UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

School of Biological Sciences

An Investigation into Novel DNA Manipulation Strategies for Forensic Applications

by

Adam S. Long BSc(hons)

Thesis for the degree of Master of Philosophy

August 2005

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

SCHOOL OF BIOLOGICAL SCIENCES

Master of Philosophy

An investigation into novel DNA manipulation strategies for forensic applications

by Adam S. Long

Over recent years there has been considerable interest in the development of faster more robust and sensitive ways of analysing DNA. In partial response to this, the Forensic Science Service® has developed a novel amplification strategy based on single nucleotide polymorphisms. In a continuing effort to speed up the analysis time, the concept of miniaturisation is beginning to receive more attention. Miniaturising the various elements required to complete a DNA analysis is going to require a considerable research effort, specifically when focusing on the areas of DNA extraction and integration as well developing new amplification strategies and their associated detection.

This study specifically looked at three areas; microfabrication for DNA extraction, duplex stability assessment with a view to developing more efficient amplification strategies and microarray technology as a replacement for conventional DNA analysis techniques.

Microfabricated silicon channels were designed and used in conjunction with a modified Qiagen™ chemistry. Through optimisation, these channels demonstrated themselves as a robust and efficient platform for DNA extraction. DNA pre-concentration and DNA purification were also demonstrated. In addition, the channels could be used as a means of quantification, by binding a fixed amount of DNA. Their simple design facilitated controllable liquid flow as well as minimising the potential for device blockage. These results demonstrated a proof of principle with regards to efficient miniaturised DNA extraction, providing a feasible first step to a fully integrated DNA analysis system.

Duplex stability is critical to the development of any novel amplification strategy. Future developments require an accurate and efficient assessment of oligonucleotide melting temperature particularly for efficient multiplex development. This investigation determined a new algorithm which calculated duplex melting temperature based on nearest neighbour thermodynamics. This algorithm was incorporated into a Visual Basic™ programme called MOSAIC. Through a series of validation experiments MOSAIC was found to be more accurate at predicting *Tm* in lower salt environments than the commercially available software Primer Express™. It is expected that the improved algorithm can be used for future primer designs, for any new developing amplification strategy.

The use of Microarray technology as an alternative detection strategy for SNP analysis was also investigated. Multiplexes consisting of up to 5 loci were detected by two different methods; direct hybridisation to the SNP site and indirect hybridisation to a universal tail. Both strategies showed equivalent levels of discrimination, reporting correct genotypes in each case. Accurate genotyping from a 10-plex PCR using the indirect approach was also demonstrated.

Contents

Summary ..... **i**

Contents ..... **ii**

Index of Figures and Tables ..... **v**

Declaration of Authorship ..... **xiii**

Acknowledgements ..... **xiv**

Abbreviations ..... **xv**

1 Introduction 1

1.1 DNA in the Forensic Science Service® 1
1.1.1 DNA analysis for forensic identification – A Brief History 1
1.1.2 DNA Profiling – A future perspective 3
1.1.3 Single Nucleotide Polymorphisms (SNPs) 4
1.1.4 Amplification Refractory Mutation System (ARMS) 5
1.1.5 Universal Reporter Primer Principle (URP) 6

1.2 Miniaturisation with a view to improved DNA analysis 9

1.3 Microfabrication 11
1.3.1 The early years 11
1.3.2 Microfabrication specifically for DNA applications 12

1.4 Duplex Stability and multiplex design 28
1.4.1 Background 28
1.4.2 Nearest Neighbour Thermodynamics Model 30
1.4.3 The application of the Nearest Neighbour Model to predicting Tm for mismatched sequences 41
1.4.4 Melting temperature for Computer simulated duplex stability 47

1.5 Microarrays for DNA analysis 52
1.5.1 Background 52
1.5.2 Early arrays 52
1.5.3 Non-porous solid support DNA arrays 54
1.5.4 Array fabrication 55
1.5.5 Attachment Chemistry 59
1.5.6 Target density and Hybridisation efficiency 61
1.5.7 Microarray Analysis 65
1.5.8 DNA Microarray applications 66
1.5.9 URP principle for Microarray SNP detection 72

2 Aims 76

3 Materials and Methods 77

3.1 Microfabrication Assays 77
3.1.1 Microfabricated Silicon Chips 77
3.1.2 DNA preparation 78
3.1.3 DNA Extraction employing the Qiagen™QiaAmp Mini DNA Extraction kit 79
3.1.4 DNA Extraction employing the Silicon Channel 79
3.1.5 DNA Amplification 80
3.1.6 Polyacrylamide Gel Electrophoresis 81
3.1.7 Analysis of Results 81

3.2 Oligonucleotide Melting Temperature Assays 82
3.2.1 Oligonucleotides 82
3.2.2 Oligonucleotide set-up for duplex melting temperature analysis 82
3.2.3 Duplex Melting Temperature (Tm) Analysis 82
3.2.4 Analysis of Results 83

3.3 Microarray Assays 85
3.3.1 DNA Preparation 85
3.3.2 PCR Amplification 85
3.3.3 Agarose Gel Electrophoresis 86
3.3.4 PCR Product Purification 87
3.3.5 Microarray Spotting 87
3.3.6 Probe Hybridisation 88
3.3.7 Stringency Washing 88
3.3.8 Slide scanning 89
3.3.9 Scan Analysis 89
3.3.10 Comparative data Analysis 89

4 Results and Discussion 92

4.1 Microfabrication Assay Results[1] 92
4.1.1 Calculating an estimate of experimental error 92
4.1.2 A comparison of silicon channel extraction of DNA diluted in (a) SDW, (b) Qiagen AL buffer 93
4.1.3 Increasing the amount of DNA recovered as compared to a control sample by increasing sample incubation time 'Y': 95
4.1.4 Optimising the Qiagen buffer to enhance extraction efficiency 98
4.1.5 A comparison of silicon channel extraction efficiency of different DNA samples concentrations: 99
4.1.6 Silicon channel extraction of DNA from whole liquid blood. 102
4.1.7 Silicon channel extraction of DNA from samples contaminated with lead (II) nitrate (PCR inhibitor). 104
4.1.8 A comparison of the amount of DNA which can be trapped from two different length channels. 106
4.1.9 DNA saturation of a silicon channel 107
4.1.10 General discussion 109

4.2 Melting Temperature (*Tm*) Assay results 112
4.2.1 Optimising experimental design 112
4.2.2 Calculating a theoretical Melting Temperature (Tm) 124

4.2.3 Duplex melting assay 125
4.2.4 Validation of MOSAIC as an accurate Tm prediction model 125
4.2.5 Defining a trend in mismatch stability 137
4.2.6 Defining a trend in the context of single mismatches 140
4.2.7 Visual Basic 6.0 Computer program (MOSAIC v1.0) 143
4.2.8 General discussion 144

4.3 Microarray Assay Results 147
4.3.1 Oligonucleotides attachment 147
4.3.2 Calculating an estimate of experimental error 148
4.3.3 Primary Amine group oligonucleotide modification 150
4.3.4 Optimum allele specific probe concentration 150
4.3.5 Duplex detection (direct approach) 151
4.3.6 Pentaplex A development (Direct vs. Indirect) 153
4.3.7 Pentaplex B development 159
4.3.8 Decaplex detection 163
4.3.9 Pentaplex C-v1 166
4.3.10 Cy3/Cy5 signal balancing 168
4.3.11 Slide type 169
4.3.12 Slide Quality 171
4.3.13 Spotting and Pen set Blockage 174
4.3.14 Primer-dimer 175
4.3.15 General discussion 180

5 Conclusions 183
5.1 Microfabrication and DNA Extraction 183
5.2 DNA Duplex stability 183
5.3 Microarray platform for DNA detection 184

6 Appendix 185

6.1 Appendix 1 – Oligonucleotide sequences 185
6.1.1 Duplex oligonucleotide sequences 185
6.1.2 Microarray PCR primer and probe sequences 186

6.2 Appendix 2 – Duplex Melting Temperature data 190

6.3 Appendix 3 - MOSAIC computer program code 194

6.4 Appendix 4 - Comparative Data Analyses Figures 204

7 References 215

Index of Figures and Tables

Figures

Figure 1.1 demonstrates a C•T Single Nucleotide Polymorphism (SNP)..... 4

Figure 1.2 Schematic of the ARMS principle. Only primer (a) with its 3' complementary base can extend. Primer (b) has a 3' mismatch and will not extend..... 5

Figure 1.3 A schematic representation of the URP/ARMS principle. In this example a SNP with a C•T polymorphism is targeted by two allele specific primers. Only the primer that has a 3' complementary base becomes extended during PCR. This is shown as the G primer in (a) carrying universal 1, and the A primer in (b) carrying universal 2..... 6

Figure 1.4 A schematic showing dye-labelled universal primers acting as amplification primers in phase 3 of the URP principle..... 8

Figure 1.5 Possible PCR product combinations pertaining to a single SNP locus following URP SNP amplification according to the ARMS principle..... 8

Figure 1.6 An electrophoretogram of a 21 SNP multiplex. Where a blue and green peak is present at a given position, the individual sample is heterozygotic for that SNP. Single green or single blue peaks represent SNPs that are homozygotic..... 9

Figure 1.7 A schematic representation of the likely processes contained within a μTAS device (proposed by Manz *et al.,* 1990).....10

Figure 1.8 Different amplification profiles for eight SNP loci amplified in singleplex using the same PCR conditions. 28

Figure 1.9 A schematic representation of five of the six strand dissociation processes. Case I - internal loop; Case II – Ends; Case III – Expansion of pre-existing coils; Case IV – Expansion of loops through to ends and Case V – coalescence of neighbouring loops. The Final case is simply separation of the two strands. (Blake *et al.,* 1999).....40

Figure 1.10 An example of a circular tiling array (Southern *et al.,* 1999). The array represents a tiling path of oligonucleotides complementary to a target sequence. Successive positions along the array, which step through the sequence are indicated by an arc shape. Areas of hybridised ^{32}P-labelled sequence are revealed by autoradiography with the darkest area indicating regions of strongest hybridisation signal. Arcs facing left (, denote sequences with matching 5' ends whereas arcs facing right), denote sequences with matching 3' ends.....57

Figure 1.11 (a) An Oligo array with unit lengths between 6 and 11 units; (b) Dye labelled template bound non-specifically and (c) A sample STR unit (9) following stringency washing. 68

Figure 1.12 Hybridisation patterns between two allele specific probes (labelled in green) hybridising to heterozygote sample on two different slides (1+2). Where probe sequences match the target sequence through their entire length (1b/2a) probes remain hybridised. Where sequences show mismatch (1a/2b) these probes detach during stringency washing. Locus specific probes (labelled in red) indicate the locus identity. The combination of signals between the two slides determines the SNP genotype. Hussain *et al., (*2003). 73

Figure 1.13 A schematic of the direct hybridisation method for two different forward probes (labelled in green) on two different slides (1+2). Only probe sequences that are complementary through their entire length (1a/2b) remain annealed. Sequences with a mismatch (as would be the case in 1b/2a), do not remain hybridised. As before, locus specific probes (labelled in red) indicate the locus. The combination of signals between the two slides determines the SNP genotype. 74

Figure 3.1 An SEM showing the deep channel morphology obtained from a cross-section through a silicon wafer. 77

Figure 3.2 A 300 mm microfabricated silicon channel showing the inlet and outlet reservoirs. 78

Figure 3.3 A schematic representation of the 5 phase duplex temperature assessment program. 83

Figure 3.4 A plot of 1st negative derivative fluorescence vs. temperature for four different oligonucleotide duplexes from phase 4 (LightCycler™ *Tm* analysis software.)....84

Figure 3.5 A plot of Test 1 ratios ([Cy5(a)/Cy3(a)] vs. [Cy5(b)/Cy3(b)]) for ten individuals, 20 replicates per individual. In the above example, the genotypes for individuals 1-10 are –ve, HET, HOM b, HET, HET, HET, HOM a, HET, HET and HET, respectively. 90

Figure 3.6 A plot of test 2 ratios (Log_{10} Cy5(a)/Cy5(b)) from the same data (figure 3.6). In this example, the genotypes for individuals 1-10 are –ve, HET, HOM b, HET, HET, HET, HOMa, HET, HET and HET, respectively. 91

Figure 4.1 Elution 1 (Q1) electrophoretograms for sample a, b, c and d respectively. Red peaks indicate size standards and were included as reference markers only. They were not included in peak area calculation. 94

Figure 4.2 Gel images showing profiles corresponding to 8 elutions for incubation times of (a) 2 minutes, (b) 5 minutes and (c) 10 minutes together with a gel profile for a positive control. 96

Figure 4.3 Graph of total peak areas for each experiment mixture compared to a 2 ng positive control. 99

Figure 4.4 Graph of total peak areas for each experiment mixture compared to a 2 ng positive control. 101

Figure 4.5 Total peak area recoveries of DNA from whole blood compared to a 2 ng positive control. 103

Figure 4.6 Gel images showing profiles corresponding to elutions 1-7, for samples 3 and 4. 105

Figure 4.7 Recovery of DNA (total peak area data) from two different channel lengths compared to a control. 107

Figure 4.8 A scatter plot of total peak areas (rfu) from eluted DNA for each sample input. ..109

Figure 4.9 demonstrate two generic 19-mer sequences. Bases shown as N, N' are changeable to create different Strand A and B sequences. Duplex sequences arise from the bringing together of a strand A molecule with a strand B molecule. Depending on the sequences selected, results in either a complementary duplex or a duplex which contains deliberate mismatch/es. 112

Figure 4.10 Fluorescence vs. temperature profiles (FAM labelled oligonucleotides only) 114

Figure 4.11 Fluorescence vs. temperature profiles (Duplexes formed between FAM labelled oligonucleotides and DABCYL labelled oligonucleotides). 114

Figure 4.12 A plot of –d(F)/dT from figure 4.10. 115

Figure 4.13 A plot of –d(F)/dT from figure 4.11. 115

Figure 4.14 A graph of *Tm* °C for four duplex pairs at varying salt concentrations 120

Figure 4.15 A plot of actual *Tm* together with plots of predicted *Tms* from Primer Express™ and MOSAIC for 64 different duplex combinations (as shown in table 4.29). 126

Figure 4.16 Artificial alignment of data from figure 4.15, between (a) Actual *Tm* and Primer Express™ *Tm*, (table 4.29), and (b) Actual and MOSAIC *Tm*, (table 4.29) after adjusting data. 127

Figure 4.17 A plot of predicted *Tm* (MOSAIC) and Actual *Tm* for each oligo duplex pair. ...129

Figure 4.18 A plot of the residuals (Actual *Tm* – MOSAIC *Tm*) for each of the 64 duplexes sorted into ascending order....130

Figure 4.19 Box plots of *Tm* residuals (°C), derived from Primer Express™ data and MOSAIC data. 131

Figure 4.20 A plot of Oligonucleotide length (bp) vs. salt correction factor. A 3 order polynomial regression line, together with its equation is overlaid to describe the data trend. 135

Figure 4.21 A plot of actual *Tm* together with plots of predicted *Tms* from Primer Express™ and MOSAIC for data from table 4.47. 136

Figure 4.22 A view of the opening window of the MOSAIC v1.0 VB6.0 program. Oligo duplex ID 20 from Validation Series II oligos is shown....143

Figure 4.23 A view of the report window of the MOSAIC v1.0 VB6.0 program. A compilation of Oligo sequences corresponding to Validation Series II are shown. 144

Figure 4.24 A schematic diagram of the DABCYL residue attached to a 3′ cytosine base. Manufactured by IBA, Germany....145

Figure 4.25. Simple comparative analysis for tsco D....152

Figure 4.26. Log_{10} Data analysis for tsco D. 152

Figure 4.27. Simple comparative analysis for tsco F. 153

Figure 4.28. Log_{10} Data analysis for tsco F. 153

Figure 4.29 A minigel image showing singleplex PCR products from each of the above loci. A low mass DNA ladder is shown (left and right) together with sizing markers to determine fragment sizes....161

Figure 4.30 shows the relative Cy5 fluorescence values for each locus within Pentaplex C-v1, for sample EA-2. 168

Figure4.31 Analysis graphs following Cy5 allele specific probe hybridisation (type 7 slides).170

Figure 4.32 Analysis graphs following Cy5 allele specific probe hybridisation (type 7* slides). 170

Figure4.33 Analyses graphs following Cy3 allele specific probe hybridisation (type 7 slides). 170

Figure 4.34 Analyses graphs following Cy3 allele specific probe hybridisation (type 7* slides). 171

Figure 4.35 Cy3 and Cy5 images of type 7* slides following hybridisation and washing. Areas associated with hydrophobicity are indicated. 172

Figure 4.36 shows differences in spot density between type 7* and type 7 slides. 173

Figure 4.37 Spot images of probes containing amine modification vs. probes carrying no modification. 174

Figure 4.38 show three arrays spotted with 100 nM Cy3 labelled probe by the same pen... 175

Figure 4.39 A minigel image of PCR products (1 to 6) following PCR amplification using pentaplex A-v4. Low mass DNA ladder (left and right) are included for determining band sizes. 177

Figure 4.40 A minigel image showing pentaplex B-v3 products amplified according to strategy 1 and 2. Primer-dimer relating to each concentration is shown within the white boxed areas. Low DNA mass ladders are shown (left and right) for band sizing comparison. 179

Figure 4.41 A minigel image showing pentaplex B-v3 products amplified according to strategy 1 and 2. Table 4.61 outlines each amplification strategy. Low mass DNA ladders are shown (left and right) for band sizing comparison. 180

Figure 6.1 - Indirect approach, pentaplex A-v1. 204

Figure 6.2- Direct approach, pentaplex A-v1. 205

Figure 6.3 - Indirect approach, pentaplex A-v2. 206

Figure 6.4 - Direct approach, pentaplex A-v2. 207

Figure 6.5 - Indirect approach, pentaplex A-v3. (5-plex) 208

Figure 6.6 - Indirect approach, pentaplex A-v3, (17-plex) 209

Figure 6.7 - Indirect approach, pentaplex B-v1. 210

Figure 6.8 - Indirect approach, pentaplex A-v3. 211

Figure 6.9 - Indirect approach, 9-plex 212

Figure 6.10 - Indirect approach, locus J2 from pentaplex B-v2 212

Figure 6.11 - Indirect approach, pentaplex B-v3. 212

Figure 6.12 - Indirect approach, pentaplex B-v2. 213

Figure 6.13 - Indirect approach, pentaplex C-v1. 214

Tables

Table 1.1 Nearest Neighbour Thermodynamic parameters determined by Breslauer *et al.,* (1986). All values assume 1 M NaCl, 25 °C and pH7. The units for ΔG° and ΔH° are kcal/mol. The units for ΔS° are cal/K per mol. Propagation sequence shows the nearest neighbour of the forward strand separated by a slash to indicate the reverse strand. ...31

Table 1.2 Unified Nearest Neighbour Thermodynamic parameters determined by Allawi *et al.,* (1997). All values assume 1 M NaCl, 37 °C and pH7. The units for ΔG° and ΔH° are kcal/mol. The units for ΔS° are cal/K per mol. 34

Table 1.3 Unified initiation parameters for sequences containing terminal CG or AT base pairs (Allawi *et al.,* 1997). 35

Table 1.4 Mismatch thermodynamic data for (a) G•T (Allawi *et al.,* 1997, 1998(d)); (b) G•A (Allawi *et al.,* 1998(a)); (c) A•C (Allawi *et al.,* 1998(b)) and (d) C•T (Allawi *et al.,* 1998(c)). All values assume 1 M NaCl, 37 °C and pH7. The units for ΔH° are kcal/mol. The units for ΔS° are cal/K per mol. 44

Table 3.1 A list of the components required for SGM*plus* PCR amplification. 80

Table 3.2 SGM*plus* PCR amplification parameters....81

Table 3.3 Duplex melting temperature parameters for duplex *Tm* analysis. 83

Table 3.4 A list of components required for the indirect method of Universal PCR amplification. NB. For the direct method, two allele specific primers (ASPs) are replaced by a single forward locus specific primer (FLSP)....85

Table 3.5 A resume of the URP PCR amplification parameters. NB. Cycle number for phases 1 and 2 should total 35 in each case....86

Table 3.6 Spotting parameters for a typical microarray spotting run....87

Table 3.7 Microarray slide stringency washing parameters following probe hybridisation. 88

Table 4.1 Replicate sample data from (1) section 4.2.2 Elution 1, 2 minute incubation; (2) Elutions 1-8, 2 minutes incubation; (3) section 4.1, positive experimental controls and (4) section 4.1.4,solution 3. In each case, the mean, standard deviation and standard error of the mean (95% confidence interval) was calculated. These values were used to calculate the % st. error of the mean for each data and finally an average % error for all experimental data. 92

Table 4.2 Silicon channel extraction protocol parameters for section 4.1.2. 93

Table 4.3 Combined totals peak areas from elutions 1-4 for each sample (a) to (d). 94

Table 4.4 Silicon channel extraction protocol parameters for section 4.1.3. 95

Table 4.5 Peak areas (rfu) and % recovery from elution 1 for each incubation time compared to positive control....97

Table 4.6 Combined total peak areas and % recovery from elutions 1-8 for each incubation time. 97

Table 4.7 Mixture ratios of Qiagen buffer : Ethanol. 98

Table 4.8 Silicon channel extraction protocol parameters for section 4.1.4. 98

Table 4.9 Combined total peak areas and % recovery from elutions 1-8 for each experimental mixture. 98

Table 4.10 DNA concentrations of different sample solutions. 99

Table 4.11 Silicon channel extraction protocol parameters for section 4.1.5. 100

Table 4.12 Total peak area and % recovery of DNA from each elution, for each sample dilution compared to the total peak area for a 2 ng control. 100

Table 4.13 Silicon channel extraction protocol parameters for section 4.1.6. 102

Table 4.14 Peak area and percentage recovery of DNA from elution 1 and elutions 2-8 as compared to a 2 ng positive control. 103

Table 4.15 Silicon channel extraction protocol parameters for section 4.1.7. 104

Table 4.16 Peak area data from samples 3 and 4. 105

Table 4.17 Silicon channel extraction protocol parameters for section 4.1.8. 106

Table 4.18 Combined total peak area and % recovery data for each elution, for each sample as compared to the 2 ng control. 106

Table 4.19 Silicon channel extraction protocol parameters from section 4.1.9. 108

Table 4.20 Total Peak areas for DNA samples processed through the channel (test) and total peak areas from equivalent control DNA sample (not processed through the channel). 108

Table 4.21 A list of oligonucleotide configurations for figures 4.10 and 4.11. 114

Table 4.22 *Tms* determined from four duplex pairs tested in the presence of 50 mM and 200 mM salt. 116

Table 4.23 *Tms* determined from four duplex pairs. Test 1 and 2 reflect proximal quenching (FRET) where as test 3 and 4 reflect collisional quenching. 118

Table 4.24 Melting temperatures for duplex pairs 1 and 3 in the presence /absence of 1.5 mM $MgCl_2$. 120

Table 4.25 Melting temperatures for stable and unstable 19-mer duplexes in the presence of differing sodium chloride concentrations (mM). 122

Table 4.26 Melting temperatures for stable and unstable 17-mer duplexes in the presence of differing sodium chloride concentrations (mM). 122

Table 4.27 Melting temperatures for the 8-mer oligonucleotide forming a duplex with two different 19-mer sequences. 123

Table 4.28 A comparison of duplex *Tm*, between Actual and predicted models. 124

Table 4.29 show a comparison of actual *Tm* °C to predicted *Tm* °C from (1) Primer Express™ and (2) MOSAIC for duplexes containing only Watson Crick base pairs. 126

Table 4.30 Calculated absolute residual total for observed (Actual *Tm* ºC) – Predicted *Tm* ºC. 128

Table 4.31 Recomputed predicted *Tms* ºC (MOSAIC) assuming the salt correction factor to be 7.962. The absolute residual total for observed (Actual) – Predicted *Tm* is also shown. 129

Table 4.32 Mean Absolute Deviation values for two duplex *Tm* prediction models; Primer Express™ and MOSAIC. 130

Table 4.33 Melting temperature data for 6 duplex pairs analysed with the standard melt assay. In each case, the predicted *Tm* was deducted from the mean *Tm* to determine the new residual. 132

Table 4.34 Actual *Tm* values compared to Primer Express™ and MOSAIC predictions. Residuals were calculated from (a) Actual *Tm* – Primer Express™ *Tm*, (b) Actual *Tm* – MOSAIC *Tm*. 133

Table 4.35 shows the adjusted salt correction factors required to transform each MOSAIC *Tm* prediction such that predicted *Tm* was equal to actual *Tm*. 134

Table 4.36 Revised salt correction factors derived from the regression equation in figure 4.20, together with the recalculated MOSAIC *Tm* prediction. 135

Table 4.37 10 novel duplexes with their associated *Tms* defined by Actual measurement, Primer Express™ and MOSAIC. 136

Table 4.38 show the Mean Absolute Deviation for the two duplex *Tm* prediction models. ... 137

Table 4.39a *Tm* values for all possible combinations for contexts where bases at positions 1 and 3 were fixed as A and A respectively. 138

Table 4.40 Combined rank values for each base pair type for all possible base pair contexts. 138

Table 4.41 Ranked mismatch type (ordered by decreasing *Tm*) for each context. 139

Table 4.42a δ*Tm* values for all possible A•G mismatch contexts. 141

Table 4.43 Combined rank values for each single mismatch context 141

Table 4.44 show the ranked positions (ordered by increasing δ*Tm*) of each mismatch context, for each single mismatch type. 142

Table 4.45 Replicate sample data from section 4.3.5 Tsco D Cy3 and Cy5 (indirect approach) for a heterozygote sample. In each case, the mean, standard deviation and standard error of the mean (95% confidence interval) was calculated. These values were used to calculate the % st. error mean for each data and finally an average % error for all experimental data. 149

Table 4.46 rfu values corresponding to average specific and background fluorescent signals for two probes at three different concentrations on a singleplex PGM+ sample. 151

Table 4.47 A list of probe mixtures used for detecting duplex PCR products. 152

Table 4.48. Genotype results for 11 individuals amplified with Pentaplex A-v1, appendix 6.4 figures 6.1. 154

Table 4.49. Genotype results for 11 individuals amplified with Pentaplex A-v1, appendix 6.4 figures 6.2. 154

Table 4.50 Genotype results for 11 individuals amplified with Pentaplex A-v2, appendix 6.4, figures 6.3. 155

Table 4.51 Genotype results for 11 individuals amplified with Pentaplex A-v2 and detected using the direct approach. Corresponding genotype results are shown from 377 gel electrophoresis analysis. See appendix 6.4, figures 6.4. 156

Table 4.52 Combined genotype results for 11 individuals amplified with either pentaplex A-v3, appendix 6.4, figures 6.5, or from the 17-plex, appendix 6.4, figures 6.6. 158

Table 4.53 Genotype results for 11 individuals amplified with pentaplex B-v1, appendix 6.4, figures 6.7. 160

Table 4.54 A comparison of the old and new amplicon sizes for each locus associated with each pentaplex. 161

Table 4.55 Genotype results for 11 individuals amplified with modified pentaplex A-v3, appendix 6.4, figures 6.8. 162

Table 4.56 Genotypes for sample EA1-20 amplified with the nine-plex (tsco F omitted). Appendix 6.4, figure 6.9. 163

Table 4.57 Genotype results for 11 individuals amplified with pentaplex B-v2, appendix 6.4, figures 6.12. 164

Table 4.58 show the relative Cy3 and Cy5 fluorescence values corresponding to alleles C and T respectively, appendix 6.4, figures 6.10. 165

Table 4.59 Genotype results for control DNA amplified with pentaplex B-v3, appendix 6.4, figures 6.11. 166

Table 4.60 Genotype results for 11 individuals amplified with pentaplex C-v1, appendix 6.4, figures 6.13. 167

Table 4.61 Amplification primer concentrations employed between the two different PCR strategies. 178

Acknowledgements

I wish to express my sincere thanks to my internal supervisor, Dr Peter Gill for his helpful advice. Particular thanks also to my Southampton University supervisors, Professor Tom Brown and Professor Keith Fox for their continued support and encouragement.

I am extremely grateful to Jenny Hughes for her technical assistance in my undertaking of the silicon channel investigation. In addition I would like to express my sincere appreciation to Dr Tim Cox of QinetiQ who provided all of the silicon channel devices for this work. His motivation persuaded me to file this piece of work as a patent application, so for that, I am particularly grateful.

To all my friends in the Forensic Science Service®, I am extremely grateful for their help and encouragement throughout. Particular thanks must go to Peter, Alex and Lindsey who made lunchtimes all the more interesting. Also thanks to Pieris Koumi for his technical assistance in Visual Basic programming.

I would also like to thank the many others who have contributed less directly through informal discussions and/or associations, also The Forensic Science Service® for their support and funding which has enabled me to undertake this course of study.

Finally I would like to thank Vicky, for giving me her love and the determination to carry on. Without her support, this thesis may never have existed.

Abbreviations

A, dATP, ddATP	adenine, deoxyadenosine triphosphate, dideoxyadenosine triphosphate
ABD	Applied Biosystems Division
AC	alternating current
AL	Qiagen Chaotrophic salt
ARMS	amplification refractory mutation system
AW1	alcohol wash 1 (50% Ethanol)
AW2	alcohol wash 2 (70% Ethanol)
bp	base pair
BOC	British Oxygen Company
BSA	bovine serum albumin
C, dCTP, ddCTP	cytosine, deoxycytidine triphosphate, dideoxycytidine triphosphate
CE	capillary electrophoresis
cm	centimetre
Cy3	indodicarbocyanine dye label – Ex. 552 nm Em. 568 nm
Cy5	indodicarbocyanine dye label – Ex. 648 nm Em. 668 nm
DABCYL	4-(4'-dimethylaminophenylazo)benzoic acid
DB	dextran blue/formamide
DC	direct current
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
FAM	5/6-carboxyfluorescein
FRET	fluorescence resonance energy transfer
G, dDTP, ddGTP	guanine, deoxyguanosine triphosphate, dideoxyguanosine triphosphate
g	gravity
GuHCl	guanadine hydrochloride
HET or het	heterozygote
HOM or hom	homozygote
kb	kilobase
L	litre
µCE	microcapillary electrophoresis

µg	microgram
µL	microlitre
µm	micrometre
µM	micromolar
µTAS	micro total analysis system
MAD	mean absolute deviation
MΩ	Mega Ohm
mg	milligram
$MgCl_2$	magnesium chloride
min	minute
mJ	millijoules
mL	millilitre
mm	millimetre
mM	millimolar
MTAP	microarray technology access program
N, dNTP, ddNTP	adenine/cytosine/guanine/thymine, deoxynucleotide triphosphate, dideoxynucleotide triphosphate
NaOH	sodium hydroxide
NaP	sodium phosphate buffer
ng	nanograms
nL	nanolitre
nM	nanomolar
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
$Pb(NO_3)_2$	lead (II) nitrate
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pg	pictogram
pL	picolitre
psi	pounds square inch
QLHS	quick light hybridisation solution
RFLP	restriction fragment length polymorphism
rfu	relative fluorescence unit
ROX	6-carboxy-X-rhodamine
SDS	sodium dodecyl sulphate
SDW	sterile distilled water
SSC	buffer containing 10 parts sodium chloride to 1 part sodium citrate, adjusted to pH 6.9 – 7.1
STR	short tandem repeat

SNP	single nucleotide polymorphism
T, dTTP, ddTTP	thymine, deoxythymidine triphosphate, dideoxythymidine triphosphate
TE	tris(HCl) EDTA
TEAC	tetraethyl ammonium chloride
TMAC	tetramethyl ammonium chloride
Tris-HCl	tris(hydroxymethyl)methylamine, pH adjusted with hydrochloric acid
URP	universal reporter primer
UV	ultra violet
V	volts
v/v	volume per volume
VNTR	variable number tandom repeat

1 Introduction

1.1 DNA in the Forensic Science Service®

1.1.1 DNA analysis for forensic identification – A Brief History

The application of DNA to Forensics was first realised by Jeffreys *et al.,* (1985); Gill *et al.,* (1985). The detection of specific hypervariable loci (minisatellites) markers, by restriction fragment length polymorphism (RFLP) analysis, generated a specific DNA pattern or 'fingerprint'. This pattern was determined by the positions at which restriction enzymes cleave DNA. Between individuals, the exact position of these cleavable sites changes as a result of different numbers of repeat sequences. It is this variable number of tandem repeats (VNTRs), within loci that are exploited for human forensic identification. Although RFLP analysis vastly improved the discrimination power compared to standard serological techniques (Gill *et al.,* 1985), several micrograms of DNA were required for analysis. This often limited the suitability of the technique for many crime scene analyses, where the amount of biological material deposited was often limited.

Then in 1987, Mullis and Falloona developed the molecular biology technique known as the Polymerase Chain Reaction or PCR. With prior knowledge of the DNA sequence which flanks the specific region of interest, small artificial fragments of DNA referred to as primers can be used to replicate the target region of the DNA strand. PCR proceeds in a cyclical manner; after each cycle or round, the number of copies of the specific target increase. Under optimal conditions, the increase is exponential with the number of template copies doubling after each complete cycle.

When PCR was discovered, the method of choice for DNA forensic analysis involved minisatellites (Restriction Fragment Length Polymorphisms - RFLP). Typically, minisatellites are comprised of a core sequence between 9 and 80 base pairs in length, that are tandemly repeated to yield stretches extending between 1000 and 20,000 base pairs in size. Amplifying regions of DNA, of this magnitude, was highly inefficient and therefore did not readily lend itself to RFLP analysis.

To address this issue, the use of microsatellites or Short Tandem Repeats (STRs) as templates for PCR were investigated (Urquhart *et al.,* 1994). STRs are smaller in base composition than minisatellites, having a core sequence between 1 and 6 base pairs. Each core sequence is repeated a number of times (typically 5-50), depending on the STR in question, resulting in a sequence between 40 and 500 base pairs in length. STRs are widely distributed throughout the human genome and are found in non-coding regions. STR loci are much more amenable to PCR amplification. In addition, it has been demonstrated that certain STRs exhibit a high degree of length polymorphism between individuals making them suitable candidates for forensic human identification (Urquhart *et al.,* 1994; Kimpton *et al.,* 1994).

The first STR based DNA profiling test, for routine use in Forensic casework, was used by the Forensic Science Service® (Kimpton and Lygo, 1994). This test consisted of 4 STR loci which were co-amplified as a quadruplex reaction using fluorescently labelled primers. Due to the amplicons being much shorter, as a result of the STR loci being smaller, PCR was found to be more efficient. This meant that the total amount of starting material (DNA) could be much smaller in quantity compared to the amount required for RFLP analysis. After separation of these smaller PCR products by denaturing polyacrylamide gel electrophoresis (PAGE), the identity of each repeat was determined and an STR profile generated. The ability for reduced amounts of DNA (0.5–2 ng) to be amplified and analysed in this way meant that the use of STRs for forensic human identification became widely accepted.

In 1996, the Quadruplex system was replaced by a second STR based test. This was known as the Second Generation Multiplex (SGM) system. It was comprised of 6 STR loci combined with a sex determining marker, amelogenin (Sparkes *et al.,* 1996). Then in 1999, this was superseded by a third STR based test known as AMP*FL*STR® SGM *plus* ™. This test was developed on the back of the previous SGM test by combining its 7 markers with an addition four STR loci, increasing the discrimination power to approximately 1 in 1 billion. Today the SGM*plus* test is routinely used for forensic casework and also in DNA profiling for the national DNA database (NDNAD).

1.1.2 DNA Profiling – A future perspective

Although current DNA profiling methods are designed to amplify DNA as little as 0.1-2 ng, this does not take into account the DNA quality. Quite often, DNA is degraded by the environmental conditions which prevail between the time of deposition and the time it is collected. Degradation results in the breakdown of the DNA structure, leading to fragmentation and loss of integrity. If this occurs within a region targeted for STR analysis, the locus fails to amplify during PCR and is lost from the final profile, resulting in a reduction in the discriminating power of the test. This effect is referred to as allele dropout and is characteristic of old and/or degraded DNA samples.

In order to maximise the discrimination power of the test, by reducing the chance of allele dropout, it is sometimes advantageous to redesign the primer sequences for loci that ordinarily give rise to longer amplification products, so that they anneal closer together (Grubwieser *et al.,* 2003). By reducing the size of the amplification product in this way, the target for the DNA template is also reduced such that the integrity of the sequence spanning a particular locus may be protected to a greater extent. Loss of DNA integrity in degraded samples will still be prevalent but the chance of this occurring within these smaller target

regions is less likely. Of course, in highly degraded samples, the DNA may have fragmented to an even greater extent. In these circumstances, redesign of the primer sequences may still not facilitate the amplification of all STR loci meaning that allele dropout may still be prevalent.

1.1.3 Single Nucleotide Polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs), consist of individual nucleotides that are characterised by the presence of two alternative bases at a particular position in the DNA sequence (see figure 1.1). They are present in the genome at a frequency of about 1 in every 500-1000 base pairs (Brookes, 1999) but in some instances, this can be as great as 1 in every 250-350 base pairs (Nikiforov *et al.,* 1994). As well as being very abundant throughout the genome, SNPs are highly conserved within degraded samples, thus attracting considerable interest from the forensics community (Syvanen *et al.,* 1993). The following web site http://snp.cshl.org/ carries a detailed reference of SNP identification to date, along with their mapped positions within the human genome.

C A T G C C G T A G A C T C C A A G C T G C A C A	Allele 1
C A T G C C G T A G A C T T C A A G C T G C A C A	Allele 2

Figure 1.1 demonstrates a C•T Single Nucleotide Polymorphism (SNP).

SNPs tend to be biallelic in nature meaning that individually, they have a low discrimination power. The suitability of SNPs as forensic markers was outlined by Gill (2001). He demonstrated that the simultaneous interrogation of approximately 50 SNP's would generate the same level of discrimination as current STR profiling methodologies. In order for this to be achievable, there has to be a prerequisite, namely a novel biochemical approach capable of co-amplifying this many loci together. This is due to the limited amount of starting

material negating the possibility of multiple singleplex reactions. The ability to efficiently multiplex, thereby obtaining maximum information from the limited amount of starting material has been the subject of much interest in the Forensic community (Shuber *et al.,* 1995; Gill *et al.,* 2000; Hussain *et al.,* 2003). Within the Biology Research team at the FSS, one particular group has been directly concerned with the development of an efficient multiplex strategy. This has led to the development of the Universal Reporter Primer (URP) Principle (Gill *et al.,* 2000; Hussain *et al.,* 2003) which is detailed in section 1.1.5, which exploits the principle of ARMS.

1.1.4 Amplification Refractory Mutation System (ARMS)

Newton *et al.,* (1989) first described a method to assess any point mutation by Amplification Refractory Mutation System (ARMS). For a given SNP locus, two alleles exist. ARMS can be used to specifically amplify one of these alleles in preference to the other by exploiting the fact that a primer can only extend during PCR if its 3′ base is complementary to the target. If this condition is fulfilled then primer extension results, however if the primer forms a 3′ mismatch then primer extension does not occur. This specificity is maintained by the absence of 3′-5′ proof reading activity in the *Taq* polymerase enzyme.

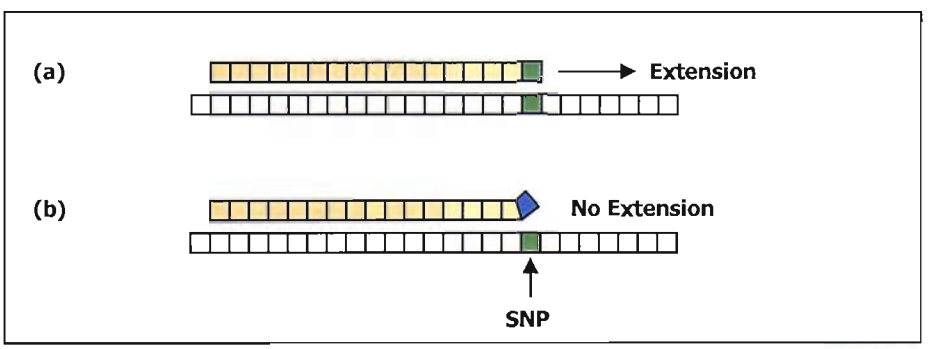


Figure 1.2 Schematic of the ARMS principle. Only primer (a) with its 3′ complementary base can extend. Primer (b) has a 3′ mismatch and will not extend.

ARMS facilitates accurate SNP characterisation. This is combined with the URP Principle as described by Gill *et al.,* (2000); Hussain *et al.,* (2003).

1.1.5 Universal Reporter Primer Principle (URP)

The URP principal is a method for co-amplifying multiple SNP loci in a single reaction. This is achieved by adopting a nested PCR approach in the same reaction tube using common or universal sequences. PCR proceeds in a cyclical manner in three distinct phases.

For each SNP, there are two allele specific primers and a single reverse primer. In phase 1, these primers, carrying unique 5′ universal sequences anneal to template DNA at the SNP position. The universal primer sequences are 20 base pairs in length and are deliberately designed to be non-complementary to genomic DNA. The forward and reverse primers contain different universal sequences but between loci these are the same. The overall configuration is shown in figure 1.3.

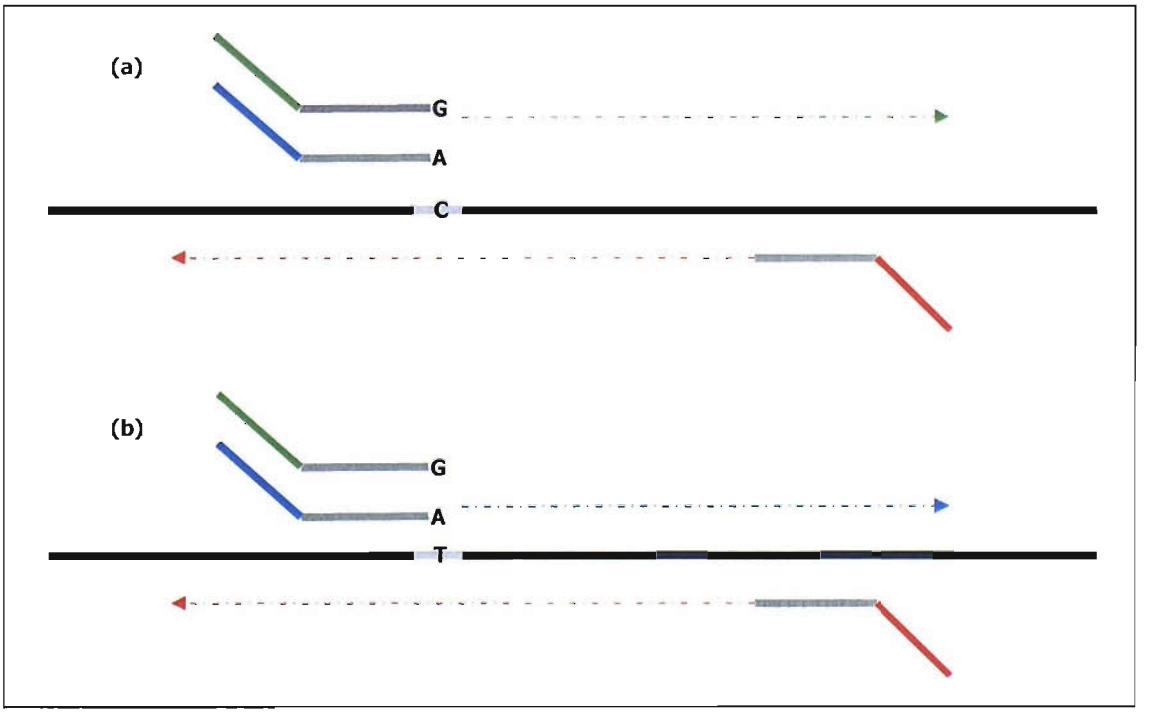


Figure 1.3 A schematic representation of the URP/ARMS principle. In this example a SNP with a C•T polymorphism is targeted by two allele specific primers. Only the primer that has a 3′ complementary base becomes extended during PCR. This is shown as the G primer in (a) carrying universal 1, and the A primer in (b) carrying universal 2.

As phase 1 amplification proceeds, primers anneal to the newly synthesised strands, and extend back, copying along the amplicon through the universal sequences. This means that PCR product derived from individual SNP sites that initially differed by only a single base now incorporate a 20 base universal sequence to discriminate between alleles, enhancing the specificity of the amplification.

In phase 2, amplicons generated in phase 1 continue to anneal to allele specific primers. However, due to the incorporation of the universal sequence in earlier rounds of replication, the primers now anneal through their entire length enabling a higher annealing temperature to be adopted. This ensures that specificity is maintained through the exponential phase of the reaction. Between loci the same universal sequences become incorporated depending on the SNP allele and this gives rise to enhanced specificity as well as a more balanced amplification between loci.

Phase 3 in the URP principle is where amplicons generated in phases 1 and 2 become fluorescently labelled to allow detection. Between loci, each allele is uniquely tagged by the incorporation of a universal tail, either universal 1 for allele 1 or universal 2 for allele 2. Since the same universal tails are present on every amplicon then it is possible to use these sequences as a target for a second round set of primers. These are identical in sequence to the universal sequences but carry one of two different fluorescent dye labels. At a specific point within the PCR when first round primer concentration becomes limited, the annealing temperature of the reaction is reduced allowing the dye labelled universal primers to anneal to complementary sequence. These primers extend during PCR resulting in the generation of fluorescent PCR product (See figure 1.4).

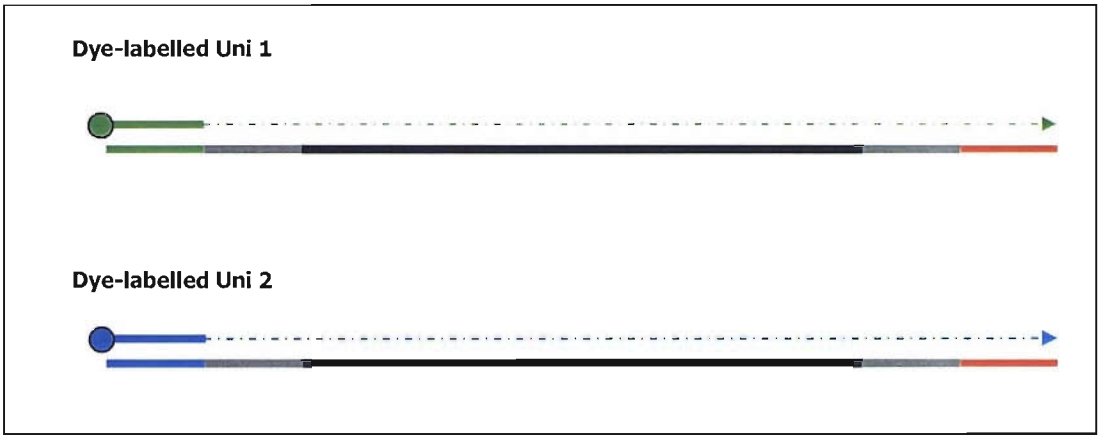


Figure 1.4 A schematic showing dye-labelled universal primers acting as amplification primers in phase 3 of the URP principle.

To prevent asymmetric amplification, the reverse locus specific primers are present in the reaction at a higher concentration than the forward locus specific primers and so act in conjunction with the universal primers in phase 3. For any given SNP therefore, one of three different types of product will result:

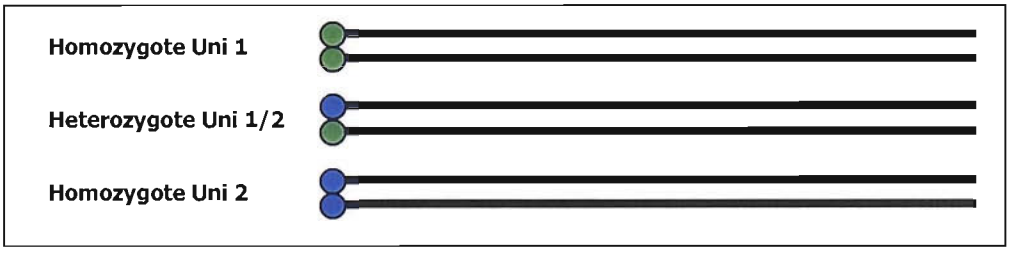


Figure 1.5 Possible PCR product combinations pertaining to a single SNP locus following URP SNP amplification according to the ARMS principle.

Samples that contain both copies of the polymorphism (heterozygotes) result in a mixture of blue and green whereas samples that have just one copy (homozygote) display only one colour.

Small amplicons, ideally up to a maximum of 250 base pairs in length are designed in order to be useful in the detection of highly degraded DNA. In addition, each of the reverse primers within a multiplex reaction can be positioned so that each locus yields amplicons of a

different size. Following PCR, each locus can then be separated out using conventional slab based gel electrophoresis as shown in figure 1.6

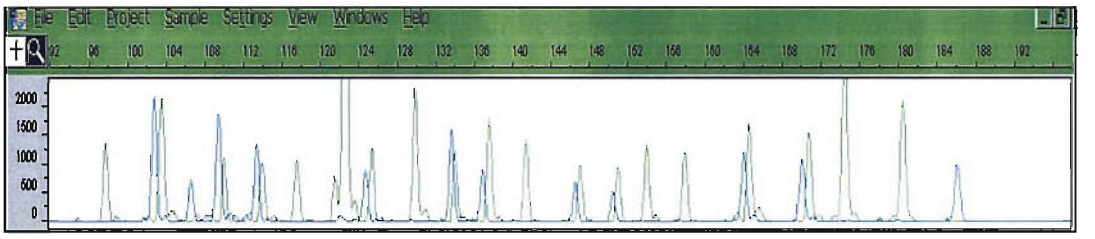


Figure 1.6 An electrophoretogram of a 21 SNP multiplex. Where a blue and green peak is present at a given position, the individual sample is heterozygotic for that SNP. Single green or single blue peaks represent SNPs that are homozygotic.

The adoption of the URP principle ensures balanced amplification between loci and the use of ARMS in combination with URP acts to maximise specificity.

As well as developing alternative amplification strategies, there is considerable interest in developing miniaturised devices for the manipulation and analysis of DNA.

1.2 Miniaturisation with a view to improved DNA analysis

The purpose of the Miniaturisation Project within the context of DNA analyses is to reduce the overall size of the analysis platform. Advantages include a reduction in the amount of consumables required and decrease in the process time (smaller reagent volumes can be manipulated quicker). Collectively, these efficiency gains enable faster DNA analysis compared to conventional means, as well as offering some degree of portability if the platform is small enough. This would potentially allow analyses to be carried out at the crime-scene as a matter of routine.

Miniaturised devices suitable for DNA analysis were first suggested by Manz *et al.,* (1990). They proposed that a Micro total analysis system (μTAS), could be used to perform all of the

fundamental steps that standard platforms could accomplish, including sample pre-treatment, separation, reaction and overall detection (see figure 1.7).

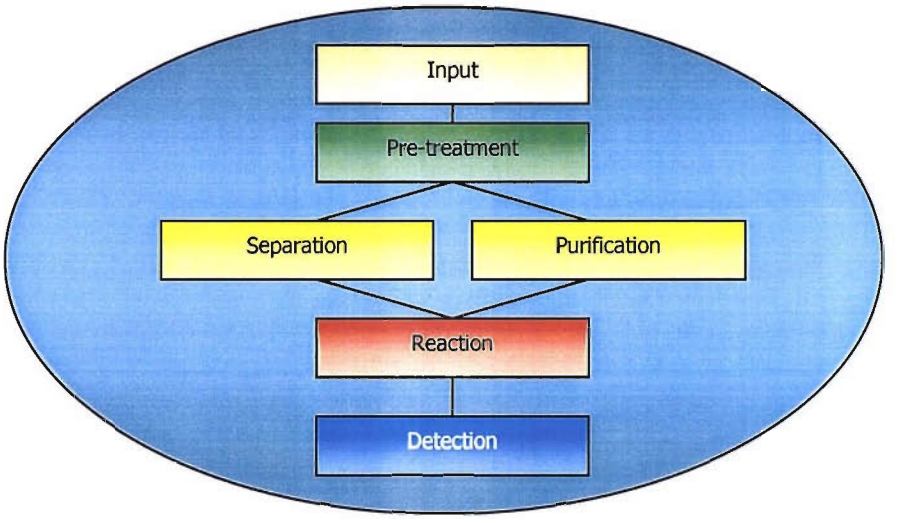


Figure 1.7 A schematic representation of the likely processes contained within a μTAS device (proposed by Manz *et al.*, 1990).

The delivery of a miniaturised total analysis system is not currently feasible due to a number of fundamental problems including manufacturing of microfabricated devices, component integration and liquid handling.

As a result, the emphasis of the miniaturisation group has been to develop specific parts of the process. Three main themes are considered here:

- Pre-treatment: the development of microfabrication as a suitable platform for processing biological material (DNA extraction).
- Reaction: a review of the principles behind oligonucleotide duplex stability with a view to developing an efficient melting temperature algorithm for use in the design of primers for inclusion into novel multiplexes.
- Detection: the development of microarray technology as an alternative means of detecting DNA polymorphisms.

1.3 Microfabrication

1.3.1 The early years

One of the earliest materials to emerge as a suitable substrate for microfabrication was silicon (Reyes *et al.,* 2002). From a historical perspective, silicon was well established within the electronics industry hence many of the fabrication techniques were already well advanced. This made the transition to adapting silicon substrates for novel analysis platforms relatively straightforward.

One of the very first devices to be fabricated out of silicon was a gas chromatographic analyser (Terry, 1975). With this device, separation of a simple mixture could be achieved in a matter of seconds. This device was one of the first to highlight the improved speed with which miniaturised components could perform routine analyses and paved the way for future developments.

When Manz *et al.,* (1990) first proposed a μTAS device, it was considered using a theoretical model. A number of important parameters were identified that were related to hydrodynamics and diffusion theories relevant to chromatographic separation, electrophoretic separation, transport time and detection. In each case, Manz *et al.,* (1990) concluded that these processes could be significantly enhanced, by virtue of increased speed of processing as well as offering improved sensitivity and efficiency. Following this pioneering research, Manz along with his fellow research colleagues devoted considerable effort to developing their own initial theories and specifications for what a μTAS device should contain. Their research and ambition was considered of paramount importance, creating the driving force for the continued development of this relatively new technology (Knapp *et al.,* 1999).

This interest continued until 1994, resulting in a steady increase in the number of publications concerning the development and use of μTAS devices, yet the total number of publications remained fairly low. After 1994, as the potential of these devices became more widely appreciated, the number or research groups exploring the concepts greatly increased (Reyes *et al.,* 2002), resulting in a massive increase in the number of research papers being published.

Rather than consider all of the various type of μTAS device that have been described in the literature, I have chosen to focus on advances relating to the manipulation of macromolecules such as nucleic acids, bearing in mind the FSS desire to develop a fully integrated portable DNA analysis device.

1.3.2 Microfabrication specifically for DNA applications

Much of the early work concerning microfabrication was targeted towards developing proof of principle with regards to a particular function or performance. These included various separations, reactions and modes of detection. Most of these early examples were manufactured out of silicon or glass. In order to improve or enhance the efficiency of these techniques, considerable research emphasis has been given to finding alternative substrates. Examples of these include the use of polymer substrates such as polymethylmethacrylate (PMMA) (Martynova *et al.,* 1997), and also polydimethylsiloxane (PDMS), developed by Dan Corning. These alternative substrates have received significant prominence in recent times due to their inherent physical properties namely, that they are easy to mould, optically transparent, durable, chemically and biologically inert, cheap, non-toxic and chemically stable between –50 °C and 200 °C (Jakeway *et al.,* 2000). Despite these obvious advantages, traditional silicon/glass devices still feature significantly as suitable manufacturing substrates.

For any microfabricated device, the architecture or chip design is regarded as the most important feature (Reyes *et al.,* 2002), as the design itself specifies the function and the sequence of actions which it is designed to perform. In 2002, Auroux *et al.,* published an article which considered some of the integral components which may be seen as prerequisites for any μTAS system. These included components like pumps and valves which enable control of the reagents within a device. In addition to these, components that perform a particular action also need to be considered. Relating to the focus for this investigation, the development and design of these various components are considered below.

1.3.2.1 Pumps and Valves

The type of material used to fabricate a device determines to a large extent what form a valve can take. Silicon is inflexible and therefore not suitable for pumps which rely on liquid displacement through changes in channel structure. A type of peristaltic pump which relies on a localised collapse of an elastomeric channel has been successfully developed (Lim *et al.,* 2001) and represents one of the simplest type of pumping mechanism.

For devices that are fabricated in rigid substrates such as glass or silicon, a number of pump designs have been considered. The first type of pump to emerge was demonstrated by Manz *et al.,* (1991). This used electro-osmotic action to move molecules along a channel through a solution. Where two channels met at a single point, the direction of flow between the carrier solution, reagents and waste reservoirs could be independently controlled by way of switching the voltage applied from one path to another. This had the effect of creating a type of valve mechanism that could be used to direct reagents down a preferred route. Although one of the first pumping mechanisms to be developed, its potential application as a method for separating molecules was quickly realised (Manz *et al.,* 1992).

As well as electro-osmotic driven pumps, simple pumps that rely on passive liquid movement have been developed. The simplest of these is the exploitation of capillary action to fill a mixing device (Seidel *et al.,* 1999). This type of pump eliminates the need for complicated mechanical pumping mechanisms but is limited in the rate and amount of liquid that can be displaced. Improvements in design have resulted in enhanced liquid displacement and by using a combination of capillary action and sample evaporation from the outlet reservoir (Goedecke *et al.,* 2001), liquid displacement is more controllable. Liquid pumping by means of hydraulic pumps has also been considered (Weigl *et al.,* 2000). These are driven by pressure differences generated as a consequence of chip orientation, where the inlet reservoir is positioned higher than the outlet reservoir. Gravity and liquid volume generates a hydraulic pressure that pushes the solution through the device.

Passive valves as demonstrated above, offer very limited liquid controllability. For this reason examples of the use of active valves have been more prominent. In one example, liquid displacement has been achieved by the growing and collapsing of vapour bubbles deliberately trapped within the device (Song *et al.,* 2001). In other adaptations, centrifugal force has been used to actively drive liquids through a device (Duffy *et al.,* 1999; Thomas *et al.,* 2000). The sample is presented to the centre of a disc before being spun, inducing a centrifugal force on the solution. The use of hydrophobic regions (Kellogg *et al.,* 2000) within these channel confines allows for some control in the form of valving to be exerted. The centrifugal pressure must be great enough to overcome the resistance imposed by the hydrophobic regions and therefore preferential liquid flow can be achieved.

More complicated pump designs have been described which include mechanically driven piezoelectric actuators. An example of this type of pump was used on a device in a chamber with two orifices (Shinohara *et al.,* 2000). Actuation of the pump caused displacement of liquid through hinged silicon rubber posts. When the device was deactivated, these posts naturally closed and so acted as gates preventing back flow of liquid.

The development of these complicated mechanical pumps allowed more control with respect to liquid handling but were seen as much more difficult to fabricate. For this reason, many of the earlier pump designs that do not rely on complicated fabrication techniques, such as electro-osmotic flow, have been favoured (Manz *et al.,* 1992).

1.3.2.2 Electrophoresis and sample separation

Since the successful demonstration of one of the earliest designs for a device capable of sample separation (Manz *et al.,* 1991), various separation modes based on electrophoresis have been developed. For many of these examples, electrophoresis of a mixture through a channel containing a separation matrix is typical.

The use of electro-osmotic flow, within this context, in a channel as a means to induce electrophoretic separation was shown to be particularly effective (Manz *et al.,* 1992; Harrison *et al.,* 1992) and rapidly led to developments that saw the integration of sample injection, separation and detection via laser induced fluorescence. These were initially used to separate chemical species (Manz *et al.,* 1991), and then simple mixtures of amino acids and dyes were also demonstrated (Effenhauser *et al.,* 1993; Harrison *et al.,* 1993; Von Herren *et al.,* 1996). These later devices showed an improved analysis time compared to the early work of Manz, resulting in complete separations in fewer than 30 seconds. In addition to the separation of amino acids, oligonucleotides (Effenhauser *et al.,* 1993), and DNA (Woolley *et al.,* 1994; 1995) were also successfully separated via electrophoretic means.

As an extension to the single channel extraction devices described above, micro-capillary array electrophoresis devices have been developed (Woolley *et al.,* 1994). This early device comprised of channels only 3.5 cm in length. They were used to separate restriction

fragments of between 70 and 1000 bp in less than 120 seconds. In later work, a device comprising 12 parallel channels was microfabricated. Each of these channels could be addressed with a different DNA sample and so provided one of the first devices that could perform parallel separations (Woolley *et al.,* 1997). This device was capable of separating a simple mixture containing two different fragments in as little as 160 seconds. Since this work, there have been a significant number of examples of high throughput separations utilising capillary array electrophoresis for genetic analysis (Simpson *et al.,* 1998; Medintz *et al.,* 2001; Berti *et al.,* 2001).

The potential for rapid separations of this nature combined with the ability to perform parallel analyses rapidly saw the technology developed to enable separation of even smaller molecules. Electrophoretic separation of single locus microsatellite markers in fewer than 30 seconds demonstrated an improved sensitivity over previous devices (Schmalzing *et al.,* 1997). The device comprised a channel 26 mm long, filled with a replaceable polyacrylamide matrix. After single locus separation, the device was successfully used to separate PCR samples containing four STR loci. Separation times of less than 2 minutes were reported, demonstrating one of the first ultra-fast allelic profiling platforms (Schmalzing *et al.,* 1997).

To increase resolution power and in an effort to improve performance, considerable effort has been expended to optimise the various parameters which control electrophoretic performance, such as the separation matrix, reaction temperature, channel dimensions and injector size. (Liu *et al.,* 1999; Ronai *et al.,* 2001). For single colour detection of sequencing products, channel lengths equal to 6.5 cm long containing 3% linear polyacrylamide gel have been successfully used to sequence fragments up to 500bp in under 10 minutes (Lui *et al.,* 1999). It was found necessary to re-optimise run conditions to increase the detection limit to 4 colours. Using a 4% gel they were able to generate a similar level of resolution but this was only possible with a longer process time. In fact, separation of these sequencing products took just under 20 minutes but still showed 99.4% accuracy when compared to

conventional sequencing analysis. This work was developed further, and in 2000, Liu *et al.,* demonstrated a 16-channel capillary electrophoresis device that was capable of parallel DNA sequencing. Through laser induced fluorescence detection, they were able to sequence up to 450 bp with 99% accuracy. The possibility of employing microchip electrophoresis for high speed SNP detection has also been investigated, however a decrease in sensitivity was reported (± 5bp), which could not be improved upon (Schmalzing *et al.,* 1999).

With the advent of alternative substrate types, there stood the possibility of manufacturing chips suitable for electrophoretic separation more cheaply than for silicon/glass devices. McCormick *et al.,* (1997) described the batch manufacture of injection moulded plastic channels for electrophoretic separation. Following exhaustive trials, their results suggested that single use, low cost chips of this fashion were suitable for electrophoretic applications. Average standard deviations in run time on the same chip and between chips were calculated to be 1% and 2-3% respectively. In another study, the use of agarose gel filled microchannels, manufactured out of PDMS were used to separate DNA molecules in the range 100-1000 bp (Hong *et al.,* 2001). These results clearly demonstrated the possibility for a cheap but fairly robust, high throughput electrophoretic analysis platform suitable for DNA analysis.

The development of improved separation devices continued with emphasis now being given to their application to the Forensics community (Mitnik *et al.,* 2002). They described the separation of eight multiplexed short tandem repeat loci in under 20 minutes. The sensitivity of this device showed single base pair resolution of between 0.75 and 1.

Despite these advances, since the mid 1990s, there has been an increasing drive towards process integration. Whilst separation efficiencies continued to improve, new developments

integrated PCR reaction chambers (outlined in section 1.3.2.3), with micro electrophoresis channels.

The integration of capillary electrophoresis with PCR was first described by Woolley and Mathies, (1995). Their device consisted of a chamber integrated to a glass capillary electrophoresis chip. This initial design was developed such that the channel was filled with a sieving matrix consisting of hydroxyethylcellulose (Woolley *et al.,* 1996). This was one of the first integrated DNA analysis devices. Ultra high speed DNA sequencing and separations mediated through the use of denaturing polyacrylamide continued to be developed, (Lagally *et al.,* 2000; Dunn *et al.,* 2000). In these devices, DNA sequences were successfully amplified in micro-PCR reactors. Differences in sequence length polymorphism were detected following electrophoretic separation. Genomic as well as plasmid DNA have all been successfully amplified and separated in various microfabricated designs (Waters *et al.,* 1998; Lagally *et al.,* 2001; Paegel *et al.,* 2003).

The development of efficient miniaturised electrophoresis devices that exhibit accurate single base pair resolution is a pre-requisite for forensic identification, and advances towards this requirement are ongoing. One promising report suggests that DNA analysis showing single base pair resolution up to 400 bp is now achievable, (Goedecke *et al.,* In preparation). As an additional requirement to portable forensic DNA identification, microfabricated devices capable of performing miniaturised PCR have attracted considerable research attention.

1.3.2.3 PCR Reaction chambers

Within any biological analysis, there is a need to undertake some form of biological reaction. This might take the form of PCR, DNA sequencing, a specific hybridisation event, various types of immunological assay or other biological reaction, (Reyes *et al.,* 2002). The earliest type of biological reactor to emerge was simply chambers that could be used to contain

reactants (Northrup *et al.,* 1993). At that time, the use of restriction enzymes to digest DNA was typical in large scale DNA manipulations. The reaction chambers suggested by Northrup were modified and used to digest DNA prior to electrophoretic separation (Jacobson *et al.,* 1996).

PCR chambers were initially fabricated out of silicon principally for two reasons. The first was the ease with which they could be micromachined and secondly, their low thermal mass (Northrup *et al.,* 1993; Daniel *et al.,* 1998). Early designs did not contain integrated heating and cooling, but instead relied upon an external heating source (Wilding *et al.,* 1994). PCR devices fabricated out of silicon could be sealed by the application of a Pyrex lid thus preventing sample evaporation (Wilding *et al.,* 1994). Although heating and cooling was not integrated at this time, some researchers found the simplicity of the design encouraging. Nagai *et al.,* (2001) showed that a 40 cycle PCR could be achieved in less than 18 minutes by means of simple temperature transitions created by moving their device to three distinct temperature regions.

In 1995, one of the very first battery powered integrated thermal cycling devices was demonstrated (Northrup *et al.,* 1995), although the PCR performance was not very reproducible. Surface passivation via the use of polymer dynamic coatings (silanisation) resulted in improved PCR performance, approaching the same quality achieved by conventional thermal cycling techniques (Cheng *et al.,* 1996; Shoffner *et al.,* 1996; Giordano *et al.,* 2001). In a comparative assay looking at the efficiency of PCR in standard propylene tubes compared to microfabricated silicon chambers, Lin *et al.,* (2002) demonstrated a significant advantages in the use of the silicon chambers for PCR. This they concluded was due to the silicon's high thermal conductivity and uniform temperature profile within the reaction chamber. With their device they were able to achieve a 30 cycle PCR in only 26 minutes.

The versatility of silicon, in terms of ease with which it can be micromachined, has enabled various designs of reaction chamber to be proposed. Where most groups have employed distinct recognisable chambers, in one particular example, the chip design was comprised of a serpentine channel fabricated in silicon which ran through three distinctive temperature zones. As the PCR reagents passed along the channel, the three zones would create the denaturation, annealing and extension conditions necessary for PCR to proceed (Kopp *et al.,* 1998; Schneegass *et al.,* 2001(a), 2001(b); Zhang *et al.,* 2002).

Alternative integrated designs for DNA amplification in silicon microchambers have been proposed (Daniel *et al.,* 1998). Their design incorporates a thin film platinum resistor as a temperature sensor and heater. The chamber volume in this device was small (approximately 2 μL), which was sealed in the device by applying a small droplet of oil.

With continuing emphasis on integration, researchers started to explore the possibility of many different design approaches. Different examples were presented as a result, including simple PCR chambers connected directly to glass capillary chips (Woolley *et al.,* 1996). Waters *et al.,* (1998) linked four separate PCR chambers to one common capillary electrophoresis channel. Further advances in integration such as that described by Burns *et al.,* (1998) were reported. In their designs, they incorporated channels for fluid handling, PCR chambers with integrated heaters and temperature sensors, gel electrophoresis channel with built in electrodes and a photo detector for fluorescence measurements.

As integration advanced, chambers that had the potential to undertake more rapid PCR became ever more desirable. Khandurina *et al.,* (2000) produced a rapid integrated compact thermal cycler that consisted of a peltier thermoelectric element. The specific configuration allowed for rapid temperature transitions resulting in fast (1 cycle per minute), and efficient amplification. Reducing the PCR volume was also known to reduce the cycling time, as the

rate of temperature change in smaller sample volumes is more rapid. Giordino *et al.,* (2001) demonstrated a 15 cycle PCR regime using an infrared mediated temperature controlled thermocouple device made out of polyimide. This device was successfully used to amplify a 500 bp fragment of lambda phage DNA in 240 seconds with a PCR volume of 1.7 μL. Lagally *et al.,* (2000) described the fabrication of a device capable of performing PCR and μCE separation. The chamber volume in this device was only 280 nL and was heated using thin film heaters. This resulted in cycling times of only 30 seconds. Rodriguez *et al.,* (2003) describes a slightly different variation of an integrated PCR chamber with an electrophoresis channel. Their chamber contained aluminium heaters and temperature sensors on a glass cover. The temperature accuracy of this device was reported to be within ±0.2 °C with typical cycling times of around 16 seconds. Although the sensitivity of the electrophoretic system was not high, the sieving matrix they used was able to differentiate fragments that differed in size by 18 bp.

In addition to integrated devices, stand alone miniature thermocycling devices have been demonstrated (Northrup *et al.,* 1998; Belgrader *et al.,* 1998; 2000; 2001). The MATCI device (miniature analytical thermal cycling instrument) was originally developed by Northrup in 1998. Based on earlier developments pioneered by his group, the device is battery operated and is completely stand alone. Contained within a brief case, the device consists of a chamber with integrated heaters, simple electronics for efficient temperature control and fluorescence based real time detection. The device is portable and can be used for various applications. Demonstration of the MATCI device has been shown by a SNP polymorphism analysis using a TaqMan™ assay. Although the device successfully identified SNP polymorphisms, its sensitivity was approximately 10 fold less that that reported from the ABI 7700 (Sofi Ibrahim *et al.,* 1998).

