Degradation, Quantification and the Theory of Nucleosome Protection

By

Phuvadol Thanakiatkrai

Centre for Forensic Science
Department of Pure and Applied Chemistry
University of Strathclyde

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy

2011

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Date:

Abstract

Partial DNA profiles are often obtained from degraded samples due to allelic and locus dropout, particularly at the high molecular weight loci. Increasing the chance of genotyping success via shortening of amplicon size has been previously demonstrated with mini-STRs. A viable alternative based on nucleosome protection influenced by base sequences was explored in this study.

Different systems for accurate quantification of degraded samples were looked at, including both single- and multi-copy targets. Optimisation was successful for GAPDH and it was compared to Plexor® HY using casework samples. Although GAPDH was more accurate, Plexor® HY was chosen and used to quantify degraded saliva samples due to its higher sensitivity and informativeness. The saliva samples had been degraded via the incubation method, which was assumed to preserve the chromatin structure.

Next-generation and mini-STR kits were assessed in terms of sensitivity and casework genotyping. All kits performed exceptionally well and were comparable in all categories.

60 sequences (58 STRs and amelogenin X and Y) were evaluated for their nucleosome-forming potential (NFP) using two computer programs. The markers were divided into three groups based on their NFPs and the findings were verified empirically by amplifying degraded saliva samples and casework samples using 14 randomly chosen primer sets from the three groups. The effect of nucleosome protection was not observed for degraded saliva and casework samples.

This is the first study that looks at an inherent property of STRs as a determinant of survivability from degradation processes. The work can be expanded to include more sample types. Other computer programs can be used, as predicting nucleosome positions is a rapidly advancing field.

Acknowledgement

Throughout these five eventful years in Glasgow, I have learnt many, many things about life, and that would not have been possible without the following people, who truly deserve these praises.

Dr. Lindsey Welch: What would I do without you? What would you do without me? You have been a dear friend and a great mentor. Thank you for your encouragements and criticisms. To you, I am truly grateful. You will always be my role model. This thesis is yours too.

My family: Thank you for always being there for me when I need you. How supportive you have been throughout these years! Uncles and aunts, thank you for watching over me. Bird and Arm, I would never have been here without you two. Four, who makes my life so colourful. Mom and dad, thank you for letting me choose my own path, for never questioning me, and for never doubting me. I love you. To you, I dedicate this doctorate.

My friends: Thank you for taking care of me and putting up with me all these time. I'll never forget the countless meals, movies, training sessions, camps, drives, and many more that we've enjoyed together. Special mention to the Clintons, Sergey, Neil, Jason, Yuva, Sara, Anika, Kevin, the F.F.B. and Nerdeus.

I might not be the best at keeping in touch, but do know that I'll never forget any of you.

Publications and presentations relating to this research

Publications

Using the Taguchi method for rapid quantitative PCR optimization with SYBR Green I

Phuvadol Thanakiatkrai and Lindsey Welch,

International Journal of Legal Medicine, In Press, 2011

Evaluation of nucleosome forming potentials (NFPs) of forensically important STRs.

Phuvadol Thanakiatkrai and Lindsey Welch,

Forensic Science International: Genetics, In Press, 2010

Oral Presentations

Want to save money? Ask an engineer

Phuvadol Thanakiatkrai

Presented at the 5th WestCHEM Research Day 2010, University of Strathclyde, Glasgow, UK, 24th June 2010.

Rapid and low-cost optimization of two qPCR assays using the Taguchi method

Phuvadol Thanakiatkrai

Presented at the 2nd International Conference on Forensic Genetics (Forensica 2010), Telc, Czech Republic, 24th-26th May 2010.

Nucleosome computation: possible implications for degraded DNA

Phuvadol Thanakiatkrai

Presented at the 31st ENFSI DNA Working Group Meetings, Edinburgh, UK, 21st-22nd October 2009.

Poster Presentations

Rapid and low-cost optimization of qPCR assays using the Taguchi method

Phuvadol Thanakiatkrai, Lindsey Welch

Presented at the University of Strathclyde Research Day 2010, Glasgow, UK, 9th June 2010

Evaluation of 59 STR markers for their NFP

Phuvadol Thanakiatkrai, Lindsey Welch, Peter Gill

Presented at the 5th European Academy of Forensic Science Conference, University of Strathclyde, Glasgow, UK, 8th-11th September 2009

List of abbreviations

°C	Degrees Celsius
η	Signal-to-noise ratio
µL	Microlitre
µM	Micromolar
A	Adenine
ALDOA	Aldolase A, fructose-biphosphate
BHQ	Black hole quencher
bp	Base pairs
BSA	Bovine serum albumin
C	Cytosine
CAD	Caspase activated DNase
CAF	Chromatin assembly factor
CODIS	Combined DNA Index System
CpG	C-phosphate-G
C_q	Quantification cycle
CXR	Carboxy-X-rhodamine
DF	Degree of freedom
DHS	DNase I hypersensitive site
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dsDNA	Double-stranded DNA
EDNAP	European DNA Profiling Group
EDTA	Ethylenediaminetetraacetic acid (disodium salt)
ENFSI	European Network of Forensic Science Institutes
EP01	PowerPlex® EP01 (ESX prototype)
EP02	PowerPlex® EP02 (ESI prototype)
EPG	Electropherogram
ESS	Mentype ESS
FRET	Fluorescence resonance energy transfer
G	Guanine
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase

HAT	Histone acetyltransferase
Hb_x	Heterozygous balance
HEO	Hexaethyleneoxide
HPLC	High-performance liquid chromatography
IPC	Internal PCR control target
LCN	Low copy number
LT-DNA	Low template DNA
MF	AmpF*l*STR® MiniFiler™
$MgCl_2$	Magnesium chloride
mL	Millilitre
mM	Millimolar
MNase	Micrococcal nuclease
mtDNA	Mitochondrial DNA
NDNAD	National DNA database
NFP	Nucleosome-forming potential
ng	Nanogram
nM	Nanomolar
NIST	National Institute of Standards and Technology (US)
NNL	Non-nucleotide linker
NPS	Nucleosome positioning signal
NPScore	Nucleosome positioning score
NTC	No-template control
PARP	Poly (ADP-ribose) polymerase
P_c	Percent contribution
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
pg	Picogram
P_m	Match probability
Pol	Polymerase enzyme
PP-16	PowerPlex® 16
qPCR	Quantitative polymerase chain reaction
RFLP	Restriction fragment length polymorphism

RFU	Relative fluorescence unit
RNU2	RNA, U2 small nuclear 1
ROX	6-carboxy-X-rhodamine
RPPH1	Ribonuclease P RNA component H1
S5	PowerPlex® S5
SD	Standard deviation
SGM+	AmpF*l*STR® SGM Plus®
SLP	Single-locus probe
SNP	Single nucleotide polymorphism
S_p	Stutter proportion
ssDNA	Single-stranded DNA
STR	Short tandem repeat
SWGDAM	Scientific Working Group on DNA Analysis Methods
T	Thymine
TAMRA	Tetramethyl-6-carboxyrhodamine
Taq	*Thermus aquaticus* DNA polymerase
TERT	Telomerase reverse transcriptase
T_m	Melting temperature
Tris	Tris(hydroxymethyl)methylamine
TSPY	Testis-specific protein, Y encoded locus
UV	Ultraviolet
VNTR	Variable number tandem repeat

Table of contents

List of figures

List of tables

1 General introduction

1.1 Scientific basis of DNA

Deoxyribonucleic acid, commonly referred to as DNA, is found in every organism and located inside the nucleus in eukaryotes. Protected inside the nucleus, DNA in almost every cell of a multi-cellular organism is the same, with some exception such as red blood cells, which lack nuclei, and B and T cells of the adaptive immune system, which have genetic recombinations. DNA codes for the structure and function of all ribonucleic acids and, subsequently, proteins required for the proper development of any organism. DNA is a polymer with subunits called nucleotides. A deoxyribose sugar, a phosphate group and a base are the components that make up a nucleotide. Four bases are found in the DNA molecule – adenine, guanine, cytosine and thymine. Adenine and guanine are made of two carbon rings and are classed as purines; while cytosine and thymine are made of one carbon ring and are classed as pyrimidines. The phosphate groups and sugars are linked together by phosphodiester bonds and together make up the backbone of DNA, with bases extending out like rungs of a ladder (Figure 1.1). Being double helical in its natural state, the bases of one strand will pair with those on the other strand; adenine pairs with thymine and guanine pairs with cytosine [1, 2].

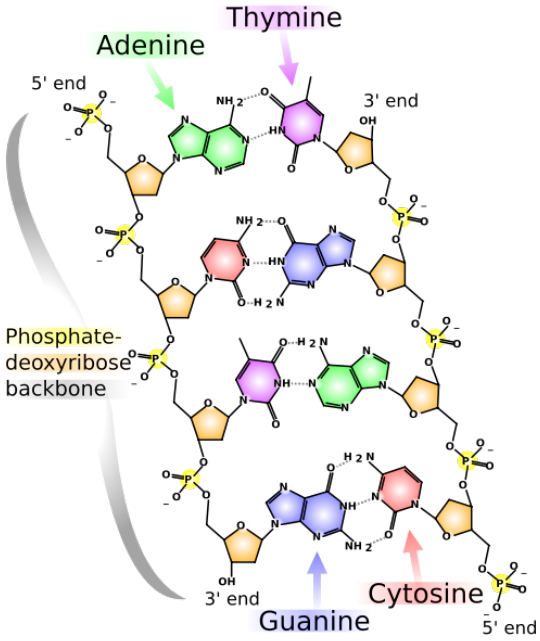


Figure 1.1. Schematic representation of DNA with colour-coded bases. Produced by Madeleine Price Ball and reprinted under GNU Free Documentation License.

1.2 Use of DNA in forensic science

Multiple reasons contribute to making DNA an excellent source of information that aids individualization in forensic science. Once a zygote has been created by the fusion of a sperm and an ovum, each and every other cell that is replicated through mitosis and subsequently differentiated still retains the same DNA sequence. This means that there is almost no difference in the sequence of DNA in an individual regardless of the type of cell it is taken from; therefore, it does not matter whether leukocytes, osteocytes or spermatozoa are used as the source of DNA. Moreover, the DNA bases currently used for an individual's identification do not change throughout his or her lifetime, barring some mutations, recombinations and modifications. Additionally, DNA can withstand more environmental degradation than other biological molecules, such as proteins. Lastly, DNA is polymorphic, which means there is enough variation between each individual for identification, except in the case of identical twins [1].

Human DNA is composed of coding and non-coding regions. Coding regions are transcribed and translated to form proteins by using tri-nucleotide units (codons), while non-coding regions include intergenic regions and introns. Intergenic regions are found between genes and gene clusters, while introns are the non-translated part of a gene. The non-coding regions can be used for human individualization [1, 3]. For example, the TH01 marker included in most DNA profiling kits is found in the first intron of the tyrosine hydroxylase gene, which is located on the short arm of chromosome 11 [1].

The official introduction of DNA in forensic science came in the year 1985, when Sir Alec Jeffreys developed a technique for human identification that started the field of forensic DNA. He published how the hypervariable regions of DNA called "minisatellites" or "variable number tandem repeats" (VNTRs) could be used for human identification [4, 5]. These regions contain DNA units of 10 to 100 bases long, which are repeated over and over in tandem. Using a combination of multiple restriction endonucleases that cut up long strands of DNA into smaller fragments at specific sites, a multi-band pattern called a "DNA fingerprint" could be visualised. This technique was termed restriction fragment length polymorphism (RFLP) [4, 6].

It would take up to 8 weeks for a sample to be processed and an intact source of at least 50 ng of DNA had to be used [1].

An assay using single-locus probes (SLP) was the first technique widely utilized for criminal investigations. Each probe, revealing a single, highly polymorphic region, simplified interpretation and the conviction of Colin Pitchfork in 1988 was the first criminal case that used this technique [7]. Two separate murders – Lynda Mann in 1983 and Dawn Ashworth in 1986 – near Narborough in Leicestershire were shown to be linked through matching DNA "fingerprints" from semen taken from both crime scenes. The man who initially confessed to one of the murders was exonerated due to non-matching DNA fingerprints. Colin Pitchfork was arrested on 19 September 1987, after he raised police suspicion by hiring another man to give samples in his place [1, 8].

1.3 Saliva as a reliable source of DNA

Saliva, a heterogeneous fluid, is composed of 98% water with the remainder made up of slightly alkaline electrolytes, mucus, antibacterial compounds, and various enzymes, such as lingual lipase, alpha-amylase, and salivary lysozymes. Three pairs of salivary glands – submandibular, parotid, and sublingual glands – secrete a total of one to two litres of saliva per day at the rate of 0.1 mL/min at rest and up to 4 mL/min during periods of active stimulation via both the sympathetic and parasympathetic nervous systems [9]. The functions of saliva are mucosal protection, pH maintenance, microbial control, re-mineralization of teeth, taste mediation and digestion [10].

On average, one mL of saliva contains 4.3 x 10^5 epithelial cells [11], while the number of nucleated cells in an equivalent amount of blood ranges from 4.5 - 11 x 10^5 [12]. A large number of bacteria, averaging 1.694 x 10^7 cells/mL of saliva, seem like a potential drawback for its use as a DNA source. Assessments by mitochondrial DNA (mtDNA) sequencing, PCR-RFLP assays of Y-chromosomal SNPs, and STR fragment analysis confirmed that although the amount of non-human DNA could be substantial in a saliva sample, it does not interfere with successful extraction and amplification of human DNA [13]. The amount of DNA in liquid saliva is approximately 1,000-10,000 ng per mL saliva, while a buccal swab contains about 100-150 ng [14] (Table 1.1). It has also been shown that saliva can be reliably used as a source of DNA for up to 30 days of storage at room temperature [13] as well as in the pellet and frozen forms [15].

Table 1.1. Biological sample types and their DNA content. Adapted from Lee and Ladd [14].

Type of sample	Amount of DNA
Saliva (liquid)	1,000-10,000 ng/mL
Oral swab	100-1,500 ng/swab
Blood (liquid)	20,000-40,000 ng/mL
Bloodstain	250-500 ng/cm^2
Semen (liquid)	150,000-300,000 ng/mL
Postcoital vaginal swab	10-3,000 ng/swab
Plucked hair	1-750 ng/root
Shed hair	1-10 ng/root
Urine	1-20 ng/mL
Bone	3-10 ng/mg
Tissue	50-500 ng/mg

Nevertheless, samples from different individuals give rise to degradation series with different properties [16]. Saliva samples have also been shown to undergo the highest level of degradation when compared to blood and semen due to the presence of various enzymes such as amylases, peroxidases and lysozymes [16, 17]. The high numbers of bacteria further add endonucleases, such as micrococcal nuclease (MNase), to hasten degradation. While it is known that the species and number of bacteria found in saliva differ between individuals [18], their cumulative effects no doubt contribute to the breakdown of DNA.

In addition to being a reliable source of human DNA, saliva can be obtained non-invasively, in contrast to liquid blood samples, which require a trained phlebotomist. Chances of finding DNA in saliva is also high, as the surface layer of epithelial cells are constantly shed and replaced every 2.7 hours on average [11]. Most importantly, the speed of DNA degradation in saliva has been shown to be higher than in blood and buccal swabs [16], which will facilitate the degradation experiment in this study.

1.4 STRs and mini-STRs

Microsatellites, more commonly known as short tandem repeats (STRs), are one of the available DNA marker regions that have gained momentum in the last twenty years of DNA testing. Each repeat unit – usually two to ten bases long depending on the locus – of simple and compound repeats is located adjacent to each other while in complex repeats, the length of the repeat units can vary and intervening sequences can interrupt the repeat units. STRs occupy about three percent of the human genome [1], of which the majority (~92%) are in the non-coding regions [19]. STRs selected for forensic purposes have been shown to have minimal linkages with diseases and are mainly in non-coding regions; these two conditions are deemed important and even required by law in Germany [3]. Various different types of samples have been used to yield DNA profiles with STR technology in the past 15 years and these include, but are not limited to, urine, blood, saliva, hairs, buccal cells on cigarette butts, ancient skeletal remains, and partially charred remains [20-29]. The varied number of repeats from different persons allows for individualization, and with multiple loci analysed at the same time (multiplexing), there is an exponential boost in discrimination power [1]. The amount of DNA required for forensic STR analysis is only 0.5 to 2 ng compared to 50 ng needed for RFLP, owing to the technique of "Polymerase Chain Reaction" (PCR). PCR amplifies the amount of DNA template in a reaction by doubling the targeted region every cycle [30].

