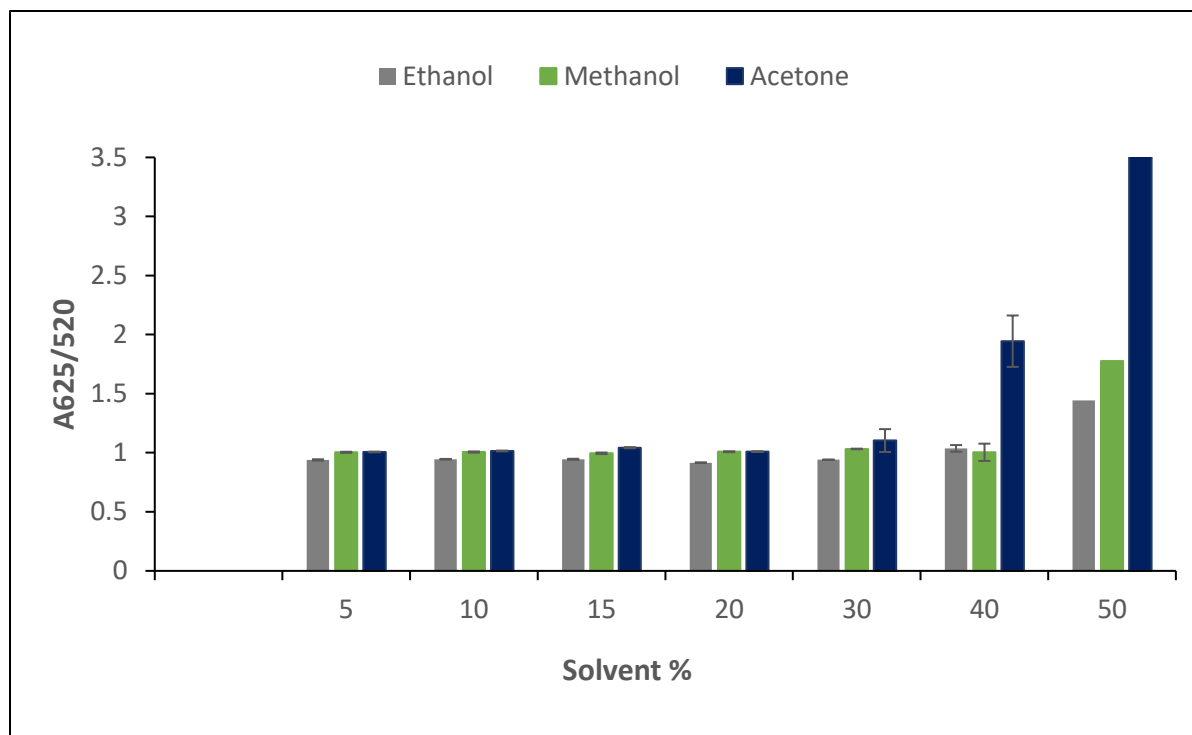
Figure 3.18: The absorbance ratio measured at 625 and 520 nm wavelength of gold nanoparticles (AuNP) resuspended in varying concentrations of ethanol, methanol, and acetone. The error bars represent the standards error of the mean (SEM) of three independent experiments.

For the solvent's ethanol and methanol, there was an increase of A625/520 ratio at a concentration of 50 % (v/v), indicating this concentration was sufficient to destabilise AuNPs. This was consistent with the visual observation demonstrated in Figure 3.19 where AuNPs resuspended in 50 % (v/v) solvent exhibited a deeper red colour, in comparison to that resuspended in lower organic solvent concentrations.

Spectroscopic analysis showed that acetone concentrations above 30 % (v/v) were sufficient to greatly destabilise the AuNPs causing immediate aggregation, as depicted by a sharp increase in 625/520 nm absorption ratio (Figure 3.18). This was also consistent with the visual view in Figure 3.19 where the AuNPs in 40 and 50 % (v/v) acetone exhibited a deep purple colour, indicative of surface plasmon resonance coupling of AuNPs from aggregation.

These results indicated that high solvent concentrations destabilised AuNPs in their native state which led to greater amounts of aggregation.
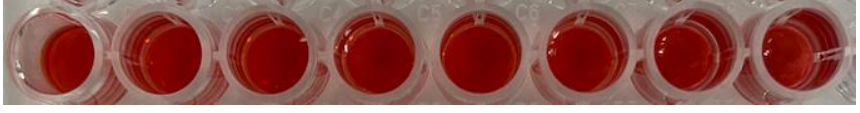
% of Solvent								
Organic solvent	0	5	10	15	20	30	40	50
Ethanol								
Methanol								
Acetone								

Figure 3.19: The visual representation of gold nanoparticles (AuNP) resuspended in varying concentrations of ethanol, methanol, and acetone. The red colour depicts non-aggregated AunPs and the blue colour depicts solvent-induced AuNP aggregation

Oligonucleotides can be characterised using an AuNPs dissociation scheme. In a dissociation scheme, oligonucleotides non-specifically bind to AuNPs, covering their surface and providing some protection against salt-induced aggregation. In the presence of a target molecule, the oligonucleotides with a binding affinity to the target molecule undergo a conformational change and dissociate from the AuNP surface to bind the target leaving the AuNPs exposed to salt-induced aggregation. In the presence of higher target concentrations, more

oligonucleotides will dissociate from the AuNPs surface, resulting in increased AuNP aggregation and a greater 625/520 nm absorbance ratio.

To determine whether differing concentrations of organic solvents interfered with this scheme, a positive control was employed using a validated aptamer that binds the target molecule methamphetamine. Aptamer-coated AuNPs were suspended in varying concentrations of organic solvents and exposed to a range of target concentrations. Based on the data presented above, the maximum concentration of solvent that can be incubated with AuNPs alone is 40 % (v/v) for ethanol and methanol and 30 % (v/v) for acetone.

Target Concentration (µM)					
Solvent + %	0	0.15	0.625	2.5	10
No Solvent					
EtOH 20%					
EtOH 40%					
MeOH 20%					
MeOH 40%					
Acetone 15%					
Acetone 30%					

Figure 3.20: The visual representation of gold nanoparticles (AuNP) coated with aptamer resuspended in varying concentrations of ethanol, methanol, and acetone and incubated with the aptamers target molecule (methamphetamine). The red colour depicts non-aggregated AuNPs and the blue colour depicts aggregated AuNPs.

For the reactions that contained 40 % (v/v) ethanol and methanol, the AuNPs exhibited a deep blue colour across all concentrations of the target molecule (methamphetamine). This suggests that this high concentration of these two solvents caused surface plasmon resonance coupling from AuNP aggregation, independent of the presence of the target molecule. The incubations that contained 20 % (v/v) ethanol and methanol and 30 % (v/v) acetone showed a deeper red colour across all target concentrations compared to the no solvent control, indicating a slight aggregation in these wells. Wells that contained 15 % (v/v) acetone, exhibited a similar colour change to the no solvent control. These observations aligned with spectroscopy data presented in Figure 3.20.

In Figure 3.20, the no solvent control showed a positive correlation between target concentration and absorbance ratio. As the target concentration increased, the 625/520 nm absorbance ratio increased.
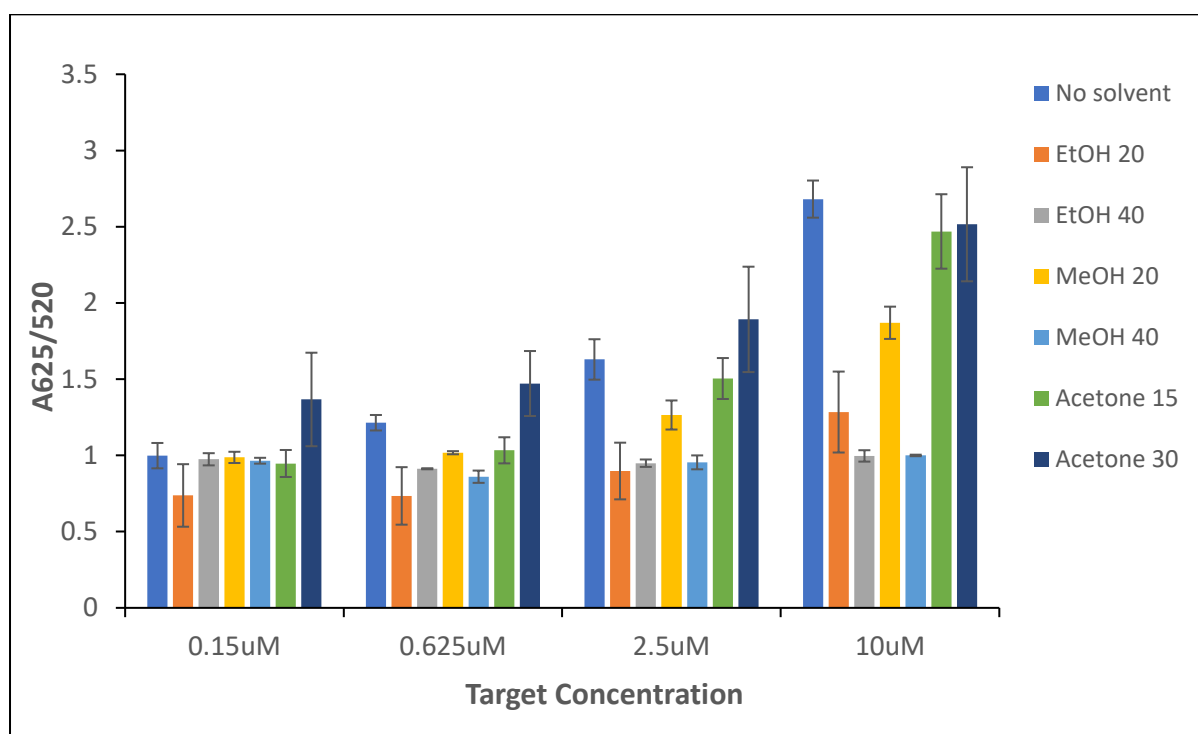
Figure 3.21: Target-induced AuNP aggregation after incubating with various concentrations of target molecule and in varying concentrations (15-40 % v/v) of solvents. Error bars represent the range of data points from the experiment conducted in duplicate. EtOH, ethanol; MeOH, methanol.

In AuNPs re-suspended with 20 and 40 % (v/v) of ethanol and methanol, there was a reduced absorbance ratio compared to the no solvent control in target concentrations above 0.15 μM.

Additionally, despite higher target concentrations, there was only a slight increase in the A625/A520 nm ratio. This suggests that the AuNPs were aggregated by a mechanism independent of the target concentration. This can be explained by the solvent interfering with the protective effect of the aptamer on the AuNP surface, or interference with oligonucleotide folding or target molecule-aptamer binding chemistry.

It was interesting to note that similar to other organic solvents, lower concentrations of acetone (15 % v/v) reduced the A625/A520 ratio compared to that in the non-solvent control and that in 30 % (v/v) acetone. In fact, the latter (*i.e.* 30 % v/v acetone) had a slightly higher A625/A520 ratio, indicating increased sensitivity to target-induced AuNP aggregation.

The solvent concentration range was then lowered to determine a concentration that interfered minimally with the target-binding assay.

Target Concentration (µM)					
Solvent + %	0	0.15	0.625	2.5	10
No Solvent					
EtOH 5 %					
EtOH 10 %					
MeOH 5 %					
MeOH 10 %					
Acetone 5 %					
Acetone 10 %					

Figure 3.22: The visual representation of gold nanoparticles (AuNP) coated with aptamer resuspended in varying concentrations of ethanol, methanol, and acetone and incubated with the aptamers target molecule (methamphetamine). The red colour depicts non-aggregated AuNPs and the blue colour depicts the coupling of surface plasmon resonance from AuNP aggregation.

72

Figure 3.21 illustrates that as the target concentration increased, there was a colour change from red to deep red to blue, which is indicative of greater amounts of AuNP surface plasmon coupling (aggregation). There were minimal visual differences between reactions containing the organic solvent at either concentration and the no solvent control under varying concentrations of the target molecule. This indicated that there were minimal effects of these organic solvents on target-induced aggregation.

The spectroscopy data (Figure 3.22) confirmed the visual colour change data. It showed that similar to the reaction of the no solvent control, there was a positive correlation between target molecule concentration and A625/A520 ratio across all three organic solvents at 5 and 10 % (v/v). However, there was a reduced absorbance ratio compared to the no solvent control in wells containing 10 % (v/v) of ethanol, 5 and 10 % (v/v) of methanol, notably in the reactions with target concentrations above 0.625 µM. This indicated that these solvent concentrations resulted in a decreased sensitivity of the assay through solvent induced destabilisation of aptamer- AuNP adsorption.
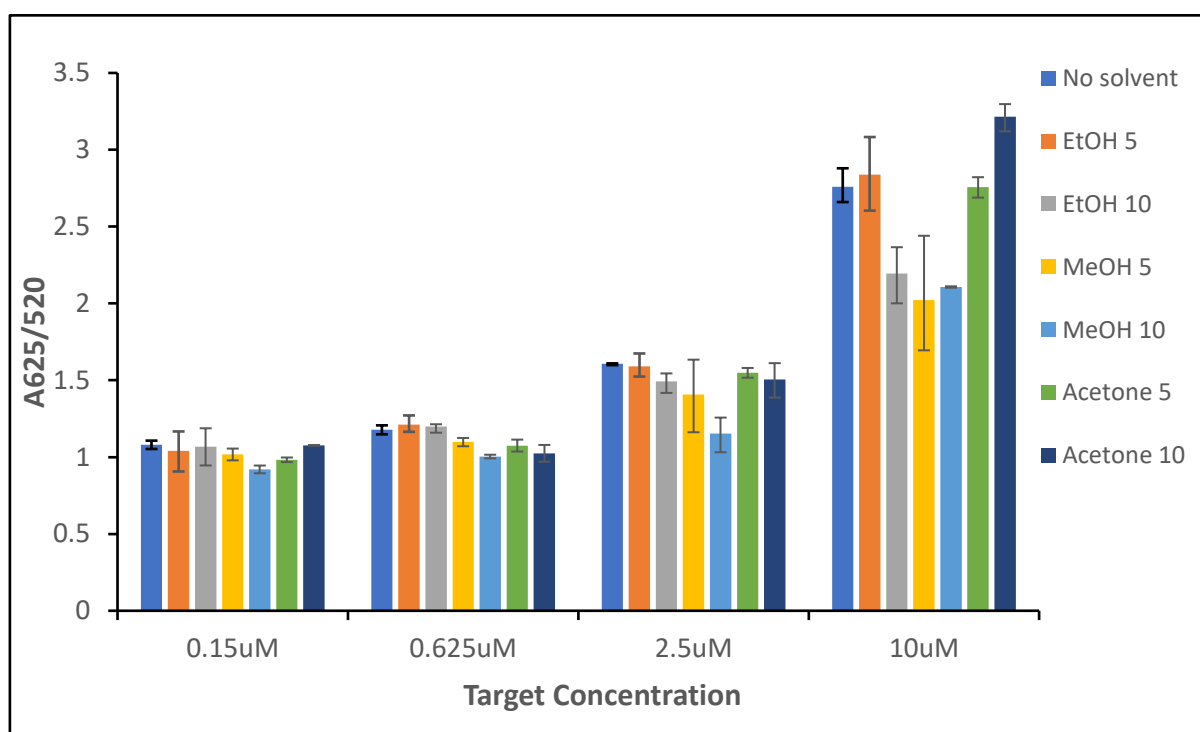
Figure 3.23: AuNP aggregation after incubating with various concentrations of target molecule and in lower concentrations (5 or 10 % v/v) of solvents. Error bars represent the range of data points from duplicate experiments. EtOH, ethanol; MeOH, methanol.