1.3.2.4 Detection

The development of integrated PCR chambers combined with electrophoresis chips has in many instances taken precedence over the development of microfabricated detectors. Jakeway (2000), report that the limitation to most μTAS device is in fact the detector. Small level detection in analytical devices is usually confined to optical based methods and of these the most prominent is that of laser induced fluorescence detection. For DNA applications, miniaturised capillary electrophoresis results in the separation of PCR products. The detection of these products via this particular detection method has been well documented, (Woolley *et al.,* 1996; Lagally *et al.,* 2000; Khandurina *et al.,* 2000; Northrup *et al.,* 1998). For this type of detection, the level of sensitivity can be variable but by confining the detection area of the electrophoretic channel the level of sensitivity has been shown to be significantly increased (Lyon *et al.,* 1997). Current DNA analysis methods rely on the amplification of STRs and so separation and detection via electrophoretic methods is considered standard practise. For this reason further consideration of alternative detection strategies in this type of miniaturised context are not considered here. Instead section 1.5 of this introduction draws attention to alternative detection methods for amplification strategies which target different DNA polymorphism markers.

1.3.2.5 Development of Extraction and sample preparation

One of the most important pre-requisites of any miniaturised device is its ability to perform some form of sample pre-treatment. In a review by Lichtenberg *et al.,* (2002) sample pre-treatment is acknowledged as one of the last remaining hurdles to be overcome in achieving μTAS. In their review, they cite four reasons for sample pre-treatment, summarised as follows:

- Separation of the sample from the remaining sample matrix. This is characterised as the specific removal of material that may affect the ability to perform the analysis. For miniaturised devices this might be the removal of particulates that may block or clog the device.
- Pre-concentration. This is necessary when the analyte is present in minute quantities.
- Derivatisation. The analyte may require some form of chemical transformation in order to render it detectable.
- Biochemical pre-treatment.

In most instances and in particular for DNA applications, both separation and pre-concentration are particularly important. Separation involves the initial processing of the sample in some way so as to make available the DNA molecule. This is often achieved by simply cell lysis. When cell lysis occurs, the integrity of the cell membrane is disrupted such that DNA from the cell nucleus is released into solution. As well as DNA being released, the sample will contain other cellular material and debris all of which could compromise subsequent analysis if ignored. As a way of preventing this, the unwanted substrates are removed from the sample prior to analysis. This act of purification is essential for maximising the efficiency of DNA analysis. With pre-concentration, this is necessary due to the often limited amount of biological material recovered form a crime scene.

The first type of miniaturised extraction device to emerge was simple microdialysis compartments. These consisted of two microdialysis chambers sandwiched between three polymer layers. The layers contained serpentine channels which set-up a counter flow liquid separation unit (Xu *et al.,* 1998; Xiang *et al.,* 1999). Early use of these devices was directed towards sample cleanup, for example desalting of DNA or protein extracts.

With particular regard to DNA extraction, initial sample processing to yield DNA molecules has been considered within the context of miniaturisation. Separation was induced with the development and use of a microlysis chamber (Belgrader *et al.,* 1999). In this device, a microsonicator acting under electrical control was used to lyse anthrax spores. Of course, this process only breaks open cells to yield DNA and therefore does not separate the DNA from other cellular material.

Traditionally, and within non-miniaturised contexts, DNA extraction and separation has involved the use of organic compounds such as phenol chloroform, combined with cold ethanol precipitation to selectively remove DNA from biological material (Holm *et al.,* 1986). Although effective, more recent reports have suggested that DNA purification techniques of this type are only really suitable for large scale extractions (Tian *et al.,* 2000).

Following on from these early extraction processes, the development of non-organic based extraction methods has been proposed (Kelly, 1995). These rely on the precipitation and selective trapping of DNA to a silica matrix contained within a column, under the influence of high ionic strength buffers, (guanadine hydrochloride). The use of guanadine derivatives (specifically guanadine thiocyanate), was first used for simple nucleic acid purification by Boom *et al.,* (1990), however, this early work demonstrated poor DNA yields of approximately 50% yield and also significant DNA shearing. In later work, the use of silica as an extraction medium stabilised the DNA molecule following adsorption (Kelly, 1995; Tian *et al.,* 2000). By reversing the ionic strength of the binding buffer, DNA could be released from the silica surface and was subsequently presented as a purified intact DNA extract. This methodology has received considerable interest and has resulted in commercial exploitation by various companies of which Qiagen™ is the most active. The application of the Qiagen extraction method for routine DNA extraction in forensic casework was first described by Greenspoon *et al.,* (1998). This method demonstrated improved extraction efficiency compared to classical extraction techniques particularly with regards to degraded samples.

Early efforts to integrate DNA extraction into miniaturised devices saw the development of novel design features being integrated into the miniaturised device. Weirs which represent narrowings inside the channel were used to trap white blood cells from a mixture of whole blood. The red blood cells also present, which do not contain DNA, could pass through this weir structure unrestricted (Wilding *et al.,* 1998). These structures were useful in the sense that they could be used to trap cells of a particular type and therefore cells containing DNA however, they were particularly susceptible to clogging during operation and were not suitable for multiple use. In addition, they did not perform actual DNA extraction and so attention was focused towards micro-solid phase extraction systems employing silica matrices.

The physical trapping of DNA to silica surfaces was already well documented but the process for these phenomena was not well understood. In 2000, Tian *et al.,* considered the mechanism for DNA adsorption to silica surfaces. They concluded that the negative potential of a silica surface would be reduced under high ionic strength and so reduce the electrostatic repulsion effect induced by the negatively charged DNA molecule. In addition, high salt levels were known to reduce the activity of water by forming hydrated ions. These ions would dehydrate DNA and the silica surface, driving the DNA towards the silica. High salt concentrations also had the effect of disrupting hydrogen bonds which Tian *et al.,* (2000) suggested would destabilise the DNA molecule, although in principle high salt concentrations were known to stabilise duplex DNA. Tian concluded that as double stranded DNA splits, the liberated single stranded molecules were free to form hydrogen bonds with the silica surface.

These findings gave considerable support to the possibility of using solid phase extraction devices within the context of device integration. The use of octadecylsilane coated silica beads were used within a weir structure by Oleschuk *et al.,* (2000), but were found to have numerous problems. These included a limited binding capacity for capture, limited ability to capture large DNA fragments (kb range), loss of DNA activity due to excessive shearing and

the possibility of irreversible DNA binding. In addition, the device itself was susceptible to clogging, at the bead input stage, leading to poor reproducibility.

An alternative solid phase extraction method involving the use of Sol-Gel was proposed by Wolf *et al.,* (2002). Sol-gel or the sol-gel process is a method for creating ceramic or glass materials. This is achieved by a chemical transition from a liquid "Sol" to a solid "gel" state. Wolfe made use of the sol-gel chemistry to create porous solid-phase extraction channels however they did not achieve very satisfying results. They found that the stability of the sol-gel was poor, being prone to drying which resulted in cracks forming both within the gel and between the gel and channel walls. Liquid flow during sample introduction was redirected through these cracks rather that through the gel resulting in poor mechanical stability and reduced extraction efficiencies. They also noted that any successfully extracted DNA could not be amplified in subsequent analysis.

In response to these early findings, Wolfe *et al.,* (2002) considered the possibility of using a hybrid approach, combining the use of silica coated beads packed into a column using sol-gel chemistry. The use of beads, they felt, would stabilise the sol-gel matrix and give rise to efficient DNA extraction. With this device, they were able to demonstrate successful DNA extraction in the order of nanograms of DNA, in less than 25 minutes. Furthermore, evaluation of the eluted DNA showed that it was of suitable quality for PCR amplification. Wolfe did however report some problems with reproducibility using this approach. In generating the sol-gel-bead mixture, the beads were added to the sol-gel precursor mixture to form a slurry. In some instances a stable suspension could not be maintained as the beads settled before the matrix had completely gelled resulting in an uneven bead/sol-gel matrix. In addition, concentrated bead masses could cause clogging of the device.

Instead of using additional support matrices to extract DNA, some researchers concentrated on using silicon alone as the extraction substrate. High aspect ratio devices were routinely fabricated which could be designed to have a very high surface area to volume ratio (Christel *et al.,* 1998). These designs consisted of silicon pillars arranged into an array, often referred to as beds of nails. Oxidation of the surface of these structures resulted in a thin layer of silicon dioxide which Christel *et al.,* (1998) concluded might provide suitable conditions for the controlled extraction of nucleic acids. From early studies with these devices, Christel attempted to capture, wash and elute small DNA fragments (500 bp) and medium fragments (48,000 bp) using glass binding chemistry. These proved successful and demonstrated DNA extraction quantities of around 40 ng/cm^2 of binding area for solutions in the 100-1000 ng/mL concentration range (Christel *et al.,* 1998; Christel *et al.,* 1999). This was the first extraction of its type using a purpose built microchip.

In later work carried out by Cox *et al.,* (In preparation), the bed of nail devices proved to be difficult to control and did not show consistent reliability. The bed of nails design could become easily blocked if particulates were present within the sample and these structures were prone to bubble formation during sample introduction. Bubble formation is undesirable within microfabricated devices since they are extremely difficult to remove. Much of this and indeed other work which assessed the effectiveness of DNA extraction devices, has resulted in very similar problems being identified, namely clogging and device reliability in terms of manufacturing reproducibility. Both problems severely impact their suitability for full system integration.

In summary, the suitability of microfabricated platforms for DNA analysis has been extensively investigated. Whilst many of the required functions of such a device have been demonstrated, a suitable extraction device is still not available and thus represents one of the major stumbling blocks in producing a fully integrated DNA analysis system.

1.4 Duplex Stability and multiplex design

1.4.1 Background

At the FSS, the requirement for an alternative DNA test has led to the development of the Universal Reporter Primer (URP) principle (section 1.1.5). This relies on the co-amplification of a selection of SNP loci, in a multiplexed PCR reaction. Multiplexing large numbers of loci together is extremely problematic due to problems associated with primer efficiency and primer compatibility. In addition, sequence differences between primers can lead to differences in duplex stability for both primer/primer and primer/template interactions. Within a PCR reaction, for a given annealing temperature, primer/template configurations will result in some duplex combinations being extremely stable whilst other combinations may be unstable. These stability differences affect the degree to which amplification can occur and lead to differences in amplification efficiency between loci. An example of this is shown in figure 1.8. Eight primer pairs directed to genomic DNA targets were amplified as singleplex reactions in the presence of SybrGreen™, on an ABI 7700 real time detection platform. As each locus amplifies, SybrGreen™ binds to the double-stranded amplicon, resulting in a large fluorescent enhancement. This enhancement is monitored throughout and reflects the progress of the PCR reaction.

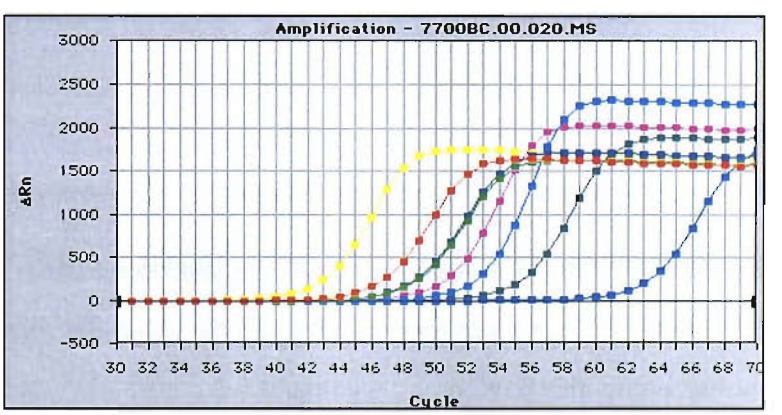


Figure 1.8 Different amplification profiles for eight SNP loci amplified in singleplex using the same PCR conditions.

In this example, the time taken (cycle number), for each locus to reach exponential amplification (linear increase phase of the fluorescent profile) was different, clearly demonstrating different amplification efficiencies. These efficiency differences are thought to be due entirely to sequence differences affecting primer/template duplex stability. One of the earliest studies within the context of oligomer/template stability came from work looking at the effect of probe stability (Wallace *et al.,* 1979). They compared the stabilities of two different oligomer sequences in DNA derived from ΦX 174 bacteriophage. One sequence from the wild type (wt), formed a complete match with the oligomer, whilst a mutated form (am-3) resulted in a single point mismatch. Following thermal stability studies, Wallace *et al.,* (1979) recorded a 10 °C difference in melting temperature between the wt and mutated forms, highlighting the effect of single base mismatches on duplex stability.

In addition to primer/template stability, if primers demonstrate sufficient sequence homology then preferential primer-dimer interaction can occur (Newton *et al.,* 1994). Primer-dimers can arise from a number of different mechanisms, including

- where 3′ nucleotides are deleted from one or both amplifiers (primers)
- where random sequences are inserted between 3′ ends
- when sequences derived from one amplifier appear between amplifiers
- where there is an overlap extension following 3′ annealing

(Brownie *et al.,* 1997)

Successful primer-dimer events result in the formation of artefacts which can preferentially amplify under certain conditions, affecting the amplification efficiency of true PCR products. To assess the stability of nucleic acids in the context of primer-dimer stability, it is possible to exploit the nearest neighbour concept.

1.4.2 Nearest Neighbour Thermodynamics Model

The application of the nearest neighbour model to nucleic acid stability was pioneered by Crothers and Zimm, (1964). This model assumes that the stability of a given sequence (nucleic acid sequence), is dependent on two factors; the identity of adjacent bases and the specific base orientation (SantaLucia *et al.,* 1998). These factors came from early observations that suggested the actual conformational state of a base pair complex exhibited a strong dependence on the states of their nearest neighbours (Poland and Scheraga, 1970).

Since these early investigations, there have been several experimental and theoretical papers concerning the application of the nearest neighbour model for determining nucleic acid stability specifically, RNA (Tinoco *et al.,* 1971; Gotoh *et al.,* 1981; Vologodskii *et al.,* 1984) and for DNA (Breslauer *et al.,* 1986; Sugimoto *et al.,* 1996; SantaLucia *et al.,* 1990, 1991 1996, 1997, 1998; Allawi *et al.,* 1997, 1998(a), 1998(b), 1998(c), 1998(d)). Of these groups, Breslauer *et al.,* (1986) was the first to consider the nearest neighbour model for specifically predicting DNA duplex stability.

1.4.2.1 The application and development of the Nearest Neighbour Model

Prior to the work of Breslauer *et al.,* (1986), the use of thermodynamic data for predicting overall stability and melting behaviour was confined to RNA (Tinoco *et al.,* 1971). Using a similar thermodynamic approach for predicting DNA stability did not occur at the same pace. As a result, early attempts to predict DNA sequence stability relied on the use of RNA thermodynamic data available from previous studies (Gotoh *et al.,* 1981; Vologodskii *et al.,* 1984; Aboul-ela *et al.,* 1985). Early comparisons suggested that this was not suitable as serious errors could be introduced. An example of this was seen when Vologodskii *et al.,* (1984) made the assumption that AT = TA and CG = GC. Oligonucleotide experiments performed using these assumptions showed that the nearest neighbour thermodynamic parameters yielded only approximations with respect to accurate definition of DNA stability.

In response to these early findings, Breslauer *et al.,* (1986) suggested that meaningful evaluation of sequence dependent DNA stability could only be achieved using a DNA specific data set and so this became the focus for their work. Breslauer's group set out with the original intention to describe the thermodynamic model for a DNA molecule in terms of the thermodynamic sum of its nearest neighbour interactions. Their approach makes use of the following thermodynamic relationship:

Equation 1

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

where ΔG° *= Free energy,* ΔH° *= Enthalpy,* T *= Temperature (°K) and* ΔS° *= Entropy*

This equation describes how the free energy change ΔG° of a reaction can be defined by its enthalpy of formation ΔH° - the sum of its entropy ΔS° at any given temperature T (°C). From calorimetric and spectroscopic measurements, Breslauer *et al.,* (1986) was able to measure the free energy, enthalpy and entropy change associated with thermally induced helix-to-coil transitions for a variety of candidate oligo sequences. From these studies, the first set of DNA specific thermodynamic data for all 10 Watson-Crick DNA nearest neighbours was determined. See Table 1.1.

Propagation sequence	ΔH°	ΔS°	ΔG°
AA/TT	-9.1	-24.0	-1.9
AT/TA	-8.6	-23.9	-1.5
TA/AT	-6.0	-16.9	-0.9
CA/GT	-5.8	-12.9	-1.9
GT/CA	-6.5	-17.3	-1.3
CT/GA	-7.8	-20.8	-1.6
GA/CT	-5.6	-16.5	-1.6
CG/GC	-11.9	-27.8	-3.6
GC/CG	-11.1	-26.7	-3.1
GG/CC	-11.0	-26.6	-3.1

Table 1.1 Nearest Neighbour Thermodynamic parameters determined by Breslauer *et al.,* (1986). All values assume 1 M NaCl, 25 °C and pH7. The units for ΔG° and ΔH° are kcal/mol. The units for ΔS° are cal/K per mol. Propagation sequence shows the nearest neighbour of the forward strand separated by a slash to indicate the reverse strand.

Using this data, it was possible to accurately predict the transition enthalpy for any given oligo sequence by calculating the sum of the individual nearest neighbours. Furthermore, these predicted ΔH° values showed excellent agreement to those determined by calorimetric means. The observation that two oligomers with the same base composition but different sequence arrangements exhibited different transition enthalpies confirmed the argument that base sequence also played a crucial role with respect to stability. This is clearly seen by the example given by Breslauer *et al.,* (1986). They showed that ΔH^{37} for the sequences GGAATTCC and GGATATCC were -60.0 kcal/mol and -56.4 kcal/mol respectively, despite the two sequences having the same base composition.

In a similar manner to that described for predicting transition enthalpies, Breslauer *et al.,* (1986) also demonstrated that transition free energies (ΔG°), could be calculated using equation 1. In addition to the 10 nearest neighbour thermodynamic parameters, two additional terms were also included. These are referred to as Δg_{sym} for symmetry correction and Δg_i as an initiation correction (Freier *et al.,* 1986). Δg_{sym} relates specifically to the entropic difference between duplexes formed from self-complementary and non-self complementary sequences and were assigned values of 0.4 kcal and 0 kcal respectively (Breslauer *et al.,* 1986). Δg_i is an initiation correction factor which takes into account the difference between duplexes starting with an AT or GC base pair. This difference leads to a slight change in the amount of energy associated with forming this first base pair and has been assigned values of Δg_i (AT) = 6 kcal and Δg_i (CG) = 5 kcal (Breslauer *et al.,* 1986). Applying these correction factors gives rise to the following equation for predicting the free energies for DNA Oligomers:

Equation 2

$$\Delta G_{total} = -(\Delta g_I + \Delta g_{sym}) + \Sigma \Delta G^{\circ}_{each\ nearest\ neighbour}$$

Using equation 2, with the appropriate correction factors, Breslauer *et al.,* (1986) was able to demonstrate excellent agreement between predicted and experimental free energies for a

series of different oligo sequences and established a growing research interest into thermodynamic data usage.

Since the publication of the Breslauer *et al.,* (1986) paper, there have been a number of papers specifying improved thermodynamic data parameters specifically for DNA (Delcourt *et al.,* 1991; Doktycz *et al.,* 1992; SantaLucia *et al.,* 1996; Allawi *et al.,* 1997). Sugimoto *et al.,* (1996) for example reported that there were occasionally large differences noted using the thermodynamic parameters of Breslauer *et al.,* (1996). They suggested that these individual nearest neighbour thermodynamics were incorrect as well as the initiation factors previously set. As a result Sugimoto *et al.,* (1996) produced a revised set of thermodynamic parameters as well as what they described as improved initiation factors. Using these modified data sets and comparing them to existing data sets for a series of duplexes, Sugimoto *et al.,* (1996) were able to demonstrate that the average errors for the predicted values fell from 23.5% to just 4.8%. For $\Delta H°$, $\Delta S°$ and *Tm*, the new parameters yielded 6.1%, 6.8% and 5.7% compared to existing 9.3%, 9.4% and 19.5%.

Direct comparisons between data from different groups were found to be extremely difficult to carry out for a number of reasons. These are summarised as:

- different oligonucleotide sequences
- different methods for determining thermodynamics
- different methods for analysing data
- different salt concentrations
- different formats for presenting nearest neighbour parameters

(SantaLucia, 1998)

In addition to these factors, the notion that there may be a length dependency to predicting accurate DNA thermodynamics has been proposed (Doktycz *et al*, 1992) however, SantaLucia, (1998) demonstrate no such length dependency for small oligonucleotides (<20 bp), for nearest neighbour thermodynamics, ruling this out as a possible source for error.

Due to the apparent discrepancies reported between the various group's data, Allawi *et al.,* (1997) generated a refined set of nearest neighbour parameters based on the work of fellow co-worker SantaLucia, (SantaLucia *et al.,* 1996). This data is referred to as the unified data set and is presented in table 1.2.

Propagation sequence	$\Delta H°$	$\Delta S°$	$\Delta G°_{37}$
AA/TT	-7.9	-22.2	-1.02
AT/TA	-7.2	-20.4	-0.88
TA/AT	-7.2	-21.3	-0.60
CA/GT	-8.5	-22.7	-1.46
GT/CA	-8.4	-22.4	-1.46
CT/GA	-7.8	-21.0	-1.29
GA/CT	-8.2	-22.2	-1.32
CG/GC	-10.6	-27.2	-2.17
GC/CG	-9.8	-24.4	-2.24
GG/CC	-8.0	-19.9	-1.83

Table 1.2 Unified Nearest Neighbour Thermodynamic parameters determined by Allawi *et al.,* (1997). All values assume 1 M NaCl, 37 °C and pH7. The units for ΔG° and ΔH° are kcal/mol. The units for ΔS° are cal/K per mol.

In 1998, SantaLucia considered each of the published data sets together with the unified parameter set by presenting them in a common format which enabled a direct comparison to be made. In this study, particular attention was paid to the ΔG^{37} parameter, as this was considered more accurate than $\Delta H°$ or $\Delta S°$ due to compensating errors linked to the initiation parameters (for most of the studies, values corresponding to the initiation parameters were not calculated). Although initially, the various data sets suggested

differences with regards to free energy stability, SantaLucia, 1998 was able to demonstrate a high degree of consensus between the different groups with the exception of the Breslauer *et al.,* (1996) data. In addition, the initiation parameters which were initially selected as the major cause of discrepancy within the Breslauer data (Sugimoto *et al.,* 1996), was confirmed, suggesting that these were an important factor in determining DNA stability. Terminal base pair identity was found to affect the initiation contribution; specifically, sequences with terminal GC base pairs had to be treated differently to sequences that contained terminal AT bases (Allawi *et al.,* 1997). As such the following initiation parameters were suggested (see table 1.3).

Type	$\Delta H°$	$\Delta S°$	$\Delta G°_{37}$
Initiation G•C	0.1	-2.8	0.97
Initiation A•T	2.3	4.1	1.03
Symmetry correction	0	-1.4	0.43

Table 1.3 Unified initiation parameters for sequences containing terminal CG or AT base pairs (Allawi *et al.,* 1997).

These revised initiation and symmetry correction values as well as the unified nearest neighbour thermodynamic parameters have been applied to the nearest neighbour model concept. These data have been used in various studies (Allawi *et al.,* 1997; SantaLucia, 1998), and demonstrate an effective way to assess DNA stability.

By way of example, using the above data parameters from tables 1.2 and 1.3, $\Delta H°$, $\Delta S°$ and ΔG_{37} values for any given oligo sequence can be calculated as follows:

For the Sequence CGTTGA, there are 5 nearest neighbours to consider; CG, GT, TT, TG and GA.

Equation 3

$\Delta H°$ = $\Delta H°$(CG/GC) + $\Delta H°$(GT/CA) + $\Delta H°$(TT/AA) + $\Delta H°$(TG/AC) +

$\Delta H°$(GA/CT) + Initiation (GC) + Symmetry

= (-10.6) + (-8.4) + (-7.9) + (-8.5) + (-8.2) + 0.1 + 0

= -43.5 kcal/mol

Equation 4

$\Delta S°$ = $\Delta S°$(CG/GC) + $\Delta S°$(GT/CA) + $\Delta S°$(TT/AA) + $\Delta S°$(TG/AC) +

$\Delta S°$(GA/CT) + Initiation (GC) + Symmetry

= (-27.2) + (-22.4) + (-22.2) + (-22.7) + (-22.2) + (-2.8) + 0

= -119.5 cal/k mol

= -0.1195 kcal/mol)

Equation 1

ΔG_{37} = $\Delta H° - T\Delta S°$ where T = temperature °K.

= (-43.5) – (310 x (-0.1195))

= -6.46 kcal/mol

1.4.2.2 Melting Temperature as a measure of duplex stability

In general nearest neighbour thermodynamics offer a quick and simple way of determining duplex stability. As $\Delta G°$, the free energy, becomes more negative, the more stable a duplex becomes. In a different approach, duplex melting temperature can be used to assess stability. The melting temperature of a duplex can be described in a number of ways. The simplest and most commonly applied formula to do this is as follows:

Equation 5

Tm = [(number of A + T) x 2 °C] + [(number of =C + G) x 4 °C]

(Suggs *et al.,* 1981)

This formula is based on an oligonucleotide dissolved in 1 M salt and is considered as a very rough approximation. This formula is not suitable for oligonucleotides greater than 20 bases in length (Newton *et al.,* 1994). For longer oligonucleotides two other formulae have been derived:

Equation 6

$$Tm = 81.5 + 16.6(\log_{10}[J^+]) + 0/41(\%G+C)-(600/l)-0.63(\%formamide)$$

(Newton *et al.,* 1994)

In this formulae, J = concentration of monovalent cations and l = oligo length. The equation is suitable for primers between 14 and 70 bases long.

Equation 7

$$Tp = 22 + (1.46[(\text{number of G+C}) \times 2] + [\text{number of A+T}])$$

(Newton *et al.,* 1994)

Tp is a term which relates to the optimum temperature at which an oligo anneals to target. The formula is suitable for oligonucleotides between 20 and 30 bases.

In a different approach for predicting *Tm*, it is possible to use the thermodynamic parameters outlined earlier, directly. The relationship between *Tm*, $\Delta H°$ and $\Delta S°$ is shown by the formula:

Equation 8

$$Tm\,°C = [\Delta H / (\Delta S + R\mathrm{Ln}(C_t))] + 273.15$$ (SantaLucia *et al.,* 1996)

In equation 8, R = Molar gas constant (1.987 cal/Kmol) and C_t = total oligo concentration (M). The units for R and ΔS are cal/Kmol where as the units for ΔH are kcal/mol, therefore for the formula to be valid ΔH values should be converted to cal/Kmol by multiplying by 1000.

This formula has been used by many groups to calculate the melting temperature of an oligo sequence (Freier *et al.,* 1986; Sugimoto *et al.,* 1986; SantaLucia *et al.,* 1990, 1991, 1996, 1997, 1998; Sugimoto *et al.,* 1994; Allawi *et al.,* 1997; 1998(a), 1998(b), 1998(c) and 1998(d)). In each instance, the actual length of the oligonucleotide is considered as an irrelevant factor however, Giesen *et al.,* (1998) suggest that only duplexes of between 20-30 bp are suitable for the application of the nearest neighbour model for predicting *Tm*. Many of these studies actually use very much smaller oligo sequences and yet still demonstrate a high degree of similarity between predicted and experimental *Tm* data. This suggests that the conclusion of Giesen *et al.,* (1998) may be extended to include all duplex sizes below 20 bp as well. Beyond 30 bp Giesen suggest that more reliable *Tm* calculations for longer oligo duplexes can be achieved using the Marmur-Schildkraut-Doty equation :

Equation 9

$$Tm\,(°C) = 193.67 - (3.09 - f_{(G+C)})(34.47 - 6.52\,log\,[Na^+])$$

(Blake *et al.,* 1998)

This equation describes *Tm* as a linear function of GC content and salt concentration. Here, the dependence of salt is expressed by the $[Na^+]$ concentration. The dependency of salt concentration on DNA melting temperature was known from as far back as 1965 (Schildkraut *et al.,* 1965). High salt concentrations lead to an increased ionic strength which increases Tm as cations stabilise the DNA duplex. Conversely, a decrease in melting temperature associated with a low salt concentration is known to cause an increase in electrostatic repulsion between negatively charged phosphate ions within the DNA backbone (Schildkraut *et al.,* 1965).

The Marmur-Schildkraut-Doty equation has been used extensively to demonstrate the relationship between salt concentration on stability for large DNAs and large mixtures of DNA with differing base composition [$f_{(C+G)}$], however, this relationship is not quantitative for short sequences (≤4000), (Blake *et al.,* 1998). For smaller oligo sequences, a number of different

correction factors for salt concentration have been proposed. Of these, one of the most frequently used is outlined in equation 10 (SantaLucia *et al.,* 1996). All of the thermodynamic data deduced by this group as well as co-workers have been recorded using oligonucleotides dissolved in 1 M sodium chloride at pH 7. To allow for approximate predictions in lower salt environments, they suggest the following correction:

Equation 10

$$Tm^{[Na]} = Tm^{1M\ Na} + 12.5\log[Na^+]$$

Where $Tm^{1M\ Na}$ is the *Tm* predicted from thermodynamics from tables 1.2, 1.3 and equation 8
$Tm^{[Na]}$ is the *Tm* predicted at the desired sodium concentration

(SantaLucia *et al.,* 1996)

The correction applied by SantaLucia *et al.,* (1996) is lower than that suggested by Breslauer *et al.,* (1986) which was $Tm^{[Na]} = Tm^{1M\ Na} + 16.5\log[Na^+]$, however this revised correction is in agreement with previously published data although the correction is only suitable for oligonucleotides dissolved in >100 mM sodium chloride. Below 100 mM this correction was shown to be unreliable (Kumur *et al.,* 1995; SantaLucia *et al.,* 1998) and led the authors to conclude that further work on the salt dependence of oligonucleotide thermal denaturation was required.

As well as corrections for salt concentration, equation 8 also uses an oligonucleotide concentration factor, as this can also influence melting temperature. Higher oligo concentrations favour duplex formation resulting in a higher melting temperature. The *Tm* of an oligonucleotide sequence therefore is the result of a number of factors influencing nucleic acid stability. For simplicity, the actual *Tm* value calculated is defined as the temperature at which half or 50% of a given sequence is hybridised to its complementary strand (this could be either a second oligonucleotide or a longer genomic DNA template). Rather than melt or

dissociate in a continuous manner, strand displacement has been shown to occur in a piecemeal fashion with increasing temperature (Blake *et al.,* 1999).

Whilst complete strand separation ultimately results, nucleic acid duplexes dissociate in a cooperative fashion by one of five types of sub-transition depending on the sequence and stability of the various domains (Blake *et al.,* 1998).

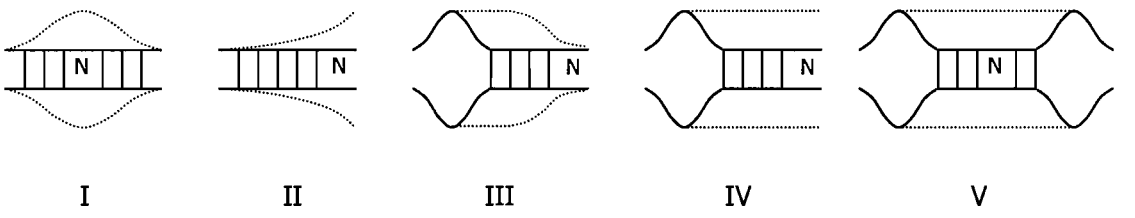


Figure 1.9 A schematic representation of five of the six strand dissociation processes. Case I - internal loop; Case II – Ends; Case III – Expansion of pre-existing coils; Case IV – Expansion of loops through to ends and Case V – coalescence of neighbouring loops. The Final case is simply separation of the two strands. (Blake *et al.,* 1999).

Since *Tm* only relates to the point at which half the strand is displaced, no account is made of the dynamics of strand separation. The melting temperature for a given duplex therefore has to be regarded as an approximation.

Despite these findings, the observed trend in the nearest neighbour stabilities at 37 °C was defined as:

GC > CG> GG> GA =GT = CA >CT> AA> AT>TA (SantaLucia *et al.,* 1996; 1998).

Although this trend is valid, it only applies to deoxyribonucleic acid base residues. Analogues of deoxycytosine, namely N-4-ethyl-2′-deoxycytidine $d^{4Et}C$ have been shown to bond with deoxyguanadine leading to a GC base pair whose stability is very much closer to that of AT according to Nguyen *et al.,* (1998). This has important consequences for assessing duplex stability however, from here on, all reference to duplex stability will assume normal deoxyribonucleic acid bases.

1.4.3 The application of the Nearest Neighbour Model to predicting Tm for mismatched sequences

So far, the application of the nearest neighbour model for predicting duplex stability in terms of ΔG° and *Tm* have been described only for complementary sequences however, there has been an increasing interest in its application for predicting stability of duplexes that contain mismatched base pairs.

In genomic DNA, mismatches occur naturally as a result of errors from the misincorporation of bases during replication (Goodman *et al.,* 1993), or from errors arising from recombination events. These can be the result of mutagenic chemicals, (Leonard *et al.,* 1990), ionising radiation (Brown *et al.,* 1995) and or spontaneous deamination (Brown *et al.,* 1995). Whilst many errors are detrimental to the host organism and are thus repaired, some confer a selective advantage leading to genetic variation and evolutionary change (provided this occurs in gametes).

The consequence of base sequence change is an important consideration in relation to the application of several molecular biological techniques including Southern Blotting (Southern 1975), single stranded conformational polymorphisms (Kanazawa *et al.,* 1986) and PCR (Saiki *et al.,* 1988). In this later example, primer sequences designed to target a specific region of the native DNA sequence may not form perfect matches as a direct consequence of changes in the DNA sequence. In addition, primers may form more stable duplexes (primer dimer) with and between themselves. Accurate prediction of DNA thermal denaturation for possible mismatched primers is therefore of utmost importance. To assess all possible types and incidences of mismatches within a duplex pair would be extremely complicated and time consuming and so much emphasis has been initially given to investigating the effect of just single mismatches.

In addition to the standard Watson and Crick base pairs, AT, TA, CG and GC, there are eight possible mismatches A•A, A•C, C•C, C•T, G•G, G•A, G•T, and T•T which can occur (SantaLucia *et al.,* 1998). In a similar way to using the nearest neighbour model for predicting DNA stability in duplexes containing only Watson Crick pairs, studies have shown that the nearest neighbour model can be extended to include internal mismatches.

The work of SantaLucia and Allawi has been principle to the development of an understanding of mismatched duplex stability. In their approach, a number of oligos were synthesised such that when brought together as a duplex, they contained a single mismatch type. The most sensitive way of assessing the effect of this mismatch is by following equilibrium denaturation by measuring UV absorbance. As the temperature is increased, there is an increase in the perturbation of stacked base pairs during helix-coil transition and these results in a large increase in absorption at 260 nm (Blake *et al.,* 1998, 1999). By measuring the absorbance of the duplex as a function of temperature in this way, and then plotting reciprocal melting temperature (Tm^{-1}) versus the natural log of oligonucleotide concentration (Ln C_T) according to equation 11, thermodynamic data values for $\Delta H°$ and $\Delta S°$ can be deduced.

Equation 11

$$Tm^{-1} = ((R/\Delta H°) \ln C_T) + (\Delta S°/\Delta H°)$$

(Borer *et al.,* 1974)

According to the nearest neighbour model previously outlined, the total energy change, $\Delta G°$ of any oligo sequence can be calculated from the sum of the energy increments for helix initiation, helix symmetry and nearest neighbour interaction between base pairs. Using this relationship allows the contribution of the mismatch to be calculated by simple subtraction of these terms from the measured thermodynamic data:

5′ GGAC<u>T</u>GACG-3′ = Initiation + symmetry + GG + GA + AC + C<u>T</u> + C<u>G</u> + GA + AC + CG
3′ CCTG<u>G</u>CTGC-5′ CC + CT + TG + G<u>G</u> + G<u>T</u> + CT + TG + GC

By rearranging the above formula, the mismatch contribution can be shown as:

$$\begin{array}{lll} \underline{C}T + C\underline{G} = & GGAC\underline{T}GACG & - \text{ Initiation - symmetry - GG - GA - AC - GA - AC - CG} \\ G\underline{G} + G\underline{T} & CCTG\underline{G}CTGC & CC - CT - TG - CT - TG - GC \end{array}$$

Assuming $\Delta G°$ for the mismatch sequence to be –8.37 kcal/mol, and using thermodynamic data values from tables 1.2 and 1.3, the mismatch contribution to duplex stability can be calculated as:

$$\underline{C}T + C\underline{G} = -8.37 - 2(0.98) - (0) - (-1.84) - (-1.30) - (-1.44) - (-1.30) - (-1.44) - (-2.17)$$
$$G\underline{G} + G\underline{T}$$

$$\Delta G° = -0.84 \text{ kcal/mol.}$$

Whilst the exact nearest neighbour values for $\underline{C}T/G\underline{G}$ and $C\underline{G}/G\underline{T}$ are unknown via this method, their overall joint contribution to duplex stability is fairly precise. In this instance, $\Delta G° = -0.84$ kcal/mol.

This approach has been adopted by a number of researchers and has resulted in the publication of a number of papers that deal specifically with particular mismatch types, including G•T (Allawi *et al.,* 1997, 1998(d)), G•A (Allawi *et al.*, 1998(a)), A•C (Allawi *et al.,* 1998(b)) and C•T (Allawi *et al.,* 1998(c)). In each case the thermodynamics of a series of control sequences each containing the single mismatch were determined from a van't Hoff analysis of absorbance versus temperature melting curves as described above. From these data a series of specific mismatch thermodynamic parameters were deduced which were then used to calculate predicted $\Delta G°_{37}$, $\Delta H°$, $\Delta S°$, and *Tm.* A summary of these data are shown in tables 1.4(a), 1.4(b), 1.4(c) and 1.4(d).

(1.4a) G•T mismatch thermodynamics

Mismatch	$\Delta H°$	$\Delta S°$
AG/TT	1.0	0.9
AT/TG	-2.5	-8.3
CG/GT	-4.1	-11.7
CT/GG	-2.8	-8.0
GG/CT	3.3	10.4
GT/CG	-4.4	-12.3
TG/AT	-0.1	-1.7
TT/AG	-1.3	-5.3

(1.4b) G•A mismatch thermodynamics

Mismatch	$\Delta H°$	$\Delta S°$
AA/TG	-0.6	-2.3
AG/TA	-0.7	-2.3
CA/GG	-0.7	-2.3
CG/GA	-4.0	-13.2
GA/CG	-0.6	-1.0
GG/CA	0.5	3.2
TA/AG	0.7	0.7
TG/AA	3.0	7.4

(1.4c) A•C mismatch thermodynamics

Mismatch	$\Delta H°$	$\Delta S°$
AA/TC	2.3	4.6
AC/TA	5.3	14.6
CA/GC	1.9	3.7
CC/GA	0.6	-0.6
GA/CC	5.2	14.2
GC/CA	-0.7	-3.8
TA/AC	3.4	8.0
TC/AA	7.6	20.2

(1.4d) C•T mismatch thermodynamics

Mismatch	$\Delta H°$	$\Delta S°$
AC/TT	0.7	0.2
AT/TC	-1.2	-6.2
CC/GT	-0.8	-4.5
CT/GC	-1.5	-6.1
GC/CT	2.3	5.4
GT/CC	5.2	13.5
TC/AT	1.2	0.7
TT/AC	1.0	0.7

Table 1.4 Mismatch thermodynamic data for (a) G•T (Allawi *et al.*, 1997, 1998(d)); (b) G•A (Allawi *et al.*, 1998(a)); (c) A•C (Allawi *et al.*, 1998(b)) and (d) C•T (Allawi *et al.*, 1998(c)). All values assume 1 M NaCl, 37 °C and pH7. The units for ΔH° are kcal/mol. The units for ΔS° are cal/K per mol.

The determination of melting temperature for duplexes containing a single mismatch of the type specified in table 1.4 is made possible by combining the thermodynamic parameters from tables 1.2 and 1.3 in a piecemeal fashion. Typically, these combined mismatch parameters predict ΔH°, ΔS°, and *Tm* for each of the four mismatch types with deviations of:

- G•T 7.5%, 8.0% and 1.4 °C;
- G•A 7.4%, 8.0% and 1.5 °C;
- A•C 11.0%, 12.2% and 1.3 °C;
- C•T 9.9%, 10.6% and 1.9 °C respectively.

Although some of these values seem to have fairly large deviations, the authors concluded that they still lie within experimental limits, allowing the joint thermodynamic parameters presented in tables 1.2, 1.3 and 1.4 to adequately predict the thermodynamics of oligonucleotides for most applications (Allawi *et al.,* 1998(a), 1998(b), 1998(c), 1998(d)). Whilst G•A and G•T mismatches are more stable than A•C and C•T mismatches, other groups have considered like for like mismatches. Peyret *et al.,* (1999) considered the affect of single like for like mismatches in all Watson and Crick scenarios deducing the following trend; G•G > T•T = A•A > C•C.

1.4.3.1 Mismatches within base context

In addition to the revised thermodynamic contribution for a particular mismatch, the immediate base stacking environment surrounding the mismatch also has an impact on the thermodynamics of stability. Generally, within the examples shown in figure 1.4, a mismatch is more stable if the adjacent bases are either C or G, with A or T residues giving rise to a greater destabilising effect from the mismatch (Allawi *et al.,* 1997, 1998(a), 1998(b), 1998 (c), 1998(d); Peyret *et al.,* 1999). This is generally true for most of the mismatch contexts however some anomalies to this rule have been observed. Looking at the context of a C•T mismatch, it would be expected that a GC base pair would confer greater stability than AT, however a T•C mismatch residing 3′ to a GC base pair destabilised the oligonucleotide to a greater extent than corresponding AT stacking. This demonstrates that stacking interactions as a result of context take on just as an important role as hydrogen bonding in determining duplex stability (Allawi *et al.,* 1998(a)).

If sequences form offset duplex structures, resulting in overhangs or dangling ends, this too affects duplex stability by reducing the free energy ($\Delta G°$). This idea was first investigated by Doktycz *et al.,* (1990) when they considered the effect of dangling ends within the context of a hairpin loop. They showed that in all cases, a dangling end stabilised the hairpin duplex compared to a corresponding blunt ended control. For single base overhangs, 5′ dangling

ends were nearly always more stabilising or at least as stabilising as their 3' counterparts and the identity of the overhang itself gave rise to a hierarchy in terms of stabilising affect. Purine base (A or G) overhangs were more stabilising than pyrimidine base overhangs in the order (A=G>T>C), (Doktycz *et al.,* 1990; Bommarito *et al.,* 2000).

1.4.3.2 Mismatch type and consequences for amplification efficiency

Within a multiplexed system, unwanted primer–primer interactions will undoubtedly result especially since the primers are in vast excess over their target sequence in the initial reaction. During PCR amplification, it is desirable for these interactions to remain unchanged however if the stringency of the reaction is low or the melting temperature of the unwanted dimers is high, there may be some dimer extension. Huang *et al.,* (1992) specifically looked at the efficiency of *Taq* polymerase induced extension of all possible match and mismatch contexts. They found that transitional mismatches (A•C, C•A, G•T and T•G) were extended between 10^{-3} and 10^{-4} fold less than corresponding Watson Crick counterparts. Transversion mismatches ranged from 10^{-4} to 10^{-5} for T•C and T•T, to 10^{-6} for A•A, to $<10^{-6}$ for G•A, A•G, G•G and C•C. The transversion mismatch C•T extended 10^{-2} compared to AT. These results suggest that the observed trend in the nearest neighbour mismatch stabilities are:

G•A = A•G = G•G = C•C>A•A>T•T>T•C>C•T Huang *et al.,* (1992)

This trend is confirmed from more recent publications (Allawi *et al.,* 1997; Allawi *et al.,* 1998(a), 1998(b), 1998(c), 1998(d)).

Typically, for these nearest neighbour thermodynamics to be applicable, and give stability data that can be used to optimise multiplex development, the same sodium chloride concentration that is assumed for *Tm* calculation must be adopted in the PCR. Changes in salt concentration will change the duplex melting temperature as this is dependent on base

sequence, oligonucleotide length and sodium chloride concentration. Salts such as tetraethyl ammonium chloride (TEAC) and tetramethylammonium chloride (TMAC) alter the salt dependency giving rise to melting temperature data that behaves independently of sequence (Record *et al.,* 1978; Wood *et al.,* 1985; Jacobs *et al.,* 1988). For a given oligonucleotide duplex, in the presence of either of these salts, duplex stability will be dependent on length and not base sequence.

1.4.4 Melting temperature for Computer simulated duplex stability

Measurement of duplex stability in terms of melting temperature has more recently been applied to a number of computer programmes to facilitate efficient multiplex primer design. Some of these programmes require simple nucleotide sequence input from the operator. This enables operators the flexibility to design primers for the same multiplex with similar melting temperatures. Other programmes select candidate primer sequences according to a user specification, from bulk sequence data.

One of the first design software packages was described by Lowe *et al.,* (1990). This paper introduced a computer program to effectively select primer pairs from bulk sequence data according to a number of defined parameters. These included primer length (ideally between 18 and 22 base pairs), percentage GC content, GC clamps at the 3' end to stabilise primer binding, no primers to contain more than 4 identical contiguous bases, primer pairs to be separated by no more that 600 base pairs. Having found a suitable candidate primer pair that fulfilled the above criteria, the predicted melting temperature of the resulting amplicon was deduced from the equation:

Equation 12

$$Tm\ (^{o}C) = 81.5 + 16.6(\log_{10}[M]) + 0.41(\%G+C)-(500/l)$$

Schildkraut and Lifson, (1965)

If the amplicon *Tm* lies between 76 °C and 82 °C, the primers are accepted as a suitable candidate pair, otherwise if the *Tm* falls outside of this range, the primer sequences are initially rejected. This is to maximise the likelihood that all amplicons within the same multiplex undergo complete denaturation within the same temperature range. Although the program is designed to provide a faster means of primer design, the authors conclude that the *Tm* calculation is only a very rough approximation and should not be used to exclude possible primer sequence designs.

Chen *et al.,* (1997) describe a computer program known as 'dPrimer', which uses the thermodynamic parameters of Breslauer *et al.,* (1986) to predict primer *Tm*. This program allows for the input of either nucleotide sequence or amino acid sequence which is then used to cluster a series of degenerate primer sequences according to *Tm.* The accuracy of *Tm* prediction can show variations from actual primer *Tm* due to the inaccuracies of the Breslauer data discussed earlier.

Schultz *et al.,* (1999) describe the development of a similar computer based program for thermodynamic computer based prediction known a 'MeltCalc©'. MeltCalc© employs the unified set of thermodynamic parameters to predict melting temperature as described by Allawi *et al.,* (1997) to predict *Tm* in a similar fashion to Chen *et al.,* (1997). MeltCalc© provides estimates of primer *Tm* that are more accurate than those from dPrimer due to the improved thermodynamic parameters.

A computer program for the construction of suitable primers for PCR applications is described by Gorelenkov *et al.,* (2001). This is called VectorNTI®. The program permits the rapid selection of multiple PCR primers designed for multiplex PCR. As with previous simulations (Chen *et al.,* 1997), Gorelenkov and co-workers use the parameters of Breslauer *et al.,* (1986), to calculate *Tm.* Although the speed of operation is greatly enhanced over existing

software programs in terms of speed and operator defined primer selection criteria, it does not take into account primers which may form secondary structures. For this reason some primer designs may fail during PCR as a result of preferential primer-dimer formation.

In 1995, Applied Biosystems launched their primer design software package known as Primer Express™. The Primer Express™ software allows for the independent design and analysis of oligonucleotides for a variety of PCR and sequencing applications. Calculation of primer melting temperature, as a measure of duplex stability, is achieved using the nearest neighbour model. Primer Express™ is capable of performing some minor intra-molecular primer comparisons for suggested primer pairs as well as a simple intermolecular comparison of primer pairs for primer-dimer propensity. Although fairly accurate, the software cannot compare multiple primer sequences for primer-dimer likelihood although possible primer-dimer interactions can be highlighted from manually entered primer sequence pairs. As with previous software programs, the salt correction factor and nearest neighbour thermodynamics are based on the Breslauer *et al.,* (1986) data. This would suggest that *Tm* prediction may deviate slightly from experimental *Tm* due to the inaccuracies of this data set.

In 1999, SantaLucia and Peyret launched a computer simulation program known as HyTher™ which is used to predict DNA hybridisation thermodynamics. SantaLucia and co-workers are responsible for many of the advancements concerning the development of the nearest neighbour model for predicting duplex melting temperature. HyTher™ is available as an internet program which can be accessed according to the following link:-

http://ozone2.chem.wayne.edu/Hyther/hythermenu.html

The program comprises a series of modular programs that can interact together or separately to predict:

- duplex hybridisation thermodynamics,
- best primer sequences of a given length complementary to a given target,

- primer/target comparisons for determining secondary sequence homology stabilities.

The program uses the most recent up to date thermodynamic parameters as well as an improved salt correction model (SantaLucia *et al.,* 1996).

Matveeva *et al.,* (2003) used the improved thermodynamic parameters outlined in table 1.2, in their calculation of $\Delta G°$, within their program OligoAnal™. This program was specifically designed to consider the hybridisation efficiencies of probes designed to target the same RNA region, by reducing the likelihood of designing poor binding oligonucleotides. As well as OligoAnal™, a program known as mfold™ (SantaLucia, 1998), was used to assess the intra-molecular binding potential of candidate primer sequences derived from OligoAnal™. Although mfold™ performs fairly well, the authors concluded that a more up to date set of thermodynamic parameters would lead to significant improvement in mfold™ prediction performance. Matveea *et al.,* (2003) concluded that the program offered benefits over existing programs in terms of a more efficient oligonucleotide selection scheme.

In summary, the development of the nearest neighbour model for predicting oligonucleotide melting temperature has received much interest. This model has been incorporated into a variety of computer programs, to aid efficient primer design and has resulted in a number of research papers that continue to improve the accuracy of the individual thermodynamic parameters required for *Tm* prediction. Much of this data was determined from plots of absorbance vs. temperature melting curves in the presence of 1 M salt. Clearly for PCR applications, the amount of salt is much lower, hence a salt correction model was employed to correct the predicted *Tm* value. The literature suggests a number of factors for correcting salt concentration, all of which seem to be valid. However, choosing the right correction factor is paramount to obtaining reliable *Tm* predictions. In addition, an assessment of the likelihood of intermolecular interactions would clearly benefit multiplex primer design.

Whilst some programs offer some degree of intra-molecular assessment, (Primer express™, mfold™ and HyTher™), a program that offers complete intra / inter-molecular assessment is not yet available probably as a result of the vast number of interactions that could potentially result. This is likely to be the focus for future research and may result in a further refinement of the thermodynamic parameter values.

1.5 Microarrays for DNA analysis

1.5.1 Background

With the elucidation of the entire human genome sequence as a result of the efforts of the human genome mapping project, there is now a rapid need to understand the underlying primary sequence in context. The motivation for this program of work is to identify new targets, for therapeutic intervention and the identification of diagnostic markers associated with disease (Anderson *et al.,* 1998). Having established this primary level of structure, the next phase is to associate functions with newly discovered genes. In turn this should enable researchers to map genes to specific genome locations; to determine the extent to which gene expression changes, thereby affecting protein expression in disease and non diseased states as well as understanding how genes of different individuals differ. It has previously been suggested that there are between 10,000 and 100,000 genes in man (Schwarzacher, 1976) although current estimates suggest between 20,000 and 30,000. Rather than acting independently from one another, gene expression is thought to exist as a complex interrelated process. With seemingly vast number of genes and their interactions to consider, the classical gene by gene approach is not suitable as a means to understand function (Anderson *et al.,* 1998). Lander, (1999) suggests that the only way to fully characterise biological processes in terms of genes and their expression, is to take a 'global approach', looking at many factors simultaneously. In recent years, the most innovative technology suitable for such an approach is DNA arrays.

1.5.2 Early arrays

The entire field of DNA arrays has evolved from pioneering work by Ed Southern and co-workers. Southern, (1975) described for the first time the detection of specific DNA sequences among fragments of DNA previously separated out by gel electrophoresis.

Initial separations were conducted on agarose gel following restriction enzyme cleavage. Following alkali denaturation to render the fragments single stranded, they were transferred from the agarose gel to a sheet of nitrocellulose membrane by means of capillary action: denatured single stranded DNA fragments bind irreversibly to the nitrocellulose membrane (Southern, 1975). This process came to be known as Southern Blotting. The nitrocellulose sheet, now containing a replica of the original DNA gel is incubated under annealing conditions with a radioisotope labelled oligonucleotide probe. Specific hybridisation of the probe to target sequences on the membrane was detected by autoradiography. A hybridisation binding event of this type is the basic principle of any array application (Bier *et al.,* 2001).

Southern blotting has been used in combination with RFLP analysis by a number or groups (Kafatos *et al.,* 1979; Rozen *et al.,* 1986; Martin *et al.,* 1987) to identify human genes associated with specific disease. This was soon followed by filter-based screening of clone libraries, which introduced a one-to-one correspondence between clones and hybridisation signals (Grunstein *et al.,* 1975).

The use of radioisotopes to label nucleic acid probes is problematical, mainly due to the wide scatter of radioactive disintegrations following probe hybridisation, reducing the sensitivity of the method (Connor *et al.,* 1984). To improve sensitivity, radioisotope labels were replaced with biotin labels, detected via colourimetric means such as that described by Leary *et al.,* (1983). In this example, biotin-labelled DNA probes were hybridized to DNA immobilized on nitrocellulose filters. After removal of residual probe, the filters were incubated with avidin-DH (or streptavidin) and biotinylated polymers of intestinal alkaline phosphatase. The filters were then incubated with a mixture of 5-bromo-4-chloro-3-indolyl phosphate and nitro blue tetrazolium, which resulted in the deposition of a purple precipitate at the sites of hybridization (Leary *et al.,* 1983).

This latter approach has been used extensively as a means of isolating species specific sequences to act as probes for bacterial pathogen identification (Picken *et al.,* 1988). Similar

techniques were used to explore expression analysis by hybridising mRNA to cDNA libraries on gridded nylon membranes. However, these techniques proved difficult to use (Lander, 1999). As methods of 'dot-blotting' became more automated and miniaturised, hybridisation in the form of dot-blots, was seen as a technology that could be used on a large scale to exploit the data emerging from the genome project (Lennon *et al.,* 1991). More recently, nylon membranes and dot-blot assays have also been successfully applied to mutation detection assays such as that described by Kerezturya *et al.,* (2002) to detect HLA polymorphisms for paternity testing.

Apart from dot-blot approaches, further interest in the development of DNA array systems was assured by two key innovations; the use of non-porous supports such as glass and the development of methods for high density spatial synthesis of oligonucleotides.

1.5.3 Non-porous solid support DNA arrays

The main distinction between dot-blots and DNA arrays is in the use of an impermeable substrate or solid support. Typically, this is glass which has a number of advantages over porous membranes. As liquid cannot penetrate the surface of the support, nucleic acid targets have immediate access to the tethered probe without diffusing into pores. Faster diffusion results in an enhanced rate of hybridisation (Southern *et al.,* 1999). Flatness, rigidity and transparency of glass supports all give rise to improved data acquisition, as the location of the probes are much better defined. Glass with its planar structure however does limit the loading capacity (density) of the array spot which can affect hybridisation sensitivity (Beier *et al.,* 1999), although the amount of material lost from the array during processing can be minimised as a result of surface modifications to promote covalent attachment. It was also noticed that glass substrates did not significantly contribute to background noise due to its low intrinsic fluorescence (Cheug *et al.,* 1999). This factor alone improved the sensitivity of assays compared to nitrocellulose membranes.

1.5.4 Array fabrication

As emphasis moved away from standard dot-blot assays, interest mounted in the generation of synthetic array platforms containing specific oligonucleotide probes. For the most part, nomenclature with regards to these arrays follows on from that described for dot-blot assays. Where a standard dot-blot assay refers to spotting of DNA template (PCR product, restriction digest etc.) followed by the detection of the template sequence using labelled probes, reverse dot-blots, as the name implies, is the exact reverse. Here probes (oligonucleotides) are sited on the solid support and then detected with labelled DNA template (Hacia *et al.,* 1998).

Reverse dot-blots are manufactured in one of two distinct ways, *in-situ* synthesis directly on the support surface or deposition of pre-synthesised oligonucleotides (Proudnikov *et al.,* 1998; Southern *et al.,* 1999).

1.5.4.1 In-situ synthesis for array fabrication

There are two recognised approaches for *in-situ* synthesis of oligonucleotides. The first involves depositing phosphoramidite monomers directly to the surface of the support by means of an inkjet (Blanchard *et al.,* 1996; Ermanntraut *et al.,* 1997). This technique has been successfully used to construct 'random access' arrays, that is, the oligonucleotide in any one position can have any chosen sequence.

The second method for generating *in-situ* synthetic arrays relies on a photolithographic process first developed by Fodor *et al.,* (1991), sometimes referred to as a photochemical deprotection method (Southern *et al.,* 1992; Fodor *et al.,* 1993; Southern *et al.,* 1996; Lipshutz *et al.,* 1999). In this approach, a glass substrate is modified with a silane reagent to

provide hydroxyalkyl active groups containing photo labile protective linkers. Upon exposure of light, through a mask, exposed regions are selectively deprotected, producing an activated site. A specific phosphoramidite monomer containing the same 5' protected group (5'-(o-methyl-6-nitropiperonyloxycarbonyl), is then washed over the array surface. Activated sites couple these protected monomers producing the first layer. The process is then repeated, cycling between photodeprotection and phosphoramidite extension, activating different sites and adding different monomers to create the desired oligonucleotide sequence. Similarly this technique has been used to construct 'random access' arrays (Fodor *et al.,* 1993; McGall *et al.,* 1997; Lipshultz *et al.,* 1999).

An extension to the photolithographic technique was described by Southern *et al.,* (1992), Maskos *et al.,* (1993). This involved localising synthesis to specific regions by means of physical barriers such as masks to localise the addition of phosphoramidite monomers. If the surface of the array is flooded through orthogonally intersecting channels, this allows for more precision and specificity as well as the ability to make arrays of a chosen length. Both circular and diagonal shaped channels have been used to make scanning or 'tilling' arrays as shown in figure 1.10.

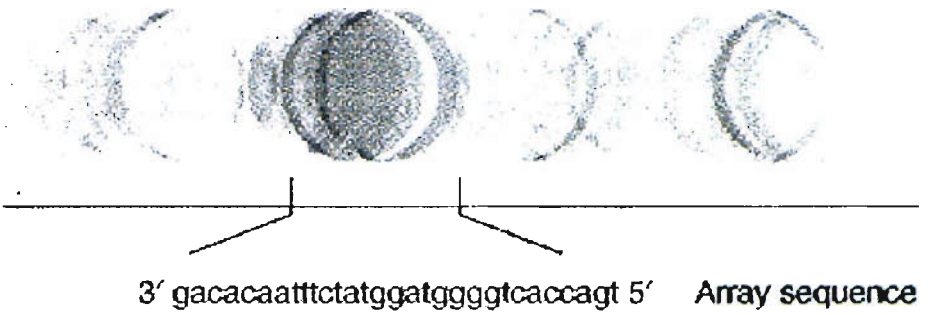


Figure 1.10 An example of a circular tiling array (Southern *et al.,* 1999). The array represents a tiling path of oligonucleotides complementary to a target sequence. Successive positions along the array, which step through the sequence are indicated by an arc shape. Areas of hybridised ^{32}P-labelled sequence are revealed by autoradiography with the darkest area indicating regions of strongest hybridisation signal. Arcs facing left (, denote sequences with matching 5′ ends whereas arcs facing right), denote sequences with matching 3′ ends.

This type of array has been useful for studies looking at hybridisation behaviour (Southern *et al.,* 1999) as well as for the selection of candidate oligonucleotides for sequence specific mutation detection.

Both *in-situ* methods synthesise oligonucleotides in the 3′ to 5′ orientation. There are however alternative ways of achieving *in-situ* synthesis by using different photoprotective groups such as those which are described by McGall and Fidanza (2001). These can change the orientation of synthesis from standard 3′-5′ to 5′-3′. These types of arrays are known specifically as 'reverse' arrays.

Overall, these later examples of *in-situ* synthesis methods have shown themselves to exhibit a high degree of efficiency, approaching 90-95%, due mainly to the efficiency of the photodeprotection method (Heller *et al.,* 2002), and it is this fact that has ensured its continued use as an array fabrication technique.

1.5.4.2 Array fabrication by deposition of pre-synthesised oligonucleotides.

Rather than define an array from the *in-situ* synthesis of oligonucleotides, much interest has been given to the deposition of pre-synthesised oligonucleotides on to a solid support. This involves delivering a small volume of aqueous material (nucleic acid), directly to the support surface (spotting).

The technology for spotting DNA material onto a substrate falls into two categories: non-contact spotting and contact spotting (Rose, 2000).

Non-contact dispensing involves the ejection of material from a delivery capillary to a solid support without contact. One method of achieving this uses technology similar to that used in inkjet printers. Typically, piezoelectric printing uses a type of electrically sensitive ceramic material closely apposed to a sample reservoir. When a voltage is applied to the piezoelectric material, it causes the ceramic to deform, squeeze the reservoir, and displace a small volume of sample from the tip (Rose, 2000). Alternatively, syringe-solenoid printing technologies have been used (Lemmo *et al.,* 1998). This combines a syringe pump with a microsolenoid valve to provide quantitative dispensing. In either case, the array develops as a result of small droplets of sample striking the solid support under gravity.

Contact dispensing on the other hand involves the ejection of material from a delivery capillary to a solid support as a result of physical contact. This technology is generally regarded as pin printing and usually involves the use of rigid pin tools. The pins are dipped into the sample solution, resulting in the transfer of a small volume of material onto the tip of the pin. Touching the pins onto a solid surface deposits some of this material in the form of a spot, the diameter of which is determined by the support's surface properties, pin properties and carrier fluid (Rose, 2000). An adaptation of the pin printing process is the Pin-and-Ring™ technique developed by Genetic MicroSystems. This involves dipping a small ring into the sample and then removing it to capture liquid. A solid metal pin is then pushed

through the centre of the ring trapping a small amount of the sample. As the pin contacts the solid surface the sample trapped on its underside is released.

In general, contact spotting is a much more simple technology than non-contact methods of delivery (Rose, 2000). Contact printing can deliver smaller sample volumes requiring less initial sample as well as generate arrays that are more densely packed, allowing more spots to be contained within a given area. For this reason, non-contact methods have not achieved the same level of interest as contact methods and so have not developed at the same pace. As a result, before the advent of commercially available microarray instrumentation, crude spotters that relied on contact printing were developed. An example of this is described by Graves *et al.,* (1998), whose machine used a single droplet tip, spotting up to 96 spots per slide and 32 slides in total. This gives an array density of approximately 400 spots/cm^2.

More recently, commercially available array spotters have followed this trend, adopting the contact spotting method for array fabrication. An example of this is the Generation III spotter by Molecular Dynamics which uses pin printing technology.

1.5.5 Attachment Chemistry

The efficiency of any array application is controlled in part by the level of interaction between the spotted sample and the solid support. This interaction usually involves the creation of some form of bridging complex or bond to link normally inactive species together (Chrisey *et al.,* 1996). The type and or strength of interaction will ultimately determine the level of sensitivity that can be achieved for the assay. For the majority of cases, glass is used as the solid support and so typically the types of interaction that can result involve either non-covalent or covalent interactions. Glass is a preferred substrate since its surface can be easily modified by silane chemistry. This can be used to introduce specific functions such as amino groups, epoxide, carboxylic acid and aldehydes (Zammatteo *et al.,* 2000).

Non covalent methods and hydrophobic interactions result from thin film deposition onto the solid support prior to spotting. Schena *et al.,* (1995) describes the use of polylysine to promote DNA binding. Although fairly efficient, these can strip array from the surface under extremes of temperature and salt concentration thus removing the spotted sample from the array surface (Allemand *et al.,* 1997). For this reason covalent binding methods are preferred.

Covalent attachment of DNA to a solid support is mediated primarily through the use of ultraviolet radiation or high temperature baking methods although these processes can give rise to ionic bond formation also (Benters *et al.,* 2002). Both methods provide the energy required to form a covalent bond between the solid support, directly, or via an intermediary such as a linker molecule (Zammatteo *et al.,* 2000). For solid supports that are functionalised with positively charged amino groups, covalent bond formation can occur between thymidine residues in the DNA chain. This is not a particularly well defined covalent attachment strategy, since it is governed by oligonucleotide base composition combined with fixation conditions. Each of these affect the way the DNA is bound to the support so that the length and sequence available for subsequent hybridisations can vary. An alternative and more predictable method involves tethering the DNA molecules at their extremities by introducing a functionalised group (Chrisey *et al.,* 1996). There are a number of different tethering chemistries that have been demonstrated. These include carboxylated (Joos *et al.,* 1997) or phosphorylated DNA molecules (Rasmussen *et al.,* 1991), which can be coupled to aminated supports and amino terminal oligonucleotides which can bind to isothiocyanate-activated supports (Guo *et al.,* 1994), aldehyde-activated glass (Schena *et al.,* 1996) or to solid supports modified with epoxides (Lamture *et al.,* 1994). Lindroos *et al.,* (2001) undertook a study which looked at various attachment chemistries. This revealed stark differences between the different approaches in terms of robustness and efficiency. Lindroos concluded that their best results were observed when disulfide-modified oligonucleotides were attached

to mercaptosilane-coated slides. Beatie *et al.,* (1995) demonstrated a simple approach by preparing simple terminal amine derived 9-mer sequences for attachment to a glass support containing epoxysilane active groups. This type of interaction is considered extremely stable as demonstrated by Jung *et al.,* (2001), when they covalently immobilised pre-assembled amino-linked oligonucleotides to heated silanised glass slides via an ink jet printing method. Jung reported that the hybridisation capacity was still achievable after 150 repeated regenerations of the surface by acid treatment and denaturing agents.

In addition to the routine attachment methods, combinatorial approaches have been demonstrated to enhance DNA attachment. Heterobifunctional crosslinking molecules which bear both amine and thiol reactive moieties have been used to link oligonucleotides to glass slides (Chrisey *et al.,* 1996). In a different approach, Lee *et al.,* (2002) demonstrate the use of a polyethylenimine coating to generate a 3-dimentional surface which enhances DNA attachment. In this example binding is mediated through both covalent and non-covalent means.

1.5.6 Target density and Hybridisation efficiency

Achieving high attachment efficiency is clearly desirable, however this does bring into play an important factor; spot/target density. Spot density is a measure of the closeness or density of target material contained within a single array feature (spot) and is governed by a number of factors including the ionic salt concentration, interfacial electrostatic potential between the array surface and target and duplex characterisation (single stranded or double stranded) (Peterson *et al.,* 2001). In addition, sample concentration will affect the density of material in any given spot. Cheung *et al.,* (1999) calculated the spot volume to be:

Equation 13

$$\text{Spot Volume } (\mu L) = (\tfrac{1}{2} \times 4/3\ \pi\ r^3) \qquad \text{(Cheung } et\ al.,\ 1999)$$

Equation 14

Total DNA (ng) = DNA concentration (ng/μL) x Spot volume (μL)

(Cheung *et al.,* 1999)

For example, if DNA concentration was 500 ng/μL and spot radius was 50 μm, then the total spot volume would be 0.00026 μL (0.26 nL), according to equation 13. This equation assumes that spot deposition will produce a perfect hemisphere although the exact shape of the hemisphere will be governed by the spot deposition rate. If this is too fast, spot size will be larger and not necessarily round due to fast deceleration and acceleration of the pen towards and away from the slide, drawing excess liquid towards the spot. Assuming a perfect hemisphere, the spot would contain approximately 0.13 ng DNA, according to equation 14. The small spot volume means that the amount of sample for hybridisation is also very small, even if the initial sample concentration is relatively high. Cheung *et al.,* (1999) conclude that this limitation should be regarded as of paramount importance when considering hybridisation efficiency, as additional factors will also act to further reduce the amount of available template for hybridisation / signal intensity. These include the sensitivity of the detection method employed following hybridisation, the specific length and activity of the targeted region, as well as the proportion of probe DNA that is complementary to the target region within each DNA strand. This later point is particularly relevant in the context of array density as it brings into question the notion of steric hindrance.

Steric hindrance is particularly problematic for arrays that have been generated by *in-situ* synthesis methods. Bases nearest the surface are less accessible than those furthest away and in addition, the packing of oligonucleotides is usually high, such that steric crowding occurs between adjacent oligonucleotides (Southern *et al.,* 1999). Steric crowding is a physical condition which limits access between the probe and its target sequence, reducing the number of potential hybridisation sites. To combat this effect, the use of spacer

molecules have been investigated (Beattie *et al.,* 1995; Shchepinov *et al.,* 1997; Southern *et al.,* 1999).

Beatie *et al.,* (1995) previously showed that the use of triethylene glycol spacers did not enhance the hybridisation efficiency for tethered oligonucleotides. Given the findings of Cheung *et al.,* (1999), this was probably the result of there not being a sufficiently dense array due to the limited amount of template contained within each spot volume. In contrast, Shchepinov *et al.,* (1997) looked at the effects of spacer molecules on hybridisation behaviour and demonstrated that their inclusion did in fact enhance hybridisation yields. Southern *et al.,* (1999) also found this to be the case when they demonstrated 2 orders of magnitude increase in hybridisation efficiency for oligos attached via ethylene glycol spacers compared to oligonucleotides that were attached directly. Southern did, however, notice a length limitation for the spacer unit which, if exceeded, resulted in a reduced hybridisation signal. They concluded that this was probably due to the oligonucleotides 'dissolving' in the linker, becoming less accessible to the target, thus reducing hybridisation efficiency.

These conflicting observations reflect the efficiencies of the two different array fabrication methods whereby *in-situ* synthesis gives rise to very dense, closely packed arrays. Clearly, the presence of spacer molecules in these circumstances aid to enhance hybridisation efficiency. Deposition methods generally give rise to less dense arrays and so hybridisation efficiency is not affected by steric hindrance to the same extent (Southern *et al.,* 1999), but more so by the factors outlined earlier.

In an effort to enhance the density of template within a given array feature, various groups have focused their research on improving the deposition and retainment of material on the solid support by adopting various surface enhancement treatments. Benters *et al.,* (2001) described the use of a polyamidoamine dendritic linker system as a mediator moiety specifically for DNA arrays. These dendrimers contain 64 primary amino groups which direct the binding of DNA molecules. Benters *et al.,* (2001) showed that for fluorescently labelled

oligonucleotides, the surface capture dendrimers displayed an approximately 2 fold greater surface coverage over conventional linear linkers resulting in a much denser cluster of oligonucleotides within a spot. Similar surface modifications have been demonstrated by Beier *et al.,* (1999) with their use of dendrimer linkers.