Due to their small size ranges (from 100 bp to 450 bp), STRs can still be PCR amplified and detected in degraded samples. This small size range also made multiplexing possible as the same dye could be used on different loci with non-overlapping allele sizes and subsequently visualised at the same time. Another advantage of STRs over larger markers, such as minisatellites which range from 400-1000 bp, is the lower rate of dropout because there is less preferential amplification of smaller alleles [1].

Commercial STR typing kits are available from companies such as Promega Corporation (WI, USA) and Applied Biosystems (CA, USA). The kits currently in use from Promega include PowerPlex® 16 and PowerPlex® Y, while the kits from Applied Biosystems are AmpF*l*STR® Identifiler®, AmpF*l*STR® SEfiler™, and AmpF*l*STR® SGM Plus® (SGM+). These kits are different in the STR loci that they

target and amplify. The current kit used for the UK National DNA Database (NDNAD) is SGM+, which amplifies ten loci: D3S1358, vWA, D16S539, D2S1338, D8S1179, D21S11, D18S51, D19S433, TH01, and FGA plus an additional sex typing locus amelogenin. The combined loci discrimination power of SGM+ is less than one in a billion. Through validation studies, the kit has been shown to be very sensitive and can work with aged and degraded samples [31]. It is also highly human specific and in general, the SGM+ multiplex is robust and reliable for forensic casework samples [31].

On the other hand, the United States Combined DNA Index System (CODIS) requires 13 STR loci, which feature the same loci as the UK-NDNAD except for D2S1338 and D19S433. The five additional loci of CODIS are CSF1PO, TPOX, D5S818, D7S820 and D13S317 [1]. The loci for the two databases are shown in Table 1.2. The main commercial kits presently in use for CODIS loci are Identifiler™ and Powerplex® 16, both of which have been validated for forensic use [32, 33].

Table 1.2. Loci included in the UK-NDNAD (NDNAD) and the US-CODIS (CODIS). A check mark (✓) indicates inclusion and a cross mark (✖) indicates exclusion.

Locus	NDNAD	CODIS	Locus	NDNAD	CODIS
D3S1358	✓	✓	TH01	✓	✓
vWA	✓	✓	FGA	✓	✓
D16S539	✓	✓	AMEL	✓	✓
D2S1338	✓	✖	CSF1PO	✖	✓
D8S1179	✓	✓	TPOX	✖	✓
D21S11	✓	✓	D5S818	✖	✓
D18S51	✓	✓	D7S820	✖	✓
D19S433	✓	✖	D13S317	✖	✓

Six additional STR loci unlinked to the CODIS ones have been made into two multiplexes in order to supplement cases where a high level of confidence is needed or when only a partial profile is obtained with conventional loci. Generating products of less than 125 bp in length, these six loci are D1S1677, D2S441, D4S2364, D10S1248, D14S1434 and D22S1045 [34].

In the case of highly degraded samples, such as those from burnt human remains, remains or stains exposed to the environment, and DNA samples comingled with environmental contaminants, conventional STR kits fail to yield desirable results [16, 35-38]. In order for amplification to proceed successfully, the DNA sequence targeted must be intact for the region bound by both forward and reverse primers [39]. Reducing the size of the PCR products by redesigning the primers to bind as close to the repeat sequence as possible has shown an improved success with these types of sample [34, 36-38, 40]. This is due to the fact that the amount of flanking region in the target sequence that must be intact is kept to a minimum and thus is less prone to random degradation. The minimum size limit of a mini-STR, with additional constraints from polymorphisms directly adjacent to the repeat units, is therefore the repeat units themselves plus the forward and reverse primers (Figure 1.2).

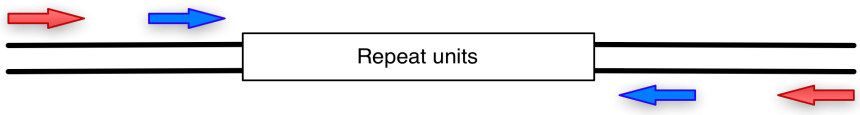


Figure 1.2. Comparison of primer binding of STR primers (red arrows) and mini-STR primers (blue arrows).

Nonetheless, the regions closer to the repeat units are more prone to mutations, of which the common polymorphisms are base polymorphisms, partial repeats, mononucleotide stretches and insertions/deletions [35], therefore, concordance and validation studies are necessary with new mini-STRs.

Larger STR multiplexes (e.g. PowerPlex® 16) move the primers away from the repeat regions in order to avoid overlapping sizes and so one dye can be used for many loci. This allows more than ten loci to be amplified and detected simultaneously for a very low match probability of less than one in a billion. Mini-STR multiplexes – generating products in the regions of 100-200 bp – sacrifice discrimination power for higher success rates in amplifying PCR inhibited or degraded DNA samples [1]. Because most of the amplicons are the same size, only one locus can be tagged with one of the four or five dyes available. This means that a maximum of four or five loci can be amplified and detected simultaneously with conventional capillary electrophoresis.

Size separation of mini-STRs can be aided by the addition of non-nucleotide linkers (NNL), which are oligomeric non-nucleotide linker hexaethyleneoxide (HEO) molecules attached to the primers between the fluorescent dye and the DNA sequence (Figure 1.3) [41]. The NNLs serve to change the mobility characteristics of the PCR products. As a consequence, similar-sized mini-STR products can be visualised as different sizes in an electropherogram (EPG) [1].

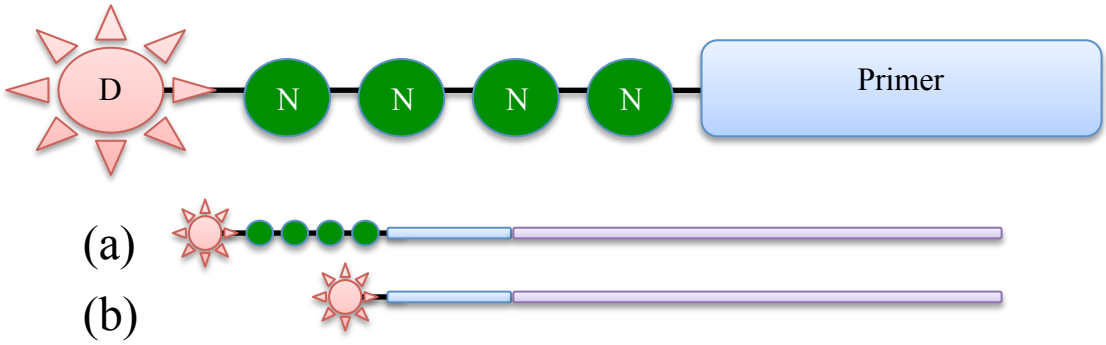


Figure 1.3. A schematic diagram of the addition of NNL. (a) Amplicon with NNLs and (b) amplicon without NNLs incorporated. Note the size and the mobility difference. D = dye and N = NNL.

The MiniFiler™ and the NGM™ kit from Applied Biosystems (CA, USA) utilizes NNLs in order to amplify and visualise many mini-STR loci using five dyes [42, 43]. The primers for CSF1PO, FGA, D16S539, D18S51, amelogenin, D2S1338, D21S11, and D7S820 in the MiniFiler™ kit contain NNLs added during primer synthesis [42].

Mini-STR kits have been shown to achieve better results in forensic casework. They were used for identification of 19,963 human remains from the World Trade Center incident [44, 45]. Other difficult sample types successfully amplified include charred femur [46], buried and exposed femurs and a tibia [47], and human telogen hairs [40].

1.5 Nucleosome theory and forensic science

1.5.1 Structure of chromatin

Nuclear DNA in eukaryotes is packed into structures called chromosomes by association with proteins. The chromosomal proteins, mainly histones and other scaffolding proteins, compact and organize DNA into different structural levels, serving to strengthen DNA during mitosis and meiosis as well as regulate gene expression and activity [48]. Histones are small, highly conserved proteins of roughly 100 to 200 amino acids and are composed of about 20 to 30 percent lysine and arginine, both positively charged, which help in binding to the negatively charged DNA molecules. Association of DNA with five major types of histone molecules – H1, H2A, H2B, H3 and H4 – gives rise to the "beads-on-a-string" structure and then the 30 nm fibre. Two dimers of H2A and H2B plus a tetramer of H3 and H4 make up the octameric histone core that winds 147 bp of DNA in a 1.75 turn [49, 50] (Figure 1.4). These units of histone and DNA – a barrel shape 5.5×10^{-9} metre high and 1.10×10^{-8} metre wide – are called the nucleosomes. In order to be able to wrap around a nucleosome, the DNA must possess some form of flexibility or tolerate bending [51].

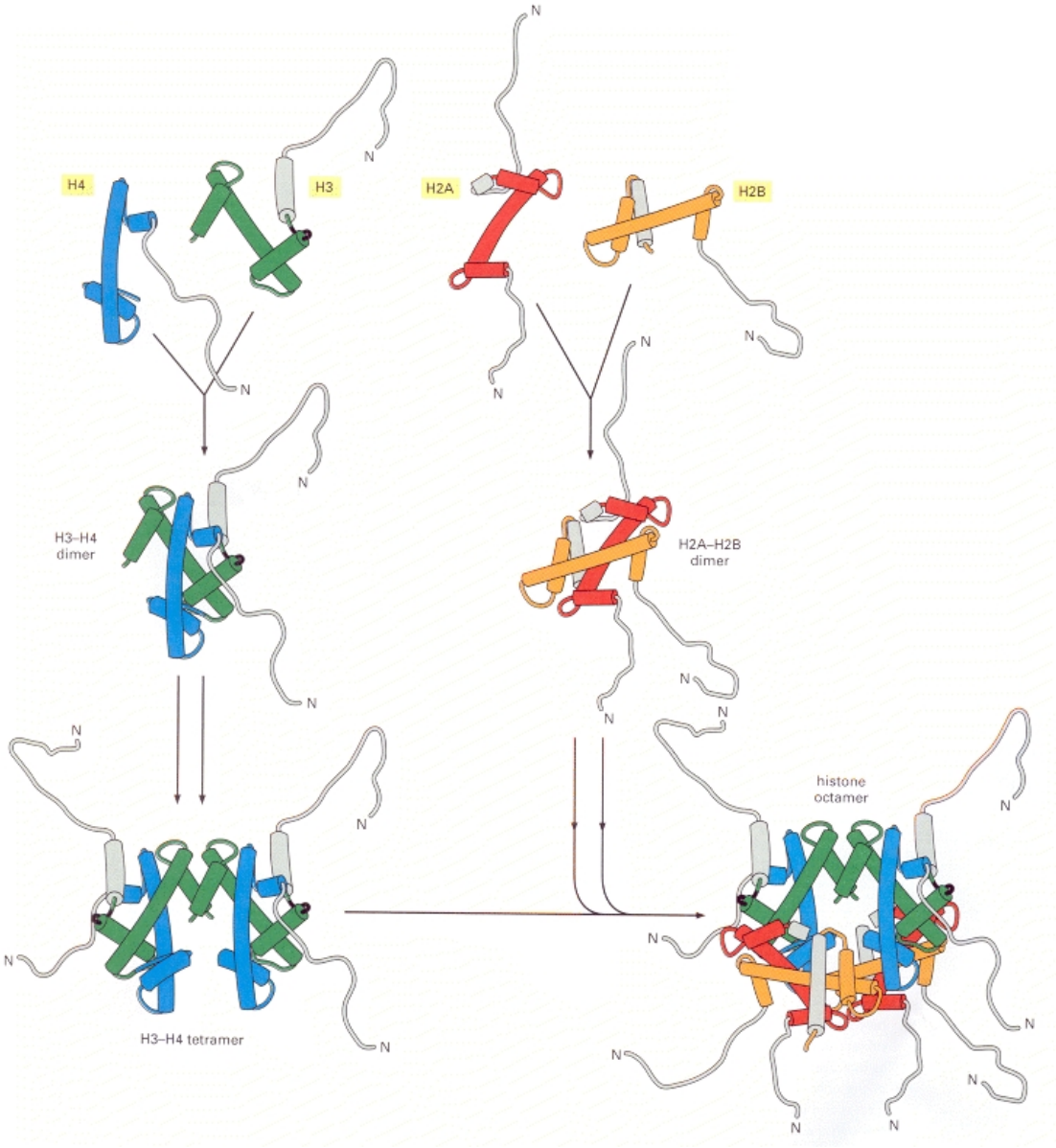


Figure 1.4. Schematic representation of the assembly of the core histones into the nucleosome. Taken from Alberts *et al.* [12].

Addition of H1 histone to the nucleosome stabilizes about 10 to 50 bp of linker DNA, depending on species, between the core particles and further compacts the DNA into the 30 nm fibre [52]. In humans, this linker DNA is on average 40 bp long [53]. The combination of H1 histones, linker DNA, and nucleosomes is termed a chromatosome.

1.5.2 Histone and nucleosome modification

After histones are synthesized, chromatin assembly factors (CAF-1) assemble and direct them to the replication fork by binding to proliferating cell nuclear antigen

(PCNA). More than 150 examples of histone modification are known [48]. These modifications change the properties of histones, one of which reduces their affinity with DNA in order to allow access to genes. For example, histone acetylation and deactylation – attachment/detachment of an acetyl group to 44 available lysine amino acids to the N-termini of each of the core molecules – is accomplished by histone acetyltransferases (HATs) [48].

The structure of the nucleosome was first inferred from nuclease protection experiments in 1973-74 [48]. Nucleosomes can be remodeled by protein complexes such as the Swi/Snf complex via sliding and eviction [54, 55]. These energy-dependent modifications allow the DNA binding proteins to gain access to the DNA sequence. Three main types of nucleosome alterations have been classified. Firstly, remodeling only results in the change of the nucleosomal structure with no change in position. The nature of this process is yet to be discovered but *in vitro* studies led to a doubling in size of the remodeled nucleosome. Secondly, *cis*-displacement or sliding moves the nucleosome along the same DNA molecule. Lastly, *trans*-displacement or transfer brings about the transfer of a nucleosome from one location to a second DNA molecule or non-adjacent part of the same molecule [48].

1.5.3 Nucleosome positioning signal (NPS)

1.5.3.1 Bendability-dependent positioning

Wrapping of DNA around a histone core requires a certain amount of bending in the DNA, therefore, it follows that if a certain pattern or sequence of DNA can tolerate bending better than another, it could be favourable for nucleosome positioning. Recent evidence has shown that the interaction between DNA and histones could be sequence specific [51, 56-63]; contrary to past indications that this interaction is a non-specific association via hydrogen bonding and static interaction with the negatively charged phosphate groups of DNA and non-polar interactions with the deoxyribose sugars [64].

Factors that influence nucleosome positions, in general, are DNA sequence, chromatin remodelers, and non-histone chromatin factors [62]. One important factor for positioning is sequence-dependent, in which DNA curvature and anisotropic nucleosome formation [59], e.g. poly-A and poly-T stretches are stiff while AT

dinucleotides bend easily [62, 65]. According to Rando and Ahmad [62], "*sequences that bend more easily are favourable for wrapping around a histone octamer*". Extensive searches, comparative genomics, and statistical models based on yeast DNA (*Saccharomyces cerevisiae*), chicken DNA, and randomly synthesized n-mers have all revealed that roughly 50 percent of nucleosomes are surprisingly positioned according to predictive models based on DNA sequence [56, 60-62, 66].

Evaluation of 204 DNA sequences regarded as nucleosome forming, derived from the literature by computer algorithms, revealed that AA and TT dinucleotides are repeated with a period of 10.3 (±0.2) bases [57]. A central section – the dyad – is not bent, because there is a lack of both an AA and TT dinucleotide signal (Figure 1.5). This lack of positioning signal at the dyad was independently observed in human DNA by Kogan *et al.* [59].