The AuNP that were re-suspended in 5 % (v/v) ethanol and 5 to 10 % (v/v) acetone exhibited an A625/A520 ratio similar to that for the no-solvent control, across the increasing target

molecule concentrations. This indicated a minimal effect of the solvents at these lower concentrations on salt-induced aggregation following aptamer-target-binding.

It was interesting to note that the addition of 10 μM of the target molecule (methamphetamine) to the well containing 10 % (v/v) acetone resulted in a greater absorbance ratio indicating the AuNPs may be more sensitive to salt-induced aggregation in the presence of acetone. Following the assessment of all of this optimisation data, the solvent selected to dissolve 4-NP and NPTE for the aptamer characterisation experiments was 5 % (v/v) of ethanol.

3.5 Candidate Aptamer Characterisation using Gold Nanoparticle Assay

Forty oligonucleotides selected from the bioinformatics data were assessed for binding activity through a medium-throughput screen using the AuNP assay. In congruity with the dissociation scheme, AuNPs were incubated with candidate 4-NP and NPTE aptamers and exposed to increasing concentrations of the target molecule(s), followed by salt-induced aggregation. All forty oligonucleotides were initially screened using a wide target range and assessed spectroscopically for binding activity. From this, candidate aptamers were selected for further validation. These are shown in Figure 3.23. See Appendices E.1 and E.2 for more information.
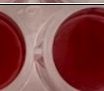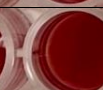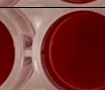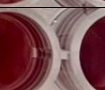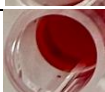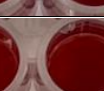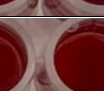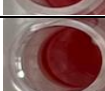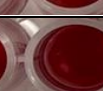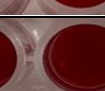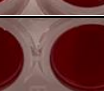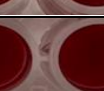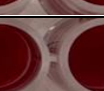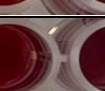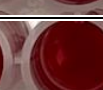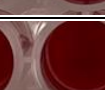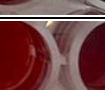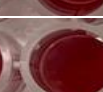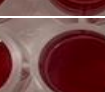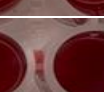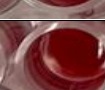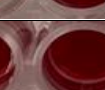
Final Target Concentration In Wells (μM)								
Oligo name	0	0.0016	0.008	0.04	0.2	1	5	25
NS								
NPTE L2								
NPTE L3								
NPTE L4								
NPTE L5								
NPTE L6								
NPTE L11								
4-NP L1								
4-NP L3								
4-NP L6								
4-NP L7								
4-NP H4								
4-NP H10								
4-NP H11								

Figure 3.24: The visual representation of gold nanoparticles (AuNP) coated with aptamer candidates from the initial screen. The red colour depicts non-aggregated AuNPs and the deeper red-blue colour depicts greater amounts of AuNP aggregation. NS represents the non-specific oligonucleotide control; NPTE denotes nonylphenol technical equivalents, 4-NP denotes 4- nonylphenol, L and H denotes aptamer candidates were from libraries subjected to low and high stringency strategies.

The AuNPs that had oligonucleotides bound and which exhibited a protective effect from salt-induced aggregation remained red with little change in absorbance ratio despite increasing concentrations of the target. As depicted in Figure 3.23, the wells that contained the non-specific oligonucleotides remained red across all target concentrations.

Additionally, some oligonucleotides illustrated in Figure 3.23 demonstrated aggregation or a gradient of aggregation (e.g. NP-eq L6, 4-NP H4) in higher target concentrations, compared to the non-specific oligonucleotides and no target wells. Therefore, the AuNP characterisation of these candidates was repeated, and the spectroscopic data is displayed in Figures 3.24 and 3.25.
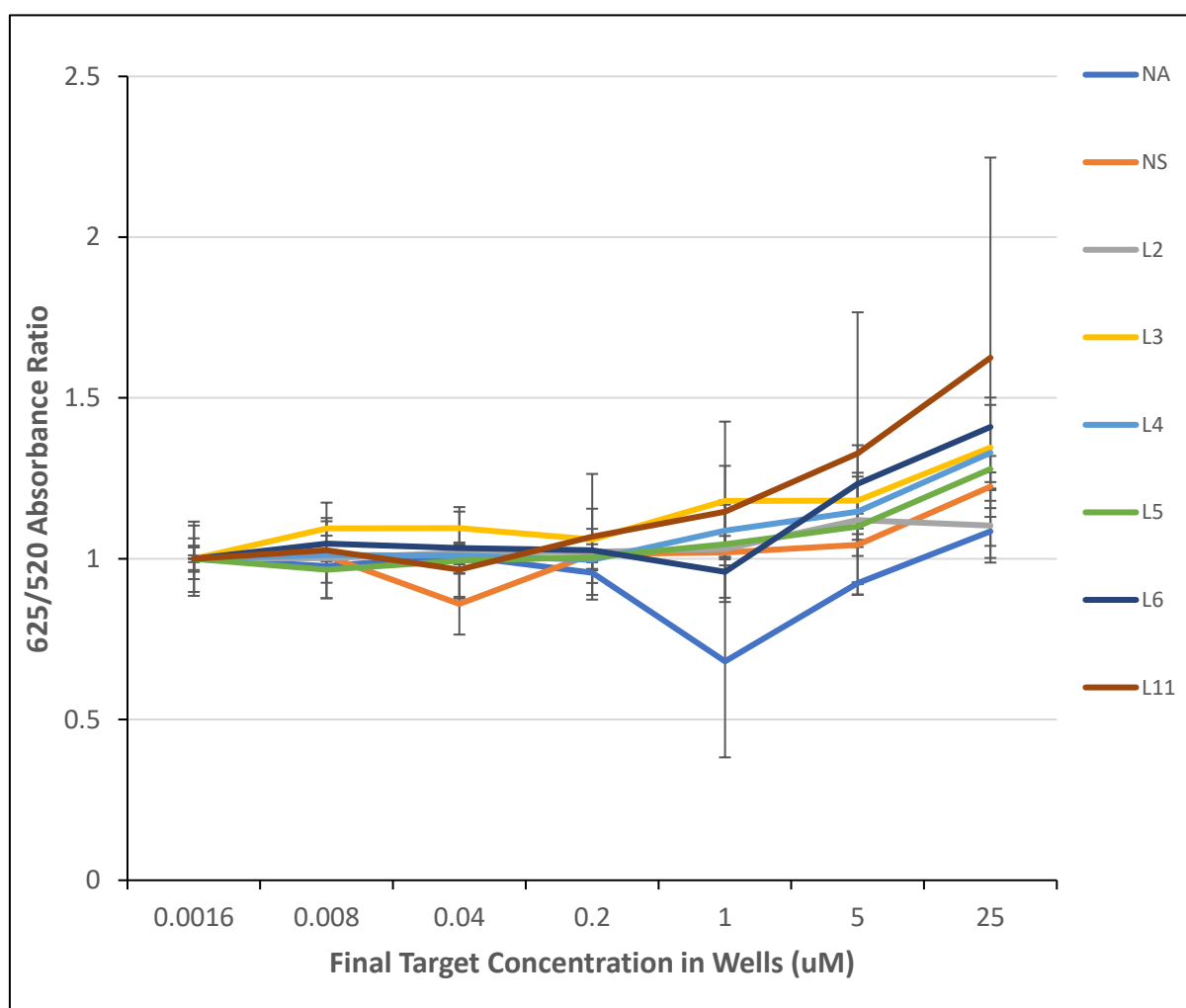
Figure 3.25: Salt-induced gold nanoparticle (AuNP) aggregation after incubating with various concentrations of target molecule (nonylphenol technical equivalents; NPTE) as measured by absorbance ratio at 625 and 520 nm wavelength. NA and NS represent the no aptamer and non-specific aptamer control, respectively. Error bars represent the SEM of three technical replicates.

As demonstrated in Figure 3.25, there was minimal difference between the A625:A520 ratio of AuNP incubated with the NPTE aptamer candidates L2, L3, L4, L5, L6 and L11 oligos compared to the non-specific aptamer control. Of these NPTE aptamer candidates, those that do present higher ratios across some of the different target molecule (NPTE) concentrations such as L3, L6 and L11, exhibit inconsistent results as demonstrated by the large error bars and therefore, formal conclusions regarding binding activity cannot be made. In retrospect, it was found that 8 sequences in the NPTE low and high stringency libraries that were characterised were the same. Thus only 12 unique sequences total from libraries enriched with sequences that bind NPTE were assessed. Conserved sequences are as follows; NPTE L1, L2, L4, L3, L5, L6, L8 and L11 were equal to NPTE H1, H2, H3, H4, H7, H5, H13, H9 respectively.

The AuNP incubated with the 4-NP aptamer candidates L1, L3, L6, L7, H4, H10 and H11, demonstrated a similar A625:A520 ratio across increasing concentrations of the target molecule (4-NP) compared to the non-specific aptamer control (Figure 3.25). This suggested that these oligos did not bind the target molecule

There was an increase in AuNP aggregation depicted by a higher A625:A520 ratio at 25 µM of 4-NP, however, this was also occurred in the non-specific aptamer control suggesting that the concentration of target alone contributed to AuNP aggregation.
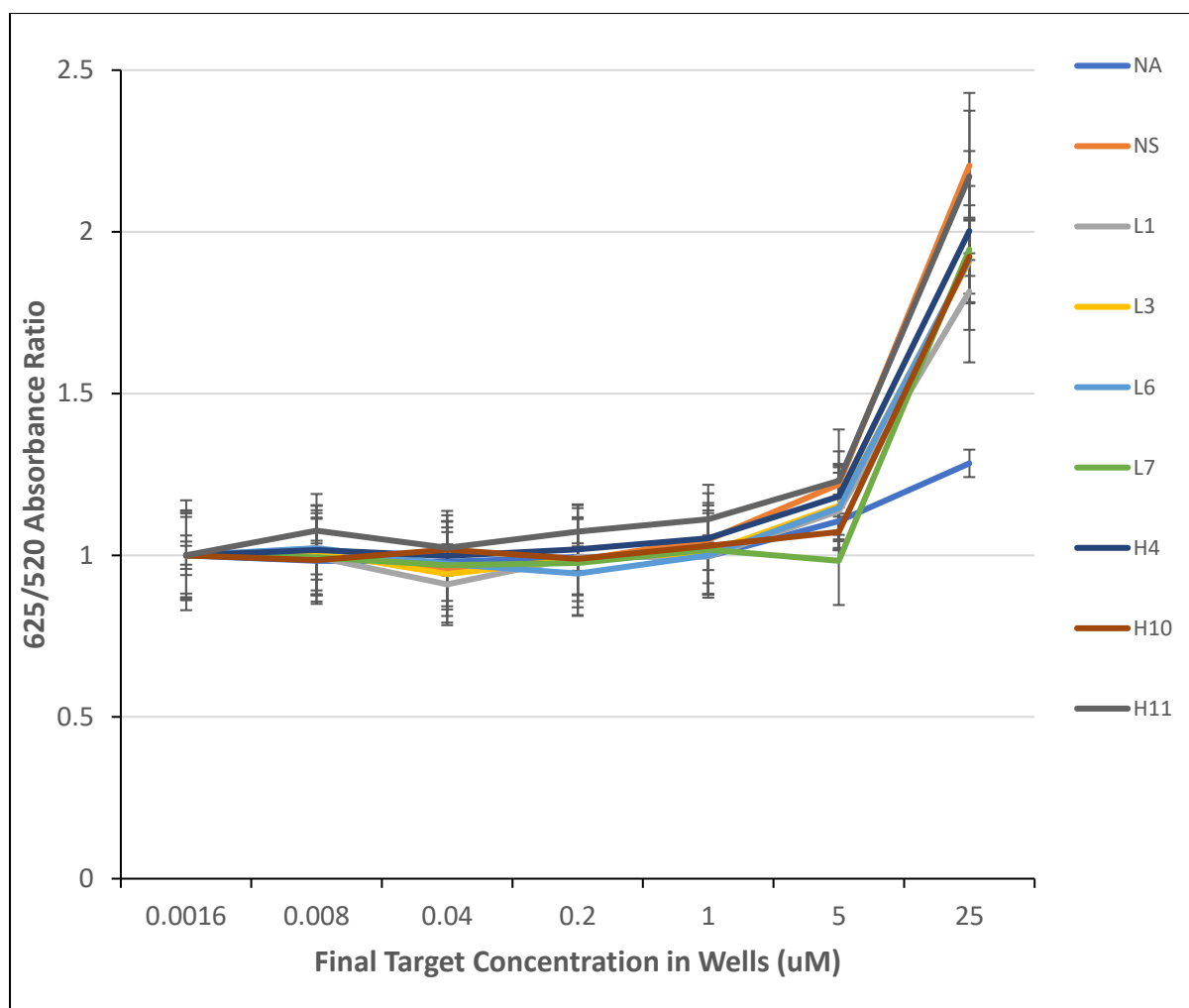
Figure 3.26: Salt-induced gold nanoparticle (AuNP) aggregation after incubating with various concentrations of the target molecule (4-nonylphenol; 4-NP) as measured by absorbance ratio at 625 and 520 nm wavelength. NA and NS represent the no aptamer and non-specific aptamer control, respectively. Error bars represent the SEM of three technical replicates.

Whilst the initial oligonucleotide candidate screening using the AuNP assay looked promising, however, data from technical repeats suggested this was not the case. The error bars presented in Figures 3.24 and 3.25 can be explained by analytical errors such as pipetting volume, inherent variation between different batches of nanoparticles or artefacts of the assay.

Chapter 4. Discussion

In this study, SELEX was performed to generate aptamers for 4-NP by incubating a random nucleotide library with the single-molecule of 4-NP or with a mix of NPTE (which also contains 4-NP). In each SELEX experiment, seven selection rounds that introduced a low or high stringency strategy in the partitioning steps were performed. As a result, four enriched oligonucleotide libraries were sequenced using HTS by Illumina. Subsequently, twenty aptamer candidates from SELEX experiments against each target molecule, 4-NP and NPTE, were selected following bioinformatics analyses and characterised using a modified AuNP assay. Unfortunately, the generation of an aptamer that was specific to 4-NP and or any of the molecules in the NPTE mix was not identified using the AuNP characterisation method. This may have been due to specific features of the target molecule(s), the selection methodology or the AuNP assay being an unsuitable characterisation method.