Clearly, target density is an important factor in determining hybridisation efficiency but having established a suitable array, there are other factors which govern hybridisation efficiency. The first of these is target length. Lipshutz *et al.,* (1999) showed that the efficiency of hybridisation was increased if the probe and target were of equivalent sizes (lengths). Larger targets resulted in reduced hybridisation efficiencies due to a reduction in the efficiency of interaction with the probe. If target and probe are similar in size, hybridisation efficiency can also be affected by base composition particularly for random access arrays where a broad range of thermodynamics might be expected due to the large number of target sequences. These thermodynamic differences are due to differences in the number of AT versus GC base pairs throughout the array sequence. AT rich targets will frequently display lower hybridisation intensities than do more stable sequences with high GC content (Heller *et al.,* 2002). For hybridisations that require high levels of stringency, perhaps to break-up secondary structure or primer-dimer, or where an increased level of discrimination is required, this can present a challenge. In these circumstances, base analogues such as those described by Gryaztiov *et al.,* (1994), or TMAC (section 1.4.3.2), can be used to negate the effect of differences in duplex melting temperature. Gryaztiov *et al.,* (1994) successfully increased the stability of A-rich sequence by substituting deoxyadenosine with 2'-deoxy-2,6-diaminopurine. In addition, factors such as salt concentration and formamide concentration will affect the level of stringency and thus change the hybridisation efficiency characteristics. Specific reference to these is made in section 1.4.

1.5.7 Microarray Analysis

DNA hybridisation analysis within the context of a microarray usually involves detecting the signal generated by the binding of a reporter probe to target DNA sequence. This reporter may be fluorescent, chemiluminescent, colorimetric or radioisotopic (Heller *et al.,* 2002). Depending on the reporter molecule determines what analysis can be performed. Historically, radioisotope labelled DNA probes were detected by autoradiography (Southern *et al.,* 1975). As radioisotopes were replaced by fluorescent probes, detection of the hybridisation event is now monitored by fluorescent imaging.

Fluorescent imaging usually falls into two broad categories which were considered by Hacia *et al.,* (1999). These are the gain of hybridisation and the loss of hybridisation. Gain in hybridisation is the classical approach for microarray detection and is the process whereby a labelled probe binds to an unlabelled template or vice versa. This produces a gain in signal following hybridisation. Loss on the other hand involves the use of two labelled probes/templates. One acts as an internal control, the other as the test. These bind to their respective targets only if the test is positive, else only the control probe binds if the test is negative. The presence/absence of test signal in the presence of a control signal indicates the test result.

The degree to which microarray data can be interpreted is dependent on a number of important factors namely, the surface chemistry, microarray printing method, labelling method and hybridisation efficiency. Some applications rely on the assessment of transcript levels calculated from ratios of one dye to another by targeting cDNA (Bilban *et al.,* 2000; Young, 2000), however it is important to demonstrate that different ratios are a true reflection of changes in expression rather than as a result of systematic changes in physical/chemical dye characteristics (Bilban *et al.,* 2002). Normalisation of signal occurs through the use of invariant genes, that is genes whose expression changes in a known way. The expression of 'housekeeping genes', as they are known, gives a true reflection of what is

happening in terms of microarray efficiency although Bilban *et al.,* (2002) showed that choosing the correct housekeeping genes was important. They noted that expression levels for some housekeeping genes could fluctuate depending on the biological pathway involved, concluding that a normalisation strategy should be routinely adopted for applications such as these.

With the huge amount of data that usually results from microarray applications, emphasis has been given to developing suitable software for automated analysis. Jain *et al.,* (2002) describes the programme UCSFSpot™ which is a software programme developed for automatic data quantification. The programme locates and identifies each sub-array/spot in a field, then calculates ratios of fluorescence based on explicit spot segmentation (number of pixels within each spot). The software has been used routinely to measure expression levels and DNA copy number in biological samples (Jain *et al.,* 2002).

ArrayVision™ developed by Imaging Research is another gene expression array analysis software package. It is used as a rapid, automated package for the analysis of array images, for batch processing, and is capable of performing comprehensive background and normalization corrections. Although other types of microarray data analysis tools have been developed, ArrayVision™ is the analysis package used in this investigation.

1.5.8 DNA Microarray applications

Over recent years, there has been a significant increase in the number of papers describing a broad range of microarray applications. Microarray analysis has been used in a number of different fields including pharmacogenomic research, infectious disease/pathogen diagnostics, genetic disease and cancer diagnostics as well as forensic and identification diagnostics (Heller *et al.,* 2002). Within these disciplines, a number of different microarray analysis approaches have been developed. Essentially, these can be split into two major applications:

(1) Determination of expression level (abundance) of genes; and

(2) Identification of sequence (gene / gene mutations).

The basic principle to gene expression analysis is to determine the relative abundance of genes within a common sample by monitoring their hybridisation behaviour against an array of target cDNA molecules. This type of approach has been considered by many groups of which a good example was demonstrated by Schena *et al.,* (1996). In their approach an array was generated containing over 1000 cDNAs of unknown sequence. These were then probed against cognate human genes under defined experimental conditions to reveal differential expression levels. Where different expression levels were detected, these targets were then characterised by sequencing resulting in the discovery of novel genes associated with human T-cells.

This type of application has shown considerable potential as a means of identifying novel genes (Schena *et al.,* 1998) and has been used in a variety of cases including:

- Gene expression changes in diseased states. Graveel *et al.,* (2001) used microarray expression profiling to identify candidate genes associated with a particular liver carcinoma and Monni *et al.,* (2001) used the same technology to determine the copy number of a sequence associated with breast cancer.

- Gene expression analysis for drug discovery, metabolism and toxicology (Service *et al.,* 1998). Gerhold *et al.,* (2001) investigated the global expression levels of genes known to be active in drug metabolism for a variety of novel drug compounds.

- Gene expression analysis for microbiological and infectious disease (Heller *et al.,* 2002). Mehrotra *et al.,* (2001) uses this technology to identify regulatory genes associated with the bacterium Mycobacterium tuberculosis. This micro-organism can survive in a diverse range of physical/environmental conditions as a result of changes

in the expression of genes during its life cycle. Mehrotra *et al.,* (2001) demonstrated through the use of expression profiling, the identification of genes (or virulence determinants) specifically expressed under these varied conditions. Similarly, Chizhikov *et al.,* (2001) used a similar approach for expression profiling of virulence factors for other microbial species.

The second major use for microarray technology is as a discovery platform for functional genomics. This involves the interrogation of the genome for the identification/genotyping of mutations (Schena *et al.,* 1998). Mutation detection falls into two categories; microarray analysis for STRs and microarray analysis for single point mutations or SNPs.

For forensic identification purposes, STRs offer a high level of discrimination between individuals and it is therefore advantageous to have a microarray platform that specifically characterises STRs. The principle to STR microarray detection is shown in figure 1.11. It involves the deposition of a series of oligonucleotide templates, each one differing by one STR repeat unit (figure 1.11a). An unknown sample is then amplified using dye labelled primers which amplify the STR region of interest. This sample is then allowed to hybridise to the array, binding fairly non-specifically to the oligonucleotide series (figure 1.11b). Through a number of stringency washes, oligonucleotides that do not form complete matches, due to overhangs/offset hybridisation, will be removed from the array surface. Oligonucleotides with the same number of repeat units as the sample and which have bound the target without offset will remain bound following stringency washing revealing the STR unit length (figure 1.11c).

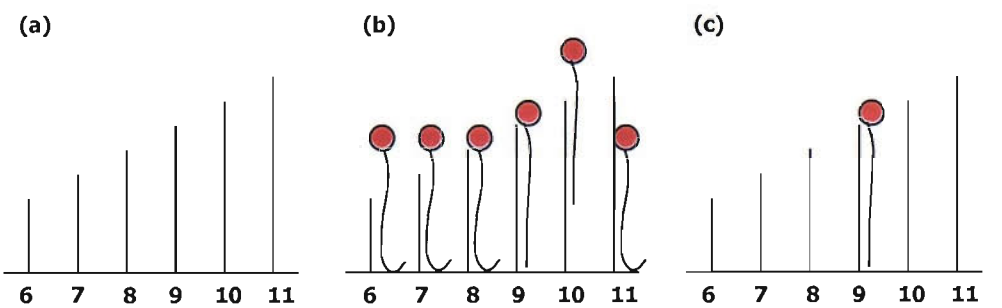


Figure 1.11 (a) An Oligo array with unit lengths between 6 and 11 units; (b) Dye labelled template bound non-specifically and (c) A sample STR unit (9) following stringency washing.

The major limitation to this type of array was outlined by Hacia *et al.,* (1999). Passive arrays such as this where the target is allowed to passively anneal to the probes required an extremely stringent set of conditions to maximise specificity. Hacia found these conditions very difficult to determine resulting in incorrect STR genotypes. Radtkey *et al.,* (2000) also showed this but demonstrated that the use of electronic arrays, where a charge bias can be applied to any of the array features (spots), increased specificity. Radtkey concludes that this type of array is more suitable for STR analysis.

The use of microarray platforms for the detection of SNPs has received considerable interest over recent years and has led to a number of different detection strategies. Historically, SNPs were detected and characterised from gel base sequencing of a number of individuals (Collins *et al.,* 1997), however more recently this has been superseded by microarray technology.

Huber *et al.,* (2001), describe a method for sequencing using gene specific tilling arrays. These arrays contain gene specific oligonucleotides which interrogate sequential single nucleotide positions within the region of interest. Amplicons generated from standard PCR hybridise to the tilling array according to their sequence. The tethered probes are then extended by a polymerase enzyme in the presence of di-deoxynucleotides. Depending on which base has been incorporated at each stage in the array, determines the sequence of the unknown region.

Direct microarray SNP characterisation is reliant on there being some underlying knowledge of the surrounding sequence. SNP determination in these instances can be achieved through a variety of different methods including;

- Direct hybridisation. Fluorescent markers generated during PCR bind to probe arrays. Individual alleles are differentiated by two different probes in different locations on the array. The presence of a fluorescent signal at the site of these probes determines SNP/allele identity.

- Probe modification to produce fluorescence reporting. This is where the probe is modified to produce fluorescence following complementary target hybridisation. An example of this is referred to as allele specific extension.

The simplest way to conduct allele specific extension is to generate a reverse dot-blot array of probes directed to each of the SNPs under investigation. It is necessary for the oligonucleotides to be oriented in the 5′ to 3′ direction with the 5′ ends closest to the support (Dubiley *et al.,* 1999). The 3′ base of each oligonucleotide should be complementary to the 5′ base adjacent to the SNP of interest in the target sequence. Following hybridisation, the tethered oligonucleotides are extended by a single base (di-deoxynucleotide) complementary to the SNP. Following washing to remove the target amplicon, the SNP identity is revealed by fluorescent detection of the extended oligonucleotides. This approach is often referred to as Genetic Bit analysis (Nikiforov *et al.,* 1994).

This type of application has been employed by a number of groups (Pastinen *et al.,* 1997; Hacia *et al.,* 1999; Modin *et al.,* 2000 and Lareu *et al.,* 2003). If the approach is used as described, four colour detection is required to elucidate all SNP possibilities although careful experimental design can reduce this to two colours by targeting SNPs with the same bi-allelic bases.

Many groups have taken this principle and adapted it to SNP analysis, but in slightly different ways. Erdogan *et al.,* (2001) noted that the technique could be improved by enhancing the fluorescence intensity generated. In their approach, an array comprising groups of four identical capture probes each carrying polymorphic sites at their free 3′ base were generated. Template was then allowed to hybridise to the array. Duplexes that had a 3′ match with the template, complementary to the SNP, were then extended by a polymerase enzyme using dNTPs of which one was dye-labelled, according to the ARMS principle. This is outlined in section 1.5.9. By undertaking a series of elongation steps resulting in the incorporation of more dNTPs, Erdogan *et al.,* (2001) was able to demonstrate enhanced sensitivity.

The position of fluorescent signals corresponded to the probe which contained the complementary 3' base, thereby identifying the SNP in the original target sequence.

In a different approach, Divne *et al.,* (2003) used a combination of dye terminator extension and direct hybridisation. Their PCRs result in small amplicons which they allowed to anneal to oligonucleotides in a liquid medium. These oligonucleotides were then extended by specific di-deoxynucleotide incorporation, complementary to the SNP in the original amplicon. The extended oligonucleotides were then hybridised to predefined arrays containing sequences complementary to the extended oligonucleotides in a direct hybridisation assay. Position of the fluorescent signal identified the SNP locus whilst the colour determined SNP designation.

Rather than relying on fluorescent primer extension assays to identify SNPs, assays that rely on specific binding events to cause fluorescent generation/enhancement have been developed. These types of assay usually rely on FRET (fluorescence resonance energy transfer) to cause either fluorescent enhancement or fluorescent quenching as a means to determine SNP characterisation. Didenko, (2001) outlines the principle of FRET. Excitation is transferred from a donor to an acceptor fluorophore through dipole-dipole interaction without the emission of a photon. Donor fluorescence is therefore quenched whilst the acceptor is excited. The FRET principle can be applied to hybridisation, although originally the concept was first used with molecular beacons (Tyagi *et al.,* 1996).

Frutos *et al.,* (2002) described the use of molecular beacons as probes to detect SNPs in an array. For each SNP under investigation, Frutos designed two allele specific molecular beacons, each containing a donor fluorophore. Complementary invader oligonucleotides containing an abasic residue at the SNP position were then designed and labelled with an acceptor (quencher) molecule. Unlabelled PCR product spanning the SNP region was then

allowed to hybridise to the molecular beacon probes in the presence of the invader targets. Duplexes formed between the invader oligo and beacon brings the two fluorophores into close proximity, resulting in quenching and no fluorescent signal. Duplexes formed between unlabelled PCR product and beacon prevents the invader oligo from binding and so prevents FRET from occurring. In this instance fluorescence is maintained. The combination of fluorescent signals, between the two molecular beacon sites, then determines the SNP genotype.

Each of the preceding examples relies on a common event. That is the hybridisation of a target to the probe. Ignoring how the fluorescent signal may be developed, (either automatically as a result of direct hybridisation of dye-labelled target, or as a result of a reporter probe modification), the specificity of the reaction is dependent entirely on the specificity of the hybridisation event. For SNP discrimination, attempting to differentiate one allele from the other can be problematic due to the small difference in sequence, affecting a single base. This is certainly true for direct hybridisation methods, however the discrimination of alleles can been enhanced by adopting the ARMS methodology combined with the URP principle previously outlined in section 1.1.4.

1.5.9 URP principle for Microarray SNP detection

As previously outlined, the URP principal was designed to co-amplify multiple loci using the same three universal sequences. These are initially combined with locus specific sequences to determine where amplification occurs. For microarray applications, two detection strategies can be employed;

1.5.9.1 URP for indirect SNP Detection (Indirect method)

The indirect method relies on the hybridisation of dye labelled probes to the universal segments of the amplicons, rather than to the SNP site it self. A reverse probe targets the locus specific region, identifying the locus, whilst allele specific probes target the forward universal/locus specific complex. Forward probes are typically 20 bases in length; 10 bases complementary to the locus specific region and 10 bases specific to one of the universal sequences. This configuration makes each forward probe specific to a particular allele.

Figure 1.12 Hybridisation patterns between two allele specific probes (labelled in green) hybridising to heterozygote sample on two different slides (1+2). Where probe sequences match the target sequence through their entire length (1b/2a) probes remain hybridised. Where sequences show mismatch (1a/2b) these probes detach during stringency washing. Locus specific probes (labelled in red) indicate the locus identity. The combination of signals between the two slides determines the SNP genotype. Hussain *et al., (*2003).

In each case the probes do not target the SNP site directly but target the universal region, hence the name indirect method.

1.5.9.2 URP for Direct SNP Detection (Direct method)

The direct approach does not rely on the ARMS principle to identify the SNP under investigation but instead relies on the assessment of the SNP by direct hybridisation to it.

The URP methodology is used as this maximises the possibility of a balanced amplification between different loci however, instead of using three universal sequences, the direct approach uses just two, universal 2 and universal 3. For detection, three different probes are used; the reverse probe is identical to the reverse probe used in the indirect method, where as the forward probes target the SNP site rather than the forward universal sequence. They differ from each other only by the base corresponding to the specific allele they are designed to detect.

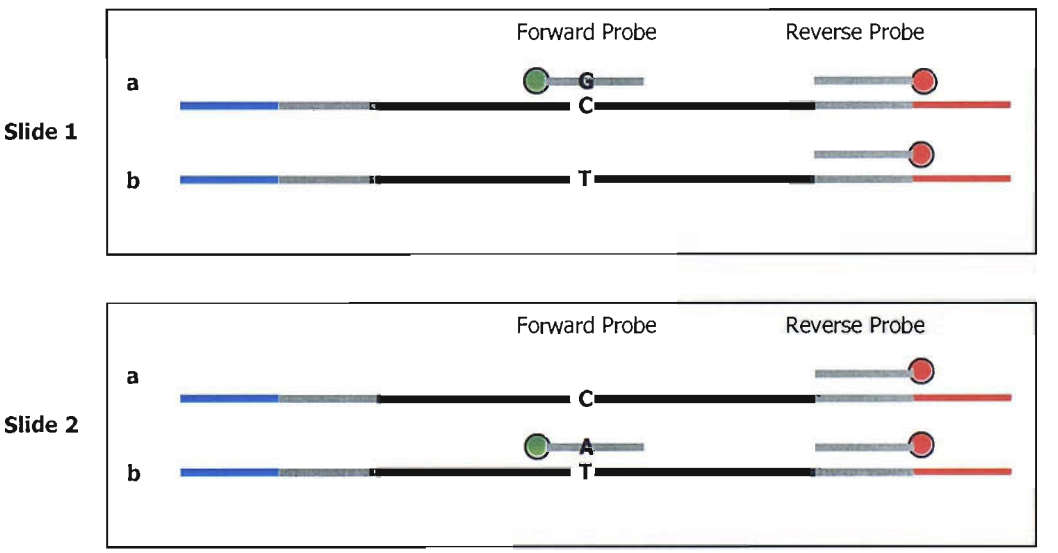


Figure 1.13 A schematic of the direct hybridisation method for two different forward probes (labelled in green) on two different slides (1+2). Only probe sequences that are complementary through their entire length (1a/2b) remain annealed. Sequences with a mismatch (as would be the case in 1b/2a), do not remain hybridised. As before, locus specific probes (labelled in red) indicate the locus. The combination of signals between the two slides determines the SNP genotype.

In all case the probes target their SNP site directly, hence the name the direct method.

The application of microarray analysis to multiplex detection has not been fully investigated especially within the context of forensic identification. An efficient amplification strategy such as URP/ARMS, enables rapid and efficient multiplex design. This permits the use of SNP markers and thus makes the technique more amenable to degraded material, however, multiplex detection of this kind on a microarray platform has yet to be realised.

2 Aims

The long term overall concept is to develop a total analysis system that will be capable of sampling crude material, extracting and retaining any DNA contained within it, amplifying the DNA using a novel, efficient biochemistry and ultimately detecting the products . The integration of this system is paramount to the successful delivery of such a system however much research is still required. For this reason, the aim of this work was to consider three specific elements of the overall concept. These are:

- The development of a suitable miniaturised DNA extraction device;
- The development of an efficient duplex stability assessment algorithm to determine accurate duplex melting temperature for use in the design of novel multiplex strategies;
- The development of microarray technology for application to multiplex DNA analysis in forensics.

Each of these elements may or may not feature in the final design however there is a requirement to assess their suitability and application to the overall concept.

3 Materials and Methods

3.1 Microfabrication Assays

3.1.1 Microfabricated Silicon Chips

Microfabricated silicon devices are manufactured and supplied to the Forensic Science Service® by the Microsystems and Microengineering (MEMS) fabrication laboratory at QinetiQ (formally Defence Evaluation Research Agency, DERA) at Malvern. The entire fabrication process is outlined in a paper by Cox *et al.,* (In preparation). Essentially the process begins with a p-type silicon wafer (resistivity 1-10 ohms/cm). The masking layer is defined in a 7 μm thick photoresist layer. A serpentine pattern in then etched into the silicon using an STS deep dry etching machine. The process cycles between first depositing a thin layer of polymer followed then by silicon etching. The polymer protects the walls of the developing structure and permits deep channels to be etched into the silicon structure. This type of manufacture allows for structure with a relatively high aspect ratio of between 2:1 and 50:1 (see figure 3.1).

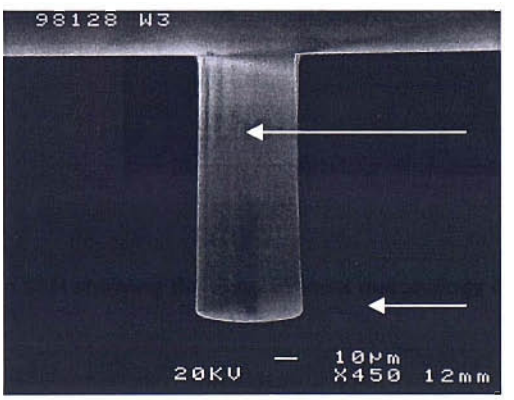


Figure 3.1 An SEM showing the deep channel morphology obtained from a cross-section through a silicon wafer.

Channels are typically between 125 µm x 50 µm. The walls and floor of the silicon 'channel' are coated with a thin layer of a fluorocarbon polymer before the masking layer is removed from the top outer surface. The silicon wafer is then anodically bonded to a 3 mm glass plate of Corning type 7740, thus creating a channel. The glass plate contains two 5 mm diameter drilled holes which permits access / egress to the ends of the channel. These act as inlet /outlet reservoirs. Channel lengths of 25 mm, 300 mm and 1000 mm were fabricated (see figure 3.2)[1]

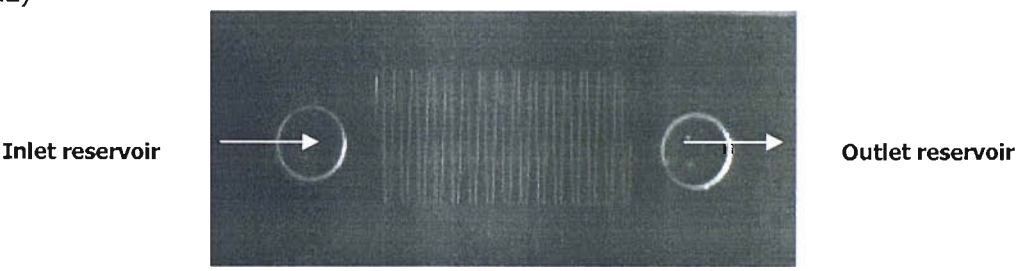


Figure 3.2 A 300 mm microfabricated silicon channel showing the inlet and outlet reservoirs.

3.1.2 DNA preparation

DNA extracts were initially prepared from standard liquid whole blood extracts. Routine DNA extraction was performed using commercially available Qiagen™ QiaAmp Mini DNA extraction kit (Cat. No. 51306). Stocks of DNA extract were quantified using a Picogreen™ quantification assay. Picogreen™ is an intercalating dye that undergoes fluorescent enhancement after binding to the minor grove of double stranded DNA. The amount of fluorescence is proportional to the amount of double stranded DNA present, as such, the assay involved comparing fluorescence attributed to the DNA extract, to fluorescence derived from a serially diluted commercially available Cambio™ DNA of known concentration. The DNA extract was then diluted to a stock of 1 ng/µL and stored at –20 °C until required for the investigations.

[1] Channels were microfabricated and supplied by Dr Tim Cox, QinetiQ.

3.1.3 DNA Extraction employing the Qiagen™ QiaAmp Mini DNA Extraction kit

For each sample, 1 µL liquid whole blood was placed into a tube containing 32 µL PBS, 4 µL Proteinase K and 32 µL AL buffer. The samples were mixed thoroughly and then incubated in a water-bath for 10 minutes at 70 °C. Following incubation, 32 µL ethanol (96-100%) was added to the sample and mixed. The sample was then decanted into an empty QIAamp spin column before being transferred to a centrifuge and spun for 1 minute at 6000 g. The spin column filter basket (containing retained DNA) was then transferred to a clean collection tube to which 80 µL AW1 was then added. The unit was again centrifuged for 1 minute at 6000 g. Once again the spin filter basket was transferred to a clean collection tube. 80 µL AW2 solution was added to the basket before being centrifuged for 3 minutes at 13000 g. The spin filter basket was subsequently transferred to a clean tube. The basket was loaded with 20 µL of 1x ABD TE at 70 °C or 20 µL SDW. Finally, the sample was centrifuged for 1 minute at 6000 g. The supernatant (DNA extract) then was then taken forward to PCR.

3.1.4 DNA Extraction employing the Silicon Channel

Many of the parameters used for DNA extraction were defined during the course of the investigation and therefore appear as generic references in the protocol to follow. These are clearly defined during each experiment. DNA extraction employing silicon channels is facilitated by using the wet chemistry available with the Qiagen™ QiaAmp Mini DNA extraction kits.

Sample introduction and incubation:

D µL of DNA extract was added to the inlet port of a silicon channel. A 5 psi gas overpressure was then applied to the inlet port to move the sample into the channel such as to fill the channel. The air pressure was then released and the sample was left to incubate inside the channel for **Y** minutes. A 10 psi gas overpressure was reapplied to the inlet port

and the sample was evacuated through the outlet port and discarded. The process of sample introduction and incubation was repeated **N** times in total.

Sample washing and channel drying:

30 µL AW2 wash buffer was loaded into the input reservoir. A 10 psi gas overpressure was applied to the inlet port to move the wash buffer through the channel into the outlet port. The solution was then removed was discarded.

Elution procedure:

20 µL 1x ABD TE (elution buffer) @ 70 °C **or** 20 µL 18.2 MΩ water @ 70 °C was introduced into the inlet reservoir. A 10psi gas overpressure was applied to the inlet reservoir to move the elution buffer into the channel so as to completely fill the channel. The channel structure was then placed into a humid hybridisation cabinet pre-set to 70 °C and left to incubate for 1 minute before being removed. A 10 psi overpressure was reapplied to the inlet port until the sample was fully evacuated from the channel. This was collected during evacuation, and stored for later analysis. The process of Eluting a DNA sample was repeated **Q** times.

3.1.5 DNA Amplification

All DNA amplifications were performed using the commercially available AMP*FL*STR® SGM *plus*™ kit. Table 3.1 shows the components of each PCR reaction whilst table 3.2 shows the amplification parameters employed for SGM*plus* amplification.

Component	Volume (µL)
AMP*FL*STR® SGM*plus* Hybridisation buffer	19.1
AMP*FL*STR® SGM*plus* Primer Mix	10.0
Ampli*Taq* GOLD (5 Units/µL)	0.9
DNA or eluted sample	20.0
TOTAL	**50.0**

Table 3.1 A list of the components required for SGM*plus* PCR amplification.

Process	Temperature	Time	Cycles
Activation	95 °C	11 minutes	1 cycle
Amplification	94 °C	60 seconds	28 cycles
	50 °C	60 seconds	
	72 °C	60 seconds	
Extension	60 °C	45 minutes	1 cycle
Hold	4 °C	Forever	

Table 3.2 SGM*plus* PCR amplification parameters.

Following PCR amplification samples were prepared for electrophoresis by preparing a 1:1 v/v ratio sample:dextran blue/formamide (minimum volume should be 3 µL). NB. Dextran blue/formamide (DB) contains a size standard, GeneScan 500XL™ ROX (Cat. No. 403039) prepared by mixing a 6:1 v/v DB:Sizestandard. Samples were denatured on a Perkin Elmer GeneAmp 9600 Thermal Cycler set at 95 °C for 2 minutes before being snap-cooled on ice in preparation for loading onto a polyacrylamide gel.

3.1.6 Polyacrylamide Gel Electrophoresis

2 µL of each sample was loaded onto a polyacrylamide gel (6% LongRanger™ Solution). All samples were run on a Perkin Elmer ABI 377 automated sequencer using the following parameters:

Plate check module:	Plate check F
Pre-run module:	GS PR 36F 2400
Run module:	GS Run 36F 2400
Well-to-read distance:	36 cm
Collect time:	2.5 hours

3.1.7 Analysis of Results

Results from polyacrylamide gels were collected using ABI Prism™ 377XL Collection™ software and analysed using ABI Prism™ Genescan™ and Genotyper™ software. Total peak areas for each test sample were compared against control samples to determine % recovery of DNA.

3.2 Oligonucleotide Melting Temperature Assays

3.2.1 Oligonucleotides

All oligonucleotides were synthesised by either IBA, Germany or the Department of Chemistry, Southampton University and supplied as either stock solutions or lyophilised powders. Lyophilised powders were reconstituted in SDW to give stock concentrations of primer. All primer stocks were then diluted in SDW to give final stock concentrations of 2 µM. These were stored at –20 °C until ready for use. A full list of oligonucleotide sequences is shown in section 6, appendix 6.1.1.

3.2.2 Oligonucleotide set-up for duplex melting temperature analysis

100 mM sodium phosphate buffer (pH 7.4) was prepared by mixing 77.4 mL 1 M di-sodium hydrogen phosphate (Na_2HPO_4) to 22.6 mL 1 M sodium di-hydrogen phosphate (NaH_2PO_4) and making up to 1 L with sterile distilled water. The resulting buffer was stored at room temperature until ready to use.

For experimental set-up, 0.5 µM forward oligonucleotide was mixed with 0.5 µM reverse oligonucleotide in the presence of 50 mM Sodium Chloride and 10 mM sodium phosphate buffer (total volume being 20 µL).

3.2.3 Duplex Melting Temperature (Tm) Analysis

All *Tm* analyses were carried out using the Roche LightCycler™ instrument. The instrument was configured with the run parameters shown in table 3.3. Figure 3.3 shows a diagrammatic representation.

Phase	Description	Target temperature	Temperature transition rate °C/s	Hold time	Data Acquisition
1	Equilibrate	30 °C	0.0	3 minutes	No
2	Melt 1	95 °C	0.1	0 minutes	Continuous
	Hold 1	95 °C	0.0	3 minutes	No
3	Anneal	30 °C	0.1	0 minutes	Continuous
	Hold 2	30 °C	0.0	3 minutes	No
4	Melt 2	95 °C	0.1	O minutes	Continuous
5	End	35 °C	20.0	1 minute	No

Table 3.3 Duplex melting temperature parameters for duplex *Tm* analysis.

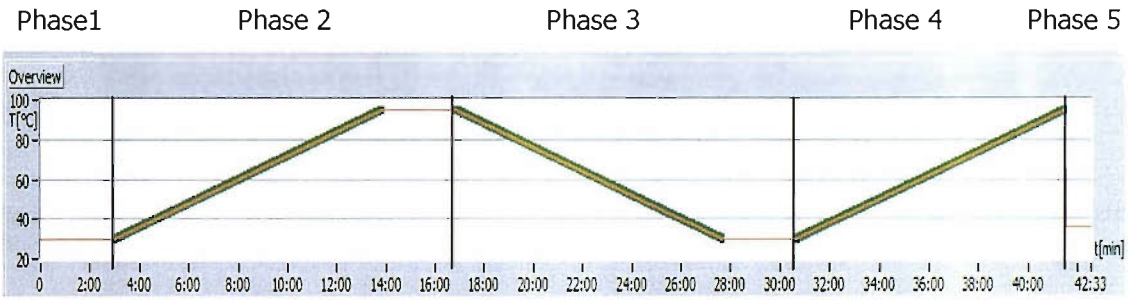


Figure 3.3 A schematic representation of the 5 phase duplex temperature assessment program.

3.2.4 Analysis of Results

Results were collected using the LightCycler™ software version 3.0. *Tm* values were estimated from the maxima in the first derivatives, using the LightCycler™ software, as illustrated in figure 3.4

Figure 3.4 shows plots from four different oligonucleotide pairs from phase 4 of the temperature assessment program. Comparing the different melting curve phases (phases 2 and 4), no hysteresis between corresponding duplex pairs was observed. No difference was observed either between the melting temperature (phases 2 and 4) and annealing temperature (phase 3) for the duplexes tested.

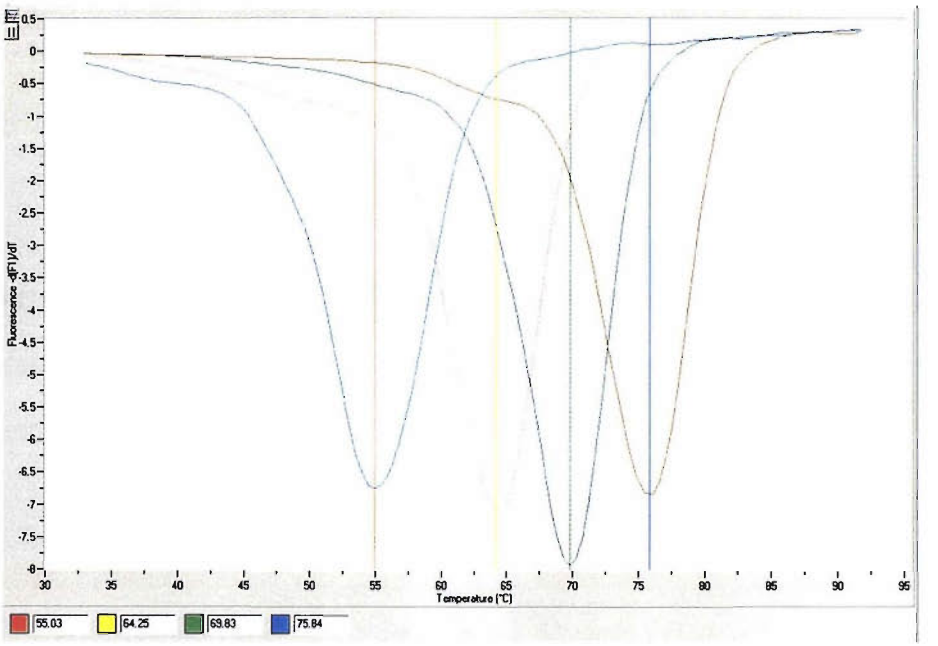


Figure 3.4 A plot of 1st negative derivative fluorescence vs. temperature for four different oligonucleotide duplexes from phase 4 (LightCycler™ *Tm* analysis software).

3.3 Microarray Assays

3.3.1 DNA Preparation

DNA extracts were initially prepared from standard liquid whole blood extracts as described in the Microfabrication assays section (3.1.2). DNA extracts stored at –20 °C until required for these investigation.

3.3.2 PCR Amplification

PCR Amplification was undertaken to produce PCR product for either the direct method or the indirect method of analysis as described in section 1.5.9. Table 3.4 shows the components of each PCR reaction whilst table 3.5 shows the amplification parameters employed for Universal Primer amplification.

Component	Manufacturer	Stock	Reaction Conc.
Universal 3 primer	Oswel DNA/ IBA	50 µM	2 µM
Forward allele specific primer 1 (FASP)	Oswel DNA/ IBA	5 µM	100 nM
Forward allele specific primer 2 (FASP)	Oswel DNA/IBA	5 µM	100 nM
Reverse locus specific primer (RLSP)	Oswel DNA/IBA	5 µM	50 nM
PCR Buffer	Perkin Elmer	x10	x1
Magnesium chloride ($MgCl_2$)	Perkin Elmer	25 mM	1.5 mM
dNTPs	Boehringer Manheim	10 mM	225 µM
Bovine Serum Albumin	Boehringer Manheim	20 µg/µL	0.4 µg/µL
Ampli*Taq* GOLD (5 Units/µL)	Perkin Elmer		10 Units
DNA			2 ng/50 µL reaction

Table 3.4 A list of components required for the indirect method of Universal PCR amplification. NB. For the direct method, two allele specific primers (ASPs) are replaced by a single forward locus specific primer (FLSP).

A full list of primer sequences for the direct and indirect methods are shown in section 6, appendix 6.1.2. PCR amplification was performed using cycling parameters developed for the URP principle as shown in table 3.5:

Process	Temperature	Time	Cycles
Activation	95 °C	11 minutes	1 cycle
Phase 1	94 °C	30 seconds	2-6 cycles
	60 °C	15 seconds	
	72 °C	15 seconds	
	60 °C	15 seconds	
	72 °C	15 seconds	
Phase 2	94 °C	30 seconds	29-33 cycles
	76 °C	105 seconds	
Phase 3	94 °C	60 seconds	3 cycles
	60 °C	30 seconds	
	72 °C	60seconds	
Hold	4 °C	Forever	

Table 3.5 A resume of the URP PCR amplification parameters. NB. Cycle number for phases 1 and 2 should total 35 in each case.

3.3.3 Agarose Gel Electrophoresis

Following PCR amplification, samples were checked for the presence of DNA product by running them out on a 3% nusieve agarose minigel. 4 µL of each PCR product was mixed with 3 µL minigel STOP mix (Bromophenol Blue, TBE (Amersco, USA) and ethidium bromide (Sigma, UK). Low Mass DNA ladders (GibCoBRL, UK, cat. No.10068013) were run as internal sizing/quantification standards. All gels were run for 30-60 minutes with an applied DC current of 120 V. Subsequent illumination of the gel by UV light identified the component bands within the samples and their intensity relative to the sizing standard gave an estimation of DNA quantity.

3.3.4 PCR Product Purification

PCR purification was performed using QiaQuick™ PCR Purification columns (Qiagen, UK). PCR product was eluted in 50 µL ultra pure (18.2 MΩ) water and stored at –20 °C until required for spotting.

3.3.5 Microarray Spotting

Purified PCR product was prepared for spotting by mixing a 1:1 v/v ratio with dimethylsulphoxide (DMSO). All samples were loaded into a 384 well microtitre plate (Amersham Bioscience, Cat. No. RPK0195). Type 7 or Type 7* microarray slides (available only through the Microarray Technology Access Program - MTAP) were then loaded into the Molecular Dynamics Generation III Microarray spotter (Amersham Biosciences) along with the pre-loaded microtitre plate. Due to the configuration of the spotting head (pen set), samples were spotted in batches of 12, referred to as a single well set. The spotting session was run using the spotting parameters outlined in table 3.6.

Program	Number of 384 well plates	Number of plate repeats	Number of well sets	Number of well repeats
Single plate, single well	1	5	1	20

Table 3.6 Spotting parameters for a typical microarray spotting run.

Following spotting, slides were allowed to air dry before being placed in a UV crosslinker set to apply UV energy at 50 mJ/cm^2. For type 7 slides only, an additional post spotting step was performed. Type 7 slides were placed in to an oven preset to 80 °C and left to incubate for 1 hour. All slides were then stored in a sealed cabinet containing a desiccant at room temperature until ready for further processing.

3.3.6 Probe Hybridisation

Two slides were required for each locus under investigation, together with two different probe mixtures (one slide per probe mixture (allele)). Probe mixtures comprised 1 allele specific probe and one locus specific probe diluted in a proprietary hybridisation buffer, Quick Light Hybridisation solution (QLHS), (Lifecodes Corp. Cat. No. 10066). Probe concentrations following dilution in the hybridisation buffer were between 5 nM and 50 nM. For each slide hybridisation, 200 µL of QLHS containing one probe set was loaded onto a 22x60 mm coverslip. The arrayed slide was inverted onto the coverslip, then flipped upright and incubated in a darkened humid hybridisation oven set at 50 °C for 30 minutes. This procedure was then repeated for the second slide using QLHS containing the alternative allele probe mixture. Hybridisations for all other loci were carried out in a similar fashion. See section 6, appendix 6.1.2 for a complete list of probe sequences.

3.3.7 Stringency Washing

Post hybridisation stringency washing involved removing excess/non specifically bound probe from the slide surface by subjecting the slide to a number of wash solutions. For each slide, the cover-slip was first removed before being immediately rinsed according to the following stringency wash protocol:

Solution	Manufacturer	Temperature	Volume	Time
2 x SSC	BDH	Room temperature	100 mL agitated	1 minutes
3 M TMAC/0.1% SDS	BDH	60 °C	100 mL agitated	3 minutes
2 x SSC	BDH	Room temperature	100 mL agitated	10 seconds
2 x SSC	BDH	Room temperature	100 mL	<5 minutes
1 x SSC	BDH	Room temperature	100 mL	3x2 seconds
0.5 x SSC	BDH	Room temperature	100 mL	3x2 seconds
18.2 MΩ SDW	BDH	Room temperature	100 mL	3x1 second

Table 3.7 Microarray slide stringency washing parameters following probe hybridisation.

Following stringency washing, each slide was rapidly dried by passing it in front of a continuous flow of ultra pure nitrogen gas (BOC, UK). Dried slides were then stored in a darkened cabinet in the presence of a desiccator until ready for scanning.

3.3.8 Slide scanning

Each slide was scanned for the presence of Cy3 and Cy5 using the Molecular Dynamics Generation III Scanner (Amersham Bioscience). Each slide within the pair was scanned twice, one for each colour producing two comparable scan files per locus.

3.3.9 Scan Analysis

Scanned images were viewed using Image Quant™ v5.0 software. Analysis of the scanned images was carried out using Array Vision™ v6.0 software. This software package produced data values for individual spot fluorescence as well as background fluorescence for each slide. Comparative data analysis was then performed to assess probe hybridisation efficacy and sample genotype as follows:

3.3.10 Comparative data Analysis

Comparative data analysis was used to compare the corrected fluorescent data, from each element within one array, (allele 1, slide 1), with its corresponding element within a duplicate array, (allele2, slide 2). This was achieved by considering two types of analysis, both of which relied on calculating the ratios of the two corresponding data sets.

3.3.10.1 Test 1 : Simple Comparative Ratios

Equation 15

$$\text{Test 1 ratio} = \left(\frac{CorrectedCy5(a)}{CorrectedCy3(a)}\right) \text{ Plotted against } \left(\frac{CorrectedCy5(b)}{CorrectedCy3(b)}\right)$$

Where (a) = allele a
(b) = allele b

The calculated test 1 ratios, according to equation 15, are plotted as a scatter graph, as shown in the following example.

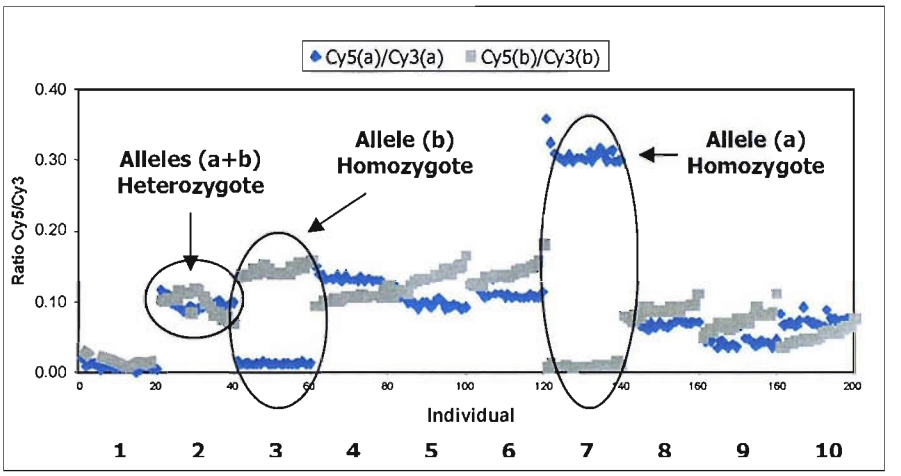


Figure 3.5 A plot of Test 1 ratios ([Cy5(a)/Cy3(a)] vs. [Cy5(b)/Cy3(b)]) for ten individuals, 20 replicates per individual. In the above example, the genotypes for individuals 1-10 are –ve, HET, HOM b, HET, HET, HET, HOM a, HET, HET and HET, respectively.

3.3.10.2 Test 2 : Log $_{10}$ Comparative Ratios

Equation 16

$$\text{Test 2 ratio} = \left(\log_{10} \left(\frac{CorrectedCy5(a)}{CorrectedCy5(b)} \right) \right)$$

Where (a) = allele a
(b) = allele b

The calculated test 2 ratios , according to equation 16, are plotted as a scatter graph, as shown in the following example.

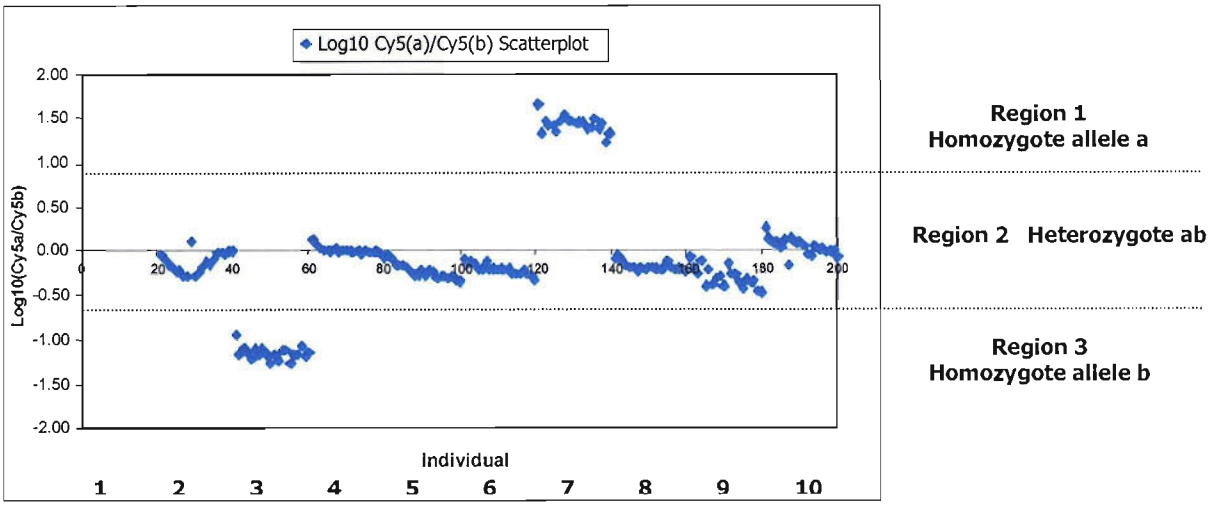


Figure 3.6 A plot of test 2 ratios (Log_{10} Cy5(a)/Cy5(b)) from the same data (figure 3.6). In this example, the genotypes for individuals 1-10 are –ve, HET, HOM b, HET, HET, HET, HOMa, HET, HET and HET, respectively.

The position of each data point in figure 3.6 relative to the y-axis determined the genotype of the sample. For samples that were homozygous for allele a, the ratio of a to b would be (a>b), resulting in a positive log_{10} value. Alternatively, for samples that were homozygous for allele b, the ratio of a to b would be (a<b), resulting in a negative log_{10} ratio. Where data sets showed equivalence, the ratio of a to b would be (a=b), resulting in a log_{10} ratio value close to zero.

All DNA samples were initially genotyped by conventional 377 gel analysis. Microarray data was then directly compared to these references to ensure accuracy of the microarray data.

4 Results and Discussion

4.1 Microfabrication Assay Results[1]

The following results describe the different assays performed to assess the viability of performing DNA extraction using a silicon channel platform.

4.1.1 Calculating an estimate of experimental error

For each of the experiments, total peak area data were summarised as mean, standard deviation, and standard error (95% confidence interval) from a number of replicate samples that were concurrently analysed.

	Sample experimental data (rfu)			
	1	**2**	**3**	**4**
Experimental replicates (rfu)	-	-	556951	-
	-	-	1040444	-
	-	-	820000	-
	279970	373543	578115	544485
	246714	390798	170645	803432
	271935	558897	554324	657680
	465174	797811	88554	923443
Mean	**315948**	**530262**	**620080**	**732260**
Standard deviation	100488	196989	292826	165769
St. Error of Mean (95% confidence)	98476	193045	216924	132640
% St. Error mean	31.2	36.4	35.0	18.1
Average % Error	**30.2**			

Table 4.1 Replicate sample data from (1) section 4.2.2 Elution 1, 2 minute incubation; (2) Elutions 1-8, 2 minutes incubation; (3) section 4.1, positive experimental controls and (4) section 4.1.4,solution 3. In each case, the mean, standard deviation and standard error of the mean (95% confidence interval) was calculated. These values were used to calculate the % st. error of the mean for each data and finally an average % error for all experimental data.

[1] Experimental work was carried out by Jenny Hughes under my direct supervision

Table 4.1 showed the average standard error of the mean (95% confidence interval), expressed as a percentage of the mean rfu value ±30.2%. This standard error for each of the following experiments was determined to be rfu ±30.2%. This error value is quite high probably due to stochastic variation between different PCR amplifications as well as a limited number of sample replicates.

In each instance control samples comprised 1.5-2 ng DNA in 20 µL SDW. These were not extracted through the silicon channel. Test samples were extracted according to section 3.1.4. NB. Protocol variables D, N, Y and Q were specified in each case. Following extraction, test and control samples were sent forward for PCR amplification (3.1.5), gel separation (3.1.6) and analysis (3.1.7).

4.1.2 A comparison of silicon channel extraction of DNA diluted in (a) SDW, (b) Qiagen AL buffer.

Two different mixtures of DNA were made up from standard Qiagen extracted DNA samples

(1) 0.13 ng/µL DNA in 18.2 MΩ SDW, prepared in duplicate and labelled a+c

(2) 0.13 ng/µL DNA Qiagen AL buffer, prepared in duplicate and labelled b+d

Samples a and c were treated as negative controls whilst samples b and d were processed as test samples using the following parameters:

Variable name	Variable code	Identification
Volume of input sample (µL)	D	15
Number of sample volumes added	N	1
Incubation time of sample (minutes)	Y	2
Number of elutions	Q	4

Table 4.2 Silicon channel extraction protocol parameters for section 4.1.2.

Figure 4.1 shows the electrophoretograms from elution 1 for samples (a) to (d) whilst table 4.3 shows the combined totals peak areas of elution 1-4 for each sample (a) to (d).

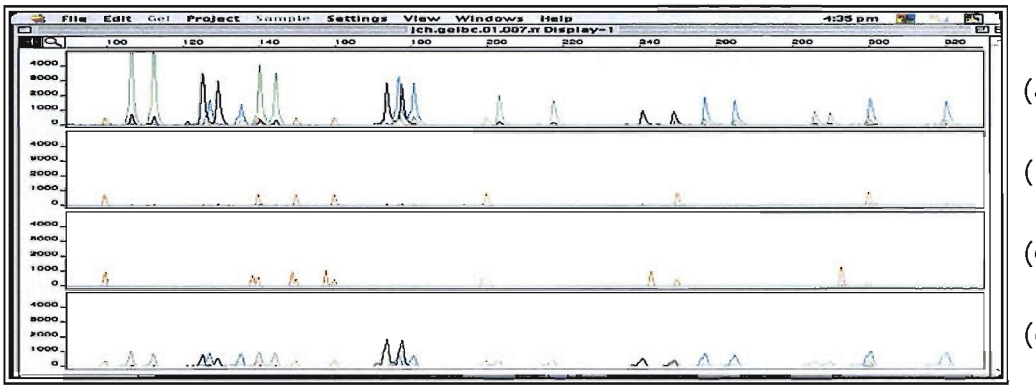


Figure 4.1 Elution 1 (Q1) electrophoretograms for sample a, b, c and d respectively. Red peaks indicate size standards and were included as reference markers only. They were not included in peak area calculation.

Sample	Total peak area (rfu)
A	556951
B	0
c	0
d	195093

Table 4.3 Combined totals peak areas from elutions 1-4 for each sample (a) to (d).

Electrophoretogram (a) showed the control profile for DNA in SDW. This represents the total amount of DNA present in the sample prior to processing and yields a total peak area of around 557000 rfu. Electrophoretogram (b) showed the profile obtained from elution 1 of DNA diluted in SDW. The lack of a profile suggested that no DNA was present in the elution and implied that the channel did not bind DNA from this particular sample.

Electrophoretogram (c) again showed no profile. As the control, for DNA diluted in Qiagen buffer, this was not surprising since the buffer contained ethanol, which is known to inhibit PCR. Electrophoretogram (d) showed the profile obtained from an elution derived from the initial addition of DNA in Qiagen buffer. The presence of a profile (peak area c.195000 rfu), indicated that DNA was present in the elution, and therefore must have been retained within

the channel during the extraction phase. Although the size of the peaks and therefore total peak area within electrophoretogram (d) were smaller than those in the control sample electrophoretogram (a), Qiagen buffer clearly encouraged DNA to bind to the channel surface. The bound DNA was subsequently released when the elution buffer was added. SDW, corresponding to sample (b), does not contain chaotrophic salt and therefore the conditions required for DNA trapping were not present. This results in no DNA being trapped during extraction and therefore no DNA being present in any subsequent elutions.

These initial results showed that DNA could be retained within a silicon channel structure using the Qiagen chemistry. By applying a low ionic strength buffer, such as SDW, conditions permitted the recovery of DNA. As eluted DNA samples were successfully amplified, SDW appears not to interfere with DNA integrity although it was not possible to ascertain the exact degree of DNA shearing, if this had occurred.

In the previous example, the trapping and release of DNA from within channels has been demonstrated although the amount of DNA recovered appears to be less than that from the control. Optimisation of the extraction technique is now demonstrated.

4.1.3 Increasing the amount of DNA recovered as compared to a control sample by increasing sample incubation time 'Y':

In this example, DNA samples were prepared as before i.e. 0.13 ng/μL in either SDW (Control) or Qiagen (AL) buffer and processed as test samples using the parameters:

Variable name	Variable code	Identification
Volume of input sample (μL)	D	15
Number of sample volumes added	N	1
Incubation time of sample (minutes)	Y	2, 5 or 10
Number of elutions	Q	8

Table 4.4 Silicon channel extraction protocol parameters for section 4.1.3.

Figure 4.2 shows the gel images for elutions 1-8 for samples incubated for 2, 5 or 10 minutes compared to a control sample.

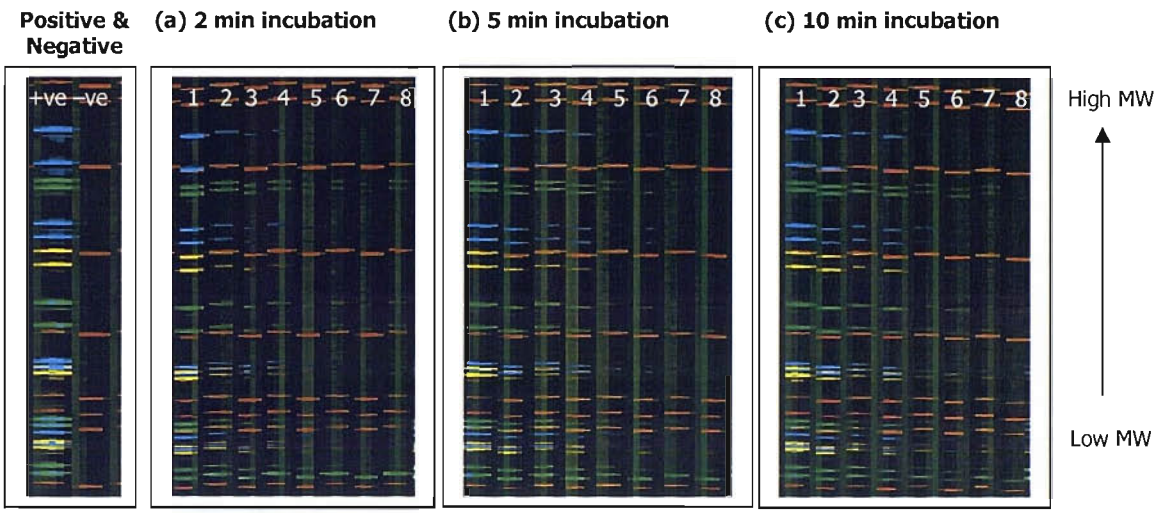


Figure 4.2 Gel images showing profiles corresponding to 8 elutions for incubation times of (a) 2 minutes, (b) 5 minutes and (c) 10 minutes together with a gel profile for a positive control.

It is clear from the gel images above that as the incubation time was increased from 2 minutes to 5 and 10 minutes, the intensity of the profiles throughout the range of molecular weight markers, within the first couple of elutions appeared to increase. The intensity of these profiles can be thought of as being directly proportional to the amount of DNA present in them and therefore indicated that more DNA had been recovered. In addition, as the incubation time was increased, profiles corresponding to the entire range of molecular weight markers were seen in later elutions indicating that more DNA had been recovered.

An electrophoretogram was created from the positive control and the summed total peak area calculated as 1040444 rfu. This corresponded to the total amount of DNA in the control sample and was considered proportional to the starting amount of DNA. Similar electrophoretograms were constructed for each of the eight elutions, for each sample. The peak area for the first elution was compared to the control value to highlight the difference in DNA concentration between each incubation time. This data is shown in table 4.5.

Incubation time (minutes)					
2		5		10	
Peak Area (rfu)	**% recovery**	**Peak Area (rfu)**	**% recovery**	**Peak Area (rfu)**	**% recovery**
315948	30.4 %	497597	47.8 %	576602	55.4 %

Table 4.5 Peak areas (rfu) and % recovery from elution 1 for each incubation time compared to positive control.

Incubation time (minutes)					
2		5		10	
Peak Area (rfu)	**% recovery**	**Peak Area (rfu)**	**% recovery**	**Peak Area (rfu)**	**% recovery**
530262	51.0 %	738778	71.0 %	979585	94.1 %

Table 4.6 Combined total peak areas and % recovery from elutions 1-8 for each incubation time.

As the sample incubation time was increased, the amount of DNA recovered in the first elution (table 4.5) and the total amount of DNA recovered, elutions 1-8 (table 4.6) increased. This increase was indicative of more DNA being retained by the channel during the incubation phase.

In summary, these result demonstrated that DNA could be incubated within a channel using the Qiagen chemistry for 10 minutes without any apparent detrimental effects. Increasing the incubation time lead to a higher % recovery overall however, not all DNA could be recovered in the first elution. This drawback may have implications for the efficiency of extraction of very low levels of DNA however multiple elutions suggest that a higher % recovery of DNA would be achievable.

4.1.4 Optimising the Qiagen buffer to enhance extraction efficiency

DNA samples were prepared as before with AL buffer containing different concentrations of ethanol as shown in table 4.7.

Mixture	Ethanol (μL)	Qiagen buffer (μL)	Ratio (EtOH:AL)
1	0	18	0:1
2	6	12	1:2
3	9	9	1:1
4	12	6	2:1
5	18	0	1:0

Table 4.7 Mixture ratios of Qiagen buffer : Ethanol.

The samples were processed as test samples using the parameters:

Variable name	Variable code	Identification
Volume of input sample (μL)	D	15
Number of sample volumes added	N	1
Incubation time of sample (minutes)	Y	10
Number of elutions	Q	8

Table 4.8 Silicon channel extraction protocol parameters for section 4.1.4.

Combined peak areas for each experiment from the resulting electrophoretograms were compared to the total peak area from a positive control (820000 rfu). The results are shown in table 4.9.

Sample dilution mixture									
1		2		3		4		5	
Peak Area (rfu)	% recovery	Peak Area (rfu)	% recovery	Peak Area (rfu)	% recovery	Peak Area (rfu)	% recovery	Peak Area (rfu)	% recovery
337217	41.1 %	429680	52.4 %	732260	89.3 %	532180	64.9 %	1640	0.2 %

Table 4.9 Combined total peak areas and % recovery from elutions 1-8 for each experimental mixture.

The peak areas for each elution were plotted in a bar graph for each experiment mixture as shown in figure 4.3.

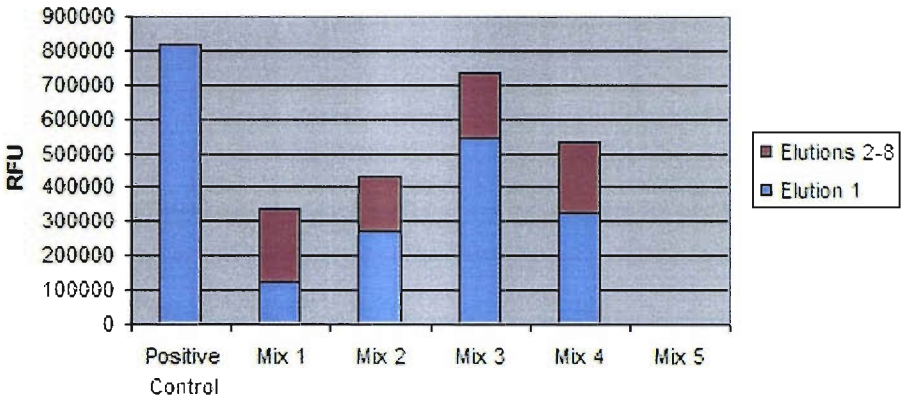


Figure 4.3 Graph of total peak areas for each experiment mixture compared to a 2 ng positive control.

The results showed that adjusting the chaotrophic mixture to a 50% ratio of Qiagen buffer (AL) and ethanol enhanced the amount of DNA which was retained in the channel. This was seen by an increased peak area arising from elution 1 and also from the combined peak areas from elutions 2-8. Diluting DNA in a mixture of 50% Qiagen Buffer and ethanol enhanced the overall recovery of DNA. It was also apparent that the ethanol reduced the viscosity of the DNA/chaotrophic solution. This enabled the solutions to be passed more easily through the channel, reducing the time taken to process each sample.

4.1.5 A comparison of silicon channel extraction efficiency of different DNA samples concentrations:

DNA samples were prepared in a mixture of 50% Qiagen buffer and Ethanol at the following concentrations (see table 4.10).

Sample	DNA amount	DNA Concentration
1	2 ng in 20 µL	0.1 ng / µL
2	2 ng in 200 µL	0.01 ng / µL
3	1 ng in 200 µL	0.005 ng / µL

Table 4.10 DNA concentrations of different sample solutions.

Samples were processed as test samples using the parameters:

Variable name	Variable code	Identification
Volume of input sample (μL)	D	15
Number of sample volumes added	N	1 or 10
Incubation time of sample (minutes)	Y	10
Number of elutions	Q	8

Table 4.11 Silicon channel extraction protocol parameters for section 4.1.5.

NB. For 200 μL sample volumes, 20 μL aliquots were sequentially incubated in the channel until the entire volume had been added.

Combined peak areas for each experiment from resulting electrophoretograms were compared to the total peak area from a positive control (578115 rfu) as shown in table 4.12.

	DNA dilution					
	1		2		3	
	Peak Area	% Recovery	Peak Area	% Recovery	Peak Area	% Recovery
Elution 1	329320	56.96 %	491631	85.04 %	280154	48.46 %
Elution 2	30669	5.30 %	24290	4.20 %	20204	3.50 %
Elution 3	14325	2.48 %	11340	1.96 %	5135	0.89 %
Elution 4	4890	0.85 %	4927	0.85 %	4517	0.78 %
Elution 5	3170	0.55 %	2865	0.50 %	5764	1.00%
Elution 6	4692	0.81 %	3942	0.68 %	4654	0.08 %
Elution 7	5094	0.88 %	3056	0.53 %	967	0.17 %
Elution 8	0	0 %	0	0 %	0	0 %
Total Peak area	**392160**	**67.8 %**	**542051**	**93.7 %**	**321395**	**55.5 %**

Table 4.12 Total peak area and % recovery of DNA from each elution, for each sample dilution compared to the total peak area for a 2 ng control.

Peak areas for elution 1 together with combined peak areas for elutions 2-8 were plotted as a bar graph shown in figure 4.4.

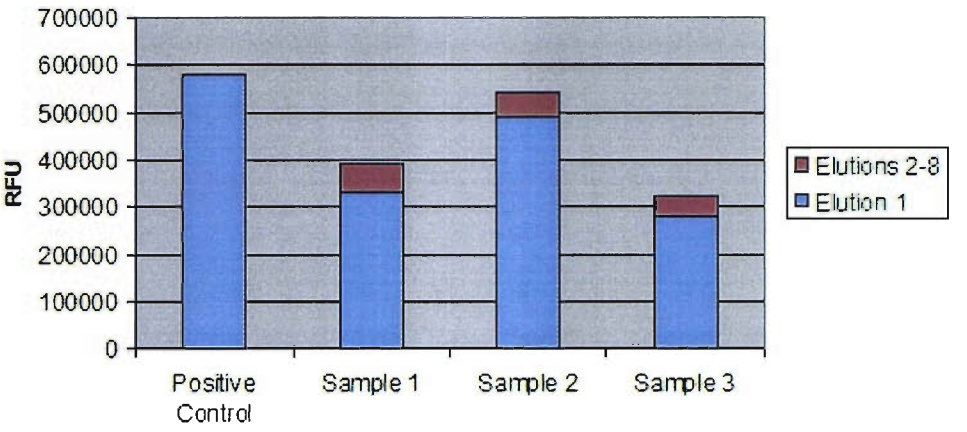


Figure 4.4 Graph of total peak areas for each experiment mixture compared to a 2 ng positive control.

The percentage recovery for sample 3 appeared to be very poor compared to samples 1 and 2 however this recovery was calculated against the 2 ng control. Sample 3 only had 1 ng of input DNA and so the true control for this sample would have been a 1 ng control sample. Given that the amount of DNA is directly proportional to peak area, to reflect the difference in DNA amounts between the experimental control and sample 3, the total peak area for sample 3 was multiplied by 2, thus normalising the data between the three concentrations.

This meant that the total % recovery and total amount of DNA recovered from each dilution, for elution 1 could be regarded as:

Sample 1	57.0 %	c. 1.14 ng DNA from an initial 2 ng total DNA
Sample 2	85.0 %	c. 1.70 ng DNA from an initial 2 ng total DNA
Sample 3	96.9 %	c. 1.94 ng DNA from a theoretical 2 ng total DNA

And that this recovery was possible within a single 20 µL elution.

These results demonstrated that recovery of DNA was possible even when the DNA concentration was very low (c.0.005 ng / µL). More importantly, very dilute DNA samples (sample 3), as well as being retained in the channel device and then eluted in a much smaller

volume showed a higher % recovery overall. In these examples, sample 3 was recovered in a volume equivalent to 10 times less than initial volume representing a 10-fold increase in concentration. This has quite important considerations with respect to forensic sample where the amount of starting material is often limited.

The % recovery of DNA from sample 1 was lower than from sample 2 despite the same amount of DNA being added initially. This was thought to be as a result of there being a 10-fold increase in the amount of time taken to pass sample 2 through the channel compared to sample 1. This increase in time suggested that a higher proportion of DNA molecules were coming into contact with the channel surface despite being present in the sample at a 10-fold less concentration. This increased time would increase the likelihood of DNA being trapped and ultimately recovered following elution.

4.1.6 Silicon channel extraction of DNA from whole liquid blood.

1 μL of whole liquid blood was taken in duplicate and processed according to experimental procedure 3.1.4 up to and including the addition of 32 μL ethanol. After this stage the samples were processed according to procedure 3.1.5 using the parameters shown in table 4.13.

Variable name	Variable code	Identification
Volume of input sample (μL)	D	15
Number of sample volumes added	N	1
Incubation time of sample (minutes)	Y	10
Number of elutions	Q	8

Table 4.13 Silicon channel extraction protocol parameters for section 4.1.6.

Combined peak areas for each experiment from resulting electrophoretograms were compared to the total peak area from a positive control (170645 rfu) as shown in table 4.14.

	Elution 1		Elutions 2-8	
Sample	Peak area (rfu)	% recovery	Peak Area (rfu)	% recovery
1	397264	232.8 %	688821	403.6 %
2	443686	160.0 %	598315	350.6 %

Table 4.14 Peak area and percentage recovery of DNA from elution 1 and elutions 2-8 as compared to a 2 ng positive control.

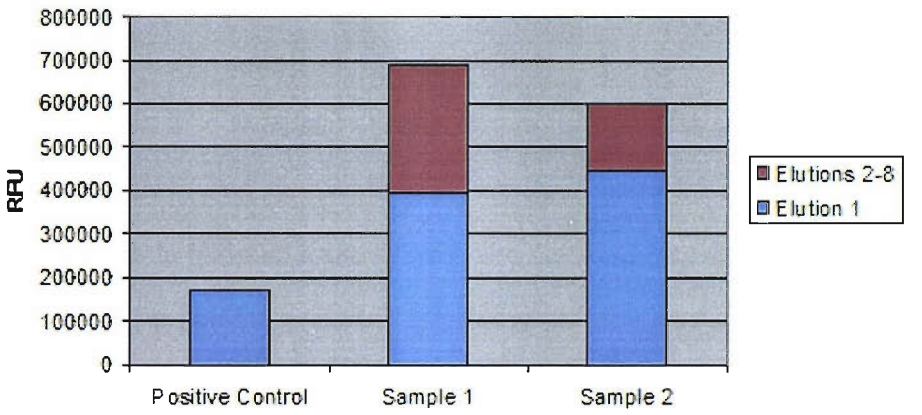


Figure 4.5 Total peak area recoveries of DNA from whole blood compared to a 2 ng positive control.

Samples 1 and 2 showed respective peak areas from elution 1 of 397264 rfu and 443686 rfu. Comparing these values with the total peak area from the 2 ng control suggested that samples 1 and 2 contained approximately 2.5 times the amount of DNA. This equates to approximately 5 ng DNA in the first elution. These results demonstrated that DNA could be extracted from liquid whole blood using the silicon channel protocol in combination with the Qiagen chemistry. The average total amount of DNA extracted between the two samples was approximately 5.5 ng of DNA.

4.1.7 Silicon channel extraction of DNA from samples contaminated with lead (II) nitrate (PCR inhibitor).

Initial investigations were carried out to determine the effect of lead nitrate on PCR efficiency. To illustrate this, a titration experiment was set up whereby 2 ng DNA was adulterated with increasing amounts of lead nitrate. The results showed that DNA samples containing up to 5 ng/µL lead nitrate, successfully amplified during PCR compared to a 2 ng control DNA. Furthermore, samples that contained greater than 12 ng/µL lead nitrate showed complete PCR inhibition as seen by the complete lack of any PCR product.

As a result of these initial findings, two samples containing 2 ng DNA were prepared as control samples (labelled 1 and 3). Two further samples containing 2 ng DNA adulterated with 2000 ng lead nitrate (labelled 2 and 4) were also prepared (equivalent to 40 ng/µL). Samples 1 and 2 were processed as control samples whilst samples 3 and 4 were processed as test samples using the parameters:

Variable name	Variable code	Identification
Volume of input sample (µL)	D	15
Number of sample volumes added	N	1
Incubation time of sample (minutes)	Y	10
Number of elutions	Q	7

Table 4.15 Silicon channel extraction protocol parameters for section 4.1.7.

The control sample containing 2 ng DNA gave a total peak area of 554324 rfu. In contrast the total peak area for the control sample adulterated with lead nitrate was 0 rfu. This demonstrated PCR inhibition as a result of the addition of 2000 ng lead nitrate.

