UPTEC X 07 062
Examensarbete 20 p
November 2007

Optimization of the selector technique for parallel sequencing applications

Anna Åsman

UPPSALA UNIVERSITET

Molecular Biotechnology Programme

Uppsala University School of Engineering

UPTEC X 07 062	**Date of issue 2007-11**

Author

Anna Åsman

Title (English)

Optimization of the selector technique for parallel sequencing applications

Title (Swedish)

Abstract

With the development of second generation sequencing platforms, there is currently a need for techniques capable of massively parallel targeting of genomic regions. The selector method attempts to do this, but suffers from uneven representation of selected regions and artifact build-up. This project aimed at improving the uniformity and finding ways of increasing the specific product yield. Reduction of artifacts was attempted by enzymatic treatment and modification of the selector probe arms. RCA was applied to out-compete unspecific products and improve uniformity. The selector technique is an affordable and efficient tool for resequencing of genomic regions and should be ready for applications such as characterization of cancer cell-lines.

Keywords

Selector technique, second generation sequencing instruments, resequencing, uniformity, artifact reduction, multiplex amplification, RCA

Supervisors

Magnus Isaksson
Mats Nilsson
Department of Genetics and Pathology, Uppsala university

Scientific reviewer

Prof. Siv Andersson
Department of Evolution, Genomics and Systematics, Uppsala university

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **47**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

Optimization of the selector technique for parallel sequencing applications

Anna Åsman

Sammanfattning

De genomiska sekvenserna hos ett stort antal prokaryota och eukaryota organismer, däribland människans, har bestämts under de senaste åren. Idag är fokus inställt på resekvensering, det vill säga sekvensering av intressanta delar av publicerade genom. Genom att bestämma sekvensen hos kandidatgener i cancercellinjer kan t.ex. gener involverade i tumörutveckling hittas och biomarkörer för olika cancerformer utvecklas.

För att kunna skilja ut de intressanta generna från de 6,4 miljarder baspar som det humana genomet utgör, behövs dock speciella metoder. Traditionella tekniker för DNA-kopiering, såsom PCR (Polymerase chain reaction), är dåligt anpassade till de moderna instrument för parallell sekvensering som utvecklats under de senaste åren.

Selektortekniken är en metod utvecklad vid Institutionen för genetik och patologi på Uppsala Universistet som möjliggör infångande av ett stort antal genomiska fragment i en och samma reaktion. Syftet med detta projekt har varit att optimera selektortekniken, med speciellt fokus på två faktorer; spridningen i representationen av produktfragment samt reducering av ospecifik produktuppbyggnad. Dessa två egenskaper är grundläggande för utfallet av selektorassayen i sig och i förlängningen för kvalitén hos de resekvenserade genomfragmenten. Resultat från detta arbete visade att förlängd ligeringstid samt behandling med enzym som bryter ner linjärt DNA (Exonukleas I) ökar selektiviteten hos tekniken. Modifiering av 5’- och 3’-ändarna hos selektorprober visade sig öka mängden specifik produkt i reaktioner innehållande få selektorer men inte i reaktioner av högre komplexitet.

Examensarbete 20 p i Molekylär bioteknikprogrammet
Uppsala universitet November 2007

Abbreviations

bp	base pairs
BSA	Bovine Serum Albumin
DMSO	Dimethyl sulfoxide
dNTP	deoxyribonucleotide triphosphate
dUTP	2'-deoxyuridine 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
DOP-PCR	Degenerate Oligonucleotide-primed PCR
FU	Fluorescence Units
IGP	Department of Genetics and Pathology
IRS-PCR	Interspersed Repetitive Sequence PCR
MDA	Multiple Displacement Amplification
MLGA	Multiplex Ligation dependent Genome Amplification
NCBI	National Centre for Biotechnology Investigation
nt	nucleotides
ON	Over Night
PCR	Polymerase Chain Reaction
RCA	Rolling circle amplification
ROI	Region of interest
UNG	Uracil N-Glycosylase

Table of contents

1. Introduction……..1
1.1 Background……..1
1.1.1 Development of DNA amplification and sequencing…….1
1.1.2 Earlier methods for parallel DNA amplification……1
1.1.3 The selector technique……2
1.1.4 Previous applications of the technique……..2
1.1.3 Principles of the technique……..…3
1.2 Selectors used……...3
1.3 Aims of the project……...3
2. Material and methods……...3
2.1 Selector design……3
2.2 Standard methods……4
2.2.1 Restriction digestion……...4
2.2.2 Ligation……...5
2.2.3 Exonuclease I treatment……..6
2.2.4 PCR……6
2.3 Analysis of PCR products……6
2.3.1 Agarose gel electrophoresis……7
2.3.2 Microfluidics-based gel electrophoresis……7
2.3.3 Capillary gel electrophoresis……...7
2.4 Inclusion of BSA……...……...7
2.5 Study of selector probe- and vector artefacts……...7
2.6 Increased vector- and selector concentrations……7
2.7 Enzymatic treatments……...8
2.7.1 Uracil N-Glycosylase (UNG) treatment……8
2.7.2 T7 exonuclease treatment……...8
2.7.3 Lambda exonuclease treatment……...8
2.7.4 $RecJ_f$ treatment……8
2.7.5 T5 exonuclease treatment……...8
2.8 Validation of selectors in five-, seven- and eight-plex assays……9
2.9 Variations in the ligation protocol……...9
2.9.1 Ligation temperature and cycle number……9
2.9.2 Additives in the ligation mix……10
2.9.3 Ligation using 9° N™ DNA ligase……...10
2.10 Modifications of 5'- and 3'-ends of selector probes……10
2.11 Uracil containing selector probes ……11
2.12 Rolling circle amplification (RCA)……11
2.11 Grouping of selectors that perform similarly……...12
3. Results and discussion ……..12
3.1 Validation of the two selector sets in simplex……...12
3.2 Protocol optimization……15
3.2.1 Genomic DNA concentration……...15
3.2.2 Selector probe concentration……15
3.2.3 Inclusion of BSA……..……15
3.2.4 Linearization of circles……17
3.3 Study of selector probe- and vector artefacts……17
3.4 Increased vector- and selector concentration……18
3.5 Enzymatic treatments……19

3.5.1 UNG treatment……..19
3.5.2 T7 exonuclease treatment…….20
3.5.3 Lambda exonuclease treatment…….20
3.5.4 $RecJ_f$ treatment……..21
3.5.5 T5 exonuclease treatment…….21
3.6 Validation of selectors in five-, seven- and eight-plex assays……...23
3.7 Variations in the ligation protocol…….26
3.7.1 Ligation temperature and cycle number……...26
3.7.2 Additives in the ligation mix……29
3.7.3 Ligation using 9°N™ DNA ligase……...29
3.8 Modification of 5'- and 3'-ends of selector probes……...29
3.8.1 Singleplex screening assay……...29
3.8.2 Multiplex assay…….30
3.8.3 Study of genomic DNA concentration……..33
3.9 Uracil containing selector probes……...33
3.10 RCA…….34
3.10.1 RCA with and without primer…….34
3.10.2 RCA and Exonuclease I treatment……..35
3.10.3 PCR cycle number……..36
3.10.4 RCA in 32-plex……...36
3.11 Grouping of selectors that perform similarly……...37
4. Conclusions……38
4.1 Enzymatic treatments to reduce selector probe- and vector artefacts……38
4.2 Ligation protocol……38
4.3 The problem of non-uniformity…….38
4.5 Modification of 5'- and 3'-ends…….39
4.6 Genomic DNA concentration study……...40
5. Other new methods for sequence capture……..40
6. Method developments……41
6.1 Future technological developments……...41
6.2 Future applications……41
7. Acknowledgements……..……42
8. References…….43
9. Appendices……45
9.1 Appendix A – Selector sequences..……...…..…….45
9.2 Appendix B- Capillary electrophoresis electropherogram…...…….47

1. Introduction

1.1 Background

1.1.1 Development of DNA amplification and sequencing

The chain termination technique of DNA sequencing using dideoxyribonucleotides was developed in 1977[1]. Since then, except for the application of capillary gel electrophoresis and fluorescent terminator molecules to the Sanger method, very little has been done to devise alternative sequencing technologies. The recent years have however seen the development of a number of high-throughput sequencing platforms. The ability of these instruments to process a large number of sequences in parallel means that a whole bacterial genome can be sequenced and assembled *de novo* in much less time than what is required using traditional sequencing methods. Likewise, large parts of mammalian genomes can be resequenced fast and efficiently, leading the way for many new applications. In cancer research, these second generation sequencing technologies, some of which produce a billion of base pairs per run[2], can e.g. be used to resequence exons from a large number of cancer-associated genes in parallel. Differences in mutational patterns observed between normal and cancerous cell lines can then give valuable information about the genetics behind a disease and lead to the development of genetic biomarkers.

To be able to selectively sequence specified pieces of a genome, techniques are however needed to sort out the sequences of interest from the large amount of genetic information it contains[3]. In addition, amplification of the genetic material is often needed, since only minute amounts of DNA is usually obtained from patient- or cell line samples. Amplification by the polymerase chain reaction (PCR) causes problems, since either a high number of separate PCRs or a multiplexed PCR need to be conducted. Performing many individual PCRs implies large investments in time, money and workload. Multiplex amplification reactions are on the other hand associated with amplification artifacts due to cross-reactivity between different primers and target molecules. Ten pairs of primers is usually the upper limit of a multiplex PCR reaction[4]. Techniques that allow highly selective targeting and amplification of genomic DNA sequences are therefore currently in demand[3, 4].

1.1.2 Earlier methods for parallel DNA amplification

Some of the first methods for parallel amplification from multiple genomic loci used primers complementary to repetitive sequences (Interspersed repetitive sequence PCR, IRS-PCR)[5]. Since IRS-PCR amplifies only sequences flanked by repeats and since the repeat structure of the human genome is non-uniform, the method suffers from sequence bias. Degenerate oligonucleotide-primed PCR (DOP-PCR)[6] on the other hand, has low selectivity. This technique makes use of partially degenerate primers having specified 5'- and 3'- ends but central degenerate motifs. In the first cycles of the PCR, a low annealing temperature is used, allowing the partially degenerate primers to anneal to many genomic locations. In later cycles, the 5' end sequence introduced by the first PCR is used to specifically generate high amounts of each target sequence.

Multiple displacement amplification (MDA)[7] allows for a more uniform genomic amplification and produces reaction products of more even length than the above mentioned methods. In this technique, Φ29 DNA polymerase and random exonuclease-resistant primers are applied to perform an isothermal amplification reaction. First, a rolling-circle amplification reaction is primed by the random primers and then strand-displacement DNA synthesis takes place. After this, the displaced product strands are used for secondary priming events forming a hyper-branched DNA structure.

1.1.3 The selector technique

The selector technique, which has been developed at the Department of Genetics and Pathology (IGP), allows simultaneous amplification of a high number of DNA sequences with low amounts of unspecific PCR products being generated. This is achieved by combining the high specificities of nucleotide hybridization- and nucleotide ligation reactions.

A selector is composed of a selector probe, which is 70-90 nucleotides (nt) long, and a vector oligonucleotide of a length of 34 nt. The central part of the selector probe is complementary to the vector sequence, which contains annealing sites for a pair of general primers (fig 1a). The 5'- and 3'- arms of each selector probe are complementary to two genomic target sequences and with the aid of a ligase enzyme, circular products can be formed from the vector and the target sequences (fig. 1b). Only target sequences complementary to the sequences of the selector arms are circularized, due to the discriminatory power of the ligation reaction.

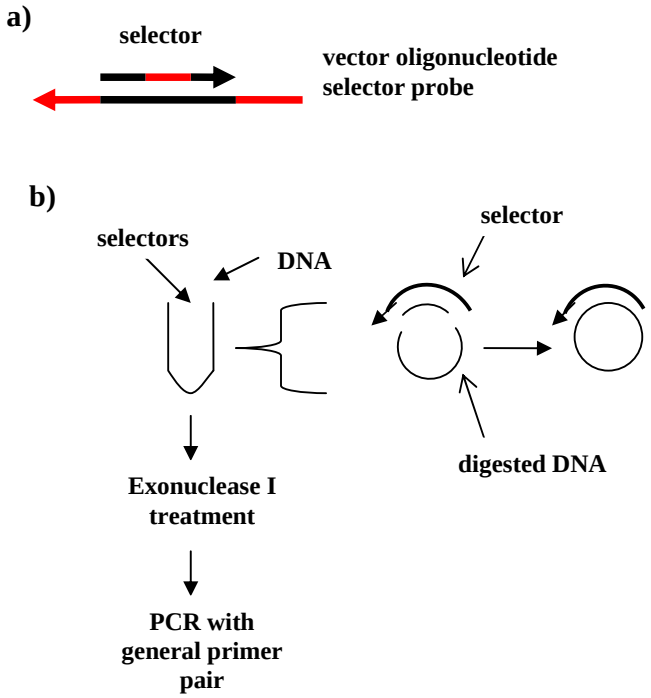

Figure 1. a. A selector is composed of a vector oligonucleotide and a selector probe. The general primer pair of the vector and the target specific arms of the selector probe are colored red.
b. The principle of the selector technique. See the text for details.

1.1.4 Previous applications of the technique

Earlier this year, a paper was published that aimed at evaluating the selector technique by resequencing of ten human genes associated with cancer[4]. This report, though it clearly demonstrated the multiplexing capacity of the selector method and its ability to explore sequence differences between different cancer cell lines, pointed towards some problems with the method. In particular the uneven sequencing depth was found to be a limitation of the technique in its present form. While some selectors produced over a thousand reads in a particular cell line, others did not generate any sequences at all. Methods that make the selectors perform more uniformly will be important in the development of the selector technique, thereby harnessing the full capacity of today's new sequencing technologies[3,4].

The above described study used DNA from seven cancer cell lines and one normal cell line and it targeted the sequences of interest using 508 selectors. A standard selector protocol was used: genomic fragments were ligated to vectors, treated with a uracil excision mix and

amplified in parallel by priming from the general motif in the vector sequence. After this, 454 sequencing of the selected fragments was performed, resulting in an average sequence coverage of 93% per sample. Although successfully demonstrating the feasibility of combining multiplex amplification and parallel sequencing, the representation of the amplified sequences was, as stated above, not uniform. The explanations for this bias have not been determined and could include the fragment selection process, the 454 amplification step, or both. Earlier experiments performed at IGP have indicated that the most bias is introduced at the ligation step, something that recently also has been seen by others[3].

1.1.5 Principles of the technique

The selector technique has several advantages compared to earlier methods for parallel amplification of genomic DNA. Due to the demand for dual recognition between the target fragment and the selector arms, the selectivity of the target recognition reaction is very high. The method can be used to amplify a high number of sequences in parallel, owing to its high selectivity and due to the application of a general primer pair. Since only one probe is needed per locus and since all probes are of the same length, their manufacturing is simple and relatively inexpensive. In addition, the uni-molecular nature of the selector probe makes the ligation reaction fast and efficient. Target sequence identification can be made in different read-out formats, like size separation by agarose- or capillary gel electrophoresis, by sequencing or by performing a TaqMan assay.

1.2 Selectors used

The selectors used in this project have the same target sequences as 24 of the selectors used in the paper by Dahl et al[4]. The central part of the selector probes (34 nucleotides) has been designed to be complementary to the vector used. The selector probe arms have the same sequences as the 24 selectors described above, but they have been extended by five nucleotides in both the 5'- and in the 3'-end. The set of 24 selectors with original arms have been used as controls.

1.3 Aims of the project

One of the aims of the project was to try to increase the uniformity in the amounts of amplification products generated from different selectors. Evidently, an assay managing to capture and amplify any of a number of sequences to an equal extent is preferable to a method that produces a large amount of products from only a limited number of target sequences. The ability to generate product fragments with an even sequence representation will be of large value for resequencing applications.

Another challenge was to find ways of reducing the production of unspecific products that are generated in the PCR step of the selector protocol. These artifacts are built up from selector probes and vectors that remain free in solution after the ligation reaction and the exonuclease treatment. Preventing the build-up of artifacts is important because this increases the sensitivity and possibly also the uniformity of the assay.

In the long term, the objective of developing and improving the selector method is to provide a competitive tool that makes the new large-scale sequencing technologies more efficient and selective.

2. Material and methods

2.1 Selector design

Selector probes were designed to have the same sequences as 24 of the selectors in the paper by Dahl et al[4], except that each arm was extended by five nucleotides in both the 5'- and in the 3'-end. Initially, the design was attempted at constructing these longer probes so that their

arms would have melting temperatures ten degrees above that of the shorter probes. Since these two alternative approaches turned out to be essentially equivalent and since the design becomes simpler when a constant increase in length of all 24 selectors is employed, this option was chosen. This increase in arm length was done as an attempt to increase the selectivity of the selectors. An increase in arm length will result in an increase in melting temperature of the selector probe-target sequence duplex.

Human genomic target sequences were downloaded from the National Centre for Biotechnology Investigation (NCBI) using the *Symphony* software (unpublished). *Symphony* is a program that has been developed at the Molecular Tools group at IGP. It uses a text file containing information about the name of a specific gene, its chromosome number, its start and stop positions on this chromosome and the start and stop positions of the region of interest (ROI). It downloads the specified sequences from the NCBI Human Genome database automatically. The ROI is the region of the gene of interest that contains the sequence targeted by the 5'- and 3'-arms of the selector probe. Sequences of the selector probes used in the paper by Dahl et al were obtained from the authors of the paper (table A1). The correct identities of the 24 downloaded genomic sequences were checked by BLAST searches of the sequences against human genomic sequences at NCBI.

The probe sequences were increased in length by both manual investigation of the target sequence and by using the *ProbeMaker* software[8]. This program takes a set of target sequences and one or more tag sequences and designs probes according to design criteria such as probe length and melting temperature. The program needs information about minimum and maximum lengths of the probes that are to be designed. These numbers were set at 20 and 25 nucleotides respectively, while the preferred hybridization temperature was set at 55˚C and the allowed melting temperature span was set to five degrees. For melting temperature calculations, the program needs values of sodium ion and selector probe concentrations. These were set at 0.2 M and 1 nM, respectively. The melting temperature was calculated by the program by using the Nearest Neighbour model[9], which takes the identity and orientation of the bases on each side of a particular nucleotide into account.

The vector sequence (table 3) was used as the probe tag sequence. This sequence contains annealing sites for the general primer pair (table 3) that is used to amplify the selected fragment. It also contains a sequence that, when combined with its complement in the selector sequence, forms a *Hind* III restriction site.

When the probe design was completed, the short as well as the long selectors were bought from Biomers (Ulm, Germany). The vector and the primers used had previously been designed by researchers at IGP and were available as stock solutions at the laboratory. The forward primer contains a 5'-motif that is not complementary to the vector sequence but which increases the chance that the polymerase used adds an adenosine phosphate to the 3'-end of the target fragment[10].

2.2 Standard methods

The quality of the two sets of selectors designed to have short or long arms were validated, first in simplex- and later in multiplex reactions. A routine protocol that included four steps was used: restriction digestion followed by ligation of target DNA to the vector, Exonuclease I treatment and finally amplification by PCR. The PCR products were analyzed by either agarose gel electrophoresis, by microfluidics-based gel electrophoresis (*Agilent 2100 bioanalyzer*) or by capillary gel electrophoresis.

2.2.1 Restriction digestion

Human genomic male DNA (Promega, Madison, WI, USA) at a final concentration of 40 ng/µl (table 1) was restriction digested by using the one of the three enzyme mixes indicated

in table 1 that generated the required genomic fragment. In all instances when the two sets of 24 short and 24 long selectors were studied, the restriction enzyme mixes denoted in table 2 were applied. Each enzyme was used at a final concentration of 0.4 U/µl. The mixes were incubated (37°C, 1 h) in the NEBuffer (1x) recommended by the manufacturer (table 2), supplemented by 0.1 µg/µl Bovine serum albumin (BSA). Finally, the enzymes were inactivated (65°C, 20 min). Restriction digestion using 10 ng/µl genomic DNA was also attempted.

Table 1. **Human male genomic DNA samples used for restriction digestion (Promega).**

DNA sample	Concentration
G147A; 22549001	233 µg/ml
G147A; 23274801	161 µg/ml

Table 2. Restriction enzymes used.
FspB I was purchased from Fermentas (Burlington, Ontario, Canada). All other enzymes were purchased from New England Biolabs (Ipswich, MA, USA). Contents of buffers are indicated in table 7.

Enzyme mix	Concentrations	NEBuffer
FspB I/ *Alu* I	10 U/µl	NEB 4
Mly I/ *Hpy188* I	10 U/µl	NEB 4
CviA II/ *Bcc* I	5 U/µl, 10 U/µl	NEB 1

2.2.2 Ligation

In all singleplex assays, a 10 µl mix was prepared containing 6.67 mM $MgCl_2$, 0.8 mM NAD, 0.2 U/µl ampligase (Epicentre, Madison, WI, USA), 0.2 U/µl *Taq* Polymerase (Invitrogen, Carlsbad, CA, USA), 0.5 nM vector, 0.1 nM selector probe and PCR buffer (Invitrogen). 0.67x PCR buffer was used when restriction digested mixes contained NEBuffer 4 and 1x PCR buffer was used when mixes contained NEBuffer 1. The different amounts of PCR buffer adjusted the KCl concentration of the samples. Restriction digested DNA (200 ng) in a volume of 5 µl was added to the 10 µl mix.