Aptamer-based technology if successful can have endless applications and its advantages are great. However, aptamer technology has often been described as a black box referring to the many uncertainties on how to generate candidates successfully binding to their target molecule within each selection experiment. In saying this, understanding aptamer development is vital for future growth and success of the field and the advancement of their applications. Thus, despite not identifying a binding aptamer to the target molecule(s) 4-NP and NPTE, many insights from this work may be made.

4.1 Modifications to Aptamer Selection Strategies

Since the first *in vitro* protocol for the development of DNA bioreceptors in 1990, there has been many studies and alterations to the original named "traditional" SELEX method to develop aptamers to a range of target molecules for various applications (Yang et al., 2011). The modifications made were largely to enhance either the suitability of the oligonucleotide binding properties for its end application, reduce the selection time frame or strengthen desirable aptamer characteristics such as increased affinity, selectivity, or specificity.

This work aimed to investigate modifications made to the selection strategy by increasing the stringency conditions and the presence of a single, versus multiple, target(s) on the

enrichment of the resultant oligonucleotide pool. Stringency is described as any selection pressure used to increase target binding affinity and specificity while reducing the number of selection rounds required. In this study, a range of detergent (Tween 20) concentrations from 0.01 to 0.05 % (v/v) and from 0.1 to 0.5 % (v/v) were used to represent low and high stringency conditions, respectively.

4.1.1 Stringency and DNA quantity

In the single-target SELEX towards 4-NP, it was found that increasing stringency led to a decreased quantity of DNA recovered after elution steps in each selection, as illustrated by the higher stringency samples having higher Ct values in each amplification round as seen in figure 3.9. It is likely the higher detergent concentrations interfered with lower levels of binding affinity leading to the removal of sequences that may have persisted in less adverse conditions. The difference in the amount of DNA recovered following low and high stringency washes decreased across selection rounds and the resultant libraries converged notably at Round 6 of SELEX. It is possible that certain thresholds may exist where the presence of certain detergent concentrations will remove non-covalent bonding to a similar extent.

It was also found that compared to SELEX Round 6, an increased quantity of DNA was recovered at SELEX Round 7 following both low and high stringency selection conditions despite a greater detergent concentration. Generally, as selection rounds progress, the oligo pool evolves, and contains fewer unique sequences, as non-specific sequences are removed leading to a pool enriched with sequences that bind the target molecule (Stoltenburg et al., 2007). It is likely that sequences with lower levels of binding were removed in previous rounds and thus, only sequences with a high-binding affinity, and specificity to the target molecule persisted. The introduction of stringency measures and negative selection steps in Rounds 4 and 6 may have greatly contributed to reducing non-specific binders. In future, it would be interesting to increase the number of selection rounds to observe whether this trend continued and whether it is indicative of convergent evolution towards high-affinity aptamers. If this were the case, we would expect a greater quantity of DNA to be recovered from both conditions following Selection Round 8.

Dissimilarly to the observations made for the SELEX against 4-NP, the effects of stringency on the library incubated with the multiple target mix (NPTE) were variable and no obvious trend was observed. However, technical issues were experienced in Selection Rounds 1 to 5. Specifically, it was retrospectively discovered that the magnetic pen used to assist the strand separation was faulty and may have led to a loss in DNA recovered in each round. Moreover, the conjugation assessment of NPTE to the affinity matrix was inconclusive. Thus, no accurate conclusions can be made regarding stringency effects on the quantity of DNA recovered within these selection rounds.

4.1.2 Stringency Effects on Libraries from HTS Data

In this study, HTS was used to further investigate the impact of low and high stringency selection pressures on the resultant oligonucleotide libraries from single versus. multiple target SELEX. The sequenced libraries were analysed through the FASTAptamer pipeline and information including clustering, rank, reads, RPM and enrichment values were determined.

In libraries enriched for sequences that bind 4-NP, high stringency selection was found to decrease the total number of unique sequences, compared to low stringency selection. Specifically, 35,152 and 53,124 clusters were present in the high and low stringency libraries, respectively. These observations, that greater stringency conditions in the form of detergent concentration resulted in fewer clusters and thus fewer unique sequences in the final oligonucleotide library, align with the expected effects of stringency on the selection pool. In previous studies, different stringency measures have been introduced in experiments in different forms such as decreasing target molecule concentrations, decreasing the incubation time of the library-target molecule interactions or increasing the number of wash steps to accelerate the level of competition amongst aptamers, all to generate fewer sequences with higher affinity by surviving in adverse conditions (Yu et al., 2021).

In contrast, the high stringency measures used on the library enriched for sequences that bind NPTE resulted in a greater number of unique sequences, in comparison to the library generated under low stringency conditions. This reveals that more variant sequences were generated and conserved throughout the high stringency selection measures. This result is most likely a repercussion of technical errors endured during the selection process. Despite

this, a greater number of clusters (64, 861) were conserved in the high stringency library for the mix of NPTE, compared to both low and high stringency libraries generated against 4-NP. The NPTE contain a large mix of isomers, thus a broad range of sequences would have been selected, in comparison to 4-NP where only one isoform target was present. However, as the conjugation assessment was inconclusive, we cannot definitively say the target molecule(s) were present in the selection and thus, no accurate conclusions can be made based on this assumption. In future, the conjugation assessment should be repeated.

Conserved sequences were found within and between the libraries generated against 4-NP and NPTE. Shared sequences between low and high stringency libraries reflect parallel sequence evolution. In these experiments, sequences were consistently evolved towards the target molecules(s) under near-identical conditions excluding the differing detergent concentrations. As mentioned previously, the NPTE contains a mix of isomers including the isomer 4-NP. It is possible that any of the shared sequences between these libraries will bind this specific isomer. Alternatively, and less ideally, this observation may be the result of non-specific sequences persisting, such as those that bind the Sepharose beads. For chemically anchored target molecules, the presence of an affinity matrix is problematic as non-specific binding can obstruct part of the target molecule and thus, hinder the evolution of target-binding sequences. The negative selection round was performed to remove non-specific binders, however, the efficiency of this is always unknown and non-specific sequences may have persisted in the pool. There is some evidence that the negative selection was effective as the DNA yield decreased in libraries generated under both low and high stringency conditions following this step. However, non-specific sequences may bind matrices in structure-dependent or structure independent modes where sequences are captured by matrices by electrostatic absorption. If the latter scenario occurs, negative selection is not as effective as anticipated (Wang et al., 2019). To confirm this, the negative selection libraries should be sequenced.

Interestingly, the bioinformatics data from the libraries generated against 4-NP showed that within the top 100 clusters resultant from the high stringency conditions, a greater range of unique sequences was present, compared to that subjected to the low stringency conditions. The accumulation of mutations through error-prone PCR should increase the library diversity

and influence the binding characteristics of the oligonucleotides (Bittker et al., 2002). As clusters from high stringency contained a greater number of unique sequences, this suggests that the mutations that occurred either enhanced or at least did not reduce their affinity for the target molecule. This excluded Cluster 1 which contained only two unique sequences indicating that the variants most likely do not exhibit affinity to the target molecule and may be the result of PCR bias and not target molecule binding. The top sequences in the high stringency library showed greater RPM values which represents a greater frequency and abundance within the population, compared to sequences in the low stringency library. It has been reported that the removal of competition from non-specific or low-affinity binders means high-affinity binders can become enriched at a greater rate within the library (Kohlberger & Gadermaier, 2021).

Within the libraries generated against NPTE, the majority of the top 100 clusters in both low and high stringency libraries contained fewer than ten unique sequences. Fewer unique sequences might indicate that either variant sequences were removed from the selection pool due to lack of affinity and increasing stringency, an inefficient separation of bound and unbound sequences or defective negative control step, or multiple targets resulted in a large number of sequences being carried over in each selection round resulting in a reduced opportunity to mutate and persist. However, to confirm these hypotheses, future studies should ensure that a proportion of each library pool is sequenced following each selection round to examine sequence evolution, compared to the final library which was used in this study.

4.2 AuNP Characterisation

To conclude this research, an AuNP assay was used to characterise ten sequences from each of the four SELEX libraries. These sequences were the parent sequences of the top ten clusters based on total sequence reads (the total reads of the parent sequence and all unique sequences within that cluster). It was retrospectively found that 8 sequences from each of the low and high stringency libraries containing sequences enriched for NPTE were conserved. Thus 20 unique sequences from the 4-NP libraries and 12 unique sequences from NPTE libraries were characterised. In future, if I had more time, I would reselect additional sequences that did not overlap and assess these for target binding for a greater range.

AuNPs exhibit unique properties that enable them to act as sensitive detectors, via changes in surface plasmon resonance which can be quantified spectroscopically through UV-vis absorption spectrum or visually through a colour change observed. In a commonly used format and the one employed in this study, termed the dissociation scheme, oligonucleotides non-specifically bind the AuNP surface exuding a protective effect from salt-induced aggregation. In the presence of the target molecule(s), oligonucleotides that exhibit target molecule binding properties dissociate from the AuNP to bind the target molecule, causing AuNP aggregation and coupling of surface plasmon resonance following the introduction of salt. For both 4-NP and NPTE oligonucleotide-binding candidates, the initial AuNP characterisation results suggested that some candidates exhibited target molecule-binding as demonstrated by a colour change and an increase in the A625/A520 ratio. However, upon repeated characterisation studies, no significant absorbance changes were observed, compared to the non-specific aptamer control, and thus no aptamers for NP were identified. The variation observed in absorbances between triplicate experiments was large suggesting inconsistent results. This may be due to pipetting error, artefacts of the assay, or batch-batch variation from AuNP synthesis. Adjustments to AuNP synthesis such as pH and reagent concentration can greatly alter the properties of AuNPs (Kimling et al., 2006). One study found gold concentrations above 0.8 mM led to the destabilisation of the AuNP particles (Schulz et al., 2014). In this study, a concentration of 0.8 mM was used to ensure a sufficient concentration of AuNPs were produced. This is at the upper limit of what is recommended which may have resulted in less consistency between AuNP batches.

A general increase of the A625/A520 ratio and aggregation existed when greater concentrations of the target molecule were present, notably 25 µM as observed in both oligonucleotide candidate and non-specific aptamer control. There is an intricate balance between oligonucleotide concentration and the amount of AuNP. If an excess amount of oligonucleotides are present, then the AuNP will reach saturation and the remaining oligonucleotides will remain in solution, free to bind the target resulting in an inaccurate representation of target-binding. Inversely, a deficiency of oligonucleotides bound to the AuNP would leave the surface area exposed to effects from the target molecule or unprotected making it susceptible to salt aggregation. It is possible the target molecule(s) were in excess concentrations and influenced the stability of the AuNPs resulting in

aggregation observed across control and test lines. Therefore, the oligonucleotide concentration should be optimised before making further conclusions.

Due to the results produced from the AuNP characterisation herein, this study can't definitively conclude on the affinity and specificity of sequences produced from the low and high stringency strategies or the multiple versus single target molecule strategy. To draw clear conclusions on aptamer characteristics from resultant oligonucleotide pools, additional or alternative characterisation methods must be employed.

4.3 Limitations for Successful Aptamer Identification

4.3.1 The Nature of Target and Selection Methodology

The main drawbacks for the successful identification of oligonucleotides that bind the target molecule(s) 4-NP and NPTE, may be attributed to the nature of the target molecule(s) and the selection methodology used. The immobilisation of the target molecule(s) onto an affinity matrix was implemented for easy partitioning of sequences and to assess the effect of stringency introduced in wash steps on the oligonucleotide pool. 4-nonylphenol is a small molecule and has only one functional group, a hydroxyl group, that was used in chemical conjugation to the affinity matrix (see Methods 2.4). More functional groups, epitopes and structural motifs give rise to more complex interactions through hydrogen bonds, hydrophobic and electrostatic interactions (McKeague & Derosa, 2012). Thus, the occupation of the only functional group limits the potential interactions between oligonucleotides and the target molecule. Additionally, immobilisation of the target molecule via the functional group restricts binding due to a decrease in accessible surface area, increased steric hindrance and lowering the number of epitopes available for oligonucleotides to bind. It is possible that the nature of the target molecule having few epitopes for more complex levels of binding prevented specific binding to occur. However, the results from this study demonstrated an increased quantity of DNA in SELEX round 7 following a harsh negative selection step in round 6, indicating retained sequences. Based on these results, it is unlikely that the oligonucleotide libraries were purely derived from non-specific binders.

Immobilising the target molecule(s) for the purpose of SELEX means the target molecule is presented in a chemically modified form and different from its native state. Chemical

alterations can influence the physiochemical properties of the target such as charge distribution. The characterisation method employed here involved the target(s) free in solution. As no binding candidates were identified through this method, it is possible that the chemical modification of the target molecule(s) during selection compromised the binding interaction between the oligonucleotides and the free target molecules. Moreover, aptamers may partially bind the matrix, therefore, the removal of the matrix results in the aptamers having reduced functionality in its applications (McKeague & Derosa, 2012), despite negative selection steps. Studies have reported aptamers that have reduced or no affinity for free target molecules when originally developed for immobilised small molecule targets (Yu et al., 2021). Selection methods such as capture SELEX and graphene-oxide SELEX have been developed to circumvent these problems associated with target immobilisation. Sequences that are more compatible with characterisation techniques and applications are selected, such as biosensor platforms that rely on structure-switching for signal transduction (Lyu et al., 2021). Therefore, alternative methods may be explored in future research to eliminate the need for target modification.

4.3.2 PCR Bias

One factor that is crucial to the success of SELEX is the enrichment of the libraries with sequences that bind the target molecule. Seven rounds of PCR amplification were completed in this study and candidates for characterisation were chosen based on total cluster frequency. The limitation of this is that sequence frequency may be a result of PCR bias rather than target molecule affinity. An inherent bias exists where sequences that are amplified preferentially under PCR conditions become overrepresented in DNA libraries and this perpetuates as more cycles continue. Multiple different mechanisms exist that can introduce bias. For example, due to the heterogenicity of the DNA libraries, amplification may form by-products, primer dimers and product-product hybridisation. In addition, sequences that are rich in G-C bases are generally less accessible and therefore sequences are not equally amplified (Kohlberger & Gadermaier, 2021). Steps were taken to promote bias-free amplification including monitoring the reaction using Real-Time PCR to ensure amplification was stopped at the exponential phase before a shift in reaction conditions. However, the hypothesis that some or many of these candidates may be a result of poor amplification

efficiency cannot be ruled out. A clear example of this was that the top two sequences from the library enriched with sequences that bind NPTE for both low and high stringency strategies had RPM values far greater than the rest of the oligonucleotide pool. Studies have reported that the most enriched sequences in aptamer libraries do not accurately predict aptamer affinity (Dao et al., 2016). Through HTS analyses, one study found that high performing sequences were enriched in earlier selection rounds, then were outcompeted by weaker-performing sequences with greater amplification efficiency (Kinghorn et al., 2017). In future to avoid this, it may prove useful to reduce the number of SELEX rounds. Employing alternative PCR techniques to promote better amplification efficiency of complex libraries such as emulsion or droplet PCR can also be employed. Moreover, sequencing libraries after each selection round can help monitor shifts in population and determine if sequence enrichment is due to target-affinity or PCR bias.

