

# **THE GENERATION AND EVOLUTIONARY STUDY OF NOVEL APTAMERS FOR ENVIRONMENTAL CONTAMINANTS**

**ZAK LANDON MURRAY**

A thesis submitted to the Victoria University  
of Wellington in fulfilment of the  
requirements for the degree of Doctor of  
Philosophy

Victoria University of Wellington  
(2020)



# Abstract

The contamination of waterways by environmental pollutants is of growing global concern. The bio-accumulative properties of these contaminants suggest long-term impacts on many species, even those not directly exposed. There is ample evidence of the presence of environmental contaminants within biological fluids of humans, but their effects on health are largely unknown. Understanding the extent of this problem is hampered by labour-intensive extraction techniques that require expensive instrumentation and highly specialised technical expertise. Due to the prohibitive nature of routine analysis, the occurrence of many of these compounds in New Zealand waterways is unknown. Thus, a robust, portable and sensitive biosensor is urgently needed to guide regulatory agencies worldwide. Aptamers are single-stranded nucleic acid molecules that can bind to a specific target molecule with high affinity. Whilst the use of aptamers presents a novel technology to monitor small molecule environmental contaminants, the generation of high affinity aptamers has been limited. The objectives of this PhD study were to: (1) measure key emerging organic contaminants (EOC's) in a selection of New Zealand waterways covering different land-use types; (2) generate and characterise aptamers that bind three key EOC's, namely glyphosate, nonylphenol and oxybenzone, and; (3) explore the evolutionary pathways that random nucleotide libraries follow when generating aptamers to a small molecule under different stringencies. A novel bioinformatics pipeline for the analysis of high throughput (HT) SELEX data from multiple selection strategies has been developed and implemented.

The conventional method of gas chromatography mass spectrometry (GC-MS) was used to analyse water samples from waterways in the North Island of New Zealand. Technical

nonylphenol equivalents and oxybenzone were detected above the minimum detection limit (7.5 ng/L and 0.5 ng/L respectively) at all sampled sites. Concentrations of nonylphenol exceeding environmental guidelines were detected in the Waiwhetu Stream, a small stream within an industrial area in Lower Hutt. The pesticide terbuthylazine was detected at all sampled sites with particular high concentrations in the Waiwhetu and Porirua Stream. Carbendazim (a fungicide) and hexazinone (a herbicide) were also highly prevalent being detected at 87.5% and 75% of sites respectively. Glyphosate was detected at 800 ng/L in both the Waiwhetu Stream and the Porirua Stream. In general, contaminant load was much higher in urban areas than rural or forested areas. These results indicate that EOCs are present in NZ waterways and are likely to be having an impact on aquatic species.

The selection of aptamers to three key EOC's was completed using standard (glyphosate and nonylphenol) and high-throughput (HT) (oxybenzone) SELEX methodologies. DNA aptamers for glyphosate and oxybenzone were successfully generated and characterised. The GLY04 (glyphosate) and OXY-ED7-C1 (oxybenzone) aptamers were characterised using micro-scale thermophoresis and exhibited a  $K_d$  of 158 and 107.5 nM, respectively. This is the first report of a glyphosate-binding aptamer in the literature. Attempts to generate a DNA aptamer for nonylphenol were unsuccessful. Whilst five aptamer candidates were generated through 20 rounds of selection, they did not show any evidence of binding to the target molecule.

A HT-SELEX approach was utilised to study the effect of different selection parameters on the same starting library during the generation of an oxybenzone aptamer. Six strategies, compared to a standard protocol, were assessed including mutation via error-prone PCR, increased washing volume, increased detergent concentration, higher incubation temperature and negative selection and counter selection. Within each strategy, the

frequency and enrichment of candidates at each SELEX round was determined using a novel bioinformatics pipeline. On average, higher frequency candidates were present at the end of SELEX within strategies using higher stringency. Higher enrichment was also observed in the strategy using the most stringent conditions.

In summary, this PhD study presents a number of novel findings. The wide-scale presence of key EOC's in New Zealand waterways was determined. The generation of aptamers that bind to glyphosate and oxybenzone with a nanomolar affinity reveals that aptamers can be generated to such small molecules. This study also resulted in the development of a novel bioinformatics pipeline for HT-SELEX analysis that resulted in a number of recommendations on the design of such experiments. The findings presented herein highlight the possibilities and pitfalls of selecting future aptamers for EOC's and for implementing HT-SELEX experiments.



# Acknowledgments

A PhD thesis is a long, often very challenging journey. I wouldn't have got to the end of this process without the support of many people. First of all I would like to thank my two supervisors, Janet and Shalen, for all of their guidance throughout my three and half years of PhD. Shalen gave me the opportunity to undertake this journey and for that I will always be thankful. At the beginning of the project, Shalen challenged me to "make the project my own" and I can definitely say that I have done that. Janet has been an ever-calm presence; helping to make sure I am on the right track, keeping me calm (and not panicking) and has also been extremely supportive during some really tough times.

I would also like to thank Dr. Grant Northcott for all his help with the work on sampling and testing of environmental contaminants. Grant enabled us to survey rivers and streams within New Zealand for contaminants using traditional methods and without his input and help with laboratory work this would not have been possible.

This journey has been the hardest time of my life due to health challenges, which I faced as a result of some of the laboratory work required for my project. I can say with absolute certainty that I would not have got through to this point without the support of my very loving partner, Aprille. Aprille has always been there for me, motivating me to keep going and being there for me when I had lost all hope that I could complete this project. Aprille was the one who had to comfort me night after night when I woke up with my hands sticking to the sheets from the bleeding that resulted from itching my hands in my sleep, as a result of contact dermatitis from working in the lab. She has put her life on hold by staying in Wellington and supporting

me and I couldn't have done this without her. Aprille, I love you very much and thank you so much.

To my close family. Mum, Dad, Adam, and Nana, thank you for all your financial support during my PhD and for being there for me during some challenging times. You have always been so kind to me, loved me and supported me no matter what I wanted to do. I really am blessed to have such a great family and I couldn't have done this without you. Also, to Aprille's family, Pat, Patrika and Elyse (plus you, Ben), the support you have given me has also been incredible and you played a big part in enabling me to complete this journey. Thank you for your support!

To my friends, Evan, Orin and Antoine. You guys have been ever-present during this journey as well and have listened to me complaining and struggling but have always been supportive. Without our WhatsApp chat group this project would have been a lot less fun and much tougher. Thanks for all of your help, advice, humour to cheer me up and for all the inspirational breakfast meetings we had at Charley Noble. All these things were instrumental in keeping me going.

To the Auramer Bio crew and my lab mates. I want to make a special thank you to Zara, Kelly, Matire and Bex who all supported me in finishing my lab work by working alongside me. I know it wasn't always easy but I will be forever thankful that you all got me through. To Odey, Hamish and Frey, thank you for always being around to bounce ideas off, for providing advice during the aptamer meetings and in the lab, and for commiserating when results didn't go our way. Hamish, thank you especially for taking a week to help me with lab work in Hamilton. Without your help, a chapter would be missing from this thesis! Sarah, thanks for being a wonderful lab manager as well as a bundle of energy in the lab! Working with you is tons of

fun and it is great to have someone like you around! To Peter, thank you very much for all your help with lab work in the early stages of my PhD. Thank you for showing me the fluorescent imaging assay and for teaching me many of the laboratory techniques I needed to complete SELEX. To the extended Pitman lab group – thank you so much for your support, help and for all the good times we had. I will miss working with all of you.

I would like to also make a special mention here to Dinindu Senanyake and Miles Benton who inspired me to take on the challenge of learning bioinformatics and R when I knew nothing about the subject. It was a steep learning curve but very enjoyable and I gained some valuable skills. Thank you also to Dini for helping to set me up on the university HPC and for lots of initial discussions and teachings around bioinformatics. Taking the time out of your busy schedule to help me was much appreciated!

Another special mention is for Dr. Valerie Guieu who hosted me at the university in Grenoble, France. My two trips to France were amongst the fondest memories of my PhD and your great sense of humour, kind attitude and amazing support with trying to characterise my aptamers was incredible. I will always remember my time in Grenoble as an amazing experience. Thank you, Valerie! Even though we couldn't get the experiments to work, I learnt some valuable skills and had the best time.

I would also like to thank Dr. Andrew Kinghorn and Dr. Julian Tanner for great discussions regarding aptamers and for advice on bioinformatic analysis.

Lastly, to all my amazing friends, every single one of you has played a part in me being able to complete this journey. Khoi – thank you so much for always being there to listen, for providing advice. I appreciate it so much. Sean, Becky, Dan, Kareena, Tiger, Liam, Anna, Matt, Nick, Ashton, Suzie, Jim and Tina, you are all wonderful and have supported me too.

If I have forgotten anyone then please forgive me. If you have helped me in some way, thank you so much.

PhD is not a journey that can be completed alone and I am so lucky to have so many amazing people in my life to support me. Thank you everyone, I am forever grateful.

# Table of Contents

Title Page .....	1
Abstract .....	2
Acknowledgments .....	5
Table of Contents .....	9
List of Figures.....	19
List of Tables.....	27
List of Abbreviations.....	29
Chapter 1: General Introduction .....	34
1.1 Water and small molecule environmental contaminants .....	34
1.1.1 The Importance of Water .....	34
1.1.2 Global Water Pollution .....	35
1.1.3 Water Pollution in New Zealand.....	36
1.1.4 Emerging Organic Contaminants.....	37
1.1.5 Key EOC's .....	38
1.1.6 EOC Summary .....	44
1.2 Bio-Recognition Aptamers for Environmental Pollution .....	45
1.2.1 Aptamers .....	45
1.2.2 Different Types of Bio-Receptors .....	47
1.2.3 Mechanisms of Aptamer Binding .....	49
1.2.4 Applications of Aptamers for Sensing Environmental Pollutants.....	51

1.3 SELEX .....	51
1.3.1 Basic Principles .....	51
1.3.2 SELEX Parameters .....	53
1.4 Aptamer Characterisation .....	70
1.5 PhD Objectives .....	73
Chapter 2: Materials & Methods.....	74
2.1 Preparation of Oligonucleotide Library .....	74
2.2 Preparation of Binding and Blocking Buffers.....	75
2.2.1 Binding and Washing Buffer Preparation .....	75
2.2.2 Blocking Buffer Preparation .....	75
2.3 PCR Reaction Conditions .....	75
2.4 Agarose Gel Electrophoresis.....	76
2.5 DNA Gel Extraction of PCR Products .....	77
2.6 Strand Separation .....	78
2.7 Negative Selection Procedure .....	78
2.8 Counter Selection Procedure.....	79
2.9 Cloning and Sequencing of Plasmid DNA.....	79
2.10 Sanger Sequencing of Aptamer Candidates .....	81
2.11 Prediction of Secondary Structure Formation .....	81
Chapter 3: A survey of prominent emerging organic contaminants in New Zealand freshwater ways.....	82

3.1 Introduction.....	82
3.2 Methodology .....	86
3.2.1 Contaminant and Site Selection .....	86
3.2.2 Chemicals & Consumables.....	87
3.2.3 Sampling Procedure.....	88
3.2.4 Analyses of Glyphosate Residues .....	89
3.2.5 Extraction of Phenolic Compounds .....	90
3.2.6 Extraction of Pesticide Compounds.....	92
3.2.7 Quality assurance for the analysis of nonylphenols and oxybenzone .....	94
3.2.8 Quality Assurance for Pesticide Analyses .....	95
3.2.8 Gas Chromatography Mass Spectrometry of Phenolic Compounds and Pesticide Residues .....	96
3.3 Results .....	97
3.3.1 Quality Assurance Results – Oxybenzone & Nonylphenol.....	97
3.3.2 Results of survey for Oxybenzone and Nonylphenols .....	98
3.3.3 Results of Survey for Glyphosate.....	98
3.3.4 Results of survey for pesticide compounds .....	100
3.3.5 Overall results of survey for EOC's .....	101
3.4 Discussion .....	104
Chapter 4: Selection of a Glyphosate-binding aptamer using standard SELEX .....	110
4.1 Background.....	110

4.1.1 Glyphosate and Current Detection Methods .....	110
4.1.2 Aptamers as Bioreceptors for Pesticides .....	110
4.2 Methodology .....	116
4.2.1 Preparation of Affinity Matrix: Glyphosate – Bead Conjugation .....	116
4.2.2 SELEX using Glyphosate Conjugated Beads .....	117
4.2.4 Washing of Affinity Beads.....	118
4.2.5 PCR Amplification of Bound Oligonucleotides.....	120
4.2.7 Counter-Selection Rounds .....	121
4.2.8 Cloning and sequencing of plasmid DNA.....	123
4.2.10 Phylogenetic Analysis .....	123
4.2.11 Secondary Structure Prediction.....	123
4.2.12 Characterisation of aptamer candidates by MST .....	123
4.3 Results .....	125
4.3.1 Glyphosate – Bead Conjugation .....	125
4.3.2 Results of Selection Round 3 .....	126
4.3.3 Results of Selection Round 6 .....	128
4.3.4 Results of Selection Round 8 .....	129
4.3.5 Results of Selection Round 9 (after Negative Selection) .....	130
4.3.6 Results from Selection Round 10 (including Counter Selection) .....	131
4.3.7 Aptamer Candidates for Glyphosate .....	131



4.3.8 Secondary Structures of Aptamer Candidates.....	136
4.3.9 Characterisation of Aptamer Candidates by MST.....	140
4.4 Discussion .....	143
Chapter 5: Selection of a nonylphenol binding aptamer using standard SELEX.....	147
5.1 Background and Introduction.....	147
5.2 Methodology .....	149
5.2.1 Synthesis of Nonylphenol-Conjugated Affinity Matrices.....	149
5.2.2 Conjugation Assessment .....	150
5.2.3 Preparation of the Affinity Matrix for SELEX .....	151
5.2.4 Washing of Sepharose Beads .....	151
5.2.5 PCR Amplification of Bound Oligonucleotides.....	152
5.2.6 Counter Selection Protocol.....	153
5.2.7 Cloning of Aptamer Candidates into a Plasmid Vector.....	155
5.2.8 Analysis of Resulting Candidate Sequences.....	156
5.2.9 Characterisation of Aptamer Candidates .....	156
5.2.10 Characterisation of NON-12 using Micro-Scale Thermophoresis .....	158
5.3 Results .....	159
5.3.1 Conjugation of Nonylphenol to the Affinity Matrix.....	159
5.3.2 Results of Selection Round 5 .....	160
5.3.3 Results of Selection Round 6 (Negative selection round).....	160
5.3.4 Results of Selection Round 7 .....	161

5.3.5 Results of Selection Round 8 .....	162
5.3.6 Results of Selection Round 16 .....	163
5.3.7 Results of Selection Round 19 .....	164
5.3.8 Results of Cloning and Sequencing .....	165
5.3.9 Secondary Structure Characterisation of Nonylphenol Aptamer Candidates .....	168
5.3.10 Fluorescent Binding Assay Characterisation of Nonylphenol Aptamer Candidates .....	174
5.3.11 Characterisation of Aptamer Candidate NON-12 using MST .....	178
5.4 Discussion .....	179
Chapter 6: Generation of an Oxybenzone-binding aptamer using HT-SELEX .....	182
6.1 Background and Introduction .....	182
6.1.2 Introduction to High Throughput SELEX .....	183
6.1.3 Aims of the study .....	188
6.2 Materials and methods .....	189
6.2.1 Affinity Matrix Preparation .....	189
6.2.3 Description of SELEX Strategies .....	189
6.2.4 Controls .....	196
6.2.5 General SELEX Methods .....	198
6.2.6 Library Preparation for Next Generation Sequencing .....	198
6.2.8 Hardware used for analysis .....	200
6.2.9 Description of Bioinformatics Pipeline .....	200

6.2.9.1 Quality Check and Filtering.....	201
6.2.9.2 Script 1: Selection of Forward Strand .....	202
6.2.9.3 Script 2: Removal of PolyX Motifs .....	203
6.2.9.4 Script 3: Removal of Primer Sequences.....	204
6.2.9.5 Script 4 and 5: Removal of Contaminant Sequences.....	204
6.2.9.6 Script 6: Removal of Reads with Primer Binding Sites within the N40 Region .	205
6.2.9.7 Script 7: Length Filtering.....	205
6.2.9.8 Script 8: Counting Number of Sequences in each Output .....	205
6.2.9.9 Script 9: Frequency Counting and Enrichment Analysis .....	206
6.2.10 Longitudinal Enrichment of the Top 50 Candidates .....	207
6.2.11 Statistical Analysis of Frequency and Enrichment .....	207
6.2.12 Frequency Tracing of the Top 50 Candidates .....	208
6.2.13 Selection of Candidates for MST Analysis.....	209
6.2.14 Characterisation of Aptamer Candidates .....	209
6.3 Results .....	211
6.3.1 Conjugation of Oxybenzone to the affinity matrix .....	211
6.3.2 Results of SELEX Rounds.....	211
6.3.3 Quality and quantity of sequencing reads.....	230
6.3.4 Read Numbers .....	231
6.3.5 PolyX Motif Occurrence.....	235

6.3.6 Counts for TES07 Homologs .....	237
6.3.7 Counts for R18C1 Homologs.....	238
6.3.8 Number of Reads with Primer-Binding Sites in the N40 Region.....	239
6.3.9 Number of Unique Sequences at each Selection Round .....	240
6.3.10 Average frequency of Top 50 candidates in Round 7 .....	242
6.3.11 Relatedness of Top 50 Sequences (frequency) from Each Strategy .....	243
6.3.12 Longitudinal Enrichment of Top 50 candidates in Selection Rounds 6 and 7 ....	244
6.3.13 Frequency of Top 50 Candidates through SELEX Rounds 1 to 7 .....	246
6.3.14 Frequency of Individual Sequences throughout SELEX .....	252
6.3.15 Top candidates from Each Strategy .....	259
6.3.16 Candidate Characterisation by Micro-Scale Thermophoresis .....	267
6.4 Discussion .....	270
Chapter 7: General Discussion.....	276
7.1 EOCs in New Zealand waterways .....	276
7.2 Potential of Aptamer Detection Methods for EOCs .....	278
7.3 Evolution of Libraries Using Different Selection Strategies .....	281
7.4 Future Directions .....	284
Reference List .....	288
Appendix A: Materials and Equipment.....	299
A1: General Materials and Equipment .....	299
A2: Survey of NZ for selected environmental contaminants.....	303

A3: Glyphosate aptamer selection .....	305
A4: Nonylphenol aptamer selection .....	306
A5: Oxybenzone aptamer selection .....	307
Appendix B: Recipes .....	308
Appendix C: Synthetic oligonucleotide sequences .....	310
Appendix D: Scripts for Bioinformatics Pipeline .....	311
D1: Script 1 – Reverse Complement and Merge.....	311
D2: Script 2 – Removal of polyX motifs .....	312
D3: Script 3 – Primer Removal.....	313
D4: Script 4 – Removal of TES07 homologous sequences .....	314
D5: Script 5 – Removal of R18C1 homologous sequences .....	315
D6: Script 6 – Removal of reads containing primer binding sites in the N40 region ....	316
D7: Script 7 – Length trimming.....	317
D8: Script 8 – Counting the number of unique sequences in each dataset.....	318
D9: Script 9 – Enrichment Analysis .....	319
D10: Script 10 – Round Splitter (any condition) .....	320
D11: Script 11 – Frequency Counts (any condition) .....	323
D12: Script 12 – Enrichment Counts (any condition).....	327
D13: Script 13 – Graphing of average frequency for each strategy.....	331
D14: Script 14 – Graphing of average enrichment for each strategy .....	333
D15: Script 15 – Candidate Tracking.....	335

Appendix E: DNA Quantification of Libraries for High Throughput Sequencing .....	344
Appendix F: Gel Electrophoresis Images from Control Library PCR's .....	347
F1: Control Library Amplification Results for Round 3 (30 PCR cycles) .....	347
F2: Control Library Amplification Results for Round 4 (30 PCR cycles) .....	348
F3: Control Library Amplification Results for Round 4 (25 PCR cycles) .....	349
F4: Control Library Amplification Results for Round 4 (22 PCR cycles) .....	350
F5: Control Library Amplification Results for Round 5 (22 PCR cycles) .....	351
F6: Control Library Amplification Results for Round 6 (25 PCR cycles) .....	352
F7: Control Library Amplification Results for Round 6 (22 PCR cycles) .....	353
F8: Control Library Amplification Results for Round 7 (22 PCR cycles) .....	354
Appendix G: GC-MS Analysis Methods .....	355

# List of Figures

Figure 1.1 Basic principles of SELEX.....	53
Figure 2.1: DNA bands from the Gene Ruler Low Range DNA Ladder (Thermo Fisher Scientific, United States) as electrophoresed on a 4% (w/v) GTG agarose gel.....	77
Figure 3.1: Flow diagram showing the process for the extraction of phenolic compounds from environmental water samples.....	91
Figure 3.2: Flow diagram showing the process for the extraction of pesticide compounds from environmental water samples.....	92
Figure 3.3:.....	103
Summary of emerging organic contaminants detected in New Zealand waterways .....	103
Figure 4.1: Scheme for conjugating glyphosate to w-aminohexyl agarose beads via amide bond formation. ....	116
Figure 4.2: Structures of the target molecule glyphosate, and the compounds used for counter selection (glutamine, glycine, leucine, alanine and valine). ....	122
Figure 4.3: SELEX Round 3 products from differing ratios of template: glyphosate-conjugated beads electrophoresed on a 2.5% (w/v) agarose gel. ....	127
Figure 4.4: SELEX Round 6 products electrophoresed on a 2.5% (w/v) agarose gel. ....	128
Figure 4.5: SELEX Round 8 products electrophoresed on a 2.5% (w/v) agarose gel. ....	129
Figure 4.6: SELEX Round 9 products electrophoresed on a 2.5% (w/v) Agarose gel.....	130
Figure 4.7: SELEX Round 10 products electrophoresed on a 2.5% (w/v) agarose gel. ....	131
Figure 4.8: Alignment of the seven candidate aptamer sequences completed using the ClustalW algorithm in MEGA 7.0. ....	134

Figure 4.9: A maximum likelihood phylogenetic tree, computed in MEGA 7.0 with 1000 bootstrap replications, indicating the relationship between the candidate aptamer sequences from the Glyphosate SELEX. ....	135
Figure 4.10: The predicted secondary structure of GLY-01. ....	136
Figure 4.11: The predicted secondary structure of GLY-02. ....	137
Figure 4.12: The predicted secondary structure of GLY-04. ....	137
Figure 4.13: The predicted secondary structure of GLY-12. ....	138
Figure 4.14: The predicted secondary structure of GLY-15. ....	138
Figure 4.15: The predicted secondary structure of GLY-27. ....	139
Figure 4.16: The predicted secondary structure of GLY-30. ....	139
Figure 4.17: Results of the MST binding assay for glyphosate aptamer candidate GLY04... ..	141
Figure 4.18: Results of the MST binding assay for glyphosate aptamer candidate GLY12... ..	142
Figure 5.1: Scheme for conjugating nonylphenol to CL-6B Sepharose beads .....	149
Figure 5.2: Structures of the target molecule nonylphenol, and the compounds used for counter selection (Triton X-100, Tween-20, Nonidet P40 and IGEPAL CA-630). ....	155
Figure 5.3: Absorption spectra of Nonylphenol-conjugated beads.....	159
Figure 5.4: Absorption spectra of control beads (blocked with ethanolamine).....	159
Figure 5.5: SELEX Round 5 products electrophoresed on a 2.5% (w/v) agarose gel. ....	160
Figure 5.6: SELEX round 6 products electrophoresed on a 2.5% (w/v) agarose gel. ....	161
Figure 5.7: SELEX round 7 products electrophoresed on a 2.5% (w/v) agarose gel. ....	162
Figure 5.8: SELEX round 8 products electrophoresed on a 2.5% (w/v) agarose gel. ....	163
Figure 5.9: SELEX round 16 products electrophoresed on a 2.5% (w/v) agarose gel. ....	164
Figure 5.10: SELEX round 19 products electrophoresed on a 2.5% (w/v) agarose gel. ....	165
Figure 5.11: Pairwise distance comparisons between nonylphenol aptamer candidates. ..	166



Figure 5.12: Alignment of nonylphenol aptamer candidates with R18C1.....	166
Figure 5.13: Maximum likelihood phylogenetic tree showing the relationship between nonylphenol aptamer candidates and R18C1. ....	167
Figure 5.14: Predicted secondary structure of aptamer candidate NON-12.....	169
Figure 5.15: Predicted secondary structure of aptamer candidate NON-14.....	170
Figure 5.16: Predicted secondary structure of aptamer candidate NON-31.....	171
Figure 5.17: Predicted secondary structure of aptamer candidate NON-33.....	172
Figure 5.18: Predicted secondary structure of aptamer candidate NON-38.....	173
Figure 5.19: Binding curve of aptamer candidate NON-12 using the fluorescent binding assay. ....	174
Figure 5.20: Binding curve of aptamer candidate NON-14 using the fluorescent binding assay. ....	175
Figure 5.21: Binding curve of aptamer candidate NON-31 using the fluorescent binding assay. ....	175
Figure 5.22: Binding curve of aptamer candidate NON-33 using the fluorescent binding assay. ....	176
Figure 5.23: Binding curve of aptamer candidate NON-38 aptamer candidate using the fluorescent binding assay. ....	176
Figure 5.24: Binding curve of a random 75mer using the fluorescent binding assay.....	177
Figure 5.25: Results of MST characterisation for NON-12.....	178
Figure 6.1: Outline of bioinformatics pipeline.....	201
Figure 6.2: UV-Visible spectrum of control beads (conjugation with Ethanolamine), oxybenzone-conjugated beads and BPA-conjugated beads. ....	211

Figure 6.3: Gel electrophoresis image showing PCR amplicons from SELEX Round 2 that were incubated with target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies.....	212
Figure 6.4: Gel electrophoresis image showing PCR amplicons from SELEX Round 2 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies.....	214
Figure 6.5: Gel electrophoresis image showing PCR amplicons from SELEX Round 3 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies.....	215
Figure 6.6: Gel electrophoresis image showing PCR amplicons from SELEX Round 3 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. ....	217
Figure 6.7: Gel electrophoresis image showing PCR amplicons from SELEX Round 4 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies.....	219

Figure 6.8: Gel electrophoresis image showing PCR amplicons from SELEX Round 5 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies..... 221

Figure 6.9: Gel electrophoresis image showing PCR amplicons from SELEX Round 6 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies..... 223

Figure 6.10: Gel electrophoresis image showing PCR amplicons from a repeat of SELEX Round 6, with lower detergent concentrations, that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. .... 224

Figure 6.11: Gel electrophoresis image showing PCR amplicons from SELEX Round 7 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies..... 226

Figure 6.12: Gel electrophoresis image showing PCR amplicons from a repeat of SELEX Round 7, using 35 cycles of PCR, that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation

(SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. ....	227
Figure 6.13: Gel electrophoresis image showing PCR amplicons from SELEX Round 7, using 40 cycles of PCR, that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Room Temperature (RT) and Counter Selection (CS) strategies. ....	228
Figure 6.14: Quality score of sequencing data (Library S01) from Gen X Pro. ....	230
Figure 6.15: The average number of polyX motifs across the seven selection rounds. ....	235
Figure 6.16: Proportion of TES07 homologs (50% sequence identity cut-off) in each selection strategy across all selection rounds. ....	237
Figure 6.17: Proportion of R18C1 homologous sequences (50% identity cut-off) in each selection strategy across seven selection rounds. ....	238
Figure 6.18: Proportion of reads that contained primer binding sites in the N40 region in each selection strategy across the seven selection rounds ....	239
Figure 6.19: The proportion of unique sequences present in each library at each selection round. ....	241
Figure 6.20: Average frequency of top 50 candidates in the final SELEX round. ....	242
Figure 6.21: Average enrichment (fold-change) of the top 50 candidates from Selection Rounds 6 to 7. ....	245
Figure 6.22: Frequency of the top 50 candidates for the Standard and Standard plus Mutation strategies for all selection rounds. ....	247
Figure 6.23: Frequency of the top 50 candidates for the Standard + Mutation, Extra Volume and Extra Detergent strategies for all selection rounds. ....	248

Figure 6.24: Frequency of the top 50 candidates for the Extra Detergent and Room Temperature strategies for all selection rounds. ....	249
Figure 6.25: Frequency of the top 50 candidates for the Extra Detergent and Negative Selection strategies for all selection rounds.....	250
Figure 6.26: Frequency of the top 50 candidates for the Negative Selection and Counter Selection strategies for all selection rounds.....	251
Figure 6.27: Frequency of the top 50 candidates for the Standard strategy compared to Counter Selection strategy for all selection rounds. ....	252
Figure 6.28: The frequency of the top 50 candidates from the Standard strategy through all SELEX rounds. ....	253
Figure 6.29: The frequency of the top 50 candidates from the Standard + Mutation strategy through all SELEX rounds.....	254
Figure 6.30: The frequency of the top 50 candidates from the Extra Volume strategy through all SELEX rounds. ....	255
Figure 6.31: The frequency of the top 50 candidates from the Extra Detergent strategy through all SELEX rounds.....	256
Figure 6.32: The frequency of the top 50 candidates from the Room Temperature strategy through all SELEX rounds.....	257
Figure 6.33: The frequency of the top 50 candidates from the Negative Selection strategy through all SELEX rounds.....	258
Figure 6.34: The frequency of the top 50 candidates from the Counter Selection strategy through all SELEX rounds.....	259
Figure 6.35: Results of the MST binding assay for oxybenzone-binding aptamer OXY-ED7-C1. ....	269

Figure F1: Gel electrophoresis image showing PCR amplicons from the Round 3 amplification of the four control libraries. ....	347
Figure F2: Gel electrophoresis image showing PCR amplicons from the Round 4 amplification (30 PCR cycles) of the four control libraries. ....	348
Figure F3: Gel electrophoresis image showing PCR amplicons from the Round 4 amplification (25 PCR cycles) of the four control libraries. ....	349
Figure F4: Gel electrophoresis image showing PCR amplicons from the Round 4 amplification (22 PCR cycles) of the four control libraries. ....	350
Figure F5: Gel electrophoresis image showing PCR amplicons from the Round 5 amplification (22 PCR cycles) of the four control libraries. ....	351
Figure F6: Gel electrophoresis image showing PCR amplicons from the Round 6 amplification (25 PCR cycles) of the four control libraries. ....	352
Figure F7: Gel electrophoresis image showing PCR amplicons from the Round 6 amplification (22 PCR cycles) of the four control libraries. ....	353
Figure F8: Gel electrophoresis image showing PCR amplicons from the Round 7 amplification (22 PCR cycles) of the four control libraries. ....	354

# List of Tables

Table 3.1: Details of sampling sites. ....	88
Table 3.2: A summary of quality assurance protocols used in sample extraction. ....	94
Table 3.3: The percentage recovered in spiked controls and the concentration measured in negative controls of target and related compounds. ....	97
Table 3.4: Concentration of oxybenzone and nonylphenols in river water samples, including the recovery of a spike control compound. ....	98
Table 3.5: Concentration of glyphosate and aminomethylphosphonic acid in river water samples collected in 2019 and 2017. ....	99
Table 3.6: Concentration of pesticide compounds in river water samples collected in 2019. ....	100
Table 4.1: A summary of pesticide aptamers in published literature ....	112
Table 4.2: Summary of Glyphosate aptamer patent claims ....	115
Table 4.3: Washing Conditions for each round of Glyphosate SELEX.....	119
Table 4.4: PCR Conditions used for each glyphosate SELEX round.....	120
Table 4.5: Glyphosate aptamer candidates.....	133
Table 5.1: Washing conditions used in the selection of an aptamer for binding nonylphenol. ....	151
Table 5.2: PCR conditions used for each nonylphenol SELEX round. ....	153
Table 5.3: Nonylphenol aptamer candidates selected for characterisation.....	168
Table 6.1: The different selection strategies applied to each selection round. ....	190
Table 6.2: Amplification conditions for each of the control libraries. ....	197

Table 6.3: Number of reads at each processing step. ....	231
Table 6.4: The percentage of sequences of the top 50 from Round 7 (based on frequency) which appear in pairwise combinations of strategies. ....	243
Table 6.5: Top candidates from the Standard strategy .....	260
Table 6.6: Top candidates from the Standard + Mutation strategy. ....	261
Table 6.7: Top candidates from the Extra Volume strategy. ....	262
Table 6.8: Top candidates from the Extra Detergent strategy. ....	263
Table 6.9: Top candidates from the Room Temperature strategy. ....	264
Table 6.10: Top candidates from the Negative Selection strategy. ....	265
Table 6.11: Top candidates from the Counter Selection strategy. ....	266
Table 6.12: Aptamer candidates chosen for characterisation by MST. ....	267
Table E1: DNA quantities of all libraries sent for high throughput sequencing .....	344



# List of Abbreviations

$\Delta G$  = Change in Gibbs free energy

$\Delta H$  = Change in enthalpy

$\Delta S$  = Change in entropy

$\mu\text{g/L}$  = micrograms per litre

$\mu\text{M}$  = micromolar

A = Adenine

ACVM = Agricultural compounds and veterinary medicines

ANOVA = Analysis of variance

ANZECC = Australian and New Zealand Environment and Conservation Council

AMPA = Aminomethylphosphonic acid

AT = Adenine and thymine

BBP = Benzyl butyl phthalate

BDDE = 1,4-Butanediol diglycidyl ether

BDE47 = 2,2',4,4'- tetrabromodiphenyl ether

BDE99 = 2,2',4,4',5-pentabromodiphenyl ether

BDE209 = Decabromodiphenyl ether

bp = Base pairs

BPA = Bisphenol A

BWB = Binding and Washing Buffer

C = Cytosine

CD = Circular dichroism spectroscopy

CS = Counter Selection strategy

CE-SELEX = Capillary electrophoresis-systematic evolution of ligands by exponential enrichment

CPU = Central processing unit

dATP = Deoxyadenosine triphosphate

DCC = N,N'-Dicyclohexylcarbodiimide

dCTP = Deoxycytidine triphosphate

DEHP = di (2- ethylhexyl) phthalate

dGTP = Deoxyguanosine triphosphate  
DNA = Deoxyribonucleic acid  
dNTP = Deoxyribonucleotide triphosphate  
dsDNA = Double stranded DNA  
dTTP = Deoxythymidine triphosphate  
EC<sub>50</sub> = Half maximal effective concentration  
ECHA = European Chemicals Agency  
ED = Extra Detergent strategy  
ELONA = Enzyme linked oligonucleotide assay  
EOC = Emerging organic contaminant  
EPA = Environmental Protection Agency  
ESPS = 5-enolpyruvylshikimate-3-phosphate  
EV = Extra Volume strategy  
FASTA = FAST-All  
FASTQ = FAST Quality  
fM = femtomolar  
FW = Formula weight  
G = Guanine  
GC = Guanine and cytosine  
GC-MS = Gas chromatography mass spectrometry  
GCMS-MS = Gas chromatography with tandem mass spectrometry  
GFC = Glass Microfiber Filter – Class C  
GNU = GNU's Not Unix!  
GPS = Global positioning system  
HDPE = High-density polyethylene  
HEX = Hexachloro-fluorescein  
HPLC = High performance liquid chromatography  
HSNO = Hazardous substances and new organisms  
HT = High-throughput  
HTS = High-throughput sequencing  
HT-SELEX = High-throughput systematic evolution of ligands by exponential enrichment

IR = Infrared

ITC = Isothermal titration calorimetry

IUPAC = International Union of Pure and Applied Chemistry

$K_d$  = Equilibrium constant of dissociation

LAS = Linear alkylbenzene sulphonate

LB = Luria Broth

LC<sub>50</sub> = Lethal concentration, 50%

LC-MS = Liquid chromatography mass spectrometry

LCMS-MS = Liquid chromatography with tandem mass spectrometry

LED = Light emitting diode

LOD = Limit of detection

LSPR = Local surface plasmon resonance

LTS = Long term support release

MDL = Minimum detection limit

MEGA = Molecular Evolutionary Genetics Analysis

mg/L = milligrams per litre

mM = millimolar

MPI = Ministry for Primary Industries

MST = Micro-scale thermophoresis

N40 = Random oligonucleotide region with a length of 40 nucleotides

NEB = New England Biolabs

ng/L = nanograms per litre

nM = nanomolar

NMR = Nuclear magnetic resonance

NHS = N-Hydroxysuccinimide

NP = Nonylphenol

NS = Negative Selection strategy

NZ = New Zealand

OD = Optical density

PAGE = Polyacrylamide gel electrophoresis

PCR = Polymerase chain reaction

PFOA = Perfluorooctanoic acid  
PFOS = Perfluorooctanesulfonic acid  
pM = picomolar  
PNEC = Predicted no-effect concentration  
POP = Persistent organic pollutant  
ppb = parts per billion  
ppt = parts per trillion  
QA = Quality assurance  
QGRS = Quadruplex forming G-Rich Sequences  
R18C1 = an oestradiol -binding aptamer  
RAM = Random access memory  
rGO = reduced Graphene Oxide  
RNA = Ribonucleic acid  
RPM = Reads per million  
RT = Room temperature  
S = Standard strategy  
SELEX = Systematic evolution of ligands by exponential enrichment  
SEM = Standard error of the mean  
SM = Standard + Mutation strategy  
SPE = Solid phase extraction  
ssDNA = Single stranded DNA  
T = Thymine  
TAE = Tris base, acetic acid and EDTA  
TCPP = Tris (1-chloro-2-propyl) phosphate  
TDCPP = Tris (1,3-dichloroisopropyl) phosphate  
TES-07 = A testosterone-binding aptamer, used as a PCR positive control  
TPP = Triphenyl phosphate  
USGS = United States Geological Society  
UV = Ultraviolet  
UVA = Ultraviolet A  
UVB = Ultraviolet B

v/v = Volume per volume

w/v = Weight per volume

XNA = Xenonucleic acid

# Chapter 1: General Introduction

## 1.1 Water and small molecule environmental contaminants

### 1.1.1 The Importance of Water

Freshwater is the most important resource on Earth. It is crucial in every aspect to the way all species utilise this planet. Virtually every living organism relies on freshwater to survive. The Earth's surface is nearly 70% water, however only 2.5% is fresh water, consisting of glaciers, ice caps and permanent snow (68.7%), ground water (30.1%) and surface/other freshwater (1.2%) (Shiklomanov, 1993). Rivers and freshwater lakes, which are main source of accessible freshwater make up only 0.007% and 0.0002% of the world's total water, respectively (Shiklomanov, 1993). The amount of freshwater is finite, and whilst it has remained relatively constant over the past 100 million years, the amount that is clean and usable is rapidly decreasing. Clean freshwater is essential for many fundamental anthropogenic activities such as human consumption and public water supply, agriculture and aquaculture, industry, recreation and electrical power generation (Carr & Neary, 2006). Determining the future demand for, and availability of, water globally is difficult due to the complex and dynamic nature of water supply and use (Vörösmarty, Green, Salisbury, & Lammers, 2000).

Water security is defined as “the availability of an acceptable quantity and quality of water for health, livelihoods, ecosystems and production, coupled with an acceptable level of water-related risks to people, environments and economies” (Grey & Sadoff, 2007, p. 547-548). Large-scale global modelling studies have indicated that water security will be a significant global threat in the near future. In fact, 80% of the world's population lives in areas

where biodiversity threat or human water security exceeds the 75% percentile, showing the dire state of river health (Vörösmarty et al., 2010). In areas of high water security threat, pollution and water resource development (*i.e.* hydro-electric dams) were the major contributor to this.

### 1.1.2 Global Water Pollution

Changes in land use, the rise of manufacturing industries and use of synthetic chemicals, together with a growing human population have all resulted in a constant global dissemination of pollutants being discharged into freshwater resources. Water pollutants can be divided into chemical and biological pollutants. Chemical pollutants can be further divided into two sub-categories including macro-pollutants occurring in mg/L concentrations and micro-pollutants which are present at  $\mu\text{g/L}$  to  $\text{ng/L}$  concentrations (Schwarzenbach et al., 2010). Macro-pollutants include the various types of nitrogen and phosphorus and elevated levels of these can lead to algal blooms and oxygen depletion. Micro-pollutants are far more diverse with  $\sim 100,000$  synthetic chemicals registered for use globally (Schwarzenbach et al., 2010). These compounds are often toxic at very low concentrations, and can be exacerbated when the chemicals are present as mixtures. Moreover, their chronic toxicity makes their effects on human and biodiversity health very difficult to quantify. A well-illustrated example of micro-pollutants are persistent organic pollutants (POP's). This group of chemicals includes a diverse range of high-volume production chemicals as well as by-products that form from various industrial processes. They exhibit common characteristics including (1) persistence in the environment which chemical, photochemical and biological processes can't remove, (2) bio-accumulation through the food-web, (3) susceptibility to long distance dissemination resulting in global distribution and (4) a broad toxicity to living organisms. Studies have shown

critical concentrations of POP's within freshwater and marine fish, and subsequently their presence in human milk and in tissues of species who consume them (Porta et al., 2008). It should be noted that many compounds have persisted in the environment for decades but could not be identified due to analytical limitations. As detection technology improves, an increasing number of these chemicals are being classified. It is imperative that sensitive and more cost-effective detection technology be developed for the vast array of micro-pollutants being discharged into the aquatic environment.

### 1.1.3 Water Pollution in New Zealand

New Zealand is an isolated country with a small population density and an abundance of freshwater resources. Land use change to intensive farming has caused significant pressure on water quality in pastoral areas (Ministry for the Environment NZ, 2015). New Zealand's total river network is 199,641 km in length (Ministry for the Environment NZ, 2015) and is comprehensively monitored for many macro-pollutants. Regional councils monitor river water quality at a large number of catchment areas (708) across the country that are dominated by agricultural land use (Ministry for the Environment NZ, 2015). Since 1989, the National Institute of Water and Atmospheric Research (NIWA) also monitors water quality on 35 major river systems at 77 sites. The water quality measurements taken are restricted to traditional measures including water clarity (or turbidity), nitrogen and phosphorous levels, presence of *Escherichia coli* (*E. coli*) and the composition of macroinvertebrates present (Ministry for the Environment NZ, 2015). Macroinvertebrate community composition is a good indicator of overall health of a river as some species are tolerant to pollution while others are not. The overall freshwater quality at these monitored sites has decreased over



time, with 60% of sites showing a significant increase in total nitrogen from 1989 – 2013 and 51% showed an increase in dissolved phosphorous (Ministry for the Environment NZ, 2015).

The routine monitoring of our rivers only measures the traditional pollutants, with little monitoring of, or policy surrounding, contamination by micro-pollutants such as synthetic organic compounds (Bruning, 2017). A lack of a nationally co-ordinated program has resulted in the current monitoring approaches being sporadic (Bruning, 2017). A heavy focus on ‘nutrient’ contamination (nitrogen and phosphorus) and the complex nature of measuring the diverse range of synthetic chemicals used in New Zealand seems to have deterred policy makers (Bruning, 2017). Therefore, the extent of chemical contamination in New Zealand waterways is unknown. Of additional concern is the expanding breadth of literature demonstrating that synthetic organic compounds do have significant effects on both environmental and human health.

#### 1.1.4 Emerging Organic Contaminants

Emerging organic contaminants (EOC's) describe molecules from a large range of chemical classes including pesticides, industrial chemicals, flame retardants, plasticizers and chemicals used in the manufacture of personal care products (Stewart, Northcott, Gaw, & Tremblay, 2016). The emergence of synthetic chemicals together with increasing levels of agriculture and an escalation in the global population have led to a crisis of global concern.

For those organic contaminants that have tests available, extensive monitoring is hampered due to high costs and specialised expertise required to detect these compounds in complex sample matrices. These barriers means that the ubiquity of these chemicals in the environment is relatively unknown.

### 1.1.5 Key EOC's

This thesis will focus on three key EOC's that are in high use globally and in New Zealand, have a likelihood of entering the aquatic environment and present significant risks for both human and environmental health, namely glyphosate, nonylphenol and oxybenzone.

#### *1.1.5.1 Introduction to Glyphosate*

The practice of managing weeds using synthetic herbicides has increased dramatically since its introduction post-World War Two. No molecule has been more heavily used than the broad-spectrum herbicide glyphosate ( $C_3H_8NO_5P$  ; N-(phosphonomethyl)glycine) following its introduction in 1974 (Duke & Powles, 2008). Glyphosate-based herbicides are unique as they are the only herbicides that target the shikimate acid pathway in plants. Glyphosate induces the inhibition of a key plant enzyme called 5-enolpyruvyl-shikimate-3-phosphate synthase (EPSPS) (Myers et al., 2016) which is essential for aromatic amino acids synthesis (Myers et al., 2016). As the EPSPS-based pathway is not present in mammals, it was initially assumed that glyphosate-based herbicides would be unable to harm mammals, leading to glyphosate being promoted as one of the least toxic pesticides to mammals (Cuhra et al., 2016; Duke & Powles, 2008). Advocates of glyphosate-based herbicides argue that it is the most studied pesticide in history and therefore, human or environmental safety concerns would be evident if they existed. A counter-argument to this is that often studies focussed on the glyphosate technical acid and not the total components of the glyphosate-based herbicidal-formulation and that the earlier studies were conducted by agrochemical-industry scientists and not independent researchers (Cuhra et al., 2016). The most commonly sold glyphosate-based herbicide formulation is Monsanto's Round-Up. Due to Monsanto's patent expiration in the United States, many other formulations now exist on the market (e.g. Vision, Glyphos Bio,

Touchdown 480 and Infosato) which contain different glyphosate salts, different concentrations of glyphosate and varying types of surfactants leading to diverse toxicities (Annett et al., 2014). This difference in toxicity was identified as early as 1979 by Folmar, Sanders, & Julin, (1979) who assessed the toxicity of technical grade glyphosate compared to the isopropylamine salt of glyphosate, as well as the interaction with the surfactant polyethoxylated tallowamine and the commercial Round-Up formulation. The surfactant was the key factor determining toxicity of the herbicide to several species of aquatic organisms, ranging from aquatic invertebrates to teleost fish (Folmar et al., 1979). A comprehensive review of the impact of glyphosate and glyphosate-based herbicides on the freshwater environment revealed that while glyphosate-based herbicides are not approved for use in the aquatic environment in most countries, measurable quantities of both active ingredient and formulation surfactants can be detected in surface waters. Thus, contamination of New Zealand freshwater is increasing and with it, the likelihood of exposure to a range of New Zealand's evolutionary unique aquatic species (Annett et al., 2014).

#### *1.1.5.2 Use and Detection of Glyphosate*

Glyphosate has been used in New Zealand since 1976 for a variety of different applications and is the most used herbicide in the country. The registration and approval of herbicides is regulated jointly by the Environmental Protection Agency, under the Hazardous Substances and New Organisms (HSNO) Act 1996, and the Ministry for Primary Industries, under the Agricultural Compounds and Veterinary Medicines (ACVM) Act 1997 (NZ EPA, 2015) . The EPA has approved approximately 60 substances containing glyphosate under the HSNO Act and there were 91 glyphosate products registered by MPI under the ACVM as at 22 July 2019 (Ministry for Primary Industries NZ, 2019) . In New Zealand, formulations of glyphosate have even been approved for use over water (Barber, 2012), with Nufarm Glyphosate Gold used

extensively in aquatic ecosystems across the country to control the native bullrush, *Typha orientalis* (Typhaceae) (Barber, 2012). Data on the overall usage of glyphosate-based herbicides in New Zealand is relatively scarce and wholly unreliable due to the inconsistent classification of pesticides in Custom's importation records and data from the New Zealand Association for Animal Health and Crop Protection (AGCARM). However there is data to suggest a 35% increase in the sale of Glyphosate herbicide between 1999-2003 (Manktelow et al., 2005)

The forestry industry uses glyphosate-based herbicides for control of highly-competitive introduced weeds in mainly *Radiata* pine forests across the country (Rolando et al., 2013). In fact, it is one of the three most used active ingredients, with 175,000 kg being applied annually, at an application rate of 3.5 kg per hectare (Rolando et al., 2013). The application route of glyphosate in this case is aerial, meaning that the risk of non-target exposure is much higher than small scale ground-based applications. The on-going trends in overall glyphosate use, and its impact on the aquatic environment, in New Zealand are difficult to determine due to a lack of more recent data. One study has investigated the distribution of emerging contaminants in sediments of an estuary in Auckland and showed that glyphosate was present at 8 of the 13 sites measured, at concentrations ranging from 58 to 950 ng/g (Stewart, Olsen, Hickey, Ferreira, Jelic, et al., 2014). The three sites with the highest concentrations were established residential areas (Stewart, Olsen, Hickey, Ferreira, Jelic, et al., 2014). This work, to our knowledge, remains the only publicized assessment of glyphosate in the New Zealand aquatic environment.

Many studies have however been conducted in other countries to assess the environmental fate of glyphosate-based herbicides. A comprehensive study carried out within 38 states of

USA between 2001 and 2010 showed that glyphosate and its metabolite, AMPA occurred widely in the environment, and were most frequently detected in sediments, ditches, drains, large rivers and streams in more than 50% of samples at these types of sites (Battaglin et al., 2014).

Currently, the most employed detection method for glyphosate in environmental water samples is USGS Method 0-2141-09, developed by Meyer et al. (2009). This method has a detection limit of 0.02 µg/L and consists of isotope dilution, followed by online solid-phase extraction and analysis of the sample using LCMS-MS. Methods such as these are time-consuming and expensive.

#### *1.1.5.4 Introduction to Nonylphenols*

Nonylphenol ethoxylates (C<sub>15</sub>H<sub>24</sub>O) are a group of amphipathic, non-ionic surfactants that are prevalent in many industrial applications and are also commonly included in formulations of “down the drain” consumer products (U.S. Environmental Protection Agency, 2010). Some of the common applications of nonylphenol ethoxylates are detergents, cleaners, degreasers, agrochemicals, paints and coatings, dust control agents and indoor pesticides (U.S. Environmental Protection Agency, 2010). When subjected to environmental conditions, nonylphenol ethoxylates degrade into nonylphenols, which are highly toxic to aquatic organisms and exhibit moderate bio-accumulation. Both nonylphenol ethoxylates and nonylphenols are produced in large volumes and have uses which lead to widespread deposition into the aquatic receiving environment (U.S. Environmental Protection Agency, 2010). In New Zealand, nonylphenol ethoxylates currently have approval for use with controls (NZ EPA, 2019).

#### *1.1.5.5 Use and Detection of Nonylphenols*

Nonylphenol was detected in 12 out of 13 sediment samples from the estuarine receiving environment in Auckland, at concentrations ranging from 145 to 32,000 ng/g. Two sites in particular, had high reported levels of nonylphenol, Puketutu Island and Milford Marina (Stewart, Olsen, Hickey, Ferreira, Jelic, et al., 2014). These sites, which had concentrations of 32,000 ng/g and 1005 ng/g respectively, both exceed the Canadian Interim Sediment Quality Guidelines (ISQG) of 1,000 ng/g. These limits are put in place to protect aquatic organisms, indicating that the levels measured through the estuarine could have negative impacts on the environment. Detection of nonylphenols in surface water has been widely reported in many countries, such as Nigeria (Oketola & Fagbemigun, 2013), China (Y.-Z. Zhang, Tang, Song, & Li, 2009), Greece (Arditsoglou & Voutsas, 2008), Korea (Duong et al., 2010), Switzerland (Jonkers et al., 2009), Italy (Vitali et al., 2004), France (Cailleaud et al., 2007) and Taiwan (Shue et al., 2010) indicating it is a widely used EOC which commonly enters the aquatic environment.

The detection of nonylphenol in water samples is typically performed using gas chromatography – mass spectrometry (GC-MS) with mass selective detection (ASTM International, 2017). The limit of detection is 0.9 µg/L.

#### *1.1.5.6 Introduction to Oxybenzone*

Oxybenzone ( $C_{14}H_{12}O_3$  ; (2-Hydroxy-4-methoxyphenyl)-phenylmethanone), also called benzophenone-3, is a UV-filtering compound which is used as an ingredient in many personal care products, especially sunscreens. In fact, it is present in approximately 3,500 brands of sunscreen worldwide at a concentration of up to 5% (w/v). Oxybenzone was included in a list of EOC's of concern for New Zealand (Stewart et al., 2016) due to it being known as a contaminant of freshwater and marine environments. The main route of entry comes from

recreational swimmers and municipal/residential discharges. Oxybenzone is known to exhibit oestrogenic effects on freshwater fish species, such as rainbow trout. Specifically, it reduces hatching and fertilization rates of eggs and increases plasma vitellogenin levels (an *in-vivo* marker of oestrogenic activity) (Coronado et al., 2008). Additionally, it damages the larval form of many coral species and has been implicated in the rapid global decline of many coral reefs, especially those that are located near tourist hotspots. The LC<sub>50</sub> of coral exposed to oxybenzone in the light was 139 µg/L, while the coral experienced deformity at only 6.5 µg/L (Downs et al., 2016). Coral reef contamination of oxybenzone ranged from 75 µg/L to 1.4 mg/L in the U.S Virgin Islands and from 0.8 µg/L to 19.2 µg/L in Hawaii. These concentrations are likely to cause significant pathologies to the coral species present in these popular tourist areas (Downs et al., 2016). Oxybenzone has also been detected in freshwater samples. For example, oxybenzone was measured at concentrations from 34 – 2,128 ng/L in influent of a waste water treatment plant in Northern China (Li et al., 2007).

#### *1.1.5.7 Use and Detection of Oxybenzone*

The detection method for oxybenzone is typically a liquid (solid)-liquid extraction, with derivatisation and quantification by GC-MS (Jeon, Chung, & Ryu, 2006). This method delivers a detection limit in water of 5 ng/L (Jeon et al., 2006).

The extent to which oxybenzone enters the New Zealand freshwater or marine environment, has not been assessed to the best of my knowledge, most likely due to the cost of analyses. Given the environmental impacts stated in other literature, an assessment of this compound in New Zealand environments and the development of more cost-effective detection methods is essential.

#### 1.1.6 EOC Summary

There is a huge number of EOCs with different chemical classes and applications that are not mentioned here . The three EOC's that I have focused on are highly-used and likely to enter the aquatic-receiving environment and thus, have potential down-stream effects on aquatic organisms and human health. Importantly, they are not included in routine screening tests using gas chromatography mass spectrometry and are therefore very good candidates for the development of a new type of detection methodology. All three compounds listed above are of concern and a quick, robust method for detecting these compounds in water is urgently needed.



## 1.2 Bio-Recognition Aptamers for Environmental Pollution

### 1.2.1 Aptamers

Aptamers are synthetic bio-receptors made from single-stranded oligonucleotides or peptides that are selected via an *in vitro* evolution process to bind a specific target molecule. In 1990, two independent lab groups first reported the use of nucleic acid ligands that bound to a molecular target. Tuerk & Gold (1990) reported the isolation of an RNA ligand that was able to bind to bacteriophage T4 DNA polymerase (gp43) from an estimated pool of 65,536 RNA species. The pool was made up of oligonucleotides containing all possible combinations of an 8-nucleotide hairpin loop that was randomized from a known ribosome-binding site of the mRNA for the T4 polymerase. This novel method demonstrated the rapid selection of two preferred binding sequences from a population of random sequences. One of the sequences was a known wild-type binding site and the other was a variant of the wild type, with sequence differences at four positions. Both sequences bound with equivalent affinity. The developed method was coined 'Systematic Evolution of Ligands by Exponential Enrichment' (SELEX) and has been used to describe aptamer selection ever since (Tuerk & Gold, 1990). A description of SELEX and its variants can be found in Section 1.3.

Meanwhile in the same year, Ellington and Szostak reported the selection of RNA molecules that bound to a range of small molecule organic dyes which were cross-linked to agarose beads. The secondary contribution of this paper was to provide a methodological description of how to synthesise large numbers of random sequence RNA molecules. Therefore, the pool which Ellington & Szostak selected from was many orders of magnitude greater than that used by Tuerk & Gold (1990), specifically approximately  $10^{13}$  different sequences compared to 65,000. The organic dyes were attached to affinity columns and initially only 0.1% of the

RNA pool bound to the columns, however after four to five cycles of selection and mutation, the columns bound over 50% of the applied RNA pool (Ellington & Szostak, 1990). They termed the individual RNA sequences that exhibited binding properties, 'aptamers', from the Latin term 'aptus' meaning 'to fit'. This second seminal paper in aptamer literature provided the foundations for many studies that followed and demonstrated that the SELEX process could be utilised to generate high-affinity aptamers to almost any target molecule.

Initial attempts to isolate oligonucleotides that bound to specific ligands were focused on RNA. In 1992, a single-stranded DNA aptamer was isolated that was capable of binding to thrombin protease with binding affinities in the 25 to 200 nM range. These were the first aptamers that exhibited binding to a target that does not physiologically interact with nucleic acids (Bock et al., 1992).

Since these early discoveries, many DNA and RNA aptamers have been selected against a wide range of targets including proteins, small molecules, whole cells, viruses, bacteria, ions and heavy metals (McKeague, McConnell, et al., 2015). The majority of aptamers reported in scientific literature between 1990 and 2011 were selected to bind to protein targets (71%), followed by small molecules (19%), cells (7%) and lastly viruses (3%) (McKeague & DeRosa; 2012). The majority of published aptamers consist of naturally occurring nucleotides however in 2012, it was shown that genetic information can be stored and recovered from six alternative genetic polymers that are not found in nature. These artificial genetic polymers were capable of specific base pairing with DNA and polymerases were engineered to synthesize XNA (xenonucleic acid) from DNA, and reverse transcribe XNA back to DNA (Pinheiro et al., 2012). In order to test if the newly created XNA's were able to undergo Darwinian evolution and perform functions such as folding and binding to specific ligands, 1,

5-anhydrohexitol nucleic acid aptamers were generated against two different target molecules: HIV trans-activating response RNA and hen egg lysosome (Pinheiro et al., 2012). The ability to replicate and evolve artificial genetic material enables the creation of aptamers with more divergent chemistry, structural motifs and physiochemical properties, including the ability to be nuclease resistant in biological fluid (Pinheiro et al., 2012; Rangel, Chen, Ayele, & Heemstra, 2018). The first XNA aptamer capable of binding to a small molecule (ochratoxin A) was reported in 2018 (Rangel et al., 2018). The threose nucleic acid aptamer demonstrated significant advantages over its DNA and RNA counterparts. Specifically, it was able to be incubated in 50% (v/v) human blood serum for seven days and retain comparable binding affinity to the target molecule whilst the DNA aptamer degraded and lost all binding affinity to ochratoxin A.

### 1.2.2 Different Types of Bio-Receptors

There are many types of compounds available that can be used as bio-receptors for diagnostic purposes such as antibodies, affimers, nanobodies and designed ankyrins repeat proteins. However, aptamers present numerous advantages over these other types of receptors. The primary advantages are the ability to chemically synthesize DNA or RNA cheaply and consistently without the use of an animal host, therefore limiting batch-to-batch variation. Functionalization of nucleic acids and their incorporation into bio-sensors is also much easier than that of proteins. When antibodies bind to their target ligand, they do not change shape or emit electrons/photons, therefore the transduction of the binding signal typically requires multi-step washing procedures which are cumbersome and time-consuming (Vallée-Bélisle & Plaxco, 2010). In contrast, nucleic acids undergo binding-induced switches in their

conformational state, which are easier to transduce into a recognisable signal (Vallée-Bélisle & Plaxco, 2010).

Moreover as antibodies are generated in a biological system, it is problematic to generate antibodies against any compound that exhibits toxicity to its host, (Nimjee et al., 2005). The generation of a bio-receptor in a biological system also has many other drawbacks compared to an *in vitro* evolutionary system. For example, antibodies are restricted to only being functional in physiological conditions whereas aptamers can be selected to be functional in a wide range of conditions. In addition, the immune system determines the site of target binding and this cannot be optimized or preferentially selected. Lastly, antibodies cannot be raised against targets that lack immunogenicity, such as very small molecules (Nimjee et al., 2005). Unlike proteins, which undergo irreversible denaturation, aptamers have an unlimited shelf life and can return to their original conformation after temperature increases. They display no evidence of immunogenicity and are amenable to a wide range of chemical modifications that enhance their function (Nimjee et al., 2005).

Aptamers have very comparable binding affinities to monoclonal antibodies from pM to nM range and exhibit very specific binding characteristics. Some highly cited examples of the discriminatory binding ability of aptamers are those of the theophylline (Jenison et al., 1994) and L-arginine (Geiger et al., 1996) aptamer. Theophylline is a bronchodilator used in the treatment of respiratory diseases and has a similar chemical structure to both theobromine and caffeine, both of which may also be found in serum. However despite only differing from caffeine by a methyl group at a nitrogen atom at position 7, the aptamer had a 10,000-fold higher binding affinity for theophylline (Jenison et al., 1994). Likewise, an L-arginine aptamer

exhibited a 12,000-fold higher binding affinity for L-arginine than its enantiomer, D-arginine (Geiger et al., 1996).

In terms of a recognition element specifically for environmental contaminants, aptamers have huge advantages over antibodies. Most environmental contaminants are pesticides or herbicides, which are highly toxic and therefore cannot be raised in an animal host. In addition, the contaminants will need to be measured in an environmental sample, most likely a sample from a river water column. The composition of this sample matrix is very different to that of the physiological medium used to raise antibodies. An *in-vitro* selection methodology allows molecules to be preferentially selected to function in these conditions.

### 1.2.3 Mechanisms of Aptamer Binding

The techniques of nuclear magnetic resonance (NMR) and X-ray crystallography, coupled with *in silico* modelling are the most common methods used to determine the mechanism of aptamer binding. These methods allow specific identification of the molecular interactions between nucleotides and functional groups of the target molecule. Other methods, including circular dichroism (CD) spectroscopy and isothermal titration calorimetry (ITC) are able to provide more general information about binding such as whether there is a conformation change induced when the target is present and how many binding sites the aptamer may possess.

The binding of an aptamer to its target molecule relies on chemical interactions as well as conformational plasticity allowing the oligonucleotide to fold into a specific 3-dimensional structure (Famulok & Mayer, 2011). Typically, aptamers bind to their cognate ligand using the same chemical interactions as proteins do during substrate recognition. These include hydrogen-bonding, van der Waals forces and electrostatic, hydrophobic and stacking

interactions or a combination of these (Roth & Breaker, 2009). Chemical interactions are facilitated by a change in conformational structure of the aptamer to form binding pockets or 'aptatopes'. Aptamer structures are dependent on the nucleotide composition of the aptamer and its interaction with the physical features of the target but can include stems, loops, hairpins, bulges, pseudoknots and G-quadruplexes (Cai et al., 2018; Tan et al., 2016). When nucleic acids are single-stranded, they undergo a cascade of structural alignments to attain an overall stable structure by following the same rules of base-pair complementarity that double-stranded nucleic acids follow. However, there are instances when non-conical base pairing exists. It is relatively simple to predict the secondary structure of an aptamer and as such, a variety of algorithms have been created for doing this which are based on the lowest free energy of each structure (Radom et al., 2013). Unfortunately, predicting the three-dimensional DNA structure has proven more difficult, thus computationally predicted structures only loosely match the solved NMR or crystal structures (Radom et al., 2013).

There are many well-characterised aptamers and riboswitches in the literature where the exact binding mechanisms have been determined. The use of NMR has solved the structures for a variety of aptamer-small molecule interactions including malachite green, neomycin B, ATP, GTP, L-arginine, L-citrulline and an aptamer for a bovine prion peptide (Sakamoto, 2017). The latter example was particularly interesting as the NMR structure was solved for both the free and bound aptamer complex. It was demonstrated that the aptamer folded into a unique quadruplex structure in both free and bound forms. The 3D structure is a dimer allowing each monomer to simultaneously bind two regions of the N-terminal half of the prion peptide (Mashima et al., 2013). Understanding the binding mechanisms enables binding improvements to be made through post-SELEX modifications or mutations as well as suitable bio-sensing platforms to be tailored to each aptamer.

#### 1.2.4 Applications of Aptamers for Sensing Environmental Pollutants

Most environmental contaminants are very small molecules of less than 1000 kDa in size. They generally don't possess immunogenic properties in low doses and therefore antibodies cannot be generated against them. This has resulted in increased interest in using aptamers as bio-receptors for environmental pollutants for a variety of bio-sensing platforms, named aptasensors. For example, a fluorescence polarization aptasensor capable of detecting four different organophosphorus pesticides with limits of quantification ranging from 13.4 to 23.4 nM has been developed (Zhang et al., 2014). The sensor was able to detect all of the pesticides in an extract from 10 g of minced vegetable sample. Other pesticide aptasensors that utilise electrochemical impedance spectroscopy have achieved even lower detection limits. For example, a label-free impedimetric aptasensor for the detection of the pesticide acetamiprid, has a detection limit of 17 fM in spiked water samples (Fei et al., 2015).

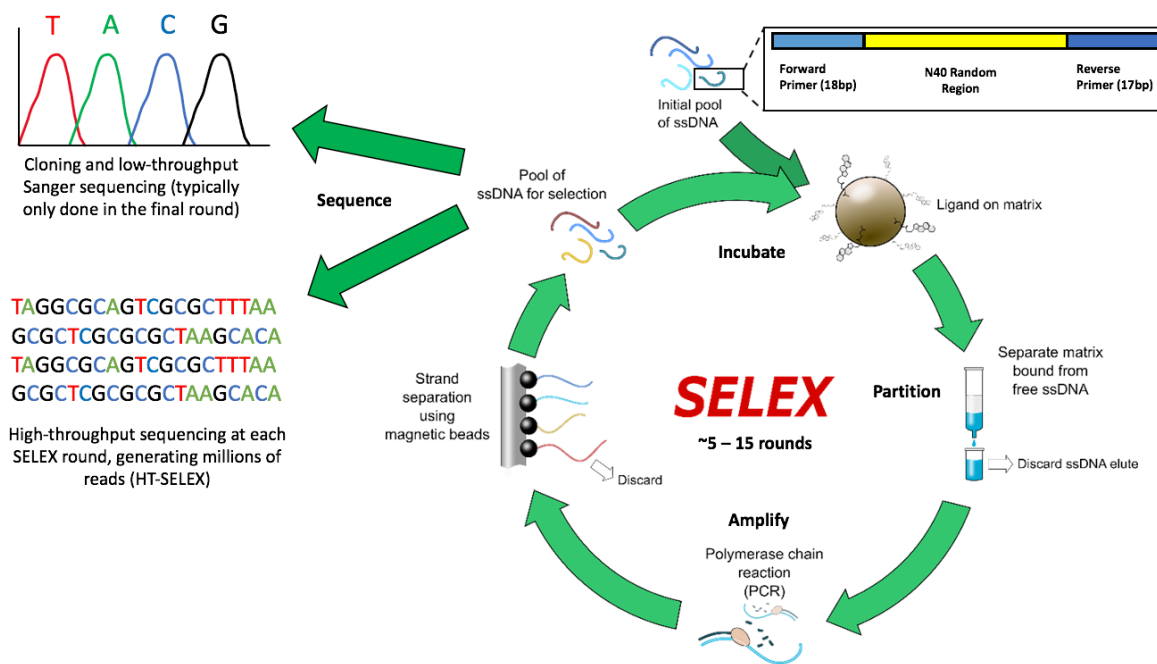
### 1.3 SELEX

#### 1.3.1 Basic Principles

At its essence, SELEX works on the principle of enrichment through Darwinian evolution by natural selection and is briefly summarised in Figure 1.1. Although Darwin's theory works on the levels of individuals and species, there are many parallels between *in vitro* and *in vivo* evolution. During SELEX, a random library of oligonucleotides containing approximately  $1 \times 10^{15}$  unique sequences are introduced to a target of interest, such as a protein, small molecule, bacteria or virus. Following incubation, oligo's that remain unbound are removed through a partitioning step such as washing. The stringency of the wash step may be altered to adjust the selection pressure and is one of a number of possible, "selection events". These

selection events remove members (*i.e.* individual sequences) from the population that do not exhibit sufficient chemical interactions with the target (*i.e.* are not fit). There are a variety of selection events that can be used to select for oligo's within the library that have high specificity to the target, without binding similarly-structured molecules. Following a selection event which essentially bottlenecks the population, the retained sequences are then amplified using polymerase chain reaction (PCR). Sequences with a fitness advantage are able to reproduce while those that do not, are eliminated from the pool and are not amplified by PCR. During the amplification process, the polymerase enzyme used to copy the DNA can be manipulated to introduce errors by introducing random mutations. As only sequences that show affinity to the target molecule are being replicated, introducing mutations into the selection pool of sequences may confer a further binding advantage. The process of random mutation increases the diversity of the population, increasing the chance that a high-affinity binder will be selected. This process of removing sequences which don't bind and amplifying those that do, is typically repeated from two up to 25 rounds. As the process is repeated, the population of oligo's becomes enriched for sequences which bind strongly to the target molecule. At SELEX completion, these sequences should make up the majority of the population. Binding conditions are typically modified to become increasingly stringent as SELEX proceeds, allowing only the "fittest" sequences to remain.





**Figure 1.1 Basic principles of SELEX.** This figure was kindly gifted to me by Dr. Shalen Kumar and his permission was given to reproduce and modify the figure.

When the library is deemed to be enriched with high affinity binders, it is sequenced to determine its composition. Specific oligo sequences are selected, often by highest frequency counts, to be synthesized and characterised for binding affinity to its target molecule. The advantage of the SELEX process is that there are so many parameters that may be modified to alter the outcome of the selection resulting in hundreds of variations to the classical SELEX procedure.

### 1.3.2 SELEX Parameters

#### 1.3.2.1 Characteristics of the Selection Pool

The first step in any *in vitro* selection experiment is to design the oligonucleotide pool such that it meets the characteristics required for the target. A typical oligo design encompasses two specific regions; a central random nucleotide region flanked on either end with known

primer regions. Typically, choosing the length of the oligo's, the length of the random region, designing the primer binding sites, and selecting the type and size of the pool are critical for success.

Selection pools can be made of RNA or DNA, and there is little evidence to suggest that either type of nucleic acid results in superior aptamers (McKeague, McConnell, et al., 2015). From 1990 to 2007, there was a large bias towards the selection of RNA aptamers with approximately 70% of selection experiments being based on RNA templates (McKeague, McConnell, et al., 2015). This shifted dramatically in 2008 to 2013, whereby 70% of experiments were based on DNA (McKeague, McConnell, et al., 2015). Typically, DNA libraries are used when the application has a requirement for a more stable chemistry that is less susceptible to nuclease degradation.

It has been shown that  $1 \times 10^{13} - 1 \times 10^{15}$  sequences is sufficient diversity to create high-affinity aptamers through SELEX. The size of the random region size may be varied from around 22 to 200 nucleotides, with the most commonly used range being between 40 and 70 nucleotides (McKeague, McConnell, et al., 2015). However, when 2,334 unique aptamer sequences with published dissociation constants were assessed, no correlation between template length and binding affinity was determined, indicating that sequence composition and structure are likely more important (McKeague et al., 2015).

#### *1.3.2.2 Incubation of Library with Target Molecule*

There are a number of variables which can be altered when incubating the library with the target molecule including the incubation, time, temperature and the amount i.e. the concentration of library that is added. Incubation temperature has been analysed as a selection parameter. When assessing aptamers incubated at 4°C, 25°C and 37°C, it was found

that aptamers selected at 37°C had a higher affinity (lower  $K_d$  value) than those selected at 4°C and 25°C (McKeague, McConnell, et al., 2015). No significant difference was reported between 4°C and 25°C (McKeague, McConnell, et al., 2015). In general, aptamers are usually selected in the temperature required for their desired application, for example if the aptamer is to be used in a physiological application, then it should be able to bind the target ligand at 37°C. To my knowledge, this meta-analysis is the only study of the effect of selection temperature on the dissociation constant. No studies on the effect of incubation time on the outcomes of SELEX have been published.

#### *1.3.2.3 Target molecule and method of partitioning*

One of the most crucial steps in determining SELEX success is the methodology used to partition the bound and un-bound oligo's. If partitioning is not effective, unbound oligos will be carried through to the next round of selection in essence diluting the pool of high affinity binders. The partitioning method used depends on the type of target molecule which the SELEX is being used for. The very first method used for partitioning bound from unbound oligos used the absorptive properties of nitrocellulose membrane for the protein target (Tuerk & Gold, 1990). This strategy works by incubating the oligonucleotide library with the target protein in solution, and then the mixture is passed over a nitrocellulose membrane for filtration (Wang et al., 2019). The proteins together with bound oligonucleotide sequences are absorbed non-specifically onto the nitrocellulose membrane while unbound oligonucleotide sequences pass over the membrane and are not retained. The limitations of this method are that it only works for proteins and that the large surface area of the membrane allows for non-specific absorption of free nucleic acid, significantly hindering enrichment of target-binding aptamers (Szeto & Craighead, 2014). Dilution of the pool by non-binders due to the non-specific properties of nitrocellulose membrane partitioning

results in a SELEX protocol requiring between 12 and 20 selection rounds (Gopinath, 2007; Wang et al., 2019).

When selecting an aptamer to a small molecule with conventional SELEX, the partitioning process is more difficult as the size difference between bound and unbound oligonucleotide is minimal. Affinity chromatography using target-conjugated beads as a solid-support matrix has been the most commonly used method to achieve partitioning for small molecule targets. Typically, researchers have been limited to targets which are able to be immobilized onto sepharose beads using common biochemical conjugation techniques. Targets that can be easily immobilized on a solid support include those with primary amines, sulfhydryl's, carbonyls, hydroxyls, carbohydrates and carboxylic acids as functional groups. A large limitation of this method is that many small molecules do not possess these functional groups. The different types of support matrices also vary with the most commonly used being agarose-based resin, sepharose, glass coverslips and magnetic beads (Wang et al., 2019). Bruno (1997) first described the use of magnetic beads as a partitioning technique in SELEX, whereby DNA aptamers for chloroamorous compounds that were conjugated to a linker arm on the magnetic beads were selected. The advantage of this method is that one can simply remove the beads together with the bound oligonucleotides using a magnet and transfer to the PCR mix, eliminating the need for an elution step (Bruno, 1997).

Filtration using affinity chromatography columns is another way to separate bound and unbound oligonucleotides. In this scheme, the target molecule is immobilized on agarose beads ranging in size between 40 and 160 micrometres, and the beads are added to a filter column. A washing buffer is passed over the beads and unbound oligonucleotides are washed away

through the pores in the filter. The beads are too large to pass through the filter and therefore bound oligonucleotides are retained and successful partitioning is achieved.

A limitation associated with targets conjugated to a solid phase is non-specific electrostatic interactions with the support matrix (Darmostuk, Rimpelova, Gbelcova, & Ruml, 2015; Wang et al., 2019). The structure of the target molecule can also be altered significantly from its native conformation during the immobilization process and some of the most important functional groups are used to form the covalent bonds. This results in the oligo library only binding to the available functional groups and the altered structure which may be a detrimental influence on the ability of the aptamer to bind to the free target molecule in solution.

A solution to the non-specific binding of oligonucleotides to the support matrix is to use 'negative selection'. This culminated from a study which attempted to select DNA aptamers for a number of organic dyes but found that after three selection cycles, there was little to no selectivity and that the majority of oligonucleotides bound non-specifically to the agarose resin (Ellington & Szostak, 1992). To address this limitation, the selection pool was incubated with the matrix itself, followed directly by incubation with the target-conjugated matrix. This step eliminated any oligonucleotides which bound to the support matrix and increased the affinity of the final aptamers by approximately 10-fold (Ellington & Szostak, 1992).

Although SELEX using immobilization of the target molecule is still the most common separation method, new variations on SELEX methodology using in-solution partitioning or library immobilization have been described in the literature and adopted by some researchers. An alternative to SELEX coined capillary electrophoresis SELEX (CE-SELEX) was introduced by Mendonsa & Bowser (2004) and partitions bound and unbound

oligonucleotides based on mobility shift. All unbound DNA migrates at a single rate while target-DNA complexes move at slower rates allowing the collection of the two separate fractions. This technique enabled the selection of high affinity aptamers to immunoglobulin E in as little as two selection rounds which was a huge improvement on the 8-12 selection rounds using conventional SELEX. In addition after four selection rounds, the oligonucleotides showed high affinity to the target. Bulk measurements showed that on average the pool had a dissociation constant ( $K_d$ ) of  $29 \pm 6$  nM (Mendonsa & Bowser, 2004). The limitation of CE-SELEX is that the target molecule must be large enough to induce a mobility shift meaning that the technique is only suitable for large molecules such as proteins.

Capture-SELEX, proposed by Stoltenburg, Nikolaus, & Strehlitz (2012), was designed to explicitly address the limitations of having to immobilize target molecules on a solid surface. Specifically designed for selecting aptamers that bind to small organic molecules, Capture-SELEX uses a short-constant region known as a “docking sequence” within the random region of the oligo, which is used to hybridize the selection pool onto magnetic beads. The basic principle behind the strategy is that any oligonucleotide that shows affinity to the target molecule and is able to undergo a target-induced structural switch will be released from the bead and remain in solution. The oligonucleotide's in solution can then be separated from the magnetic beads and amplified using PCR, as in conventional SELEX. Using this technique, aptamers to kanamycin A were selected with dissociation constants in the low micromolar range (Stoltenburg et al., 2012). Another advantage of Capture-SELEX is the ability to select aptamers against multiple targets in a single SELEX experiment (Stoltenburg et al., 2012; Lauridsen, Doessing, Long, & Nielsen, 2018). Lauridsen et al. (2018) expanded on the work completed in the seminal paper by successfully selecting RNA aptamers to rebaudioside A and carminic acid, highlighting that the approach can be used for different nucleic acid

chemistries. Capture-SELEX represents an exciting approach for the selection of aptamers to small molecule targets, which cannot be immobilized easily on a support matrix or which are required to have structure-switching properties for their implementation into a biosensor platform.

There are hundreds of different types of SELEX reported in the literature. In this dissertation, the most important variations are discussed. Most variations focus on the partitioning step, as this is crucial for selecting successful aptamer candidates. However there are other steps in SELEX which may also influence the resulting candidates. These include the type of PCR, and the methodologies used to generate, sequence and characterise these single-stranded DNA aptamer candidates.

#### *1.3.2.4 Selection Pool Amplification*

The PCR is used to amplify the selection pool after unbound oligonucleotides have been removed. This enriches the library for sequences which show affinity to the target molecule. In order to select the best aptamer candidates, the composition of the library after PCR should faithfully reflect that of the input DNA however it has been shown that this is far from the case in reality. Each individual sequence within a random library has different characteristics, such as GC content, secondary structures and melting temperatures meaning that the efficiency of amplification for each sequence will vary (Wang et al., 2019). This can lead to the enrichment of “parasitic” sequences that have not evolved to be the best binders but have instead best adapted to the PCR conditions (Wang et al., 2019). In addition, shorter sequences or by-products are generally favoured in standard PCR and amplify with a higher efficiency (Head et al., 2014). Both of these factors combine to result in a large loss of sequence information during SELEX.

By-product formation is a common problem when amplifying libraries during SELEX experiments. Musheev & Krylov (2006) showed that standard procedures for PCR amplification of homogenous DNA samples should not be applied to the amplification of random DNA libraries. When amplifying homogenous DNA, product formation proceeds until primers are exhausted; however when using a random library, product formation ceases when primers are still far in excess of the product (Musheev & Krylov, 2006). It is suggested that this is due to the mechanism underlying the formation of by-products. In the case of homogenous DNA, by-products are formed by primer-primer hybridization and are dsDNA. In contrast, when amplifying random libraries which are highly heterogeneous, by products are formed by product-product hybridization and are a combination of single-stranded and double-stranded DNA. Once a threshold amount of product is reached, then by-product formation is enabled and proceeds extremely rapidly. Product-product hybridization occurs, rapidly depleting the amount of correctly-sized product. Interestingly, Musheev & Krylov (2006) found that initial template concentration had no effect on by-product formation.

Further analyses showed that two types of unique by-products form when repeatedly amplifying libraries during SELEX. Ladder type by-products were formed when the 3' end of the random region annealed to the reverse primer binding site forming a ss-dsDNA product. After the next round of SELEX, a dsDNA product is formed with two reverse primer binding sites (Tolle et al., 2014). This allows the formation of progressively longer products with increasing numbers of reverse primer binding sites, creating a "ladder" effect. Secondly, non-ladder by-products were also formed. If the priming event occurs in the random region, then the resulting template will have random gaps meaning that it can't form ladder type products as it can only be extended once. Thus, both primer design and the target can play a role in by-product formation (Tolle et al., 2014).



When taking these studies into consideration, it can be concluded that there is a high chance of by-product formation occurring during PCR on random libraries. The risk can be limited by using analytical PCR to determine the optimum number of cycles for amplifying the library (Wang et al., 2019) as by-product formation is library and target-dependant. Ensuring libraries are not over-amplified is absolutely crucial in order to avoid by-product formation. Deep sequencing of libraries over multiple rounds using next generation sequencing (NGS) could provide further insight into how by-products form.

The type of PCR used can also have an effect on the resulting aptamers. Emulsion PCR removes the formation of by-product template heterodimers that occurs in traditional PCR. This is often used when conventional PCR failed to generate enrichment of the target library despite multiple attempts (Yufa et al., 2015). Each reaction for emulsion PCR is compartmentalized and miniaturized *in vitro* by mixing mineral oil and surfactants with traditional PCR reagents. This mixture creates micro-reactors containing ideally a single piece of template DNA and the relevant PCR components to copy it. By having each reaction isolated in its own micro-reactor, product-product hybridization is prevented. Thus, typical PCR by-products cannot form, thus improving library enrichment (Yufa et al., 2015). When comparing these two PCR methods for amplifying a heterogeneous DNA library, conventional PCR produced a significant number of by-products by the 10<sup>th</sup> cycle, but of greater concern was that by-product formation actually exceeded product generation by the 15<sup>th</sup> cycle. By the 30<sup>th</sup> cycle, all products were converted to by-products. In contrast, emulsion PCR produced no by-products and no product loss was observed (Yufa et al., 2015).

A comparison between traditional solution-driven PCR and digital droplet PCR, a type of emulsion PCR, was performed using RNA libraries that were independently progressed

through SELEX for seven rounds using the two different PCR methodologies (Takahashi et al., 2016). High-throughput sequencing compared the libraries at Rounds 1, 3, 5, 6 and 7. The traditional solution-driven, compared to digital droplet, PCR resulted in a rapid convergence in library diversity (> 70-fold and 2-fold by Round 5, respectively). Molecular evolution proceeded more slowly in the digital droplet PCR SELEX leading to the preservation of a higher sequence diversity throughout the selection experiment. This was demonstrated by the lower frequencies in the top 1000 unique sequences (Takahashi et al., 2016). When aptamers from both selection schemes were compared, the top aptamer candidate selected from the traditional solution-driven PCR showed a higher binding affinity than the non-overlapping candidates generated using digital droplet PCR. This indicates the reduced rate of molecular evolution during digital droplet PCR may limit the binding affinity of aptamers selected (Takahashi et al., 2016). Moreover, whilst both methodologies resulted in a gradual reduction in adenosine incorporation to only ~10% by Round 7, the traditional solution-driven PCR showed a bias towards thymidine nucleobases from 28.1% in the starting library to 41.1% by Round 7 whilst digital droplet PCR showed a slight bias towards guanine-rich sequences. An additional two selection rounds did not result in increased binding affinity of the aptamers generated (Takahashi et al., 2016).