Sequences of the vector and of the selector probes are indicated in table 3 and in table A1, respectively. The ligation reactants were incubated in a thermal cycler with heated lid (PTC-200, MJ Research) according to table 4 (short probes) or table 5 (long probes). Alternative ligation temperatures that were tried are described in later paragraphs. Decreasing the selector concentration from 0.1 nM to 0.05 nM was also attempted.

The procedure described above was used also in multiplexed (pooled) assays, except for the fact that the final concentration of *each* selector probe was 0.1 nM and that the vector concentration was adjusted so that the ratio of selector probe to vector would be the same as in singleplex assays. The reason for adding an excess of vector to selector probe to the ligation mix, is to make sure that the amounts of vector-selector probe duplexes are sufficient to allow efficient ligation.

Table 3. **Vector and primer sequences.** The recognition site for the *Hind* III restriction enzyme in the vector sequence is indicated by letters in bold italics. The tag IDs refer to identities of the oligonucleotides in OligoDB (IGP, Uppsala University).

Oligonucleotide	Tag ID	Nucleotide sequence
Vector	X01702	5'-CTCGACCGTTAGCA***A|AGCTTT***CTACCGTTATCGT-3'
Primer forward	P2938	5'-GTTTCTTAGCTTTGCTAACGGTCGAG-3'
Primer reverse	P2937	5'-AGCTTTCTACCGTTATCGT-3'
Primer reverse	P01711	5'-FAM-AGCTTTCTACCGTTATCGT-3'
Primer forward	P3535	5'-AAAGTTTCTTAGCTTTGCTAACGGTCGAG-3'
Primer reverse	P3534	5'-AAAAGCTTTCTACCGTTATCGT-3'

Table 4. Temperature scheme for ligation of short selectors.

Temperature (ºC)	Time (min)	
95	5	
75	5	
65	5	
60	5	x3
55	5	
50	10	

Table 5. Temperature scheme for ligation of long selectors.

Temperature (ºC)	Time (min)	
95	5	
85	5	
75	5	
70	5	x3
65	5	
60	10	

2.2.3 Exonuclease I treatment

To remove uncircularized vectors and free selector probes remaining in the ligation mix, both which give rise to artifacts in the PCR step, 10 µl of ligated DNA fragments at a final concentration of 6.67 ng/µl were added to a 10 µl mixture of 0.25 U/µl Exonuclease I (Fermentas), 67 mM Tris-HCl (pH 9.0), 1.7 mM $MgCl_2$ and 0.01 µg/µl BSA. Tris buffer was added to adjust the pH of the reaction. The samples were incubated at 37°C for either 1 h or 30 min without any noticeable differences observed between the incubation times. Then the enzyme was inactivated (70°C, 10 min).

2.2.4 PCR

PCR mixes were prepared by adding 6 µl of the mix containing exonuclease treated DNA to a 19 µl mix of 0.7x PCR buffer (Invitrogen; table 7), 0.25 mM uracil-containing deoxyribonucleotide triphosphates (dNTPs), 0.5 mM $MgCl_2$ (Invitrogen), 0.5 µM of each forward and reverse primer (table 5), 0.2 U/µl *Hind* III restriction enzyme (Fermentas) and 0.02 U/µl Platinum *Taq* DNA polymerase (Invitrogen). The final concentration of DNA in the reaction was 1.6 ng/µl. The restriction enzyme was added to linearize the template as to prevent the polymerase from creating multiple linked copies of the target molecules. Temperature cycling was performed in a thermal cycler with heated lid according to table 6. A PCR without added *Hind* III restriction enzyme but with all other concentrations left unchanged was also performed.

Table 6. **Temperature scheme for PCR.**

Temperature (ºC)	Time	
37	30 min	
95	2 min	
95	15 s	
55	30 s	x33
72	30 s	
72	5 min	

2.3 Analysis of PCR products

The products from a typical PCR reaction were analyzed by gel electrophoresis; either by agarose gel electrophoresis, by using an *Agilent 2100 bioanalyzer* instrument or by capillary gel electrophoresis.

2.3.1 Agarose gel electrophoresis

A 1.5 % agarose gel was prepared in 1xTAE buffer (pH 8) and 1 µl ethidium bromide (1 %) was added. Typically, the gel was run at 90 V for 40 min.

2.3.2 Microfluidics-based gel electrophoresis

The *Agilent 2100 bioanalyzer* is an instrument for capillary gel electrophoresis that separates proteins or fragments of DNA or RNA that are loaded onto specific analysis chips. It can separate DNA fragments that differ in length by a few base pairs. No labeling of the DNA samples is needed, since the gel dye contains a laser-induced fluorescent molecule that intercalates into the DNA samples and makes it possible to determine the length of these according to their retention times in the gel.

Samples were loaded onto a DNA analysis chip according to instructions provided by the manufacturer and the data generated was analyzed using the *2100 Expert* computer software. The limit of peak calling was typically set to 5 Fluorescence Units (FU).

2.3.3 Capillary gel electrophoresis

High-plex samples were analyzed by capillary electrophoresis instead of *Agilent 2100 bioanalyzer*- or agarose gel electrophoresis. When this was the case, samples were sent to *Uppsala Genome Center* at the Rudbeck laboratory. In capillary gel electrophoresis, fragments are separated in a gel instead of directly in the buffer of the capillaries. The technique has higher resolution power than ordinary gel electrophoresis; it can separate DNA sequences that have size differences of no more than one base pair. Labeling of the samples is required and herein this was achieved by using a reverse primer containing a 5'-FAM (6-carboxyfluorescein) modification in a standard PCR reaction.

2.4 Study of inclusion of BSA

Coating of the walls of plastic reaction tubes with BSA is a standard procedure employed to make DNA more available to the reactants in a reaction mix. The effect of adding BSA to reaction mixes was studied by comparing products that had been ligated and Exonuclease I treated in the presence of 0.1 µg/µl BSA with products that had been treated the same way but in reaction mixes without BSA. The rolling circle amplification (RCA) reaction was studied in the same way. Except for BSA addition, the reaction conditions were as described (2.2.2, 2.2.3 and 2.12).

2.5 Study of selector probe- and vector artifacts

To determine which of the PCR artifacts that were vector- respectively selector probe dependent, ligation reactions including only vectors or only selector probes were performed. A ligation reaction including both vector and selector probe was also executed and all reactions were conducted both in the absence and in the presence of added restriction digested genomic DNA. The standard vector- and selector probe concentrations were used in all reactions; 0.5 and 0.1 nM, respectively. In the samples incubated without DNA, $MgCl_2$ was added at a final concentration of 10 mM instead of 6.67 mM, to compensate for the $MgCl_2$ present in the NEBuffer used in the restriction digestion step. All other reaction conditions were as described in paragraph 2.2.2. Two selectors from the five-plex set (selectors 190, 155; table 8) were used and the reactions were performed in singleplex.

2.6 Increased vector- and selector concentrations

As a means to study how the build-up of specific- and unspecific products depends on the concentrations of vectors and selector probes, these concentrations were increased ten or then 100 times. Three selectors were used, all able to produce specific products as shown by

previous experiments (selectors 190, 155, 236; table 8) and the reactions were performed in singleplex.

2.7 Enzymatic treatments

2.7.1 Uracil N-Glycosylase (UNG) treatment

Exonuclease I treated products were treated with UNG (Fermentas) to examine whether artifacts that were seen in PCR products were produced in the PCR or if they were contaminants produced in an earlier PCR and then accidentally introduced into the present amplification mix. UNG was added to the PCR mix at a final concentration of 0.004 U/µl and the samples were incubated (10 min, 37°C) before the standard PCR program was initiated (2.2.4). The time of the first PCR step at 95°C was increased from 2 min to 10 min to fully denature the UNG enzyme.

2.7.2 T7 exonuclease treatment

Two different buffer conditions were tried using T7 exonuclease (New England Biolabs). In the first attempt, the reaction conditions were essentially the same as those of the Exonuclease I treatment (2.2.3). The samples were however incubated at 25°C instead of 37°C and the amount of enzyme was adjusted to match the final concentration of Exonuclease I (0.25 U/µl). In the second approach, the Tris buffer was changed to 1x NEB4 buffer (New England Biolabs) as recommended by the manufacturer. This changed the pH of the reaction mixture from pH 9.0 to pH 7.9, which is more favourable for the T7 exonuclease. Three selectors were used in singleplex and each ligation product was treated with either T7 exonuclease, Exonuclease I or with both enzymes (at final concentration 0.25 U/µl of each). Ligation had been carried out using the protocol of the long selectors (table 5). The samples treated with both exonucleases were incubated at 25°C for 30 min followed by 37 °C for 30 min.

2.7.3 Lambda exonuclease treatment

Treatment with λ exonuclease (New England Biolabs) was performed as described above for the Exonuclease I treatment except that the Tris buffer was changed to 1x λ exonuclease reaction buffer (New England Biolabs; table 7). 0.25 U/µl of enzyme was used. Three selectors were used in singleplex and each ligation product was treated with either λ exonuclease, Exonuclease I or with both enzymes. Ligation had been carried out using the protocol of the long selectors (table 5).

2.7.4 $RecJ_f$ treatment

The NEBuffer 2 was used (at a final concentration of 1x) instead of Tris buffer when samples were treated with $RecJ_f$ (New England Biolabs). Otherwise the reaction conditions were the same as for the Exonuclease I treatment. Two different enzyme concentrations were used (0.25 U/µl and 2.5 U/µl). Three selectors were used in singleplex and ligation had been carried out using the protocol of the long selectors (table 5).

2.7.5 T5 exonuclease treatment

Treatment with T5 exonuclease (Epicentre) was carried out in 1x T5 exonuclease buffer (Epicentre; table 7). Otherwise the reaction conditions were the same as for the Exonuclease I treatment. The five-plex selector set in table 8 was used in a five-plex reaction and ligation was carried out over night (ON, 16 h, 60°C).

Table 7. Contents of buffers.

Buffer	Contents (1x buffer)	pH at 25°C
NEBuffer 1	10 mM Bis-Tris-Propane-HCl,10 mM $MgCl_2$, 1 mM DTT	7.0
NEBuffer 2	50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT	7.9
NEBuffer 4	50 mM KAc, 20 mM Tris-Ac, 10 mM $MgAc_2$, 1 mM DTT	7.7
9°NTH buffer	10 mM Tris-HCl, 600 µM ATP, 2.5 mM $MgCl_2$, 2.5 mM DTT, 0.1 % Triton X-100	7.5
λ exonuclease buffer	67 mM Glycine-KOH, 2.5 mM $MgCl_2$, 50 µg/ml BSA	9.4
T5 exonuclease buffer	33 mM Tris-acetate, 66 mM potassium acetate, 10 mM magnesium acetate, 5.0 mM DTT, pH 7.8 at 25°C.	7.8
UNG buffer	20 mM Tris-HCl, 1 mM EDTA, 10 mM NaCl	8.2
Uracil excision buffer	50 mM Tris-HCl, 20 mM $[NH_4]_2SO_4$, 10 mM EDTA	9.0
PCR buffer	20 mM Tris, 50 mM KCl	8.3

2.8 Validation of selectors in five-, seven- and eight-plex assays

As a first attempt to evaluate the performance of the selectors in multiplex, five selectors from the 24-set having long selector arms were chosen (table 8). To study the effect of addition of each individual selector, five selector mini-pools were prepared; to the first mix, only one selector probe was added, to the second mix two selector probes were added and so on. Each selector probe was added at a final concentration of 0.1 nM, summing up to a total of 0.5 nM selector probe in the fifth mix. Seven- and eight-plex mixes were prepared in an analogous fashion. Ligation was then performed as described for multiplex assays above (2.2.2).

Table 8. Selectors used in a five-plex assay.
Indicated are the lengths of the selected genomic fragments and the number of counted reads reported for the short counterparts of the selectors (Dahl et al[4]).

Selector	Counts	Length (bp)
80	3879	152
190	3001	160
155	626	175
236	7371	179
316	0	200

2.9 Variations in the ligation protocol

2.9.1 Ligation temperature and cycle number

Three rounds of temperature cycling were used in the standard ligation protocols (tables 4, 5). Increasing the number of cycles is expected to result in more specific ligation product being produced, since the arms of the selector probes then get more chances of annealing to the target sequences. The number of cycles was increased to 30 in some assays. The effect of ON incubation was also validated. Different ligation temperatures were tried (tables 9, 10) in both the temperature cycling protocol and in the ON ligation protocol, by making use of the temperature gradient function of a thermal cycler. The selector set in table 8 was used. The results were analyzed on the *Agilent 2100 bioanalyzer* and the peak heights of the generated electropherograms were compared.

Table 9. Temperature protocol for cycled ligation. The indicated temperature intervals were partitioned over six different samples. Thus, six different ligation temperatures were tried.

Temperature (ºC)	Time (min)	
95	5	
65 – 89	5	
55 – 79	5	
50 – 74	5	x3/x30
45 – 69	5	
40 - 64	10	

Table 10. Temperature protocol for ON ligation. As indicated, six different ligation temperatures were tried

Temperature (ºC)	Time
95	15 min
60	20 min
46.8/ 49.9/ 54.3 / 60.3/ 65.0/ 68.1	ON (16 h)

2.9.2 Additives in the ligation mix

PCR additives were added to the ligation mixture in an attempt to decrease the formation of secondary structures in the target molecules. Using the PCR additives dimethyl sulfoxide (DMSO) and formamide, ligation was performed according to the standard protocol (2.2.2), but including 5% DMSO in one tube and 2.5% formamide in another tube. The five-plex selector mix in table 8 was used.

2.9.3 Ligation using 9° N™ DNA ligase

The thermostable 9°N DNA ligase (New England Biolabs) was used in some assays instead of the standard ampligase enzyme. In a typical five-plex assay using the selectors in table 8, a 10 µl mix was prepared containing 0.2 U/µl 9°N DNA ligase, 1x 9°N DNA ligase buffer (New England Biolabs; table 7), 0.2 U/µl *Taq* Polymerase (Invitrogen), 3 nM vector and 0.1 nM of each selector probe. Then 200 ng of restriction digested DNA in a volume of 5 µl was added to the 10 µl mix.

2.10 Modifications of 5'- and 3'-ends of selector probes

Short tails of nucleotides can be added to the 5'- and 3'-ends of oligonucleotides to prevent their ends from annealing to each other and causing build up of unspecific products in PCRs.

To perform a first pilot study, two selectors (selectors 190, 155; table 8) were redesigned to have 5'-tails of three adenosine phosphates and 3'-tails of three thymidine phosphates. The general primer pair (table 3) was also redesigned; a 5'-tail of three adenosine phosphates was added to each primer. Both the modified and the unmodified primers were tested together with the products from the modified selectors in the PCR. The selectors and primers were purchased from Biomers.

When this first screening experiment had been performed, the principle of 5'- and 3'-modification needed testing in a higher-plexed assay. A set of 32 selectors that had been used for another application in the lab (table A2) were therefore manufactured having these 5'- and 3'- modifications. Ligation and Exonuclease I treatment were performed as described, using the modified selectors as well as the unmodified 32-set. The ligation reactions were incubated at 60˚ ON. To generate the target sequences for ligation, the restriction enzyme *Mnl* I was used at a final concentration of 0.4 U/µl. Both the modified and the unmodified forward primer were tested together with the products from the modified selectors in the PCR. A reverse primer labeled with a FAM molecule in its 5'-end was used and the results were analyzed by capillary electrophoresis. All primers were included at a final concentration of 0.1 µM and duplicate samples were prepared. The standard PCR program (table 6) was used, but now including 35 rounds of thermal cycling.

The effect of changing the number of cycles in the PCR was studied by comparing the results from amplification reactions including 30 or 40 cycles. A dilution series was made of restriction digested genomic DNA to study the effect of decreasing the DNA concentration. The PCR products were analyzed by capillary gel electrophoresis and by using the in-house analysis software *SeQuanter* (unpublished). After allocation of specific and unspecific peaks, the peak areas of the specific products at each genomic DNA concentration were summed. The ratio between the summed peak areas of the specific peaks and the total peak area of all peaks was then calculated. As a measure of the variation of the assay, the minimum- and maximum values of the peak areas of the duplicates were used. The summed peaks areas of the artifact peaks and the percentage of peaks called at each genomic DNA concentration were also calculated.

2.11 Uracil containing selector probes

Sixteen of the selectors in the 32-selector set described (2.10) were manufactured containing uridine phosphates instead of thymidine phosphates. These were validated in 16-plex using the standard protocol described above but using a step of UNG treatment instead of the standard Exonuclease I treatment. Ligated DNA fragments in a volume of 5 µl were added to a 10 µl mix containing 0.1 U/µl UNG (Fermentas) in 1x UNG reaction buffer (Fermentas; table 7). The final concentration of DNA was 6.67 ng/µl. The samples were incubated (1 h, 37˚C) and the enzyme was inactivated (10 min, 95˚C). A dNTP mix containing dTTPs was used in the subsequent PCR instead of the standard dUTP-containing mix. Control ligation-, exonuclease- and PCR reactions were run in parallel using the corresponding 16 thymine-containing selectors. A dilution series was made of restriction digested genomic DNA to study the effect of decreasing the DNA concentration.

The experiment was repeated using Uracil-DNA Excision Mix (Epicentre) instead of UNG enzyme. This mixture contains two enzymes: heat-killable UNG (HK-UNG), which creates apyrimidinic sites wherever a uridine phosphate is present, and Endonuclease IV, which cleaves the sugar backbone of DNA (Epicentre). Ligated DNA fragments in a volume of 5 µl were added to a 10 µl mix containing 1x Uracil-Excision Enzyme buffer (table 7), 0.05 U/µl Uracil Excision Enzyme mix and 0.01 µg/µl BSA and 5 mM $MgCl_2$. The samples were incubated (1 h, 37˚C) and the enzyme was inactivated (20 min, 80˚C).

2.12 Rolling circle amplification (RCA)

Isothermal amplification of the ligation products using Φ29 polymerase (Fermentas) was applied to enrich for circularized products. This enzyme produces over a thousand copies of a DNA circle in an hour at 37°C[11]. Ligated DNA fragments in a volume of 10 µl were added to a 15 µl mix containing 0.25 mM dNTPs, 0.1 µg/µl BSA, 0.3 U/µl Φ29 polymerase and 0.65x Φ29 reaction buffer (Fermentas). If primer (P2938) was added, its final concentration was 50 nM. The final concentration of DNA was 5.3 ng/µl. The mix was incubated (1 h, 37°C) and the enzyme was inactivated (10 min, 80°C). After this, the product was amplified by PCR as described (2.2.4). In the first experiment performed, no Exonuclease I treatment was carried out, but in a later experiment, some samples were treated with Exonuclease I before RCA. The Exonuclease I treatment was performed as described (2.2.3). Exonuclease treated product in a volume of 10 µl was then added to a 10 µl mix containing 0.25 mM dNTPs, 0.1 µg/µl BSA, 0.3 U/µl Φ29 polymerase and 0.7x Φ29 reaction buffer. The PCR buffer was used to adjust the concentrations of KCl and $MgCl_2$. The final concentration of DNA was 3.3 ng/µl.

To control for between- experimental variations, samples that were either only treated with Exonuclease I or only rolling circle amplified were also included in the experiment. One sample that was neither amplified by Φ29 nor treated with Exonuclease I was also prepared.

The number of PCR cycles performed after the rolling circle amplification step was altered in order to optimize the selector protocol. Ten and 20 cycles were tried and the amounts of specific PCR products were compared between these reactions and the standard 33-cycle protocol.

2.11 Grouping of selectors that perform similarly

Fifteen selectors from the described 32-plex set (2.10) were divided into three subsets. The first group included the five selectors targeting the longest genomic fragments in the set, the second group targeted the five shortest fragments and the third group targeted five fragments in the middle of the span. These were ligated, treated with Exonuclease I and amplified in separate five-plex assays. The amounts of products produced in each case were compared to the amounts produced in a 32-plex reaction. The ligation reactions were incubated at 60˚ ON. To generate the target sequences for ligation, restriction enzyme *Mnl* I was used at a final concentration of 0.4 U/µl.

3. Results and discussion

3.1 Validation of the two selector sets in simplex

To asses the quality and performance of the newly designed long (84 nt, 25 nt arms) and short (74 nt, 20 nt arms) selector probes adapted from the paper by Dahl et al[4], singleplex reactions were performed. This corresponds to matching each single selector probe with restriction digested genomic DNA containing its target. A selector that has successfully found the right sequence is characterized by giving rise to a strong specific band upon gel electrophoresis analysis.

An agarose gel image representing the PCR products from an assay performed using the long selectors is seen in figure 2. As can be seen, all selectors give rise to unspecific product bands in the PCR (the lowest band in each lane, corresponding to a product less than 100 bp long). Depending on the performance of the selector, weaker or stronger specific product bands can be seen. In large, the pattern corresponds to the results presented in the paper by Dahl et al[4]. Selectors that gave rise to few (or no) sequences in that study fail to produce specific fragments in this run. At the same time, those selectors that have specific bands in this assay have many or moderate counts in the mentioned study. It is also observed that the unspecific band is weak in cases where a specific band is present. This is due to competition between the templates that give rise to specific and unspecific products in the PCR.