Combined peak areas from electrophoretograms of the two test samples appear in table 4.16.

Sample	Peak Area (rfu)
3	487251
4	51733

Table 4.16 Peak area data from samples 3 and 4.

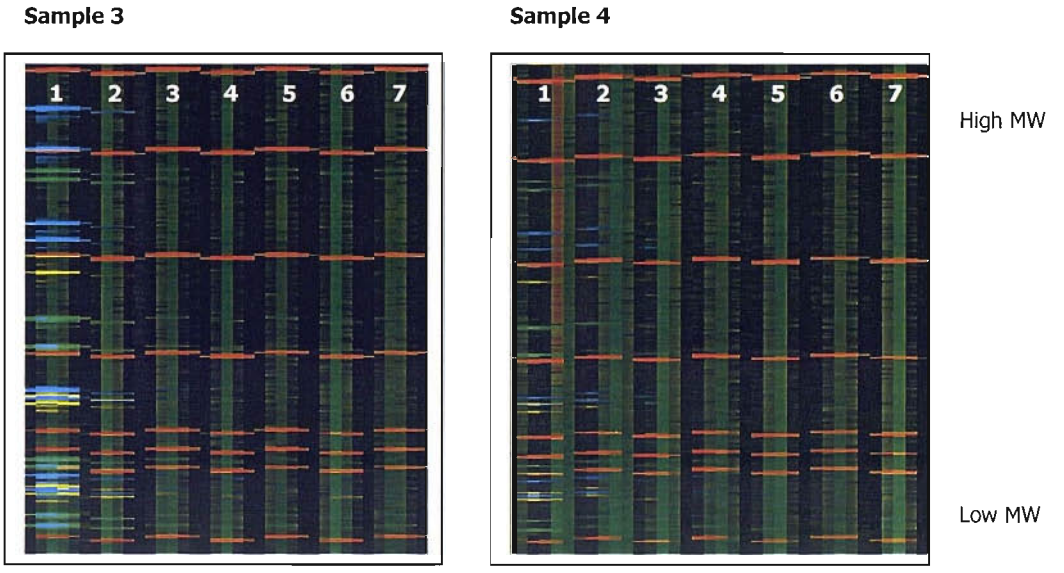


Figure 4.6 Gel images showing profiles corresponding to elutions 1-7, for samples 3 and 4.

Results for sample 3 showed the profile obtained for the control sample containing no lead nitrate. The combined total peak area value when compared to the control showed an 87.9 % recovery. Sample 4 showed the profile for the sample which originally contained lead nitrate. In this instance the presence of only a weak DNA profile was indicative of reduced amplification efficiency, probably as a result of inhibition. The sample was original adulterated with lead nitrate (40 ng/μL), more than 3 times the minimum amount required to cause complete PCR inhibition. The presence of a small DNA profile however suggested that the amount of lead nitrate in the sample following elution must have been less than 12 ng/μL. This was equivalent to a loss of >1400 ng lead nitrate during the extraction phase. The exact amount of lead still remaining in the sample was not calculated but was known to be within the range 5 ng/μL to 12 ng/μL. These results suggest that contaminants known to inhibit PCR can be effectively removed from a DNA sample using the silicon channel

extraction process, and that the washing step can selectively remove inhibitors and contaminants without disrupting DNA integrity.

4.1.8 A comparison of the amount of DNA which can be trapped from two different length channels.

DNA samples were prepared as before (0.13 ng/µL in AL Buffer) and then processed as test samples using the parameters below in either a 300 mm or 1000 mm channel.

Variable name	Variable code	Identification
Volume of input sample (µL)	D	15
Number of sample volumes added	N	1-3
Incubation time of sample (minutes)	Y	10
Number of elutions	Q	7

Table 4.17 Silicon channel extraction protocol parameters for section 4.1.8.

NB. The number of volumes of sample to add was 3 for 300 mm channel and 1 for the 1000 mm channel due to the difference in volume capacity between the two devices.

A control sample containing 2 ng DNA gave a total Peak area of 88554 rfu. Combined peak areas from electrophoretograms of the test samples were:

	300 mm channel		1000 mm channel	
	Peak Area (rfu)	% Recovery	Peak Area (rfu)	% Recovery
Elution 1	38162	43.09 %	77581	87.61 %
Elution 2	1273	1.44 %	3367	3.80 %
Elution 3	0	-	258	0.29 %
Elution 4	0	-	0	-
Elution 5	0	-	0	-
Elution 6	0	-	0	-
Elution 7	0	-	0	-
TOTAL	39435	44.53 %	81206	91.7 %

Table 4.18 Combined total peak area and % recovery data for each elution, for each sample as compared to the 2 ng control.

The peak areas for elutions 1-7 for each DNA sample were as follows:

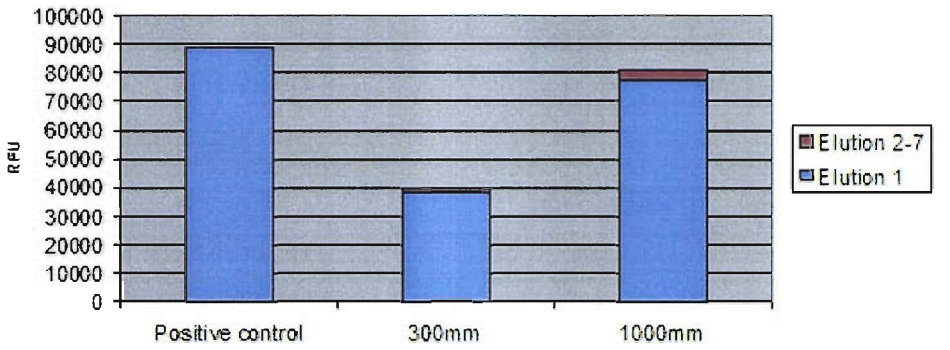


Figure 4.7 Recovery of DNA (total peak area data) from two different channel lengths compared to a control.

The 300 mm channel successfully extracted and recovered approximately 45 % of the total amount of DNA added. This value increased to 91.7 % recovery when extraction was carried out in the longer (1000 mm) channel. This demonstrated that the channel length was a contributing factor to the amount of DNA that could be recovered from a sample. Furthermore, these results suggested that shorter channels may have become saturated and therefore could not retain as much DNA as longer channels.

4.1.9 DNA saturation of a silicon channel

In the previous example, an increased channel length gave rise to an increased amount of DNA recovery. This suggested that channels of a fixed dimension and predefined surface area would bind a specific quantity of DNA so that as the channel surface area decreased, the amount of DNA which could be retained in the device would also decrease in a proportional manner. This reasoning implied that it would be possible to saturate a channel with a given surface area to the point at which no further DNA binding would occur.

A number of different DNA concentrations within the range 0.1 ng to 0.5 ng/µL were prepared in duplicate, one set for the test and one set for the control. Test samples were processed using the parameters shown below on a 300 mm channel.

Variable name	Variable code	Identification
Volume of input sample (µL)	D	10
Number of sample volumes added	N	1
Incubation time of sample (minutes)	Y	10
Number of elutions	Q	6

Table 4.19 Silicon channel extraction protocol parameters from section 4.1.9.

Amplifications performed in this study were optimised for 2 ng DNA. Attempting to amplify more that 2 ng would not yield a quantitative linear relationship between peak area and total DNA amount, therefore samples were split prior to PCR. For samples whose concentration was 0.25, 0.3, 0.35 and 0.4 ng/µL, elution 1 was split into two aliquots prior to PCR. For starting DNA concentrations of 0.45 and 0.5 ng/µL, elution 1 was split into three aliquots prior to PCR. Exactly the same methodology was given to the control DNA samples.

Combined peak areas from electrophoretograms of the samples were:

Total Peak Area (rfu)	DNA Concentration ng/µL									
	0.1	0.15	0.175	0.20	0.25	0.3	0.35	0.40	0.45	0.50
Test samples	33752	62797	78618	123240	147213	261443	277625	267858	288099	280895
Control samples	41179	98327	119536	190136	392458	399294	540813	543739	672657	800911

Table 4.20 Total Peak areas for DNA samples processed through the channel (test) and total peak areas from equivalent control DNA sample (not processed through the channel).

Results from table 4.20 appear in figure 4.8.

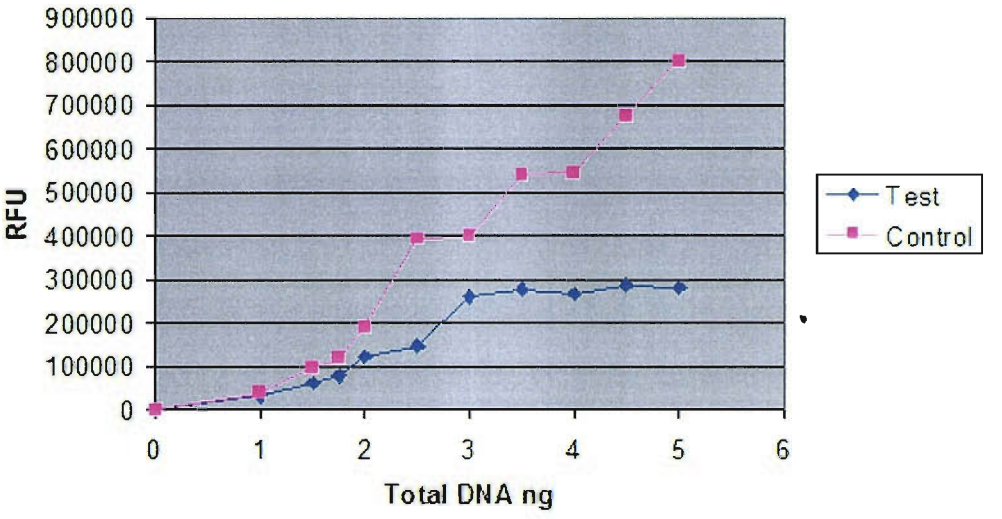


Figure 4.8 A scatter plot of total peak areas (rfu) from eluted DNA for each sample input.

The results showed that the amount of expected PCR product indicated by the total peak area for the control samples, continued to rise in a proportional manner. Results for the test samples however, increased up to approximately 3 ng total DNA. Above 3 ng total DNA, no increase in peak area was observed suggesting that the saturation limit for a 300 mm channel was 3 ng.

These results demonstrated that a small fixed channel length would provide a maximum binding capacity to trap DNA from very concentrated samples. If this amount was optimised to be 2 ng, which is known to be optimum for these PCR reactions, then the possibility of having a compromised PCR result in subsequent analysis is reduced. In effect the channel is acting as a quantitative device.

4.1.10 General discussion

The use of microfabricated silicon channels as a means of performing DNA extraction has been successfully demonstrated. Whilst many of the literature examples have relied on the

fabrication of complex miniaturised features, such as the bed of nails, serpentine channels etched into silicon wafers offer a structurally, more simple alternative.

The specific channel design employed here occupied a relatively small surface area yet the serpentine shape ensured that the internal surface area of the channel remained high. This was an important attribute as it ensured a consistently high % DNA recovery following elution. As well as acting as an extraction device, the silicon channels were also shown to be capable of concentrating very dilute DNA samples; a factor that is of particular relevance in forensic casework. This ability ensures that enough DNA can be collected for further downstream processes such as PCR. In many conventional and larger scales applications, the amount of DNA extracted from a crude sample is unknown, requiring additional analytical techniques to determine DNA concentration. The microfabricated devices outlined here had a predefined surface area which was shown to bind a particular and constant amount of DNA. This eliminated the need for a post extraction quantification step which would otherwise consume some of the DNA sample.

Although eluted DNA was capable of being amplified through PCR, the effect of the device together with the extraction chemistry, on DNA integrity was not determined. The fact that DNA was amplifiable suggested that the DNA was not degraded or compromised to any great extent. The lack of any physical barriers in the channel, such as a sieving matrix or additional structures, meant that the DNA solution could pass through the device unhindered. This minimised the shearing forces which would serve to disrupt the DNA molecule causing excessive degradation.

Blockage of microfabricated devices has been well documented in the literature and remains one of the biggest challenges for integrated systems as these occurrences often result in the device being rendered inoperable. Many of the cited examples (e.g. bed of nails) operate by

passing liquid through a wide chamber containing additional structures. These structures create multiple flow paths which mean debris or air bubbles can become trapped in the device creating blockages or voids, which cannot be cleared. The silicon channels presented here, did not have multiple flow path as a consequence of their geometry and lack of any inner chamber structures. Their single flow path allowed for a much greater degree of control. Air bubbles or small particulates were easily removed such that the device remained operable for longer periods of time. This advantage meant that the devices could be re-used many times during the course of the investigation. An efficient cleaning regime was implemented that ensured no cross contamination between extraction runs.

The simplicity of the design combined with the high efficiency of extraction created a system that could be easily incorporated into a miniaturised system. The single input and output nature of its operation meant that future integration to a millifluidic system should prove to be relatively straightforward.

4.2 Melting Temperature (*Tm*) Assay results

4.2.1 Optimising experimental design

Initial investigations concerned optimising the experimental design to account for representative duplex interactions. The LightCycler™ instrument can generate temperatures within the range, room temperature to 100 °C, therefore it was necessary that duplex melting temperature fell within this window for the broad range of duplexes being tested. In addition, the sensitivity of the assay needed to be fairly high, being able to differentiate duplexes that differed from one another by a minimum of a single base mismatch. This would reflect differences in duplex stability for probes annealing to targets in microarray applications, for example, or pairs of competing primers binding to template in an ARMS assay.

As a suitable starting position, two complementary 19-mer sequences were designed to act as the test oligo duplex. These were referred to as strand A and strand B as shown in figure 4.9.

Strand A	**5'- C G A C G T G C N N N A T G T G C T G** **G C T G C A C G N'N'N'T A C A C G A C - 5'**	**Strand B**

Figure 4.9 demonstrate two generic 19-mer sequences. Bases shown as N, N' are changeable to create different Strand A and B sequences. Duplex sequences arise from the bringing together of a strand A molecule with a strand B molecule. Depending on the sequences selected, results in either a complementary duplex or a duplex which contains deliberate mismatch/es.

For simplicity, only the strand type (A or B), together with the three variable base names will be used in the naming convention for each of the sequences, such that '5' FAM, A, GCC' refers to the sequence 5' FAM-CGACGTGC**GCC**ATGTGCTG and 'B, CTG-3' DABCYL' refers to the sequence CAGCACAT**CTG**GCACGTCG-3' DABCYL. Labelling oligonucleotides with 5' FAM was done by adding a labelling phosphoramidite monomer in the final step of oligonucleotide

synthesis. For oligonucleotides labelled with 3' DABCYL, the solid support for oligonucleotide synthesis (resin) was prefunctionalised with the label prior to synthesis.

19-mer sequences were selected as they contained an odd number of bases. This meant that the variable bases at positions 9, 10 and 11 were centralised; position 10 residing at the very centre of the sequence. A deliberate mismatch at these positions will cause a maximum destabilising effect compared to the same mismatch within either of the two tails. Having designed strand A and strand B sequences to have three variable bases, this could introduce up to three consecutive mismatches depending on the sequence of the two strands. This allowed not only complementary sequences to be tested but also sequences which resulted in deliberate mismatches.

4.2.1.1 The FRET Principle

The basic principle to FRET relies on energy passing from a donor molecule to an acceptor molecule through dipole-dipole interactions. Ordinarily, when a donor molecule such as FAM is excited, energy is given out in the form of photons. In the presence of an acceptor molecule such as DABCYL, this energy is harnessed by the acceptor resulting in a diminished or quenched donor (FAM) signal. The degree of quenching achieved is proportional to the spatial distance between the donor and acceptor. This principle is exploited in oligonucleotide melting temperature assays where one oligo in a duplex is labelled with a donor and the other is labelled with an acceptor. At low temperatures, the donor and acceptor are brought into close proximity such that following excitation, FRET can occur resulting in a quenched donor signal. As the temperature is increased beyond the duplex melting temperature, the strands separate causing the donor and acceptor to become spatially separated. FRET no longer occurs resulting in an enhanced donor signal. This transition can be monitored as a function of temperature to determine the exact duplex melting temperature.

4.2.1.2 FRET Controls

It is expected that the *Tm* of a duplex can be ascertained by monitoring the amount of fluorescence given off by a donor in the presence of an acceptor molecule during temperature transition. Temperature changes are known to alter the amount of fluorescence produced by a donor without an acceptor molecule being present, leading to the possibility of obtaining erroneous melting temperature results. As a control, three different FAM-labelled oligonucleotides were exposed to the melt assay in two different ways; firstly on their own without an acceptor labelled oligo present, figure 4.10, and secondly in the form of a duplex with an acceptor labelled molecule present, figure 4.11. For each assay, buffer comprising 500 mM sodium chloride and 10 mM sodium phosphate had into it dissolved 0.5 µM of each oligonucleotide. The fluorescence intensity was monitored throughout as shown below. Oligonucleotide sequences are shown in table 4.21. A full list of oligonucleotide sequences are shown in section 6, appendix 6.1.1.2.

Figure 4.10	Figure 4.11
5' FAM, A, AAA	5'FAM, A, AAA / B' AAA – 3' DABCYL
5' FAM, A, TTT	5'FAM, A, TTT / B' AAA – 3' DABCYL
5'FAM, A, GGG	5'FAM, A, GGG / B' CCC – 3' DABCYL

Table 4.21 A list of oligonucleotide configurations for figures 4.10 and 4.11.

Figure 4.10

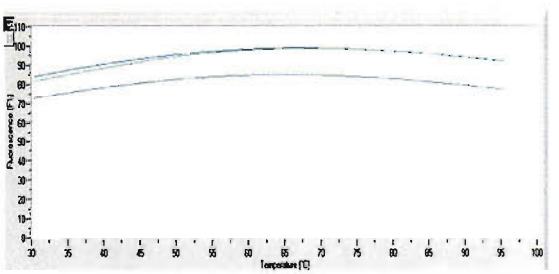


Figure 4.11

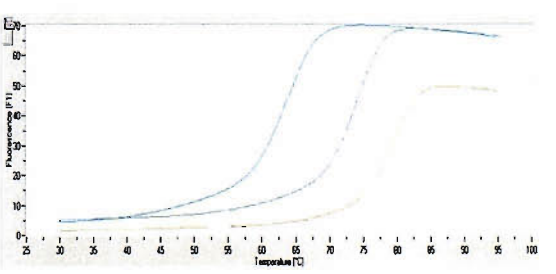


Figure 4.10 Fluorescence vs. temperature profiles (FAM labelled oligonucleotides only)
Figure 4.11 Fluorescence vs. temperature profiles (Duplexes formed between FAM labelled oligonucleotides and DABCYL labelled oligonucleotides).

The results showed that there was a small temperature dependent change in the absolute fluorescence of the FAM molecule. Figure 4.11 showed the classical temperature vs. fluorescence plot for duplexes exhibiting FRET. As the temperature was increased, a clear transition from quenched state to unquenched state corresponding to the point at which the duplex became separated was observed. Figure 4.10 showed a different profile altogether. In this example, no such transition was observed, although in general, as the temperature was increased, there was a steady rise in the amount of fluorescence detected. In this situation, in the absence of a complementary oligonucleotide, some primer-dimer formation may occur, bringing FAM fluorophores closer together, inducing a small FAM/FAM quenching. This serves to suppress some fluorescent signal, however the starting fluorescent intensity was much higher in this example. As the temperature is increased, any primer-dimer which resulted will be disrupted, moving closely aligned fluorophores apart. This reduces the amount of quenching which can occur, resulting in a slight increase in the amount of fluorescence, although the absence of a clear sigmoidal transition more likely indicates a small temperature dependent change in fluorescence.

To further demonstrate this difference, plots of -d(F)/dT (1st negative derivative) fluorescence vs. temperature for each of the previous plots were constructed, as shown in figures 4.12 and 4.13.

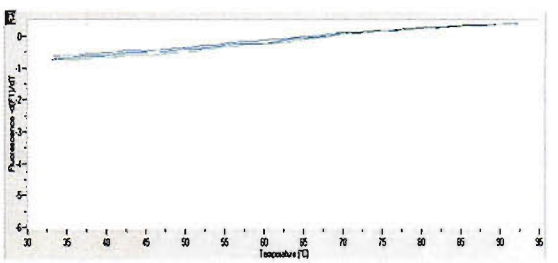


Figure 4.12 A plot of –d(F)/dT from figure 4.10.

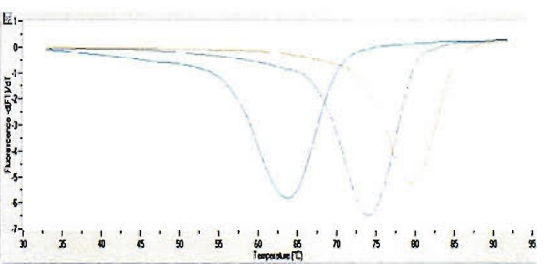


Figure 4.13 A plot of –d(F)/dT from figure 4.11.

The absence of melting temperature profiles in the first example compared to the second example demonstrated that there was very little change in fluorescence as the temperature was increased for FAM labelled oligonucleotides alone (figure 4.12). This was in contrast to

the second plot (4.13) where clear changes in fluorescence, depicted by the profiles were seen.

This result demonstrated that any changes in fluorescent intensity could be solely attributed to changes in duplex conformation, affecting the degree to which FRET could occur rather than be attributed to temperature induced changes in FAM fluorescence alone.

4.2.1.3 Quencher type

All Strand A sequences contained a 5' FAM fluorophore as the dye label. In a typical FRET system, FAM fluorescence is quenched by either DABCYL or ROX, which acts as the acceptor molecule. Initial investigations concerned determining which acceptor was the most suitable for quenching FAM fluorescence.

For this assay four sequences were selected; two strand A sequences and two strand B sequences. These sequences were designed according to the generic sequences in figure 4.9.

Strand A – oligo 1	5' FAM, A, GGG
Strand A – oligo 2	5' FAM, A, TTT
Strand B – oligo 3	B, CCC, 3' ROX
Strand B – oligo 4	B, CCC, 3' DABCYL

In each case a strand A oligo was mixed with a strand B oligo in the presence of 50 mM or 200 mM salt containing 10 mM sodium phosphate. These duplex combinations were then analysed using the melt analysis program. The results are shown in table 4.22.

Test	Strand A Oligo	Strand B Oligo	*Tm*°*C* (50 mM)	*Tm* °*C* (200 mM)
1	5' FAM, A, GGG	B, CCC-3' ROX	69.8	77.0
2	5' FAM, A, TTT	B, CCC-3' ROX	48.0	55.4
3	5' FAM, A, GGG	B, CCC-3' DABCYL	69.8	77.0
4	5' FAM, A, TTT	B, CCC-3' DABCYL	49.0	55.4

Table 4.22 *Tms* determined from four duplex pairs tested in the presence of 50 mM and 200 mM salt.

These results demonstrated that the *Tm* of a duplex pair, for a given salt concentration was similar regardless of the type of acceptor molecule used. The change in fluorescence during the different duplex melts showed that the relative fluorescent values corresponding to the FAM labels in each context were very similar (c.5 rfu rising to c.50 rfu). As the salt concentration was increased from 50 mM to 200 mM, there was an increase in melting temperature as expected. This increase was the same for both acceptor molecule types. For duplexes containing ROX, channel 2 of the LightCycler™ was also examined as this records the temperature dependent decrease in ROX fluorescence as the *Tm* is increased. No differences in Tm between the two different measurements were noted. These results support the previous observation demonstrating that the same *Tm* result can be obtained regardless of the type of acceptor molecule chosen.

4.2.1.4 ROX / DABCYL (acceptor) positioning

The position of the acceptor relative to the fluorophore can affect its ability to accept the energy given off from the fluorophore. In the previous example, the fluorophore and quencher were deliberately designed to sit adjacent to one another in the duplex configuration, causing collisional quenching. The converse of this is that the fluorophore and quencher reside at opposite ends of the duplex. This type of quenching is termed proximal quenching or FRET. Differences between the two types may be important, as the state of the duplex during a melt may yield different results as the fluorophore and quencher come apart at different times even though the duplexes are identical. To determine the degree to which results would change, as a result of differences in quenching mode, an assay using five sequences were selected; one strand A sequence and four strand B sequences. These sequences were designed according to the generic sequences in figure 4.9.

Strand A – oligo 1	5′ FAM, A, GGG
Strand B – oligo 2	5′ ROX, B, CCC
Strand B – oligo 3	5′ DABCYL, B, CCC
Strand B – oligo 4	B, CCC-3′ ROX
Strand B – oligo 5	B, CCC-3′ DABCYL

In each case, the strand A oligo was mixed with a strand B oligo in the presence of 200 mM salt and 10 mM sodium phosphate, prior to assessment through melt analysis. The results appear in table 4.23.

Test	Strand A Oligo	Strand B Oligo	*Tm* °C
1	5′ FAM, A, GGG	5′ ROX, B, CCC	75.3
2	5′ FAM, A, GGG	5′ DABCYL, B, CCC	75.8
3	5′ FAM, A, GGG	B, CCC-3′ ROX	76.1
4	5′ FAM, A, GGG	B, CCC-3′ DABCYL	76.3

Table 4.23 *Tms* determined from four duplex pairs. Test 1 and 2 reflect proximal quenching (FRET) where as test 3 and 4 reflect collisional quenching.

As with the previous result, there was virtually no difference in *Tm* between different fluorophore and quencher types (1v2 and 3v4). Comparing the quenching mode, (1v3 and 2v4 – proximal (FRET) versus collisional), there was a very slight increase in *Tm*, equivalent to less than 1 °C for the collisionally quenched duplexes. This increase suggested that the fluorophore and quencher were closer, for a slightly longer period of time during the melt transition stage, however, given that quenching was mediated through a collisional interaction, thereby bringing the fluorophore and quencher naturally closer together, this fact was not unsurprising.

For these candidate oligonucleotide sequences, the placement of the acceptor relative to the donor seems to demonstrate little difference in terms of the resulting *Tm,* however, there was a noticeable difference in the intensity of the FAM molecule between the different types of quenching. In tests that resulted in collisional quenching, initial FAM fluorescence was recorded as <5 rfu. During the melt, this fluorescence value rose to c.60 rfu as the quencher

and fluorophore became spatially separate. In contrast, tests that resulted in proximal FRET showed the initial FAM fluorescence to be at a much higher initial level, around c.50 rfu rising to c.60 rfu during melting. This elevated level of initial fluorescence was as a direct result of the quencher molecule being further away from the fluorophore and therefore not quenching the fluorescence to the same extent. This difference may have been greater had the duplexes been longer, placing the quencher even further away from the fluorophore in proximally quenched pairs.

In summary, the relative change in FAM fluorescence during melting was much lower for duplexes that exhibited proximal FRET compared to duplexes that exhibited collisional quenching. Adopting this later approach would suggest that the sensitivity of the assay could be increased as the relative change between quenched and unquenched states would be greater. For this reason, collisional quenching was chosen as the preferred method.

4.2.1.5 Salt dependence

As discussed in the introduction, the influence of salt on duplex melting temperature is of critical importance, particular for PCR/hybridisation applications. For a melting temperature investigation such as this, oligonucleotide duplexes must exhibit a salt dependency with respect to melting temperature in order to be of interest. In order to assess this, duplexes were created in the presence of different salt concentrations before undergoing a melt analysis. For the assay, four duplex types were created; two complementary pairs (1 and 4), and two unstable pairs each carrying three consecutive mismatches (2 and 3), as follows:

Duplex 1	5′ FAM, A, GGG / B, CCC-3′ DABCYL
Duplex 2	5′ FAM, A, GGG / B, AAA-3′ DABCYL
Duplex 3	5′ FAM, A, TTT / B, CCC-3′ DABCYL
Duplex 4	5′ FAM, A, TTT / B, AAA-3′ DABCYL

Duplex pairs were dissolved in either 50 mM, 200 mM or 500 mM sodium chloride containing 10 mM sodium phosphate. Results are shown in figure 4.14.

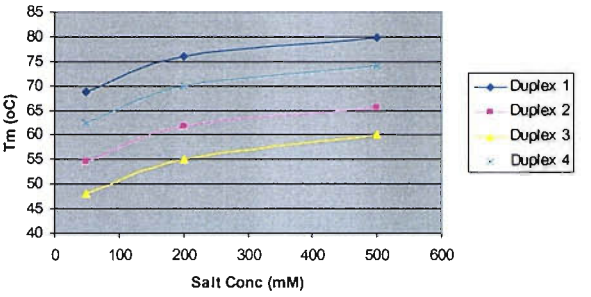


Figure 4.14 A graph of *Tm* °C for four duplex pairs at varying salt concentrations

These results showed a salt dependence over the range of salt concentrations tested. This dependence was the same for complementary duplex pairs and duplex pairs that contained deliberate mismatches. This demonstrated that the mode of testing was sufficient to characterise the affect of varying monovalent salt concentrations.

It is well characterised that divalent ions such as magnesium chloride also serve to stabilise duplex structure. As such the degree to which magnesium chloride altered melting temperature was also considered. Duplex pairs 1 and 3 from the previous example were prepared in 50 mM sodium chloride, 10 mM sodium phosphate. Duplicate samples were setup, containing 1.5 mM magnesium chloride. Following a melt analysis the following results were recorded:

Duplex	*Tm* °C (- $MgCl_2$)	*Tm* °C (+ $MgCl_2$)
1	69.6	71.6
3	48.4	51.0

Table 4.24 Melting temperatures for duplex pairs 1 and 3 in the presence /absence of 1.5 mM $MgCl_2$.

The addition of magnesium chloride resulted in a small increase in melting temperature as indicated by the two results. This increase was similar for both duplexes, indicating that the

degree of stability was the same for complementary and non-complementary states. These results demonstrated that magnesium chloride does serve to stabilise a duplex however, the degree of stabilisation is small compared to the previous result for sodium chloride, where a much larger dependence was demonstrated.

4.2.1.6 Determining a suitable melting temperature range (maximum vs. minimum)

The introduction of deliberate mismatches to a duplex will naturally cause some degree of destabilisation, reducing its associated melting temperature. The greater the number of mismatches the more unstable the duplex will become. For the test oligonucleotide duplex, the maximum possible number of mismatches that can be deliberately introduced is three. The stability of this duplex therefore is determined by the ratio of mismatch base pairs to matched base pairs (3 from 19). Assuming the test oligonucleotide duplex was shorter (17 base pairs), three mismatches should have a greater destabilising effect (3 from 17), as the proportion of mismatches to matches increases. This difference may have important consequences for being able to accurately differentiate duplex melting temperatures. For this reason, a back to back assay was set up to compare the effect of oligonucleotide duplex length on maximum/minimum melting temperatures and therefore determine a melting temperature range.

Maximum melting temperatures were derived from perfectly complementary duplexes, containing three consecutive CG base pairs at the three variable positions. Minimum melting temperature were derived from duplexes containing three consecutive mismatches, of the form C•T, as this was previously reported as being the most unstable (SantaLucia *et al.,* 1998). Two different duplex lengths were tested (sequences shown in appendix 6.1.1.1 / 6.1.1.3);

- stable and unstable 19-mers and
- stable and unstable 17-mers.

These were subjected to melt analysis in the presence of 50, 100, 200 and 400 mM sodium chloride, 10 mM sodium phosphate, as shown in tables 4.25 and 4.26.

Oligo duplex (19-mer)	*Tm* °C			
	50 mM	**100 mM**	**200 mM**	**400 mM**
5′ FAM, A, GGG / B, CCC-3′ DABCYL	69.5	73.2	77.5	79.4
5′ FAM, A, TTT / B, CCC-3′ DABCYL	47.5	52.2	56.5	58.7
Temperature window	**22.0**	**21.2**	**21.00**	**20.7**

Table 4.25 Melting temperatures for stable and unstable 19-mer duplexes in the presence of differing sodium chloride concentrations (mM).

Oligo duplex (17-mer)	*Tm* °C			
	50 mM	**100 mM**	**200 mM**	**400 mM**
5′ FAM, A, GGG / B, CCC-3′ DABCYL	65.8	69.6	73.2	76.2
5′ FAM, A, TTT / B, CCC-3′ DABCYL	40.0	44.6	47.9	51.8
Temperature window	**25.8**	**25.0**	**25.3**	**24.4**

Table 4.26 Melting temperatures for stable and unstable 17-mer duplexes in the presence of differing sodium chloride concentrations (mM).

The result showed that the temperature range for the 19-mer duplex was c. 22 °C. For the 17-mer this was slightly higher (c. 25 °C), as predicted. The ranges were consistent for each salt concentration tested.

These results demonstrated that is were possible to obtain a larger melting temperature range by decreasing the duplex length. This window operated between 40 °C and 65 °C compared to the longer duplexes whose window lay between 47 °C and 70 °C. The measurable range for the assay, as defined by the melt analysis program operated between 35 °C and 95 °C. As a consequence, the 19-mer oligonucleotide length was selected as the more suitable oligonucleotide length since the *Tm* window between stable and un-stable duplexes fell centrally within this range.

The original 19-mer oligonucleotides were designed such that the 5′ and 3′ regions should have equivalent *Tm*s. A duplex which contains three unpaired bases in the centre will melt according to the *Tm* of the two ends (which are equivalent) and therefore melt at the same temperature. The minimum melting temperature that could be expected therefore would result from one of the strands (A or B), forming a duplex with a shortened 8-mer oligonucleotide complementary to either end of this stand.

As such, the following 5′ FAM labelled oligonucleotide was synthesised:

5'FAM-CGACGTGC same as 5′ region of strand A; therefore complementary to Strand B

This 8-mer oligonucleotide was mixed with two different strand B oligonucleotides, at two different salt concentrations (containing 10 mM sodium phosphate) as follows:

Duplex	50 mM (NaCl)	200 mM (NaCl)
5′ FAM, A, 8-mer / B, AAA-3′ DABCYL	42.3	50.7
5′ FAM, A, 8-mer / B, CCC-3′ DABCYL	39.0	48.1

Table 4.27 Melting temperatures for the 8-mer oligonucleotide forming a duplex with two different 19-mer sequences.

Although the 8-mer sequences interact with the two 19-mer oligos in the same place, there is a clear difference in *Tm* between the two different 19-mer sequences tested. This difference can be attributed to the stabilisation effects of overhangs as described by Doktycz *et al.,* (1990); Bommarito *et al.,* (2000) and suggests that proximal purine overhangs give greater stability than pyrimidines.

The minimum *Tm* that could be expected from this investigation would be from a duplex containing 3 unpaired bases in the centre; in the context above, this would be from the duplex containing three C-base overhangs. At 50mM sodium chloride, the minimum *Tm* threshold was 39.0 °C. This easily falls within the *Tm* windows previously described, demonstrating that the 19-mer duplex configuration would be suitable for the investigation.

These preliminary results demonstrated the viability of the assay for determining the stability of duplexes containing multiple mismatches. The next section compares the experimental melting temperatures with those derived from theoretical considerations.

4.2.2 Calculating a theoretical Melting Temperature (Tm)

For comparison, two different predicted *Tm* calculations were used. These were based on the nearest neighbour models outlined in section 1.4, as follows:

1. **Primer Express™ (based on the parameters of Breslauer *et al.*, 1986)**

   $Tm\ ^{o}C = ((\Delta H^{o}/\Delta S^{o}+(R\ln(C_T))) - (273.15 + 16.6\mathrm{Log}[Na^{+}])$

2. **MOSAIC (based on the parameters of Allawi and SantaLucia, 1997)**

   $Tm\ ^{o}C = ((\Delta H^{o}/\Delta S^{o}+(R\ln(C_T))) - (273.15 + 12.5\mathrm{Log}[Na^{+}])$

In each case, R = Molar gas constant (1.987 cal/kmol), C_T = Oligo concentration ($5x10^{-7}$ M), $[Na^{+}]$ = sodium chloride concentration (50 mM). **NB. ΔH^{o}, ΔS^{o} and initiation parameters were different between the Breslauer and SantaLucia Data.**

Using previously synthesised oligonucleotides, a comparison of predicted melting temperature with actual melting temperature was performed as shown below:

Duplex	Actual *Tm* °C	Primer Express™ *Tm* °C	MOSAIC *Tm* °C
5'FAM,A, GGG/B, CCC-3' DABCYL	68.6	69.7	62.5
5'FAM,A, TTT/B, AAA-3' DABCYL	62.5	59.0	56.6

Table 4.28 A comparison of duplex *Tm*, between Actual and predicted models.

Melting temperature values derived from actual *Tm* melts were different to the values predicted by Primer Express™. Although the stable duplex (containing central GC pairs) prediction was within 1.5 °C, the prediction for the less stable duplex (containing AT pairs)

was nearly 3.5 °C out. Predictions from MOSAIC were even further away from actual melting temperatures, with both duplex predictions falling below their actual melting temperatures by c.6.0 °C.

It was important that *Tm* predictions were as close to actual melting temperatures as possible, at least for complementary sequences. This would provide a suitable base to calculate duplex stability as a function of melting temperature for mismatched duplexes. To address this issue, design of a more accurate duplex *Tm* prediction model was investigated. This involved adapting the MOSAIC *Tm* calculator by identifying trends in melting temperature through a series of duplex melting assays.

4.2.3 Duplex melting assay

Using the generic sequences for strand A and strand B, all possible variant oligonucleotides were constructed. This comprised the manufacture of 64-strand A sequences and 64-strand B sequences. These are shown in appendix 6.1.1.1.

Each Strand A oligonucleotide (0.5 µM), in turn, was mixed with each strand B oligonucleotide (0.5 µM) in the presence of 50 mM sodium chloride and 10 mM sodium phosphate buffer. This resulted in 4096 different combinations, indicated by the sequence (1)NN' (2)NN' (3)NN', where N refers to strand A sequence and N' refers to strand B sequence. Each combination was subjected to a melt analysis.

A full list of experimental *Tm* values for all combinations is shown in appendix 6.2.

4.2.4 Validation of MOSAIC as an accurate Tm prediction model

For the MOSAIC validation, only perfectly matching sequences from the melting analyses were considered, of which there were 64 different data values. These are shown in table 4.29. NB. Duplexes are referred to by the base positions 1,2 and 3 of the variable three

central bases of strand A alone; (duplex AAA refers to strand A, sequence AAA forming a duplex with strand B, sequence TTT).

ID	Duplex	Actual Tm	Primer Exp.	MOSAIC
1	AAA	62.39	62.5	56.61
2	AAC	64.53	63.7	58.52
3	AAG	63.84	62.7	57.98
4	AAT	61.16	59.9	55.56
5	ACA	63.00	63.7	58.52
6	ACC	65.16	66.2	60.53
7	ACG	65.44	66.3	60.35
8	ACT	60.66	60.1	57.41
9	AGA	64.20	62.7	57.98
10	AGC	66.13	66.4	60.79
11	AGG	65.98	65.2	60.00
12	AGT	62.65	60.1	57.41
13	ATA	61.75	59.9	55.56
14	ATC	64.84	63.7	58.06
15	ATG	63.64	63.7	58.06
16	ATT	60.66	59.9	55.56
17	CAA	64.54	64.9	58.61
18	CAC	66.43	66.2	60.53
19	CAG	66.46	65.2	60.00
20	CAT	63.00	62.3	57.58
21	CCA	66.03	67.2	60.64
22	CCC	67.73	69.6	62.69
23	CCG	68.22	69.7	62.48
24	CCT	64.44	63.8	59.53
25	CGS	66.51	67.3	60.46
26	CGC	69.08	70.8	63.23
27	CGG	68.24	69.7	62.48
28	CGT	66.22	64.9	59.90
29	CTA	63.55	61.4	57.50
30	CTC	65.83	65.2	60.00
31	CTG	65.87	65.2	60.00
32	CTT	62.85	61.4	57.50

ID	Duplex	Actual Tm	Primer Exp.	MOSAIC
33	GAA	64.74	65.0	58.47
34	GAC	67.23	66.3	60.35
35	GAG	66.45	65.3	59.83
36	GAT	64.07	62.5	57.44
37	GCA	67.13	68.5	61.23
38	GCC	68.72	70.8	63.23
39	GCG	69.32	70.9	63.02
40	GCT	66.43	65.1	60.16
41	GGA	66.63	67.3	60.46
42	GGC	68.92	70.8	63.23
43	GGG	67.73	69.7	62.48
44	GGT	65.96	64.9	59.90
45	GTA	63.93	62.5	57.90
46	GTC	66.36	66.3	60.35
47	GTG	66.23	66.3	60.35
48	GTT	63.55	62.5	57.90
49	TAA	61.55	59.0	55.48
50	TAC	63.55	60.1	57.41
51	TAG	63.55	59.2	56.87
52	TAT	61.31	56.4	54.43
53	TCA	64.24	62.7	57.98
54	TCC	66.06	65.2	60.00
55	TCG	66.63	65.3	59.83
56	TCT	63.00	59.2	56.87
57	TGA	64.36	62.7	57.98
58	TGC	67.45	66.4	60.79
59	TGG	65.54	65.2	60.00
60	TGT	62.05	60.1	57.41
61	TTA	61.14	59.0	55.41
62	TTC	64.38	62.7	57.98
63	TTG	64.86	62.7	57.98
64	TTT	61.04	59.0	55.48

Table 4.29 show a comparison of actual *Tm* °C to predicted *Tm* °C from (1) Primer Express™ and (2) MOSAIC for duplexes containing only Watson Crick base pairs.

As previously outlined in section 4.2.2, there were noticeable differences between predicted *Tm* values and actual *Tm* values. For Primer Express™, the predicted values were much closer than those predicted from MOSAIC as shown in figure 4.15.

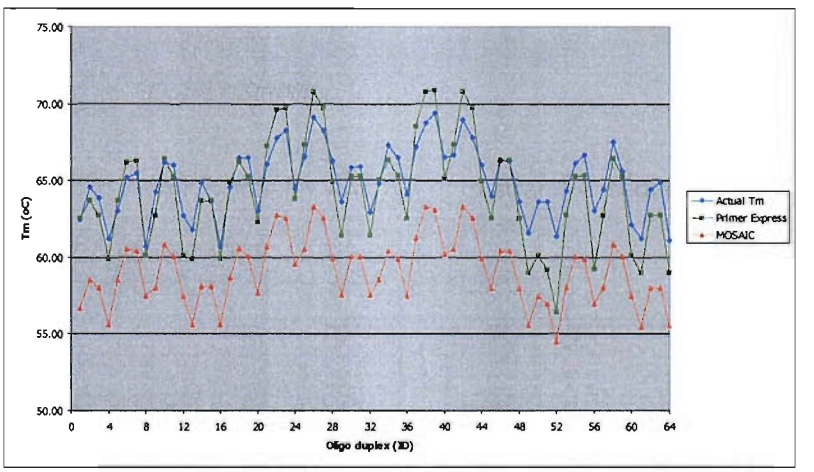


Figure 4.15 A plot of actual *Tm* together with plots of predicted *Tms* from Primer Express™ and MOSAIC for 64 different duplex combinations (as shown in table 4.29).

Despite these initial observations, the MOSAIC data seemed to align itself with a high degree of fit to the actual *Tm* data (after adjusting the data using method described in 4.2.4.1), compared to a similar alignment between the actual *Tm* and Primer Express™ *Tm* data. (See figure 4.16).

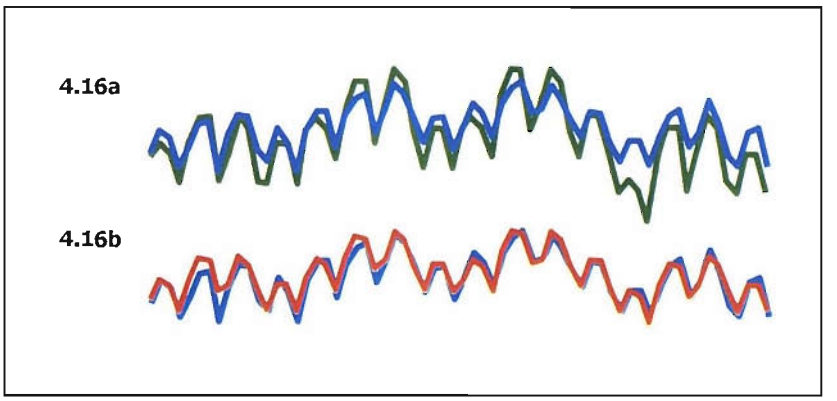


Figure 4.16 Artificial alignment of data from figure 4.15, between (a) Actual *Tm* and Primer Express™ *Tm*, (table 4.29), and (b) Actual and MOSAIC *Tm* , (table 4.29) after adjusting data.

This demonstrated that the original *Tm* prediction model (MOSAIC) was able to predict a similar trend to the experimental data except for the fact that it was shifted down by a few degrees. This shift was thought to arise from an overestimate of the salt correction factor, which reduces the calculated melting temperature as the salt concentration decreases. According to SantaLucia, their salt correction factor was only suitable for salt concentrations above 300 mM sodium chloride, hence to accurately predict the melting temperature of duplexes in lower salt environments, a new modified salt correction factor would need to be identified.

4.2.4.1 Adjusting the salt correction factor using Solver™.

Using the melting temperature prediction equation of Allawi and SantaLucia;

$$Tm\ °C = ((\Delta H°/\Delta S°+(R\ln(C_T))) - (273.15 + \mathbf{12.5}\ \mathrm{Log}[Na^+])$$

An attempt was made to optimise the salt correction value (highlighted in red), using the experimental data for complementary duplexes.

All of the data were loaded into a Microsoft Excel spreadsheet. For each of the 64 duplex melting temperatures, the absolute differential (actual *Tm* - predicted *Tm*) was calculated as follows:

e.g. Duplex AAA = Actual *Tm* °C (62.39) – Predicted *Tm* °C (56.61)

Differential (*diff.*) = +5.78 °C

Absolute differential (*Abs.diff.*) = 5.78 (ignore sign)

Summing the individual (*Abs.diff.*) values for each data point gave the total *Abs.diff.* value for the data set; equal to 374.86 as shown in table 4.30.

ID	Duplex	Actual Tm	MOSAIC	Absolute Residual
1	AAA	62.39	56.61	5.78
2	AAC	64.53	58.52	6.01
3	AAG	63.84	57.98	5.86
4	AAT	61.16	55.56	5.60
5	ACA	63.00	58.52	4.48
6	ACC	65.16	60.53	4.63
7	ACG	65.44	60.35	5.09
8	ACT	60.66	57.41	3.25
9	AGA	64.20	57.98	6.22
10	AGC	66.13	60.79	5.34
11	AGG	65.98	60.00	5.98
12	AGT	62.65	57.41	5.24
13	ATA	61.75	55.56	6.19
14	ATC	64.84	58.06	6.78
15	ATG	63.64	58.06	5.58
16	ATT	60.66	55.56	5.10
17	CAA	64.54	58.61	5.93
18	CAC	66.43	60.53	5.90
19	CAG	66.46	60.00	6.46
20	CAT	63.00	57.58	5.42
21	CCA	66.03	60.64	5.39
22	CCC	67.73	62.69	5.04
23	CCG	68.22	62.48	5.74
24	CCT	64.44	59.53	4.91
25	CGS	66.51	60.46	6.05
26	CGC	69.08	63.23	5.85
27	CGG	68.24	62.48	5.76
28	CGT	66.22	59.90	6.32
29	CTA	63.55	57.50	6.05
30	CTC	65.83	60.00	5.83
31	CTG	65.87	60.00	5.87
32	CTT	62.85	57.50	5.35

ID	Duplex	Actual Tm	MOSAIC	Absolute Residual
33	GAA	64.74	58.47	6.27
34	GAC	67.23	60.35	6.88
35	GAG	66.45	59.83	6.62
36	GAT	64.07	57.44	6.63
37	GCA	67.13	61.23	5.90
38	GCC	68.72	63.23	5.49
39	GCG	69.32	63.02	6.30
40	GCT	66.43	60.16	6.27
41	GGA	66.63	60.46	6.17
42	GGC	68.92	63.23	5.69
43	GGG	67.73	62.48	5.25
44	GGT	65.96	59.90	6.06
45	GTA	63.93	57.90	6.03
46	GTC	66.36	60.35	6.01
47	GTG	66.23	60.35	5.88
48	GTT	63.55	57.90	5.65
49	TAA	61.55	55.48	6.07
50	TAC	63.55	57.41	6.14
51	TAG	63.55	56.87	6.68
52	TAT	61.31	54.43	6.88
53	TCA	64.24	57.98	6.26
54	TCC	66.06	60.00	6.06
55	TCG	66.63	59.83	6.80
56	TCT	63.00	56.87	6.13
57	TGA	64.36	57.98	6.38
58	TGC	67.45	60.79	6.66
59	TGG	65.54	60.00	5.54
60	TGT	62.05	57.41	4.64
61	TTA	61.14	55.41	5.73
62	TTC	64.38	57.98	6.40
63	TTG	64.86	57.98	6.88
64	TTT	61.04	55.48	5.56
		Total Abs(residual)		374.86

Table 4.30 Calculated absolute residual total for observed (Actual *Tm* °C) – Predicted *Tm* °C.

In order to transform the data, so that predicted *Tm* fell more inline with actual *Tm*, the analysis software Solver™, within Microsoft Excel© was employed. Within the equation for *Tm* prediction, Solver™ was used to automatically iterate the target value (salt correction factor), to return an answer based on the minimum residuals calculated from Abs. Σ(observed *Tm* °C – expected *Tm* °C); in this case, the minimum possible total *Abs.diff.*

value. The returned value from Solver™ represented the optimum salt correction value which would give rise to the smallest summed difference between predicted *Tm* and observed *Tm* across the data set. In this example, Solver™ returned the value 7.962. This gave an overall total *Abs.diff.* value equal to 30.23 following recomputation of the predicted melting temperature, for each of the 64 duplex pairs. See table 4.31/figure 4.17.

ID	Duplex	Actual Tm	MOSAIC	Absolute Residual
1	AAA	62.39	62.51	0.12
2	AAC	64.53	64.42	0.11
3	AAG	63.84	63.89	0.05
4	AAT	61.16	61.47	0.31
5	ACA	63.00	64.42	1.42
6	ACC	65.16	66.43	1.27
7	ACG	65.44	66.26	0.82
8	ACT	60.66	63.31	2.65
9	AGA	64.20	63.89	0.31
10	AGC	66.13	66.69	0.56
11	AGG	65.98	65.91	0.07
12	AGT	62.65	63.31	0.66
13	ATA	61.75	61.47	0.28
14	ATC	64.84	63.96	0.88
15	ATG	63.64	63.96	0.32
16	ATT	60.66	61.47	0.81
17	CAA	64.54	64.52	0.02
18	CAC	66.43	66.43	0.00
19	CAG	66.46	65.91	0.55
20	CAT	63.00	63.48	0.48
21	CCA	66.03	66.55	0.52
22	CCC	67.73	68.60	0.87
23	CCG	68.22	68.39	0.17
24	CCT	64.44	65.44	1.00
25	CGS	66.51	66.37	0.14
26	CGC	69.08	69.14	0.06
27	CGG	68.24	68.39	0.15
28	CGT	66.22	65.80	0.42
29	CTA	63.55	63.40	0.15
30	CTC	65.83	65.91	0.08
31	CTG	65.87	65.91	0.04
32	CTT	62.85	63.40	0.55

ID	Duplex	Actual Tm	MOSAIC	Absolute Residual
33	GAA	64.74	64.37	0.37
34	GAC	67.23	66.26	0.97
35	GAG	66.45	65.74	0.71
36	GAT	64.07	63.35	0.72
37	GCA	67.13	67.14	0.01
38	GCC	68.72	69.14	0.42
39	GCG	69.32	68.92	0.40
40	GCT	66.43	66.06	0.37
41	GGA	66.63	66.37	0.26
42	GGC	68.92	69.14	0.22
43	GGG	67.73	68.39	0.66
44	GGT	65.96	65.80	0.16
45	GTA	63.93	63.80	0.13
46	GTC	66.36	66.26	0.10
47	GTG	66.23	66.26	0.03
48	GTT	63.55	63.80	0.25
49	TAA	61.55	61.39	0.16
50	TAC	63.55	63.31	0.24
51	TAG	63.55	62.77	0.78
52	TAT	61.31	60.33	0.98
53	TCA	64.24	63.89	0.35
54	TCC	66.06	65.91	0.15
55	TCG	66.63	65.74	0.89
56	TCT	63.00	62.77	0.23
57	TGA	64.36	63.89	0.47
58	TGC	67.45	66.69	0.76
59	TGG	65.54	65.91	0.37
60	TGT	62.05	63.31	1.26
61	TTA	61.14	61.31	0.17
62	TTC	64.38	63.89	0.49
63	TTG	64.86	63.89	0.97
64	TTT	61.04	61.39	0.35
		Total Abs (residual)		30.23

Table 4.31 Recomputed predicted *Tms* °C (MOSAIC) assuming the salt correction factor to be 7.962. The absolute residual total for observed (Actual) – Predicted *Tm* is also shown.

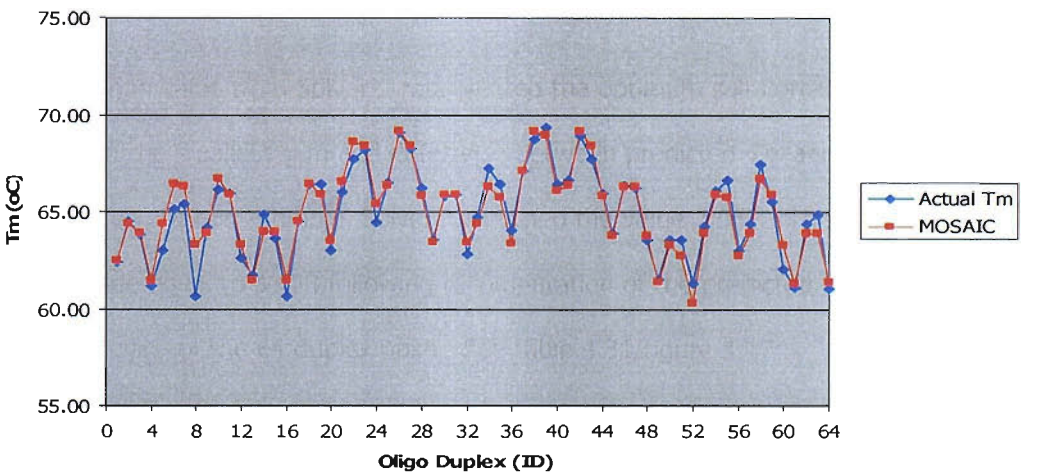


Figure 4.17 A plot of predicted *Tm* (MOSAIC) and Actual *Tm* for each oligo duplex pair.

Substituting the adjusted salt correction value (7.962) into the *Tm* prediction model (MOSAIC), resulted in a high degree of alignment between the observed and predicted data sets.

To determine the closeness of fit between actual *Tm* and Primer Express™ and MOSAIC predicted *Tms*, the mean absolute deviation (MAD), for each prediction model was calculated according to the equation:

Equation 17

$$MAD = \frac{1}{N}\sum_{x=N}^{x=1}\left(actualTm_x - predictedTm_x\right)^2$$

The calculated values appear below.

Prediction model	Mean Absolute Deviation
Primer Express™	2.78
MOSAIC	0.42

Table 4.32 Mean Absolute Deviation values for two duplex *Tm* prediction models; Primer Express™ and MOSAIC.

The MAD results demonstrated that there was a much better fit from the MOSAIC model compared to Primer Express™ for the 64 duplexes tested, indicated by the smaller MAD value. Despite the smaller mean absolute deviation from the MOSAIC data, some differences between the actual and predicted *Tms* were evident. To depict the range of residual values from the MOSAIC program, the 64 data points were sorted according to ascending residual (Actual *Tm* – MOSAIC *Tm*), as shown in figure 4.18.

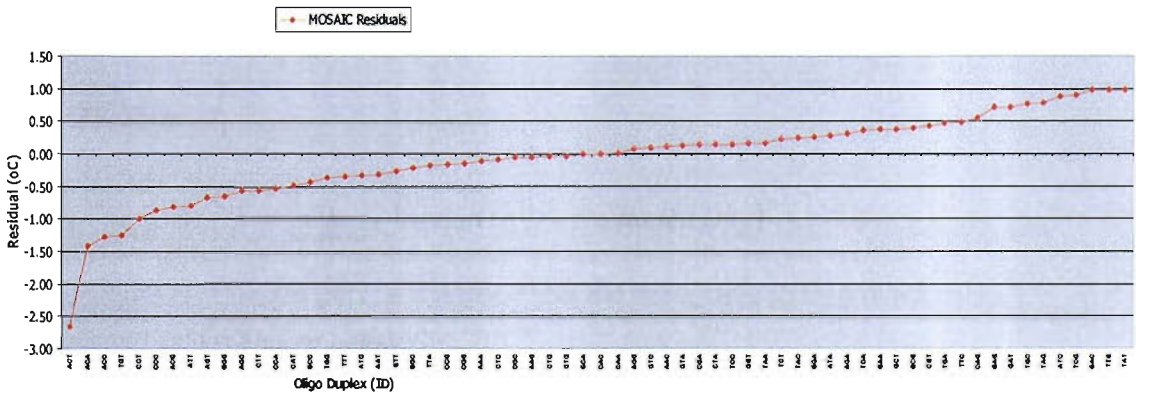


Figure 4.18 A plot of the residuals (Actual *Tm* – MOSAIC *Tm*) for each of the 64 duplexes sorted into ascending order.

The residuals for each of the duplexes ranged from -2.65 °C to +1 °C, although in total, 41 from the 64 fell within the range ± 0.5 °C. To further illustrate the data, a box plot of residuals derived from Primer Express™, and MOSAIC was constructed as follows:

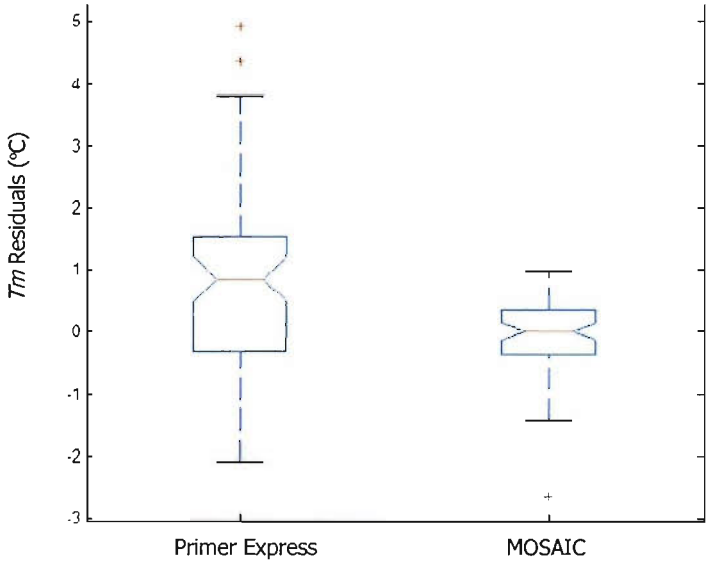


Figure 4.19 Box plots of *Tm* residuals (°C), derived from Primer Express™ data and MOSAIC data.

Figure 4.19 showed the median residual for the MOSAIC data to be 0, with greater than 50% of the data points lying within ± 0.36 °C. The data point, -2.65, lay outside the lower whisker confidence limit, indicating that it could be regarded as an outlier. In contrast, the median residual for the Primer Express™ data was higher than that for MOSAIC, equal to 0.84. The overall spread of data was also much greater with a total of two data points identified as suspect outliers.

Overall, these results indicated that the MOSAIC prediction was better than Primer Express™, despite the prevalence of small differences as indicated by the box plots. Considering only duplexes whose residual fell outside the box, there did not appear to be a relationship between them and the discrepancy in predicted *Tm*. For these duplexes, one factor that may have caused the predicted *Tm* to be different from actual *Tm* was the accuracy of the experimental *Tm* data (experimental error). This was tested by undertaking a reproducibility study.

4.2.4.2 Experimental reproducibility

To confirm the validity of the experimental procedure, pairs of complementary duplexes were retested in replicates of ten, using the standard oligo melt assay. In addition, an extended oligo melt assay was performed. This comprised reducing the temperature transition rate from 0.1 to 0.05 °C/s, to ensure melting temperatures were indicative of thermodynamics and not as a result of kinetic differences. Oligonucleotide duplexes that appeared to predict more or less exact *Tm* s were selected as were duplexes that appeared at both extremes (where maximum/minimum residuals were observed).

	Duplex pair					
Replicate	**ACT/TGA**	**ACA/TGT**	**TCT/AGA**	**TAC/ATG**	**GAC/CTG**	**TAT/ATA**
1	60.8	62.7	63.1	63.5	67.1	61.3
2	60.9	62.9	63.0	63.8	67.1	61.2
3	60.7	62.9	62.9	63.8	66.9	61.3
4	60.7	62.9	63.2	63.6	67.2	61.3
5	60.8	62.8	63.0	63.6	67.2	61.3
6	60.5	62.7	63.0	63.8	67.2	61.5
7	60.7	62.9	62.9	63.5	67.1	61.3
8	60.7	62.8	63.2	63.9	67.1	61.3
9	60.8	62.9	62.9	63.6	67.2	61.3
10	60.8	62.8	63.2	63.7	67.1	61.3
Mean *Tm(°C)*	**60.8**	**62.8**	**63.0**	**63.7**	**67.2**	**61.3**
St. Deviation	**0.106**	**0.110**	**0.119**	**0.137**	**0.092**	**0.109**
MOSAIC *Tm(°C)*	**63.3**	**64.4**	**62.8**	**63.3**	**66.3**	**60.3**
New Residual	**-2.55**	**-1.60**	**0.27**	**0.36**	**0.89**	**0.98**
Old Residual	**-2.65**	**-1.42**	**0.23**	**0.24**	**0.97**	**0.98**

Table 4.33 Melting temperature data for 6 duplex pairs analysed with the standard melt assay. In each case, the predicted *Tm* was deducted from the mean *Tm* to determine the new residual.

Table 4.33 shows data from the standard oligo melt assay only. *Tms* for each duplex pair in this assay were directly compared to the same duplex pairs in the long oligo melt assay and no differences were seen. This confirmed that the actual *Tm* observations were the result of duplex thermodynamics and not duplex melting kinetics. The data in table 4.33 showed that

experimental *Tm* reproducibility was high, given the small standard deviations for each duplex pair. These findings confirmed that the residual values were in fact the result of true experimental observations and not experimental errors. The average standard deviation for the 6 duplex replicates was calculated as 0.109. This was multiplied by 3 to determine the threshold limits for experimental variation; in this case determined to be ± 0.33 °C. This confidence level tied in very well with the box plot result from earlier, which had suggested that 50% of the data fell between the limits of ± 0.36 °C. This meant that for a large number of the differences between observed and predicted *Tm,* the difference could be explained by experimental variation and not from errors in *Tm* measurement. For duplexes whose actual *Tm* was found to be outside the predicted *Tm* range, their *Tms* were shown to be real observations. In addition, the outlier for the duplex ACT was repeated 10 times, producing a series of actual *Tms* that were consistently 2.5 to 2.65 °C below the MOSAIC prediction.

Generally, these findings confirmed that the improved salt correction factor led to a reliable prediction of *Tm* and that this prediction was better than Primer Express™. The next phase was to test this improved model, with truly unknown duplexes. This was achieved by constructing a series of novel oligonucleotide duplexes referred to as validation(I) duplexes (see appendix 6.1.1.4). These varied in both GC content and duplex length from the original set of 64. All duplexes were tested to determine their actual melting temperature; the results are shown in table 4.34.

Duplex ID	Length	Actual Tm	Primer Exp.	MOSAIC	(a) Residual	(b) Residual
1	10	46.96	26.90	41.83	20.06	5.13
2	10	41.40	15.30	35.40	26.10	6.00
3	15	57.68	52.70	57.69	4.98	0.00
4	15	50.41	40.30	49.60	10.11	0.81
5	20	66.88	66.40	68.12	0.48	-1.25
6	20	63.75	62.90	63.58	0.85	0.17
7	25	71.96	75.30	73.53	-3.34	-1.57
8	26	67.80	75.60	70.47	-7.80	-2.67
9	28	73.53	79.10	75.59	-5.57	-2.06
10	28	63.41	66.60	68.09	-3.19	-4.68

Table 4.34 Actual *Tm* values compared to Primer Express™ and MOSAIC predictions. Residuals were calculated from (a) Actual *Tm* – Primer Express™ *Tm*, (b) Actual *Tm* – MOSAIC *Tm*.

The predicted *Tm* from MOSAIC was calculated using the modified salt correction factor (7.962). In each case, the predicted *Tm* was different to the actual *Tm* as shown by the calculated residuals. This suggested that the salt correction factor derived earlier was not accurate for the duplex lengths tested. In addition, the spread of the data suggested that duplex *Tms* with lengths closest to 19 bp were more accurately predicted, whilst smaller and longer duplexes showed more extreme deviations. This observation implied a length dependency to the salt correction factor as originally indicated by Giesen *et al.,* (1998). In a similar approach to that adopted for determining the optimum salt correction factor, Solver™ was used to transform the predicted *Tm* to exactly match actual *Tm* for each duplex length as shown in table 4.35.

Duplex ID	Length	Actual Tm	MOSAIC	Salt Corr.
1	10	46.96	46.96	4.02
2	10	41.40	41.40	3.35
3	15	57.68	57.68	7.97
4	15	50.41	50.41	7.34
5	20	66.88	66.88	8.92
6	20	63.75	63.75	7.83
7	25	71.96	71.96	9.17
8	26	67.80	67.80	10.02
9	28	73.53	73.53	9.55
10	28	63.41	63.41	11.56

Table 4.35 shows the adjusted salt correction factors required to transform each MOSAIC *Tm* prediction such that predicted *Tm* was equal to actual *Tm.*

Previously it had been shown that for a 19 bp duplex, the optimised salt correction factor was 7.962. The closest duplex size tested in this example (20 bp), showed similar salt correction factor values. For smaller duplexes, the salt correction factor decreased indicating that the predicted *Tm* more closely matched the actual *Tm.* In contrast, longer duplexes required higher salt correction values indicative of actual *Tms* that were lower than predicted estimates. Different salt correction value estimates for duplexes of the same length reflected differences in the %GC content of the specific duplex; each pair of duplexes, of the same length, represented a high (>60%) GC content and low (<60%) GC content, respectively.

To establish a relationship between oligonucleotide length and salt correction factor, a third order polynomial regression was fitted to data from table 4.35, columns 2 and 5. This type of regression was selected as best fit to the data in preference to a simple linear regression as this method was shown to minimise/optimise the residuals between the observed data and fitted line. Higher order regression showed no improvement.

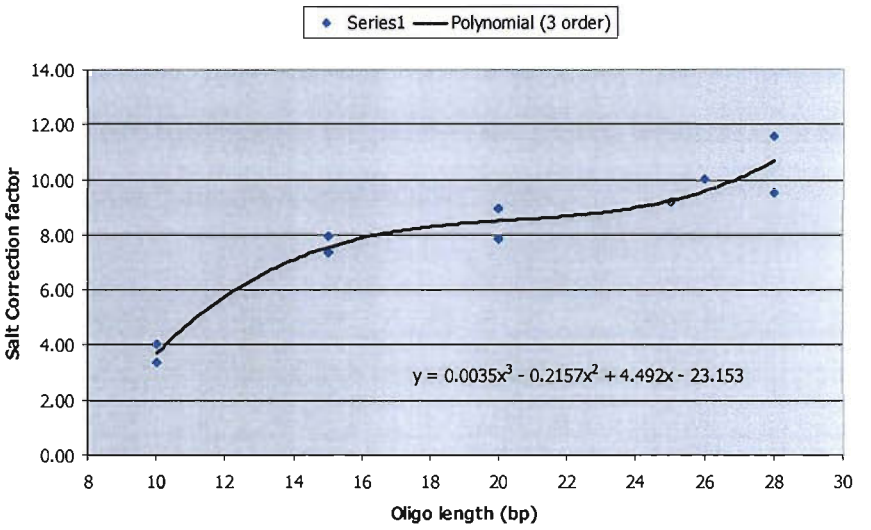


Figure 4.20 A plot of Oligonucleotide length (bp) vs. salt correction factor. A 3 order polynomial regression line, together with its equation is overlaid to describe the data trend.

The regression equation depicted in figure 4.20 was used to calculate the salt correction factor for each duplex length. This revised factor was then used to calculate the revised MOSAIC *Tm* prediction as shown in table 4.36.

Duplex ID	Length	Actual Tm	MOSAIC	Salt Corr.	Residual
1	10	46.96	47.38	3.70	-0.42
2	10	41.40	40.95	3.70	0.45
3	15	57.68	58.28	7.51	-0.60
4	15	50.41	50.19	7.51	0.22
5	20	66.88	67.54	8.41	-0.67
6	20	63.75	63.00	8.41	0.75
7	25	71.96	72.15	9.02	-0.19
8	26	68.51	68.68	9.34	-0.17
9	28	73.53	72.49	10.35	1.04
10	28	63.41	64.99	10.35	-1.58

Table 4.36 Revised salt correction factors derived from the regression equation in figure 4.20, together with the recalculated MOSAIC *Tm* prediction.

The residual values in table 4.36 were compared directly with the residual values (b), table 4.34. The revised salt correction factors calculated from the regression equation gave a more accurate *Tm* prediction as indicated by the smaller residual values. Of course the regression equation was derived from these data and consequently it would be expected that residuals would be minimised. To test the validity of the revised *Tm* prediction equation, a second series of novel oligonucleotide duplexes referred to as validation (II) duplexes (see appendix 6.1.1.5) were constructed. These duplexes were tested to determine their actual melting temperature which was subsequently compared to two melting temperature prediction models; Primer Express™ and the revised MOASIC model.

Duplex ID	Length	Salt Corr.	Actual Tm	Primer Exp.	MOSAIC
11	19	8.33	59.68	53.90	60.76
12	19	8.33	65.92	64.10	64.56
13	19	8.33	62.05	57.80	61.71
14	19	8.33	66.77	63.90	66.02
15	15	7.51	57.54	49.50	57.54
16	15	7.51	60.49	56.20	60.74
17	15	7.51	60.37	56.20	60.74
18	25	9.02	69.9	74.00	70.30
19	25	9.02	71.36	76.70	72.03
20	25	9.02	70.49	75.80	71.87

Table 4.37 10 novel duplexes with their associated *Tms* defined by Actual measurement, Primer Express™ and MOSAIC.

Figure 4.21 shows a diagrammatic representation of this data.