Inducing mutation using error-prone PCR is another way PCR can be used to aid *in vitro* selection experiments. By introducing mutation, the sequence space of a DNA library can be increased, introducing variants that exhibit better binding affinity. *Thermus aqauticus* (*Taq*) polymerase is a low-fidelity polymerase that lacks proof reading capability. The intrinsic error rate of *Taq* polymerase varies considerably dependant on the conditions used for the PCR. Error rates as low as 1/300,000 has been observed using low magnesium chloride concentrations (4 mM) and a pH of 7.2 whilst the same conditions but with a high magnesium

concentration (20 mM) resulted in an error rate of 1/4,200 (a 71-fold decrease in fidelity) (Eckert & Kunkel, 1990). Another study was congruent, demonstrating that the optimal Mg<sup>++</sup> concentration for *Taq* fidelity was 2mM and that fidelity decreased from 2 mM up to 10 mM in a relatively linear fashion (Ling et al., 1991). A notable caveat of the standard mutation PCR approach is that it is biased towards A to G and T to C transitions, which results in a strong GC bias (Cadwell & Joyce, 1992). An approach to improve this mutational bias is to alter the ratio of dNTP's in the PCR mixture, with the optimal ratio being 5:1 ratio of dGTP and dATP to dCTP and dTTP. This ratio provided amplification which was free of GC or AT bias (Cadwell & Joyce, 1992). Error-prone PCR is common in directed evolution protocols and experiments, however it is not commonly used in SELEX. Adding additional mutational adds to the uncertain and complex outcomes of SELEX and it has been demonstrated that the 10<sup>15</sup> sequences present in a normal SELEX library provides adequate sequence diversity (Wang et al., 2019). However, it has been argued that enhanced mutation followed by progressive rounds of positive selection might be necessary in cases where the initial library does not contain a binding sequence (Lin et al., 2016). As the initial library composition is random, the authors argue that using mutagenic PCR to increase the sequence space should result in a better chance of successfully selecting a binding sequence (Lin et al., 2016). Mutation via error-prone PCR remains an interesting option in regards to increasing sequence diversity and improving aptamer binding affinity.

In conclusion, PCR bias, the formation of by-products and the type of PCR used all need to be considered when designing a SELEX experiment. Conventional PCR is by far the most commonly used method and in most cases, has resulted in the generation of high-affinity aptamers. However, in some circumstances and with certain targets, it is not successful.

#### *1.3.2.5 Creating Single Stranded DNA*

The process of PCR produces double-stranded DNA. However for the purposes of bio-recognition through incubation with the target molecule, the DNA needs to be single-stranded so that it can fold into a flexible 3D architecture. There are various methods of strand separation which can be used to achieve this. The four most commonly used methods are asymmetric PCR (Ellington & Szostak, 1992), lambda exonuclease digestion (Higuchi & Ochman, 1989), a denaturing urea-polyacrylamide gel (Williams & Bartel, 1995) and alkaline denaturation followed by separation with streptavidin-coated magnetic beads (Espelund et al., 1990).

In asymmetric PCR, one strand of the DNA template is amplified more than the other by using an unequal ratio of forward and reverse primer during PCR. Once the restricted primer is exhausted, the reaction proceeds to create only single stranded DNA in the following PCR cycles (Svobodová et al., 2012). The main advantage of asymmetric PCR is its low cost method. However although this method appears to be straightforward, the resulting PCR product contains both ssDNA, dsDNA and free primer sequences (Wang et al., 2019). An additional purification step then must be used to separate the different types of PCR products using a non-denaturing PAGE gel so the ssDNA product can be eluted from the gel. The separation and elution protocols are time-consuming and inefficient, resulting in a large sequence information loss (Marimuthu et al., 2012). In addition, the mobility of DNA in a non-denaturing gel is dependant not only on molecular weight but also on electrical charge and tertiary structure. The result is that ssDNA product appears as a smear rather than a clear band and it can be difficult to excise from the gel (Wang et al., 2019).

Exonuclease digestion (*e.g. lambda* or *T7 Gene 6*) is a method whereby one of the two DNA strands has a phosphate group introduced using the primer. The enzyme is then used to digest this strand following PCR amplification. Unfortunately this costly method is dependent upon enzyme incubation conditions and results in a varying amount of dsDNA remaining after incubation. Moreover, the purification steps required to then remove the enzyme before the selection incubation step also removes a large amount of sequence information. Finally, PCR by-product formation can incorporate the phosphate group onto the wrong strand resulting in the desired strand being digested. The advantages of the exonuclease digestion method are that it is easy to perform and results in a good yield of single-stranded (ss)DNA, with almost double the recovery of asymmetric PCR and slightly higher recovery than magnetic bead separation (Svobodová et al., 2012).

Alkaline denaturation is by far the most common method of generating ssDNA during SELEX. The PCR primer of the undesired strand is biotinylated so that when the double stranded (ds)DNA product is incubated with streptavidin-coated magnetic beads, the biotin tag upon the undesired strand covalently binds to the streptavidin molecule. The dsDNA is denatured into ssDNA through treatment with a highly alkaline solution and non-biotinylated DNA is left free in solution. The beads, together with the attached biotinylated undesired DNA strands, are removed through the use of magnets. The ssDNA in solution is neutralized and then used for the next round of selection. The method is very simple and efficient, taking approximately 30 minutes to complete. However, recent literature has demonstrated the dissociation of streptavidin from the magnetic beads during the alkaline treatment (Paul et al., 2009). Thus, the undesired ssDNA remains in solution and can reanneal to the complementary strand forming dsDNA products. In addition, streptavidin is also left in solution providing another target for aptamers to bind during the selection step (Paul et al., 2009). This method also

produces potential by-products that have multiple primers incorporated, and thus multiple biotin tags meaning that these sequences will have a higher chance of binding to the streptavidin-conjugated beads (Tolle et al., 2014; Wang et al., 2019).

The final method of strand separation which has been demonstrated to generate ssDNA from PCR products is a denaturing urea-PAGE gel separation. This method involves using a PCR primer with a spacer at the end (*e.g.* polyethylene glycol) that links it to a polynucleotide tail (*e.g.* polyA 20) to create a PCR product in which the two strands have different lengths (Wang et al., 2019). The strands are then separated using denaturing gel electrophoresis and the shorter desired strand is excised and extracted from the gel for the next round of SELEX. This method has some significant advantages. For example, all DNA including unwanted by-products, longer oligos containing the polyA tail, the expected ssDNA and free PCR primers are able to be separated. PCR by-products can therefore be removed after each round of SELEX as they should not match the length of the desired template. In addition, precise separation of ssDNA from other PCR material may be achieved resulting in higher yields of product (Svobodová et al., 2012; Rahimizadeh et al., 2017). The only disadvantage of this method is that extraction of ssDNA from the PAGE gel block is time-consuming.

As discussed above, these four methods of generating ssDNA from dsPCR product all have advantages and disadvantages. Depending on the requirements of each individual experiment, different methods may be employed. It is important however to consider the limitations of each method when designing a SELEX protocol and the effect that this may have on the resultant selection pool. Identifying these risks at the beginning of the selection process can help to prevent selection failure.

#### *1.3.2.6 Sequencing SELEX Libraries*

When sufficient enrichment is achieved, deemed by comparing the binding of the library to a target and non-target matrix, the final SELEX library is sequenced in one of two ways. The most common method is via low-throughput cloning and subsequent Sanger sequencing of successfully cloned candidates. For this approach, typically between 20 and 100 clones are sequenced upon SELEX completion to identify clones that are predominant within the library. The drawback of this approach is that a large number of rounds is required to achieve enrichment to a level where specific sequences dominate the library population. As discussed earlier, increasing the number of rounds can favour the presence of “parasitic” sequences within the library that are not true binders but are unfortunately favourably amplified during PCR.

However, high-throughput sequencing (HTS) using next generation sequencing techniques such as Illumina or Ion Torrent is an approach that is gaining momentum. HTS allows the analysis of millions of sequences from multiple selection rounds enabling the detection of highly-enriched sequences present at very low proportions (Nguyen Quang et al., 2016). This sequencing approach presents numerous advantages over low-throughput Sanger sequencing including fewer selection rounds required to identify aptamer candidates, assessing enrichment of sequences over multiple rounds and between different selection conditions and a better knowledge of aptamer binding sites and motifs (Nguyen Quang et al., 2016). Selection parameters may also be studied in order to optimize selection conditions to generate the highest affinity candidates. An example of the superiority of HTS is eloquently demonstrated in a study that used two libraries, one each from Company’s A and B. The library from Company A had four consensus sequence groups which were predominant in the top 50 unique sequences prior to selection. The most prevalent sequence group comprised

87% of reads within the top 1000 sequences, indicating that it may have been favoured during solid phase chemical synthesis. The library from Company B only had one consensus sequence group predominating, which comprised < 5% of the top 1000 unique sequences, indicating that this library had a higher sequence complexity than the library from Company A (Takahashi et al., 2016). These types of insights cannot be gained with low-throughput sequencing, as the quality of the initial library cannot be adequately assessed with such a small number of sequences. Conversely, HTS generated ~30-50 million total reads for each library.

Given the amount of information acquired from HTS, it may be used to examine the evolution of an RNA library undergoing selection. Such an example was the evolution of a library incubated with the target molecule murine Interleukin-10 receptor. At Rounds 5, 6, 9, 12 and 16, the resultant library was assessed by HTS and identified candidates were tested for binding affinity. A high-affinity aptamer that bound the target with a  $K_d$  of 27 nM was identified after 5 rounds of selection. This candidate had a much higher frequency with 283,345 copies compared to 85,037 copies for the second most frequent candidate. Interestingly, as SELEX proceeded to the point where Sanger sequencing could be used to identify candidates (Round 16), the high affinity aptamer was no longer the most represented sequence and was replaced by a moderate affinity binder (Berezhnoy et al., 2012). In addition, many high affinity binders from Round 5 were completely removed from the selection pool by Round 16. Therefore, a higher number of SELEX cycles resulted in a loss of high-affinity aptamers (Berezhnoy et al., 2012). These two studies demonstrate the powerful insights that can be obtained using HTS. These insights may then be used to further optimise SELEX methodology to guarantee the best aptamers generated through the SELEX process are



identified. Thus, high-throughput SELEX (HT-SELEX) promises to be a powerful tool for investigating SELEX parameters and selecting better aptamers in the future.

## 1.4 Aptamer Characterisation

Aptamer characterisation generally refers to the calculation of an equilibrium dissociation constant to determine binding affinity. To measure binding affinity, either the aptamer or target concentration is kept constant and titrated against an increasing concentration of the other molecules, producing a binding isotherm.

However, aptamer characterization for small molecules (< 1000 g/mol) is notoriously difficult. Depending on the number of nucleotides, aptamers can range from 10,000 – 30,000 g/mol in size. Therefore, the difference in mass often used to deduce target binding between free and bound aptamer is very small. Thus the techniques used to detect target binding must be very sensitive. Techniques such as fluorescence polarization, surface plasmon resonance , quartz crystal microbalance, high performance liquid chromatography (HPLC) and capillary electrophoresis all rely on a change in mass and are therefore not effective for small molecule-aptamer binding studies (McKeague & DeRosa, 2012) .

Many characterisation techniques require the immobilization of either the target molecule or the aptamer which can influence binding due to steric hinderance or changes in aptamer conformation. Additionally, immobilization usually requires chemical modification of either the aptamer or the target which can have unintended effects on binding affinity. Therefore, it is best to avoid immobilization if possible.

It has been demonstrated that no characterisation technique is applicable to all aptamers. Seven previously characterised aptamers selected to bind to ochratoxin A (FW 403.81 g/mol) were tested for binding affinity using a range of low-cost, common characterisation assays including equilibrium dialysis, ultra-filtration, affinity chromatography with magnetic beads,

fluorescence polarization, surface plasmon resonance, *DNase I* assay, SYBR green assay and a qualitative assay using gold nanoparticles. The results showed highly variable  $K_d$  values for each aptamer across all different methods, with no aptamer demonstrating binding across all assays. The best method that worked for six of the seven aptamers tested was affinity chromatography using magnetic beads (McKeague, De Girolamo, et al., 2015). The key finding was that multiple characterisation assays are required to draw clear conclusions about binding affinity. This study focused on cost-effective assays that could be performed in most molecular biology laboratories. Other techniques, such as isothermal titration calorimetry (ITC) and micro-scale thermophoresis (MST) that require expensive instrumentation are suggested to be more amenable to small molecule aptamer characterisation.

MST works on the principle known as thermophoresis, which is the directed movement of molecules across a temperature gradient depending on physical characteristics such as size, charge and hydration shell (Breitsprecher et al., 2016). At least one of these characteristics will change upon target binding meaning the movement of the aptamer is altered and this is used to deduce its binding affinity. The aptamer and increasing concentrations of target molecule are loaded into glass capillaries. The initial fluorescence of the aptamer is measured and then heat is applied in the form of a laser, creating a micro-temperature gradient across the capillary (Breitsprecher et al., 2016). The primary disadvantage of MST is that the aptamer has to be labelled with a fluorophore in order to track its movement within the capillary, which may influence target binding. The technique is effective for small molecule-aptamer characterization as it is not solely reliant on a change in mass upon binding and is sensitive to other changes induced when the aptamer-target complex is formed.

ITC is a technique for determining the  $K_d$  of aptamers by measuring a change in enthalpy caused by target binding. The technique is label-free and performed in-solution so binding isn't altered by immobilization or labelling. The technique works by employing two identical cells into which the sample cell is filled with either target ligand or aptamer and the reference cell with the desired application matrix (*e.g.* water, buffer) (Vogel & Suess, 2016). The aptamer or target ligand, respectively is then automatically titrated into the sample cell using a series of injections at evenly-spaced intervals. The instrument is programmed to keep both the sample cell and the reference cell at the same temperature and it is the amount of power applied to compensate for heat gain or loss with each injection that is plotted (Vogel & Suess, 2016). Aside from ligand binding, a heat change upon ligand injection may also be caused by a dilution effect or stirring, thus it is imperative that appropriate controls are in place to ensure any heat change is due to ligand binding (Vogel & Suess, 2016). As ITC measures the change in enthalpy ( $\Delta H$ ) directly,  $K_d$  can be calculated by fitting a binding isotherm. Additional thermodynamic parameters such as Gibbs free energy ( $\Delta G$ ) and the change in entropy ( $\Delta S$ ) can also be calculated using the appropriate equations (Vogel & Suess, 2016). Although ITC is a powerful technique for elucidating aptamer-target molecule interactions, there are some common problems associated with the technique. Mismatches between the buffer used to prepare the aptamer and ligand, air bubbles in the titration syringe, the presence of organic solvents to dissolve ligands and low binding heats can make it difficult to determine binding affinity (Johnson & Slavkovic, 2018).

In summary, aptamer characterization for small molecule targets is complicated and no assay has been proven to work effectively for a wide range of aptamers. Two techniques that require costly instruments and specialised training, MST and ITC, appear the most promising.

In many cases, simple techniques that can be performed with commonly available equipment have proven inconsistent.

## 1.5 PhD Objectives

The overall hypothesis of this PhD study was that a number of EOC's are present in New Zealand's freshwater environments due to urban, agricultural and manufacturing outputs. An overarching objective of this study is to generate aptamers for a range of EOC's that may be adapted into biosensor to address the void of simple and affordable detection methods to enable the extent of contamination to be realised. During the completion of this thesis, the difficulty of generating aptamers to bind small target molecules was realised and HT-SELEX was employed to demonstrate the best SELEX conditions.

Therefore, the specific objectives of this thesis are to:

1. Determine the common EOC's in a range of NZ freshwater environments using traditional sampling techniques and analyses;
2. Generate and characterise aptamers for both glyphosate and nonylphenol using SELEX with Sanger sequencing, and;
3. Compare different selection strategies for selecting the best aptamer to a small molecule target (oxybenzone) using HT-SELEX for the purpose of identifying the optimal selection strategy for generating aptamers against small molecule targets

# Chapter 2: Materials & Methods

This chapter describes the general materials and methods used for all SELEX experiments in this study. These methods were adapted from Kumar (2014), Lin et al. (2016) and Li (2016). The generation of each set of aptamer candidates for the selected target molecules varied depending on the characteristics of the targeted molecule and the results of each selection round. These variations are described in each relevant chapter (*i.e.* Chapters 4, 5 and 6 for glyphosate, nonylphenol and oxybenzone aptamer selections, respectively).

## 2.1 Preparation of Oligonucleotide Library

For all aptamers generated herein, an oligonucleotide library with a central randomized region of 40 nucleotides (N40 region) flanked by two known primer sequences of 18 (forward) and 17 (reverse) base pairs (bp) in length were used as the starting input for SELEX (see Appendix C for sequences). The oligonucleotide library was synthesized and purchased from Thermo Fisher Scientific (Waltham, Massachusetts, United States of America). Previous studies from our research group have demonstrated the selection of aptamers for small molecules using a similar library, indicating that 40 randomized nucleotides provides sufficient diversity to create aptamers with high binding affinities. The reverse primer region of each oligonucleotide is modified by the addition of a biotin tag at the 5' end. This tag is used to separate the double-stranded PCR product into single stranded DNA at the end of each SELEX round. All vials of commercially-synthesized oligonucleotides (*i.e.* primers and N40 library) were rehydrated using the same method. Each vial was centrifuged for 2 minutes at 13,300 x g and then opened in a Type 2 Biosafety Cabinet (Labconco, Kansas City, Missouri, United States of America). The oligonucleotides were rehydrated to a concentration of 100

μM using Ultra-Pure water (Thermo Fisher Scientific, Waltham, Massachusetts, United States of America). The primer sequences were further diluted to a concentration of 10 μM for use in all PCR reactions. All oligonucleotides were stored at -20°C until required.

## **2.2 Preparation of Binding and Blocking Buffers**

### **2.2.1 Binding and Washing Buffer Preparation**

The binding and washing buffer (BWB) was prepared as a 10X stock in a final volume of 100 mL of 100 mM NaCl, 5 mM KCl, 2 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub> and 20 mM Tris-HCl at a pH of 7.5. The BWB was autoclaved prior to use. For use in washing procedures, BWB was diluted to 1X using distilled water and a reagent grade, non-ionic surfactant, IGEPAL CA-630 (Sigma Aldrich, St Louis, Missouri, United States of America) was added at different concentrations depending on the desired selection pressure.

### **2.2.2 Blocking Buffer Preparation**

A blocking buffer of 5% (w/v) of skim milk powder (98%) in 1X BWB was prepared immediately prior to use.

## **2.3 PCR Reaction Conditions**

All PCR were performed on a Master Cycler Vapo Protect PCR machine (Eppendorf, Hamburg, Germany) with a final reaction volume of 50μL, unless otherwise specified. All reagents were purchased from Thermo Fisher Scientific (Waltham, Massachusetts, United States of America). The final reaction consisted of Ultra-Pure water, 1X PCR Buffer, Mg<sup>2+</sup> (concentration varied depending on the desired mutation rate), 240 μM of dNTP's, 0.22 μM of forward primer, 0.22 μM of reverse primer and 1 unit of DNA *Taq* Polymerase. Three different

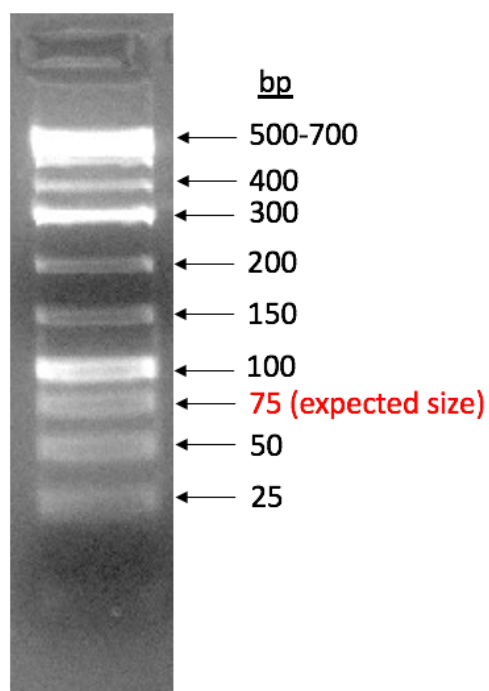
template amounts of 1  $\mu\text{L}$ , 4  $\mu\text{L}$  or 10  $\mu\text{L}$  were typically used during each SELEX round. The cycling conditions were kept consistent throughout but the number of cycles varied in accordance with the amount of DNA observed in the previous round and the wash stringency prior to PCR. The PCR cycles were initiated with a 95°C denaturation step for 5 minutes followed by 25 to 35 cycles (depending on the result of the amplification for each template amount) of a 94°C denaturation step for 40 seconds, a 53°C annealing step for 30 seconds and a 72°C extension step for 15 seconds. The PCR was terminated following a final 72°C extension step for 60 seconds. Following PCR, the reaction products were stored at 4°C prior to loading onto a gel for electrophoretic separation.

## **2.4 Agarose Gel Electrophoresis**

Agarose gel electrophoresis was used to visualize the size and density of PCR products following amplification and analyse the results of each SELEX round. An aliquot of 25  $\mu\text{L}$  of PCR product was first incubated with 1.6  $\mu\text{L}$  of 100X SYBR Green for 8 minutes at room temperature. A 6X DNA loading dye (Qiagen, Hilden, Germany) was then added to the PCR products to give a final concentration of 1X and mixed by repeated pipetting. The products were then loaded into wells of a 2.5% (w/v) agarose gel made with 1X TAE buffer (Appendix B). The gels were loaded into a Mini-Sub Cell GT Gel Tank (Bio-Rad, Hercules, California, USA), submerged in fresh 1X TAE buffer and electrophoresed for 40 minutes at 90 Volts. Gels were visualized on an Omega Lum™ G Imaging System (San Francisco, California, United States of America). The location of dsDNA bands were compared with a GeneRuler Low Range DNA Ladder (Thermo Fisher Scientific, Waltham, Massachusetts, United States) to confirm a 75-base pair size (see Figure 2.1 for a description of the size of different DNA bands). The bands



of interest were then excised from the agarose gel using a scalpel blade and stored at -20°C until required.



**Figure 2.1:** DNA bands from the Gene Ruler Low Range DNA Ladder (Thermo Fisher Scientific, United States) as electrophoresed on a 4% (w/v) GTG agarose gel.

## 2.5 DNA Gel Extraction of PCR Products

Double stranded DNA was extracted from the excised gel bands using a Qiagen Min Elute Gel Extraction Kit. The manufacturer's guidelines were followed but some modifications were made to the protocol. Specifically, to maximise the retention of DNA sequences on the column, the dissolved gel (after the addition of QG Buffer and isopropanol) was passed through the column 3-5x before it was discarded. Following this, 500  $\mu$ L of QG buffer was added to the column prior to centrifugation at 6000  $\times g$  for 1 minute as an additional wash step. The dsDNA was eluted from the column using a 1:50 dilution of Elution Buffer and the elution was carried out in two steps. Firstly, 30  $\mu$ L of the diluted buffer was added to the

column, incubated for one minute and centrifuged at  $13,300 \times g$  for 1 minute. A second elution using 20  $\mu\text{L}$  Elution Buffer was then performed to give a total volume of 50  $\mu\text{L}$  of eluted dsDNA. The dsDNA was then stored in a 650  $\mu\text{L}$  microcentrifuge tube at  $-20^{\circ}\text{C}$  until required for further use.

## 2.6 Strand Separation

Double stranded DNA was strand separated to produce ssDNA using an alkaline denaturation protocol. An aliquot of 30  $\mu\text{L}$  of streptavidin magnetic beads (New England Biolabs, Ipswich, Massachusetts, United States) were pipetted into a 650  $\mu\text{L}$  microcentrifuge tube. The bead portion of the mixture was collected using a magnetic Pick pen and washed three times in 30  $\mu\text{L}$  of BWB. The magnetic beads were then transferred into the dsDNA, released and incubated at room temperature for 30 minutes. Following incubation, the beads were collected and washed twice with 30  $\mu\text{L}$  of 1X BWB and once with 30  $\mu\text{L}$  of distilled water. After washing, the beads were placed in 30  $\mu\text{L}$  of 0.3 M NaOH for 3 minutes. While incubating, the beads were stirred using the silicone tip of the pic pen. Following removal of the beads from the NaOH solution, an aliquot of 7.8  $\mu\text{L}$  of 1 M HCl was added to neutralise the pH. To further stabilise the ssDNA prior to overnight incubation with the target molecule, 60  $\mu\text{L}$  of 0.1% (v/v) IGEPAL in 1X BWB was added to the ssDNA.

## 2.7 Negative Selection Procedure

A negative selection step was performed in order to remove any DNA sequences that bind the control matrix. The negative selection step used for all SELEX experiments followed the same basic protocol. The ssDNA from the previous selection round was divided evenly in two portions. The portions were incubated overnight with either the control matrix or the target-

conjugated matrix. Following incubation, the control beads are spun down for 4 minutes at 13,300 x g. From the control bead incubation, 150  $\mu$ L of the supernatant was removed without disturbing the bead pellet and split into two separate tubes containing either control beads or target-conjugated beads in 60  $\mu$ L of 1X BWB for a second incubation step. This step was to confirm that any sequences that bind specifically to the control beads were removed. An additional 150  $\mu$ L of 1X BWB was added to the control bead pellet in order to re-suspend the pellet. The centrifugation step was repeated, and the supernatant was once more removed and added into the two portions which were removed previously for a second incubation. The wash step was then repeated once more and the two tubes were put on a shaking incubator at 500 rpm for 4 hours at room temperature. The two additional washing steps were to ensure all of the sequences that do not bind to the control beads were collected. Both the control and target beads from the original incubation were washed as previously described and then these, together with the two second incubation tubes from the control beads were amplified by PCR. The results of these gels were compared to ascertain if the negative selection was successful.

## **2.8 Counter Selection Procedure**

Each counter selection protocol was specific to the aptamer being generated. Specific details of the reagents and protocols for counter selection are provided in each relevant aptamer generation chapter (*i.e.* Chapters 4, 5 and 6).

## **2.9 Cloning and Sequencing of Plasmid DNA**

The gel-extracted PCR products were ligated into a pGEM-Teasy plasmid vector (Promega, Madison, Wisconsin, United States of America) according to the manufacturer's instructions.

Briefly, whilst keeping all reagents on ice, 3  $\mu$ L of PCR product was added to 5  $\mu$ L of Ligase buffer, 1  $\mu$ L of the plasmid vector and 1  $\mu$ L of *T4 ligase*. The ligation reaction was incubated at room temperature for two hours.

The ligated plasmid vector was transformed into *E. coli* DH5-alpha competent cells (Thermo Fisher Scientific, Waltham, Massachusetts, United States). An aliquot of 5  $\mu$ L of the ligation mix was added to 50  $\mu$ L of DH5-alpha cells and incubated for 30 minutes on ice. Following this incubation, the cells were subjected to heat shock at a temperature of 37°C in a water bath for 2 minutes before being placed on ice for an additional 2 minutes. The cells were grown in 950  $\mu$ L of Luria Broth (LB; recipe in Appendix B) for 2 hours at 37°C at a shaking speed of 250 rpm. Thereafter, the cells were centrifuged at 3,000 rpm for 5 minutes and the supernatant was carefully removed and discarded. An additional 100  $\mu$ L of LB was added to re-suspend the bacterial pellet and the re-suspended mixture was pipetted onto LB agar plates (Appendix B) containing 2% (v/v) X-gal and 50  $\mu$ g/mL of ampicillin. The LB agar plates were incubated overnight at 37°C. The transformed colonies, indicated by a white colour, were selected and added to 10 mL of LB containing 10  $\mu$ L of 50  $\mu$ g/mL ampicillin. These vials were incubated for a further 18 hours at 37°C with a shaking speed of 250 rpm. Following incubation, the plasmid DNA was purified from the bacterial cells using a Qiagen Miniprep purification kit, according to the manufacturer's instructions. The quantity and quality of the DNA was assessed using a Nanodrop (Thermo Fisher Scientific, Waltham, Massachusetts, United States). The purified dsDNA was stored at -20°C until DNA sequencing was performed. Samples were sent for Sanger sequencing at the Massey University DNA sequencing facility.

## **2.10 Sanger Sequencing of Aptamer Candidates**

Raw sequence data was received from the Massey University sequencing facility in a FASTA format. The first step of the analyses performed was to manually edit the sequences to remove the plasmid DNA regions by searching for the flanking primers. Any sequences in which the insert could not be located or was of poor quality were not used in further analyses. The sequences were then imported in MEGA 7.0 and were aligned using the ClustalW algorithm with the default alignment settings (Kumar et al., 2016). Sequences which were identical were removed and the relationship between the remaining sequences was analysed using a Maximum Likelihood phylogenetic tree with 1000 bootstrap replications (Kumar et al., 2016). The parameters for generating the phylogeny were the default parameters aside from the addition of bootstrapping.

## **2.11 Prediction of Secondary Structure Formation**

Aptamer secondary structures were predicted using the 'Structure Display and Free Energy Determination' function on the mFold web server. The default parameters were used with exception of temperature at 23°C, Na<sup>+</sup> ion concentration at 10 mM and the Mg<sup>2+</sup> ion concentration at 0.2 mM (Zuker, 2003). These conditions were to simulate the environment in which the aptamer was generated in.

# **Chapter 3: A survey of prominent emerging organic contaminants in New Zealand freshwater ways**

## **3.1 Introduction**

Environmental monitoring data, which is collected by regional councils across New Zealand and collated by the Ministry for the Environment, has shown that water quality is highly variable. Waterways in farming areas are polluted, while both air, water and land are polluted in urban areas (Ministry for the Environment NZ, 2019). The water quality monitoring currently undertaken in New Zealand focuses on macro-environmental pollutants such as pathogenic bacteria (or indicators of), phosphorous and nitrogen. Unfortunately, monitoring of chemical contamination is sporadic, leading to an information void on the presence of environmental contaminants in New Zealand's waterways. It could be argued that the breadth of potential chemical contaminants and the technologically demanding methods required for their detection, makes the task of monitoring their occurrence in the environment too expensive for local or central government to address. Thus, a lack of legislative pressure and cost-effective testing methods has culminated in a situation where the occurrence of chemical contaminants, termed EOC's, in the New Zealand freshwater environment is largely unknown. The lack of data on these contaminants within the

environment presents a growing concern, not only for the unique flora and fauna of New Zealand, but also for public health.

A limited number of studies have provided some preliminary evidence that at least some of these contaminants are present with New Zealand's waterways. The presence of 32 EOC's within four Dunedin waterways were assessed based on their high occurrence in comparable surveys in similar countries and on the availability of analytical methods. The results were compared to a recent freshwater survey in the United States (Bernot et al., 2019). Four compounds (caffeine, carbamazepine, cotinine and diethyltoluamide) were detected in at least one water sample and diethyltoluamide was the only compound detected in a sediment sample (Bernot et al., 2019). The other 28 compounds were not present above the detection level. Diethyltoluamide, an active ingredient in insect repellents, was detected in 75% of water samples at concentrations ranging between 5.5 - 510 ng/L. Overall, the average concentrations of the four detected compounds were comparable to those detected in US freshwaters (Bernot et al., 2019).

In another study, the presence of seven common pesticides in 36 streams within agricultural areas in the Waikato, Canterbury, Otago and Southland regions was examined (Hageman et al. (2019). These sites were selected based on having high nutrient contamination, however the study found that there was no correlation with high nutrient and pesticide load (Hageman et al., 2019). Multiple pesticides were found at most sites, with two or more pesticides being detected at 78% of sites. Chlorpyrifos was the most commonly detected compound (83% of sites), followed by atrazine (72%), diazinon (67%) , 2,4-Dichlorophenoxyacetic acid (53%) and the three neonicotinoids imidacloprid, clothianidin and thiamethoxam (22%, 8% and 3%, respectively) (Hageman et al., 2019). Chlorpyrifos was detected at two sites at concentrations

higher (*i.e.* both 180 ng/L) than that deemed to have no observable effect (*i.e.* the 'No Observable Effect Concentration' of 140 ng/L), indicating that the compound could be having an effect of the biodiversity of aquatic organisms within the streams (Hageman et al., 2019).

The presence of EOC's within 13 estuarine receiving environment sites around Auckland City have also been assessed. The contaminants tested included flame-retardants, plasticisers, alkylphenols, pesticides, pharmaceuticals and oestrogens. Of note, the phthalate plasticiser DEHP was measured at 11,500 ng/g at one site, while nonylphenol was detected at 32,000 ng/g at a site adjacent to the city's largest wastewater treatment plant (Stewart et al., 2014). Glyphosate concentrations of up to 950 ng/g were also detected at residential sites, with a median concentration of 120 ng/g (Stewart et al., 2014).

These previous studies reveal that EOC's can be detected in New Zealand freshwater systems in very different environmental settings (urban vs. rural catchments) and at environmentally relevant concentrations. In addition, these studies identified a lack of monitoring of EOC's in New Zealand and advocated the need for comprehensive data on the abundance of EOC's in the New Zealand environment.

Stewart et al. (2016) recommended a tiered monitoring approach whereby measurement of representative "EOC markers" is performed at a large number of sites. The first tier of monitoring would assess the occurrence of the "EOC markers" and then the most highly impacted sites could then be regularly monitored for a refined list of EOC's (Stewart, Northcott, Gaw, Tremblay, et al., 2016). The "EOC markers", based on a number of criteria, were suggested to include a single compound from a wide range of chemical classes including flame retardants (*e.g.* BDE47, BDE99, BDE209, TDCPP, TPP and TCPP), plasticisers (*e.g.* DEHP, BBP and Bisphenol A), surfactants (*e.g.* nonylphenol and LAS), perfluorinated compounds (*e.g.*



PFOS and PFOA), musk fragrances (*e.g.* galaxolide and tonalide), pesticides (*e.g.* glyphosate), neonicotinoid insecticides (*e.g.* imidacloprid), pyrethroid insecticides (*e.g.* bifenthrin and permethrin), pharmaceuticals (*e.g.* acetaminophen, diclofenac, ibuprofen and carbamazepine), steroids (*e.g.* estrone), personal care products (*e.g.* triclosan and methyltriclosan), preservatives (*e.g.* methylparaben), corrosion inhibitors (*e.g.* benzotriazole), antifouling agents (*e.g.* diuron and isoproturon) and UV-filters (*e.g.* oxybenzone) (Stewart et al., 2016).

The study presented in this chapter tested the presence of the six EOC's at a variety of freshwater sites associated with different land uses across the lower and middle regions of North Island, New Zealand. The six EOC's selected were glyphosate (a herbicide), nonylphenol (an industrial detergent used in common 'down the drain' applications), oxybenzone (a UV-filter which is commonly used in sunscreens), and cypermethrin, chlorpyrifos and fipronil (all insecticides). My hypothesis was that waterways in urban areas would have the highest concentration of contaminants from personal care products (*i.e.* oxybenzone and nonylphenol), whilst waterways in rural areas would have higher concentrations of pesticides (*i.e.* glyphosate, cypermethrin, chlorpyrifos and fipronil).

## 3.2 Methodology

### 3.2.1 Contaminant and Site Selection

The six contaminants (*i.e.* oxybenzone, nonylphenol, glyphosate, cypermethrin, chlorpyrifos and fipronil) selected for testing in this study were based on a combination of characteristics including their ability to bio-accumulate in the environment, likelihood to enter the aquatic environment, association with eco-toxicity or human health concerns, high usage in New Zealand, documented detection in the freshwater environment of New Zealand or a similar country and/or being identified as a ‘contaminant of concern’.

Eight sampling sites representing a variety of different land use types (lowland forest, urban and rural) were selected across four different catchment areas in the lower and middle regions of North Island, New Zealand (Table 3.1). The Hutt River catchment had three sampling sites: (1) Hutt River at Boulcott, (2) Hutt River at Te Marua Intake and (3) Waiwhetu Stream at Parkside Road. The Te Marua intake site was chosen as it is one of the drinking water intake sites for the Wellington region and is located downstream of native forest so we would not expect contamination with any of the targeted molecules. The Hutt River is heavily used for recreational swimming and the Boulcott Street site is downstream of some large residential suburbs, thus was chosen to represent a typical urban New Zealand river. The Waiwhetu Stream site is downstream of a large industrial area with many chemical manufacturers. It is a small stream with a low flow rate meaning contaminants are concentrated. This site is representative of streams flowing through industrialized areas. The site on the Porirua Stream was chosen, as there was a desire to include a roughly equal split of urban, rural and forested sites. The Porirua Stream is another small stream that represents a mixture of residential and industrial inputs in an area with high population density. The two

sites on the Manawatu River are both rural sites. Upstream of both sites, the primary land use is dairy farming and other types of agriculture. The Manawatu River is also used for recreational boating and swimming at the Foxton site, thus a higher input of phenolic compounds at this site would be expected. The Waikato River has both agricultural and urban inputs. It is used for recreational swimming and boating but also sits adjacent to some of the largest dairy farms in the country. Most importantly, the level of EOC contamination in the water column at these sites is unknown. Collectively, assessing EOC load at all of these sites should give a good indication of the presence of these compounds in the North Island of New Zealand. . All sites are at or near regularly monitored sites by the governing regional councils. The sampling dates for each site are presented in Table 3.1. Sampling was completed during the summer months, as this was the safest time to access waterways. A set of initial samples was taken in November 2017 however due to some of the samples being compromised in storage; another set of samples had to be taken in February 2019.

### 3.2.2 Chemicals & Consumables

All information regarding the chemicals and consumables used in this chapter are listed in Appendices 1 and 2, respectively.

**Table 3.1: Details of sampling sites.**

<b>Name of Site</b>	<b>Catchment</b>	<b>GPS Co-ordinates</b>	<b>Land-Use Type</b>	<b>Sampling Dates (DD/MM/YYYY)</b>
Hutt at Boulcott	Hutt River	41°11'51.21"S 174°55'9.33"E	Lowland forest	29/11/2017 23/02/2019
Hutt at Te Marua Intake Site	Hutt River	41°04'47.7"S 175°08'37.1"E	Lowland forest	29/11/2017 23/02/2019
Waiwhetu Stream at Parkside Road	Hutt River	41°14'13.1"S 174°54'33.3"E	Lowland urban	29/11/2017 23/02/2019
Porirua Stream at Dog Park	Porirua Stream	41°08'41.3"S 174°50'35.0"E	Lowland urban	23/02/2019
Manawatu at Foxton (Whirokino)	Manawatu River	40°30'36.1"S 175°15'56.2"E	Lowland rural	24/02/2019
Manawatu at Teachers College	Manawatu River	40°22'24.4"S 175°37'37.2"E	Lowland rural	29/11/2017 24/02/2019
Waikato River at Narrows Boat Ramp	Waikato River	37°50'22.5"S 175°20'51.7"E	Lowland forest	29/11/2017 24/02/2019
Waikato at Hamilton Traffic Bridge	Waikato River	37°47'29.0"S 175°17'22.2"E	Lowland forest	29/11/2017 24/02/2019

### 3.2.3 Sampling Procedure

Two grab samples were taken at each site which consisted of collecting water samples at approximately 1 m from the edge of the riverbank where the depth of the water was approximately 40 cm. The sampling bottle (described below) was submerged into the water at a depth of approximately 20 cm from the surface and was removed when completely full. A 250 mL water sample was collected into a pre-cleaned Nalgene HDPE bottle for the analysis of glyphosate and a 4 L sample was collected into a pre-rinsed 4 L amber glass bottle with a

Teflon-lined lid for the analysis of oxybenzone, nonylphenol, cypermethrin, chlorpyrifos and fipronil. Pre-cleaning of the Nalgene HDPE bottles were performed by Asure Quality. The 4 L glass amber bottles were pre-rinsed three times with 100 mL of acetone, followed by rinsing three times with 100 mL dichloromethane and rinsing an additional three times with 100 mL methanol. After sampling at each site, all samples were immediately stored on ice for transportation to the laboratory. Water samples destined for glyphosate analyses were frozen at -20°C upon arrival back to the laboratory. These samples were transported to Asure Quality in Lower Hutt the week after sampling (in the case of the 2019 samples) and remained frozen in transport. In the case of the two samples from 2017, they were frozen directly after sampling and stored until February 2019; when they were transported to Asure Quality (while still frozen) with the samples collected in 2019. Water samples to be analysed for phenolic compounds and general pesticides were stored at 4°C for less than 24 hours prior to extraction.

#### 3.2.4 Analyses of Glyphosate Residues

The analyses of glyphosate and its primary degradation product, Aminomethylphosphonic acid (AMPA), were completed by Asure Quality in Lower Hutt, New Zealand using liquid chromatography–tandem mass spectrometry (LCMS-MS). This methodology (developed in-house by Asure Quality) has a detection limit of 0.001 mg/kg. All samples taken during February 2019 were analysed for glyphosate residues, as well as two samples from 2017 (Waiwhetu Stream and Manawatu River at Teachers College). These two samples were chosen from the set due to their urban location, meaning glyphosate contamination from residential weed control would be likely. Glyphosate samples taken during 2017 were not

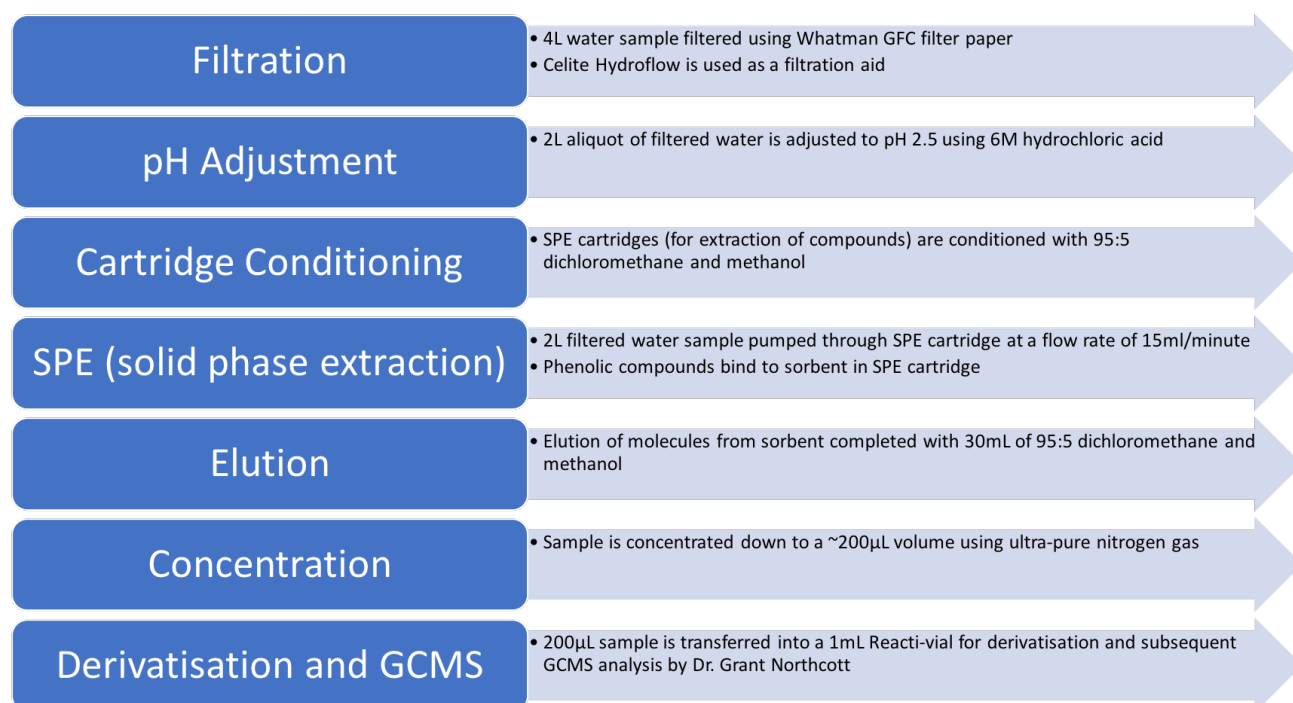
compromised due to being stored in a different location to samples intended for the analysis of phenolic compounds and general pesticides.

### 3.2.5 Extraction of Phenolic Compounds

The eight 4L water samples taken during February 2019 were all analysed for phenolic compounds. Samples taken during November 2017 were not analysed as they were compromised during storage (due to the unnoticed failure of a storage freezer). The 4 L water samples were filtered through pre-conditioned 47 mm Whatman GFC filters in a sintered funnel under vacuum. The GFC filters were conditioned by washing with 50 mL of dichloromethane, followed by 50 mL methanol and 50 mL acetone, and then washed with MilliQ water. All glassware was pre-cleaned by washing three times with the solvents listed above and in the same order. Approximately 2 g of Celite® Hyflo Super-cell (Sigma Aldrich, St Louis, Missouri, United States of America) was added to the sintered funnel as a filtration aid, mixed with MilliQ water and a vacuum was applied to spread the wet mixture over the filter.

A 2 L aliquot of filtered water was used for the extraction of phenolic compounds. The extraction of the phenolic compounds oxybenzone and nonylphenol from filtered samples consisted of solid-phase extraction, followed by derivatisation. A solid phase extraction (SPE) method using pre-conditioned Oasis Hydrophilic-Lipophilic-Balanced (HLB) cartridges (Waters Corp, Milford, Massachusetts, United States of America) was used to extract phenolic compounds from the samples. This method was adapted from Zaugg, Sandstrom, Smith, & Fehlberg (1995) and Madsen, Sandstrom, & Zaugg (2002). After compressing the sorbent bed by gently pressing down with a plastic cylinder, cartridges were conditioned by soaking in 10 mL of 95:5 dichloromethane: methanol solution for three minutes. The solvent mixture was passed through under gravity and the cartridges were left on the bench to air-dry until all of

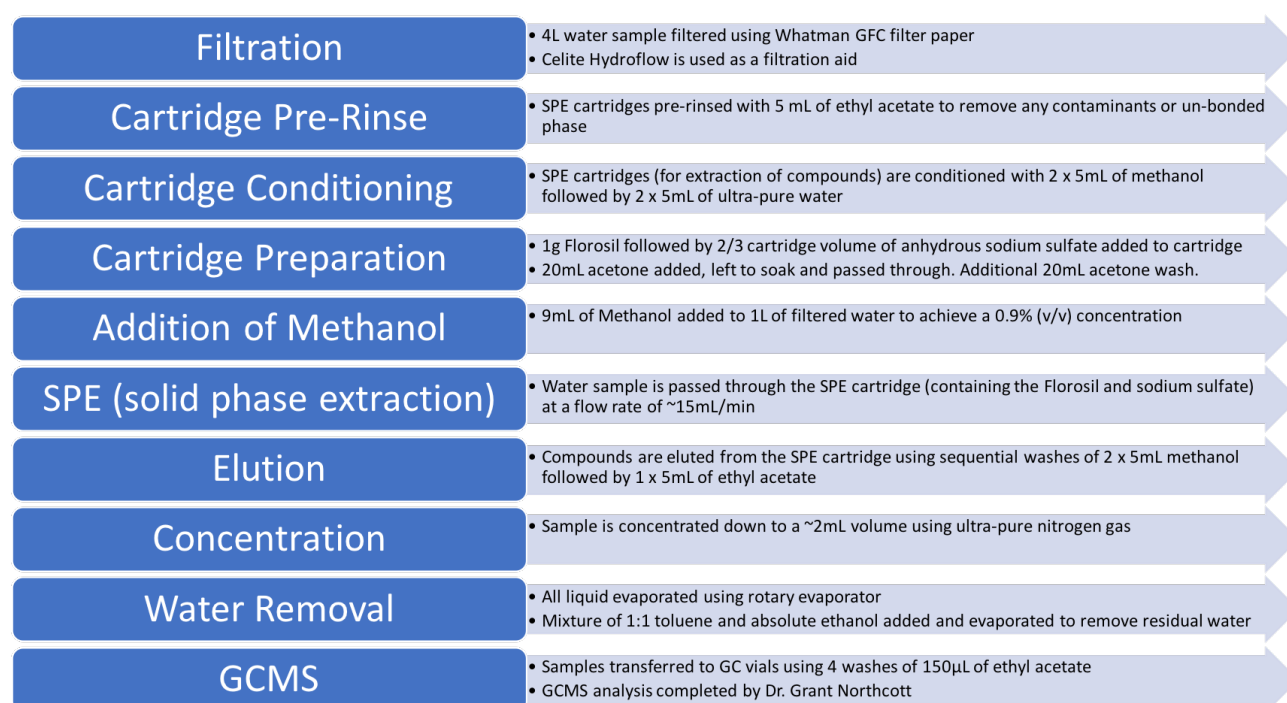
the solvent had evaporated. To increase the hydrophobicity of molecules, all samples were adjusted to pH 2.5 using a stepwise addition of 6 M hydrochloric acid. Teflon tubing was used to pump the sample through the SPE cartridges at a flow rate of ~15 mL/min until the entire sample had passed through. The flow rate was controlled using a vacuum pump. An SPE blank, whereby all of the solvent steps were included but no sample was passed through was included as a negative control. Molecules of interest were eluted from the SPE cartridge with 30 mL of 95:5 dichloromethane: methanol solution. The elutes were stored in amber glass bottles with Teflon-lined lids at 4 °C until they were concentrated to a ~200 µL volume under ultrapure nitrogen gas. Reacti-vials were washed four times with 250 µL of acetone and the remaining 200 µL of sample was transferred into them. The samples were then given to Dr Grant Northcott at Plant and Food Research, Hamilton, New Zealand for derivatisation and GC-MS analysis. A summary of the extraction process is depicted below in Figure 3.1.



**Figure 3.1: Flow diagram showing the process for the extraction of phenolic compounds from environmental water samples.**

### 3.2.6 Extraction of Pesticide Compounds

The eight 4 L water samples taken during February 2019 were all analysed for pesticide compounds. Samples taken during November 2017 were not analysed as they were compromised during storage (due to the unnoticed failure of a storage freezer). The extraction and analyses of these compounds was performed using modifications of a previously-published method (Zaugg, Sandstrom, Smith, & Fehlberg, 1995). A flow diagram below summarises the overall process:



**Figure 3.2: Flow diagram showing the process for the extraction of pesticide compounds from environmental water samples.**

Briefly, IST Isolute-XL C18 (EC) SPE cartridges were pre-rinsed with 5 mL of ethyl acetate to remove any contaminants or un-bonded phase. The cartridges were then pre-conditioned with two 5 mL aliquots of methanol, followed by two 5 mL aliquots of MilliQ water. An aliquot of 9 mL of methanol was added to the 1 L filtered sample to achieve a ~0.9% (v/v)



concentration. Approximately 1 g of Florisil (a form of magnesium silicate) was added to the SPE cartridges to remove any polar contaminants and compressed by gently pressing down using a plastic cylinder. Anhydrous sodium sulfate (dried overnight at 600°C) was then added on top so that it filled 2/3 of the cartridge by volume. Approximately 20 mL of acetone was added to the cartridge and left to soak for ten minutes. Following a final wash step of passing 20 mL of acetone through the cartridge, the column was now ready for sample loading.

A 1 L aliquot of filtered water (described above) was used for the extraction of the pesticide compounds. The entire 1 L water sample was passed through the SPE cartridge at a flow rate of ~15 mL/min. The flow rate was controlled using a pressure valve on a vacuum pump. Once all of the sample was passed through, compounds were eluted using two washes of 5 mL methanol and a final wash using 5 mL of ethyl acetate (15 mL total elution volume) at a flow rate of ~15 mL/min (also controlled by vacuum). Elutes were then concentrated to a volume of ~2 mL under ultrapure nitrogen gas. Any remaining solvent was evaporated from the sample using a rotary evaporator. Then, ~1 mL of 1:1 mixture of toluene: absolute ethanol was added to the dried sample and again evaporated using a rotary evaporator to remove any final traces of water. The elutes were transferred to gas chromatography vials by washing the elute vials four times with 150 µL of ethyl acetate. The 600 µL of ethyl acetate containing the compounds of interest was provided to Dr. Grant Northcott for subsequent GC-MS analysis.

### 3.2.7 Quality assurance for the analysis of nonylphenols and oxybenzone

A summary of the protocols used for quality assurance is displayed below in Table 3.2. A more detailed description including concentrations of standards is provided below in Sections 3.2.7.1 – 3.8.2.2.

**Table 3.2: A summary of quality assurance protocols used in sample extraction.**

All Environmental Samples	Milli-Q Water Positive Control	Milli-Q Water Laboratory Blank	Comparative Standard
<ul style="list-style-type: none"><li>Spiked with Carbon-13 labelled surrogate standard (to compare recovery between different samples)</li></ul> <p><i>Expect: between 80-100% recovery of the surrogate standard (if extraction was successful)</i></p>	<ul style="list-style-type: none"><li>Addition of analytical standard identical to target compound (to measure recovery of target compound using comparative standard)</li><li>Spiked with Carbon-13 labelled surrogate standard (as a control to ensure extraction worked correctly)</li><li>Extracted through the same protocol as samples</li></ul> <p><i>Expect: between 80-100% recovery of the target compound and the surrogate standard</i></p>	<ul style="list-style-type: none"><li>Spiked with Carbon-13 labelled surrogate standard (as a control to ensure extraction worked correctly and to check whether any contaminating compound is present in lab/equipment)</li><li>Extracted through the same protocol as samples</li></ul> <p><i>Expect: undetectable level of target compound</i></p>	<ul style="list-style-type: none"><li>Addition of analytical standard identical to target compound (same volume and concentration as MilliQ water positive control)</li><li>Spiked with Carbon-13 labelled surrogate standard (same volume and concentration as other spikes)</li><li>Not put through extraction procedure (added directly to GC-MS)</li></ul> <p><i>Expect: value of 100% standard recovery</i></p>

#### *3.2.7.1 Surrogate standard spike*

All 8 x 2 L samples of filtered water, together with a blank sample of 2 L Milli-Q laboratory water, and another 2 L Milli-Q laboratory water sample used as a positive control were spiked with a surrogate standard solution containing  $^{13}\text{C}_6$ -labelled oxybenzone and 4-n-nonylphenol. The spiking procedure was as follows: each of the ten samples (*i.e.* the 8 collected samples and the two control samples described above) were spiked with 20  $\mu\text{L}$  of a mixed solution containing 5 ng/ $\mu\text{L}$  of  $^{13}\text{C}_6$ -oxybenzone and 5 ng/ $\mu\text{L}$  of  $^{13}\text{C}_6$ -4-n-nonylphenol in acetone. Therefore, the 2 L sample volumes were spiked with 100 ng of each  $^{13}\text{C}_6$ -labelled surrogate standard to give an equivalent concentration of 50 ng/L (*i.e.* 50 parts per trillion; ppt). A 20  $\mu\text{L}$  aliquot of this surrogate standard solution was also added to a 1 mL reacti-vial as a comparative standard for determining the percentage recovery of the surrogate compounds.

#### *3.2.7.2 Target Analyte Addition*

One of the spiked 2 L Milli-Q laboratory water sample had a solution containing oxybenzone and a technical nonylphenol mix added to act as a positive control. The addition of the target analyte was completed as follows: a 20  $\mu\text{L}$  aliquot of a solution containing 5 ng/ $\mu\text{L}$  oxybenzone and 5 ng/ $\mu\text{L}$  technical nonylphenol equivalents (prepared in acetone) was added to the spiked Milli-Q water sample and the vial containing the comparative standard. The corresponding concentration in the spiked 2 L Milli-Q water positive control was 50 ng/L (*i.e.* 50 parts per trillion; ppt).

### 3.2.8 Quality Assurance for Pesticide Analyses

#### *3.2.8.1 Surrogate Standard Spike*

All 8 x 1 L samples of filtered water, together with a blank sample of 1 L Milli-Q laboratory water, and another 1 L Milli-Q laboratory water sample used as a positive control were spiked

with a surrogate standard solution containing  $^{13}\text{C}_6$ -cis-permethrin. The spiking procedure was as follows: each of the ten samples (*i.e.* the 8 collected samples and the two control samples) were spiked with 40  $\mu\text{L}$  of a 2 ng/ $\mu\text{L}$  solution of  $^{13}\text{C}_6$ -cis-permethrin in acetone. Therefore, the 1 L sample volume was spiked with 80 ng of  $^{13}\text{C}_6$ -cis-permethrin to give an equivalent concentration of 80 ng/L (*i.e.* 80 parts per trillion; ppt). A 40  $\mu\text{L}$  aliquot of the  $^{13}\text{C}_6$ -cis-permethrin surrogate standard solution was also added to a fresh 1 mL reacti-vial as a comparative pesticide standard.

#### *3.2.8.2 Target Analyte Addition*

One of the spiked 1 L Milli-Q laboratory water sample had a solution containing chlorpyrifos and cypermethrin added to act as a positive control. The addition of the target analyte was completed as follows: a 40  $\mu\text{L}$  solution containing 2 ng/ $\mu\text{L}$  of chlorpyrifos and 2 ng/ $\mu\text{L}$  of cypermethrin in acetone was added into the spiked Milli-Q water positive control and the vial containing the comparative standard. The resulting concentration in the 1L positive control sample was 80 ng/L (*i.e.* 80 parts per trillion; ppt).

### 3.2.8 Gas Chromatography Mass Spectrometry of Phenolic Compounds and Pesticide Residues

Both GC-MS runs as well as subsequent calculations to quantify the concentration of each compound in the sample were completed by Dr Grant Northcott at Plant and Food Research, Hamilton, New Zealand using previously published methods (Madsen et al., 2002; Zaugg et al., 1995).

A brief description of the instrument settings used in the analysis was provided by Dr Grant Northcott and can be found in Appendix G.

## 3.3 Results

### 3.3.1 Quality Assurance Results – Oxybenzone & Nonylphenol

The percentage recovered from the spiked control and the concentrations measured in the negative control of each of compound are shown in Table 3.3. Most of these compounds were below or close to the minimum detection limit in the negative control, with the exception of Technical-nonylphenol-15.

**Table 3.3: The percentage recovered in spiked controls and the concentration measured in negative controls of target and related compounds. Values preceded by a less than sign are below the minimum detection limit.**

Compound	Spiked Control (% recovered)	Negative Control (ng/L)
4-tert-amylphenol	88	< 0.5
4-n-amylphenol	90	< 0.5
4-tert-octylphenol	79	< 0.5
4-tert-heptylphenol	83	< 0.5
Technical-nonylphenol-1	80	< 0.5
Technical-nonylphenol-2	78	< 0.5
Technical-nonylphenol-3	76	< 0.5
Technical-nonylphenol-4	74	< 0.5
Technical-nonylphenol-5	82	< 0.5
Technical-nonylphenol-6	84	0.52
Technical-nonylphenol-7	74	< 0.5
Technical-nonylphenol-8	82	0.68
Technical-nonylphenol-9	75	< 0.5
Technical-nonylphenol-10	81	0.84
Technical-nonylphenol-11	83	< 0.5
Technical-nonylphenol-12	73	0.51
Technical-nonylphenol-13	77	< 0.5
Technical-nonylphenol-14	78	< 0.5
Technical-nonylphenol-15	79	1.6
Technical-nonylphenol equivalents	74	< 7.5
4-n-octylphenol	81	< 0.5
4-n-nonylphenol	78	< 0.5
4-n-nonylphenol-13C6	96	< 0.5
Oxybenzone	104	0.52

### 3.3.2 Results of survey for Oxybenzone and Nonylphenols

The concentrations of target and related compounds in samples from the eight river sites are shown in Table 3.4. Technical nonylphenol equivalents and oxybenzone were detected in all eight water samples at concentrations above the minimum detection limit. The mean  $\pm$  SEM recovery of spiked  $^{13}\text{C}_6$  4-n-nonylphenol in samples was  $103.1 \pm 7.5\%$ .

**Table 3.4: Concentration of oxybenzone and nonylphenols in river water samples, including the recovery of a spike control compound. ND indicates the compound was not detected above the minimum detection limit. The percentage recovered refers to recovery of  $^{13}\text{C}_6$  4-n-nonylphenol which was spiked into the samples.**

Compound	Concentration in ng/L (ppt)								
	Blank	Wellington				Manawatu		Waikato	
		Boulcott	Waiwhetu	Te Marua	Porirua	Foxton	College	Narrows	Bridge
4-tert-amylphenol	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>
4-n-amylphenol	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>
4-tert-octylphenol	ND <sub>1</sub>	ND <sub>1</sub>	3.16	ND <sub>1</sub>	0.80	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>
4-tert-heptylphenol	ND <sub>1</sub>	ND <sub>1</sub>	1.13	ND <sub>1</sub>	1.33	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>
Tech nonylphenol equivalents	ND <sub>2</sub>	27.8	167	64.4	48.9	68.8	56.2	20.9	61.2
4-n-Octylphenol	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>
4-n-nonylphenol	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>	ND <sub>1</sub>
Oxybenzone	0.52	5.52	7.98	4.29	9.24	8.75	3.47	19.3	4.37
% recovery	N/A	102	106	92	93	102	109	113	108

\*<sub>1</sub> indicates a minimum detection limit of 0.5 ng/L, <sub>2</sub> indicates a minimum detection limit of 7.5 ng/L

### 3.3.3 Results of Survey for Glyphosate

The concentrations of glyphosate and its primary environmental metabolite, aminomethylphosphonic acid (AMPA) in samples collected in February 2019 from the eight river sites and in November 2017 from two river sites are shown in Table 3.5. None of the samples collected contained concentrations of glyphosate or AMPA above Asure Quality's

reporting limit of 1000 ng/L but samples from several sites in the Wellington region had measurable concentrations. The reporting limit developed byASURE Quality is designed to assess samples against guideline drinking water values for Glyphosate provided by the US Environmental Protection Agency (0.7 mg/L) and the World Health Organisation (0.9 mg/L). This is why the reporting limit is much higher than the method used for the other pesticide molecules and the phenolic compounds, as these are designed to detect molecules that may be harmful to aquatic organisms at much lower concentrations (in the order of ng/L).

**Table 3.5: Concentration of glyphosate and aminomethylphosphonic acid in river water samples collected in 2019 and 2017. ND indicates the compound was not detected above the minimum detection limit of 200 ng/L.**

Compound	Concentration in ng/L (ppt)								
	Blank	Wellington				Manawatu		Waikato	
		Boulcott	Waiwhetu	Te Marua	Porirua	Foxton	College	Narrows	Bridge
Glyphosate (2019)	ND	ND	800	ND	800	ND	ND	ND	ND
Glyphosate (2017)	ND		400				ND	ND	ND
AMPA (2019)	ND	ND	200	ND	400	ND	ND	ND	ND
AMPA (2017)	ND		300				ND	ND	ND

### 3.3.4 Results of survey for pesticide compounds

The concentrations of a range of pesticides in samples collected in February 2019 from the eight sample sites are presented in Table 3.6.

**Table 3.6: Concentration of pesticide compounds in river water samples collected in 2019. ND indicates the compound was not detected above the minimum detection limit of 0.2 ng/L.**

Compound	Concentration in ng/L (ppt)								
	Blank	Wellington				Manawatu		Waikato	
		Boulcott	Waiwhetu	Te Marua	Porirua	Foxton	College	Narrows	Bridge
Atrazine	ND	ND	ND	ND	ND	0.216	0.360	0.491	0.518
Atrazine-desethyl	ND	ND	ND	ND	ND	ND	ND	0.327	0.337
Atrazine-desisopropyl	ND	ND	0.621	ND	1.638	ND	ND	ND	ND
Carbendazim	0.219	0.394	158.180	ND	39.380	3.993	1.504	1.417	1.427
Chlorpyrifos	ND	0.806	ND	ND	ND	ND	ND	ND	ND
Chlorpyrifos-methyl	ND	ND	ND	ND	ND	ND	ND	ND	ND
Cyfluthrin	ND	ND	0.247	ND	ND	ND	ND	ND	ND
Cyhalothrin (lambda)	ND	ND	ND	ND	ND	ND	ND	ND	ND
Cypermethrin	ND	ND	ND	ND	ND	ND	ND	ND	ND
Fipronil	ND	ND	ND	ND	ND	ND	ND	ND	ND
Fluvalinate (tau)	ND	ND	ND	ND	ND	ND	ND	ND	ND
Hexazinone	ND	0.847	5.401	ND	24.790	0.205	ND	2.816	3.032
<sup>13</sup> C <sub>6</sub> -cis-permethrin	ND	68.464	72.690	71.456	77.450	75.31	76.113	74.398	73.005
Permethrin A	0.296	ND	ND	ND	ND	ND	0.382	ND	ND
Permethrin B	0.343	ND	0.263	ND	ND	ND	0.370	ND	ND
Simazine	ND	0.230	4.257	ND	2.618	0.479	0.934	ND	ND
Tebuconazole	ND	ND	0.308	ND	1.145	0.275	ND	ND	ND
Terbutylazine	0.247	11.438	1199.700	0.710	2355.4	4.486	4.748	5.703	7.714
Terbutylazine-desethyl	ND	1.498	259.210	ND	977.70	0.710	1.154	2.761	2.964
Terbutryn	ND	ND	3.487	ND	6.739	ND	ND	ND	ND
Thiacloprid	ND	ND	ND	ND	ND	ND	ND	ND	ND
% recovery		82.8	87.9	86.4	93.6	91.0	92.0	89.9	88.2

The 95% confidence interval for the mean recovery of the <sup>13</sup>C<sub>6</sub>-cis-permethrin surrogate standard was 88.5% ± 2.3% indicating consistent recovery across all samples. Similarly, the

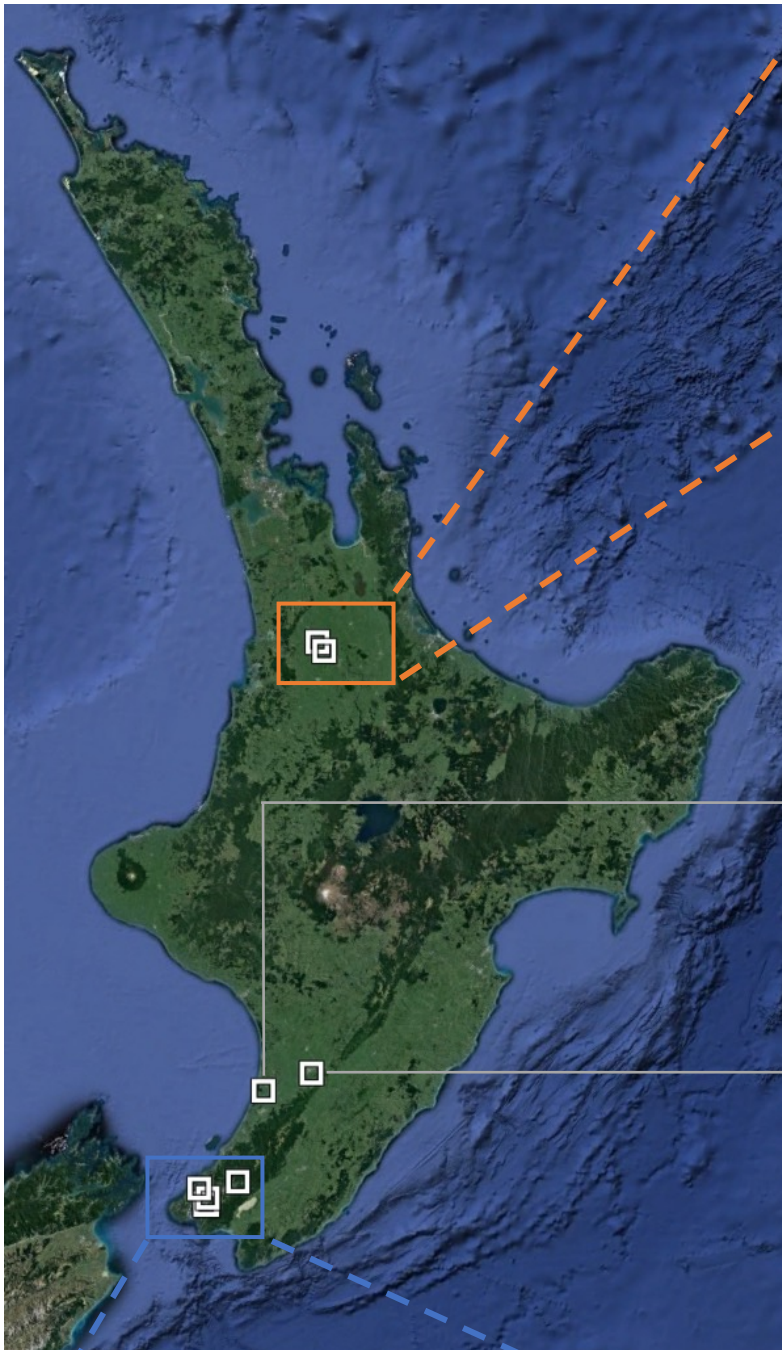


recovery of chlorpyrifos and cypermethrin from the spiked Milli-Q water QA sample was 80 and 82% respectively.

Fipronil, a common insecticide used in New Zealand was not measured above the detection limit at any of the eight sites. Chlorpyrifos was only measured above the detection limit at a single site (Hutt River at Boulcott Street) at a concentration of 0.806 ng/L. Cypermethrin was not detected at any of the sampling sites. Terbutylazine was detected at all sampling sites but of note was the high concentrations of 1,199.7 and 2,555.4 ng/L at two sites in the Wellington region (Waiwhetu Stream and Porirua Stream).

### 3.3.5 Overall results of survey for EOC's

The location of sampling sites and the number of compounds detected at each site is shown in Figure 3.3. The highest number of compounds (15) were detected in the Waiwhetu Stream in Lower Hutt. The site with the lowest number of compounds present above the minimum detection limit (3) was in the Hutt River at the Te Marua drinking water intake site. The two urban sites both had higher numbers of contaminants detected than the rural and forested sites.



#### **Manawatu River at Foxton (Whirokino)**

Land Type: Rural

Compounds Detected (9):

- Technical nonylphenol equivalents
- Oxybenzone
- Atrazine
- Carbendazim
- Hexazone
- Simazine
- Tebuconazole
- Terbutylazine
- Terbutylazine-desthyl

#### **Manawatu River at Teachers College**

Land Type: Rural

Compounds Detected (9):

- Technical nonylphenol equivalents
- Oxybenzone
- Atrazine
- Carbendazim
- Permethrin A
- Permethrin B
- Simazine
- Terbutylazine
- Terbutylazine-desthyl



#### **Hutt River at Te Marua**

Land Type: Forest

Compounds Detected (3):

- Technical nonylphenol equivalents
- Oxybenzone
- Terbutylazine

#### Waikato River at Narrows Boat Ramp

Land Type: Forest

Compounds Detected (8):

- Technical nonylphenol equivalents
- Oxybenzone
- Atrazine
- Atrazine-desethyl
- Carbendazim
- Hexazione
- Terbutylazine
- Terbutylazine-desethyl

#### Waikato River at Hamilton Traffic Bridge

Land Type: Forest

Compounds Detected (8):

- Technical nonylphenol equivalents
- Oxybenzone
- Atrazine
- Atrazine-desethyl
- Carbendazim
- Hexazione
- Terbutylazine
- Terbutylazine-desethyl

#### Porirua Stream

Land Type: Urban

Compounds Detected (13):

- 4-tert-octylphenol
- 4-tert-heptylphenol
- Technical nonylphenol equivalents
- Oxybenzone
- Glyphosate and AMPA
- Atrazine-desisopropyl
- Carbendazim
- Hexazione
- Simazine
- Tebuconazole
- Terbutylazine
- Terbutylazine-desethyl
- Terbutryn

### Figure 3.3:

## Summary of emerging organic contaminants detected in New Zealand waterways

#### Waiwhetu Stream

Land Type: Urban

Compounds Detected (15):

- 4-tert-octylphenol
- 4-tert-heptylphenol
- Technical nonylphenol equivalents
- Oxybenzone
- Glyphosate and AMPA (2017 & 2019)
- Atrazine-desisopropyl
- Carbendazim
- Cyfluthrin
- Hexazione
- Permethrin B
- Simazine
- Tebuconazole
- Terbutylazine
- Terbutylazine-desethyl
- Terbutryn

#### Hutt River at Boulcott

Land Type: Forest

Compounds Detected (8):

- Technical nonylphenol equivalents
- Oxybenzone
- Carbendazim
- Chlorpyrifos
- Hexazinone
- Simazine
- Terbutylazine
- Terbutylazine-desethyl

## 3.4 Discussion

There are few recommended guideline values for EOC's and no minimum acceptable limit in the New Zealand Drinking Water Standards for glyphosate, nonylphenols or oxybenzone. As a result, these compounds are rarely tested. However, the National Water Quality Management Strategy ANZECC guidelines for freshwater (ANZECC, 2000) define the level that protects 95% of a species (*i.e.* trigger value) as 100 ng/L for nonylphenols and 1200 µg/L for glyphosate (ANZECC, 2000). A trigger value for oxybenzone is not listed. The concentrations of technical nonylphenol equivalents detected in the Waiwhetu Stream in Wellington exceeded the trigger value (167 *c.f.* 100 ng/L, respectively) indicating a risk to the species present within the stream.

When comparing the levels of glyphosate detected in New Zealand waterways in this study to those in overseas studies, the levels in New Zealand were comparably lower. The highest glyphosate concentration measured in our study was from the Waiwhetu Stream in February, 2019 at 0.8 µg/L. In comparison, waterway concentrations of glyphosate were reported at 0.03 – 73 µg/L from the Bogue Phalia Stream in the US, <0.02 – 5.7 µg/L from the South Fork Iowa River and <0.1 – 86 µg/L from Rouffach in France (Coupe et al., 2012). In France specifically, 303 samples were collected from the Rouffach Basin during 58 run-off events and the median concentrations of glyphosate and AMPA were 4.7 and 1.9 µg/L, respectively. All but one sample had a concentration greater than the detection limit of 0.1 µg/L (Coupe et al., 2012).

In regards to acute eco-toxicity values for glyphosate, the samples measured in the study presented herein had glyphosate concentrations that were many orders of magnitude lower

than the  $LC_{50}$  values for crustacea ( $LC_{50} = 1770-5390 \mu\text{g/L}$ ) and for fish ( $LC_{50} = 1700-6100 \text{ mg/L}$ ). However, it should be acknowledged that there have been reports of negative effects to aquatic invertebrates when long-term chronic exposure has been considered. For example, a significant reduction in juvenile size of *Daphnia magna* has been demonstrated following exposure to glyphosate concentrations of  $50 \mu\text{g/L}$  (Cuhra et al., 2013). This exposure level is nearly 50x higher than what was detected in our study, therefore it can be concluded that at these levels, the risk to aquatic species is relatively negligible. It should also be noted a growing concern of a link between glyphosate exposure and increased risk of developing non-Hodgkin's lymphoma (Zhang et al., 2019).

Nonylphenols are the breakdown products of nonylphenol polyethoxylates which are a widely used non-ionic surfactant that have been widely detected in surface waters and sediments (Mao et al., 2012). Nonylphenols are hydrophobic and are therefore often absorbed by sediment particles, resulting in higher concentrations in the sediment layer than surface water (Mao et al., 2012). The highest concentrations of technical nonylphenol equivalents detected in our study was  $167 \text{ ng/L}$ , with the mean concentration being  $58 \text{ ng/L}$ , which is comparable to other countries (Mao et al., 2012). For example, mean nonylphenol concentrations in surface water from rivers in South Korea ranged from  $95.7 - 163.5 \text{ ng/L}$ , and from rivers in Vietnam, Thailand and Indonesia were  $1761.9$ ,  $958.5$  and  $238.5 \text{ ng/L}$ , respectively (Duong et al., 2010). In contrast, the Pearl River Delta in China had extremely high levels of nonylphenol concentrations detected, ranging from  $36 - 33,231 \text{ ng/L}$  at nine sampling sites (Peng et al., 2008).

The median lethal concentration of nonylphenol for fish and aquatic vertebrates after 96 hours exposure, was  $128$  and  $69 \mu\text{g/L}$ , respectively, as reported in a European Union risk

assessment report (European Chemical Bureau, 2002). These values are much higher than most reported environmental concentrations. As nonylphenols are constantly being replenished in the environment, a more accurate representation of their effects can be assessed by analysing chronic effects. The concentration of nonylphenol that exhibited no observed effect for fish over a 33-day period was 7.4 µg/L. A decreased survival of fish embryos was observed above this concentration. For aquatic invertebrates, the highest no-observed-effect dose was 24 µg/L, indicating that fish are more sensitive to nonylphenol (European Chemical Bureau, 2002). Based on the 33-day chronic toxicity data for fish, an overall predicted-no-observed-effect-concentration (PNEC) of nonylphenol was 0.074 µg/L. This value was calculated based on the fact that nonylphenol is moderately bio-accumulative across trophic levels, and therefore a factor of 100 was applied to the toxicity data of the most sensitive species, *i.e.* fish (European Chemical Bureau, 2002). The concentration of technical nonylphenol equivalents in the Waiwhetu Stream in the Wellington region (*i.e.* 0.167 µg/L) exceeds both the PNEC (0.074 µg/L) and trigger (0.10 µg/L) values for this chemical, indicating there is a risk to aquatic organisms in the Waiwhetu Stream from the chronic effects of nonylphenols. This study would suggest that further sampling at multiple time points should be completed at this site, and at other sites where a high concentration of technical nonylphenol equivalents were found. Other rivers or streams in highly urbanised locations with industrial processing facilities should also be assessed for environmental nonylphenol levels. It is not unreasonable to propose that due to increased run-off events, the nonylphenol concentrations might be higher during periods of higher rainfall than what was detected in our summer samples.

Oxybenzone (benzophenone-3) is an aromatic hydrocarbon that is used as a UV-filter in sunscreen and other personal care product formulations. Oxybenzone filters both UVB light

(290 nm to 320 nm) and UVA light in the range of 320-340 nm. This compound is a common ingredient in sunscreens at concentrations of up to 6% (w/v), and in concentrations of up to 0.5% (w/v) in other personal care products (DiNardo & Downs, 2018). Oxybenzone is known to bio-accumulate in fish species, has been demonstrated to be an endocrine disruptor exhibiting estrogenic activity in both fish and rats, and is genotoxic to larval coral (Downs et al., 2016). It has been detected in multiple surface water samples from lakes in Switzerland, ranging in concentrations from 10 – 35 ng/L (Balmer et al., 2005). These concentrations are comparable to what was detected in our samples (3.5 – 19.3 ng/L). The highest concentration of 19.3 ng/L was detected at the Narrows boat ramp on the Waikato River. This is likely due to this site having the highest recreational usage out of all the sites sampled herein, and therefore where the risk of sunscreen contamination is high. Unfortunately, even higher concentrations have been detected in the marine environment, with measurements of 75 µg/L to 1.4 mg/L at sites on the coral reefs of the US Virgin Islands and of 0.8 to 19.2 µg/L in Hawaii (Downs et al., 2016). The European Chemicals Agency (ECHA) reported that the PNEC for oxybenzone in both freshwater and marine water was 0.67 µg/L (ECHA, 2019). The concentrations in our freshwater samples and from multiple surface water samples from lakes in Switzerland (Balmer et al., 2005) were much lower than the PNEC concentration, which is many orders of magnitude lower than acute toxicity values for aquatic macroinvertebrates (24 hour  $EC_{50}$  of 2.34mg/L), fish (24 hour  $LC_{50}$  of 5.5mg/L) or aquatic algae and cyanobacteria (72 hour  $EC_{50}$  of 670µg/L) (ECHA, 2019). It can therefore be concluded that based on current eco-toxicological data, it is unlikely that there is any risk to the environment from oxybenzone, at least at the sites assessed herein. No assessment of the marine environment was conducted as part of this study. Popular swimming beaches and water adjacent to wastewater

discharges should be also assessed for oxybenzone concentrations, as these are the most likely areas where environmentally-damaging concentrations may be detected.

The three insecticide molecules which we were interested in were either not present above the detection limit (fipronil and cypermethrin) or were present at concentrations that are not harmful to aquatic biodiversity (chlorpyrifos). It should be noted that cypermethrin is extremely toxic to aquatic organisms. The no observable effect concentration (21 day) for fish is 30 ng/L and for aquatic invertebrates is 40 ng/L (IUPAC, 2020). Due to the sporadic nature of pesticide spraying and conditions required for leeching into waterways, not detecting the presence of the compound above the MDL in our grab sample at a single time point does not preclude the possibility of higher, harmful concentrations being present at these sampling sites at other time points. Deploying a passive sampler to determine the pesticide load over a longer duration or sampling at multiple time points would overcome this limitation.

Although terbuthylazine was detected at high concentrations compared to other pesticides, its toxicity is comparably lower. The  $LC_{50}$  (96 h) for rainbow trout and mirror carp is 2.2 and > 5.7 mg/L respectively (MacBean, 2012). For algal species, the  $EC_{50}$  (72 h) ranges from 0.012 – 0.102 mg/L (MacBean, 2012). In terms of chronic toxicity to aquatic invertebrates (specifically *Daphnia magna*), the no observable effect concentration for 21 days (19,000 ng/L) was roughly ten times greater than the highest concentration measured in our study (IUPAC, 2020). The measured concentrations at the Waiwhetu Stream (1199 ng/L) and Porirua Stream (2555.4 ng/L) are well below the acute and chronic toxicity values for these aquatic organisms, indicating there is likely not a threat to biodiversity at these concentrations. Passive sampling over a longer period would give greater certainty as to the maximum concentration that this molecule reaches in these urban streams.



This study provides a first assessment of these prominent emerging contaminants in the New Zealand freshwater environment. In summary, oxybenzone and technical nonylphenol equivalents were detected at all sites sampled herein. Additional compounds from the nonylphenol family were also detected at some of the sites. Although the concentrations were low compared to acute toxicity limits, the nonylphenol concentration in the Waiwhetu stream was above both the trigger and PNEC values. At most other sites, the nonylphenol concentrations neared the PNEC limit. Glyphosate was also detected at two sites but the sensitivity of the assay used was limited and evidence is only now emerging on the effects of glyphosate on human health. Oxybenzone concentrations within the environment were well below levels deemed to harm the aquatic environment. This study provides sufficient evidence to warrant a monitoring programme to assess all of the compounds mentioned in the Tier 1 approach postulated by Stewart et al. (2016). Expanding the number of EOC's assessed at a larger number of targeted sites throughout the year would add to the knowledge of the state of EOC's in the New Zealand freshwater environment, one of New Zealand's most precious yet vulnerable natural resources.

# Chapter 4: Selection of a Glyphosate-binding aptamer using standard SELEX

## 4.1 Background

### 4.1.1 Glyphosate and Current Detection Methods

Glyphosate ( $C_3H_8NO_5P$  ; N-(phosphonomethyl)glycine) is the most commonly-used herbicide in the world and has the potential to contaminate the freshwater environment through spray-drift, run-off and direct application over waterways. The latter of which is approved with controls in New Zealand. The current usage, concentrations within environmental samples and the potential health and environmental impacts of glyphosate have been comprehensively reviewed in Chapter 1. Moreover, measurements of glyphosate in a number of New Zealand streams are reported in Chapter 3.

The most employed detection method for glyphosate in environmental water samples is the USGS Method 0-2141-09 developed by Meyer et al. (2009). This consists of isotope dilution, followed by online solid-phase extraction and analyses using LCMS-MS with a detection limit of 0.02  $\mu\text{g/L}$  (Meyer et al., 2009). Methods such as these are time-consuming and expensive. Due to emerging health and environmental concerns of glyphosate, there is an urgent need to develop an easier, cost-effective and sensitive detection method.