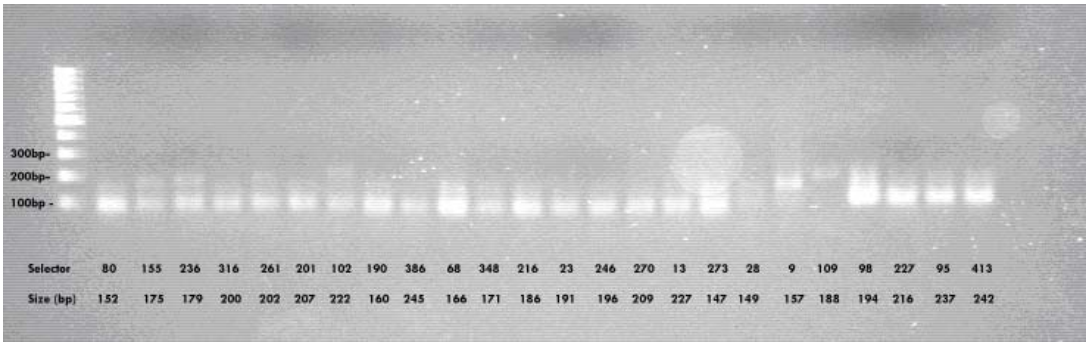

Figure 2. Agarose gel image showing the PCR products produced in validation of 24 of the long (84 bp) selectors in singleplex. Selector number and expected length of specific product are indicated. Lane **1**: 100 bp DNA ladder.

The performances of the long and short versions of a particular selector were compared by running their respective PCR products on the same agarose gel (fig. 3). Since the two sets of experiments have been performed at two separate occasions, no firm conclusions can be

drawn concerning which selector length that is more advantageous. In addition, it is the ratio between specific and unspecific PCR products that is of interest on this gel, not the strength of any particular specific band.

One interesting observation is that the unspecific bands of the short selector probes are shorter than the unspecific bands of the long probes. The reason for this is that a large part of the artifacts contained in the unspecific product smear is produced from selector probes that anneal to each other through their 3'-ends. These 3'-ends are extended in the subsequent PCR and the extended forms of the probes are targeted by the forward primer. The lengths of the artifacts thus created differ by ten base pairs, since they consist of the sequences of two selector arms and two forward primers (figure 4).

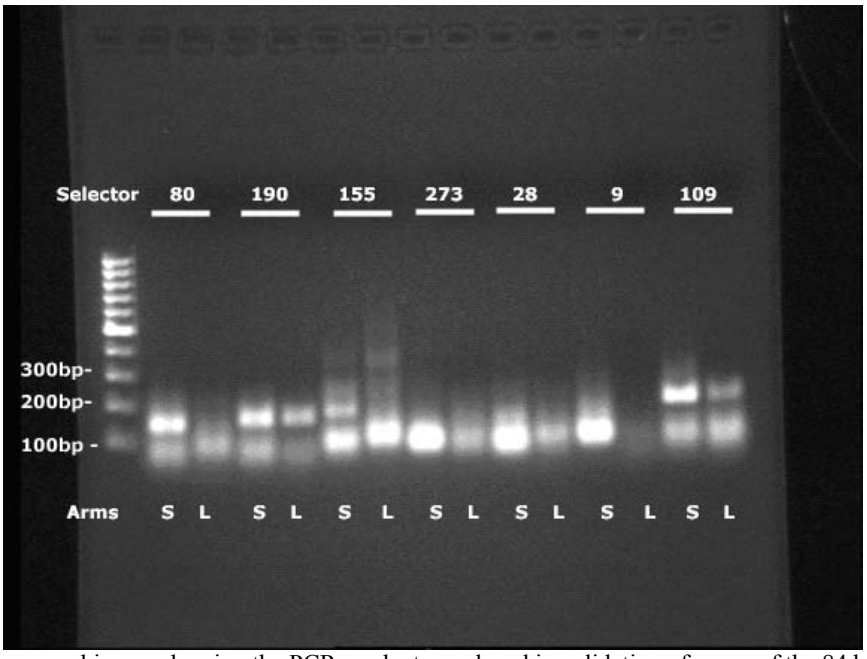

Figure 3. Agarose gel image showing the PCR products produced in validation of seven of the 84 bp- and seven of the 74 bp selectors in singleplex reactions. Selector number and selector probe arm length in nt (S = 20, L = 25) are indicated. Expected lengths of specific products; selector 80: 152 bp, selector 190: 160 bp, selector 155: 175 bp, selector 273: 147 bp, selector 28: 149 bp, selector 9: 157 bp, selector 109: 188 bp. **Lane 1**: 100 bp DNA ladder.

The performances of the long and short selectors were further compared using the five selectors in table 8 in singleplex ligation reactions. Analysis of the PCR products was carried out on the *Agilent 2100 bioanalyzer*. This instrument allows better resolution of unspecific and specific product bands than an agarose gel. As is seen in figure 5, four out of the five selectors produced specific products. Selector 316, which did not give any product, performed very poorly also in the paper by Dahl et al[4], being unable to produce any sequence information after amplification and 454 sequencing (table 8). The outcome of this assay is thereby consistent with those previous results.

Three out of the four successful selectors produced more specific product in their shorter- compared to their longer version. The remaining selector worked better in its long format. This result was intriguing, since ligation was performed using an annealing temperature protocol designed to suit the melting temperatures of the long selectors (table 4). As stated in the introduction, the idea behind increasing the length of each selector arm is to make the ligation reaction more specific, by increasing the ligation temperature. In this way, less

unspecific hybridizations should take place and more specific product should be formed. This seems however not to be the case in this assay.

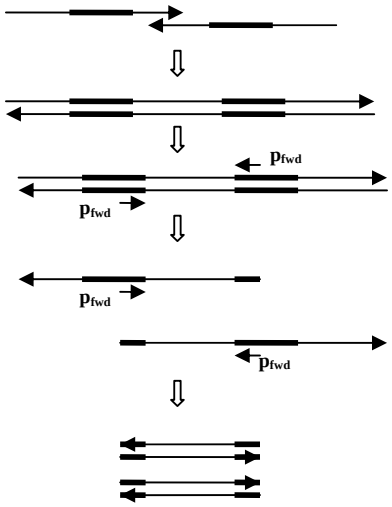

Figure 4. A model of how the 90 bp long artifact is produced. Arrows designate 3'-ends of oligonucleotides. Two selector probes anneal through their 3'-ends and are extended in the PCR. The central vector-complementary sequence (bold) contains sites for annealing of a general primer pair. See the text for further information.

Problems in the chemical synthesis step could be a reason for bad performance of the selectors. Since the amount of wrongly coupled nucleotides increases with the length of a probe, the quality of long selectors is more likely, in general, to be inferior to the quality of shorter selectors (Integrated DNA Technologies). It is however premature to draw any firm conclusions concerning this issue, since only five selectors were tried.

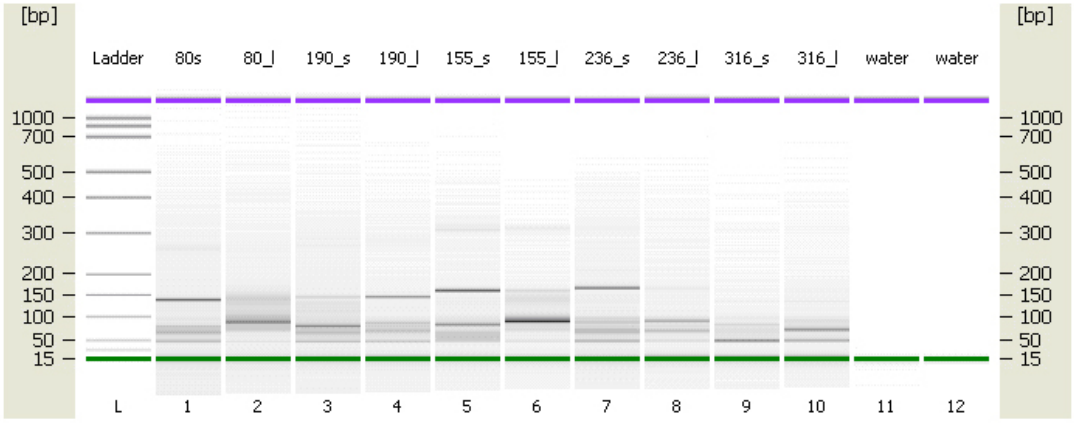

Figure 5. *Agilent 2100 bioanalyzer* gel image showing PCR products produced in validation of five of the 84 bp- and five of the 74 bp selectors in singleplex.
L: ladder. Selector number, arm length and expected length of specific product: **1**:selector 80, 20 nt arms, 152 bp **2**: selector 80, 25 nt arms, 152 bp, **3**: selector 190, 20 nt arms, 160 bp, **4**: selector 190, 25 nt arms, 160 bp, **5**: selector 155, 20 nt arms, 175 bp, **6**: selector 155, 25 nt arms, 175 bp, **7**: selector 236, 20 nt arms, 179 bp, **8**: selector 236, 25 nt arms, 179 bp, **9**: selector 316, 20 nt arms, 200 bp, **10**: selector 316, 25 nt arms, 200 bp, **11-12**: Water loaded. The lengths reported by the instrument deviate by ± 10% (Agilent Technologies) and depend on the internal standard used as well as the length of the detected fragment.

3.2 Protocol optimization

3.2.1 Genomic DNA concentration

Decreasing the genomic DNA concentration from 40 ng/µl to 10 ng/µl was attempted, but this resulted in production of less specific product and more selector probe- and vector dependent artifacts. The production of unspecific products is due to the selector probes annealing to each others (giving rise to the artifacts in fig. 4 above) and to vectors when genomic DNA target fragments are scarce. At the particular selector probe- and vector concentrations used here, this genomic DNA concentration (10 ng/µl) seems to give rise to specific product close to the lower limit of detection.

3.2.2 Selector probe concentration

In an attempt to reduce the production of unspecific products, three selectors were used in singleplex ligation reactions conducted at reduced selector probe concentration (0.05 nM instead of 0.1 nM). PCR products from these reactions were run on the same agarose gel as products from ligations using the normal concentration (fig. 6). The PCR product bands on the gel indicate that performing the ligation reaction using a reduced amount of selector probe decreases the generation of specific product. Two of the selectors failed to give rise to any specific fragment band at all and the third selector showed only a weak specific band on the gel. It is more difficult to see any differences between the amounts of unspecific products. Reducing the amount of selector in the ligation step from 0.1 nM to 0.05 nM does however not seem to improve the performance of the selector assay.

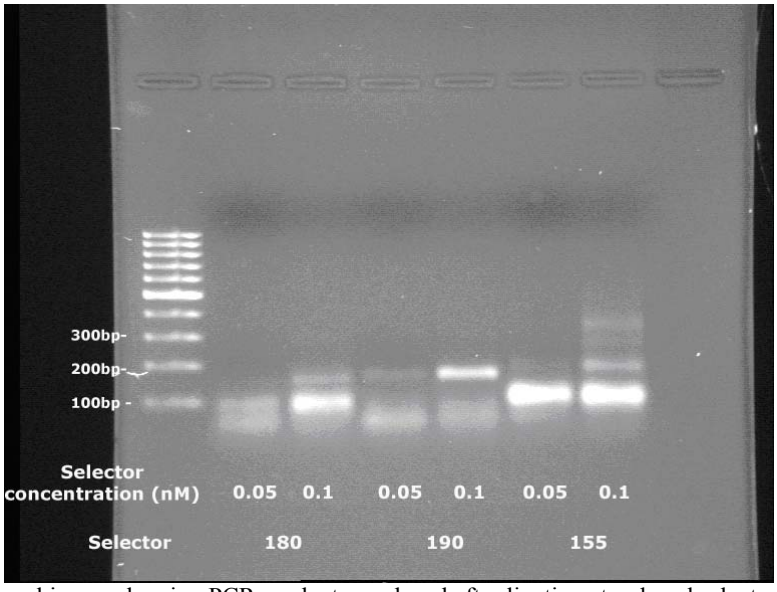

Figure 6. Agarose gel image showing PCR products produced after ligation at reduced selector probe concentration. Concentration of selector and selector number are indicated. Expected lengths of specific products; selector 80: 152 bp, selector 190: 160 bp, selector 155: 175 bp. **Lane 1**: 100 bp DNA ladder.

3.2.3 Inclusion of BSA

BSA inclusion in the ligation- and exonuclease reactions did not seem to have any effect on the generation of specific products as judged by the peak areas of the *Agilent 2100 bioanalyzer* electropherograms.

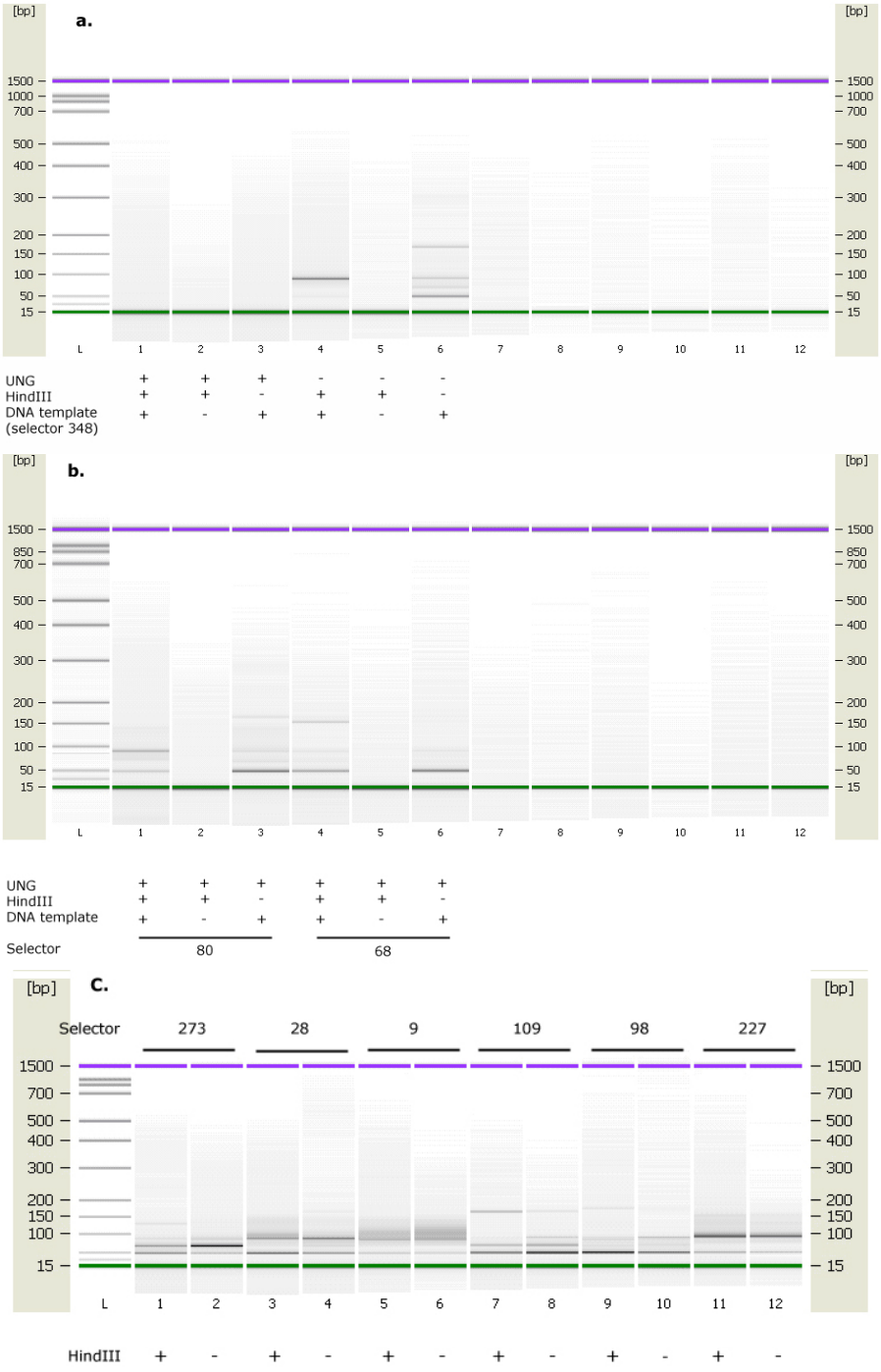

Figure 7a. *Agilent 2100 bioanalyzer* gel image showing the products from PCRs performed with or without added UNG and *Hind* III restriction enzyme. **L**: ladder.
Addition of UNG enzyme, *Hind* III enzyme and genomic DNA is indicated. Selector 348 was used for the ligation reaction. The expected length of the specific fragment is 171 bp.
b. Gel image s showing the products from PCRs performed with or without added UNG and *Hind* III restriction enzyme. Addition of UNG enzyme, *Hind* III enzyme and genomic DNA is indicated. The expected length of the specific product of selector 80 is 152 bp and that of selector 68 is 166 bp.
c. *G*el image showing the products from PCRs performed with or without added *Hind* III restriction enzyme. Selector number and expected length of specific fragment; selector 273, 147 bp, selector 28, 149 bp, selector 9, 186 bp, selector 109, 188 bp, selector 98, 194 bp, selector 227, 216 bp.

3.2.4 Linearization of circles

Omission of *Hind* III restriction enzyme from the PCR was attempted with the aim of studying the importance of linearizing the ligated circles before amplification. For the application of sequencing, opening of the circles is not needed, since formation of concatemers by the *Taq* polymerase presumably does not influence the sequencing reaction. Identification of amplification products by size separation is on the contrary dependent on *Hind* III addition.

The results from assays performed with or without added *Hind* III enzyme were inconclusive. The *Agilent 2100 bioanalyzer* gel images showed different band patterns for treated and untreated samples, indicating differences in the production of the various artifacts and specific products (fig. 7). In most cases, the specific product band was stronger for the treated samples. (This is the case for all three samples that have specific products in fig. 7c and for the one that has specific product in fig. 7b. The reverse is true for the samples in figure 7a.) Concerning the strengths of the artifact bands at 50 and 90 bp in the two cases, no conclusions could be drawn. One complicating fact is that the *Hind* III enzyme was added to each tube separately. This may have resulted in more than the recommended 0.1 µl enzyme per tube being added in each case, due to problems of pipetting this small amount of enzyme in glycerol.

No products longer than 200 bp can be observed on the gels, indicating that the polymerase has not created any multiple-copy products or that these have been out-competed in the PCR.

3.3 Study of selector probe- and vector artifacts

When agarose gels were used to analyze the PCR products, a smear containing fragments of lengths between 50 and 100 bp was seen. This smear consists of various vector- and selector probe dependent artifacts that are produced in the PCR. The different products could be resolved when samples were run on the *Agilent 2100 bioanalyzer* and two bands in particular (50 and 90 bp) were noticeable because of their presence in almost every sample. Because these artifacts were so prevalent, it was of interest to study their origin, i.e. whether they were selector probe- or vector dependent or if they were produced from joining of vectors and selector probes. Reduction of such unspecific products is of importance since they compete with specific targets in the PCR.

Three different types of ligation reactions were performed; without added vector, without added selector probe and with addition of both vector and selector probe. All samples were also prepared in versions including or excluding genomic DNA. The results indicated that the fragment at 50 bp is vector dependent while a fragment at about 90 bp is produced both in ligation mixes containing only vector and in mixes containing only selector probe (fig. 8). The mechanism of production of a band of this length by annealing of two selector probes has been described (fig. 4). Results from this experiment indicate that vectors are able to produce an artifact of similar length. Various weaker unspecific bands were also seen in lanes where samples containing selector probe had been loaded. The exact mechanism behind formation of the different artifacts is hard to determine, but they are most probably produced from selectors or vectors that anneal to each other and form partly double stranded structures. Some of these structures can be degraded by Exonuclease I or another exonuclease, while others are left intact. In the PCR, the universal primer pair anneals to complementary sequences in vectors and selectors and the artifacts are amplified alongside with the specific products.