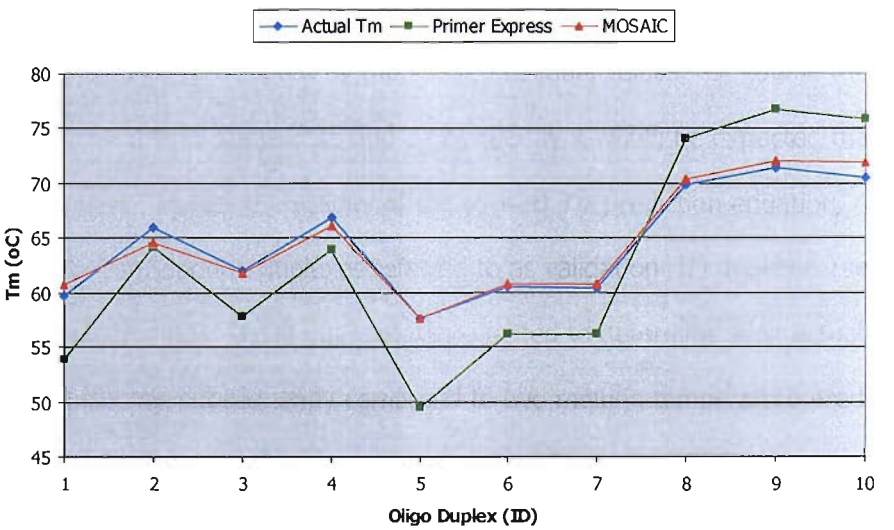


Figure 4.21 A plot of actual *Tm* together with plots of predicted *Tms* from Primer Express™ and MOSAIC for data from table 4.47.

MOSAIC prediction of *Tm* aligned itself with a high degree of fit to actual *Tm* for all duplexes tested. In contrast, *Tms* derived from Primer Express™ showed a poor correlation. To illustrate this difference, mean absolute deviations were calculated for both prediction models. These were as follows:

Prediction model	Mean Absolute Deviation
Primer Express™	23.70
MOSAIC	0.64

Table 4.38 show the Mean Absolute Deviation for the two duplex *Tm* prediction models.

The MAD results (table 4.38) demonstrated an extremely high closeness of fit from the MOSAIC model compared to Primer Express™ confirming the validity of the revised MOSAIC algorithm.

In conclusion, the revised algorithm for calculating duplex melting temperature for complementary sequences can be defined as:

$$\boldsymbol{Tm}\ ^{\circ}\boldsymbol{C} = ((\Delta H^{\circ}/\Delta S^{\circ}+(R\ln(C_T))) - (273.15 + \mathbf{X}\ \mathrm{Log}[Na^{+}])$$

Where $\mathbf{X} = [0.0035(\mathbf{I})^3-0.2157(\mathbf{I})^2+4.492(\mathbf{I})-23.154]$

$\mathbf{I}$ = Oligo length (bp)

4.2.5 Defining a trend in mismatch stability

From the data, trends in mismatch stability were derived from a sub-set of the data, exploring the effects of different Watson Crick base pairs on either side of the central mismatched base.

In this analysis, the positions of all three central bases were fixed in a particular context e.g. A,A,A; where A,A,A refer to the sequence of strand A. Duplexes which formed complementary Watson Crick base pairing at positions 1 and 3 were selected giving rise to 4 duplex types; resulting from variant bases at position 2; A,C,G or T. Each of these variant

bases formed either a complementary base pair or a mismatched base pair depending on the exact sequence of strand B, resulting in 16 possible duplex variants for any one context. An example of this is shown in table 4.39a. Each of the 16 variants were sorted according to ascending *Tm* and then assigned an arbitrary rank number ranging from 1 (lowest *Tm)*, to 16 (highest *Tm*) as shown in table 4.39b.

Table 4.39a

Base position			*Tm*
1	2	3	
AT	AA	AT	55.11
AT	AC	AT	54.73
AT	AG	AT	54.50
AT	AT	AT	62.39
AT	CA	AT	52.79
AT	CC	AT	51.67
AT	CG	AT	63.00
AT	CT	AT	54.04
AT	GA	AT	57.88
AT	GC	AT	64.20
AT	GG	AT	55.77
AT	GT	AT	56.95
AT	TA	AT	61.75
AT	TC	AT	55.68
AT	TG	AT	56.18
AT	TT	AT	56.08

Table 4.39b

Base position			*Tm*	Rank
1	2	3		
AT	GC	AT	64.20	16
AT	CG	AT	63.00	15
AT	AT	AT	62.39	14
AT	TA	AT	61.75	13
AT	GA	AT	57.88	12
AT	GT	AT	56.95	11
AT	TG	AT	56.18	10
AT	TT	AT	56.08	9
AT	GG	AT	55.77	8
AT	TC	AT	55.68	7
AT	AA	AT	55.11	6
AT	AC	AT	54.73	5
AT	AG	AT	54.50	4
AT	CT	AT	54.04	3
AT	CA	AT	52.79	2
AT	CC	AT	51.67	1

Table 4.39a *Tm* values for all possible combinations for contexts where bases at positions 1 and 3 were fixed as A and A respectively.

Table 4.39b Data sorted according to *Tm* together with assigned rank values.

NB. Yellow indicates a Watson Crick match. For each base position, the first letter refers to strand A, whilst the second letter refers to strand B.

The rank number for each base pair type, for this context was noted. This type of analysis was repeated for each of the other base 1 and 3 contexts, resulting in 16 separate analyses, each generating a set of rank values for each particular base pair type. These rank values were summed together for each context resulting in the following table.

	Base pair type															
Base Context	GC	CG	AT	TA	GA	GT	GG	AG	TG	AA	TT	AC	TC	CA	CT	CC
Rank Total	251	243	222	212	170	166	158	150	141	105	101	72	61	58	50	16

Table 4.40 Combined rank values for each base pair type for all possible base pair contexts.

This data suggested that the overall stability (from high to low) of single mismatches (ignoring their sequence context) was as follows:

$$G•A = G•T > G•G \geq A•G \geq T•G > A•A \geq T•T > A•C > T•C = C•A > C•T > C•C$$

This trend in stability fitted very well with previously published data where purine mismatches of the form G•A, G•G and A•G tended to be more stable than pyrimidine mismatches, T•T, T•C, C•T and C•C. Mismatches that involved both purine pyrimidine bases tended to fall within the middle of the stability trend.

The trend shown in the previous example was based solely on the use of an arbitrary ranking system to indicate the strength of a particular mismatch type. The more destabilising the mismatch type, the lower the assigned rank number. Although the data were analysed systematically, including mismatches within all contexts, the actual context of the mismatch was ignored. This is shown in table 4.41.

Position 1 bp	AT				CG				GC				TA			
Position 3 bp	AT	CG	GC	TA	AT	CG	GC	TA	AT	CG	GC	TA	AT	CG	GC	TA
Position 2 bp (mismatch or match)	GC	GC	GC	GC	GC	GC	GC	GC	CG	GC	CG	CG	GC	GC	CG	CG
	CG	CG	CG	AT	CG	CG	CG	CG	GC	CG	GC	GC	CG	CG	GC	GC
	AT	TA	AT	CG	AT	AT	AT	AT	AT	AT	AT	AT	AT	TA	TA	AT
	TA	AT	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	AT	AT	TA
	GA	GA	GA	GA	GT	AG	GT	GT	GA	GG	GA	GG	GT	AG	TG	AG
	GT	GG	TG	GT	GA	GT	GA	GG	GG	AG	GG	GA	GA	GG	GT	GT
	TG	TG	GT	GG	AG	GA	AG	GA	GT	GA	TG	AG	AG	GT	AG	GG
	TT	AG	GG	AA	TG	GG	TG	AG	TG	GT	AG	GT	TG	TG	GG	AA
	GG	GT	TT	TT	GG	TG	GG	TT	AG	TG	GT	TG	GG	GA	GA	GA
	TC	AA	AG	AG	TT	AA	TT	TG	AA	AC	CA	AA	AA	AA	TT	TG
	AA	TT	TC	TC	AA	AC	AA	AA	TT	TT	AA	CA	TC	AC	AA	TT
	AC	AC	AA	AC	AC	TT	AC	TC	CA	TC	TT	AC	TT	CA	CA	CT
	AG	CA	AC	TG	CT	CA	CT	CT	AC	AA	AC	TT	AC	TT	TC	AC
	CT	TC	CT	CT	TC	CT	CA	CA	TC	CT	CT	CT	CT	TC	CT	CA
	CA	CT	CA	CA	CA	TC	TC	AC	CT	CA	TC	TC	CA	CT	AC	TC
	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC

Table 4.41 Ranked mismatch type (ordered by decreasing *Tm*) for each context.

Table 4.41 suggests that the context of a single mismatch; that is the identity of the two base pairs either side of the mismatch affects the degree to which the mismatch affects duplex stability. For very unstable mismatches such as C•C, the context was less important, as this always resulted in the lowest recorded duplex melting temperature. For less unstable mismatches such as that highlighted in red (T•G), the context of the mismatch did affect the

degree to which the mismatch destabilised the duplex. Within the context AT—TA (column 5), a T•G mismatch (highlighted in red),appears low down in the list indicative of a fairly high destabilising context; in contrast, within the context TA--GC (column 16), the T•G mismatch is much higher up in the list indicating that the mismatch within that context is less destabilising. This meant that there was a possibility that a trend in the context of a single mismatch might exist.

4.2.6 Defining a trend in the context of single mismatches

Melting temperatures derived from duplexes containing single mismatch could be directly compared to two corresponding data values for duplexes not containing mismatches. These represented duplexes that could be viewed as corrections to the mismatched sequence.

If the context for the mismatch was AT--AT and the mismatch type was A•G, then the two correcting duplex contexts would be:

- AT-CG-AT correcting strand A from A to C – *Tm* °C = 63.00;
- AT-AT-AT correcting strand B from G to T – *Tm* °C = 62.39.

The change in *Tm* (δTm), between each comparison for such corrections would be:

Mismatch melting temperature AT,A•G,AT *Tm* °C = 54.50

AT-CG-AT	= 54.50 – 63.00	AT-AT-AT	= 54.50 – 62.39
$\delta Tm1$	= -8.40	$\delta Tm2$	= -7.89

The average δTm is therefore = - 8.15°C for an A•G mismatch within the context AT--AT.

Similar comparisons were made for all other contexts of an A•G mismatch, as shown in table 4.42a. Each of the 16 variants were sorted according to ascending δTm and then assigned

an arbitrary rank number ranging from 1 (smallest δTm), to 16 (highest δTm) as shown in table 4.42b.

Table 4.42a

Base Position			*dTm*
1	2	3	
AT	AG	AT	-8.20
AT	AG	CG	-4.89
AT	AG	GC	-7.27
AT	AG	TA	-7.02
CG	AG	AT	-5.81
CG	AG	CG	-3.53
CG	AG	GC	-5.19
CG	AG	TA	-4.79
GC	AG	AT	-5.78
GC	AG	CG	-2.83
GC	AG	GC	-5.35
GC	AG	TA	-3.93
TA	AG	AT	-5.75
TA	AG	CG	-3.16
TA	AG	GC	-5.43
TA	AG	TA	-4.99

Table 4.42b

Base Position			*dTm*	Rank
1	2	3		
GC	AG	CG	-2.83	16
TA	AG	CG	-3.16	15
CG	AG	CG	-3.53	14
GC	AG	TA	-3.93	13
CG	AG	TA	-4.79	12
AT	AG	CG	-4.89	11
TA	AG	TA	-4.99	10
CG	AG	GC	-5.19	9
GC	AG	GC	-5.35	8
TA	AG	GC	-5.43	7
TA	AG	AT	-5.75	6
GC	AG	AT	-5.78	5
CG	AG	AT	-5.81	4
AT	AG	TA	-7.02	3
AT	AG	GC	-7.27	2
AT	AG	AT	-8.20	1

Table 4.42a δTm values for all possible A•G mismatch contexts.

Table 4.42b Data sorted according to increasing δTm together with assigned rank values

The rank number for each context was noted. This type of analysis was repeated for each of the other mismatch types as well as for Watson and Crick base pairing at position 2, resulting in 16 separate analyses, each generating a set of rank values for each particular context. These rank values were summed together for each context resulting in the following table.

	Context															
Position 1	GC	GC	GC	CG	CG	TA	CG	GC	CG	TA	TA	TA	AT	AT	AT	AT
Position 2	mismatched basepair															
Position 3	GC	CG	AT	CG	GC	TA	TA	TA	AT	CG	AT	GC	GC	CG	TA	AT
Rank total	163	134	121	117	113	113	110	108	106	102	102	95	70	69	57	52

Table 4.43 Combined rank values for each single mismatch context.

A general trend exists for the context of a single mismatch as indicated from table 4.43. Adjacent base pairs that typically assume either CG or GC, serve to stabilise the mismatch whilst base pairs that assume either AT or TA do not stabilise the mismatches to the same extent. This was consistent with the fact that CG base pairs form three hydrogen bonds whilst AT base pairs form only two hydrogen bonds. Where adjacent base pairs were a

mixture of AT, TA with CG or GC, the stability of the mismatch was much less predictable, residing somewhere in between the two extreme cases. As with the prediction of trends in mismatch stability, the trends shown in this example were based solely on the use of an arbitrary ranking system to indicate the strength of a particular mismatch context. The more destabilising the mismatch context, the lower down the order it appeared and therefore the lower the assigned rank number. The actual position of each mismatch context, for each mismatch type is shown in table 4.44.

	Mismatch type											
	TG	GG	GT	GA	AG	TT	CT	TC	AA	CA	AC	CC
Mismatch context	GC--CG	GC--CG	CG--TA	GC--GC	GC--CG	CG--TA	TA--TA	TA--AT	GC--GC	GC--GC	GC--GC	GC--GC
	GC--GC	GC--TA	TA--GC	GC--AT	TA--CG	GC--GC	CG--TA	AT--AT	CG--CG	CG--CG	GC--CG	TA--TA
	AT--CG	GC--GC	GC--CG	GC--TA	CG--CG	CG--GC	CG--AT	GC--AT	TA--AT	GC--TA	TA--CG	CG--GC
	TA--GC	CG--TA	TA--TA	GC--CG	GC--TA	CG--AT	TA--AT	GC--GC	TA--TA	GC--AT	CG--CG	CG--AT
	TA--CG	AT--CG	CG--GC	AT--CG	CG--TA	AT--GC	TA--GC	AT--GC	GC--TA	TA--CG	GC--TA	TA--GC
	GC--AT	TA--CG	GC--TA	CG--GC	AT--CG	TA--GC	CG--GC	GC--CG	TA--CG	CG--AT	TA--TA	GC--AT
	CG--GC	GC--AT	GC--GC	AT--GC	TA--TA	TA--AT	CG--CG	AT--TA	GC--AT	TA--TA	GC--AT	TA--AT
	AT--GC	CG--CG	CG--CG	AT--TA	CG--GC	AT--AT	GC--CG	CG--AT	CG--TA	CG--TA	CG--AT	CG--CG
	CG--AT	AT--TA	CG--AT	CG--CG	GC--GC	GC--AT	GC--GC	TA--TA	CG--AT	CG--GC	TA--AT	GC--CG
	CG--CG	TA--TA	GC--AT	AT--AT	TA--GC	GC--CG	GC--AT	TA--GC	AT--TA	GC--CG	CG--GC	GC--TA
	TA--AT	TA--GC	TA--CG	TA--AT	TA--AT	CG--CG	TA--CG	CG--GC	CG--GC	TA--AT	AT--TA	TA--CG
	CG--TA	CG--GC	TA--AT	CG--AT	GC--AT	TA--TA	AT--AT	CG--TA	AT--CG	AT--CG	AT--GC	CG--TA
	AT--AT	CG--AT	AT--GC	CG--TA	CG--AT	AT--TA	AT--GC	TA--CG	TA--GC	TA--GC	AT--CG	AT--AT
	GC--TA	AT--GC	AT--TA	TA--TA	AT--TA	TA--CG	GC--TA	GC--TA	AT--GC	AT--GC	TA--GC	AT--GC
	TA--TA	TA--AT	AT--AT	TA--CG	AT--GC	AT--CG	AT--TA	CG--CG	GC--CG	AT--TA	AT--AT	AT--TA
	AT--TA	AT--AT	AT--CG	TA--GC	AT--AT	GC--TA	AT--CG	AT--CG	AT--AT	AT--AT	CG--TA	AT--CG

Table 4.44 show the ranked positions (ordered by increasing δTm) of each mismatch context, for each single mismatch type.

Table 4.44 shows the variation in position of particular contexts for each mismatch type, according to the ranking assignments. Clearly the general trend eluded too earlier is much less obvious from this table. Four different contexts are indicated shown in blue, green, peach and yellow. For the mismatches G•T and T•T, the most stable context was CG--TA, however this was the least stabilising context for the A•C mismatch.

Despite these findings, the general trend observed earlier has been demonstrated previously (Allawi *et al.,* 1997, 1998(a), 1998(b), 1998 (c), 1998(d); Peyret *et al.,* 1999). Interestingly, C•T and T•C mismatches were stabilised more so by AT sequence rather than GC. This was consistent with previously published data (Allawi *et al.,* 1998(a)).

4.2.7 Visual Basic 6.0 Computer program (MOSAIC v1.0)

The manual calculation of melting temperature using the revised MOSAIC *Tm* algorithm (page 137) proved to be time consuming and laborious particularly if using it within the context of primer design. For this reason, the computer program MOSAIC was written in Visual Basic 6.0. The underlying code for this program is shown in appendix 6.3. The program operates from a windows based environment as shown in figure 4.22/4.23. Each session can be loaded with up to 12 different sequences; each sequence displaying an associated delta*H* and delta*S* value derived from nearest neighbour thermodynamics (calculated from data in tables 1.2 and 1.3). In addition, sequence composition (%GC), and individual base counts are included. Finally, the melting temperature is calculated according to the revised MOSAIC *Tm* algorithm, in real time.

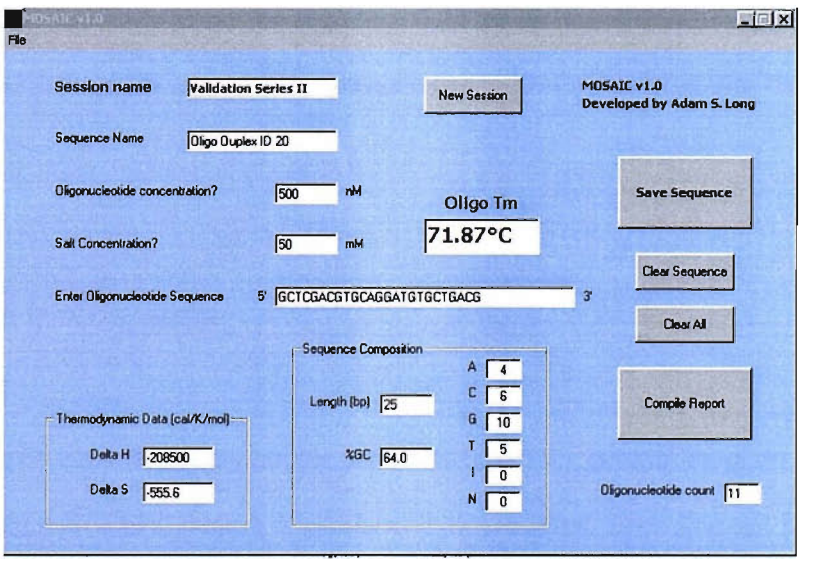


Figure 4.22 A view of the opening window of the MOSAIC v1.0 VB6.0 program. Oligo duplex ID 20 from Validation Series II oligos is shown.

The output from each session is a report containing a list of the sequences tested. This is shown in figure 4.23.

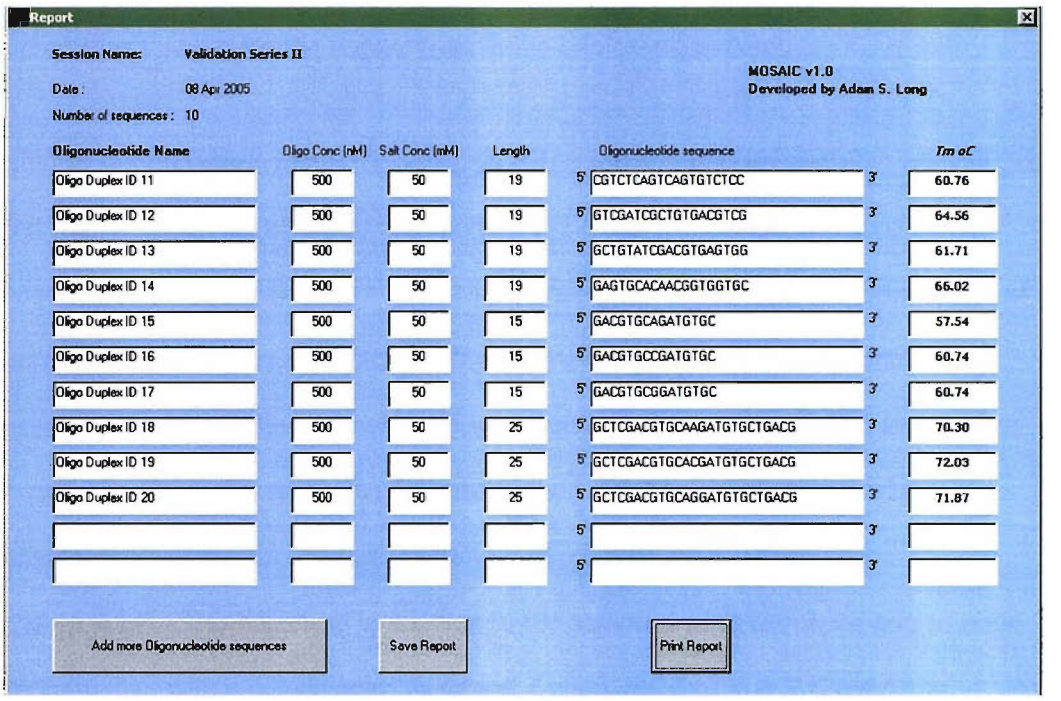


Figure 4.23 A view of the report window of the MOSAIC v1.0 VB6.0 program. A compilation of Oligo sequences corresponding to Validation Series II are shown.

The interface automatically saves the sequence data for archiving or retrieval at a later date. The MOSAIC v1.0 program was used to calculate the MOSAIC predicted *Tm* throughout the validation.

4.2.8 General discussion

Analysis of the 64 complementary duplex pairs resulted in the generation of an algorithm which described with a high degree of precision their predicted *Tms.* The salt correction value associated with this model was determined to be 7.962. Following the subsequent identification of a length dependency associated with the original salt correction value, the revised algorithm calculated a different salt correction value. This was found to be 8.333 (derived from figure 4.20). Changing the salt correction value from 7.962 to 8.333 reduced the overall predicted *Tm* by 0.4 °C. This meant that predicted *Tms* for each of the 64 duplex pairs was decreased by 0.4 °C. Although this shift in *Tm* was only small, the use of the

revised salt correction factor for calculating predicted *Tm,* appeared to make it less accurate compared to actual *Tm* for the duplex pairs in the original sample set.

The shift in salt correction factor for duplexes of the same length between the original data set and the validation set of oligos was thought to be due to different oligonucleotide manufacturers. All of the original duplex oligos were synthesised from the same batch of consumables, by the same manufacturer (Department of Chemistry, Southampton University). In contrast the oligonucleotides manufactured for the validation series were synthesised by IBA in Germany. In the case of the validation sequences, the DABCYL quencher residue was attached to the Strand B oligos through a cytosine base (see figure 4.24). This base was in addition to the standard oligo sequence and so upon duplex formation with a complementary strand A oligo, resulted in a 3′ C overhang.

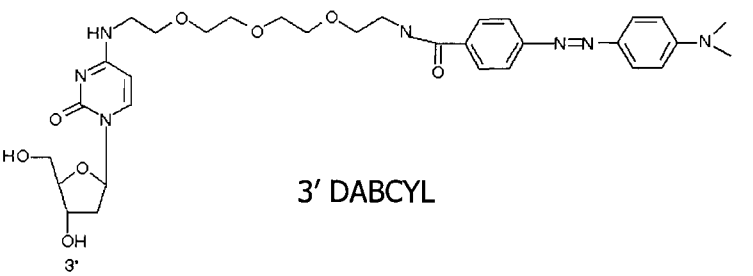


Figure 4.24 A schematic diagram of the DABCYL residue attached to a 3′ cytosine base. Manufactured by IBA, Germany.

The original oligo set did not have this additional cytosine base incorporated onto the '3 end of strand B, therefore, no 3′ base overhangs resulted. It was expected that base overhangs of this type would confer a small stabilisation effect on the duplex from observations by Doktycz *et al.,* (1990); Bommarito *et al.,* (2000), although the authors showed that C residues were the least stabilising compared to A, T or G bases. This being the case, the validation series of oligos should be slightly more stable than predictions would otherwise have stated and so a salt correction factor lower than 7.962 be implied. Actually, the salt

correction factor for the validation series was optimised at 8.333. Actual *Tms were* 0.4 °C below what was predicted by the original salt correction factor. This suggested that the additional cytosine residue did not serve to stabilise the duplex, but rather destabilised it, as indicated by the decreased melting temperature. Although an important observation, the change in salt correction factor from 7.962 to 8.333, only resulted in a 0.4 °C shift in *Tm.* This lay extremely close to the standard error attributed to experimental variation (± 0.36 °C), and therefore could be regarded as insignificant.

The oligo length/salt correction factor dependency model was validated for duplexes between 10 and 28 nucleotides long. Beyond 28 nucleotides, the salt correction factor model predicts a decreasing *Tm* trend with increasing oligonucleotide length. Clearly the effect of salt on longer duplexes (>28 bp) would have some degree of an effect, however this was not determined in this investigation.

Although melting temperature data were available for duplexes containing double and triple mismatches, they were not analysed in this investigation due to time constraints. It is envisaged that this data will eventually be used to ascertain trends in mismatch stability which will then be used to inform the MOSAIC v1.0 programme so that it accurately predicts melting temperature for mismatched duplexes as well as matched duplexes.

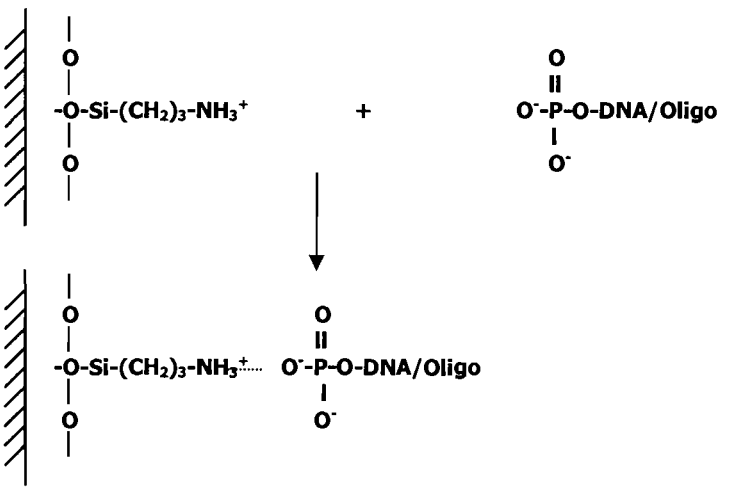
4.3 Microarray Assay Results

4.3.1 Oligonucleotides attachment

There are several different approaches to binding DNA onto microscope slides, which can be divided into two main groups; those which involve hydrogen/ionic bonding or those which involve covalent linkage.

4.3.1.1 Simple Ionic Bonding mediated through aminosilane coated slides

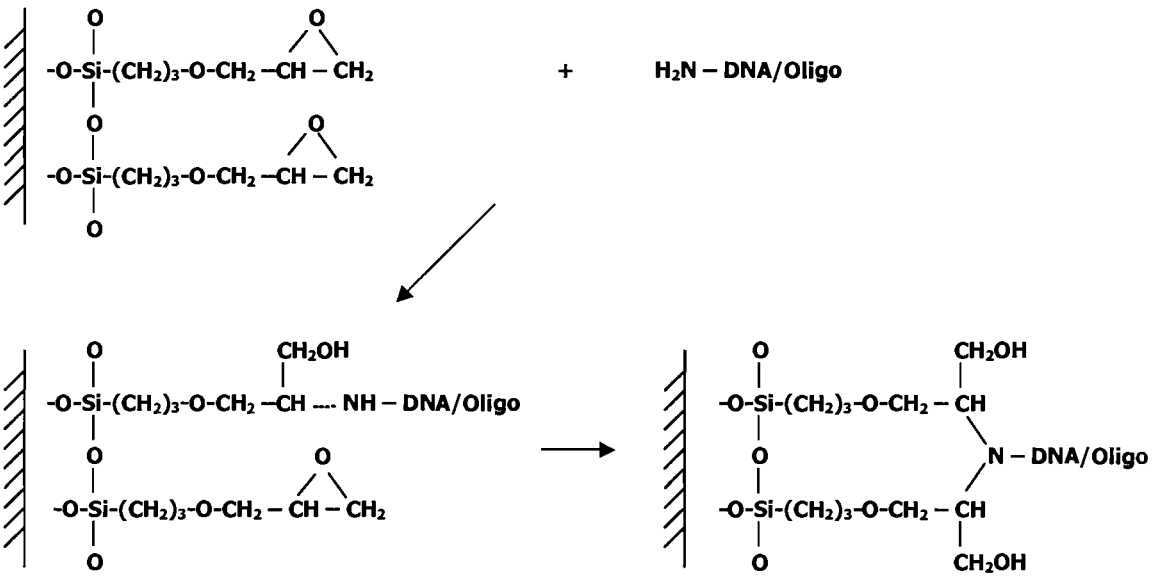
Slides coated in an aminosilane monolayer provide available amine groups which form initial ionic bonds with negatively charged phosphate groups in the DNA/Oligo backbone:



Subsequent exposure to UV radiation or elevated temperatures (c.80°C) for 1 hour causes the formation of a covalent bond tethering the oligonucleotide to the slide surface.

4.3.1.2 Covalent bond attachment mediated through epoxysilane coated slides

Type 7 and type 7* (star) slides were provided with an epoxysilane monolayer surface. These active groups interact directly with primary amine groups present on the DNA molecules being spotted;



To encourage covalent bond formation, slides were subjected to UV radiation.

For the majority of this investigation, epoxysilane coated slides (type 7/type 7*) were used, therefore it was anticipated that the later example of slide attachment would have occurred if the spotted template contained free amine groups.

4.3.2 Calculating an estimate of experimental error

In a similar approach to that adopted for silicon channel data, a value for the standard error of the mean (95% confidence) was calculated for the microarray data from a sample data set as follows:

	Sample experimental data (rfu) from section 4.3.5			
	Tsco D (C) Cy3 (rfu)	**Tsco D (C) Cy5 (rfu)**	**Tsco D (T) Cy3 (rfu)**	**Tsco D (T) Cy3 (rfu)**
Experimental replicates (rfu)	560194	709517	448652	443011
	609712	664973	401666	402718
	686542	679215	427792	413911
	682380	697226	424568	421751
	681787	688440	419178	385391
	688684	673795	394802	381383
	686553	679127	419223	390914
	700323	689495	392904	384281
	690852	688938	415215	383093
	678915	660125	415388	372822
	723550	678106	432478	365336
	703079	677206	401917	379530
	687286	666602	428780	372008
	709675	663360	417916	373361
	695549	655575	419011	370979
	741728	676984	423443	361122
	749197	680848	422066	365582
	701017	663469	417918	349620
	700345	665064	384011	328604
	777925	680342	389267	328446
Mean	**692765**	**676920**	**414810**	**378693**
Standard deviation	45526	13477	16124	27878
St. Error of Mean (95% confidence)	19954	5906	7067	12218
% St. Error mean	2.9	0.9	1.7	3.2
Average % Error	**2.2**			

Table 4.45 Replicate sample data from section 4.3.5 Tsco D Cy3 and Cy5 (indirect approach) for a heterozygote sample. In each case, the mean, standard deviation and standard error of the mean (95% confidence interval) was calculated. These values were used to calculate the % st. error mean for each data and finally an average % error for all experimental data.

Table 4.45 showed that the average standard error of the mean (95% confidence), expressed as a percentage of the mean rfu value was 2.2%. This indicated that for each of the average rfu values stated in the experimental data, average variation was within ±2.2% rfu of that value.

4.3.3 Primary Amine group oligonucleotide modification

Two Cy3 dye labelled artificial templates were synthesised (see appendix 6.1.2), that mimicked the Gc1s locus amplicon generated through normal PCR. One template contained an additional 5′ primary amine group whilst the second template did not contain any additional amine modification. Both templates were spotted onto type 7 slides as described in section 2.3.5 and then left to incubate for 30 minutes at 37 °C. Following incubation, slides were washed in 90 °C water before being scanned. In both cases, a comparable Cy3 signal was observed suggesting that both templates showed equivalent binding efficiencies irrespective of having an amine modification. The fact that high temperature washing in water (high stringency) failed to cleave either templates from the slide surface demonstrated that attachment was more likely mediated through covalent bonding. This was thought to be the result of interactions between amine groups already present within the adenine and cytosine molecules in the DNA chain. Additional amine modification was excluded as a consequence of these findings and all subsequent primers were not modified in this manner.

4.3.4 Optimum allele specific probe concentration

Initial investigations utilised a Cy3 dye labelled reverse Universal 3 primer to generate dye labelled PCR product. Following spotting, determination of the specific product type (allele) was achieved using Cy5 dye labelled specific probes which hybridised to the amplicon of interest. Typically 1-5 pmol Cy5 labelled probe were used per hybridisation reaction. Following hybridisation and stringency washing, detection of bound probe was achieved by scanning for the presence of Cy3 and Cy5.

A singleplex PCR was set up using URP primers specific for PGM +. Following spotting, slides were hybridised with probes specific for either PGM+ or PGM– at 5 nM, 10 nM and 20 nM. Average fluorescence data was as follows:

		Relative Fluorescent Units (rfu)			
Slide Number	**Probe**	**Consensus Cy3 signal**	**Specific Cy5 Signal (S)**	**Background Cy5 Noise (N)**	**S:N ratio**
Slide 1	PGM+ (5 nM)	191505	713983	151491	4.7
Slide 2	PGM- (5 nM)	202903	6915	109597	0.06
Slide 3	PGM+ (10 nM)	209624	994128	183498	5.4
Slide 4	PGM- (10 nM)	242936	11260	165483	0.06
Slide 5	PGM+ (20 nM)	184584	950851	1192252	0.7
Slide 6	PGM- (20 nM)	209403	11313	176223	0.06

Table 4.46 rfu values corresponding to average specific and background fluorescent signals for two probes at three different concentrations on a singleplex PGM+ sample.

Average Cy5 fluorescence for each of the PGM+ slides remained relatively constant for each of the concentrations tested. For slides probed for PGM-, extremely low Cy5 signals were detected, demonstrating a very low level of non-specific binding probably as a result of the overall efficiency of the stringency washing. One noticeable feature however was that as the probe concentrations were increased, there was a slight increase in Cy5 background fluorescence leading to a reduced signal to noise ratio for the higher concentration of Cy5 probe. Increased background fluorescence was undesirable since it would cause ambiguity to genotype calling. As a direct consequence of this result, probe concentrations for Cy5 labelled probes were fixed to 5 nM.

Cy3 signals were detected in all PCR reactions indicating that the negative control samples contained either Cy3 labelled universal 3, carried over from the purification or alternatively an artefact of the PCR reaction. The Cy3 signals in these negative controls were much less than those in the positive reactions and so their presence was thought to originate from primer dimer interactions involving universal 3.

4.3.5 Duplex detection (direct approach).

During the course of investigation, candidate loci have been tested by the indirect approach, classically targeting markers associated with the blood group proteins such as Gc and PGM.

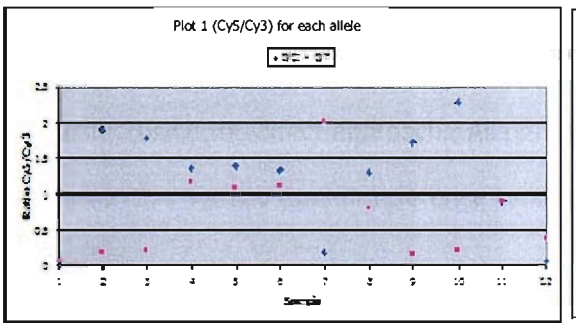
Alternative SNP loci have been located through publicly available databases and these have been tested using the direct approach. All primer and probe sequences are shown in appendix 6.1.2. All subsequent reference to SNP loci in this paper uses the FSS class designations.

Primers for tsco D and tsco F were amplified according to the direct approach using standard DNA extracts. Following spotting the slides were hybridised according to the parameters:

Slide	Allele specific probe	Probe concentration	locus specific probe	Probe concentration
1	tsco D/C	5 nM	tsco D/RLS	50 nM
2	tsco D/T	5 nM	tsco D/RLS	50 nM
3	tsco F/A	5 nM	tsco F/RLS	50 nM
4	tsco F/C	5 nM	tsco F/RLS	50 nM

Table 4.47 A list of probe mixtures used for detecting duplex PCR products.

Following stringency washing the slides were scanned for the presence of Cy3 and Cy5. Comparative data analysis was performed and the resulting figures appear below.



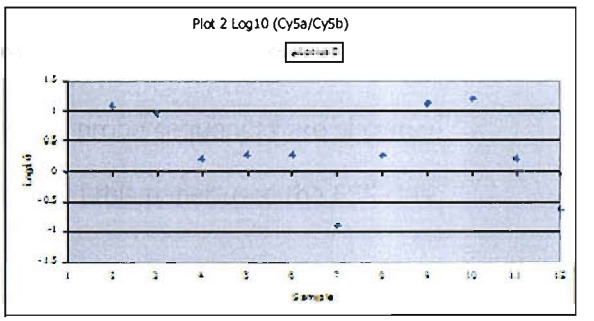


Figure 4.25. Simple comparative analysis for tsco D. **Figure 4.26. Log_{10} Data analysis for tsco D.**

Genotypes were assigned to all samples with respect to tsco D, figures 4.25 and 4.26 respectively.

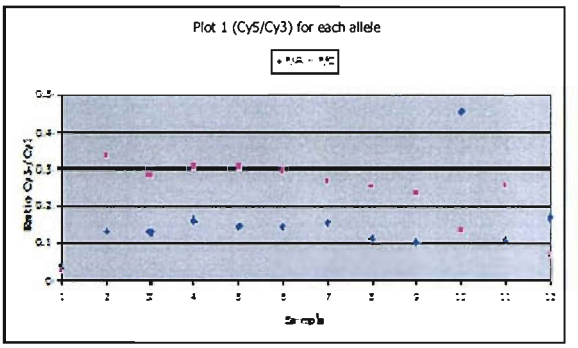


Figure 4.27. Simple comparative analysis for tsco F.

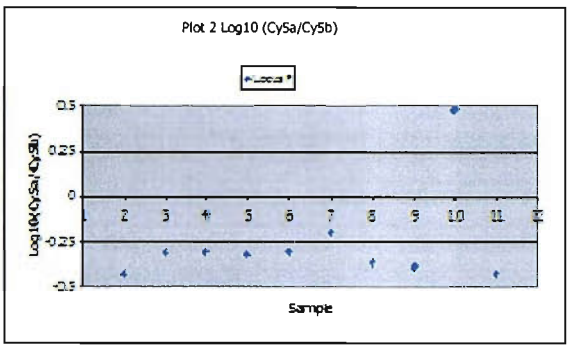


Figure 4.28. Log_{10} Data analysis for tsco F.

As with the previous chart, genotypes were assigned to all individuals although the mean fluorescent values were considerably lower for tsco F than for tsco D, demonstrating a slight decrease in locus amplification efficiency.

Accurate duplex detection using the direct approach suggested that this technique could be used successfully for SNP detection and further effort was given to increasing the multiplex size with respect to detection via the direct and indirect approaches.

4.3.6 Pentaplex A development (Direct vs. Indirect)

Increasing the multiplex size was necessary in order to maximise the discrimination power of the SNP assay. Multiplex analysis has never been demonstrated on this type of platform previously, therefore an attempt was made to detect products derived from a 5plex (pentaplex) in an assay that directly compared both detection methodologies.

Primers for pentaplex A (tsco D, F, X, L2 and Z2), were amplified according to both methodologies using standard DNA extracts. Following spotting the slides were hybridised with probe sets designed for both detection strategies for tsco D/C, D/T, F/A, F/C, X/A, X/C, L2/C, L2/T, Z2/C and Z2/T. After stringency washing the slides were scanned for the presence of Cy3 and Cy5. Comparative data analysis was performed and the resulting figures (appendix 6.4, figures 6.1/6.2 respectively), for the direct and indirect approaches used to determine sample genotypes.

4.3.6.1 Indirect approach Pentaplex A-v1

Sample ID	Genotype									
	tsco D		tsco F		tsco X		tsco L2		tsco Z2	
	Indirect	377	Indirect	377	Indirect	377	Indirect	377	Indirect	377
EA1-3	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/T	C/C	C/C
EA1-7	T/T	T/T	C/C	C/C	C/C	C/C	C/T	C/T	C/T	C/T
EA1-10	C/C	C/C	C/A	C/A	C/C	C/C	C/T	C/T	C/T	C/T
EA1-11	C/T	C/T	C/C	C/C	C/C	C/C	C/T	C/T	T/T	T/T
EA1-20	C/C	C/C	C/A	C/A	C/C	C/C	C/C	C/C	T/T	T/T
EA1-22	C/T	C/T	A/A	A/A	C/C	C/C	C/C	C/C	C/C	C/C
EA1-24	T/T	T/T	C/C	C/C	-	C/A	C/C	C/C	C/T	C/T
EA1-26	C/T	C/T	C/C	C/C	-	C/C	C/T	C/T	C/T	C/T
EA1-27	T/T	T/T	C/A	C/A	C/C	C/C	T/T	T/T	C/C	C/C
EA1-31	C/T	C/T	C/C	C/C	-	C/A	C/C	C/C	T/T	T/T
EA4-25	T/T	T/T	C/C	C/C	-	A/A	C/C	C/C	C/T	C/T

Table 4.48. Genotype results for 11 individuals amplified with Pentaplex A-v1, appendix 6.4 figures 6.1.

4.3.6.2 Direct approach Pentaplex A-v1

Sample ID	Genotype									
	tsco D		tsco F		tsco X		tsco L2		tsco Z2	
	Direct	377	Direct	377	Direct	377	Direct	377	Direct	377
EA1-3	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/T	C/C	C/C
EA1-7	T/T	T/T	C/C	C/C	C/C	C/C	C/T	C/T	C/T	C/T
EA1-10	C/C	C/C	C/A	C/A	C/C	C/C	C/T	C/T	C/T	C/T
EA1-11	C/T	C/T	C/C	C/C	C/C	C/C	C/T	C/T	T/T	T/T
EA1-20	C/C	C/C	C/A	C/A	C/C	C/C	C/C	C/C	T/T	T/T
EA1-22	C/T	C/T	A/A	A/A	C/C	C/C	C/C	C/C	C/C	C/C
EA1-24	T/T	T/T	C/C	C/C	C/A	C/A	C/C	C/C	C/T	C/T
EA1-26	C/T	C/T	C/C	C/C	C/C	C/C	C/T	C/T	-	C/T
EA1-27	T/T	T/T	C/A	C/A	C/C	C/C	T/T	T/T	C/C	C/C
EA1-31	C/T	C/T	C/C	C/C	C/A	C/A	C/C	C/C	T/T	T/T
EA4-25	T/T	T/T	C/C	C/C	A/A	A/A	C/C	C/C	C/T	C/T

Table 4.49. Genotype results for 11 individuals amplified with Pentaplex A-v1, appendix 6.4 figures 6.2.

Most of the microarray data in tables 4.48 and 4.49 were confirmed by comparison to reference 377 data. There were slight differences between the different detection approaches as seen by the subtle difference in genotype results. For the indirect approach, all samples amplified with the exception of four for tsco X. Genotypes for these individuals could not be ascertained due to the rather weak signals from allele A, in contrast to the stronger signals from allele C. For the direct approach these signals are much more balanced resulting in accurate genotypes been called for each sample.

Due to the poor nature and irreproducibility of tsco X, particularly for the indirect approach, this locus was substituted for another, tsco I3. The newly formed pentaplex A-v2 (D, F, L2, Z2 and I3) was tested as in the previous example for both methodologies. All resulting comparative data analyses figures are shown in appendix 6.4 figures 6.3 and 6.4.

4.3.6.3 Indirect approach Pentaplex A-v2

Sample	Genotype									
	tsco D		tsco F		tsco L2		tsco Z2		tsco I3	
	Indirect	377	Indirect	377	Indirect	377	Indirect	377	Indirect	377
EA1-3	C/C	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/G	C/G
EA1-8	C/T	C/T	C/C	C/C	C/C	C/C	C/T	C/T	C/C	C/C
EA1-12	T/T	T/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
EA1-13	C/C	C/C	C/C	C/C	C/T	C/T	T/T	T/T	C/C	C/C
EA1-15	C/T	C/T	C/A	C/A	C/T	C/T	C/T	C/T	C/G	C/G
EA1-20	C/C	C/C	C/A	C/A	C/C	C/C	T/T	T/T	C/G	C/G
EA1-22	C/T	C/T	A/A	A/A	C/C	C/C	C/C	C/C	C/G	C/G
EA1-24	T/T	T/T	C/C	C/C	C/C	C/C	C/T	C/T	G/G	G/G
EA1-25	C/C	C/C	C/A	C/A	C/C	C/C	T/T	T/T	C/G	C/G
EA1-27	T/T	T/T	C/A	C/A	T/T	T/T	C/C	C/C	G/G	G/G
EA1-29	C/T	C/T	C/C	C/C	C/C	C/C	C/C	C/C	G/G	G/G

Table 4.50 Genotype results for 11 individuals amplified with Pentaplex A-v2, appendix 6.4, figures 6.3.

4.3.6.4 Direct approach Pentaplex A-v2

Sample	Genotype									
	tsco D		tsco F		tsco L2		tsco Z2		tsco I3	
	Direct	377	Direct	377	Direct	377	Direct	377	Direct	377
EA1-3	C/C	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/G	C/G
EA1-8	C/T	C/T	C/C	C/C	C/C	C/C	C/T	C/T	C/C	C/C
EA1-12	T/T	T/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
EA1-13	C/C	C/C	C/C	C/C	C/T	C/T	T/T	T/T	C/C	C/C
EA1-15	C/T	C/T	C/A	C/A	C/T	C/T	C/T	C/T	C/G	C/G
EA1-20	C/C	C/C	C/A	C/A	C/C	C/C	T/T	T/T	C/G	C/G
EA1-22	C/T	C/T	A/A	A/A	C/C	C/C	C/C	C/C	C/G	C/G
EA1-24	T/T	T/T	C/C	C/C	C/C	C/C	C/T	C/T	G/G	G/G
EA1-25	C/C	C/C	C/A	C/A	C/C	C/C	T/T	T/T	C/G	C/G
EA1-27	T/T	T/T	C/A	C/A	T/T	T/T	C/C	C/C	G/G	G/G
EA1-29	C/T	C/T	C/C	C/C	C/C	C/C	C/C	C/C	G/G	G/G

Table 4.51 Genotype results for 11 individuals amplified with Pentaplex A-v2 and detected using the direct approach. Corresponding genotype results are shown from 377 gel electrophoresis analysis. See appendix 6.4, figures 6.4.

Pentaplex A-v2 was amplified efficiently by both methodologies and through detection, the results showed complete concordance between each other. Similarly, concordance was reached when the results were compared to 377 data.

These results demonstrated that both methodologies could be used to effectively multiplex loci together and ultimately provide a valid method for genotyping unknown DNA samples.

Due to the nature of the indirect approach and its successful development within the FSS (patent pending), this methodology was considered of greater interest. As a result, continuation to develop larger multiplexes, by the direct approach, was seen as a duplication of effort and so further microarray development only considered the indirect approach utilising the URP principle with ARMS.

4.3.6.5 Pentaplex A-v3 and detection within a larger multiplex

Detection of each locus within a pentaplex via the indirect approach had been demonstrated however there was some variability associated with the previously substituted locus tsco I3. For that reason multiplex A-v2 was modified again by substituting tsco I3 for tsco G, producing a third version of the original multiplex, known as pentaplex A-v3. Following testing, the modified pentaplex showed similar results to pentaplex A-v2 but for the substituted locus which showed much more consistency with respect to accurate genotyping of samples.

To understand the limitations of genotyping multiplexed PCR samples on a microarray platform, it was decided that a larger multiplex PCR should be performed, and then detected in part by probing for the five candidate loci within pentaplex A-v3. As such a 17-plex which had been previously developed in house was used. Five of its constituent loci were the same as those used in pentaplex A-v3. Following amplification, the 17-plex samples were spotted and hybridised using probes specific to the 5 loci in question. A similar amplification was conducted using only the loci associated with pentaplex A-v3 as a back to back comparison. Following stringency washing the slides were scanned for the presence of Cy3 and Cy5. Comparative data analysis was performed (resulting figures appear in appendix 6.4, figures 6.5 and 6.6). From these figures the following genotypes were produced.

Sample	Genotype of 5 loci form pentaplex A-v3 and a 17-plex									
	tsco D		tsco F		tsco G		tsco L2		tsco Z2	
	5-plex	17-plex	5-plex	17-plex	5-plex	17-plex	5-plex	17-plex	5-plex	17-plex
EA1-5	C/T	-	C/C	-	C/C	-	C/C	-	T/T	-
EA1-7	T/T	-	C/C	-	T/T	-	C/T	-	C/T	-
EA1-10	C/C	-	C/A	-	T/T	-	C/T	-	C/T	-
EA1-11	C/T	-	C/C	C/C	T/T	-	C/T	-	T/T	-
EA1-14	C/T	-	C/C	C/C	C/T	-	C/C	-	C/T	-
EA1-16	C/T	-	C/A	C/A	T/T	-	C/C	-	C/T	-
EA1-17	C/T	-	C/C	-	T/T	-	C/C	-	C/T	-
EA1-22	C/T	-	A/A	A/A	T/T	-	C/C	-	C/C	-
EA1-23	T/T	-	C/A	C/A	C/T	-	C/C	-	C/C	-
EA1-30	T/T	-	C/C	C/C	T/T	-	C/C	-	C/C	-
EA1-31	C/T	-	C/C	C/C	T/T	-	C/C	-	T/T	-

Table 4.52 Combined genotype results for 11 individuals amplified with either pentaplex A-v3, appendix 6.4, figures 6.5, or from the 17-plex, appendix 6.4, figures 6.6.

All samples amplified with pentaplex A-v3 were successfully genotyped and showed concordance with previous results. The same five loci within the 17-plex samples however, could not be accurately interpreted. The exception to this was tsco F, where six from 11 samples were scored correctly. These results suggested that the size of the multiplex may govern the sensitivity of the technique with respect to accurate genotype calling.

As the multiplex size increased, the relative amounts of each component amplicon would reduce in a proportional fashion as a consequence of competition between the different sites during amplification. During spotting, the same volume of PCR product was deposited, regardless of multiplex size and so the relative concentration of amplicon associated with each individual locus would be reduced also. This made detection difficult as the sensitivity was compromised as a direct result of reduced specific signal compared to an increase in background noise (non-specific probe hybridisation to other amplicon targets). Overall this result demonstrated that there would be inevitable limitations to the technique.

4.3.7 Pentaplex B development

All of the loci used for microarray detection had previously been shown to work well within a larger multiplex. For microarray applications, these loci could be simply included into any developing multiplex as and when required, however, the previous result suggested that larger multiplex sizes may prove more difficult to detect. Despite this, detecting pentaplexes was seen as being relatively straightforward and so rather than continually add loci to pentaplex A-v3, a different strategy was adopted. This involved developing additional pentaplexes as separate, unique multiplexes with a view to subsequently combining these together once finalised to generate a larger multiplex.

4.3.7.1 Pentaplex B-v1

Primers for pentaplex B-v1 (tsco J2, I3, K3, O3 and W3) were amplified according to the indirect approach using standard DNA extracts. Following spotting the slides were hybridised with allele specific/locus specific probes sets. After stringency washing the slides were scanned for the presence of Cy3 and Cy5. Comparative data analysis was performed and the resulting figures appear in appendix 6.4, figures 6.7. From these figures the following genotypes were produced.

Sample	Genotype				
	tsco J2	**tsco K3**	**tsco I3**	**tsco O3**	**tsco W3**
EA1-5	C/T	C/C	C/G	C/G	C/C
EA1-7	C/C	C/G	C/G	C/C	C/G
EA1-15	C/C	C/G	C/G	C/G	C/G
EA1-17	C/C	C/G	C/G	C/G	C/C
EA1-18	C/C	C/G	C/G	C/G	C/G
EA1-19	C/C	C/G	C/G	C/G	C/C
EA1-22	C/C	C/G	C/G	C/G	C/C
EA1-23	C/C	C/G	G/G	G/G	C/C
EA1-24	C/T	C/G	G/G	G/G	C/G
EA1-29	C/C	C/G	G/G	G/G	C/C
EA1-30	C/C	C/G	C/C	G/G	C/C

Table 4.53 Genotype results for 11 individuals amplified with pentaplex B-v1, appendix 6.4, figures 6.7.

All of the loci within pentaplex B-v1 were successfully genotyped for each of the samples tested. These were consistent with previously determined 377 data and demonstrated a second working pentaplex.

Having successfully demonstrated two different pentaplexes; pentaplex A-v3 and pentaplex B-v1 the aim was to then combine the two pentaplexes together to form a decaplex comprising loci for tsco D, F, G, L2, Z2, J2, I3, K3, O3 and W3. Before this was done, it was anticipated that different amplicon lengths might alter the relative efficiency of primer amplification between loci. It was suggested that primer sets that give rise to longer amplicons would not amplify with the same efficiency as primers that give rise to shorter amplicons. Between the two Pentaplexes A-v3 and Bv1, amplicons within the range 123bp-183bp were produced as a result of using the exact same primer sequences used for gel based detection. This difference in size was necessary for gel based detection so that amplicons could be effectively separated by length. As a pre-requisite for microarray detection, it was decided that differences in amplicon length between loci were less important

since their individual detection was not length dependent, therefore it was suggested that it might be possible to redesign the reverse locus specific primers so that all amplicons within the same multiplex yield amplicons of a similar size. As such, reverse locus specific sequences were redesigned to yield amplicon lengths as follows:

Pentaplex	Locus ID	Old amplicon length	New amplicon length
A version 3	tsco D	157 bp	118 bp
	tsco F	168 bp	122 bp
	tsco G	173 bp	119 bp
	tsco L2	178 bp	121 bp
	tsco Z2	136 bp	121 bp
B version 1	tsco J2	145 bp	-
	tsco I3	150 bp	124 bp
	tsco K3	140 bp	119 bp
	tsco O3	123 bp	124 bp
	tsco W3	183 bp	-

Table 4.54 A comparison of the old and new amplicon sizes for each locus associated with each pentaplex.

The efficiency of the newly synthesised reverse primers was determined by carrying out a series of singleplex PCR amplifications. All samples were run on a minigel to determine primer function.

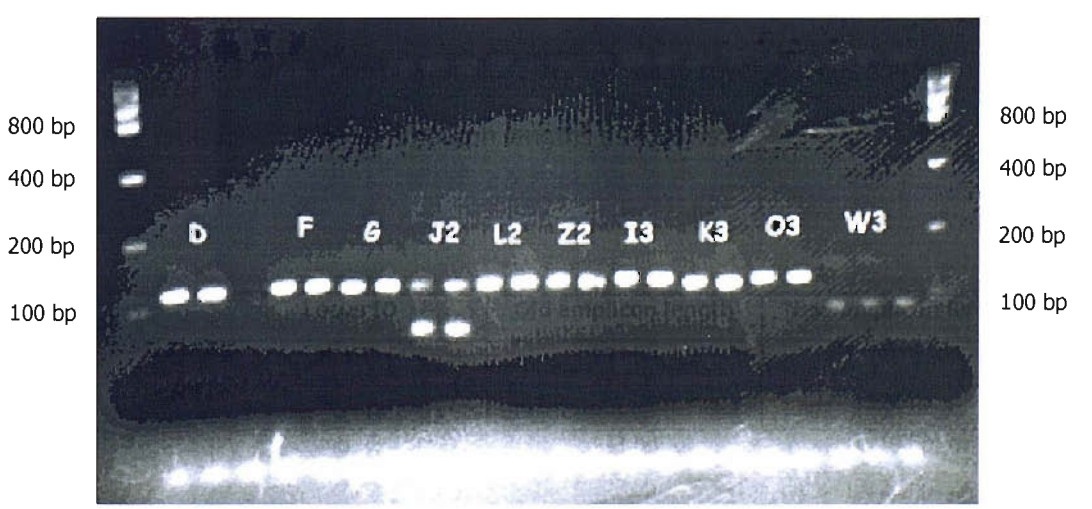


Figure 4.29 A minigel image showing singleplex PCR products from each of the above loci. A low mass DNA ladder is shown (left and right) together with sizing markers to determine fragment sizes.

The minigel image (figure 4.29), showed that all loci had amplified efficiently with the exception of locus J2 and W3 where additional bands (>100 bp) were present. As the primer redesigns for these two loci yielded spurious products following PCR amplification, it was decided that for future amplifications the original reverse primer sequence sets for these two loci would be used instead. In all other instances the new reverse primer sets were used.

The redesigned Pentaplex A-v3 (shorter amplicons) was re-run and analysed using the same methodology as described earlier. Comparative data analysis was performed and the resulting figures appear in appendix 6.4, figures 6.8. From these figures the following genotypes were produced.

Sample	Genotype				
	tsco D	tsco F	tsco G	tsco L2	tsco Z2
EA1-4	C/T	C/C	C/T	C/T	C/C
EA1-5	C/T	C/C	C/C	C/C	T/T
EA1-6	C/T	C/C	T/T	C/C	T/T
EA1-17	C/T	C/C	T/T	C/C	C/T
EA1-18	C/T	C/C	C/T	C/C	T/T
EA1-19	C/T	A/C	T/T	C/C	C/T
EA1-22	C/T	A/A	T/T	C/C	C/C
EA1-23	T/T	A/C	C/T	C/C	C/C
EA1-24	T/T	C/C	C/T	C/C	C/T
EA1-30	T/T	C/C	T/T	C/C	C/C
EA1-31	C/T	C/C	T/T	C/C	T/T

Table 4.55 Genotype results for 11 individuals amplified with modified pentaplex A-v3, appendix 6.4, figures 6.8.

All samples were correctly genotyped using the revised Pentaplex. Emphasis now was given to combining the redesigned pentaplex with pentaplex B-v1, to produce a working decaplex.

4.3.8 Decaplex detection

From initial investigations concerning the decaplex, results showed that loci from pentaplex A-v3 consistently outperformed those from pentaplex B-v1 in terms of accurate genotyping and reduced cross hybridisation. Singleplex results for each locus demonstrated a high degree of specificity across all loci, however when these were amplified together in the same multiplex, some cross hybridisation was seen following hybridisation. Singleplex PCR amplifications were run for each locus and spotted onto the same slide. Hybridisation and washing, followed by scanning for the presence of Cy3 and Cy5 demonstrated that an accurate genotype could still be assigned in each instance. These result demonstrated that the cross hybridisation seen in the multiplex result was probably the consequence of some form of primer-dimer interaction during amplification. To determine if primer-dimer was a cause of the slight cross hybridisation, a series of 10 multiplex PCR amplifications were set up each containing nine out of the ten original loci, (in each case a different locus was omitted). Following amplification all PCR products were spotted onto slides prior to hybridisation then, following hybridisation and stringency washing, the slides were scanned for the presence of Cy3 and Cy5 before comparative data analysis was conducted.

From these results it was possible to see that multiplex 9 (omission of tsco F), was the most consistent multiplex. A figure showing the genotype for each locus within the nine-plex is shown in appendix 6.4, figure 6.9. From this figure, the following genotype table was produced.

Sample	Genotype								
	tsco D	tsco G	tsco L2	tsco Z2	tsco J2	tsco I3	tsco K3	tsco O3	tsco W3
EA1-20	C/C	G/G	C/C	T/T	C/T	C/G	G/G	C/G	C/G

Table 4.56 Genotypes for sample EA1-20 amplified with the nine-plex (tsco F omitted). Appendix 6.4, figure 6.9.

Accurate genotypes could be assigned for each of the 9 loci demonstrating a working 9-plex or nonaplex amplified and detected according to the indirect approach.

The result of the nonaplex meant that the classification of Pentaplex A had to be changed from A-v3 to A-v4. This modified pentaplex now contained loci for tsco D, G, L2, Z2 and I3. The remaining 4 loci were taken forward and combined with an additional locus, tsco N4 to replace tsco F, to form Pentaplex B-v2. To maximise the discrimination power of the SNP multiplex systems, by ruling out possible linkage association, tsco O3 was substituted for tsco P5 as tsco O3 resided on the same chromosome as tsco J2. Pentaplex B-v2 therefore contained primers for tsco J2, K3, W3, N4 and P5.

Pentaplex B-v2 was amplified according to the indirect approach using standard DNA extracts. Following spotting, probe hybridisation and stringency washing the slides were scanned for the presence of Cy3 and Cy5. Comparative data analysis was performed and the resulting figures appear in appendix 6.4, figures 6.12. From these figures the following genotypes were produced.

Sample	Genotype				
	tsco J2	tsco K3	tsco W3	tsco N4	tsco P5
EA1-1	C/T ?	C/G	G/G	T/T	A/T
EA1-2	C/T ?	C/C	C/C	A/T	A/A
EA1-3	C/T ?	C/G	C/C	T/T	T/T
EA1-4	C/T ?	C/G	C/C	A/T	T/T
EA1-5	C/T ?	C/G	C/C	A/T	A/T
EA1-6	C/T ?	C/C	C/C	A/T	A/T
EA1-7	C/T ?	C/C	G/G	A/A	A/A
EA1-8	C/T ?	C/C	C/G	A/T	A/T
EA1-9	C/T ?	C/G	C/C	A/T	A/A
EA1-10	C/T ?	C/G	C/G	A/T	A/A
EA1-11	C/T ?	C/G	C/C	A/T	A/T

Table 4.57 Genotype results for 11 individuals amplified with pentaplex B-v2, appendix 6.4, figures 6.12.

For loci K3, W3, N4 and P5, genotypes could be accurately assigned to all samples but there were inconsistencies with locus J2. From the Log_{10} results, all samples appeared to be heterozygotic for locus J2, however looking at previous results it was shown that some samples were in fact homozygotic with respect to allele C. As such, singleplex PCR amplifications were run for each of the component loci within pentaplex B-v2. Samples of each product were spotted on to the same slide before being hybridised with an allele/locus specific probe set specific for one locus. This was repeated in parallel for all loci. Following hybridisation and stringency washing, all slides were scanned for the presence of Cy3 and Cy5. Comparative data analysis was performed and figures corresponding to locus J2 appear in appendix 6.4, figures 6.10.

Previously determined J2 Genotype	Sample ID	Singleplex Products	Allele C		Allele T	
			Cy3	Cy5	Cy3	Cy5
C/C	JG-1	J2	1931126	392525	1774564	116958
C/C	JG-2	J2	1417593	393246	1345956	176602
C/C	JG-3	J2	1702455	311434	1585556	216483
-	JG-4	K3	35471	4195	2679	**142610**
-	JG-5	K3	9567	1880	14520	**105705**

Table 4.58 show the relative Cy3 and Cy5 fluorescence values corresponding to alleles C and T respectively, appendix 6.4, figures 6.10.

From these results is was possible to determine that Cy5 signals specific to allele T, for locus J2, were also binding to PCR product amplified with primers specific to locus K3. This was evident by the large Cy5 fluorescence values for sample JG-4 and JG-5. This cross hybridisation event was seen at both K3 allele positions suggesting that it was this that was giving the appearance of a heterozygotic sample with respect to locus J2. For this reason Pentaplex B-v2 was modified by substituting tsco J2 for a new locus, tsco L6. This new pentaplex was called Pentaplex B-v3.

4.3.8.1 Pentaplex B-v3

Primers for pentaplex B-v3 (tsco K3, W3, N4, P5 and L6) were amplified according to the indirect approach using control DNA supplied by Cambio™. Following spotting the slides were hybridised with allele specific/locus specific probes sets. After stringency washing the slides were scanned for the presence of Cy3 and Cy5. Comparative data analysis was performed and the resulting figures are shown in appendix 6.4, figures 6.11. From these figures the following genotypes were produced.

Sample	**Genotype**				
	tsco K3	**tsco W3**	**tsco N4**	**tsco P5**	**tsco L6**
Cambio™ DNA	C/G	C/C	A/A	A/T	A/T

Table 4.59 Genotype results for control DNA amplified with pentaplex B-v3, appendix 6.4, figures 6.11.

For Cambio™ control DNA, genotypes were accurately assigned following pentaplex B-v3 amplification and detection via the indirect approach. Although pentaplex B-v3 was tested on only a small sample set, correct genotypes for all loci were scored.

During the course of this investigation it was anticipated that a series of pentaplexes be developed as separate entities before being combined together to form a larger common multiplex. As such a third pentaplex was developed.

4.3.9 Pentaplex C-v1

Pentaplex C-v1 comprised primers for tsco A4, K4, Z5, B6 and O6. These were amplified according to the indirect approach using standard DNA extracts. Following spotting the slides were hybridised with allele specific/locus specific probes sets. After stringency washing the slides were scanned for the presence of Cy3 and Cy5. Comparative data analysis was

performed; resulting figures appear in appendix 6.4, figures 6.13. From these figures the following genotypes were produced.

Sample	Genotype									
	tsco A4		tsco K4		tsco Z5		tsco B6		tsco O6	
	5-plex	377	5-plex	377	5-plex	377	5-plex	377	5-plex	377
EA1-1	C/C	C/C	G/G	G/G	A/T	A/T	A/T	A/T	A/T	A/T
EA1-2	C/C	C/C	G/G	G/G	A/A	A/A	A/A	A/A	A/A	A/A
EA1-3	C/C	C/C	G/G	G/G	A/A	A/A	A/A	A/A	A/A	A/A
EA1-4	C/C	C/C	C/G	C/G	A/T	A/T	A/A	A/A	A/A	A/A
EA1-5	C/C	C/C	G/G	G/G	A/T	A/T	A/T	A/A	A/T	A/T
EA1-6	C/C	C/C	G/G	G/G	A/T	A/T	T/T	T/T	A/T	A/T
EA1-7	C/C	C/C	G/G	G/G	A/T	A/T	A/T	A/T	A/T	A/T
EA1-8	C/C	C/C	G/G	G/G	A/A	A/T	A/T	A/T	A/A	A/A
EA1-9	C/C	C/C	G/G	G/G	A/A	A/A	A/T	A/A	A/T	A/T
EA1-10	C/G	C/G	G/G	G/G	T/T	T/T	A/T	A/A	A/T	T/T
EA1-11	C/G	C/G	C/G	C/G	A/T A/A?	T/T	A/T	A/T	A/A	A/A

Table 4.60 Genotype results for 11 individuals amplified with pentaplex C-v1, appendix 6.4, figures 6.13.

Table 4.60 showed some inconsistencies with respect to accurate genotyping for the pentaplex result compared to 377 sequencer results. Instances where the individual appeared to be a homozygote such as for EA1-8, tsco Z5 were in fact known to be heterozygotic. This apparent error can be explained by the concept known as allele dropout. This occurs for a variety of reasons and results in one of the alleles failing to amplify during PCR. In such instances simply increasing the respective allele primer concentration may serve to prevent allele drop out and thus improve amplification efficiency and ultimately heterozygote balance.

For EA-11, tsco Z5, the pentaplex result is ambiguous, suggesting the sample to be either a type A homozygote or an A/T heterozygote. This was actually confirmed as being a type T homozygote. Clearly, the pentaplex result was inaccurate and was most likely caused by the

presence of cross hybridisation of the A allele probe. If particularly strong, this cross hybridisation could mask the presence of the alternative probe.

4.3.10 Cy3/Cy5 signal balancing

Over the course of the investigation, additional research concerning the developing multiplex for conventional gel based detection demonstrated the need for some degree of primer balancing prior to PCR amplification. This was considered inevitable as gel based detection relied upon discerning different amplicons apart by length. Different sized amplicons means that amplification efficiency between loci can be somewhat variable however this can be easily remedied by applying balancing where the less efficient loci have their primer sets added a higher concentration.

Taking the example from table 4.60, sample EA-2, Cy5 signals were plotted for each locus following pentaplex amplification as shown in figure 4.30.

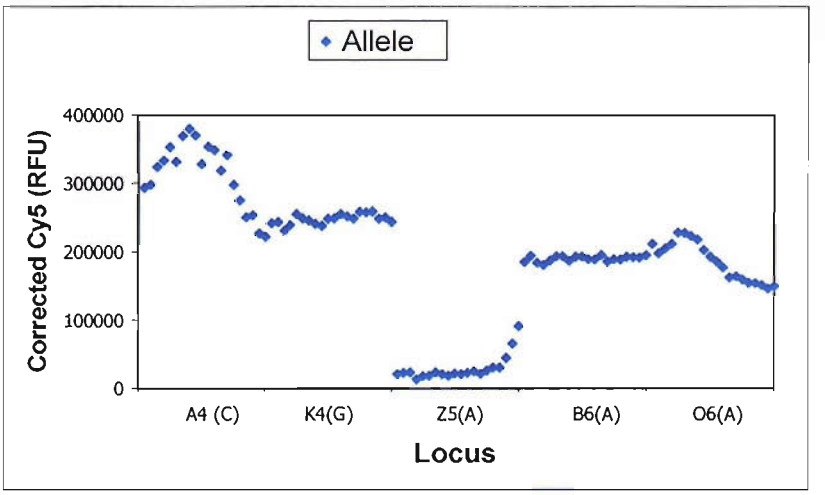


Figure 4.30 shows the relative Cy5 fluorescence values for each locus within Pentaplex C-v1, for sample EA-2.

Figure 4.30 showed that there was clearly a difference in Cy5 intensity between each locus particularly with locus Z5 which showed the least fluorescence. This result implied that the

reduced fluorescence be attributed to either a failure to amplify due to lack of starting template, or alternatively, a failure to amplify with the same efficiency as the other loci within the pentaplex. An improved fluorescent signal brought about by the need for some primer balancing would improve this result ruling out the possibility of locus failure and should be considered as a matter or routine practise for future multiplex development.