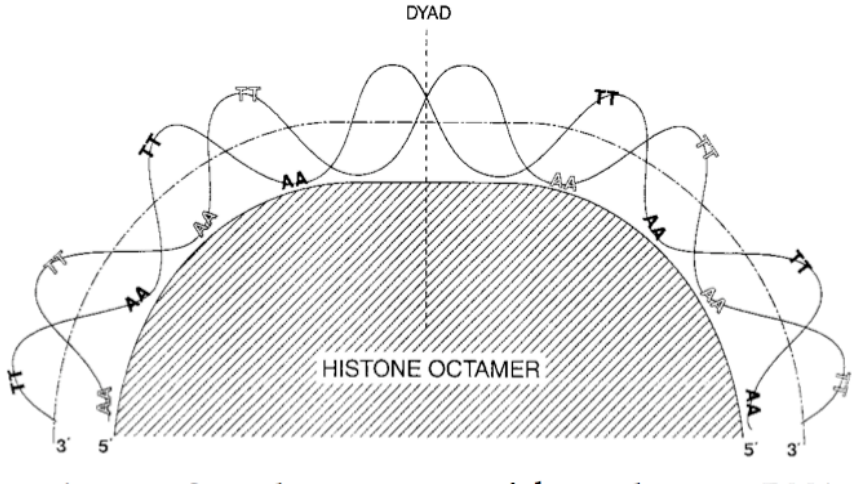


Figure 1.5. Bending of DNA around a histone octamer. AA and TT dinucleotides are spaced periodically, roughly 10 to 11 bp apart. Adapted from Figure 6 in Richmond and Davey [50].

Kiyama and Trifonov proposed that these "dispersed signals" of two to four repeating nucleotides are amplified by their repeated occurrences [58]. In yeast, AA/TT dinucleotides with periodic repeats of approximately 10 to 11 bp of DNA can be used to predict 30-50 percent of nucleosome positions [56, 63]. The reason for this positioning capability was again attributed to the flexibility of AA/TT, allowing bending of DNA around histones. In contrast, Peckham *et al.* used an SVM (support vector machine – a data analysis and pattern recognition method) and found that CG, GC and GG/CC steps are flexible while AT and AA/TT are rigid [61]. Thus, according to them, GC favours and AT inhibits nucleosome formation.

An argument was put forward that a similar rule governing nucleosome positions exists for higher mammals. Segal *et al.* backed this with the inherent nature of their study, in which a computational model built from data derived from experiments with synthesized DNA and chicken DNA successfully predicted almost half of the *in vivo* locations of yeast nucleosomes [63]. Similar strict periodicity has been observed with 13 other organisms, with differences in the dinucleotide pairs depending on the organism [67, 68]. While AA and TT are strong governing signals in yeast and other species, they are only weak contributors to nucleosome positioning in humans [59, 68]. In humans, the NPS is largely dependent on GG and CC dinucleotide repeats every 10.4 base pairs instead of AA and TT, with other dinucleotides being interchangeable – AA, AG, GA and GG oscillating in counterphase with TT, CT, TC and CC [59]. The dinucleotide repeats (GG and CC) oscillate counterphase, which means their locations are shifted about one half-period from one another [59] (Figure 1.5 – AA and TT counterphased). These differences between species suggest that there are other factors besides sequence signals influencing the exact positions of nucleosomes.

One possible explanation of bendability or flexibility of RR/YY periodicity is the preference for strong stacking interaction between purines [69]. Multiple models have been constructed in recent years to predict nucleosome positions based on this bendability pattern taking into account the newer information [56, 57, 63, 65, 67, 70]. The latest universal pattern published is $(GGAAATTTTCC)_n$ [71] based on the universal bendability matrix derived from *Caenorhabditis elegans* [70]. This pattern is deduced from binary positioning signals (strong/weak and purine/pyrimidine) and eleven positioning patterns published from 1983 to 2009.

1.5.3.2 Additional factors affecting nucleosome positioning

Nucleosome positioning is not an entirely fixed event. As discussed in Section 1.5.2, nucleosomes can be moved by an ATP-dependent chromatin remodeling enzyme; for instance, the functions of the Snf2 protein family in budding yeast include nucleosome sliding, nucleosome eviction, histone subunit exchange, and histone variant replacement [62]. Genomes, therefore, tend to retain some flexibility or adaptability regarding the positions of nucleosomes. With the increased complexity in higher eukaryotes, it is likely that additional levels of control over nucleosome

positioning do exist [62]. The *in vivo* assembly of histones by a chromatin assembly factor (CAF) can override the sequence-dependent positioning of nucleosomes [72], and ISWI can move the nucleosome to unfavourable sequences [73]. It is also known that acetylation of histones and methylation of DNA affect nucleosome positioning [55, 74-77], e.g. methylation of CpG dinucleotides hinders the bendability of DNA and thus moves nucleosomes away from such methylated sites [65, 77]

Statistical positioning via physical forces has been proposed as an alternative model to sequence-dependent positioning (Figure 1.6). In the barrier model, only the first (+1) nucleosome is well positioned through the proximity and interactions with transcription factor binding sites and the enzyme Pol II [72, 78, 79]. Other nucleosomes downstream are then juxtaposed to the +1 nucleosome with increasing uncertainty (fuzziness) further downstream [78].

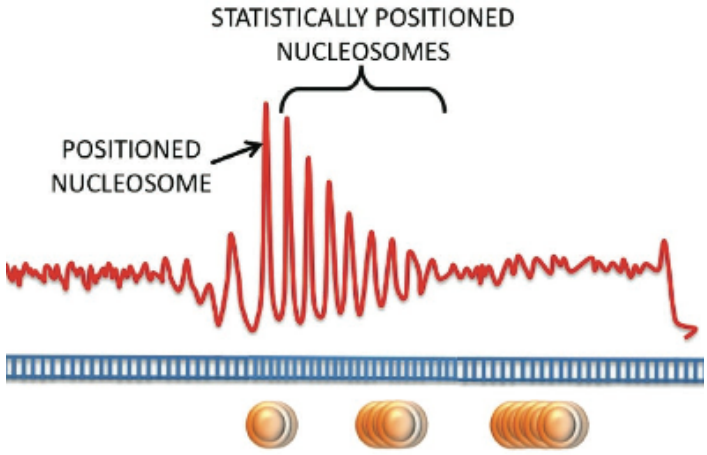


Figure 1.6. Statistical positioning of nucleosomes based on the barrier model. Adapted from Figure 3 in Arya *et al.* [80].

In the absence of Pol II, this statistical positioning has not been observed in the human genome [53]. Moreover, transcription-factor binding sites that are located near each other can induce cooperative displacement of nucleosomes [65], even when the transcription-factors do not interact directly [81]. Cui and Zhurkin [82] even suggested that there might be two separate rules governing nucleosome positions near the promoter regions and those far away from the promoters. Even a low-affinity, unfavourable sequence can be bound as a nucleosome when there is a constraint from other nearby nucleosomes [65].

Reasons for the specific positioning of nucleosomes have been put forward as efficiency and gene regulation [62]. If nucleosomes were specifically positioned, it would help to efficiently form higher order chromatin structure, such as the 30 nm fibre. Being specifically positioned also allows regulation of access to genes or protein-DNA interactions to be controlled by bending and twisting DNA and preventing binding of protein to DNA [50]. For example, promoter or enhancer regions lack NPS and thus are void of nucleosomes so that they remain accessible to DNA binding proteins or transcription factors [58, 63]. High-resolution nucleosome mappings of many species have confirmed that the areas near transcription start and stop sites are generally regarded as nucleosome free [72, 83-86].

Recently there has been an on-going debate regarding this "nucleosome code" based on intrinsic sequence preference [87, 88]. Using the same protocols involving MNase digestions, nucleosome reconstructions and parallel sequencing, the two "schools" disagree on the degree of contribution of the intrinsic DNA sequence to nucleosome positioning *in vivo*. While Kaplan *et al.* [89] proposed that approximately 70% of *in vivo* nucleosomes are positioned based on DNA sequence, Zhang *et al.* limited this contribution to a modest 20% [72]. The root of this debate has stemmed from the ambiguity of the term "nucleosome positioning" and "nucleosome occupancy". Pugh [90] recently clarified the two terms in order to prevent further confusions: "*Occupancy is a measure of histone or nucleosome density*" while "*positioning [is] a measure of the extent to which a population of nucleosomes resists deviating from its consensus location along the DNA*". Nonetheless, what is agreed upon is the complexity and dynamic nature of *in vivo* nucleosome positions (Figure 1.7). This is best explained by Segal and Widom [65]: "*Several decades of chromatin studies collectively show that many nucleosomes change their positions between biological conditions, cell-cycle timing and cell types, whereas the positions of many other nucleosomes remain unchanged*".

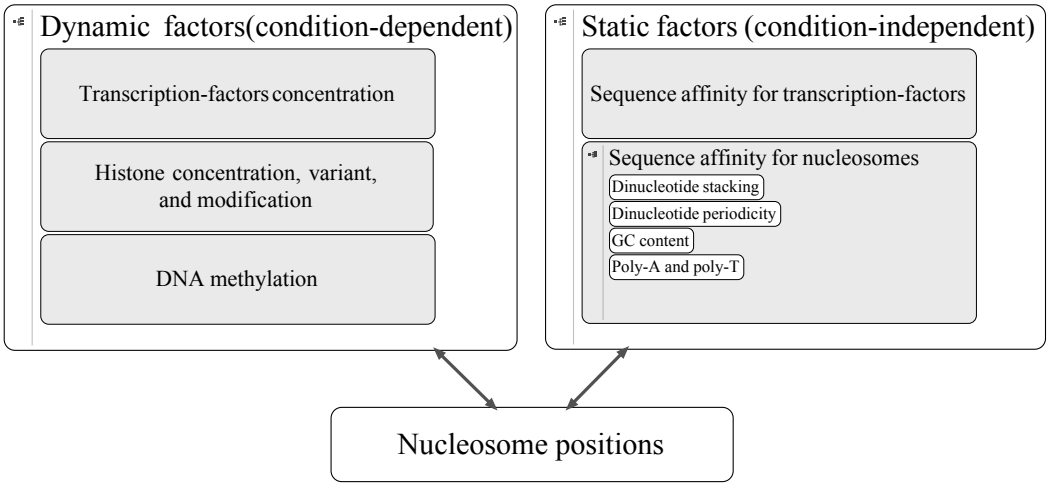


Figure 1.7. Factors that affect nucleosome positions. Dynamic factors are condition-dependent and can be changed according to the needs of each cell or tissue. Static factors, on the other hand, are based on the intrinsic DNA sequence and are consistent for all cells in an organism. Adapted from Figure 1 of Segal and Widom [65].

1.5.4 Nucleosome protection and its forensic implications

Each subunit of the octameric core particle has a specialized function. H2A and H2B dimers do not mainly serve to maintain the structure of the core particle, but possibly to modulate higher order structure, such as the 30 nm fibre. Optical melting showed that each dimer of H2A and H2B binds to and protects about 22 bp of DNA [49]. Additional tetrameric protection from $(H3/H4)_2$ covers about 50 bp in the centrally bound core region with about 10 bp of DNA flanking on both sides [49]. The whole human genome was recently mapped (using a software) for nucleosome exclusion regions, which were correlated with various factors, e.g. tissue specificity, gene density, gene expression levels, and DNase I hypersensitive sites (DHSs) [91]. Of note to this study is the finding that nucleosome excluded regions were correlated with DHSs, which could be inferred that nucleosomes offer some kind of protection to the bound DNA.

Dixon *et al.* [16] suggested that nucleosomes could offer protection to the 147 bp of DNA in the nucleosome core regions against the activity of endonucleases, which would attack DNA at exposed sites. Linker DNA associated with H1 histone, however, is more susceptible to degradation as these regions are not bound and hence not protected [16] (Figure 1.8).

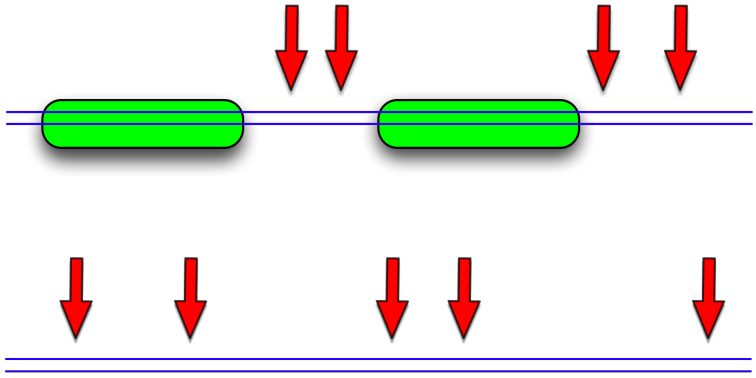


Figure 1.8. Schematic diagram of DNA degradation (a) with the protection of nucleosomes and (b) without the protection of nucleosomes. Green blocks represents nucleosome bounded DNA, red arrows represent attack sites and blue lines represent double-stranded DNA.

With the precise steps of DNA degradation not yet characterized [92], it is a possibility that smaller target sequences less than 147 bp in length could be more likely to remain intact and thus would be an ideal target for amplification and subsequently used for human identification. Identification of the sequences that are more likely to be bound to nucleosomes based on sequence signals would increase the chances of successful PCR amplification. Recent studies used three sequence-dependent approaches, periodic dinucleotide distributions, intrinsic structural properties and universal bendability matrix, to evaluate the nucleosome-forming potentials (NFPs) of any DNA sequence [93-95].

1.6 Degradation of DNA

1.6.1 Cell death pathways

There are two recognized pathways of cell death; apoptosis and necrosis. Apoptosis, otherwise referred to as programmed cell death (PCD), is classified by nuclear condensation and fragmentation into internucleosomal fragments. While the plasma membrane of a cell undergoing apoptosis blebs, numerous apoptotic bodies are formed, which are then phagocytosed and removed by phagocytic cells such as macrophages. Other defining characteristics of apoptosis are exposure of phosphatidylserine (death signal), cytosolic condensation, protein cross-linking and cell shrinkage [96]. The most prominent sign of apoptosis is the lack of inflammation (Figure 1.9). Apoptosis is an energy dependent process and thus ATP is necessary for it to be carried out successfully [97].

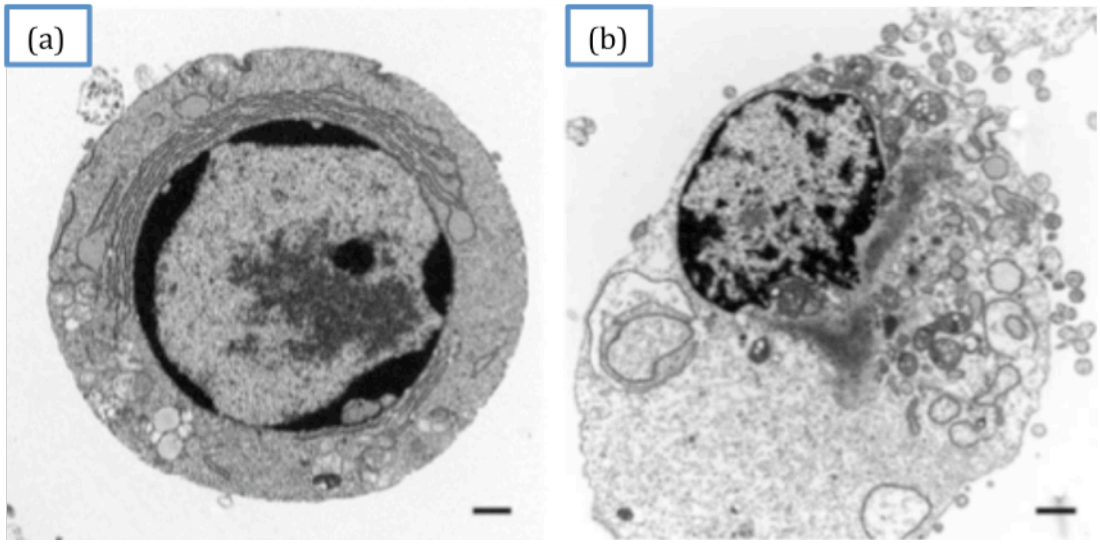


Figure 1.9. Morphological features of a cell undergoing (a) apoptosis and (b) necrosis. The black scale represents 1 μM. Adapted from Edinger and Thompson [97].