4.3.3 Bioinformatic Analyses

The limitations in computational modelling should also be considered while assessing experimental results. The bioinformatics pipeline used in this study was FASTAptamer which was primarily concerned with frequency and enrichment and does not account for motifs or secondary structure predictions. Thus, although HTS data provided insightful information to compare sequence libraries from different selection conditions, this was purely based on primary sequence information and oligonucleotides selected for characterisation were based on frequency counts. Alternatively, capturing different aspects of aptamers including sequence or structural motif searches may prove useful in identifying binding candidates as highly conserved regions within a population can indicate key areas for target-binding (Beier et al., 2014). A more comprehensive bioinformatics analysis accounting for sequence or structural motifs within or between oligonucleotide populations has been demonstrated as an effective approach to identifying binding candidates. One study revealed a binding aptamer in one SELEX round from sequence motifs based on secondary structure, removing the need for further laborious selection rounds and the limitations that come with this (Kinghorn et al., 2017). Available aptamer-specific software packages analyse different aspects and no gold standard pipeline exists to process HTS data. Thus, caution must be used when selecting certain pipelines depending on the information desired.

4.3.4 Characterisation Method

The characterisation method employed to identify target-binding oligonucleotides may have prevented the successful identification of aptamers due to the nature of the assay. The AuNP dissociation scheme depends on two variables, namely aptamer binding and structure switching. Structure switching is the basis of many aptamer-based biosensors as a structure switch induced by target recognition can be converted into a detectable signal. However, the selection method employed in this research did not select for structure-switching abilities. Affinity matrix SELEX is non-discriminatory, and the library will contain a pool of all types of binding sequences including those that fold before binding and those that structure switch in the presence of the target molecule. The problem with the former is that sequences remain non-specifically bound to the AuNP surface whilst binding target(s) and thus won't adsorb from the AuNP surface. It is possible that the oligonucleotides present in the assay were able to bind the target molecules(s), however, an affinity for the target molecule could not be detected due to the assays requirement for structure switching molecules.

Although previously, AuNP assays have been used to screen for different characteristics such as binding and specificity for the target molecule, it is prone to producing erroneous results (Kwon et al., 2015). Many variables within the assay can cause non-specific AuNP aggregation such as the ratios between an aptamer-AuNP-target molecule, the structure of the aptamer, properties of the target molecule, and solution composition. Additionally, the binding affinity of the aptamer can be masked by the incomplete dissociation of aptamers from the surface of AuNPs (Thevendran & Citartan, 2022). Moreover, the effects of solvent on the assay performance was unknown and it was possible that it could influence interactions between the aptamers and the target molecule or aptamers and AuNP. Numerous controls must be put in place and vigilance should be exercised when interpreting results. Initially, ITC was intended to screen aptamer candidates for this project, which would have circumvented the limitations of the AuNP assay however, unfortunately, the ITC was having technical issues and due to time constraints, had to be abandoned.

It is likely that the lack of binding candidates identified was the result of limitations in the selection or characterisation methodology. This raises the issue that many aptamer researchers face which is that the selection environment is often different from the

88

characterisation and application environments. Thus, oligonucleotides may behave differently in different conditions. To vastly improve screening methods for assessing aptamer candidates for future research, alterations in the selection methodology which reflect the chemistries of the intended characterisation assay or end applications need to be considered.

4.4 Future Directions

In future research to evaluate the binding characteristics of aptamer candidates from different selection libraries, I would use alternative label-free methods of characterisation that are not dependent on structure-switching capabilities. ITC is a robust and sensitive characterisation method that would provide an accurate thermodynamic profile of the molecular interactions that occur between the aptamer and target molecule and would circumvent the problems associated with candidate screening using an AuNP assay. Originally ITC was intended for use in this study, however, due to technical issues and time, it was not utilised, but is something I would prioritise in the future. The gold standard in aptamer technology is to utilise two characterisation methods to validate aptamer binding. If binding was observed using ITC, I would apply this rationale to further validate promising aptamer candidates.

Once target-binding abilities were confirmed, the next step would be to explore aptamer-based biosensor platforms. The main platforms include colorimetric, fluorescent and electrochemical sensors. Fluorescent aptasensors have been developed to compounds that are structurally similar to NP including E_2 and BPA. These aptasensors exhibit analyses time under two hours and high sensitivity as illustrated by their limit of detection of 3.5 x10^{-10} and 4.1 x10^{-10} M, for E_2 and BPA, respectively (Kim et al., 2019). The former aptasensor has been applied to successfully determine targets in biological, environmental and food samples (Zhang et al., 2018). Depending on the aptamer-target recognition mode, I would explore fluorescent or electrochemical aptasensors for NP detection in complex environmental matrices. This sensor would overcome limitations in current environmental monitoring techniques for inexpensive, simple, and time-effective detection of this harmful contaminant found across Aotearoa and the world.

4.5 Conclusions

In conclusion, potential aptamer candidates were generated towards 4-NP and a mix of NPTE under increasing low and high stringency conditions. Selection libraries were analysed using NGS data and based on cluster frequency, ten candidates from each library were characterised with an AuNP assay. It was found that the high stringency, compared to the low stringency, conditions reduced the quantity of DNA following each SELEX round, as expected. Moreover, the high stringency libraries contained fewer unique sequences in comparison. The data from the libraries selected for NPTE binders was variable which is likely due to technical issues experienced during SELEX strand separation and suboptimal NPTE conjugation to the Sepharose beads as evidenced by the inconclusive conjugation assessment. Nonetheless, a greater number of unique sequences were found in the high stringency library, compared to both low and high stringency libraries selected to 4-NP, which may be attributable to the fact that multiple target isomers were present on the beads. Alternatively, it could be due to non-specific binders owing to unconjugated beads. Though no binding candidates were successfully identified in any enriched library in this study, this may be due to discrepancies in the selection methodology and characterisation technique. Thus, the first hypothesis remains unanswered regarding aptamer characteristics (affinity, specificity) resulting from a single or multiple target approach SELEX. However, the second hypothesis can be accepted as greater detergent concentrations increased selection stringency resulting in fewer nucleotide sequences. Whether these stringency measures resulted in high affinity binders is yet to be confirmed. Overall, some findings of this study can provide insight into the future experimental design of small molecule SELEX.

Bibliography

4-Nonylphenol | C15H24O - PubChem. Retrieved, from
    https://pubchem.ncbi.nlm.nih.gov/compound/4-Nonylphenol#section=Use-and-
    Manufacturing. Accessed April 15, 2021

Ahamad, A., Madhav, S., Singh, A. K., Kumar, A., & Singh, P. (2020a). *Types of Water
    Pollutants: Conventional and Emerging* (pp. 21–41). Springer, Singapore.
    https://doi.org/10.1007/978-981-15-0671-0_3

Ahamad, A., Madhav, S., Singh, A. K., Kumar, A., & Singh, P. (2020b). *Types of Water
    Pollutants: Conventional and Emerging* (pp. 21–41). Springer, Singapore.
    https://doi.org/10.1007/978-981-15-0671-0_3

Alam, K. K., Chang, J. L., & Burke, D. H. (2015). FASTAptamer: A Bioinformatic Toolkit for
    High-throughput Sequence Analysis of Combinatorial Selections. *Molecular Therapy.
    Nucleic Acids*, *4*(3), e230. https://doi.org/10.1038/MTNA.2015.4

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Isolating,
    Cloning, and Sequencing DNA*. https://www.ncbi.nlm.nih.gov/books/NBK26837/

Aquino-Jarquin, G., & Toscano-Garibay, J. D. (2011). RNA Aptamer Evolution: Two Decades
    of SELEction. *International Journal of Molecular Sciences 2011, Vol. 12, Pages 9155-
    9171*, *12*(12), 9155–9171. https://doi.org/10.3390/IJMS12129155

Arnold, F. H., Georgiou, G., Cirino, P. C., Mayer, K. M., & Umeno, D. (2003). Generating
    Mutant Libraries Using Error-Prone PCR. In *Directed Evolution Library Creation* (pp. 3–
    10). Humana Press. https://doi.org/10.1385/1-59259-395-x:3

Avci-Adali, M., Paul, A., Wilhelm, N., Ziemer, G., & Wendel, H. P. (2009). Upgrading SELEX
    Technology by Using Lambda Exonuclease Digestion for Single-Stranded DNA
    Generation. *Molecules*, *15*(1), 1–11. https://doi.org/10.3390/molecules15010001

Bai, N., Abuduaini, R., Wang, S., Zhang, M., Zhu, X., & Zhao, Y. (2017). Nonylphenol
    biodegradation characterizations and bacterial composition analysis of an effective

consortium NP-M2. *Environmental Pollution*, *220*, 95–104.
https://doi.org/10.1016/j.envpol.2016.09.027

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition*, *98*(6), 236–238.
https://doi.org/10.1136/archdischild-2013-304340

Beier, R., Boschke, E., & Labudde, D. (2014). New strategies for evaluation and analysis of SELEX experiments. *BioMed Research International*, *2014*.
https://doi.org/10.1155/2014/849743

Belay, K. (2016). *Advanced Analytical Microextraction Techniques and There Applications: A Review*. *6*(7). www.iiste.org

Bhandari, G., Bagheri, A. R., Bhatt, P., & Bilal, M. (2021). Occurrence, potential ecological risks, and degradation of endocrine disrupter, nonylphenol, from the aqueous environment. *Chemosphere*, *275*, 130013.
https://doi.org/10.1016/J.CHEMOSPHERE.2021.130013

Bhardwaj, T., Rathore, A. S., & Jha, S. K. (2020a). The selection of highly specific and selective aptamers using modified SELEX and their use in process analytical techniques for Lucentis bioproduction. *RSC Advances*, *10*(48), 28906–28917.
https://doi.org/10.1039/D0RA03542D

Bhardwaj, T., Rathore, A. S., & Jha, S. K. (2020b). The selection of highly specific and selective aptamers using modified SELEX and their use in process analytical techniques for Lucentis bioproduction. *RSC Advances*, *10*(48), 28906–28917.
https://doi.org/10.1039/d0ra03542d

Bittker, J. A., Le, B. V., & Liu, D. R. (2002). Nucleic acid evolution and minimization by nonhomologous random recombination. *Nature Biotechnology*, *20*(10), 1024.
https://doi.org/10.1038/NBT736

Blind, M., & Blank, M. (2015). Aptamer Selection Technology and Recent Advances. In *Molecular Therapy - Nucleic Acids* (Vol. 4, Issue 1, p. e223). Nature Publishing Group.

https://doi.org/10.1038/mtna.2014.74

Boretti, A., & Rosa, L. (2019). Reassessing the projections of the World Water Development Report. *Npj Clean Water*, *2*(1), 1–6. https://doi.org/10.1038/s41545-019-0039-9

Cai, S., Yan, J., Xiong, H., Liu, Y., Peng, D., & Liu, Z. (2018). Investigations on the interface of nucleic acid aptamers and binding targets. In *Analyst* (Vol. 143, Issue 22, pp. 5317–5338). Royal Society of Chemistry. https://doi.org/10.1039/c8an01467a

Cherniaev, A. P., Kondakova, A. S., & Zyk, E. N. (2016). Contents of 4-Nonylphenol in Surface Sea Water of Amur Bay (Japan/East Sea). *Achievements in the Life Sciences*, *10*(1), 65–71. https://doi.org/10.1016/j.als.2016.05.006

Chokwe, T. B., Okonkwo, J. O., & Sibali, L. L. (2017a). Distribution, exposure pathways, sources and toxicity of nonylphenol and nonylphenol ethoxylates in the environment. In *Water SA* (Vol. 43, Issue 4, pp. 529–542). South African Water Research Commission. https://doi.org/10.4314/wsa.v43i4.01

Chokwe, T. B., Okonkwo, J. O., & Sibali, L. L. (2017b). Distribution, exposure pathways, sources and toxicity of nonylphenol and nonylphenol ethoxylates in the environment. In *Water SA* (Vol. 43, Issue 4, pp. 529–542). South African Water Research Commission. https://doi.org/10.4314/wsa.v43i4.01

Clean Water Crisis Facts and Information. Retrieved from https://www.nationalgeographic.com/environment/article/freshwater-crisis. Accessed April 15, 2021.

Close, M. E., Humphries, B., & Northcott, G. (2021). Outcomes of the first combined national survey of pesticides and emerging organic contaminants (EOCs) in groundwater in New Zealand 2018. *Science of The Total Environment*, *754*, 142005. https://doi.org/10.1016/J.SCITOTENV.2020.142005

Cowperthwaite, M. C., & Ellington, A. D. (2008). Bioinformatic analysis of the contribution of primer sequences to aptamer structures. *Journal of Molecular Evolution*, *67*(1), 95–102. https://doi.org/10.1007/S00239-008-9130-4/FIGURES/5

Daems, E., Moro, G., Campos, R., & De Wael, K. (2021). Mapping the gaps in chemical analysis for the characterisation of aptamer-target interactions. *TrAC Trends in Analytical Chemistry*, *142*, 116311. https://doi.org/10.1016/J.TRAC.2021.116311

Daniels, J. S., & Pourmand, N. (2007). Label-free impedance biosensors: Opportunities and challenges. In *Electroanalysis* (Vol. 19, Issue 12, pp. 1239–1257). NIH Public Access. https://doi.org/10.1002/elan.200603855

Dao, P., Hoinka, J., Takahashi, M., Backofen, R., Burnett, J., Zhou, J., Ho, M., Wang, Y., Costa, F., Rossi, J. J., & Przytycka, T. M. (2016). AptaTRACE Elucidates RNA Sequence-Structure Motifs from Selection Trends in HT-SELEX Experiments. *Cell Systems*, *3*, 62–70. https://doi.org/10.1016/j.cels.2016.07.003

Darmostuk, M., Rimpelova, S., Gbelcova, H., & Ruml, T. (2014). Current approaches in SELEX: An update to aptamer selection technology. In *Biotechnology Advances* (Vol. 33, Issue 6, pp. 1141–1161). Elsevier Inc. https://doi.org/10.1016/j.biotechadv.2015.02.008

de Araujo, F. G., Bauerfeldt, G. F., & Cid, Y. P. (2018a). Nonylphenol: Properties, legislation, toxicity and determination. *Anais Da Academia Brasileira de Ciencias*, *90*(2), 1903–1918. https://doi.org/10.1590/0001-3765201720170023

de Araujo, F. G., Bauerfeldt, G. F., & Cid, Y. P. (2018b). Nonylphenol: Properties, legislation, toxicity and determination. *Anais Da Academia Brasileira de Ciencias*, *90*(2), 1903–1918. https://doi.org/10.1590/0001-3765201720170023

DeWitt, J. C., & Patisaul, H. B. (2018). Endocrine disruptors and the developing immune system. In *Current Opinion in Toxicology* (Vol. 10, pp. 31–36). Elsevier B.V. https://doi.org/10.1016/j.cotox.2017.12.005

Dong, C. Di, Chen, C. W., & Chen, C. F. (2015). Seasonal and spatial distribution of 4-nonylphenol and 4-tert-octylphenol in the sediment of Kaohsiung Harbor, Taiwan. *Chemosphere*, *134*, 588–597. https://doi.org/10.1016/j.chemosphere.2014.10.082

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

https://doi.org/10.1038/346818a0

EPA. (2005). *Aquatic Life Ambient Water Quality Criteria Nonylphenol* . https://www.epa.gov/sites/default/files/2019-03/documents/ambient-wqc-nonylphenol-final.pdf

Famulok, M. (2002). Molecular Recognition of Amino Acids by RNA-Aptamers: An L-Citrulline Binding RNA Motif and Its Evolution into an L-Arginine Binder. *Journal of the American Chemical Society*, *116*(5), 1698–1706. https://doi.org/10.1021/JA00084A010

Ferrey, M. L., Hamilton, M. C., Backe, W. J., & Anderson, K. E. (2018). Pharmaceuticals and other anthropogenic chemicals in atmospheric particulates and precipitation. *Science of The Total Environment*, *612*, 1488–1497. https://doi.org/10.1016/J.SCITOTENV.2017.06.201

Freshwater Values • Environment Guide. Retrieved from https://www.environmentguide.org.nz/issues/freshwater/the-freshwater-environment/freshwater-values/. Accessed November 1, 2021.