4.3.11 Slide type

During the course of the investigation, the manufacturers of the type 7 slides made substantial changes to the slide surface chemistry in an effort to enhance the amount of Cy3 fluorescence seen following hybridisation. These new slides were termed type 7*. Routinely for the type 7 slides, Cy3 probes were added at a concentration equal to 10 fold more than that of the Cy5 probe. Preliminary studies using dye labelled probes showed that an increased Cy3 signal was evident with the new type 7* slides but that the corresponding Cy5 signal was reduced. This was not surprising given that the Cy5 probe was added as a 10 fold weaker concentration than the Cy3 probe. For conventional hybridisation using type 7 slides, the Cy5 label was always designed to be present on the allele specific probe. Given that the new type 7* slides enhance the Cy3 signal, the sensitivity of the technique may be improved by switching the Cy3 label to the allele specific probe. This would increase the amount of allele-specific hybridisation seen.

Primers specific to tsco D were amplified in singleplex before being spotted onto Type 7 and Type 7* slides. Allele specific probes labelled with Cy5 together with locus specific probes labelled with Cy3 were hybridised to half the type 7 and half the type 7* slides. Similarly, allele specific probes labelled with Cy3 together with locus specific probes labelled with Cy5 were hybridised to the remaining type 7 and type 7* slides. Following hybridisation and stringency washing all slides were scanned for the presence of Cy3 and Cy5. Comparative data analysis was performed producing figures 4.31 – 4.34.

Figure 4.31a Cy5/Cy3 Type 7

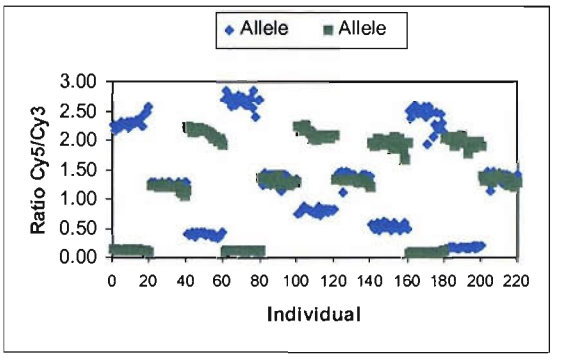


Figure 4.31b Log_{10}(Cy5a/Cy5b) Type 7

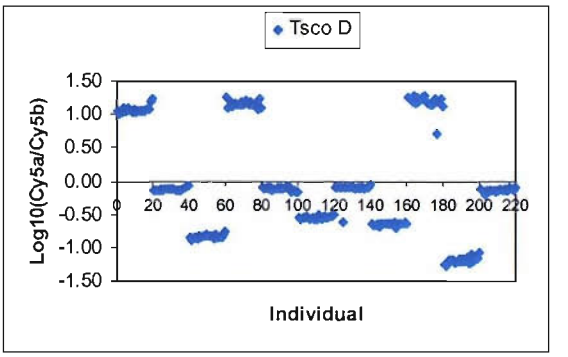


Figure4.31 Analysis graphs following Cy5 allele specific probe hybridisation (type 7 slides).

Figure 4.32a Cy5/Cy3 Type 7*

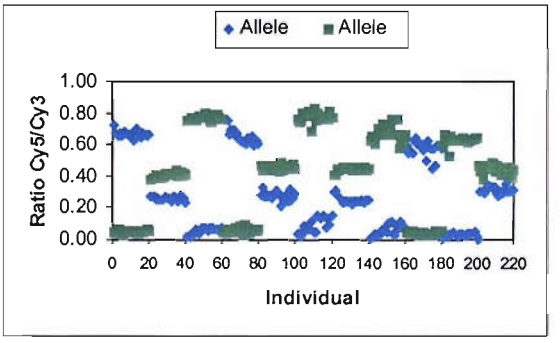


Figure 4.32b Log_{10}(Cy5a/Cy5b) Type 7*

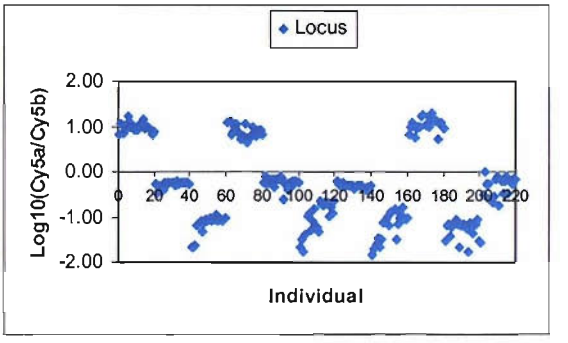


Figure 4.32 Analysis graphs following Cy5 allele specific probe hybridisation (type 7* slides).

Samples from both slide types could be easily scored. There was however a larger degree of variability seen in the Log_{10} data for the Type 7* slides given by the fact that the data were more spread out between the three genotype classes. An increased spread of data in this way may result in an incorrect genotype classification from this particular slide type.

Figure 4.33a Cy3/Cy5 Type 7

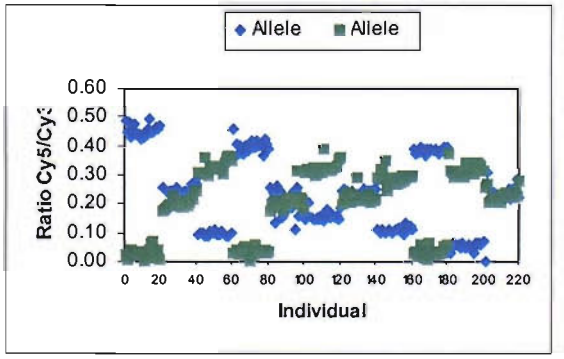


Figure 4.33b Log_{10}(Cy3a/Cy3b) Type 7

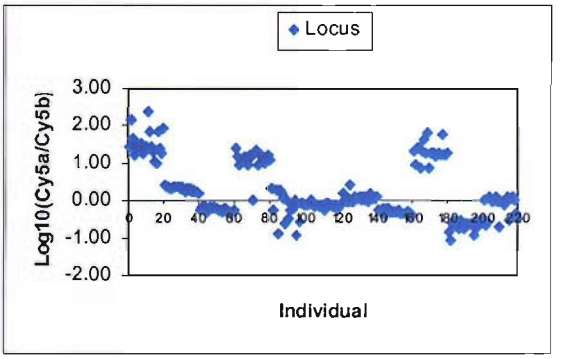


Figure4.33 Analyses graphs following Cy3 allele specific probe hybridisation (type 7 slides).

Figure 4.34a Cy3/Cy5 Type 7 *

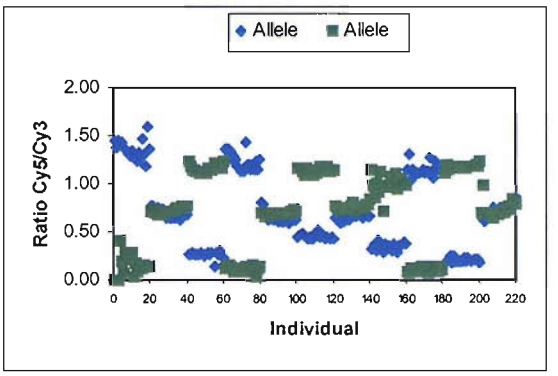


Figure 4.34b Log_{10}(Cy3a/Cy3b) Type 7 *

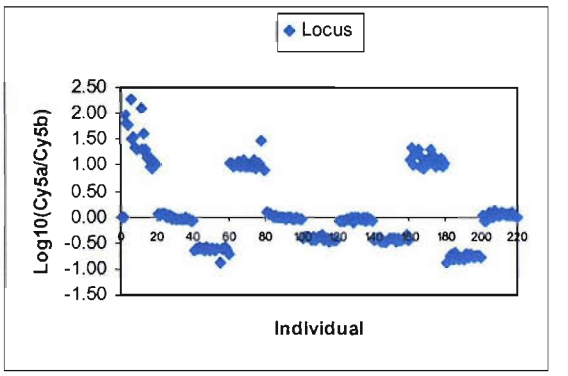


Figure 4.34 Analyses graphs following Cy3 allele specific probe hybridisation (type 7* slides).

The type 7* slides appeared to give better Log_{10} results using Cy3 allele-specific probes compared to the type 7 slides. In addition, the type 7* slides hybridised with Cy3 allele specific probes gave overall better log_{10} ratios than the type 7* slides hybridised with Cy5 allele specific probes.

Type 7 slides hybridised with Cy5 allele specific probes and type 7* slides hybridised with Cy3 allele specific probes showed comparable results therefore it would be advantageous to consider the specific slide type prior to hybridisation as this should determine which dye label should be present on the allele specific probe.

4.3.12 Slide Quality

With the introduction of type 7* slides to replace type 7 slides there was a noticeable increase in the degree of variability seen between slides. The detection strategy employed for microarray analysis in this investigation relied on the sequential spotting and hybridisation of a number of slides, each one being hybridised with a different probe mixture. Any variation between slides would inevitably result in a breakdown of the analysis process requiring repeat hybridisations to produce accurate genotype results. Typically the type of variation seen with the new slide type affected the degree to which the samples adhered to

the glass surface following spotting. Scanning images demonstrated that there were areas of extreme hydrophobicity where samples were unable to adequately attach (see figure 4.35)[1]

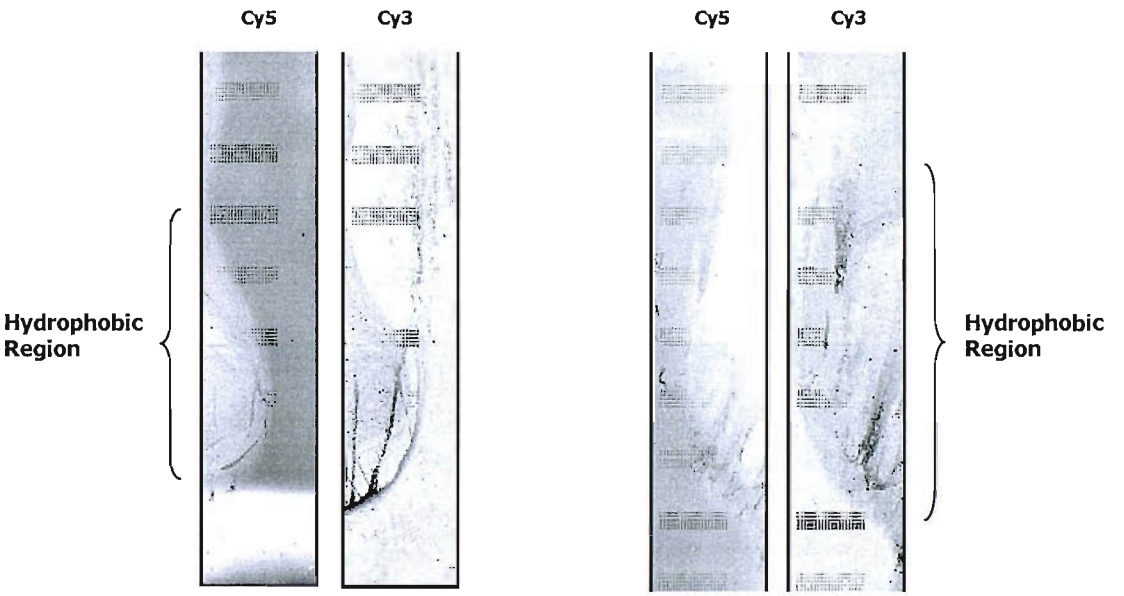


Figure 4.35 Cy3 and Cy5 images of type 7* slides following hybridisation and washing. Areas associated with hydrophobicity are indicated.

This observation was in keeping with that reported by Gabig *et al.,* (2001). They showed that aminosilanes enhanced the hydrophobicity of their slide surfaces; a factor that would explain these observations.

The consequence of this type of variation was all the more apparent given the way genotypes were assigned. For a given pair of slides, representing each allele of a particular locus, the combined result of the two slides at any given position determined that particular samples genotype. If at a given position, one of the slides exhibited the hydrophobic nature, then the sample would not bond efficiently and may be lost altogether during hybridisation and washing. Following scanning, a reduced Cy3 and/or Cy5 signal would mean that the sample may be classified as a negative with respect to that particular allele. If the signal was seen on the other slide, indicating the presence of the alternative allele, then the sample would be

[1] Images provided by Lindsey Dixon

incorrectly classified as a homozygote when in fact it should have been classified as a heterozygote.

In addition, the hydrophobic nature of the type 7* slide surface meant that following spotting, the sample dried in an uneven manner, concentrating the sample as the spot size reduced during evaporation. DMSO was mixed with the PCR product to prevent this phenomenon from occurring. However of the two slide types, type 7* slides regularly showed this occurrence. Type 7 slides showed a more even spot drying pattern (see figure 4.36).

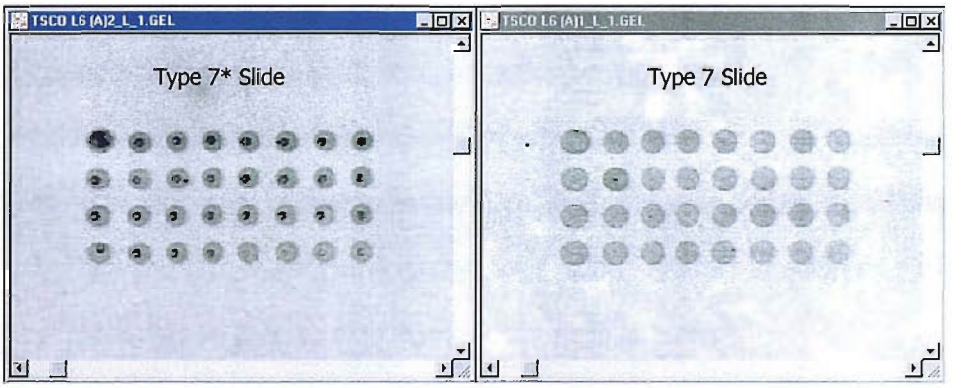


Figure 4.36 shows differences in spot density between type 7* and type 7 slides.

An even spot density gave rise to more accurate fluorescent readings at the analyses phase as there was less variation in pixel density across the spot. This type of variation in slide quality was never observed with the type 7 slides. For type 7* slides this factor should be considered when undertaking data interpretation and genotype scoring.

Given the variation seen in the type 7* slides with regards to product attachment, a reinvestigation of additional amine groups to enhance covalent attachment was undertaken. Two Cy3 labelled probes were taken; one of which was modified by the incorporation of a 5′ amino group; the other did not contain any amino modification. The two probe types were spotted onto a type 7* slides and processed as though undertaking a normal hybridisation and stringency washing protocol. Following processing, the slide was scanned for the presence of Cy3. The resulting images were obtained.

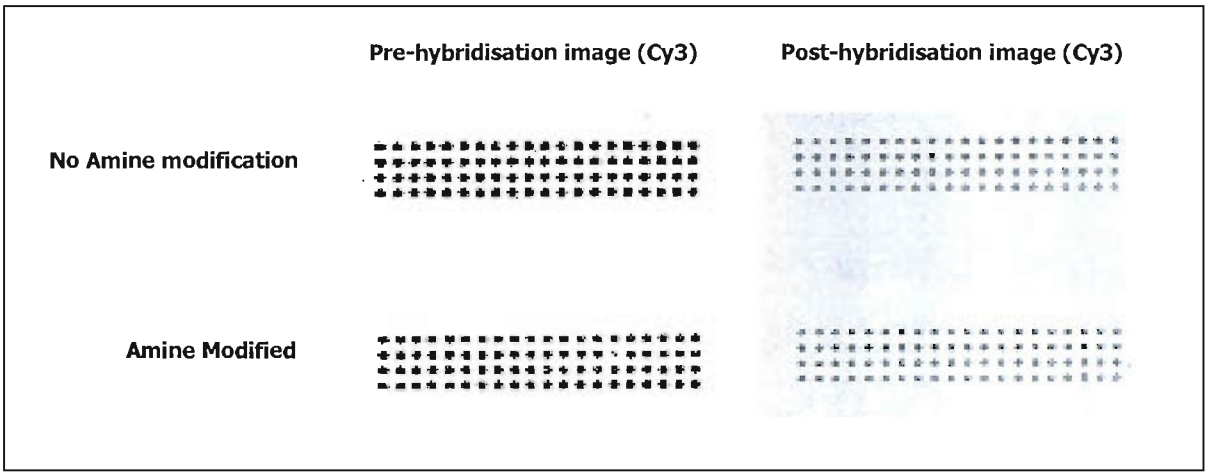


Figure 4.37 Spot images of probes containing amine modification vs. probes carrying no modification.

The images in figure 4.37 together with actual fluorescent intensity readings suggested that there was very little difference in spot intensity/morphology between probes containing an amine modification and probes not carrying any modification. Product attachment was not enhanced by this additional process step during probe manufacture.

4.3.13 Spotting and Pen set Blockage

For operation, the generation III spotter used a proprietary pen set comprising 12 quilled needles. During sample take up, approximately 300 nL sample was drawn into each needle via surface tension. During spotting, when the needles strike the surface of the slide, the surface tension of the contained liquid breaks. As the pen set moves away from the slide surface, the surface tension of the deposited liquid was restored. Each resulting spot carries approximately 500 pL of the sample. During each run, the pen set was automatically cleaned between each sample pick up to prevent sample carryover. In addition, routine cleaning is required comprising of a thorough wash in 18.2 MΩ water contained in a sonicating incubator. This was to ensure each needle was thoroughly cleaned however, it was possible for the pens to become blocked by the accumulation of salt within the quilled pen structure, usually between spotting runs.

A restored Pen set function was achieved only after an effective service of the wash station and thorough soaking in DMSO, figure 4.38.

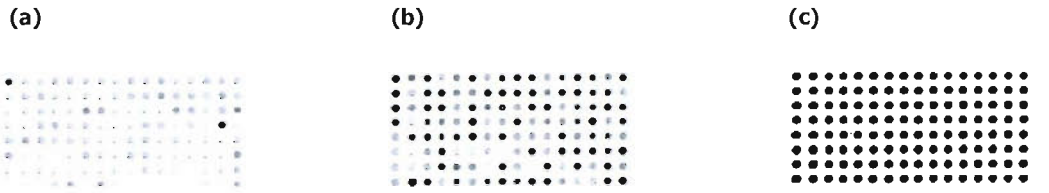


Figure 4.38 show three arrays spotted with 100 nM Cy3 labelled probe by the same pen.
(a) prior to service/cleaning
(b) immediately following a wash station service
(c) following a wash station service and a 24 hr soak in 25% DMSO

All of the spots in each array are replicates of the same probe sample and should generate a consistent spot morphology and intensity throughout the array. This was only evident in array (c). If left unchecked, this problem would severely affect the interpretation of results and would lead to unreliable genotyped data.

4.3.14 Primer-dimer

Throughout the investigation, there was an underlying theme which was particularly problematic with respect to microarray detection. This was the presence of primer-dimer following PCR amplification.

Primer-dimers occur as a result of non specific interactions between primers in the multiplex reaction. If such interactions give rise to 3' homology then this would have the potential to be acted upon by the *Taq* enzyme resulting in an extension of the dimer interaction. For primers that were present in very high concentrations in the multiplex, the amount of possible interactions would be extremely high and would rapidly take precedence as the prominent amplification sites. This would cause a rapid depletion of the essential resources of PCR, namely *Taq* and dNTPs and would result in a dramatic reduction of true PCR product formation. Similarly, if the multiplex size was increased, the number of different amplification

primers would also increase and so too would the number of possible unwanted interactions. If any 3' homology resulted from these interactions, then the same phenomena would occur.

For conventional gel based detections, this problem is overcome in part by the separation of primer-dimer from true PCR product. Electrophoresis separates products according to size and as primer-dimers tends to be smaller than true amplicons, they can be easily differentiated. For microarray analyses, separation of product by size does not occur. Since the targets for detection were the primer sequences themselves, it was not possible to differentiate true PCR product from primer-dimer and thus could potentially compromise the generation of an accurate result.

In an attempt to reduce the amount of the primer-dimer, PCR product was first purified through PCR clean-up columns. These contained a separation matrix that would bind or trap DNA products of a particular length depending on the column specification. For our purposes, the columns were designed to trap fragments greater than 100 bp. Primer-dimer sizes were dependent on the combination of the initial primer lengths forming the primer-dimer. For the URP methodology combined with ARMS, primers were typically around 35-45 bp meaning that primer-dimer sizes varied between 75-85 bp. These values fell below the minimum size specification for the column (100 bp) and so primer-dimers should pass through the column unhindered. In practice, this was not the case. Secondary structures between neighbouring primer-dimers as well as primer stacking could occur. This would mean that some primer- dimer, although much smaller than the 100 bp cut-off, may become trapped in the column during purification and ultimately be eluted along with true amplicon prior to spotting. Purification did reduce the amount of primer-dimer carried through to spotting however it did not eliminate it from the purified sample entirely.

Figure 4.39 shows an example of primer-dimer following pentaplex A-v4 amplification.

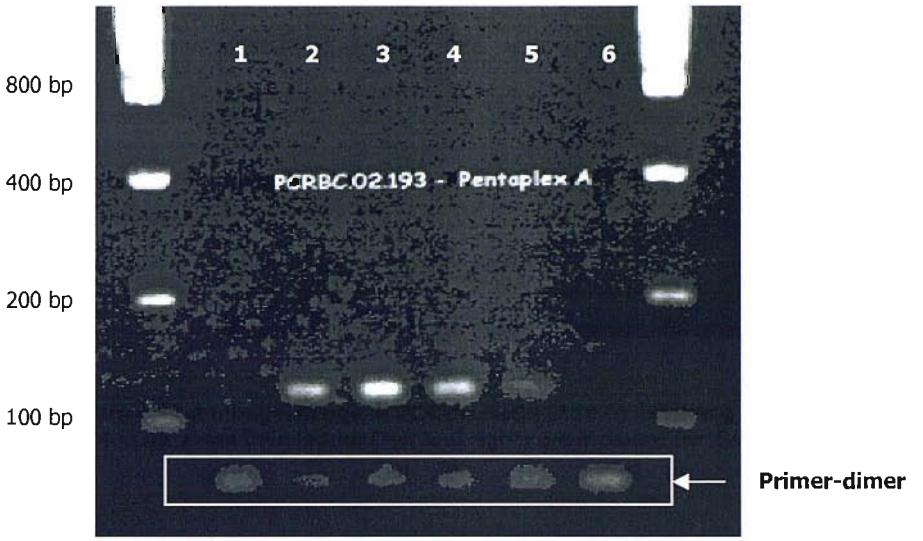


Figure 4.39 A minigel image of PCR products (1 to 6) following PCR amplification using pentaplex A-v4. Low mass DNA ladder (left and right) are included for determining band sizes.

Primer-dimer formation became problematic as the multiplex size was increased as in the above example. No visible primer-dimer was seen when amplification occurred as singleplex reactions (see figure 4.29, page 161).

From previous research work concerning the URP principle, gel based detection employed a slightly different amplification approach that had the effect of reducing the amount of primer-dimer following PCR. In those amplifications, a type of asymmetric amplification was set up whereby, the reaction was initially driven in the reverse direction by having the forward primer/s at a reduced concentration compared to the reverse primer/s. The reverse amplicons were then targeted by dye labelled forward universals thus driving the reaction in the forward direction. In the microarray approach, the reaction was initially driven in the forward direction producing a higher percentage of forward amplicons. These were then targeted by the reverse universal which produced reverse amplicon. During probe detection, it was these reverse amplicons which were subsequently targeted. Table 4.59 shows the different amplification strategies employed.

	Amplification Strategy 1 Gel based detection			Amplification Strategy 2 Microarray based detection		
	PCR for Gel based detection			**PCR for Microarray based detection**		
	Primer conc.	**PCR Phase 1**	**PCR Phase 2**	**Primer conc.**	**PCR Phase 1**	**PCR Phase 2**
Forward locus specific 1	50 nM	Drives Reverse	-	100 nM	Drives Forward	-
Forward locus specific 2	50 nM			100 nM		
Reverse locus specific 3	100 nM			50 nM		
Forward Universal 1	2 µM	-	Drives Forward	-	-	Drives Reverse
Forward Universal 2	2 µM			-		
Reverse Universal 3	-			2 µM		

Table 4.61 Amplification primer concentrations employed between the two different PCR strategies.

As a result of these two different strategies, different universal sequences were present within the PCR at different concentrations suggesting that that this would lead to a noticeable difference in primer-dimer accumulation following PCR.

To determine if primer-dimer content could be reduced whilst still maintaining the correct amplification orientation, a titration experiment was performed whereby amplification was carried out according to strategy 2 using primers for pentaplex B-v3, except that each amplification set contained a different concentration of universal 3. Following amplification, all samples were run on a 3% nusieve minigel as shown overleaf.

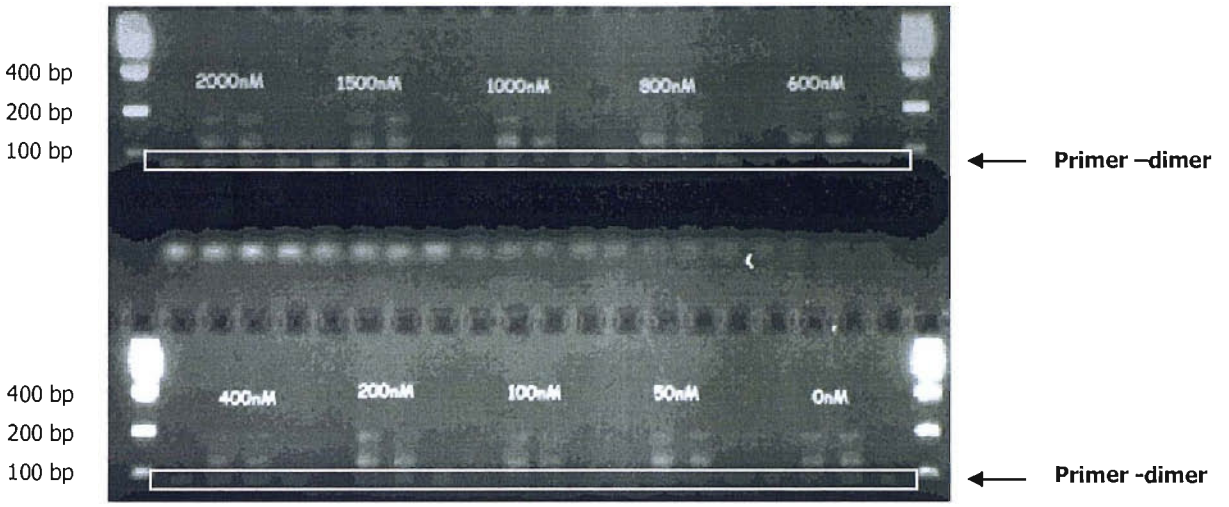


Figure 4.40 A minigel image showing pentaplex B-v3 products amplified according to strategy 1 and 2. Primer-dimer relating to each concentration is shown within the white boxed areas. Low DNA mass ladders are shown (left and right) for band sizing comparison.

As the concentration of universal 3 was reduced, there was a noticeable decrease in the amount of primer-dimer present following amplification. This primer-dimer did not reduce entirely even when no universal 3 was present in the multimix suggesting that this primer-dimer was derived from a different universal 3 substrate. A likely candidate would be the universal 3 sequence present on the 5' ends of the reverse locus specific primers.

Universal 3 shared a very similar sequence to universal 1 and as such were believed to be the principle cause of the primer-dimer phenomena. To combat this, various universal 3 sequence redesigns were undertaken and tested however, none of the redesigns successfully eliminated primer-dimer altogether.

As a consequence, attention was turned to the alternative amplification strategy where reduced primer-dimer was routinely seen. Although this method drove PCR in the opposite orientation to the normal microarray approach, the universal 3 sequence was not present as

a separate primer, but instead substituted by universals 1 and 2. This reduced the likelihood of universal 1 and 3 interacting and therefore mispriming to produce primer-dimer.

To test this theory, primers for pentaplex B-v3 were amplified according to the two amplification strategies. Following PCR, all samples were run on a 3% nusieve minigel for direct comparison of primer-dimer.

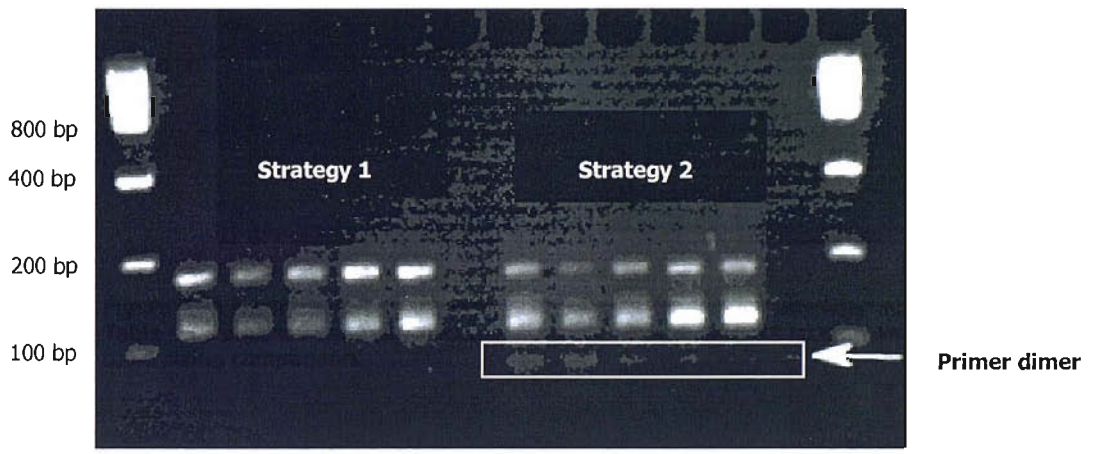


Figure 4.41 A minigel image showing pentaplex B-v3 products amplified according to strategy 1 and 2. Table 4.61 outlines each amplification strategy. Low mass DNA ladders are shown (left and right) for band sizing comparison.

From these results it was clear that the strategy employed to carry out PCR amplification directly accounted for the amount of primer-dimer seen in the resulting products. The only difference between the two strategies was the universal sequences employed in the third phase of PCR. This suggested that the substitution of universal 3 for universals 1 and 2 should be adopted for efficient amplification for future microarray experiments.

4.3.15 General discussion

The overall aim was to establish whether a microarray platform was a suitable alternative to conventional gel based SNP detection. This had to be considered in the context of forensic samples where only small quantities of biological material, capable of yielding enough

amplifiable DNA, are encountered. The microarray platform had the potential to offer a very high throughput solution for either intelligence or databasing purposes, however, both examples are heavily dependent on there being a suitable and efficient multiplex strategy.

The URP principle has been previously demonstrated as a means to rapidly construct large multiplexes and so in part fulfils the prerequisite for microarray detection. However, the results presented here suggest that the size of the multiplex was a limiting factor with regards to accurate genotyping. At best, a working multiplex comprising of up to nine loci, analysed by the indirect approach has been demonstrated. This is clearly someway short of the 50 SNPs required to generate the same level of discrimination as current STR based tests.

One of the continuing problems with this work has concerned the different slide types used for spotting, and the overall robustness of the generation III spotter. Initially, all work was undertaken with the type 7 slides until these were subsequently replaced by type 7* slides. Unfortunately, this second slide type did not perform as well as the type 7 slides due to variation in slide surface combined with regions of high hydrophobicity. In addition, the pen set used on the Generation III spotter was prone to blockage resulting in an uneven deposit of material.

Although the direct approach for genotyping was abandoned quite early on in the investigation, due to a shift in business emphasis, duplicate pentaplexes amplified according to the direct and the indirect approach seemed to offer comparable levels of sensitivity. Determining the limits of sensitivity for the direct approach would have been very useful and may have highlighted differences between the levels of discrimination between the two approaches. A possible reason for this was the means by which the amplicons were detected in the two strategies. The direct approach used probes that were completely different between loci and so the possibility of cross hybridisation with other loci in the multiplex minimised. For the indirect approach however, each probe pair shared a common sequence with the other probes represented and so the technique was more reliant on efficient

stringency washing to prevent cross hybridisation. Whilst the direct approach served to minimise the possibility of inter-locus cross hybridisation, this advantage might be offset by the increased likelihood of inter-allele cross hybridisation. Conversely, the indirect approach maximised the differentiation of alleles within a given locus but offered a reduced level of discrimination between loci.

As previously considered, the size of the multiplex was the biggest limiting factor for microarray detection. As the multiplex size increased, the overall amount and density of the constituent product (amplicons) also increased, meaning that more and more template was contained within each spot array. Steric hindrance and secondary structure exhibited by the template DNA can subsequently prevent probe sequences from finding their target site, resulting in an overall reduction in signal intensity for any given probe locus. The only way to overcome this problem was to limit the size of any one multiplex so that the amount of amplicon within a PCR product was not over-concentrated. This conclusion was in agreement with that of Hacia *et al.,* (1998) and Pastinen *et al.,* (2000), who concluded that the detection of PCR product from multiplexes of greater than 10 markers was extremely difficult to achieve. Where sample quantities are not limited, such as with Criminal Justice samples (buccal swabs or liquid blood), an improved strategy would be to generate a series of smaller multiplexes, each containing a different 3-5 locus complement. Then at the spotting stage, instead of having a single spot per individual, containing products from up to a 50-plex SNP amplification, multiple spots per individual, each containing product from a different multiplex amplification would be represented. Although the capacity of the array would be reduced, the sensitivity of the detection should be enhanced.

Due to problems with slide quality, equipment reliability and a shift in business focus further development of this methodology was halted in preference for more conventional DNA detection methodologies.

5 Conclusions

Conclusions for this work are separated into three areas, each specific to one of the three main themes investigated:

5.1 Microfabrication and DNA Extraction

Silicon microfabrication resulted in a number of serpentine channels, each one a fixed length of either 300 mm or 1000 mm long. The channel architecture was successfully combined with a modified Qiagen™ chemistry to act as a support for DNA extraction. The results obtained from the investigation suggested that the devices were suitable for performing DNA extraction. In addition the devices demonstrated that both DNA sample clean-up and DNA sample pre-concentration were possible. The efficiency and ease of operation minimised the need for complicated experimental protocols and the geometry of the device minimised the possibility of sample blockage. These advantages together with the reliability of the device suggested that integration to a millifluidic system would be relatively straightforward. The novel aspect of the investigation resulted in the submission of a patent application.

5.2 DNA Duplex stability

Through assessment of carefully designed oligonucleotide duplexes, a reliable algorithm for predicting duplex melting temperature for complementary sequences was deduced. Compared to the commercially available software, Primer Express™, MOSAIC yielded more accurate *Tm* predictions for a variety of oligonucleotide sequences at salt concentration more in line with those required for PCR amplification. Trends in single mismatch stability confirmed previously reported data. In addition these trends were ordered according to

adjacent Watson Crick contexts. In most though not all instances, G•C and C•G contexts stabilised the mismatch to a greater extent than A•T or T•A.

The algorithm was successfully incorporated into a visual Basic computer program, MOSAIC to enable rapid melting temperature assessment for novel primer designs. This programme should facilitate faster more reliable multiplex primer design in the future.

5.3 Microarray platform for DNA detection

The development of microarray technology for DNA analysis proved to be challenging, however, the platform could be used to detect single nucleotide polymorphisms derived from multiplex amplifications. Two different strategies for detection were investigated; direct hybridisation to the SNP target and indirect hybridisation to a universal reporter tail. Both methods demonstrated equivalent discrimination efficiencies for small multiplexes (<6 loci). Bringing together larger numbers of loci (up to 10), in a multiplex was achieved using the URP principle but their detection was problematic on this scale. Background was greatly increased in these situations making genotyping more difficult. In summary, these results suggested that a microarray approach for sample genotyping was probably more suited to databasing rather than intelligence. In these circumstances, DNA material would not be the limiting factor and so would allow for smaller component multiplexes to be used. For intelligence purposes, where the amount of starting material is limited and therefore a requirement for larger multiplexes, conventional analytical techniques such as capillary electrophoresis should be preferentially used.

6 Appendix

6.1 Appendix 1 – Oligonucleotide sequences

6.1.1 Duplex oligonucleotide sequences

6.1.1.1 19-mer sequences

Generic sequence for Strand A: 5′ FAM-CGACGTGC**NNN**ATGTGCTG 3′

where N = A, C, G or T

Generic sequence for Strand B: 5′ CAGCACAT**NNN**GCACGTCG-3′ MR

where N = A, C, G or T

6.1.1.2 Fret Test oligos

Single oligos	**Duplex oligo pairs**
5′FAM, A AAA = FAM-CGACTGC**AAA**ATGTGCTG	5′FAM, A AAA = FAM-CGACTGC**AAA**ATGTGCTG B′ AAA – 3′ DABCYL = CAGCACAT**AAA**GCACGTCG-DABCYL
5′FAM, A TTT = FAM-CGACTGC**TTT**ATGTGCTG	5′FAM, A TTT = FAM-CGACTGC**TTT**ATGTGCTG B′ AAA – 3′ DABCYL = CAGCACAT**AAA**GCACGTCG-DABCYL
5′FAM, A GGG = FAM-CGACTGC**GGG**ATGTGCTG	5′FAM, A GGG = FAM-CGACTGC**GGG**ATGTGCTG B′ CCC – 3′ DABCYL = CAGCACAT**CCC**GCACGTCG-DABCYL

6.1.1.3 17-mer sequences

5′ FAM, A, GGG = FAM-GACGTGC**GGG**ATGTGCT

5′ FAM, A, TTT = FAM-GACGTGC**TTT**ATGTGCT

B, CCC-3′ DABCYL = AGCACAT**CCC**GCACGTC-DABCYL

6.1.1.4 Duplex Validation (I) sequences

	Strand A	**Strand B**
Duplex 1	FAM-GCAGGTCGTC	GACGACCTGC-DABCYL
Duplex 2	FAM-GCAGGTAGTC	GACTACCTGC-DABCYL
Duplex 3	FAM-GCTGCTTGCTGAACG	CGTTCAGCAAGCAGC-DABCYL
Duplex 4	FAM-GACTCTTGCTGAATG	CATTCAGCAAGAGTC-DABCYL
Duplex 5	FAM-GCAGTAGGCTGCACCGAACG	CGTTCGGTGCAGCCTACTGC-DABCYL
Duplex 6	FAM-GACTGATGCTGCATCGAACG	CGTTCGATGCAGCATCAGTC-DABCYL
Duplex 7	FAM-CGTGGCAGCACTACGCAGAGCCATG	CATGGCTCTGCGTAGTGCTGCCACG-DABCYL
Duplex 8	FAM-CGGTGCATTTCGGATGCATACGCATG	CATGCGTATGCATCCGAAATGCACCG-DABCYL
Duplex 9	FAM-CGTCCGAGCACGTGTACGCAGAGCCATG	CATGGCTCTGCGTACACGTGCTCGGACG-DABCYL
Duplex 10	FAM-CGGTACAGCACATGTACGTAGAGTCATG	CATGACTCTACGTACATGTGCTGTACCG-DABCYL

6.1.1.5 Duplex Validation (II) sequences

	Strand A	**Strand B**
Duplex 11	FAM-CGTCTCAGTCAGTGTCTCC	GGAGACACTGACTGAGACG-DABCYL
Duplex 12	FAM-GTCGATCGCTGTGACGTCG	CGACGTCACAGCGATCGAC-DABCYL
Duplex 13	FAM-GCTGTATCGACGTGAGTGG	CCACTCACGTCGATACAGC-DABCYL
Duplex 14	FAM-GAGTGCACAACGGTGGTGC	GCACCACCGTTGTGCACTC-DABCYL
Duplex 15	FAM-GACGTGCAGATGTGC	GCACATCTGCACGTC-DABCYL
Duplex 16	FAM-GACGTGCCGATGTGC	GCACATCGGCACGTC-DABCYL
Duplex 17	FAM-GACGTGCGGATGTGC	GCACATCCGCACGTC-DABCYL
Duplex 18	FAM-GCTCGACGTGCAAGATGTGCTGACG	CGTCAGCACATCTTGCACGTCGAGC-DABCYL
Duplex 19	FAM-GCTCGACGTGCACGATGTGCTGACG	CGTCAGCACATCGTGCACGTCGAGC-DABCYL
Duplex 20	FAM-GCTCGACGTGCAGGATGTGCTGACG	CGTCAGCACATCCTGCACGTCGAGC-DABCYL

6.1.2 Microarray PCR primer and probe sequences

Prefix 'p' denotes probe. Suffices '(i)' and '(d)' denote primers for the indirect and direct approaches respectively.

Gc1s Artificial template

+ Amine NH_2-CGACGTGGTGGATGTGCTAGGTTCCGTGGGTGTGGCC$(N)_{58}$TTTGCTTTTAGTCG
CTCTGCCCTAGCACATCCACCACGTCG
- Amine CGACGTGGTGGATGTGCTAGGTTCCGTGGGTGTGGCC$(N)_{58}$TTTGCTTTTAGTCGCTC
TGCCCTAGCACATCCACCACGTCG

PGM

PGM + (i)	CGACGTGGTGGATGTGCTAG
PGM – (i)	TGACGTGGCTGACCTGAGAC
PGM RLS (i)	CGACGTGGTGGATGTGCTAG
Uni 3 Cy3	Cy3-CGACGTGGTGGATGTGCTAG
pPGM + (i)	Cy5-GATGTGCTAG
pPGM – (i)	Cy5-GACCTGAGAC

PENTAPLEX A-v1 (Indirect approach)

tsco D/T(i)	CGACGTGGTGGATGTGCTAGGGGAAACTGCTGGGTCTGT	
tsco D/C(i)	TGACGTGGCTGACCTGAGACGGGAAACTGCTGGGTCTGC	
tsco D/RLS (i)	CTAGCTGGTGGCTGTGCTAGCAGCTCTTGACCCTGGCTG	157bp
tsco F/C (i)	CGACGTGGTGGATGTGCTAGCCTGGAGCATGAGCTGACCC	
tsco F/A (i)	TGACGTGGCTGACCTGAGACCCTGGAGCATGAGCTGACCA	
tsco F/RLS (i)	CTAGCTGGTGGCTGTGCTAGGGCTCTGAAGAACAATGGGGA	168bp
tsco X/C (i)	CGACGTGGTGGATGTGCTAGCTGTGCATCCACTGCGCC	
tsco X/A (i)	TGACGTGGCTGACCTGAGACCTGTGCATCCACTGCGCA	
tsco X/RLS (i)	CTAGCTGGTGGCTGTGCTAGTCTAGGCTGGTGCCAGCCC	162bp
tsco L2/C (i)	CGACGTGGTGGATGTGCTAGGCATGCCATTGCCAAATTCC	
tsco L2/T (i)	TGACGTGGCTGACCTGAGACGCATGCCATTGCCAAATTCT	
tsco L2/RLS (i)	CTAGCTGGTGGCTGTGCTAGGATGCCCATGGCAGGACC	178bp
tsco Z2/C (i)	CGACGTGGTGGATGTGCTAGCATTGTGTTTCAAACGCGTGCC	
tsco Z2 /T (i)	TGACGTGGCTGACCTGAGACCATTGTGTTTCAAACGCGTGCT	
tsco Z2/RLS (i)	CTAGCTGGTGGCTGTGCTAGCAGCTGGTCAAGCCCAGGA	136bp
ptsco D/T (i)	Cy5-GATGTGCTAGGGGAAACTGC	
ptsco D/C (i)	Cy5-GACCTGAGACGGGAAACTGC	
ptsco D/RLS (i)	Cy3-CAGCCAGGGTCAAGAGCTG	
ptsco F/A (i)	Cy5-GACCTGAGACCCTGGAGCAT	
ptsco F/C (i)	Cy5-GATGTGCTAGCCTGGAGCAT	
ptsco F/RLS (i)	Cy3-TCCCCATTGTTCTTCAGAGCC	
ptsco X/C (i)	Cy5-GATGTGCTAGCTGTGCATCC	
ptsco X/A (i)	Cy5-GACCTGAGACCTGTGCATCC	
ptsco X/RLS (i)	Cy3-GGGCTGGCACCAGCCTAGA	
ptsco L2/C (i)	Cy5-GATGTGCTAGGCATGCCATT	
ptsco L2/T (i)	Cy5-GACCTGAGACGCATGCCATT	
ptsco L2/RLS (i)	Cy3-TCCCCACACCTTGGCTCA	
ptsco Z2/C (i)	Cy5-GATGTGCTAGCATTGTGTTTT	
ptsco Z2/T (i)	Cy5-GACCTGAGACCATTGTGTTTT	
ptsco Z2/RLS (i)	Cy3-CCAGCTGCACTTTCTCTGATCCT	

PENTAPLEX A-v1 (Direct approach)

tsco D/F (d)	TGACGTGGCTGACCTGAGACTCTAGGTCCATCCCAGCATG	186bp
tsco D/RLS (d)	CGACGTGGTGGATGTGCTAGCAGCTCTTGACCCTGGCTG	
tsco F/F (d)	TGACGTGGCTGACCTGAGACAGGGTCTGTCTTCTTGCCTGG	184bp
tsco F/R (d)	CGACGTGGTGGATGTGCTAGGGCTCTGAAGAACAATGGGGA	
tsco X/F (d)	TGACGTGGCTGACCTGAGACCCGAATGCTTCAGTGAAATGTGT	162bp
tsco X/R (d)	CGACGTGGTGGATGTGCTAGTCTAGGCTGGTGCCAGCCC	
tsco L2/F (d)	TGACGTGGCTGACCTGAGACATCCGTTTTGCATGCCATTG	178bp
tsco L2/R (d)	CGACGTGGTGGATGTGCTAGTGAGCCAAGGTGTGGGGA	
tsco Z2/F (d)	TGACGTGGCTGACCTGAGACGGTCACATTGTGTTTCAAACG	136bp
tsco Z2/F (d)	CGACGTGGTGGATGTGCTAGGGATCAGAGAAAGTGCAGCTGG	
ptsco D/C (d)	Cy5-CTGGGTCTGCACCCCCACC	
ptsco D/T (d)	Cy5-CTGGGTCTGTACCCCCACC	
ptsco D/RLS (d)	Cy3-CAGCCAGGGTCAAGAGCTG	
ptsco F/A (d)	Cy5-AGCTGACCAAAGGGGGTCC	
ptsco F/C (d)	Cy5-AGCTGACCACAGGGGGTCC	
ptsco F/RLS (d)	Cy3-TCCCCATTGTTCTTCAGAGCC	
ptsco X/A (d)	Cy5-CCACTGCGCACTCTGCCCT	
ptsco X/C (d)	Cy5-CCACTGCGCCCTCTGCCCT	
ptsco X/RLS (d)	Cy3-GGGCTGGCACCAGCCTAGA	
ptsco L2/C (d)	Cy5-GCCAAATTCCTCCCAGAGC	
ptsco L2/T (d)	Cy5-GCCAAATTCTTCCCAGAGC	
ptsco L2/RLS (d)	Cy3-TCCCCACACCTTGGCTCA	
ptsco Z2/C (d)	Cy5-AACGCGTGCCGCTGTCTGT	
ptsco Z2/T (d)	Cy5-AACGCGTGCTGCTGTCTGT	
ptsco Z2/RLS (d)	Cy3-CCAGCTGCACTTTCTCTGATCC	

PENTAPLEX A-v2 (Indirect approach)

As for A-v1 except tsco X substituted for tsco I3

tsco I3/F (d)	TGACGTGGCTGACCTGAGACGCTTCATTTCACAAAGCTGATCTG	150bp
tsco I3/R (d)	CGACGTGGTGGATGTGCTAGTGCCTTTGAGAACTCAGTGCCT	
ptsco I3/C (d)	Cy5-CACCTATTACGTCTACAAG	
ptsco I3/G (d)	Cy5-CACCTATTAGGTCTACAAG	
ptsco I3/RLS (d)	Cy3- AGGCACTGAGTTCTCAAAGGCA	

PENTAPLEX A-v2 (Direct approach)

As for A-v1 except tsco X substituted for tsco I3

tsco I3/C (i)	CGACGTGGTGGATGTGCTAGCACAAAGCTGATCTGAGCACCTATTAC	
tsco I3/G (i)	TGACGTGGCTGACCTGAGACCACAAAGCTGATCTGAGCACCTATTAG	
tsco I3/RLS (i)	CTAGCTGGTGGCTGTGCTAGTGCCTTTGAGAACTCAGTGCCT	150bp
ptsco I3/C (i)	Cy5-GATGTGCTAGCACAAAGCTG	
ptsco I3/G (i)	Cy5-GACCTGAGACCACAAAGCTG	
ptsco I3/RLS (i)	Cy3-AGGCACTGAGTTCTCAAAGGCA	

PENTAPLEX A-v3 (Indirect approach)

As for A-v2 except tsco I3 substituted for tsco G

tsco G/G (i)	CGACGTGGTGGATGTGCTAGCCATGCCTCACCTCCTGCATC	
tsco G/C (i)	TGACGTGGCTGACCTGAGACCCATGCCTCACCTCCTGCATG	
tsco G/RLS (i)	CTAGCTGGTGGCTGTGCTAGCCTGCCCCTTCCTGGAAAG	119bp
ptsco G/C (i)	Cy5-GACCTGAGACCCATGCCTCA	
ptsco G/T (i)	Cy5-GATGTGCTAGCCATGCCTCA	
ptsco G/RLS (i)	Cy3-CTTTCCAGGAAGGGGCAGG	

PENTAPLEX A-v3 Smaller amplicons (Indirect approach)

tsco D/RLS (i)*	CTAGCTGGTGGCTGTGCTAGTTTTAGCTACAAATGACCTGCCCC	118bp
tsco F/RLS (i)*	CTAGCTGGTGGCTGTGCTAGGGAAGTCCCTTCCAGCATCTG	122bp
tsco L2/RLS (i)*	CTAGCTGGTGGCTGTGCTAGTGAGCCAAGGTGTGGGGA	121bp
tsco Z2/RLS (i)*	CTAGCTGGTGGCTGTGCTAGCAGCTGGTCAAGCCCAGGA	121bp
tsco I3/RLS (i)*	CTAGCTGGTGGCTGTGCTAGGGCAAGGGTCCTGTCTTGCT	124bp
ptsco D/RLS (i) *	Cy3-GGGGCAGGTCATTTGTAGTCTAAA	
ptsco F/RLS (i) *	Cy3-CCTTCAGGGAAGGTCGTAGAC	
ptsco L2/RLS (i) *	Cy3-TCCCCACACCTTGGCTCA	
ptsco Z2/RLS (i) *	Cy3-TCCTGGGCTTGACCAGCTG	
ptsco I3/RLS (i) *	Cy3-AGCAAGACAGGACCCTTGCC	

PENTAPLEX B-v1 (Indirect approach)
Including tsco I3 + redesigned RLS to give a smaller amplicon

tsco J2/C (i)	CGACGTGGTGGATGTGCTAGCTGCCTTGGCTCCCAGCC	
tsco J2/T (i)	TGACGTGGCTGACCTGAGACCTGCCTTGGCTCCCAGCT	
tsco J2/RLS (i)	CTAGCTGGTGGCTGTGCTAGCCTGAACATCCCTGAAGGTATTTCG	145bp
tsco K3/G (i)	CGACGTGGTGGATGTGCTAGTGCCACTCTGACACTGATGCTTG	
tsco K3/C (i)	TGACGTGGCTGACCTGAGACTGCCACTCTGACACTGATGCTTC	
tsco K3/RLS (i)	CTAGCTGGTGGCTGTGCTAGTCTCTGGATGGGCCCAATG	119bp
tsco O3/C (i)	CGACGTGGTGGATGTGCTAGCTTAGATGTCACCTCTTCCATGCAC	
tsco O3/G (i)	TGACGTGGCTGACCTGAGACCTTAGATGTCACCTCTTCCATGCAG	
tsco O3/RLS (i)	CTAGCTGGTGGCTGTGCTAGTTTGAAAGCACCCAGGAATGG	124bp
tsco W3/G (i)	CGACGTGGTGGATGTGCTAGGCCAACCAGACCTCCCAGG	
tsco W3/C (i)	TGACGTGGCTGACCTGAGACGCCAACCAGACCTCCCAGC	
tsco W3/RLS (i)	CTAGCTGGTGGCTGTGCTAGAAACACAGGTCTCCAGCTTGAGC	183bp
tsco I3/RLS (i)	CTAGCTGGTGGCTGTGCTAGGGCAAGGGTCCTGTCTTGCT	124bp
ptsco I3/RLS (i) *	Cy3-AGCAAGACAGGACCCTTGCC	
ptsco J2/C (i)	Cy5-GATGTGCTAGCTGCCTTGGC	
ptsco J2/T (i)	Cy5-GACCTGAGACCTGCCTTGGC	
ptsco J2/RLS (i)	Cy3-CGAAATACCTTCAGGGATGTTCAGG	
ptsco K3/C (i)	Cy5-GACCTGAGACTGCCACTCTG	
ptsco K3/G (i)	Cy5-GATGTGCTAGTGCCACTCTG	
ptsco K3/RLS (i)	Cy3-CATTGGGCCCATCCAGAGA	
ptsco O3/C (i)	Cy5-GATGTGCTAGCTTAGATGTC	
ptsco O3/G (i)	Cy5-GACCTGAGACCTTAGATGTC	
ptsco O3/RLS (i)	Cy3-CCATTCCTGGGTGCTTTCAAA	
ptsco W3/G (i)	Cy5-GATGTGCTAGGCCAACCAGA	
ptsco W3/C (i)	Cy5-GACCTGAGACGCCAACCAGA	
ptsco W3/RLS (i)	Cy3-GCTCAAGCTGGAGACCTGTGTTT	

PENTAPLEX A-v4 (Indirect approach)
Same as Av3 except redesigned reverse locus specific primers for smaller amplicons
tsco D, G, L2, Z3, I3

PENTAPLEX B-v2 (Indirect approach)
Same as Bv1 except tsco I3 replaced with tsco N4, tsco O3 replaced by tsco P5

tsco N4/A (i)	CGACGTGGTGGATGTGCTAGCAGAAAAGGCAGGAACCTGGACA
tsco N4/T (i)	TGACGTGGCTGACCTGAGACCAGAAAAGGCAGGAACCTGGACT
tsco N4/RLS (i)	CTAGCTGGTGGCTGTGCTAGGACGGGGGTTGAGTGGTTCA
tsco P5/A (i)	CGACGTGGTGGATGTGCTAGGGGGGTACTGGGGAGACCAA
tsco P5/T (i)	TGACGTGGCTGACCTGAGACGGGGGTACTGGGGAGACCAT
tsco P5/RLS (i)	CTAGCTGGTGGCTGTGCTAGGGGAAGGCGGATCCTGGA
ptsco N4/A (i)	Cy5-GATGTGCTAGCAGAAAAGGCA
ptsco N4/T (i)	Cy5-GACCTGAGACCAGAAAAGGCA
ptsco N4/RLS (i)	Cy3-TGAACCACTCAACCCCCGTC
ptsco P5/A (i)	Cy5-GATGTGCTAGGGGGGTACTGG
ptsco P5/T (i)	Cy5-GACCTGAGACGGGGGTACTGG
ptsco P5/RLS (i)	Cy3-TCCAGGATCCGCCTTCCC

PENTAPLEX B-v3 (Indirect approach)
Same as Bv2 except tsco J2 was replaced with tsco L6

tsco L6/A (i)	CGACGTGGTGGATGTGCTAGTGTGCATGTTCCCTGGTGTTCA
tsco L6/T (i)	TGACGTGGCTGACCTGAGACTGTGCATGTTCCCTGGTGTTCT
tsco L6/RLS (i)	CTAGCTGGTGGCTGTGCTAGACTCGGCACAAGAGAGGCCA
ptsco L6/A (i)	Cy5-GATGTGCTAGTGTGCATGTTC
ptsco L6/T (i)	Cy5-GACCTGAGACTGTGCATGTTC
ptsco L6/RLS (i)	Cy3-TGGCCTCTCTTGTGCCGAGT

PENTAPLEX C-v1 (Indirect approach)

tsco A4/C (i)	CGACGTGGTGGATGTGCTAGGATGCCTCTTGCATTGTGAACG
tsco A4/G (i)	TGACGTGGCTGACCTGAGACGATGCCTCTTGCATTGTGAACC
tsco A4/RLS (i)	CTAGCTGGTGGCTGTGCTAGCAACGTCAACAGCACAACTCTGC
tsco K4/G (i)	CGACGTGGTGGATGTGCTAGGCGGGAGGAAGGAAGGGAGG
tsco K4/C (i)	TGACGTGGCTGACCTGAGACGCGGGAGGAAGGAAGGGAGC
tsco K4/RLS (l)	CTAGCTGGTGGCTGTGCTAGAGAGGAAGCCGCCCCTG
tsco Z5/A (i)	CGACGTGGTGGATGTGCTAGGGCATACCCATGTGACTAAATGTTGA
tsco Z5/T (i)	TGACGTGGCTGACCTGAGACGGCATACCCATGTGACTAAATGTTGT
tsco Z5/RLS (i)	CTAGCTGGTGGCTGTGCTAGGAGCAAAGAAGGGTCGGCA
tsco B6/A (i)	CGACGTGGTGGATGTGCTAGGGGAGACAGGCCCATGCA
tsco B6/T (i)	TGACGTGGCTGACCTGAGACGGGAGACAGGCCCATGCT
tsco B6/RLS (i)	CTAGCTGGTGGCTGTGCTAGGCCATTCAGAACTAACTAGTCTGGGA
tsco O6/A (i)	CGACGTGGTGGATGTGCTAGGAGCCAAGAATCGCAGGGAA

tsco O6/T (i)	TGACGTGGCTGACCTGAGACGAGCCAAGAATCGCAGGGAT
tsco O6/RLS (i)	CTAGCTGGTGGCTGTGCTAGGCTAAAGCAGCTCTGAAACCCA
ptsco A4/C (i)	Cy5-GATGTGCTAGGATGCCTCTTG
ptsco A4/G (i)	Cy5-GACCTGAGACGATGCCTCTTG
ptsco A4/RLS (i)	Cy3-GCAGAGTTGTGCTGTTGACGTTG
ptsco K4/G (i)	Cy5-GATGTGCTAGGCGGGAGGAA
ptsco K4/C (i)	Cy5-GACCTGAGACGCGGGAGGAA
ptsco K4/RLS (i)	Cy3-CAGGGGCGGCTTCCTCT
ptsco Z5/A (i)	Cy5-GGATGTGCTAGGGCATACCCA
ptsco Z5/T (i)	Cy5-GGACCTGAGACGGCATACCCA
ptsco Z5/RLS (i)	Cy3-TGCCGACCCTTCTTTGCTC
ptsco B6/A (i)	Cy5-GGATGTGCTAGGGGAGACAGG
ptsco B6/T (i)	Cy5-GGACCTGAGACGGGAGACAGG
ptsco B6/RLS (i)	Cy3-TCCCAGACTAGTTAGTTCTGAATGGC
ptsco O6/A (i)	Cy5-GGATGTGCTAGGAGCCAAGAATC
ptsco O6/T (i)	Cy5-GGACCTGAGACGAGCCAAGAATC
ptsco O6/RLS (i)	Cy3-TGGGTTTCAGAGCTGCTTTAGC

Pentaplex A-v1 Cy3 Cy5 reversed dye labelled

ptsco D/T (i)Rev	Cy3-GATGTGCTAGGGGAAACTGC
ptsco D/C (i)Rev	Cy3-GACCTGAGACGGGAAACTGC
ptsco D/RLS (i) Rev	Cy5-CAGCCAGGGTCAAGAGCTG

Pentaplex A-v1 Amine modified Cy3 dye labelled

ptsco Z2/C (i) Mod	NH_2-GATGTGCTAGCATTGTGTTTT-Cy3
ptsco Z2/C (i) Mod	GATGTGCTAGCATTGTGTTTT-Cy3

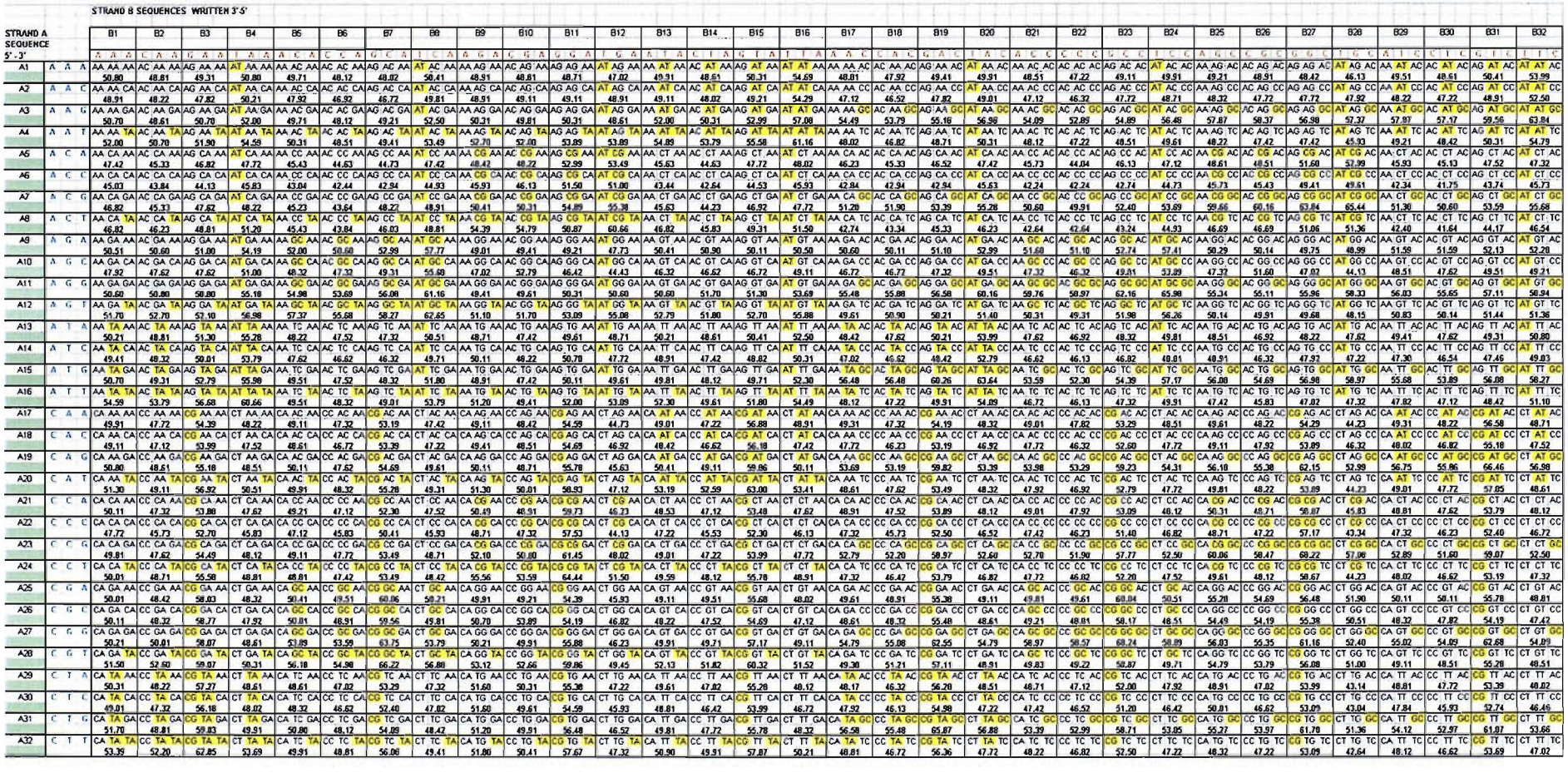
Unlabelled Universals

Universal 1	CGACGTGGTGGATGTGCTAG
Universal 2	TGACGTGGCTGACCTGAGAC
Universal 3	**CTAGCTGGTGGCTGTGCTAG**

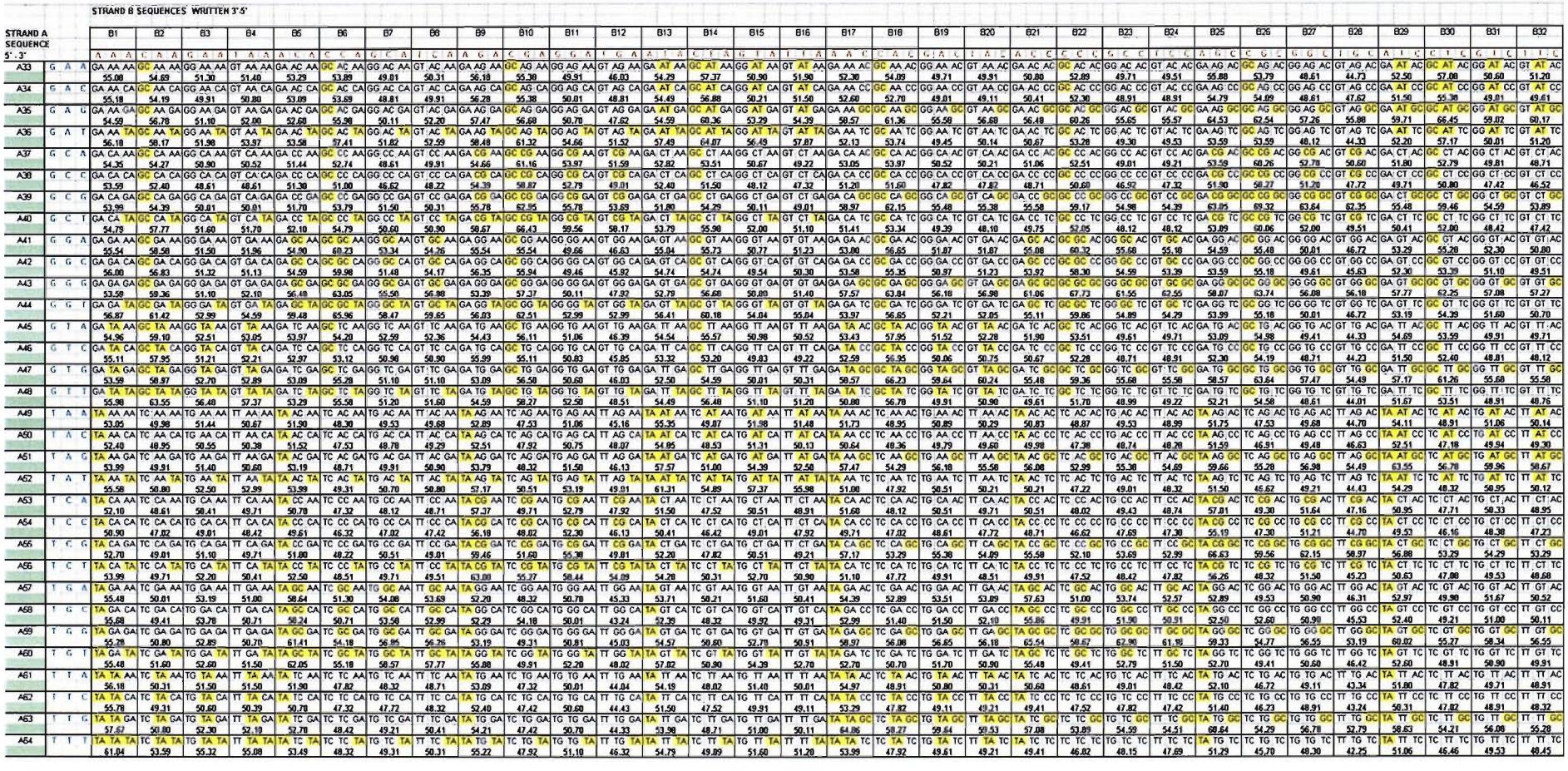
6.2 Appendix 2 – Duplex Melting Temperature data

Data sequences A1-A32/B1 to B32.

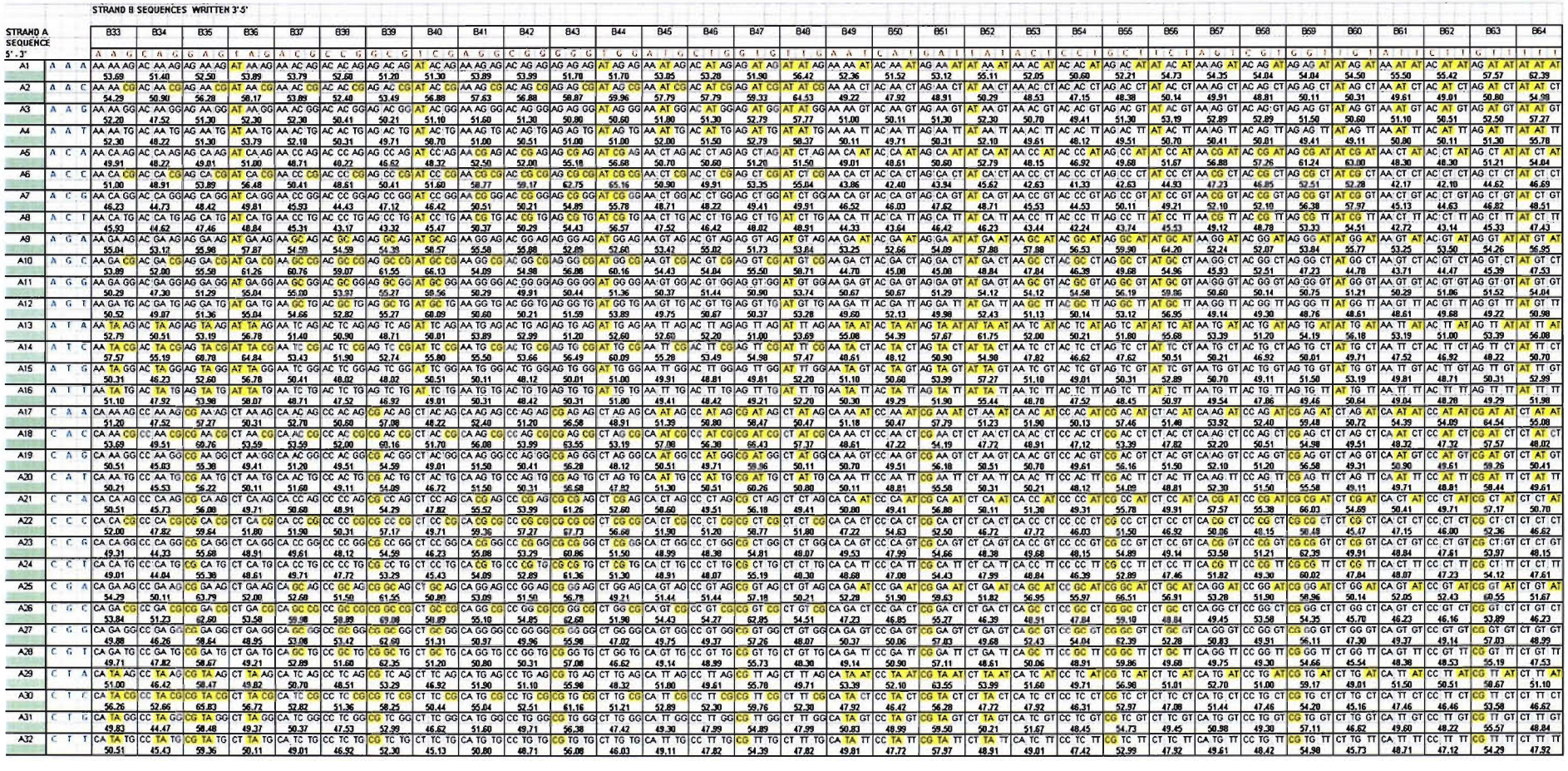
NB. *Tms* are shown in °C. Only the three central bases are shown as doublets. Each doublet is the result of Strand A base/Strand B base. Yellow indicates a complementary base pair.