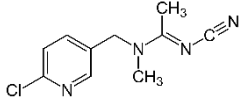
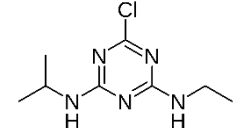
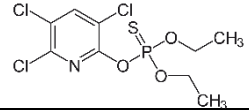
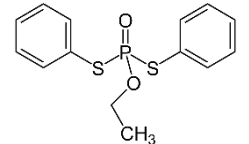
### 4.1.2 Aptamers as Bioreceptors for Pesticides

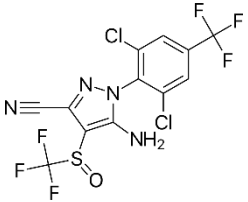
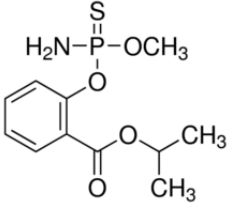
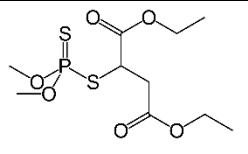
Aptamers have been used as recognition molecules for pesticides and integrated into biosensors (aptasensors) in previous studies. Aptasensors may be designed for on-site or

continuous pesticide monitoring and are affordable enabling routine monitoring of compounds.

Several aptamers for pesticide molecules are now published, with several now having been used for detection in real environmental samples (Kim, Thanh Thoa, & Gu (2019)). Targets which have been employed in real-world sensing schemes include acetamaprid (Madianos et al., 2018; Qi et al., 2016), atrazine (Madianos et al., 2018), edifenphos (Arvand & Mirroshandel, 2017), fipronil (Hong & Sooter, 2017) and malathion (Bala et al., 2016). Liu et al. (2020) reports a broad spectrum aptamer that is able to bind to organophosphorus pesticides, which would therefore include detection of glyphosate. Chlorpyrifos-specific aptamers coupled with silver nanozymes were able to detect chlorpyrifos in river water samples with a response time of only two minutes (Weerathunge et al., 2019). A summary of current pesticide aptamers, their target molecules, their chemical structure and properties, the desired application matrix and detection limit/ $K_d$  (if detection limit is not applicable) are listed in Table 4.1. It should be noted that the number of published aptamers for pesticides compared to other targets is very small. Many aptamers are being developed for commercial applications and therefore their sequences or sensing capabilities are not disclosed.

**Table 4.1: A summary of pesticide aptamers in published literature**

Target pesticide	Chemical Structure	Molecular Weight (g/mol)	Application	LOD/ $K_d$	Aptamer Sequence	References
Acetamaprid		222.67	Detection in waste water	$K_d = 4.98 \mu\text{M}$	5'-CTGACACCATATTATGAAGA-3'	Qi et al. (2016) He, Liu, Fan, & Liu (2011)
Atrazine		215.68	Environmental samples	$K_d = 0.62 \text{ nM}$	5'-(SH)-(CH <sub>2</sub> ) <sub>6</sub> -TACTGTTTGCCTGGCGGATTAGCCAGTCAGTG-[Flc]-3' (R12.23)	Williams, Crieffeld, Gattu, Holland, & Sooter (2014)
Broad-spectrum organophosphorus pesticides	Not applicable	Not applicable	Analysis of food residues	Variable for each compound	5'AAGCTTGCTTTATAGCCTGCAGCGATTCTTGATCGGAAAAGGCTGAGAGCTACGC-3' (SS2-55) 5'AAGCTTTTTTGACTGACTGCAGCGATTCTTGATCGCCACGGTCTGGA AAAAGAG-3' (SS4-54)	Wang et al. (2012) Liu et al. (2020)
Chlorpyrifos		350.59	Intended application not stated	$K_d$ is not reported	5'CCTGCCACGCTCCGAAGCTTAGGGTTACGCCTGCAGCGATTCTGATCGCGCTGCTGGTAATCCTTCTTTAAGCTTGGCACCCGCATCGT-3'	Lei et al. (2012)
Edifenphos		310.37	Food samples (rice)	38 nM	5'CGTACGGAATTCGCTAGCTAAGGGATTCTGTAGA AGGAGCAGTCTGGATCCGAGCTCCACGTG-3'	Kwon, Nguyen, Park, & Gu (2015)

Fipronil		437.15	River water	48 nm ± 8 nM	5' <b>TGTACCGTCTGAGCGATTCTG</b> ACAGTTTCTGGAGGACTGGGCGGGG TGACGGTTATG <b>AGCCAGTCAGTGTTAAGGAGTGC</b> -3'	Hong & Sooter, (2017)
Isocarbophos		289.29	Intended application not stated	0.83 μM	5'AAGCTTGCTTTATAGCCTGCAGCGATTCTTGATCGGAAAAGGCTGA GAGCTACGC-3'	Wang et al. (2012)
Malathion		330.36	Detection using surface enhanced Ramen scattering	K <sub>d</sub> not reported	5'-SH- ATCCGTCACACCTGCTCTTATACACAATTGTTTTCTCTTAACCTCTTGA CTGCTGGTGTGGCTCCCGTAT-3'	Nie et al. (2018)

\*primer regions are highlighted in red

To the best of my knowledge, there are no published specific glyphosate-binding aptamers in the literature. However, there are patents for sequences which bind to glyphosate. Specifically, CN105039347A, a patent for “aptamer specifically bound with glyphosate and application” which was filed on 10/08/2015 and granted on the 23/10/2018 (Patent No. CN105039347 (A), 2015). The patent specifies a DNA sequence which is 84 nucleotides long and contains a G-quadruplex structure. The authors of the patent report a change in Gibbs free energy ( $\Delta G$ ) of -16.78

kJ/mol. The  $\Delta G$  value indicates how favourable a chemical reaction is and in the case of single stranded DNA, how stable the predicted secondary structure is. A positive  $\Delta G$  value indicates an unfavourable reaction whilst a negative value indicates a favourable reaction. The more negative the value, the more favourable the reaction is and the more stable the secondary structure. The value reported in the patent is highly negative compared to values for other published aptamers. This indicates that the predicted secondary structure is extremely favourable and the aptamer would be highly stable. A stable secondary structure is not indicative of a binding aptamer and without binding studies there is no evidence to suggest that this sequence would bind to glyphosate. There are other patent applications that appear in database searches of literature, such as KR2019073327-A, KR2018054018-A, KR2013044981-A and KR1338520-B1. However, no further details about the patents were able to be located, despite searching multiple patent databases.

**Table 4.2: Summary of Glyphosate aptamer patent claims**

Aptamer Sequence	Reported or Calculated $\Delta G$ (kcal/mol)	$K_d$	Application	References
5'-TGCTAGACGATATTCGTCCATCCGAGCCCGTGGCGGG CTTTAGGACTCTGCGGGCTTCGCGGCGCTGTCAGACTGA ATATGTCA -3'	-16.78	Not reported	ELONA for the detection of Glyphosate	Patent No. CN105039347 (A), (2015)
Not reported. Sequence length is 100 nucleotides.	Not reported	Not reported	Kit for the detection, recovery and removal of glyphosate	Patent No. KR2019073327 (A), (2019)

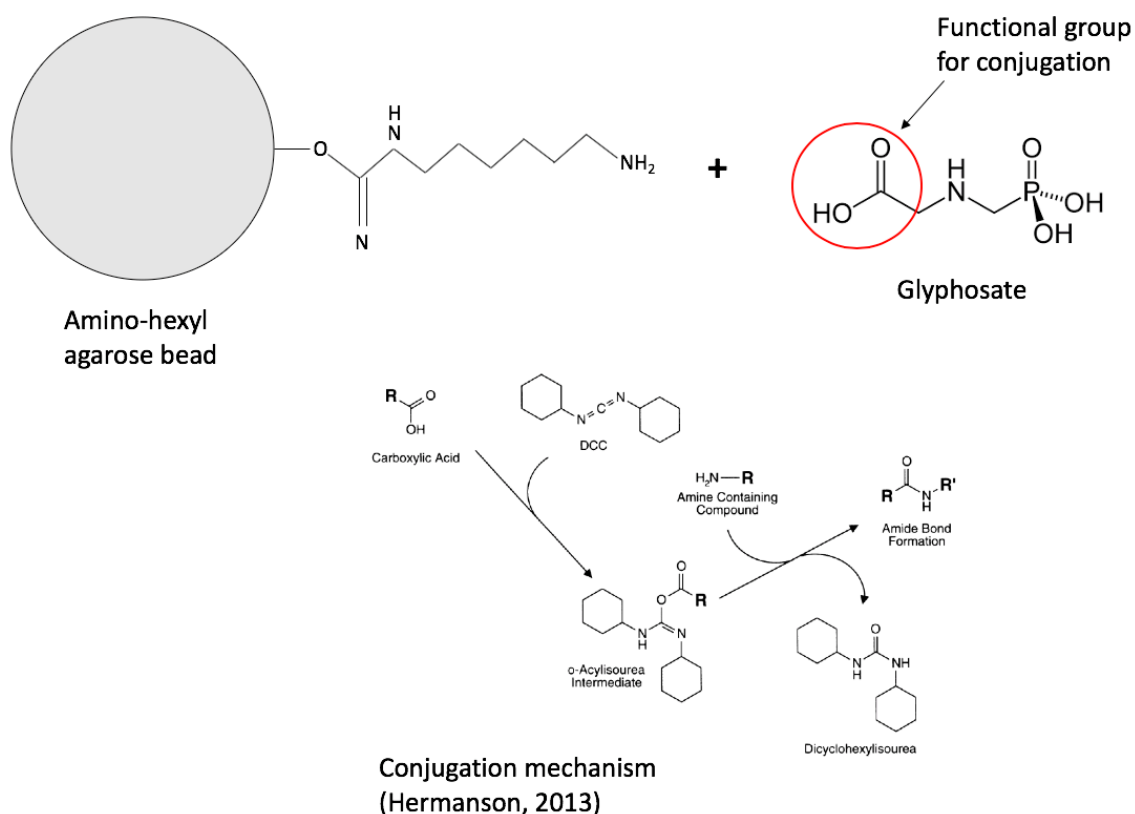
It appears that other research groups and commercial entities are interested in selecting aptamers for glyphosate due to the demand for an easy, cost-effective detection solution. However, as Glyphosate is a small molecule (169.07 g/mol), generating an aptamer against it is very challenging.

The aims of this chapter were to generate and characterise an aptamer that can bind glyphosate with at least a nanomolar affinity as determined by its equilibrium constant of dissociation ( $K_d$ ).

## 4.2 Methodology

### 4.2.1 Preparation of Affinity Matrix: Glyphosate – Bead Conjugation

In order to select aptamers against glyphosate, the molecule first needed to be conjugated onto an affinity matrix to enable separation of bound and un-bound oligonucleotides during the SELEX process (Figure 4.1).



**Figure 4.1: Scheme for conjugating glyphosate to w-amino-hexyl agarose beads via amide bond formation. Figure is adapted from Hermanson (2013).**

For this target molecule, w-amino-hexyl-agarose beads were selected as a suitable affinity matrix. Glyphosate, w-amino-hexyl-Agarose beads, N,N'-dicyclohexylcarbodiimide (DCC) and N-hydroxysuccinimide (NHS) were purchased from Sigma Aldrich (St Louis, Missouri, United States of America).



The addition of an NHS group to the carboxyl group on glyphosate was carried out using the following procedure. Glyphosate (0.5 mmol), NHS (1 mmol) and DCC (1 mmol) were combined in a glass vial in a 1:2:2 ratio. An aliquot of 10 mL of tetrahydrofuran was added and the mixture was then incubated at room temperature whilst being stirred at a speed of 480 rpm for four hours. After incubation, the mixture was aliquoted into 1mL aliquots by transferring into 1.7 mL micro centrifuge tubes (Eppendorf, Hamburg, Germany) and centrifuged at 13,300 x g for 5 minutes.

Three grams of w-aminoethyl-agarose beads were prepared by washing with 500 mL of distilled water in a sintered glass funnel under vacuum. The beads were removed from the sintered funnel, and halved into two separate glass vials. The supernatant of the glyphosate/NHS/DCC mixture was then added to one of glass vials containing washed w-aminoethyl agarose beads. A volume of 4.5 mL of Borate buffer (pH 9) was added to each vial containing the beads (control and target beads) and incubated overnight at room temperature with shaking at a speed of 250 rpm. The control and target beads were then each washed in a sintered glass funnel with 500 mL of distilled water and stored dry in a micro centrifuge tube at 4°C until further use.

#### 4.2.2 SELEX using Glyphosate Conjugated Beads

Approximately 1 mg of dry glyphosate-conjugated beads were transferred to a new 1.7 mL micro-centrifuge tube and re-suspended in 30 µL of 1X BWB using a pipette-tip which had been wet with 1X BWB. In order to reduce the likelihood of non-specific binding of oligonucleotides to the surface of the agarose beads, blocking buffer at a volume of 60µL was added to the beads and incubated for five minutes at room temperature. For the first selection round, 4 µL of 100 µM N40 oligonucleotide library was added to the mixture and

incubated overnight at 4°C on a shaker at a speed of 500 rpm. For Selection Rounds 1-5, up to 30 µL of ssDNA from the previous selection round was added to glyphosate-conjugated matrix for overnight incubation in place of the N40 oligonucleotide library. Starting with selection round six, an activated control matrix was used to ensure that the amplified oligonucleotides were being selected for their affinity to glyphosate and not their affinity to any other component of the matrix. The activated control matrix was prepared as described above, with the same blocking procedure. From Selection Round 6, 15 µL of ssDNA from the previous selection round was added to both the activated control and the glyphosate-conjugated beads. For selection round 10, which included a counter selection round, 7.5 µL of ssDNA from selection round 9 was added to four tubes. Two of the tubes contained the activated control matrix and two contained the glyphosate-conjugated matrix. A counter selection was then performed against one pair of control and glyphosate matrices and the other pair was treated normally, so a comparison could be made to see if the counter selection successfully removed non-specific oligonucleotide sequences. The counter selection procedure is described in Section 4.2.7. Incubation conditions are as described above, except for selection round 9, in which the control and glyphosate-conjugated matrices were incubated for 2.5 hours at room temperature (with shaking at approximately 500 rpm) to encourage selection pressure for sequences which bound more rapidly.

#### 4.2.4 Washing of Affinity Beads

Following overnight incubation, the agarose bead mixture was transferred into a filter column so that unbound oligonucleotides could be removed through washing. Using a pipette, the mixture was added to a filter column. The empty micro centrifuge tube that contained the beads was rinsed with 1 mL of 1X BWB containing 0.1% (v/v) IGEPAL and also added to the

top of filter column. The washing procedure was completed by pipetting the desired washing buffer into the filter column and allowing the liquid to flow over the beads using gravity capillary action. The beads are retained on the membrane of the filter column and collected after washing, using a transfer pipette. The beads were washed with 5 mL of 1X BWB containing 0.1% (v/v) IGEPAL for the first selection round and then varying volumes of 1X BWB, depending on the desired selection pressure. The amount of non-ionic surfactant and the incubation time was varied to induce the desired selection pressure. An outline of the washing and incubation conditions for each SELEX round is listed in Table 4.3. Following washing, the beads were removed from the filter column and were re-suspended in approximately 75  $\mu$ L of distilled water for storage at 4°C, prior to amplification of the remaining oligonucleotide sequences using PCR.

**Table 4.3: Washing Conditions for each round of Glyphosate SELEX**

Round No.	Washing Conditions
1	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
2	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
3	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
4	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
5	2 mL of 1X BWB with 0.5% (v/v) IGEPAL, followed by 3 mL of 1X BWB with 0.1% (v/v) IGEPAL
6	2 mL 1X BWB with 0.5% (v/v) IGEPAL, then 3 mL of 1X BWB with 0.1% (v/v) IGEPAL
7	2 mL 1X BWB with 0.5% (v/v) IGEPAL, then 3 mL of 1X BWB with 0.1% (v/v) IGEPAL
8	2 mL 1X BWB with 0.5% (v/v) IGEPAL, then 3 mL of 1X BWB with 0.1% (v/v) IGEPAL
9	2 mL 1X BWB with 0.5% (v/v) IGEPAL, then 3 mL of 1X BWB with 0.1% (v/v) IGEPAL
10	<b>Counter Selection:</b> 5 mL of amino acid mixture, then 1 mL of 1X BWB with 0.1% (v/v) IGEPAL <b>Normal Selection:</b> 5 mL of 1X BWB with 0.1% (v/v) IGEPAL

#### 4.2.5 PCR Amplification of Bound Oligonucleotides

A thermal cycling machine was used to amplify oligonucleotides by PCR. The amount of template added, and therefore the total reaction volume varied between cycles, depending on the desired level of amplification for each selection round. The concentration of magnesium ions in the reaction was also varied to induce mutation when desired. The different PCR conditions for each round of SELEX is listed in Table 4.4. A negative and positive control were included in each reaction. The negative control was ultra-pure water and the positive control was a characterised oligonucleotide (TES-07) of a known concentration which contained the same primer sequences. In some rounds, if there was not enough DNA produced from the initial amplification reaction, then an additional enrichment reaction was carried out (data not shown).

**Table 4.4: PCR Conditions used for each glyphosate SELEX round**

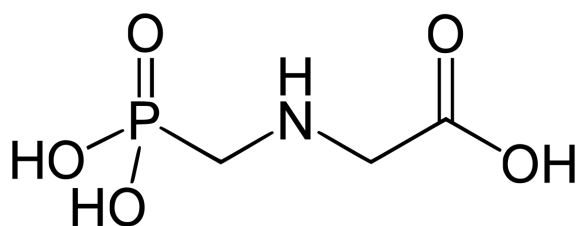
Round Number	Number of Cycles	Volume of 25 mM Magnesium Added ( $\mu\text{L}$ )
1	40	0
2	35	0
3	28	0.5
4	28	0
5	28	0.3
6	35	0
7	35	0
8	35	0.5
9	35	0
10	35	0

The PCR were performed on a Rotor-Gene Q (Qiagen, Hilden, Germany) with a final reaction volume of 25  $\mu$ L. The final reaction consisted of ultra-pure water, 1X PCR Buffer,  $Mg^{2+}$  (concentration was varied depending the desired mutation rate), 240  $\mu$ M of dNTP's, 0.22  $\mu$ M of forward primer, 0.22  $\mu$ M of reverse primer, 2X SYBR Green and 1 unit of DNA *Taq* Polymerase. All reagents were purchased from Thermo Fisher Scientific (Waltham, Massachusetts, United States of America). Three different template amounts were used during each SELEX round : 0.5  $\mu$ L, 2  $\mu$ L and 5  $\mu$ L. PCR cycles were performed as follows: an initial denaturation step of 5 minutes at 95°C to ensure that the template was completely denatured; followed by between 25 and 35 cycles (depending on the result of the amplification for each template amount) of 40 seconds at 94°C (denaturing), 30 seconds at 53°C (annealing), 15 seconds at 72°C (extension); and a final extension step of 60 seconds at 72°C. Following PCR, the reaction products were stored at 4°C until being used for gel electrophoresis.

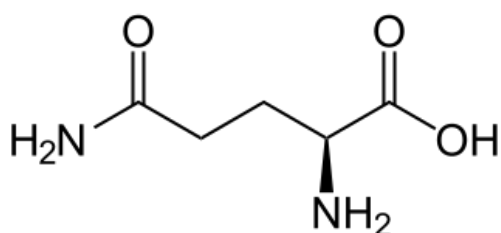
#### 4.2.7 Counter-Selection Rounds

Counter-selection was used to reduce the number of DNA sequences that exhibited non-specific binding by incubating with similarly-structured molecules to the target of interest. The sequences that bound to the similar compounds were then eluted off the matrix. For the glyphosate SELEX, five amino acids with similar structures to Glyphosate were selected, namely alanine, glutamine, leucine, valine and glycine (Figure 4.2).

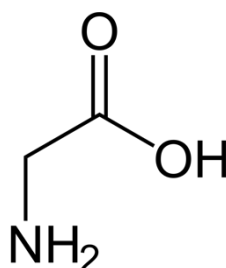
Glyphosate:



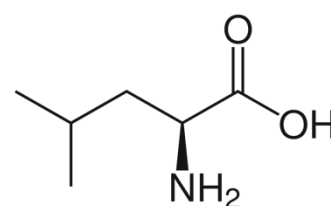
Glutamine:



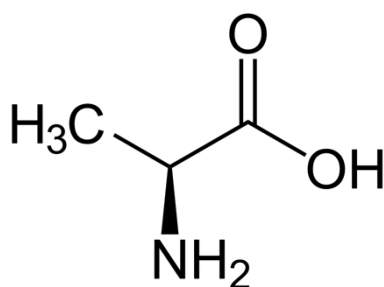
Glycine:



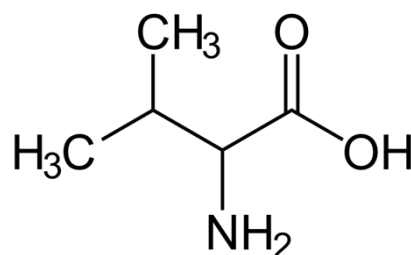
Leucine:



Alanine:



Valine:



**Figure 4.2: Structures of the target molecule glyphosate, and the compounds used for counter selection (glutamine, glycine, leucine, alanine and valine).**

All amino acids were made up to a 100 mM concentration in 20 mL of 1X BWB containing 5% (v/v) ethanol to assist solubilisation. Instead of washing the glyphosate-conjugated and control beads with 1X BWB (as in a normal selection round), the beads are added to the affinity column and 1 mL of the counter-selection amino acid mix was added. The mixture was incubated in the column with the beads for 5 minutes at room temperature. Following incubation, the beads were washed with an additional 5 mL of the amino acid mixture. This

competitive elution should eliminate any sequences which bind to similarly structured compounds from the aptamer library. Finally, the beads were washed with 1 mL of 1X BWB containing 0.1% (v/v) IGEPAL, collected and re-suspended in approximately 70  $\mu$ L of 1X BWB for PCR.

#### 4.2.8 Cloning and sequencing of plasmid DNA

Following the tenth selection round, PCR products from Selection Rounds 8, 9 and 10 were ligated into a plasmid, the plasmid DNA was extracted from the transformed colonies and sent for Sanger sequencing. The cloning procedure is described in the general materials and methods (Section 2.9). Thirty six plasmids were sent for sequencing at the Massey University sequencing facility in Palmerston North, New Zealand.

#### 4.2.10 Phylogenetic Analysis

The protocol used for alignment and phylogenetic analysis of sequences is described in the general materials and methods (Section 2.10).

#### 4.2.11 Secondary Structure Prediction

The protocol used for the prediction of the single stranded secondary structure for each aptamer candidate is described in the general materials and methods (Section 2.11).

#### 4.2.12 Characterisation of aptamer candidates by MST

Due to cost, only two aptamer candidates were able to be characterised for binding affinity by micro scale thermophoresis (MST). The first and second most frequent sequences (*i.e.* GLY04 and GLY12) were selected. Characterisation using MST was completed entirely by 2Bind (Regensburg, Germany). The composition of the assay buffer, ligand solubility instructions and potential binding affinity advice were provided to 2Bind. The protocol for

MST was designed and implemented by 2Bind as follows. A 10 mM stock solution of glyphosate was prepared in 1X BWB containing 0.02% (v/v) Tween 20. A 1 mM working solution was prepared by diluting the stock solution at a 1:10 dilution using 1X BWB containing 0.02% (v/v) Tween 20. The instrument used for MST was a Monolith NT.115 (Nanotemper, Munchen, Germany). The concentration of the aptamer was kept constant at 100 nM and two-fold serial dilutions of the ligand were prepared from 10  $\mu$ M to 305 pM. Both the ligand and aptamer were made up using 1X BWB containing 0.02% (v/v) Tween 20. An aliquot of 5  $\mu$ L of the oligonucleotide was mixed with 5  $\mu$ L of the ligand at each concentration and the mixture was added to a standard capillary for analyses on the instrument. The instrument was set to 40% LED power and 60% laser power and the temperature was set to 25°C. Two technical repeats were completed. Data analyses and the fitting of a binding isotherm was completed by 2Bind using the software supplied with the instrument.



## 4.3 Results

### 4.3.1 Glyphosate – Bead Conjugation

As glyphosate does not contain a phenolic ring, UV-visible spectroscopy could not be used to confirm whether the compound had successfully conjugated to the amino-hexyl agarose beads. Instead, IR-spectroscopy was used to compare 2 mg of dried control beads to 2 mg of dried prospective glyphosate-conjugated beads. Due to the sensitivity of this technique and the presence of a large number of hydroxyl groups on the surface of the beads dominating the spectrum, the results of the spectrum were inconclusive. As no other methods were available at the time, SELEX was commenced and positive selection of oligonucleotides to the bound beads was used as confirmation of a successful glyphosate-bead conjugation. It is further confirmed in later SELEX rounds where sequences were not amplified when the single-stranded DNA library was incubated with control-beads indicating that the library was enriched for sequences that bound to glyphosate and not the affinity matrix.

#### 4.3.2 Results of Selection Round 3

Different ratios of bead: PCR template were assessed due to the observation of double banding in the initial PCR products. It was postulated that this was caused by the presence of too much template in the PCR or the presence of too many Sepharose beads influencing the PCR. Comparisons of differing amounts of template and beads in the third selection round are illustrated in Figure 4.3. When 10  $\mu$ L of template was used, the highest (~4 mg) amount of beads produced a single strong band. When 4  $\mu$ L of template was tested, the lowest (~1 mg) bead amount produced the strongest result, but a double band was observed. A double band was also observed with 1  $\mu$ L of template added to the highest bead amount. Testing different bead: template ratios was largely inconclusive i.e. there wasn't any consistent pattern. The decision was made to continue the SELEX with the PCR product from the well with 10  $\mu$ L of template using ~4 mg of beads. These results demonstrated a sufficient amount of library is being retained on the glyphosate-conjugated beads after three rounds of selection.

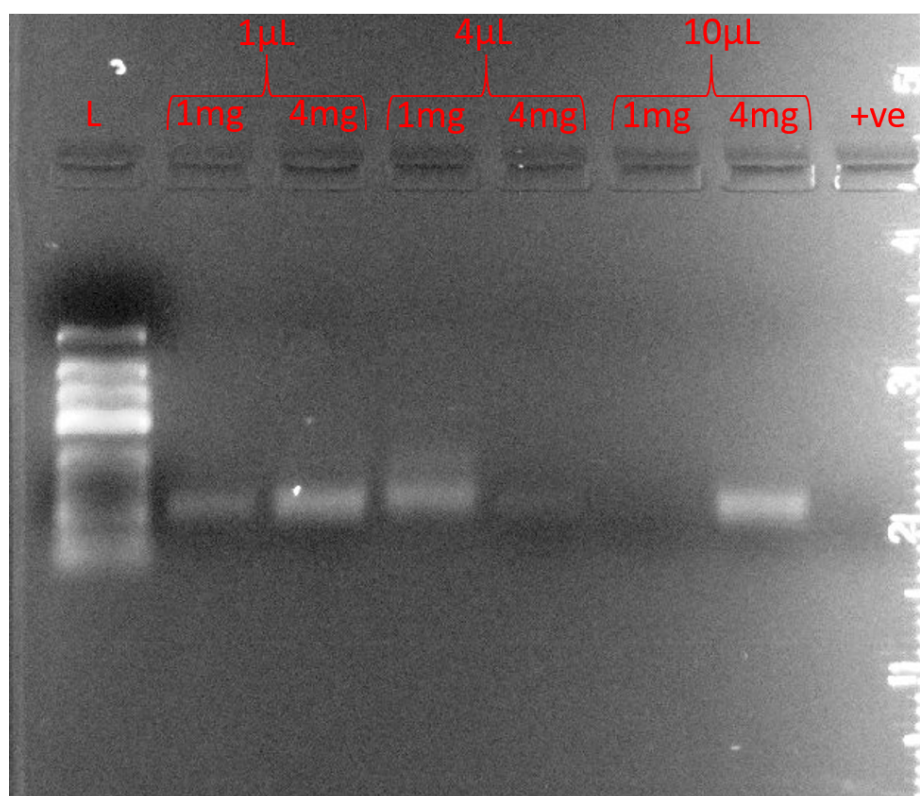


Figure 4.3: SELEX Round 3 products from differing ratios of template: glyphosate-conjugated beads electrophoresed on a 2.5% (w/v) agarose gel.

### 4.3.3 Results of Selection Round 6

The PCR products following the sixth selection round are illustrated in Figure 4.4. Positive selection towards the glyphosate-conjugated beads was evident with no visible DNA bands resulting from control bead incubations at any template volume. As expected, visible 75 bp-sized bands were observed following the amplification of the N40 library that had incubated with the glyphosate-conjugated beads, with band intensity correlated to template volume.

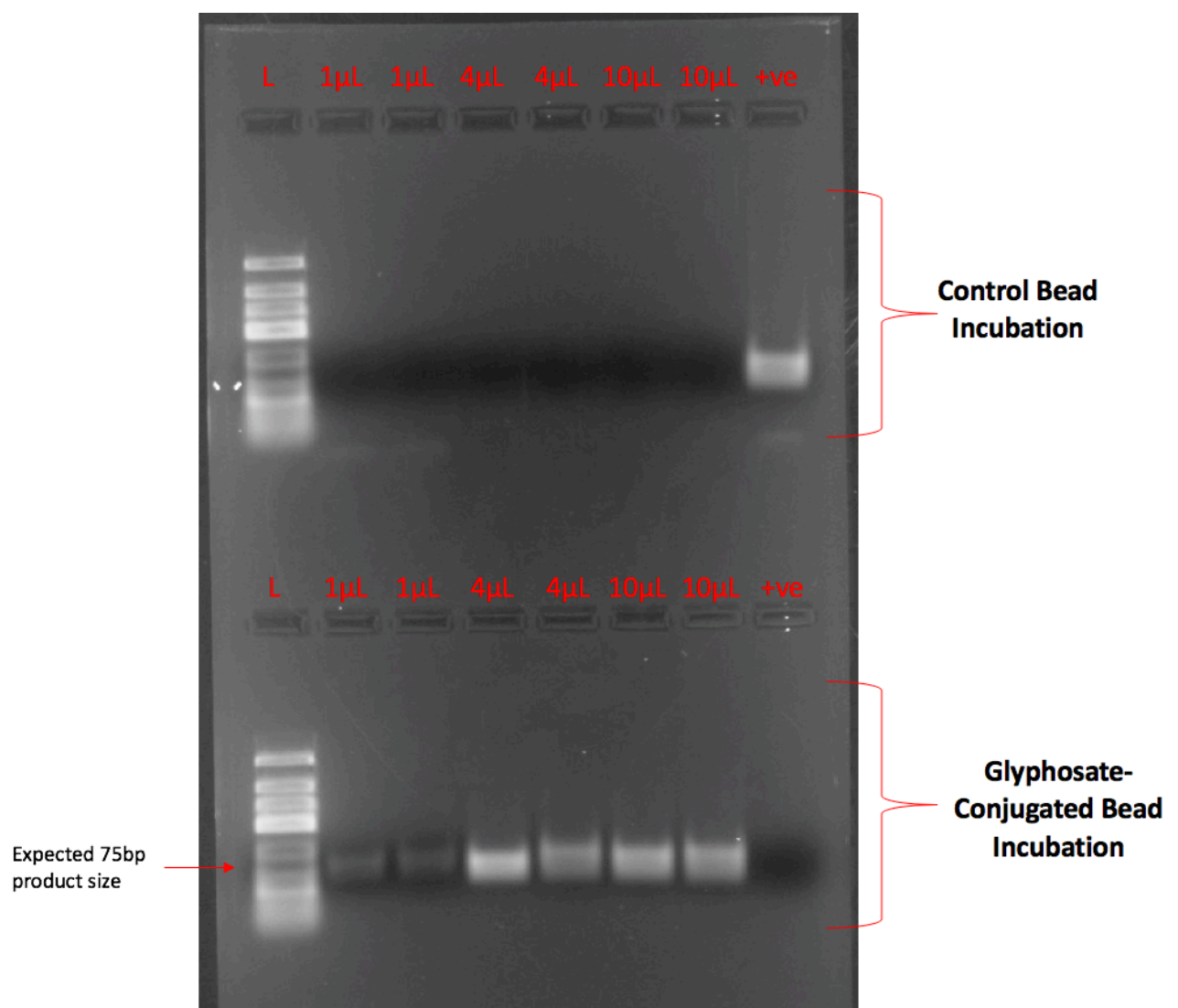


Figure 4.4: SELEX Round 6 products electrophoresed on a 2.5% (w/v) agarose gel.

#### 4.3.4 Results of Selection Round 8

The PCR products following the eighth selection round are illustrated in Figure 4.5. Strong bands are evident even in template incubated with control beads indicating binding to the control affinity matrix. The bands are not as clear or bright as those from template incubated with glyphosate-conjugated beads. None-the-less, a high level of non-specific binding was observed.

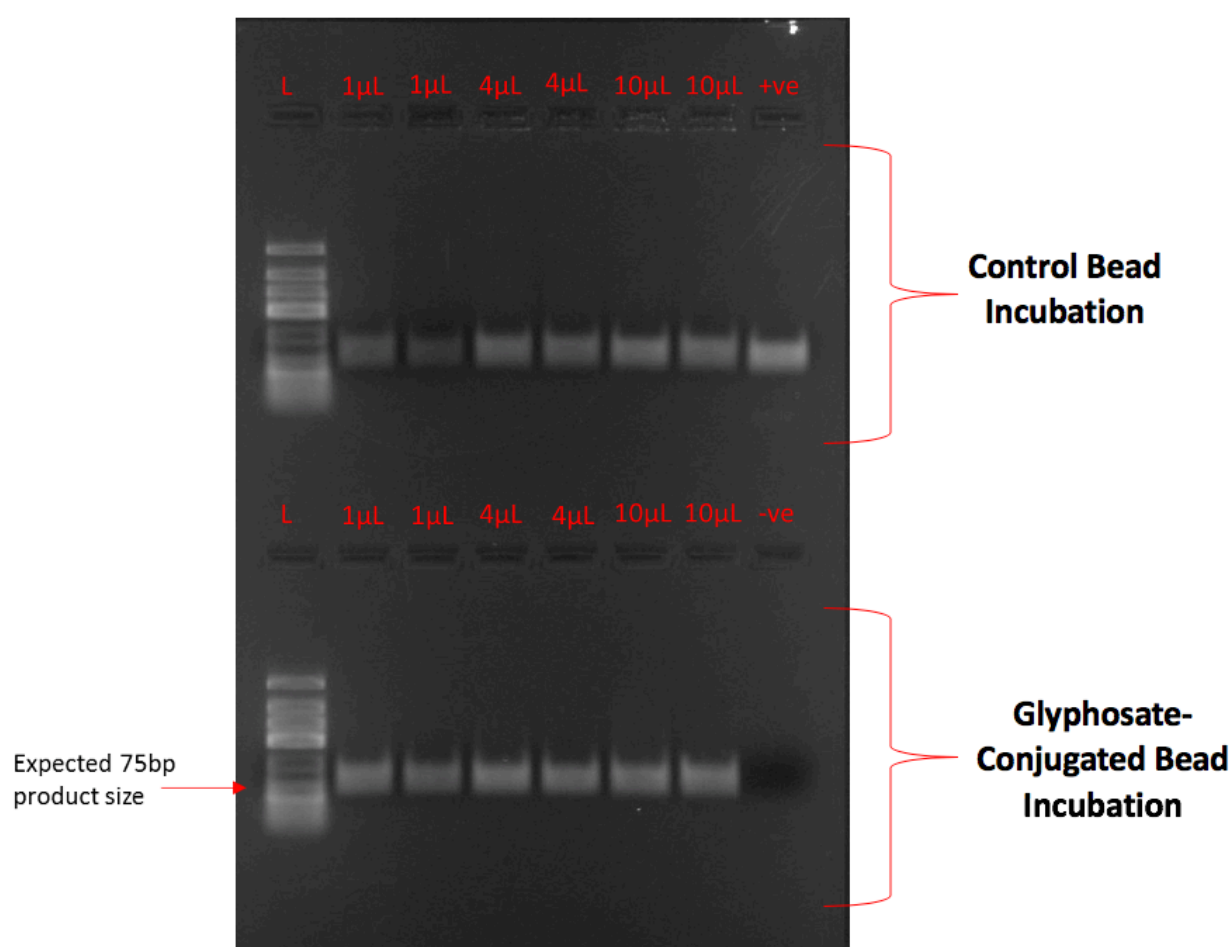


Figure 4.5: SELEX Round 8 products electrophoresed on a 2.5% (w/v) agarose gel.

#### 4.3.5 Results of Selection Round 9 (after Negative Selection)

Due to the presence of oligonucleotides binding to control beads, a negative selection round was performed in Selection Round 9, and is illustrated in Figure 4.6. No bands were evidence following the negative selection round indicating that sequences demonstrating non-specific binding had been removed from the library.

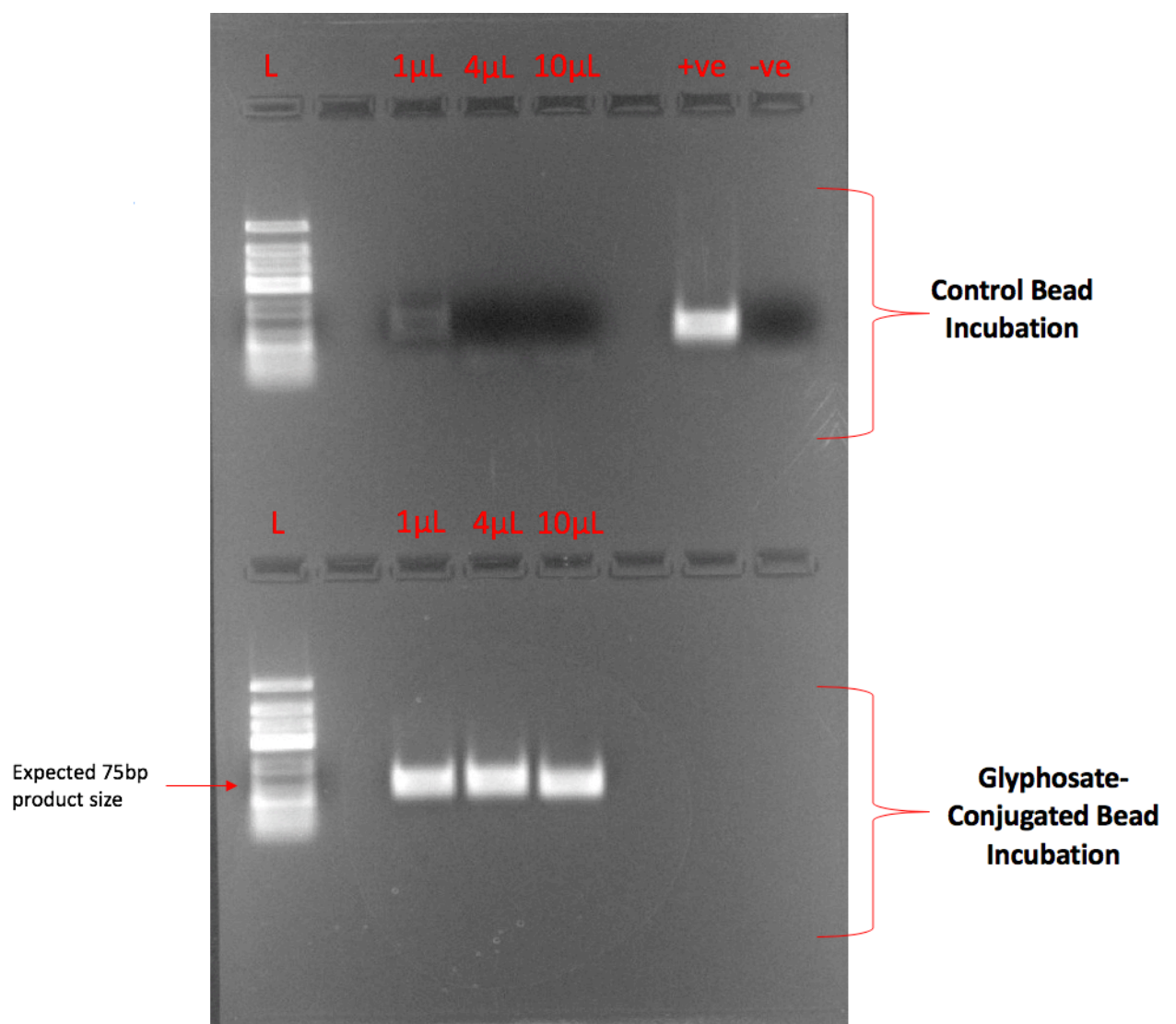
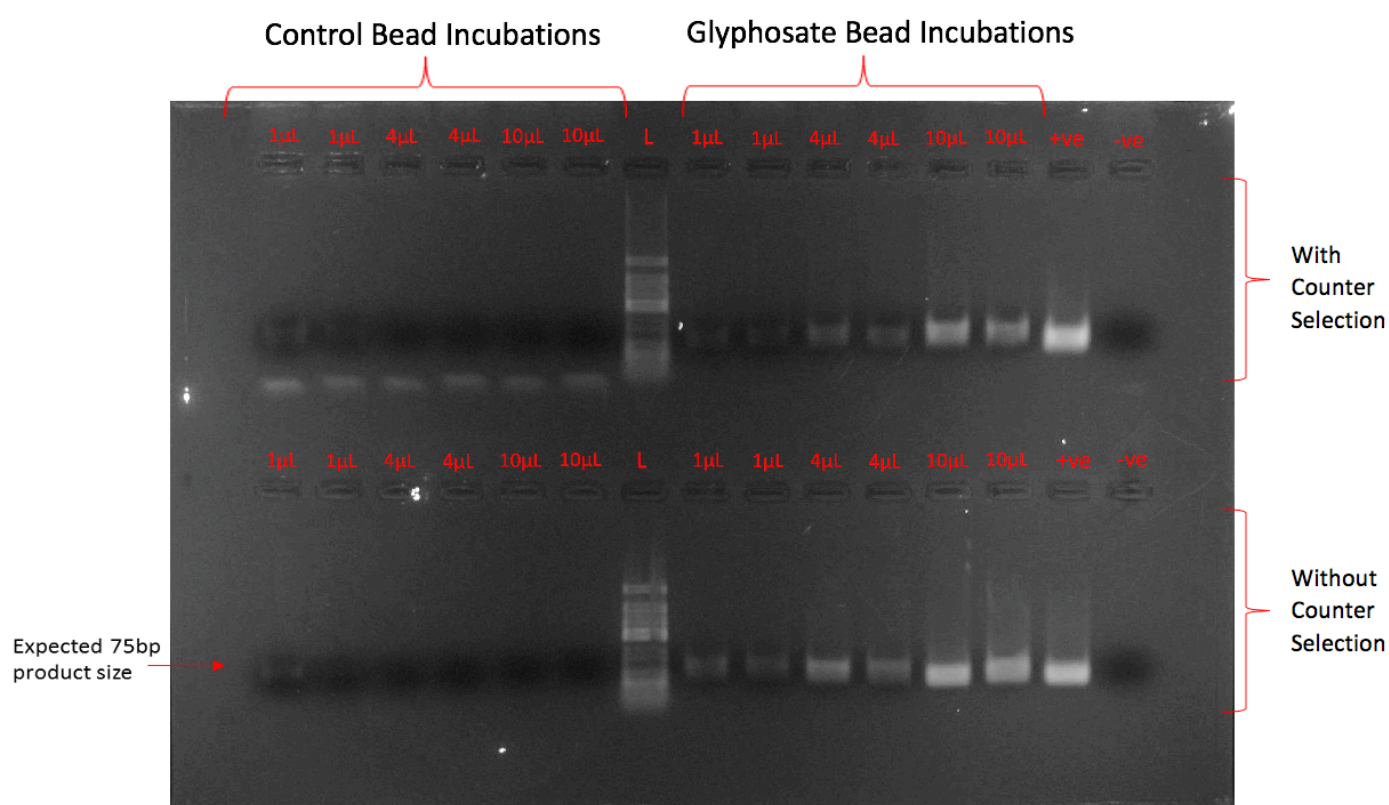


Figure 4.6: SELEX Round 9 products electrophoresed on a 2.5% (w/v) Agarose gel.

#### 4.3.6 Results from Selection Round 10 (including Counter Selection)

Selection Round 10 incorporated a counter selection step involving adding an amino acid mixture to the aptamer library and results are illustrated in Figure 4.7. Only a very small proportion of sequences within the library display non-specific binding as evidenced by the absence of, or faint, bands in control lanes. As expected, a reduction in band intensity was observed after counter selection indicating that sequences demonstrating non-specific binding were removed from the library.



**Figure 4.7: SELEX Round 10 products electrophoresed on a 2.5% (w/v) agarose gel.**

#### 4.3.7 Aptamer Candidates for Glyphosate

Of the 36 plasmid-incorporated oligonucleotides sent for sequencing, 30 were successfully sequenced. Out of these 30 sequences, 23 sequences were identical indicating a strong selection toward this sequence had occurred. The sequence information of the six remaining

unique sequences including length, nucleotide composition and the delta Gibbs free energy of the most stable secondary structure is displayed in Table 4.5. The sequences are aligned in Figure 4.8 with the conserved regions highlighted. A phylogenetic tree showing the relationship between the eight aptamer candidates is shown in Figure 4.9.



A high level of similarity between the seven sequences is evident. The sequence of GLY-04 made up the highest frequency of sequenced clones, accounting for 76.7%. The sequence of GLY-12 was the second highest frequency sequence, with 2 copies out of 30 (6.67%). GLY-15 appears to be the most divergent sequence, with the highest number of unique sites.

**Table 4.5: Glyphosate aptamer candidates**

Sequence Identifier	Sequence (5' to 3')	Frequency	Length (nt)	ΔGibbs Free Energy (kcal/mol)	Nucleotide Composition (N40 region)				
					%A	%C	%G	%T	%GC
GLY-01	ATACGAGCTTGTTCAATANNTGATAGTAAGAGCAATC	1	75	-7.56	17.5	40.0	22.5	20.0	62.5
GLY-02	ATACGAGCTTGTTCAATANNTGATAGTAAGAGCAATC	1	75	-3.21	20.0	35.0	20.0	25.0	55.0
GLY-04	ATACGAGCTTGTTCAATANNTGATAGTAAGAGCAATC	23	75	-4.85	20.0	37.5	20.0	22.5	57.5
GLY-12	ATACGAGCTTGTTCAATANNTGATAGTAAGAGCAATC	2	75	-4.85	15.0	37.5	25.0	22.5	62.5
GLY-15	ATACGAGCTTGTTCAATANNTGATAGTAAGAGCAATC	1	75	-2.58	20.0	37.5	15.0	27.5	52.5
GLY-27	ATACGAGCTTGTTCAATANNTGATAGTAAGAGCAATC	1	75	-5.42	17.5	35.0	22.5	25.0	57.5
GLY-30	ATACGAGCTTGTTCAATANNTGATAGTAAGAGCAATC	1	75	-5.72	17.5	37.5	22.5	22.5	60.0

\*primer regions are highlighted in red. \*\* N40 region sequence information has been redacted for intellectual property protection.

This figure has been redacted for intellectual property protection.

**Figure 4.8: Alignment of the seven candidate aptamer sequences completed using the ClustalW algorithm in MEGA 7.0.**

**Regions of difference are highlighted.**

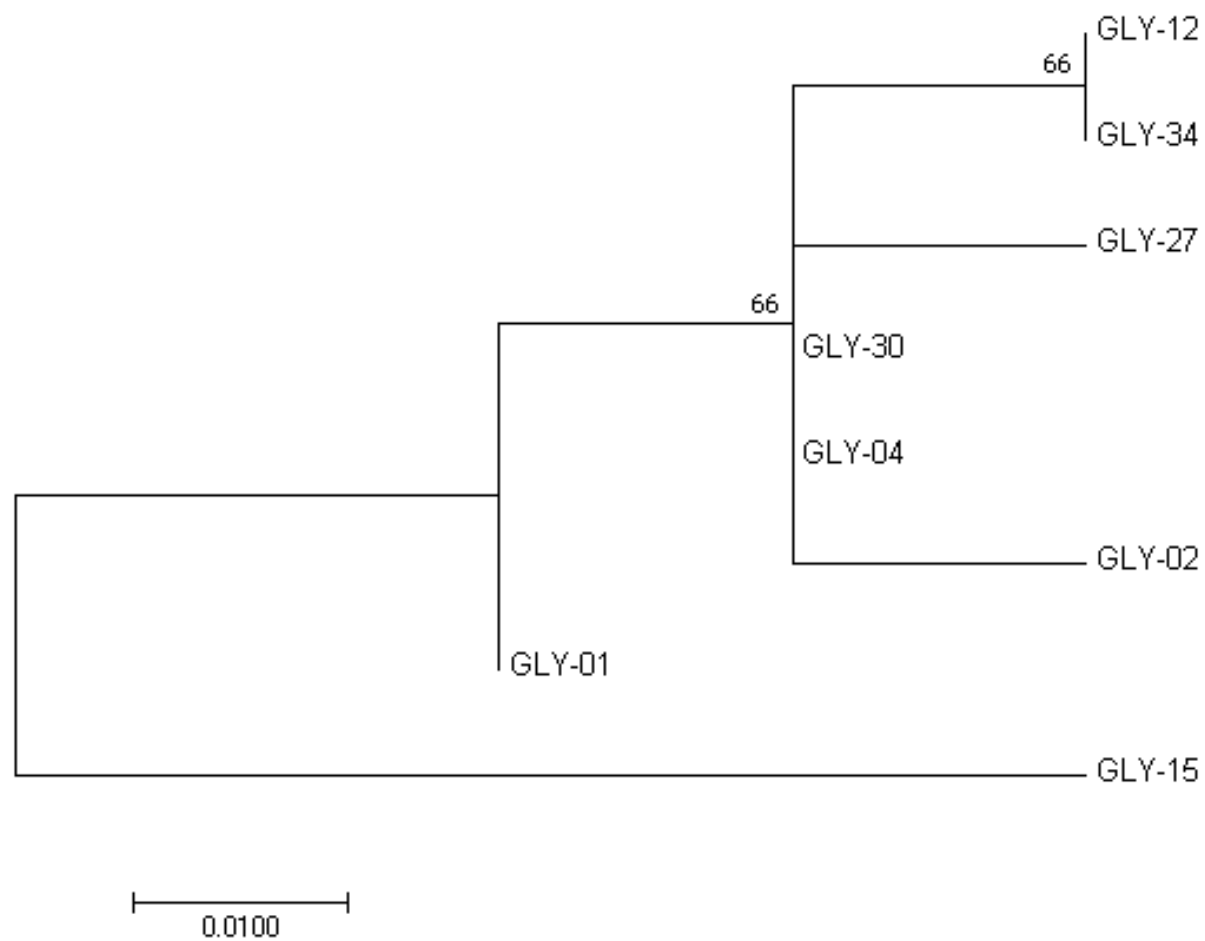


Figure 4.9: A maximum likelihood phylogenetic tree, computed in MEGA 7.0 with 1000 bootstrap replications, indicating the relationship between the candidate aptamer sequences from the Glyphosate SELEX.

#### 4.3.8 Secondary Structures of Aptamer Candidates

The predicted secondary structures of the single-stranded DNA for all eight aptamer candidates indicate that all form stable stem and loop structures in the randomized N40 region under the buffer conditions used during selection as illustrated in Figures 4.10 to 4.16. Stem and loop structures are one of the main motifs used in aptamer-target binding and is a good indication that these oligonucleotides should bind to the target molecule.

This figure has been redacted for intellectual property protection.

**Figure 4.10: The predicted secondary structure of GLY-01.**

This figure has been redacted for intellectual property protection.

**Figure 4.11: The predicted secondary structure of GLY-02.**

This figure has been redacted for intellectual property protection.

**Figure 4.12: The predicted secondary structure of GLY-04.**

This figure has been redacted for intellectual property protection.

**Figure 4.13: The predicted secondary structure of GLY-12.**

This figure has been redacted for intellectual property protection.

**Figure 4.14: The predicted secondary structure of GLY-15. Sequence information has been redacted for intellectual property protection.**

This figure has been redacted for intellectual property protection.

**Figure 4.15: The predicted secondary structure of GLY-27. Sequence information has been redacted for intellectual property protection.**

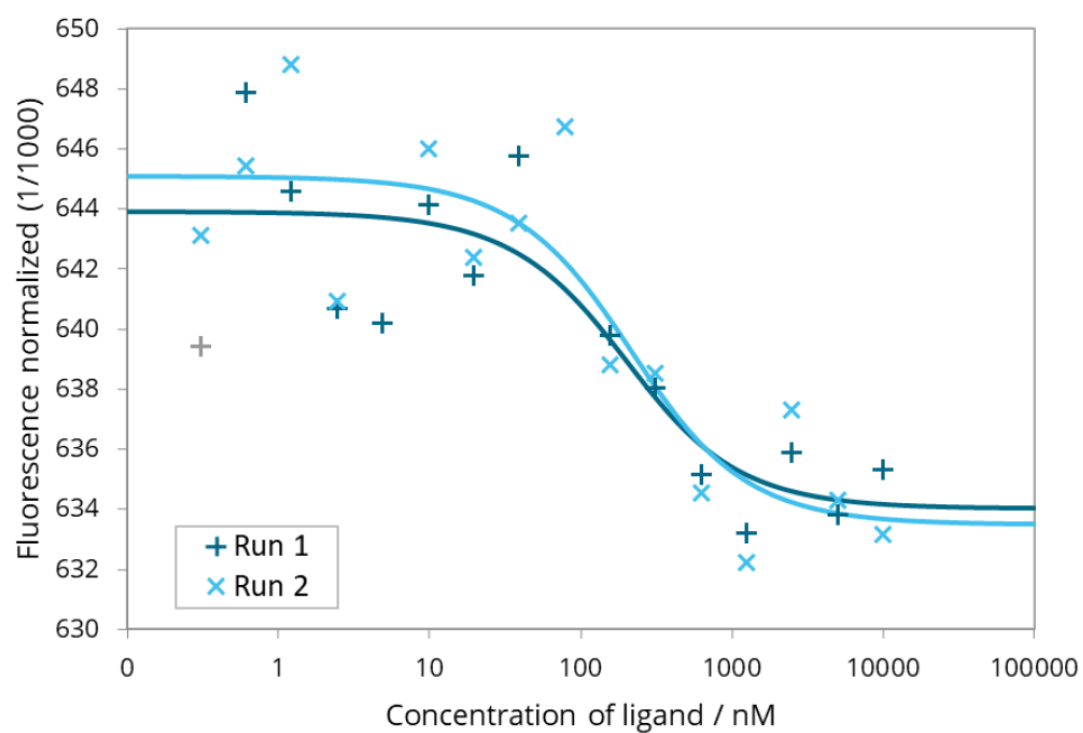
This figure has been redacted for intellectual property protection.

**Figure 4.16: The predicted secondary structure of GLY-30. Sequence information has been redacted for intellectual property protection.**

#### 4.3.9 Characterisation of Aptamer Candidates by MST

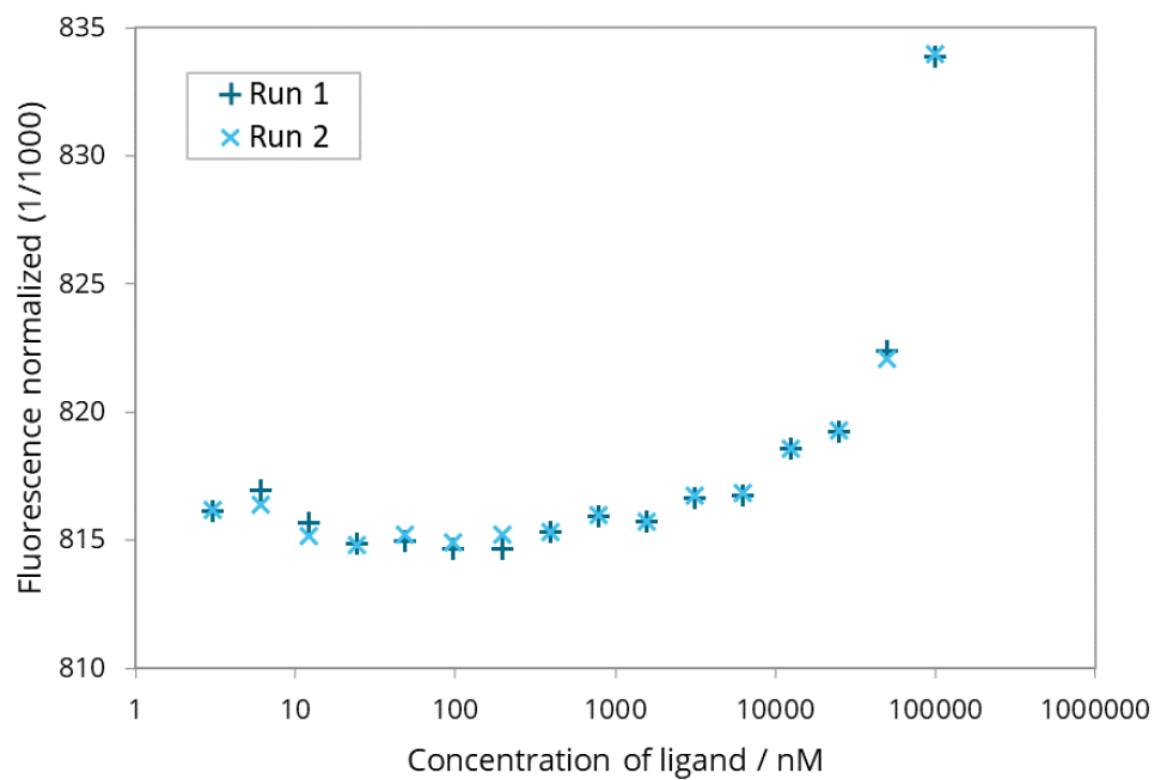
Both aptamer candidates demonstrated binding to glyphosate in the MST assay. The binding interaction of GLY04 with glyphosate showed a typical sigmoidal binding isotherm and had an average  $K_d$  value of 156 nM across the two technical repeats (Figure 4.17). The data for GLY12 indicated a binding interaction however, no plateau in signal was reached at the highest glyphosate concentration (10  $\mu$ M) and therefore a  $K_d$  value was unable to be calculated (Figure 4.18). The binding interaction for GLY12 occurred at much higher ligand concentrations (> 10,000 nM) than for GLY04 (between 100 and 1000 nM). It should be noted that in the assay for GLY04, the signal to noise ratio was low especially at concentrations of less than 100 nM.





	$K_D$	Amplitude	S/N
<b>Run 1</b>	148 nM	-9.9	4.6
<b>Run 2</b>	164 nM	-11.6	4.8

Figure 4.17: Results of the MST binding assay for glyphosate aptamer candidate GLY04.



	$K_D$	Amplitude	S/N
Run 1	---	---	---
Run 2	---	---	---

Figure 4.18: Results of the MST binding assay for glyphosate aptamer candidate GLY12.

## 4.4 Discussion

This study presents the first published report in scientific literature (to our knowledge) of an aptamer that demonstrates binding to glyphosate. Although binding has been demonstrated for GLY04, a low signal-to-noise ratio in the MST assay indicates that the assay should be repeated using a second independent experiment. I would recommend repeating the MST experiment in triplicate to ascertain whether the aptamer truly binds to glyphosate and also validating the binding with another characterisation method such as isothermal titration calorimetry. Unfortunately, due to cost and time considerations, this was unable to be completed during the course of the present study. A  $K_d$  in the mid-nanomolar range is low enough that the aptamer would be useful for monitoring glyphosate at environmentally relevant concentrations. Both of the aims of this study (stated in the introduction) have been achieved.

The results of this SELEX were quite unusual in that out of the 30 clones which were successfully sequenced, 23 of the sequences were identical. The remaining sequences had very high sequence similarity to each other as well. It is unusual for the frequency of any sequence to be this high after 10 rounds selection. This could indicate that the selection was very successful and the library was highly enriched for these sequences or that the sequences are PCR artefacts. This SELEX was completed in a short number of rounds (10, compared to 12-18 rounds in typical affinity-matrix SELEX), therefore it is unlikely that these sequences were favoured due to amplification efficiency rather than affinity to the target molecule. In addition, GLY04 (the most represented sequence) has demonstrated binding to the target molecule indicating it is likely highly enriched due to binding affinity rather than being a favourably amplified sequence. GLY12 demonstrated a response in the characterisation assay

but appeared to have a low affinity to glyphosate. This may be the result of a non-specific interaction. The only sequence differences between GLY04 and GLY12 are two nucleotide substitutions. GLY12 contains Guanine residues at positions 40 and 41, while GLY04 contains Adenine at both positions. Position 40 and 41 are within a predicted stem and loop structure for both GLY04 and GLY12 (specifically, the second and third residue of the 8nt loop, see Figure 4.12 and 4.13). The secondary structure remains the same despite the different sequence. I hypothesize that the two Adenine residues give GLY04 a binding advantage and that this stem and loop is involved in target binding. Further characterisation of the aptamer, such as a ligand binding domain assay and/or mutation of these residues followed by MST or an equivalent characterisation method are needed to determine whether this is the case. Characterisation of a truncated sequence which includes this stem and loop would also be beneficial as truncated aptamer sequences have been shown to have improved binding affinity.

There are some limitations associated with this study. Cloning and Sanger sequencing was used to provide sequence information on the aptamer candidates. This method is low throughput meaning that only 30 individual clones were able to be assessed. Only sequences from the final three rounds were cloned (12 from each) providing very limited information on the evolution of the library. Methods for sequencing aptamer libraries with high-throughput sequencing were not established in our lab at the time of this study. In addition, it was deemed unnecessary to use HTS as many successful aptamers had been selected with cloning and Sanger sequencing. HTS data of multiple SELEX rounds would have enabled easier selection of candidates (through the use of enrichment data) and also provided information about the evolution of the library. It is possible that higher affinity binders appeared early in

the library and were then diluted as moderate affinity binders that were efficiently amplified dominated the library. Without the use of HTS, this cannot be determined.

The SELEX strategy employed here was able to select a glyphosate-binding aptamer from a random nucleic library and the aptamer was able to bind the target molecule in laboratory conditions, with the presence of a salt buffer. Further characterisation is needed to determine if the aptamer will function in a complicated freshwater sample matrix.

Characterisation using other methods was attempted for this set of aptamers however, the results were inconsistent. Fluorescence polarisation/anisotropy, a fluorescent bead assay and a SYBR Green assay were all attempted (data not shown). MST is considered one of the most reliable methods for small molecule aptamer characterisation as it is not dependant on the mass change of the aptamer-target complex compared to the aptamer alone (which can be difficult to detect for small molecule aptamer interactions). MST does have some limitations such as the requirement for the aptamer to be labelled with a fluorescent dye, which could have an impact on binding affinity. In order to provide further evidence that GLY04 is able to bind glyphosate, characterisation using isothermal titration calorimetry alongside the existing MST results is recommended. During ITC, the aptamer is unlabelled meaning that a dye cannot interfere with target binding. The target is unconjugated and free in solution, as is the aptamer, limiting the effect of steric hindrance and meaning all functional groups of both molecules are available for binding interactions. Although a binding interaction is evident in the MST data, validation with another reliable characterisation method such as ITC would provide further evidence of target binding. Structural characterisation of aptamer-target complexes using X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy

would enhance our understanding of the nature of the interaction and are recommended as a future step.

GLY04 is the first aptamer reported to show a response to glyphosate in MST binding studies. There is evidence to suggest that GLY04 is able to bind glyphosate with a high affinity. Integration of the aptamer into a bio-sensing platform would enable more robust testing of this. By coupling GLY04 with an appropriate bio-sensing platform, near real-time, onsite detection of water samples could be achieved.

# Chapter 5: Selection of a nonylphenol binding aptamer using standard SELEX

## 5.1 Background and Introduction

Nonylphenol ( $C_{15}H_{24}O$  ; 4-(2,4-dimethylheptan-3-yl)phenol) is part of a growing class of emerging organic contaminants (EOC), which are potential endocrine disruptors. Technical nonylphenol equivalents, defined as the total nonylphenol load including all potential isomers, was measured in the Waiwhetu Stream at above environmental guidelines. The level of nonylphenol contamination in other freshwater ways in New Zealand is unknown. An extensive review of nonylphenol is found in Chapter 1.1.5 and additional information is provided in Chapter 3.

Detecting phenolic compounds such as nonylphenol using GC-MS is very labour intensive and costly. In order to routinely monitor nonylphenol in the New Zealand aquatic receiving environment, a cost-effective method of detection is required. An aptasensor provides a solution to this problem.

To my knowledge, there are no published aptamers which bind specifically to technical nonylphenol equivalents, *i.e.* a mixture of isomers, or any of the individual isomers. Nonylphenols are produced by nonylphenol ethoxylates breaking down in the environment. Earlier this year, a ssDNA aptamer specific for nonylphenol ethoxylate 12 was generated with a  $K_d$  of  $100.9 \pm 13.2$  nM (Kim et al., 2019). The selection was completed using reduced graphene oxide (rGO) SELEX. The aptamer showed specific binding to nonylphenol ethoxylate

12 and not to similarly structured compounds such as nonylphenol, bisphenol A (BPA), dibutyl phthalate or phenyl phenol. Reduced graphene oxide was also used in a sensing system to detect nonylphenol ethoxylate 12 in spiked water samples and achieved a limit of detection (LOD) of 694 pM (Kim et al., 2019). Two of the other aptamer candidates bound to a technical nonylphenol mixture but were not specific and showed a response to BPA, dibutyl phthalate and phenyl phenol (Kim et al., 2019). These aptamers would not be suitable for assessing nonylphenol contamination due to their ability to detect other similarly structured targets. Aptamers to similar molecules such as BPA, 17 $\beta$ -oestradiol and various poly-chlorinated biphenyls have already been published (Alsager et al., 2014; Jo et al., 2011; Mehta et al., 2012; Xu et al., 2012) .

There is a gap in the literature which shows proof-of-concept of using aptasensors for nonylphenols. This gap is demonstrated by the variety of aptasensors which have been made for oestradiol or BPA detection, while none have been produced for nonylphenol. This can be attributed to the fact that no aptamers have been published which bind solely to nonylphenols. Nonylphenols only have a single functional group (a hydroxyl group) making them difficult to conjugate to an affinity matrix. A lack of functional groups also means that selecting an aptamer for a compound is much more difficult.

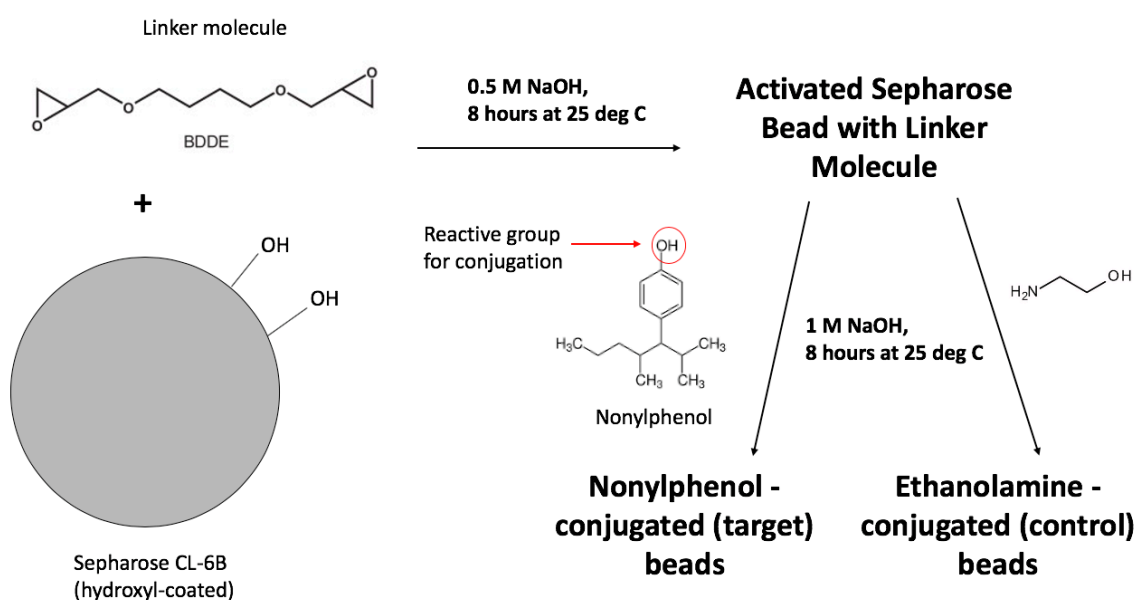
The aim of this chapter was to develop an aptamer that was specific for technical nonylphenol equivalents and bound with a nanomolar affinity.



## 5.2 Methodology

### 5.2.1 Synthesis of Nonylphenol-Conjugated Affinity Matrices

In order to partition target-bound from unbound oligonucleotides, an affinity matrix needs to be used to which the target molecule is conjugated. Sepharose CL-6B (Sigma Aldrich) was used as the affinity matrix and BDDE (a linker molecule) was used to activate the matrix. Conjugation of technical nonylphenol equivalents to the linker molecule was achieved via the hydroxyl group. Blocking of unused sites was completed using ethanolamine (Figure 5.1).



**Figure 5.1: Scheme for conjugating nonylphenol to CL-6B Sepharose beads**

An 8 mL aliquot of CL-6B Sepharose beads was pipetted into a sintered glass funnel in a fume hood. The beads were washed sequentially with 500 mL of 0.5 M NaCl, 1000 mL of distilled water and finally, 200 mL of 1M NaOH. Beads were then dried under vacuum for 10 minutes and weighed. For the addition of the linker molecule, a 15 mL mixture containing 8 mL of 1 M NaOH, 4.6 mL of distilled water and 2.4 mL of 1, 4-butanediol diglycidyl ether was added to

the washed beads in a round bottomed flask. The beads and activation mixture were then incubated on a shaking incubator at room temperature for 8 hours at 300rpm.

Following incubation, the beads were washed with 500 mL of 1 M NaOH followed by 2000 mL of distilled water and dried under vacuum for 10 minutes. The beads were then divided into two portions for either the addition of the target molecule and as a control matrix. A 1.33 mM solution of technical nonylphenol equivalents in 15 mL of 1 M NaOH was added to one portion of the beads for addition of the target molecule to the BDDE linker. A 15 mL solution of 1 M ethanolamine was added to the other portion that acts as a control matrix. Both the control and target bead mixtures were incubated overnight at room temperature on the shaking incubator at 300 rpm. The control beads were then washed with 2000 mL of distilled water, dried under vacuum and then stored dry at 4°C until further use. Conversely, the target beads were washed with 500 mL of 1 M NaOH followed by a wash in 2000 mL distilled water. The target beads were then blocked by adding 15 mL of 1 M ethanolamine and incubated in a shaking incubator at 300 rpm overnight at room temperature. Finally, the target beads were washed with 2000 mL distilled water and stored in the same manner as described above for the control beads.

### 5.2.2 Conjugation Assessment

In order to confirm the conjugation of technical nonylphenol equivalents onto the sepharose matrix, approximately 200 mg of both control and target-conjugated beads were resuspended in 1 mL of 20% ethanol (v/v). The beads were then gently pipetted into a Quartz cuvette with a path length of 1 mm (Starna, Essex, United Kingdom) and allowed to settle to the bottom of the cuvette to form a homogenous layer. Any liquid on top of the beads was removed and this process was repeated until a homogenous layer of beads filled

approximately 75% of the height of the cuvette. A UV-Vis spectra of the beads was then measured from 200 nm to 800 nm using a Cary 500 UV-Visible Photo Spectrometer (Agilent Technologies, Santa Clara, California, United States of America). The presence of an absorption peak at 225 and 280 indicates the presence of nonylphenol molecules on the beads.

### 5.2.3 Preparation of the Affinity Matrix for SELEX

Preparation of the affinity matrix was completed using the same method outlined in Section 4.2.1.

### 5.2.4 Washing of Sepharose Beads

Beads were washed using the same procedure as described in Section 4.2.4. The volume and percentage of non-ionic detergent contained in the binding and washing buffer (BWB) was varied according to the desired selection conditions. An outline of the washing conditions for the selection of a nonylphenol aptamer is listed in Table 5.1.

**Table 5.1: Washing conditions used in the selection of an aptamer for binding nonylphenol.**

<b>Selection Round</b>	<b>Washing Conditions</b>
1	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
2	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
3	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
4	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
5	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
6	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
7	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
8	2 mL of 1X BWB with 0.5% (v/v) IGEPAL followed by 3 mL of 1X BWB with 0.1% (v/v) IGEPAL

9	5 mL of 1X BWB with 0.1% (v/v) IGEPAL
10	3 mL of 1X BWB (No detergent)
11	3 mL of 1X BWB (No detergent)
12	3 mL of 1X BWB (No detergent)
13	3 mL of 1X BWB (No detergent)
14	3 mL of 1X BWB (No detergent)
15	3 mL of 1X BWB (No detergent)
16	3 mL of 1X BWB (No detergent)
17	<b>Counter selection:</b> 1 mL of detergent cocktail followed by 5 mL of 1X BWB (no detergent) and 5 mL of distilled water <b>Normal selection:</b> 3 mL of 1X BWB (No detergent)
18	3 mL of 1X BWB (No detergent)
19	5 mL of 1X BWB (No detergent)
20	<b>Counter selection:</b> 1 mL of detergent cocktail followed by 5 mL of 1X BWB (no detergent) and 5 mL of distilled water <b>Normal selection:</b> 5 mL of 1X BWB (No detergent)
21	2 mL of 1X BWB with 0.01% (v/v) IGEPAL followed by 3 mL 1X BWB (No detergent)

### 5.2.5 PCR Amplification of Bound Oligonucleotides

Beads that were recovered from the filter column and re-suspended in distilled water were added directly to the PCR reaction as template. The template amounts were varied to gauge the binding of the oligonucleotide library to the control and target matrices. Additionally, the number of cycles was modified depending on the desired level of amplification for each selection round. The concentration of magnesium ions was also varied to induce mutation when desired. The different PCR conditions for each round of SELEX is listed in Table 5.2. The general cycling conditions and PCR reaction reagents are as detailed in the General Materials and Methods chapter (Section 2.3).

**Table 5.2: PCR conditions used for each nonylphenol SELEX round.**

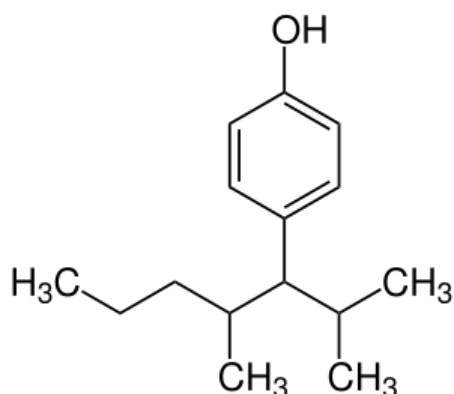
<b>Selection Round</b>	<b>Number of Cycles</b>	<b>Volume of 25 mM Magnesium Added (μL)</b>	<b>Template Amounts Used (μL)</b>
1	35	0	1, 4 and 10
2	35	0	1, 4 and 10
3	33	0.5	1, 4 and 10
4	35	1	1, 4 and 10
5	30	0	5 and 10
6	32	0	5 and 10
7	32	1	3 x 10 (enrichment)
8	32	0	5 and 10
9	35	0	3 x 10 (enrichment)
10	35	0	5 and 10
11	35	0.3	5 and 10
12	35	0	5 and 10
13	35	0.3	5 and 10
14	35	0.3	5 and 10
15	35	0.3	5 and 10
16	35	0	5 and 10
17	35	0	5 and 10
18	35	0	1, 5 and 10
19	35	0	5 and 10
20	35	0	5 and 10
21	35	0	5 and 10

#### 5.2.6 Counter Selection Protocol

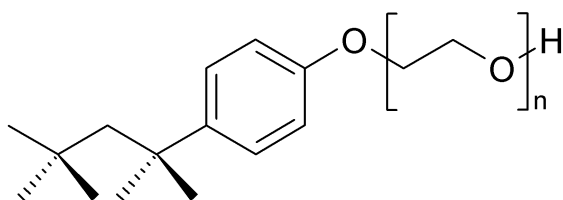
Counter-selection was used during the 17<sup>th</sup> and 20<sup>th</sup> selection round to ensure that oligonucleotides with affinity to molecules that have a similar structure to the technical nonylphenol equivalents were removed from the selection pool. A counter selection step

ensures that oligonucleotides selected for further characterisation are very specific to the target of interest. In this case, four similar detergent molecules were selected to provide a stringent counter-selection. The chosen molecules were IGEPAL, Nonidet P40, Tween 20 and Triton X-100 (Figure 5.2). All four molecules were added to 5 mL of distilled water to create a solution that contained all four molecules at a concentration of 0.01% (v/v). The pool of ssDNA from the prior selection round was divided into four equal volumes. Two of the ssDNA aliquots were added to a control bead matrix and two were added to the target bead matrix, for overnight incubation. Following incubation, one of each matrix type (control and target-conjugated) was washed as per normal (see Table 5.1 for specific washing conditions) and the other was washed with 1 mL of the detergent mixture (described above), followed by 5 mL of BWB and 5 mL of distilled water. The detergent mixture was left in the column to incubate with the beads for 5 minutes before being flowed through; ensuring that oligonucleotides which bound quickly (and therefore should have high affinity) to the counter-selection agents would be removed from the selection pool. The additional washing step is to remove any residual detergent, which may interfere with PCR amplification.

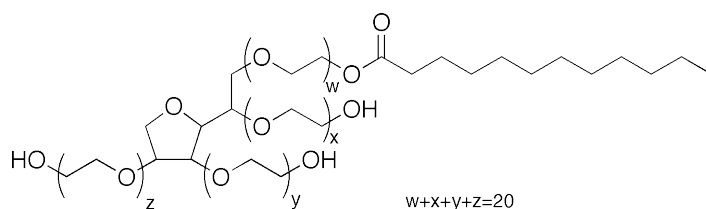
Nonylphenol:



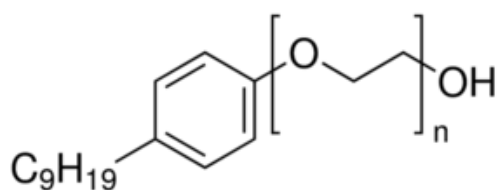
Triton X-100:



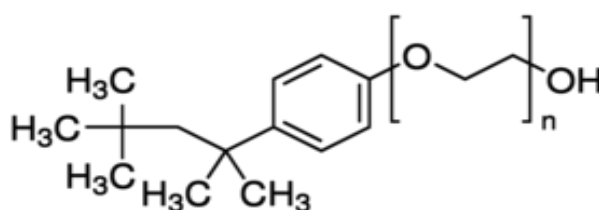
Tween-20:



Nonidet P40:



IGEPAL CA-630:



**Figure 5.2: Structures of the target molecule nonylphenol, and the compounds used for counter selection (Triton X-100, Tween-20, Nonidet P40 and IGEPAL CA-630).**

### 5.2.7 Cloning of Aptamer Candidates into a Plasmid Vector

It should be noted that although 21 rounds of selection were completed, PCR products from Round 20 and 21 were not cloned into a plasmid vector for sequencing. Products from these rounds appeared to bind to the control matrix as DNA was able to be amplified from the

control beads. This was evident as the gel bands from the control bead PCR reaction were of a similar brightness to the DNA from the target beads. Therefore, the decision was taken to only clone product from Selection Round 19, the last round which appeared to be enriched for oligonucleotides that preferentially bound the target matrix. The cloning procedure is described previously in the General Materials and Methods chapter (Section 2.9)

#### 5.2.8 Analysis of Resulting Candidate Sequences

Analyses of the resultant DNA sequences were completed as described previously in the General Materials and Methods chapter (specifically, section 2.10 and 2.11) with some exceptions. In the case of these aptamer candidates, the similarity of the nonylphenol aptamer candidates to an aptamer commonly used in our laboratory (R18C1) was assessed. To do this, the sequence of R18C1 was imported into MEGA 7.0 and a pairwise distance matrix was used to compare homology between sequences. Sequences that showed homology to R18C1 were excluded from further analyses (see Section 5.3.7). R18C1 was also included when generating a phylogenetic tree to analyse the relationship between aptamer candidates.

#### 5.2.9 Characterisation of Aptamer Candidates

A fluorescent binding assay that followed a similar methodology to that previously described by Li (2016) was used to assess the binding affinity of aptamer candidates to technical nonylphenol equivalent-conjugated sepharose beads. Synthesized oligonucleotides were ordered from Sangon Biotech (Shanghai, China) at a synthesis scale of 10 OD. Ordered oligonucleotides were HPLC-purified and labelled with a HEX fluorophore at the 5' end. Nonylphenol-conjugated beads were prepared as described in Section 5.2.1. 600  $\mu$ L of a 2  $\mu$ M working solution of each candidate aptamer was made up in 1X BWB from the 100  $\mu$ M stock



solution. The working solution was denatured for 5 minutes at 95°C and then cooled on ice for 10 minutes. Following denaturation, the DNA was left at room temperature for 10 minutes. 200 µL of the following concentrations of the oligonucleotides were made up using 1X BWB, from the 2 µM working solution: 2 µM, 1.5 µM, 1 µM, 500 nM, 250 nM and 100 nM. All oligonucleotides were made up in amber 1.7 mL microcentrifuge tubes. Nonylphenol beads, stored at a concentration of 200 mg/mL in 20% (v/v) Ethanol were aliquoted out in a 20 µL volume into amber 1.7 mL micro centrifuge tubes. The beads were then washed by adding 200 µL of 1X BWB, shaking for 30 seconds at 750 rpm at room temperature followed by centrifugation for 2 minutes at 13,300 x g. After centrifugation, 170 µL of supernatant was removed to leave the beads suspended in a 50 µL volume of 1X BWB. 50 µL of oligonucleotides at the concentrations specified above was added to separate 50 µL aliquots of bead suspension and incubated for 1 hour at room temperature on a rotator. Following incubation, the beads were washed three times with 1X BWB and resuspended in a volume of 200 µL. The entire volume of beads was transferred into a black, clear-bottom 96-well microtitre plate (NUNC, Roskilde, Denmark) using a 200 µL micropipette with the tip cut (to widen the bore size) and the fluorescence of each well was read using a Clariostar microplate reader (BMG Lab Tech, Ortenberg, Germany). The instrument gain was set to 1000 and the fluorescence was read from the bottom at an excitation wavelength of 535 nm and an emission wavelength of 556 nm. A control with no oligonucleotides was also prepared where instead, 50 µL of 1X BWB was added to the beads and used as a blank to account for the fluorescence of the beads alone. This blank reading was subtracted from the raw fluorescence values to give a corrected fluorescence level. Analyses of binding affinities were completed in the software programme Prism 5. A non-linear regression using a saturation binding model (one site – total) was used

to fit a binding isotherm and calculate a  $K_d$ . The background was constrained to zero as the background fluorescence of the beads had already been subtracted from the data.