Mediation of apoptosis is carried out by a family of enzymes called caspases (cysteine proteases), for which humans carry 13 genes [98]. Two main systems are responsible for DNA degradation in apoptosis: caspase activated DNase (CAD) and lysosomal DNase II [99]. In the first process, CAD is mainly responsible for the breakdown of DNA, and other factors such as endonuclease G and AIF-activated nuclease preferably target the linker DNA before the nucleosome-wrapped DNA. At the completion of this first stage, DNA is degraded into 50 kb to 200 kb fragments. After the apoptotic bodies have been engulfed by macrophages, which recognize the

phosphatidylserine on the cell surface, the DNase II present in the macrophage lysosomes then further digests the nuclear DNA into even smaller fragments of about 180 bp. Because DNA is immunogenic, the two systems working together may be beneficial in infections, as phagocytosis might not be able to clear everything up quickly enough [99].

In order for proper mammalian development to proceed during embryogenesis, apoptosis is crucial. Digit formation and sexual differentiation are two examples of apoptosis at work. Different mechanisms exist for different types of programmed cell deaths – lens differentiation, erythropoiesis and embryogenesis [99].

Another form of cell death is termed necrosis. It is traditionally recognized as a passive form, meaning no energy is required [97]. Energy depletion is one of the causes of necrosis, which is mainly associated with toxicity or physical damage to the cell. Unlike apoptosis, the defining characteristics of necrosis are cytoplasmic vacuolation, plasma membrane breakage and inflammatory response induction in the vicinity via the release of cellular contents and proinflammatory molecules (Figure 1.9), which could serve to fight microbial infection as well. DNA damage, especially in a proliferating cell, could also trigger necrosis via the DNA repair protein PARP. Obviously, this reduces the risk of amplifying mutation [97]. In necrotic cells, cellular DNA has been observed to fragment into about 200 bp units [100], which is roughly the size of degraded DNA in both artificially degraded environments and casework samples [16, 34, 35, 38, 101]. Therefore, it is possible that the same mechanisms that exist for necrotic DNA damage could play a role in DNA stability of dried stains, as after death the cells cease to function. Determination of whether a cell undergoes apoptosis or necrosis depends on many factors, including intracellular ATP levels [102].

Autophagy (from Greek "*autos*" and "*phagia*" meaning "self-consumed"), otherwise known as "programmed necrosis", is another process related to cell death. In a condition of energy crisis where normal functions cannot be maintained, cellular components are used as energy source. It is essentially an alternative death pathway when apoptosis is disabled, but its main purpose is actually a survival mechanism [97].

1.6.2 Factors that affect DNA degradation

Sunlight, moisture, temperature and growth of microorganisms could reduce the amount of intact DNA available for PCR by accelerating DNA degradation or inherently inhibiting the amplification reaction [101]. The first and foremost factor that degrades DNA is the cellular (cytosolic) nucleases themselves [103]. These cytosolic nucleases, whose main function is to attack and digest foreign sources of DNA, work non-specifically and nuclear DNA suffers the same fate once they have encountered these enzymes. Bacterial nucleases, produced and released by both normal flora (microorganisms co-existing internally and externally in a human body) and those in the environment, also degrade DNA [39, 104]. Other sources of bacterial nucleases are from *Staphylococcus spp*, *Streptococci spp* and members of the Enterobacteriaceae family, which penetrate the gut wall of a human corpse twelve to fifteen hours after death [105].

The DNA fragments remaining after digestions from cellular then bacterial nucleases are further subjected to chemical degradation [106] (Figure 1.10). Hydrolysis at the glycosidic bonds between the sugar and the base, especially of purines, results in an abasic site that will undergo phosphate-deoxyribose backbone cleavage. Hydrolysis can also cause deamination of DNA bases with single amino groups, such as adenine, cytosine, 5-methylcytosine, and guanine, changing the bases to hypoxanthine, uracil, thymine, and xanthine, respectively [107]. The rate of hydrolytic degradation has been shown to be slower in dry state when compared to hydrated state [108].

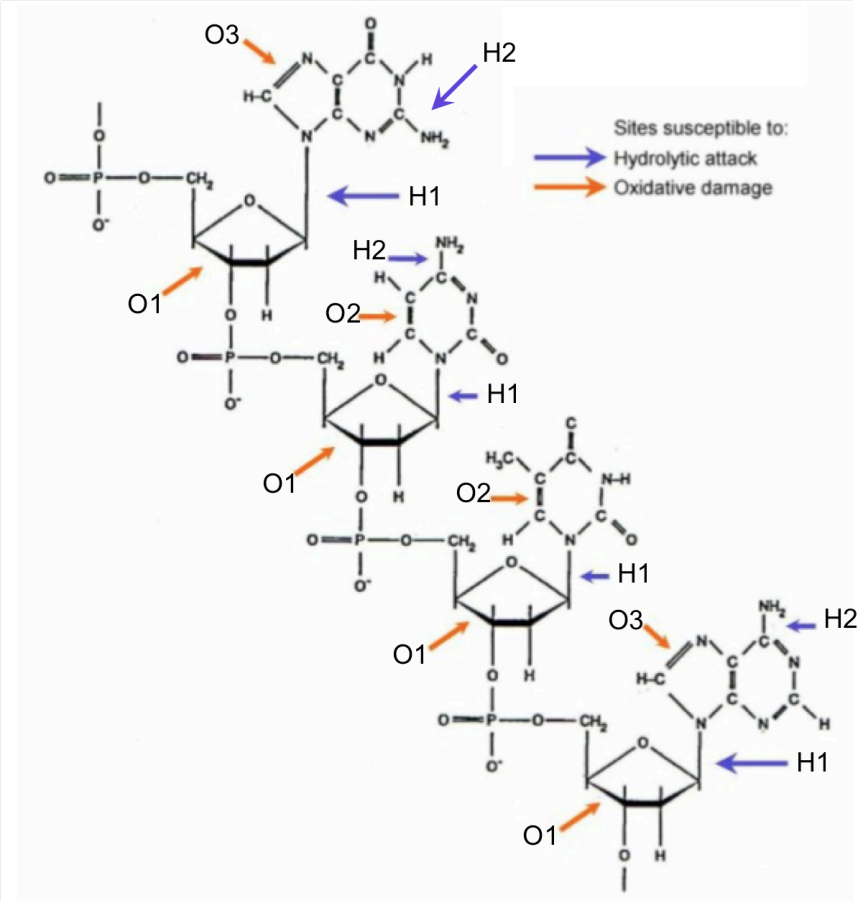


Figure 1.10. Chemical degradation sites of DNA. H1's are glysodic bond sites and H2's are demination sites. O1's are 3'-4' carbon bond sites, O2's are double carbon bonds of pyrimidines, and O3's are imidazole ring targets. Adapted from Lindahl [107] and Dixon [109].

Oxidative damage is another chemical alteration that attacks DNA (Figure 1.10). The 3'-4' carbon bond of the deoxyribose sugar can be oxidized, fragmenting the ring structure and leading to strand scission. Another main attack site is the 5'-6' double carbon bond of the pyrimidines and the imidazole ring found in purines, which also lead to ring fragmentation [107].

UVC radiation has also been shown to induce single-strand breaks in dried bloodstains as the main pathway of degradation, but the possibility of additional damage from double-strand breaks and DNA-DNA cross-linking cannot be excluded [110]. Cross-linking between apurinic and apyrimidinic sites of opposite DNA strands can also happen in a temperature-dependent manner [111].

The rate of DNA degradation differs depending on the cell type, tissue type and activity [17, 112], but in general the tissues with higher activity have a faster rate of

DNA degradation [92, 112]. The degradation process in different forensic stains and samples has not been characterized to date [92].

1.6.3 Micrococcal nuclease

Micrococcal nuclease (MNase) possesses both endonucleolytic and exonucleolytic activity. Produced by *Staphylococcus aureus*, this calcium-dependent enzyme attacks the 5'-phosphodiester bond of both single-stranded and double-stranded DNA [113]. The mechanism of attack is hydrolysis, which yields 3' mononucleotides and oligonucleotides and a free 5'-hydroxyl group [114]. MNase preferentially digests linker DNA, cleaving it to produce multimers of nucleosomes and then monomers of nucleosomes plus linker DNA. The weak exonucleolytic activity then digests the ends of the linker DNA from the 160 bp unit to produce a final unit of 147 bp core nucleosome of protected DNA [62, 114]. Extended digestion times can result in fragments of DNA in the range of 50 to 70 bp [49] by cutting off 10 bp of DNA at a time [115].

Although MNase is relatively non-specific, experiments with mouse satellite DNA showed that 5'CATA and 5'CTA were preferentially attacked, especially when they were adjacent to a GC-rich region. In general, MNase prefers an alternating sequence of A and T preceded by C or G [115]. Specific cleaving based on sequence preference is more pronounced at lower temperatures due to exonucleolytic capability being suppressed [116].

1.6.4 Deoxyribonuclease I

Deoxyribonuclease (DNase) is a family of enzyme that non-specifically digests both double-stranded DNA (dsDNA) and single-stranded DNA (ssDNA) to yield 5'-phosphorylated di-, tri-, and oligonucleotide products [117] via hydrolytic cleavage of phosphodiester bonds. The activity of DNase I depends on magnesium, manganese, and calcium ions [118, 119]. DNase I is often used to obtain artificially degraded DNA for forensic studies [120-122] as well as for preparation of purified RNA for reverse-transcriptase PCR [123]. Preparation of artificially degraded DNA product is rapid in the early phases followed by a plateauing effect in the later stages [101]. DNase I can be used by itself [120] or in combination with other degradation techniques such as sonication [101].

1.7 Real-time quantitative PCR (qPCR)

1.7.1 Real-time analysis vs. endpoint analysis

PCR can be used to quantify the amount of initial DNA in a sample in a process of quantitative PCR (qPCR) [124]. Measurements can be taken at each cycle (real-time analysis), i.e. at the end of each annealing step or extension step in the cycle [125], or it can be taken at the end of the whole PCR process (endpoint analysis) (Figure 1.11). Real-time measurements are more accurate, as they are taken during the exponential phase of the reaction, which is less influenced by limiting reagents, small differences in reaction components, and cycling conditions [126]. The rationale behind using real-time analysis is that the exponential gain phase, in which a linear relationship is observed between log 10 of starting template copies and the number of cycles, happens in the early to mid cycles of PCR [126]. Moreover, the effect of sample variation on amplification efficiencies remains small when the concentration of the products remains low [127].

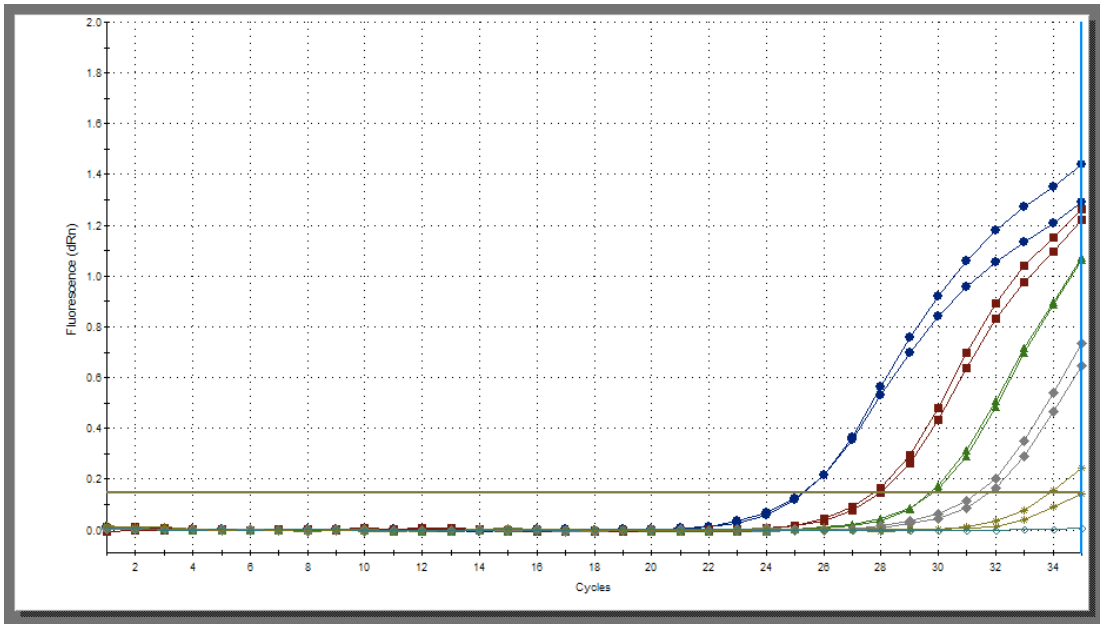


Figure 1.11. The amplification plot (normalised) of a five-fold serial dilution of DNA in duplicate over approximately three orders of magnitude. Note the deviation in the endpoint value of blue lines and lack of repeatability at very low starting template (yellow lines).

Endpoint analysis, on the other hand, is inherently inaccurate because endpoint data are affected by the aforementioned factors. Further inaccuracy in quantification data can also arise from the accumulation of inhibitors. Furthermore, chemical analysis shows that at the end of a typical PCR, there are plenty of dNTPs and primers left.

The real reason for slowing down of amplification in the late cycles is dsDNA inhibiting DNA polymerase, resulting in early saturation. Polymerase tends to bind the amplicon duplex rather than the small quantity of primer-target duplex at the later stages of amplification [128]. The advantages of real-time analysis over endpoint quantification include the capability to measure over a wide dynamic range, high accuracy, and good sensitivity [129].

1.7.2 Real-time analysis data and interpretation

qPCR can measure both in absolute quantity or relative quantity, i.e. how many folds relative to another target. Absolute quantification requires good optimisation prior to use, i.e. PCR efficiency must be similar across a range of samples and, more importantly, in the standards. The quantification cycle (C_q) is used in qPCR to determine the initial concentration of a sample. Other non-standard abbreviations of quantification cycle devised by different manufacturers include threshold cycle (C_t), crossing point (C_p) and take-off point (TOP) [130]. C_q is the cycle at which the fluorescence is statistically significantly different from the background noise [130] determined by various algorithms dependent on the qPCR machine manufacturer. For instance, Stratagene's instruments (Cedar Creek, Texas) threshold can be set by base-line method, e.g. three standard deviations above the baseline mean, or by determining the second derivative maximum for each curve describing fluorescence versus cycle [129].

C_q is inversely proportional to the log of the initial quantity of DNA template [131], which means that the C_q value will be low when the initial template amount is high. With a small amount of starting template, it takes more cycles of PCR to generate enough product, and subsequently fluorescent signal, to cross the quantification threshold [131]. In combination with a calibration curve generated from a set of known standards, absolute quantification can be achieved by comparing the C_q value of the unknown samples against those from the calibration curve. Generating a calibration curve requires the target sequence of samples with known concentrations to be amplified [130]. The concentrations of the standards should encompass the range that is expected to be found in the experimental tubes, i.e. 10 ng to 10 pg of DNA in standards cannot be used to accurately quantify an unknown sample of 2 pg. Extrapolation of the C_q value when the C_q value of an unknown sample does not fall

within the full range of the standards is known to compromise the accuracy of the test [129].

Certain requirements are taken into account to properly construct a calibration curve. A triplicate over several orders of magnitude is recommended [129]. A high R-squared value and efficiency must be observed, as these indicate a linear dynamic range and doubling of PCR products at every cycle [130] (Figure 1.12). While running a set of standards with every assay is desirable, it can be costly. Periodically constructing a new set is acceptable, but a highly accurate, quantified standard must be used to determine if there is any "drift". Freeze-thaw cycles should be avoided for DNA standards as much as possible and only dilutions freshly prepared from a stock solution should be used [129].