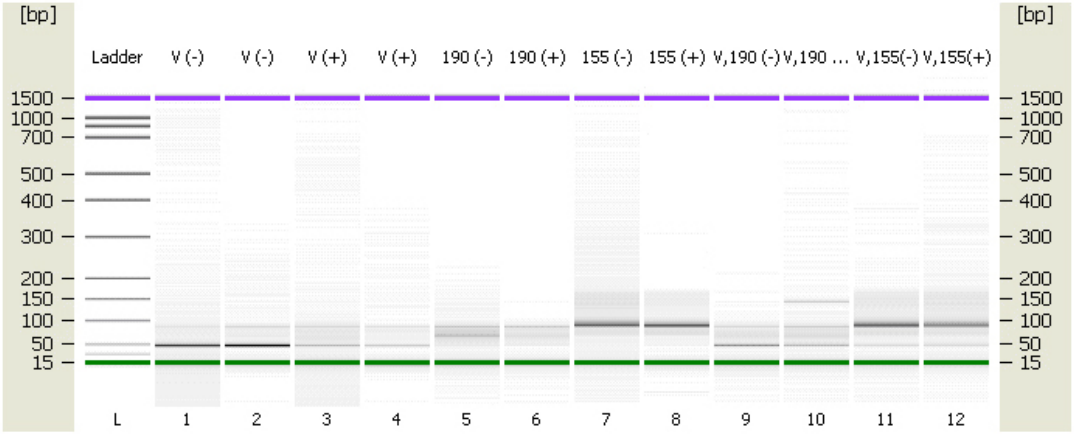

Figure 8. *Agilent 2100 bioanalyzer* gel image showing PCR products from samples incubated with or without vector and selector probe. **L**: ladder, **1-2**: (+) vector, (-) DNA, **3-4**: (+) vector, (+) DNA, **5**: (+) selector 190, (-) DNA, **6**: (+) selector 190, (+) DNA, **7**: (+) selector 155, (-) DNA, **8**: (+) selector 155, (+) DNA, **9**: (+) vector, (+) selector 190, (-) DNA, **10**: (+) vector, (+) selector 190, (+) DNA, **11**: (+) vector, (+) selector 155, (-) DNA, **12**: (+) vector, (+) selector 155, (+) DNA. Expected lengths of specific products; selector 190: 160 bp, selector 155: 175 bp. Specific product is expected in lanes 10 and 12 only.

3.4 Increased vector- and selector concentration

In order to study the effect of the vector- and selector probe concentrations on the amount of specific- and unspecific PCR products, both concentrations were increased ten times or 100 times. In the assay where the concentrations were increased ten times, positive effects on the production of specific products could be observed for all three selectors used (fig. 9). The largest difference was seen for selector 155, which in all previously performed experiments had given rise to a strong unspecific band at 90 bp. This band was much weaker when the concentrations were increased, while at the same time the amount of specific product increased.

At the same time as the 90 bp band of the sample containing selector 155 decreased in strength, the amounts of shorter artifacts increased. This is probably due to competition between the different fragments in the PCR. Some long products with lengths of over 300 bp were formed in samples containing both high and low concentrations of selector 155. These products have been observed in other experiments performed using this selector and are formed from specific products which anneal through their 3'-ends in the PCR and are extended in the way described for selector probes in fig. 4.

Increasing the concentrations 100 times resulted in the build-up of more selector probe- and vector dependent artifacts. In the cases of selectors 190 and 236 this resulted in a considerable decrease in the amounts of specific products formed.

Although it is not possible to draw any firm conclusions from an experiment performed on such a small scale, these results indicate that an increased selector probe- and vector concentration reduces the amount of artifact at 90 bp, while the production of other unspecific products increases. The optimal selector probe- and vector concentrations might be somewhere between ten times and 100 times the one normally used, but further experiments are needed to establish this.

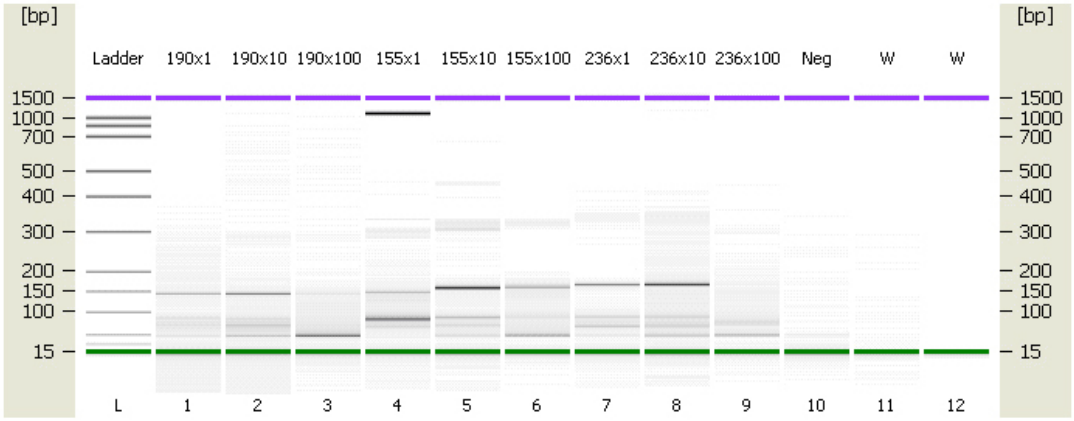

Figure 9. *Agilent 2100 bioanalyzer* gel image showing PCR products from reactions performed with ten and 100 times increased selector probe and vector concentration. **L**: ladder. **1**: 1x selector 190, **2**: 10x selector 190, **3**: 100x selector 190, **4**: 1x selector 155, **5**: 10x selector 155, **6**: 100x selector 155, **7**: 1x selector 236, **8**: 10x selector 236, **9**: 100x selector 236, **10**: negative PCR control, **11-12**: water. The vector concentration is increased to the same extent as indicated for the selector probe in each case. The very sharp band at ~1000 bp in lane 4 is due to the presence of an air bubble in the *Agilent* gel. Expected lengths of specific products; selector 190: 160 bp, selector 155: 175 bp, selector 236: 179 bp.

3.5 Enzymatic treatments

Since the smear of unspecific products seen upon agarose gel electrophoresis was present in all samples analyzed, the possibility existed that some of these products were PCR contaminants. This prompted further investigations using uridine phosphate-containing dNTPs and treatment with UNG enzyme.

The standard way of removing unwanted by-products formed from free selector probes and vectors is treatment with Exonuclease I. However, the artifacts can have partial single- or double-stranded regions, which mean that a mix of exonucleases should be more effective at degrading them than one single enzyme would be. A blend of 5'- to 3'- exonucleases and 3'- to 5'- exonucleases that optimally target both single-stranded and double-stranded DNAs would be preferred. Different enzyme combinations have previously been tried at the lab without success. Some exonucleases have not been able to efficiently degrade their target structures, while others have gained endonucleolytic activities and thereby targeted the circles when they have been incubated under non-optimal buffer conditions (unpublished data).

3.5.1 UNG treatment

All PCR reactions were performed using a dNTP mix containing uridine triphosphates (dUTPs) instead of thymidine triphosphates (dTTPs). This allowed for treatment of the samples with UNG to eliminate PCR carry-over contaminations. UNG cuts uridine monophosphates out of single- or double stranded DNA and leaves an apyrimidinic site which can be cleaved at the backbone by e.g. alkali treatment or, as in this case, heating the sample[12].

Some samples that had previously been Exonuclease I treated were picked randomly and reacted with UNG before PCR. Untreated samples from the same selectors were included in the PCR as controls. As is seen in figure 7a, UNG treatment did not reveal the presence of any uracil-containing PCR contaminant. Instead, the enzyme prevented the build-up of products in the PCR, probably because incubation at 95°C for 10 min does not fully inactivate the enzyme. Indeed, there have been reports of regained activity of UNG after denaturation under certain conditions[13]. In addition, there is a possibility that the amount of enzyme added into each reaction tube was too high, since the pipettes used are not precise enough to pipette 0.1

µl. No firm conclusions could be drawn from this experiment concerning the origin of the supposed contamination bands. Other measures were therefore needed to investigate this.

3.5.2 T7 exonuclease treatment

T7 exonuclease is an enzyme that digests double stranded DNA from the 5'- to the 3'-end. This enzyme should, if incubated under the right buffer conditions and the right temperature, degrade structures such as the double stranded part of selectors. Combining this enzyme with Exonuclease I would provide a means to degrade both single-stranded and double-stranded parts of fragments in the same reaction.

In a first assay, samples were incubated with T7 exonuclease using the above described Exonuclease I protocol (2.2.3). A negative control was prepared by incubating the samples using the Exonuclease I protocol but by adding water to the mix instead of enzyme. The results indicated that, under these reaction conditions, neither Exonuclease I nor T7 exonuclease was able to reduce the amount of artifacts as compared to the negative control.

In the next assay, the reaction conditions of the T7 exonuclease treatment were adjusted by exchanging the Tris buffer (pH 9) with NEB 4 buffer (pH 7.9). The amounts of unspecific products were not significantly different compared to the untreated sample under these conditions either (fig. 10). Possible degradation of circularized product was however observed. The specific product of selector 190 was very weak in the T7 exonuclease treated samples and absent in the samples treated with both T7 exonuclease and Exonuclease I. This could on the one hand be due to a problem in some other step of the selection process. On the other hand are exonucleases sensitive of the secondary structure of the target molecule, which would explain why only one out of the three ligation products was degraded.

Since the combination of Exonuclease I and T7 exonuclease was not successful at reducing the amounts of artifacts, other enzyme combinations were tried.

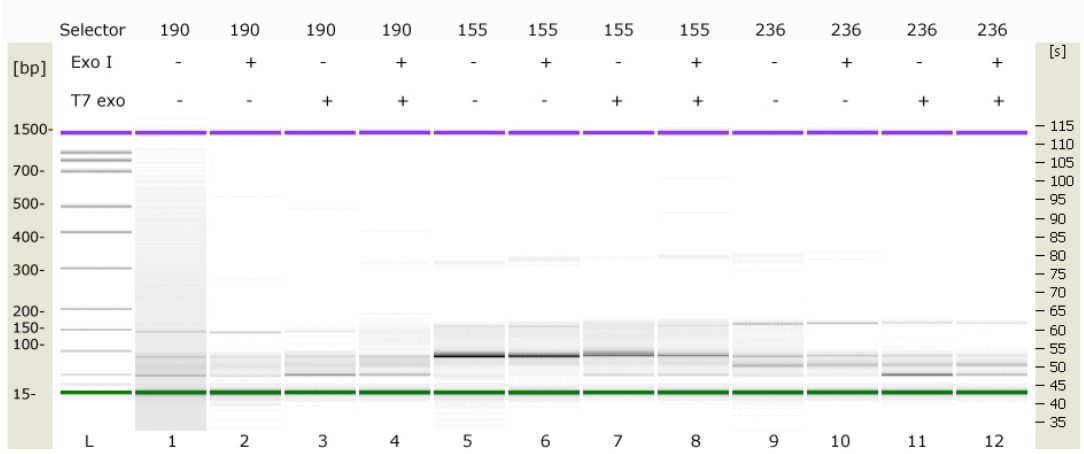

Figure 10. *Agilent 2100 bioanalyzer* gel image showing PCR products from samples treated with Exonuclease I, T7 exonuclease, both enzymes or left untreated. **L**: ladder. Expected lengths of the specific products: selector 190: 160 bp, selector 155: 175 bp, selector 236: 179 bp.

3.5.3 Lambda exonuclease treatment

The next enzyme to be tested was λ exonuclease. Just like T7 exonuclease, this enzyme degrades double stranded DNA from the 5'- to the 3'-end. It has however some characteristics that could make it a better match with Exonuclease I than was T7 exonuclease. λ exonuclease and Exonuclease I are both preferentially incubated at pH 9 and at a temperature of 37°C. In addition, λ exonuclease has higher processivity than T7 exonuclease[14]. When comparing the

samples treated with one or both enzymes to an untreated sample, no differences in the amounts of specific and unspecific products were however seen.

3.5.4 RecJ$_f$ treatment

RecJ$_f$ and Exonuclease I are both single-strand specific exonucleases but they have opposite polarities, RecJ$_f$ degesting from the 5'- towards the 3'-end. The combination of these two enzymes would therefore work well for the purpose degrading single-stranded DNA molecules such as free selector probes and vectors. No effect of RecJ$_f$ treatment was however seen on the total amounts of unspecific products, neither at enzyme concentration 0.25 U/µl nor at 2.5 U/µl. Nonetheless, the artifact band at 50 bp was stronger in the RecJ$_f$ treated samples than in the Exonuclease I treated samples (fig. 11). At the same time, the unspecific product bands of other lengths were weaker, something that would be due to competition in the PCR. The reason for why the relative amounts of unspecific products differ between the two enzymes is unknown, but the same pattern was seen when samples were treated with T5 exonuclease (3.5.5 below).

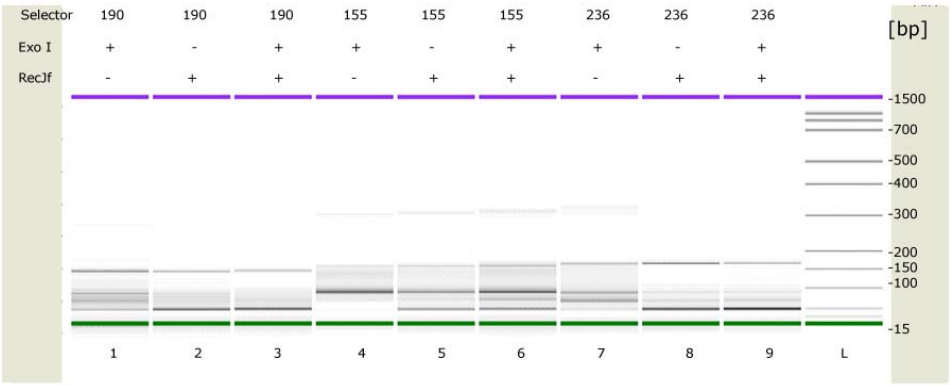

Figure 11. *Agilent 2100 bioanalyzer* gel image showing PCR products from samples treated with Exonuclease I, RecJ$_f$ or both enzymes. **L**: ladder. Expected lengths of the specific products: Selector 190: 160 bp, Selector 155: 175 bp, Selector 236: 179 bp.

3.5.5 T5 exonuclease treatment

Combining the single-strand specific exonucleolytic activities of Exonuclease I (3'- to 5'-polarity) and T5 exonuclease (5'- to 3'-polarity) would in theory be an efficient way of reducing artifact production. These two enzymes would digest all single stranded fragments in the ligation mix, leaving the double stranded molecules intact. T5 exonuclease has however been reported to gain endonucleolytic activity at a magnesium ion concentration of 1-10 mM, a characteristic that could potentially lead to degradation of formed vector-target fragment circles[15]. Indeed, *Agilent 2100 bioanalyzer* analysis of exonuclease treated products that had been PCR amplified showed that less specific products were formed from samples that had been incubated with T5 exonuclease than from samples that had been Exonuclease I treated. The same pattern was seen whether only T5 exonuclease had been used or both T5 exonuclease plus Exonuclease I had been applied.

The amounts of unspecific artifacts were about the same for the three sample sets (Exo I treated-, T5 exo treated- and Exo I plus T5 exo treated; fig. 12), except for the vector-dependent unspecific product at 50 bp, which was reduced to a larger extent in the samples where Exonuclease I had been added. The ligation products treated with T5 exonuclease produced less specific products than a negative control containing all reagents of the Exonuclease I incubation mix except enzyme (fig. 12b). Exonuclease I treated products contained less unspecific- and more specific products than this negative control (fig 12c), indicating that treatment with Exonuclease I alone is the preferred alternative.

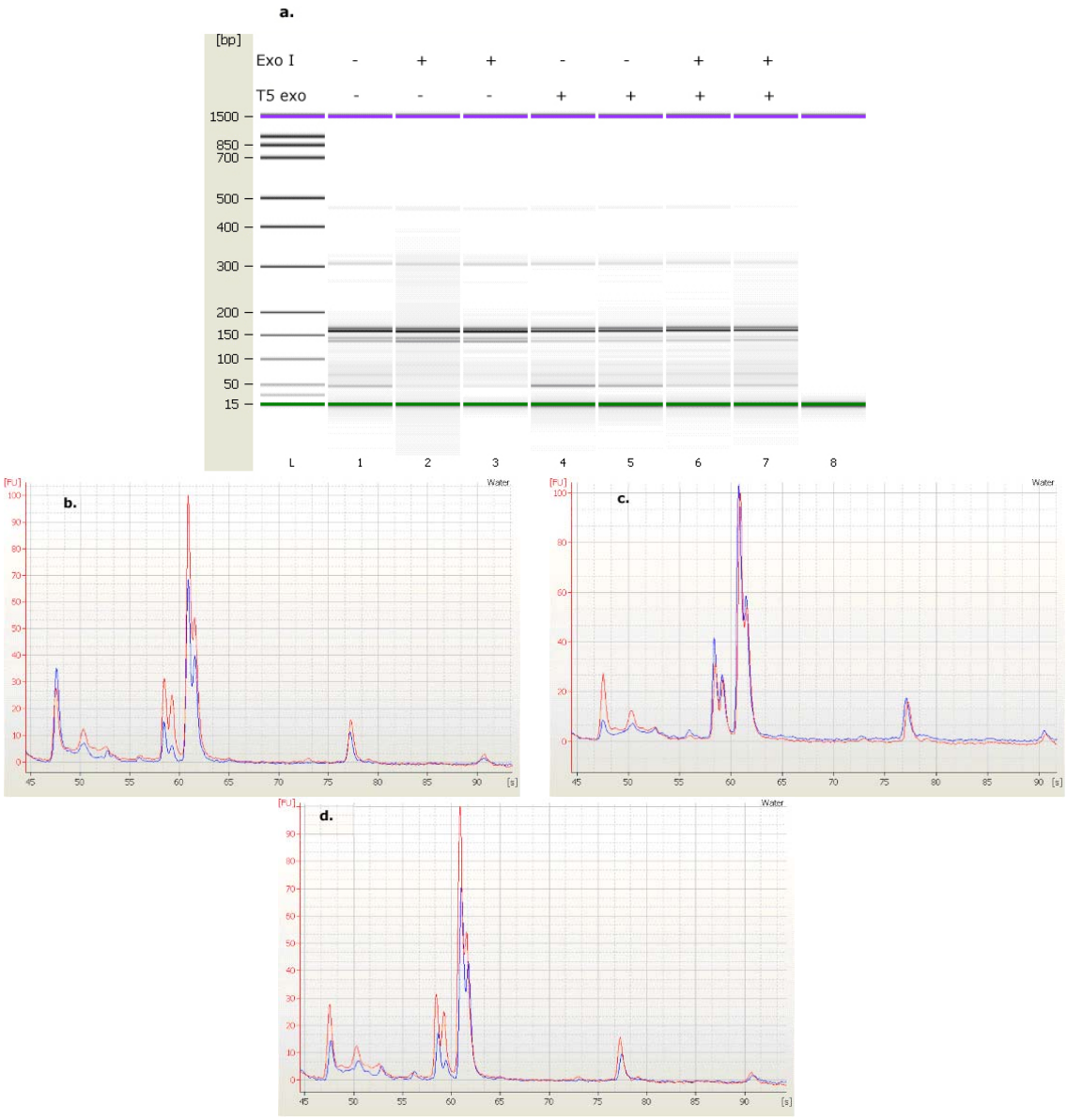

Figure 12 a. *Agilent 2100 bioanalyzer* gel image showing PCR products from samples treated with Exonuclease I, T5 exonuclease, both enzymes or left untreated. **L**: ladder. **8**: Negative PCR control (water added in the PCR mix instead of exonuclease treated ligation product). Expected lengths of specific fragments: 152 bp, 160 bp, 175 bp, 179 bp and 200 bp.
b-d. *Agilent 2100 bioanalyzer* electropherograms showing gel electrophoresis elution peaks. **Red** = untreated sample, **blue** = treated sample. **b** Sample treated with T5 exonuclease. **c.** Sample treated with Exonuclease I. **d.** Sample treated with T5 exonuclease and Exonuclease I. The peaks at 58-60 s correspond to the specific products of selectors 80 (~137 bp) and 190 (~144 bp) and the peaks at 61-64 s correspond to the specific products of selectors 155 (~160 bp) and 236 (~166 bp). The peaks at 47-53 s correspond to unspecific products (50-90 bp).

It was interesting to see that the band at 50 bp was reduced in strength in the Exonuclease I treated sample compared to the negative control, but that T5 exonuclease was not able to break down this artifact. This is in accordance with the pattern seen when the $RecJ_f$- and

Exonuclease I treated samples were compared (3.5.4). Both $RecJ_f$ and T5 exonuclease are single strand specific 5'- to 3'- exonucleases while Exonuclease I has the opposite polarity. Since the mechanism behind production of the 50 bp artifact is not known, it is hard to find a reason for the differences seen between the two classes of enzymes. As will be seen in experiments described later however, the product at 50 bp is a vector dependent artifact.

In conclusion, none of the four enzymes tried (Exonuclease I, T7 exonuclease, λ exonuclease and $RecJ_f$) was able to reduce the amounts of selector probe- and vector dependent artifacts to any significant extent in simplex reactions. It was interesting to see that the Exonuclease I enzyme was ineffective, since this enzyme has long been used routinely at the lab to degrade oligonucleotides before the PCR. On the other hand have results obtained by other researchers at the lab indicated that Exonuclease I treatment is more important in multiplex reactions (unpublished data). This was also seen in multiplexed reactions performed later herein (3.8.2).

There are different explanations for the inability of the four enzymes to eliminate the unspecific products. None of the enzyme combinations tried is able to target all kinds of structures, leaving either single-stranded or double-stranded 5'- or 3'- ends that can hybridize to other molecules or be extended in the subsequent PCR. In addition, the efficiency of removing the protecting group used in the protocol of chemical oligonucleotide synthesis is not 100% (Integrated DNA Technologies). Artificially produced DNA, like the selectors, can therefore be harder to degrade enzymatically than genomic DNA. A factor making the interpretation of the results somewhat vague is the fact that the bands on the *Agilent 2100 bioanalyzer* gel images used to analyze the results were quite weak. Another factor making interpretations difficult is sample-to-sample variations, which make it difficult to draw any firm conclusions concerning product formation in different lanes.