#### 5.2.10 Characterisation of NON-12 using Micro-Scale Thermophoresis

Due to none of the aptamer candidates demonstrating affinity in the fluorescent binding assay, the candidate with the most stable secondary structure (NON-12) was selected for characterisation using micro-scale thermophoresis (MST). The characterisation of this aptamer using MST was completed entirely by 2Bind (Regensburg, Germany). The composition of the assay buffer, ligand solubility instructions and potential binding affinity advice were provided to 2Bind by myself. The protocol for MST was designed and implemented by 2Bind as follows. A 10 mM stock solution of technical nonylphenol equivalents were prepared in 100% ethanol and diluted to a 1 mM working solution in 100% ethanol. The concentrations of technical nonylphenol equivalents were made up using 5% ethanol to ensure that the ligand remained in-solution. The instrument used for MST was a Monolith NT.115 (Nanotemper, Munchen, Germany). The concentration of the aptamer was kept constant at 5nM and the serial dilutions of the ligand were prepared using two-fold dilutions from 100  $\mu$ M to 3.05 nM. Both the ligand and aptamer were made up in 1X BWB with 5% (v/v) ethanol. A aliquot of 5  $\mu$ L of the oligonucleotide was mixed with 5  $\mu$ L of the ligand at each concentration and the mixture was added to a standard capillary for analyses on the instrument. The instrument was set to 80% LED power and 60% laser power and the temperature was set to 25°C. Two technical repeats were completed. Data analyses and the fitting of a binding isotherm was completed by 2Bind using the software supplied with the instrument.

## 5.3 Results

### 5.3.1 Conjugation of Nonylphenol to the Affinity Matrix

A UV-Visible spectroscopy confirmed that technical nonylphenol equivalents were successfully conjugated to the affinity matrix as evidenced by absorption peaks at both 225 and 280 nm (Figure 5.3), compared to the control (Figure 5.4).

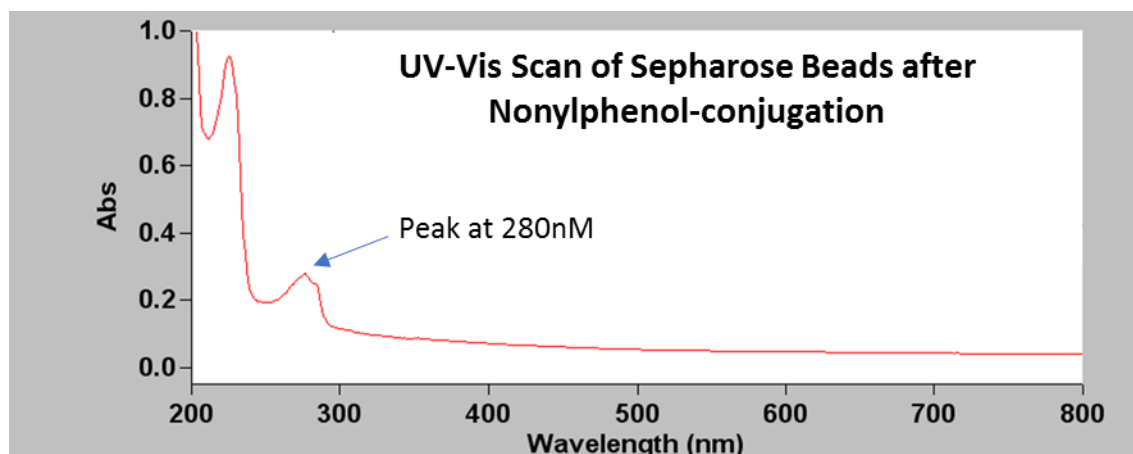


Figure 5.3: Absorption spectra of Nonylphenol-conjugated beads.

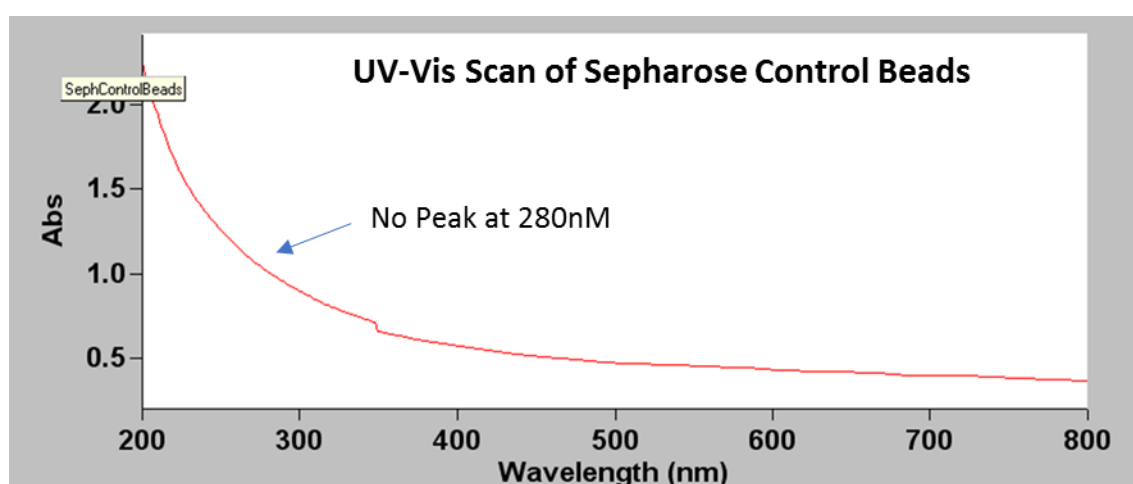
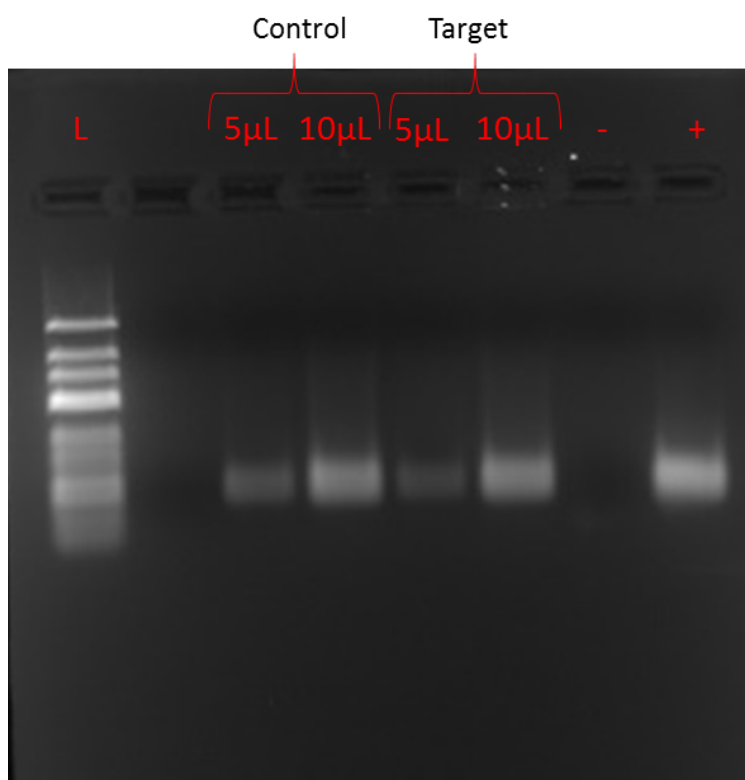


Figure 5.4: Absorption spectra of control beads (blocked with ethanolamine).

### 5.3.2 Results of Selection Round 5

After Selection Round Five, DNA was amplified from both the control matrix and the nonylphenol-conjugated (target) matrix (Figure 5.5). The band intensities were similar when using the same amount of template indicating that the library has equal affinity to the control and target beads. Thus, the library is not enriched by SELEX Round Five for sequences that preferentially bind to the target-conjugated beads.

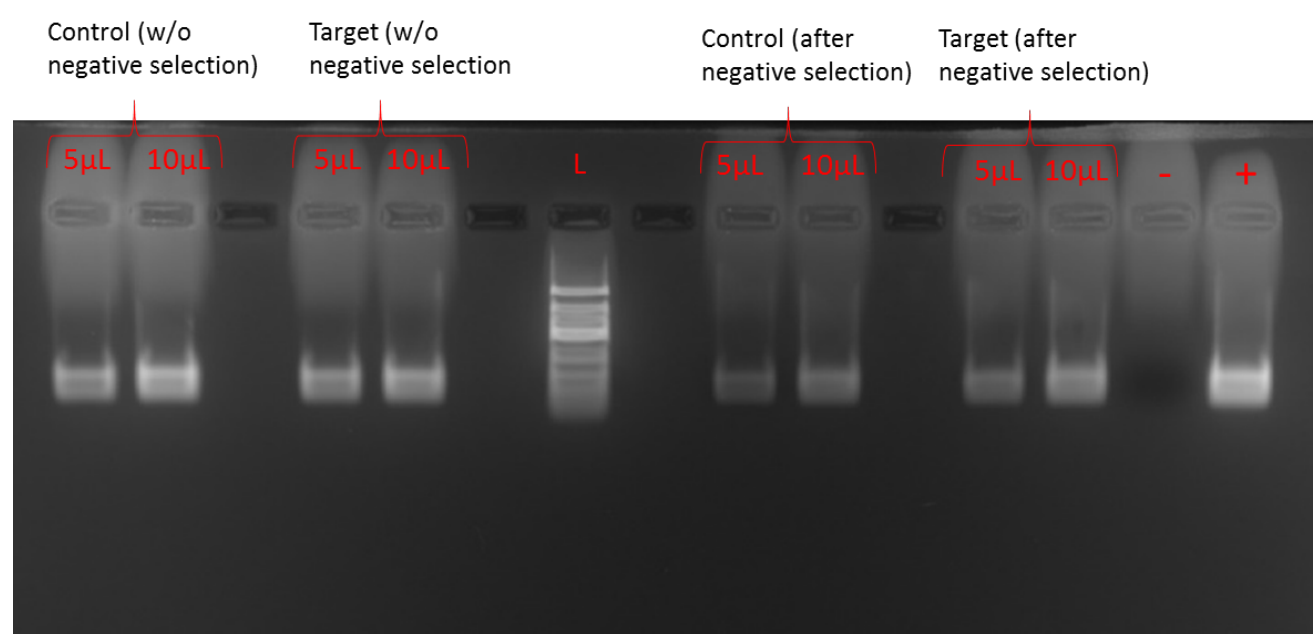


**Figure 5.5: SELEX Round 5 products electrophoresed on a 2.5% (w/v) agarose gel.**

### 5.3.3 Results of Selection Round 6 (Negative selection round)

In Selection Round Six, negative selection was introduced to remove the oligonucleotides in the library that showed affinity to the control bead matrix. As evidenced by the weaker bands

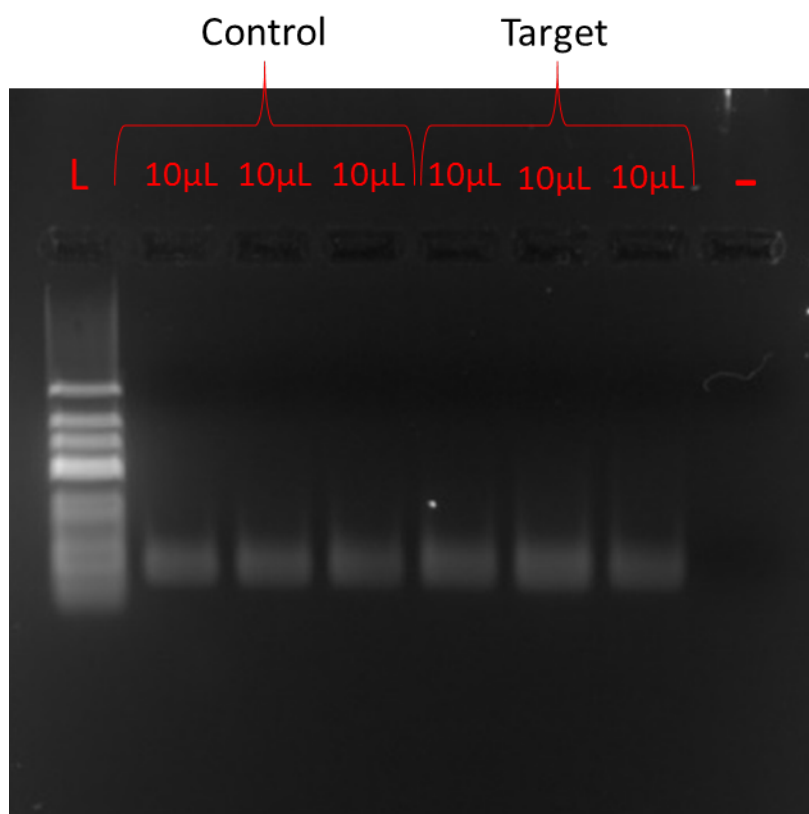
in the control lanes following negative selection, the negative selection step was successful in reducing the number of sequences retained on both the control and target beads (Figure 5.6). The negative selection did not completely remove the portion of the library which bound to the control beads so the process was repeated in later selection rounds.



**Figure 5.6: SELEX round 6 products electrophoresed on a 2.5% (w/v) agarose gel.**

#### 5.3.4 Results of Selection Round 7

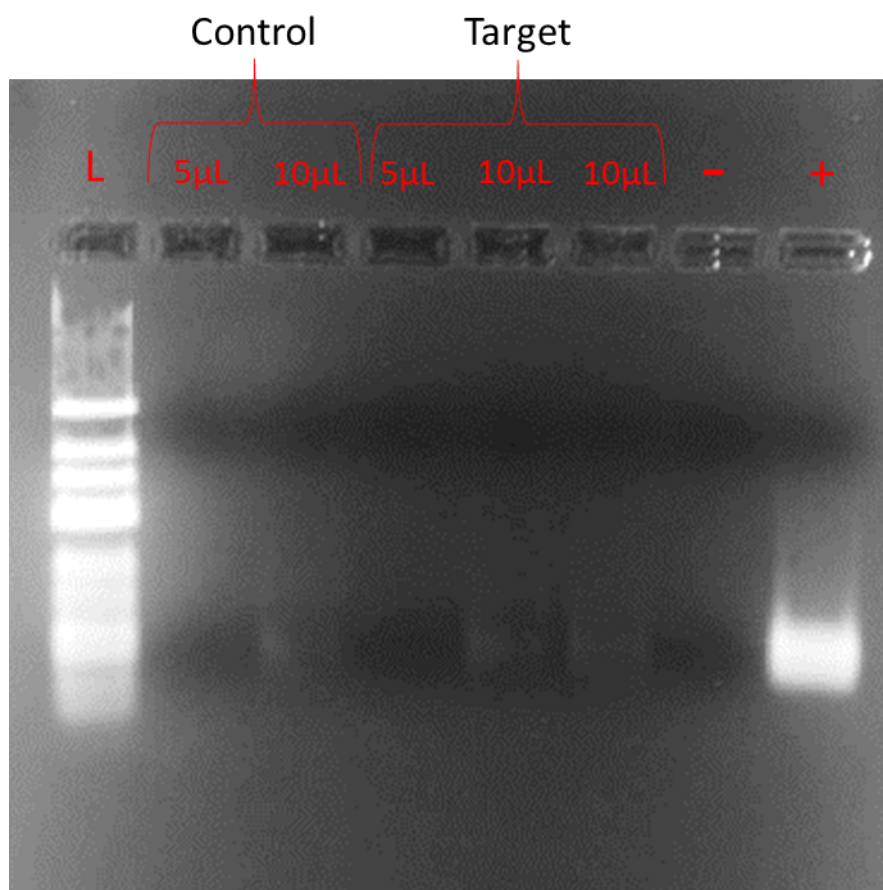
Selection Round Seven results demonstrated that the negative selection step did not remove all sequences that bind to the control matrix. The intensity of the bands from incubating oligonucleotides from SELEX Round 7 with the control beads was again equivalent to that with the nonylphenol-conjugated beads (Figure 5.7).



**Figure 5.7: SELEX round 7 products electrophoresed on a 2.5% (w/v) agarose gel.**

### 5.3.5 Results of Selection Round 8

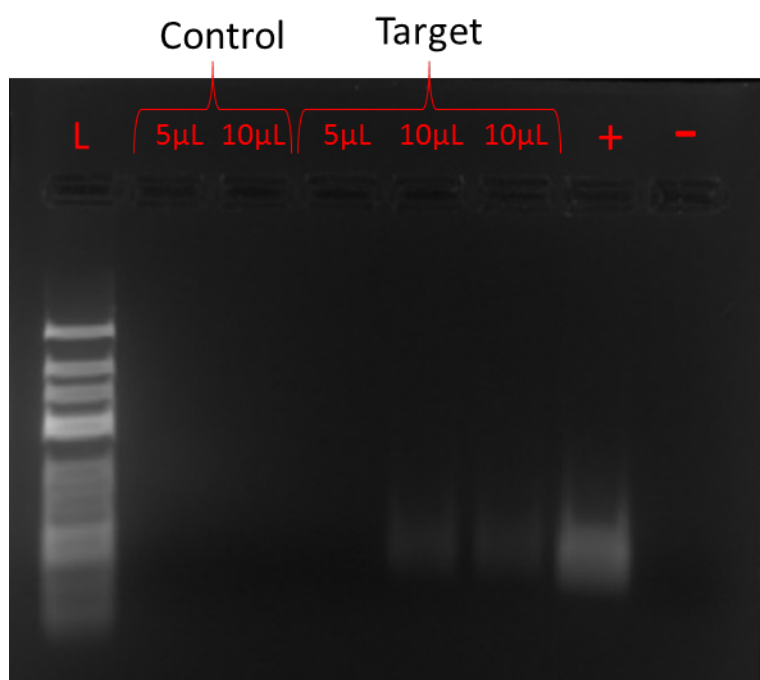
In an attempt to increase the selection pressure on the oligonucleotide library, an increased detergent concentration was added to the washing buffer for Selection Round Eight. An increased detergent concentration of 0.5% (v/v) IGEPAL almost completely removed the library from the beads (Figure 5.8) resulting in weak bands suggesting a low amount of DNA. Subsequent selection Rounds 9-16 were predominantly used to re-enrich the library (gel images not shown).



**Figure 5.8: SELEX round 8 products electrophoresed on a 2.5% (w/v) agarose gel.**

### 5.3.6 Results of Selection Round 16

After 16 selection rounds, a differential enrichment between the control and target-conjugated bead incubations was evident (Figure 5.9). The bands which represented the amplified product were still very weak in comparison with the bands which were produced by amplifying the library prior to using increased detergent in the wash step.

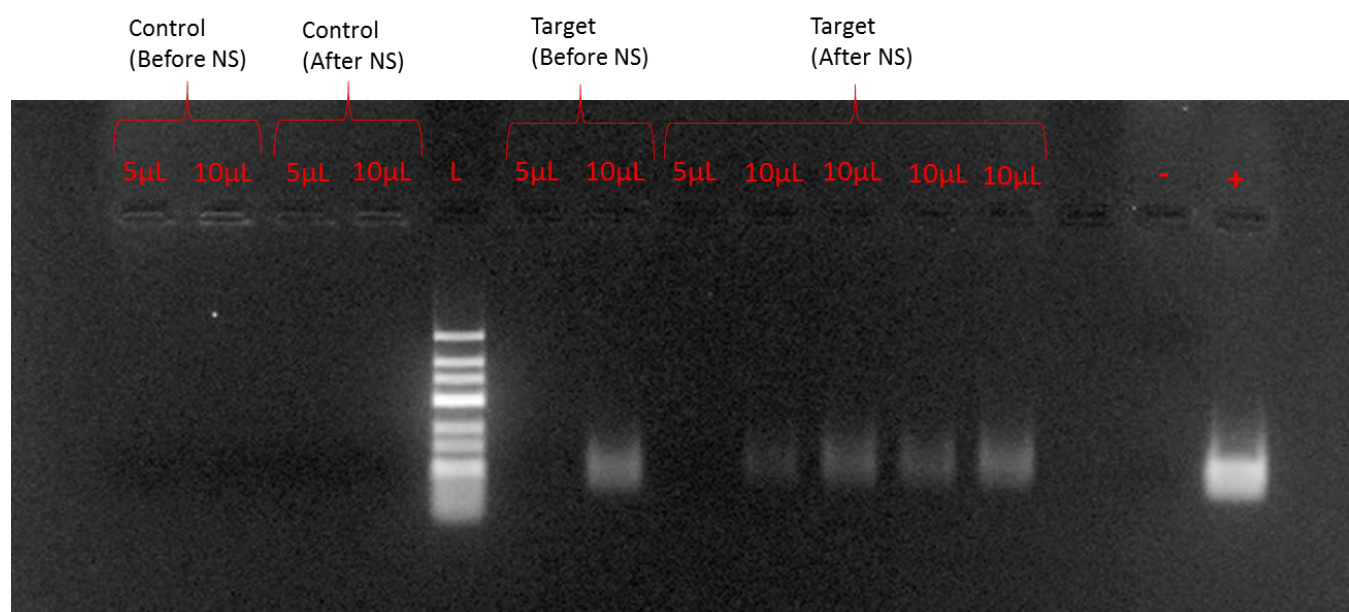


**Figure 5.9: SELEX round 16 products electrophoresed on a 2.5% (w/v) agarose gel.**

### 5.3.7 Results of Selection Round 19

A second round of negative selection was completed during Selection Round 19. No visible bands were produced from incubating the oligonucleotide library with the control beads and then amplifying the library (Figure 5.10). A reduction in the amount of DNA being amplified from the target-conjugated beads before and after negative selection can be seen, indicating that some non-specific oligonucleotides were removed from the library using negative selection. Multiple template aliquots from the target-conjugated bead incubation were amplified in order to have enough DNA for cloning and subsequent sequencing.





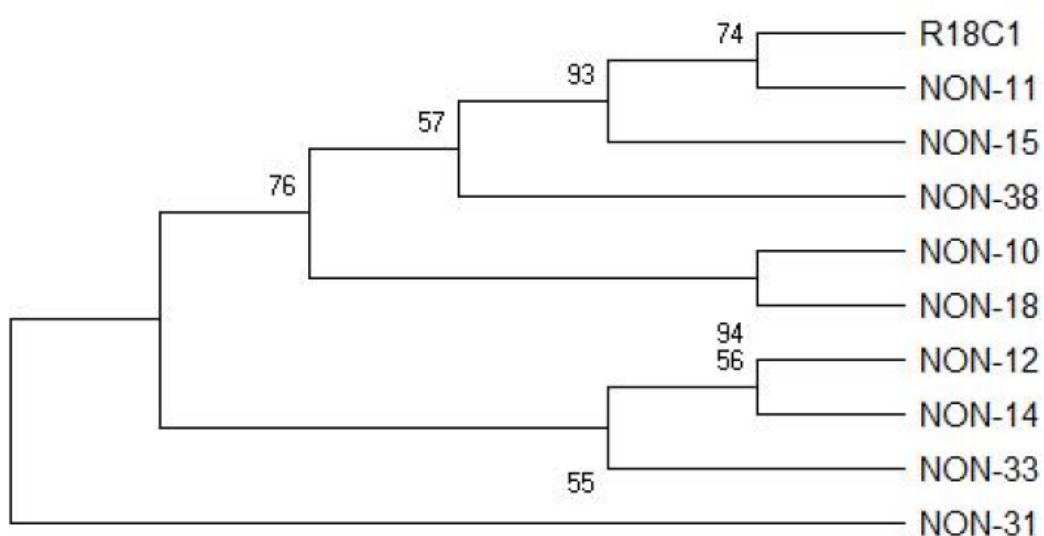
**Figure 5.10: SELEX round 19 products electrophoresed on a 2.5% (w/v) agarose gel.**

### 5.3.8 Results of Cloning and Sequencing

A total of 40 individual sequences were cloned into plasmid vectors and sent for Sanger sequencing. Of these 40 sequences, only 9 were of acceptable quality. When importing FASTA files into Geneious, the chromatogram produced by the sequencer for each base is binned into three categories: low quality, medium quality or high quality. Acceptable quality in this case was defined as a minimum of 60% of the bases being binned as high quality. All candidates were assessed in terms of their similarity to R18C1 (an oestradiol-binding aptamer which is frequently used in our laboratory) using a basic pairwise distance matrix (Figure 5.11).



A phylogenetic tree (Figure 5.13) showed that aptamer candidates NON-12, NON-14, NON-33 and NON-31 were the most divergent sequences when compared to R18C1. These four candidates plus NON-38 were selected for further characterisation. The sequences, delta Gibbs free energy and nucleotide composition of these candidates are listed in Table 5.3.



**Figure 5.13: Maximum likelihood phylogenetic tree showing the relationship between nonylphenol aptamer candidates and R18C1.**

**Table 5.3: Nonylphenol aptamer candidates selected for characterisation.**

Sequence Identifier	Sequence (5' to 3')	Length (nt)	$\Delta$ Gibbs Free Energy (kcal/mol)	Nucleotide Composition (N40 region)
NON-12	ATACGAGCTTGTTCAATACCCACAATGACGTTTGTATATGGTATATTCTTCGTTGTGGTGATAGTAAGAGCAATC	75	-5.62	A = 20.0% C = 17.5% G = 22.5% T = 40.0% GC = 40.0%
NON-38	ATACGAGCTTGTTCAATAGGGGGCCGTGCTGATGGGGCTGGTGGATTGTAGTGTGGGGTGATAGTAAGAGCAATC	75	-3.31	A = 21.3% C = 12.0% G = 40.0% T = 26.7% GC = 52.0%
NON-33	ATACGAGCTTGTTCAATATCCCTATCTATCATGTTCTTACGGACTATTTCTTTGGCGTGATAGTAAGAGCAATC	75	-3.31	A = 15.0% C = 27.5% G = 15.0% T = 42.5% GC = 42.5%
NON-31	ATACGAGCTTGTTCAATATTTCTAATGCTTTTTTATGTTCCGGCTACCATGGTCTGTGTGATAGTAAGAGCAATC	75	-2.02	A = 12.5% C = 20.0% G = 20.0% T = 47.5% GC = 40.0%
NON-14	ATACGAGCTTGTTCAATAGTGACGAGGTTTATATATTGGGTATAATGGGTCTTAGTTGGGTGATAGTAAGAGCAATC	77	-1.91	A = 21.4% C = 4.8% G = 35.7% T = 38.1% GC = 40.5%

\*primer sequences are highlight in red; \*G-quadruplex motif highlighted in yellow

### 5.3.9 Secondary Structure Characterisation of Nonylphenol Aptamer Candidates

Secondary structure predictions for each candidate were completed using the mFold web server. The most likely secondary structure for each candidate is displayed in Figures 5.14- 5.18. Aptamer Candidate NON-12 had the lowest delta Gibbs free energy (-5.62 kcal/mol) indicating the most stable secondary structure. Aptamer candidate NON-38 was predicted to

form a G-quadruplex structure with a G-score of 65 using the QGRS mapper tool and had a higher GC content (52%) than the other candidates.

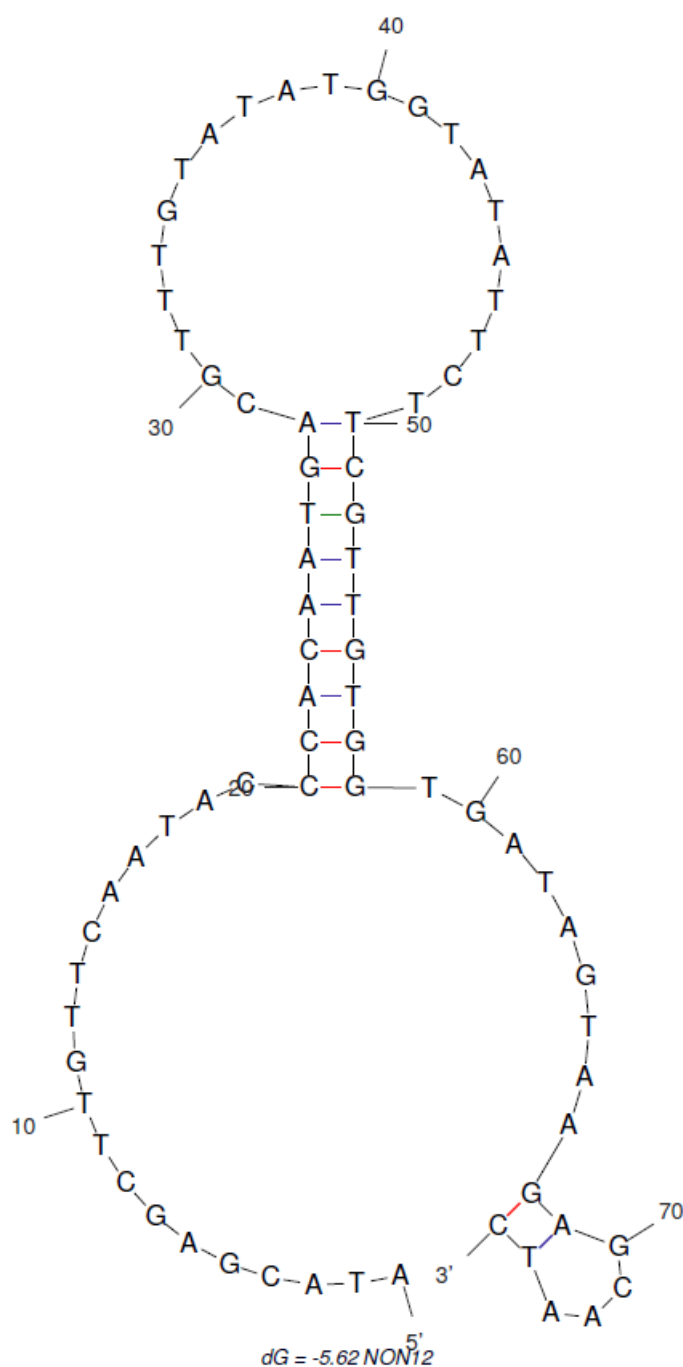


Figure 5.14: Predicted secondary structure of aptamer candidate NON-12.





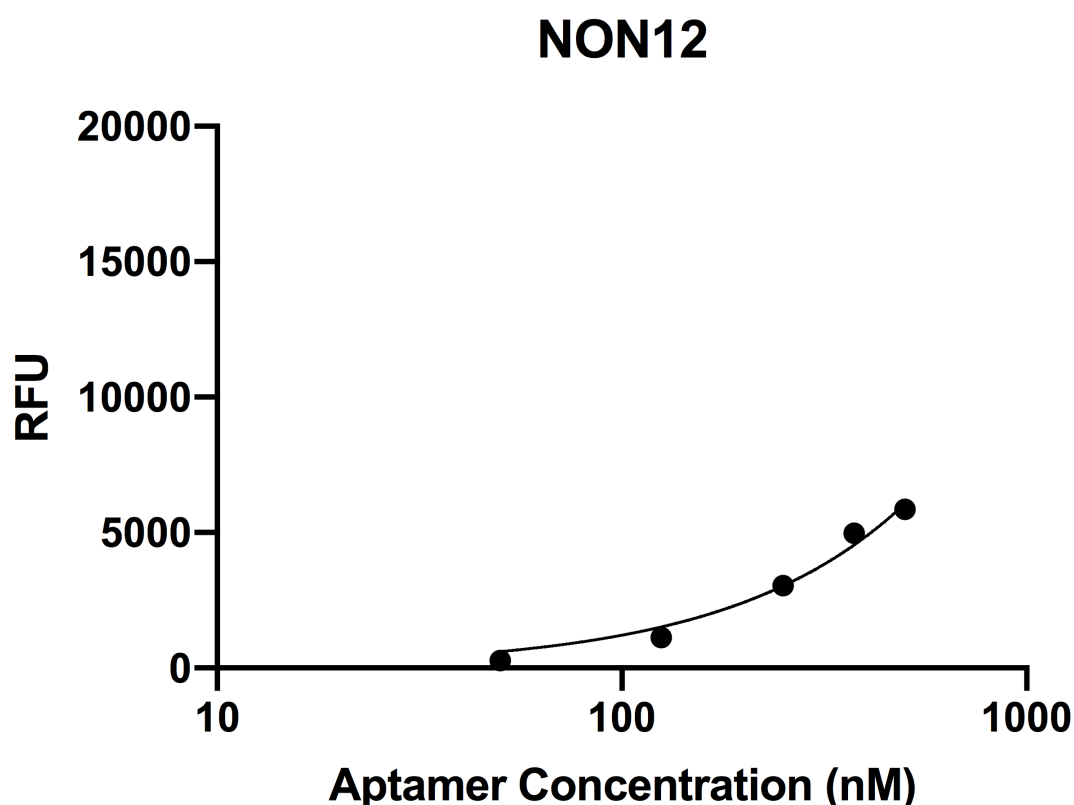






### 5.3.10 Fluorescent Binding Assay Characterisation of Nonylphenol Aptamer Candidates

All candidates were assessed for binding affinity using a fluorescent binding assay whereby a HEX label was added to each aptamer candidate. Different concentrations of HEX-labelled aptamer candidates were incubated with target-conjugated beads and the unbound aptamer was removed via washing. A  $K_d$  could not be calculated for any of the candidates from the binding curves generated using this assay (Figure 5.19-5.24). In addition, a random HEX-labelled sequence of the same length (75 nucleotides) was used as a control and displayed a similar result (see Figure 5.24).



**Figure 5.19: Binding curve of aptamer candidate NON-12 using the fluorescent binding assay.**

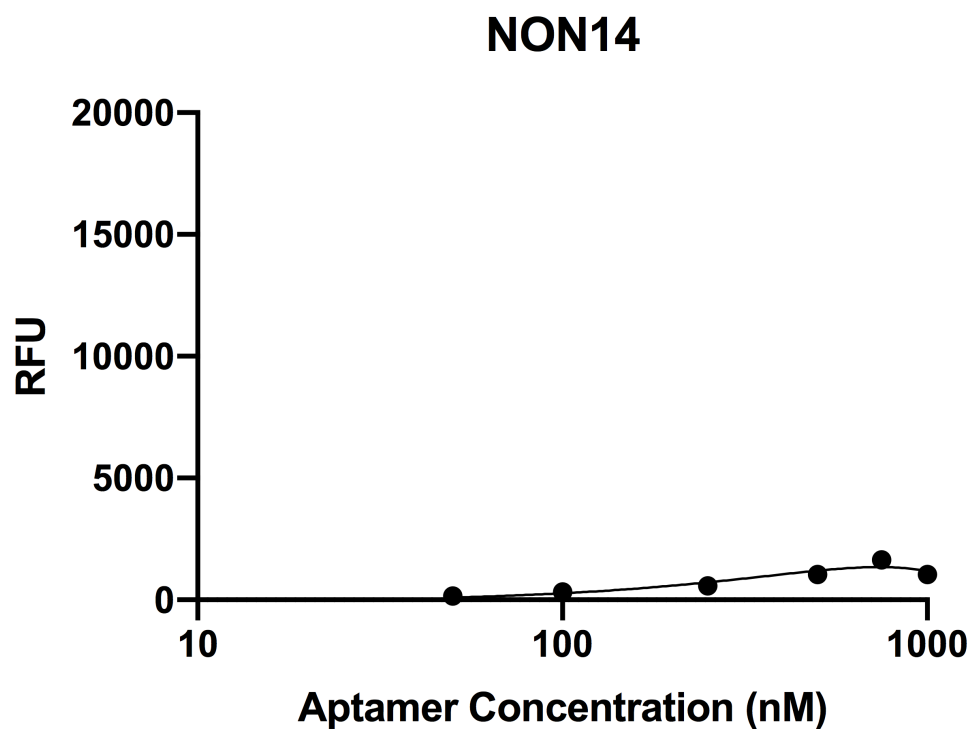


Figure 5.20: Binding curve of aptamer candidate NON-14 using the fluorescent binding assay.

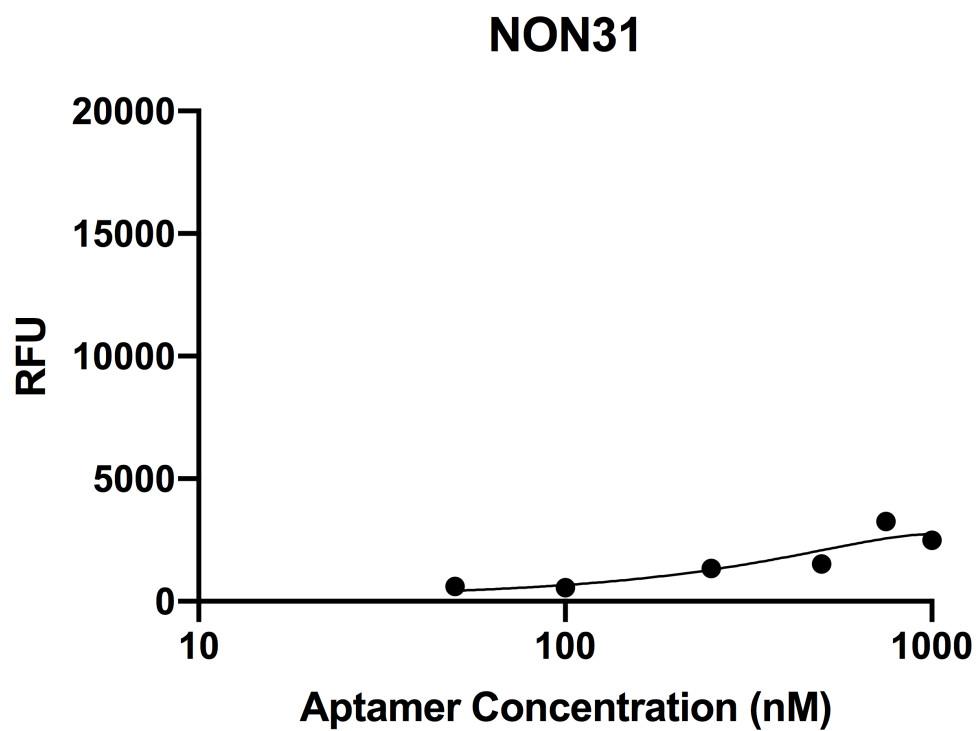


Figure 5.21: Binding curve of aptamer candidate NON-31 using the fluorescent binding assay.

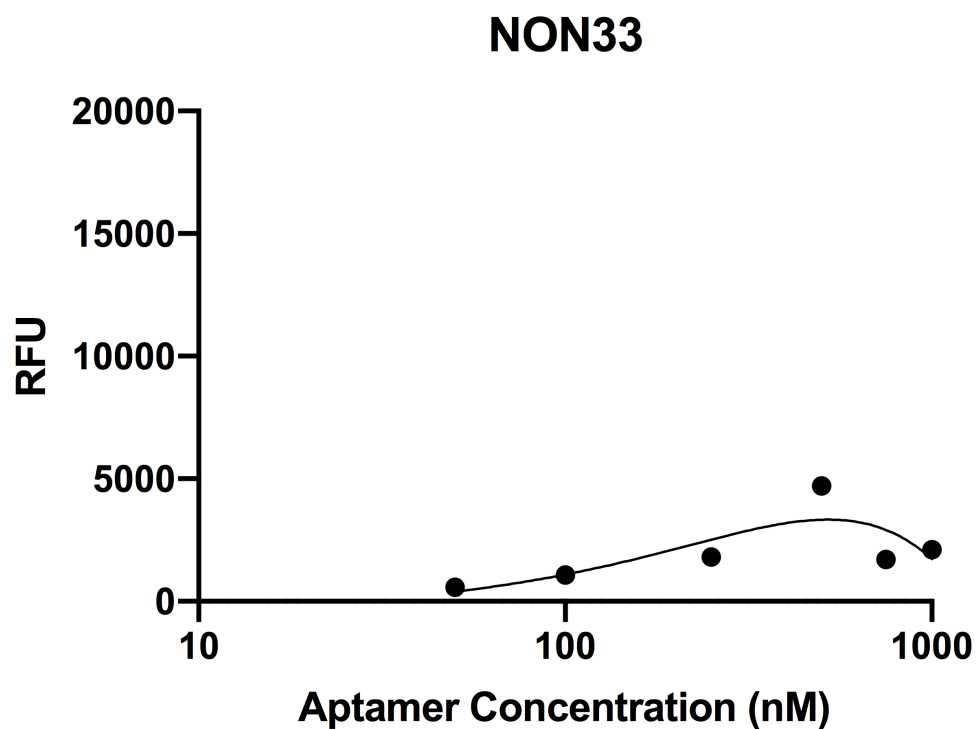


Figure 5.22: Binding curve of aptamer candidate NON-33 using the fluorescent binding assay.

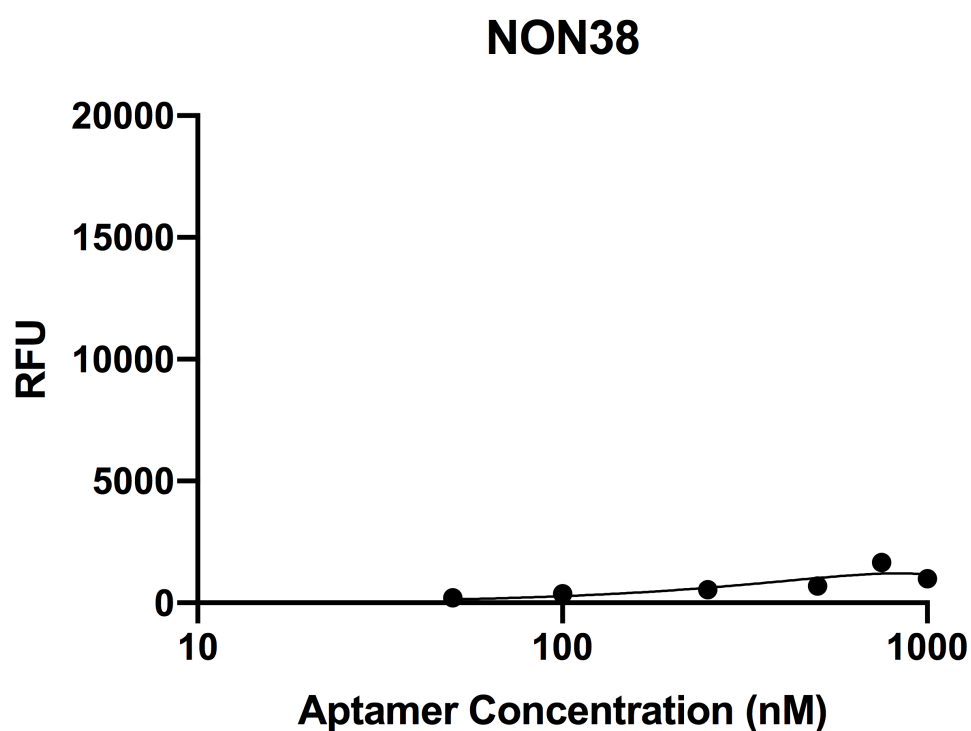


Figure 5.23: Binding curve of aptamer candidate NON-38 aptamer candidate using the fluorescent binding assay.

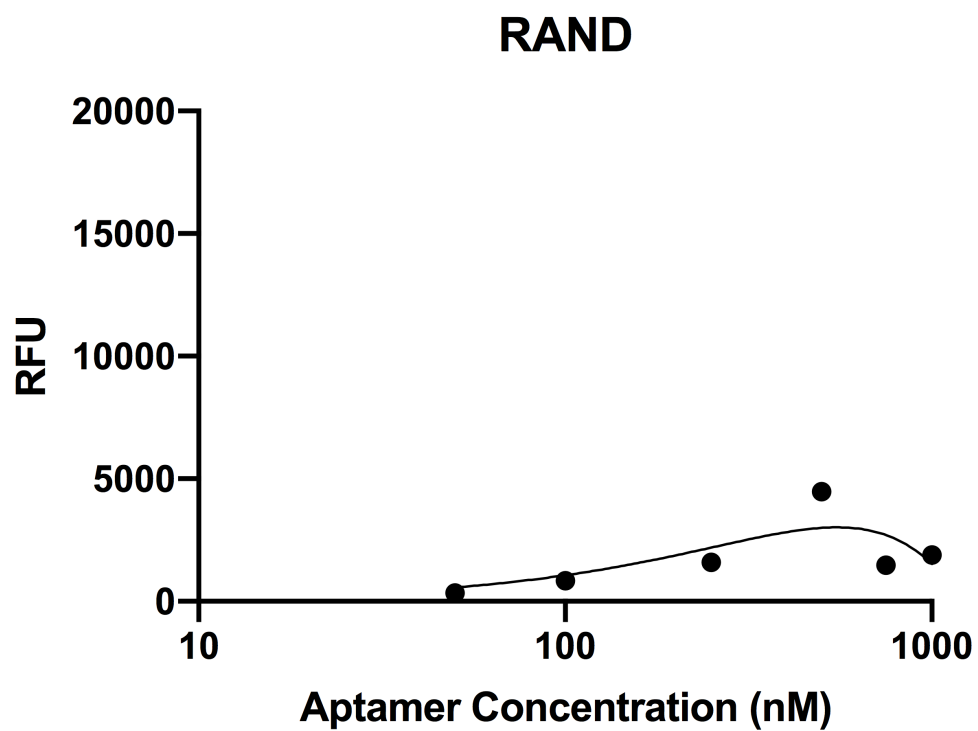


Figure 5.24: Binding curve of a random 75mer using the fluorescent binding assay.

### 5.3.11 Characterisation of Aptamer Candidate NON-12 using MST

MST characterisation of NON-12 showed no binding of the oligonucleotide with the target ligand (Figure 5.25). A  $K_d$  curve was not able to be fitted to the data.

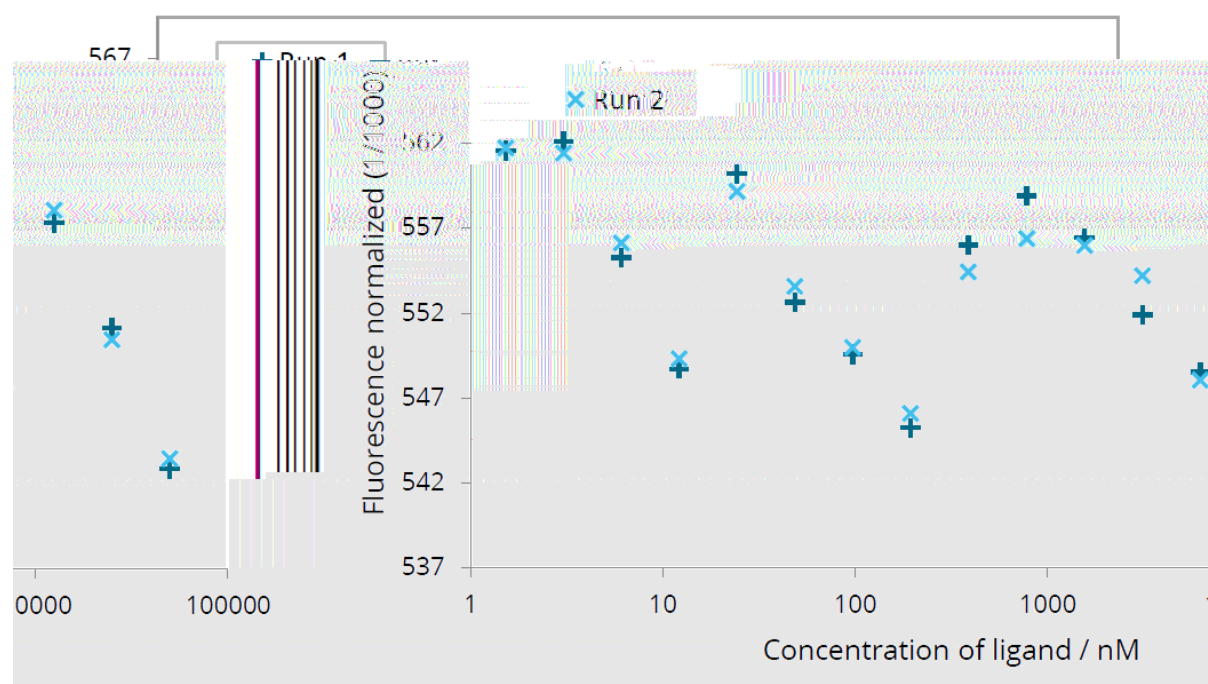


Figure 5.25: Results of MST characterisation for NON-12.

## 5.4 Discussion

In conclusion, aptamer candidates for technical nonylphenol equivalents were generated using affinity matrix SELEX. The binding affinity of these candidates to the target molecule could not be determined with the methods available in our laboratory. The top candidate, Aptamer NON-12 was chosen using in-silico analyses for further characterisation using MST. Unfortunately, it appeared to exhibit little binding affinity to nonylphenol using MST analyses.

There were numerous limitations to our approach for selecting an aptamer to nonylphenol. Upon reflection, a better design of our selection methodology may have resulted in superior aptamer candidates. For future experiments, several modifications could be tested. Selecting a single nonylphenol isomer (*e.g.* the most common isomer NP-04), instead of the mix of technical nonylphenol equivalents, will provide a focused selection effort and may increase the chance of generating a successful aptamer for a nonylphenol molecule. A capture-SELEX based methodology that ensures the aptamer can bind to the target molecule in-solution could be used such that the entirety of the molecule is available for target-aptamer interactions. The concentration of detergent in washing buffers or the selection environment could be limited to provide less stringent conditions. Similarly, a less-stringent counter selection could be used earlier in the SELEX process. Finally, using deep high-throughput sequencing to characterise the oligonucleotide library at every round of selection, or at the very least after key selection events, would provide information on more candidates.

It was evident from the gel images of the PCR amplification step for Selection Round Seven and Eight that there was a huge reduction in library size following the use of 0.5% (v/v) IGEPAL CA-630 in the wash buffer. IGEPAL CA-630 is a very similar molecule to nonylphenol and therefore it can be hypothesized that the majority of the library had some level of affinity to

IGEPAL CA-630, in addition to affinity for nonylphenol. Repeated rounds of positive selection were not enough to enrich the library to the same level as before the wash stringency was increased. It was assumed that even though a small amount of DNA was being amplified each round, the oligonucleotide pool would be free of sequences which bound to IGEPAL CA-630. Thus, if the library could be amplified to a point where it could be sequenced, then the aptamer candidates would be specific to nonylphenol.

The sequencing method used during this SELEX was low-throughput Sanger sequencing, therefore the sequence content of the library at each round was unable to be determined. As nonylphenol is a very small and potentially difficult molecule to select an aptamer to, using high-throughput (HT) SELEX would have provided a better chance to generate a successful aptamer candidate. By characterising the library at multiple rounds, we could have isolated the sequences that bound specifically to nonylphenol much earlier in the selection. Limiting the number of selection rounds has been proven to increase the chances of successfully isolating a high affinity candidate (Berezhnoy et al., 2012). This experiment involved 19 rounds of selection before candidates were cloned for sequencing. As the number of selection rounds increases, the chance of isolating oligonucleotides which are favourably amplified in PCR and have weak affinity is increased. It is possible that the sequences isolated during this selection experiment are not binders, but in fact favourably amplified sequences.

Although the two counter selection protocols were unsuccessful (Rounds 17 and 20), a wash step using IGEPAL CA-630 was effectively a counter-selection. As discussed above, any oligonucleotides which remained to the end of the selection should be specific enough to bind only to nonylphenols, and not octylphenols such as IGEPAL CA-630. In order to prove this



assumption, cross-reactivity testing to assess the binding of aptamer candidates to similarly structured molecules is recommended but was beyond the scope of this study.

In summary, the generation of an aptamer candidate that bound to nonylphenol was unsuccessful using this SELEX approach. However, by following the SELEX recommendations listed above, the likelihood of selecting a sequence that binds nonylphenol using a modified SELEX approach should be increased.

# Chapter 6: Generation of an Oxybenzone-binding aptamer using HT-SELEX

## 6.1 Background and Introduction

Oxybenzone ( $C_{14}H_{12}O_3$  ; (2-Hydroxy-4-methoxyphenyl)-phenylmethanone) is an EOC that is present at ever-increasing concentrations in marine and freshwater receiving environments. Detection of oxybenzone using traditional methods involve mass spectrometry and are time-consuming and expensive. In addition, this technology is not portable and cannot be used to measure concentrations in the field.

At the time of this study, there are no published aptamers that bind to oxybenzone. This may in part be due to the difficult nature of selecting aptamers to small molecules, and of selecting aptamers to molecules with limited functional groups. In the previous chapter describing the generation of candidate aptamers that bind to nonylphenol, the pitfalls associated with the generation of aptamers to these types of phenolic compounds were identified. Thus, some of the recommendations made in the discussion of Chapter 5 were applied herein. Specifically to utilise deep high-throughput (HT) sequencing to characterise the library at every selection round, and to compare seven different selection strategies along a stringency gradient to determine the best strategy for selecting aptamers against target molecules of this type. These selection strategies included using no or limited detergent concentrations in the

washing buffer, limiting the selection rounds to seven to limit the possibility of “parasitic” sequences that are favoured in PCR, and using a control library which was under no selection pressure as a comparison.

## 6.1.2 Introduction to High Throughput SELEX

### 6.1.2.1 What is High Throughput SELEX?

Fundamentally, HT-SELEX is methodologically the same as traditional SELEX except for the sequencing method used to characterise the oligonucleotide library. Sanger sequencing is used to analyse the sequence content of a small number of clones, and HT sequencing is used to analyse millions of sequences. A variety of HT sequencing technologies are available, including Illumina, Ion Torrent and Minion.

There are a range of advantages to analysing a larger sequence content. Preparing samples for HT sequencing is less labour intensive than traditional cloning methodologies, and the ability to sequence a larger population of oligos from every selection round allows researchers to analyse the effect of specific evolutionary processes during SELEX. This greatly improves the likelihood for the identification of aptamers with a high specificity for the target molecule. Analyses of the nucleotide library from every SELEX round allows for the identification of sequences which are enriched over the course of a SELEX experiment. Thus, sequences may be selected based on their enrichment (*i.e.* increase in frequency) rather than end frequency count. In addition, the identification of these enriched sequences can be detected much earlier resulting in less selection rounds being required to identify potential binders. By taking a much larger sample of the oligonucleotide library, such as millions of sequences as opposed to approximately one hundred, it is possible to detect enriched sequences at lower percentages (< 1%) of the library. Additionally, a higher number of sequences analysed also

allows for the identification of enriched primary and secondary sequence motifs with more robust statistical analyses. Increased statistical robustness enables researchers to compare different selection parameters and determine their effects on the resultant oligonucleotide library (Nguyen Quang et al., 2016).

The myriad of benefits of using HT-SELEX to select aptamers do not come without challenges. Datasets containing millions of sequences are computationally intensive to analyse and require specialised bioinformatic skills. There are also few programmes available to analyse these sorts of datasets.

#### *6.1.2.2 Aptamer Selection analysis using HT-SELEX*

The use of HT-SELEX to study evolutionary patterns within oligonucleotide libraries which are subjected to *in-vitro* selection has gained more prominence in recent years as HT sequencing has become more affordable. Schütze et al. (2011) compared the use of cloning with Sanger sequencing to HT sequencing when selecting a DNA aptamer to streptavidin. One of the central findings from this study was that the highest frequency sequences in the final selection round did not have the highest affinity. It was discovered that high affinity binders began to enrich as early as the first and second selection rounds, and that the frequency of some binding sequences decreased in later selection rounds. As mentioned previously, the chance of sequences that are favourably amplified in PCR and dilute “true” binders increases with the number of selection rounds.

Hoinka et al. (2015) expanded on this point through use of two algorithms to analyse sequencing data from five rounds of selection against the interleukin-10 receptor  $\alpha$  chain. AptaCluster and AptaMut were jointly used to analyse the sequencing data. By comparing the size of clusters from the fifth and final selection round together with binding experiments, the

study confirmed that the most abundant sequences were not the best binders (Hoinka et al., 2015). An experimental proof-of-concept for the hypothesis that cycle-to-cycle enrichment is a better predictor of binding propensity was demonstrated by showing that in general, sequences from clusters with higher enrichment had lower dissociation constants (Hoinka et al., 2015).

#### *6.1.2.3 Tools for Analysing HT-SELEX Data*

There are few tools available for analysing HT-SELEX data. Of those that are available, some are freely available (FASTAptamer, AptaSuite, APTANI and MPBind) whilst others must be purchased (PATTERNITY-seq, COMPAS). The two most comprehensive and readily useable software tools, and thus the two that are used herein, are FASTAptamer and AptaSUITE.

FASTAptamer is an open-source tool written by the Burke lab and has a variety of useful functions (Alam et al., 2015). The primary input for FASTAptamer is a FASTQ file. FASTAptamer is divided into modular functions that each perform a primary task. The first function, *count*, is able to count the number of times a sequence appears in a given population, normalise the count based on the population size and then rank all the sequences in a population based on abundance. This is the simplest form of analysis and a pre-requisite for the other tools. The *cluster* function groups sequences into closely related families based on sequence similarities. This is achieved using a user-specific Levenshtein edit distance, which is the total number of insertions, deletions, or substitutions required to transform one character string into another (Alam et al., 2015). The advantage of the Levenshtein distance algorithm over the alternate Hamming distance employed in other analysis software is that the character string (*i.e.* nucleotide sequences) may be dissimilar lengths (Alam et al., 2015). Insertions or deletions are common in SELEX, especially when low-fidelity amplification is used to create variants

within the selection process, therefore being able to cluster sequences of unequal length is extremely useful.

The *enrich* function calculates sequence enrichment over two or three sequential selection rounds. This function calculates the fold-enrichment of unique sequences by locating those which are present in both input files (selection rounds) and then dividing the normalized read count of the sequence in one round by the normalized read count in the second round. By assessing enrichment as opposed to just abundance, it is possible to identify high affinity binders and eliminate sequences which may just be favoured in amplification reactions.

Two other functions present within the FASTAptamer tool kit are the *search* and *compare* functions. The *search* command allows for the searching of known binding motifs within sequence datasets and the *compare* function enables the user to generate snap-shots of how the genotypic frequencies of sequences in one population compares to that of another population (Alam et al., 2015) .

FASTAptamer is a simple collection of scripts written in Perl which are easy to use with a basic knowledge of UNIX command-lines. The output of each script provides valuable information for the analysis of HT-SELEX data and for identifying high affinity aptamer candidates within these datasets.

In comparison, AptaSUITE is a comprehensive software suite for managing and analysing aptamer data which is written in the Java programming language. It integrates a number of individual programmes written and published for HT-SELEX data analysis previously, including AptaCLUSTER, AptaPLEX, AptaTRACE, AptaMUT and AptaSIM (Hoinka et al., 2018). AptaPLEX is a de-multiplexer designed specifically for HT-SELEX data. AptaCLUSTER uses the Hamming distance to cluster sequences into families and derive a consensus sequence for each cluster.

Hamming distance is a simple metric used to compare two strings of equal length for differences in their corresponding symbols (Hamming, 1950). Hamming distance is restricted to analysing sequences of equal length and if insertions or deletions are possible then a more sophisticated algorithm such as Levenshtein distance should be employed. The Hamming distance clustering algorithm employed by AptaCLUSTER relies on the random region of the library being a constant size which is a disadvantage however it claims to outperform “traditional clustering algorithms” (Hoinka et al., 2018). AptaMUT allows the user to perform analyses of the mutational landscape within a cluster to differentiate mutations that are favourable for target binding from those which are detrimental. This is achieved by assessing enrichment of mutants over multiple selection cycles (Hoinka et al., 2018). AptaTRACE is used to identify sequence-structure patterns (*i.e.* motifs) that are beneficial for target binding. It should be noted that only secondary structure prediction for RNA is available within the framework. AptaSIM is designed to provide a realistic recreation of the SELEX process *in-silico*. AptaSUITE has the option to use both the command line if the dataset size is large and requires the use of a high-performance computer or a graphical user interface for small to medium size datasets. The ability to use a graphical user interface introduces the possibility of dynamic data visualization and is easier to use for researchers who are not familiar with the UNIX command-line, a pre-requisite for most bioinformatic analyses.

### 6.1.3 Aims of the study

As discussed above, HT-SELEX provides the analytical potential to assess the evolution of an oligonucleotide library under different selection pressures. Although the number of tools available to analyse HT sequence data for SELEX experiments is limited, the two tools described above were considered adequate to perform this proposed study. It is assumed that methodological changes during SELEX may have significant effects on the evolution of the library.

The specific aims of this study were to:

- (1) generate aptamer candidates for oxybenzone;
- (2) develop a suitable bioinformatics pipeline for comparing multiple selection strategies across multiple selection rounds;
- (3) assess the effects of a variety of selection pressures implemented during SELEX on an oligonucleotide library using bioinformatic tools;
- (4) determine the strategy that resulted in the highest average cycle-to-cycle enrichment rates as a good predictor of binding affinity, and;
- (5) characterise the binding affinity of the lead aptamer candidate from each selection strategy using MicroScale Thermophoresis (MST).

My hypothesis was that strategies with increased stringency would result in higher average enrichment rates during SELEX, indicating that sequences selected under these strategies may have improved affinity for the target molecule.



## 6.2 Materials and methods

### 6.2.1 Affinity Matrix Preparation

Oxybenzone was conjugated to an affinity matrix as previously described (see Section 5.2.1), with the modification that a higher concentration of oxybenzone was used, in comparison to that for nonylphenol. A 15 mL volume of 40 mM oxybenzone (analytical standard, Sigma Aldrich) was prepared in 1 M NaOH. A control matrix conjugated to ethanolamine was also prepared as described in Section 5.2.1.

A Shimadzu UV-Visible spectrophotometer was used to confirm the conjugation of oxybenzone onto the affinity matrix. The affinity matrix was loaded into a quartz cuvette (Starna, Essex, United Kingdom) using a 1000  $\mu$ L micropipette. The pipette tip was cut at the end to increase the bore size and prevent beads from clogging the tip. The UV-Visible spectrum of both the control and oxybenzone-conjugated matrix was measured at wavelengths from 200 to 800 nm. The presence of absorption peaks at 240 and 280 nm would indicate the successful conjugation of oxybenzone, and the absence of these peaks would confirm the control matrix.

### 6.2.3 Description of SELEX Strategies

Seven selection strategies were designed to determine the optimal strategy for selecting high affinity aptamers to oxybenzone. Each selection strategy had one methodological change that increased selection pressure from the previous strategy. The conditions for all of the different selection strategies that were carried out over seven selection rounds are outlined in Table 6.1. The first selection strategy (Standard, Std) did not include mutagenic PCR at any selection round and used no detergent in the BWB and a consistent wash volume of 3 mL. Strategy Two

(Standard + Mutation) used mutagenic PCR in two of the selection rounds (Rounds 2 and 5). Strategy Three (Extra Volume) introduced an increased washing stringency in the form of a linear increase in the wash volume in each selection round (from 3 mL in Round 1 to 15 mL in Round 7). Strategy Four (Extra Detergent) also had increased stringency but in the form of adding a non-ionic surfactant to the BWB. Strategy Five (Room Temperature) was identical to Strategy Four with the exception that the incubation was carried out over four hours at room temperature, as opposed to overnight at 4°C. Strategy Six (Negative Selection) included the use of negative selection at Rounds 3 and 6, whilst Strategy Seven (Counter Selection) included this and the addition of a counter selection step in the final round.

**Table 6.1: The different selection strategies applied to each selection round.**

<b>SELEX Round 1</b>							
<b>SELEX Strategy Number</b>	<b>Strategy Name</b>	<b>Strategy Description</b>	<b>Mutation</b>	<b>Washing</b>	<b>Incubation</b>	<b>Negative Selection</b>	<b>Counter Selection</b>
1	Standard (S)	Standard	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
2	Standard+ Mutation (SM)	Standard with mutation	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
3	Extra Volume (EV)	Mutation and increased wash volume	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
4	Extra Detergent (ED)	Mutation and detergent usage	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
5	Room Temperature (RT)	Mutation, detergent usage and short RT incubation	No	3 mL 1X BWB (No Detergent)	4 hours at RT	No	No
6	Negative Selection (NS)	Mutation, detergent usage and negative selection	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
7	Counter Selection (CS)	Mutation, detergent usage, negative selection and counter-selection	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No

SELEX Round 2							
SELEX Strategy Number	Strategy Name	Strategy Description	Mutation	Washing	Incubation	Negative Selection	Counter Selection
1	Standard (S)	Standard	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
2	Standard+ Mutation (SM)	Standard with mutation	Yes - 0.7 $\mu$ L of 25 mM MgCl <sub>2</sub> per rxn	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
3	Extra Volume (EV)	Mutation and increased wash volume	Yes - 0.7 $\mu$ L of 25 mM MgCl <sub>2</sub> per rxn	5 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
4	Extra Detergent (ED)	Mutation and detergent usage	Yes - 0.7 $\mu$ L of 25 mM MgCl <sub>2</sub> per rxn	1 mL 1X BWB (0.1% IGEPAL) 2 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
5	Room Temperature (RT)	Mutation, detergent usage and short RT incubation	Yes - 0.7 $\mu$ L of 25 mM MgCl <sub>2</sub> per rxn	1 mL 1X BWB (0.1% IGEPAL) 2 mL 1X BWB (No Detergent)	4 hours at RT	No	No
6	Negative Selection (NS)	Mutation, detergent usage and negative selection	Yes - 0.7 $\mu$ L of 25 mM MgCl <sub>2</sub> per rxn	1 mL 1X BWB (0.1% IGEPAL) 2 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
7	Counter Selection (CS)	Mutation, detergent usage, negative selection and counter-selection	Yes - 0.7 $\mu$ L of 25 mM MgCl <sub>2</sub> per rxn	1 mL 1X BWB (0.1% IGEPAL) 2 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No

SELEX Round 3							
SELEX Strategy Number	Strategy Name	Strategy Description	Mutation	Washing	Incubation	Negative Selection	Counter Selection
1	Standard (S)	Standard	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
2	Standard+ Mutation (SM)	Standard with mutation	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
3	Extra Volume (EV)	Mutation and increased wash volume	No	7 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
4	Extra Detergent (ED)	Mutation and detergent usage	No	2 mL 1X BWB (0.1% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
5	Room Temperature (RT)	Mutation, detergent usage and short RT incubation	No	2 mL 1X BWB (0.1% IGEPAL) 1 mL 1X BWB (No Detergent)	4 hours at RT	No	No
6	Negative Selection (NS)	Mutation, detergent usage and negative selection	No	2 mL 1X BWB (0.1% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	Yes	No
7	Counter Selection (CS)	Mutation, detergent usage, negative selection and counter-selection	No	2 mL 1X BWB (0.1% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	Yes	No

SELEX Round 4							
SELEX Strategy Number	Strategy Name	Strategy Description	Mutation	Washing	Incubation	Negative Selection	Counter Selection
1	Standard (S)	Standard	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
2	Standard+ Mutation (SM)	Standard with mutation	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
3	Extra Volume (EV)	Mutation and increased wash volume	No	9 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
4	Extra Detergent (ED)	Mutation and detergent usage	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
5	Room Temperature (RT)	Mutation, detergent usage and short RT incubation	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	4 hours at RT	No	No
6	Negative Selection (NS)	Mutation, detergent usage and negative selection	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
7	Counter Selection (CS)	Mutation, detergent usage, negative selection and counter-selection	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No

SELEX Round 5							
SELEX Strategy Number	Strategy Name	Strategy Description	Mutation	Washing	Incubation	Negative Selection	Counter Selection
1	Standard (S)	Standard	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
2	Standard+ Mutation (SM)	Standard with mutation	Yes - 0.5 µL of 25 mM MgCl <sub>2</sub> per rxn	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
3	Extra Volume (EV)	Mutation and increased wash volume	Yes - 0.5 µL of 25 mM MgCl <sub>2</sub> per rxn	11 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
4	Extra Detergent (ED)	Mutation and detergent usage	Yes - 0.5 µL of 25 mM MgCl <sub>2</sub> per rxn	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
5	Room Temperature (RT)	Mutation, detergent usage and short RT incubation	Yes - 0.5 µL of 25 mM MgCl <sub>2</sub> per rxn	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	4 hours at RT	No	No
6	Negative Selection (NS)	Mutation, detergent usage and negative selection	Yes - 0.5 µL of 25 mM MgCl <sub>2</sub> per rxn	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
7	Counter Selection (CS)	Mutation, detergent usage, negative selection and counter-selection	Yes - 0.5 µL of 25 mM MgCl <sub>2</sub> per rxn	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No

SELEX Round 6							
SELEX Strategy Number	Strategy Name	Strategy Description	Mutation	Washing	Incubation	Negative Selection	Counter Selection
1	Standard (S)	Standard	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
2	Standard+ Mutation (SM)	Standard with mutation	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
3	Extra Volume (EV)	Mutation and increased wash volume	No	13 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
4	Extra Detergent (ED)	Mutation and detergent usage	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
5	Room Temperature (RT)	Mutation, detergent usage and short RT incubation	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	4 hours at RT	No	No
6	Negative Selection (NS)	Mutation, detergent usage and negative selection	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	Yes	No
7	Counter Selection (CS)	Mutation, detergent usage, negative selection and counter-selection	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	Yes	No

SELEX Round 7							
SELEX Strategy Number	Strategy Name	Strategy Description	Mutation	Washing	Incubation	Negative Selection	Counter Selection
1	Standard (S)	Standard	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
2	Standard+ Mutation (SM)	Standard with mutation	No	3 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
3	Extra Volume (EV)	Mutation and increased wash volume	No	15 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
4	Extra Detergent (ED)	Mutation and detergent usage	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
5	Room Temperature (RT)	Mutation, detergent usage and short RT incubation	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	4 hours at RT	No	No
6	Negative Selection (NS)	Mutation, detergent usage and negative selection	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	No
7	Counter Selection (CS)	Mutation, detergent usage, negative selection and counter-selection	No	2 mL 1X BWB (0.3% IGEPAL) 1 mL 1X BWB (No Detergent)	O/N @ 4 degrees	No	Yes

#### 6.2.4 Controls

Four control libraries were used to gauge the enrichment of “parasitic” sequences that favourably amplified without any selection pressures. For this purpose, two monoclonal and two polyclonal libraries were used. The monoclonal library consisted of an aptamer that had been synthesized previously in our lab. The polyclonal library was the same N40 library used for the selection experiments. For the first PCR, 10 µL of 20 nM polyclonal or monoclonal library was added as the template in a 50 µL reaction volume. The cycling conditions and reaction composition was similar to that specified in Section 2.3. The number of cycles and



input concentration was varied each round to prevent non-specific amplification of PCR product that would result in a large smear on the gel. The conditions for each round are presented in Table 6.2. The dilutions and number of cycles were optimised for each individual round and each condition (data not shown).

**Table 6.2: Amplification conditions for each of the control libraries.**

<b>Round 1</b>	<b>Mono – (M-)</b>	<b>Mono + (M+)</b>	<b>Poly – (P-)</b>	<b>Poly + (P+)</b>
Input	10 µL of 20 nM	10 µL of 20 nM	10 µL of 20 nM	10 µL of 20 nM
PCR Cycles	35	35	35	35
Mutation	No	No	No	No
<b>Round 2</b>	<b>Mono – (M-)</b>	<b>Mono + (M+)</b>	<b>Poly – (P-)</b>	<b>Poly + (P+)</b>
Input	10 µL of gel extracted product from R1	10 µL of gel extracted product from R1	10 µL of gel extracted product from R1	10 µL of gel extracted product from R1
PCR Cycles	35	35	35	35
Mutation	No	Yes	No	Yes
<b>Round 3</b>	<b>Mono – (M-)</b>	<b>Mono + (M+)</b>	<b>Poly – (P-)</b>	<b>Poly + (P+)</b>
Input	1 µL of gel extracted product from R2	1 µL of gel extracted product from R2	1 µL of gel extracted product from R2	1 µL of gel extracted product from R2
PCR Cycles	30	30	30	30
Mutation	No	No	No	No
<b>Round 4</b>	<b>Mono – (M-)</b>	<b>Mono + (M+)</b>	<b>Poly – (P-)</b>	<b>Poly + (P+)</b>
Input	1 µL of 0.1 ng/µL of gel extracted product from R3	1 µL of 0.1 ng/µL of gel extracted product from R3	1 µL of 0.1 ng/µL of gel extracted product from R3	1 µL of 0.1 ng/µL of gel extracted product from R3
PCR Cycles	22	22	22	22
Mutation	No	No	No	No
<b>Round 5</b>	<b>Mono – (M-)</b>	<b>Mono + (M+)</b>	<b>Poly – (P-)</b>	<b>Poly + (P+)</b>
Input	1 µL of 0.1 ng/µL of gel extracted product from R4	1 µL of 0.1 ng/µL of gel extracted product from R4	1 µL of 0.1 ng/µL of gel extracted product from R4	1 µL of gel extracted product from R4, diluted 1 in 100
PCR Cycles	22	22	22	22
Mutation	No	Yes	No	Yes

<b>Round 6</b>	<b>Mono – (M-)</b>	<b>Mono + (M+)</b>	<b>Poly – (P-)</b>	<b>Poly + (P+)</b>
Input	10 µL of gel extracted product from R5, diluted 1/100	10 µL of gel extracted product from R5, diluted 1/100	10 µL of gel extracted product from R5, diluted 1/100	10 µL of gel extracted product from R5, diluted 1/100
PCR Cycles	22	22	22	22
Mutation	No	No	No	No
<b>Round 7</b>	<b>Mono – (M-)</b>	<b>Mono + (M+)</b>	<b>Poly – (P-)</b>	<b>Poly + (P+)</b>
Input	1 µL of gel extracted product from R6, diluted 1/10	1 µL of gel extracted product from R6, diluted 1/10	1 µL of gel extracted product from R6, diluted 1/10	1 µL of gel extracted product from R6, diluted 1/10
PCR Cycles	22	22	22	22
Mutation	No	No	No	No

### 6.2.5 General SELEX Methods

The selection methodology including washing the affinity matrix, PCR amplification of bound oligonucleotides, gel extraction, strand separation of dsDNA PCR product and the incubation procedure are specified in Section 2.1-2.6, with the modifications listed in Table 6.1.

### 6.2.6 Library Preparation for Next Generation Sequencing

Initial tests using gel electrophoresis followed by gel extraction using the Qiagen Min Elute kit for purification of DNA demonstrated that the yield, in terms of DNA quantity, would not meet the requirements of our sequencing providers (data not shown). The minimum requirement was 10 ng of total DNA to complete library preparation and sequencing. Therefore, another method was developed to increase the yield of DNA. During each selection round, two 10 µL aliquots of template from each bead incubation were amplified via PCR to yield 100 µL of PCR product per bead incubation. Specifically, 100 µL of product for both the control and target bead incubations. Subsequently, 25 µL of this product was loaded into the gel to visualize the library and 75 µL was stored at -80°C until required. Upon completion of all seven selection

rounds, the 75  $\mu$ L aliquots were purified using the Monarch PCR and DNA Clean-up Kit (New England Biolabs, Ipswich, Massachusetts, United States). The manufacturer's guidelines were followed and an elution volume of 33  $\mu$ L was used to concentrate the DNA but ensure sufficient volume as specified by sequencing supplier's guidelines. This purification was completed for each round of the seven libraries in which each of the seven selection strategies (49 samples) and each of the four control strategies (28 samples) were tested, as well as all seven aliquots of starting library (7 samples).

As the aliquots from the starting library were ssDNA, for sequencing purposes they were converted to dsDNA by PCR amplification. Specifically, 1  $\mu$ L of the starting library was amplified using 5 cycles of PCR and the PCR product was concentrated using the NEB Monarch PCR Clean-up kit described above. All 84 samples were quantified using a Qubit 2.0 and a Qubit high sensitivity DNA kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) as per the manufacturer's instructions. The quantity of DNA in all 84 samples is shown in Appendix E.

All 84 samples were prepared for the Illumina NextSeq platform by Gen X Pro (Frankfurt, Germany) using proprietary methods. Generally, template DNA was phosphorylated and adapters were ligated to the DNA. The adapters were designed by the sequencing provider and contained an integrated 8 nucleotide barcode plus an 8 nucleotide unique molecular identifier in both index reads.

Quantitative PCR was performed to optimise the minimal cycle number for sufficient library amplification. Following amplification, the PCR product was purified using SPRI (solid phase reversible immobilization) beads to remove the PCR primers and quantified using Qubit.

Quality control was performed using LabChip assays prior to sequencing using Illumina NextSeq500 and a v2 75 cycle sequencing kit.

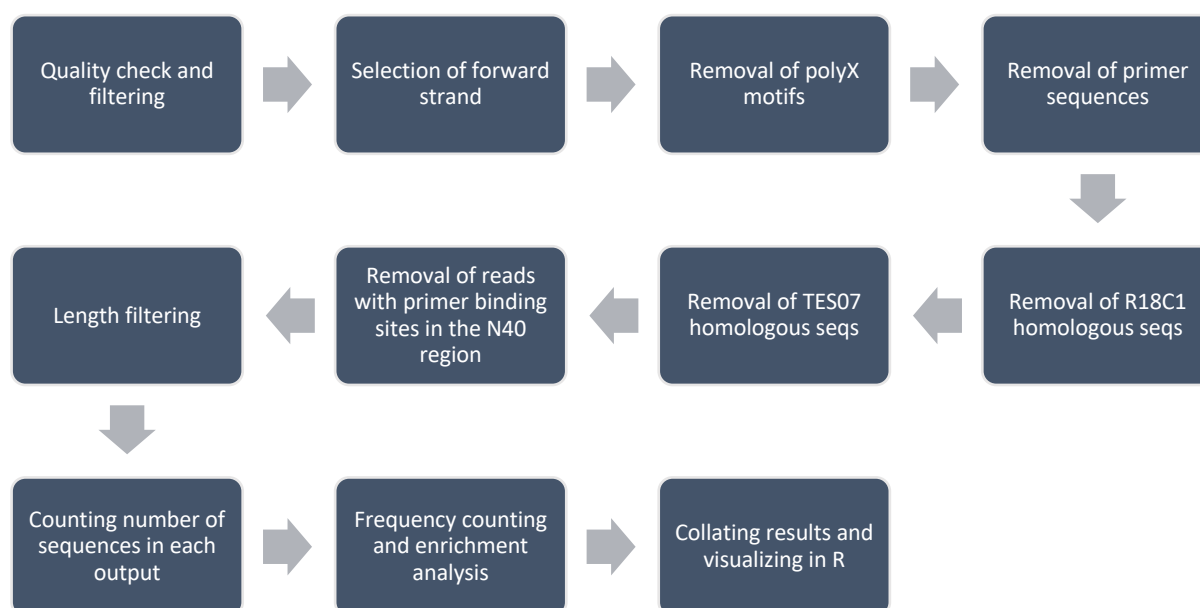
Gen X Pro edited the generated data by removing the adapter sequences, qPCR duplicates and tag counting to ensure their intellectual property was retained. All sequencing files were provided in FASTQ format and were downloaded directly from the Gen X Pro server for further bioinformatic analyses.

#### 6.2.8 Hardware used for analysis

All computational analyses were completed on a high performance computer based at Victoria University of Wellington using the Ubuntu 16.04.5 LTS – GNU/Linux 4.4.0-139-generic x86\_64 operating system. The server had 48 CPU cores available and approximately 200 GB of RAM.

#### 6.2.9 Description of Bioinformatics Pipeline

The bioinformatics pipeline used is presented in Figure 6.1 and details are described further below.



**Figure 6.1: Outline of bioinformatics pipeline.**

This pipeline was designed to automate the processing of a large number of sequencing files where each sequencing file represents a round of selection. The core set of 63 files (7 different SELEX strategies and two control libraries, each containing seven selection rounds) were analysed and results visualised.

#### 6.2.9.1 Quality Check and Filtering

To assess the quality of the sequencing data, the FastQC program from the Brabham Bioinformatics group was used. All 84 FASTQ files were analysed using FastQC version 0.11.7 to check the sequence quality, base content, length distribution and duplication levels. Initial early exploratory analyses revealed that all of the libraries consisted of an unusually high proportion of thymines in the N40 random region of the aptamer. This is discussed in later sections of this chapter and influenced the design of the analyses pipeline and justified the use of FastQC as an initial quality check. All reads with a Phred quality score of below 30 were removed from the dataset. Phred is a computer program that assigns a quality score to every

nucleobase from an automated DNA sequencing run using the properties of the peak shape and resolution from the DNA trace coupled with error probabilities (Ewing & Green, 1998). The Phred score relates to how likely it is that there is an error in the identification of the nucleobase which is called. For example, a Phred score of 30 equates to a probability of 1 in 1000 that the base is incorrect (or 99.9% accuracy) (Ewing & Green, 1998).

#### 6.2.9.2 Script 1: Selection of Forward Strand

During the SELEX process, only the forward strand was carried through to every SELEX round. Prior to sequencing the aptamer libraries, ssDNA was converted to dsDNA by PCR amplification and thus, the resultant data contained reads pertaining to both the forward and reverse strand. As a result of the library preparation and sequencing kit used, the raw sequencing data consisted of single-end reads with a length of 58 nucleotides. This meant that the resulting reads could be in one of two formats. Namely, either a forward primer region containing 18 nucleotides followed by a 40 nucleotide random region and no reverse primer region, or a reverse primer region containing 17 nucleotides followed by a 40 nucleotide random region and a 1 nucleotide forward primer region.

Therefore, sequences containing the forward primer at the beginning of the read originate from the forward strand and those containing the reverse primer originate from the reverse strand. All reads from the reverse strand were processed by generating the reverse complement to give the corresponding forward strand sequence. This process was completed using a script written in the bash programming language which utilised tools from the Fast-X Toolkit, specifically FASTQ/A Barcode Splitter and Reverse Complement. This script can be found in Appendix D1. In detail, reads were binned into their respective 'forward' or 'reverse' strand groups using the presence of either the forward or reverse primer as a barcode for

matching. Reads which didn't match to either primer were discarded. Those that matched to the reverse strand were piped into the reverse complement tool to produce the complementary forward strand. All forward strand reads were then combined and used for subsequent analyses. It was assumed that the reverse complement contained no mismatches, *i.e.* that the reverse complement of the reverse strand reads would match identically to the forward strand. An additional function of the FastX tool kit which was utilised here was 'FastX Stats' which enabled the calculation of global nucleotide counts to identify whether nucleotide bias was introduced during the designed selection steps or whether bias was present in the starting library.

#### 6.2.9.3 Script 2: Removal of PolyX Motifs

Further analysis of high frequency reads within the libraries from the early rounds of SELEX, revealed that the overabundance of thymine (T) nucleotides was due to a large number of polyT sequences. In order to determine how frequently polyT motifs were occurring within reads for each dataset, AfterQC may be used to remove any reads that contained polyX (consecutive and identical nucleotides) motifs at user provided thresholds (Chen et al., 2017). After some exploratory analyses (data not shown), the minimum length for polyX motifs was selected to be 10 nucleotides with zero tolerance for mismatches. This meant that if a read contained a motif of 10 of any nucleotide in a row and had no mismatches (*e.g.* TTTTTTTTTT or longer; or AAAAAAAAAA or longer), then it would be removed from the library and not included for deep data interrogation. The bash script used to complete length trimming, quality filtering and polyX identification and removal can be found in Appendix D2.

#### 6.2.9.4 Script 3: Removal of Primer Sequences

As the read length from HT sequencing was insufficient to cover both primers and the N40 region, the next step of data processing required the removal of the flanking primer sequences. This was completed prior to any downstream analyses. Primer removal was completed using a bash script which utilised Cutadapt 1.18 (Martin, 2011). Specifically, the non-internal 5' adapter option and the non-internal 3' adapter option were used to trim the forward and reverse primers, respectively. The script used for primer removal can be found in Appendix D3.

#### 6.2.9.5 Script 4 and 5: Removal of Contaminant Sequences

A monoclonal aptamer library (TES07) was used as a positive control during the amplification stage of SELEX. Additionally, another monoclonal aptamer library (R18C1) is used extensively in our laboratory and during the time these SELEX experiments were being undertaken. Thus, to ensure that the analysed data set only contained sequences from the selection pool for oxybenzone, and not those of TES07 or R18C1, a contamination check was undertaken using a custom designed bash script. This comprised of a combination of bash commands as well as Cutadapt 1.18 (Martin, 2011). The script was designed to remove any sequence that had  $\geq 50\%$  sequence similarity to either TES07 or R18C1 sequences.