Gao, H., Tao, S., Wang, D., Zhang, C., Ma, X., Cheng, J., & Zhou, Y. (2003). Comparison of Different Methods for Preparing Single Stranded DNA for Oligonucleotide Microarray. *Analytical Letters*, *36*(13), 2849–2863. https://doi.org/10.1081/AL-120025260

Gas Chromatography Mass Spectrometry (GC-MS) Information | Thermo Fisher Scientific - NZ. Retrieved from https://www.thermofisher.com/nz/en/home/industrial/mass-spectrometry/mass-spectrometry-learning-center/gas-chromatography-mass-spectrometry-gc-ms-information.html. Accessed April 15, 2021.

Hanson, A. M., Kittilson, J. D., Martin, L. E., & Sheridan, M. A. (2014). Environmental estrogens inhibit growth of rainbow trout (Oncorhynchus mykiss) by modulating the growth hormone-insulin-like growth factor system. *General and Comparative Endocrinology*, *196*, 130–138. https://doi.org/10.1016/j.ygcen.2013.11.013

Harmsworth, G., Awatere, S., & Robb, M. (2016). *Indigenous Māori values and perspectives to inform freshwater management in Aotearoa-New Zealand*. *21*(4).

https://doi.org/10.5751/ES-08804-210409

Inoue, K., Yoshida, S., Nakayama, S., Ito, R., Okanouchi, N., & Nakazawa, H. (2006). Development of stable isotope dilution quantification liquid chromatography-mass spectrometry method for estimation of exposure levels of bisphenol A, 4-tert-octylphenol, 4-nonylphenol, tetrabromobisphenol A, and pentachlorophenol in indoor air. *Archives of Environmental Contamination and Toxicology*, *51*(4), 503–508. https://doi.org/10.1007/s00244-005-0236-z

Ji, X., Li, N., Yuan, S., Zhou, X., Ding, F., Rao, K., Ma, M., & Wang, Z. (2019). A comparison of endocrine disruption potential of nonylphenol ethoxylate, vanillin ethoxylate, 4-n-nonylphenol and vanillin in vitro. *Ecotoxicology and Environmental Safety*, *175*, 208–214. https://doi.org/10.1016/j.ecoenv.2019.03.060

Kalra, P., Dhiman, A., Cho, W. C., Bruno, J. G., & Sharma, T. K. (2018). Simple Methods and Rational Design for Enhancing Aptamer Sensitivity and Specificity. *Frontiers in Molecular Biosciences*, *0*(MAY), 41. https://doi.org/10.3389/FMOLB.2018.00041

Karpińska, J., & Kotowska, U. (2019). Removal of Organic Pollution in the Water Environment. *Water*, *11*(10), 2017. https://doi.org/10.3390/w11102017

Kim, A. R., Kim, S. H., Kim, D., Cho, S. W., Son, A., & Yoon, M. Y. (2019). Detection of Nonylphenol with a Gold-Nanoparticle-Based Small-Molecule Sensing System Using an ssDNA Aptamer. *International Journal of Molecular Sciences 2020, Vol. 21, Page 208*, *21*(1), 208. https://doi.org/10.3390/IJMS21010208

Kim, S. H., Thoa, T. T. T., & Gu, M. B. (2019). Aptasensors for environmental monitoring of contaminants in water and soil. *Current Opinion in Environmental Science & Health*, *10*, 9–21. https://doi.org/10.1016/J.COESH.2019.09.003

Kim, S. H., Nam, K. H., Hwang, K. A., & Choi, K. C. (2016). Influence of hexabromocyclododecane and 4-nonylphenol on the regulation of cell growth, apoptosis and migration in prostatic cancer cells. *Toxicology in Vitro*, *32*, 240–247. https://doi.org/10.1016/j.tiv.2016.01.008

Kimling, J., Maier, M., Okenve, B., Kotaidis, V., Ballot, H., & Plech, A. (2006). Turkevich Method for Gold Nanoparticle Synthesis Revisited. *Journal of Physical Chemistry B*, *110*(32), 15700–15707. https://doi.org/10.1021/JP061667W

Kinghorn, A. B., Fraser, L. A., Lang, S., Shiu, S. C. C., & Tanner, J. A. (2017). Aptamer Bioinformatics. *International Journal of Molecular Sciences*, *18*(12). https://doi.org/10.3390/IJMS18122516

Kohlberger, M., & Gadermaier, G. (2021a). SELEX: Critical factors and optimization strategies for successful aptamer selection. *Biotechnology and Applied Biochemistry*. https://doi.org/10.1002/BAB.2244

Kohlberger, M., & Gadermaier, G. (2021b). SELEX: Critical factors and optimization strategies for successful aptamer selection. *Biotechnology and Applied Biochemistry*. https://doi.org/10.1002/BAB.2244

Komarova, N., Barkova, D., & Kuznetsov, A. (2020). Implementation of High-Throughput Sequencing (HTS) in Aptamer Selection Technology. *International Journal of Molecular Sciences*, *21*(22), 1–22. https://doi.org/10.3390/IJMS21228774

Komarova, N., & Kuznetsov, A. (2019). Inside the black box: What makes Selex better? In *Molecules* (Vol. 24, Issue 19, p. 3598). MDPI AG. https://doi.org/10.3390/molecules24193598

Kwon, Y. S., Nguyen, V. T., Park, J. G., & Gu, M. B. (2015). Detection of Iprobenfos and Edifenphos using a new Multi-aptasensor. *Analytica Chimica Acta*, *868*, 60–66. https://doi.org/10.1016/J.ACA.2015.02.020

Lee, J.-W., Han, H.-K., Park, S., & Moon, E.-Y. (2017). Nonylphenol increases tumor formation and growth by suppressing gender-independent lymphocyte proliferation and macrophage activation. *Environmental Toxicology*, *32*(6), 1679–1687. https://doi.org/10.1002/tox.22385

Leggett, R. M., Ramirez-Gonzalez, R. H., Clavijo, B. J., Waite, D., & Davey, R. P. (2013). Sequencing quality assessment tools to enable data-driven informatics for high

throughput genomics. *Frontiers in Genetics*, *4*(DEC).
https://doi.org/10.3389/FGENE.2013.00288

León, V. M., Viñas, L., Concha-Graña, E., Fernández-González, V., Salgueiro-González, N.,
Moscoso-Pérez, C., Muniategui-Lorenzo, S., & Campillo, J. A. (2020). Identification of
contaminants of emerging concern with potential environmental risk in Spanish
continental shelf sediments. *Science of the Total Environment*, *742*, 140505.
https://doi.org/10.1016/j.scitotenv.2020.140505

Li, B., Du, Y., Wei, H., & Dong, S. (2007). Reusable, label-free electrochemical aptasensor for
sensitive detection of small molecules. *Chemical Communications*, *0*(36), 3780–3782.
https://doi.org/10.1039/b707057h

Liu, L. S., Wang, F., Ge, Y., & Lo, P. K. (2020). Recent Developments in Aptasensors for
Diagnostic Applications. In *ACS Applied Materials and Interfaces* (Vol. 13, pp. 9329–
9358). American Chemical Society. https://doi.org/10.1021/acsami.0c14788

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., &
Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology : AMB*,
6(1), 26. https://doi.org/10.1186/1748-7188-6-26

Lyu, C., Khan, I. M., & Wang, Z. (2021). Capture-SELEX for aptamer selection: A short review.
Talanta, *229*, 122274. https://doi.org/10.1016/J.TALANTA.2021.122274

Mao, Z., Zheng, X.-F., Zhang, Y.-Q., Tao, X.-X., Li, Y., & Wang, W. (2012). Occurrence and
Biodegradation of Nonylphenol in the Environment. *International Journal of Molecular
Sciences*, *13*(1), 491. https://doi.org/10.3390/IJMS13010491

Māori values: Protecting NZ's rivers. Retrieved from https://www.doc.govt.nz/about-
us/statutory-and-advisory-bodies/nz-conservation-authority/publications/protecting-
new-zealands-rivers/02-state-of-our-rivers/maori-values/. Accessed November 1, 2021.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing
reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/EJ.17.1.200

McConnell, E. M., Nguyen, J., & Li, Y. (2020). Aptamer-Based Biosensors for Environmental

Monitoring. In *Frontiers in Chemistry* (Vol. 8, p. 434). Frontiers Media S.A.
https://doi.org/10.3389/fchem.2020.00434

McKeague, M., & Derosa, M. C. (2012). Challenges and opportunities for small molecule
aptamer development. In *Journal of Nucleic Acids* (Vol. 2012).
https://doi.org/10.1155/2012/748913

McKeague, M., McConnell, E. M., Cruz-Toledo, J., Bernard, E. D., Pach, A., Mastronardi, E.,
Zhang, X., Beking, M., Francis, T., Giamberardino, A., Cabecinha, A., Ruscito, A., Aranda-
Rodriguez, R., Dumontier, M., & DeRosa, M. C. (2015). Analysis of In Vitro Aptamer
Selection Parameters. *Journal of Molecular Evolution*, *81*(5–6), 150–161.
https://doi.org/10.1007/s00239-015-9708-6

Mishra, G. K., Sharma, V., & Mishra, R. K. (2018). Electrochemical aptasensors for food and
environmental safeguarding: A review. In *Biosensors* (Vol. 8, Issue 2). MDPI AG.
https://doi.org/10.3390/bios8020028

Moreau, M., Hadfield, J., Hughey, J., Sanders, F., Lapworth, D. J., White, D., & Civil, W.
(2019a). A baseline assessment of emerging organic contaminants in New Zealand
groundwater. *Science of the Total Environment*, *686*, 425–439.
https://doi.org/10.1016/j.scitotenv.2019.05.210

Moreau, M., Hadfield, J., Hughey, J., Sanders, F., Lapworth, D. J., White, D., & Civil, W.
(2019b). A baseline assessment of emerging organic contaminants in New Zealand
groundwater. *Science of The Total Environment*, *686*, 425–439.
https://doi.org/10.1016/J.SCITOTENV.2019.05.210

Moreno, M. (2015). Aptasensor. In *Encyclopedia of Astrobiology* (pp. 114–115). Springer
Berlin Heidelberg. https://doi.org/10.1007/978-3-662-44185-5_5167

Murray, Z., & Zak. (2020). *The generation and evolutionary study of novel aptamers for
environmental contaminants*. http://researcharchive.vuw.ac.nz/handle/10063/9188

Nguyen, V. T., Kwon, Y. S., & Gu, M. B. (2017). Aptamer-based environmental biosensors for
small molecule contaminants. In *Current Opinion in Biotechnology* (Vol. 45, pp. 15–23).

Elsevier Ltd. https://doi.org/10.1016/j.copbio.2016.11.020

Noorimotlagh, Z., Mirzaee, S. A., Ahmadi, M., Jaafarzadeh, N., & Rahim, F. (2018). The possible DNA damage induced by environmental organic compounds: The case of Nonylphenol. In *Ecotoxicology and Environmental Safety* (Vol. 158, pp. 171–181). Academic Press. https://doi.org/10.1016/j.ecoenv.2018.04.023

Noorimotlagh, Z., Mirzaee, S. A., Martinez, S. S., Rachoń, D., Hoseinzadeh, M., & Jaafarzadeh, N. (2020). Environmental exposure to nonylphenol and cancer progression Risk–A systematic review. In *Environmental Research* (Vol. 184, p. 109263). Academic Press Inc. https://doi.org/10.1016/j.envres.2020.109263

Nowak, K., Jabłońska, E., & Ratajczak-Wrona, W. (2019). Immunomodulatory effects of synthetic endocrine disrupting chemicals on the development and functions of human immune cells. In *Environment International* (Vol. 125, pp. 350–364). Elsevier Ltd. https://doi.org/10.1016/j.envint.2019.01.078

Ozer, A., Pagano, J. M., & Lis, J. T. (2014). New technologies provide quantum changes in the scale, speed, and success of SELEX methods and aptamer characterization. In *Molecular Therapy - Nucleic Acids* (Vol. 3, p. e183). Nature Publishing Group. https://doi.org/10.1038/mtna.2014.34

Popkin, B. M., D'Anci, K. E., & Rosenberg, I. H. (2010). Water, hydration, and health. In *Nutrition Reviews* (Vol. 68, Issue 8, pp. 439–458). Blackwell Publishing Inc. https://doi.org/10.1111/j.1753-4887.2010.00304.x

Priac, A., Morin-Crini, N., Druart, C., Gavoille, S., Bradu, C., Lagarrigue, C., Torri, G., Winterton, P., & Crini, G. (2017). Alkylphenol and alkylphenol polyethoxylates in water and wastewater: A review of options for their elimination. In *Arabian Journal of Chemistry* (Vol. 10, pp. S3749–S3773). Elsevier B.V. https://doi.org/10.1016/j.arabjc.2014.05.011

Puteh, M. H., Stuckey, D. C., & Othman, M. H. D. (2015). Direct Measurement of Anaerobic Biodegradability of Nonylphenol Ethoxylates (NPEOs). *International Journal of Environmental Science and Development*, *6*(9), 660–663.