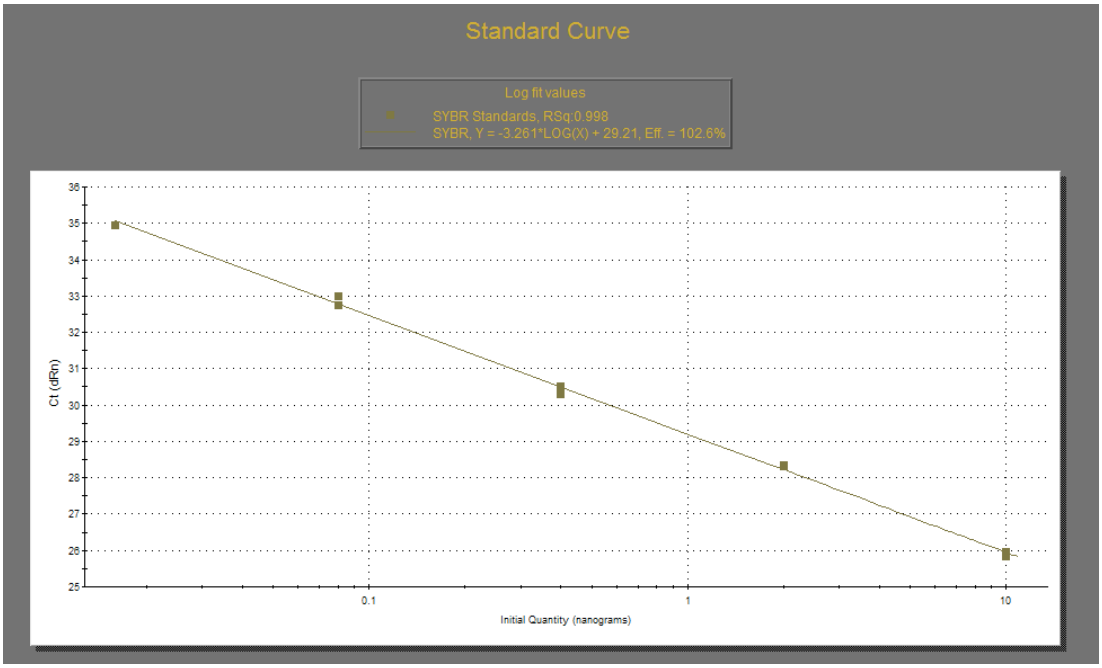


Figure 1.12. A calibration curve generated from 10 ng to 0.016 ng of DNA in duplicate. Note a high r^2 value of 0.998 and a satisfactory efficiency of 102.6%. The y-axis is shown as Ct due to Stratagene software using a non-standard abbreviation of threshold cycle instead of quantification cycle [130].

qPCR has a dynamic range of up to 10^{10} copies of DNA template, but the linearity of the assay will break down at a low template level [124, 131]. Artefacts such as double-stranded primer-dimers can produce an apparent increase in fluorescence when an intercalating dye is used, while in fact the efficiency is decreased because resources are depleted as artefacts accumulate. The accuracy of the assay depends on the base-line noise and noise-signal ratio, both of which depend on the chemistries

and machine used [130]. Variations in sample, albeit small, could have a cumulatively significant effect on the amplification efficiency. This problem can be overcome or minimized by working with many replicates of a sample [127].

The reaction efficiency of an assay is estimated from the slope of starting template versus PCR cycle. This value should be as close to 100 percent as possible, and is often accepted when above 95 or 98 percent, depending on the type of work being carried out [132]. In theory, a log 10 increase should occur approximately every 3.32 cycles (PCR efficiency = $10^{(-1/slope)} - 1$) [130]. The reason for an efficiency of above 100 percent is difficult to specify, as this could be the result of degraded DNA standard, poor optimisation, template secondary structure, and even inappropriate temperature settings. A reduced efficiency is generally observed due to mismatched base incorporation, ultimately resulting in early chain termination [127].

While the presence of inhibitors - substances that reduce the activity of polymerase or compete with target template for resources – initially has no net effect because every reaction component is added in excess; accurate quantification is not possible if inhibitors have an effect on the PCR before the threshold cycle is crossed. For this reason, it is important to extract and purify template carefully. Controls can be used to reveal the presence of inhibitors [130]. Addition of another PCR target, or "spiking", can act as an external control [133], while modifying the inter-primer region of the original target and detecting with a differently coloured probe can serve as an internal control [127].

1.7.3 Detection of target

1.7.3.1 Intercalators

Currently there are many detection systems available for use with qPCR, with the most commonly used being SYBR® Green I, hydrolysis probes (TaqMan®), Molecular Beacons and Scorpions. SYBR® Green I is an intercalating dye, which integrates itself into the minor grooves of any double-stranded DNA. SYBR® Green I, an asymmetrical cyanine dye (Figure 1.13) absorbs blue light (λ = 497 nm) and emits green light (λ = 520 nm). In the intercalated state, its fluorescence increases about 1000-fold [134]. It is often recommended for preliminary qPCR experiments

due to its high signal to noise ratio leading to a smooth amplification/detection curve as well as eliminating the need for probe optimisation.

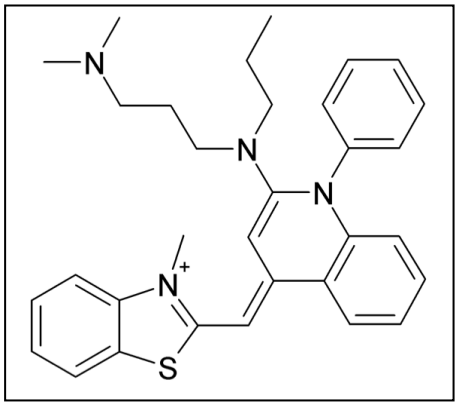


Figure 1.13. Chemical structure of SYBR® Green I dye [2-[*N*-(3-dimethylaminopropyl)-*N*-propylamino]-4-[2,3-dihydro-3-methyl-(benzo-1,3-thiazol-2-yl)-methylidene]-1-phenyl-quinolinium] [134].

SYBR® Green I is used to monitor the accumulation of PCR products in qPCR during the annealing and elongation steps, in which double stranded products are formed and extended, allowing the SYBR® Green I dye to bind (Figure 1.14). For long amplicons, fluorescence data collection is recommended at the elongation step [129], as this allows more dyes to intercalate into the products.

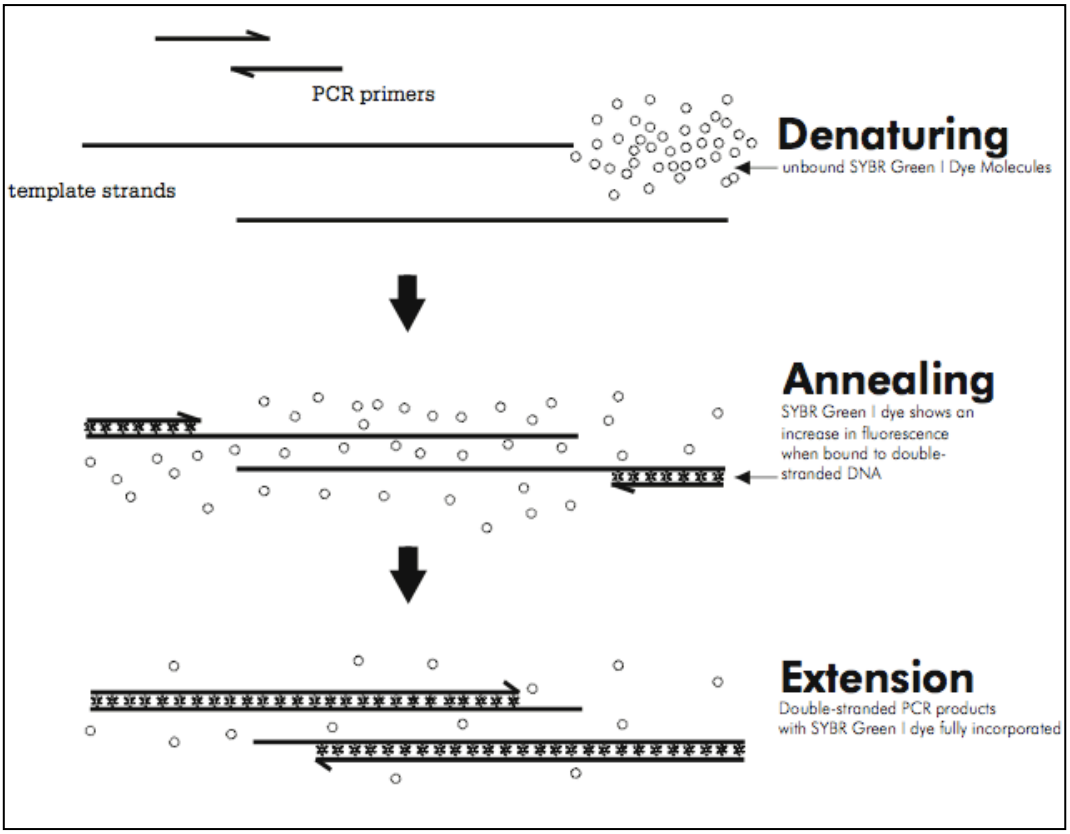


Figure 1.14. The mechanisms of qPCR utilizing the SYBR® Green I detection system. Replicated from Brilliant® SYBR® Green qPCR Master Mix Manual [127].

A problem with SYBR® Green I is its non-specificity, as it intercalates into any double-stranded DNA from any source, including animals and bacteria [135]. Multiplexing is not recommended with SYBR® Green I as it is not specific. It can be used with multiple products that have at least 5°C difference in temperature, although this should only be done for a preliminary study before advancing to other more specific detection methods such as hydrolysis probes.

Different products, non-specific dsDNA such as primer-dimers, and spurious PCR products are detected, but can be distinguished from the each other by dissociation curve analysis [136]. Dissociation curve analysis is performed by using a step-wise increase in temperature from approximately 55 to 95°C while monitoring the decrease in fluorescence from the unbinding of SYBR® Green I dye to the DNA as the dsDNA dissociates to be single-stranded [129] (Figure 1.15).

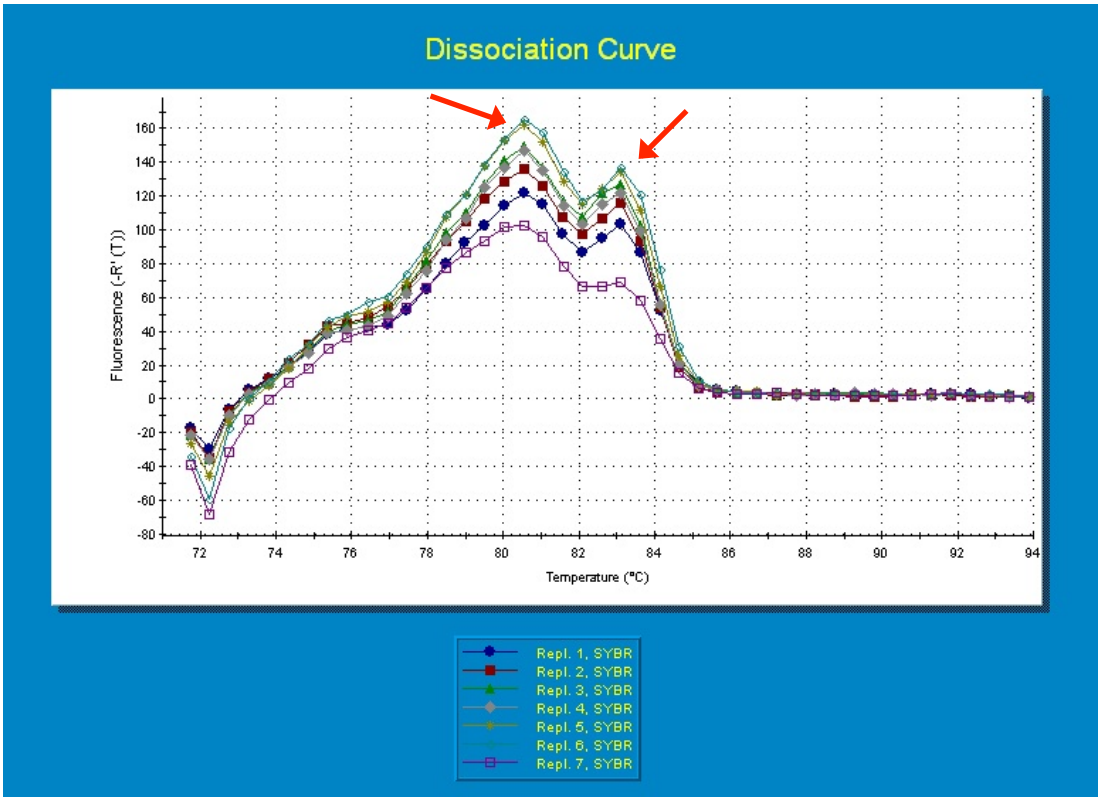


Figure 1.15. A dissociation curve with two distinct melting peaks, suggesting that two products are detected in this qPCR experiment.

Interpretation is performed by graphing the negative first derivative (-R'(T) or -Rn'(T) if a normaliser has been used) against temperature. Primer-dimers and non-specific products usually melt at a lower temperature than the desired products and will show up as an extra peak in the graph [137] (Figure 1.15). In general, if more than one peak is seen in the dissociation curve in a singleplex assay, it is certain that non-specific products are present in that reaction. These profiles can then be used to optimise the assay by adjusting the relevant parameters. For example, if a primer-dimer melts at 70°C, the temperature in the extension step can be raised to 72°C or 74°C. While this does not interfere with the ability of the *Taq* polymerase, there is a higher chance that the primer-dimer will dissociate during the extension step. With SYBR® Green I, the specificity of the assay depends solely on primer specificity. Therefore, primer purification and careful primer design to prevent primer-dimer, secondary structure formation, and non-specific binding can minimize the effects of "side-reaction products" [129].

1.7.3.2 Probes technology

Probe-based qPCR methods are specific to each target, which means that in the absence of a target, the probe will remain quenched and no fluorescence will be detected. Linear hydrolysis probes, such as TaqMan®, are labeled with a fluorescent dye at the 5' end and a quencher attached to the 3' end. Historically TAMRA is used as a universal quencher, but now dark quenchers, such as a Black Hole Quencher (BHQ) are more widely utilized [138]. In its intact state, the fluorophore and the quencher are in close proximity and consequently no fluorescence can be detected based on fluorescence resonance energy transfer (FRET). FRET, first described in 1948, is a process in which an excited fluorophore transfers its energy via a non-radioactive, dipole-dipole coupling mechanism to a nearby (less than 1 x 10^{-8} m) acceptor fluorophore or chromophore [139, 140]. These chromophores in modern probe technology are replaced by a quencher, a dye that accepts energy of overlapping wavelength with the fluorophore but does not result in fluorescence itself (Figure 1.16). Multiple probes can be labeled with different dyes and quenchers; hence, multiplexing is possible [129].

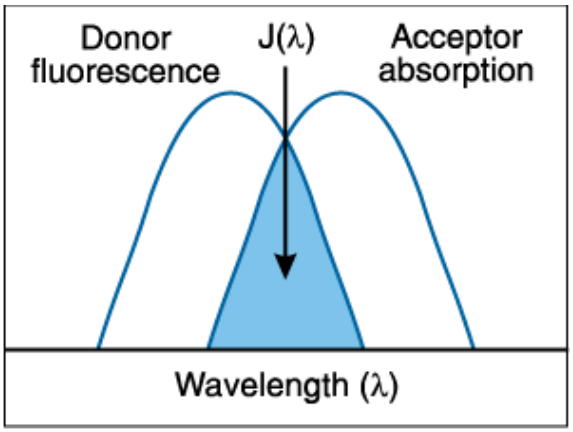


Figure 1.16. Spectral diagram of a fluorophore and a quencher (acceptor). Replicated from Haughland [141].

Probes are designed to bind to a single-stranded DNA sequence downstream from the primer. During extension by *Taq* polymerase, the 5' to 3' exonuclease activity of the *Taq* enzyme will digest the 5' end of the probe leaving the fluorophore free in solution, resulting in an increase in fluorescence. With linear probes, PCR is often performed using two-step reactions, comprising of a denaturation step at 95°C and a combined annealing/extension step at 60°C. If the normal extension temperature of

72°C is used, the probes are usually strand-displaced (no hybridization with target) and thus probe digestion will not occur [138].

Structured probes such as Molecular Beacons can also be used with qPCR chemistries. They form a secondary structure, usually a hair-pin loop, that increases specificity to the target region. The fluorophore and the quencher are located on each end of the two arms, which are immediately adjacent to each other due to the loop structure [142]. If the target sequence is available, these probes will bind to the target and lose their self-complementarity, otherwise, they energetically favour the stem-loop conformation over non-specific binding. Binding to a target in turn separates the fluorophore from the quencher and hence fluorescence can be detected (Figure 1.17).