A bash command utilising the *Agrep* function was used to ensure that Cutadapt had successfully removed the contaminating TES07 and R18C1 homologs. *Agrep* is a simple approximate pattern matching tool, however it is limited to looking for strings which contain a maximum of 8 mismatches. Therefore, *Agrep* was used as a basic control to ensure that Cutadapt had functioned correctly by searching for TES07, R18C1 and similar homologs. The script used for TES07 and R18C1 removal can be found in Appendix D4 and D5 respectively.



Counts of the number of TES07 and R18C1 sequences within each dataset at each round were also generated and graphed in the software package *R* Version 3.4.3 (R Core Team, 2019) using *Ggplot2* Version 3.1.0 (Wickham, 2016).

#### 6.2.9.6 Script 6: Removal of Reads with Primer Binding Sites within the N40 Region

During exploratory data interrogation, a very uncommon phenomena was detected whereby the flanking primer-binding sites were found to also be present in the N40 region. These reads were removed also, using the same methodology as in Section 6.2.9.5. The script for this can be found in Appendix D6.

#### 6.2.9.7 Script 7: Length Filtering

The random region of the selection library consisted of 40 nucleotides. Therefore, the last part of the processing pipeline was designed to filter reads based on length. Cutadapt Version 1.18 was again utilised to perform this function. Reads of 36 to 44 nucleotides in length were accepted in the filtering criteria, allowing for a maximum of four insertions or deletions. The script for length filtering can be found in Appendix D7.

#### 6.2.9.8 Script 8: Counting Number of Sequences in each Output

The first step of the downstream analyses was to count the number of unique sequences within each starting library prior to the occurrence of any selection. This was then repeated for every subsequent selection round (Rounds 1 – 7). The counting of unique sequences at each round, for each of the seven strategies being tested, was completed using the ‘count’ function from FASTAptamer Version 1.0.14 (Alam et al., 2015). All of the figures for the downstream analysis section of this chapter were generated using the software package *R*

Version 3.4.3 (R Core Team, 2019) and *Ggplot2* (v3.1.0; (Wickham, 2016)). The count files generated from this analysis were used as the input for the subsequent enrichment and comparative analyses detailed in Section 6.2.9.9. The script for generating counts of unique sequences can be found in Appendix D8.

#### 6.2.9.9 Script 9: Frequency Counting and Enrichment Analysis

The first indication of a selection regime successfully creating a better affinity aptamer is the level of enrichment observed in aptamer candidates in the final round of selection. In order to determine this, the enrichment that occurred between Rounds 6 and 7 was calculated by using the fold-change in the reads per million (RPM) between the two rounds. The 'enrich' function within FASTAptamer was used to calculate these values for every individual sequence within both datasets, for all seven selection regimes. The script for completing enrichment calculations may be found in Appendix D9. The enrichment values within each strategy were then ranked from highest to lowest and the top 50 individual sequences were selected for further analysis. The average enrichment value of the top 50 aptamer candidates, as well as the standard error, was then calculated using a script written in R software and plotted for each selection regime for comparison (see Appendix D12 and D14). Another common method of analysing sequencing data from SELEX is to look at the abundance (or frequency) of monoclonal aptamer candidates in the final round of selection. The output from FASTAptamer was used to visualize the frequency of top candidates from the seven different selection strategies as well as the two polyclonal control libraries. The output from the FASTAptamer enrichment analysis was split into files for individual rounds (see Appendix D10). Each individual round file was then utilized to rank and filter the top 50 candidates from each strategy in terms of normalized frequency (reads per million). The script for completing

this can be found in Appendix D11. Visualizing frequency trends at each selection round was completed with another script written in R (see Appendix D13).

#### 6.2.10 Longitudinal Enrichment of the Top 50 Candidates

The highest levels of enrichment of a monoclonal oligo within each selection regime might not necessarily occur in the final round of selection, but in an earlier round or rounds. Therefore, in order to identify changes throughout the length of the SELEX process, the cycle-to-cycle enrichment for each selection regime was calculated as described above, and the patterns visualized. The script for analysing the enrichment patterns in any round can be found in Appendix D14.

#### 6.2.11 Statistical Analysis of Frequency and Enrichment

All statistical analyses were completed using R Version 3.4.3 (R Core Team, 2019). The frequency of the top 50 aptamer candidates in Selection Round 7 were compared for each selection strategy and the two control strategies. Histograms and quantile-quantile plots (Q-Q plot) of the non-transformed data were generated using the *'hist'* and *'qqnorm'* functions. Visualization of these plots indicated that the data was not normally distributed and needed to be logarithmically transformed for statistical analyses. Data was logarithmically transformed using the *'log'* function. The logarithmically transformed data was also visualized using a histogram and a Q-Q plot. The Q-Q plot of the log-transformed data indicated a normal distribution. Levene's test was performed using the *'levene.test'* function to check homogeneity of variance. This test indicated that there was not homogeneity of variance ( $p < 0.05$ ). Due to these characteristics, a Welch's ANOVA was chosen as an appropriate method to assess the difference in group means. Welch's ANOVA can be used when data violate the assumption of homogeneity of variance (Welch, 1951). A Welch's ANOVA was performed in

R using the *'welch.test'* function from the *'onewaytests'* package (Dag et al., 2018; Welch, 1951). The Games-Howell test was chosen to perform post-hoc comparisons as it can also be used when group means have unequal variance (Ruxton & Beauchamp, 2008). The test was performed using the *'one.way'* function from the *'userfriendlyscience'* package (Peters et al., 2018 ; Games & Howell, 1976). Statistical significance was denoted on figures using letters, where all variables with the same letter indicate no statistically significant difference in group means. Variables with different letters are significantly different. The mean cycle-to-cycle enrichment of the top 50 candidates in Selection Round 7 from each selection strategy was also compared using the same method described above. Comparisons were not made against the control libraries in this case as some of the oligonucleotides in the control libraries had enrichment values of zero meaning that they could not be logarithmically transformed.

#### 6.2.12 Frequency Tracing of the Top 50 Candidates

In order to understand the evolutionary route of the top 50 aptamer candidates ranked by frequency in the final SELEX, the normalized frequency (reads per million) of each individual aptamer was tracked back through the selection process using scripts in the R programming language that take the output from FASTAptamer, process it and perform tracing. Specifically, the first script (see Appendix D10), takes the output from the cycle-to-cycle enrichment analysis, splits the output into frequency counts for each individual round and then saves the data from each round in a separate .csv file. The second script (see Appendix D15) then ranks the candidates in terms of normalized frequency in Round 7, selects the top 50 rows (sequences) from this data frame and then searches for each individual sequence in each of the selection rounds (1-7) and returns their frequency in each round. As long as the file architecture is set-up correctly, the name of the selection regime (named 'condition' in the

script) is all that needs to be modified in order to run the analysis again. The second script also graphs the “fate” of all of these top 50 sequences by plotting their frequency over different selection rounds. All analyses were run using the same packages specified above, however *dplyr* Version 0.7.8 (Wickham, Francois, Henry, & Muller, 2018) was also used to filter specific variables for data visualisation.

#### 6.2.13 Selection of Candidates for MST Analysis

Due to the high costs, only the top candidates, ranked by normalised frequency, from each selection strategy at Round 7 were selected for binding affinity analyses by micro-scale thermophoresis (MST). In addition, two other aptamer candidates were selected based on having the highest enrichment from Rounds 6 to 7. These candidates were from Selection Strategy 6 (negative selection) and 7 (counter selection). The most common method for selecting candidates is frequency count, and as such was chosen as the first selection criteria. As previously mentioned, enrichment has been suggested to be a better predictor of binding affinity and therefore, two candidates were selected based on enrichment from the two strategies that were hypothesised to produce the best candidates. My rationale was that the higher stringency conditions imposed by negative and counter selection steps would produce candidates with improved binding affinity.

#### 6.2.14 Characterisation of Aptamer Candidates

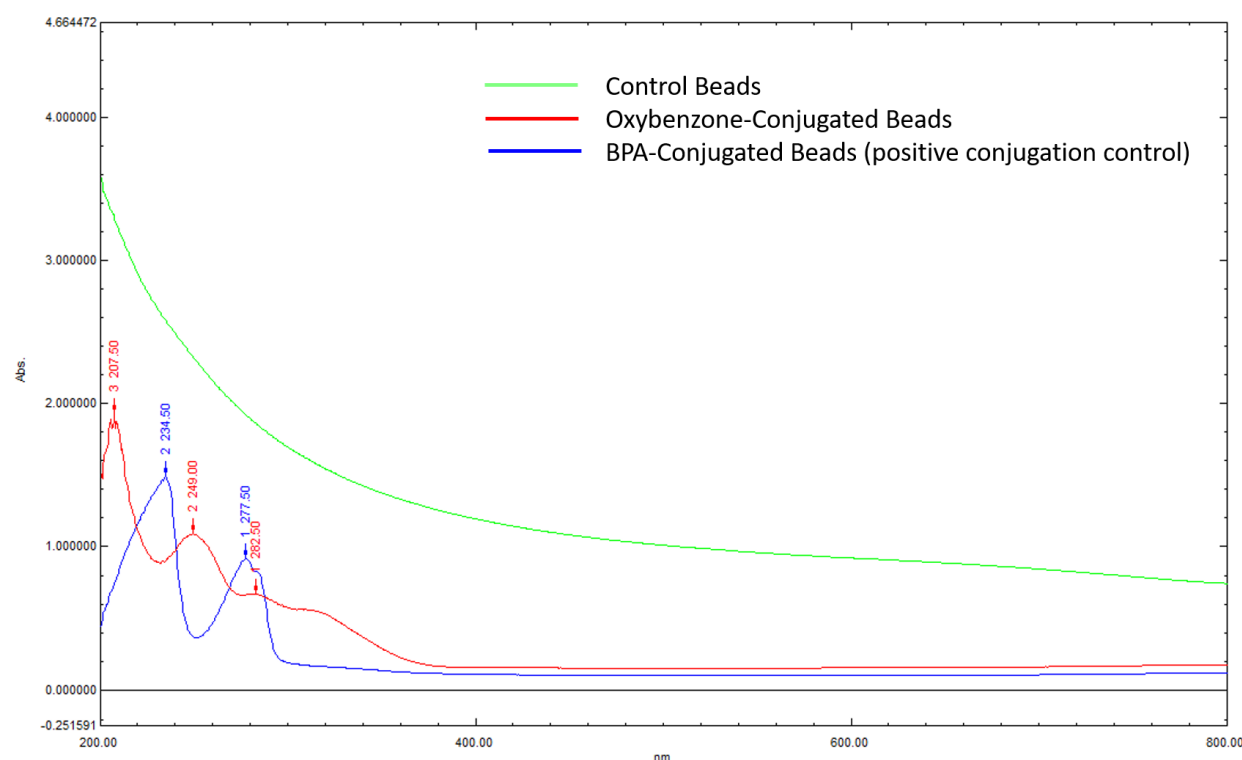
The characterisation of aptamer candidates for oxybenzone using MST was completed entirely by 2Bind (Regensburg, Germany). The composition of the assay buffer, ligand solubility instructions and potential binding affinity advice were provided to 2Bind. The protocol for MST was designed and implemented by 2Bind as follows. A 10 mM stock solution of oxybenzone was prepared in 100% ethanol and diluted to a 1 mM working solution in 100%

ethanol. The concentrations of oxybenzone were made up using 5% ethanol to ensure that the ligand remained in-solution. The instrument used for MST was a Monolith NT.115 (Nanotemper, Munchen, Germany). The concentration of the aptamer was kept constant at 5nM and the serial dilutions of oxybenzone were prepared using two-fold dilutions from 100  $\mu$ M to 3.05 nM. Both oxybenzone and aptamer were made up in 1X BWB with 5% (v/v) ethanol. A aliquot of 5  $\mu$ L of oligonucleotide was mixed with 5  $\mu$ L of oxybenzone at each concentration and the mixture was added to a standard capillary for analyses on the instrument. The instrument was set to 80% LED power and 60% laser power at a temperature of 25°C. Two technical repeats were completed. Data analyses and the fitting of a binding isotherm was completed by 2Bind using the software supplied with the instrument.

## 6.3 Results

### 6.3.1 Conjugation of Oxybenzone to the affinity matrix

Successful conjugation of oxybenzone to the CL-6B Sepharose beads was indicated by the peaks at 240 and 280 nm on the UV-visible spectrum (Figure 6.2). As expected, the ethanolamine-conjugated control matrix had no notable absorption peaks.

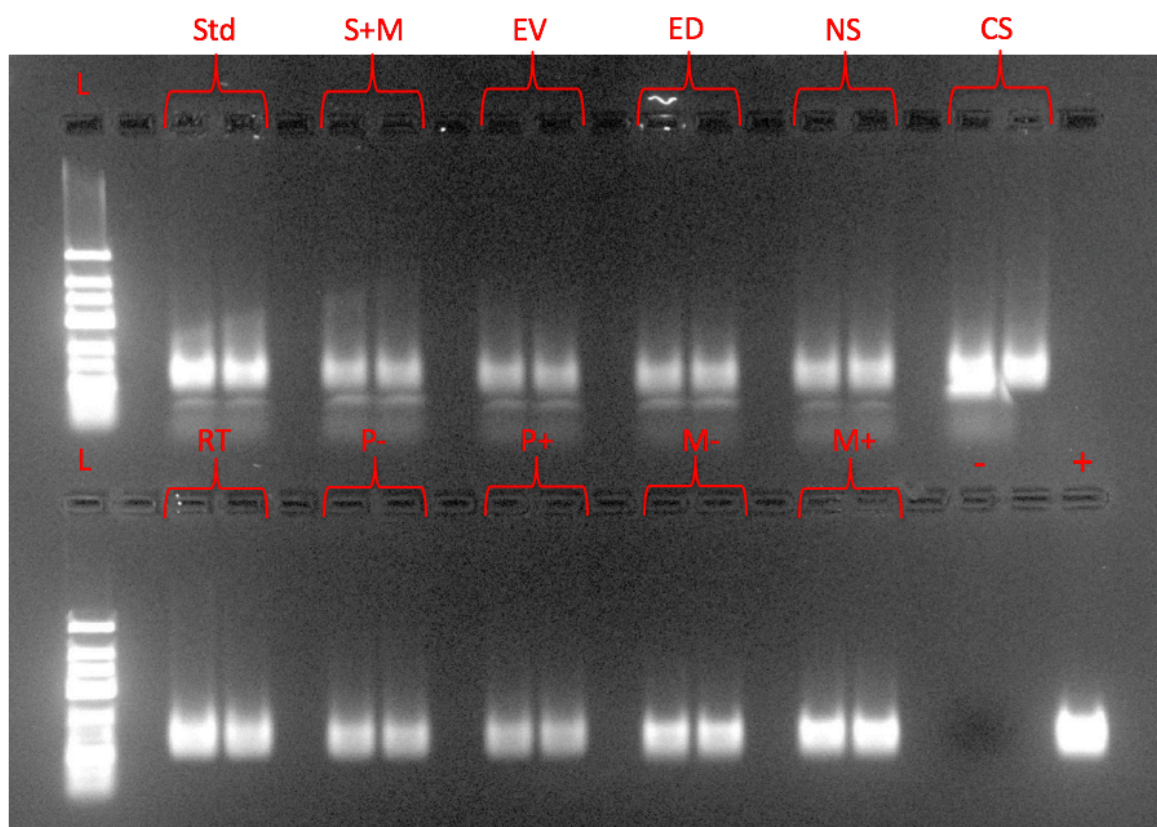


**Figure 6.2: UV-Visible spectrum of control beads (conjugation with Ethanolamine), oxybenzone-conjugated beads and BPA-conjugated beads.**

### 6.3.2 Results of SELEX Rounds

For the first selection round, all libraries were incubated with target-conjugated beads whilst the four control libraries were not incubated with target beads to determine parasitic sequences that are enriched by PCR-bias. As the selection pressures were the same on all of the libraries for the first selection round, the band intensity, and thus the amount of library

being retained and subsequently amplified, was similar across all strategies (Figure 6.3). It is also important to note that none of the strategies employed mutation in the first selection round.



**Figure 6.3:** Gel electrophoresis image showing PCR amplicons from SELEX Round 2 that were incubated with target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. Amplicons from a monoclonal (M; R18C1) and polyclonal (P; N40) library are also depicted following the omission (-) or use (+) of mutagenesis PCR. The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.



During Round 2, selection pressures were varied as per each selection strategy condition. In addition, each aliquot of ssDNA from the previous selection round was split into two, with each half of the ssDNA being incubated with either control or oxybenzone-conjugated beads. It can be observed that for most strategies, the bands depicting the ssDNA that had been incubated with oxybenzone-conjugated beads were brighter (Figure 6.4). This is especially obvious in the selection strategies where the wash stringency was increased (*i.e.* Extra Volume, Extra Detergent (ED), Negative Selection (NS) and Counter Selection (CS)) and where no mutation was employed (*i.e.* Standard (S)). Monoclonal and polyclonal control libraries amplified using mutagenic PCR (M+ and P+, respectively) resulted in higher band densities than those without mutagenesis (M- and P-, respectively).

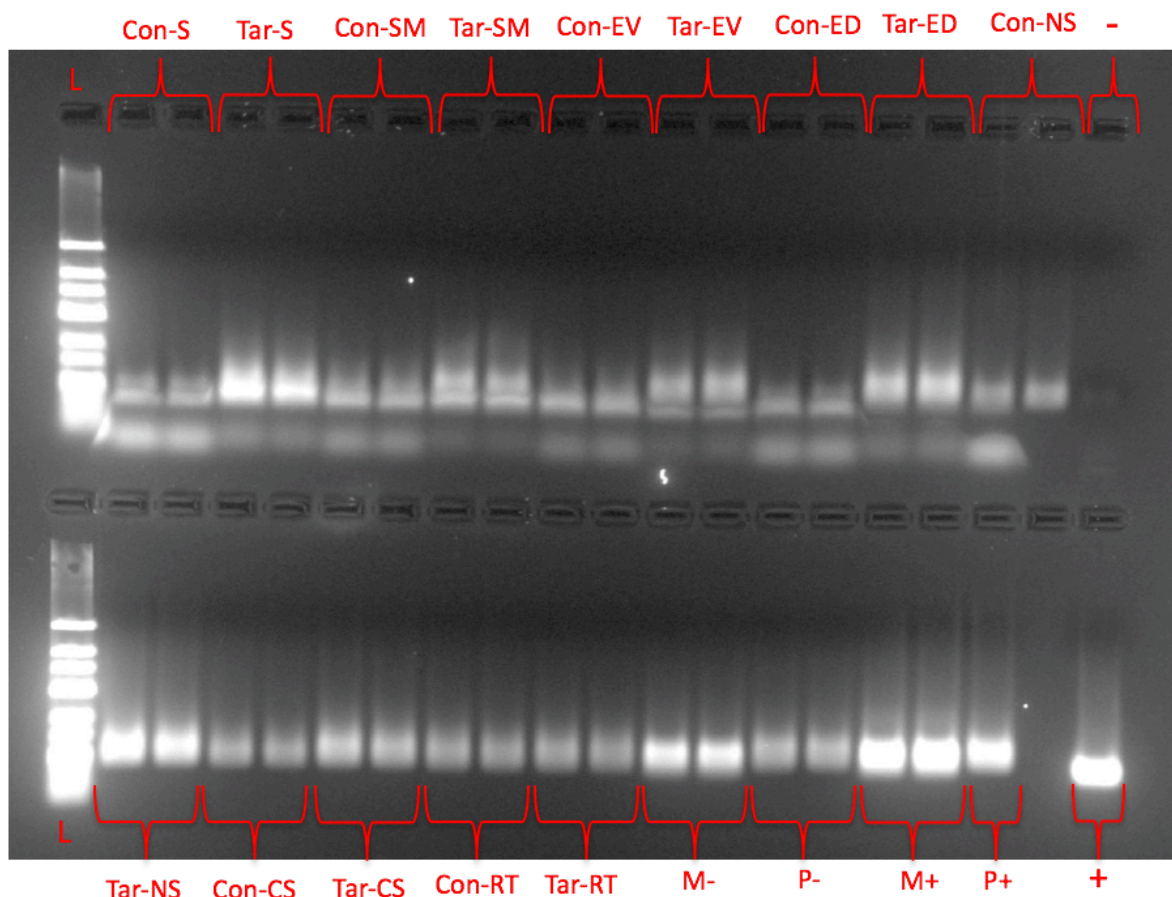
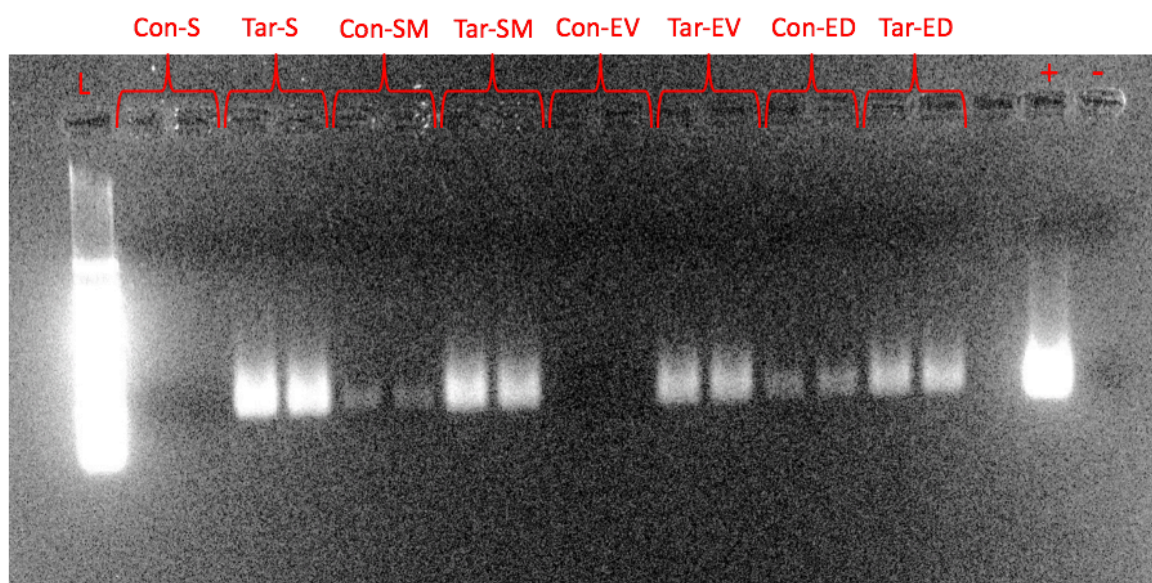


Figure 6.4: Gel electrophoresis image showing PCR amplicons from SELEX Round 2 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. Amplicons from a monoclonal (M; R18C1) and polyclonal (P; N40) library are also depicted following the omission (-) or use (+) of mutagenesis PCR. The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

During Selection Round 3, a negative selection step was used in both the Negative Selection and Counter Selection strategies in an attempt to remove oligonucleotides within the library that bound specifically to the Sepharose-bead matrix (Figure 6.5). Even without the use of

negative selection, it can be observed that libraries preferentially bind to the target beads. There was not enough DNA retained on the control beads to produce visible bands in the gel for the Standard and Extra Volume strategies. Moreover, the bands from the Standard + Mutation and Extra Detergent strategies were more dense following incubation with target-conjugated, than control, beads.



**Figure 6.5:** Gel electrophoresis image showing PCR amplicons from SELEX Round 3 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

Negative selection was clearly effective during Round 3 at removing oligonucleotides which bind to the control bead matrix (Figure 6.6). Prior to the negative selection step, the PCR bands, regardless of being incubated with either control or oxybenzone-conjugated beads,

were dense. After the negative selection step, the bands representing DNA amplified from the control bead templates, Con-NS (A) and Con-CS (A) were much less dense. After three rounds of selection, oligo's within the Room Temperature strategy still appeared to bind to both the control and target bead matrices with similar affinity. Following three successive rounds of PCR, amplification of the control libraries (M-, M+, P- and P+) was causing smeared product in the gel (particularly for M- and P+, see Figure 6.6 and 6.7). At the time this was believed to be caused by excess sample due to either a higher concentration of template being introduced into the PCR reaction or an excess number of PCR cycles. Following Round 4, the control libraries were diluted prior to amplification and amplified separately to the SELEX libraries.

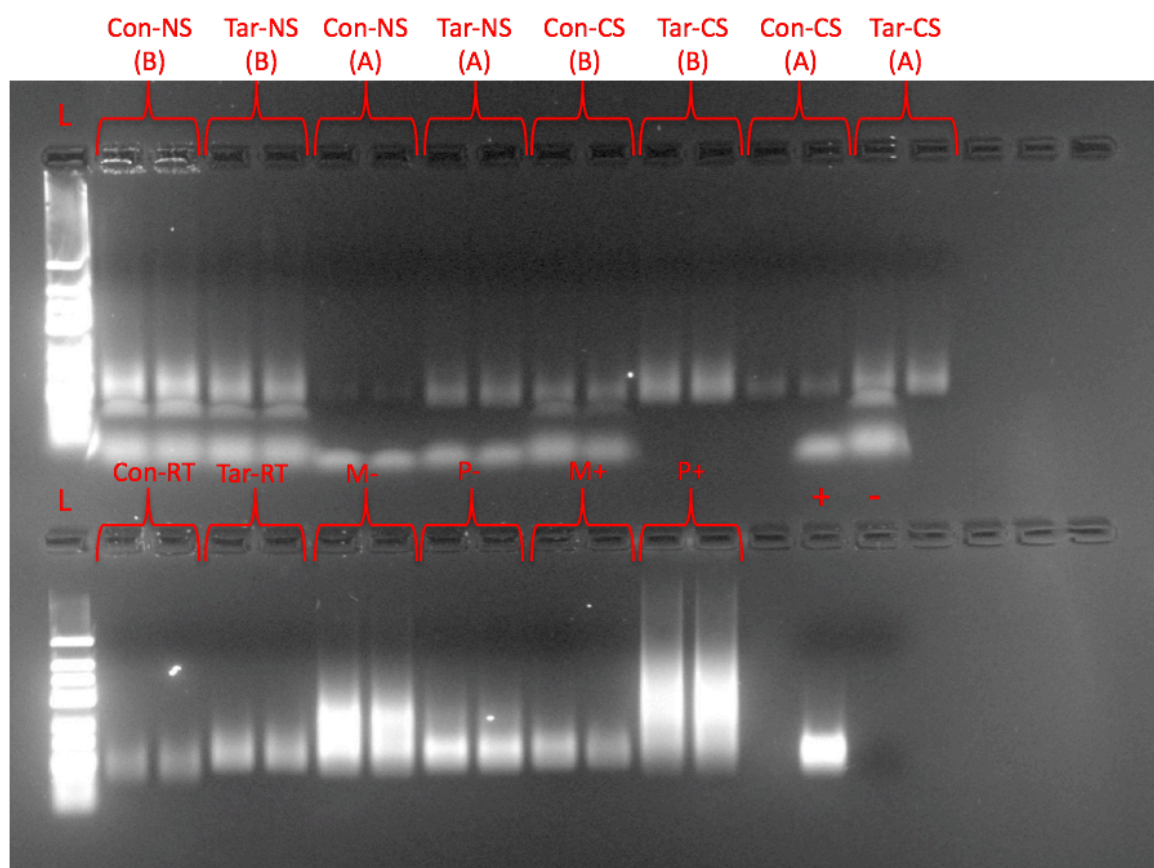


Figure 6.6: Gel electrophoresis image showing PCR amplicons from SELEX Round 3 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. Amplicons from a monoclonal (M; R18C1) and polyclonal (P; N40) library are also depicted following the omission (-) or use (+) of mutagenesis PCR. (A) represents libraries amplified after negative selection and (B) represents the same library amplified before negative selection. The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

The effect of negative selection can be observed more prominently at the fourth selection round (Figure 6.7). In all of the strategies that did not include a negative selection step (*i.e.*

Standard, Standard + Mutation, Extra Volume, Extra Detergent and Room Temperature strategies), there was evidence of some binding to the control matrix. For the Standard + Mutation strategy, the amount of DNA amplified from the control bead incubation is almost identical to that from the target bead incubation. In the Extra Detergent strategy, the band relating to the control bead incubation was much fainter than that for the target beads, indicating much less DNA is being retained. Interestingly in the Negative Selection strategy, the band for the control bead incubation was brighter than the target bead band. In comparison, in the Counter Selection strategy, the negative selection step seems to have been more effective compared to that in the Negative Selection strategy, with no visible band present from the control bead incubation. Finally, bands from the Room Temperature strategy were comparable between control and target bead incubations, but overall were much fainter than the other strategies.

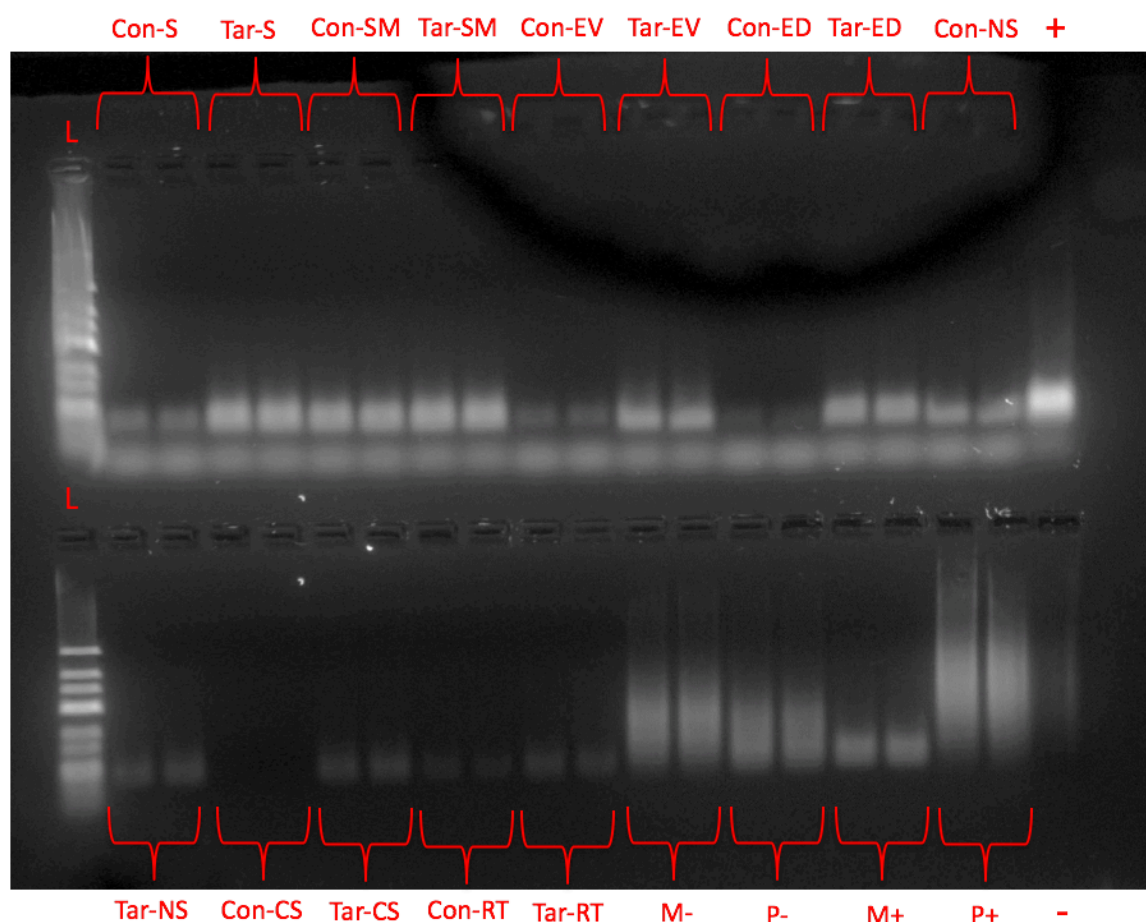


Figure 6.7: Gel electrophoresis image showing PCR amplicons from SELEX Round 4 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. Amplicons from a monoclonal (M; R18C1) and polyclonal (P; N40) library are also depicted following the omission (-) or use (+) of mutagenesis PCR. The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.



The results from Round 5 SELEX are illustrated in Figure 6.8 and show similar trends to that of Round 4. For the Standard, Extra Detergent, Extra Volume, Counter Selection and Room Temperature strategies, more DNA was retained from the target bead, than the control bead, incubations indicating preferential binding to the oxybenzone-conjugated beads. In fact, for the latter three strategies, there was no visible band from the control bead incubation indicating a strong preference of the library towards the oxybenzone-conjugated beads. This is likely to indicate an evolutionary trajectory towards a library enriched for oligonucleotides that bind to oxybenzone. In the Standard + Mutation and the Negative Selection strategies, the template from both bead incubations was amplified to a similar level.



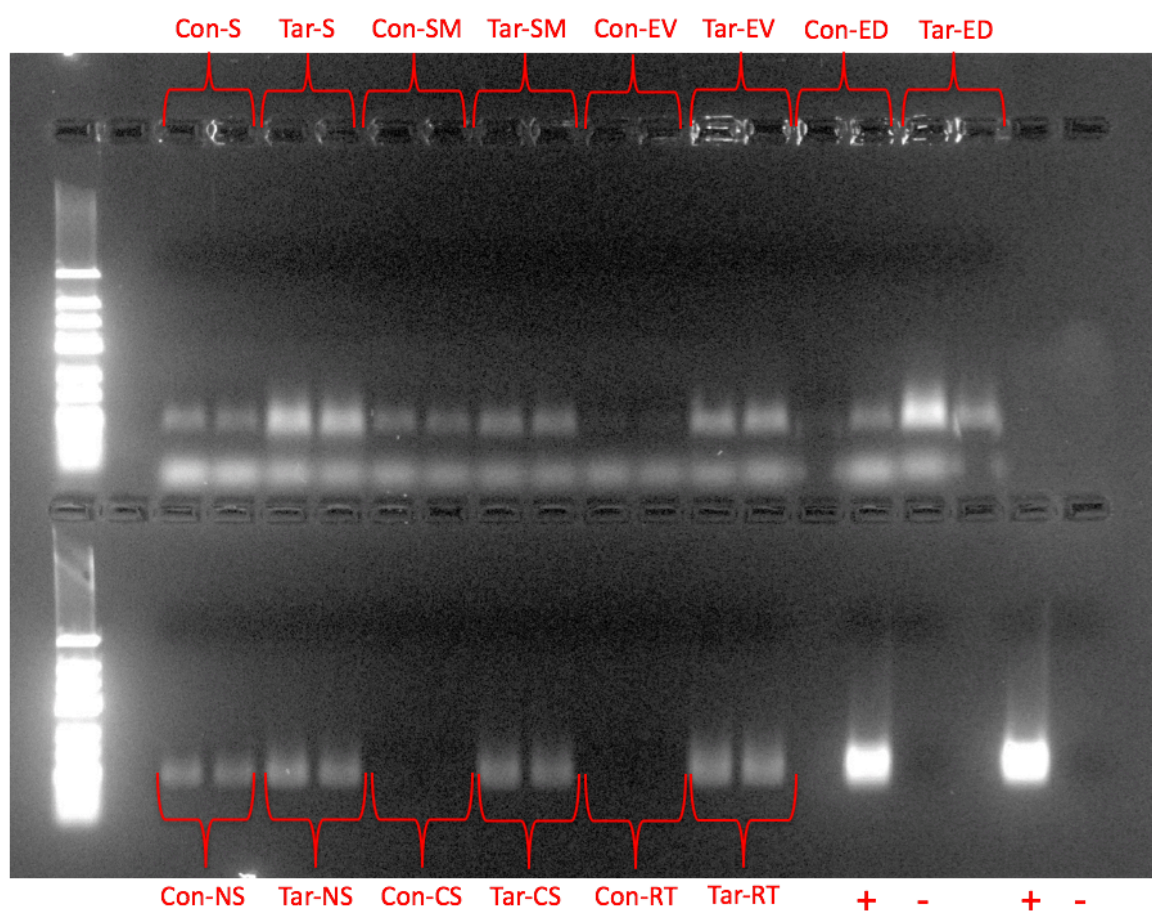


Figure 6.8: Gel electrophoresis image showing PCR amplicons from SELEX Round 5 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

Round 6 included a second iteration of negative selection and for those strategies using detergent, an increase in wash stringency to 0.5% IGEPAL in BWB. The three strategies which did not use detergent in the BWB resulted in bands corresponding to the target bead incubations and no visible bands in the control bead incubations (Figure 6.9). For all strategies which used increased detergent and negative selection (*i.e.* the Negative Selection and Counter Selection strategies), no visible bands were observed when amplifying DNA which had been incubated with the target-conjugated beads. It was hypothesized that this was due the increased detergent concentration being too stringent, and thus resulted in the removal of all DNA from the beads. The washing stringency associated with these strategies (*i.e.* Extra Detergent, Negative Selection, Counter Selection and Room Temperature) was therefore reduced and Round 6 was repeated, as illustrated in Figure 6.10.

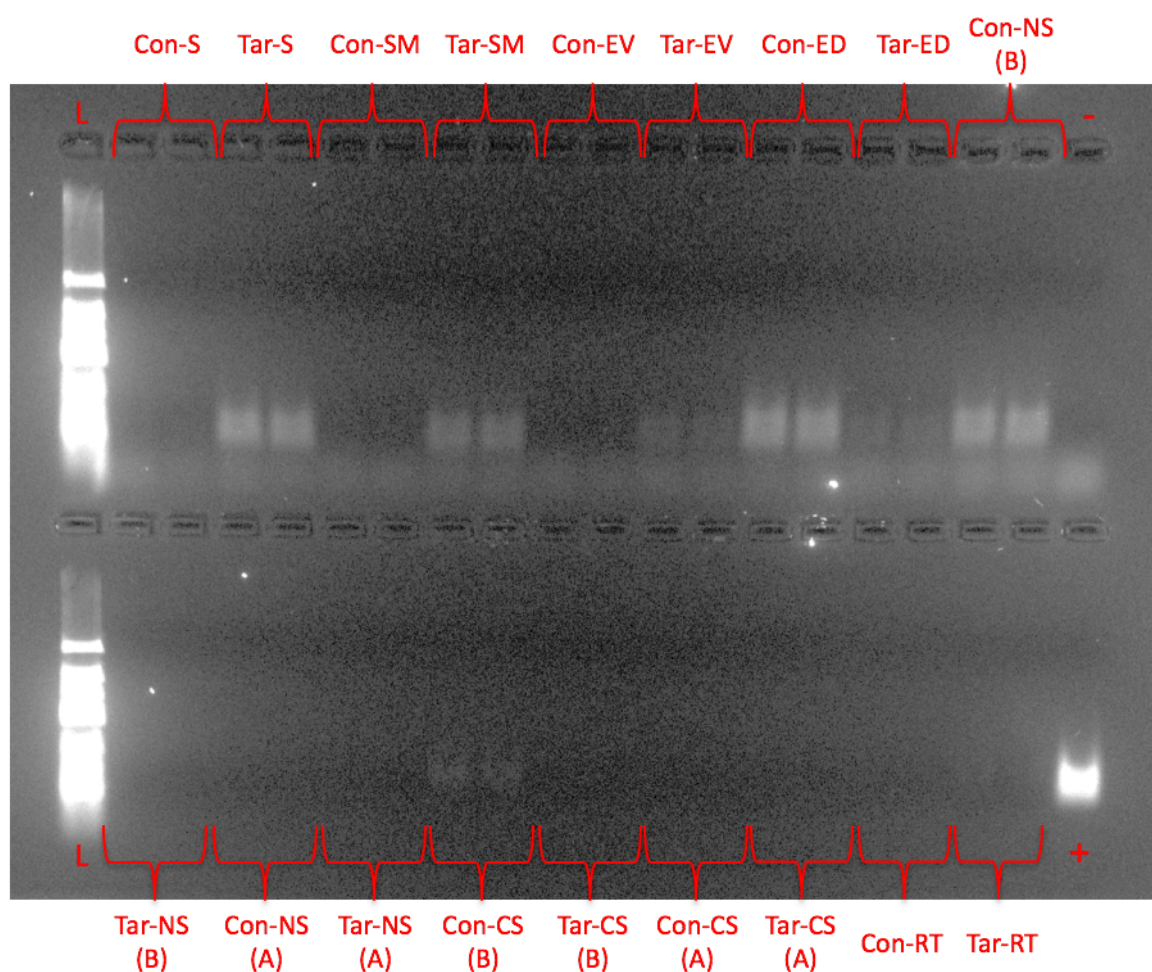
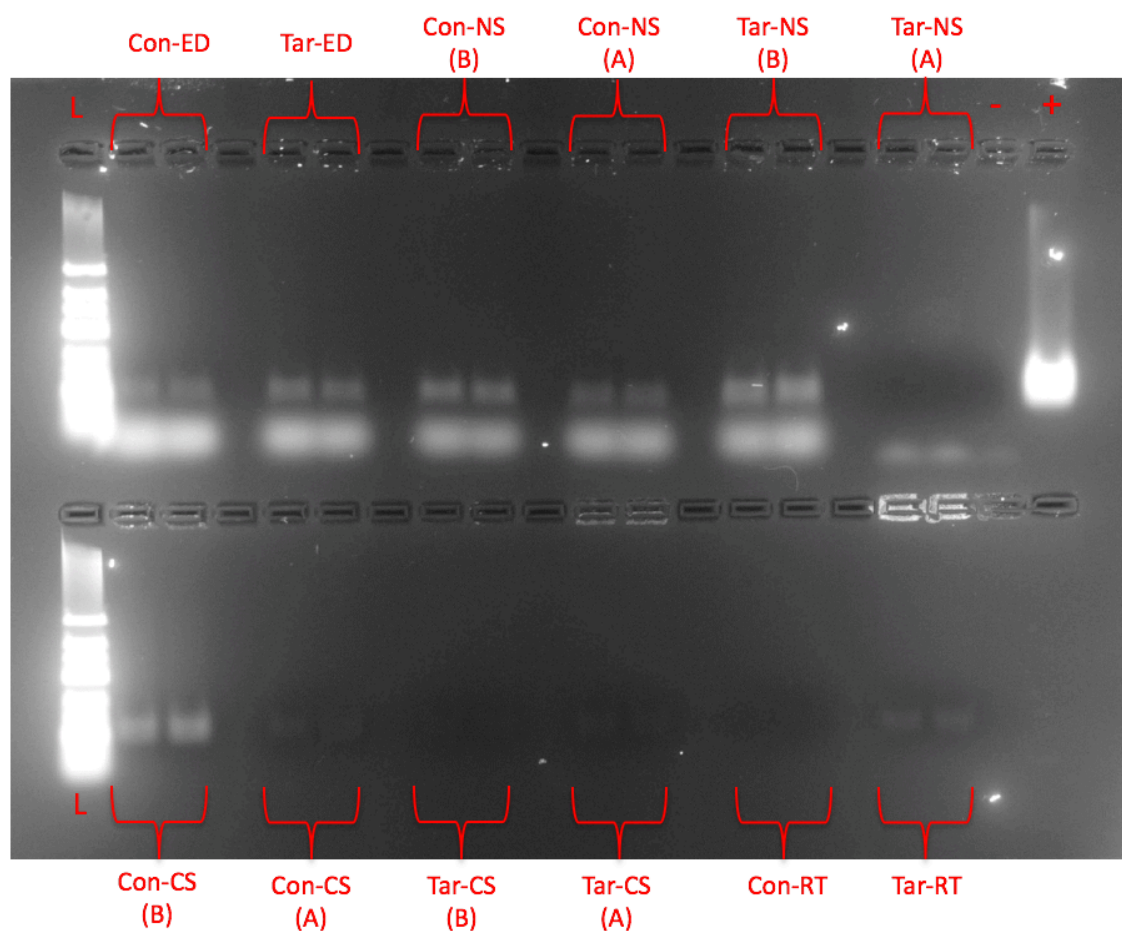


Figure 6.9: Gel electrophoresis image showing PCR amplicons from SELEX Round 6 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. (A) represents libraries amplified after negative selection and (B) represents the same library amplified before negative selection. The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

Repeating with a lower stringency wash still resulted in no visible bands corresponding to the target bead incubation for strategies involving negative selection. For these strategies (Counter Selection and Negative Selection), the template was amplified using a higher number of PCR cycles (*i.e.* 35 versus 30) to attain sufficient DNA for the final selection round (Figure 6.11).

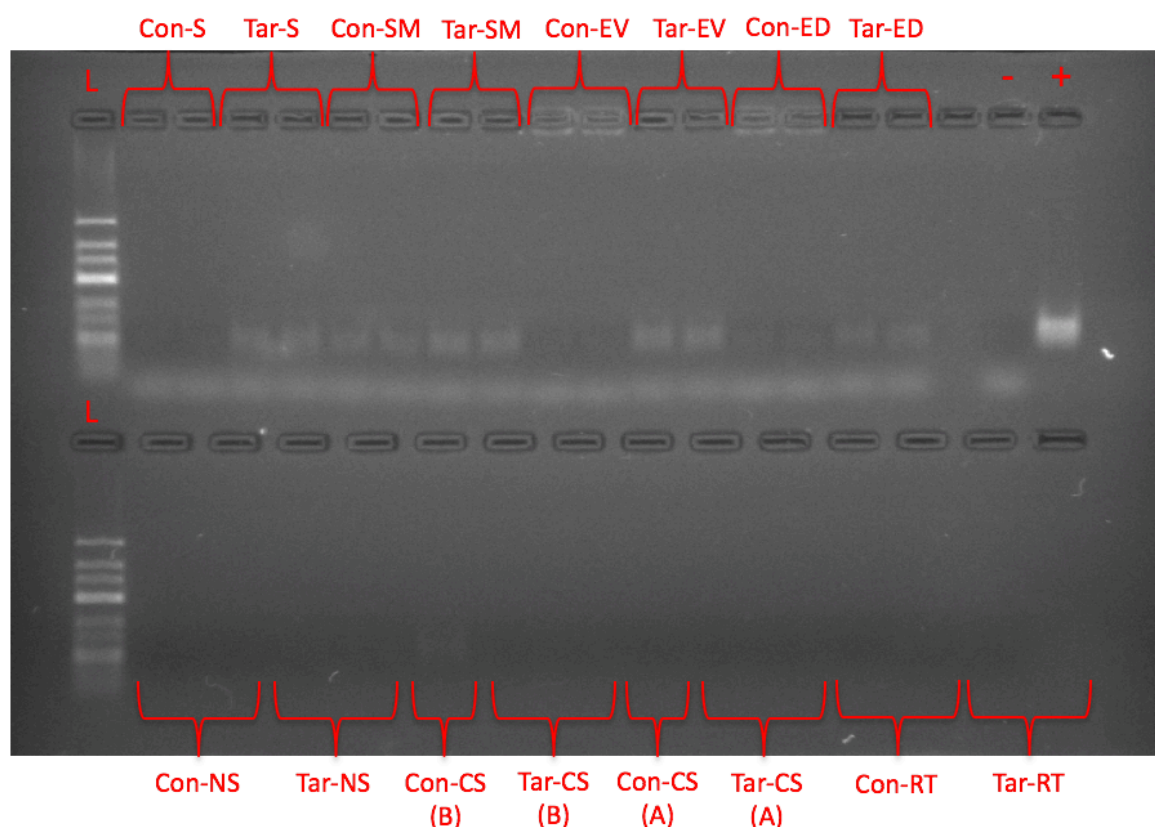


**Figure 6.10:** Gel electrophoresis image showing PCR amplicons from a repeat of SELEX Round 6, with lower detergent concentrations, that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. (A) represents libraries amplified after negative selection and (B)

represents the same library amplified before negative selection. The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

Selection Round 7 involved counter selection for the Counter Selection strategy only. Visible bands with a sufficient quantity of DNA for high-throughput sequencing were generated for the Standard, Extra Volume, Extra Detergent and Standard + Mutation strategies. The first three strategies produced sufficient differentiation in the visual brightness of the PCR bands when comparing the control to the target bead incubations. Visible bands were unable to be generated for the Negative Selection, Counter Selection and Room Temperature strategies for either bead incubation (Figure 6.11).





**Figure 6.11:** Gel electrophoresis image showing PCR amplicons from SELEX Round 7 that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. (A) represents libraries amplified after counter selection and (B) represents the same library amplified with normal selection conditions as a comparison. The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

To generate a sufficient quantity of DNA for these strategies, the number of PCR cycles was increased from 30 to 35 and results are depicted in Figure 6.12.

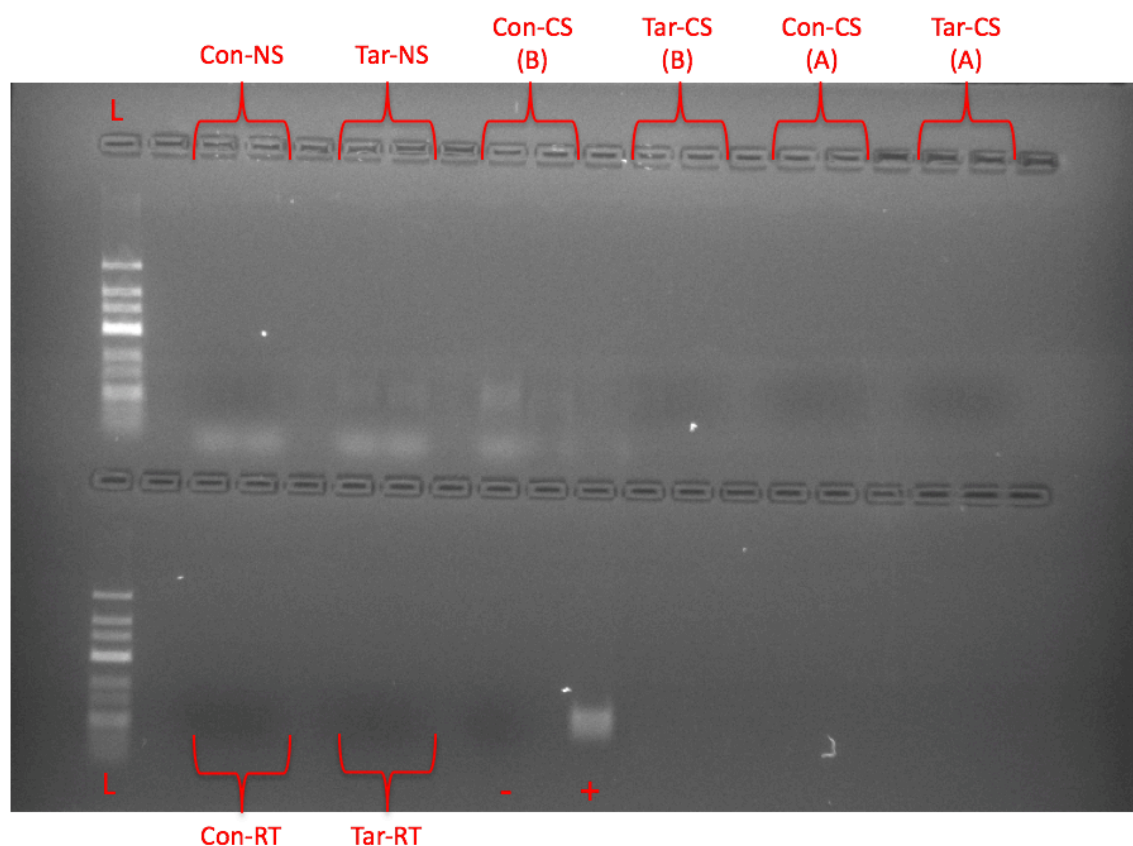
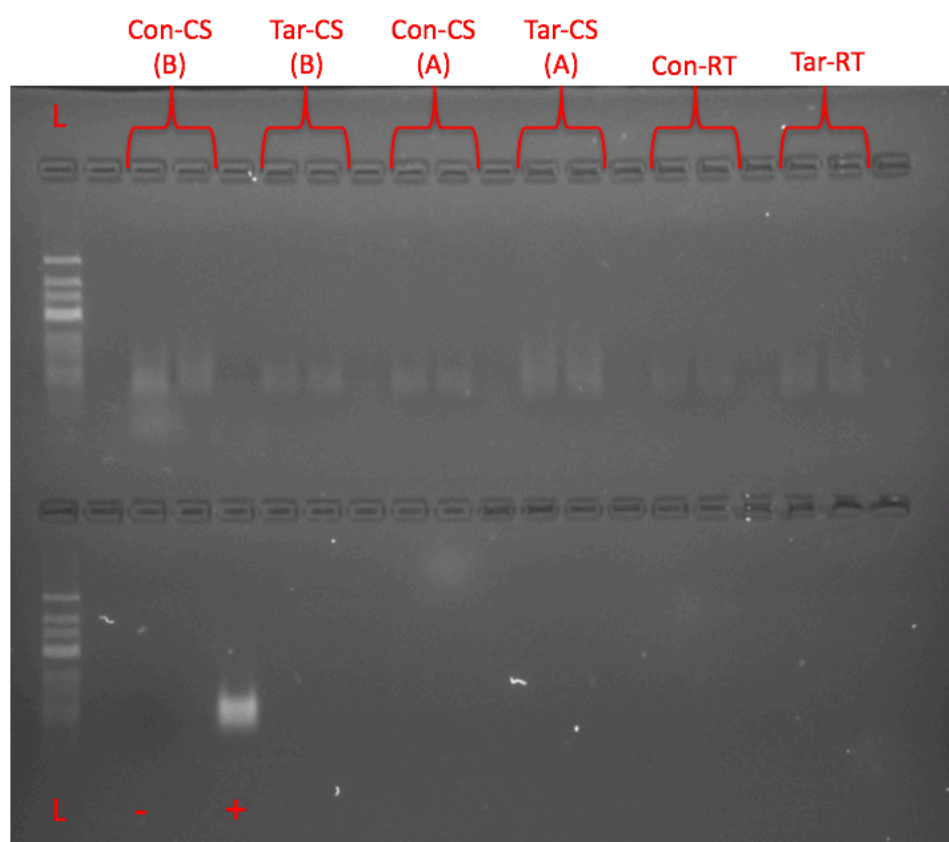


Figure 6.12: Gel electrophoresis image showing PCR amplicons from a repeat of SELEX Round 7, using 35 cycles of PCR, that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Standard (S), Standard + Mutation (SM), Extra Volume (EV), Extra Detergent (ED), Room Temperature (RT), Negative Selection (NS) and Counter Selection (CS) strategies. (A) represents libraries amplified after counter selection and (B) represents the same library amplified with normal selection conditions as a comparison.

Using 35 cycles, suitable quantities of DNA for sequencing were only able to be generated for the Negative Selection strategy but not for the Counter Selection or Room Temperature strategies. For these strategies, the PCR was repeated and the number of cycles was increased to 40 (Figure 6.13). This resulted in the sufficient amplification of both strategies for

sequencing . The bands produced by amplifying libraries with (Figure 6.13; Lanes Con-CS (A) and Tar-CS (A)) and without (Figure 6.13; Lanes Con-CS (B) and Tar-CS (B)) a counter selection step in the Counter Selection (CS) strategy were of similar brightness, indicating that this library was specific for the target molecule. It should be noted that DNA from the control bead incubation produced a band equally as bright as the target bead incubation for both the Counter Selection and Room Temperature strategies.



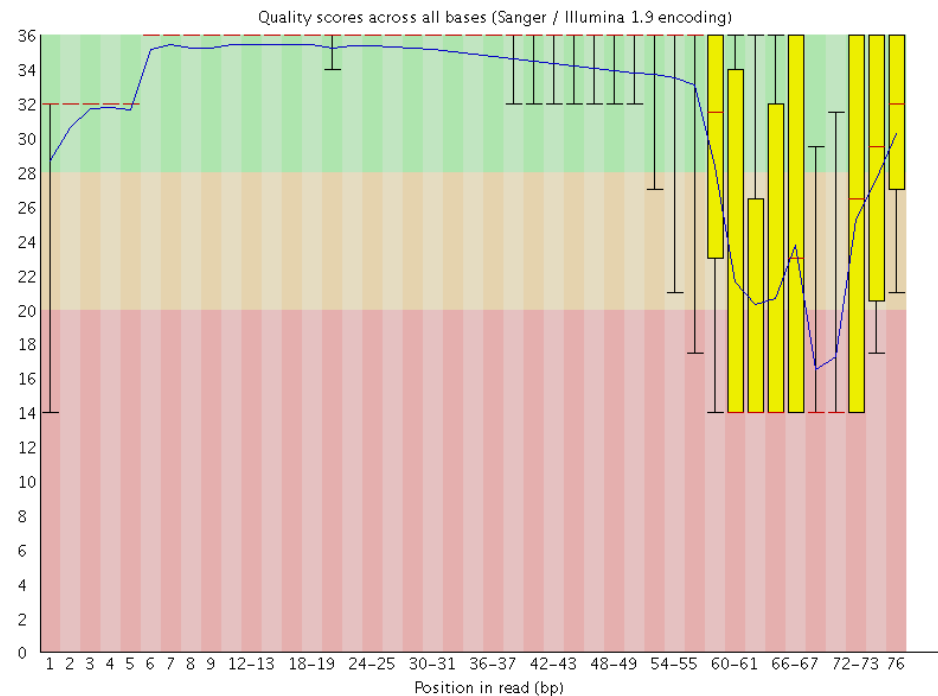
**Figure 6.13:** Gel electrophoresis image showing PCR amplicons from SELEX Round 7, using 40 cycles of PCR, that were incubated with either control (ethanolamine-conjugated, Con) or target (oxybenzone-conjugated) matrix following Room Temperature (RT) and Counter Selection (CS) strategies. (A) represents libraries amplified after counter selection and (B) represents the same library amplified with normal selection conditions as a comparison.



Control libraries for testing the ability of PCR to enrich a library without any other selective pressure were diluted prior to PCR to prevent the smearing displayed in Figure 6.7 after Selection Round 4. Various dilutions were tested to ensure that clear bands were produced (data not shown). The libraries were then taken through to the completion of the seventh round and the resultant gels may be found in Appendix F.

### 6.3.3 Quality and quantity of sequencing reads

Data received from the sequencing supplier (Gen X Pro) was of acceptable quality as indicated in Figure 6.14. An average Phred score of above 30 (meaning a 99.9% chance that the base has been called correctly) is maintained across the desired read length up until 58 nucleobases, where after the Phred score decreased slightly. All of the received datasets were similar to the representative chart displayed here for the first library of the Standard strategy. Sequences with an average Phred score of lower than 30 were removed before subsequent analyses.



**Figure 6.14: Quality score of sequencing data (Library S01) from Gen X Pro. Green indicates a Phred score that is acceptable (a 99.9% or greater chance that the base has been called correctly), yellow indicates a 90% or greater chance of correct base calling and red indicates a less than 90% chance. An average Phred score of above 30 is maintained across the read length of 58 nucleobases.**

### 6.3.4 Read Numbers

The numbers of raw reads and of reads remaining after each processing step, as well as the percentage of the original reads remaining, is displayed in Table 6.3.

**Table 6.3: Number of reads at each processing step.**

Library	Selection Strategy	Round	Raw Reads	# After PolyX Removal	%After PolyX Removal	# After Primer Removal	% After Primer Removal	# After TES07 Removal	% After TES07. Removal	After R18C1 Removal	% After R18C1 Removal	# After N40 Primer Removal	% After N40 Primer Removal	Processed Reads
S1	Standard	1	845,231	501,219	59%	501,219	59%	377,542	45%	375,684	44%	375,353	44%	375,353
S2	Standard + Mutation	1	780,357	460,133	59%	460,133	59%	343,553	44%	341,829	44%	341,497	44%	341,497
S3	Extra Wash Volume	1	1,050,453	615,583	59%	615,583	59%	462,761	44%	460,482	44%	460,050	44%	460,050
S4	Extra Wash Detergent	1	957,892	575,620	60%	575,620	60%	432,895	45%	430,721	45%	430,323	45%	430,323
S5	Negative Selection	1	1,046,729	612,374	59%	612,374	59%	459,875	44%	457,616	44%	457,152	44%	457,152
S6	Counter Selection	1	927,889	540,626	58%	540,626	58%	406,929	44%	404,816	44%	404,418	44%	404,418
S7	Room Temperature	1	925,947	543,032	59%	543,032	59%	409,741	44%	407,732	44%	407,349	44%	407,349
S8	Standard	2	504,345	297,390	59%	297,390	59%	220,392	44%	219,283	43%	219,038	43%	219,038
S9	Standard + Mutation	2	700,672	415,454	59%	415,454	59%	314,378	45%	312,980	45%	312,678	45%	312,678
S10	Extra Wash Volume	2	627,607	372,359	59%	372,359	59%	280,194	45%	278,835	44%	278,573	44%	278,573
S11	Extra Wash Detergent	2	538,087	288,801	54%	288,801	54%	219,456	41%	218,538	41%	218,329	41%	218,329
S12	Negative Selection	2	553,779	303,534	55%	303,534	55%	231,127	42%	230,142	42%	229,864	42%	229,864
S13	Counter Selection	2	466,767	240,767	52%	240,767	52%	184,867	40%	184,178	39%	183,967	39%	183,967
S14	Room Temperature	2	409,881	246,576	60%	246,576	60%	183,234	45%	182,179	44%	181,958	44%	181,958
S15	Standard	3	270,913	159,143	59%	159,143	59%	118,847	44%	118,277	44%	118,169	44%	118,169
S16	Standard + Mutation	3	571,582	360,758	63%	360,758	63%	263,368	46%	261,964	46%	261,574	46%	261,574

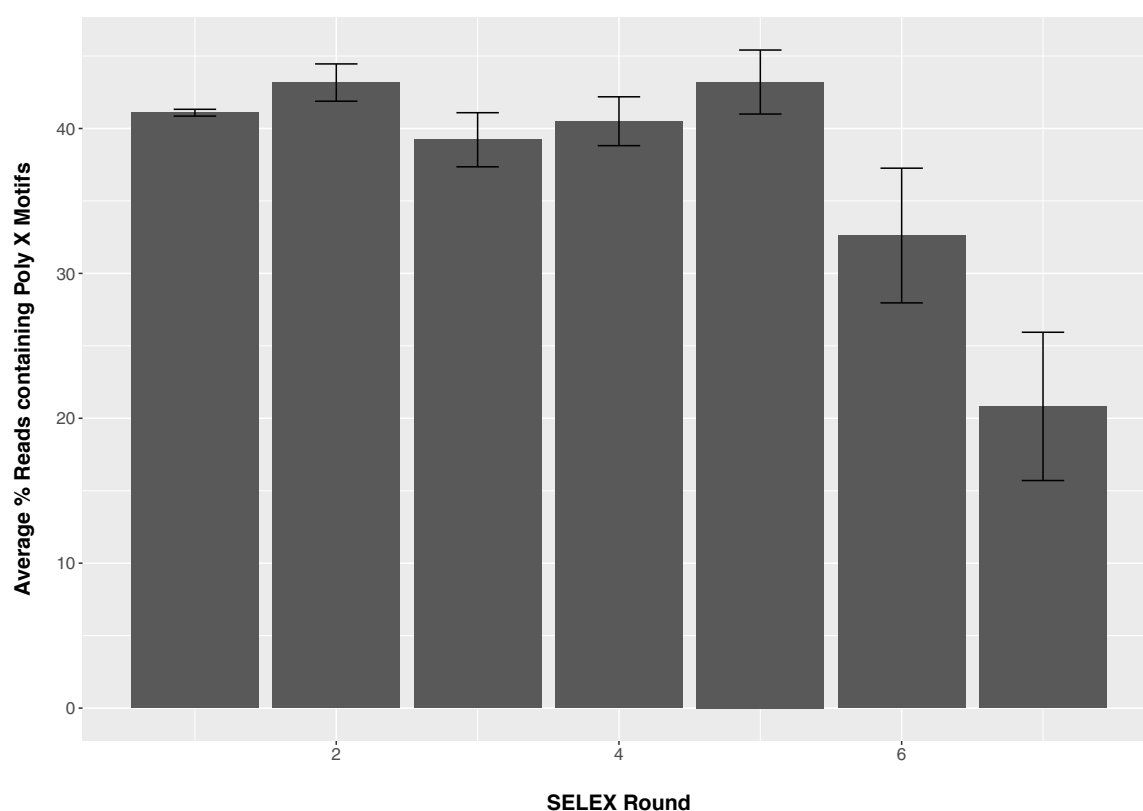
S17	Extra Wash Volume	3	1,401,797	903,079	64%	903,079	64%	662,516	47%	658,928	47%	658,151	47%	658,151
S18	Extra Wash Detergent	3	2,013,188	1,278,569	64%	1,278,569	64%	945,370	47%	935,001	46%	933,908	46%	933,908
S19	Negative Selection	3	1,383,235	777,382	56%	777,382	56%	580,869	42%	578,208	42%	577,455	42%	577,455
S20	Counter Selection	3	938,218	496,390	53%	496,390	53%	379,820	40%	378,251	40%	377,475	40%	377,475
S21	Room Temperature	3	1,895,654	1,261,400	67%	1,261,400	67%	904,565	48%	897,121	47%	892,861	47%	892,861
S22	Standard	4	1,176,186	760,007	65%	760,007	65%	559,085	48%	555,069	47%	554,203	47%	554,203
S23	Standard + Mutation	4	401,449	250,810	62%	250,810	62%	186,370	46%	185,264	46%	185,007	46%	185,007
S24	Extra Wash Volume	4	474,829	288,041	61%	288,041	61%	218,776	46%	217,583	46%	217,317	46%	217,317
S25	Extra Wash Detergent	4	1,265,845	681,993	54%	681,993	54%	519,559	41%	514,770	41%	513,999	41%	513,999
S26	Negative Selection	4	1,652,596	1,029,840	62%	1,029,840	62%	745,331	45%	736,991	45%	735,923	45%	735,923
S27	Counter Selection	4	433,484	229,377	53%	229,377	53%	171,934	40%	170,057	39%	169,664	39%	169,664
S28	Room Temperature	4	195,447	116,546	60%	116,546	60%	86,628	44%	85,527	44%	85,259	44%	85,259
S29	Standard	5	2,096,774	1,367,305	65%	1,367,305	65%	846,039	40%	794,790	38%	748,020	36%	748,020
S30	Standard + Mutation	5	1,218,614	657,766	54%	657,766	54%	510,990	42%	505,497	41%	503,422	41%	503,422
S31	Extra Wash Volume	5	780,883	425,493	54%	425,493	54%	330,598	42%	328,240	42%	327,498	42%	327,498
S32	Extra Wash Detergent	5	1,587,646	886,360	56%	886,360	56%	674,256	42%	669,315	42%	668,036	42%	668,036
S33	Negative Selection	5	971,861	508,925	52%	508,925	52%	389,840	40%	385,963	40%	385,219	40%	385,219
S34	Counter Selection	5	2,170,562	1,103,583	51%	1,103,583	51%	836,762	39%	828,762	38%	827,009	38%	827,009
S35	Room Temperature	5	3,940,147	2,555,033	65%	2,555,033	65%	1,886,870	48%	1,871,211	47%	1,868,132	47%	1,868,132
S36	Standard	6	1,500,719	900,406	60%	900,406	60%	663,673	44%	655,481	44%	652,325	43%	652,325
S37	Standard + Mutation	6	959,412	543,105	57%	543,105	57%	412,154	43%	409,693	43%	408,987	43%	408,987
S38	Extra Wash Volume	6	416,571	216,878	52%	216,878	52%	163,389	39%	160,654	39%	160,264	38%	160,264
S39	Extra Wash Detergent	6	542,051	379,423	70%	379,423	70%	219,245	40%	186,039	34%	184,171	34%	184,171
S40	Negative Selection	6	1,216,965	834,543	69%	834,543	69%	405,181	33%	360,892	30%	354,315	29%	354,315
S41	Counter Selection	6	911,016	702,531	77%	702,531	77%	365,276	40%	246,306	27%	242,568	27%	242,568
S42	Room Temperature	6	1,844,231	1,611,063	87%	1,611,063	87%	535,374	29%	415,592	23%	407,422	22%	407,422
S43	Standard	7	1,443,008	1,232,629	85%	1,232,629	85%	671,322	47%	593,597	41%	565,708	39%	565,708

S44	Standard + Mutation	7	962,791	646,366	67%	646,366	67%	440,171	46%	430,747	45%	430,297	45%	430,297
S45	Extra Wash Volume	7	1,281,320	777,912	61%	777,912	61%	548,301	43%	543,524	42%	542,842	42%	542,842
S46	Extra Wash Detergent	7	822,125	569,704	69%	569,704	69%	380,919	46%	370,371	45%	368,180	45%	368,180
S47	Negative Selection	7	4,622,789	3,910,035	85%	3,910,035	85%	1,400,606	30%	1,045,659	23%	986,787	21%	986,787
S48	Counter Selection	7	1,752,287	1,711,679	98%	1,711,679	98%	142,286	8%	72,050	4%	68,172	4%	68,172
S49	Room Temperature	7	3,605,788	3,224,425	89%	3,224,425	89%	673,388	19%	559,218	16%	531,239	15%	531,239

The number of reads for each round was highly variable due to the different selection pressures on each library, but ranged from approximately 190,000 up to 4.622 million reads. The number of reads removed due to various processing steps was also highly variable. In some cases, contamination with TES07 and R18C1 was highly prevalent. In most of the libraries, approximately 40% of reads were retained after all processing steps were completed. The worst case of contamination was in library S48 (Round 7 of the Counter Selection strategy) where only 4% of the raw reads remained after all processing steps were completed. There was a decrease in retained reads in Rounds 6 and 7 of selection, as contamination seemed to be more prevalent in these rounds.

### 6.3.5 PolyX Motif Occurrence

Initial exploratory analyses demonstrated that the random starting library supplied by Thermo Fisher Scientific was significantly biased towards thymine (T) nucleotides. Further exploration of the data revealed the presence of a large number of reads containing polyT motifs (*i.e.* reads of ten or more consecutive thymines). Figure 6.15 illustrates the average number of polyX motifs which were detected at each selection round across the seven different selection strategies.



**Figure 6.15: The average number of polyX motifs across the seven selection rounds.**

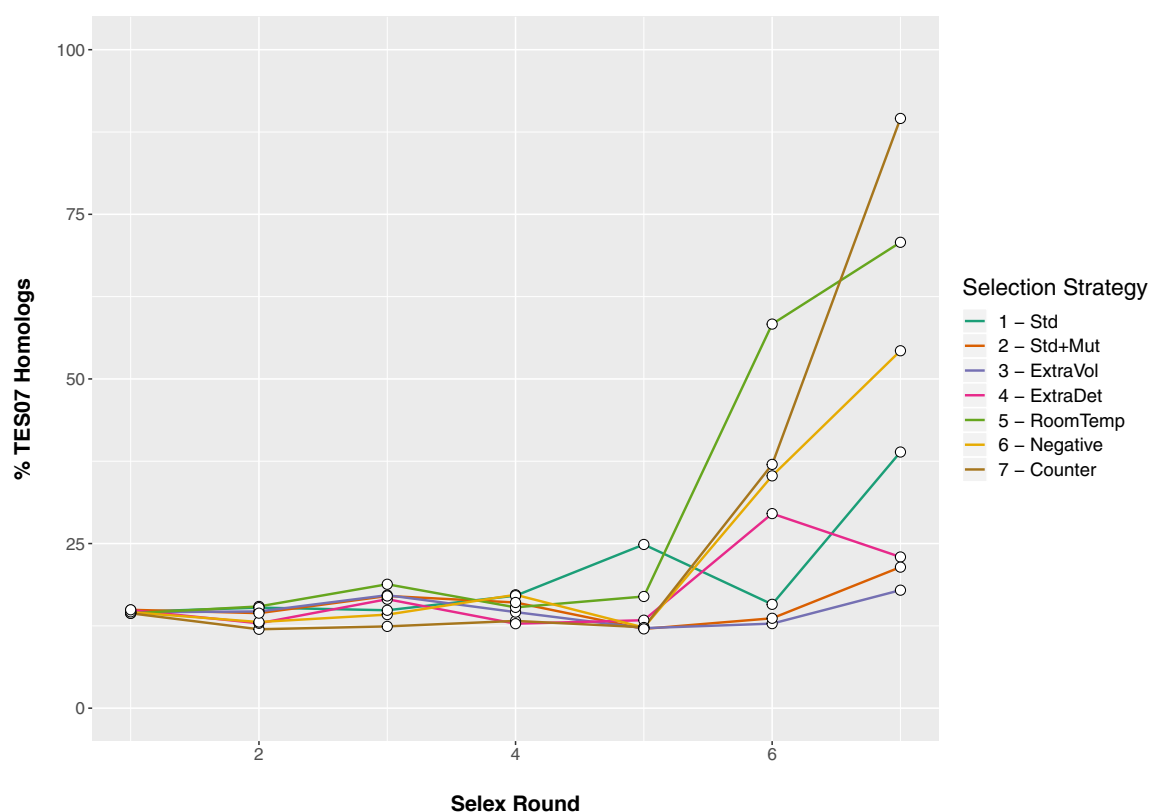
After one round of selection, approximately 41% of the reads contained polyX motifs. These motifs were pervasive within the oligonucleotide libraries during selection and only decreased in Selection Rounds 6 (33%) and 7 (21%). The error bars were much larger in later selection rounds, which was to be expected as the libraries diverge due to differing selection

pressures within each of the strategies tested. It appears that these different selection pressures did not initially influence the numbers of polyX motifs within the respective library as the proportions of polyX motifs remains relatively stable from Selection Rounds 1 to 5.



### 6.3.6 Counts for TES07 Homologs

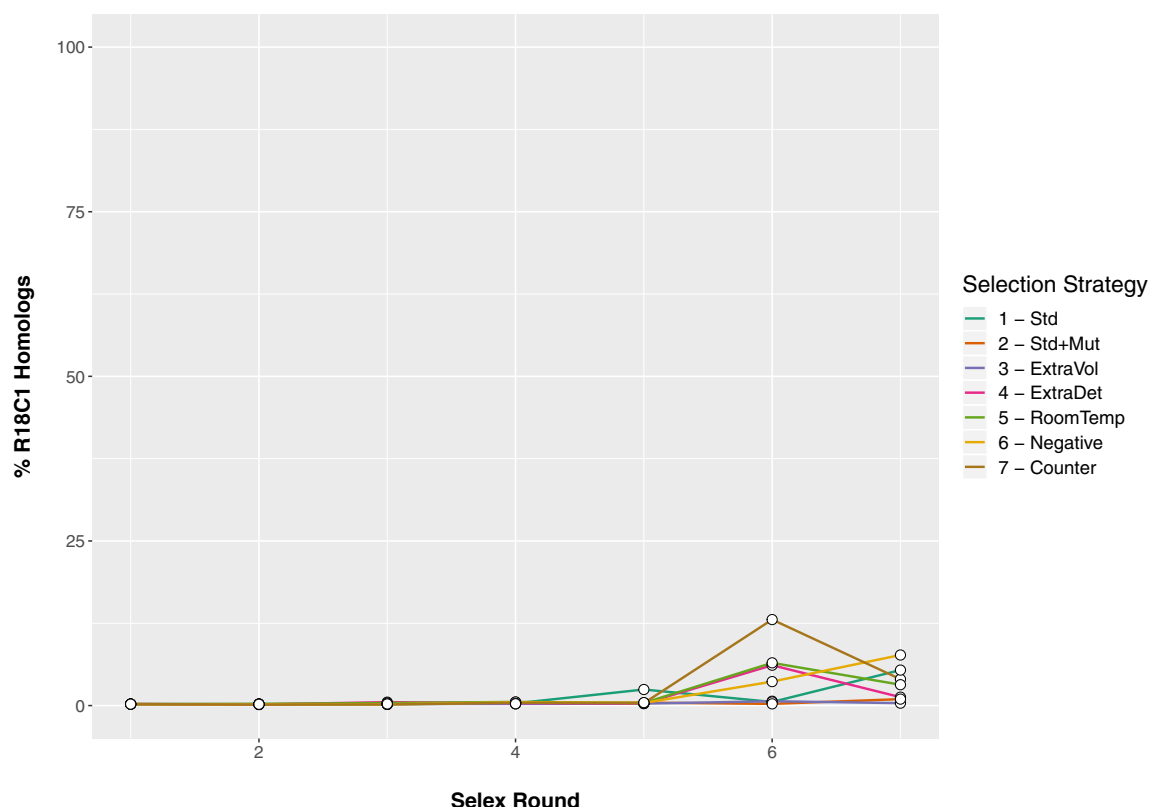
The percentage of reads that contained sequences with a 50% or higher sequence similarity to the TES07 aptamer increased as SELEX progressed (Figure 6.16). Selection Strategies 5, 6 and 7 had the highest proportion of TES07 homologs in the last two selection rounds. After the first selection round, libraries from all strategies contained TES07 homologs at a frequency of >10% of reads, indicating that contamination occurred early in the selection process. In the final selection round of the Counter Selection strategy, the frequency of TES07 homologs was 89.56%. All reads containing TES07 homologs were discarded before further analyses.



**Figure 6.16: Proportion of TES07 homologs (50% sequence identity cut-off) in each selection strategy across all selection rounds.**

### 6.3.7 Counts for R18C1 Homologs

R18C1 is an oestradiol-binding aptamer commonly used in our laboratory. As R18C1 uses the same flanking primers as the libraries used in this HT-SELEX experiment, it was pertinent to check for R18C1 homolog contamination within the dataset (Figure 6.17).



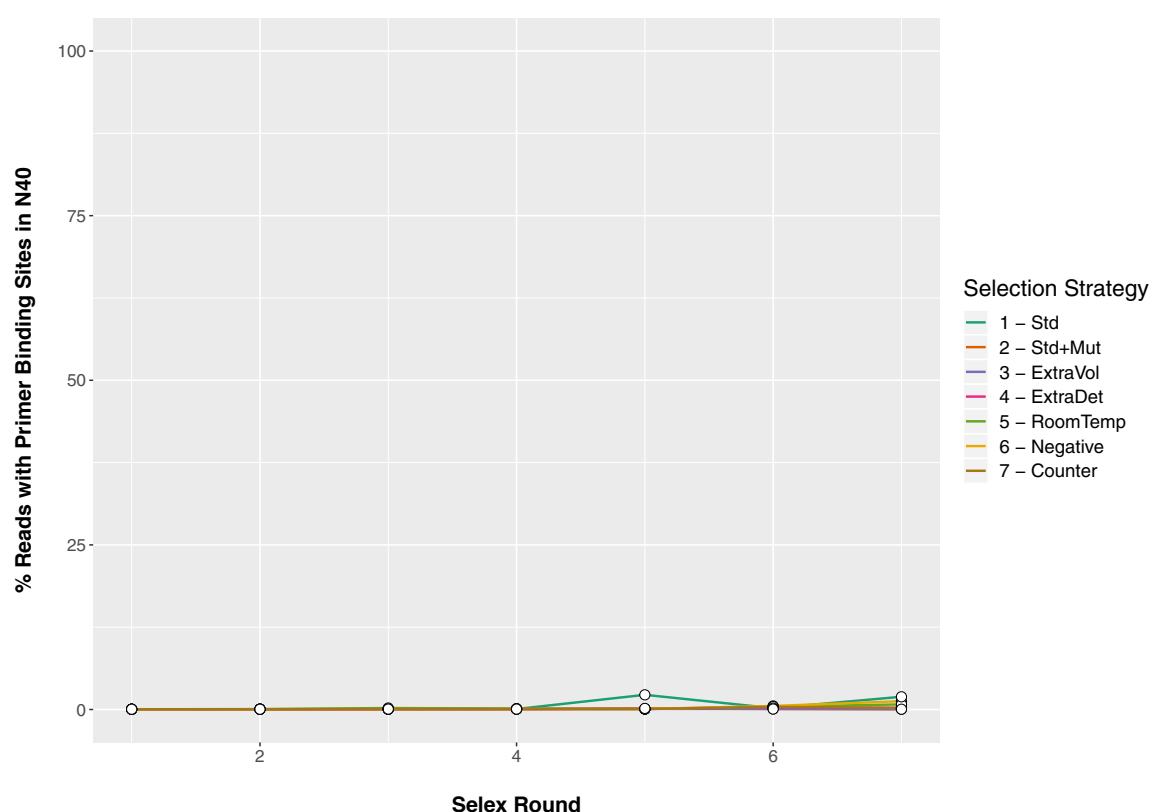
**Figure 6.17: Proportion of R18C1 homologous sequences (50% identity cut-off) in each selection strategy across seven selection rounds.**

The prevalence of R18C1 homologs was relatively low compared to TES07. The highest frequency of 13.06% of reads was recorded in selection round 6 of the Counter Selection strategy. The frequency of homologs increased from nearly zero to >5% of reads in Selection Round 6 for the Extra Detergent, Room Temperature and Counter Selection strategies. The frequency decreased from Rounds 6 to 7 in all of these strategies. For the Negative Selection

strategy, a linear increase in frequencies were observed from Rounds 5 to 7. All reads containing R18C1 homologs were discarded before further analyses of the data.

#### 6.3.8 Number of Reads with Primer-Binding Sites in the N40 Region

Analyses of the starting library composition (prior to any selection) revealed that there were reads within the starting library that contained primer binding sites that were 100% homologous to the binding sites in the flanking region. It was assumed that this would result in truncated PCR by-products. The proportion of reads containing these binding sites was determined for each selection strategy across all selection rounds (Figure 6.18).

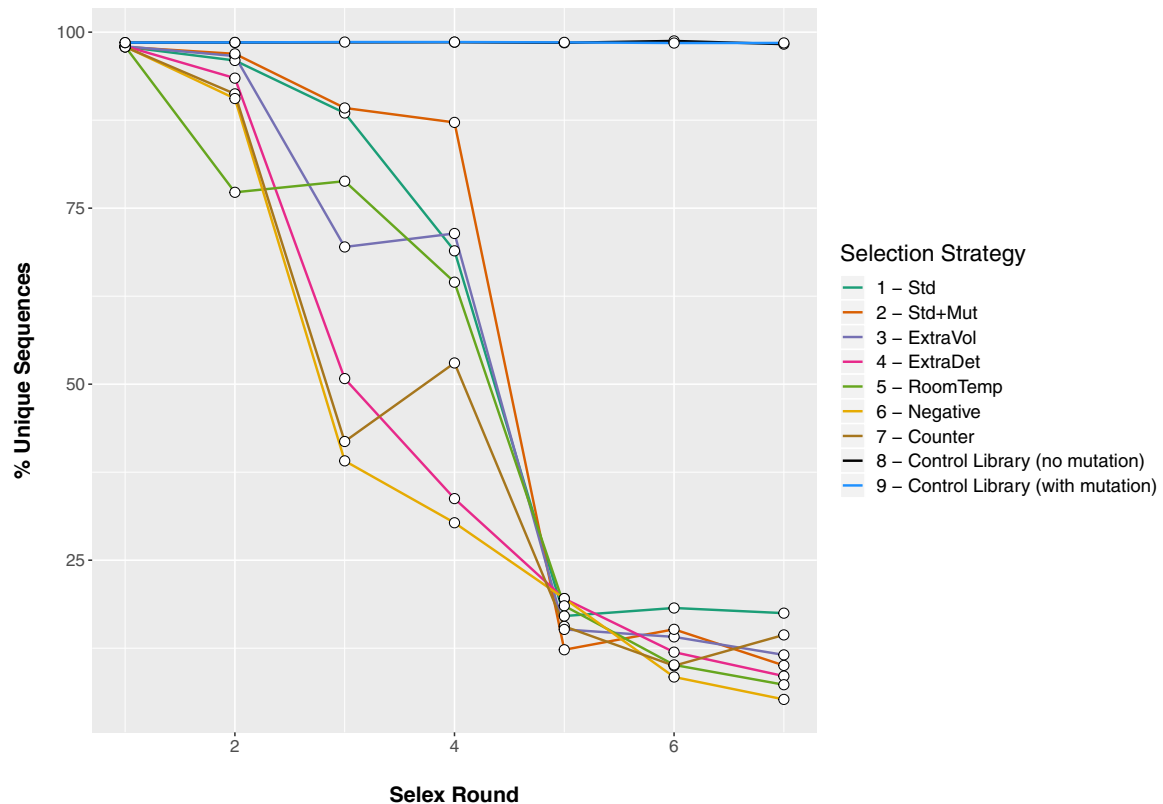


**Figure 6.18: Proportion of reads that contained primer binding sites in the N40 region in each selection strategy across the seven selection rounds**

The proportion of reads containing primer binding sites was < 2% in all strategies for all rounds with one exception. In Selection Round 5 for Standard strategy, there was a small increase in reads containing primer binding sites in both Rounds 5 and 7. All reads containing primer binding sites were removed from the dataset.