https://doi.org/10.7763/IJESD.2015.V6.676

Raju, S., Sivamurugan, M., Gunasagaran, K., Subramani, T., & Natesan, M. (2018). Preliminary studies on the occurrence of nonylphenol in the marine environments, Chennai—a case study. *The Journal of Basic and Applied Zoology 2018 79:1*, *79*(1), 1–7. https://doi.org/10.1186/S41936-018-0063-1

Rosenfeld, P. E., & Feng, L. G. H. (2011). Emerging Contaminants. In *Risks of Hazardous Wastes* (pp. 215–222). Elsevier. https://doi.org/10.1016/b978-1-4377-7842-7.00016-7

Rougée, L. R. A., Collier, A. C., & Richmond, R. H. (2021). Chronic Exposure to 4-Nonylphenol Alters UDP-Glycosyltransferase and Sulfotransferase Clearance of Steroids in the Hard Coral, Pocillopora damicornis. *Frontiers in Physiology*, *0*, 118. https://doi.org/10.3389/FPHYS.2021.608056

Ruscito, A., & DeRosa, M. C. (2016). Small-molecule binding aptamers: Selection strategies, characterization, and applications. In *Frontiers in Chemistry* (Vol. 4, Issue MAY, p. 14). Frontiers Media S. A. https://doi.org/10.3389/fchem.2016.00014

Ruscito, A., McConnell, E. M., Koudrina, A., Velu, R., Mattice, C., Hunt, V., McKeague, M., & DeRosa, M. C. (2017). In Vitro Selection and Characterization of DNA Aptamers to a Small Molecule Target. *Current Protocols in Chemical Biology*, *9*(4), 233–268. https://doi.org/10.1002/cpch.28

Rudel, A. R., Camann, D. E., Spengler, J. D., Korn, L. R., & Brody, J. G. (2003). Phthalates, Alkylphenols, Pesticides, Polybrominated Diphenyl Ethers, and Other Endocrine-Disrupting Compounds in Indoor Air and Dust. *Environmental Science and Technology*, *37*(20), 4543–4553. https://doi.org/10.1021/ES0264596

Saha, K., Agasti, S. S., Kim, C., Li, X., & Rotello, V. M. (2012). Gold Nanoparticles in Chemical and Biological Sensing. *Chemical Reviews*, *112*(5), 2739. https://doi.org/10.1021/CR2001178

Sakamoto, T., Ennifar, E., & Nakamura, Y. (2018). Thermodynamic study of aptamers binding to their target proteins. In *Biochimie* (Vol. 145, pp. 91–97). Elsevier B.V.

https://doi.org/10.1016/j.biochi.2017.10.010

Sampson, T. (2003). Aptamers and SELEX: The technology. *World Patent Information*, *25*(2), 123–129. https://doi.org/10.1016/S0172-2190(03)00035-8

Sanger Sequencing Steps & Method. Retrieved from https://www.sigmaaldrich.com/NZ/en/technical-documents/protocol/genomics/sequencing/sanger-sequencing. Accessed October 27, 2021.

Sayed, A. E. D. H., & Hamed, H. S. (2017). Induction of apoptosis and DNA damage by 4-nonylphenol in African catfish (Clarias gariepinus) and the antioxidant role of Cydonia oblonga. *Ecotoxicology and Environmental Safety*, *139*, 97–101. https://doi.org/10.1016/j.ecoenv.2017.01.024

Schulz, F., Homolka, T., Bastús, N. G., Puntes, V., Weller, H., & Vossmeyer, T. (2014). Little adjustments significantly improve the Turkevich synthesis of gold nanoparticles. *Langmuir*, *30*(35), 10779–10784. https://doi.org/10.1021/LA503209B/SUPPL_FILE/LA503209B_SI_001.PDF

Song, S., Wang, L., Li, J., Fan, C., & Zhao, J. (2008). Aptamer-based biosensors. *TrAC - Trends in Analytical Chemistry*, *27*(2), 108–117. https://doi.org/10.1016/j.trac.2007.12.004

Stewart, M., Olsen, G., Hickey, C. W., Ferreira, B., Jelić, A., Petrović, M., & Barcelo, D. (2014). A survey of emerging contaminants in the estuarine receiving environment around Auckland, New Zealand. *Science of The Total Environment*, *468–469*, 202–210. https://doi.org/10.1016/J.SCITOTENV.2013.08.039

Stoltenburg, R., Reinemann, C., & Strehlitz, B. (2007). SELEX—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering*, *24*(4), 381–403. https://doi.org/10.1016/J.BIOENG.2007.06.001

Sun, D., Chen, Q., He, N., Diao, P. P., Jia, L. X., & Duan, S. S. (2017). Effect of environmentally-relevant concentrations of nonylphenol on sexual differentiation in zebrafish: A multi-generational study. *Scientific Reports*, *7*(1), 1–8. https://doi.org/10.1038/srep42907

Sun, H., & Zu, Y. (2015). A Highlight of recent advances in aptamer technology and its application. In *Molecules* (Vol. 20, Issue 7, pp. 11959–11980). MDPI AG. https://doi.org/10.3390/molecules200711959

Svobodová, M., Pinto, A., Nadal, P., & O' Sullivan, C. K. (2012). Comparison of different methods for generation of single-stranded DNA for SELEX processes. *Analytical and Bioanalytical Chemistry*, *404*(3), 835–842. https://doi.org/10.1007/s00216-012-6183-4

Tackling global water pollution | UNEP - UN Environment Programme. Retrieved from https://www.unep.org/explore-topics/water/what-we-do/tackling-global-water-pollution. Accessed April 15, 2021.

Tan, C. S., Citartan, M., Tang, T.-H., Tan, S.-C., Hoe, C.-H., Saini, R., Tominaga, J., & Gopinath, S. C. B. (2012). Asymmetric PCR for good quality ssDNA generation towards DNA aptamer production. In *Article in Songklanakarin Journal of Science and Technology*. https://www.researchgate.net/publication/225292962

Thevendran, R., & Citartan, M. (2022). Assays to Estimate the Binding Affinity of Aptamers. *Talanta*, *238*, 122971. https://doi.org/10.1016/J.TALANTA.2021.122971

Thiviyanathan, V., & Gorenstein, D. G. (2012). Aptamers and the next generation of diagnostic reagents. In *Proteomics - Clinical Applications* (Vol. 6, Issues 11–12, pp. 563–573). NIH Public Access. https://doi.org/10.1002/prca.201200042

Tolle, F., & Mayer, G. (2016). Preparation of SELEX samples for next-generation sequencing. In *Methods in Molecular Biology* (Vol. 1380, pp. 77–84). Humana Press Inc. https://doi.org/10.1007/978-1-4939-3197-2_6

Tombelli, S., Minunni, M., & Mascini, M. (2007). Aptamers-based assays for diagnostics, environmental and food analysis. In *Biomolecular Engineering* (Vol. 24, Issue 2, pp. 191–200). Elsevier. https://doi.org/10.1016/j.bioeng.2007.03.003

Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)*, *249*(4968), 505–510. https://doi.org/10.1126/SCIENCE.2200121

United Nations Department of Economic and Social Affairs. (2017). *World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100 | UN DESA | United Nations Department of Economic and Social Affairs*. https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html

Valasek, M. A., & Repa, J. J. (2005). The power of real-time PCR. *American Journal of Physiology - Advances in Physiology Education*, *29*(3), 151–159. https://doi.org/10.1152/advan.00019.2005

Vargas-Berrones, K., Bernal-Jácome, L., Díaz de León-Martínez, L., & Flores-Ramírez, R. (2020). Emerging pollutants (EPs) in Latin América: A critical review of under-studied EPs, case of study -Nonylphenol-. In *Science of the Total Environment* (Vol. 726, p. 138493). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2020.138493

Villalonga, A., Pérez-Calabuig, A. M., & Villalonga, R. (2020). Electrochemical biosensors based on nucleic acid aptamers. In *Analytical and Bioanalytical Chemistry* (Vol. 412, Issue 1, pp. 55–72). Springer. https://doi.org/10.1007/s00216-019-02226-x

Wai Māori-Māori values in Water. (2010). https://doi.org/10.449

Wang, C., & Yu, C. (2013). Detection of chemical pollutants in water using gold nanoparticles as sensors: a review. *Reviews in Analytical Chemistry*, *32*(1), 1–14. https://doi.org/10.1515/REVAC-2012-0023

Wang, T., Chen, C., Larcher, L. M., Barrero, R. A., & Veedu, R. N. (2019). Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. In *Biotechnology Advances* (Vol. 37, Issue 1, pp. 28–50). Elsevier Inc. https://doi.org/10.1016/j.biotechadv.2018.11.001

Why freshwater matters | Ministry for the Environment. Retrieved from https://environment.govt.nz/what-government-is-doing/areas-of-work/freshwater/why-freshwater-matters/. Acessed November 1, 2021.

Wuyts, S., & Segata, N. (2019). At the forefront of the sequencing revolution - Notes from

the RNGS19 conference. *Genome Biology*, *20*(1), 93. https://doi.org/10.1186/s13059-019-1714-3

Yang, R., Liu, J., Song, D., Zhu, A., Xu, W., Wang, H., & Long, F. (2019). Reusable chemiluminescent fiber optic aptasensor for the determination of 17β-estradiol in water samples. *Microchimica Acta*, *186*(11), 1–9. https://doi.org/10.1007/s00604-019-3813-y

Yang, X., Li, N., & Gorenstein, D. G. (2011). Strategies for the discovery of therapeutic Aptamers. *Expert Opinion on Drug Discovery*, *6*(1), 75. https://doi.org/10.1517/17460441.2011.537321

Yu, H., Alkhamis, O., Canoura, J., Liu, Y., & Xiao, Y. (2021). Advances and Challenges in Small-Molecule DNA Aptamer Isolation, Characterization, and Sensor Development. *Angewandte Chemie*, *133*(31), 16938–16961. https://doi.org/10.1002/ANGE.202008663

Yu, Z., Luan, Y., Li, H., Wang, W., Wang, X., & Zhang, Q. (2019). A disposable electrochemical aptasensor using single-stranded DNA–methylene blue complex as signal-amplification platform for sensitive sensing of bisphenol A. *Sensors and Actuators, B: Chemical*, *284*, 73–80. https://doi.org/10.1016/j.snb.2018.12.126

Esaka, Y., Tanaka, K., Uno, B., Goto, M., & Kano, K. (1997). Sodium Dodecyl Sulfate–Tween 20 Mixed Micellar Electrokinetic Chromatography for Separation of Hydrophobic Cations: Application to Adrenaline and Its Precursors. *Analytical Chemistry*, *69*(7), 1332–1338. https://doi.org/10.1021/AC960731A

Zhang, F., & Liu, J. (2021). Label-Free Colorimetric Biosensors Based on Aptamers and Gold Nanoparticles: A Critical Review. *Analysis & Sensing*, *1*(1), 30–43. https://doi.org/10.1002/ANSE.202000023

Zhang, F., Yang, P., Qin, L., & Zhang, J. (2017). Adverse stimulation of 4-nonylphenol in abnormal reproductive organs of female chickens. *Oncotarget*, *8*(66), 110029–110038. https://doi.org/10.18632/oncotarget.21858

Zhang, G., Li, T., Zhang, J., & Chen, A. (2018). A simple FRET-based turn-on fluorescent aptasensor for 17β-estradiol determination in environmental water, urine and milk samples. *Sensors and Actuators B: Chemical*, *273*, 1648–1653. https://doi.org/10.1016/J.SNB.2018.07.066

Zhu, G., & Chen, X. (2018). Aptamer-based targeted therapy. In *Advanced Drug Delivery Reviews* (Vol. 134, pp. 65–78). Elsevier B.V. https://doi.org/10.1016/j.addr.2018.08.005

Zhuo, Z., Yu, Y., Wang, M., Li, J., Zhang, Z., Liu, J., Wu, X., Lu, A., Zhang, G., & Zhang, B. (2017). Recent advances in SELEX technology and aptamer applications in biomedicine. In *International Journal of Molecular Sciences* (Vol. 18, Issue 10, p. 2142). MDPI AG. https://doi.org/10.3390/ijms18102142

Appendix A. Reagents and Consumables

Table A.1: Reagents used in systemic evolution of ligands by exponential enrichment (SELEX). Supplier address is only given on first use.

Reagent/ Equipment	Abbreviation or Common Name	Catalogue No.	Manufacturer/ Supplier
Polysorbate 20	Tween 20	85115	Thermofisher Scientific, Waltham, Massachusetts, United States
Sepharose CL-6B		CL6B200-100ML	Sigma Aldrich, St Louis, Missouri, United States
1,4-Bis(2,3-epoxypropyloxy)butane	1,4-Butanediol diglycidyl ether	220892-50G	Sigma Aldrich
Sodium Chloride	NaCl	BP358-1	Thermofisher Scientific
Sodium Hydroxide (Pellets/Certified ACS), Fisher Chemical	Sodium Hydroxide NaOH	S318-500	Thermofisher Scientific
Potassium Chloride	KCL	P9541-500G	Sigma Aldrich
Calcium Chloride	$CaCl_2$	102378	Merck Millipore, Burlington, Massachusetts, United States
Tris-Hydrochloride > 99%	Tris-HCL	1185-53-1	Thermofisher Scientific
Nonylphenol PESTANAL, analytical standard, technical mixture	Nonylphenol technical equivalents	46018-1G	Sigma Aldrich
4-Nonylphenol, PESTANAL, analytical standard	4-Nonylphenol	46405-100MG	Sigma Aldrich
Ethanolamine ≥ 98%		E9508-100ML	Sigma Aldrich
UltraPure™ DNase/RNase-Free Distilled Water	Ultra-pure Water (up$H_2$0)	10977015	Thermofisher Scientific
Ethanol, Absolute (Analytical grade)	Ethanol (EtOH)	64-17-5	Thermofisher Scientific

Taq DNA Polymerase PCR Buffer (10X)	PCR Buffer	18-067-017	Thermofisher Scientific
DNTP set 100 mM (dATP, dGTP, dCTP, dTTP)	dNTP's	R0181	Thermofisher Scientific
DNA Gel Loading Dye (100X)		R0611	Thermofisher Scientific
Taq DNA Polymerase, recombinant (5 U/µL)	*Taq* Polymerase	EP0401	Thermofisher Scientific
Magnesium Chloride (25 mM)	MgCl2	R0971	Thermofisher Scientific
Buffer PB		19066	Qiagen, Hilden, Germany
SYBR Green I Nucleic Acid Gel Stain, 10,000X Concentrate in DMSO	SYBR Green	S7567	Thermofisher Scientific
TriTrack DNA Loading Dye (6X)		R1161	Thermofisher Scientific
GeneRuler low range DNA ladder	DNA Ladder	SM1193	Thermofisher Scientific
QG solubilisation buffer	QG Buffer	19063	Qiagen
Fisher Scientific 2-Propanol, Fisher Chemical	Isopropanol	FSBA464-4	Thermofisher Scientific
PE Wash Buffer	PE buffer	19065	Qiagen
Elution Buffer	EB Buffer	19086	Qiagen
Streptavidin Magnetic Beads	SMBs	S1420S	New England Biolabs, Ipswich, Massachusetts, United States
Buffer QX1	QX1 Solubilisation Buffer	20912	Qiagen
Agarose Beads		20021	Qiagen
Bisphenol A (BPA)	BPA	239658-50-G	Sigma Aldrich
Dibutyl Phthalate	DP	524980-100ML	Sigma Aldrich
2-Phenylphenol	2-P	P28263-500G	Sigma Aldrich
Monarch PCR and DNA Clean-up Kit		T1030S	New England Biolabs
MinElute Gel Extraction Kit		28604	Qiagen

Table A.2: Reagents used in gold nanoparticle characterisation. Supplier address is only given on first use.