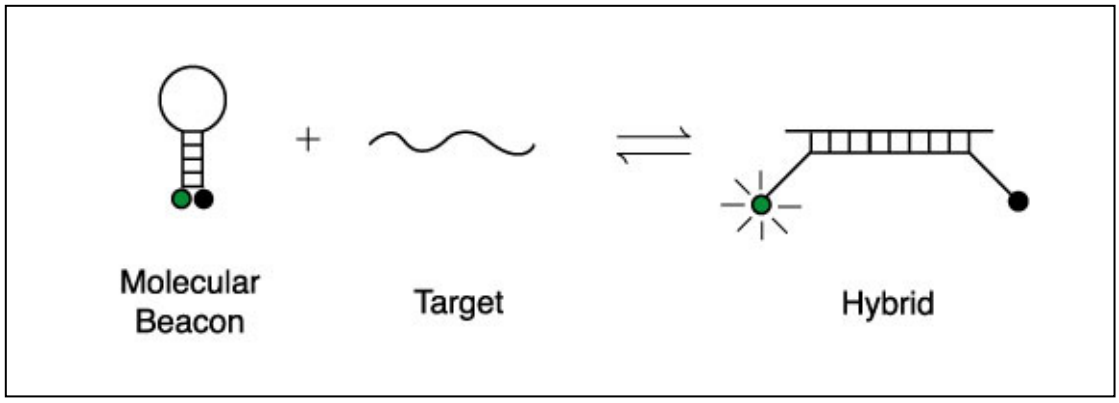


Figure 1.17. A schematic diagram of a molecular beacon and its working mechanism. Green dots represent fluorophores and black dots are quenchers. Replicated from Haugland [141].

A molecular beacon fluoresces when the hairpin loop structure is disrupted and so there is no need for the fluorophore to be liberated from the probe as in TaqMan® technology. Since no hydrolysis is needed for a molecular beacon, a traditional three-step qPCR or a two-step qPCR can be performed [143]. However, in order to obtain that specificity, designing is critical and extensive tests must be carried out to ensure that they behave as expected [129].

1.8 Statistical analysis

Selecting appropriate statistical methods for analysing data is of extreme importance in any scientific work. Exploratory data analysis must first be performed to get the general descriptive statistics and overview of the trends and patterns in a dataset, which defines the type of further hypothesis tests. Parametric methods assume that the samples come from a normally distributed (Gaussian) population, that the variances are homogeneous, and that the errors are independent. Compared to non-parametric methods, which are distribution-free, the parametric methods have more statistical power, but can only be applied in limited cases.

To determine whether the data follow a normal distribution, the Shapiro-Wilk test [144] can be used. The null hypothesis for the Shapiro-Wilk test states that the samples are normally distributed. A p-value smaller than the alpha-value is evidence for rejection of the null hypothesis. In the case of non-normal data, the homogeneity of variances of the different loci can be determined using the Fligner-Killeen (median) test, which is robust against departures from normality [145]. Similarly, a small p-value indicates that the null hypothesis can be rejected.

When the samples are related, the Friedman's two-way analysis of variance (Friedman's test) [146] can be used to determine whether the samples in each group are different. Ranking is done for each sample across the different groups, and the distributions of the ranks in each treatment group are then compared. A small p-value in the Friedman's test is grounds for rejection of the null hypothesis, suggesting that the groups are different.

Determination of correlations can be done using Spearman's rank correlation methods (Spearman's ρ). This non-parametric method measures the degree of association between two variables without any assumption on the distribution of the variables being investigated, and the relationship between the two variables does not need to be linear, as in the case of Pearson's correlation [146]. A positive correlation coefficient indicates that as one variable increases, the other also increases. A strong correlation is supported when the coefficient is close to one or negative one, while a weak correlation is supported when the coefficient is close to zero.

1.9 Aims

The main objective of this study is to determine the protective capabilities of nucleosomes against DNA degradation in degraded DNA samples. This will be carried out in four main stages.

- A robust and sensitive method to quantify the amount of DNA in degraded samples will be identified. Candidate targets include *Alu* elements, single-copy and multi-copy loci. A comparison with a commercially available quantification kit will also be made. An optimisation process will be studied in order for subsequent qPCR assay optimisations.
- The performance of various next-generation and mini-STR kits will be evaluated to provide a basis for an informed decision in choosing kits to adopt for laboratory use.
- A method to degrade DNA quickly while preserving the chromatin structure naturally will be identified and assessed. This method should provide degraded samples that can be used to differentiate between a locus that will have been protected from degradation by nucleosomes and one that will not have been protected.
- Forensically important STRs will be evaluated for their nucleosome forming potentials (NFPs) using computer software. Primers will be designed for selected loci and a comparison of their performances on artificially degraded DNA samples and simulated casework samples will be carried out.

If some STR loci could withstand degradation better due to the protective effects of nucleosomes, further work could be carried out to multiplex them into a specialized "degraded sample kit", designed specifically to increase the chances of being able to genotype highly degraded samples. Alternatively, they could be kept as high molecular weight loci while the low molecular weight loci could be ones that are more discriminatory but less able to withstand degradation. This means that a single DNA profiling kit can be highly informative due to the low molecular weight loci (mini-STR), and at the same time the probability of allelic dropout in the high molecular weight loci is minimized, thereby overcoming the limitations of current STR and mini-STR kits.

2 Investigation of quantification methods using real-time qPCR

2.1 Introduction

During the past few years, both singleplex and multiplex qPCR assays have been developed to quantify and assess the degree of degradation in forensic samples [147-153]. These assays are advantageous because they attempt to quantify 'amplifiable DNA' instead of total human DNA, and hence they serve as a guide to what downstream protocols are most appropriate. The loci employed by these studies include STR loci (TH01 and CSF1PO), amelogenin, mtDNA and human specific *Alu*-elements. The amplicons are generally between 60 and 200 bp. The most recent efforts have resulted in two commercial kits validated for forensic use, the Quantifiler® Duo and Plexor® HY [121, 122], although with their own caveats, limitations, and lack of reproducibility between the two kits [154].

To assess the degree of degradation in DNA samples, the ratio of amplified DNA from two targets can be used. The ratio of mtDNA to nuclear DNA was used in early studies [147, 151], but more recently the ratio of two nuclear DNA targets, namely CSFP1O and TH01, was employed [149, 150]. This concept is based on the different amplicon size for the two targets. Shorter amplicons are more likely to be found intact in degraded samples. As a result, the ratio of the quantity of the short amplicon over the long amplicon provides a rough estimate as to the amount of degradation of the template DNA [150].

To determine the degree of inhibition, a synthetic internal PCR control target sequence (IPC) can be included in the multiplex. In the case of degraded DNA, the IPC should still be amplifiable and its signal detected, while in the presence of inhibitors both IPC and template DNA amplification should be hindered, resulting in a dampened signal for both [149, 150].

2.1.1 *Alu*-elements

Alu-elements have been selected as a potential target in this study due to their abundance in nuclear DNA. For this reason, they would be most appropriate for quantifying artificially degraded DNA in the later stages of this work. With a consensus sequence of 280 bp and various different subspecies, up to one million

copies of *Alu*-elements exist in a single nuclear DNA copy [155, 156]. Subspecies that are specific to human (Yb8 and Yd6) have been identified and utilized for forensic purposes with an extremely low sensitivity of 0.001 ng of DNA [148, 153]. The downside of *Alu* quantification is the highly stringent PCR conditions required. Non-specific products are almost unavoidable as multiple subspecies with similar sequences are available [148].

2.1.2 Single-copy and multi-copy loci

In contrast to *Alu* elements, there are only two copies of a single-copy gene in a diploid cell. This means that a quantification assay targeting them can be reliably used to quantify STR loci, as the copy number of both a single-copy and an STR locus is identical. Unlike using STR loci as targets [149, 150], variation in the length of the coding region of a single-copy gene in humans is rare; hence, quantification results calculated from calibration curves are more accurate because there is no difference in product size between the DNA standards and unknown samples. Furthermore, the length variation of an STR target, e.g. the nuTH01 target varies from 170-190 bp [149], can lead to multiple products detected with the SYBR® Green I dye. It can also affect the amplification efficiency of the assay [122]. A forensic quantification kit that uses a single-copy gene as a target is the Quantifiler® Duo by Applied Biosystems [122].

Housekeeping genes are genes expressed in all cell types. Their main functions are usually related to basic cellular maintenance and sustenance [157]. Due to these basic functionalities, a housekeeping gene is usually highly conserved and can have pseudogenes – genomic sequences highly similar to the parent gene that have limited functions due to insertion/deletion and nonsense mutation [158]. Targeting the sequences found in both the parent gene and pseudogenes can result in a multi-copy locus quantification assay. As housekeeping genes are popularly used in gene expression qPCR studies [159], 13 housekeeping genes and two other single-copy genes (ribonuclease P RNA component H1 and telomerase reverse transcriptase) were evaluated using Primer3 [160]. Four potential candidates that could be used for quantification of degraded DNA were identified (bolded in Table 2.1).

Table 2.1. Fifteen candidate genes evaluated for further primer design and optimisation.

RefSeq ID	Description
NM_198253	**Homo sapiens telomerase reverse transcriptase (TERT)**
NM_000034	**Homo sapiens aldolase A,fructose-bisphosphate (ALDOA)**
NR_002312	**Homo sapiens ribonuclease P RNA component H1 (RPPH1)**
NM_002046	**Homo sapiens glyceraldehyde-3-phosphate dehydrogenase (GAPDH)**
NM_00110	Homo sapiens actin, beta (ACTB)
NM_000291	Homo sapiens phosphoglycerate kinase 1 (PGK1)
NM_005566	Homo sapiens lactate dehydrogenase A (LDHA)
NM_002954	Homo sapiens ribosomal protein S27a (RPS27A)
NM_000981	Homo sapiens ribosomal protein L19 (RPL19)
NM_000975	Homo sapiens ribosomal protein L11 (RPL11)
NM_007363	Homo sapiens non-POU domain containing, octamer-binding (NONO)
NM_004309	Homo sapiens Rho GDP dissociation inhibitor (GDI) alpha (ARHGDIA)
NM_000994	Homo sapiens ribosomal protein L32 (RPL32)
NM_022551	Homo sapiens ribosomal protein S18 (RPS18)
NM_007355	Homo sapiens heat shock 90kDa protein 1, beta (HSPCB)

Telomerase reverse transcriptase (TERT), located on chromosome 5 at 5p15.33, is a subunit of the telomerase enzyme [161, 162] that catalyses the addition of telomeric repeats (5′-TTAGGG-3′), thereby protecting the integrity of the chromosomes from shortening of the chromosomal ends [163]. The shortening is a normal process, and this loss is between 30 to 150 bp per cell division [164]. This gene is expressed only in cells with telomerase activity, such as cancer cells [161], and is repressed in normal somatic tissues [165].

The second candidate gene, human aldolase A, fructose-biphosphate (ALDOA), is located on chromosome 16 at 16p11.2. It codes for the enzyme aldolase A or fructose-biphosphate aldolase, a glycolytic enzyme catalysing the conversion of fructose-1,6-biphosphate to glyceradehyde 3-phosphate and dihydroxyacetone phosphate, as well as the reverse reaction [166]. Expression of ALDOA is found in developing embryos and in adult muscles and liver [167]. ALDOA has two pseudogenes in humans [158].

The third candidate gene is human ribonuclease P RNA component H1 (RPPH1) located on chromosome 14 at 14q11.2. RPPH1 is the RNA component of the

ribonucleoprotein RNase P, which functions as an endoribonuclease cleaving tRNA precursor molecules to 5'-termini of their tRNA sequences [168, 169].

The fourth and last candidate gene is glyceraldehyde-3-phosphate dehydrogenase (GAPDH), a well-recognized glycolytic enzyme involved in the metabolic pathway. The gene is located on chromosome 12 at 12p13.31. In addition, GAPDH is also classed as a multifunctional protein [170], performing various roles in endocytosis, vesicular secretory transport and translational control, nuclear tRNA transport, DNA replication and repair, apoptosis [171], recognition of mismatch DNA nucleotides [172], regulation of telomere structure [173], and nuclear membrane fusion [174]. There are 62 pseudogenes of GAPDH in the human genome [158]. It was speculated that these numerous non-glycolytic functions of GAPDH are correlated to its numerousness [158].

2.1.3 Plexor® HY

The Plexor® HY kit (Promega Corporation, WI, USA) is a DNA quantification kit optimised for forensic purposes. The kit simultaneously amplifies three targets (triplex), which are used to quantify total human DNA, human male DNA and an internal PCR control (IPC) [121].

The 99 bp total human DNA target is located inside the 6.1 kb RNA, U2 small nuclear 1 (RNU2) gene, which belongs to a multi-gene family with 5-25 repeats found in each cell [175]. The gene encodes human U small nuclear RNA (snRNA) and is mapped to 17q21-q22 [176]. For human male DNA quantification, a 133 bp target located inside the 20 kb repeat motif of testis-specific protein, Y encoded locus (TSPY) mapped to the DYZ5 region of the Y chromosome, is used [121].

Using a multi-copy target, the Plexor® HY kit was reported to be highly sensitive, yielding consistent results down to 8 pg of DNA [154, 177]. Plexor® HY has higher sensitivity but lower accuracy for the ratio of autosomal to Y DNA compared to another commercial kit based on single-copy targets, the Quantifiler® Duo [154].

2.1.4 Aims

Five candidate targets (*Alu*, TERT, ALDOA, RPPH1 and GAPDH) will be evaluated in order to determine the most appropriate assay that can accurately quantify degraded samples. Optimisation is therefore necessary to ensure that the primers

work and that the assay is capable of yielding consistent results for low level or fragmented DNA. Primer concentrations, annealing temperature, reproducibility of calibration curves, and non-specific products will be looked at. Standardized genomic DNA template (CAMBIO, Cambridge, UK) will be used to determine the sensitivity of the assay. Simulated casework samples will be quantified with the best target to assess its ability to work with forensic samples.

2.2 Materials and methods

2.2.1 Primer design

Alu-elements forward and reverse primer published by Nicklas and Buel [148] were used to amplify a section of the human-specific Ya5 *Alu*-subfamily of over 2000 genomic copies to generate an amplicon size of 124 bp (Table 2.2). The amplicon size was verified to be correct by performing a nucleotide BLAST search for short sequences on the NCBI website [178].

Primers for the four other candidate genes were designed using Primer3 [160] (Table 2.2) using the following parameters:

- Product size between 130 and 170 bp
- Primer length between 18 and 27 bases (optimal at 20 bases)
- Melting temperature between 57°C and 63°C (optimal at 60°C).
- GC content at 20-80% (optimal at 50%)
- Max hairpin score at 8
- Max primer-dimer score at 3
- Max poly-x at 5
- Max 3' stability at 9

NCBI Primer-BLAST [178] was used to check for cross-reactivity to other species as well as unintended targets in the human genome. All primers were resuspended and diluted in ddH2O (pH 7.0) to the desired concentrations.

Table 2.2. Basic information for the five candidate primer sets. Size = expected product size in base pairs, F = forward and R = reverse.

Name	Length	T_m (°C)	Dimer	2° Structure	GC%	Sequence (5'-3')	Size
Alu-F	22	68.2	No	Weak	59	GTCAGGAGATCGAGACCATCCC	124
Alu-R	20	71.1	No	Weak	65	TCCTGCCTCAGCCTCCCAAG	
TERT-F	20	63.7	No	Strong	45	GGGATGTCTTTTCCCCATTT	137
TERT-R	20	63.9	No	Strong	55	CCTCACCCATGCTAGGTGTT	
ALDOA-F	20	63.7	No	Weak	50	AGGGCCTGAGCTTTAACCAT	168
ALDOA-R	20	64.1	No	Moderate	50	TGTGGCTTCTGGAAGGAATC	
RPPH1-F	19	64.3	No	Weak	58	CATCTCCTGCCCAGTCTGA	137
RPPH1-R	20	64.6	No	None	60	GTCACTCCACTCCCATGTCC	
GAPDH-F	20	63.8	No	Strong	60	GGCCTCCAAGGAGTAAGACC	147
GAPDH-R	20	63.9	No	None	55	AGGGGTCTACATGGCAACTG	

2.2.2 *Alu* optimisation

Preliminary experiments for primer optimisation used a modified protocol from Nicklas and Buel [148]. A 20 μL reaction volume was used containing 10 μL 2X Brilliant® SYBR® Green Master Mix (Agilent Technologies, CA), 0.10-0.45 μM forward and reverse primer, 1 μL DNA Control 9947A (1 ng/μL) (Promega, WI, USA) and variable amounts of ddH_2O to make up the final 20 μL volume.