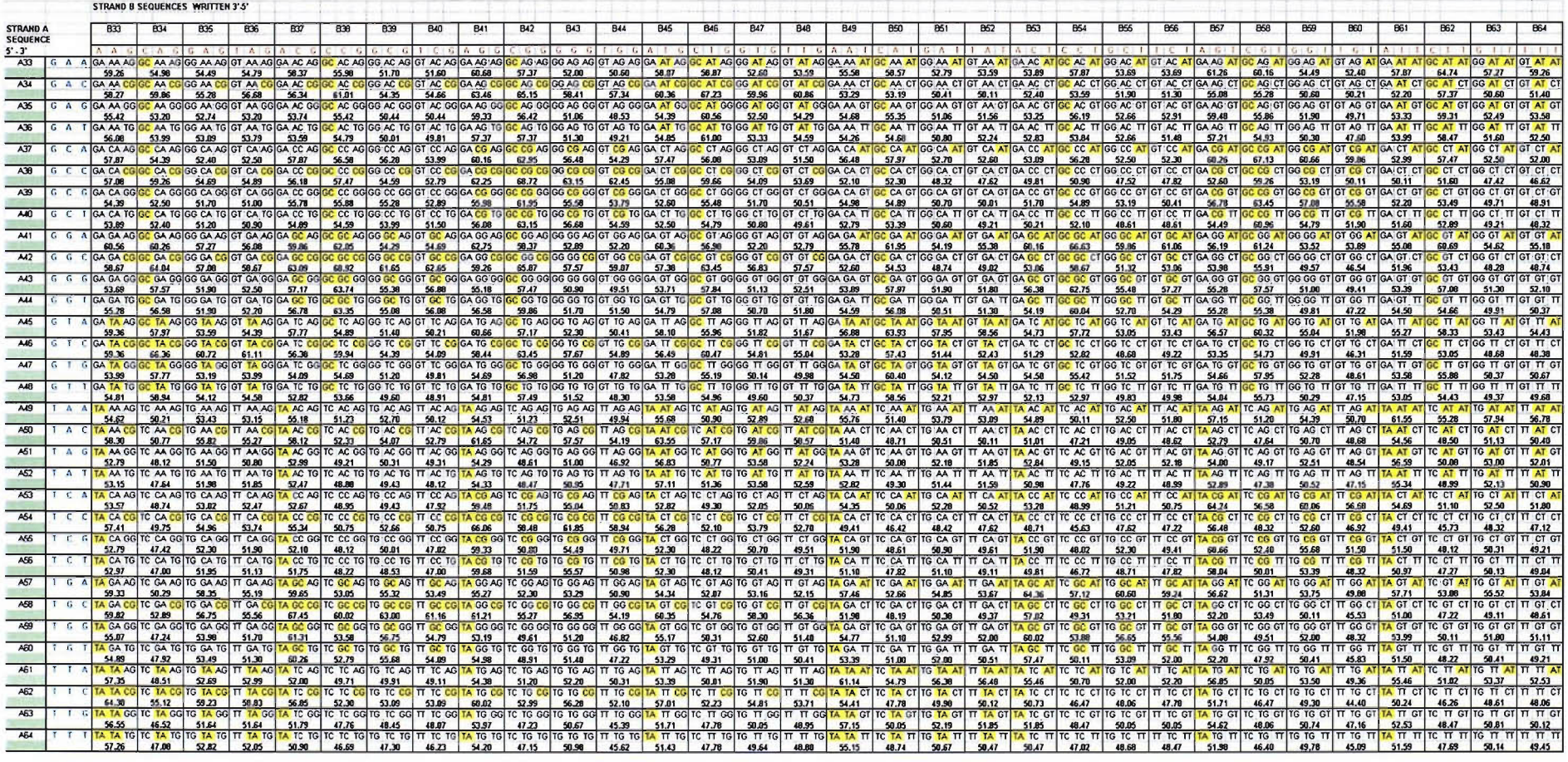
Data sequences A33-A64/B1 to B32.



Data sequences A1-A32/B33 to B64.



Data sequences A33-A64/B33 to B64.



6.3 Appendix 3 - MOSAIC computer program code

The following code is the programming syntax for the Visual Basic Programme MOSAIC.

```
'#####################################################################
'MOSAIC Version 1.0     completed 19/11/04
'Programming developed, written and compiled by Adam S. Long
'#####################################################################

MODULE 1

Option Explicit
Dim logSalt As Variant                  'Salt concentration Log(M)
Dim OligoConc As Variant                'Oligonucleotide concentration
Dim Primerconc As Single                'Calculated primer conc in
Dim Doublets As Integer                 'doublet position
Dim Sequence As String                  'Oligonucleotide Sequence
Dim checkInputText As Boolean           'Determines if input oligoconc is numeric
Dim BaseDoublet(40)                     'Sequence as doublets in an array
Dim SaltF As Single                     'Salt correction factor
Dim BaseDuplex(26), dH(26)              'Array for DeltaH
Dim BaseDuplex2(26), dS(26)             'Array for DeltaS
Dim DelH As Integer                     'Array counter for loading DelH data
Dim DelS As Integer                     'Array Counter for loading DelS data
Dim DeltaH As Single                    'Thermodynamic value for deltaH
Dim DeltaS As Single                    'Thermodynamic value for deltaS
Dim EndOfData1 As Integer               'Number of pieces of thermodynamic data for DeltaH
Dim EndofData2 As Integer               'Number of pieces of thermodynamic data for DeltaS
Dim RlnPrimerConc                       'GasConstant, and natural log (ln) Primer conc(uM)
Public StrPrinters(10)                  'Global variable for storing printers from
Public SessionName As String            'SessionName of run
Public strName As String                'Name of session
Private Const TWIP_FACTOR = 1440        'Number of twips per inch (cmdPrint_Click)
Private Const PAGE_WIDTH = 10           'Width of A4 in inches (cmdPrint_Click)
Private Const PAGE_LENGTH = 6.75        'Length of A4 in inches (cmdPrint_Click)
Public strFilePath As String            'Pathway (C:\Programe Files etc) to saved data

'The next set of variables and Function enable the transfer of the _
frmReport.Image to the PrintPicture.picture for printing.

Private Const WM_PAINT = &HF
Private Const WM_PRINT = &H317
Private Const PRF_CLIENT = &H4&
Private Const PRF_CHILDREN = &H10&
Private Const PRF_OWNED = &H20&
Private Declare Function SendMessage Lib "user32" Alias _
  "SendMessageA" (ByVal hwnd As Long, ByVal wMsg As Long, _
  ByVal wParam As Long, ByVal lParam As Long) As Long


Sub CheckOligoInput()       'Ensures Oligo concentration is an integer '
   OligoConc = FrmMOSAIC.txtOligoConc.Text
   checkInputText = IsNumeric(OligoConc)
   If checkInputText = False Then
      MsgBox "You must enter an integer for Oligo Concentration", vbOKOnly, "Warning"
      Reset1
   ElseIf checkInputText = True Then
   End If
   CalculateEquationFactors
End Sub
```

```
Sub CheckSaltInput()        'Ensures Salt concentration is an integer
Dim Saltconc
   Saltconc = FrmMOSAIC.txtSaltConc.Text
   checkInputText = IsNumeric(Saltconc)
   If checkInputText = False Then
      MsgBox "You must enter an integer for Salt Concentration", vbOKOnly, "Warning"
      Reset2
   ElseIf checkInputText = True Then
   End If
   CalculateEquationFactors
End Sub


Sub CalculateEquationFactors()            'Computes salt and conc factors for Tm equation
   RlnPrimerConc = FrmMOSAIC.txtOligoConc.Text
   logSalt = FrmMOSAIC.txtSaltConc.Text
   Sequence = FrmMOSAIC.txtSequence.Text
   If RlnPrimerConc = "" Or logSalt = "" Or Sequence = "" Then
      FrmMOSAIC.txtLength.Text = "0"
      FrmMOSAIC.txtGC.Text = "0.0"
      FrmMOSAIC.txtTm.Text = ""
      FrmMOSAIC.txtDelH.Text = ""
      FrmMOSAIC.txtDelS.Text = ""
      FrmMOSAIC.txtA.Text = "0"
      FrmMOSAIC.txtC.Text = "0"
      FrmMOSAIC.txtG.Text = "0"
      FrmMOSAIC.txtT.Text = "0"
      FrmMOSAIC.txtN.Text = "0"
      Exit Sub
   End If
   If RlnPrimerConc < "1" Or logSalt < "1" Or Sequence < "1" Then
      FrmMOSAIC.txtLength.Text = "0"
      FrmMOSAIC.txtGC.Text = "0.0"
      FrmMOSAIC.txtTm.Text = ""
      FrmMOSAIC.txtDelH.Text = ""
      FrmMOSAIC.txtDelS.Text = ""
      FrmMOSAIC.txtA.Text = "0"
      FrmMOSAIC.txtC.Text = "0"
      FrmMOSAIC.txtG.Text = "0"
      FrmMOSAIC.txtT.Text = "0"
      FrmMOSAIC.txtN.Text = "0"
      Exit Sub
   End If
   Primerconc = 1 * 10 ^ 9
   RlnPrimerConc = Log(RlnPrimerConc / Primerconc) * 1.987
   logSalt = Log(logSalt / 1000) / Log(10#)
   LoadThermodynamicsData
End Sub


Sub LoadThermodynamicsData()    'Loads thermodynamics data from DeltaH and DeltaS
   DelH = 0
   Open strFilePath & "DeltaH.asl" For Input As #1 'Opens Reference File for DeltaH
   Do Until EOF(1)
      Input #1, BaseDuplex(DelH), dH(DelH)
      DelH = DelH + 1
   Loop
   EndOfData1 = DelH
   DelS = 0
   Open strFilePath & "DeltaS.asl" For Input As #2 'Opens Reference File for DeltaS
   Do Until EOF(2)
      Input #2, BaseDuplex2(DelS), dS(DelS)
      DelS = DelS + 1
   Loop
   EndofData2 = DelS
   SequenceComposition
End Sub
```

```
Sub SequenceComposition()              'determines sequence composition
Dim Base
Dim Obase
Dim CountA As Integer
Dim CountC As Integer
Dim CountG As Integer
Dim CountT As Integer
Dim CountN As Integer
Dim Number As Integer
   Sequence = FrmMOSAIC.txtSequence.Text
   FrmMOSAIC.txtLength.Text = Len(Sequence)
   CountA = 0
   CountC = 0
   CountG = 0
   CountT = 0
   CountN = 0
   Number = FrmMOSAIC.txtLength.Text
   For Base = 1 To Number              'Counter for number of bases
      Obase = Mid(Sequence, Base, 1)
         If Obase = "A" Then
            CountA = CountA + 1
         ElseIf Obase = "C" Then
            CountC = CountC + 1
         ElseIf Obase = "G" Then
            CountG = CountG + 1
         ElseIf Obase = "T" Then
            CountT = CountT + 1
         ElseIf Obase = "N" Then
            CountN = CountN + 1
         End If
   Next
   FrmMOSAIC.txtA.Text = CountA       'Number of A bases
   FrmMOSAIC.txtC.Text = CountC       'Number of C bases
   FrmMOSAIC.txtG.Text = CountG       'Number of G bases
   FrmMOSAIC.txtT.Text = CountT       'Number of T bases
   FrmMOSAIC.txtN.Text = CountN       'Number of N bases
   If CountC > "0" Or CountG > "0" Then
      FrmMOSAIC.txtGC.Text = ((CountC + CountG) / Number) * 100 ' %GC Content
      FrmMOSAIC.txtGC.Text = Format(FrmMOSAIC.txtGC.Text, "0.0")
   Else: End If
   AsymmetryThermodynamics
End Sub


Sub AsymmetryThermodynamics()  'Loads additional thermodynamic factors according to sequence
Dim Oligo
   DeltaH = 0
   DeltaS = 0
   Oligo = FrmMOSAIC.txtSequence.Text
   Oligo = Left(Oligo, 1)                 ' 5'region
   If Oligo = "C" Or Oligo = "G" Then
      DeltaH = 0.1
      DeltaS = DeltaS + -2.8
   Else: DeltaH = 2.3
      DeltaS = DeltaS + 4.1
   End If
   Oligo = FrmMOSAIC.txtSequence.Text
   Oligo = Right(Oligo, 1)                ' 3'region
   If Oligo = "C" Or Oligo = "G" Then
      DeltaH = DeltaH + 0.1
      DeltaS = DeltaS + -2.8
   Else: DeltaH = DeltaH + 2.3
      DeltaS = DeltaS + 4.1
   End If
   LoadDoublets
End Sub
```

```
Sub LoadDoublets()              'determines each nearest neighbour in an array format
   Sequence = FrmMOSAIC.txtSequence.Text
   For Doublets = 1 To 39
      BaseDoublet(Doublets) = Mid(Sequence, Doublets, 2)
   Next
   CompareDoublets
End Sub


Sub CompareDoublets()  'compares each doublet in the array with thermodynamic arrays
Dim Base
   For Doublets = 1 To 39
      Base = BaseDoublet(Doublets)
      For DelH = 0 To EndOfData1
         If Base = BaseDuplex(DelH) Then
            DeltaH = DeltaH + dH(DelH)
         End If
      Next DelH
   Next Doublets
   FrmMOSAIC.txtDelH.Text = DeltaH * 1000
   Close #1
   For Doublets = 1 To 39
      Base = BaseDoublet(Doublets)
      For DelS = 0 To EndofData2
         If Base = BaseDuplex2(DelS) Then
            DeltaS = DeltaS + dS(DelS)
         End If
      Next DelS
   Next Doublets
   FrmMOSAIC.txtDelS.Text = DeltaS
   Close #2
   FinalTm
End Sub


Sub FinalTm()        'computes the final Tm using the regression equation for Salt
Dim Lenbp
Dim CalcTm
   Lenbp = FrmMOSAIC.txtSequence.Text
   Lenbp = Len(Lenbp)
   SaltF = ((0.0035 * Lenbp ^ 3) + (-0.2157 * Lenbp ^ 2) + (4.492 * Lenbp) + (-23.153))
                                                  ' based on salt dependance versus length formulae
   CalcTm = ((DeltaH * 1000) / (DeltaS + RlnPrimerConc) - 273.15) + (logSalt * SaltF)
   CalcTm = Format(CalcTm, "0.00" & Chr(176) & "C")
   FrmMOSAIC.txtTm.Text = CalcTm
End Sub


Sub Reset1()         'Resets the oligo concentration to ""
   FrmMOSAIC.txtOligoConc.Text = ""
End Sub


Sub Reset2()         'Resets the Salt concentration to ""
   FrmMOSAIC.txtSaltConc.Text = ""
End Sub


Sub PrintOptions()              'Prints completed form to printer selected by user in frmSetPrinter module
Dim ErrorMsg As String
Dim Message As Long
   Unload FRMSetPrinter
   Printer.PaperSize = vbPRPSA4          'Ensures printer is setup for A4 paper
   Printer.Orientation = 2               'Ensures printer is setup for portrait orientation
   frmReport.Height = PAGE_LENGTH * TWIP_FACTOR
   frmReport.Width = PAGE_WIDTH * TWIP_FACTOR
   FrmMOSAIC.PrintPicture.Height = PAGE_LENGTH * TWIP_FACTOR
   FrmMOSAIC.PrintPicture.Width = PAGE_WIDTH * TWIP_FACTOR
                          'Will resize form and picture box, from which form is printed, to A4 sized.
```

```
    frmReport.SetFocus
    FrmMOSAIC.PrintPicture.AutoRedraw = True
    Message = SendMessage(frmReport.hwnd, WM_PAINT, FrmMOSAIC.PrintPicture.hDC, 0)
    Message = SendMessage(frmReport.hwnd, WM_PRINT, FrmMOSAIC.PrintPicture.hDC, _
     PRF_CHILDREN + PRF_CLIENT + PRF_OWNED)
    FrmMOSAIC.PrintPicture.Picture = FrmMOSAIC.PrintPicture.Image
    FrmMOSAIC.PrintPicture.AutoRedraw = False
    'Paints picture of FrmMain and adds it to frmMOSAIC. PrintPicture.picture _
    for printing, because forms larger than the screen will only print area _
    visible on screen.
    Printer.Print ""
    Printer.PaintPicture FrmMOSAIC.PrintPicture, 1000, 1000    'position picture page centered
    Printer.EndDoc
    'frmmosaic.PrintPicture is printed
    'Once form has been printed the program will close.
    GoTo Finish
ErrorHandler:
    'If there is a printer error the form will be minimised for printing _
    at a later time. This is only needed if there is no printer or there _
    is a printer error.
    ErrorMsg = "The form cannot be printed at the moment"
    MsgBox ErrorMsg
    frmReport.WindowState = vbMinimized
    'CMDPrint.Visible = True
Finish:
End Sub

Sub CompileReport(NoofSeq)              'Compiles a report from previous Sessions
Dim SequenceReport(12)
Dim files
Dim Counter
Dim myDate
  frmLoadFile.Hide
      For files = 1 To NoofSeq
        Open SessionName & "\" & files & ".mos" For Input As #2
           Counter = 0
              Input #2, SequenceReport(Counter)
              frmReport.txtReportcomp(files).Text = SequenceReport(Counter)
              Counter = Counter + 1
              Input #2, SequenceReport(Counter)
              frmReport.txtOligoConc(files).Text = SequenceReport(Counter)
              Counter = Counter + 1
              Input #2, SequenceReport(Counter)
              frmReport.txtSalt(files).Text = SequenceReport(Counter)
              Counter = Counter + 1
              Input #2, SequenceReport(Counter)
              frmReport.txtTm(files).Text = SequenceReport(Counter)
              frmReport.txtTm(files) = Format(frmReport.txtTm(files), "0.00")
              Counter = Counter + 1
              Input #2, SequenceReport(Counter)
              frmReport.txtSeq(files).Text = SequenceReport(Counter)
           Close #2
        Next
    myDate = Format(Date, "dd mmm yyyy")
    frmReport.lblDate = myDate
    frmReport.lblName = strName
    frmReport.lblNoSequences = NoofSeq
    frmReport.cmdReturn.Visible = False
    frmReport.cmdSave.Visible = False
    frmReport.cmdLoad.Visible = True
    frmReport.cmdClose.Visible = True
    frmReport.Show
End Sub
`#########################################################################################
```

```
'##########################################################################
FRMMOSAIC

Private Sub cmdClearAll_Click()          'Resets all input files on frmMOSAIC
   FrmMOSAIC.txtSeqName.Text = ""
   FrmMOSAIC.txtOligoConc.Text = ""
   FrmMOSAIC.txtSaltConc.Text = ""
   FrmMOSAIC.txtSequence.Text = ""
   FrmMOSAIC.txtSeqName.SetFocus
End Sub

Private Sub cmdNewSession_Click()        'Resets all input files on frmMOSAIC + Session
   Query = MsgBox("Are you sure you want to continue with a New Session?", vbYesNo, "Warning!")
      If Query = vbNo Then
         Exit Sub
      ElseIf Query = vbYes Then
         FrmMOSAIC.txtNumber = "1"
         FrmMOSAIC.Text1.Enabled = True
         FrmMOSAIC.Text1.TabStop = True
         FrmMOSAIC.Text1.Text = ""
         FrmMOSAIC.txtSeqName.Enabled = True
         FrmMOSAIC.txtSeqName.Text = ""
         FrmMOSAIC.txtOligoConc.Enabled = True
         FrmMOSAIC.txtOligoConc.Text = ""
         FrmMOSAIC.txtSaltConc.Enabled = True
         FrmMOSAIC.txtSaltConc.Text = ""
         FrmMOSAIC.txtSequence.Enabled = True
         FrmMOSAIC.txtSequence.Text = ""
         FrmMOSAIC.cmdClearAll.Enabled = True
         FrmMOSAIC.cmdResetSeq.Enabled = True
         FrmMOSAIC.cmdReport.Enabled = False
         FrmMOSAIC.cmdNewSession.TabStop = False
         FrmMOSAIC.Text1.SetFocus
      End If
End Sub

Private Sub cmdReport_Click()            'Compiles report during a session
Dim SequenceReport(12)
   NoofSeq = FrmMOSAIC.txtNumber.Text - 1
   If NoofSeq >= 12 Then
      frmReport.cmdReturn.Enabled = False
   ElseIf NoofSeq < 12 Then
      frmReport.cmdReturn.Enabled = True
   End If
   For files = 1 To NoofSeq
      Open strFilePath & "DATA\" & SessionName & "\" & files & ".mos" For Input As #2
         a = 0
            Input #2, SequenceReport(a)
            frmReport.txtReportcomp(files).Text = SequenceReport(a)
            a = a + 1
            Input #2, SequenceReport(a)
            frmReport.txtOligoConc(files).Text = SequenceReport(a)
            a = a + 1
            Input #2, SequenceReport(a)
            frmReport.txtSalt(files).Text = SequenceReport(a)
            a = a + 1
            Input #2, SequenceReport(a)
            frmReport.txtTm(files).Text = SequenceReport(a)
            frmReport.txtTm(files) = Format(frmReport.txtTm(files), "0.00")
            a = a + 1
            Input #2, SequenceReport(a)
            frmReport.txtSeq(files).Text = SequenceReport(a)
         Close #2
      Next
   myDate = Format(Date, "dd mmm yyyy")
   frmReport.lblDate = myDate
   frmReport.lblName = SessionName
   frmReport.lblNoSequences = FrmMOSAIC.txtNumber.Text - 1
   frmReport.cmdClose.Visible = False
   frmReport.cmdLoad.Visible = False
   frmReport.Show
End Sub
```

```
Private Sub cmdResetSeq_Click()          'Clears sequence and sequence name boxes
    FrmMOSAIC.txtSeqName.Text = ""
    FrmMOSAIC.txtSequence.Text = ""
    FrmMOSAIC.txtSeqName.SetFocus
End Sub


Private Sub cmdSaveSeq_Click()           'Saves sequence data as a .MOS file in "DATA" file
    FrmMOSAIC.cmdReport.Enabled = True
    FrmMOSAIC.Text1.TabStop = False
    NoofSeq = FrmMOSAIC.txtNumber.Text
    FileName = NoofSeq
    SessionName = FrmMOSAIC.Text1
       If SessionName = "" Then
          Response3 = InputBox("Please enter a session name!", "Session name", "Session name")
          SessionName = Response3
          FrmMOSAIC.Text1.Text = SessionName
       Else:
       End If
    On Error Resume Next
    MkDir strFilePath & "DATA\" & SessionName
    Open strFilePath & "DATA\" & SessionName & "\" & NoofSeq & ".mos" For Output As #1
    Print #1, txtSeqName.Text
    Print #1, txtOligoConc.Text
    Print #1, txtSaltConc.Text
    Print #1, txtTm.Text
    Print #1, txtSequence.Text
    Print #1, "End"
    Close #1
    NoofSeq = NoofSeq + 1
    If NoofSeq > 12 Then
       Rep = MsgBox("You have now added the maximum number of oligonucleotides in this session." & (Chr(13)) & _
       "Compile the Report and then save before continuing with a New Session", vbCritical, "WARNING!")
       FrmMOSAIC.cmdClearAll.Enabled = False
       FrmMOSAIC.cmdSaveSeq.Enabled = False
       FrmMOSAIC.cmdResetSeq.Enabled = False
       FrmMOSAIC.cmdNewSession.Enabled = False
       FrmMOSAIC.Text1.Enabled = False
       FrmMOSAIC.txtSeqName.Enabled = False
       FrmMOSAIC.txtOligoConc.Enabled = False
       FrmMOSAIC.txtSaltConc.Enabled = False
       FrmMOSAIC.txtSequence.Enabled = False
       FrmMOSAIC.txtNumber.Text = NoofSeq
       Exit Sub
    Else: End If
    FrmMOSAIC.txtNumber.Text = NoofSeq
    FrmMOSAIC.txtSeqName.SetFocus
End Sub


Private Sub File_Click(Index As Integer)          'File menues
    If Index = 0 Then                             'to loads existing report
       Load frmLoadFile
       frmLoadFile.Show vbModal
    ElseIf Index = 1 Then                             'to quit programme
       Response = MsgBox("Are you sure you want to quit?", vbYesNo, "Warning!")
          If Response = vbYes Then
          End
          Else
          End If
    End If
End Sub


Private Sub Form_Load()                  'Sets the strFilePath as the beginning of session
    strFilePath = "C:\Program Files\MOSAIC\"
    FrmMOSAIC.cmdSaveSeq.Enabled = False
    FrmMOSAIC.cmdReport.Enabled = False
End Sub
```

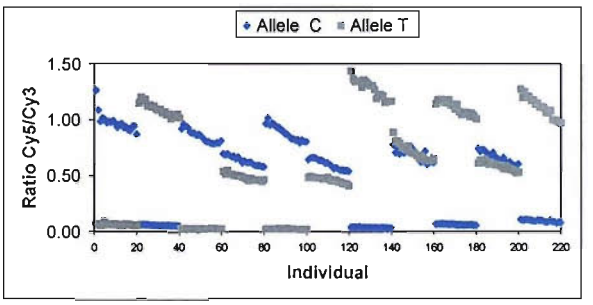
```
Private Sub Form_Unload(Cancel As Integer)          'X button error escape on control menu
   Response = MsgBox("Are you sure you want to quit?", vbYesNo, "Warning!")
      If Response = vbYes Then
      End
      ElseIf Response = vbNo Then
      Cancel = 1
      End If
End Sub

Private Sub txtOligoConc_Change()          'error trap to ensure all variables are loaded prior to Tm
   If FrmMOSAIC.txtOligoConc.Text = "" Then
      FrmMOSAIC.txtDelH.Text = ""
      FrmMOSAIC.txtDelS.Text = ""
      FrmMOSAIC.txtTm.Text = ""
      Exit Sub
   End If
   CheckOligoInput
End Sub

Private Sub txtSeqName_GotFocus()      'highlights existing text in SeqName as a group
   txtSeqName.SelStart = 0
   txtSeqName.SelLength = (Len(txtSeqName.Text))
End Sub

Private Sub txtOligoConc_GotFocus()      'highlights existing text in OligoConc as a group
   txtOligoConc.SelStart = 0
   txtOligoConc.SelLength = (Len(txtOligoConc.Text))
End Sub


Private Sub txtSaltConc_GotFocus()       'highlights existing text in SaltConc as a group
   txtSaltConc.SelStart = 0
   txtSaltConc.SelLength = (Len(txtSaltConc.Text))
End Sub

Private Sub txtSequence_GotFocus()       'highlights existing text in Sequence as a group
   txtSequence.SelStart = 0
   txtSequence.SelLength = (Len(txtSequence.Text))
End Sub

Private Sub txtSaltConc_Change()          'error trap to ensure all variables are loaded prior to Tm
   If FrmMOSAIC.txtSaltConc.Text = "" Then
      FrmMOSAIC.txtDelH.Text = ""
      FrmMOSAIC.txtDelS.Text = ""
      FrmMOSAIC.txtTm.Text = ""
      Exit Sub
   End If
   CheckSaltInput
End Sub

Private Sub txtSequence_Change()          'Auto update calculation of Tm
      CalculateEquationFactors
End Sub

Private Sub txtSequence_KeyPress(KeyAscii As Integer)          'Key stroke determination
   If KeyAscii = 22 Then
      Sequence = Clipboard.GetText
      Sequence = UCase(Sequence)
      Lenbp = Sequence
      Lenbp = Len(Lenbp)
      For X = 1 To Lenbp
          Y = Mid(Sequence, X, 1)
          If Y = "C" Or Y = "G" Or Y = "T" Or Y = "A" Or Y = "N" Or Y = "c" Or Y = "g" Or Y = "t" Or _
          Y = "a" Or Y = "n" Then
          Else: MsgBox "Invalid base at position " & X
                Sequence = ""
          Exit For
          End If
      Next
      FrmMOSAIC.txtSequence.Text = Sequence
   End If
   Char = Chr(KeyAscii)
```

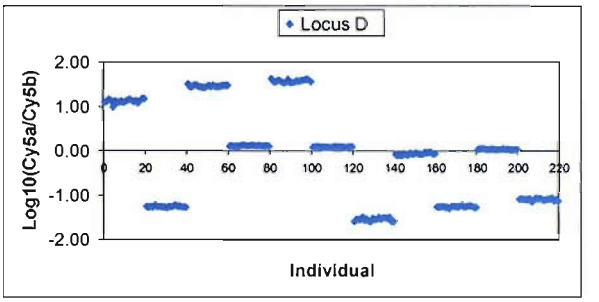
```
    If KeyAscii = 65 Or KeyAscii = 67 Or KeyAscii = 71 Or KeyAscii = 78 Or KeyAscii = 84 Or KeyAscii = 97 Or _
       KeyAscii = 99 Or KeyAscii = 103 Or KeyAscii = 110 Or KeyAscii = 116 Or KeyAscii = 8 Then
       KeyAscii = Asc(UCase(Char))
    Else: KeyAscii = 0
       Exit Sub
    End If
End Sub

Private Sub txtTm_Change()              'Enables the save button
    If FrmMOSAIC.txtTm.Text = "" Then
       FrmMOSAIC.cmdSaveSeq.Enabled = False
    ElseIf FrmMOSAIC.txtTm.Text <> "" Then
       FrmMOSAIC.cmdSaveSeq.Enabled = True
    End If
End Sub
'################################################################################

'################################################################################

FRMLOADFILE

Private Sub cmdLoad_Click()             'Activates compile report
    SessionName = Dir1.Path
    strName = (Dir(Dir1.Path, vbDirectory))
    NoofSeq = File1.ListCount
    Call CompileReport(NoofSeq)
    Unload frmLoadFile
End Sub

Private Sub cmdCancel_Click()           'Cancels request to load form
    Unload frmLoadFile
End Sub

Private Sub Dir1_Change()               'Update path and files viewed when directory change
    File1.Path = Dir1.Path
    If File1.ListCount > 0 Then
       frmLoadFile.cmdLoad.Enabled = True        'enables Load botton
    Else
       frmLoadFile.cmdLoad.Enabled = False       'disables load button
    End If
End Sub


Private Sub Drive1_Change()             'Update path when different drive selected
    On Error Resume Next
       Dir1.Path = Drive1.Drive          'defalut set to C:\ drive
    If Err.Numer = 68 Then
       MsgBox "Please place disk in Drive " & Drive1.Drive
       Drive1.Drive = "C:"
       Dir1.Path = strFilePath & "DATA"
    Else
    End If
    On Error GoTo 0
End Sub


Private Sub Form_Load()                 'Loads file path to saved data upon opening form
       Dir1.Path = strFilePath & "DATA"
       frmLoadFile.cmdLoad.Enabled = False
End Sub
'################################################################################

'################################################################################

FRMREPORT

Private Sub cmdClose_Click()            'Unloads Report form
    Unload frmReport
End Sub
```

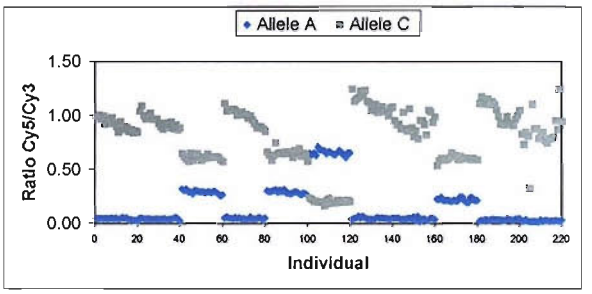
```
Private Sub cmdLoad_Click()                'Loads frmloadFile form
   Unload frmReport
   Load frmLoadFile
   frmLoadFile.Show vbModal                'must have a user response
End Sub

Private Sub cmdPrint_Click()               'Loads frmSetPrinter from
   Load FRMSetPrinter
   FRMSetPrinter.Show vbModal  'must have a user response
End Sub


Private Sub cmdReturn_Click()              'Closes Report form and returns to frmMOSAIC
   frmReport.Hide
   FrmMOSAIC.cmdNewSession.Enabled = True
   FrmMOSAIC.txtSeqName.SetFocus
End Sub

Private Sub cmdSave_Click()                'Closes Report form and returns to frmMOSAIC
   Unload frmReport
   NoofSeq = FrmMOSAIC.txtNumber.Text
   FrmMOSAIC.cmdNewSession.Enabled = True
   FrmMOSAIC.cmdNewSession.TabStop = True
   If NoofSeq >= 12 Then
      FrmMOSAIC.Text1.Enabled = False
      FrmMOSAIC.txtOligoConc.Enabled = False
      FrmMOSAIC.txtSaltConc.Enabled = False
      FrmMOSAIC.txtSeqName.Enabled = False
      FrmMOSAIC.txtSequence.Enabled = False
   ElseIf NoofSeq < 12 Then
      FrmMOSAIC.txtSeqName.SetFocus
   End If
End Sub
'##########################################################################

'##########################################################################

FRMSETPRINTER

Private Sub cmdPrint_Click()               'Launches Printer mode
   Call PrintOptions
   Unload FRMSetPrinter
End Sub


Private Sub cmdQuit_Click()                'Cancels printer request
   Unload FRMSetPrinter
End Sub


Private Sub Form_Load()                    'Displays printer select form
   FRMSetPrinter.CMDPrint.Enabled = False
   FRMSetPrinter.LBLPrinterSelected.Caption = "   Select Printer from list below"
   Y = 0
   For Each Prn In Printers
      FRMSetPrinter.PrinterList.AddItem (Prn.DeviceName)
      StrPrinters(Y) = Prn.DeviceName
      Y = Y + 1
   Next
End Sub


Private Sub PrinterList_DblClick()         'determines which primer to use
   Item = FRMSetPrinter.PrinterList.ListIndex
   FRMSetPrinter.LBLPrinterSelected.Caption = StrPrinters(Item)
   Set Printer = Printers(Item)            'sets the printer according to users selection
   FRMSetPrinter.CMDPrint.Enabled = True
End Sub
'##########################################################################
```

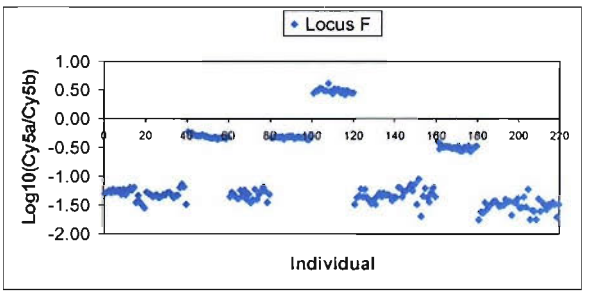
6.4 Appendix 4 - Comparative Data Analyses Figures

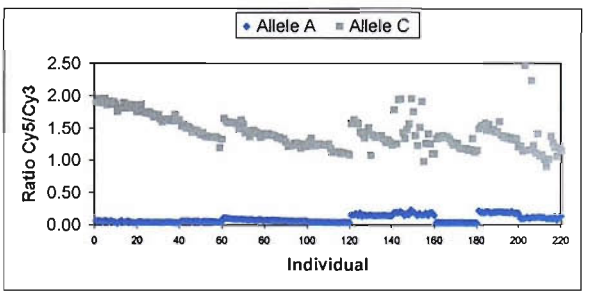
Figure 6.1 - Indirect approach, pentaplex A-v1.

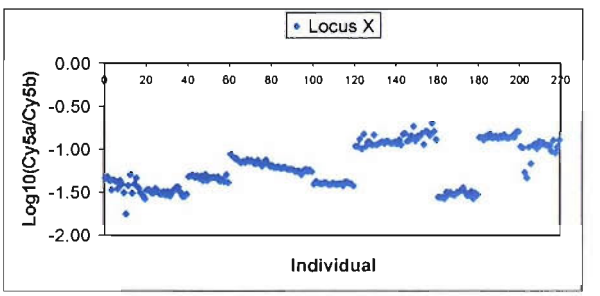


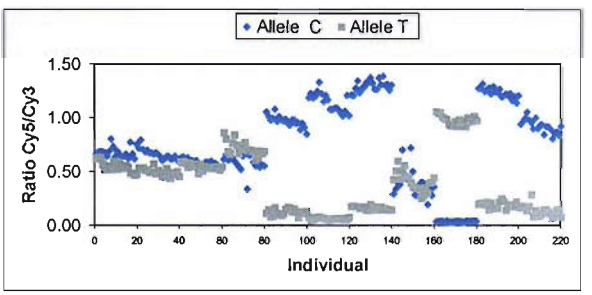


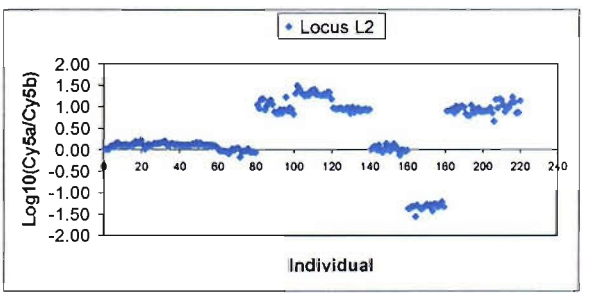


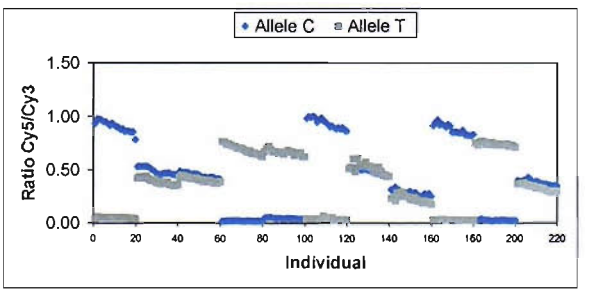












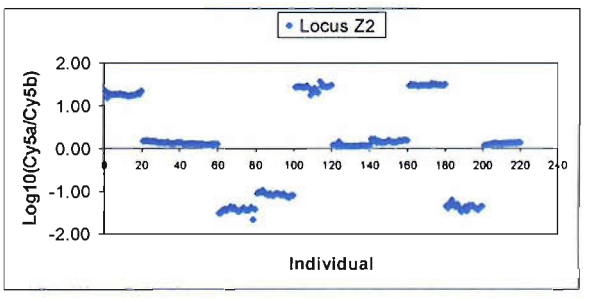
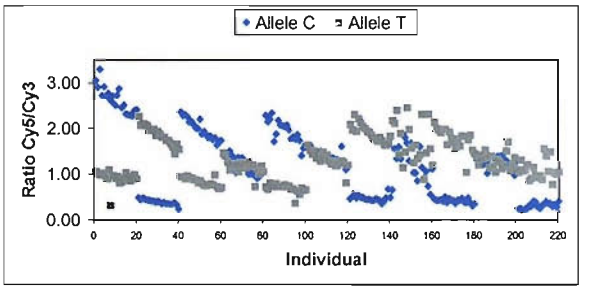
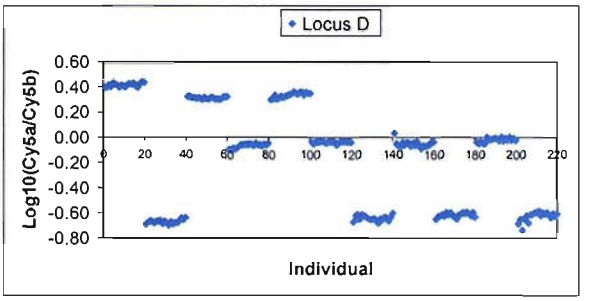
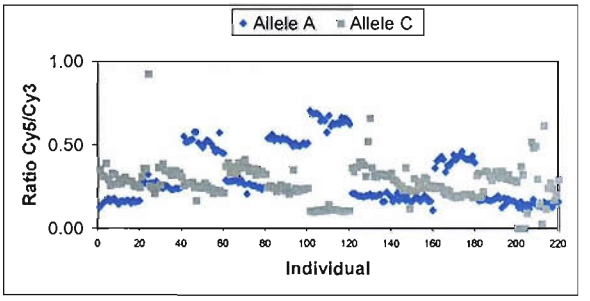
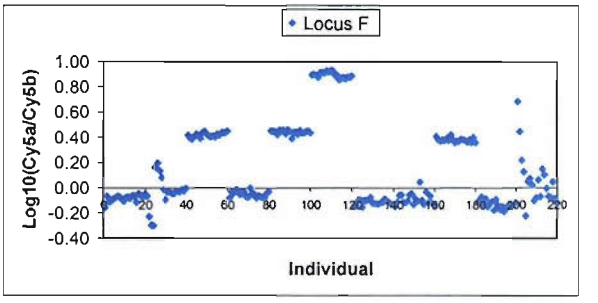


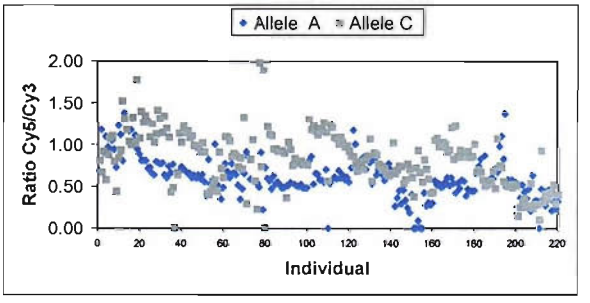
Figure 6.2- Direct approach, pentaplex A-v1.

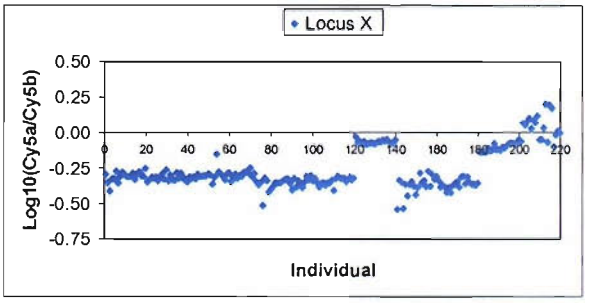


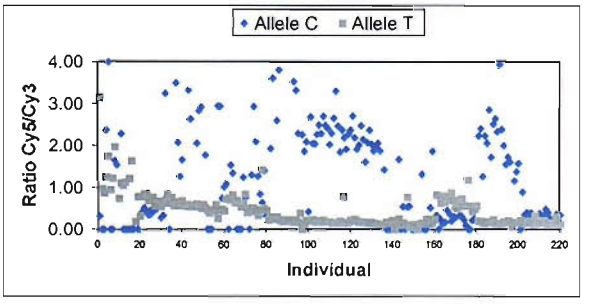


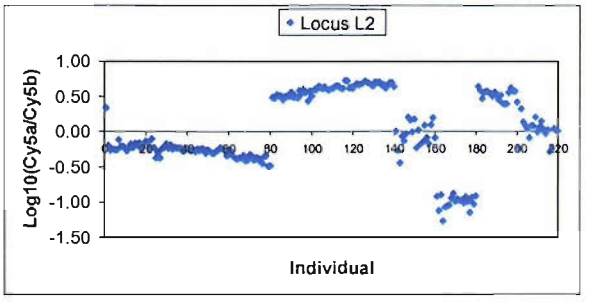


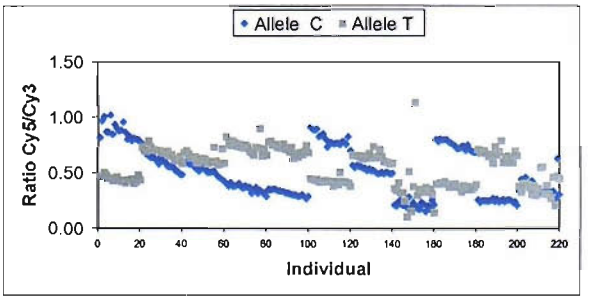












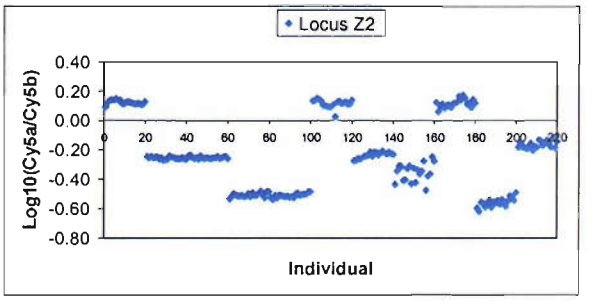
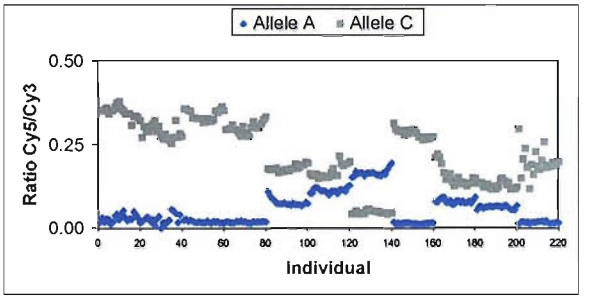
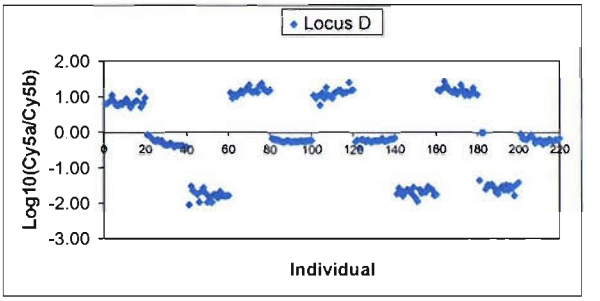
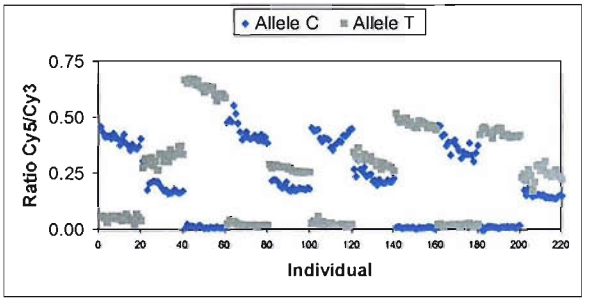
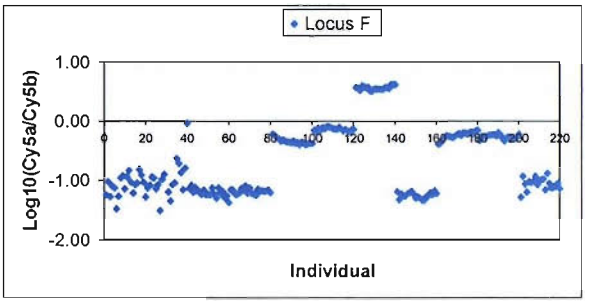


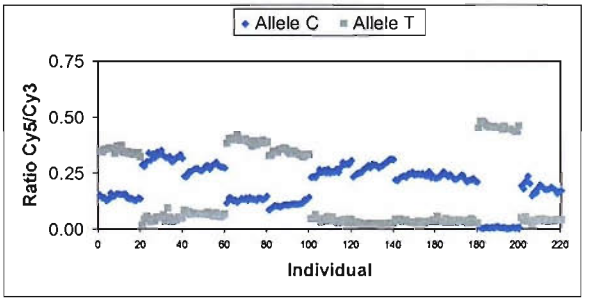
Figure 6.3 - Indirect approach, pentaplex A-v2.

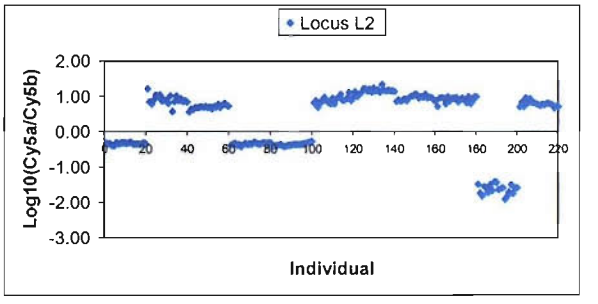


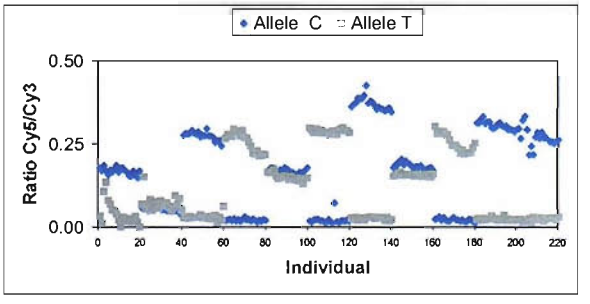


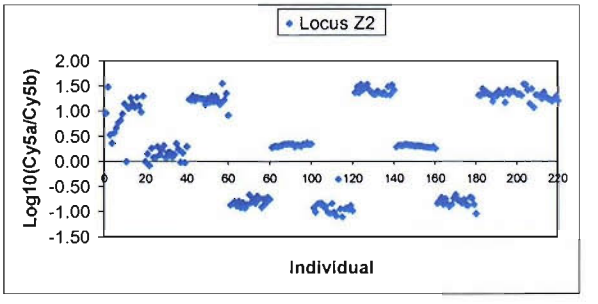


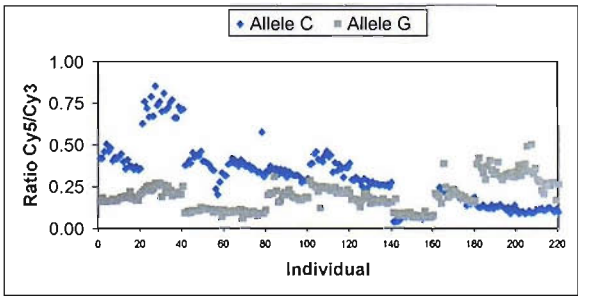












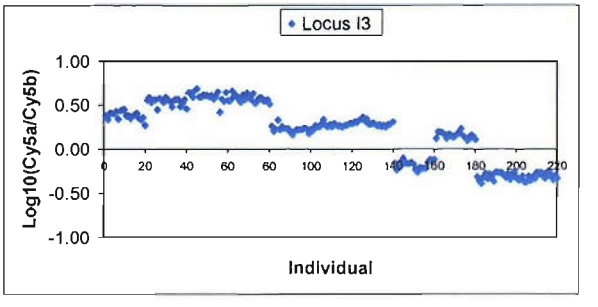
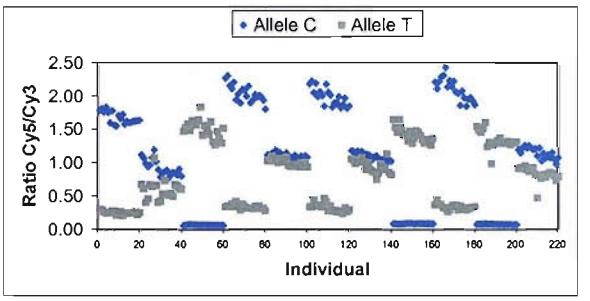
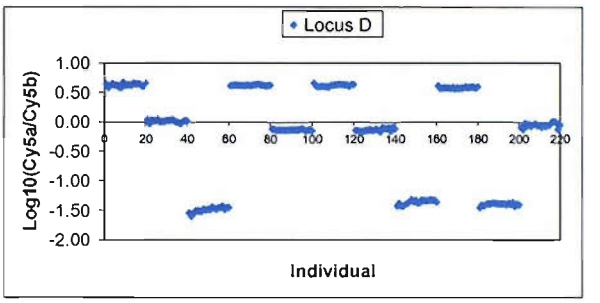
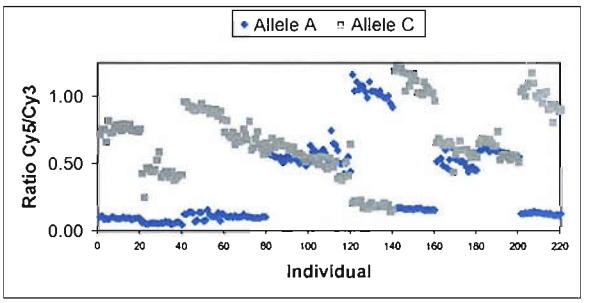
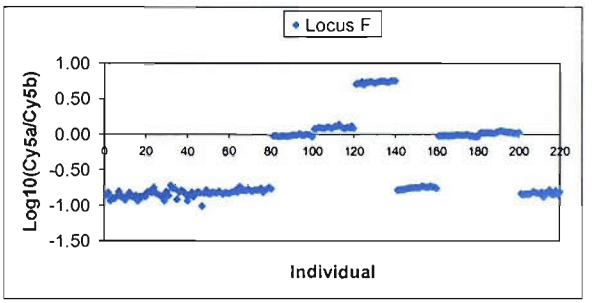


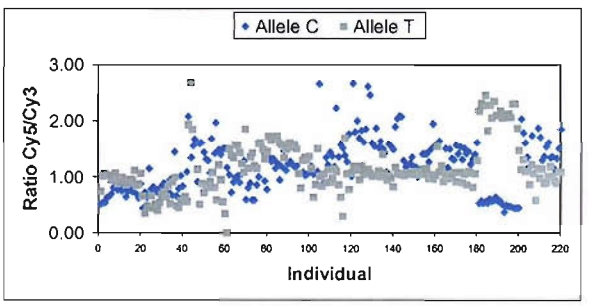
Figure 6.4 - Direct approach, pentaplex A-v2.

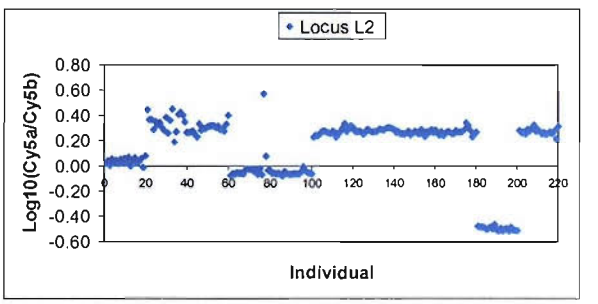


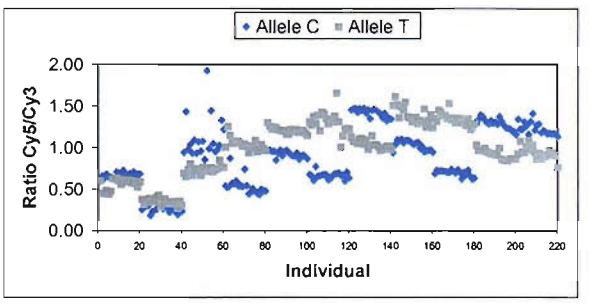


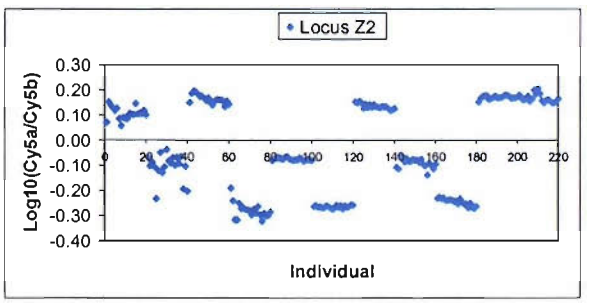


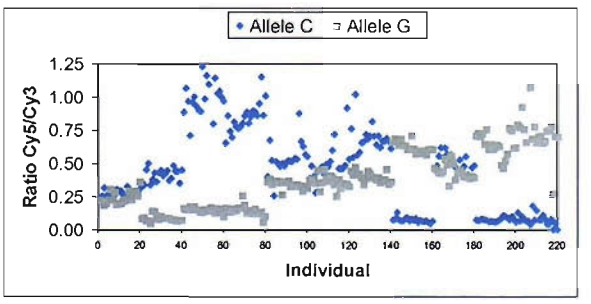












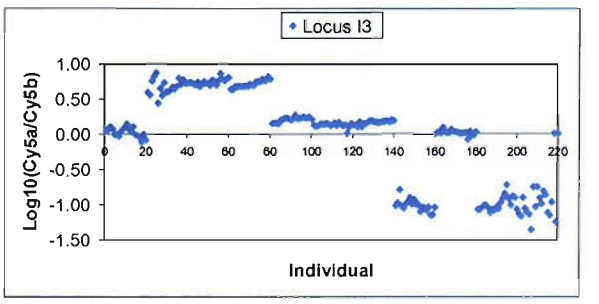
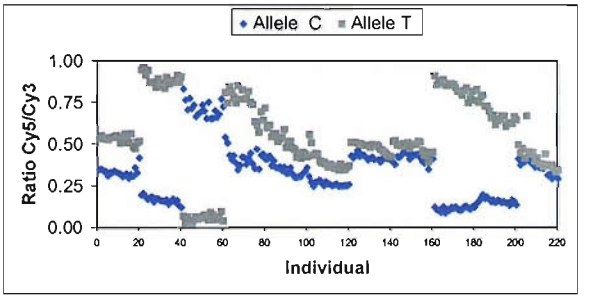
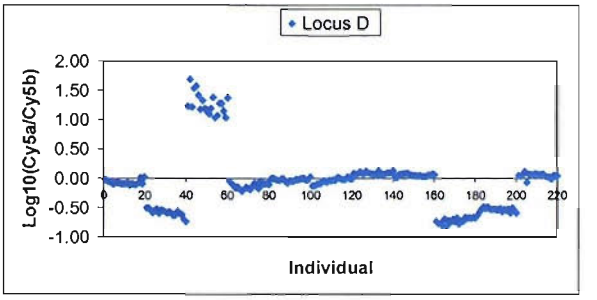
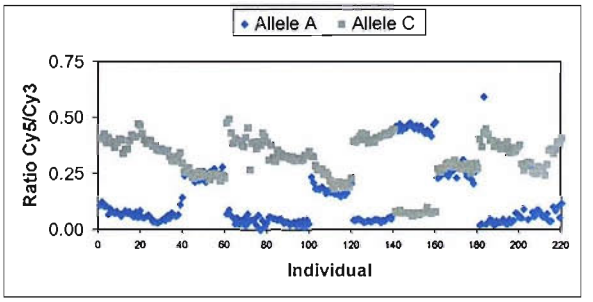
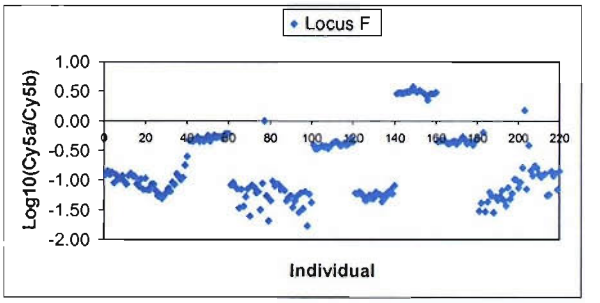


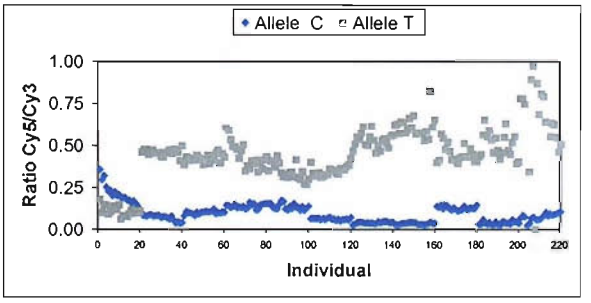
Figure 6.5 - Indirect approach, pentaplex A-v3. (5-plex)

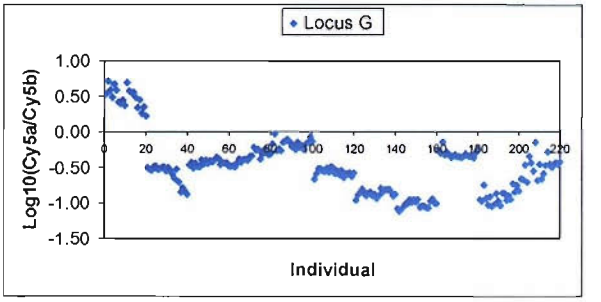


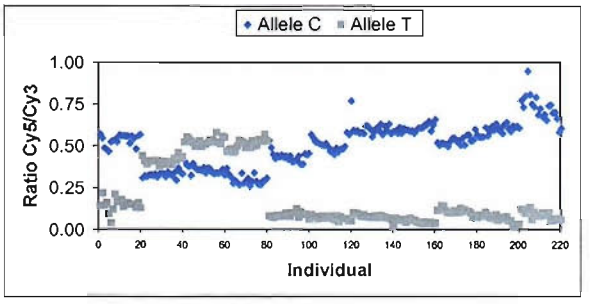


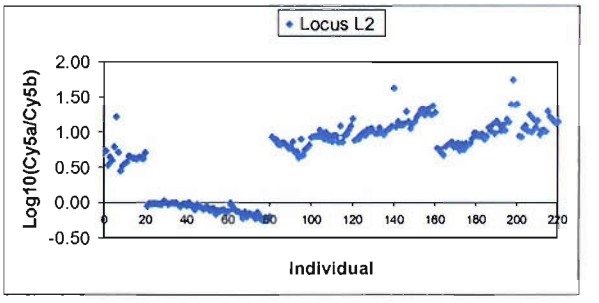


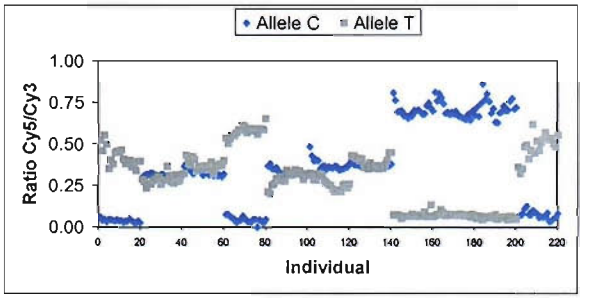












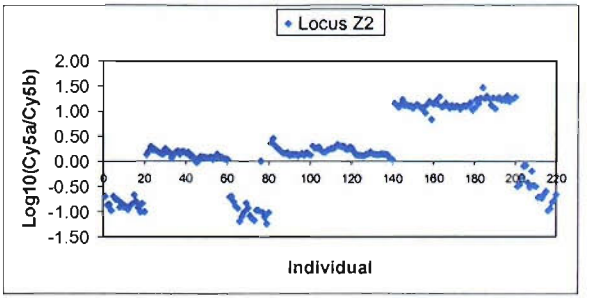
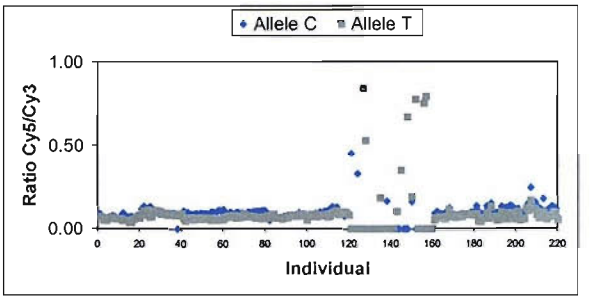
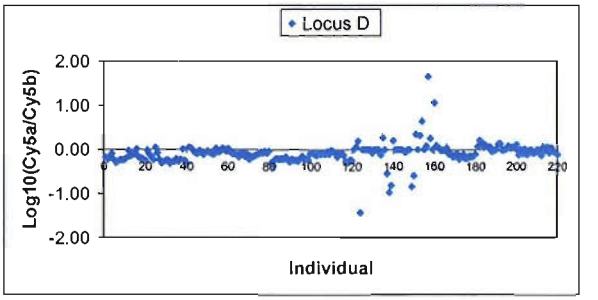
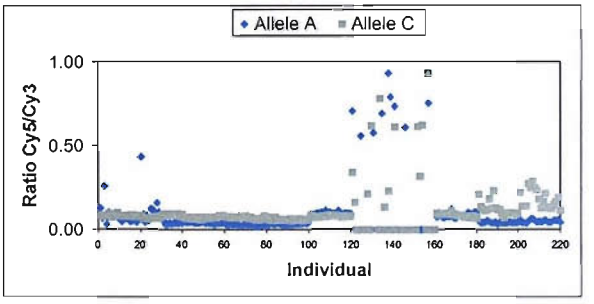
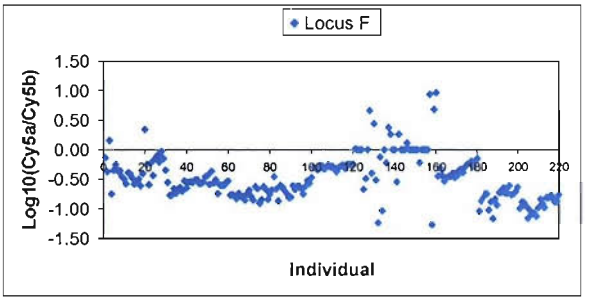


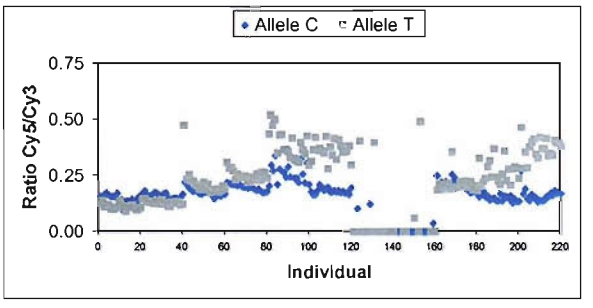
Figure 6.6 - Indirect approach, pentaplex A-v3, (17-plex).

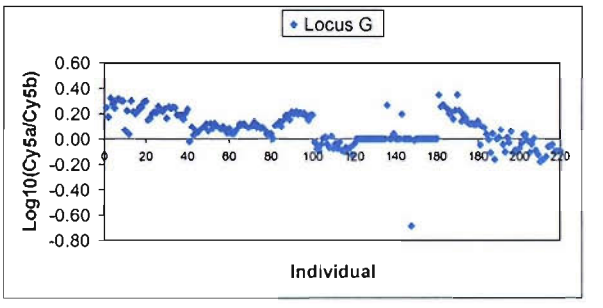


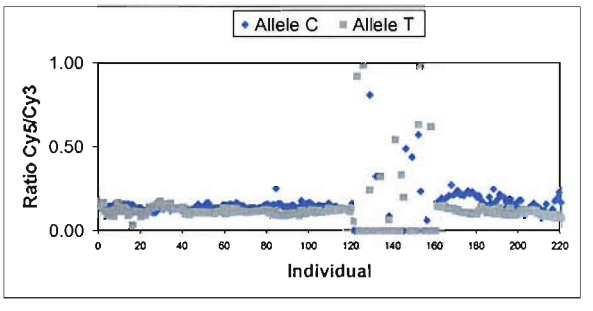


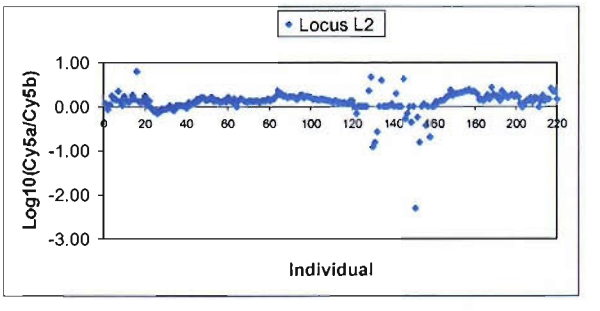


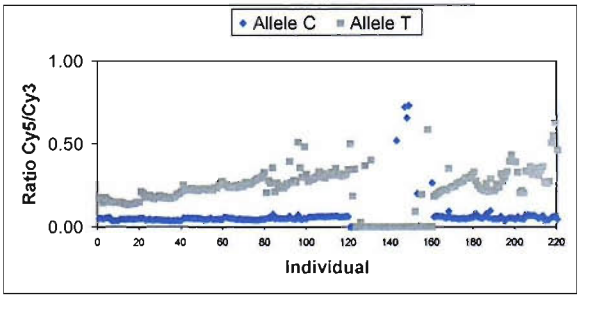












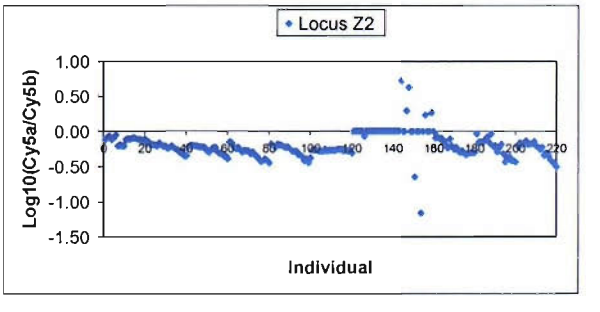
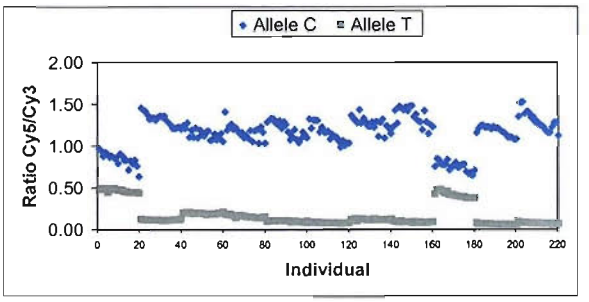
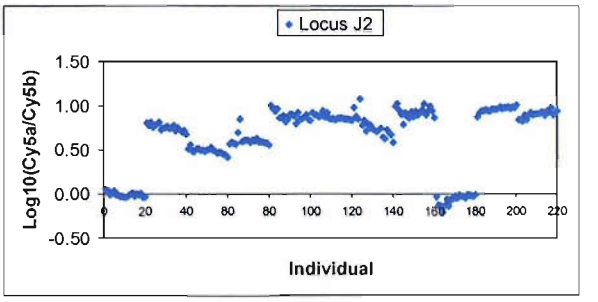
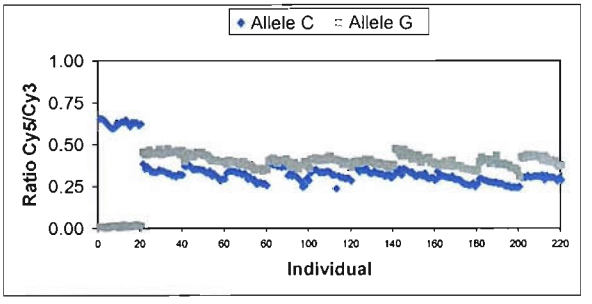
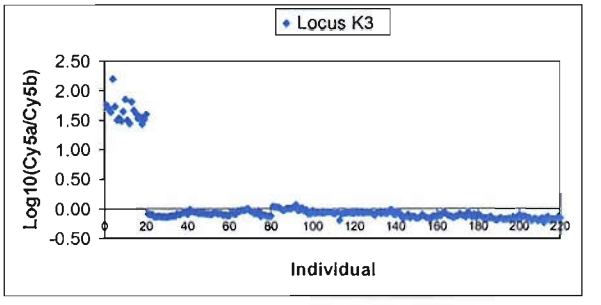


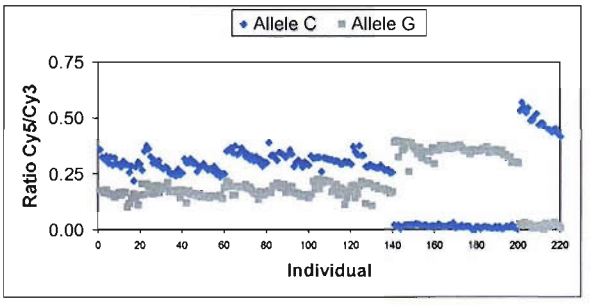
Figure 6.7 - Indirect approach, pentaplex B-v1.