Reagent/ Equipment	Abbreviation or Common Name	Catalogue No.	Manufacturer/Supplier
Acetone		67-64-1	Thermofisher Scientific, Waltham, Massachusetts, United States
Ethanol, Absolute (Analytical grade)	EtOH	64-17-5	Thermofisher Scientific
Methanol > 99.8%	MeOH	20847.307	VWR international, Radnor, Pennsylvania, United States
Nitric Acid	HNO_3	7697-37-2	Thermofisher Scientific
Hydrochloric Acid S.G. 1.18 (approximately 37%)	HCL	H/1200/PB17	Thermofisher Scientific
Gold (III) Chloride Trihydrate	$HAuCl_4$	520918-1G	Sigma Aldrich, St Louis, Missouri, United States
Trisodium Citrate Dihydrate, ACS, 99.0% min	Sodium Citrate	S1804	Thermofisher Scientific
OM5-C6 Aptamer			Integrated DNA technologies, Coraville, IA, United States
Methamphetamine		51-57-0	BDG synthesis, New Zealand

Table A.3: General Consumables. Supplier address is only given on first use.

Equipment	Common Name	Model	Manufacturer/Supplier
Axygen 1.5 mL MaxyClear Snaplock Microcentrifuge Tube, Polypropylene, Clear, Nonsterile	1.5 mL Eppendorf tube	MCT-150-C	Corning Inc., Corning, New York, United States
Corning 15 mL PP Centrifuge Tubes, CentriStar Cap, Sterile	15 mL Falcon tubes	430791	Corning Inc.
MinElute Spin Column		1026476	Qiagen, Hilden, Germany
Strip Tubes and Caps 0.1 mL	PCR tubes	981106	Qiagen
Disposable Syringe 50 mL/60 mL Luer lock		SG-503P-LL	Interlab Ltd, Wellington, New Zealand
Minisart single use filter unit. Non-pyrogenic. Filter pore size 0.45 μm	Minisart Filter	16555	Sartorius AG, Gottingen, Germany
Tissue Culture Plate 96 well, polystyrene, non-treated, flat bottom, single-use	96 well-plates	TCP000096	Interlab Ltd
PickPen 1-M	Pick Pen	23001	Bio Nobile, Pargas, Finland
SPECTROstar Nano	UV-Spec	601-1227	BMG Labtech, Ortenberg, Germany
Mini-Sub Cell GT Gel Tank	Gel Electrophoresis Tank	1704487EDU	Bio-rad, Hercules, California, United States
Omega Lum G Imaging System	Gel Imaging System	81-12100-00	Aplegen, Inc. Pleasanton, United States
Ultraviolet Transilluminator (Bio-Rad)		1708110EDU	Bio-rad
Thermo Shaker	Heat Block	MS-100	ALLSHENG, Hangzhou, China
Qubit 3.0 Fluorometer	Qubit 3	Q33216	Thermofisher Scientific

Corbett RotorGene 6000 instrument	PCR Machine	RG-6000	Qiagen
Nanodrop ND-1000 Spectrophotometer	Nanodrop	ND-1000	Thermofisher Scientific

Appendix A.4: Certificate of analysis (COA) for Nonylphenol technical grade mix of nonylphenol isomers (NPTE).

SIGMA-ALDRICH®

Certificate of Analysis

Product Name:
Nonylphenol – technical grade, mixture of ring and chain isomers

Product Number:	**290858**
Batch Number:	**MKCG3412**
Brand:	ALDRICH
CAS Number:	84852-15-3
MDL Number:	MFCD00002396
Formula:	C15H24O
Formula Weight:	220.35 g/mol
Quality Release Date:	26 APR 2018
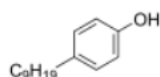
Test	Specification	Result
Appearance (Color)	Colorless	Colorless
Appearance (Form)	Liquid	Liquid
Infrared Spectrum	Conforms to Structure	Conforms

Michael Grady, Manager
Quality Control
Milwaukee, WI   US

112

Appendix A.5: Certificate of analysis (COA) for 4-Nonylphenol (4-NP)
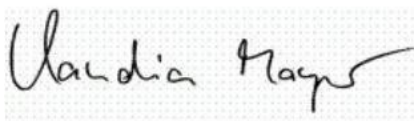
SIGMA-ALDRICH®

Certificate of Analysis

Product Name:	4-Nonylphenol
	PESTANAL™, analytical standard
Product Number:	46405
Batch Number:	STBJ4928
Brand:	Sigma-Aldrich
CAS Number:	104-40-5
Formula:	$C_{15}H_{24}O$
Formula Weight:	220.35
Expiration Date:	OCT 2024
Quality Release Date:	14 NOV 2019

TEST	SPECIFICATION	RESULT
PURITY (GC AREA %)	≥ 98.0 %	> 99.9 %
MELTING POINT	42.0 - 47.0 C	43.1 - 44.7 C
WATER	≤ 1.0 %	< 0.1 %
PROTON NMR SPECTRUM	CONFORMS TO STRUCTURE	CONFORMS

Claudia Mayer

Claudia Mayer
Manager Quality Control
Steinheim, Germany

Appendix B.   DNA Quantification of Libraries in Preparation for High-Throughput Sequencing

Table B.1: Final DNA concentrations and volumes for libraries sent to Gen X Pro for Illumina (NGS) sequencing. Abbreviations: N40 for starting library; 4-NP for 4-nonylphenol; NPTE for nonylphenol technical replicates.

SELEX strategy description	SELEX Round	Concentration (ng/µL)	Sample Volume (µL)
N40 Library	Pre-SELEX	13.3	19
NPTE Low stringency	7	8.77	19
NPTE High stringency	7	16.9	19
4-NP Low stringency	7	4.48	18
4-NP High stringency	7	4.80	18

Appendix C.   Bioinformatics Scripts

The following bioinformatics pipeline was applied to all SELEX libraries.

C.1: Script 1 – Quality control

#FastQC was used to assess the quality of the raw high-throughput sequencing data.

Command-line: fastqc – open file

This provided the base quality, Phred score, GC content, length distribution, duplication levels and overrepresented sequences for each library.

C.2: Script 2 – Removal of primer sequences

#Cutadapt was used to remove reverse complement sequences and flanking primer regions from the sequences to investigate the diverse N40 region. This was performed on all sequence libraries.

Command-line: cutadapt -j 0 -m 37 -M 41 - - discard – untrimmed -a "TAACCACATAACCGCAAGA;min_overlap=14;max_error_rate=0.3"…"TATTGTGCTACTCTCC;min_overlap=14;max_error_rate=0.3" -o N40_trimmed.fastq

Filters are applied to process data, sequences that fall outside these filters are discarded. -m and -M specify the minimum and maximum read lengths retained. These numbers were selected to allow minimal insertions and deletions to the N40 region. A minimum overlap of 14 was specified for the adaptor or primer sequences to be found.

FASTAptamer Pipeline

C.3: Script 3 – Preliminary processing and conversion of files

#FASTAptamer-Count was used to convert FASTq files to the non-redundant FASTA format required for the FASTAptamer script function. The read count, normalised reads per million and rank of each sequence was calculated and inserted as a description line.

Command-line: fastaptamer_count -i N40_trimmed.fastq -o N40_count.fasta

C.4: Script 4 – Comparing conserved sequences and sequence distribution between library populations

#FASTAptamer-Compare was employed to identify sequences that were conserved between populations and each populations sequence distribution. Data was outputted to a plain-text file in the form of (.TSV). All libraries were compared.

Command-line: fastaptamer_compare -x N40_count.fasta -y LSA_count.fasta -o N40_LSA_compare.fasta

C.5: Script 5 – Clustering sequences based on sequence similarity

#FASTAptamer-cluster was applied to group closely-related sequences. A Levenshtein edit distance of 7 was applied. Cluster identify information including cluster, rank within that cluster and edit distance from the parent sequence was encompassed in the description line.

Command-line: fastaptamer_cluster_xs -i N40_count.fasta -o N40_cluster.fasta -d 7 > N40_cluster.tsv

C.6: Script 6 – Calculating fold enrichment of sequences across populations

#FASTAptamer-Enrich was used to calculate the enrichment of sequences across the starting library and libraries following low and high stringency selection.

Command-line: fastaptamer_enrich -x N40_cluster.fasta -y LSA_cluster.fasta -z HSA_cluster.fasta -o NP-eq_ENRICH.tsv

fastaptamer_enrich -x N40_cluster.fasta -y 4NP_L_cluster.fasta -z 4NP_H_cluster.fasta -o 4NP_ENRICH.tsv

Output contained each sequence, sequence length, reads, RPM, rank and cluster information.

C.7: Script 7 – ViennaRNA sequence folding and reliability plot

#RNAfold was used to fold sequences selected for characterisation in predicted secondary structures.

Command-line: RNAfold --MEA--noconv -T 21 -p < 4NP_L10.fasta > L10

Relplot.pl 4NP_L1_ss.ps 4NP_L1_dp.ps > 4NP_L1_rss.ps

--MEA calculates the maximum expected accuracy of the structures. --Nonconv prevented the automatic conversion of "T" to "U" bases. T specified the temperature of the fold. Lastly, -p calculates the base-pairing probability matrix.
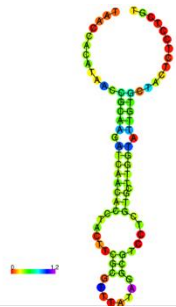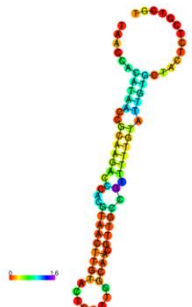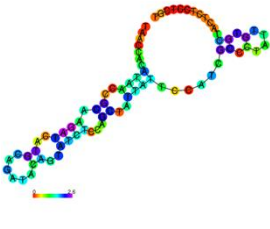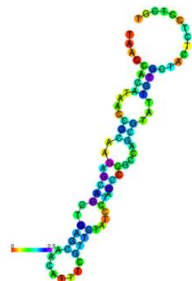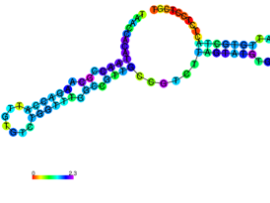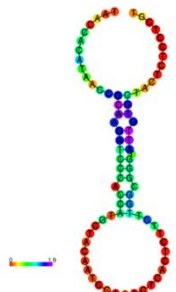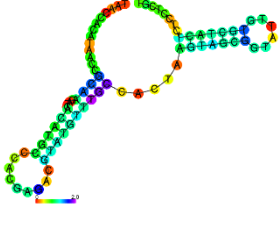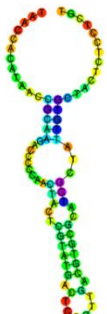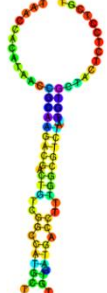
Appendix D.   Candidate Aptamers for Characterisation
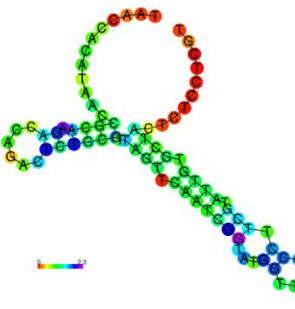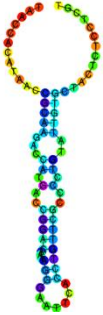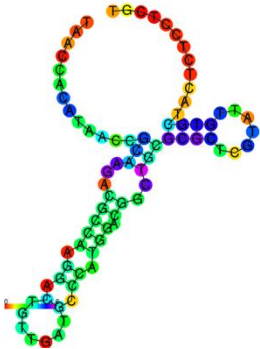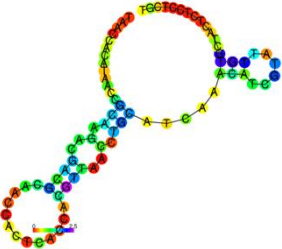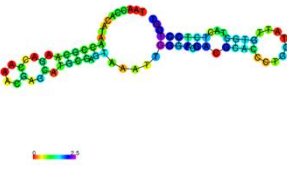
Table D.1: Sequence identifier table. Listed are the candidate oligonucleotides name, sequence, length and Gibbs free energy. Abbreviations: 4-NP for 4-nonylphenol; NPTE for nonylphenol technical replicates; L for low stringency; H for high stringency. Numbers refer to Cluster number.