The reactions were amplified using a Stratagene MX3005P (Agilent Technologies, CA) with the following parameters: an initial denaturation at 95°C for 10 minutes followed by 35 cycles of 95°C for 15 seconds, 68°C for 30 seconds, and 72°C for 30 seconds. A dissociation curve was generated by ramping up the temperature from 55 to 95°C at 0.2°C per second.

Further optimisations were carried out as follows:

- Annealing temperature at 60, 64 and 68 °C (0.30 μM primer concentrations)
- Increased primer concentrations at 0.50, 0.75, 1.00, 1.25, and 1.50 μM
- Shorter dissociation curve (72 to 95°C) [148]

Serially diluted DNA samples (CAMBIO) (1, 0.5, 0.25, 0.125, 0.063 and 0.031 ng) were amplified with the optimised primer concentrations and annealing temperature to test the sensitivity of the assay and for calibration curve construction.

2.2.3 Single-copy and multi-copy loci optimisation

The optimal primer concentrations for the forward and reverse primers were identified using a matrix of recommended concentrations (0.05-0.15 μM) and the Brilliant® SYBR® Green Kit. The annealing temperature was optimised when required. Calibration curves were constructed repeatedly over varying orders of magnitude (maximum of 50 ng to minimum of 0.003 ng) to check the amplification efficiency and linearity (r^2) (see Table 2.4 in Section 2.3).

Each reaction included 12.5 μL 2X Brilliant® SYBR® Green Master Mix, 0.375 μL ROX reference dye, and variable amounts of primer and DNA template depending on the experiment being carried out. ddH_2O was added when necessary to make a final reaction volume of 25 μL. At least one no-template control (NTC) was included with every experiment.

Thermal cycling parameters were as follows (unless stated otherwise): a hot-start of 95°C for 10 minutes followed by 40 cycles of 95°C denaturation for 30 seconds; 55°C annealing for 30 seconds; 72°C extension for 30 seconds. Dissociation curve analysis was completed by soaking at 95°C for 60 seconds then ramping up the temperature from 55°C to 95°C.

2.2.4 Plexor® HY

The reaction mix for Plexor® HY included 10 µL Plexor® HY 2X Master Mix, 7 µL amplification-grade water (Promega Corporation, WI, USA), and 1 µL Plexor® HY 20X Primer/IPC Mix. The final volume was made up to 20 µL with 2 µL DNA, either standard or unknown sample. A calibration curve was constructed using serial dilutions (1:5) of Plexor® HY Male Genomic DNA Standard from 50 ng/µL to 0.0032 ng/µL in TE buffer. All standards and unknown samples were amplified in duplicate unless stated otherwise.

The thermal cycling parameters used were as follows: initial denaturation at 95°C for two minutes followed by 38 cycles of denaturation at 95°C for 5 seconds, and then a combined annealing-extension at 60°C for 40 seconds. Dissociation curve analysis was completed with 48 cycles of 0.6°C increments every 40 seconds from 65°C.

2.2.5 qPCR data collection and analysis

Fluorescence data were collected at the annealing step, extension step and during the dissociation curve temperature ramp up. C_q values were determined by adaptive baseline settings with amplification-based threshold automatically calculated by the MxPro™ software (version 3.20 for *Alu* experiments and version 4.10 for other experiments). Raw data of Plexor® HY experiments were exported for analysis using Plexor® Analysis Software Version 1.5.4.18. Images and data were exported and manipulated with Microsoft® Excel. Statistical data analysis were carried out using JMP® 7.0 (SAS Institute Incorporated, NC, USA) and PASW Statistics 18.0 (SPSS Incorporated, IL, USA).

2.2.6 Agarose gel electrophoresis (*Alu* assay)

Amplified products of *Alu* assays were gel electrophoresed in order to determine if the correct amplicon size was generated. A 3% agarose mini-gel (Agarose A9539,

Sigma-Aldrich, Dorset, UK) was prepared in TBE buffer. Hyperladder™ V (Bioline, MA, USA) was used to size the amplicons. 4 μL DNA and 1 μL 5X loading buffer provided with the Hyperladder™ were mixed together and loaded onto the gel. Electrophoresis was carried out at 100 volts for approximately 40 minutes, followed by 20 minutes immersion in ethidium bromide solution for DNA visualisation in a UV transilluminator fitted with a digital camera. The digital image obtained was enhanced and labelled by Aperture 2 (Apple Inc., CA, USA) and Microsoft® PowerPoint.

2.2.7 Simulated casework samples: GAPDH vs. Plexor® HY

Using the optimal conditions for GAPDH and the recommended settings for Plexor® HY, five buccal swab reference samples and twenty simulated casework samples (see Table 3.3 on page 69) were quantified in duplicate. The quantification results of the GAPDH assay were compared to results from the autosomal target (RNU2) of Plexor® HY. The accuracy of quantification was determined using the interpolated DNA input data from the DNA profiles genotyped using the AmpF*l*STR® MiniFiler™ (MF) and PowerPlex® S5 (S5) kits (see Chapter 3 for amplification parameters and results).

2.3 Results and discussion

2.3.1 Primer specificity

The *Alu* primer set was shown by Nicklas and Buel [148] to be specific to human and other higher primates. NCBI Primer-BLAST was used to check for primer specificity of the other four primer sets. The TERT primer set had imperfect matches to pigs and dogs with product sizes of 4155 (pig), 1417 (dog), 1750 (dog), and 2356 (dog) bp. It was deemed species- and locus-specific due to the great difference to the intended product size of 137 bp. The ALDOA primer set had no other match besides the intended locus even with two pseudogenes; therefore, it was also both species- and locus-specific and treated as a single-copy locus.

The RPPH1 primer set had perfect matches to other higher primates (orang-utan, macaque, chimpanzee and gorilla) with the same product size as the intended locus. This was observed empirically with a kit that targets the same gene – the Quantifiler® Duo. Similarly, the GAPDH primer set had imperfect matches to other higher primates (macaque, chimpanzee, marmoset and orang-utan). It also had imperfect matches to many GAPDH pseudogenes in the human genome, yielding a highly similar product size of ±10 bp. These imperfect matches could interfere with the amplification efficiency and dissociation curve analysis.

2.3.2 *Alu* optimisation

2.3.2.1 Primer concentrations

C_q data showed that 0.30 and 0.45 μM were the optimal concentrations as the C_q values were the lowest (21.41 and 20.12, respectively). The reactions with 0.1 μM primer and the NTC did not cross the fluorescence threshold. Although different C_q values were obtained from different primer concentrations, dissociation curve analysis led to the conclusion that no specific product of 124 bp was obtained from any of these reactions. Two melting peaks, one at 59.5°C and another at 70.2°C, were observed (Figure 2.1) but they did not correspond to the 87°C peak obtained by Nicklas and Buel [148]. The dissociation curves of the no-primer control (purple) gave no dissociation peak, while the NTC (teal) showed two peaks similar to the other reactions. Since the peaks were not observed in the no-primer control, this

indicated that the primers formed multiple non-specific products in the absence of DNA template.

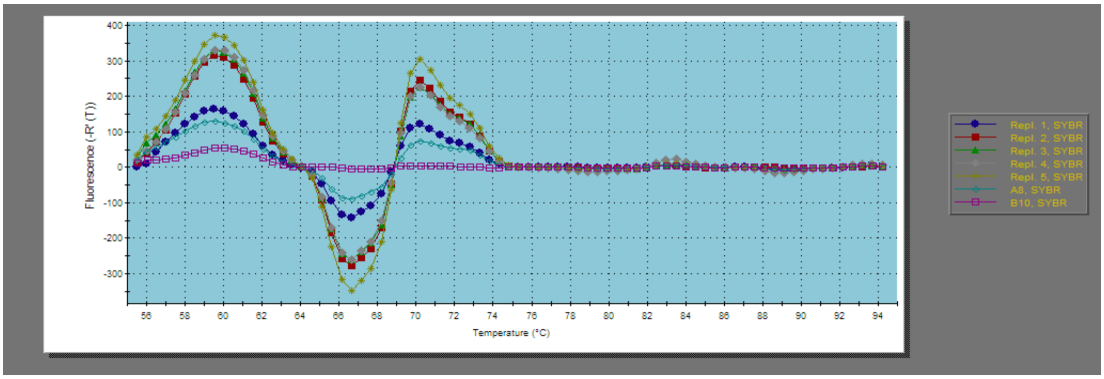


Figure 2.1. Dissociation curve for different primer concentrations analysed in duplicate (0.1 to 0.45 µM – Replicate 1 to 5 respectively). A8 = NTC (teal) and B10 = no primer control (purple).

In contrast to Nicklas and Buel [148], no additional SYBR® Green I dye was added to these reactions and hence this could result in this difference. Moreover, the primers were purified by desalting, as opposed to the recommend HPLC purification for use with SYBR® Green I [179]. The higher percentages of failed synthesis (missing 5' end and base jumping) could have led to the forward and/or reverse primers binding to other subspecies of *Alu* elements. The annealing temperature of 68°C was also suspected to be too high because the melting temperature of the *Alu* forward primer was 68.2°C.

2.3.2.2 Annealing temperature

Further attempts at optimisation were focused on determining the optimal annealing temperature for this assay. 1 ng of DNA in duplicate amplification at 68, 64 and 60°C demonstrated that lowering the annealing temperature consequently lowered the C_q values to 24.85, 18.21 and 16.57, respectively. However, the lower temperatures had a higher amount of non-specific products (data not shown) and so 68°C was chosen as the optimal annealing temperature. The dissociation curves for the three temperatures now displayed two peaks at approximately 80°C and 82°C (Figure 2.2). However, these peaks still did not correspond to the 87°C peak described by Nicklas and Buel [148] and the NTC still showed a similar dissociation curve. In an effort to minimize the amplification of non-specific products, the primer concentrations were re-optimised.

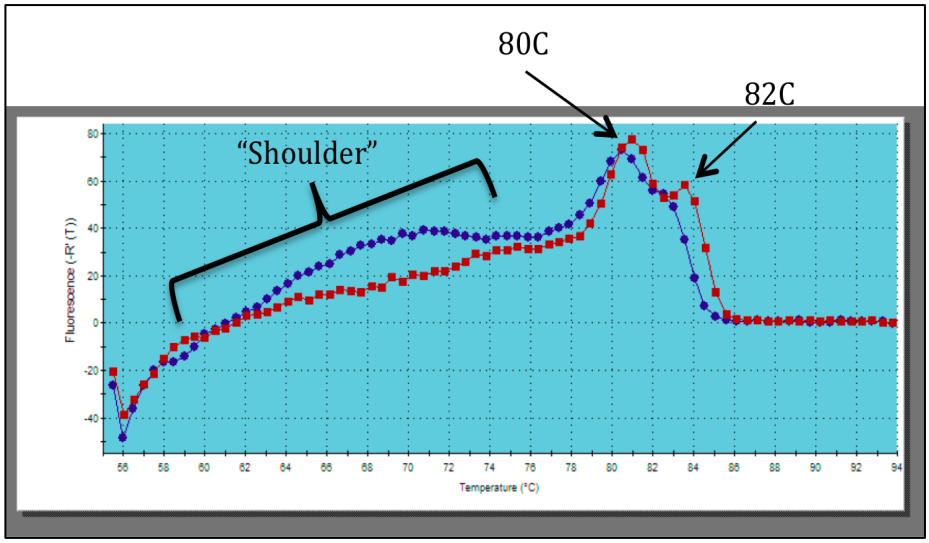


Figure 2.2. Dissociation curve for 68°C samples in duplicate (blue) and NTC (red).

2.3.2.3 Re-optimisation of primer concentration

As there are more than 2000 copies of the *Alu* target in a single genome [148] and the control DNA is intact, there might not be enough primer to bind and amplify the desired PCR products. Although the recommended concentration range of 0.05-0.30 μM was suggested in the Stratagene MX3005P Manual [129], Walker *et al.* [153] discussed that a sufficient amount of primer is needed to amplify the multiple copies of *Alu* present, therefore, higher concentrations of both forward and reverse primers were tried in an attempt to minimize non-specific products. The lowest C_q value of 16.98 was achieved by using 1.25 μM of both forward and reverse primers with an annealing temperature of 68°C (Figure 2.3). The second lowest C_q value of 17.69 was obtained with 1.50 μM of primers, while the NTC had a C_q value of 32.45. From this initial experiment, a concentration of 1.25 μM for both the forward and reverse primers and an annealing temperature of 68°C were chosen as the optimal conditions for the assay.

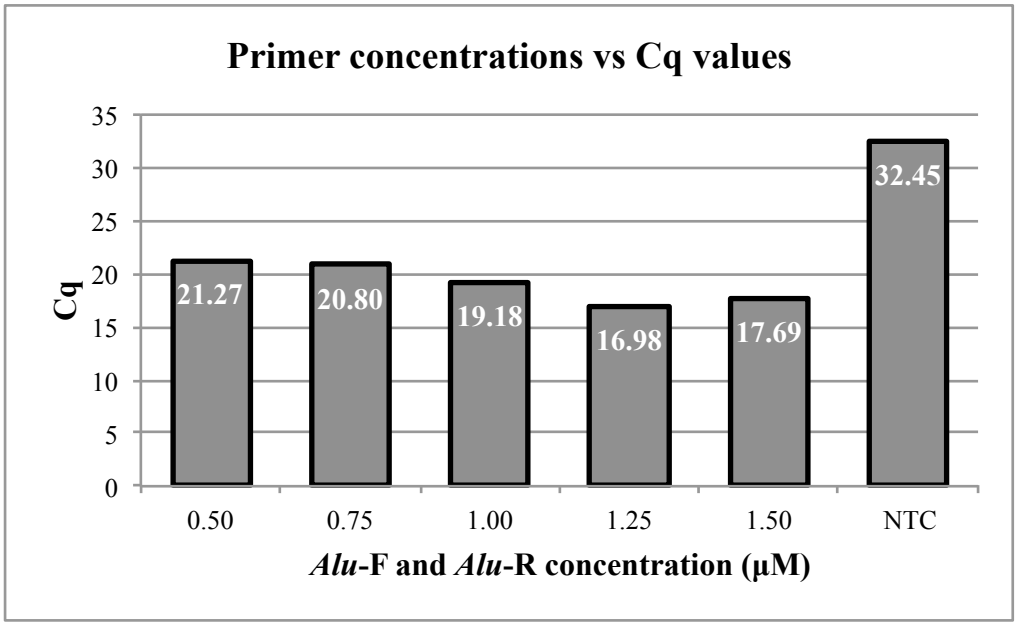


Figure 2.3. Bar graph representing the C_q values for different primer concentrations (*Alu*-F = forward primer and *Alu*-R = reverse primer).

2.3.2.4 Calibration curve

Standard dilutions of DNA were subsequently amplified to construct a calibration curve and determine the reaction efficiency and r^2 value (Figure 2.4). The analysis algorithm settings of the MxPro™ software allowed the operator to select the fluorescence data collected during either the annealing step or extension step for subsequent downstream analysis. With SYBR® Green I, the sensitivity of the assay can be improved by collecting fluorescence data at the extension step, as specific products usually have a higher melting temperature than non-specific products [137]. Efficiencies of over 160% when fluorescence data were collected during the annealing step and 130% during the extension step were seen. The r^2 values were 0.978 when the samples were analysed during the annealing step and 0.987 for the extension step. These differences were due to the higher temperature of 72°C used during extension, causing non-specific bindings to dissociate and thus the efficiency and r^2 value better reflected the specific products. As a result, further experimental analyses were performed on data collected during the extension step.