#### 6.3.9 Number of Unique Sequences at each Selection Round

Selection occurring within a population for a specific trait may be indicated by a change in population diversity. The strategies that employed the strongest selection pressures (*i.e.*, Extra Detergent, Negative Selection and Counter Selection strategies) displayed a more rapid decrease in diversity than other strategies. Figure 6.19 illustrates that the libraries subjected to these two selection strategies experienced a large decline in diversity, as represented by a decrease in the number of unique sequences, at the completion of Selection Round 3.

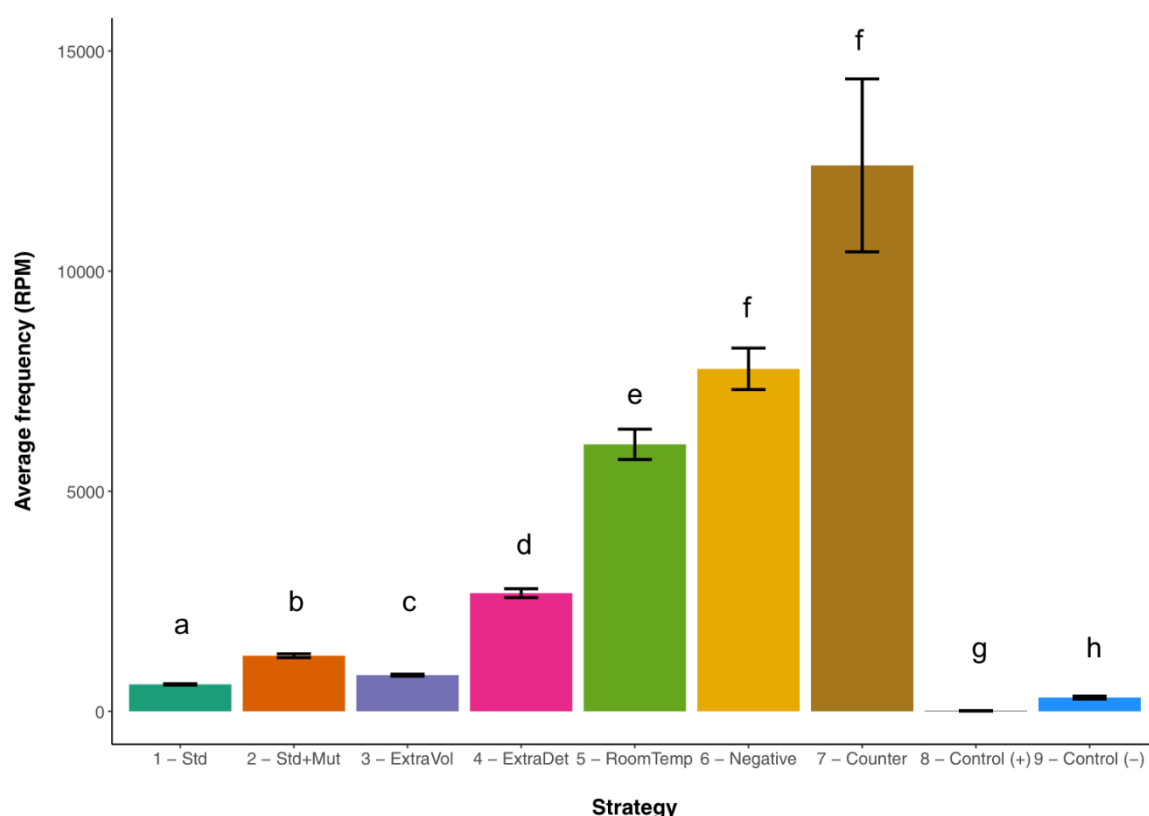


**Figure 6.19:** The proportion of unique sequences present in each library at each selection round. Control (+) and control (-) refer to the polyclonal control libraries with induced mutation (+) and no additional mutation (-).

The Standard + Mutation strategy maintained a high diversity for the longest period before the number of unique sequences declined from Rounds 4 to 5. This was expected as this strategy had the weakest selection pressures and increased mutations, thus enabling a longer maintenance of diversity. After seven rounds of selection, libraries subjected to all strategies experienced a large decrease in diversity from > 95% to 10 - 20% of sequences being unique. This is an indication that selection occurred in all strategies tested herein. Within the two control libraries (subjected to PCR but no additional selection pressure) there was virtually no decrease in the number of unique sequences from rounds 1 to 7.

### 6.3.10 Average frequency of Top 50 candidates in Round 7

The abundance of sequences in the final selection round is the traditional way in which candidates are selected for further characterisation. All selection strategies revealed a difference in the normalised frequency of sequences in the final selection round (Figure 6.20). There were statistically significant differences between group means as determined by Welch's ANOVA ( $F(8,182.72) = 3697.11$ ,  $p < 0.01$ ). Post hoc comparisons using the Games Howell test identified significant pairwise differences in the mean frequency between all strategies ( $p < 0.05$ ), except Negative and Counter ( $p = 0.518$ ).



**Figure 6.20: Average frequency of top 50 candidates in the final SELEX round. Control (+) and control (-) refer to the polyclonal control libraries with induced mutation (+) and no additional mutation (-).**

The most stringent strategy (Counter Selection) had the highest average frequency within the top 50 sequences. The least stringent strategy (Standard) had the lowest average frequency. This result is congruent with our hypothesis that as the oligonucleotide populations are subjected to stronger selection pressures, candidates become more abundant. Oligonucleotides within the control libraries, which were amplified with PCR but had no additional selection pressure and weren't incubated with the target had a lower average frequency than all libraries from the different selection strategies.

### 6.3.11 Relatedness of Top 50 Sequences (frequency) from Each Strategy

The relatedness of sequences from different selection strategies was analysed by determining if the same sequences appeared in the multiple selection strategies. Table 6.4 displays the percentage of identical sequences which were shared between different strategies in a pairwise manner.

**Table 6.4: The percentage of sequences of the top 50 from Round 7 (based on frequency) which appear in pairwise combinations of strategies.**

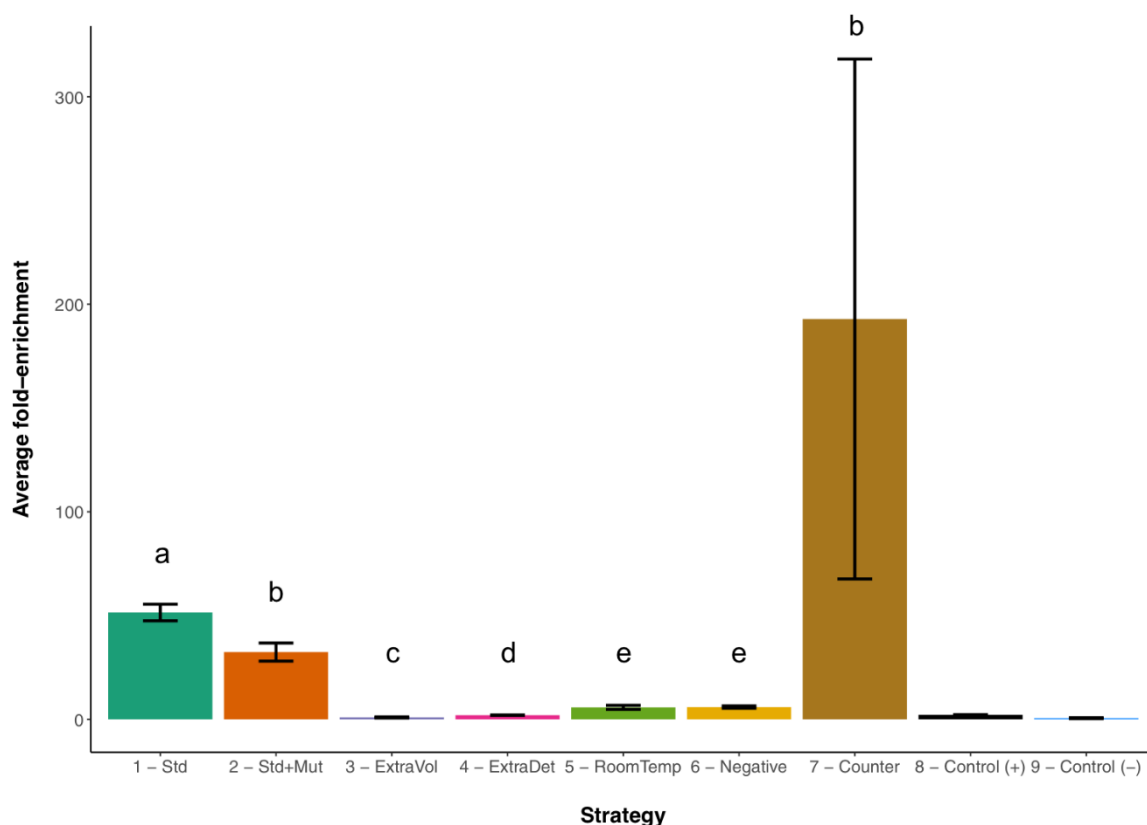
Strategy	Standard	Standard + Mutation	Extra Volume	Extra Detergent	Room Temperature	Negative	Counter
Standard							
Standard + Mutation	0%						
Extra Volume	0%	0%					
Extra Detergent	0%	0%	0%				
Room Temperature	0%	0%	0%	0%			
Negative	4%	0%	0%	0%	0%		
Counter	4%	0%	0%	0%	0%	4%	

There are very few common sequences across the multiple selection strategies. Two sequences from the standard strategy are also abundant in the negative and counter selection strategies. In addition, two abundant sequences in the negative selection strategy also appear in the top 50 sequences of the counter selection strategy. The same two sequences are shared across all three strategies (standard, negative selection and counter selection) and the two sequences are only one nucleotide different to each other (a single deletion). The presence of these sequences in multiple strategies could indicate a significant binding advantage or alternatively that this sequence is parasitic and is able to be favourably amplified in PCR.

#### 6.3.12 Longitudinal Enrichment of Top 50 candidates in Selection Rounds 6 and 7

Cycle-to-cycle enrichment has been demonstrated in previous studies as a more accurate predictor of binding propensity, than absolute frequency. The average cycle-to-cycle enrichment of the top 50 candidates from Rounds 6 to 7 is shown in Figure 6.21. There was a statistically significant difference between group means of the seven selection strategies as determined by Welch's ANOVA ( $F(6,149.14) = 482.27, p < 0.01$ ). Post hoc comparisons using the Games Howell test determined that there were pairwise differences in mean enrichment between all strategies ( $p < 0.05$ ) except Standard + Mutation/Counter ( $p = 1.00$ ) and Room Temp/Negative ( $p = 0.536$ ). Post hoc comparisons were not completed between the seven selection strategies and the two control libraries as the control libraries had some candidates with enrichment values of 'zero' meaning these could not be log-transformed.





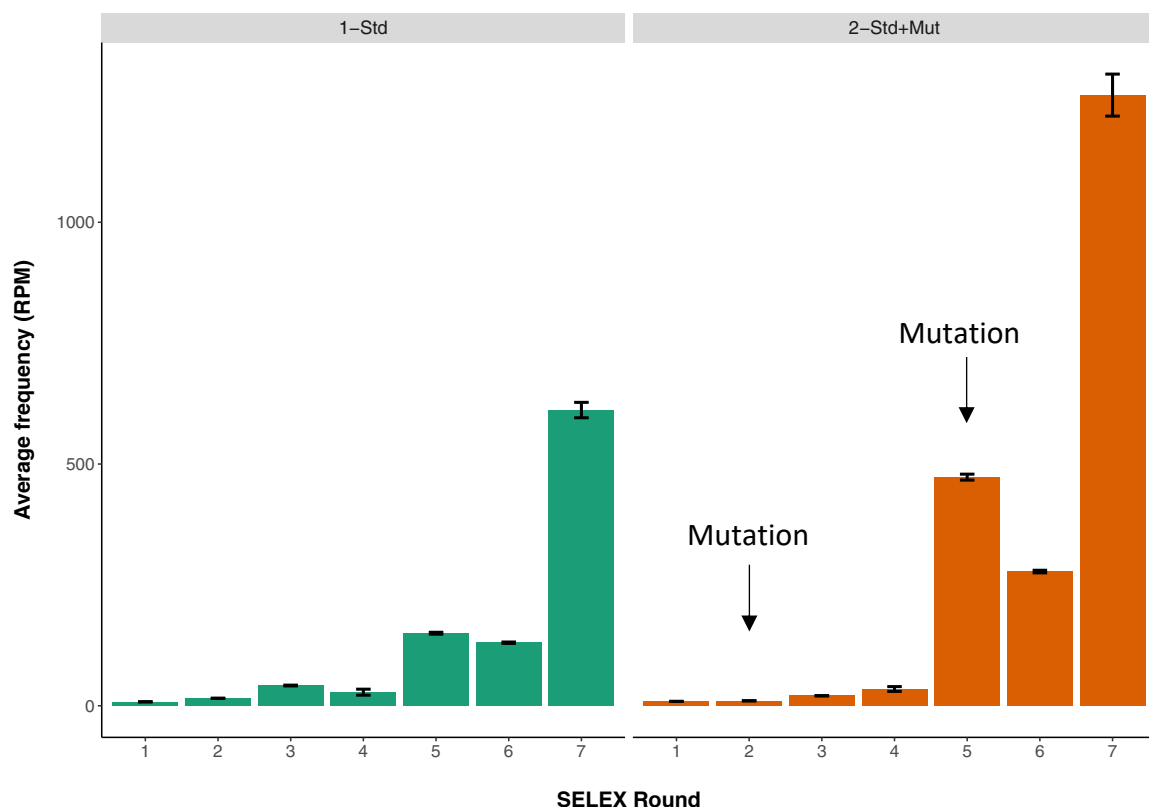
**Figure 6.21: Average enrichment (fold-change) of the top 50 candidates from Selection Rounds 6 to 7. Control (+) and control (-) refer to the polyclonal control libraries with induced mutation (+) and no additional mutation (-).**

The most stringent selection strategy (Counter Selection) displayed the highest cycle-to-cycle enrichment. The large standard error associated with this strategy possibly indicates that a small number of candidates within the top 50 are highly enriched, while most candidates are not. Interestingly, the least stringent strategy (Standard) has the next highest enrichment from Rounds 6 to 7, whilst the Extra Volume, Extra Detergent, Room Temperature and Negative Selection strategies have the lowest. The control libraries (polyclonal libraries which were subject to PCR with same conditions as libraries under selection but with no additional selection pressure) showed virtually no enrichment.

### 6.3.13 Frequency of Top 50 Candidates through SELEX Rounds 1 to 7

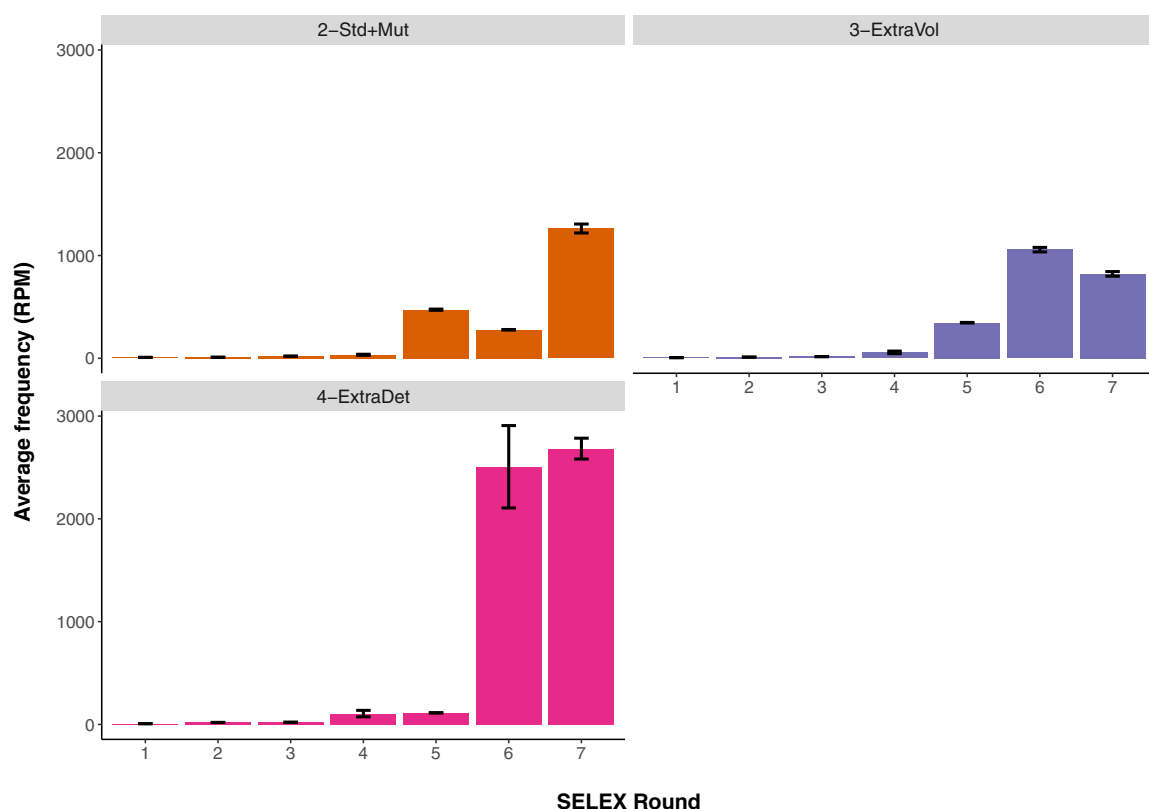
Although sequence abundance in the final round is the most commonly used criteria for selecting aptamer candidates, HT-SELEX allows for the assessment of trends throughout the selection process.

To assess these trends, the average frequency of the top 50 candidates, ranked by reads per million, was graphed at each selection round with specific selection events annotated. As each selection strategy differed from the previous strategy by one common methodological modification, graphing these strategies side by side illustrates the effects of that change on the abundance of the top candidates. Figure 6.22 shows the frequency trends for the Standard and Standard + Mutation strategies. The effect of inducing mutation can be observed when comparing the frequency trends between these two strategies. Mutation was induced in Rounds 2 and 5 as indicated above. There was an apparent higher frequency of sequences in the strategy where mutation was used, indicating that it may be beneficial for selecting higher affinity aptamers.



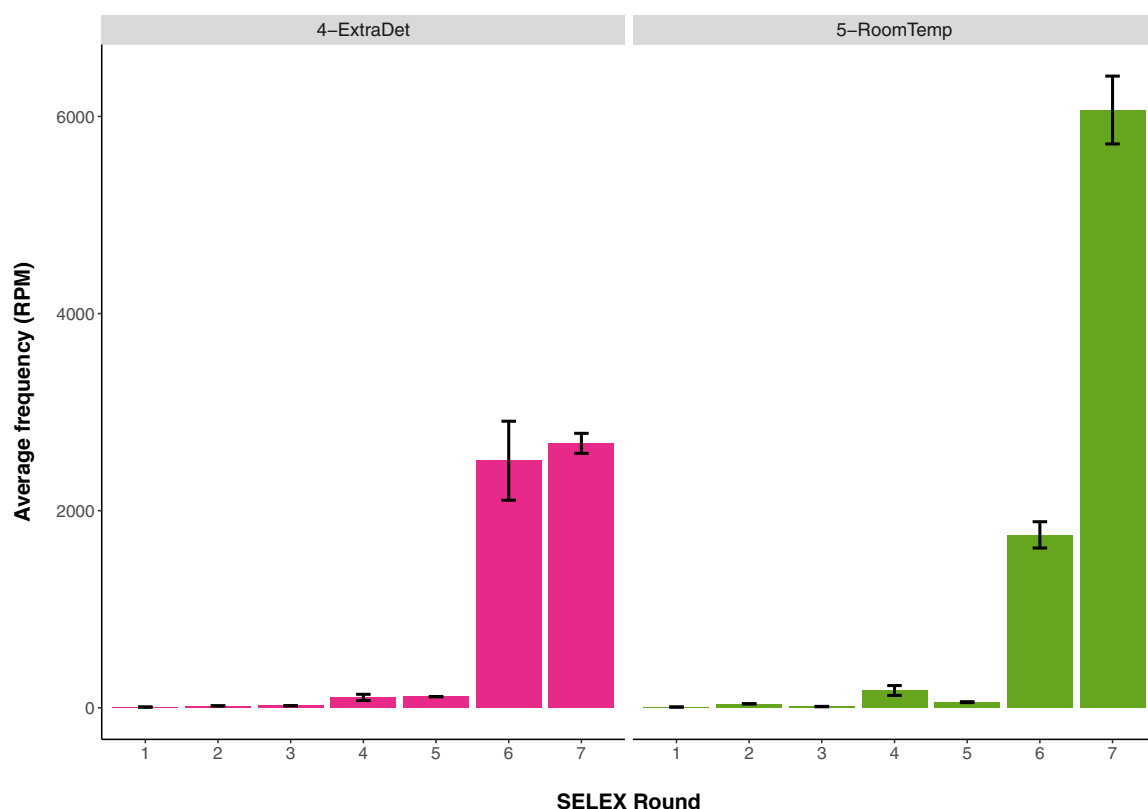
**Figure 6.22: Frequency of the top 50 candidates for the Standard and Standard plus Mutation strategies for all selection rounds.**

The effect of increasing the wash stringency via two different methods, *i.e.* iteratively increasing the washing volume and detergent concentration, is illustrated in Figure 6.23. Increasing the wash volume reduced the abundance of top 50 candidates as evidenced by a lower abundance in the final selection round of the Extra Volume strategy, compared to the Standard + Mutation strategy. Using a higher detergent concentration when partitioning aptamers from beads, as in the Extra Detergent strategy, increased the abundance of the top 50 candidates in Selection Rounds 6 and 7 compared to the Standard + Mutation strategy.



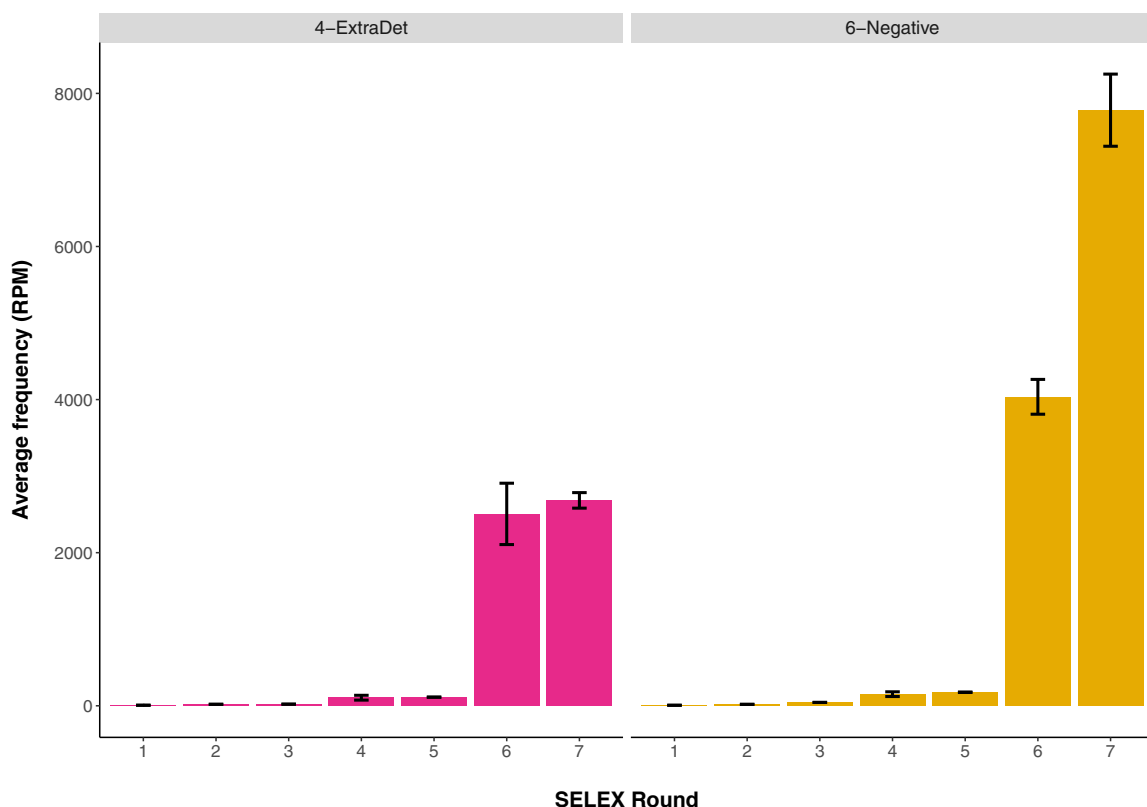
**Figure 6.23: Frequency of the top 50 candidates for the Standard + Mutation, Extra Volume and Extra Detergent strategies for all selection rounds.**

Incubating the aptamer and target at room temperature for a duration of four hours as opposed to an overnight incubation at 4°C increased the frequency of the top 50 candidates. in the final selection round as illustrated in Figure 6.24.



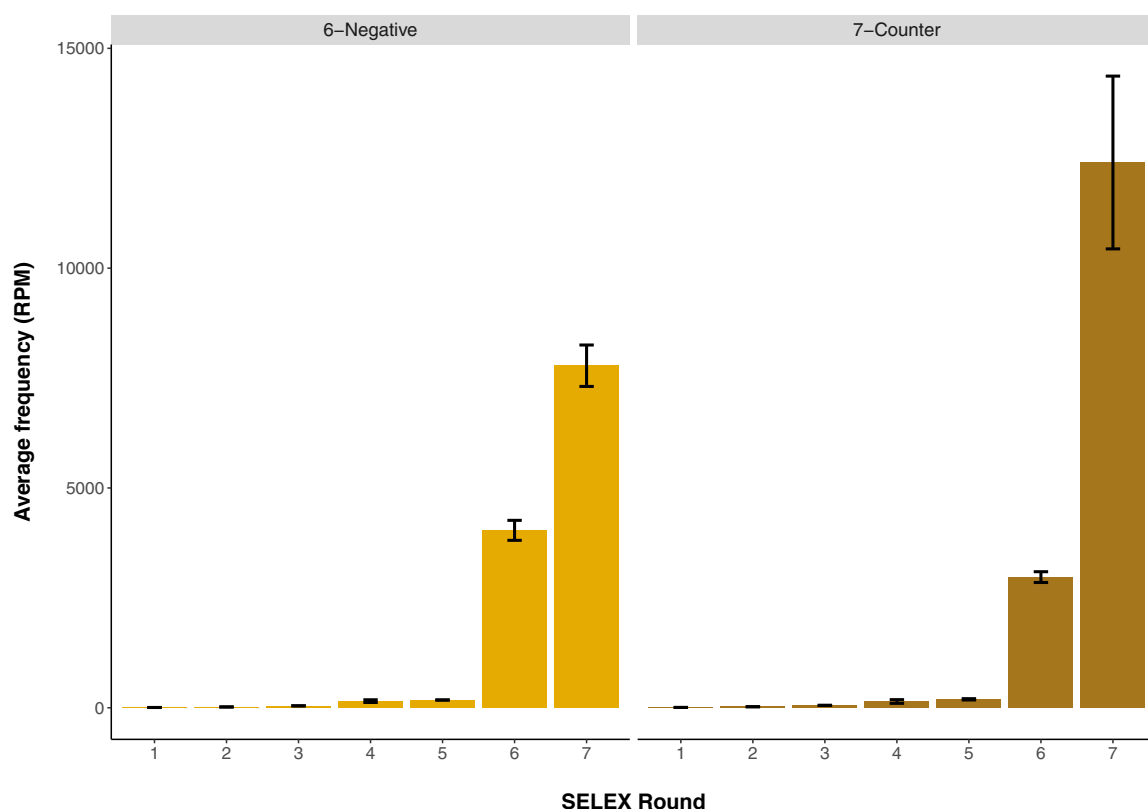
**Figure 6.24: Frequency of the top 50 candidates for the Extra Detergent and Room Temperature strategies for all selection rounds.**

Using negative selection to remove sequences that specifically bind to the control matrix also influenced unique sequence abundance. Figure 6.25 shows a comparison of the Negative Selection strategy with the Extra Detergent strategy, which had the same conditions but without the negative selection step. Sequences in the Negative Selection strategy at Rounds 3 and 6 had a higher average frequency in Selection Rounds 6 and 7, than those in the strategy without negative selection.



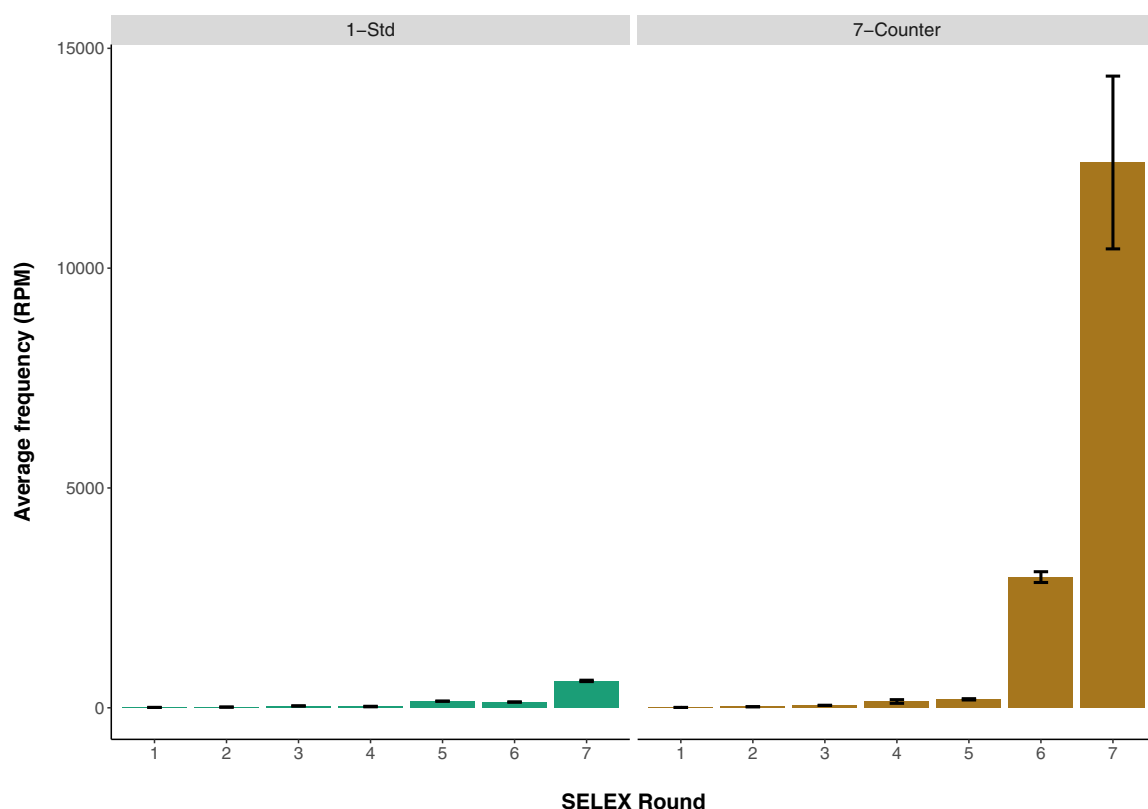
**Figure 6.25: Frequency of the top 50 candidates for the Extra Detergent and Negative Selection strategies for all selection rounds.**

Using counter selection against molecules with a similar structure to oxybenzone in the final selection round also increased the abundance of sequences compared to the previous strategy that did not contain this counter selection step. Figure 6.26 shows a comparison between the Counter Selection and Negative Selection strategies. There was an increase in the abundance of sequences in Round 7 which occurred immediately following the counter selection step in the Counter Selection strategy. It can be hypothesized that this counter selection step removed binders that exhibited weak to moderate affinity to oxybenzone, reducing the size of the population and enriching the sequences that exhibit stronger binding properties to the target molecule.



**Figure 6.26: Frequency of the top 50 candidates for the Negative Selection and Counter Selection strategies for all selection rounds.**

Finally, it is useful to look at the cumulative effect of all methodological changes on the frequency of aptamer candidates. There was a ~20-fold increase in the average abundance of sequences in the Counter Selection strategy compared to the Standard strategy (Figure 6.27). This increased enrichment was visible at Round 7 and may be attributed to the cumulative effects of all selection pressures imposed including mutation, increased detergent concentration, negative selection and counter selection.

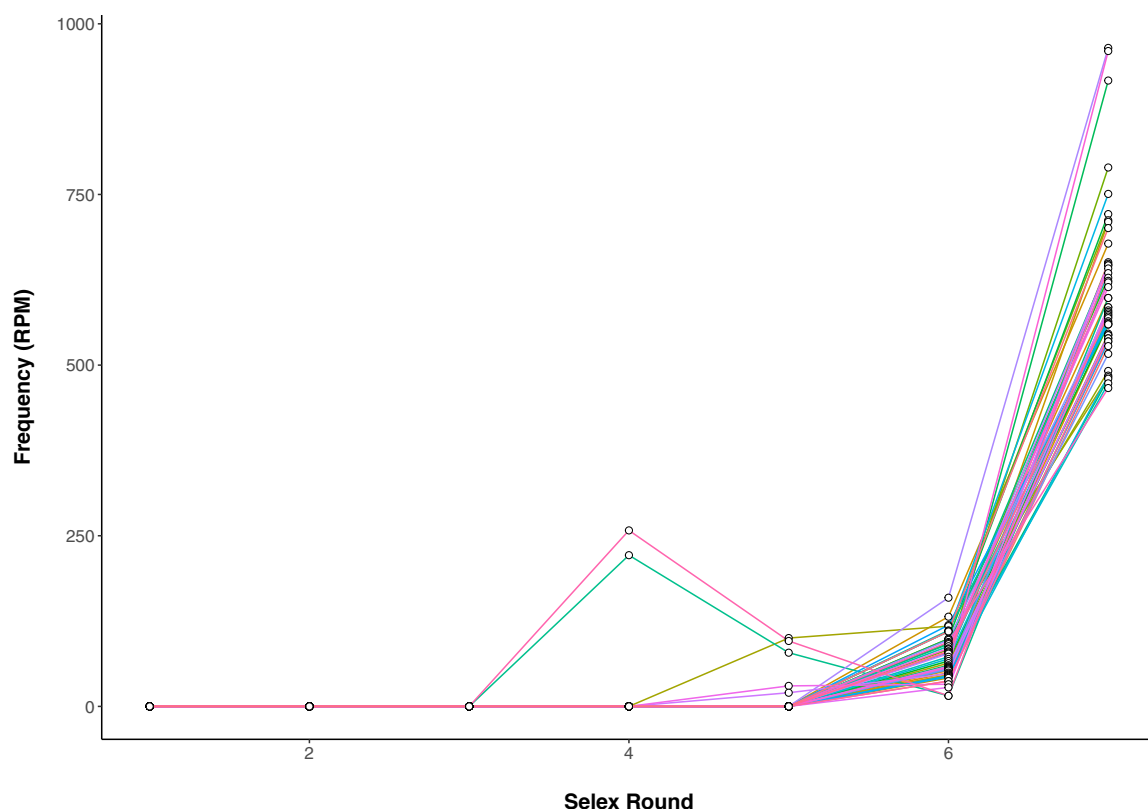


**Figure 6.27: Frequency of the top 50 candidates for the Standard strategy compared to Counter Selection strategy for all selection rounds.**

#### 6.3.14 Frequency of Individual Sequences throughout SELEX

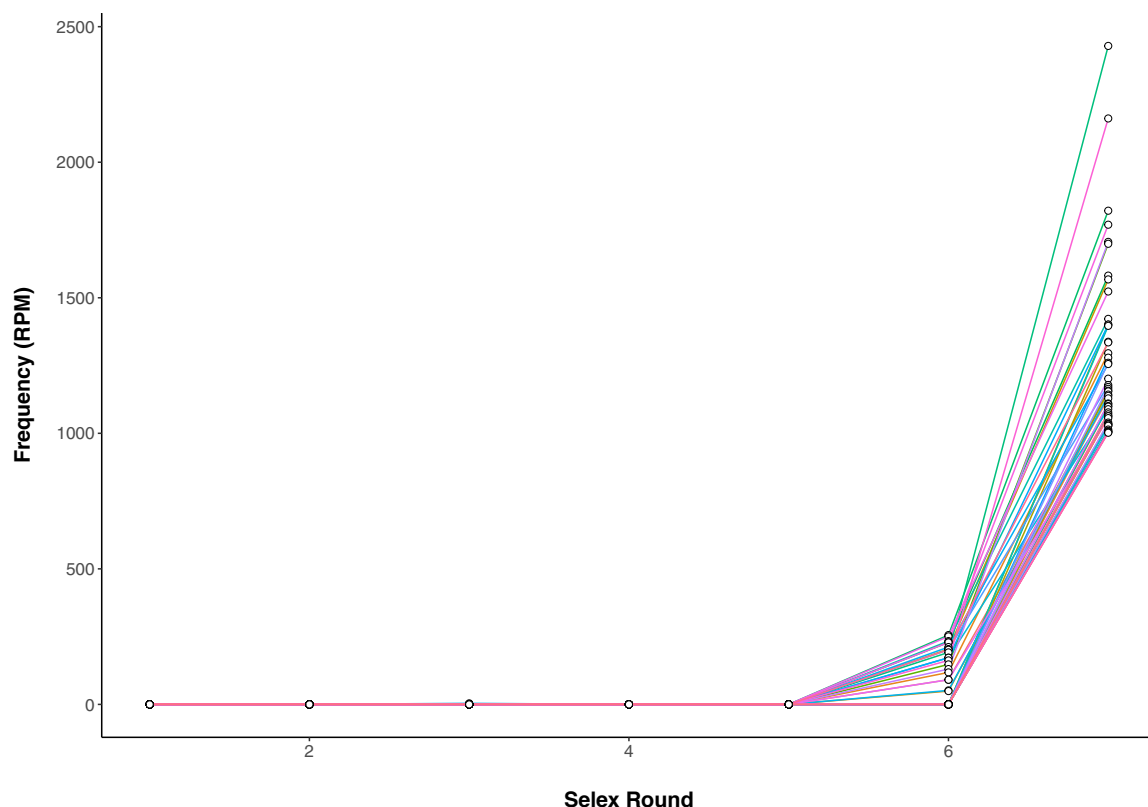
When assessing the HT-SELEX data, the overall frequency trends of individual aptamer candidates were used to select those aptamers to be characterised for binding affinity. Therefore, the fate of each of the top 50 individual candidates from the final selection round for each selection strategy was tracked through all SELEX rounds to determine enrichment trends. This identified if sequences were present at higher frequencies earlier in the SELEX process, but then decreased in later rounds, as has been described in previous literature.





**Figure 6.28:** The frequency of the top 50 candidates from the Standard strategy through all SELEX rounds.

From the Standard strategy library, two sequences were highly represented at 258 and 221 reads per million at Round 4, and then their frequency decreased over the two subsequent selection rounds. Most of the top 50 abundant sequences in Round 7 only become abundant from Round 5. For characterisation, the two sequences with the highest frequency in Round 4 were selected, as well as the highest frequency sequence at Round 7. Under the Standard + Mutation strategy, the fate of the top 50 most abundant sequences showed a similar trend and is displayed in Figure 6.29.

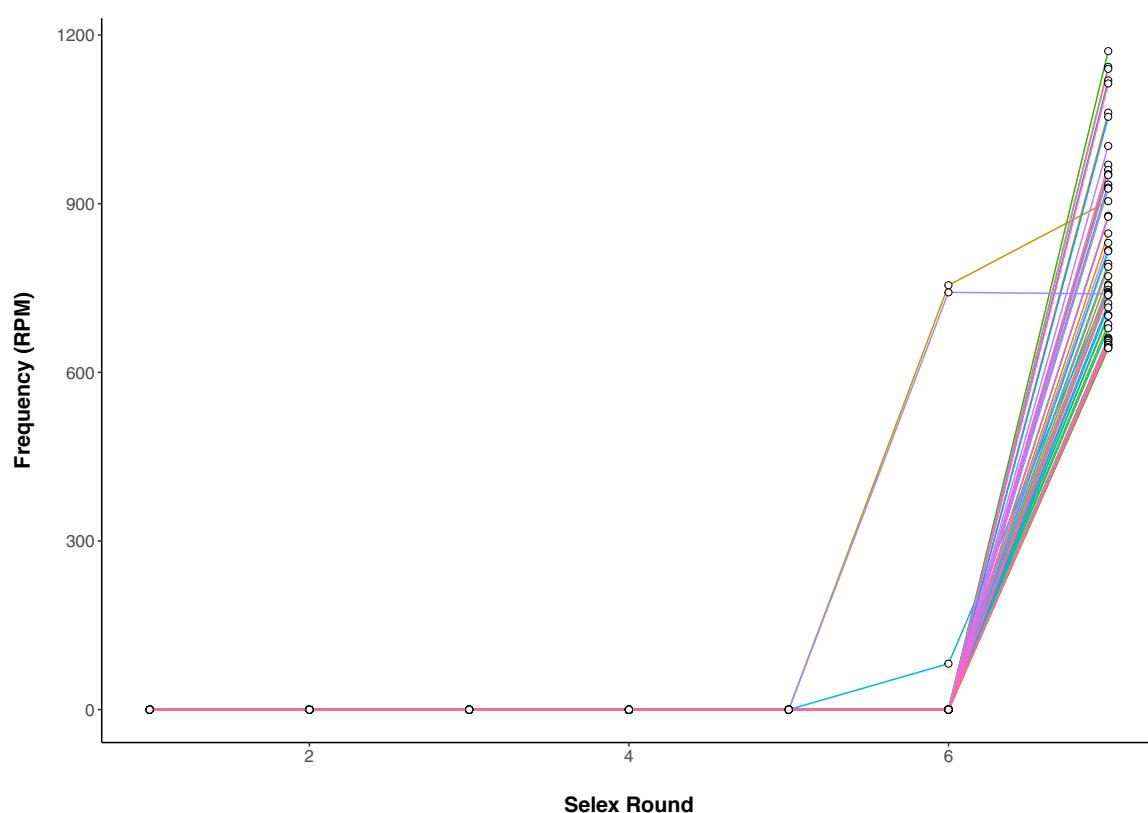


**Figure 6.29: The frequency of the top 50 candidates from the Standard + Mutation strategy through all SELEX rounds.**

Individual sequences were not present at detectable levels in the library until Rounds 5 or 6, and then became abundant during the last selection round. This strategy included a mutation step which may mean that those sequences that were abundant at Round 7 were present at high levels in earlier rounds prior to the mutation step. The sequences that were abundant after Round 7 were tracked back through previous selection rounds. There was no evidence that these abundant Round 7 candidates were present at high frequencies in previous rounds for this strategy. Thus, the most abundant sequence at Round 7 was chosen from this selection strategy for MST characterisation.

The Extra Volume strategy produced interesting results when assessing the fates of individual sequences and are displayed in Figure 6.30. The majority of sequences were undetectable

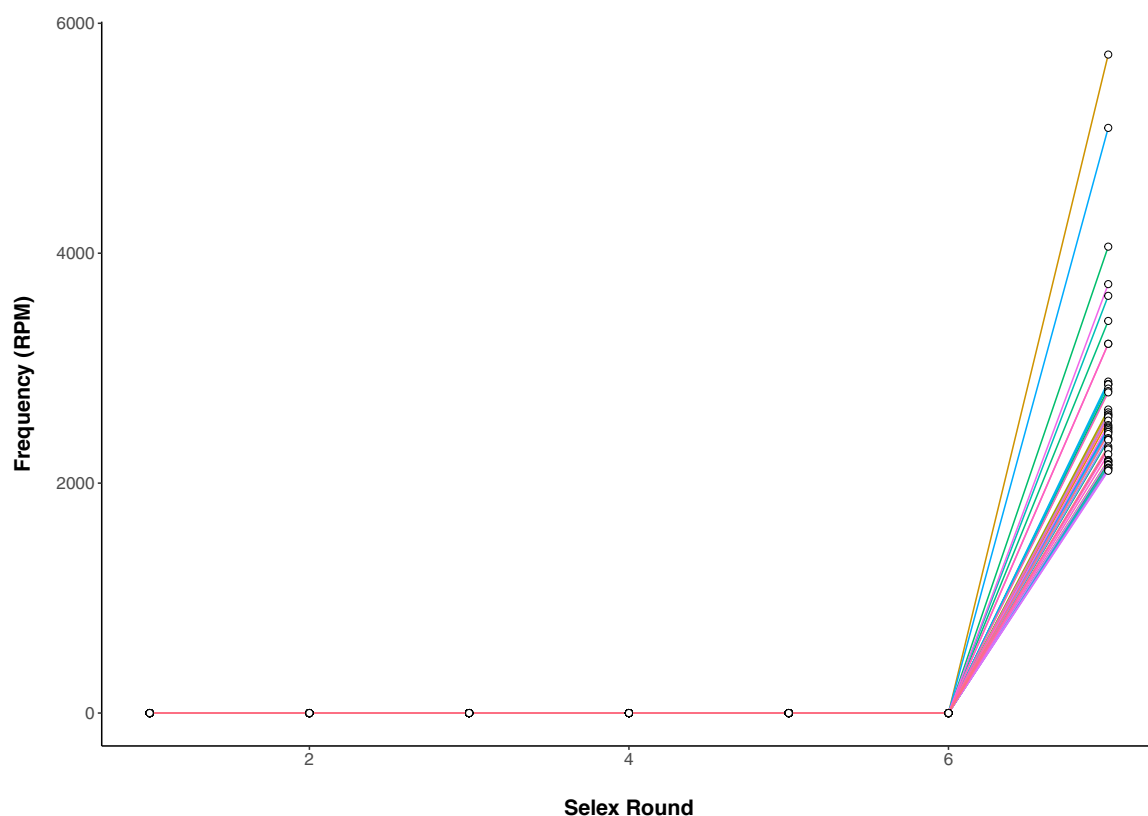
until the final round, however two sequences underwent marked enrichment at Round 6. These two sequences had a frequency of ~750 reads per million and comprised 0.08% of the total population. This highlights the advantages of selecting candidates by enrichment rather than final frequency counts. Neither of these Round 6 sequences had a chance of being identified using the later selection criteria. The most frequent sequence in Round 7 had a frequency of 1171 reads per million and comprised 0.12% of the total population.



**Figure 6.30: The frequency of the top 50 candidates from the Extra Volume strategy through all SELEX rounds.**

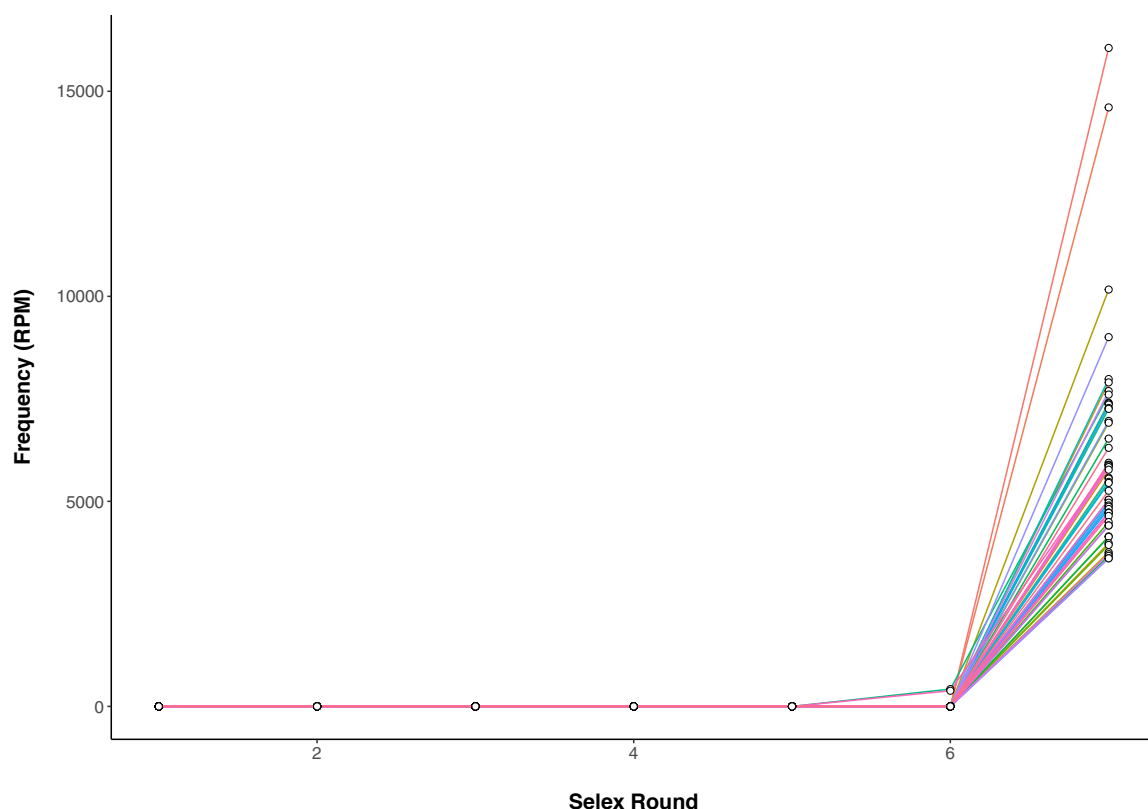
The Extra Detergent strategy should have removed weakly bound sequences and therefore increased the frequency of sequences that bound with a greater affinity to oxybenzone. The frequencies of specific nucleotide sequences were low from Selection Rounds 1 to 6, and

enrichment of these top 50 candidates is only observed at Round 7 (Figure 6.31). The most abundant clone had a frequency of 5,728 RPM and comprised 0.57% of the library.



**Figure 6.31: The frequency of the top 50 candidates from the Extra Detergent strategy through all SELEX rounds.**

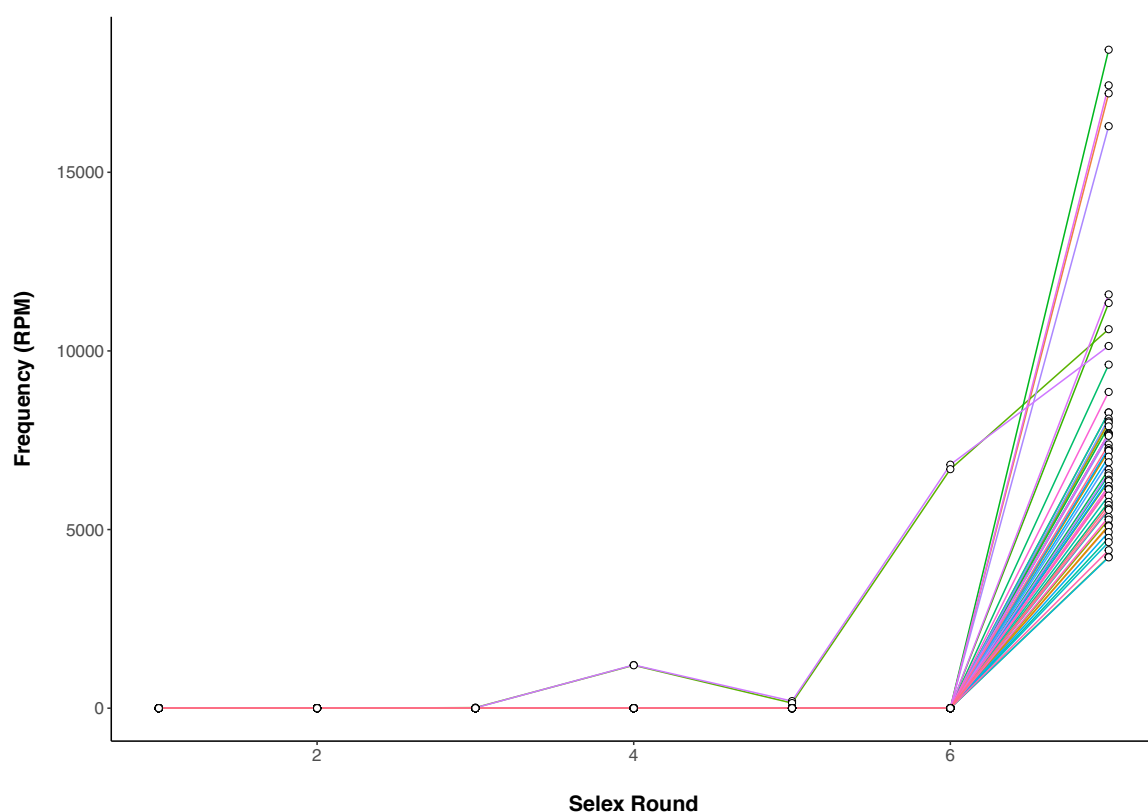
The Room Temperature strategy showed a similar trend to the Extra Detergent strategy, with the exception that minor enrichment was observed for a small number of the top candidates at Selection Round 6 (Figure 6.32). Thus, enrichment was mostly observed in the last round of selection. The most abundant sequence had a frequency of 16,057 RPM and comprised 1.61% of the total library.



**Figure 6.32: The frequency of the top 50 candidates from the Room Temperature strategy through all SELEX rounds.**

Sequences from the Negative Selection strategy show an enrichment trend that correlates with the selection pressures imposed by these conditions. A negative selection step was performed at Rounds 3 and 6 to remove sequences that showed affinity to the control bead matrix. Two sequences show enrichment at Round 4 immediately after the first negative selection step and showed further enrichment following the second negative selection step at Round 6 (Figure 6.33). Interestingly, enriched sequences at Round 4 showed a decreased abundance at Round 5, which was when mutation PCR was introduced. All other sequences exhibited a marked enrichment from Round 6 to 7. It can be hypothesized that the negative selection step in Round 6 was sufficient to remove a large number of non-binders or

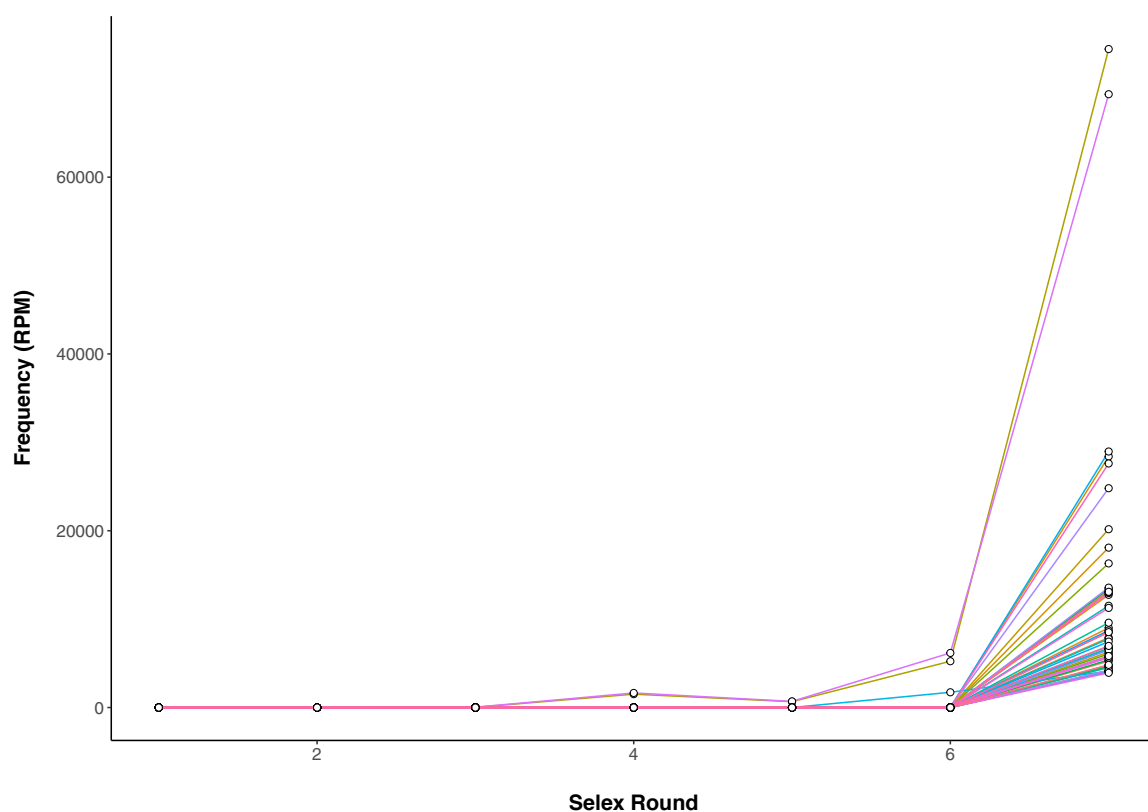
sequences with weak affinity to oxybenzone and this resulted in increased enrichment of successful sequences in the last round. The top candidate in this selection strategy had a frequency of 18,431 RPM and comprised 1.84% of the library.



**Figure 6.33: The frequency of the top 50 candidates from the Negative Selection strategy through all SELEX rounds.**

The Counter Selection strategy results in the highest average frequency in Round 7. Similar to the Standard and Negative Selection strategies, two sequences show enrichment at Round 4 (Figure 6.34). The same two sequences also have the highest enrichment, and the highest overall abundance, in Round 7. The majority of the remaining sequences were undetectable

until Round 7. The two top sequences had frequencies of 74,490 and 69,381 RPM and comprised 7.45% and 6.94%, respectively of the library.



**Figure 6.34: The frequency of the top 50 candidates from the Counter Selection strategy through all SELEX rounds.**

### 6.3.15 Top candidates from Each Strategy

The sequences of the top candidates from each selection strategy are listed in Tables 6.5 – 6.11. The top ten candidates displayed were based on the highest frequencies in the final selection round, as well as the top ten candidates based on cycle-to-cycle enrichment in the final two rounds. An interesting observation was that across nearly all strategies, there was

no overlap between the most enriched candidates and the most abundant. The only two overlaps occurred in the Room Temperature strategy where the eighth most abundant candidate was the fourth most enriched candidate, and during the Counter Selection strategy where the third most abundant candidate was also the most enriched candidate. Another finding was that all of the candidates displayed showed a thymine bias which is more likely due to the T-rich starting library than due to T-rich sequences exhibiting the best affinity to oxybenzone.

**Table 6.5: Top candidates from the Standard strategy**

Top Candidates (Frequency)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
TAAATTTTCGTCCTTTTTTCTTTGTTCTTTCC	37	2	103	159.26	1	424	964.67	6.05720206
TCGATTGTATTAATCTCAATTTATTTCTCTTG	40	2236	37	57.21	2	422	960.12	16.7823807
ATCGATTGTATTAATCTCAATTTATTTCTCTTG	41	1879	39	60.3	3	403	916.89	15.20547264
ATAATTTTAATTTTTGTCTAATTTTATATGTATTG	41	512	54	83.49	4	347	789.49	9.456102527
ATTGCTTTGTTCTGTAGTTTTGGACAAATTTGTTAG	41	92	71	109.78	5	330	750.81	6.839223902
ATCAATTTCAATTTGATTTTTTAATTCCTTTCTATGTG	41	209	64	98.96	6	317	721.23	7.2880962
ATAAAGATTCCTTTTACTTTTCTTTTATCTTTTAAATG	41	92	71	109.78	7	313	712.13	6.486882857
ACCACTTTTCTTTGTTTTAATTTATTTTCAGAACATTGTA	41	3763	30	46.39	8	312	709.85	15.30178918
TTGCTTTGTCTGTAGTTTTGGACAAATTTGTTT	38	92	71	109.78	9	308	700.75	6.383220987
ACAGTTATTTTTATTTTGTTAGTCTATCCTATTTCTGTG	41	18	85	131.43	10	298	678	5.15863958
Top Candidates (Enrichment)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
TTATCGTAATACGTTAATTGTAGATTTTATTGTGATAGT	40	48283	1	1.55	97	157	357.2	230.4516129
ACITCCTTATTACGTTTTTACTTATTTTATTTCTGG	40	40393	2	3.09	242	112	254.82	82.46601942
AATTTTCCTTATTTAATATTACTCTATTTTTTGT	37	48283	1	1.55	1389	50	113.76	73.39354839
ACACTTTTCTTCTTAAATCTTCTTTTAAACTTCTTTGT	41	40393	2	3.09	333	99	225.24	72.89320388
ATTCTCTTTAATTATCTATACATGATTGTTTCTTTTCG	40	48283	1	1.55	1539	47	106.93	68.98709677
ATTTCTTTATATTATCTATTTTTTTATATTTTGTAG	40	48283	1	1.55	1812	43	97.83	63.11612903
AACTTTTCTTTTCTATTAATTATTTTTCGATTATG	40	48283	1	1.55	1812	43	97.83	63.11612903
TGACGTTGTTCATTCCTTTTAAAAAGGTTGATAGTA	38	48283	1	1.55	1871	42	95.56	61.6516129
ATTACGTAAATTTTATACTGAGATTTATTTCAATTTCTA	39	40393	2	3.09	535	82	186.56	60.37540453
TCATTAATTTTTTCTCAATCATTTTTCTTTTTCTAT	40	48283	1	1.55	1953	41	93.28	60.18064516



**Table 6.6: Top candidates from the Standard + Mutation strategy.**

Top Candidates (Frequency)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
ATTTAGGTTTATGTTTAGTAATCTTTTCTAGTGTTTCCG	41	262	78	191.82	1	1035	2429.01	12.66296528
TTTAGGTTTATGTTTAGTAATCTTTTCTAGTGTTTCCG	40	502	66	162.31	2	921	2161.47	13.3169244
ATTGTATCTGGCTCTTTCTGGTTTTTTTATTTTCCTTGGG	41	45	104	255.76	3	776	1821.17	7.120620895
TTGTATCTGGCTCTTTCTGGTTTTTTTATTTTCCTT	37	94	93	228.71	4	754	1769.54	7.737046915
TTAGTTGTTATTGTTTGCTATTGTTTTCTGTTTCGTTGTG	40	1040	53	130.34	5	727	1706.18	13.09022556
ATTAGTTGTTATTGTTTGCTATTGTTTTCTGTTTCGTTGTG	41	704	60	147.56	6	724	1699.14	11.51490919
ATTGATTTTCTTGCTTATTTTTCTTTTTTGTTTTTG	41	94	93	228.71	7	674	1581.79	6.916138341
ATAACCTCTTCTGCTTTTTTACGATTTTTTATTTTTTG	41	191	83	204.12	8	668	1567.71	7.680335097
TTGATTTTCTTGCTTATTTTTCTTTTTTGTTTTTG	40	50	102	250.85	9	649	1523.12	6.071835758
ATTTTTGTCTGTTATTTTATTATTGGTTTTTTTGGTG	41	79	95	233.63	10	606	1422.2	6.087403159
Top Candidates (Enrichment)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
ACTTATTGTGTTTTTTTCTTAATATAATTTGTTTTCTA	40	17710	2	4.92	100	357	837.83	170.2906504
TTTTTATTCTTAAGTTCTTCTTTTTTTGTTCTGTTTTG	40	17710	2	4.92	320	235	551.52	112.097561
TTTTTTTTGTGTATTTCTTATTTTATTCTTCTGCTTTT	40	21498	1	2.46	1091	107	251.12	102.0813008
ACTTTTTCTAGTTTGTGGTTTTGTTTTTTTCGTTTAA	40	15919	3	7.38	293	249	584.37	79.18292683
ATTCATTATTTTATGTTTTTTCTTTTTTTTATTCCC	40	21498	1	2.46	1634	74	173.67	70.59756098
GATATTGTTTTTTGTTAATGTTTTTTTCTCTGCTTCT	40	14770	4	9.84	229	273	640.7	65.11178862
ATTATCTTATATTCTTTGTTTTTTTATCTTTTATCAG	40	14770	4	9.84	259	262	614.88	62.48780488
CTTTTTCTAGTTTGTGGTTTTGTTTTTTTCGTTAAT	40	14770	4	9.84	314	237	556.21	56.5254065
TTTTTTTTATGTTTGTTTTTTATTTTTTCTCAAACGT	40	15919	3	7.38	541	175	410.7	55.6504065
AGATATTGTTTTTTGTTAATGTTTTTTTCTCTGCTTC	40	13935	5	12.3	200	283	664.16	53.99674797

**Table 6.7: Top candidates from the Extra Volume strategy.**

Top Candidates (Frequency)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
ATCGATTATTTTTATATTTTTGTTATAACTTTTCG	41	0	0	0	1	632	1171.36	0
ATACTATTCGGTTATTGTTTTATTTAACTTTTCTTATG	41	0	0	0	2	617	1143.55	0
TCGATTATTTTTATATTTTTGTTATAACTTTTCG	40	0	0	0	3	615	1139.85	0
ATGCATTTGGTTTTTCGTTTTGTTTTCTTTTTTCG	41	0	0	0	4	604	1119.46	0
TGCATTTGGTTTTTCGTTTTGTTTTCTTTTTTCG	40	0	0	0	5	601	1113.9	0
AATTATTGTACTTGGTTTTCTTTGTTCTGTTTTGG	41	0	0	0	6	573	1062	0
CATTTTTAGTTGTTTTTTTACTTTTTTGTCG	37	0	0	0	7	569	1054.59	0
TACTTATTCGGTTATTGTTTTATTTAACTTTTCTTATG	40	0	0	0	8	541	1002.69	0
TCTCATTTTTAGTTGTTTTTTTACTTTTTTGTCG	40	0	0	0	9	523	969.33	0
AGTTATTTAATTTCTTTTCTTTTTTTATACCTTTGG	41	0	0	0	10	518	960.07	0
Top Candidates (Enrichment)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
ATTTAATTTTAATTTTGTTTTGTCTTTTCTTTTCG	41	2941	13	81.78	28	403	746.92	9.133284422
AGTTTTTTTCTCATCTTTGTTTTTTTGATGTTGT	40	6440	1	6.29	4589	29	53.75	8.545310016
TTTAATTTTAATTTTGTTTTGTCTTTTCTTTTCG	40	2852	14	88.07	79	310	574.56	6.523901442
GTTTTTTTCTCATCTTTGTTTTTTTGATGTTGT	40	3896	5	31.45	4177	33	61.16	1.944674086
GTTCTTTTTGTTTTCTTTTCTTTTGTCGTATTTGT	39	4513	3	18.87	6667	13	24.09	1.276629571
AGTTCTTTTTGTTTTCTTTTCTTTTGTCGTATTTGTG	41	85	120	754.9	16	488	904.46	1.198118956
TTCTTTTTTAATTTTTTATTTTTACTTTTTTCCA	40	5073	2	12.58	7820	8	14.83	1.178855326
GTTCTTTTTGTTTTCTTTTCTTTTGTCGTATTTGTG	40	89	118	742.32	31	399	739.51	0.996214571
ATTCTTTTTAATTTTTTATTTTTACTTTTTTCCA	40	6440	1	6.29	10633	3	5.56	0.883942766

**Table 6.8: Top candidates from the Extra Detergent strategy.**

Top Candidates (Frequency)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
Sequence information has been redacted for intellectual property protection.	41	0	0	0	1	2074	5728	0
CATTTTTTTATTTATTTCTATTTTATTATTGTTTCTG	40	0	0	0	2	1843	5090.02	0
ATTTATTTTATGTTTCGTATTTTATATTTTCTTTG	41	0	0	0	3	1469	4057.1	0
TTTATTTTATGTTTCGTATTTTATATTTTCTTTG	40	0	0	0	4	1351	3731.21	0
ATTTGTTTTATTTTTATTTTTCTTTTAATTTTTG	41	0	0	0	5	1314	3629.02	0
ATTCATGTTTGTTTTTCTTTGATTTTTCTTTGTA	41	0	0	0	6	1235	3410.84	0
TTTCATGTTTGTTTTTCTTTGATTTTTCTTTGT	39	0	0	0	7	1163	3211.99	0
TTTGTTTTATTTTTATTTTTCTTTTAATTTTTG	40	0	0	0	7	1163	3211.99	0
ATTTTATTCTATTTCTTTTTCTGTTATTTAATTTG	41	0	0	0	9	1044	2883.33	0
ATTCCTCTTTTGTTTTTCTTTTTATTTTTATGTG	41	0	0	0	10	1036	2861.24	0
Top Candidates (Enrichment)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
ATTATTCACTTTTAATTTATTTTTTATTTTGT	38	5401	1	5.62	2288	9	24.86	4.423487544
TTAATATTCCTTCATTCTGTTATTTTATAATTTTC	38	4042	2	11.25	1845	18	49.71	4.418666667
CGCAATTTGGATATTTTGTCTGATTTTTCTCTA	40	5401	1	5.62	2372	8	22.09	3.930604982
ATGTGTTCTAATTTAATATTTTTTGTTTTTTAATG	40	5401	1	5.62	2467	7	19.33	3.439501779
TTTATTTAATTGTTTTCTGTGTTAATTTCTTATTT	39	5401	1	5.62	2467	7	19.33	3.439501779
TTAATATTCCTTCATTCTGTTATTTTATAATTTCCA	40	5401	1	5.62	2607	6	16.57	2.948398577
AGTCTATTTTATGGTTTCTTTTTGTTCATGTTTGT	41	5401	1	5.62	2607	6	16.57	2.948398577
TTAATATTCCTTCATTCTGTTATTTTATAATTTCC	39	512	84	472.36	138	464	1281.48	2.712930815
TTTTTGTTTTTTTTGTTCTGTTTTGTTATTTTCATG	39	1637	22	123.71	805	118	325.89	2.634306038
TTTTTTTATGTTTCATACATATTTATTTTTTTCATG	40	2645	7	39.36	1491	36	99.43	2.526168699

**Table 6.9: Top candidates from the Room Temperature strategy.**

Top Candidates (Frequency)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
AAACAATTGAGTTTTTTGTTGTTATTTTTGTTTTTCG	41	0	0	0	1	7505	16057.32	0
AACAATTGAGTTTTTTGTTGTTATTTTTGTTTTTCG	40	0	0	0	2	6827	14606.71	0
ATGTTTTTTTCCTTTATTTTTGACTGAATACTTTTTTG	41	0	0	0	3	4751	10165	0
TGTTTTTTTCCTTTATTTTTGACTGAATACTTTTTTG	40	0	0	0	4	4209	9005.37	0
ATTTATTTTTCTTCTATTAATTTTTTTGTTTCTGTAG	41	0	0	0	5	3731	7982.66	0
ACTCTTTCTATTTAATTTTTCAATTCATTTTTTTGTG	41	0	0	0	6	3693	7901.36	0
TTCAGCAAATTTTTTTCTTTTATTTACTTACCGTTGT	39	0	0	0	7	3595	7691.68	0
ATTTCTTTTATTATTTTTTTCTTCCATTCTACATTTGGG	41	567	157	422.39	8	3554	7603.96	18.00222543
ATTCAGCAAATTTTTTTCTTTTATTTACTTACCGTTGTA	41	0	0	0	9	3460	7402.84	0
ATTTCTTTTGTACTTTTTTTGTCAAGTTTAATTAATTCG	41	0	0	0	10	3452	7385.73	0
Top Candidates (Enrichment)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
TTTCTTTTATTATTTTTTTCTTCCATTCTACATTTGGGT	40	5441	3	8.07	570	123	263.16	32.60966543
TTTCTTTTATTATTTTTTTCTTCCATTCTACATTTGGG	40	6375	2	5.38	715	72	154.05	28.633829
ATTTCTTTTATTATTTTTTTCTTCCATTCTACATTTGGG	40	4584	5	13.45	480	166	355.17	26.40669145
ATTTCTTTTATTATTTTTTTCTTCCATTCTACATTTGGG	41	567	157	422.39	8	3554	7603.96	18.00222543
TTTCTTTTATTATTTTTTTCTTCCATTCTACATTTT	37	683	142	382.03	21	2745	5873.06	15.37329529
TTTCTTTTATTATTTTTTTCTTCCATTCTACATTTGG	40	8847	1	2.69	1268	15	32.09	11.92936803
TTTCTTTTATTATTTTTTTCTTCCGTTCTACATTTT	37	8847	1	2.69	1306	14	29.95	11.133829
TTTACATTTTTTTTATATTCATTTATTGTTGTTTTTG	39	8847	1	2.69	1549	10	21.4	7.955390335
ACTTCTTTTATTATTTTTTTCTTCCATTCTACATTTGGG	41	8847	1	2.69	1549	10	21.4	7.955390335
ATTTACATTTTTTTTATATTCATTTATTGTTGTTTTTG	40	2832	23	61.88	403	213	455.72	7.3645766

**Table 6.10: Top candidates from the Negative Selection strategy.**

Top Candidates (Frequency)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
ATGTTTTATGTTTTATTATTTATTTCTTTTTCTGTA	41	0	0	0	1	15529	18430.82	0
TGTTTTATGTTTTATTATTTATTTCTTTTTCTGT	39	0	0	0	2	14689	17433.86	0
AATTTTTTGATTTTTGTTTCATATTTGTCATTCTG	37	0	0	0	3	14500	17209.54	0
TCTAATTTTTGATTTTTGTTTCATATTTGTCATTCTG	40	0	0	0	4	13726	16290.91	0
TGTGTTCTAATTTAATATTTTTTTGTTTTTTAATGT	40	0	0	0	5	9758	11581.43	0
ATGTGTTCTAATTTAATATTTTTTTGTTTTTTAATG	40	0	0	0	6	9555	11340.49	0
ATGTGCATGTTTTTTGTTTGATCATCACTTTCCCTTACT	41	6	2288	6690.39	7	8938	10608.2	1.585587686
TGTGCATGTTTTTTGTTTGATCATCACTTTCCCTTACT	40	5	2331	6816.13	8	8542	10138.2	1.487383603
ATTCTATTACTTTTGTCTTTCTTTTTCTTTTCTTAGTG	41	0	0	0	9	8102	9615.98	0
TTCTATTACTTTTGTCTTTCTTTTTCTTTTCTTAGTG	40	0	0	0	10	7458	8851.64	0
Top Candidates (Enrichment)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
TAACGCTGATTGATCGTGTTTTGTGCTTTGAAGTG	36	6234	1	2.92	699	62	73.59	25.20205479
ATAACGCTGATTGATCGTGTTTTGTGCTTTGAAGTACCGG	41	6234	1	2.92	950	30	35.61	12.19520548
GTTTTTTTCTTTTGTGTAATTTTTACTTTATATT	40	1265	23	67.25	255	656	778.58	11.57739777
TTTTTATTTATTTATCCTTTGCTATTTTACTCTTT	37	789	75	219.31	109	1823	2163.65	9.865715198
TAACGCTGATTGATCGTGTTTTGTGCTTTGAAGCGCCGG	40	6234	1	2.92	1057	24	28.48	9.753424658
TAACGCTGATTGATCGTGTTTTGTGCTTTGAAGTGCCTG	40	6234	1	2.92	1057	24	28.48	9.753424658
ATAACGCTGATTGATCGTGTTTTGTGCTTTGAAGTGCCTG	41	3782	2	5.85	770	48	56.97	9.738461538
TGTGCATGTTTTTTGTTTGGTCATCACTTTCCCTTACT	40	6234	1	2.92	1092	23	27.3	9.349315068
GTTTTTTTCTTTTGTGTAATTTTTACTTTATATTG	40	6234	1	2.92	1092	23	27.3	9.349315068
TCTTTTTATTTATTTATCCTTTGCTATTTTACTCTTT	40	782	76	222.23	119	1718	2039.03	9.175313864

**Table 6.11: Top candidates from the Counter Selection strategy.**

Top Candidates (Frequency)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
ATGTGCATGTTTTTTGTTTGATCATCACTTCCCTTTACT	41	2	1114	5236.64	1	3966	74490.06	14.22478154
TGTGCATGTTTTTTGTTTGATCATCACTTCCCTTTACT	40	1	1314	6176.79	2	3694	69381.32	11.23258521
CAATATCCATATTTTGTGTTTTTTCGTCCTTGTTTTTTG	41	4624	1	4.7	3	1542	28962.1	6162.148936
ATAATCTTTGTATAAAATTTTTTGTGTTTTTACATTCGG	41	0	0	0	4	1512	28398.63	0
TTCCATATTTGTTTTTTCGTCCTTGTTTTTTGTGAT	40	0	0	0	5	1470	27609.78	0
TAATCTTTGTATAAAATTTTTTGTGTTTTTACATTCGG	40	0	0	0	6	1321	24811.24	0
ATGTCATTTGTTTTTATTTTTTTGTTCAATTATCTTTTCG	41	0	0	0	7	1074	20172.04	0
ATAAGTCTAAAGGTTTTTCAATTGTTTTATTTTTCTTGG	41	0	0	0	8	963	18087.22	0
ATTAATTTATTCGTCCTTGCTTTTGTGTTTATTTCTTG	41	0	0	0	9	868	16302.92	0
TAAGTCTAAAGGTTTTTCAATTGTTTTATTTTCTTGG	40	0	0	0	10	722	13560.72	0
Top Candidates (Enrichment)								
Sequence	Length	Rank (R6)	Reads (R6)	RPM (R6)	Rank (R7)	Reads (R7)	RPM (R7)	Enrichment (R6/R7)
CAATATCCATATTTGTTTTTTCGTCCTTGTTTTTTG	41	4624	1	4.7	3	1542	28962.1	6162.148936
ATAGCTTTGTATTTTTTGGATATTTTGTCTTTGT	40	4624	1	4.7	33	334	6273.24	1334.731915
ATGTGCATGTTTTTTGTTTGATCATCACTTCCCTTTACT	40	4624	1	4.7	73	137	2573.16	547.4808511
ATATTTTCGGTTTCTGTTGATTTTAATTTTGCTTATAGT	41	4624	1	4.7	97	82	1540.14	327.6893617
AAGTGATATGCTTCTATTTTTATTTGTAATTTCTTGTTG	41	3126	2	9.4	67	159	2986.36	317.6978723
ATGTGCATGTTTTTTGTTTGATCATCACTCTCCCTTTACT	41	4624	1	4.7	142	34	638.59	135.8702128
TGTGCATGTTTTTTGTTTGATCATCACTCTCCCTTTACT	40	3126	2	9.4	151	32	601.03	63.9393617
ATGTGCATGTTTTTTGTTTGACCATCACTTCCCTTTACT	41	4624	1	4.7	231	12	225.39	47.95531915
ATGTGCATGTTTTTTGTTTGATCATCACTTCCCTTTACT	41	4624	1	4.7	257	9	169.04	35.96595745
TGTGCATGTTTTTTGTTTGATCATCACTTCCCTTTACT	40	4624	1	4.7	257	9	169.04	35.96595745

### 6.3.16 Candidate Characterisation by Micro-Scale Thermophoresis

As the number of candidates for characterisation was limited due to cost, initially nine candidates were selected. The sequence, sequence length,  $\Delta$  Gibbs free energy, nucleotide composition and the reason for selection for each candidate characterised are presented in Table 6.12.

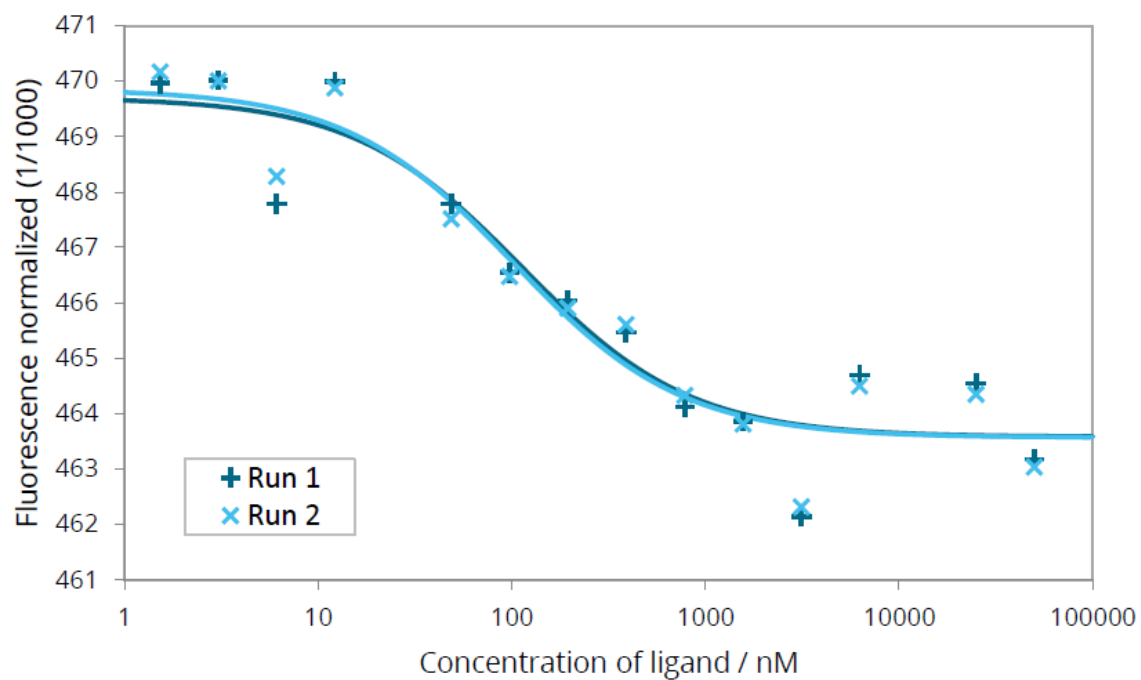
**Table 6.12: Aptamer candidates chosen for characterisation by MST**

Sequence Identifier	Sequence (5' to 3')	Length (nt)	$\Delta$ Gibbs Free Energy (kcal/mol)	Nucleotide Composition (N40 region)	Reason for selection
OXY-S7-C1	ATACGAGCTTGTTCAATAAAATTTTCGTCCCTTTTTTCTTTGTTCTTTTCCGTAGTAGTAAGAGCAATC	72	-2.31	A = 8.1% C = 21.6% G = 5.4% T = 64.9% GC = 27.0%	Top ranked candidate in selection round 7 (RPM) from the standard strategy
OXY-SM7-C1	ATACGAGCTTGTTCAATAATTAGGTTTATGTTTAGTAATCTTTTCTAGTGTTTCCGTAGTAGTAAGAGCAATC	76	-0.81	A = 17.1% C = 9.8% G = 17.1% T = 56.1% GC = 26.8%	Top ranked candidate in selection round 7 (RPM) from the standard + mutation strategy
OXY-EV7-C1	ATACGAGCTTGTTCAATAATCGATTATTTTTTATATTTTTTGTATAACTTTTCGTAGTAGTAAGAGCAATC	76	-3.02	A = 19.5% C = 7.3% G = 7.3% T = 65.9% GC = 14.6%	Top ranked candidate in selection round 7 (RPM) from the extra volume strategy
OXY-ED7-C1	ATACGAGCTTGTTCAATANNTGATAGTAAGAGCAATC (sequence information has been redacted for intellectual property protection)	76	-2.05	A = 17.1% C = 7.3% G = 4.9% T = 70.7% GC = 12.2%	Top ranked candidate in selection round 7 (RPM) from the extra detergent strategy

OXY-RT7-C1	ATACGAGCTTGTTCAATAAAACAATTGAGTTTTTTGTTGTTATTTTTGTTTTTCGTGATAGTAAGAGCAATC	76	-4.11	A = 17.1% C = 4.9% G = 14.6% T = 63.4% GC = 19.5%	Top ranked candidate in selection round 7 (RPM) from the room temp strategy
OXY-N7-C1	ATACGAGCTTGTTCAATAATGTTTTATGTTTTTATTATTTATTTCTTTTTCTGTATGATAGTAAGAGCAATC	76	-1.24	A = 14.6% C = 4.9% G = 7.3% T = 73.2% GC = 12.2%	Top ranked candidate in selection round 7 (RPM) from the negative selection strategy
OXY-CT7-C1	ATACGAGCTTGTTCAATATGTGCATGTTTTTTGTTTGATCATCACTTCCCTTTACTTGATAGTAAGAGCAATC	75	-2.00	A = 12.5% C = 20.0% G = 12.5% T = 55.0% GC = 32.5%	Top ranked candidate in selection round 7 (RPM) from the counter selection strategy
OXY-N7-E1	ATACGAGCTTGTTCAATAAACGCTGATTGATCGTGTGTTTTTGCTTTGAAGTGTGATAGTAAGAGCAATC	71	-3.29	A = 16.7% C = 11.1% G = 27.8% T = 44.4% GC = 38.9%	Top ranked candidate based on cycle-to-cycle enrichment from the negative selection strategy
OXY-CT7-E1	ATACGAGCTTGTTCAATACAATATTCCATATTTGTTTTTTTCGTCTTTGTTTTTTGTGATAGTAAGAGCAATC	76	-1.85	A = 12.2% C = 12.2% G = 9.8% T = 65.9% GC = 22.0%	Top ranked candidate based on cycle-to-cycle enrichment from the counter selection strategy

The only aptamer candidate for which binding was evident in the MST assays was OXY-ED7-C1. This aptamer was generated from the Extra Detergent strategy and exhibited a calculated  $K_d$  value of 113 and 102 nM in the first and second technical run, respectively (Figure 6.35), for an average  $K_d$  of 107.5 nM. The signal-noise ratio was sufficiently high for reliable data evaluation. The amplitude of the curve was relatively small (average of -6.2) but was a factor of 3 above the initial project noise level test.