3.6 Validation of selectors in five-, seven- and eight-plex assays

Not all selectors were able to produce specific products in singleplex reactions. Whether this is due to the selectors failing to find the right target fragments in the ligation step or due to PCR amplification problems is currently not known. The selector design step and the chemical synthesis step are also possible sources of error. Chemically synthesized oligonucleotides might be truncated or the protecting group used might not be successfully removed (Integrated DNA Technologies). Since the idea behind the selector technique is to be able to select and to amplify a large number of fragments in parallel, validation of the selectors in higher-plexed reactions is important. In addition, results obtained by others at the lab have indicated that some selectors that do not work well in singleplex reactions perform better in multiplex assays (unpublished data).

In a first attempt to evaluate the performance of the selectors in multiplex, the five selectors in table 8 (in their long versions) were tested in a pooled assay. Specific product bands of the expected lengths were seen for four out of the five selectors added (fig. 13). The fifth selector (selector 316), in its shorter version, performed poorly also in the study performed by Dahl et al[4] (table 8).

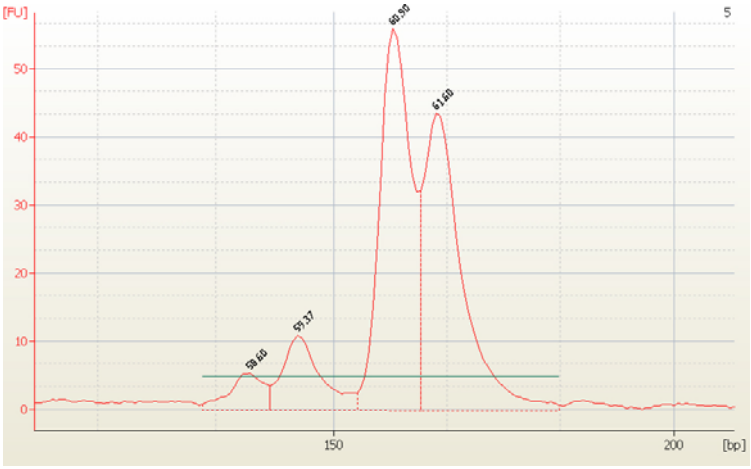

Figure 13. *Agilent 2100 bioanalyzer* electropherograms showing gel electrophoresis elution peaks. Specific products: Selector 80: ~138 bp, Selector 190: ~145 bp, Selector 155: ~159 bp, Selector 236: ~165 bp. Expected lengths of specific fragments: 152 bp, 160 bp, 175 bp, 179 bp, 200 bp.

A seven-plex and an eight- plex assay were also performed (tables 11, 12) using the standard ligation protocol including three rounds of cycling. In these cases, three and four specific products were observed, respectively. As the probes of the absent products probably were inefficient at finding their target fragments, increasing the number of cycles or the ligation time could be ways of obtaining specific products from a higher number of selectors. Incubation of the seven-plex ON at 60°C did however not help the non-working selectors. On the other hand did the bands of the products already present increase in strength, at the expense of artifact production (fig. 14).

Table 11. Selectors used in a 7-plex assay.
The numbers of the wells correspond to the wells in figure 14. Counted reads of the short counterparts of the selectors (Dahl et al[4]) and lengths of the selected genomic fragments are indicated. Selectors 68, 23, 348 and 13 generated specific fragments in singleplex assays.

Well	Selectors	Counts	Lengths (bp)
1, 2	23	1373	191
3, 4	23 + 348	724	191 + 171
5, 6	23 + 348 + 68	4726	191 + 171 + 166
7, 8	23 + 348 + 68 + 13	1876	191 + 171 + 166 + 227
9, 10	23 + 348 + 68 + 13 + 216	9	191 + 171 + 166 + 227 + 186
11, 12	23 + 348 + 68 + 13 + 216 + 270	6	191 + 171 + 166 + 227 + 186 + 209
13, 14	23 + 348 + 68 + 13 + 216 + 270 +346	0	191 + 171 + 166 + 227 + 186 + 209 + 196

Table 12. Selectors used in an 8-plex assay.
Counted reads of the short counterparts of the selectors (Dahl et al[4]) and lengths of the selected genomic fragments are indicated. Selectors 273, 28, 9, 109, 98 and 227 generated specific fragments in singleplex assays.

Selector	Counts	Length (bp)
273	2	147
28	542	149
9	887	157
109	777	188
98	0	194
227	688	216
95	700	237
413	590	242

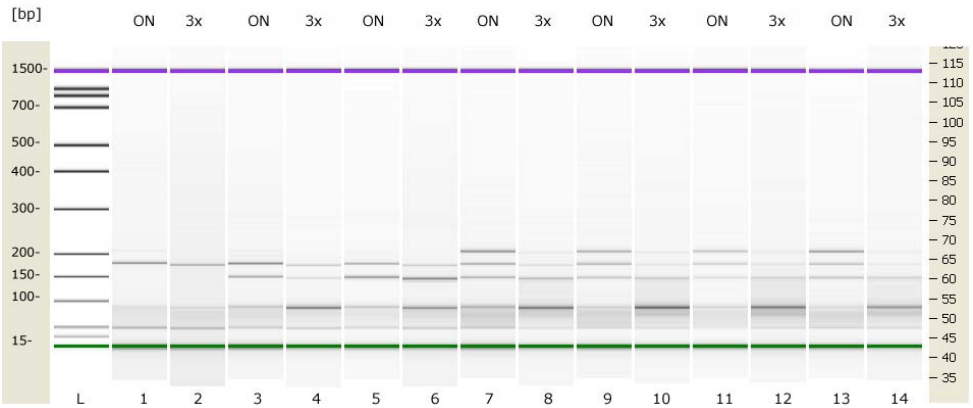

Figure 14. *Agilent 2100 bioanalyzer* gel image showing PCR products from seven-plex samples incubated using a ligation protocol including incubation ON or three rounds of temperature cycling. Selectors and lengths of specific fragments loaded into each well are indicated in table 11. Products having lengths of 50 and 90 bp represent unspecific products while the rest of the bands represent specific products.

Ligation ON at 60°C was also performed using the selectors having short arms in the corresponding five-plex assay (table 8). Samples containing long and short selector probes were treated in identical ways and analyzed on the *Agilent 2100 bioanalyzer*. Even though it is not possible to draw any firm conclusions from an experiment that has not been repeated, a few observations can be made. It seems like the two selectors targeting the longest fragments perform better in their longer version while the two selectors targeting the shorter fragments perform better in their shorter form (fig. 15). Since pipetting errors during preparation of the two five-plex selector mixes could skew the concentrations of the individual selectors, the experiment should have been repeated using newly prepared selector mixes.

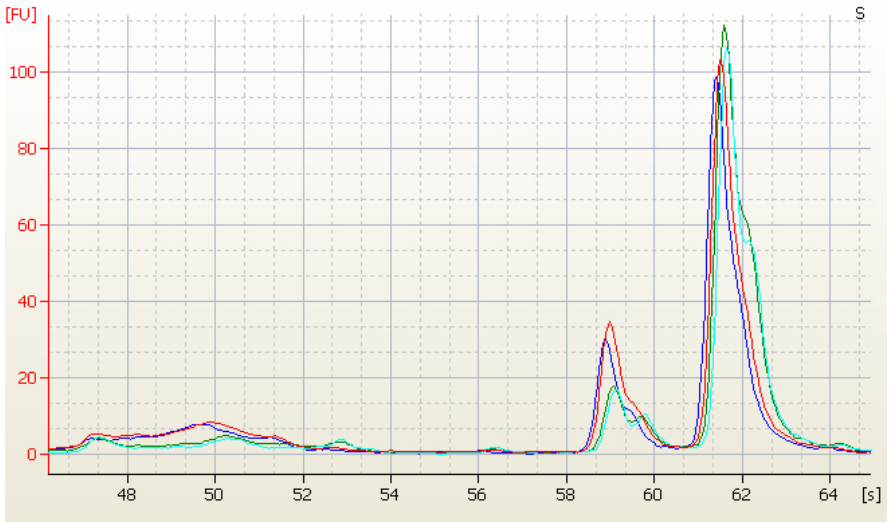

Figure 15. *Agilent 2100 bioanalyzer* electropherograms showing gel electrophoresis elution peaks. Ligation reactions have been performed using long or short selectors**.** Samples were prepared in duplicates. **Red**, **blue** = short arms of probes, **green**, **turquoise** = long arms of probes. Peaks at 59-60 s correspond to the specific products of selectors 80 and 190 (expected lengths of specific products: 152 bp, 160 bp). Peaks at 61-63 s correspond to the specific products of selectors 155 and 236 (expected lengths of specific products: 175 bp, 179 bp).

3.7 Variations in the ligation protocol

The performance of the selector assay is as expected dependent on all of its four parts; restriction fragment generation, ligation, exonuclease treatment and amplification by PCR. The two steps that however are likely to contribute the most to the difference seen in performance between different selectors in a set are the ligation- and the PCR steps. The restriction digestion is in general assumed not to be a problem, something that can be checked by running the cleaved products on an agarose gel. This was performed using the restriction enzyme mixes of table 2 and a smear was seen for each mix, indicating efficient digestion. As stated (3.5.2), exonucleases are sensitive to secondary structures in DNA, which means that the Exonuclease I treatment step could introduce some bias into the process. This potential non-uniformity should however have a much smaller effect than bias in the ligation- and PCR steps.

While bias in the PCR is due chiefly to aspects of the target, such as its GC content and its ability to form secondary structures, unequal ligation efficiencies can have a number of different reasons. Formation of secondary structures in the target- or selector probe sequences, hybridization of selector probes to each other and to vectors, unfavourable annealing temperatures, breakdown of formed circles due to application of an exaggerated annealing temperature and cross-reactivity between selectors and targets are a number of things that can lower the performance of the ligation step. High or low GC content of the arms of the targeting molecule or of the targeted fragment has been observed by others to decrease the performance of the ligation step[3].

3.7.1 Ligation temperature and cycle number

In order to examine the optimal temperature for ligation, a temperature gradient was programmed on a thermal cycler machine. Six identical samples were incubated at six different temperatures. In a first assay, a temperature gradient was applied according to table 9. Using the ligation protocol including three rounds of cycling, the largest amounts of specific products were produced in the samples incubated at the three highest temperatures (fig. 16). When the temperature scheme included 30 cycles, the most products were produced at the second and third highest temperatures. The reason for why less specific product was observed at the highest temperature is probably that incubation at elevated temperature for 30 cycles makes ligated circles break down.

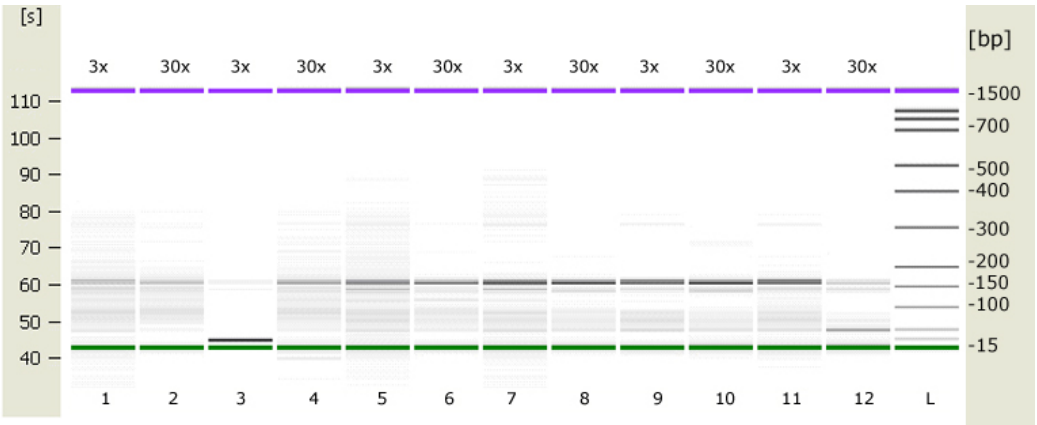

Figure 16. *Agilent 2100 bioanalyzer* gel image showing PCR products from samples incubated at different ligation temperatures for three or 30 cycles. All samples are identical but have been incubated at different ligation temperatures. Temperature 1 is the lowest and temperature 6 is the highest. **1,2**: temperature 1, **3,4**: temperature 2, **5,6**: temperature 3, **7,8**: temperature 4, **9,10**: temperature 5, **11,12**: temperature 6. Temperatures 1-6 correspond to the six ligation temperatures indicated in table 9. Expected lengths of specific fragments; 152 bp, 160 bp, 175 bp, 179 bp and 200 bp.

When ON incubation at a constant temperature was applied, the same pattern was seen as for the samples ligated using the protocol including three cycles; the three highest temperatures (table 10) were most favourable (fig. 17a). The results from these experiments make sense in that these higher temperatures match the melting temperatures of the duplexes formed between the target DNAs and the arms of the selector probes used in the assay (table 13). In addition, the ampligase enzyme used to circularize the targets has an activity optimum at 70°C (Epicentre). As can be seen in figure 17b, incubation using the standard ligation protocol including three cycles (table 5) is as ineffective as incubation ON at the lowest temperature. Ligation at constant temperature was also found to be superior to ligation using 30 cycles (fig. 18).

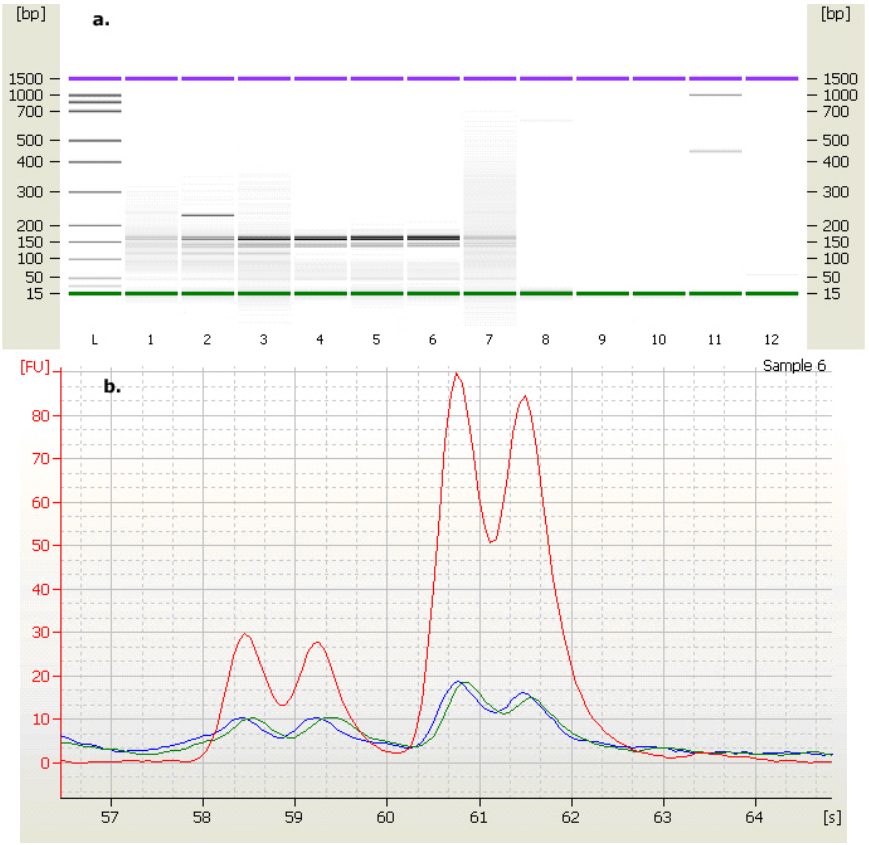

Figure 17a. *Agilent 2100 bioanalyzer* gel image showing PCR products from samples incubated at different ligation temperatures ON. All samples are identical. **L**: ladder, **1**: 46.8˚ sample, **2**: 49.9˚ sample, **3**: 54.3˚ sample, **4**: 60.3˚ sample, **5**: 65.0˚ sample, **6**: 68.1˚ sample, **7**: Sample ligated using the protocol including three cycles (table 5), **8**: Negative PCR control (water added in the PCR mix instead of exonuclease treated ligation product), **9-12**. Water loaded. Expected lengths of specific fragments; 152 bp, 160 bp, 175 bp, 179 bp and 200 bp.

b. *Agilent 2100 bioanalyzer* electropherogram showing gel electrophoresis elution peaks. The peaks at 58-60 s correspond to the specific products of selectors 80 (~137 bp) and 190 (~144 bp) and the peaks at 60-62 s correspond to the specific products of selectors 155 (~160 bp) and 236 (~166 bp). **Red** = 68.1˚C ON**, blue** = 46.8˚C ON**, green** = 3 cycles. Expected lengths of specific fragments; 152 bp, 160 bp, 175 bp, 179 bp and 200 bp.

Table 13. Melting temperatures of selectors used in a five-plex assay.
The melting temperatures have been calculated by using the Nearest Neighbour model as described by Breslauer et al[9].

Selector	T_m 5'(ºC)	T_m 3'(ºC)
80	55.44	57.34
190	59.41	66.34
155	61.43	64.89
236	68.18	64.97
316	68.18	68.99

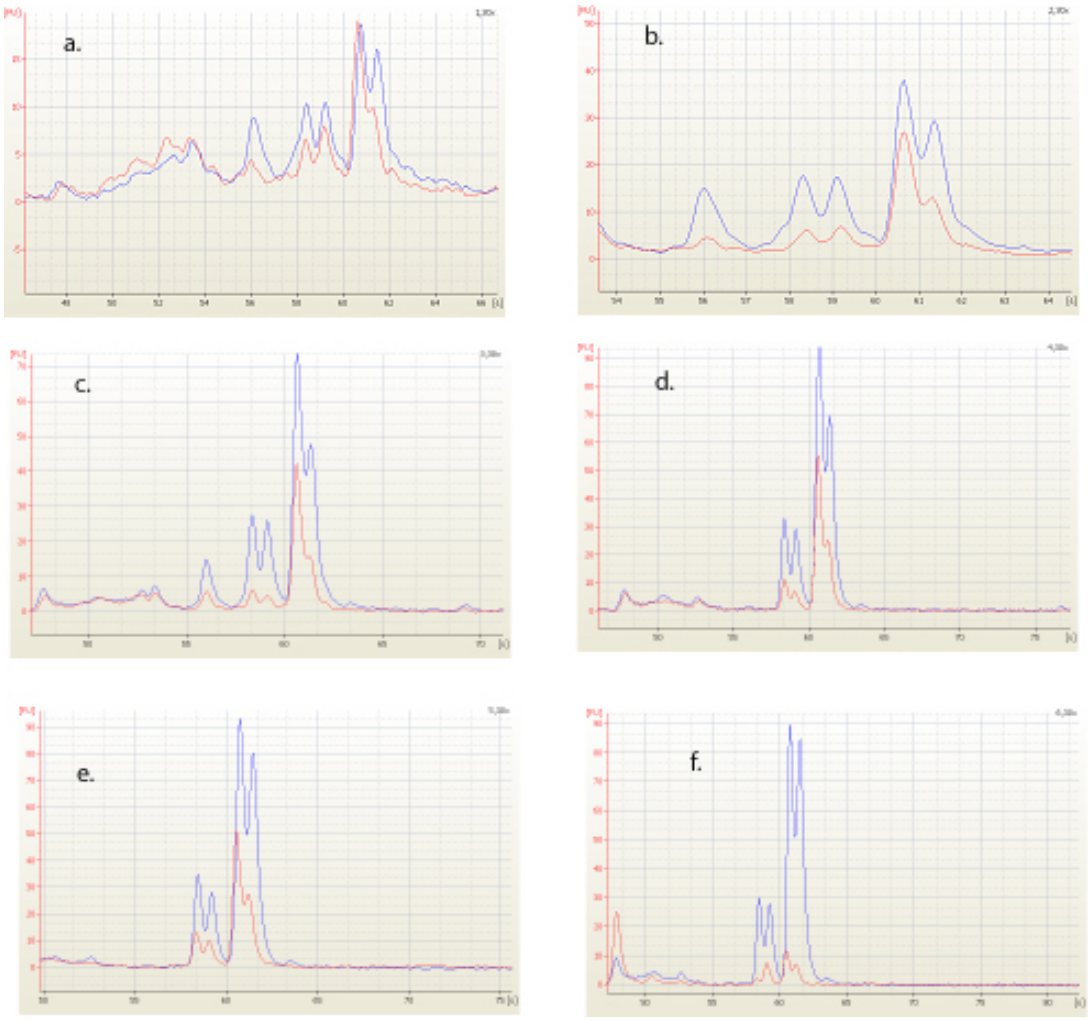

Figure 18. *Agilent 2100 bioanalyzer* electropherograms showing gel electrophoresis elution peaks. Temperature 1 is the lowest and temperature 6 is the highest. **a.** Temperature 1. **b.** Temperature 2. **c.** Temperature 3. **d.** Temperature 4. **e.** Temperature 5. **f.** Temperature 6. **Red** peaks correspond to samples incubated using the protocol including 30 cycles (table 9). **Blue** peaks correspond to samples incubated at constant temperature ON (table 10). The peaks at 58-60 s correspond to the specific products of selectors 80 (~137 bp) and 190 (~144 bp) and the peaks at 60-63 s correspond to the specific products of selectors 155 (~160 bp) and 236 (~166 bp). Expected lengths of specific fragments; 152 bp, 160 bp, 175 bp, 179 bp and 200 bp.