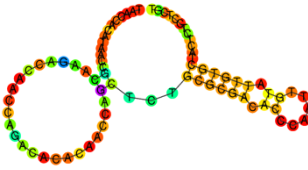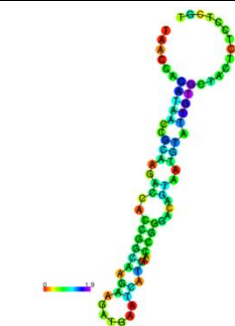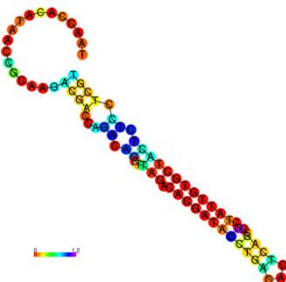| Name	Sequence	Length	Gibbs free energy
NPTE L1/H1	TAACCACATAACCGCAAGACGACGCAACCACTCACCACGTTAACCTGCATCAAACATCGTATTGTGCTACTCTCCTCGT	79	-9.85
NPTE L2/H2	TAACCACATAACCGCAAGACCGAACGCCTAACAAAACACTTGGTTCCGCAACATCGTAATATTGTGCTACTCTCCTCGT	79	10.08
NPTE L4/H3	TAACCACATAACCGCAAGACGACGCGATACAAAACACTACTAATCTCTCGTCGCACTAGTATTGTGCTACTCTCCTCGT	79	-17.85
NPTE L3/H4	TAACCACATAACCGCAAGATCATGCAGATACAGTATCTCCAGCTATTATTCCATCGCCCTATTGTGCTACTCTCCTCGT	79	-10.58
NPTE L5/H7	TAACCACATAACCGCAAGACCGTGGTGACCTATCACCCTATCAAAACCCAATCGCTCGGTATTGTGCTACTCTCCTCGT	79	-17.32
NPTE L6/H5	TAACCACATAACCGCAAGACCAAACGAGCATGCGAGTAAATTGGGACGACGCACCCTGCTATTGTGCTACTCTCCTCGT	79	-19.51
NPTE L8/H13	TAACCACATAACCGCAAGACCAACAGTCGCCCCGCCCAAACGCACTAGCCCGCTCCTCCTATTGTGCTACTCTCCTCGT	79	-10.18
NPTE L9	TAACCACATAACCGCAAGACCAACCAGACACACAACCAGCTCTGCGCGACACCCAATTGTATTGTGCTACTCTCCTCGT	79	-10.73
NPTE L7	TAACCACATAACCGCAAGACGCACGGAATCATACAATCACAGCCAGCCAAGATGACGGGTATTGTGCTACTCTCCTCGT	79	-14.92
NPTE H6	TAACCACATAACCGCAAGACCACCGGCAAGAAGATGAATCATGAACCGGGACAGTAATGTATTGTGCTACTCTCCTCGT	79	-15.29
NPTE H10	TAACCACATAACCGCAAGACGACCAGACAGCTTAGACACGATACCTGACACTCAGCCGCTATTGTGCTACTCTCCTCGT	78	-18.48
4-NP L1	TAACCACATAACCGCAAGATCAACACCTACTTCGCGTTTATAGGCGTCCTCGTGCTTGGTATTGTGCTACTCTCCTCGT	79	-20.51
4-NP L4	TAACCACATAACCGCAAGACCCACGTAACTTGTACTGATTGGCAACGTTGCCGTTTTTGTATTGTGCTACTCTCCTCGT	79	-19.47
4-NP L3	TAACCACATAACCGCAAGATCATGCAGATACAGTATCTCCAGCTATTATTCCATCGCCCTATTGTGCTACTCTCCTCGT	79	-10.58
4-NP L2	TAACCACATAACCGCAAGACCCAGTCACATGCAATTCATAGATTACGCCATGGCTCGTGGTATTGTGCTACTCTCCTCGT	80	-18.06
4-NP L5	TAACCACATAACCGCAAGACGGATAAGATTGGGACTACGTTTCAGATTCGTTGGTCTTGTATTGTGCTACTCTCCTCGT	79	-27.38

4-NP L6	TAACCACATAACCGCAAGAGCAGCTGAGCAACATTTCGTTCTATCCTAGCCCGCCAGCGTATTGTGCTACTCTCCTCGT	79	-16.54
4-NP L7	TAACCACATAACCGCAAGACCATTGTGTCTGGTTTGGCCGTTGCCGTCTTAGTATGTGTTATTGTGCTACTCTCCTCGT	79	-20.59
4-NP L11	TAACCACATAACCGCAAGATCAGCCGCGGTCTTCCCTGTAGTTGTTCTTTCTCGATTGATATTGTGCTACTCTCCTCGT	79	-19.83
4-NP L10	TAACCACATAACCGCAAGATCCCACCATGCTACAATCCTCTCTCACTCTTCTTGGCGGGTATTGTGCTACTCTCCTCGT	79	-17.37
4-NP L8	TAACCACATAACCGCAAGATCACATGCCCACGAGACGTATGTTTGCCACTAAGTAGCGGTATTGTGCTACTCTCCTCGT	79	-20.98
4-NP H1	TAACCACATAACCGCAAGACCACCCACAGCTTACTTCACTGGATAGTCGCGTGAACGGTTATTGTGCTACTCTCCTCGT	79	-18.21
4-NP H2	TAACCACATAACCGCAAGACCCCCAACTACTCGTATGACTCCGTTGACGTGCGCATGGCTATTGTGCTACTCTCCTCGT	79	-14.63
4-NP H3	TAACCACATAACCGCAAGACCCTACAGACTCCCCGGCCTTCCTGGTCGCCTGCCCTTCGTATTGTGCTACTCTCCTCGT	79	-14.90
4-NP H4	TAACCACATAACCGCAAGACGGGTGCAGACGTTAAACTATACCGCGTCTACGTTGGGGCTATTGTGCTACTCTCCTCGT	79	-26.16
4-NP H5	TAACCACATAACCGCAAGACCAGGTCAGTACCCACATCCTTACCTTTCGCCTCTCGTGGTATTGTGCTACTCTCCTCGT	79	-19.28
4-NP H6	TAACCACATAACCGCAAGACGACTGTCGGCCATGCTGATCTGTGATGACCTTTGGCGTCTATTGTGCTACTCTCCTCGT	79	-21.17
4-NP H7	TAACCACATAACCGCAAGACCGTGTGCCGTCGCCAATTGTGCTACACTCGCTTGTGTCGTATTGTGCTACTCTCCTCGT	79	-23.03
4-NP H8	TAACCACATAACCGCAAGACCAGACTCTGCGTTAGTTCAATCGGTATGGTTTGCCTTCGTATTGTGCTACTCTCCTCGT	79	-18.59
4-NP H10	TAACCACATAACCGCAAGACCATGACCGGACAATTAGGGAATTCACCTGTTCGCCCCTGTATTGTGCTACTCTCCTCGT	79	-17.22
4-NP H11	TAACCACATAACCGCAAGACGCCAAGGACTGTTGATGCCCATGGACGGCTGCGCGCTCGTATTGTGCTACTCTCCTCGT	79	-22.94

118

Table D.2 Secondary Structure Prediction for Aptamer Candidates. Sequences were computationally folded with primer regions through ViennaRNA. Secondary structure information was produced and is depicted below. Abbreviations: 4-NP for 4-nonylphenol; NPTE for nonylphenol technical replicates; L for low stringency; H for high stringency. Numbers refer to Cluster number.

4-NP L1	4-NP L4	4-NP L3	4-NP L2
4-NP L5	**4-NP L6**	**4-NP L7**	**4-NP L11**
4-NP L10	**4-NP L8**	**4-NP H1**	**4-NP H2**
4-NP H3	**4-NP H4**	**4-NP H5**	**4-NP H6**

4-NP H7	4-NP H8	4-NP H10	4-NP H11
NPTE L1/H1	NPTE L2/H2	NPTE L4/H3	NPTE L3/H4
NPTE L5/H7	NPTE L6/H5	NPTE L8/H13	NPTE L9
NPTE L7	NPTE H6	NPTE H10	

Appendix E.   Candidate Aptamer Characterisation Initial Screen

Oligonucleotide candidates from four libraries: low and high stringency selection for 4-NP or NPTE, were selected from bioinformatics data for characterisation. An initial medium through-put screen was conducted using AuNPs to assess binding activity. Spectrophotometry results from the initial screen are presented in in Figures E.1 and E.2.
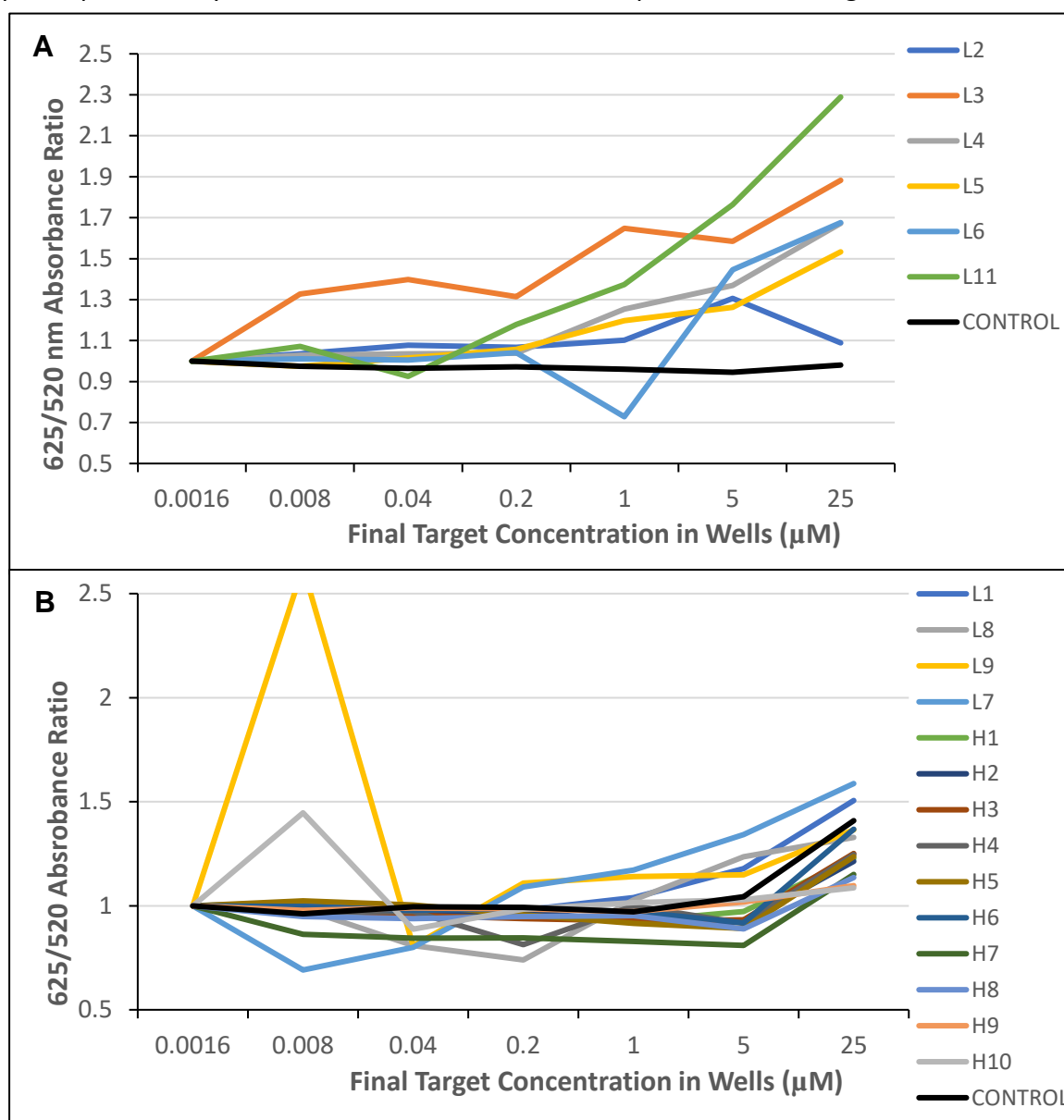

Figure E.1: Salt-induced gold nanoparticle (AuNP) aggregation after incubating with various concentrations of target molecule (nonylphenol technical equivalents; NPTE) as measured by the absorbance ratio at 625 and 520 nm wavelength. A) represents the candidates selected for further validation and B) represents the remaining oligonucleotide candidates that were not perused further. Control as seen by the black line represents the non-specific aptamer control.
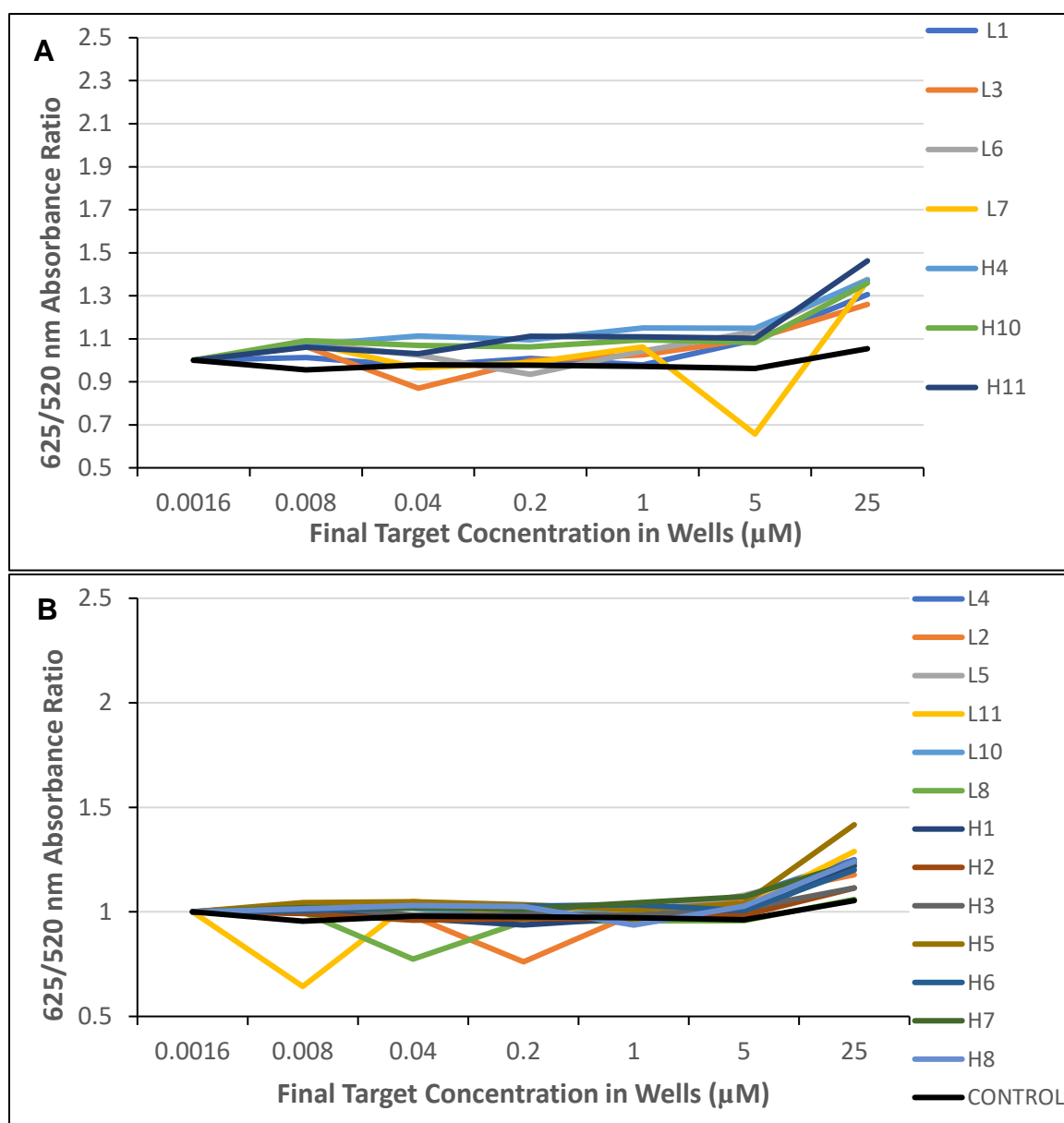
Figure E.2: Salt-induced gold nanoparticle (AuNP) aggregation after incubating with various concentrations of the target molecule (4-nonylphenol; 4-NP) as measured by the absorbance ratio at 625 and 520 nm wavelength. A) represents the candidates selected for further validation and B) represents the remaining oligonucleotides that were not perused further. Control depicted by the black line represents the non-specific aptamer control.