	$K_D$	Amplitude	S/N
Run 1	113 nM	-6.1	7.4
Run 2	102 nM	-6.3	8.7

Figure 6.35: Results of the MST binding assay for oxybenzone-binding aptamer OXY-ED7-C1.

## 6.4 Discussion

This chapter presents a novel pipeline for the analysis and visualization of data from an HT-SELEX experiment and demonstrates that different selection strategies can result in vastly different SELEX outcomes. Concerns regarding the quality of synthetic random libraries generated by suppliers and the impact that poor quality libraries can have on a SELEX experiment are raised. Many insights about the design of future HT-SELEX experiments and the selection of aptamers for small molecule phenolic compounds, such as oxybenzone have been generated from this work and are discussed here.

Currently, there are many tools available for the analysis of HT-SELEX data however most of these focus on the analysis of multiple rounds of the same SELEX experiment. The aim of these software packages is to use the sequence data to find the best potential aptamer candidates by calculating the frequency of unique sequences and calculating enrichment. FASTAptamer (Alam et al., 2015) is a good example of this, it is brilliant tool for performing these functions. Currently lacking in scientific literature and in the toolbox of aptamer scientists are solutions for comparing multiple selection strategies against each other. The output of FASTAptamer is not amenable to performing these functions. Presented in this work are a set of scripts, written in Bash and R, which given the correct file naming conventions can complete these tasks for a large number of files across many SELEX experiments. It should be noted that AptaSUITE (Hoinka et al., 2018) should be able to perform these functions but due to issues with Java dependencies and the complex nature of using the graphical user interface on a high performance computer is difficult to use. The specific functions enabled through the pipeline described in this chapter are as follows: (1) removal of contaminating sequences, (2) comparisons of frequency and enrichment across multiple SELEX rounds and multiple

selection strategies, (3) easy manipulation and visualization of .tsv files produced by FASTAptamer, and (4) semi-automated analysis of large numbers of files (corresponding to different SELEX rounds) through the entire pipeline. The collection of scripts are available in Appendix D.

The synthetic library used for this experiment, provided by Thermo Fisher Scientific, was shown to have significant bias in nucleotide composition before any selection pressure was applied to the library. Commercial libraries are generated using solid-phase synthesis which relies on phosphoramidite coupling chemistry to produce random synthetic oligonucleotides. An identified problem with this method is that the nucleotide composition of a synthesized library can be skewed towards phosphoramidites with higher coupling efficiencies, for example Guanine and Thymine phosphoramidites (Hall et al., 2009). Guanine phosphoramidites have the highest coupling efficiency (1.00) followed by Thymine (0.83), Adenine (0.67) and Cytosine (0.67) (Hall et al., 2009). In order to compensate for unequal coupling efficiency and to produce libraries which have a statistically random distribution of nucleotides, phosphoramidites have to be manually mixed at specified ratios (Hall et al., 2009). A suggestion molar ratio to create a random library is 1.5 : 1.5 : 1.0 : 1.2 (A : C : G : T) (Hall et al., 2009). The library used in this selection experiment had an extreme bias towards Thymine nucleotides in the N40 region of the library. After consultation with the supplier, even though a “random” library was ordered, it was learned that the phosphoramidites are not manually mixed at specified ratios unless requested by the customer. Even so, if the bias was based solely on coupling efficiency of different phosphoramidites, then I would expect the library to have a higher composition of Guanine than Thymine which was not the case. The most likely conclusion is therefore that the supplier made an error when synthesizing the library. Interestingly, even though the library was composed of approximately 70% Thymine

in the N40 region at the beginning of SELEX and had low complexity due to the presence of large amounts of polyT sequences, enrichment was evident in all strategies and a high affinity aptamer for oxybenzone (OXY-ED7-C1) was able to be isolated. This attests to how robust the SELEX strategies employed here are and how robust the process of *in-vitro* selection is in general. Other studies have assessed starting libraries for SELEX and there is even a software programme (COMPAS) with functionality to check the randomness of a starting library (Blank, 2016). It is noted that analyses of a multitude of starting libraries provided by different sources (academic laboratories and commercial companies) revealed that truly random libraries with an even distribution of nucleotides are actually rare (Blank, 2016). The author expands upon this point by stating that library randomness directly correlates with successful SELEX outcomes as libraries which are more random have a larger sequence space and a better distribution of motifs (Blank, 2016). The loss of sequence diversity due to the overwhelming presence of polyT motifs in the starting library (which was only discovered upon sequencing, after completing the experiment) would have almost certainly had a detrimental effect on the chance of producing successful aptamer candidates.

The normalized frequency of unique sequences within an oligonucleotide pool as well as cycle-to-cycle enrichment are good indicators of whether positive selection is acting on a particular oligonucleotide. It was hypothesized that on average, oligonucleotides within more stringent selection strategies would have higher frequencies and higher enrichment at the end of SELEX. This hypothesis was confirmed for frequency, the Counter Selection strategy had candidates with the highest average frequency in Round 7, followed by the Negative Selection, Room Temperature and Extra Detergent strategies in that order. The more selection pressure put on a library, the higher the number of non-binding oligonucleotides which are eliminated each round. As a result, the input into the PCR is much lower meaning

that the oligonucleotides which remain make up a much larger portion of the pool. The Counter Selection strategy had an approximately 20-fold higher frequency of top candidates compared to the Standard strategy. A high frequency can indicate binding potential but it can also be a result of selection for sequences which are favourably amplified in PCR. To address this, it is recommended to use cycle-to-cycle enrichment, which has been shown in one study to be a better predictor of binding (Hoinka et al., 2015). In terms of enrichment, oligonucleotides within the Counter Selection strategy showed the highest enrichment from Round 6 to 7, followed by the Standard strategy. The Counter Selection strategy having the highest enrichment in the final round makes logical sense. Washing with similarly structured compounds to oxybenzone would have removed a large number of sequences with weak binding affinity and would have also removed oligo's which had some affinity to structural motifs present in both oxybenzone and the counter selection compounds. By removing a large number of sequences through counter selection, the remaining oligonucleotides (which had sufficient binding affinity to remain) would be highly enriched in the following PCR. Other high stringency strategies (Negative Selection, Room Temperature and Extra Detergent) had comparable lower enrichment when compared to a Standard strategy. There are a couple of possible hypotheses for this. It is possible that within the starting library aliquot for these particular strategies, there was no sequence which had the capability to bind to oxybenzone and the stochastic process of induced mutation did not produce a favourable candidate in these strategies. As the libraries were of a low complexity to begin with (due to the nucleotide composition bias), it is possible that the sequence space was not large enough to have a candidate capable of binding the target. A second hypothesis is that the mutation induced by error-prone PCR was unfavourable to the selection pool. Mutation can produce candidates with better binding capabilities but it can also reduce the competitiveness of low frequency,

high-affinity binders by creating oligonucleotides which have characteristics for favourable PCR amplification. These parasitic sequences are then able to out-compete true binders. From the enrichment results presented herein, it is clear that different selection strategies produce different SELEX outcomes. It is difficult to determine if induced mutation is advantageous or not. It is pertinent to note here that other studies (Allnutt et al., 2018; Blank, 2016; Hoinka et al., 2015) use a deconvolution step to cluster similar aptamers into families and then assess the enrichment of that particular family whereas I am assessing the enrichment of individual oligonucleotides.

The only strategy which produced an aptamer candidate that showed binding to oxybenzone in an MST assay was the Extra Detergent strategy. This differed from my initial hypothesis that the highest affinity candidates would be present in the most stringent selection strategies. In contrast, the two candidates tested from the Negative Selection and Counter Selection strategies didn't show any binding to oxybenzone. One of these candidates (OXY-CT7-C1) was present in the top 50 candidates (by frequency) of three different strategies (standard, negative selection and counter selection). As no binding affinity was demonstrated for OXY-CT7-C1, this indicates to me that this oligonucleotide was a parasitic sequence which was able to be favourably amplified in PCR. The reason for its abundant presence across multiple selection strategies is therefore due to PCR amplification efficiency rather than target binding. A large limitation of this study in terms of assessing binding affinity of candidates was that only one or two candidates from each selection strategy was characterised using MST: the highest frequency sequence in round 7 from each strategy and the most enriched sequence in the Negative and Counter selection strategies. In order to determine the strategy which produces the most successful binding sequences, it is likely that at least 10 sequences (and potentially up to 100) from each strategy would need to be

characterised. From this information, the average affinity of candidates from each strategy could be compared to provide a true representation of how successful each strategy was. Small molecule aptamer characterisation is notoriously difficult; most of the methods are unreliable, inconsistent and low-throughput. To continue assessing the effectiveness of each strategy, a high-throughput characterisation method that can assess 10-100 candidates in an economical and timely manner is required. This was beyond the scope of this project and is a future direction that we would like to pursue.

HT-SELEX is a complex and costly method, which has huge potential to unlock the “black box” of SELEX. There are important and relevant lessons learned from this study which would be useful to apply to future studies of this type. Quality control of the initial starting library should have been performed before starting this experiment. The quality control performed by suppliers of oligonucleotide libraries is limited to ensuring that the oligonucleotides are of the correct size and that there is a sufficient quantity of DNA. It is recommended that high throughput sequencing of any “random” library used in an experiment similar to this be performed prior to beginning SELEX. The library should be analysed for nucleotide composition, sequence diversity and the occurrence of polyX motifs (that is, a motif containing any nucleotide consecutively in more than 3 positions). The random region should also be searched for the presence of any primer binding sites, which can result in PCR by-products. As long as there are no major deviations from statistical randomness then the library is suitable for a HT-SELEX experiment.

# Chapter 7: General Discussion

Overall, this thesis presents a number of novel findings that advance our understanding of generating aptamers (SELEX) for small molecules, and in particular environmental contaminants. This PhD study has demonstrated for the first time, the presence of a number of EOC's in waterways at different locations in the North Island of New Zealand. Furthermore, aptamers that bind to glyphosate and oxybenzone were generated for the first time through the use of SELEX. Unfortunately, the SELEX methodology for generating aptamers that bind to technical nonylphenol equivalents was not successful and the limitations of generating aptamers to this type of molecule are discussed. The use of HT-SELEX enabled a novel bioinformatics pipeline to be established for the comparison of seven different selection strategies. Insights regarding the effectiveness of these different selection strategies have been discussed as well as the limitations of studies of this type.

## 7.1 EOCs in New Zealand waterways

There is little information on the presence of EOC's in New Zealand waterways. Using traditional sampling and detection methods, an analyses of the occurrence of glyphosate (and metabolites), phenolic compounds including a range of nonylphenols and oxybenzone and a number of other pesticides were examined in a variety of different New Zealand waterways. This work was undertaken under the guidance of Dr Grant Northcott at Plant and Food Research (Hamilton, New Zealand) andASUREQuality (Wellington, New Zealand). Drawing correlations between EOC type and land use was originally planned but many more waterways would have to be sampled before any definitive conclusions could be presented.



Technical nonylphenol equivalents were detected at concerning concentrations in a small stream in a highly industrial area of Lower Hutt (Waiwhetu Stream), near the New Zealand capital city of Wellington. The concentrations exceeded the recommended guidelines set by the Australian and New Zealand Environment and Conservation Council (ANZECC) for species protection. It is recommended that sampling and testing of nonylphenols be conducted at multiple time points or a passive sampler is employed to give an estimated profile of concentrations over time. This suggested testing regime would provide longitudinal data on the occurrence of this compound. This is the first assessment of the presence of nonylphenols in the water column in New Zealand. However previous studies have assessed the concentrations in sediment; see Stewart et al. (2014). The results herein raise questions regarding whether the monitoring of nonylphenols should be included in regional council monthly monitoring programs as an indicator EOC.

Concentrations of glyphosate and its metabolites at the eight sampling sites tested were well below the ANZECC guidelines and comparably lower than those reported in similar studies overseas. Glyphosate is used heavily in New Zealand for weed control. Laboratories with testing capabilities for monitoring glyphosate in New Zealand at environmentally relevant concentrations are severely lacking. The testing of glyphosate at established water and environmental testing laboratories is expensive at an approximate cost of NZ\$400 per sample. Moreover, the minimum detection limit of these tests are 200 ng/L which is not sensitive enough to detect concentrations that are known to detrimentally impact the environment. WithASURE Quality's assistance, the glyphosate levels were measured herein using a test designed for assessing drinking water quality. As glyphosate is such a common herbicide and glyphosate's effects on human and environmental health are currently being debated, it is recommended that testing capabilities in New Zealand laboratories need to be improved. In

addition, the need for a more economical and sensitive detection technology is clearly highlighted.

Oxybenzone was present above the minimum detection level at all eight sampling sites. After reviewing eco-toxicology data, it was concluded that the concentrations present in our study were not a risk to aquatic organisms. Authorities in multiple states in the United States of America (specifically Hawaii and Florida) as well as the Pacific island nation of Palau banned sunscreens containing oxybenzone during the period of this study, highlighting the risk that the compound presents to coral. Increased monitoring of oxybenzone is needed in New Zealand. We only monitored fresh water sites in this study. Monitoring of oxybenzone in seawater around popular swimming beaches would be beneficial to establishing if there is a risk to New Zealand marine species.

By assessing the occurrence of common EOC's in New Zealand freshwater ways, this study has provided a snapshot of the potential risks and added to the knowledge of the environmental concentrations of these chemicals in a New Zealand context. The risk these compounds present to aquatic organisms through assessing eco-toxicological data and comparing detected concentrations to guidelines levels has provided evidence that further monitoring should be being carried out in New Zealand.

## **7.2 Potential of Aptamer Detection Methods for EOCs**

Environmental monitoring using aptamer-based biosensors (aptasensors) is rapidly growing as aptasensors represent a robust, sensitive and economical method of detecting contaminants. Gas or liquid chromatography coupled with mass spectrometry (GC-MS/LC-MS) is the current method for detecting these types of molecules in a freshwater matrix. GC-

MS/LC-MS is extremely sensitive however, it is expensive, is not portable, requires highly trained technical expertise and different methods are required for different types of molecules. Chapter four describes the generation of aptamer candidates for glyphosate. There are no reported glyphosate aptamers in scientific literature to our knowledge. Glyphosate has a very low molecular weight (169.07 g/mol). Typically smaller molecules are more difficult to generate aptamers against and this maybe one reason why no glyphosate aptamers have been reported to date.

This study resulted in the generation of GLY04, an oligonucleotide that showed binding to glyphosate at an equilibrium dissociation constant of  $\sim 150$  nM. It is feasible that GLY04 could be integrated into a biosensing system for rapid detection of glyphosate in the field. Due to the small molecular size of glyphosate, it would be recommended to use an aptasensor design that does not rely on a mass difference of the target-aptamer complex to the aptamer alone. A colorimetric sensor utilising gold nanoparticles (AuNP) may provide such a system. Gold nanoparticles have some unique characteristics that lend themselves well to colorimetric sensor platforms, such as local surface plasmon resonance (LSPR). The LSPR of AuNP facilitate a phenomenon whereby clearly distinguishable colours in the visible spectrum are produced dependant on whether are AuNP are dispersed in solution or aggregated (Alsager et al., 2018). As ssDNA absorbs non-specifically onto the AuNP's surface, this sensor works on the principle that absorption of ssDNA onto the AuNP surface results in the nanoparticles being protected from salt-induced aggregation. When the target molecule is introduced into the solution, the DNA dissociates from the surface of the nanoparticle due to binding affinity for the target molecule. This creates different levels of aggregation, and hence colour, for the same level of salt concentrations depending on the concentration of the target molecule being present. A scheme such as this could be implemented for the detection of glyphosate.

Attempts to generate aptamers that could bind to technical nonylphenol equivalents were unsuccessful in this study. Five candidate oligonucleotide sequences were selected as a result of twenty rounds of selection. No binding to the target molecule was evident in a fluorescent bead assay or using MST (although only the top candidate was assessed using this method). It was hypothesized that the lack of binding of the chosen candidates was due to these candidates being favourably amplified sequences as opposed to true binders. The hypothesized reasons for the SELEX resulting in unsuccessful outcomes were (1) the nature of the target molecule, (2) the type of SELEX methodology employed, (3) methodological flaws in the washing step and (4) the number of selection cycles needed to generate candidates. The unsuccessful generation of candidates for nonylphenol and the difficulties experienced during this selection prompted the use of HT-SELEX for the selection of oxybenzone aptamer candidates.

An aptamer to oxybenzone which had an equilibrium dissociation constant of  $\sim 107.5$  nM was generated using the extra detergent strategy. Based on frequency, the OXY-ED7-C1 aptamer was the highest ranked candidate from this strategy in Round 7. OXY-ED7-C1 had a high proportion of thymine nucleotides in the N40 region (70.7%) which was notably higher than most other candidates which were characterised using MST. A high proportion of thymine was expected in final candidates due to the biased composition of the starting library. The proportion of thymine in OXY-ED7-C1 was similar to the average proportion in the starting library. This indicates that a high proportion of thymine was potentially beneficial for binding the target. If high thymine was not beneficial for binding oxybenzone then oligonucleotides with this characteristic should have been selected against and removed from the pool. If it was not beneficial then we would also expect the proportion of thymine nucleotides in highly abundant sequences to be closer to random (25%) near the end of SELEX, which was not the

case. An interesting observation was that another candidate (OXY-N7-C1) which was tested for binding affinity had a similar sequence composition to OXY-ED7-C1. No binding was evident of OXY-N7-C1 to its target molecule and it was noticed that a six nucleotide motif (TTATTG) which formed a key structure in OXY-ED7-C1 was missing from OXY-N7-C1. This may indicate that this motif and subsequent secondary structure might be involved in target binding. OXY-ED7-C1 has a  $K_d$  which is sufficient for detecting environmentally relevant concentrations of oxybenzone in water samples. Further characterisation of the aptamer to determine its binding mechanism, such as nuclear magnetic resonance spectroscopy, x-ray crystallography and circular dichroism spectroscopy are required. Additional characterisation methods for aptamer binding are recommended to validate the  $K_d$ . If OXY-ED7-C1 was to be used in a sensing system for detecting oxybenzone levels in seawater then studies of the aptamer function in different salt concentrations and sample matrices would be necessary. While successful target binding was demonstrated in laboratory conditions, more research would be needed to validate that the aptamer also works in complex field matrices.

## **7.3 Evolution of Libraries Using Different Selection Strategies**

Through the use of HT-SELEX, this PhD study makes a novel and significant contribution to the scientific literature on generating small molecule aptamers. Although limitations of the study were identified, the lessons learned from the experimental design and the development of a bioinformatics pipeline provide useful recommendations to any researcher attempting to select aptamers of this type. For any HT-SELEX experiment, initial quality control in the form of HT sequencing of a starting library aliquot is essential. This is to ensure that the nucleotide composition of the starting library is roughly equal. Other quality control measures include assessing the occurrence of polyX motifs and the occurrence of primer binding sites

in the random region of the library. It is recommended that when oligonucleotide libraries are intended to be used in SELEX, manually mixing of the phosphoramidites is requested from the supplier. Moreover, a ratio that takes into account the different coupling efficiency and molecular mass of each phosphoramidite should be specified. These steps should ensure a roughly equal composition of each nucleotide. Another recommendation from this study would be to use a longer read length for the high throughput sequencing step of the process, such as 100 bp compared to 58 bp. This enables the reads to encompass both of the flanking primers as opposed to a single primer. A longer read length would make data processing much easier. Lastly, the design of a new set of flanking primer sequences which is not similar to any previously used primer sequences is also recommended. This would ensure that previously generated oligonucleotides being used in the laboratory can't be amplified and thus contaminate the experimental library.

Higher enrichment of oligonucleotides was demonstrated in the selection strategy with the highest stringency. The strategy that resulted in an aptamer that showed affinity to its target was another high stringency strategy using extra detergent in the wash step. In this strategy, the concentration of IGEPAL-CA630 (a non-ionic surfactant) in the washing buffer was iteratively increased as SELEX progressed. It was hypothesized that increasing detergent concentration would increase selection pressure on the library eliminating oligonucleotides with low or moderate affinity to the target at a greater rate and thus increasing the frequency of sequences which do bind. From the results of this study, a hypothesis is presented that the strategies with the higher stringencies generate more candidates that would exhibit binding to oxybenzone and that the binding affinity would be better than oligonucleotides generated using the standard strategy. This is based on the higher enrichment and resultant aptamer demonstrated in this strategy. By testing only one sequence from the standard strategy and

a few from the other strategies, it is difficult to definitely determine which strategy resulted in better aptamer candidates. Additional research is required to determine whether higher stringency strategies repeatedly result in better SELEX outcomes.

It should be noted that the room temperature, negative selection and counter selection strategies had an identical detergent concentration to the extra detergent strategy but also had additional selection pressures. Therefore, aptamers that successfully bind to oxybenzone would have been expected using these strategies also. An alternative hypothesis to the one presented above is that it is possible that the selection pressure in these strategies was too stringent and therefore binders at a low level were eliminated before they had a chance to enrich. Both the negative and counter selection strategies had a negative selection event at Round 3, which may have produced a bottleneck effect on the library. This is supported by the fact that additional PCR cycles were needed to produce enough DNA for the next selection round in these strategies. Additional PCR cycles may have also favoured the selection of parasitic sequences which favourably amplified but did not bind with high affinity to the target molecule. Neither of the above hypotheses can be confirmed without additional characterisation of a larger number of sequences from each strategy.

There is no doubt that the bias in the nucleotide composition of the starting library will have had an influence on the experimental results presented. The sequence space which was explored would have been limited compared to a library where the nucleotide composition was roughly equal. A smaller sequence space certainly limits the chance of successful SELEX outcomes. However that being said, the enrichment and frequency of candidates in the selection strategies compared to the control libraries was markedly higher. The number of unique sequences in the libraries reduced as SELEX progressed, an indication that selection

for favourable sequences was occurring. This same pattern was not observed in the control libraries. These results indicate that SELEX is a robust process. An aptamer with high affinity was selected from one of the strategies even though the initial library was highly biased. If the library had a roughly even distribution of nucleotides then multiple strategies may have yielded binding aptamers. The only way to determine if the results presented here apply to an evenly distributed library are to repeat the experiment with a library where extensive quality control has been performed prior to SELEX.

## 7.4 Future Directions

Future directions would include repeating the HT SELEX study using the quality control measures described above. I would recommend comparing three strategies: the standard strategy (least stringent), the extra detergent strategy (medium stringency and resulted in the only successful aptamer) and the counter selection strategy (most stringent). I would run the same experiment but in triplicate for each strategy. The triplicate would consist of three different starting libraries from three different manufacturers. Completing the experiment in triplicate would allow the assessment of the average enrichment across the three libraries and would provide further statistical robustness. For aptamer characterisation, I would test at least ten candidates from each strategy for binding affinity and compare the average binding affinity. By showing repeatedly that a certain selection strategy results in higher enrichment is not enough as enrichment does not always correlate with binding affinity. If higher affinity aptamers were able to be selected from the same strategy consistently (*i.e.* in triplicate), then this would definitively indicate the most effective strategy.

The bioinformatic pipeline developed as part of this study was effective for processing large numbers of FASTQ files (pertaining to large numbers of individual libraries) and then analysing



the enrichment and frequency of candidates. The pipeline relies on FASTAptamer for the enrichment and frequency calculations but enables comparisons of these metrics across different SELEX strategies. Assessing the available software packages for analysing HT-SELEX data, it was determined that many of these programs have significant functionality issues. Some programs such as COMPAS and PATERNITY-Seq are proprietary, limiting their use in research. Many open-source programs (such as APTANI or AptaSUITE) are not regularly maintained or updated, have functionality issues and the documentation is not comprehensive. The use of high-throughput sequencing data to select aptamers is only going to increase in the future and so further research needs to be completed on developing new solutions for analysing this data. The pipeline developed herein is only concerned with primary sequence data and does not use 2D or 3D structure prediction to help select better aptamer candidates. Three-dimensional structure prediction of ssDNA is not well developed. With advancements in this area, molecular docking simulations could enable SELEX to be performed truly *in-silico*. An easy, consistent and reliable method for predicting 3D ssDNA structure is essential for the realization of *in-silico* SELEX .

In order to generate a successful aptamer for nonylphenol, there are many steps which could be taken to improve the outcomes of SELEX. A high number of selection rounds (20) and the use of strong selection pressure early in SELEX were two critical mistakes when attempting selection for a nonylphenol aptamer. In addition, a technical mixture of different nonylphenol molecules was used making it difficult to isolate binding candidates. I believe these decisions resulted in sequences which were favourable for PCR but did not bind to nonylphenol. If repeated, I would recommend completing two SELEX experiments in parallel, an experiment with no detergent in the binding and washing buffer and a second experiment with a very low concentration of detergent in the last two SELEX rounds only. The chemical structure of

nonylphenol is nearly identical to that of the detergent which was used (IGEPAL-CA630) and therefore I would select a different detergent which would increase the selection pressure without removing sequences which bind. I would limit the number of SELEX rounds to a maximum of ten, but would attempt to select the aptamer in fewer selection rounds through the use of HT-SELEX and the bioinformatics pipeline generated. The coupling of nonylphenol to an affinity matrix was also detrimental as the molecule only has a single functional (hydroxyl) group for conjugation. If the above steps were not successful in generating an aptamer capable of binding nonylphenol, then I would use capture-SELEX which relies on immobilization of the library rather than the target molecule. By adopting these recommendations, I believe an aptamer capable of binding nonylphenol would be generated.

As glyphosate is the most used agricultural chemical in the world and its effects are highly debated, it presents a very good emerging organic contaminant which could be monitored on a regular basis as an indicator of chemical load in a waterway. The development of an appropriate bio-sensor platform for GLY-04 and testing of the aptasensor in the field would enable regional councils and other stakeholders to regularly monitor the glyphosate concentration in our environment. By validating an aptasensor in the field and demonstrating its efficacy then future adoption of aptasensors for environmental monitoring would be possible. At this stage, aptamers cannot compete with large scale screening of molecules using chromatography coupled with mass spectrometry. Screening of hundreds, even thousands of similarly structured compounds is possible using gas chromatography or liquid chromatography coupled with mass spectrometry. This level of multiplexing is not achievable using aptamer-based technology at present. That is not to say that aptamers are not a useful technology for monitoring environmental compounds. They fit in a particular niche where they can be used for compounds that are difficult to monitor using mass spectrometry due to

their complex extraction procedures. Glyphosate is a perfect example of this. Glyphosate (and metabolites) don't fit into standard screens for environmental compounds as they are difficult to extract, meaning it is cost prohibitive to test for them. An aptasensor works very well in this situation because knowledge of only a single compound (or possibly 3-4 compounds if including metabolites) is required. There are other examples of this, *for example*, nonylphenol and oxybenzone don't fall into standard screening methods either. It is in these situations where aptamers present a novel detection technology which can satisfy the requirements of being quick, cost effective, portable and robust. I am confident aptasensors will complement traditional detection methods used for detecting EOC's in the future and will provide a vital tool in our understanding of the level of chemical contamination in our natural environment.

# Reference List

- 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>
- Allnutt, T. R., Wade, B., Quinn, T. P., Richardson, M. F., & Crowley, T. M. (2018). Shortlisting Aptamer Candidates from HT-SELEX data. *Aptamers*, 2, 36–44.
- Alsager, O. A., Alotaibi, K. M., Alswieleh, A. M., & Alyamani, B. J. (2018). Colorimetric Aptasensor of Vitamin D3: A Novel Approach to Eliminate Residual Adhesion between Aptamers and Gold Nanoparticles. *Scientific Reports*, 8(1), 1–12. <https://doi.org/10.1038/s41598-018-31221-y>
- Alsager, O. A., Kumar, S., Willmott, G. R., McNatty, K. P., & Hodgkiss, J. M. (2014). Small molecule detection in solution via the size contraction response of aptamer functionalized nanoparticles. *Biosensors and Bioelectronics*, 57, 262–268. <https://doi.org/10.1016/j.bios.2014.02.004>
- Annett, R., Habibi, H. R., & Hontela, A. (2014). Impact of glyphosate and glyphosate-based herbicides on the freshwater environment. *Journal of Applied Toxicology: JAT*, 34(5), 458–479. <https://doi.org/10.1002/jat.2997>
- ANZECC. (2000). *Australian and New Zealand Guidelines for Fresh and Marine Water Quality: Aquatic Ecosystems—Rationale and Background Information*. <https://www.waterquality.gov.au/sites/default/files/documents/anzecc-armcanz-2000-guidelines-vol2.pdf>
- Arditsoglou, A., & Voutsas, D. (2008). Determination of phenolic and steroid endocrine disrupting compounds in environmental matrices. *Environmental Science and Pollution Research*, 15(3), 228–236. <https://doi.org/10.1065/espr2007.12.459>
- Arvand, M., & Mirroshandel, A. A. (2017). Highly-sensitive aptasensor based on fluorescence resonance energy transfer between l-cysteine capped ZnS quantum dots and graphene oxide sheets for the determination of edifenphos fungicide. *Biosensors and Bioelectronics*, 96, 324–331. <https://doi.org/10.1016/j.bios.2017.05.028>
- ASTM International. (2017). *ASTM D7065-17, Standard Test Method for Determination of Nonylphenol, Bisphenol A, p-tert-Octylphenol, Nonylphenol Monoethoxylate and Nonylphenol Diethoxylate in Environmental Waters by Gas Chromatography Mass Spectrometry*. ASTM International, West Conshohocken, PA. <http://www.astm.org/cgi-bin/resolver.cgi?D7065>
- Bala, R., Kumar, M., Bansal, K., Sharma, R. K., & Wangoo, N. (2016). Ultrasensitive aptamer biosensor for malathion detection based on cationic polymer and gold nanoparticles. *Biosensors and Bioelectronics*, 85, 445–449. <https://doi.org/10.1016/j.bios.2016.05.042>
- Balmer, M. E., Buser, H.-R., Müller, M. D., & Poiger, T. (2005). Occurrence of Some Organic UV Filters in Wastewater, in Surface Waters, and in Fish from Swiss Lakes. *Environmental Science & Technology*, 39(4), 953–962. <https://doi.org/10.1021/es040055r>

- Barber, T. (2012). *The effects of Nufarm Glyphosate Gold on freshwater invertebrate communities: MSc in Ecology, Massey University, Palmerston North* [Thesis, Massey University]. <https://mro.massey.ac.nz/handle/10179/4238>
- Battaglin, W. A., Meyer, M. T., Kuivila, K. M., & Dietze, J. E. (2014). Glyphosate and Its Degradation Product AMPA Occur Frequently and Widely in U.S. Soils, Surface Water, Groundwater, and Precipitation. *JAWRA Journal of the American Water Resources Association*, 50(2), 275–290. <https://doi.org/10.1111/jawr.12159>
- Berezhnoy, A., Stewart, C. A., Mcnamara II, J. O., Thiel, W., Giangrande, P., Trinchieri, G., & Gilboa, E. (2012). Isolation and Optimization of Murine IL-10 Receptor Blocking Oligonucleotide Aptamers Using High-throughput Sequencing. *Molecular Therapy*, 20(6), 1242–1250. <https://doi.org/10.1038/mt.2012.18>
- Bernot, M. J., Bernot, R. J., & Matthaei, C. D. (2019). Emerging organic contaminants (EOCs) in freshwaters in Dunedin, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, 53(1), 3–14. <https://doi.org/10.1080/00288330.2018.1457062>
- Blank, M. (2016). Next-Generation Analysis of Deep Sequencing Data: Bringing Light into the Black Box of SELEX Experiments. In G. Mayer (Ed.), *Nucleic Acid Aptamers: Selection, Characterization, and Application* (pp. 85–95). Springer. [https://doi.org/10.1007/978-1-4939-3197-2\\_7](https://doi.org/10.1007/978-1-4939-3197-2_7)
- Bock, L., Griffin, L., Latham, J., Vermaas, E., & Toole, J. (1992). Selection of Single-Stranded-Dna Molecules That Bind and Inhibit Human Thrombin. *Nature*, 355(6360), 564–566. <https://doi.org/10.1038/355564a0>
- Breitsprecher, D., Schlinck, N., Witte, D., Duhr, S., Baaske, P., & Schubert, T. (2016). Aptamer Binding Studies Using MicroScale Thermophoresis. In G. Mayer (Ed.), *Nucleic Acid Aptamers: Selection, Characterization, and Application* (pp. 99–111). Springer. [https://doi.org/10.1007/978-1-4939-3197-2\\_8](https://doi.org/10.1007/978-1-4939-3197-2_8)
- Bruning, J. (2017). *Chemical Drains: New Zealand's new NPS-FW freshwater policy is not watertight*. Bay of Plenty Regional Council.
- Bruno, J. G. (1997). In Vitro Selection of DNA to Chloroaromatics Using Magnetic Microbead-Based Affinity Separation and Fluorescence Detection. *Biochemical and Biophysical Research Communications*, 234(1), 117–120. <https://doi.org/10.1006/bbrc.1997.6517>
- Cadwell, R. C., & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods and Applications*, 2(1), 28–33. <https://doi.org/10.1101/gr.2.1.28>
- Cai, S., Yan, J., Xiong, H., Liu, Y., Peng, D., & Liu, Z. (2018). Investigations on the interface of nucleic acid aptamers and binding targets. *Analyst*, 143(22), 5317–5338. <https://doi.org/10.1039/C8AN01467A>
- Cailleaud, K., Forget-Leray, J., Souissi, S., Lardy, S., Augagneur, S., & Budzinski, H. (2007). Seasonal variation of hydrophobic organic contaminant concentrations in the water-column of the Seine Estuary and their transfer to a planktonic species *Eurytemora affinis* (Calanoïd, copepod). Part 2: Alkylphenol-polyethoxylates. *Chemosphere*, 70(2), 281–287. <https://doi.org/10.1016/j.chemosphere.2007.06.012>
- Carr, G. M., & Neary, J. P. (2006). *Water Quality for Ecosystem and Human Health*.
- Chen, S., Huang, T., Zhou, Y., Han, Y., Xu, M., & Gu, J. (2017). AfterQC: Automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics*, 18(3), 80. <https://doi.org/10.1186/s12859-017-1469-3>
- Coronado, M., De Haro, H., Deng, X., Rempel, M. A., Lavado, R., & Schlenk, D. (2008). Estrogenic activity and reproductive effects of the UV-filter oxybenzone (2-hydroxy-

- 4-methoxyphenyl-methanone) in fish. *Aquatic Toxicology (Amsterdam, Netherlands)*, 90(3), 182–187. <https://doi.org/10.1016/j.aquatox.2008.08.018>
- Coupe, R. H., Kalkhoff, S. J., Capel, P. D., & Gregoire, C. (2012). Fate and transport of glyphosate and aminomethylphosphonic acid in surface waters of agricultural basins. *Pest Management Science*, 68(1), 16–30. <https://doi.org/10.1002/ps.2212>
- Cuhra, M., Bøhn, T., & Cuhra, P. (2016). Glyphosate: Too Much of a Good Thing? *Frontiers in Environmental Science*, 4. <https://doi.org/10.3389/fenvs.2016.00028>
- Cuhra, M., Traavik, T., & Bøhn, T. (2013). Clone- and age-dependent toxicity of a glyphosate commercial formulation and its active ingredient in *Daphnia magna*. *Ecotoxicology (London, England)*, 22(2), 251–262. <https://doi.org/10.1007/s10646-012-1021-1>
- Dag, O., Dolgun, A., & Konar, N., Meric. (2018). onewaytests: An R Package for One-Way Tests in Independent Groups Designs. *The R Journal*, 10(1), 175. <https://doi.org/10.32614/RJ-2018-022>
- Darmostuk, M., Rimpelova, S., Gbelcova, H., & Ruml, T. (2015). Current approaches in SELEX: An update to aptamer selection technology. *Biotechnology Advances*, 33(6, Part 2), 1141–1161. <https://doi.org/10.1016/j.biotechadv.2015.02.008>
- DiNardo, J. C., & Downs, C. A. (2018). Dermatological and environmental toxicological impact of the sunscreen ingredient oxybenzone/benzophenone-3. *Journal of Cosmetic Dermatology*, 17(1), 15–19. <https://doi.org/10.1111/jocd.12449>
- Downs, C. A., Kramarsky-Winter, E., Segal, R., Fauth, J., Knutson, S., Bronstein, O., Ciner, F. R., Jeger, R., Lichtenfeld, Y., Woodley, C. M., Pennington, P., Cadenas, K., Kushmaro, A., & Loya, Y. (2016). Toxicopathological Effects of the Sunscreen UV Filter, Oxybenzone (Benzophenone-3), on Coral Planulae and Cultured Primary Cells and Its Environmental Contamination in Hawaii and the U.S. Virgin Islands. *Archives of Environmental Contamination and Toxicology*, 70(2), 265–288. <https://doi.org/10.1007/s00244-015-0227-7>
- Duke, S. O., & Powles, S. B. (2008). Glyphosate: A once-in-a-century herbicide. *Pest Management Science*, 64(4), 319–325. <https://doi.org/10.1002/ps.1518>
- Duong, C. N., Ra, J. S., Cho, J., Kim, S. D., Choi, H. K., Park, J.-H., Kim, K. W., Inam, E., & Kim, S. D. (2010). Estrogenic chemicals and estrogenicity in river waters of South Korea and seven Asian countries. *Chemosphere*, 78(3), 286–293. <https://doi.org/10.1016/j.chemosphere.2009.10.048>
- ECHA. (2019). *Oxybenzone: Ecotoxicological summary*. <https://echa.europa.eu/registration-dossier/-/registered-dossier/5515/6/1>
- Eckert, K. A., & Kunkel, T. A. (1990). High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Research*, 18(13), 3739–3744. <https://doi.org/10.1093/nar/18.13.3739>
- Ellington, A. D., & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346(6287), 818–822. <https://doi.org/10.1038/346818a0>
- Ellington, Andrew D., & Szostak, J. W. (1992). Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature*, 355(6363), 850. <https://doi.org/10.1038/355850a0>
- Espelund, M., Stacy, R. A., & Jakobsen, K. S. (1990). A simple method for generating single-stranded DNA probes labeled to high activities. *Nucleic Acids Research*, 18(20), 6157–6158.

- European Chemical Bureau. (2002). *European Union Risk Assessment Report for 4-nonylphenol (branched) and nonylphenol*.  
[https://mobil.bfr.bund.de/cm/343/4\\_nonylphenol\\_und\\_nonylphenol.pdf](https://mobil.bfr.bund.de/cm/343/4_nonylphenol_und_nonylphenol.pdf)
- Ewing, B., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, 8(3), 186–194.  
<https://doi.org/10.1101/gr.8.3.186>
- Famulok, M., & Mayer, G. (2011). Aptamer Modules as Sensors and Detectors. *Accounts of Chemical Research*, 44(12), 1349–1358. <https://doi.org/10.1021/ar2000293>
- Fei, A., Liu, Q., Huan, J., Qian, J., Dong, X., Qiu, B., Mao, H., & Wang, K. (2015). Label-free impedimetric aptasensor for detection of femtomole level acetamiprid using gold nanoparticles decorated multiwalled carbon nanotube-reduced graphene oxide nanoribbon composites. *Biosensors and Bioelectronics*, 70, 122–129.  
<https://doi.org/10.1016/j.bios.2015.03.028>
- Folmar, L. C., Sanders, H. O., & Julin, A. M. (1979). Toxicity of the herbicide glyphosate and several of its formulations to fish and aquatic invertebrates. *Archives of Environmental Contamination and Toxicology*, 8(3), 269–278.  
<https://doi.org/10.1007/BF01056243>
- Games, P. A., & Howell, J. F. (1976). Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study. *Journal of Educational Statistics*, 1(2), 113–125. JSTOR. <https://doi.org/10.2307/1164979>
- Geiger, A., Burgstaller, P., von der Eltz, H., Roeder, A., & Famulok, M. (1996). RNA Aptamers That Bind L-Arginine with Sub-Micromolar Dissociation Constants and High Enantioselectivity. *Nucleic Acids Research*, 24(6), 1029–1036.  
<https://doi.org/10.1093/nar/24.6.1029>
- Gopinath, S. C. B. (2007). Methods developed for SELEX. *Analytical and Bioanalytical Chemistry*, 387(1), 171–182. <https://doi.org/10.1007/s00216-006-0826-2>
- Grey, D., & Sadoff, C. W. (2007). Sink or Swim? Water security for growth and development. *Water Policy*, 9(6), 545–571. <https://doi.org/10.2166/wp.2007.021>
- Hageman, K. J., Aebig, C. H. F., Luong, K. H., Kaserzon, S. L., Wong, C. S., Reeks, T., Greenwood, M., Macaulay, S., & Matthaei, C. D. (2019). Current-use pesticides in New Zealand streams: Comparing results from grab samples and three types of passive samplers. *Environmental Pollution*, 254, 112973.  
<https://doi.org/10.1016/j.envpol.2019.112973>
- Hall, B., Micheletti, J. M., Satya, P., Ogle, K., Pollard, J., & Ellington, A. D. (2009). Design, Synthesis, and Amplification of DNA Pools for In Vitro Selection. *Current Protocols in Molecular Biology*, 88(1), 24.2.1–24.2.27.  
<https://doi.org/10.1002/0471142727.mb2402s88>
- Hamming, R. W. (1950). Error Detecting and Error Correcting Codes. *Bell System Technical Journal*, 29(2), 147–160. <https://doi.org/10.1002/j.1538-7305.1950.tb00463.x>
- Han, Q., Qiao, P., Song, Y., Zhang, J., & Chen, Q. (2015). *Aptamer specifically bound with glyphosate and application* (Patent No. CN105039347 (A)).  
<https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20151111&DB=&locale=&CC=CN&NR=105039347A&KC=A&ND=1>
- He, J., Liu, Y., Fan, M., & Liu, X. (2011). Isolation and identification of the DNA aptamer target to acetamiprid. *Journal of Agricultural and Food Chemistry*, 59(5), 1582–1586. Scopus. <https://doi.org/10.1021/jf104189g>

- Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*, 56(2), 61–64, 66, 68, passim. <https://doi.org/10.2144/000114133>
- Hermanson, G. (2013). *Bioconjugate Techniques*. Elsevier. <https://doi.org/10.1016/C2009-0-64240-9>
- Higuchi, R. G., & Ochman, H. (1989). Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. *Nucleic Acids Research*, 17(14), 5865–5865. <https://doi.org/10.1093/nar/17.14.5865>
- Hoinka, J., Backofen, R., & Przytycka, T. M. (2018). AptaSUIITE: A Full-Featured Bioinformatics Framework for the Comprehensive Analysis of Aptamers from HT-SELEX Experiments. *Molecular Therapy. Nucleic Acids*, 11, 515–517. <https://doi.org/10.1016/j.omtn.2018.04.006>
- Hoinka, J., Berezhnoy, A., Dao, P., Sauna, Z. E., Gilboa, E., & Przytycka, T. M. (2015). Large scale analysis of the mutational landscape in HT-SELEX improves aptamer discovery. *Nucleic Acids Research*, 43(12), 5699–5707. <https://doi.org/10.1093/nar/gkv308>
- Hong, K. L., & Sooter, L. J. (2017). In Vitro Selection of a Single-Stranded DNA Molecular Recognition Element against the Pesticide Fipronil and Sensitive Detection in River Water. *International Journal of Molecular Sciences*, 19(1). <https://doi.org/10.3390/ijms19010085>
- IUPAC. (2020a). *cypermethrin* (Ref: OMS 2002). <https://sitem.herts.ac.uk/aeru/ppdb/en/Reports/197.htm>
- IUPAC. (2020b). *Terbutylazine*. University of Hertfordshire. <https://sitem.herts.ac.uk/aeru/ppdb/en/Reports/623.htm>
- Jenison, R. D., Gill, S. C., Pardi, A., & Polisky, B. (1994). High-resolution molecular discrimination by RNA. *Science*, 263(5152), 1425–1429. <https://doi.org/10.1126/science.7510417>
- Jeon, H.-K., Chung, Y., & Ryu, J.-C. (2006). Simultaneous determination of benzophenone-type UV filters in water and soil by gas chromatography-mass spectrometry. *Journal of Chromatography. A*, 1131(1–2), 192–202. <https://doi.org/10.1016/j.chroma.2006.07.036>
- Jo, M., Ahn, J.-Y., Lee, J., Lee, S., Hong, S. W., Yoo, J.-W., Kang, J., Dua, P., Lee, D.-K., Hong, S., & Kim, S. (2011). Development of single-stranded DNA aptamers for specific Bisphenol a detection. *Oligonucleotides*, 21(2), 85–91. <https://doi.org/10.1089/oli.2010.0267>
- Johnson, P. E., & Slavkovic, S. (2018). Isothermal titration calorimetry studies of aptamer-small molecule interactions: Practicalities and pitfalls – Aptamers. *Aptamers*, 2, 45–51.
- Jonkers, N., Kohler, H.-P. E., Dammshäuser, A., & Giger, W. (2009). Mass flows of endocrine disruptors in the Glatt River during varying weather conditions. *Environmental Pollution*, 157(3), 714–723. <https://doi.org/10.1016/j.envpol.2008.11.029>
- Kim, A.-R., Ha, N.-R., Jung, I.-P., Kim, S.-H., & Yoon, M.-Y. (2019). Development of a ssDNA aptamer system with reduced graphene oxide (rGO) to detect nonylphenol ethoxylate in domestic detergent. *Journal of Molecular Recognition*, 32(3), e2764. <https://doi.org/10.1002/jmr.2764>



- Kim, S. H., Thanh Thoa, T. T., & Gu, M. B. (2019). Aptasensors for Environmental Monitoring of Contaminants in Water and Soil. *Current Opinion in Environmental Science & Health*. <https://doi.org/10.1016/j.coesh.2019.09.003>
- Kumar, S. (2014). *Developing aptamers capable of binding oestradiol or bisphenol A: A thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of Doctor of Philosophy*. Thesis PhD--Victoria University of Wellington.
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- 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>
- Lauridsen, L. H., Doessing, H. B., Long, K. S., & Nielsen, A. T. (2018). A Capture-SELEX Strategy for Multiplexed Selection of RNA Aptamers Against Small Molecules. In M. K. Jensen & J. D. Keasling (Eds.), *Synthetic Metabolic Pathways: Methods and Protocols* (pp. 291–306). Springer New York. [https://doi.org/10.1007/978-1-4939-7295-1\\_18](https://doi.org/10.1007/978-1-4939-7295-1_18)
- Lei, Z., Zhang, C., Liu, Y., Wang, L., Hu, Q., & Liu, X. (2012). Selection of chlorpyrifos-binding ssDNA aptamer by SELEX. *Jiangsu Journal of Agricultural Sciences*. [http://en.cnki.com.cn/Article\\_en/CJFDTotat-JSNB201201035.htm](http://en.cnki.com.cn/Article_en/CJFDTotat-JSNB201201035.htm)
- Li, S. (2016). *Selection and characterisation of single-stranded DNA aptamers for triclosan*. <http://researcharchive.vuw.ac.nz/handle/10063/5017>
- Li, W., Ma, Y., Guo, C., Hu, W., Liu, K., Wang, Y., & Zhu, T. (2007). Occurrence and behavior of four of the most used sunscreen UV filters in a wastewater reclamation plant. *Water Research*, 41(15), 3506–3512. <https://doi.org/10.1016/j.watres.2007.05.039>
- Lin, J. S., Kauff, A., Diao, Y., Yang, H., Lawrence, S., & Juengel, J. L. (2016). Creation of DNA aptamers against recombinant bone morphogenetic protein 15. *Reproduction, Fertility and Development*, 28(8), 1164–1171. <https://doi.org/10.1071/RD14409>
- Ling, L. L., Keohavong, P., Dias, C., & Thilly, W. G. (1991). Optimization of the polymerase chain reaction with regard to fidelity: Modified T7, Taq, and vent DNA polymerases. *PCR Methods and Applications*, 1(1), 63–69. <https://doi.org/10.1101/gr.1.1.63>
- Liu, D.-L., Li, Y., Sun, R., Xu, J.-Y., Chen, Y., & Sun, C.-Y. (2020). Colorimetric Detection of Organophosphorus Pesticides Based on the Broad-Spectrum Aptamer. *Journal of Nanoscience and Nanotechnology*, 20(4), 2114–2121. <https://doi.org/10.1166/jnn.2020.17358>
- MacBean, C. (2012). *The Pesticide Manual* (16th ed.). British Crop Protection Council Publications.
- Madianos, L., Tsekenis, G., Skotadis, E., Patsiouras, L., & Tsoukalas, D. (2018). A highly sensitive impedimetric aptasensor for the selective detection of acetamiprid and atrazine based on microwires formed by platinum nanoparticles. *Biosensors and Bioelectronics*, 101, 268–274. <https://doi.org/10.1016/j.bios.2017.10.034>
- Madsen, J. F., Sandstrom, M. W., & Zaugg, S. D. (2002). *Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory; a method supplement for the determination of Fipronil and degradates in water by gas chromatography/mass spectrometry* (USGS Numbered Series No. 2002–462; Open-File Report). <http://pubs.er.usgs.gov/publication/ofr02462>

- Manktelow, D., Stevens, P., J. W., S. G., Park, N. M., Zabkiewicz, J., Teulon, D., & A. R. (2005). *Trends in Pesticide Use in New Zealand: 2004*.
- 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), 491–505. <https://doi.org/10.3390/ijms13010491>
- Marimuthu, C., Tang, T.-H., Tominaga, J., Tan, S.-C., & B. Gopinath, S. C. (2012). Single-stranded DNA (ssDNA) production in DNA aptamer generation. *Analyst*, 137(6), 1307–1315. <https://doi.org/10.1039/C2AN15905H>
- 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>
- Mashima, T., Nishikawa, F., Kamatari, Y. O., Fujiwara, H., Saimura, M., Nagata, T., Kodaki, T., Nishikawa, S., Kuwata, K., & Katahira, M. (2013). Anti-prion activity of an RNA aptamer and its structural basis. *Nucleic Acids Research*, 41(2), 1355–1362. <https://doi.org/10.1093/nar/gks1132>
- McKeague, M., De Girolamo, A., Valenzano, S., Pascale, M., Ruscito, A., Velu, R., Frost, N. R., Hill, K., Smith, M., McConnell, E. M., & DeRosa, M. C. (2015). Comprehensive analytical comparison of strategies used for small molecule aptamer evaluation. *Analytical Chemistry*, 87(17), 8608–8612. <https://doi.org/10.1021/acs.analchem.5b02102>
- McKeague, M., & DeRosa, M. C. (2012). Challenges and Opportunities for Small Molecule Aptamer Development. *Journal of Nucleic Acids*, 1–20. <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>
- Mehta, J., Rouah-Martin, E., Van Dorst, B., Maes, B., Herrebout, W., Scippo, M.-L., Dardenne, F., Blust, R., & Robbens, J. (2012). Selection and characterization of PCB-binding DNA aptamers. *Analytical Chemistry*, 84(3), 1669–1676. <https://doi.org/10.1021/ac202960b>
- Mendonsa, S. D., & Bowser, M. T. (2004). In Vitro Selection of High-Affinity DNA Ligands for Human IgE Using Capillary Electrophoresis. *Analytical Chemistry*, 76(18), 5387–5392. <https://doi.org/10.1021/ac049857v>
- Meyer, M., Loftin, K., Lee, E., Hinshaw, G., Dietze, J., & Scribner, E. (2009). *Determination of Glyphosate, its Degradation Product Aminomethylphosphonic Acid, and Glufosinate, in Water by Isotope Dilution and Online Solid-Phase Extraction and Liquid Chromatography/Tandem Mass Spectrometry* (Techniques and Methods 5–A10). US Geological Survey. <https://pubs.usgs.gov/tm/tm5a10/pdf/tm5a10.pdf>
- Ministry for Primary Industries NZ. (2019). *Glyphosate*. Ministry for Primary Industries NZ. <https://www.mpi.govt.nz/food-safety/food-safety-for-consumers/whats-in-our-food-2/chemicals-and-food/agricultural-compounds-and-residues/glyphosate/>
- Ministry for the Environment NZ. (2015). *Environment Aotearoa 2015*. Ministry for the Environment NZ.
- Ministry for the Environment NZ. (2019). *Environment Aotearoa 2019*. <https://www.mfe.govt.nz/environment-aotearoa-2019>

- Musheev, M. U., & Krylov, S. N. (2006). Selection of aptamers by systematic evolution of ligands by exponential enrichment: Addressing the polymerase chain reaction issue. *Analytica Chimica Acta*, 564(1), 91–96. <https://doi.org/10.1016/j.aca.2005.09.069>
- Myers, J. P., Antoniou, M. N., Blumberg, B., Carroll, L., Colborn, T., Everett, L. G., Hansen, M., Landrigan, P. J., Lanphear, B. P., Mesnage, R., Vandenberg, L. N., vom Saal, F. S., Welshons, W. V., & Benbrook, C. M. (2016). Concerns over use of glyphosate-based herbicides and risks associated with exposures: A consensus statement. *Environmental Health*, 15(1). <https://doi.org/10.1186/s12940-016-0117-0>
- Nguyen Quang, N., Perret, G., & Ducongé, F. (2016). Applications of High-Throughput Sequencing for In Vitro Selection and Characterization of Aptamers. *Pharmaceuticals*, 9(4), 76. <https://doi.org/10.3390/ph9040076>
- Nie, Y., Teng, Y., Li, P., Liu, W., Shi, Q., & Zhang, Y. (2018). Label-free aptamer-based sensor for specific detection of malathion residues by surface-enhanced Raman scattering. *Spectrochimica Acta. Part A, Molecular and Biomolecular Spectroscopy*, 191, 271–276. <https://doi.org/10.1016/j.saa.2017.10.030>
- Nimjee, S. M., Rusconi, C. P., & Sullenger, B. A. (2005). Aptamers: An Emerging Class of Therapeutics. *Annual Review of Medicine*, 56(1), 555–583. <https://doi.org/10.1146/annurev.med.56.062904.144915>
- NZ EPA. (2015). *About glyphosate*. NZ EPA. <https://www.epa.govt.nz/assets/Uploads/Documents/Everyday-Environment/Publications/9b7cc8389c/About-Glyphosate-information-sheet.pdf>
- NZ EPA. (2019). *Nonylphenol, 4-branched*. <https://www.epa.govt.nz/database-search/approved-hazardous-substances-with-controls/view/8271>
- Oketola, A. A., & Fagbemigun, T. K. (2013). Determination of Nonylphenol, Octylphenol and Bisphenol-A in Water and Sediments of Two Major Rivers in Lagos, Nigeria. *Journal of Environmental Protection*, 4(7), 38–45. <https://doi.org/10.4236/jep.2013.47A005>
- Paul, A., Avci-Adali, M., Ziemer, G., & Wendel, H. P. (2009). Streptavidin-coated magnetic beads for DNA strand separation implicate a multitude of problems during cell-SELEX. *Oligonucleotides*, 19(3), 243–254. <https://doi.org/10.1089/oli.2009.0194>
- Peng, X., Yu, Y., Tang, C., Tan, J., Huang, Q., & Wang, Z. (2008). Occurrence of steroid estrogens, endocrine-disrupting phenols, and acid pharmaceutical residues in urban riverine water of the Pearl River Delta, South China. *Science of The Total Environment*, 397(1), 158–166. <https://doi.org/10.1016/j.scitotenv.2008.02.059>
- Peters, G.-J., Verboon, P., & Green, J. (2018). *userfriendlyscience: Quantitative Analysis Made Accessible* (Version 0.7.2) [Computer software]. <https://CRAN.R-project.org/package=userfriendlyscience>
- Pinheiro, V. B., Taylor, A. I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J. C., Wengel, J., Peak-Chew, S.-Y., McLaughlin, S. H., Herdewijn, P., & Holliger, P. (2012). Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science*, 336(6079), 341–344. <https://doi.org/10.1126/science.1217622>
- Porta, M., Puigdomènech, E., Ballester, F., Selva, J., Ribas-Fitó, N., Llop, S., & López, T. (2008). Monitoring concentrations of persistent organic pollutants in the general population: The international experience. *Environment International*, 34(4), 546–561. <https://doi.org/10.1016/j.envint.2007.10.004>
- Qi, Y., Xiu, F.-R., Zheng, M., & Li, B. (2016). A simple and rapid chemiluminescence aptasensor for acetamiprid in contaminated samples: Sensitivity, selectivity and

- mechanism. *Biosensors and Bioelectronics*, 83, 243–249.  
<https://doi.org/10.1016/j.bios.2016.04.074>
- R Core Team. (2019). *R* 3.4.3.
- Radom, F., Jurek, P. M., Mazurek, M. P., Otlewski, J., & Jeleń, F. (2013). Aptamers: Molecules of great potential. *Biotechnology Advances*, 31(8), 1260–1274.  
<https://doi.org/10.1016/j.biotechadv.2013.04.007>
- Rahimizadeh, K., AlShamaileh, H., Fratini, M., Chakravarthy, M., Stephen, M., Shigdar, S., & Veedu, R. N. (2017). Development of Cell-Specific Aptamers: Recent Advances and Insight into the Selection Procedures. *Molecules (Basel, Switzerland)*, 22(12).  
<https://doi.org/10.3390/molecules22122070>
- Rangel, A. E., Chen, Z., Ayele, T. M., & Heemstra, J. M. (2018). In vitro selection of an XNA aptamer capable of small-molecule recognition. *Nucleic Acids Research*, 46(16), 8057–8068. <https://doi.org/10.1093/nar/gky667>
- Rolando, C. A., Garrett, L. G., Baillie, B. R., & Watt, M. S. (2013). A survey of herbicide use and a review of environmental fate in New Zealand planted forests. *New Zealand Journal of Forestry Science*, 43(1), 17. <https://doi.org/10.1186/1179-5395-43-17>
- Roth, A., & Breaker, R. R. (2009). The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annual Review of Biochemistry*, 78(1), 305–334.  
<https://doi.org/10.1146/annurev.biochem.78.070507.135656>
- Ruxton, G. D., & Beauchamp, G. (2008). Time for some a priori thinking about post hoc testing. *Behavioral Ecology*, 19(3), 690–693. <https://doi.org/10.1093/beheco/arn020>
- Sakamoto, T. (2017). NMR study of aptamers. *Aptamers*, 1, 13–18.
- Schütze, T., Wilhelm, B., Greiner, N., Braun, H., Peter, F., Mörl, M., Erdmann, V. A., Lehrach, H., Konthur, Z., Menger, M., Arndt, P. F., & Glöckler, J. (2011). Probing the SELEX process with next-generation sequencing. *PloS One*, 6(12), e29604.  
<https://doi.org/10.1371/journal.pone.0029604>
- Schwarzenbach, R. P., Egli, T., Hofstetter, T. B., von Gunten, U., & Wehrli, B. (2010). Global Water Pollution and Human Health. *Annual Review of Environment and Resources*, 35(1), 109–136. <https://doi.org/10.1146/annurev-enviro-100809-125342>
- Shiklomanov, I. (1993). World fresh water resources. In *Water in Crisis: A Guide to the World's Fresh Water Resources*. Oxford University Press: New York.
- Shue, M.-F., Chen, F.-A., & Chen, T.-C. (2010). Total estrogenic activity and nonylphenol concentration in the Donggang River, Taiwan. *Environmental Monitoring and Assessment*, 168(1), 91–101. <https://doi.org/10.1007/s10661-009-1093-4>
- Stewart, M., Northcott, G., Gaw, S., & Tremblay, L. (2016). *An Update on Emerging Organic Contaminants of Concern for New Zealand with Guidance on Monitoring Approaches for Councils. Prepared by Streamlined Environmental Ltd, Northcott Research Consultants Ltd, University of Canterbury, Cawthron Institute and the University of Auckland for Auckland Council, Greater Wellington Regional Council and Environment Canterbury Regional Council. Auckland Council technical report, TR2016/006.*
- Stewart, M., Olsen, G., Hickey, C. W., Ferreira, B., Jelic, A., Petrovic, M., & Barcelo, D. (2014). A survey of emerging contaminants in the estuarine receiving environment around Auckland, New Zealand. *Science of the Total Environment*, 468, 202–210.  
<https://doi.org/10.1016/j.scitotenv.2013.08.039>
- Stoltenburg, R., Nikolaus, N., & Strehlitz, B. (2012). Capture-SELEX: Selection of DNA Aptamers for Aminoglycoside Antibiotics. *Journal of Analytical Methods in Chemistry*, 2012. <https://doi.org/10.1155/2012/415697>

- 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>
- Szeto, K., & Craighead, H. G. (2014). Devices and approaches for generating specific high-affinity nucleic acid aptamers. *Applied Physics Reviews*, 1(3), 031103. <https://doi.org/10.1063/1.4894851>
- Takahashi, M., Wu, X., Ho, M., Chomchan, P., Rossi, J. J., Burnett, J. C., & Zhou, J. (2016). High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency. *Scientific Reports*, 6. <https://doi.org/10.1038/srep33697>
- Tan, S. Y., Acquah, C., Sidhu, A., Ongkudon, C. M., Yon, L. S., & Danquah, M. K. (2016). SELEX Modifications and Bioanalytical Techniques for Aptamer–Target Binding Characterization. *Critical Reviews in Analytical Chemistry*, 46(6), 521–537. <https://doi.org/10.1080/10408347.2016.1157014>
- Tolle, F., Wilke, J., Wengel, J., & Mayer, G. (2014). By-Product Formation in Repetitive PCR Amplification of DNA Libraries during SELEX. *PLOS ONE*, 9(12), e114693. <https://doi.org/10.1371/journal.pone.0114693>
- 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.
- U.S. Environmental Protection Agency. (2010). *Nonylphenol (NP) and Nonylphenol Ethoxylates (NPEs) Action Plan*. [https://www.epa.gov/sites/production/files/2015-09/documents/rin2070-za09\\_np-npes\\_action\\_plan\\_final\\_2010-08-09.pdf](https://www.epa.gov/sites/production/files/2015-09/documents/rin2070-za09_np-npes_action_plan_final_2010-08-09.pdf)
- Vallée-Bélisle, A., & Plaxco, K. W. (2010). Structure-switching biosensors: Inspired by Nature. *Current Opinion in Structural Biology*, 20(4), 518–526. <https://doi.org/10.1016/j.sbi.2010.05.001>
- Vitali, M., Ensabella, F., Stella, D., & Guidotti, M. (2004). Nonylphenols in freshwaters of the hydrologic system of an Italian district: Association with human activities and evaluation of human exposure. *Chemosphere*, 57(11), 1637–1647. <https://doi.org/10.1016/j.chemosphere.2004.06.029>
- Vogel, M., & Suess, B. (2016). Label-Free Determination of the Dissociation Constant of Small Molecule-Aptamer Interaction by Isothermal Titration Calorimetry. In G. Mayer (Ed.), *Nucleic Acid Aptamers: Selection, Characterization, and Application* (pp. 113–125). Springer. [https://doi.org/10.1007/978-1-4939-3197-2\\_9](https://doi.org/10.1007/978-1-4939-3197-2_9)
- Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., & Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, 467(7315), 555–561. <https://doi.org/10.1038/nature09440>
- Vörösmarty, Charles J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global Water Resources: Vulnerability from Climate Change and Population Growth. *Science*, 289(5477), 284–288. <https://doi.org/10.1126/science.289.5477.284>
- Wang, L., Liu, X., Zhang, Q., Zhang, C., Liu, Y., Tu, K., & Tu, J. (2012). Selection of DNA aptamers that bind to four organophosphorus pesticides. *Biotechnology Letters*, 34(5), 869–874. <https://doi.org/10.1007/s10529-012-0850-6>
- Wang, T., Chen, C., Larcher, L. M., Barrero, R. A., & Veedu, R. N. (2019). Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on

- aptamer development. *Biotechnology Advances*, 37(1), 28–50.  
<https://doi.org/10.1016/j.biotechadv.2018.11.001>
- Weerathunge, P., Behera, B. K., Zihara, S., Singh, M., Prasad, S. N., Hashmi, S., Mariathomas, P. R. D., Bansal, V., & Ramanathan, R. (2019). Dynamic interactions between peroxidase-mimic silver NanoZymes and chlorpyrifos-specific aptamers enable highly-specific pesticide sensing in river water. *Analytica Chimica Acta*, 1083, 157–165. <https://doi.org/10.1016/j.aca.2019.07.066>
- Welch, B. L. (1951). On the Comparison of Several Mean Values: An Alternative Approach. *Biometrika*, 38(3/4), 330–336. JSTOR. <https://doi.org/10.2307/2332579>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <http://ggplot2.org>
- Wickham, H., Francois, R., Henry, L., & Muller, H. (2018). *dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>
- Williams, K. P., & Bartel, D. P. (1995). PCR product with strands of unequal length. *Nucleic Acids Research*, 23(20), 4220–4221.
- Williams, R. M., Crieffield, C. L., Gattu, S., Holland, L. A., & Sooter, L. J. (2014). In vitro selection of a single-stranded DNA molecular recognition element against atrazine. *International Journal of Molecular Sciences*, 15(8), 14332–14347. Scopus. <https://doi.org/10.3390/ijms150814332>
- Xu, S., Yuan, H., Chen, S., Xu, A., Wang, J., & Wu, L. (2012). Selection of DNA aptamers against polychlorinated biphenyls as potential biorecognition elements for environmental analysis. *Analytical Biochemistry*, 423(2), 195–201. <https://doi.org/10.1016/j.ab.2012.01.026>
- Yufa, R., Krylova, S. M., Bruce, C., Bagg, E. A., Schofield, C. J., & Krylov, S. N. (2015). Emulsion PCR Significantly Improves Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures-Based Aptamer Selection: Allowing for Efficient and Rapid Selection of Aptamer to Unmodified ABH2 Protein. *Analytical Chemistry*, 87(2), 1411–1419. <https://doi.org/10.1021/ac5044187>
- Zaugg, S. D., Sandstrom, M. W., Smith, S. G., & Fehlberg, K. M. (1995). *Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory; determination of pesticides in water by C-18 solid-phase extraction and capillary-column gas chromatography/mass spectrometry with selected-ion monitoring* (USGS Numbered Series No. 95–181; Open-File Report). U.S. Geological Survey : Open-File Reports Section/ESIC [distributor]. <http://pubs.er.usgs.gov/publication/ofr95181>
- Zhang, C., Wang, L., Tu, Z., Sun, X., He, Q., Lei, Z., Xu, C., Liu, Y., Zhang, X., Yang, J., Liu, X., & Xu, Y. (2014). Organophosphorus pesticides detection using broad-specific single-stranded DNA based fluorescence polarization aptamer assay. *Biosensors and Bioelectronics*, 55, 216–219. <https://doi.org/10.1016/j.bios.2013.12.020>
- Zhang, L., Rana, I., Shaffer, R. M., Taioli, E., & Sheppard, L. (2019). Exposure to glyphosate-based herbicides and risk for non-Hodgkin lymphoma: A meta-analysis and supporting evidence. *Mutation Research/Reviews in Mutation Research*, 781, 186–206. <https://doi.org/10.1016/j.mrrev.2019.02.001>
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), 3406–3415. <https://doi.org/10.1093/nar/gkg595>

# Appendix A: Materials and Equipment

## A1: General Materials and Equipment

Reagent/equipment	Abbreviation/Common Name	Catalogue number/model	Manufacturer/supplier
Corning® 15 mL PP Centrifuge Tubes, Bulk Packed with CentriStar™ Cap, Sterile, 50/Sleeve, 500/Case	15 mL centrifuge tubes	430791	Corning Inc, Corning, New York, United States
Corning® 50 mL PP Centrifuge Tubes, Conical Bottom with CentriStar™ Cap, Bulk Packed, Sterile, 25/Pack, 500/Case	50 mL centrifuge tubes	430829	Corning Inc, Corning, New York, United States
Axygen® 1.5 mL MaxyClear Snaplock Microcentrifuge Tube, Polypropylene, Clear, Nonsterile, 500 Tubes/Pack, 10 Packs/Case	1.5 ml microcentrifuge tube	MCT-150-C	Corning Inc, Corning, New York, United States
Axygen® 0.6 mL MaxyClear Snaplock Microcentrifuge Tube, Polypropylene, Clear, Nonsterile, 1000 Tubes/Pack, 10 Packs/Case	0.6 mL microcentrifuge tube	MCT-060-C	Corning Inc, Corning, New York, United States
Axygen® 1.7 mL MaxyClear Snaplock Microcentrifuge Tube, Polypropylene, Amber Nonsterile,	1.7 mL amber microcentrifuge tube	MCT-175-X	Corning Inc, Corning, New York, United States

500 Tubes/Pack, 10 Packs/Case			
UltraPure™ DNase/RNase-Free Distilled Water	Ultra-pure water	10977015	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Sodium Chloride	NaCl	BP358-1	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Magnesium Chloride (anhydrous)	MgCl	M8266-100G	Sigma Aldrich, St Louis, Missouri, United States
Calcium Chloride	CaCl	102378	Merck Millipore, Burlington, Massachusetts, United States
Potassium Chloride	KCl	P9541-500G	Sigma Aldrich, St Louis, Missouri, United States
Tris-Hydrochloride Molecular Biology Grade	Tris-HCL	A3452,1000	PanReac AppliChem, Darmstadt, Germany
IGEPAL® CA-630	IGEPAL	I8896	Sigma Aldrich, St Louis, Missouri, United States
<i>Taq</i> DNA Polymerase PCR Buffer (10X)	PCR Buffer	18067017	Thermo Fisher Scientific, Waltham, Massachusetts, United States
MgCl <sub>2</sub> (magnesium chloride) (25 mM)	MgCl <sub>2</sub>	R0971	Thermo Fisher Scientific, Waltham, Massachusetts, United States
dNTP Set 100 mM Solutions	dNTP's	R0181	Thermo Fisher Scientific, Waltham, Massachusetts, United States
<i>Taq</i> DNA Polymerase, recombinant (5 U/μL)	<i>Taq</i> polymerase	EP0401	Thermo Fisher Scientific, Waltham, Massachusetts, United States
DNA Gel Loading Dye (6X)		R0611	Thermo Fisher Scientific, Waltham, Massachusetts, United States



SYBR™ Green I Nucleic Acid Gel Stain, 10,000X concentrate in DMSO	SYBR Green	S7563	Thermo Fisher Scientific, Waltham, Massachusetts, United States
SeaKem® LE Agarose	Agarose gel	50004	Lonza Group AG, Basel, Switzerland
GeneRuler Low Range DNA Ladder, ready-to-use	DNA ladder	SM1193	Thermo Fisher Scientific, Waltham, Massachusetts, United States
TAE Buffer (Tris-acetate-EDTA) (50X)	Gel electrophoresis buffer	B49	Thermo Fisher Scientific, Waltham, Massachusetts, United States
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	Gel Company, San Francisco, California, United States
MinElute Gel Extraction Kit	Gel extraction kit	28604	Qiagen, Hilden, Germany
Fisher Scientific 2-Propanol (Optima™), Fisher Chemical	Isopropanol	FSBA464-4	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Streptavidin Magnetic Beads	SMB's	S1420S	New England Biolabs, Ipswich, Massachusetts, United States
PickPen 1-M	Pick pen	23001	Bio Nobile, Pargas, Finland
PickPen tips bulk 96		34096	Bio Nobile, Pargas, Finland
Hydrochloric acid S.G. 1.18 (approximately 37%)	Hydrochloric Acid (HCl)	H/1200/PB17	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Sodium Hydroxide (Pellets/Certified ACS), Fisher Chemical	Sodium Hydroxide	S318-500	Thermo Fisher Scientific, Waltham, Massachusetts, United States

pGEM®-T Easy Vector Systems		A1360	Promega, Madison, Wisconsin
Subcloning Efficiency™ DH5α Competent Cells	Competent E. coli DH5-α cells	18265017	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Ampicillin anhydrous, 96.0-100.5 %	Ampicillin	69-53-4	Sigma Aldrich, St Louis, Missouri, United States
5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside	X-Gal	B1690	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Bacto™ Tryptone		211705	Becton Dickinson, Franklin Lakes, New Jersey, United States
Bacto™ Yeast Extract		212750	Becton Dickinson, Franklin Lakes, New Jersey, United States
Bacto™		214050	Becton Dickinson, Franklin Lakes, New Jersey, United States
QIAprep Spin Miniprep Kit (50)		27104	Qiagen, Hilden, Germany
Eppendorf Micro Centrifuge	Micro-centrifuge	5415R	Eppendorf, Hamburg, Germany
Eppendorf MasterCycler 6321 Pro Vapo Protect 96-well	Thermo-cycler	6321	Eppendorf, Hamburg, Germany
Eppendorf Mini-Spin Plus		5920R	Eppendorf, Hamburg, Germany
NanoDrop® ND-1000 UV-Vis Spectrophotometer	Nanodrop	ND-1000	Thermo Fisher Scientific, Waltham, Massachusetts, United States

## A2: Survey of NZ for selected environmental contaminants

Reagent/equipment	Abbreviation/Common Name	Catalogue number/model	Manufacturer/supplier
Bottle 250 mL HDPE wide mouth with PP screw cap (Nalgene)		NAL312104-0008	Thermo Fisher Scientific, Waltham, Massachusetts, United States (supplied to us by Asure Quality)
Dichloromethane OPTIMA grade ACS 4 L for HPLC GC ICP UV Spectrophotometry	Dichloromethane	FSBD151-4	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Methanol OPTIMA LC/MS grade EcoSafPak 4 L stabilized	Methanol	FSBA456-4	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Acetone		HPLC grade	Thermo Fisher Scientific Ltd, New Zealand
Toluene		Nanograde	Thermo Fisher Scientific Ltd, New Zealand
Celite® Hyflo Supercel. CAS 68855-54-9, pH 10 (100 g/l, H <sub>2</sub> O, 20 °C).	Celite® Hyflo Supercel	1.02688	Sigma Aldrich, St Louis, Missouri, United States
Vacuum Pump		KNF Neuberger Laboport N820.3FT.18	HiVac, Silverdale, Auckland, New Zealand
Masterflex Transfer Tubing, PFA, 1/16" ID" ID x 1/8" OD; 25 ft	Teflon transfer tubing	EW-06375-01	Cole-Parmer, Vernon Hills, Illinois, United States
Oasis HLB 20cc (1g) LP Extraction Cartridge		186000117	Waters Corp, Milford, Massachusetts, United States

Whatman GF/C 47 mm		1822-047	GE Healthcare, Chicago, Illinois, United States
ISOLUTE® C18(EC) 500 mg/10 mL		221-0050-H	Biotage, Uppsala, Sweden
Ethyl Acetate, Optima™ for HPLC and GC, Fisher Chemical	Ethyl Acetate	E196SK-4	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Florisil®		60-100 mesh	Sigma Aldrich, St Louis, Missouri, United States
Sodium Sulfate Anhydrous (Granular/Certified ACS), Fisher Chemical	Sodium Sulfate	S421-500	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Gas chromatography vials	Agilent 2 mL wide mouth amber glass GC vials		DKSH NZ Ltd, Auckland, New Zealand
4 mL React-Vials		TS-13223	Thermo Fisher Scientific Ltd, New Zealand
MilliQ Water System		Millipore MilliQ Integral 5	Thermo Fisher Scientific, Auckland, New Zealand
GC-MS System		Agilent 6890N series GC and 5975MSD	Agilent Technologies, Santa Clara, California, United States
GCMS-MS System		Agilent 7000 series Triple Quad GC-MS	Agilent Technologies, Santa Clara, California, United States

### A3: Glyphosate aptamer selection

Reagent/equipment	Abbreviation/Common Name	Catalogue number/model	Manufacturer/supplier
$\omega$ -Aminohexyl–Agarose saline suspension	Amino-hexyl beads	A6017	Sigma Aldrich, St Louis, Missouri, United States
N,N'-Dicyclohexylcarbodiimide 99%	DCC	D80002	Sigma Aldrich, St Louis, Missouri, United States
N-Hydroxysuccinimide 98%	NHS	130672	Sigma Aldrich, St Louis, Missouri, United States
Tetrahydrofuran anhydrous, $\geq 99.9\%$ , inhibitor-free	THF	401757	Sigma Aldrich, St Louis, Missouri, United States
Glyphosate PESTANAL <sup>®</sup> , analytical standard	Glyphosate	45521-250MG	Sigma Aldrich, St Louis, Missouri, United States
L-Alanine, non-animal source	Alanine	A7469-25G	Sigma Aldrich, St Louis, Missouri, United States
L-Glutamine	Glutamine	G5763	Sigma Aldrich, St Louis, Missouri, United States
L-Leucine	Leucine	L8912-25G	Sigma Aldrich, St Louis, Missouri, United States
L-Valine	Valine	V0513-25G	Sigma Aldrich, St Louis, Missouri, United States
Glycine		G8790-100G	Sigma Aldrich, St Louis, Missouri, United States

## A4: Nonylphenol aptamer selection

Reagent/equipment	Abbreviation/Common Name	Catalogue number/model	Manufacturer/supplier
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, St Louis, Missouri, United States
Nonylphenol PESTANAL®, analytical standard, technical mixture	Nonylphenol	46018-1G	Sigma Aldrich, St Louis, Missouri, United States
Ethanolamine ≥ 98%		E9508-100ML	Sigma Aldrich, St Louis, Missouri, United States
Sodium Hydroxide (Pellets/Certified ACS), Fisher Chemical	Sodium Hydroxide	S318-500	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Shimadzu UV-1800 UV/Visible Scanning Spectrophotometer	UV-Vis Spectrophotometer	220-92961-01	Shimadzu, Kyoto, Kyoto Prefecture, Japan
Nunc™ MicroWell™ 96-Well Optical-Bottom Plates with Polymer Base	96-well black, clear-bottom plates	265301	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Clariostar Plus Platereader	Fluorescent plate reader		BMG Labtech, Offenburg, Germany

## A5: Oxybenzone aptamer selection

Reagent/equipment	Abbreviation/Common Name	Catalogue number/model	Manufacturer/supplier
Oxybenzone pharmaceutical secondary standard	Oxybenzone	PHR1074	Sigma Aldrich, St Louis, Missouri, United States
Monarch® PCR & DNA Cleanup Kit (5 µg)	PCR Clean-up Kit	T1030L	New England Biolabs, Ipswich, Massachusetts, United States

## Appendix B: Recipes

Reagent	Abbreviation or Common Name	Recipe	Preparation Instructions/Storage Conditions
Aptamer binding and washing buffer	BWB	For 1X BWB: 10 mM NaCl 0.5 mM KCl 0.2 mM MgCl <sub>2</sub> 0.1 mM CaCl <sub>2</sub> 0.2 mM Tris-HCl (pH 7.5) IGEPAAL CA-630 (at the required concentration)	Make up to 1 L using MilliQ water. pH adjusted to 7.5 using NaOH or HCl. Sterilize by autoclaving at 121°C for 20 minutes prior to adding IGEPAAL. Store at room temperature.
5% (w/v) Blotto Solution	Blocking buffer	For a 5 mL volume: 0.25 g of 98% skim milk powder 5 mL of 1X BWB	Make fresh and use immediately.
1X TAE Buffer	TAE buffer	TAE Buffer (Tris-acetate-EDTA) (50X) diluted to 1X using MilliQ water.	For use as the running buffer in gel electrophoresis. Store at room temperature.
Luria Broth	LB	To make a 1 L volume: 10 g NaCl 10 g Bacto™ Tryptone 5 g Bacto™ Yeast Extract 50 µg/mL Ampicillin (if required)	Make up to 1 L using MilliQ water. Sterilize by autoclaving at 121°C for 20 minutes prior to use. Store at 4°C.
Luria Agar	LA	To make a 1 L volume: 15 g Bacto™ Agar 1 L Luria Broth 2% (v/v) X-gal 50 µg/mL Ampicillin	Add 15 g of Bacto™ Agar to an empty 1L Schott bottle. Add 1L of Luria Broth. Mix well. Autoclave for 20 minutes at 121°C to sterilize and then place in a water bath at 56°C for at least 30 minutes to cool. Add Ampicillin and X-gal. Mix well and pour into



			plates using a sterile technique (approximately 20 mL per plate). Leave plates overnight at room temperature to set and then store at 4°C until use.
--	--	--	--

## Appendix C: Synthetic oligonucleotide sequences

Oligonucleotide	Sequence (5' to 3')	Supplier
Forward primer	ATACGAGCTTGTTCATA	Thermo Fisher Scientific, Waltham, Massachusetts, United States
Reverse primer	GATTGCTCTTACTATCA	
Random Library	ATACGAGCTTGTTCATA-N40-TGATAGTAAGAGCAATC	
TES07 (Testosterone binding aptamer – used as a PCR positive control)	ATACGAGCTTGTTCATAGCCCTTTACACCATTCATGCCGCTCTTATCGGTAGTCGCGTGATAGTAAGAGCAATC	