3.7.2 Additives in the ligation mix

DMSO and formamide, two PCR additives, were included in two separate ligation mixes in an attempt to decrease the formation of secondary structures in the target DNA fragments. According to Frackman et al[16], DMSO disrupts base paring, thereby facilitating melting of duplex DNA. Formamide has been reported to increase the specificity of PCRs by decreasing the melting temperature of DNA[17]. There were no clear differences seen between the samples that included any of the additives and a control sample. DMSO at 5% and formamide at 2.5 % did therefore not seem to affect the performance of the ligation reaction. To draw any firm conclusions concerning the effect of these additives, the experiment would however need to be repeated, trying different concentrations of DMSO and formamide.

3.7.3 Ligation using 9°N™ DNA ligase

Ligation was performed using the 9°N™ DNA ligase instead of the ampligase enzyme used in routine experiments. Though both enzymes are thermostable, the 9°N ligase is active at even higher temperatures than ampligase and can survive PCR conditions. The source of the enzyme is the hyperthermophilic archaeon *Thermococcus*, which was discovered at a latitude of 9 °N and consequently named strain 9°N (New England Biolabs). There were no significant differences seen in the amounts of specific products produced using the two different enzymes. The enzymes therefore seem to be as effective at ligating when performing ON ligation at 60°C. The standard ampligase enzyme was accordingly used in following experiments.

3.8 Modification of 5'- and 3'-ends of selector probes

Addition of three adenosine phosphates to the 5'-ends and three thymidine phosphates to the 3'-ends of two selectors was made to prevent the build-up of selector-dependent unspecific products in the PCR. The three thymidine phosphates should in theory be able to prevent the 3'-ends of two selector probes from annealing to each other and giving rise to the 90 bp artifact band seen after PCR (fig. 4). Modification of the primers (2.10) was also made to prevent the build-up of artifacts.

3.8.1 Singleplex screening assay

The modified- as well as the unmodified selectors were tested together with on the one hand the unmodified- and on the other hand the modified primer pair. This was first done in singleplex reactions and the results were analyzed on the *Agilent 2100 bioanalyzer* instrument. Comparisons of electropherograms showed that the modification did indeed increase the production of specific products (fig. 19). The largest positive effect was seen when both the selector probe and the primer were modified. Furthermore, some long unspecific products that were produced from selector 155 were only seen when the unmodified primer pair was used. These artifacts are formed from specific products and unmodified primers as described (3.4). The primers carrying three adenosine phosphates in their 5'-ends will however prime production of fragments having three thymidine phosphates in their 3'-ends. This three-nucleotide tail will prevent the fragments from annealing to each other in subsequent PCR cycles.

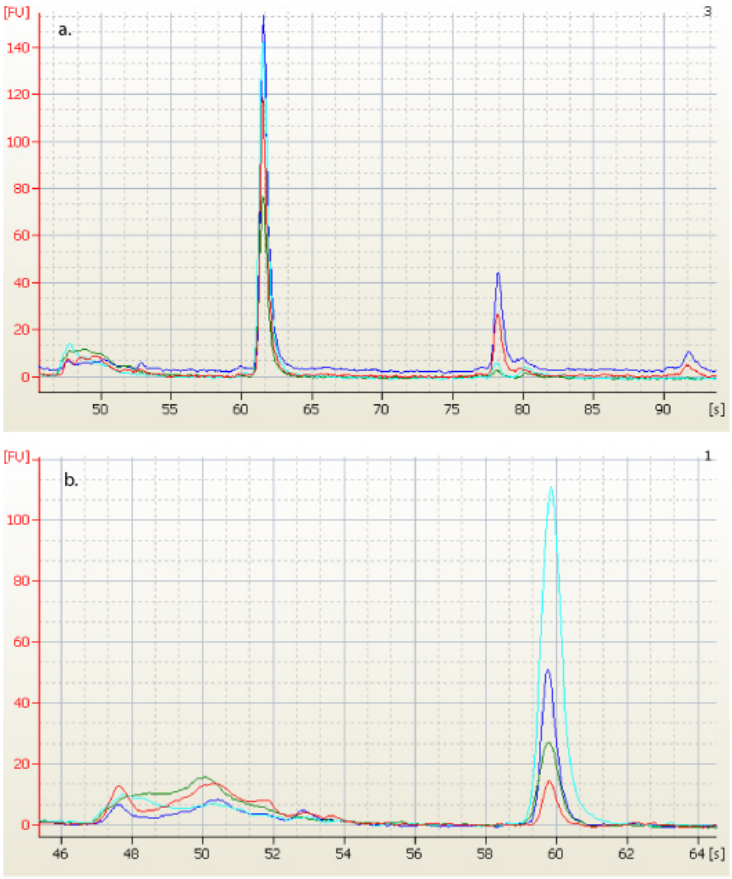

Figure 19. *Agilent 2100 bioanalyzer* electropherograms showing gel electrophoresis elution peaks.
a. The peak at 62 s is the specific peak of selector 155. (Expected length of the specific fragment: 160 bp.)
Red = unmodified selector 155 and unmodified primer, **green** = unmodified selector 155, modified primer,
blue = modified selector 155, unmodified primer, **turquoise** = modified selector 155, modified primer.
b. The peak at 60 s is the specific peak of selector 190. (Expected length of the specific fragment: 160 bp.)
Red = Unmodified selector 190 and unmodified primer, **green** = unmodified selector 190 and modified primer,
blue = modified selector 190 and unmodified primer, **turquoise** = modified selector 190 and modified primer.
In both images, the peaks at 47-54 s correspond to unspecific products.

The results concerning reduction of unspecific products were hard to interpret, but it seems like the largest effect was obtained by modifying the primers, at least for selector 155. The artifacts from selector 190 appeared largely refractory to modification of primer or selector probe (fig. 20). The results from this first pilot assay indicate that 5'- and 3'- addition of a few nucleotides to selectors and primers can be used as a means to enhance the performance of poor selectors. If further studies show this approach to be feasible, modifications of this kind could potentially be used as a routine procedure in selector design.

3.8.2 Multiplex assay

To examine the effect of modification of a larger number of selectors in a higher-plex assay, the selector probes from a 32-plex mix used previously at the lab were redesigned having the above described 5'- and 3'- end modifications. These selector probes were tested at eight different genomic DNA concentrations. The first time the experiment was conducted, the unmodified selectors performed overall better than the modified counterparts (fig. 21a). The highest genomic DNA concentration was an exception; here the modified selectors produced peaks that summed up to a larger peak area. The sum of the artifact peaks was higher for the modified selectors in all cases but at the highest DNA concentration (figure 21b). Thus, there

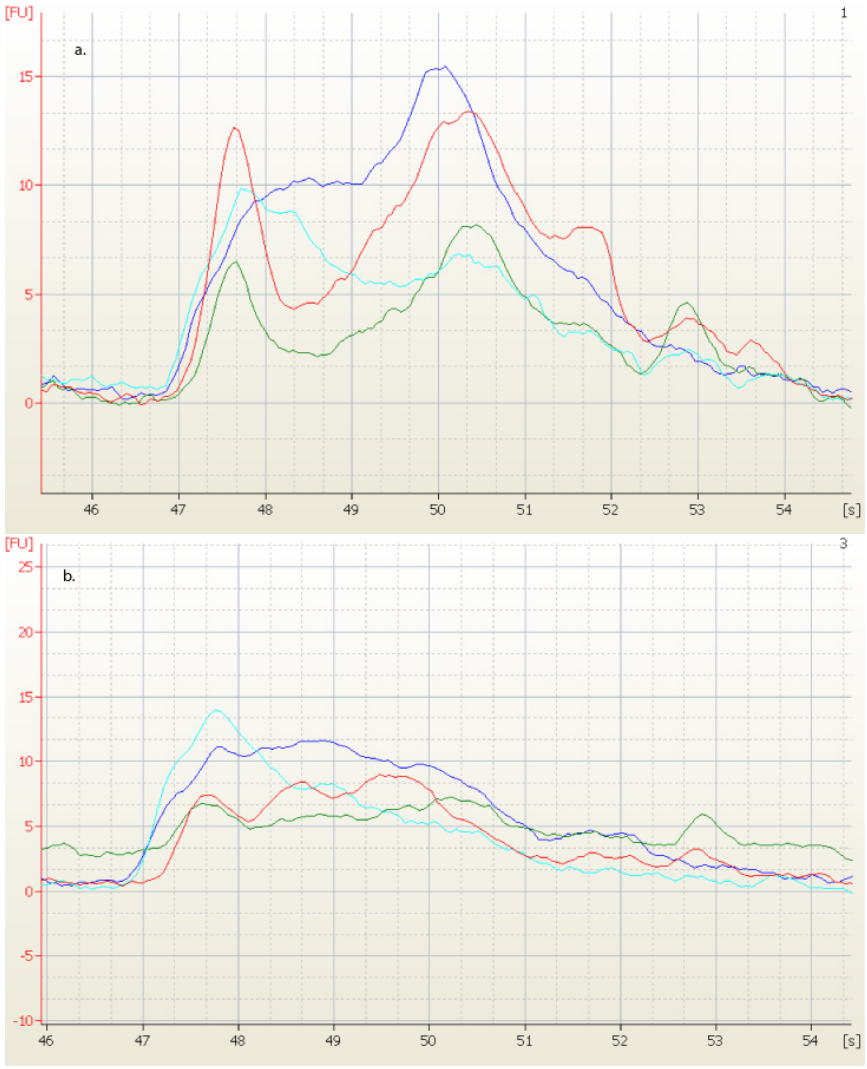

Figure 20. *Agilent 2100 bioanalyzer* electropherograms showing gel electrophoresis elution peaks.
a. Artifacts of selector 155. **Red** = unmodified selector 155 and unmodified primer, **green** = unmodified selector 155, modified primer, **blue** = modified selector 155, unmodified primer, **turquoise** = modified selector 155, modified primer.
b. Artifacts of selector 190. **Red** = Unmodified selector 190 and unmodified primer, **green** = unmodified selector 190 and modified primer, **blue** = modified selector 190 and unmodified primer, **turquoise** = modified selector 190 and modified primer.

was a correlation between the formation of specific products and artifacts, irrespective of genomic DNA concentration or type of selector probe used.

The results from the second experiment performed using these sets of modified- and unmodified selector probes confirmed what had been indicated in the first experiment. The modified probes performed poorly compared to the original selector set. The effect of increasing the number of cycles in the PCR program was also studied. It was expected that altering the number of cycles would influence the proportion of artifacts to specific products. As can be seen in figure 22, there seem however not to be any significant difference between

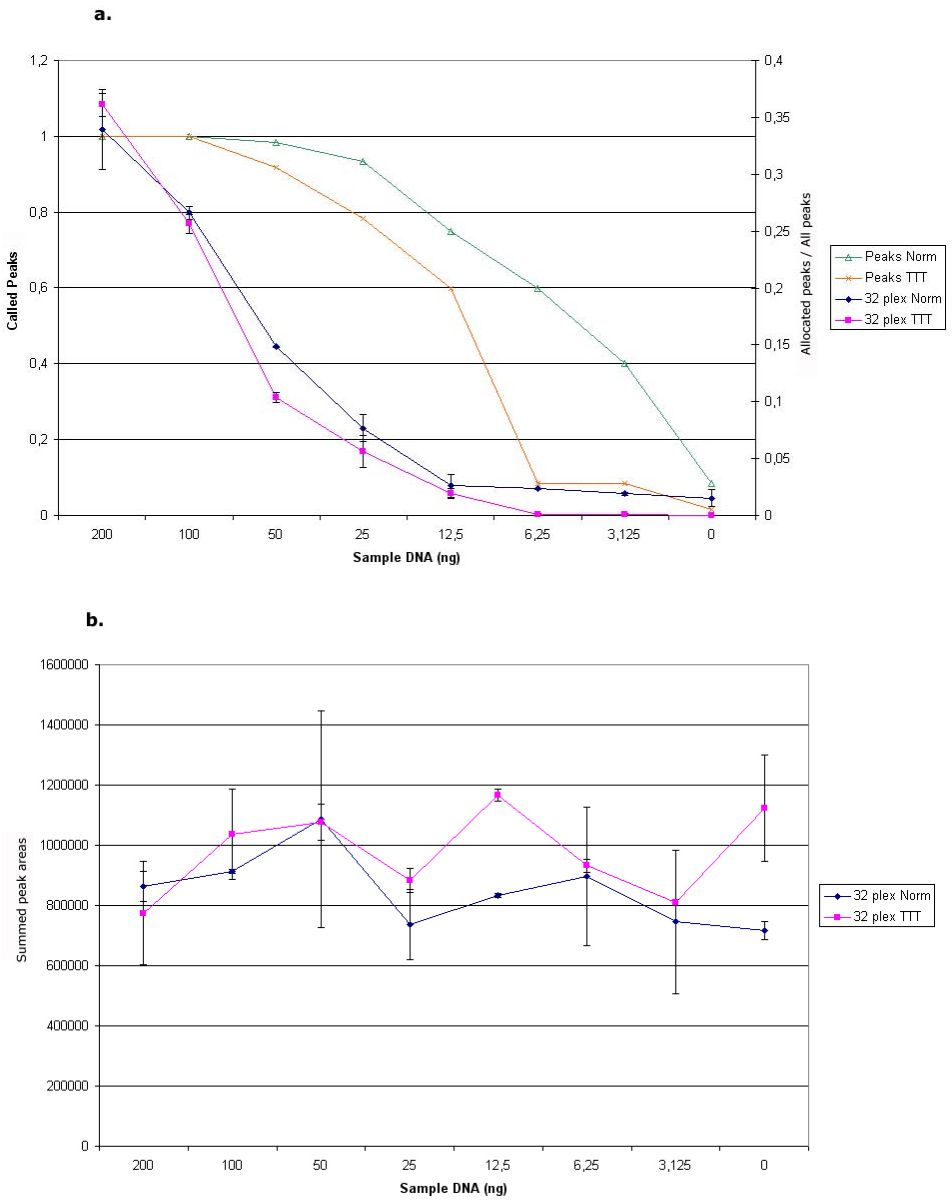

Figure 21a. Total peak areas for modified- and unmodified probes.
Blue and **pink** lines: ratio between summed peak areas of specific products and summed areas of all peaks as a function of genomic DNA concentration (right y axis). **Yellow** and **green** lines: the proportion of peaks called as a function of genomic DNA concentration (left y axis). **b.** Total peak areas for artifacts of modified- and unmodified probes as a function of genomic DNA concentration.

the samples amplified during 30 cycles *versus* the 40-cycle samples. Exonuclease I treatment, on the other hand, seem to be of importance. The non-Exonuclease I treated samples have a much lower ratio of specific- to total product. It is interesting to see that Exonuclease I treatment affects the production of artifacts in this 32-plex assay, since treatment with this enzyme seemed to have no effect in singleplex (3.5). This result is explained by the fact that there is a larger number of different probes in a multiplexed- than in a singleplexed reaction. The probability that probes, vectors and genomic fragments will find partially complementary sequences increases as a function of the number of selector probes.

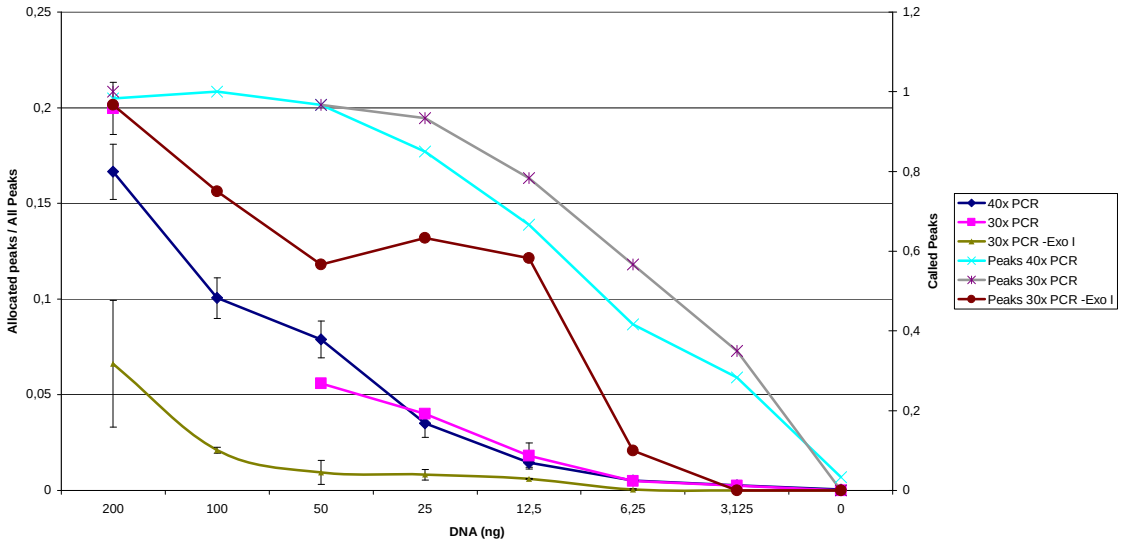

Figure 22. Total peak areas for unmodified probes as a function of genomic DNA concentration. **Blue**, **pink** and **green** lines: ratio between summed peak areas of specific products and summed areas of all peaks as a function of genomic DNA concentration (left y axis). The large min-max number of the first value of the "30x PCR (−) Exo I" sample is due to one of the duplicates having a peak area of zero. The absence of the second value of the "30x PCR" sample is due to both duplicates having zero peak area. **Turquoise**, **grey** and **brown** lines: the proportion of peaks called as a function of genomic DNA concentration (right y axis).

3.8.3 Study of genomic DNA concentration

By performing the ligation reaction using eight different concentrations of genomic DNA, the lower limit of detection of the specific products was determined using the set of 32 modified- and 32 unmodified selectors. As is seen in figure 21a, the number of called peaks drops the fastest for the modified selectors, which implies that the limit of detection is lower for the unmodified selectors in this assay. The unmodified selectors were able to produce over 90% of the 32 specific products when 200-25 ng genomic DNA was added to the ligation mix and 75% of the products at 12.5 ng DNA.

Since decreasing the genomic DNA concentration should result in more selector probes and vectors being freed from their genomic target molecules, increased amounts of artifacts would be expected when this is attempted. Such a relationship was however not observed (fig 21b). This contrasts to the results of a singleplex assay described previously (3.2.1). Single-plex assays seem to be more sensitive to DNA concentration change; in the case described above a decrease in the DNA concentration to a fourth resulted in an increase in the production of unspecific products at the expense of specific products. Concerning multiplex assays, it seems like the amounts of unspecific products are fairly constant as DNA concentration is changed.

3.9 Uracil containing selector probes

UNG enzymes from two different manufacturers were applied to degrade uracil-modified probes. Treatment of a 16-plex selector mix with UNG from Fermentas did not result in any improvement in the proportions of specific and unspecific products as compared to samples treated with Exonuclease I (fig. 23). A possible explanation for the observed inefficient degradation of uracil-modified selector probes is the low enzyme concentration used.

The assay performed with Uracil-DNA Excision Mix (Epicentre) failed of unknown reasons. No specific peaks were observed, neither in samples treated with enzyme nor in untreated controls.

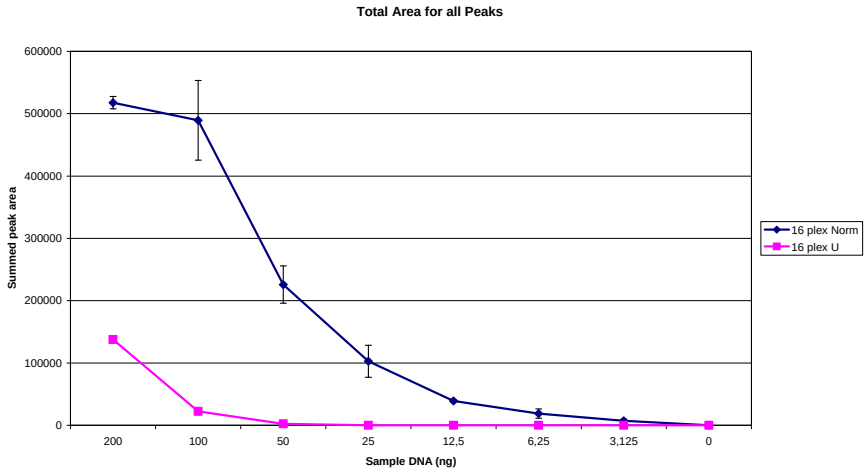

Figure 23. Total peak areas for uridine phosphate-modified selector probes and for unmodified probes as a function of genomic DNA concentration. UNG treatment does not seem to be a good alternative to Exonuclease I treatment.

3.10 RCA

3.10.1 RCA with and without primer

The effect of including step of RCA after the ligation reaction was evaluated using the five-plex selector mix in table 8. RCA by Φ29 polymerase produces long concatemers of DNA containing the specific product sequence and sites for annealing of primers. The presence of large amounts of specific template sequences should compete out the unspecific templates in the PCR.