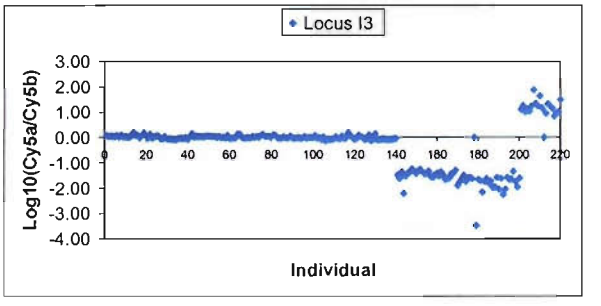


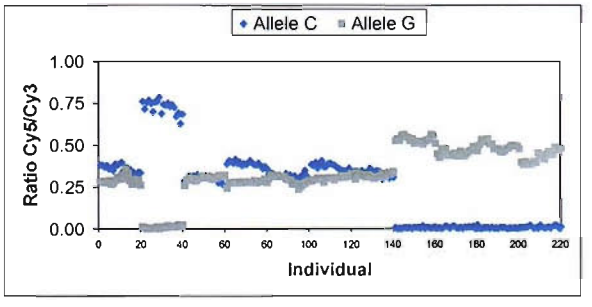


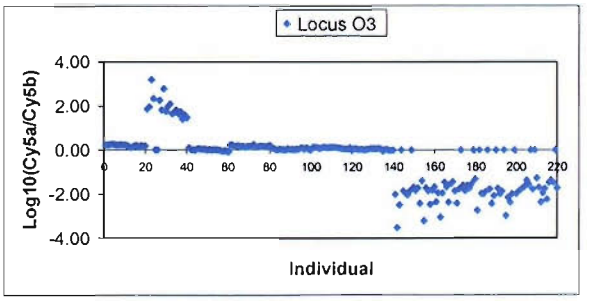


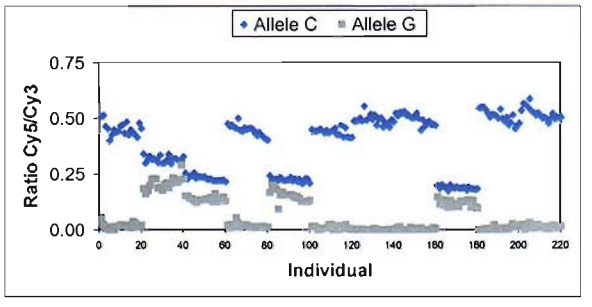












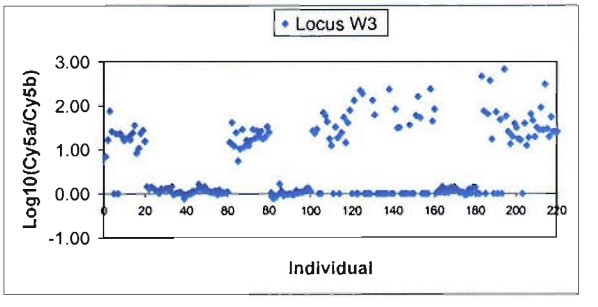
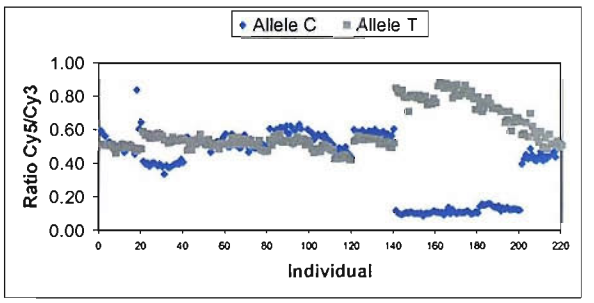
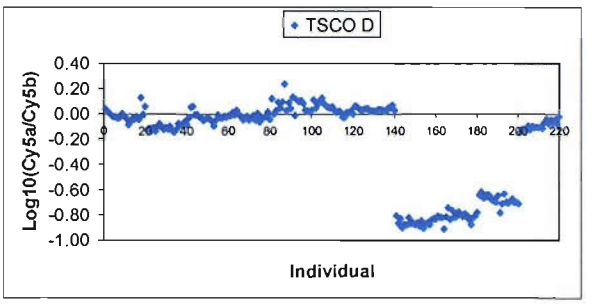
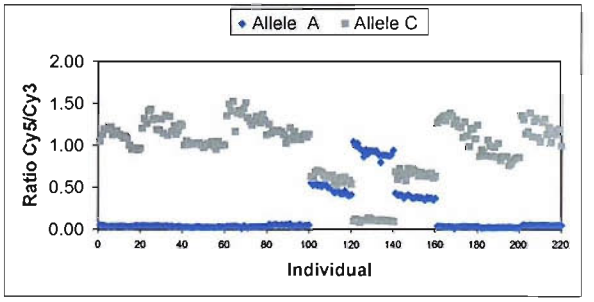
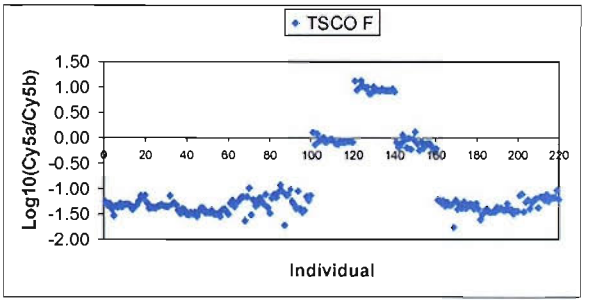


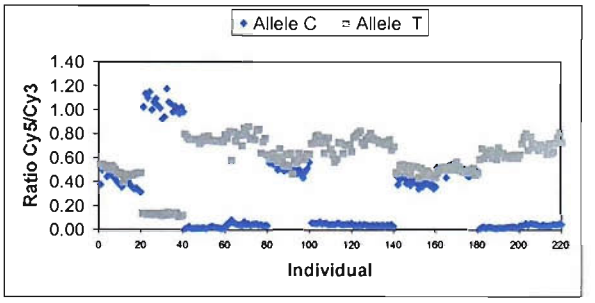
Figure 6.8 - Indirect approach, pentaplex A-v3.

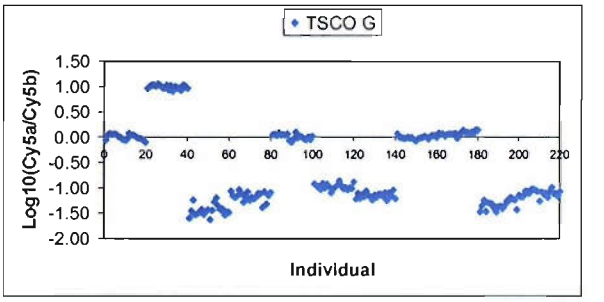


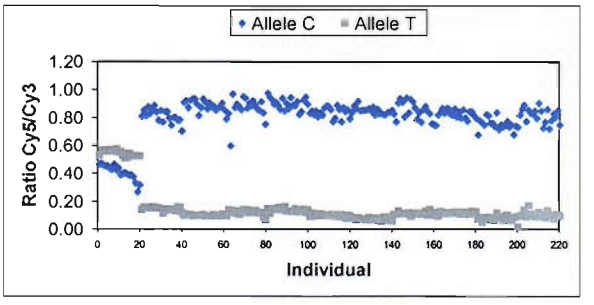


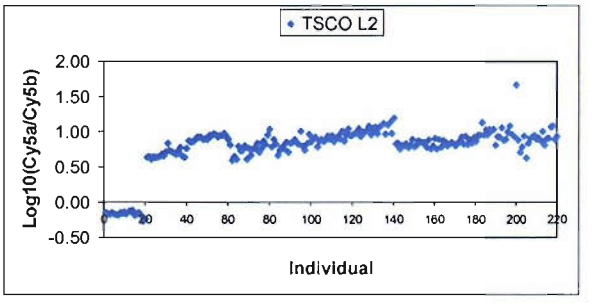


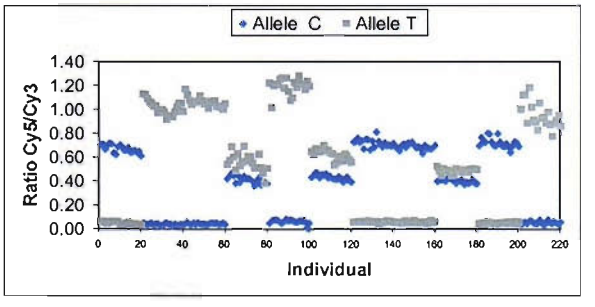












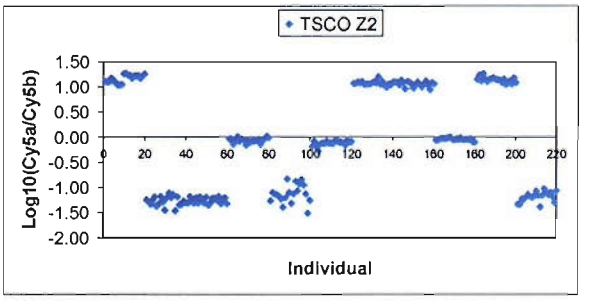
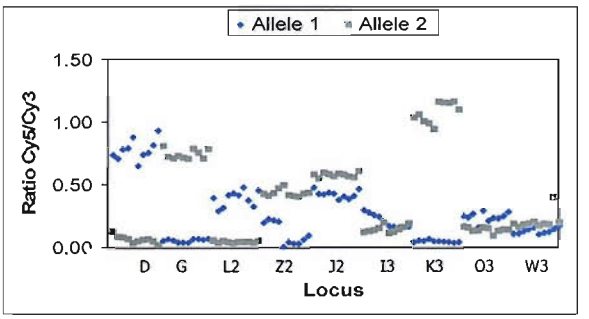


Figure 6.9 - Indirect approach, 9-plex



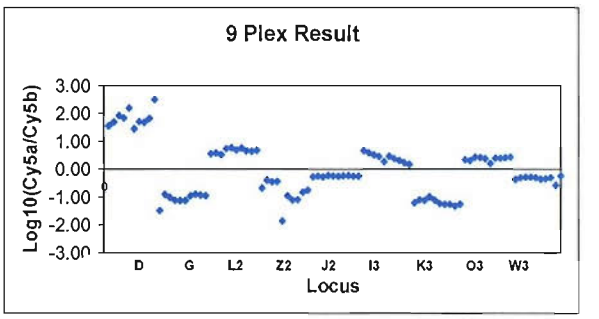
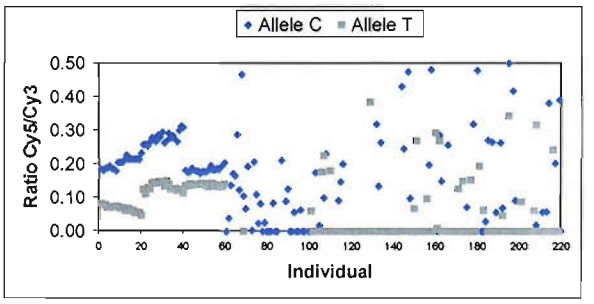
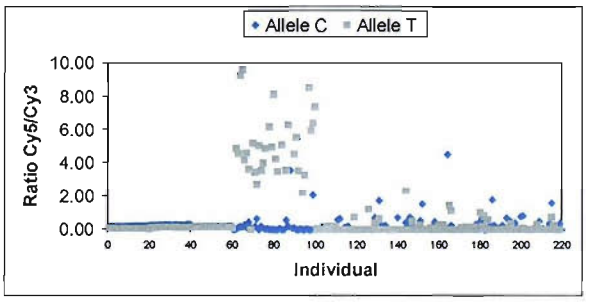


Figure 6.10 - Indirect approach, locus J2 from pentaplex B-v2.





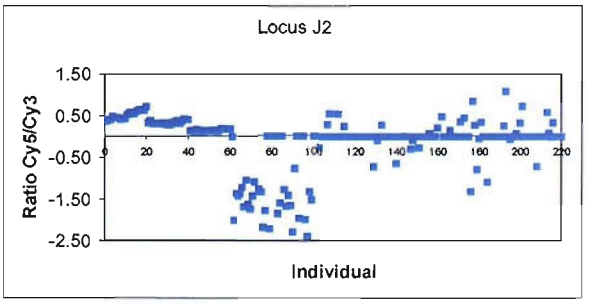
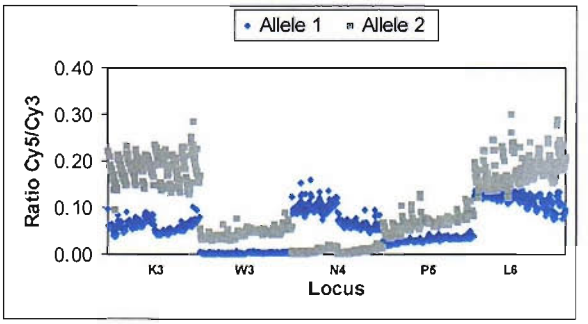


Figure 6.11 - Indirect approach, pentaplex B-v3.



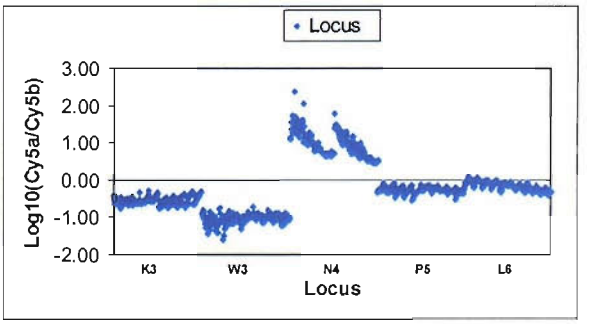
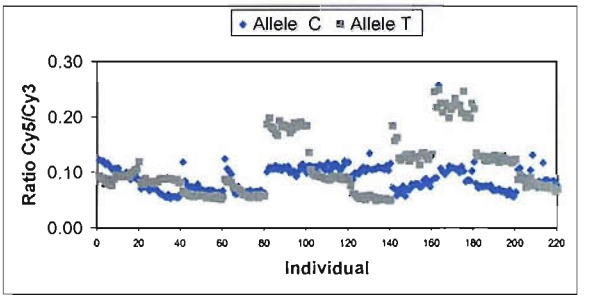
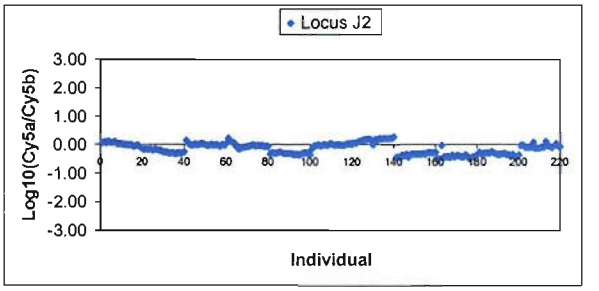
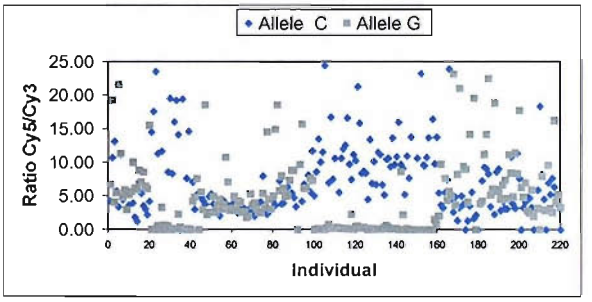
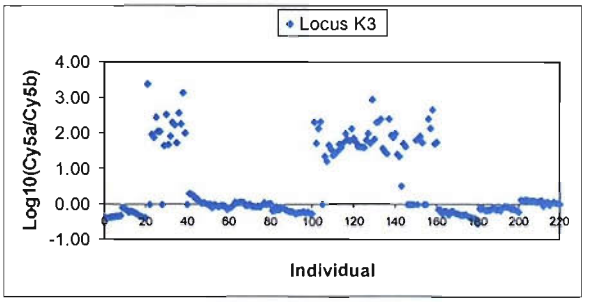


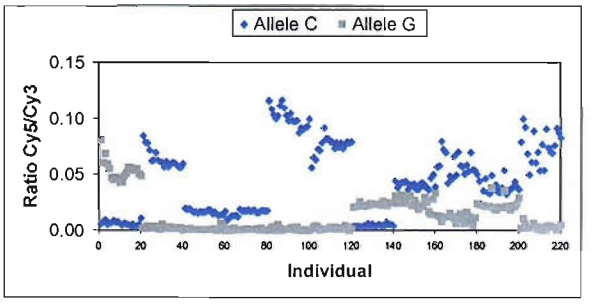
Figure 6.12 - Indirect approach, pentaplex B-v2.

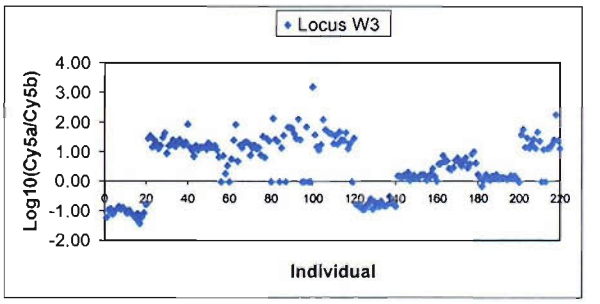


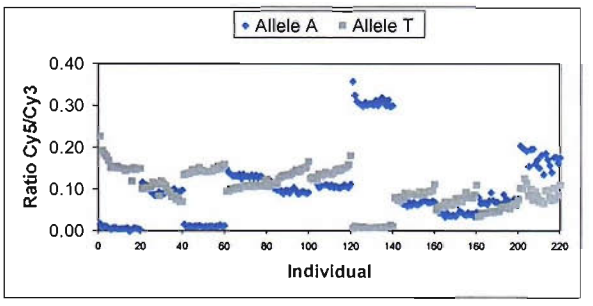


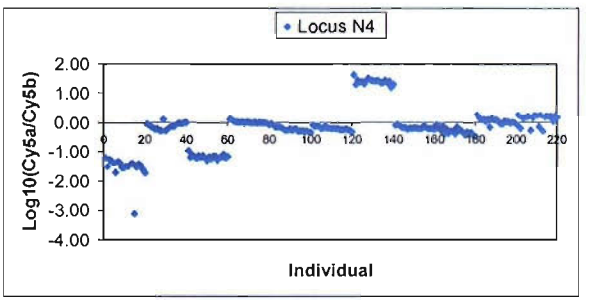


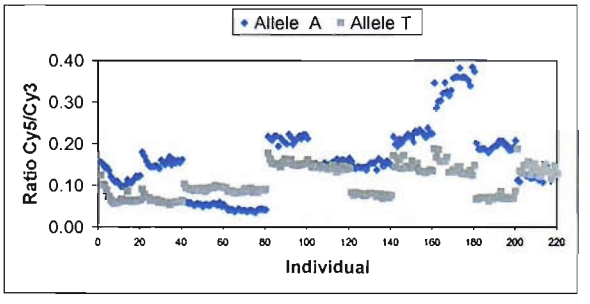












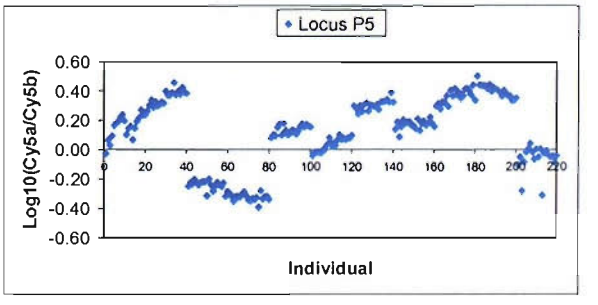
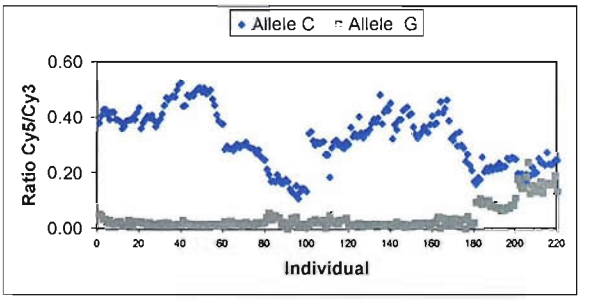
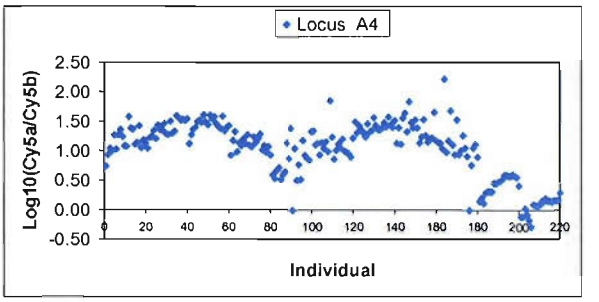
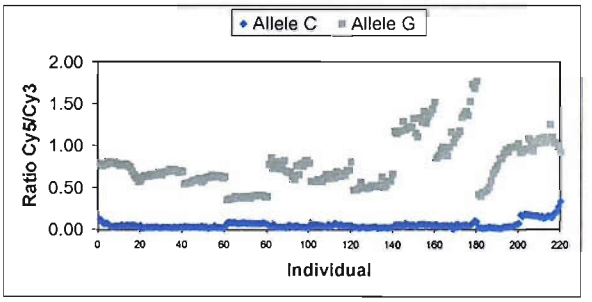
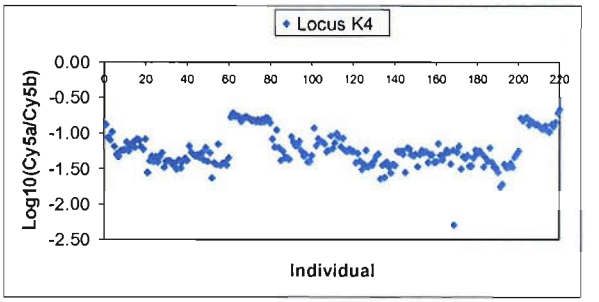


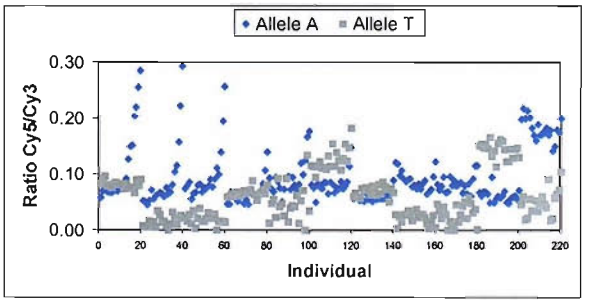
Figure 6.13 - Indirect approach, pentaplex C-v1.

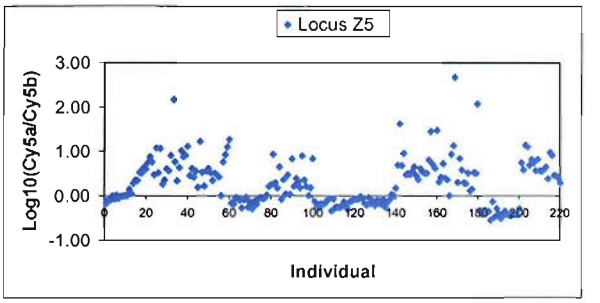


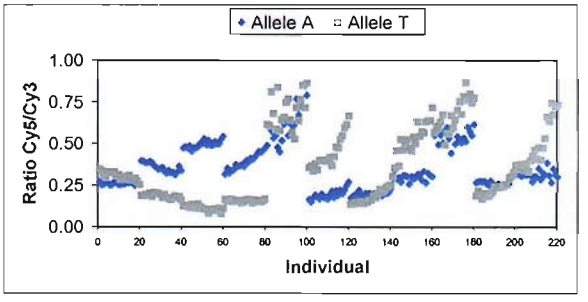


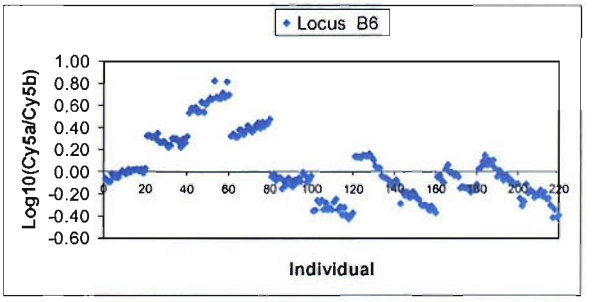


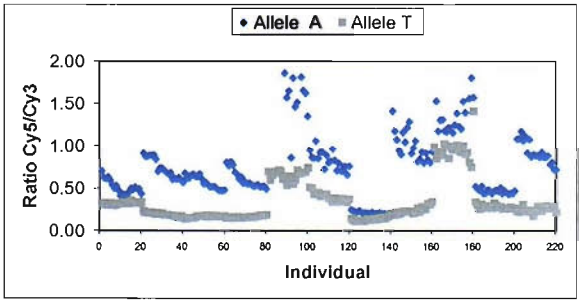


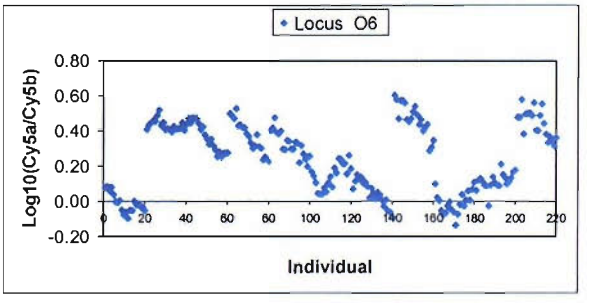












7 References

Aboul-ela, F., Koh, D., Tinocco, I Jr., Martin, F.H., (1985). Base-base mismatches. Thermodynamics of double helix formationb for dCA3XA3G + dCT3YT3G (X, Y = A, C, G, T). *Nucleic Acids Res. 13:4811-4824.*

Anderson, R.C., McGall, G., Lipshutz, R.J., (1998). Polynucleotide Arrays for Genetic Sequence Analysis. *Topics in Current Chemistry, 194:117-129.*

Allawi, H.T., SantaLucia Jr, J., (1997). Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry, 36:10581-10594.*

Allawi, H.T., SantaLucia Jr, J., (1998a). Nearest-Neighbor thermodynamic parameters for internal G·A mismatches in DNA. *Biochemistry, 37:2170-2179.*

Allawi, H.T., SantaLucia Jr, J., (1998b). Nearest-Neighbor thermodynamics of internal A·C mismatches in DNA. Sequence dependence and pH Effects. *Biochemistry, 37:9435-9444.*

Allawi, H.T., SantaLucia Jr, J., (1998c). Thermodynamics of internal C·T mismatches in DNA. *Nucleic Acids Res. 26:2694-2701.*

Allawi, H.T., SantaLucia Jr, J., (1998d). NMR solution structure of a DNA dodecamer containing single G·T mismatches. *Nucleic Acids Res. 26:4925-4934.*

Allemand, J.F., Bensimon, D., Jullien, L., Bensimon, A., Croquette, V., (1997). pH dependent specific binding and combing of DNA. *Biophys. J. 73:2064-2070.*

Auroux, P. A., Iossifidis D., Reyes D. R., Manz A., (2002). Micro Total Analysis Systems. 2. Analytical Standard Operations and Applications. *Anal. Chem. 74:2637-2652.*

Beattie, W.G., Meng, L., Turner, S.L., Varma, R.S., Dao, D.D., Beattie, K.L., (1995). Hybridisation of DNA targets to glass-tethered oligonucleotide probes. *Mol. Biotec. 4:213-225.*

Beier, M., Hoheisel, J.D., (2000). Production by quantitative photolithographic synthesis of individually quality checked DNA Microarrays. *Nucleic Acids Res. 28:e11.*

Beier, M., Hoheisel, J.D., (1999). Versatile derivatisation of solid support media for covalent bonding on DNA-microchips. *Nucleic Acids Res. 27:1970-1977.*

Belgrader, P., Young, S., Yuan, B., Primeau, M., Christel, L.A., Pourahmadi, F., Northrup, M.A., (2001). A battery powered notebook thermal cycler for rapid multiplex real-time PCR analysis. *Anal. Chem. 73:286-289.*

Belgrader, P., Joshi, R., Ching, J., Zaner, S., Borkholder, D.A., Northrup, M.A., (2000). Real time PCR analysis on nucleic acid purified from plasma using a silicon chip. *Micro Total Analysis Systems, 2000:525-528.*

Belgrader, P., Smith, J.K., Weedn, V.W., Northrup, M.A., (1998). Rapid PCR for identity testing using a battery-powered miniature thermal cycler. *J. Forensic Science, 43:315-319.*

Benters, R., Niemeyer, C.M., Drutschmann, D., Blohm, D., Wohrle, D., (2002). DNA microarrays with PAMAM dendritic linker systems. *Nucleic Acids Res. 30:e10.*

Benters, R., Niemeyer, C.M., Wohrle, D., (2001). Dendrimer-activated solid supports for nucleic acid and protein microarrays. *ChemBiochem. 2:686-694.*

Berti, L., Medintz, I.L., Glazer, A.N., Mathies, R.A., (2001). Energy Transfer cassettes for facile labeling of sequencing and PCR primers. *Anal. Biochem. 292:188-197.*

Bier, F.F., Kleinjung, F., (2001). Feature-size limitations of microarray technology – a critical review. *Fresenius J. Anal. Chem. 371:151-156.*

Bilban, M., Buehler, L.K., Head, S., Desoye, G., Quaranta, V., (2002). Normalising DNA Microarray Data. *Curr. Issues Mol. Biol. 4:57-64.*

Bilban, M., Head, S., Desoye, G., Quaranta, V., (2000). DNA Microarrays: a novel approach to investigate genomics in trophoblast invasion – a review. *Placenta. 21:99-105.*

Blake, R.D., Bizzaro, J.W., Blake, J.D., Day, G.R., Delcourt, S.G., Knowles, J., Marx, K.A., SantaLucia J, Jr., (1999). Statistical mechanical simulation of polymeric DNA melting with MELTSIM. *Bioinformatics, 15:370-375.*

Blake, R.D., Delcourt, S.G., (1998). Thermal stability of DNA. *Nucleic Acids Res. 26:3323-3332.*

Blanchard, A.P., Keiser, R.J., Hood, L.E., (1996). Synthetic DNA arrays. *Biosensors and Bioelectronics 11:687-690.*

Bommarito, S., Peyret, N., SantaLucia J.S. Jr, (2000). Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res. 28:1929-1934.*

Boom, R., Sol, C.J., Salimans, M.M., Jansen, C.L., Wertheim-van Dillen, P.M., van der Noordaa, J.J., (1990). Rapid and simple method for purification of nucleic acids. *Clinical Microbiol. 28:495-503.*

Borer, P.N., Dengler, B., Tinoco I Jr., Uhlenbeck, O.C., (1974). Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol. 86:843-853.*

Breslauer, K.J., Frank, R., Blocker, H., Marky, L.A., (1986). Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. 83:3746-3750.*

Brookes, A.J., (1999). The essence of SNPs. *Gene, 234:177-186.*

Brown, T., (1995). *Aldrichimica Acta. 28:15-20.*

Brownie, J., Shawcross, S., Theaker, J., Whitcombe, D., Ferrie, R., Newton, C., Little, S., (1997). The elimination of primer–dimer accumulation in PCR. *Nucleic Acids Res. 25:3235-3241*

Burns, M.A., Johnson, B.N., Brahmasandra, S.N., Handique, K., Webster, J.R., Krishnan, M., Sammarco, T.S., Man, P.M., Jones, D., Heldsinger, D., Mastrangelo, C.H., Burke, D.T., (1998). An integrated nanoliter DNA analysis device. *Science, 282:484-487.*

Chen, H., Zhu, G., (1997). Computer program for calculating the melting temperature of degenerate oligonucleotides used in PCR or hybridisation. *BioTechniques, 22:1158-1160.*

Cheng, J., Shoffner, M.A., Hvichia, G.E., Kricka, L.J., Wilding, P., (1996). Chip PCR II: Investigation of different PCR amplification systems in microfabricated silicon glass chips. *Nucleic Acids Res., 24:380-385.*

Cheung, V.G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., Childs, G., (1999). Making and reading microarrays. *Nature Genetics Suppl. 21:15-19.*

Chizhikov, V., Rasooly, A., Chumakov, K., Levry, D.D., (2001). Microarray aznalysis of microbial virulence factors. *Appl. Environ. Microbiol. 67:3258-3263.*

Chrisey, L.A., Lee, G.U., O'Ferrall, E., (1996). Covalent attachment of synthetic DNA to self-assembled monolayer films. *Nucleic Acids Res. 24:3031-3039.*

Christel, L.A., Petersen, K., McMillan, W., Kovacs, G.T.A., (1998). High Aspect ratio silicon microstructures for nucleic acid extraction. *Proceedings for Solid State Actuator Workshop, 1998:363-366.*

Christel, L.A., Petersen, K., McMillan, W., Northrup, M.A., (1999). Rapid, automated nucleic acid probe assay using silicon microstrauctures for nucleic acid concentration. *J. Biomechanical Engineerining, 121:22-27.*

Collins, F.S., Guyer, M.S., Chakravarti, A., (1997). Variations on a theme: Cataloging human DNA sequence variation. *Science, 278:1580-1581.*

Connor, J.M., Ferguson-Smith, M.A., (1984). Essential Medical Genetics. *Blackwell Scientific Publications.*

Cox, T., Millington, S., Fleming, K., Gill, P., *(In preparation).* Transportation of DNA and PCR reagents through silicon micro-channels: effect on subsequent PCR analysis.

Crothers, D.M., Zimm, B.H., (1964). Theory of the melting transition of synthetic polynucleotides: Evaluation of the stacking free energy. *J. Mol. Biol. 116;1-9.*

Daniel, J.H., Iqbal, S., Millington, R.B., Moore, D.F., Lowe, C.R., Leslie, D.L., Lee, M.A., Pearce, M.J., (1998). Silicon microchambers for DNA amplification. *Sensors and Actuators A 71:81-88.*

Delcourt, S.G., Blake, R.D., (1991). Stacking energies in DNA. *J. Biol. Chem. 266:15160-15169.*

Didenko, V.V., (2001). DNA probes using Flourescence Resonance Energy Transfer (FRET): Designs and applications. *Biotechniques, 31:1106-1121*

Divne, A. M., Lundstrom, A., Gyllensten, U., Allen, M., (2003). A novel DNA microarray system for the analysis of limited forensic evidence material. *International Congress Series 1239:9-10*

Doktycz, M.J., Goldstein, R.F., Paner, T.M., Gallo, F.J., Benight, A.S., (1992). Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: evealuation of the nearest-neighbor stacking interactions in DNA. *Biopolymers, 32:849-864.*

Doktycz, M.J., Paner, T.M., Amaratunga, M., Benight, A.S., (1990). Thermodynamic stability of the 5′ dangling-ended DNA hairpins formed from sequences 5′-(XY)2GGATAC(T)4GTATCC-3′, where X, Y = A, T, G, C. *Biopolymers, 30:829-845.*

Dubiley, S., Kirillov, E., Mirzabekov, A., (1999). Polymorphism analysis and gene detection by minisequencing on an array of gel-immobilised primers. *Nucleic Acids Res. 27:e19.*

Duffy, D.C., Gillis, H.L., Sheppard, N.F.J., Kellogg, G.J., (1999). *Anal. Chem. 71:4669-4678.*

Dunn, W.C., Jacobson, S.C., Waters, L.C., Kroutchininia, N., Khandurina, J., Foote, R.S., Justice, M.J., Stubbs, L.J., Ramsey, J.M., (2000). PCR amplification and analysis of simple sequence length polymorphisms in mouse DNA using a single microchip device. *Anal. Biochem. 277:157-160.*

Effenhauser, C.S., Manz, A., Widmer, H.M., (1993). *Anal. Chem. 65:2637-2642.*

Erdogen, F., Kirchner, R., Mann, W., Ropers, H., Nuber, U.A., (2001). Detection of mitochondrial single nucleotide polymorphisms using primer elongation reaction on oligonucleotide microarrays. *Nucleic Acids Res. 29:e36.*

Ermmanntraur, E., Kohler, J.M., Schulz, T., Wohlfart, K., Wolfl, S., (1997). *German Patent, DE 197 06 570 C1.*

Fodor, S.P., Rava, R.P., Huang, X.C., Pease, A.C., Holmes, C.P., Adams, C.L., (1993). Multiplexed biochemical assays with biological chips. *Nature, 364:555-556.*

Fodor, S.P. Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., Solas, D., (1991). Light directed, spacially addressable parallel chemical synthesis. *Science, 251:767-773.*

Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T., Turner, D.H., (1986). Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. 83:9373-9377.*

Frutos, A.G., Pal, S., Quesada, M., Lahari, J., (2002). Methods for detection of single-base mismatches using bimolecular beacons. *J. Am. Chem. Soc. 124:2396-2397.*

Gerhold, D., Lu, M., Xu, J., Austin, C., Caskey, C.T., Rushmore, T., (2001). Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays. *Physiol. Genomics, 5:161-170.*

Giesen, U., Kleider, W., Berding, C., Geiger, A., Orum, H., Nielsen, P.E., (1998). A formula for thermal stability (T_m) prediction PNA/DNA duplexes. *Nucleic Acids Res. 26:5004-5006.*

Gill, P., Jeffreys, A.J., Werret, D.J., (1985). Forensic application of DNA 'fingerprints'. *Nature, 318:577-579.*

Gill, P., (2001). An assessment of the utility of single nucleotide polymorphisms (SNPs) for Forensic purposes. *Int. J. Leg. Med. 114:204-210.*

Gill, P., Hussain, J., Millington, S., Long, A.S., Tully, G., (2000). An assessment of the utility of SNPs. *Progress in Forensic Genetics, 8:405-407.*

Giordano, B.C., Ferrence, J., Swedberg, S., Huhmner, A.F.R., Landers, J.P., (2001). Polymerase chain reaction in polymeric Microchips: DNA amplification in less than 240 seconds. *Analytical Biochem. 291:124-132.*

Goedecke, N., Manz, A., (2001). *Preceedings of Micro Total Analysis Systems 2001. Kluwer Academic Publishers: Dordrecht, The Netherlands, 375-376.*

Goedecke, N., *(Personal Communication).*

G**oodman**, M.F., Creighton, S., Bloom, L.B., Petruska, J., (1993). Biochemical basis of DNA replication fidelity. *Crit. Rev. Biochem. Mol. Biol. 28:83-126.*

Gorelenkov, V., Antipov, A., Lejnine, S., Daraselia, N., Yyryev, A., (2001). Set of novel tools for PCR primer design. *Biotechniques, 31:1326-1330.*

Gotoh, O., Tagashira, Y., (1981). Locations of frequently opening regions on natural DNAs and their relation to functional loci. *Biopolymers, 20:1043-1058.*

Graveel, C.R., Jatkoe, T., Madore, S.J., Holt, A.L., Farnham, P.J., (2001). Expression profilling and identification of novel genes in heptocellular carcinoma. *Oncogene, 20:2704-2712.*

Graves, D.J., Hung-Ju Su, McKenzie, S.E., Surrey, S., Fortina, P., (1998). System for preparing microhybridisation arrays on glass slides. *Anal. Chem. 70:5085-5092.*

Greenspoon, S.A., Scarpetta, M.A., Drayton, M.L., Turek, S.A., (1998). QIAamp spin columns as a method of DNA isolation for forensic casework. *J. Foren. Casework, 43:1024-1030.*

Grubwieser, P., Muhlmann, R., Parson, W., (2003). New sensetive primers for the STR locus D2S1338 for degraded casework DNA. *Int. J. Leg. Med. 117:185-188.*

Grunstein M., Hogness D.S., (1975). Colony hybridisation: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. 72:3961-3965.*

Gryatiov, S., Schultz, R.G., (1994). Stabilisation of DNA:DNA and DNA:RNA duplexes by substitution of 2'-deoxy-adenosine with 2'-deoxy-2aminoadenosine. *Tetrahedron Lett. 35:2492-2499.*

Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R., Smith, L.M., (1994). Direct fluorescence analysis of genetic polymorphisms by hybridisation with oligonucleotide arrays on glass supports. *Nucleic Acids Res. 22:5456-5465.*

Hacia, J.G., (1999). Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genetics Supp. 21:42-47.*

Hacia, J.G., Brody, L.C., Collins, F.S., (1998). Applications of DNA chips for genomic analysis. *Molecular Psychiatry, 3:483-492.*

Harrison, D.J., Manz, A., Fan, Z.H., Ludi, H., Widmer, H.M., (1992). *Anal. Chem. 64:1926-1932.*

Harrison, D.J., Fluri, K., Seiler, K., Fan, Z.H., Effenhauser, C.S., Manz, A., (1993). *Science, 261:895-897.*

Heller, M.J., (2002). DNA Microarray Technology: eDevices, Systems, and Applications. *Annual review Biomedical Eng. 4:129-153.*

Holm, C., Meeks-Wagner, D.W., Fangman, W.L., Botstein, D., (1986). A rapid efficient method for isolating DNA from yeast. *Gene, 42:169-173.*

Hong, J.W., Hosokawa, K., Fujii, T., Seki, M., Endo, I., (2001). Microfabricated polymer chip for capillary gel electrophoresis. *Biotechnol. Prog. 17:958-962.*

Huang, M.M., Arnheim, N., Goodman, M.F., (1992). Extension of base mispairs by *Taq* Polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res. 20:4567-4573.*

Huber, M., Losert, D., Hiller, R., Harwanegg, C., Mueller, M.W., Schmidt, W.M., (2001). Detection of Single Base alterations in genomic DNA by solid phase polymerase chain reaction on Oligonucleotide microarrays. *Anal. Biochem. 299:24-30.*

Hussain, J.I., Gill, P., Long, A.S., Dixon, L., Hinton, K., Hughes, J., Tully, G., (2003). Rapid preparation of SNP multiplexes utilising universal reporter primers and their detection by gel electrophoresis and microfabricated arrays. *International Congress Series, 1239:5-8.*

Jacobs, K.A., Rudersdorf, R., Neill, S.D., Dougherty, J.P., Brown, E.L., Fritsch, E.F., (1988). The thermal stability of oligonucleotide duplexes is sequence independent in tetramethylammonium salt solutions: application to identifying recombinant DNA clones. *Nucleic Acids Res. 16:4637-4650.*

Jacobson, S.C., Ramsey, J.M., (1996). Integrated microdevise for DNA restriction fragment analysis. *Anal. Chem. 68:720-723.*

Jain, A.N., Tokuyasu, T.A., Snijders, A.M., Segraves, R., Albertson, D.G., Pinkel, D., (2002). Fully automatic quantification of Microarray image data. *Genome Research, 12:325-332.*

Jakeway S. C., de Mello A. J., Russell E. L., (2000). Miniaturised total analysis systems for biological analysis. *Fresenius J. Anal. Chem. 366:525-539.*

Jefferys, A.J., Wilson, V., Thein, S.L., (1985a). Hypervariable 'minisatellite' regions in human DNA. *Nature, 314:67-73.*

Jefferys, A.J., Wilson, V., Thein, S.L., (1985b). Individual - specific 'fingerprints' of human DNA. *Nature, 316:76-79.*

Joos, B., Kuster, H., Cone, R., (1997). Covalent attachment of hybridisable oligonucleotides to glass supports. *Anal. Biochem. 247:96-101.*

Jung, A., Stemmler, I., Brecht, A., Gauglitz, G., (2001). Covalent strategy for immobilisation of DNA microspots suitable for microarrays with label-free and time resolved optical detection and hybridisation. *Frsenius J. Anal. Chem. 371:128-136.*

Kanazawa, H., Noumi, T., Futai, M., (1986). Analysis of Escherichia coli mutants of the H(+)-transporting ATPase: determination of altered site of the structural genes. *Methods Enzymol. 126:595-603.*

Kafatos, F.C., Jones, C.W., Efstratiadis, A., (1979). Determination of nucleic acid sequence homologies and relative concentrations by a dothybridisation procedure. *Nucleic Acids Res. 24:1541-1552.*

Kellogg, G.J., Arnold, T.E., Carvalho, B.L., Duffy, D.C., Sheppard, N.F., (2000). *Proceedings of Micro Total Analysis 2000; Kluwer Academic Publishers, Dordrecht, The Netherlands, 239-242.*

Kelley, M.R., (1995). Rapid genomic DNA purification from Drosophila melanogaster for restriction and PCR. *Qiagen News, 1:8-9.*

Kereszturya, L., Rajczya, K., Laszikb, A., Gyodia, E., Penzes, M., Falus, A., Petranyia, G.G., (2002). Combination of DNA-Based and conversational methods to detect human leukocyte antigen polymorphism and its use for paternity testing. *Am. J. Forensic Med. and Pathology, 23:57-62.*

Khandurina, J., Guttman, A., (2002). Microchip based high throughput screening analysis of combinatorial libraries. *Combinantorial Chemistry, 6:359-366.*

Kimpton, C.P., Fisher, D., Watson, S., Adams, M., Urquhart, A., Lygo, J., Gill, P., (1994). Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR Loci. *Int. J. Leg. Med. 106:302-311.*

Knapp, M. R., Sundberg S., Kopf-Sill A., Nagle R., Gallagher S., Chow. C., Wada G., Nikiforov T., Cohen C., Parce J. W., (1999). Laboratory on a chip: A new experimentation format. *Lab on a Chip, January 1999, 24-26.*

Kopp, M.U., de Mello, A.J., Manz, A., (1998). Chemical Amplification: Continuous-Flow PCR on a chip. *Science, 28:1046-1048.*

Kumar, A., (1995). Alternative view on thermal stability of the DNA duplex. *Biochemistry, 34:12921-12925.*

Lareu, M., Sobrino, B., Phillips, C., Torres, M., Brion, M., Carracedo, A., (2003). Typing Y-chromosome single nucleotide polymorphisms with DNA microarray technology. *International Congress Series, 1239:21-25*

Lagally, E.T., Simpson, P.C., Mathies, R.A. (2000). Monolithic integrated microfluidic DNA amplification and capillary electrophoresis analysis system. *Sensors and Actuators B, 63:138-146.*

Lagally, E.T., Emich, C.A., Mathies, R.A., (2001). Fully integrated PCR-capillary electrophoresis microsystem for DNA analysis. *Lab Chip, 73:565-570.*

Lamture, J.B., Beattie, K.L., Burke, B.E., Eggers, M.D., Ehlich, D.J., Fowler, R., Hollis, M.A., Kosicki, B.B., Reich, R.K., Smith, S.R., (1994). Direct detection of nucleic acids hybridisation on the surface of a charge coupled device. *Nucleic Acids. Res. 22:2121-2125.*

Lander, E.S., (1999). Array of Hope. *Nature Genetics Suppl. 21:3-4.*

Leary, J.J., Brigati, D.J., Ward D.C., (1983). Rapid and sensitive colorimetric method for visualising biotin-labelled DNA probes hybridised to DNA or RNA immobilised on nitocellose: Bio-blots. *Proc. Natl. Acad. Sci. 80:4045-9.*

Lee, P.H., Sawan, S.P., Modrusan, Z., Arnold, L.J.Jr. Reymolds, M.A., (2002). An efficient binding chemistry for Glass polynucleotide microarrays. *Bioconjugate Chem. 13:97-103.*

Lemmo, A.V., Rose D.J., Tisone, T.C., (1998). Injet delivery technology: applications in drug discovery. *Current Opin. Biotechnol. 9:615-617.*

Lennon, G.G., Lehrach, H., (1991). Hybridisation analyses of arrayed cDNA libraries. *Trends in Genetics, 7:314-317.*

Leonard, G.A., Booth, E.D., Brown, T., (1990). Structural and thermodynamic studies on the adenine•guanine mismatch in B-DNA. *Nucleic Acids Res. 18:5617-5623.*

Lichtenberg, J., de Rooij, N.F., Verpoorte, E., (2002). Sample pretreatment on microfabricated devices. Review. *Talanta, 56:233-266.*

Lim, K., Kim, S., Na, K., Park, J.K., Hahn, J.H., (2001). *Proceedings of Micro Total Analysis Systems; Kluwer Academic Publishers: Dordrecht, The Netherlands, 401-402.*

Lin, Y.C., Li, M., Chung, M.T., Wu, C.Y., Young, K.C., (2002). Real-time Microchip Polymerase Chain Reaction System. *Sensors and Materials, 14:199-208.*

Lindroos, K., Liljedahl, U., Raitio, M., Syvanen, AC., (2001). Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. *Nucleic Acids Res. 29:e69-79.*

Lipshutz, R.J., Fodor, S.A., Gingeras, T.R., Lockhart, D.J., (1999). High density synthetic oligonucleotide arrays. *Nature Genetics suppl. 21:20-24.*

Liu, S., Shi, Y., Ja, W.W., Mathies, R.A., (1999). Optimisation of High-Speed DNA sequencing on microfabricated capillary electrophoresis channels. *Anal. Chem. 71:556-573.*

Liu, S., Ren, H., Gao, Q., Roach, D.J., Loder, R.T.J., Armstrong, T.H., Mao, Q., Blaga, I., Barker, D.L., Jovanovich, B., (2000). *Proc. Natl. Acad. Sci. 97:5369-5374.*

Lowe, T., Sharefkin, J., Qi Yang, S., Dieffenbach, C. W., (1990). A computer program for the selection of oligonucleotide primers for polymerase chain reactions. *Nucleic Acids Res. 18:1757-1761.*

Lygo, J.E., Johnson, P.E., Holdaway, D.J., Woodroffe, S., Whitaker, J.P., Clayton, T.M., Kimpton, C.P., Gill, P., (1994). The validation of short tandem repeat (STR) loci for use in forensic casework. *Int. J. Leg. Med. 107:77-89.*

Lyon, W.A., Nie, S., (1997). Confinement and detection of single molecules in submicrometer channels. *Anal. Chem. 69:3400-3405.*

Manz, A., Graber N., Widmer H. M., (1990). Miniaturised Total Chemical Analysis Systems: a novel concept for chemical sensing. *Sensors and Actuators, B1:244-248.*

Manz, A., Harrison, D.J., Fettinger, J.C., Verpoorte, E., Ludi, H., Widmer, H.M., (1991). *Transducers 91:939-941.*

Manz, A., Harrison, D.J., Verpoorte, E., Fettinger, J.C., Paulus, A., Ludi, H., Widmer, H.M., (1992). *Chromatography, 593:253-258.*

Martin, J. B., (1987). Molecular Genetics: Applications to the clinical neurosciences. *Science, 238:765-772.*

Martynova, L., Locascio L.E., Gaitan, M., Kramer, G.W., Christensen, R.G., MacCrehan, W.A., (1997). Fabrication of plastic microfluid channels by imprinting methods. *Anal. Chem. 69:4783-4789.*

Maskos, U., Southern, E.M., (1993). A novel method for the analysis of multiple sequence variants by hybridisation to oligonucleotide arrays. *Nucleic Acids Res. 21:2267-2268.*

Matveeve, O.V., Shabalina, S.A., Nemtsov, V.A., Tsodikov, A.D., Gesteland, R.F., Atkins, J.F., (2003). Thermodynamic calculations and statistical correlations for oligo-probes design. *Nucleic Acids Res. 3114:4211-4217.*

McCormick, R.M., Nelsen, R.J., Alonso-Amigo, M.G., Benvegnu, D.J., Hooper, H.H., (1997). Microchannel electrophoretic separations of DNA in injection-molded plastic substrates. *Anal. Chem. 69:2626-2630.*

McGall, G.H., Barone, A.D., Diggelmann, M., Fodor, S.P.A., Gentalen, E., Ngo, N., (1997). The eficiency of light directed sunthesis of DNA arrays on glass substrates. *J. Am. Chem. Soc. 119:5082-5090.*

McGall, G.H., Fidanza, J.A., (2001). Photolithographic synthesis of high density oligonucleotide arrays. *Methods Mol. Biol. 170:71-101.*

Medintz, I.L., Paegel, B.M., Blazej, R.G., Emrich, C.A., Berti, L., Scherer, J.R., Mathies, R.A., (2001). High-performance genetic analysis using microfabricated capillary array electrophoresis microplates. *Electrophoresis, 22:3845-3856.*

Mehrotra, J., Bishai, W.R., (2001). Regulation of virulence genes in Mycobacterium tuberculosis. *Int. J. Med. Microbiol. 291:171-182.*

Mitnik, L., Carey, L., Burger, R., Desmarais, S., Koutny, L., Wernet, O., Matsudaira, P., Ehrlich, D., (2002). High-speed analysis of multiplexed short tandem repeats with an electrophoretic microdevice. *Electrophoresis, 23:719-726.*

Monni, O., Burland, M., Mousses, S., Kononen, J., Sauter G., (2001). Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. 98:5711-16.*

Modin, C., Pedersen, F.S., Duch, M., (2000). Comparison of DNA polymerase for quantification of single nucleotide differences by Primer Extension assays. *BioTechniques, 28:48-51.*

Mullis, K.B., Faloona, F.A., (1987). Specific synthesis of DNA in vitro via a polymerase chain reaction. *Methods Enzymol. 155:335-350.*

Nagai, H., Murakami, Y., Yokayama, K., Tamiya, E., (2001). High throughput PCR in silicon based microchamber array. *Biosensors and Bioelectronics, 16:1015-1019.*

Newton, C.R., Graham, A., (1994). PCR; Introduction to Biotechniques. *Bios. Scientific Publishers, Oxford, UK.*

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C., Markham, A.F., (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res. 17:2503-2516.*

Nguyen, H-K., Bonfils, E., Auffray, P., Costaglioli, P., Schmitt, P., Asseline, U., Durand, M., Maurizot, J-C., Dupret, D., Thuong, N.T., (1998). The stability of duplexes involving AT and/or $G^{4Et}C$ base pairs is not dependent on their AT/$G^{4Et}C$ ratio content. Implication for DNA sequencing by hybridisation. *Nucleic Acids Res. 26:4249-4258.*

Nikiforov, T.T., Rendle, R.B., Goelet, P., Rogers, Y.H., Kotewicz, M.L., Anderson, S., Trainor, G.L., Knapp, M.R., (1994). Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Res. 22:4167-4175.*

Northrup, M.A., Ching, M.T., White, R.M., Watson, R.T., (1993). *Transducers 93:924-926.*

Northrup, M.A., Gonzalez, C., Hadley, D., Hills, R.F., Landre, P., Lehew, S., Saika, R., Sinski, J.J., Watson, R., Whatman, J.R., (1995). *Transducers 95:746-767.*

Northrup, M.A., (1998). A miniature analytical instrument for nucleic acids based on micromachined silicon reaction chambers. *Anal. Chem. 70:918-922.*

Oleschuk, R.D., Shultz-Lockyear, L.L., Ning, Y., Jed Harrison, D., (2000). Trapping of bead-based reagents within microfluidic Systems: On-chip Solid-Phase extraction and electrochromatography. *Anal. Chem. 72:585-590.*

Paegel, B.M., Emrich, C.A., Wedemayer, G.J., Scherer, J.R., Mathies, R.A., (2002). High throughput DNA sequencing with a microfabricated 96-lane capillary electrophoresis bioprocessor. *Proc. Natl. Acad. Sci. 99:574-579.*

Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L., Syvanen, A.C., (1997). Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res. 7:606-614.*

Peterson, A.W., Heaton, R.J., Georgiadis, R.M., (2001). The effect of surface probe density on DNA hybridisation. *Nucleic Acids Res. 29:5163-5168.*

Peyret, N., Ananda Seneviratne, P., Allawi, H.T., SantLucia Jr, J., (1999). Nearest-Neighbor thermodynamics and NMR of DNA sequences with internal A·A, C·C, G·G and T·T mismatches. *Biochemistry, 38:3468-3477.*

Picken, R.N., Plotch S.J., Wang, Z., Lin, B.C., Donegan, J.J., Yang, H.L., (1988). DNA probes for mycobacterium. I. Isolation of DNA probes for the identification of Mycobacterium tuberculosis complex and for mycobacteria other than tuberculosis (MOTT) *Molecular Cell Probes, 2:111-124.*

Poland, D., Scheraga, H., (1970). Theory of helix-coil transitions in biopolymers. *Academic Press, New York.*

Proudnikov, D., Timofeev, E., Mirzabekov, A., (1998). Immobilisation of DNA in polyacrylamide gel for the manufacture of DNA and DNA-oligonucleotide microchips. *Anal. Biochem. 259:34-41.*

Radtkey, R., Feng, L., Muralhidar, M., Duhon, M., Canter, D., DiPierro, D., Fallon, S., Tu, E., McElfresh, K., Nerenberg, M., Sosnowski, R., (2000). Rapid, high fidelity analysis of simple sequence repeats on an electronically active DNA microchip. *Nucleic Acids Res. 28:e17.*

Rasmussen, S.R., Larsen, M.R., Rasmussen, S.E., (1991). Covalent immobilization of DNA onto polystyrene microwells; the molecules are only bound at the 5' end. *Anal. Biochem. 198:138-142.*

Record, M.T., Anderson, C.F., Lohman, T.M., (1978). Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association of release, screening, and ion effects on water activity. *Q. Rev. Biophys. 11:103-178.*

Reyes, D. R., Iossifidis D., Auroux P-A, Manz A., (2002). Micro Total Analysis Systems. 1. Introduction, Theory, and Technology. *Anal. Chem. 74:2623-2636.*

Rodriguez, I., Lesaicherre, M., Tie, Y., Zou, Q., Yu, C., Singh, J., Meng, L.T., Uppili, S., Li, S.F.Y., Gopalakrishnakone, P., Selvanayagam, Z.E., (2003). Practical integration of polymerase chain reaction amplification and electrophoretic analysis in microfluidic devices for genetic analysis. *Electrophoresis, 24:172-178.*

Ronai, Z., Barta, C., Sasvari-Szekely, M., Guttman, A., (2001). DNA analysis on electrophoretic microchips:effect of operational variables. *Electrophoresis, 22:294-299.*

Rose, D., (2000). Microfluidic Technologies and instrumentation for printing DNA microarrays. *Microarray Biochip Technology, Biotechniques Book Division, Eaton Publishing, USA.*

Rozen, R. Fox, J.E., Hack, A.M., Fenton, W.A., Horwich, A.L., Rosenberg, L.E., (1986). DNA analysis for ornithine transcarboxylse deficiency. *J. Inherit. Metab. Dis. 9:49-57.*

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A., (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science, 239:487-491.*

SantaLucia Jr, J., (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. 95:1460-1465.*

SantaLucia Jr, J., Allawi, H.T., Ananda Seneviratne, P., (1996). Improved nearest-Neighbor parameters for predicting DNA duplex stability. *Biochemistry, 35:3555-3562.*

SantaLucia Jr, J., Kierzek, R., Turner, D.H., (1990). Effects of GA mismatches on the structure and thermodynamics of RNA internal loops. *Biochemistry, 29:8813-8819.*

SantaLucia Jr, J., Kierzek, R., Turner, D.H., (1991). Stabilities of consecutive A·C, C·C, G·G, U·C and U·U mismatches in RNA internal loops: Evidence for stable hydrogen-bonded U·U and C·C+ pairs. *Biochemistry, 30:8242-8251.*

SantaLucia Jr, J., Turner, D.H., (1997). Measuring the thermodynamics of RNA secondary structure formation. *Biopolymers, 44:309-319.*

Schena, M., Hellor, R., Theriult, T.P., Konrad, K., Lachenmeier, E., Davis, R.W., (1998). Microarrays: biotechnology's discovery platform for functional genomics.. *Trends Biotechnol. 16:301-306.*

Schena, M., Shalon, D., Hellor, R., Chai, A., Brown, P.O., Davis, R.W., (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. 93:10614-10619.*

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science, 270:467-470.*

Schildkraut, C., Lifson, S., (1965). Dependence of the melting temperature of DNA on Salt concentration. *Biopolymers, 3:195-208.*

Schmalzing, D., Koutny, L., Adouian, A., Belgrader, P., Matsudaira, P., Earlich, D., (1997). DNA typing in thirty seconds with a microfabricated device. *Proc. Natl. Acad. Sci. 94:10273-10278.*

Schmalzing, D., Koutny, L., Salas-Solano, O., Adourian, A., Matsudaira, P., Ehrlich, D., (1999). Recent developments in DNA sequencing by capillary and microdevice electrophoresis. *Electrophoresis, 20:3066-3077.*

Schneegass, I., Brautigam, R., Kohler, J.M., (1991a). Miniaturised flow-through PCR with different template types in a silicon chip thermocycler. *Lab Chip. 1:42-49.*

Schneegass, I., Kohler, J.M., (1991b). Flow-through Polymerase Chain Reactions in chip thermocyclers. *J. Biotechnol. 82:101-121.*

Schutz, E., Von Ahsen, N., (1999). Spreadsheet software for thermodynamic melting point prediction of oligonucleotide hybridisation with and without mismatches. *BioTechniques, 27:1218-1224.*

Schwarzacher, H. G., (1976). Modern ideas on Chromosome structure. *Path. Eur. 11:5-13.*

Seidel, R.U., Sim, D.Y., Menz, W., Esashi, M., (1999). *Transducers 99:438-441.*

Service, R.F., (1998). Microchip arrays put DNA on the spot. *Science, 282:396-399.*

Shchepinov, M.S., Udalova, I.A., Brigman, A.J., Southern, E., (1997). Oligonucleotide dendrimers: synthesis and use as polylabelled DNA probes. *Nucleic Acids Res. 25:4447-4454.*

Shinohara, J., Suda, M., Furuta, K., Sakuhara, T., (2000). *MEMS 2000:86-91.*

Shoffner, M.A., Cheng, J., Hvichia, G.E., Kricka, L.J., Wilding, P., (1996). PCR Chip I: Surface passivation of microfabricated silicon-glass chips for PCR. *Nucleic Acids Res. 24:375-379.*

Shuber, A.P., Grondin, V.J., Klinger, K.W., (1995). A simplified procedure for developing multiplex PCRs. *Genome Res. 5:488-493.*

Simpson, P.C., Roach, D., Woolley, A.T., Thorsen, T., Johnston, R., Sensabaugh, G.F., Mathies, R.A., (1998). High-throughput genetic analysis using microfabricated 96-sample capillary array electrophoresis microplates. *Proc. Natl. Acad. Sci. 95:2256-2261.*

Sofi Ibrahim, M., Lofts, R.S., Jahrling, P.B., Henchal, E.A., Weedn, V.W., Northrup, M.A., Belgrader, P., (1998). Real-time microchip PCR for detecting single-base differences in viral and human DNA. *Anal. Chem. 70:2013-2017.*

Song, Y.J., Zhao, T.S., (2001). *J. Micromech. Microeng. 11:713-719.*

Southern, E.M., (1996). DNA chips: analysing sequence by hybridisation to oligonucleotides on a large scale. *Technical Focus TIG. 12:110-115.*

Southern, E.M., (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol. 98:503-517.*

Southern, E.M., Maskos, U., Elder, J.K., (1992). Analysing and comparing nucleic acid sequences by hybridisation to arrays of oligonucleotides: evaluation using experimental models. *Genomics, 13:1008-1017.*

Southern, E., Mir, K., Shchepinov, M. (1999). Molecular interactions on microarrays. *Nature Genetics Suppl. 21:5-9.*

Sparkes, R., Kimpton, C.P., Watson, S., Oldroyd, N., Clayton, T., Barnett, L., Arnold, J., Thompson, C., Hale, R., Chapman, J., Urquhart, A., Gill, P., (1996). The validation of a 7-locus multiplex STR test for use in forensic casework. (I) Mixtures, ageing, degredation and species studies. *Int. J. Leg. Med, 109:186-194.*

Suggs, S.V., Wallace, R.B., Hirose, T., Kawashima, E.H., Itakura, K., (1981). Use of synthetic oligonucleotides as hybridisation probes: isolation of cloned cDNA sequences for human beta-microglobulin. *Proc. Natl. Acad. Sci. 78:6613-6617.*

Sugimoto, N., Kierzek, R., Freier, S.M., Turner, D.H., (1986). Energetics of internal GU Mimatches in Ribonucleotide Helixes. *Biochemistry, 25:5755-5759.*

Sugimoto, N., Honda, K., Sasaki, M., (1994). Application of the thermodynamic parameters of DNA stability prediction to double-helix formation of Deoxyribooligonucleotides. *Nucleosides and Nucleotides, 13:1311-1317.*

Sugimoto, N., Nakano, S., Yoneyama, M., Honda, K., (1996). Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res. 24:4501-4505.*

Syvanen, A.C., Sajantila, A., Lukka, M., (1993). Identification of individuals by analysis of biallelic DNA markers, using PCR and Solid-Phase Minisequencing. *Am. J. Hum. Genet. 52:46-59.*

Terry, S.C., (1975). *PhD Thesis, Stamford, Stamford CA.*

Thomas, N., Ocklind, A., Blikstad, I., Griffiths, S., Kenrick, M., Derand, H., Ekstrand, G., Ellstrom, C., Larsson, A., Anderson, P., (2000). *Proceedings of Micro Total Analysis 2000; Kluwer Academic Publishers, Dordrecht, The Netherlands, 249-252.*

Tian, H., Huhmer, A.F.R., Landers J.P., (2000). Evaluation of silica resins for direct and efficient extraction of DNA from complex biological matrices in a miniaturised format. *Anal. Biochem. 283:176-191.*

Tinoco, I Jr., Uhlenbeck, O.C., Levine, M.D., (1971). Estimation of secondary structure in ribonucleic acids. *Nature, 230:362-367.*

Tyagi, S., Kramer, F.R., (1996). Molecular beacons: Probes that fluoresce upon hybridisation. *Nat. Biotechnol. 14:303-308.*

Urquhart, A., Kimpton, C.P., Downes, T.J., Gill, P., (1994). Variation in short tandem repeat sequences – a survey of twelve microsatellite loci for use as forensic identification markers. *Int. J. Leg. Med. 107:13-20.*

Vologodskii, A.V., Amirikyan, B.R., Lyubchenko, Y.L., Frank-Kamenetskii, M.D., (1984). Allowances for heterogeneous stacking in the DNA helix-coil transition theory. *J. Biomol. Struct. Dyn. 2:1005-1012.*

Von Heeren, F., Verpoote, E., Manz, A., Thormann, W.J., (1996). Micellular electrokinetic chromatography separations and analyses of biological samples on a cyclic planar microstructure. *Anal. Chem. 68:2044-2053.*

Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T., Itakura, K., (1979). Hybridisation of synthetic oligodeoxyribnucleotides to phi chi 174 DNA: the effect of a single base pair mismatch. *Nucleic Acids Res. 6:3543-3557.*

Waters, L.C., Jacobson, S.C., Kroutchinina, N., Khandurina, J., Foote, R.S., Ramsey, J.M., (1998). Multiple sample PCR amplification and electrophoretic analysis on a microchip. *Anal. Chem. 70:5172-5176.*

Weidl, B.H., Bardell, R., Schulte, T., Williams, C., (2000). *Proceedings of MicroTotal Analysis Systems 2000; Kluwer Academic Publishers; Dordrecht, The Netherlands, 299-302.*

Wilding, P., Kricka, L.J., Cheng, J., Hvichia, G., (1998). Integrated cell isolation and Polymerase Chain Reaction analysis using silicon microfilter chambers. *Anal. Biochem. 257:95-100.*

Wilding, P., Shoffner, M.A., Kricka, L.J., (1994). PCR in a silicon microstructure. *Clinical Chem. 40:1815-1818.*

Wolfe, K.A., Breadmore, M.C., Ferrence, J.P., Power, M.E., Conroy, J.F., Norris, P.M., Landers, J.P., (2002). Towards a microchip-based solid-phase extraction method for isolation of nucleic acids. *Electrophoresis, 23:727-733.*

Wood, W.I., Gitschier, J., Lasky, L.A., Lawn, R.M., (1985). Base composition-independent hybridisation in tetramethylammonium chloride: A method for oligonucleotide screening of highly complex gene libraries. *Proc. Natl. Acad. Sci. 82:1585-1588.*

Woolley, A.T., Mathies, R.A., (1994). Ultra-high-speed DNA fragment separations using microfabricated capillary array electrophoresis chips. *Proc. Natl. Acad. Sci. 91:11348-11352.*

Woolley, A.T., Mathies, R.A. (1995). Ultra-high-speed DNA sequencing using capillary electrophoresis chips. *Anal. Chem. 67:3367-3680.*

Woolley, A.T., Hadley, D., Landre, P., de Mello, A.J., Mathies, R.A., Northrup, M.A., (1996). Functional integration of PCR amplification and capillary electrophoresis in a microfabricated DNA analysis device. *Anal. Chem. 68:4081-4086.*

Woolley, A.T., Sensabaugh, G.F., Mathies, R.A., (1997). High-speed DNA genotyping using capillary array electrophoresis chips. *Anal. Chem. 69:2181-2186.*

Xiang, F., Lin, Y., Wen, J., Matson, D.W., Smith, R.D., (1999). An integrated microfabricated device for duel microdialysis and on-line ESI-ion trap mass spectrometry for analysis of complex biological samples. *Anal. Chem. 71:1485-1490.*

Xu, N., Lin, Y., Hofstadler, S.A., Matson, D., Call, C., Smith, R.D., (1998). A microfabricated dialysis device for sample cleanup in electrospray ionisation mass spectrometry. *Anal. Chem. 70:3553-3556.*

Young, R.A., (2000). Biomedical discovery with DNA arrays. *Cell, 102:201-205.*

Zammatteo, N., Jeanmart, L., Hamels, S., Courtois, S., Louette, P., Hevesi, L., Remacle, J., (2000). Comparison between different strategies of covalent attachment of DNA to glass surfaces to build DNA microarrays. *Anal. Biochem. 280:143-150.*

Zhang, Q., Wang, W., Zhang, H., Wang, Y., (2002). *Sens. and Actuators B, 82:75-81.*