\* primer binding sites are highlighted in red

# Appendix D: Scripts for Bioinformatics

## Pipeline

### D1: Script 1 – Reverse Complement and Merge

```
#Copyright Zak Murray 2019
#Designates which files are being processed
for sample in S01
do
echo ${sample} is being processed by script 1
# Copy the barcode file into the directory where the raw .fastq files are
cp barcode_file.txt /mnt/RosalindData/murrayza/new_analysis_August2019/raw_data/
#Moves into the directory where all the raw .fastq files are
cd /mnt/RosalindData/murrayza/new_analysis_August2019/raw_data/
#Makes a new directory for that particular file to put the processed data into
mkdir processed_data_${sample}
#Pipes the designated file into the fastx barcode splitter tool and divides the file into two,
depending on which primer is present (BC1 for the forward strand and BC2 for the reverse)
cat ${sample}_R1.Aptamer.cleaned.fastq | fastx_barcode_splitter.pl --bcfile barcode_file.txt -
-prefix ${sample} --bol --mismatches 2
#Moves the BC1, BC2 and unmatched files into the processed data directory
mv ${sample}BC1 ${sample}BC2 ${sample}unmatched processed_data_${sample}
#Moves into the processed data directory
cd processed_data_${sample}
#Use the fastx toolkit to reverse complement everything which has BC2 at the beginning (the
reverse primer and therefore the reverse strand)
fastx_reverse_complement -i ${sample}BC2 -o ${sample}BC2_reverseComp
#Merges the BC1 and BC2 reverse complement files together to make a new .fastq file with all
forward strand reads
cat ${sample}BC1 ${sample}BC2_reverseComp > ${sample}_merged.fastq
#Use the fastx toolkit to calculate nucleotide percentages and quality statistics
fastx_quality_stats -i ${sample}_merged.fastq -o ${sample}_stats
#Counts the number of lines in the all files to ensure that everything has been processed
#Move all the files to another directory
cd ..
mv processed_data_${sample}
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script1/output_script
1_${sample}/
cd /mnt/RosalindData/murrayza/new_analysis_August2019/pipeline_scripts/
mv output.txt
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script1/output_script
1_${sample}/
#End program message
echo ${sample} has been processed by script 1
#End loop
done
```

## D2: Script 2 – Removal of polyX motifs

```
#Copyright Zak Murray 2019
#Designates the files for polyX removal
for sample in S47
do
echo ${sample} is being searched to locate polyX sequences and they are being removed
#Navigates to the directory where the output of script 1 is located
cd
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script1/output_script
1_${sample}
#Runs the AfterQC script which removes the polyX sequences. Settings can be modified -
currently anything with a stretch of 10 of the same bases in a row will be removed
python /home/murrayza/AfterQC/after.py --read1_file=${sample}_merged.fastq --
poly_size_limit=10 --allow_mismatch_in_poly=0 --trim_front=0 --trim_tail=0
#Navigates to the directory where the "good" reads are located i.e. those that do not contain
polyX sequences
cd good
#Moves the good reads file to the output of script 2 directory
mv ${sample}_merged.good.fq ${sample}_merged.noPolyX.fq
mv ${sample}_merged.noPolyX.fq
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script2/
# End of program message
echo ${sample}
echo PolyX removal complete
# Ends the loop
done
```

## D3: Script 3 – Primer Removal

```
#Copyright Zak Murray 2019
#Designate the files to the run the loop on
for sample in S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40
S41 S42 S43 S44 S45 S46 S47 S48 S49
#for sample in S01 S02 S03 S04 S05 S06 S07 S08 S09 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20
do
#Tell the user which file is having the primers removed from
echo ${sample}
echo Primers are being removed from ${sample}
#Navigate to the directory where the output of script 2 is located
cd /mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script2/
#Remove the forward primer sequence using cutadapt
cutadapt -m 5 -g XATACGAGCTTGTCAATA --overlap 13 -o ${sample}_noFwdPrimer.fq
${sample}_merged.noPolyX.fq
#Remove the reverse primer sequence using cutadapt
cutadapt -m 5 -a TGATAGTAAGAGCAATCX --overlap 13 -o ${sample}_noPrimers.fq
${sample}_noFwdPrimer.fq
#Count the number of lines in the pre and post-processed files to check they are equivalent
wc -l ${sample}_merged.noPolyX.fq
wc -l ${sample}_noPrimers.fq
#Calculate the nucleotide composition of the N40 region
fastx_quality_stats -i ${sample}_noPrimers.fq -o ${sample}_noPrimers_stats.txt
#Move the output of script 3 to the appropriate directory
mv ${sample}_noPrimers.fq
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script3/
mv ${sample}_noFwdPrimer.fq
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script3/
mv ${sample}_noPrimers_stats.txt
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script3
cd /mnt/RosalindData/murrayza/new_analysis_August2019/pipeline_scripts/
mv output.txt
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script3/
#Indicate to the user that the script has finished running and end the loop
echo ***Primer Removal Complete***
#Ends loop
done
```

## D4: Script 4 – Removal of TES07 homologous sequences

```
#Copyright Zak Murray 2019
#Designates which files to run the script on
for sample in S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40
S41 S42 S43 S44 S45 S46 S47 S48 S49
#for sample in S02 S03 S04 S05 S06 S07 S08 S09 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20
#Initiates the loop
do
#States which sample the loop is being run on
echo ${sample} is being searched for TES07 or homologous sequences and they are being removed
#Navigates to the directory containing the output of script 3
cd /mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script3/
#Use cut adapt to remove any sequence matching the sequence of interest or that has 50% or
more homology
cutadapt -a "GCCCTTTACACCATTCATGCCGCTCTTATCGGTAGTCGCG;max_error_rate=0.50;min_overlap=40" --
discard-trimmed ${sample}_noPrimers.fq > ${sample}_noPrimers_TES07removed.fq
#Search for approximate matches to the specified sequence using "agrep" - if the above command
worked correctly then the search should return 0
agrep -c -8 "GCCCTTTACACCATTCATGCCGCTCTTATCGGTAGTCGCG" ${sample}_noPrimers_TES07removed.fq
#Move the output to a folder designated for the output of Script 4
mv ${sample}_noPrimers_TES07removed.fq
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script4/
#A message stating that removal of the contaminating sequence has been completed
echo ${sample}
echo TES07 removal complete
#Ends the loop
done
```

## D5: Script 5 – Removal of R18C1 homologous sequences

```
#Copyright Zak Murray 2019
#Designates which files to run the script on
for sample in S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40
S41 S42 S43 S44 S45 S46 S47 S48 S49
#for sample in S02 S03 S04 S05 S06 S07 S08 S09 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20
#for sample in S01
#Initiates the loop
do
#States which sample the loop is being run on
echo ${sample} is being searched for R18C1 or homologous sequences and they are being removed
#Navigates to the directory containing the output of script 4
cd /mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script4/
#Use cut adapt to remove any sequence matching the sequence of interest or that has 50% or
more homology
cutadapt -a "CGAAGGGATGCCGTTTGGGCCCAAGTTCGGCATAGTGTGG;max_error_rate=0.50;min_overlap=40" --
discard-trimmed ${sample}_noPrimers_TES07removed.fq > ${sample}_noPrimers_R18C1removed.fq
#Search for approximate matches to the specified sequence using "agrep" - if the above command
worked correctly then the search should return 0
agrep -c -8 "CGAAGGGATGCCGTTTGGGCCCAAGTTCGGCATAGTGTGG" ${sample}_noPrimers_R18C1removed.fq
#Move the output to a folder designated for the output of Script 5
mv ${sample}_noPrimers_R18C1removed.fq
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script5/
#A message stating that removal of the contaminating sequence has been completed
echo ${sample}
echo R18C1 removal complete
#Ends the loop
done
```

## D6: Script 6 – Removal of reads containing primer binding sites in the N40 region

```
#Copyright Zak Murray 2019
#This script is designed to remove reads which contain forward primer sequences in the N40
region#
#Designates which files to run the script on
#for sample in S01
#for sample in S02 S03 S04 S05 S06 S07 S08 S09 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20
for sample in S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40
S41 S42 S43 S44 S45 S46 S47 S48 S49
#Initiates the loop
do
#States which sample the loop is being run on
echo ${Sample}
echo Script is removing sequences which have primers in the N40 region
#Navigates to the directory containing the output of script 5
cd /mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script5/
#Use cut adapt to remove any reads which contains a match to the specified sequence of
interest
cutadapt -a "ATACGAGCTTGTCAATA;max_error_rate=0.17;min_overlap=18" --discard-trimmed
${sample}_noPrimers_R18Clremoved.fq > ${sample}_contaminantsRemoved.fq
#Move the output to a folder designated for the output of Script 6
mv ${sample}_contaminantsRemoved.fq
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script6//
#A message stating that removal of the contaminating sequence has been completed
echo ${sample}
echo Removal of N40 primers is complete
#Ends the loop
done
```



## D7: Script 7 – Length trimming

```
# Copyright Zak Murray 2019
# This script is designed to filter out any reads shorter than 36 nt or longer than 44 nt
# Designates the samples which the script will be run on
#for sample in S01
#for sample in S02 S03 S04 S05 S07 S08 S09 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20
#for sample in S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40
S41 S42 S43 S44 S45 S46 S47 S48 S49
for sample in S06
#Initiates the loop
do
#States which sample the loop is currently being run on
echo ${sample} is being filtered to give only reads between 36bp and 44bp
#Navigates to the directory containing the output of script 6
cd /mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script6/
#Uses cut adapt to filter out reads which are too long or too short
#-m is the shortest read that will be accepted (length of 36 nt in this case)
#-M is the longest read that will be accepted (length of 44 nt in this case)
cutadapt -m 36 -M 44 ${sample}_contaminantsRemoved.fq > ${sample}_final.fq
#Move the resulting output to a folder for the output of script 7
mv ${sample}_final.fq
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script7/
#Messages stating that the length filtering has been completed
echo ${sample} has been filtered to the provide reads between 36-44bp
#Ends the loop
done
```

## D8: Script 8 – Counting the number of unique sequences in each dataset

```
#Copyright Zak Murray 2019
#This script is designed to count the number of unique sequences in each of the datasets
#for sample in S01
#for sample in S02 S03 S04 S05 S06 S07 S08 S09 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20
#Designates the sample(s) which the loop will be run on
for sample in S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40
S41 S42 S43 S44 S45 S46 S47 S48 S49
#Initiates the loop
do
#States which sample the loop is currently being run on
echo FASTAptamer is counting the unique sequences within ${sample}
#Navigates to the directory containing the output of the previous script in the pipeline
cd /mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script7/
#Counts the number of unique reads using the FastAptamer Count perl script
perl /mnt/RosalindData/murrayza/new_analysis_August2019/pipeline_scripts/fastaptamer_count -i
${sample}_final.fq -o ${sample}_counts.fa
#Message stating that counting has been completed
echo Counts for ${sample} have been completed
#Ends the loop
done
```

## D9: Script 9 – Enrichment Analysis

```
#Copyright Zak Murray 2019
#This script is designed to run an enrichment analysis on two files#
#Designate the first selection round
sample1="S42"
#Designate the second selection round
sample2="S49"
#States which files the enrichment analysis is assessing
echo FASTAptamer is calculating enrichment between ${sample1} and ${sample2}
#Navigate to the directory containing the output of the previous script
cd /mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script8/
#Calculate enrichment of sequences between the two different selection rounds using the
FastAptamer Enrich perl script
perl /mnt/RosalindData/murrayza/new_analysis_August2019/pipeline_scripts/fastaptamer_enrich -x
${sample1}_counts.fa -y ${sample2}_counts.fa -o ${sample1}_${sample2}_enrichment.tsv
#Move the output to a directory containining all the outputs for script 9
mv ${sample1}_${sample2}_enrichment.tsv
/mnt/RosalindData/murrayza/new_analysis_August2019/processed_data/output_script9/
#State that the enrichment analysis has been completed
echo Enrichment analysis has been completed
```

## D10: Script 10 – Round Splitter (any condition)

```
#Script 10 - "Round Splitter"
#This is an R-script to separate the enrichment files into two separate
files (one for the first round counts and one for the second)
#Copyright Zak Murray 2019

### Round 1 and 2 ###

rm(list = ls())
library(dplyr)

#The name of the condition (which is specified in the file name) needs to
be inserted here
ConditionName <- "polyMinus"

# Reading the .tsv generated by FastAptamer for the enrichment analysis and
creating a new data frame
r1and2 <- read.table(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/enrichment_",ConditionNa
me,"_R1+2.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
r1and2
head(r1and2)
r1_counts <- dplyr::select(r1and2, "Sequence", "Length", "Rank..x.",
"Reads..x.", "RPM..x.", "Enrichment..y.x.")
r2_counts <- dplyr::select(r1and2, "Sequence", "Length", "Rank..y.",
"Reads..y.", "RPM..y.", "Enrichment..y.x.")
# Renaming the column headings for Round 1 to make them more readable
names(r1_counts)[3] = "Rank"
names(r1_counts)[4] = "Reads"
names(r1_counts)[5] = "RPM"
names(r1_counts)[6] = "Enrichment"
# Renaming the column headings for Round 2 to make them more readable
names(r2_counts)[3] = "Rank"
names(r2_counts)[4] = "Reads"
names(r2_counts)[5] = "RPM"
names(r2_counts)[6] = "Enrichment"

head(r1_counts)
head(r2_counts)
# Writing the sequences and information for the two different rounds into
their own .csv files
write.csv(r1_counts, file= paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r1_counts.csv",sep=""))
write.csv(r2_counts, file= paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r2_counts.csv",sep=""))

### Round 3 and 4 ###
# Completes the same process but for the files from round 3 and 4

library(dplyr)

r3and4 <- read.table(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/enrichment_",ConditionNa
me,"_R3+4.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
head(r3and4)
r3_counts <- dplyr::select(r3and4, "Sequence", "Length", "Rank..x.",
"Reads..x.", "RPM..x.", "Enrichment..y.x.")
r4_counts <- dplyr::select(r3and4, "Sequence", "Length", "Rank..y.",
"Reads..y.", "RPM..y.", "Enrichment..y.x.")
```

```

names(r3_counts)[3] = "Rank"
names(r3_counts)[4] = "Reads"
names(r3_counts)[5] = "RPM"
names(r3_counts)[6] = "Enrichment"

names(r4_counts)[3] = "Rank"
names(r4_counts)[4] = "Reads"
names(r4_counts)[5] = "RPM"
names(r4_counts)[6] = "Enrichment"

head(r3_counts)
head(r4_counts)

write.csv(r3_counts, file= paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r3_counts.csv",sep=""))
write.csv(r4_counts, file= paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r4_counts.csv",sep=""))

### Round 5 and 6 ###

library(dplyr)
# Completes the same process but for the files from round 5 and 6

r5and6 <- read.table(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/enrichment_",ConditionNa
me,"_R5+6.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
head(r5and6)
r5_counts <- dplyr::select(r5and6, "Sequence", "Length", "Rank..x.",
"Reads..x.", "RPM..x.", "Enrichment..y.x.")
r6_counts <- dplyr::select(r5and6, "Sequence", "Length", "Rank..y.",
"Reads..y.", "RPM..y.", "Enrichment..y.x.")
names(r5_counts)[3] = "Rank"
names(r5_counts)[4] = "Reads"
names(r5_counts)[5] = "RPM"
names(r5_counts)[6] = "Enrichment"

names(r6_counts)[3] = "Rank"
names(r6_counts)[4] = "Reads"
names(r6_counts)[5] = "RPM"
names(r6_counts)[6] = "Enrichment"

head(r5_counts)
head(r6_counts)

write.csv(r5_counts, file= paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r5_counts.csv",sep=""))
write.csv(r6_counts, file= paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r6_counts.csv",sep=""))

### Round 7 ###

library(dplyr)
# Completes the same process but for the files from round 7

r6and7 <- read.table(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/enrichment_",ConditionNa
me,"_R6+7.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
head(r6and7)
r7_counts <- dplyr::select(r6and7, "Sequence", "Length", "Rank..y.",
"Reads..y.", "RPM..y.", "Enrichment..y.x.")

```

```
names(r7_counts)[3] = "Rank"
names(r7_counts)[4] = "Reads"
names(r7_counts)[5] = "RPM"
names(r7_counts)[6] = "Enrichment"

head(r7_counts)

write.csv(r7_counts, file= paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/r7_counts.csv", sep=""))
```

## D11: Script 11 – Frequency Counts (any condition)

```
#Script 11 - "Frequency Counts (any condition)"
#This is an R-script is designed to take the counts from each round
#(produced using Script 10),
#replace all the NA values with 0 and then sort the "RPM" column based from
#largest to smallest
#Copyright Zak Murray 2019

rm(list = ls())

#The name of the condition (which is specified in the file name) needs to
#be inserted here
ConditionName <- "polyMinus"

library(dplyr)

# Round 1 count file - outputting a ranked file and replacing the NA values
#with 0
r1_counts <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/r1_counts.csv", sep=""),
header = TRUE)
head(r1_counts)
r1_counts[is.na(r1_counts)] <- 0
head(r1_counts)
r1_counts_ranked <- r1_counts[order(r1_counts$RPM, decreasing = TRUE),]
head(r1_counts_ranked)
write.csv(r1_counts_ranked, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/r1_counts_ranked.csv",
sep=""))

# Round 2 count file - outputting a ranked file and replacing the NA values
#with 0
r2_counts <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/r2_counts.csv", sep=""),
header = TRUE)
head(r2_counts)
r2_counts[is.na(r2_counts)] <- 0
head(r2_counts)
r2_counts_ranked <- r2_counts[order(r2_counts$RPM, decreasing = TRUE),]
head(r2_counts_ranked)
write.csv(r2_counts_ranked, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/r2_counts_ranked.csv",
sep=""))

# Round 3 count file - outputting a ranked file and replacing the NA values
#with 0
r3_counts <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/r3_counts.csv", sep=""),
header = TRUE)
head(r3_counts)
r3_counts[is.na(r3_counts)] <- 0
head(r3_counts)
r3_counts_ranked <- r3_counts[order(r3_counts$RPM, decreasing = TRUE),]
head(r3_counts_ranked)
write.csv(r3_counts_ranked, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/r3_counts_ranked.csv",
sep=""))
```

```

# Round 4 count file - outputting a ranked file and replacing the NA values
with 0
r4_counts <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r4_counts.csv", sep=""),
header = TRUE)
head(r4_counts)
r4_counts[is.na(r4_counts)] <- 0
head(r4_counts)
r4_counts_ranked <- r4_counts[order(r4_counts$RPM, decreasing = TRUE),]
head(r4_counts_ranked)
write.csv(r4_counts_ranked, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r4_counts_ranked.csv",
sep=""))

# Round 5 count file - outputting a ranked file and replacing the NA values
with 0
r5_counts <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r5_counts.csv", sep=""),
header = TRUE)
head(r5_counts)
r5_counts[is.na(r5_counts)] <- 0
head(r5_counts)
r5_counts_ranked <- r5_counts[order(r5_counts$RPM, decreasing = TRUE),]
head(r5_counts_ranked)
write.csv(r5_counts_ranked, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r5_counts_ranked.csv",
sep=""))

# Round 6 count file - outputting a ranked file and replacing the NA values
with 0
r6_counts <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r6_counts.csv", sep=""),
header = TRUE)
head(r6_counts)
r6_counts[is.na(r6_counts)] <- 0
head(r6_counts)
r6_counts_ranked <- r6_counts[order(r6_counts$RPM, decreasing = TRUE),]
head(r6_counts_ranked)
write.csv(r6_counts_ranked, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r6_counts_ranked.csv",
sep=""))

# Round 7 count file - outputting a ranked file and replacing the NA values
with 0
r7_counts <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r7_counts.csv", sep=""),
header = TRUE)
head(r7_counts)
r7_counts[is.na(r7_counts)] <- 0
head(r7_counts)
r7_counts_ranked <- r7_counts[order(r7_counts$RPM, decreasing = TRUE),]
head(r7_counts_ranked)
write.csv(r7_counts_ranked, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r7_counts_ranked.csv",
sep=""))

list.files(path = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/", sep=""), pattern =
"ranked.csv")

```



```

# Combining the top 50 sequences from each round into one single file (the
top 50 sequences from Round 1 to Round 7 for a given condition)
final_file = NULL
for (round_file in list.files(path = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", sep=""), pattern =
"ranked.csv")){
  print(round_file)
  loop_file <- read.csv(paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", round_file, sep=""))
  loop_file_top50 <- loop_file[1:50,]
  loop_file_top50
  loop_file_top50$round <- substr(round_file, 2, 2)
  loop_file_top50$condition <- "polyMinus"
  loop_file_top50_5cols <- loop_file_top50 %>% dplyr::select(Sequence,
Length, RPM, round, condition)
  final_file <- rbind(final_file, loop_file_top50_5cols)
}
final_file
write.csv(final_file, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_freque
ncy_top50.csv", sep=""))

#Combining all the frequency files from the different conditions

ConditionName <- "Std"
Std_frequency_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_freque
ncy_top50.csv", sep=""), header = TRUE)

ConditionName <- "Std+Mut"
StdwithMut_frequency_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_freque
ncy_top50.csv", sep=""), header = TRUE)

ConditionName <- "ExtraVol"
ExtraVol_frequency_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_freque
ncy_top50.csv", sep=""), header = TRUE)

ConditionName <- "ExtraDet"
ExtraDet_frequency_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_freque
ncy_top50.csv", sep=""), header = TRUE)

ConditionName <- "RoomTemp"
RoomTemp_frequency_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_freque
ncy_top50.csv", sep=""), header = TRUE)

ConditionName <- "Negative"
Negative_frequency_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_freque
ncy_top50.csv", sep=""), header = TRUE)

ConditionName <- "Counter"
Counter_frequency_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_freque
ncy_top50.csv", sep=""), header = TRUE)

frequency_top50_allConditions <- rbind(Std_frequency_allRounds,
StdwithMut_frequency_allRounds, ExtraVol_frequency_allRounds,

```

```
ExtraDet_frequency_allRounds, RoomTemp_frequency_allRounds,  
Negative_frequency_allRounds, Counter_frequency_allRounds)  
write.csv(frequency_top50_allConditions, file = "~/PhD-  
work/HTSELEX_analysis_August2019/frequency_top50_allConditions.csv")
```

## D12: Script 12 – Enrichment Counts (any condition)

```
#Script 12 - "Enrichment Counts (any condition)"
#This is an R-script is designed to take the enrichment values from
each round (produced using Script 10),
#replace all the NA values with 0 and then sort the "Enrichment" column
based from largest to smallest
#Copyright Zak Murray 2019

rm(list = ls())

#The name of the condition (which is specified in the file name) needs
to be inserted here
ConditionName <- "polyMinus"

library(dplyr)

#Round 1 to 2 Enrichment
rland2 <- read.table(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/enrichment_",Condi
onName,"_R1+2.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
head(rland2)
rland2_enrichment <- dplyr::select(rland2, "Sequence", "Length",
"Enrichment..y.x.")
head(rland2_enrichment)
names(rland2_enrichment)[3] = "Enrichment"
rland2_enrichment[is.na(rland2_enrichment)] <- 0
head(rland2_enrichment)
rland2_enrichment_ranked <-
rland2_enrichment[order(rland2_enrichment$Enrichment, decreasing =
TRUE),]
head(rland2_enrichment_ranked)
write.csv(rland2_enrichment_ranked, file= paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r2_enrichment.csv",s
ep=""))

#Round 2 to 3 Enrichment
r2and3 <- read.table(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/enrichment_",Condi
onName,"_R2+3.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
head(r2and3)
r2and3_enrichment <- dplyr::select(r2and3, "Sequence", "Length",
"Enrichment..y.x.")
head(r2and3_enrichment)
names(r2and3_enrichment)[3] = "Enrichment"
r2and3_enrichment[is.na(r2and3_enrichment)] <- 0
head(r2and3_enrichment)
r2and3_enrichment_ranked <-
r2and3_enrichment[order(r2and3_enrichment$Enrichment, decreasing =
TRUE),]
head(r2and3_enrichment_ranked)
write.csv(r2and3_enrichment_ranked, file= paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/r3_enrichment.csv",s
ep=""))

#Round 3 to 4 Enrichment
```

```

r3and4 <- read.table(file = paste("~/Phd-
work/HTSELEX_analysis_August2019/", ConditionName, "/enrichment_", Condi
onName, "_R3+4.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
head(r3and4)
r3and4_enrichment <- dplyr::select(r3and4, "Sequence", "Length",
"Enrichment..y.x.")
head(r3and4_enrichment)
names(r3and4_enrichment)[3] = "Enrichment"
r3and4_enrichment[is.na(r3and4_enrichment)] <- 0
head(r3and4_enrichment)
r3and4_enrichment_ranked <-
r3and4_enrichment[order(r3and4_enrichment$Enrichment, decreasing =
TRUE),]
head(r3and4_enrichment_ranked)
write.csv(r3and4_enrichment_ranked, file= paste("~/Phd-
work/HTSELEX_analysis_August2019/", ConditionName, "/r4_enrichment.csv", s
ep=""))

#Round 4 to 5 Enrichment
r4and5 <- read.table(file = paste("~/Phd-
work/HTSELEX_analysis_August2019/", ConditionName, "/enrichment_", Condi
onName, "_R4+5.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
head(r4and5)
r4and5_enrichment <- dplyr::select(r4and5, "Sequence", "Length",
"Enrichment..y.x.")
head(r4and5_enrichment)
names(r4and5_enrichment)[3] = "Enrichment"
r4and5_enrichment[is.na(r4and5_enrichment)] <- 0
head(r4and5_enrichment)
r4and5_enrichment_ranked <-
r4and5_enrichment[order(r4and5_enrichment$Enrichment, decreasing =
TRUE),]
head(r4and5_enrichment_ranked)
write.csv(r4and5_enrichment_ranked, file= paste("~/Phd-
work/HTSELEX_analysis_August2019/", ConditionName, "/r5_enrichment.csv", s
ep=""))

#Round 5 to 6 Enrichment
r5and6 <- read.table(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/enrichment_", Condi
onName, "_R5+6.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
head(r5and6)
r5and6_enrichment <- dplyr::select(r5and6, "Sequence", "Length",
"Enrichment..y.x.")
head(r5and6_enrichment)
names(r5and6_enrichment)[3] = "Enrichment"
r5and6_enrichment[is.na(r5and6_enrichment)] <- 0
head(r5and6_enrichment)
r5and6_enrichment_ranked <-
r5and6_enrichment[order(r5and6_enrichment$Enrichment, decreasing =
TRUE),]
head(r5and6_enrichment_ranked)
write.csv(r5and6_enrichment_ranked, file= paste("~/Phd-
work/HTSELEX_analysis_August2019/", ConditionName, "/r6_enrichment.csv", s
ep=""))

#Round 6 to 7 Enrichment

```

```

r6and7 <- read.table(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/enrichment_", Condi-
onName, "_R6+7.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
head(r6and7)
r6and7_enrichment <- dplyr::select(r6and7, "Sequence", "Length",
"Enrichment..y.x.")
head(r6and7_enrichment)
names(r6and7_enrichment)[3] = "Enrichment"
r6and7_enrichment[is.na(r6and7_enrichment)] <- 0
head(r6and7_enrichment)
r6and7_enrichment_ranked <-
r6and7_enrichment[order(r6and7_enrichment$Enrichment, decreasing =
TRUE),]
head(r6and7_enrichment_ranked)
write.csv(r6and7_enrichment_ranked, file= paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/r7_enrichment.csv", s
ep=""))

#Combining all the files together using list.files and a loop to
produce a single .csv file containing the top sequences by enrichment
from each selection round

list.files(path = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", sep=""), pattern
= "enrichment.csv")

final_file = NULL
for (round_file in list.files(path = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", sep=""), pattern
= "enrichment.csv")){
  print(round_file)
  loop_file <- read.csv(paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", round_file, sep="")
)
  loop_file_top50 <- loop_file[1:50,]
  loop_file_top50
  loop_file_top50$Round <- substr(round_file, 2, 2)
  loop_file_top50$condition <- "polyMinus"
  loop_file_top50_5cols <- loop_file_top50 %>% dplyr::select(Sequence,
Length, Enrichment, condition, Round)
  final_file <- rbind(final_file, loop_file_top50_5cols)
}
final_file
write.csv(final_file, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_en
richment_top50.csv", sep=""))

#Combining all the enrichment files from the different conditions

ConditionName <- "Std"
Std_enrichment_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_en
richment_top50.csv", sep=""), header = TRUE)

ConditionName <- "Std+Mut"
StdwithMut_enrichment_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_en
richment_top50.csv", sep=""), header = TRUE)

```

```

ConditionName <- "ExtraVol"
ExtraVol_enrichment_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/",ConditionName,"_en
richment_top50.csv", sep=""), header = TRUE)

ConditionName <- "ExtraDet"
ExtraDet_enrichment_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/",ConditionName,"_en
richment_top50.csv", sep=""), header = TRUE)

ConditionName <- "RoomTemp"
RoomTemp_enrichment_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/",ConditionName,"_en
richment_top50.csv", sep=""), header = TRUE)

ConditionName <- "Negative"
Negative_enrichment_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/",ConditionName,"_en
richment_top50.csv", sep=""), header = TRUE)

ConditionName <- "Counter"
Counter_enrichment_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/",ConditionName,"_en
richment_top50.csv", sep=""), header = TRUE)

enrichment_top50_allConditions <- rbind(Std_enrichment_allRounds,
StdwithMut_enrichment_allRounds, ExtraVol_enrichment_allRounds,
ExtraDet_enrichment_allRounds, RoomTemp_enrichment_allRounds,
Negative_enrichment_allRounds, Counter_enrichment_allRounds)
write.csv(enrichment_top50_allConditions, file = "~/PhD-
work/HTSELEX_analysis_August2019/enrichment_top50_allConditions.csv")

```

## D13: Script 13 – Graphing of average frequency for each strategy

```
# Script 13 - This script is designed to graph the average frequency
for the top 50 sequences from each selection strategy.
# The desired round can be selected in Line 33
# It is also possible to run an ANOVA analysis using this script
# Copyright Zak Murray 2019

# Clears the R environment
rm(list = ls())

# Loads the relevant libraries
library(dplyr)
library(ggplot2)
library(s20x)

# Load the relevant dataset (produced using script 11)
data <- read.csv(file = "~/PhD-
work/HTSELEX_analysis_August2019/frequency_top50_allConditions.csv")
# Prints the loaded data and checks the data structure
data
str(data)

# Conversion of condition to a simplified name
data$condition_new <- data$condition
data$condition_new <- as.character(data$condition_new)
data$condition_new[data$condition_new %in% "Std"] <- "1 - Std"
data$condition_new[data$condition_new %in% "Std+Mut"] <- "2 - Std+Mut"
data$condition_new[data$condition_new %in% "ExtraVol"] <- "3 -
ExtraVol"
data$condition_new[data$condition_new %in% "ExtraDet"] <- "4 -
ExtraDet"
data$condition_new[data$condition_new %in% "RoomTemp"] <- "5 -
RoomTemp"
data$condition_new[data$condition_new %in% "Negative"] <- "6 -
Negative"
data$condition_new[data$condition_new %in% "Counter"] <- "7 - Counter"
data$condition_new <- as.factor(data$condition_new)

# Calculates mean, sd, se and IC
# The desired round number needs to be selected here
data_subset <- subset(data, data$round == '7')
data_subset
str(data_subset)
my_sum <- data_subset %>%
  group_by(round, condition_new) %>%
  summarise(
    n=n(),
    mean=mean(RPM),
    sd=sd(RPM)
  ) %>%
  mutate( se=sd/sqrt(n)) %>%
  mutate( ic=se * qt((1-0.05)/2 + .5, n-1))

# Prints the data subset

data_subset
```

```

# Graphs the average enrichment showing the standard error as the error
bars
# The graph title is disabled by default but can be enabled by
activated the commented line of code (line 52)
ggplot(my_sum) +
  theme(plot.margin = unit(c(1,1,1,1), "cm")) +
  geom_bar(aes(x=condition_new, y=mean, fill=condition_new),
stat="identity", alpha=1.0) +
scale_fill_manual(values=c("#1B9E77", "#D95F02", "#7570B3", "#E7298A", "#66
A61E", "#E6AB02", "#A6761D")) +
  geom_errorbar(aes(x=condition_new, ymin=mean-se, ymax=mean+se),
width=0.3, colour="black", alpha=1.0, size=1.0) +
  #ggtitle("Average frequency of top 50 individual candidates under
\ndifferent selection strategies from SELEX round 6 to 7") +
  xlab("Library") +
  ylab("Average RPM") +
  labs(fill="Condition") +
  theme(legend.position = "none") +
  theme(axis.text.x = element_text(size = 12.0)) +
  theme(axis.text.y = element_text(size = 12.0)) +
  theme(axis.title.x = element_text(size = 15.0)) +
  theme(axis.title.y = element_text(size = 15.0)) +
  theme(axis.title.x = element_text(face = "bold")) +
  theme(axis.title.y = element_text(face = "bold")) +
  theme(axis.title.x = element_text(vjust = -5.0)) +
  theme(axis.title.y = element_text(vjust = +5.0)) +
  theme(text = element_text(size = 16.0)) +
  theme(plot.title = element_text(vjust = +5.0)) +
  theme(panel.background = element_blank()) +
  theme(panel.border = element_blank()) +
  theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank(), axis.line = element_line(colour = "black"))

```



## D14: Script 14 – Graphing of average enrichment for each strategy

```
# Script 14 - This script is designed to graph the average enrichment for
the top 50 sequences from each selection strategy.
# The desired round can be selected in Line 33
# It is also possible to run an ANOVA analysis using this script
# Copyright Zak Murray 2019

# Clears the R environment
rm(list = ls())

# Loads the relevant libraries
library(dplyr)
library(ggplot2)
library(s20x)

# Load the relevant dataset (produced using script 11)
data <- read.csv(file = "~/PhD-
work/HTSELEX_analysis_August2019/enrichment_top50_allConditions.csv")
# Prints the loaded data and checks the data structure
data
str(data)

# Conversion of condition to a simplified name
data$condition_new <- data$condition
data$condition_new <- as.character(data$condition_new)
data$condition_new[data$condition_new %in% "Std"] <- "1 - Std"
data$condition_new[data$condition_new %in% "Std+Mut"] <- "2 - Std+Mut"
data$condition_new[data$condition_new %in% "ExtraVol"] <- "3 - ExtraVol"
data$condition_new[data$condition_new %in% "ExtraDet"] <- "4 - ExtraDet"
data$condition_new[data$condition_new %in% "RoomTemp"] <- "5 - RoomTemp"
data$condition_new[data$condition_new %in% "Negative"] <- "6 - Negative"
data$condition_new[data$condition_new %in% "Counter"] <- "7 - Counter"
data$condition_new <- as.factor(data$condition_new)

# Calculates mean, sd, se and IC
# The desired round number needs to be selected here
data_subset <- subset(data, data$Round == '7')
data_subset
str(data_subset)
my_sum <- data_subset %>%
  group_by(Round, condition_new) %>%
  summarise(
    n=n(),
    mean=mean(Enrichment),
    sd=sd(Enrichment)
  ) %>%
  mutate( se=sd/sqrt(n)) %>%
  mutate( ic=se * qt((1-0.05)/2 + .5, n-1))

# Prints the data subset
data_subset

# Graphs the average enrichment showing the standard error as the error
bars
# The graph title is disabled by default but can be enabled by activated
the commented line of code (line 52)
ggplot(my_sum) +
  theme(plot.margin = unit(c(1,1,1,1), "cm")) +
```

```

    geom_bar(aes(x=condition_new, y=mean, fill=condition_new),
stat="identity", alpha=1.0) +
scale_fill_manual(values=c("#1B9E77", "#D95F02", "#7570B3", "#E7298A", "#66A61E",
"#E6AB02", "#A6761D")) +
    geom_errorbar(aes(x=condition_new, ymin=mean-se, ymax=mean+se),
width=0.3, colour="black", alpha=1.0, size=1.0) +
    #ggtitle("Average enrichment of top 50 individual candidates under
\ndifferent selection strategies from SELEX round 6 to 7") +
    xlab("Library") +
    ylab("Average fold-enrichment") +
    labs(fill="Condition") +
    theme(legend.position = "none") +
    theme(axis.text.x = element_text(size = 12.0)) +
    theme(axis.text.y = element_text(size = 12.0)) +
    theme(axis.title.x = element_text(size = 15.0)) +
    theme(axis.title.y = element_text(size = 15.0)) +
    theme(axis.title.x = element_text(face = "bold")) +
    theme(axis.title.y = element_text(face = "bold")) +
    theme(axis.title.x = element_text(vjust = -5.0)) +
    theme(axis.title.y = element_text(vjust = +5.0)) +
    theme(text = element_text(size = 16.0)) +
    theme(plot.title = element_text(vjust = +5.0)) +
    theme(panel.background = element_blank()) +
    theme(panel.border = element_blank()) +
    theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank(), axis.line = element_line(colour = "black"))

```

## D15: Script 15 – Candidate Tracking

```
#Script 15 - Candidate Tracking
#This is an R-script designed to track the frequency of individual
aptamer candidates throughout SELEX
#Specifically, the script will search files from Round 1 to 6 for a
given condition for the top 50 sequences (based on RPM) in Round 7 and
look for matches
#It will return the frequency of each individual candidate in every
SELEX round so the frequency can be tracked
#Copyright Zak Murray 2019

# Clears the R environment
rm(list = ls())
library(dplyr)

#The name of the condition (which is specified in the file name) needs
to be inserted here
ConditionName <- "Counter"

#Loop to trace the frequency of top 50 candidates during SELEX
#Here is what the script is designed to do in order
#1 read the .csv file containing all the frequency counts and
enrichment values
#2 rank the top 50 candidates based on enrichment from round 6 to 7
#3 copy the top 50 rows to a new .csv called "top 50_condition"
#4 search the other .csv files for these sequences and return their
frequencies in each round
#5 put these frequencies in a separate .csv file
#6 add rounds where there are zero counts using another loop which
checks if rounds are present
#7 MANUAL WORK REQUIRED: delete the first two columns of this final and
replace all NA's with 0
#8 graph the frequencies using ggplot2

#1 Reading the .csv file from the last round of SELEX
enrichment_data <- read.table(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/enrichment_",Condi
onName,"_R6+7.tsv", sep=""), sep="\t", header = TRUE, fill = TRUE)
head(enrichment_data)

#2 Ranking the rows based on RPM and then writing the top 50 rows to a
new .csv file
enrichment_data_ranked <-
enrichment_data[order(enrichment_data$Rank..y., decreasing = FALSE),]
head(enrichment_data_ranked)
enrichment_data_ranked_top50 <- enrichment_data_ranked[1:50,]
head(enrichment_data_ranked_top50)
write.csv(enrichment_data_ranked_top50, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/enrichment_",Condi
onName,"_top50.csv", sep=""))

#3 Reading the all the .csv files for a certain condition and then
counting the frequency of each of the top 50 candidates
```

```

list.files(path = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", sep=""), pattern =
"counts.csv")

final_file = NULL
for (round_file in list.files(path = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", sep=""), pattern =
"counts.csv")) {
  loop_file <- read.csv(paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", round_file, sep="")
)
  loop_file
  #This is to match sequences from the ranked enrichment data against
the loop_file, starting with Round 1 and then going through to Round 7
  enrichment_data_allrounds <-
loop_file[na.omit(match(enrichment_data_ranked_top50$Sequence, loop_file
$Sequence)),]
  if(nrow(enrichment_data_allrounds) <1) next
  enrichment_data_allrounds$round <- substr(round_file, 2, 2)
  final_file <- rbind(final_file, enrichment_data_allrounds)
}
enrichment_data_allrounds
head(final_file)
write.csv(final_file, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/", ConditionName, "/", ConditionName, "_fi
nal_file.csv", sep=""))

#4 Adding zero counts for the rounds that don't have any matches

final_file_rounds = NULL
for (aptaboi in unique(final_file$Sequence)) {
  print(aptaboi)
  seq_subset <- subset(final_file, final_file$Sequence == aptaboi)
  rounds_present <- nrow(seq_subset)
  round_check <- if_else(seq_subset$round == 1, TRUE, FALSE, missing =
NULL)
  round_check2 <- if_else(seq_subset$round == 2, TRUE, FALSE, missing =
NULL)
  round_check3 <- if_else(seq_subset$round == 3, TRUE, FALSE, missing =
NULL)
  round_check4 <- if_else(seq_subset$round == 4, TRUE, FALSE, missing =
NULL)
  round_check5 <- if_else(seq_subset$round == 5, TRUE, FALSE, missing =
NULL)
  round_check6 <- if_else(seq_subset$round == 6, TRUE, FALSE, missing =
NULL)
  round_check7 <- if_else(seq_subset$round == 7, TRUE, FALSE, missing =
NULL)

  if(rounds_present == 1) {
    print("Using 1 round parameters")
    if(round_check[1] == FALSE) {
      line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 1)
      seq_subset <- rbind(seq_subset, line)
    }
    if(round_check2[1] == FALSE) {

```

```

    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 2)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check3[1] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 3)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check4[1] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 4)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check5[1] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 5)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check6[1] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 6)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check7[1] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 7)
    seq_subset <- rbind(seq_subset,line)
  }
}

if(rounds_present == 2) {
  print("Using 2 round parameters")
  if(round_check[1] == FALSE & round_check[2] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 1)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check2[1] == FALSE & round_check2[2] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 2)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check3[1] == FALSE & round_check3[2] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 3)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check4[1] == FALSE & round_check4[2] == FALSE) {

```

```

    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 4)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check5[1] == FALSE & round_check5[2] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 5)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check6[1] == FALSE & round_check6[2] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 6)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check7[1] == FALSE & round_check7[2] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 7)
    seq_subset <- rbind(seq_subset,line)
  }
}

if(rounds_present == 3) {
  print("Using 3 round parameters")
  if(round_check[1] == FALSE & round_check[2] == FALSE &
round_check[3] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 1)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check2[1] == FALSE & round_check2[2] == FALSE &
round_check2[3] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 2)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check3[1] == FALSE & round_check3[2] == FALSE &
round_check3[3] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 3)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check4[1] == FALSE & round_check4[2] == FALSE &
round_check4[3] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 4)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check5[1] == FALSE & round_check5[2] == FALSE &
round_check5[3] == FALSE) {

```

```

        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 5)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check6[1] == FALSE & round_check6[2] == FALSE &
round_check6[3] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 6)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check7[1] == FALSE & round_check7[2] == FALSE &
round_check7[3] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 7)
        seq_subset <- rbind(seq_subset,line)
    }
}

if(rounds_present == 4) {
    print("Using 4 round parameters")
    if(round_check[1] == FALSE & round_check[2] == FALSE &
round_check[3] == FALSE & round_check[4] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 1)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check2[1] == FALSE & round_check2[2] == FALSE &
round_check2[3] == FALSE & round_check2[4] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 2)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check3[1] == FALSE & round_check3[2] == FALSE &
round_check3[3] == FALSE & round_check3[4] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 3)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check4[1] == FALSE & round_check4[2] == FALSE &
round_check4[3] == FALSE & round_check4[4] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 4)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check5[1] == FALSE & round_check5[2] == FALSE &
round_check5[3] == FALSE & round_check5[4] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 5)
        seq_subset <- rbind(seq_subset,line)
    }
}

```

```

    if(round_check6[1] == FALSE & round_check6[2] == FALSE &
round_check6[3] == FALSE & round_check6[4] == FALSE) {
      line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 6)
      seq_subset <- rbind(seq_subset,line)
    }
    if(round_check7[1] == FALSE & round_check7[2] == FALSE &
round_check7[3] == FALSE & round_check7[4] == FALSE) {
      line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 7)
      seq_subset <- rbind(seq_subset,line)
    }
  }

  if(rounds_present == 5) {
    print("Using 5 round parameters")
    if(round_check[1] == FALSE & round_check[2] == FALSE &
round_check[3] == FALSE & round_check[4] == FALSE & round_check[5] ==
FALSE) {
      line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 1)
      seq_subset <- rbind(seq_subset,line)
    }
    if(round_check2[1] == FALSE & round_check2[2] == FALSE &
round_check2[3] == FALSE & round_check2[4] == FALSE & round_check2[5]
== FALSE) {
      line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 2)
      seq_subset <- rbind(seq_subset,line)
    }
    if(round_check3[1] == FALSE & round_check3[2] == FALSE &
round_check3[3] == FALSE & round_check3[4] == FALSE & round_check3[5]
== FALSE) {
      line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 3)
      seq_subset <- rbind(seq_subset,line)
    }
    if(round_check4[1] == FALSE & round_check4[2] == FALSE &
round_check4[3] == FALSE & round_check4[4] == FALSE & round_check4[5]
== FALSE) {
      line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 4)
      seq_subset <- rbind(seq_subset,line)
    }
    if(round_check5[1] == FALSE & round_check5[2] == FALSE &
round_check5[3] == FALSE & round_check5[4] == FALSE & round_check5[5]
== FALSE) {
      line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 5)
      seq_subset <- rbind(seq_subset,line)
    }
  }

```



```

    if(round_check6[1] == FALSE & round_check6[2] == FALSE &
round_check6[3] == FALSE & round_check6[4] == FALSE & round_check6[5]
== FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 6)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check7[1] == FALSE & round_check7[2] == FALSE &
round_check7[3] == FALSE & round_check7[4] == FALSE & round_check7[5]
== FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 7)
        seq_subset <- rbind(seq_subset,line)
    }
}

if(rounds_present == 6) {
    print("Using 6 round parameters")
    if(round_check[1] == FALSE & round_check[2] == FALSE &
round_check[3] == FALSE & round_check[4] == FALSE & round_check[5] ==
FALSE & round_check[6] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 1)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check2[1] == FALSE & round_check2[2] == FALSE &
round_check2[3] == FALSE & round_check2[4] == FALSE & round_check2[5]
== FALSE & round_check2[6] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 2)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check3[1] == FALSE & round_check3[2] == FALSE &
round_check3[3] == FALSE & round_check3[4] == FALSE & round_check3[5]
== FALSE & round_check3[6] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 3)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check4[1] == FALSE & round_check4[2] == FALSE &
round_check4[3] == FALSE & round_check4[4] == FALSE & round_check4[5]
== FALSE & round_check4[6] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 4)
        seq_subset <- rbind(seq_subset,line)
    }
    if(round_check5[1] == FALSE & round_check5[2] == FALSE &
round_check5[3] == FALSE & round_check5[4] == FALSE & round_check5[5]
== FALSE & round_check5[6] == FALSE) {
        line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 5)

```

```

    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check6[1] == FALSE & round_check6[2] == FALSE &
round_check6[3] == FALSE & round_check6[4] == FALSE & round_check6[5]
== FALSE & round_check6[6] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 6)
    seq_subset <- rbind(seq_subset,line)
  }
  if(round_check7[1] == FALSE & round_check7[2] == FALSE &
round_check7[3] == FALSE & round_check7[4] == FALSE & round_check7[5]
== FALSE & round_check7[6] == FALSE) {
    line <- data.frame("X"= NA, "Sequence" = aptaboi,
"Length"=seq_subset$Length[1], "Rank" = 0, "Reads"= 0, "RPM"= 0,
"Enrichment" = 0, "round" = 7)
    seq_subset <- rbind(seq_subset,line)
  }
}

if(rounds_present == 7) {
  print("Using 7 round parameters")
}
final_file_rounds <- rbind(final_file_rounds, seq_subset)
}

```

```

write.csv(final_file_rounds, file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/",ConditionName,"_f
inal_withAddedRounds.csv",sep=""))

```

## #5 Graphing the Top 50 candidates

### ### IMPORTANT NOTE ###

# Before graphing data, the above .csv file needs to be opened and the following tasks need to be performed manually:

# (1) The first two columns need to be deleted

# (2) You must use the 'find and replace' function in your spreadsheet software to replace all 'NA' values with 0

# Read edited .csv file

```

top50_RPM_allRounds <- read.csv(file = paste("~/PhD-
work/HTSELEX_analysis_August2019/",ConditionName,"/",ConditionName,"_fi
nal_withAddedRounds.csv",sep=""), header = TRUE)
head(top50_RPM_allRounds)

```

# Selected the relevant columns from the dataset

```

data <- top50_RPM_allRounds %>% dplyr::select(Sequence, RPM, round)
data

```

# Makes a line graph of the RPM of all 50 candidates through SELEX #

```

library(ggplot2)
ggplot(data) +
  theme(plot.margin = unit(c(1,1,1,1), "cm")) +
  geom_line(aes(x=round, y=RPM, colour=Sequence), stat="identity",
alpha=1.0) +
  geom_point(aes(x=round, y=RPM), size=2, shape=21, fill="white") +

```

```

xlab("Selex Round") +
ylab("Frequency (RPM)") +
theme(legend.position = "none") +
theme(axis.text.x = element_text(size = 12.0)) +
theme(axis.text.y = element_text(size = 12.0)) +
theme(axis.title.x = element_text(size = 15.0)) +
theme(axis.title.y = element_text(size = 15.0)) +
theme(axis.title.x = element_text(face = "bold")) +
theme(axis.title.y = element_text(face = "bold")) +
theme(axis.title.x = element_text(vjust = -5.0)) +
theme(axis.title.y = element_text(vjust = +5.0)) +
theme(text = element_text(size = 16.0)) +
theme(plot.title = element_text(vjust = +5.0)) +
theme(panel.background = element_blank()) +
theme(panel.border = element_blank()) +
theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank(), axis.line = element_line(colour = "black"))

# Make a line graph of the average RPM of all 50 candidates through
SELEX #

# Calculates mean, sd, se and IC
my_sum <- data %>%
  group_by(round) %>%
  summarise(
    n=n(),
    mean=mean(RPM),
    sd=sd(RPM)
  ) %>%
  mutate( se=sd/sqrt(n)) %>%
  mutate( ic=se * qt((1-0.05)/2 + .5, n-1))

str(my_sum)
my_sum
?summarise

ggplot(my_sum) +
  theme(plot.margin = unit(c(1,1,1,1), "cm")) +
  geom_line( aes(x=round, y=mean), stat="identity", alpha=1.0,
size=0.8) +
  geom_errorbar( aes(x=round, ymin=mean-se, ymax=mean+se), width=0.1,
alpha=1.0, size=0.8) +
  geom_point( aes(x=round, y=mean), size=4, shape=21, fill="white") +
  #ggtitle("Average RPM of the top 50 aptamer candidates across all
SELEX rounds for different SELEX schemes") +
  xlab("Selex Round") +
  ylab("Average Frequency (RPM)\n")

```

# Appendix E: DNA Quantification of Libraries for High Throughput Sequencing

Table E1: DNA quantities of all libraries sent for high throughput sequencing

Sample ID	SELEX Strategy Description	SELEX Round	Concentration (ng/ $\mu$ L)	Sample Volume ( $\mu$ L)	Amount of DNA (ng)
SL1 dsDNA	Standard	Pre-SELEX	0.42	33	13.73
SL2 dsDNA	Standard + Mutation	Pre-SELEX	0.27	33	9.04
SL3 dsDNA	Extra Volume	Pre-SELEX	0.25	33	8.38
SL4 dsDNA	Extra Detergent	Pre-SELEX	0.21	33	6.93
SL5 dsDNA	Negative Selection	Pre-SELEX	0.23	33	7.66
SL6 dsDNA	Counter Selection	Pre-SELEX	0.19	33	6.34
SL7 dsDNA	Room Temperature Incubation	Pre-SELEX	0.20	33	6.60
S1	Standard	1	1.82	33	60.06
S2	Standard + Mutation	1	2.10	33	69.14
S3	Extra Volume	1	2.03	33	66.83
S4	Extra Detergent	1	2.05	33	67.49
S5	Negative Selection	1	1.79	33	58.91
S6	Counter Selection	1	1.97	33	65.01
S7	Room Temperature Incubation	1	2.12	33	69.96
S8	Standard	2	1.61	33	52.97
S9	Standard + Mutation	2	1.92	33	63.20
S10	Extra Volume	2	2.17	33	71.45
S11	Extra Detergent	2	1.90	33	62.70
S12	Negative Selection	2	2.09	33	68.81
S13	Counter Selection	2	1.92	33	63.36
S14	Room Temperature Incubation	2	2.17	33	71.45
S15	Standard	3	1.52	33	50.16
S16	Standard + Mutation	3	1.66	33	54.78
S17	Extra Volume	3	2.60	33	85.64
S18	Extra Detergent	3	2.17	33	71.45
S19	Negative Selection	3	1.99	33	65.51
S20	Counter Selection	3	2.08	33	68.48
S21	Room Temperature Incubation	3	1.95	33	64.19
S22	Standard	4	1.38	33	45.38
S23	Standard + Mutation	4	1.34	33	44.06

S24	Extra Volume	4	1.73	33	56.93
S25	Extra Detergent	4	1.13	33	37.29
S26	Negative Selection	4	1.20	33	39.44
S27	Counter Selection	4	0.96	33	31.52
S28	Room Temperature Incubation	4	0.84	33	27.64
S29	Standard	5	0.28	33	9.12
S30	Standard + Mutation	5	0.45	33	14.97
S31	Extra Volume	5	1.35	33	44.55
S32	Extra Detergent	5	0.81	33	26.80
S33	Negative Selection	5	0.88	33	28.97
S34	Counter Selection	5	1.62	33	53.30
S35	Room Temperature Incubation	5	0.56	33	18.45
S36	Standard	6	0.46	33	15.15
S37	Standard + Mutation	6	0.12	33	3.91
S38	Extra Volume	6	0.49	33	16.07
S39	Extra Detergent	6	0.39	33	13.00
S40	Negative Selection	6	0.62	33	20.43
S41	Counter Selection	6	1.08	33	35.64
S42	Room Temperature Incubation	6	0.74	33	24.39
S43	Standard	7	0.74	33	24.35
S44	Standard + Mutation	7	0.75	33	24.68
S45	Extra Volume	7	0.93	33	30.71
S46	Extra Detergent	7	0.34	33	11.06
S47	Negative Selection	7	0.64	33	20.99
S48	Counter Selection	7	1.68	33	55.44
S49	Room Temperature Incubation	7	0.95	33	31.30
S50	Monoclonal Library Control (no mutation)	1	5.38	33	177.54
S51	Polyclonal Library Control (no mutation)	1	1.66	33	54.78
S52	Monoclonal Library Control (with mutation)	1	0.14	33	4.46
S53	Polyclonal Library Control (with mutation)	1	0.27	33	8.98
S54	Monoclonal Library Control (no mutation)	2	5.32	33	175.56
S55	Polyclonal Library Control (no mutation)	2	1.80	33	59.40
S56	Monoclonal Library Control (with mutation)	2	6.41	33	211.53
S57	Polyclonal Library Control (with mutation)	2	1.21	33	39.93
S58	Monoclonal Library Control (no mutation)	3	6.54	33	215.82
S59	Polyclonal Library Control (no mutation)	3	2.43	33	80.19
S60	Monoclonal Library Control (with mutation)	3	2.17	33	71.61
S61	Polyclonal Library Control (with mutation)	3	7.55	33	249.15
S62	Monoclonal Library Control (no mutation)	4	2.39	33	78.87
S63	Polyclonal Library Control (no mutation)	4	2.81	33	92.73
S64	Monoclonal Library Control (with mutation)	4	1.91	33	63.03
S65	Polyclonal Library Control (with mutation)	4	3.22	33	106.26

S66	Monoclonal Library Control (no mutation)	5	3.03	33	99.99
S67	Polyclonal Library Control (no mutation)	5	2.59	33	85.47
S68	Monoclonal Library Control (with mutation)	5	3.53	33	116.49
S69	Polyclonal Library Control (with mutation)	5	4.05	33	133.65
S70	Monoclonal Library Control (no mutation)	6	1.12	33	36.96
S71	Polyclonal Library Control (no mutation)	6	1.84	33	60.72
S72	Monoclonal Library Control (with mutation)	6	1.49	33	49.17
S73	Polyclonal Library Control (with mutation)	6	2.02	33	66.66
S74	Monoclonal Library Control (no mutation)	7	2.32	33	76.56
S75	Polyclonal Library Control (no mutation)	7	2.64	33	87.12
S76	Monoclonal Library Control (with mutation)	7	2.25	33	74.25
S77	Polyclonal Library Control (with mutation)	7	2.26	33	74.58

# Appendix F: Gel Electrophoresis Images

## from Control Library PCR's

F1: Control Library Amplification Results for Round 3 (30 PCR cycles)

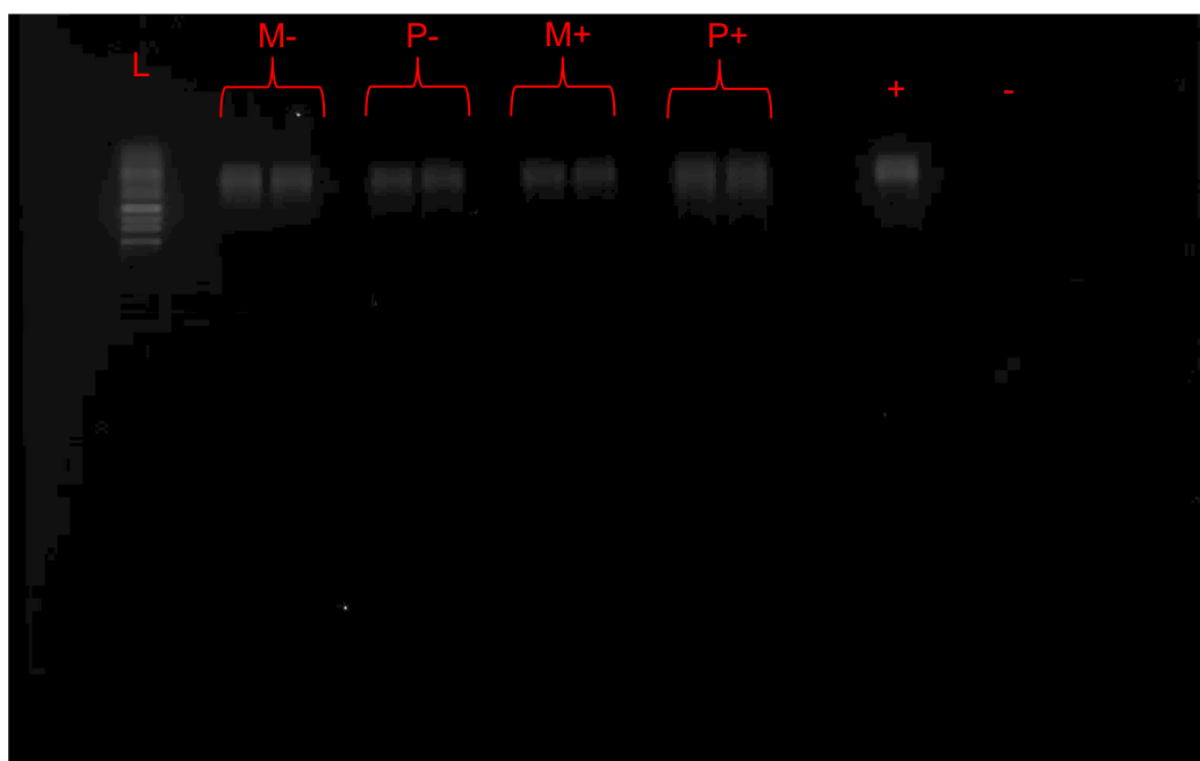


Figure F1: Gel electrophoresis image showing PCR amplicons from the Round 3 amplification of the four control libraries. Lanes are defined as follows: M- (monoclonal library without additional mutation), P- (polyclonal library without additional mutation), M+ (monoclonal library with induced mutation at selective rounds) and P+ (polyclonal library with induced mutation at selective rounds). The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

## F2: Control Library Amplification Results for Round 4 (30 PCR cycles)

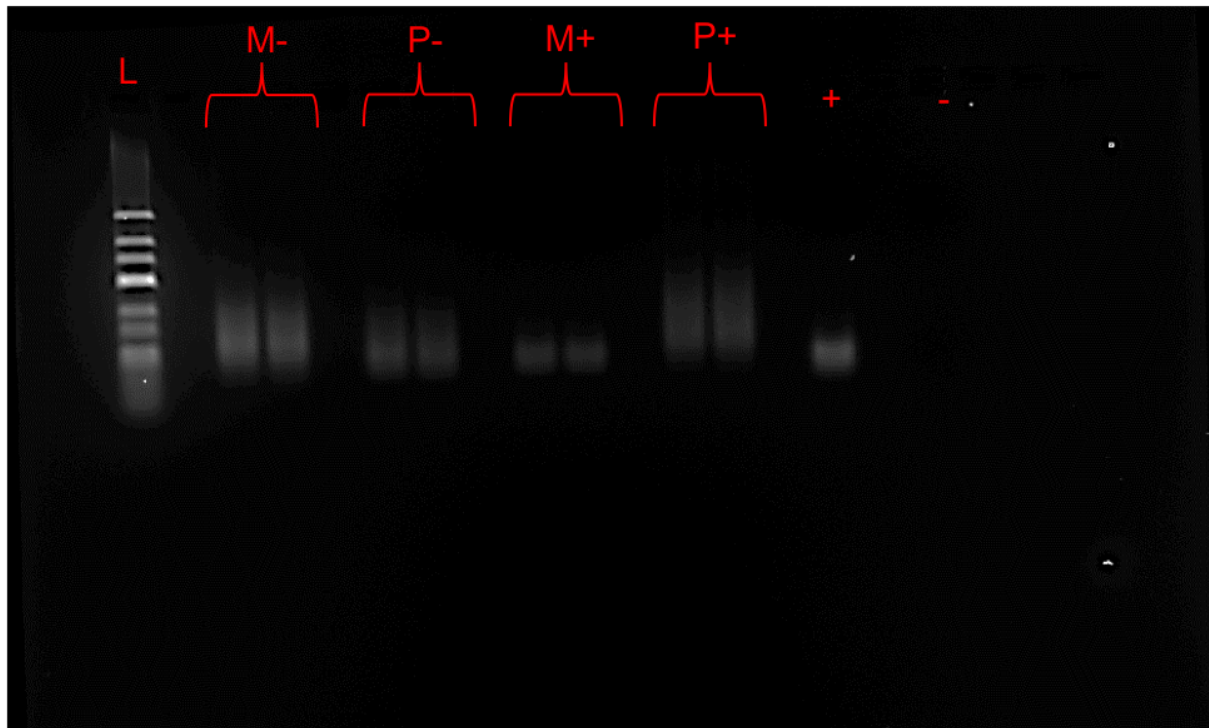


Figure F2: Gel electrophoresis image showing PCR amplicons from the Round 4 amplification (30 PCR cycles) of the four control libraries. Lanes are defined as follows: M- (monoclonal library without additional mutation), P- (polyclonal library without additional mutation), M+ (monoclonal library with induced mutation at selective rounds) and P+ (polyclonal library with induced mutation at selective rounds). The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.



### F3: Control Library Amplification Results for Round 4 (25 PCR cycles)

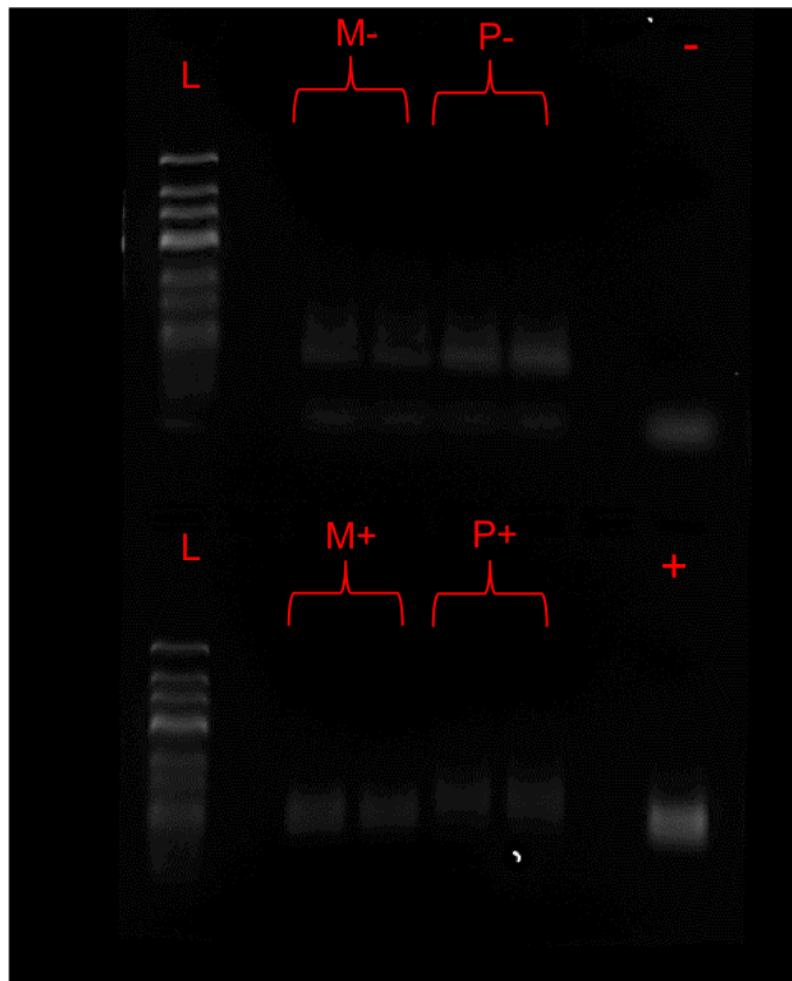


Figure F3: Gel electrophoresis image showing PCR amplicons from the Round 4 amplification (25 PCR cycles) of the four control libraries. Lanes are defined as follows: M- (monoclonal library without additional mutation), P- (polyclonal library without additional mutation), M+ (monoclonal library with induced mutation at selective rounds) and P+ (polyclonal library with induced mutation at selective rounds). The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

#### F4: Control Library Amplification Results for Round 4 (22 PCR cycles)

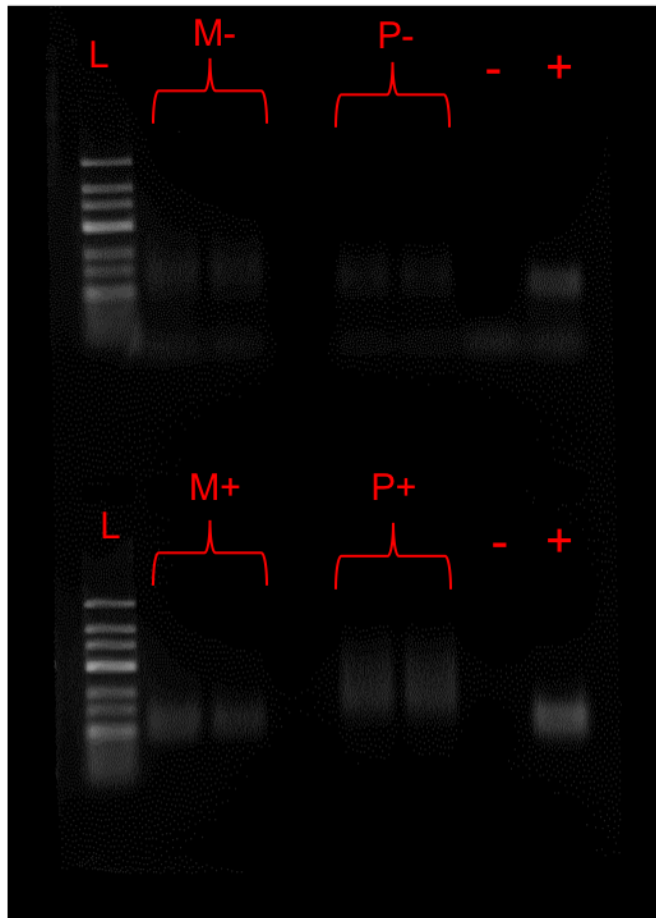


Figure F4: Gel electrophoresis image showing PCR amplicons from the Round 4 amplification (22 PCR cycles) of the four control libraries. Lanes are defined as follows: M- (monoclonal library without additional mutation), P- (polyclonal library without additional mutation), M+ (monoclonal library with induced mutation at selective rounds) and P+ (polyclonal library with induced mutation at selective rounds). The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

## F5: Control Library Amplification Results for Round 5 (22 PCR cycles)

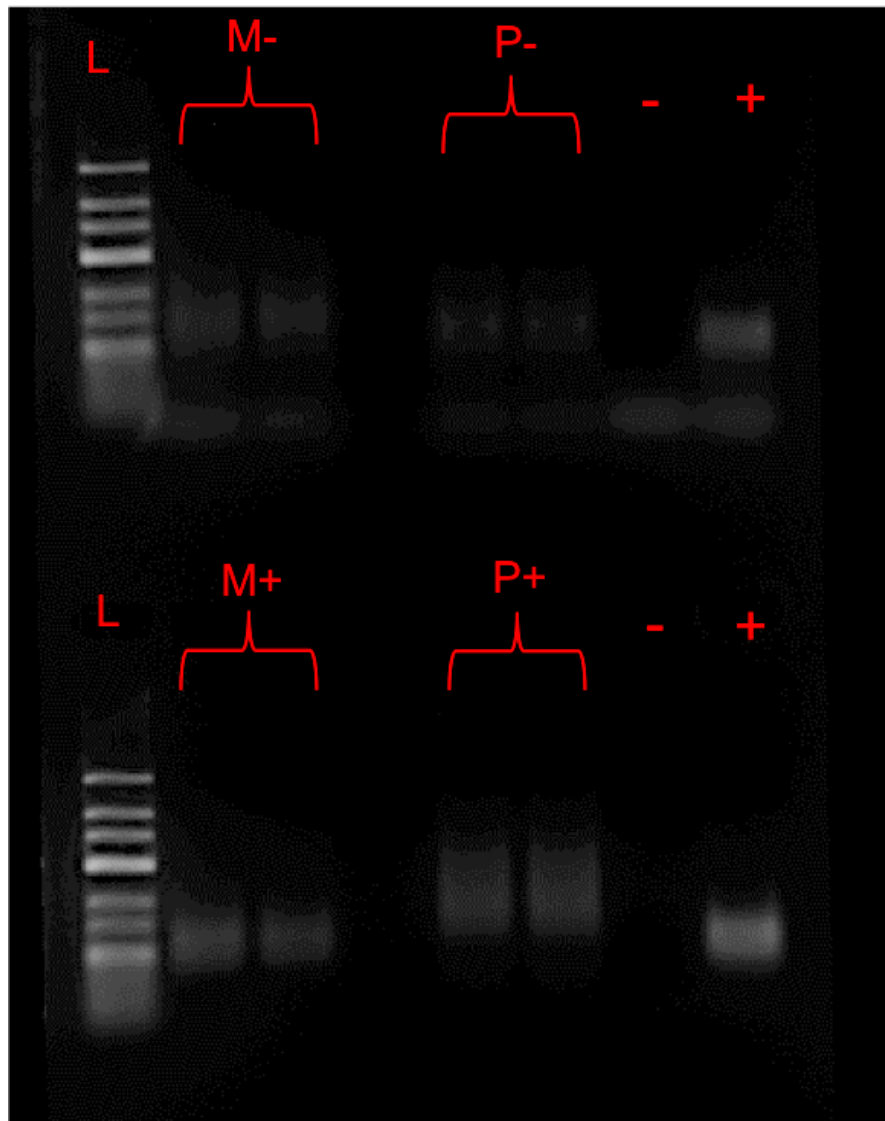


Figure F5: Gel electrophoresis image showing PCR amplicons from the Round 5 amplification (22 PCR cycles) of the four control libraries. Lanes are defined as follows: M- (monoclonal library without additional mutation), P- (polyclonal library without additional mutation), M+ (monoclonal library with induced mutation at selective rounds) and P+ (polyclonal library with induced mutation at selective rounds). The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

## F6: Control Library Amplification Results for Round 6 (25 PCR cycles)

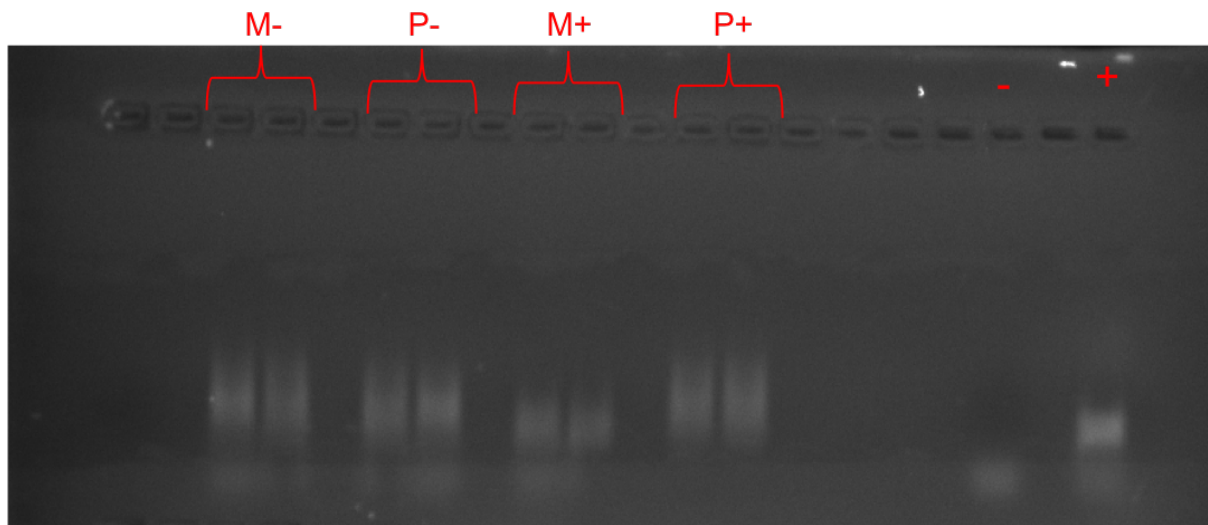
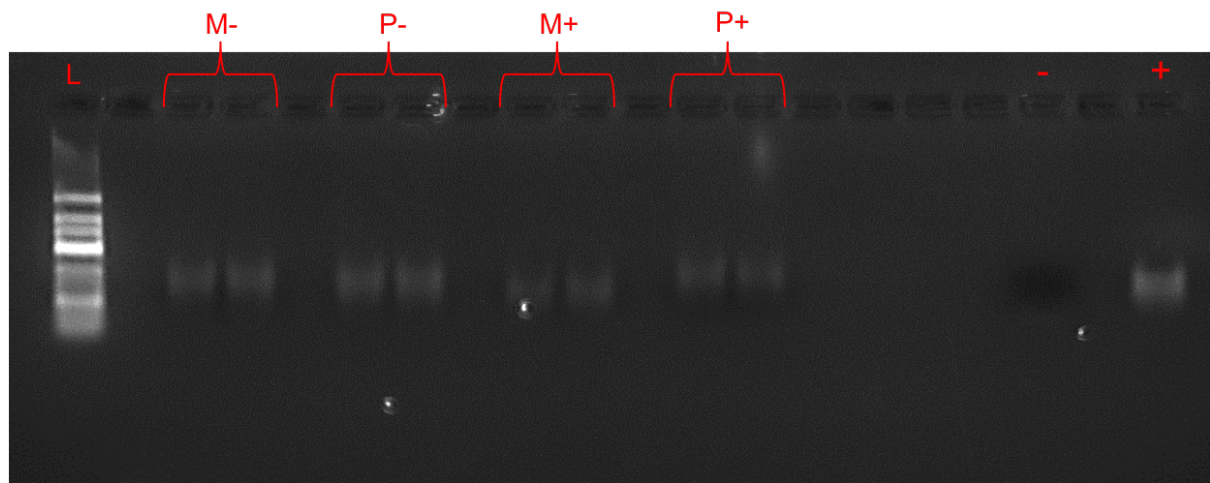


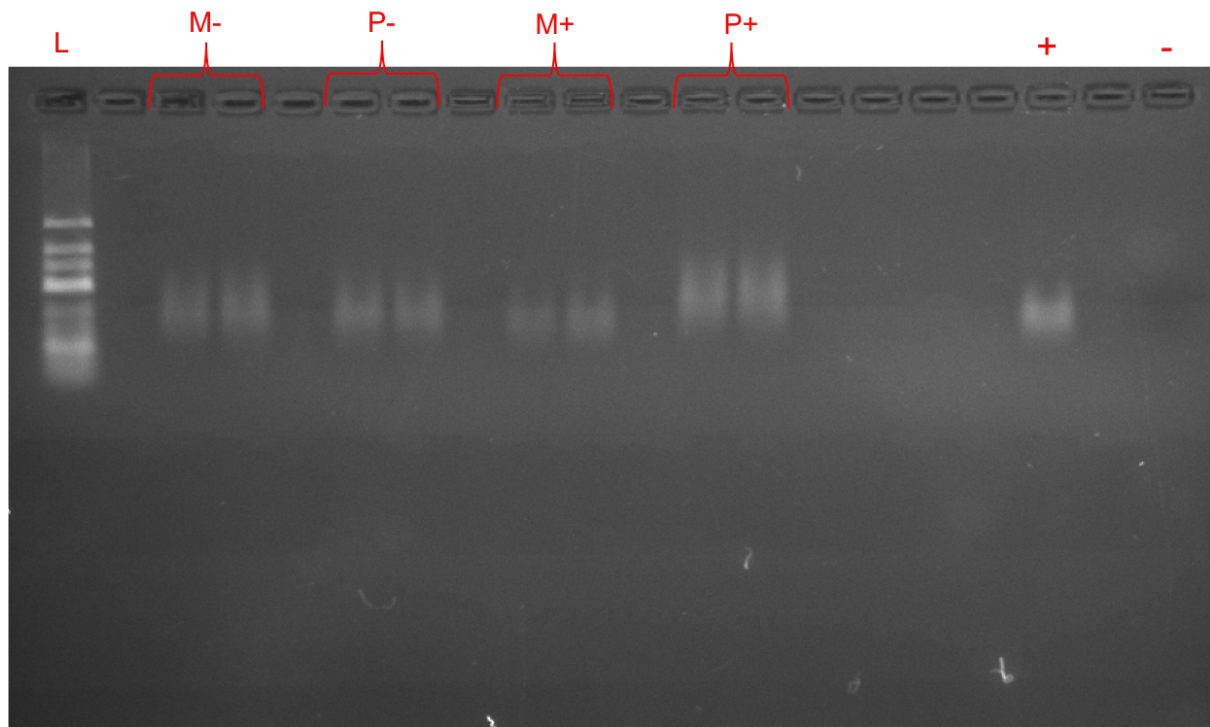
Figure F6: Gel electrophoresis image showing PCR amplicons from the Round 6 amplification (25 PCR cycles) of the four control libraries. Lanes are defined as follows: M- (monoclonal library without additional mutation), P- (polyclonal library without additional mutation), M+ (monoclonal library with induced mutation at selective rounds) and P+ (polyclonal library with induced mutation at selective rounds). The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

## F7: Control Library Amplification Results for Round 6 (22 PCR cycles)



**Figure F7:** Gel electrophoresis image showing PCR amplicons from the Round 6 amplification (22 PCR cycles) of the four control libraries. Lanes are defined as follows: M- (monoclonal library without additional mutation), P- (polyclonal library without additional mutation), M+ (monoclonal library with induced mutation at selective rounds) and P+ (polyclonal library with induced mutation at selective rounds). The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.

## F8: Control Library Amplification Results for Round 7 (22 PCR cycles)



**Figure F8: Gel electrophoresis image showing PCR amplicons from the Round 7 amplification (22 PCR cycles) of the four control libraries. Lanes are defined as follows: M- (monoclonal library without additional mutation), P- (polyclonal library without additional mutation), M+ (monoclonal library with induced mutation at selective rounds) and P+ (polyclonal library with induced mutation at selective rounds). The (-) lane refers to the PCR negative control and the (+) lane refers to the PCR positive control.**

# Appendix G: GC-MS Analysis Methods

Dr Grant Northcott from Plant and Food Research (Hamilton, New Zealand) completed the analysis and quantification of extracted compounds using gas chromatography mass spectrometry (GC-MS). The instrument settings and standards (written by Grant) are outlined below:

Pesticide residues were analysed by high resolution gas chromatography mass-spectrometry (GC-MS) using an Agilent 7000 series triple quadrupole GC-MS operating in single ion monitoring mode. Aliquots of samples and calibration standards (1  $\mu$ L) were injected into an Agilent split/splitless injector held at a temperature of 275°C. The inlet was operated in pressure pulsed injection mode (50 psi for 1. minute) with a split-less time of 1 minute and split flow of 50 mL/min.

The oven temperature was programmed at 80°C (1 min. hold) then increased to 120°C at 10°C/min, increased at 3°C/min to 200°C and held for 5 minutes, increased at 3°C/min to 219°C, increased at 10°C/min to 300°C where it was held for 10 minutes.

The injected samples were separated on an Agilent HP-5MS capillary column (30 m x 0.25 mm x 0.25  $\mu$ m) using helium as carrier gas at a constant flow of 1.2 mL/min. The MSD was operated at 70 eV with a solvent delay of 6.5 mins and ion source, quadrupole and interface temperatures of 230°C, 150°C, and 280°C respectively.

Seven calibration standards spanning a concentration range of 5 to 500 ng/mL were prepared in ethylacetate. Quantitation of pesticide residues was achieved by internal standard quantitation using the Agilent Mass Hunter MSMS Quantitative Analysis Software. The

deuterated chemicals used as internal standards were Bifenthrin-d<sub>5</sub>, acenaphthene-d<sub>10</sub>, phenanthrene-d<sub>10</sub> and pyrene-d<sub>10</sub>.

Residues of industrial alkylphenols and oxybenzone were derivatized to their respective trimethylsilyl ethers by reaction with activated N-Trimethylsilyl-N-methyl trifluoroacetamide(MSTFA) using the method of Budzinski et al (2006).

The derivatised sample extracts and calibration solutions were analysed using an Agilent 6890 gas chromatograph with an Agilent 5975 mass selective detector (MSD) and ATAS Multipurpose auto-sampler. Aliquots of samples and calibration standards (1 µL) were injected into an Agilent split/splitless injector held at 280°C. The inlet was operated in pressure pulsed injection mode (30 psi for 1.1 mins) with a splitless time of 1 minute and split flow of 50 mL/min. The oven temperature was programmed at 90°C (1.5min. hold) then increased to 130°C at 20°C/min, followed by increasing to 280°C at 4°C/min, and to 320°C at 50°C/min where it was held for 5 min. hold, giving a total run time of 47 minutes.

The injected samples were separated on an Agilent HP-5MS capillary column (30 m x 0.25 mm x 0.25 µm) using helium as carrier gas at a constant flow of 1 mL/min. The MSD was operated at 70 eV with a solvent delay of 4.5 mins and ion source, quadrupole and interface temperatures of 230°C, 150°C, and 280°C respectively. Mass spectral data was acquired in synchronous scan/single ion monitoring mode using compound specific mass/charge ions.

Eight calibration standards spanning a concentration range of 5 to 1000 ng/mL were prepared in ethylacetate and used for residue identification and quantitation. Residues in the sample extracts were quantitated by internal standard quantitation using Agilent Chemstation Data Analysis software. The deuterated chemicals used as internal standards were methylparaben-d<sub>4</sub>, isopropylparaben-d<sub>4</sub>, and 4-n-nonylphenol-d<sub>4</sub>.