The ligation reaction was performed ON at 60°C. In the first experiment that was conducted, one ligation product was treated with Exonuclease I, while two identical ligation products were rolling circle amplified. Primer (P2938) was added to one of the samples before RCA. In theory, a primer is not needed for the Φ29 polymerase to start amplifying the formed vector-target circles, since the selector probe can provide a free 3'-end for the polymerase. This was experimentally verified, since there were no large differences in the positions or the heights of the peaks of the two samples in the *Agilent 2100 bioanalyzer* electropherograms (fig. 24). The deviations seen are probably due to pipetting errors and the bad resolution of the analysis instrument.

The differences observed between the Exonuclease I treated sample and the rolling circle amplified samples are on the other hand more striking (fig. 24). First, products of lengths corresponding to doubles and triples of the specific products are present in the RCA samples. These are formed when the primers in the PCR anneal at multiple sites in the concatemer produced by RCA. The presence of these products does not pose a problem for sequencing purposes. Second, the baseline of the RCA products is higher than that of the Exonuclease I products. This could be due to the Φ29 polymerase producing branched structures from the concatemer DNA.

Third, the peaks corresponding to the products of the RCA are broader than are the peaks of the Exonuclease I treated samples. This could be due to the presence of free selector probes and vectors that anneal to the produced concatemer and generate unspecific products of various lengths. The concentration of selector probes and vectors is expected to be higher in the RCA samples than in the samples where Exonuclease I has digested these single stranded substrates. The peak broadening could also be due to impaired migration in the capillaries of the *Agilent 2100 bioanalyzer* instrument, caused by the long threads of DNA that are formed in the RCA.

Finally, the peak heights of the RCA products are more even than those of the Exonuclease I treated products. The first two peaks of the RCA samples have an increased height while the two later peaks have decreased in height. As one of the aims of this project was to make the production of specific amplification products of the different selectors more uniform, performing RCA after ligation needed further investigation.

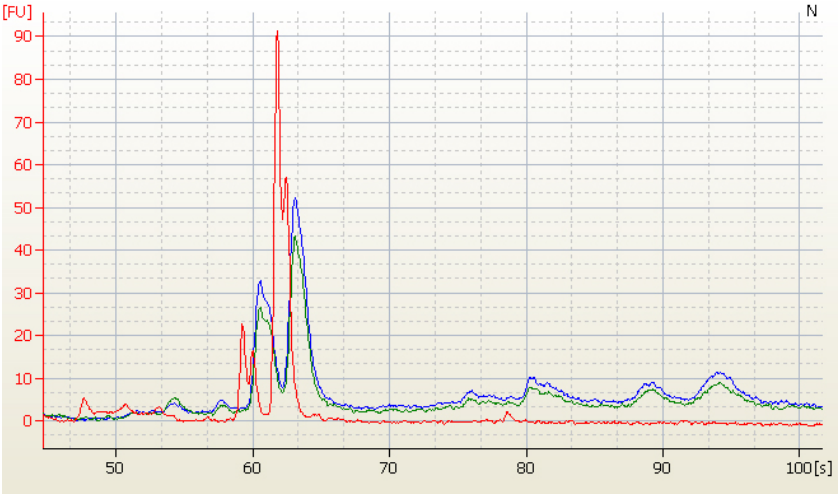

Figure 24. *Agilent 2100 bioanalyzer* electropherogram showing gel electrophoresis elution peaks. **Red** = Exonuclease I treated sample, **blue** = RCA sample without primer, **green** = RCA sample with primer. The red peaks at 58-60 s correspond to the specific products of selectors 80 (~137 bp) and 190 (~144 bp) and the red peaks at 61-64 s correspond to the specific products of selectors 155 (~160 bp) and 236 (~166 bp). Expected lengths of specific fragments: 152 bp, 160 bp, 175 bp, 179 bp and 200 bp.

3.10.2 RCA and Exonuclease I treatment

In the next experiment performed, treatment with Exonuclease I and RCA were combined. The exonuclease treatment should remove free selector probes and unligated vectors before the RCA, giving fewer opportunities for annealing of these to the concatemer produced by Φ29, thus giving less unspecific products in the PCR. This was believed to potentially reduce the width of the peaks in the *bioanalyzer* electropherogram. The Exonuclease I treatment had to be performed before the amplification step, because the long single-stranded concatemer produced in the RCA is a substrate for the Exonuclease I enzyme.

The results from the experiment can be seen in the *Agilent 2100 bioanalyzer* electropherograms in figure 25. There were no major differences concerning specific product generation between the untreated sample- and the Exonuclease I treated samples. Neither was there any evident effect of Exonuclease I treatment on the sample that had been amplified by RCA. It seemed however like Exonuclease I might be able to decrease the peak heights of the unspecific products. The baseline of the Exonuclease I treated sample was somewhat lower than that of the sample that was only rolling circle amplified. This could be due to Exonuclease I having reduced the amount of background DNA molecules in that sample. In addition, Exonuclease I treatment seemed to have had the effect of sharpening the electropherogram peaks a bit. This can be explained by the ability of the enzyme to digest unligated vectors and free selector probes.

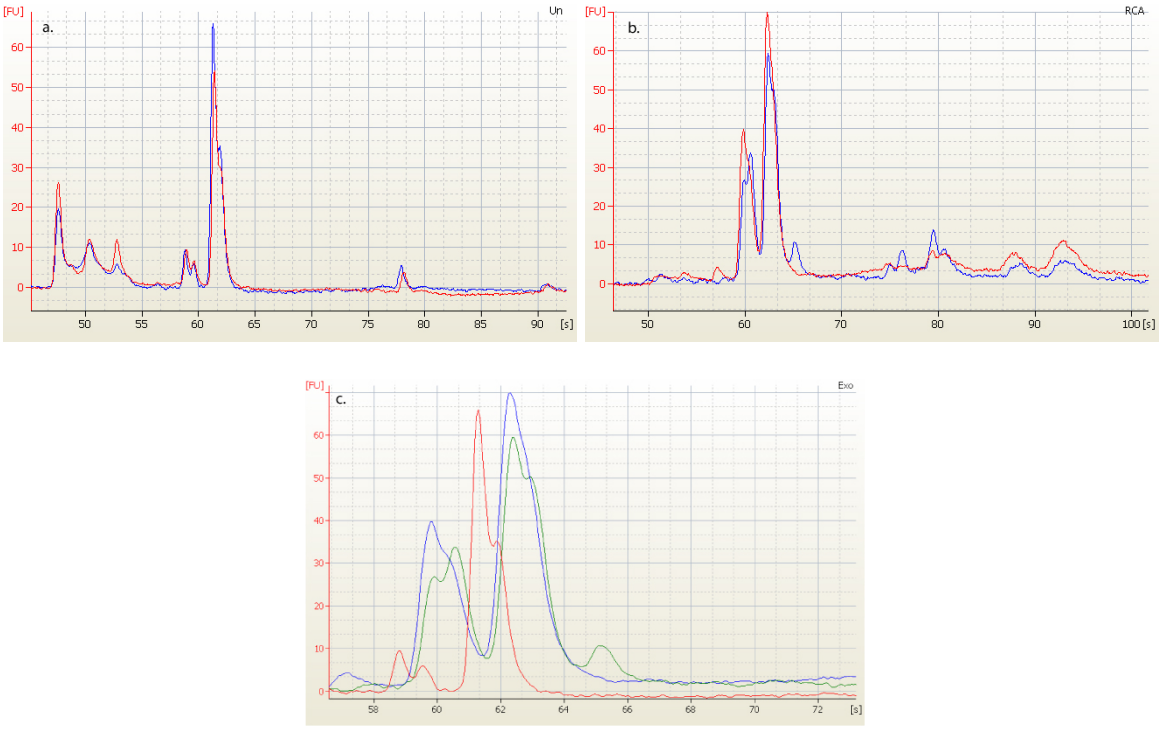

Figure 25. *Agilent 2100 bioanalyzer* electropherograms showing gel electrophoresis elution peaks. **a. Red** = untreated sample, **blue** = Exo I treated sample. **b. Red** = RCA sample, **blue** = Exo I + RCA treated sample. **c. Red** = Exo I treated sample, **blue** = RCA sample, **green** = Exo I + RCA treated sample. Expected lengths of specific fragments: 152 bp, 160 bp, 175 bp, 179 bp and 200 bp.

3.10.3 PCR cycle number

As described (3.10.1), the base-line of the RCA samples was high due to the production of branched structures from the concatemer DNA in the PCR. The standard PCR program used includes 33 rounds of thermal cycling. Reducing the number of cycles is supposed to decrease the background signal, so PCR programs including ten and 20 cycles were tried. The reaction performed using ten cycles failed of unknown reasons. When the samples were cycled for 20 rounds, both the amounts of specific products and the background signal decreased (fig. 26). The protocol including 20 cycles was found to be preferred, since the lower background in combination with reasonable amounts of specific products makes a good template for PCR. Therefore, it was used in subsequent experiments.

3.10.4 RCA in 32-plex

Since the results from the experiment just described were promising but needed further validation, the same experimental set-up was applied again, now using a larger selector set including 32 selectors. As is seen in figure B, the peak heights of the rolling circle amplified samples were very uneven. The reason for this bias in amplification success is the large length difference between the 32 selected fragments (102-358 bp). When performing one hour of RCA, more copies of the smaller circles have time to form compared to the larger ones, so there are more templates of the shorter fragments in the PCR. The Exonuclease I treated samples, on the other hand, showed a less biased distribution of peak heights, even though the shorter fragments had been somewhat more successfully amplified in this case too.

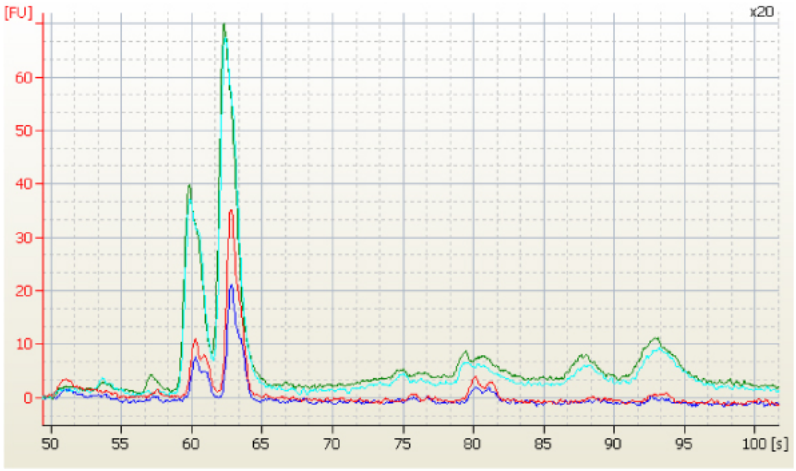

Figure 26. *Agilent 2100 bioanalyzer* electropherograms showing gel electrophoresis elution peaks. **Blue**, **red** = 20 cycles, **green**, **turquoise** = 33 cycles. All samples were prepared in duplicates. Peaks at 60-62 s correspond to the specific products of selectors 80 and 190 and the red peaks at 62-64 s correspond to the specific products of selectors 155 and 236. Expected lengths of specific fragments: 152 bp, 160 bp, 175 bp, 179 bp and 200 bp.

It thus seems like Exonuclease I treatment is more advantageous than RCA in multiplexed assays where the lengths of the selected fragments differ much. In cases where the fragments are designed to be of uniform size, RCA could still be a means to increase the production of unspecific products and to make the amplification success more even. In general cases, though, one should be aware of this risk of increased non-uniformity when using RCA.

In the five-plex assay described above, the peak heights of the two shorter products increased and the peak heights of the two longer products decreased relative to the Exonuclease I treated sample. This could also, at least in part, be due to the effect just described.

3.11 Grouping of selectors that perform similarly

When 32-plex assays were performed (3.8.2 and 3.10.4), it was observed that selectors targeting short genomic fragments performed better than selectors targeting longer fragments. This was suspected to be due to the shorter fragments containing less secondary structures and being more efficiently amplified during PCR. The performance of the selectors having long targets was studied in both a 32-plex selector mix and in a five-plex mix containing only selectors targeting long fragments (table 13). The same analysis was performed using selectors having short and medium-sized targets. This way of attempting to increase the product formation from poorly-performing selectors has recently been suggested by others[3].

By visual inspection of the capillary electrophoresis electropherograms, it was observed that in general, more specific products were produced in the three five-plexes than in the 32-plex. A quantitative comparison could however not be performed, since the ratio of selector probe to vector was not the same in the five-plexes as in the 32-plex. There were also differences in the amounts of each individual probe that had been added in the two cases. Nevertheless, it seems like dividing selectors into groups and performing ligation reactions in separate mixes can be used as a means to increase the performance of poorly working selectors.

Table 14. Selectors used in five-plex assays and their corresponding fragment lengths in bp.
The three ligation mixes contained 0.1 nM of each selector.

5 shortest fragments		*5 midmost fragments*		*5 longest fragments*	
Selector	**Fragment size**	**Selector**	**Fragment size**	**Selector**	**Fragment size**
P2320	102	P2333	184	P2350	319
P2321	107	P2334	190	P2351	324
P2322	112	P2335	196	P2352	334
P2323	119	P2338	213	P2353	339
P2324	125	P2339	223	P2354	358

4. Conclusions

4.1 Enzymatic treatments to reduce selector probe- and vector artifacts

Different kinds of exonucleases were evaluated for their ability to reduce the production of selector probe- and vector artifacts. These tests were performed in singleplex reactions and failed to find any single enzyme or any combination of enzymes that exceeded the performance of the standard Exonuclease I treatment. There were only minor differences seen between untreated samples, Exonuclease I treated samples and samples that had been treated with the various new enzymes. Later results obtained using Exonuclease I indicated that the effect of adding this enzyme was larger in multiplexed assays than in singleplex reactions. This was seen in both the experiment with 32 modified selectors (3.8.2) and in the experiment where RCA was attempted (3.10.2). It is therefore possible that treatment with the new exonucleases would have had larger impact if evaluation of these had been performed in more higher-plexed assays. Testing of these enzymes in samples containing pooled selectors is therefore something that should be attempted in the future.

The results from the 32-plex assay performed using modified and unmodified selector probes show that treatment with Exonuclease I increases the performance of multiplexed assays (fig. 27). This is something that has been proposed earlier (unpublished data), but the results obtained herein are the first ones actually showing this experimentally.

4.2 Ligation protocol

In an attempt to find the most favourable conditions for ligation of the studied selectors, the products generated from ligations using three different ligation temperature protocols were compared as described. Two of the protocols included steps of thermal cycling, while in the third protocol the samples were incubated at constant temperature ON. The reason for using a temperature cycling scheme is that it is very hard to construct a large number of selectors that will all have the same melting temperature. By cycling for three rounds, the arms of the selector probe get more chances of finding and hybridizing to its target sequence than they would get if incubated at constant temperature. This means that a shorter total incubation time will be needed when temperature cycling is applied. An annealing scheme including a constant ligation temperature ON is on the one hand time consuming but ON incubation might perform very well in a multiplex reaction if the right temperature is chosen. Indeed, this was exactly what was seen herein. An additional advantage of constant incubation temperature is that the equipment needed is simpler and less expensive than that needed for thermal cycling.

4.3 The problem of non-uniformity

Increasing the uniformity in the performances of different selectors is an important part of the optimization of the selector technique. The work presented here has to a large extent been focused on reducing the non-uniformity by limiting the production of non-specific products. Reduction of the amounts of artifacts was attempted by means such as Exonuclease I treatment and modification of the ends of the selector probes. Instead of trying to reduce the

amounts of artifacts, RCA was used to out-compete them. These approaches are however only steps on the way towards obtaining a more uniform product distribution; they do not solve the problem. Some selectors do not seem to be able to find their target sequences at all, and these selectors are not helped by enzymatic treatment of the ligation product, by the addition of tails to the selectors or by RCA.

Careful bioinformatic analysis of the target sequences is already applied at the stage of selector probe design, but the methods of analysis might need to be refined. The time invested at evaluating the selector probes might also need to be increased. Statistical analysis of features such as GC content, melting temperature, ability to form internal secondary structures and the presence of repeats in the target sequences is presently at use. Maybe this step of the selector design needs to take other factors into account too. As stated (3.11), the length of the selected sequence is a parameter that influences the outcome of the selector assay; short sequences being more successfully amplified than long ones. Preferential ligation of particular sequences by the ampligase enzyme could potentially also be a factor explaining why some selectors perform poorly.

Characterization of the large number of probes existing today could be a way of learning more about which qualities that characterize a good or a bad selector. For this to be possible, a large database containing all the selector sequences that are at use or that have previously been used, is needed. Setting up such a database is something researchers at IGP are planning to do.

As stated previously (1.1.4), the problem of uneven representation of PCR products generated from different selectors is believed to be predominantly due to differences in ligation efficiencies between different targets and selectors. The possibility still remains however that the PCR step is responsible for some of the unevenness seen and this question will need to be looked into further. Quantitative Real-Time PCR (Q-RT-PCR) could be used to give a measure of the bias in the PCR, by using dilution series of PCR products from different selectors and comparing their amplification success.

4.5 Modification of 5'- and 3'-ends

Attaching short tails of adenosine- and thymidine phosphates to the 5'- and 3'- ends of selector probes, respectively, seemed to be an efficient way of increasing the amount of specific products produced in singleplex reactions. When the same modifications were applied to the selectors in a 32-plex assay, the positive effect was however not observed. This is due to the complexity of the multiplexed reaction, which contains a large number of selector probes. When the number of probes is high, there are many possibilities of probes finding partially complementary sequences and hybridizing to each other. The addition of tails of adenosine- and thymidine phosphates should theoretically reduce the possibilities of such hybridization events, but as the number of selectors increases, the probability of finding a short complementary sequence also increases. Finally, the favourable interactions between selectors that have pieces of complementarity exceed the unfavourable interactions between their 3'-tails. In a singleplex assay, on the other hand, there is only one kind of selector probe present and annealing of the 3'-ends of two probes can be counteracted by the thymidine phosphate tail.

In figure 4, a model for how the 90 bp artifact is produced is proposed. By showing that the extension of the 3'-ends of selector probes can be efficiently reduced by the addition of a short homonucleotide tail, the results from the singleplex assay using modified selectors (3.7.1) show for the first time that this model is correct.

Since the addition of three thymidine phosphates was unable to reduce the production of artifacts in multiplex, other solutions to the artifact problem are needed. The most critical part of the process depicted in figure 4 is the initial annealing and extension of the 3'-ends of the

two selectors. One way of preventing the extension in the PCR would be to add a modified nucleotide to the 3'-end. A 2'-O-methyl nucleotide would efficiently block polymerization at the 3'-end, but manufacturing a large number of selector probes carrying this modification would be expensive. A more cost-effective solution would be to modify the probes in the lab, by using dideoxy-nucleotides and an exonuclease-free polymerase. The efficiency of such a reaction would not be 100%[18], but this approach would anyhow be a good alternative to more laborious and costly modifications.

4.6 Genomic DNA concentration study

For some future applications of the selector technique, the amount of added genomic DNA will be of critical importance for the utility of the method. For example, forensic samples or samples of ancient origin contain limited amounts of DNA. In addition, this DNA is often of bad quality, and highly sensitive methods for genomic targeting are consequently required. Determining the lower limit of detection of the selectors used herein was therefore of interest. Using the 32-plex of unmodified selectors, 93 % of the fragments were targeted when 25 ng of DNA was used. The proportion of amplified fragments dropped below 50% at a genomic concentration of about 5 ng DNA (fig. 25). Though obtaining amounts of DNA in the range successfully used in this study (25-200 ng) poses no problems in applications like mammalian exon resequencing, higher sensitivity will be needed for other applications. Further developments of the selector technique should therefore include efforts to increase the sensitivity.

5. Other new methods for sequence capture

As stated earlier, there is currently a need for efficient and high-throughput methods for sequence capture that match the demands of the modern parallel sequencing technologies. To meet this challenge, the selector technique, but also a number of other methods, have been developed and are still under development. One such method uses so called molecular inversion probes to selectively target genomic DNA sequences without fragmenting the DNA. This method has recently been used in a highly multiplexed assay, amplifying about 10 000 exons in a single reaction. Poor uniformity was however a problem in this study too: only 80% of the targets were successfully captured and amplified[3].

Another class of methods for exon amplification is based on sequence capture on high-density microarrays. In one microarray application, 385 000 long oligonucleotides were used to capture about 300 bp long DNA fragments that had been generated by shearing. After elution from the array and amplification by PCR, resequencing was performed by using so called resequencing arrays[19]. In another study, 6726 about 500 bp long fragments that had been generated by sonication were hybridized to a microarray, eluted and amplified by ligation-mediated PCR. Sequencing was performed on a 454 instrument [20]. Both microarray-based methods showed high selectivity in target capture and are possible to scale-up by using microarrays with a higher number of features.

A technique making use of biotinylated bacterial artificial chromosomes has been used to hybridize genomic fragments that after two rounds of selection and PCR were enriched 10 000-fold. Sequencing of these fragments confirmed single nucleotide polymorphisms that were previously known as well as revealed new ones. This method is cost-effective but not so selective[21].

A method called gene-collector resembles the selector technique, in that it makes use of selective ligation and circularization of the target DNA. This method, in which a multiplexed PCR is followed by circularization to the collector probe and degradation of irrelevant PCR products by exonuclease treatment, has been reported to produce products of high uniformity[22].

6. Method developments

6.1 Future technological developments

The selector technique is dependent on restriction enzymes for the generation of suitable genomic DNA fragment for ligation, something that possibly can limit the applicability of the method. The 3'-end of the selectable fragment must be precisely generated by a restriction enzyme, while the 5'-end is adjustable due to the possibility of using the flap cleavage activity of *Taq* polymerase[23]. For the generation of the large number of fragments needed in a high-plex assay, application of a large number of enzymes might be required. It is therefore of interest to find alternative ways of preparing the target fragments and initial attempts using sheared DNA have given positive results (unpublished data). The combination of DNA glycosylases and enzymes that break the DNA backbone at abasic sites has also been shown to generate usable fragments (unpublished data).

Another limitation of the selector technique is the considerable cost of probe production in higher-plexed assays. By using microarray-based probe synthesis, this cost can however be reduced, as recently demonstrated by Porreca et al[3],

As a means to reduce artifact production, the application of a vector oligonucleotide containing an internal hairpin has been proposed (unpublished data). This kind of vector would include a looped out sequence that is not complementary to the selector probe but that contains sites for the annealing of primers. It would not give rise to the kinds of artifacts that are dependent on vector-selector probe duplex formation. In addition, since the selector probe would not contain the primer annealing sites, the artifact described previously (fig. 4) would not be formed. Another way of getting around the problem of artifact build-up would be to perform ligation using a vector-free selector. The hairpin vector and the vector-free selector could be used in homogeneous assays in the way described herein. Alternatively, heterogeneous assays could be developed. Selectors could be placed on solid support, for example on beads, and RCA could be performed after free vectors and probes have been washed away. This approach would be more laborious than the solution-based protocol, but it could potentially increase the ratio of specific- to unspecific products.

6.2 Future applications

The selector technique should now be ready for evaluation in a real resequencing application. It could be used in cancer research to characterize different cell lines by sequencing exons of candidate genes. A large number of selectors would then be designed that each targets a part of an exon of interest and the selected fragments would be sequenced in parallel, by using a second generation sequencing platform. This approach would generate a large amount of valuable sequence information per run.

Another interesting application of the selector technique would be to sequence selected parts of prokaryotic genomes. Differences between pathogenic- and non-pathogenic strains of a species could be analyzed by designing selectors that target a particular bacterial genomic region. Not only base-pair differences, but also gene duplications could be studied, by using the MLGA (Multiplex ligation dependent genome amplification) method developed at IGP[24].

The selector technique could also be a valuable tool in projects aiming at sequencing of ancient DNA. Old biological samples usually contain a mixed population of DNA fragments, originating from a number of different species. Selective targeting of sequences from a specific species therefore poses a large problem for a researcher studying ancient DNA. The selector technique, being both highly selective and sensitive, would suit well for this application. It could for example be used to enrich for cave bear sequences in a sample highly contaminated by microorganisms.

Acknowledgements

My thanks go to Magnus Isaksson for being a great supervisor; helpful, patient and encouraging. I likewise thank Mats Nilsson who has given me much inspiration and cleared out many clouds in my mind. Warm thanks to Ulf Landegren for his encouragement and concern.

I further would like to thank Lena for all the help in the lab and her good advices. I am also grateful for the good advises I have got from Jenny, Henrik and Johan, both in the lab and concerning the report. Special thanks to my scientific reviewer, Siv Andersson at EBC.

Finally, thanks to my lovely family and to Jon for helping me with the report and for always supporting me.

References

1. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *PNAS*. Vol. 74, No. 12: 5463-7.
2. Blow, N. (2007) The personal side of genomics. *Nature*. Vol. 449: 627-630.
3. Porreca, G.J., Zhang, K., Li, J.B., Xie, B., Austin, D., Vassallo, S.L., LeProust, E.M., Peck, B.J., Emig, C.J., Dahl, F. Gao, Y., Church, G.M. & Shendure, J. (2007) Multiplex amplification of large sets of human exons. *Nature Methods*. Vol. 4, No. 11: 931-936.
4. Dahl, F., Stenberg, J., Fredriksson, S., Welch, K., Zhang, M., Nilsson, M., Bicknell, D., Bodmer, W.F., Davis, R.W., and Ji, H. (2007) Multigene amplification and massively parallel sequencing for cancer mutation discovery. *PNAS*. Vol. 104, No. 22.
5. Nelson, D.L., Ledbetter, S.A., Corbo L., Victoria, M.F., Ramírez-Solis, R., Webster, T.D., Ledbetter, D.H. & Caskey, C.T. (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *PNAS*. Vol. 86, No. 17: 6686-90.
6. Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjöld, M., Ponder, B.A.J and Tunnacliffe, A. (1992) Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics*. Vol 13, No. 3: 718-725.
7. Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z. Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M. & Lasken, R.S. (2002) Comprehensive human genome amplification using multiple displacement amplification. *PNAS*. Vol. 99, No. 8: 5261-5266.
8. Stenberg, J., Nilsson, M., and Landegren, U. (2005) ProbeMaker: an extensible framework for design of sets of oligonucleotide probes. *BMC Bioinformatic*s, Vol. 6, No. 229.
9. Breslauer, K.J., Frank, R., Blöcker, H. & Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence. *PNAS*. Vol. 83: 3746-3750.
10. Brownstein, M.J., Carpten, J.D. & Smith, J.R. (1996) Modulation of Non-Templated Nucleotide Addition by *Taq* DNA Polymerase: Primer Modifications that Facilitate Genotyping. *BioTechniques.* Vol. 20, No. 6: 1004-1010.
11. Dahl, F., Banér, J., Gullberg, M., Mendel-Hartvig, M., Landegren, U. & Nilsson, M. (2004) Circle-to-circle amplification for precise and sensitive DNA analysis. *PNAS*. Vol. 101, No. 13: 4548–4553.
12. Vanghan, P., McCarthy, T.V. (1998) A novel process for mutation detection using uracil DNA glycosylase. *Nucleic Acids Res*., 26: 810-815. Schmidt, M., Frey, B., Kaluza, K & Sobek, H. (1996)
13. Schmidt, M., Frey, B., Kaluza, K & Sobek, H. (1996) Application of heat-labile Uracil-DNA Glycosylase in improved carryover prevention technique. *Biochemica*, No. 2: 13-15.
14. Thomas, K.R. & Olivera, B.M. (1978) Processivity of DNA exonucleases. *JBC*, Vol. 253, Issue 2: 424-429.
15. Sayers, J. R. and Eckstein, F. (1991.) A single-strand specific endonuclease activity copurifies with overexpressed T5 D15 exonuclease. *Nucleic Acids Res*. Vol. 19, No. 15: 4127-4132.
16. Frackman, Kobs, Simpson & Storts, (1998) *Promega notes*. Nr. 65: 27.
17. Sarkar, G., Kapelner, S. & Sommer, S.S. (1990) Formamide can dramatically improve the specificity of PCR. *Nucleic. Acids Res*. Vol. 18, No. 24: 7465.

18. Zhou, G.H., Gotou, M., Kajiyama, T. & Kambara, H. (2005) Multiplex SNP typing by bioluminometric assay coupled with terminator incorporation (BATI). *Nucleic Acids Res*. Vol. 33, No. 15.
19. Okou, D.T., Steinberg, K.M, Middle, C., Cutler, D.J., Albert, T.J. & Zwick, M.E. (2007) Microarray-based genomic selection for high-throughput resequencing. *Nature Methods*. Vol. 4, No. 11: 907-909.
20. Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X., Richmond, T.A., Middle, C.M., Rodesch, M.J., Packard, C.J., Weinstock, G.M. & Gibbs, R.A. (2007) Direct selection of human genomic loci by microarray hybridization. Vol. 4, No. 11: 903-905.
21. Bashiardes, S., Veile, R., Helms, C., Mardis, E.R., Bowcock, A. & Lowett, M. (2005) Direct genomic selection. *Nat. Methods*. Vol. 2, No. 1: 63–69.
22. Fredriksson, S., Banér, J, Dahl, F., Chu, A., Ji, H., Welch, K. & Davis, R.W. (2007) Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector. *Nucleic Acids Res*. Vol. 35, No. 7.
23. Lyamichev V., Brow M.A., Dahlberg J.E. (1993) Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. Science. No 260: 778–783.
24. Isaksson, M., Stenberg, J., Dahl, F., Thuresson, A.C., Bondeson, M.L., & Nilsson, M. (2007) MLGA—a rapid and cost-efficient assay for gene copy-number analysis. *Nucleic Acids Res*. Vol. 35, No 17.

Appendix A

Table A1. Selector probe sequences of the long versions of the 24 probes used in Dahl et al[4]. Indicated are the number of counted reads reported for the short counterparts of the selectors (Dahl et al) and the lengths of the specific product of each selector (in bp).

Selector	T_m 5'	T_m 3'	5' to 3' sequence	Counts	Length
98	55.3	64.3	TTGAGAAGTCTTCTGGTTTTTTCCT|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|GAAGATCTGCGCCAGGATGAGCGTG	0	194
201	76.8	67.6	CCCCGGATCGCGCCCCGGACCCCGC|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CTCTTCGGGGAGCAGCGATGCGACC	0	207
316	68.3	68.5	TGCCGAAAGTGCCCTTGCCCAGCAG|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CTGCCACCCCGCACCCTCATCTCCA	0	200
346	56.8	57.6	GTGCATGTTTCAAAGCAAATGTGAA|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CGAGTCCTTCATCTTTTGGAACAGA	0	196
273	46.5	58.1	TAATTAAGATGAAAATAACTTTGTT|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|TTCTGATGGCAGTATTGAACTGCAG	2	147
270	47.4	50.2	TCAATTTGATAATTTATCTCACTTA|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|AGAAGTAGGATTTCTACTTACTTTT	6	209
386	55.7	59.2	TCAGGAAAATGACAATGGGAATGAA|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CTGCAGTAAATGCTGCAGTTCAGAG	6	245
216	68.2	67.5	TCTCGGCTGCTGAGGCTGGGCTCTC|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CGAGCCTGTGGGGCCGACAGCTATG	9	186
28	57	59.9	CCTGAGATCTGTGCCAATTTTTTGT|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|GTGTTCCTGTAGCTCTGATGAGTGC	542	149
413	59.6	60.1	ATTGTCTCGGCTCTTGAGTACTTGC|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|GATACAGGTGTCACCACATCACTGC	590	242
155	61.5	63.7	TTCCTCTGAGTCCCTGGGTGATCAG|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CTGATCCCACACATGCTGCGTGTCT	626	175
227	56.4	70.9	AAAAATAAGGCTCCAGGTTGTTGTT|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|GCCTTGTGCTCCCCCGAGGGCTGCT	688	216
95	54.8	60.8	GTACCGTAAAGAGAGTATACCCTTG|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|GATGCGATGCTCGATGTTGAGAAGG	700	237
348	51	60.4	ATAGACAATTACTGGTGGATTTTTA|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|AGATCTGTCCTGCTGTGTGTGTTCT	724	724
109	60.4	64.5	CCCAGTTTTTGCCTGCCTGTTTTTC|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|TCTGGATGGCTTTCACCCCCTCCAC	777	188
9	60.1	58.3	GCTCTGACTGTACCACCATCCACTA|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|AAAAAAAAAAAAAGGCCTCCCCTGC	887	157
23	75.3	67.7	AGCGCAGCGGACGGCGCCTTCCCGG|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|TGAACGTGCGGTGGGATCGTGCTGG	1373	191
13	64.8	61.5	GTAGGGGGGCTTTCTCCTGCTGCTT|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CGAGTGGAAGGAAATTTGCGTGTGG	1876	227
102	56	60.3	AAGTTATCAAAGTCTCAACCAACCC|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CTGCTTTGAGATTCGTCGGAACACA	2573	222
190	62.2	63.9	AGGCGAGATCCATGTCCTCATGGAG|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CTGGTCCGAATCAGCAGCATGGAGG	3001	160
261	55.6	64.1	ACCCTTAGGTATTCTGCATTTTCAG|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CTAATGCGGGCATGGCTGTTGGGAT	3094	202
80	59.7	59.9	AACACATGACACACTCACCAGTGAG|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CTCGTCACTAACACCACTGGACATG	3879	152
68	53.5	49	GGAATAAAGGTCATGTTGAAAACTC|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|AGATAGTTACTTTAAAAAATTGAGC	4726	166
236	68.5	62.9	GGTGCCAGGACACGGCACTTGCCAG|ACGATAACGGTAGAAAGCTTTGCTAACGGTCGAG|CTGGGGGGGACAAGAACACAGAGAC	7371	179

Table A2. Selector probe sequences of the unmodified probes used in 32-plex reactions. Modified versions of the probes have the same sequences plus a three nt 5'-extension of adenine phosphates and a three nt 3'-extension of thymine phosphates. The 16 probes having odd tag ID numbers were also manufactured in a version containing uracil phosphates instead of thymidine phosphates (but no 5'- or 3'- extensions). Lengths of the of the specific product of each selector is given in bp.

Selector	Sequence (5' to 3')	Length
P2320	GAAATCCTACCCTCCTCTTTACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGTCTGTAAGTCAAACATTAA	102
P2321	CACTTTGACAGCCCAATCAACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGTGACTCTTCGTCCCCGCC	107
P2322	AGCCGAAAAATGGCCATTACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGGCGATCAGAGGCGCAAGA	112
P2323	TTAAACAGGCTGAATACTGGACGATACGGTAAAAGCTTTGCTAACGGTCGAGTGCTATTAATTGTAAGCTGT	119
P2324	GCTGGGCTCTGGGCGCGAACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGGGAATGTGAGGGTCGCTC	125
P2325	GCTGGAACATCCTCCTAAAAACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGCTCCAGAGGCGGTGGCTC	132
P2326	TTCTAGTTTCTCTGCCATCTACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGGATTTCAACTTCGTTTTGAC	136
P2327	AACCAACTCCTCTTCTGCACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGGGGACATGAGGCCATGAC	141
P2328	CAGCGTTTCACCCTGGGCACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGCAACACAGATTGACTGCTC	148
P2329	TAGAGCGCTGAAGCCGGAACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGTAGAACAGAGGCCAGCAA	157
P2330	TCTAAGCAGAAAGGTGGGTACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGCCGCACTCGCTTGTGGTA	164
P2331	CACTTTGACAGCCCAATCAACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGTGACTCTTCGTCCCCGCC	169
P2332	ACCAATCCATTGCCTTTATGACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGAAATCTAGAGGTATTCTTTT	176
P2333	GATTCTCAAACCTCTCCTTAACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGTAACAGAGAGGGTCTGATTT	184
P2334	CTCTTGTACCTCTGCCCTACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGCATGCCCCCTTTTTATATAA	190
P2335	GGGTTTTCCCCTCATTCTTACGATAACGGAGAAAGCTTTGCTAACGGTCGAGTGCTGTTGAGGTACATACAG	196
P2338	TTGGCACAGGGAAGGAAGACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGCAGGTATACCTGTTGTGAAT	213
P2339	AGCCATGTCAAATAATGCTGACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGAGACTAAGAGGCTACTAAGG	223
P2340	CTTATGCCCCCCAACCTGACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGAAAAATGTGGGTCTTGAAGG	229
P2341	ACATATTCTTCCTCATGTTGACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGACAAAGGGAGGTGATCTAAG	236
P2342	TCATCCATCACCTCATCTTAACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGGTTTGGCTAGGATAGTAAAG	242
P2343	TCATGGTGATGGTGAAGAAAACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGAGTTGAAGAGGTTTGGGC	252
P2344	GGCCATACCAGGAAGTAGAACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGTGAAGGTGAGGTCCATAAAT	257
P2345	CCAGTGCTCCTCACTGGCACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGACATTCTGCCCAAGTAACC	264
P2347	CTGGCTATCCTCCATGGGACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGTAGTGGTCTGAGATTTGGC	283
P2348	ACTTACAGCCCTCACTTTCACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGAGGCGAAGATGCTGCCGA	290
P2349	AAGTAAGTAAGTTTCATTCTACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGGTTTTGGGGTGTTCCTGAT	300
P2350	AAATCCAAAGATCTTTCAATACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGATACAACCAGCTGACAGC	319
P2351	CCGTCGGTGGCTGCAGTGACGATAACGGTAGAAAGTTTGCTAACGGTCGAGTTGCAGGGAGGAGAAACA	324
P2352	TTCTACTTGTTTTAGCAATCACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGAGAATATGATCGATTTGGGT	334
P2353	TTACTATCAGCCTCACATTTACGATAACGGTAAAAGCTTTGCTAACGGTCGAGACCCCAGGTTGCTTACAT	339
P2354	CCTGGGGAATTCAGGGGCACGATAACGGTAGAAAGCTTTGCTAACGGTCGAGGGTATTTTCAAAGCCACTTG	358

Appendix B

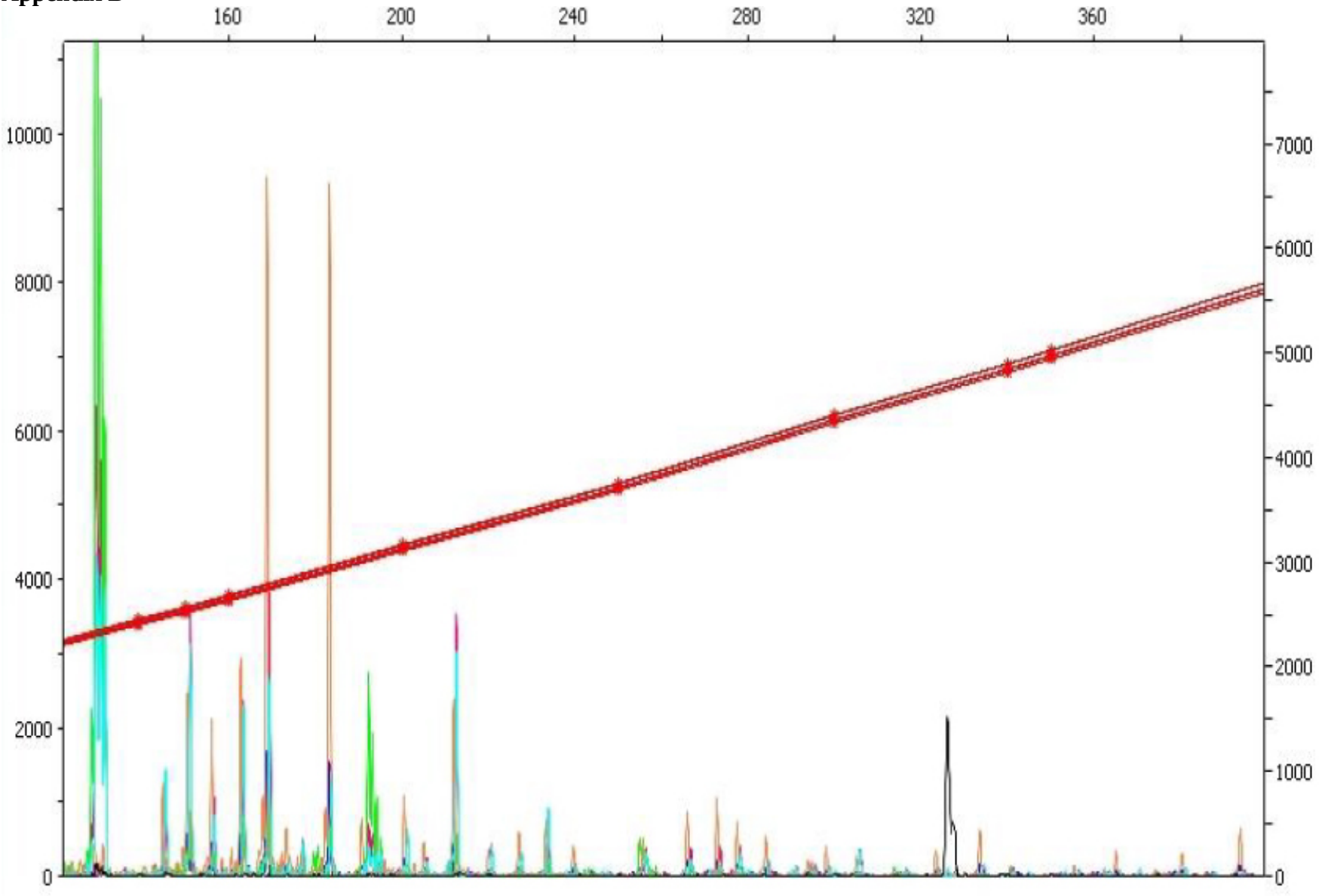

Figure B. Capillary electrophoresis electropherograms showing elution peaks. **Blue** = untreated sample, **orange** = Exo I treated sample, **turquoise**, **pink** = RCA duplicate samples, **yellow**, **green** = RCA + Exo I treated duplicate samples, **black** = negative PCR control.