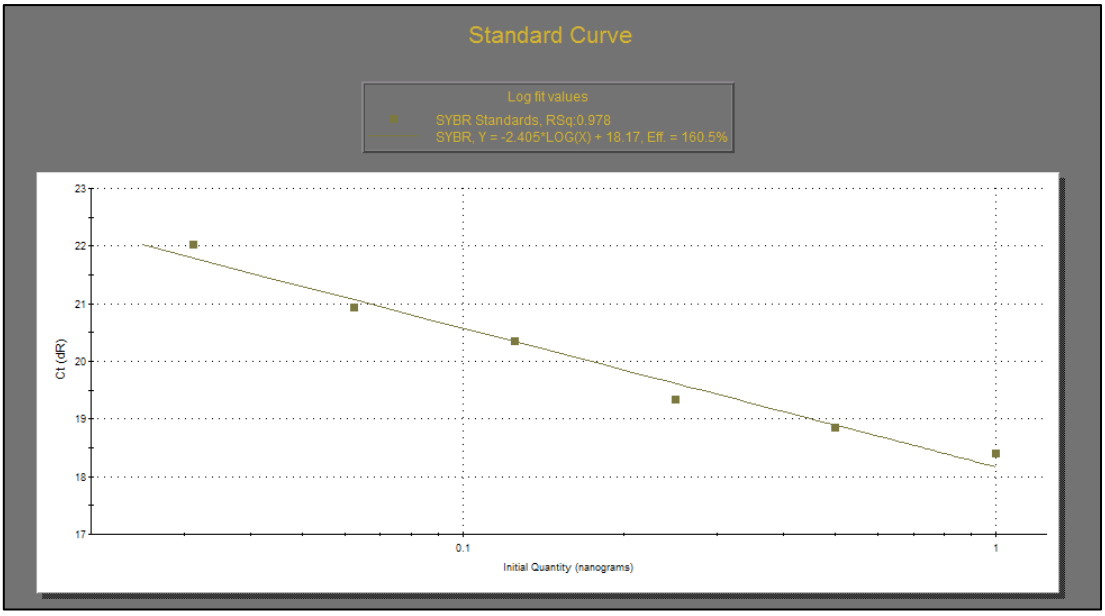


Figure 2.4. The calibration curve for serially diluted DNA samples (1, 0.5, 0.25, 0.125, 0.063 and 0.031 ng) collected during annealing step.

A doubling of the targeted product at every cycle of PCR would yield a reaction efficiency of 100% [130]. The excessively high efficiencies in this calibration curve were most probably due to the effect of non-specific products increasing the fluorescent signal ("Shoulders" in Figure 2.2 on page 47) as a result of primer-dimer and mis-priming of primers to non-target DNA sequence [180]. An NCBI BLAST search confirmed that multiple sites could be amplified by the *Alu* primer set to yield different amplicons (see Appendix A.1). All samples exhibited two dissociation peaks at 80.5 and 83.1°C (Figure 2.5). The NTC showed a similar dissociation curve to every other sample, even though the reported initial DNA quantity was less than 0.0001 ng. Differentiating the lowest amount of DNA (0.031 ng) from the NTC based on C_q values was conspicuous as the difference between them was over 7 cycles (see Appendix Table 1). This implied that the assay was sensitive enough to quantify 0.031 ng of DNA.

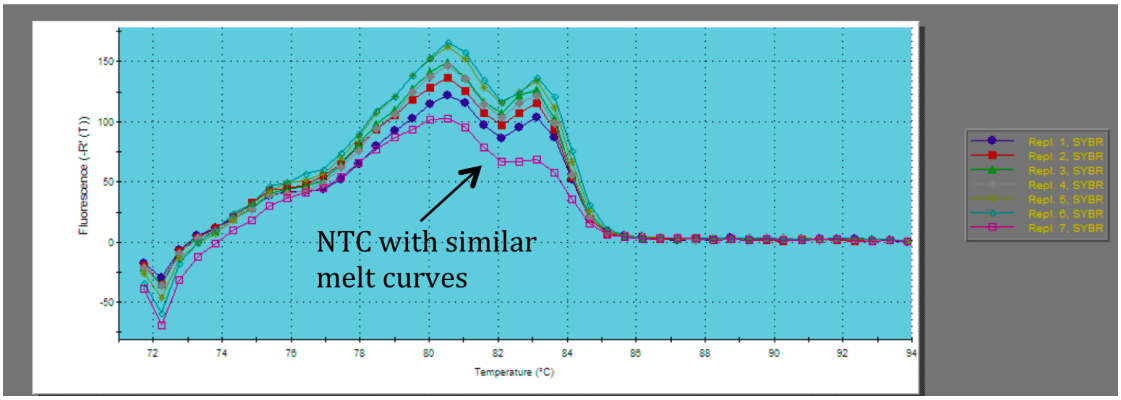


Figure 2.5. Dissociation curve for serial dilutions of DNA (Replicate 1 to 6 correspond to 1, 0.5, 0.25, 0.125, 0.063 and 0.031 ng of initial DNA amount, respectively). Replicate 7 = NTC.

Further efforts were made to determine if the increase in fluorescence of the NTC was the result of contamination in the reagents or primer interactions. Contaminating DNA in the reagents could be amplified, resulting in a detectable fluorescence but with a higher C_q value than the DNA standards. Additionally, primer-dimerization due to the excess concentrations of primers could give the same result [181]. Repeated amplifications of negative controls using different sources of water and buffer showed amplification of non-specific products in the later cycles (data not shown), indicating that the increase in fluorescent signals was not from contamination of reagents.

The excessive amplification efficiency due to primer-dimers precluded accurate DNA quantification using the *Alu* assay. As a result, it was excluded as a potential candidate.

2.3.2.5 Agarose gel electrophoresis

Serially diluted DNA samples amplified with the *Alu* primers were gel electrophoresed on a 3% agarose gel and then visualised with ethidium bromide (Figure 2.6). A band at approximately 125 bp was seen with every dilution, in agreement with the 124 bp amplicon size specified by Nicklas and Buel [148]. A faint band was also detected in the two NTCs (lane 13 and 14 in Figure 2.6).

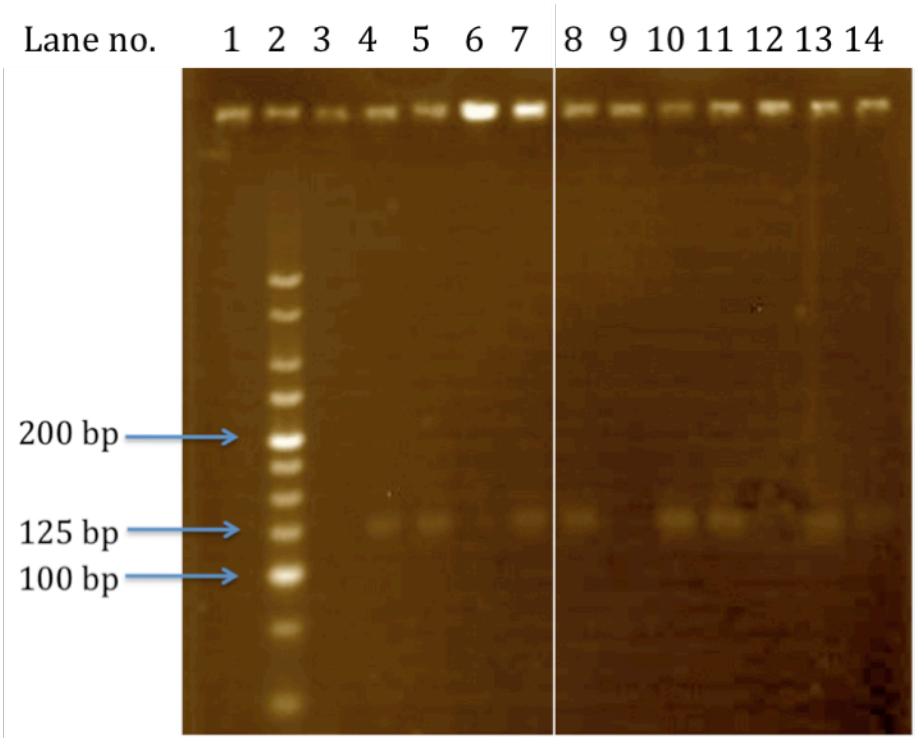


Figure 2.6. 3% gel electrophoresis of serially diluted DNA samples. From left to right: lane no. (2) hyperladder V, (4) 1 ng, (5) 0.5 ng, (7) 0.25 ng, (8) 0.125 ng, (10) 0.063 ng, (11) 0.031 ng, (13-14) NTCs. Other lanes were not loaded.

2.3.3 TERT optimisation

Using an initial primer concentration of 0.10 μM for both forward and reverse primers and 1 ng of CAMBIO DNA, the products of the TERT amplification had high C_q values of 34.62 and 34.68 in a duplicate experiment. Two distinct melting peaks were seen in the dissociation curve (Figure 2.7), which was indicative of non-specific product formation and could be due to the strong secondary structure of both forward and reverse primers (Table 2.2 on page 41). TERT was excluded as a potential candidate from further studies.

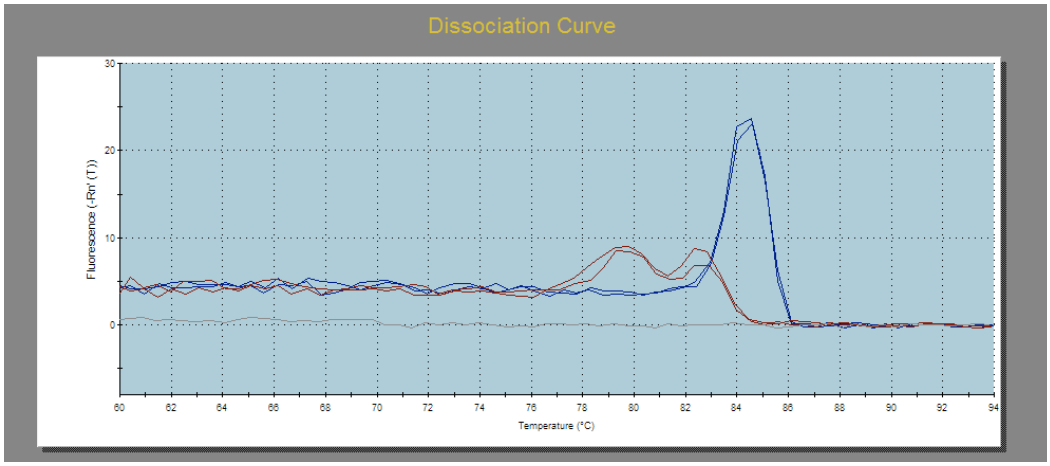


Figure 2.7. The dissociation curves of the products amplified with the TERT assay (red) compared to those of GAPDH assay (blue). Two distinct melting peaks are clearly seen for TERT.

2.3.4 ALDOA optimisation

Unlike TERT, the ALDOA amplification products gave a single melting peak of 79.3°C in all experiments. The C_q values of the duplicate set-up ranged from 28.47 to 33.42 (averaged), with the lowest seen with 0.10 μM forward primer – 0.15 μM reverse primer, followed by the 0.15 μM – 0.15 μM combination (Table 2.3).

Table 2.3. C_q values of 1 ng of CAMBIO DNA amplified in duplicate with the ALDOA primer set with varying forward and reverse primer concentrations (μM).

Reverse primer	Forward primer		
	0.05	0.10	0.15
0.05	33.42	32.57	31.15
0.10	30.40	29.49	28.93
0.15	29.88	28.47	28.49

A triplicate calibration curve showed low amplification efficiency at 74.9% with an r^2 value of 0.921 (Table 2.4). The limit of detection for the ALDOA assay was between 0.080 and 0.016 ng of initial DNA template, as all triplicate samples at 0.080 ng gave C_q values but only one of out three did at 0.016 ng. Due to the low amplification efficiency and low sensitivity, ALDOA was excluded as a potential candidate from further studies.

Table 2.4. Mean, standard deviation (SD), and relative standard deviation (RSD) of ALDOA, RPPH1, and GAPDH calibration curves. The top row shows the initial DNA template in nanograms. "-" = not available. NTC = no-template control.

DNA amount	NTC	0.003	0.016	0.031	0.063	0.080	0.125	0.250	0.400	0.500	1	2	10	50
ALDOA														
N	2	0	1 (of 3)	0	0	3	0	0	3	0	0	3	3	3
Mean	-	-	37.54	-	-	37.30	-	-	31.08	-	-	28.49	26.22	25.00
SD	-	-	-	-	-	0.73	-	-	0.15	-	-	0.145	0.15	0.30
RSD (%)	-	-	-	-	-	1.96	-	-	0.50	-	-	0.511	0.574	1.2
RPPH1														
N	8	0	5	6	4	3	6	6	3	6	6	3	3	3
Mean	-	-	37.22	37.29	35.33	37.60	33.28	32.46	32.02	31.14	30.08	29.53	27.60	26.94
SD	-	-	1.62	1.84	1.14	0.76	0.98	0.37	0.95	0.15	0.24	0.19	0.50	1.34
RSD (%)	-	-	4.36	4.93	3.23	2.02	2.95	1.15	2.98	0.48	0.79	0.65	1.82	4.97
GAPDH														
N	1 (of 3)	11	11	0	0	11	0	0	11	0	0	11	11	11
Mean	34.92	35.20	32.62	-	-	30.27	-	-	27.87	-	-	25.71	23.14	21.55
SD	-	1.47	0.63	-	-	0.31	-	-	0.16	-	-	0.10	0.16	0.16
RSD (%)	-	4.19	1.93	-	-	1.01	-	-	0.58	-	-	0.39	0.68	0.75

2.3.5 RPPH1 optimisation

An initial optimisation of primer concentrations was carried out at a 60°C annealing temperature with forward and reverse primer concentrations from 0.05 to 0.15 μM. A forward primer concentration of 0.15 μM and reverse primer concentration of 0.10 μM gave the lowest C_q value of 28.14 (singly amplified). Minor non-specific products were observed in the dissociation curve at the temperature region of 64°C to 72°C (Figure 2.8).

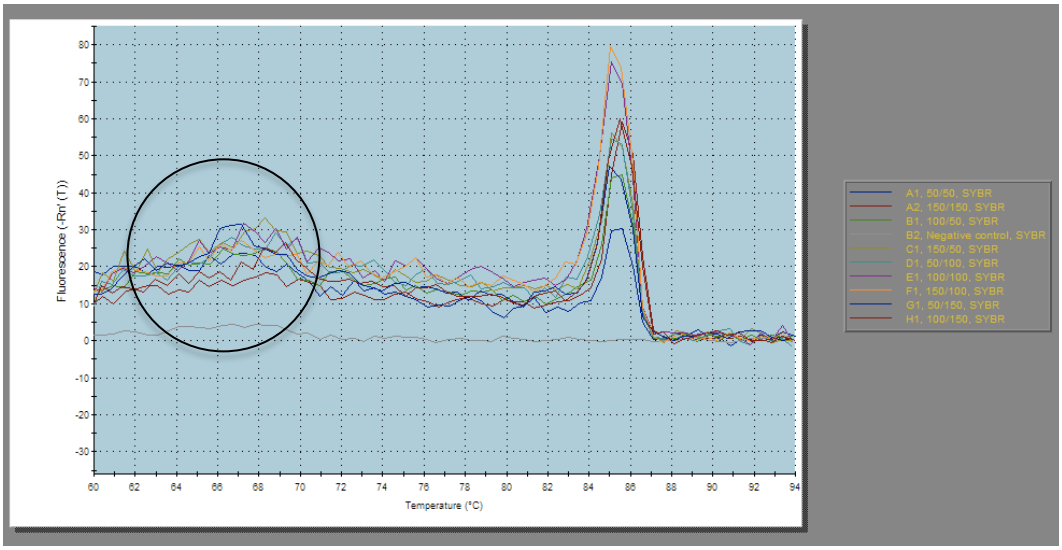


Figure 2.8. The dissociation curve of RPPH1 primer matrix experiment. Circled area indicates dissociation of minor non-specific products.

Increasing the annealing temperature to 64°C resulted in an average C_q value of 27.84 and a reduction in non-specific products. Further optimisation attempts to decrease both the annealing time and extension time to 15 seconds failed to show any amplification product, therefore, the optimal annealing temperature for RPPH1 primer set was set at 64°C with 30 seconds annealing and extension time.

Four calibration curves were constructed to determine the possibility of using RPPH1 as a DNA quantification target. The first three experiments (each one a duplicate) used DNA dilutions from 1 ng to 0.016 ng. The last calibration curve was constructed from a triplicate run of 50 ng to 0.031 ng of DNA to extend the linear range to accommodate forensic samples with higher DNA concentrations (e.g. buccal swabs). A meta-analysis of these four experiments gave an r^2 value of 0.854 and

amplification efficiency of 95.7% (Figure 2.9). The mean, standard deviation and relative standard deviation of the C_q values are shown earlier in Table 2.4 (page 53).

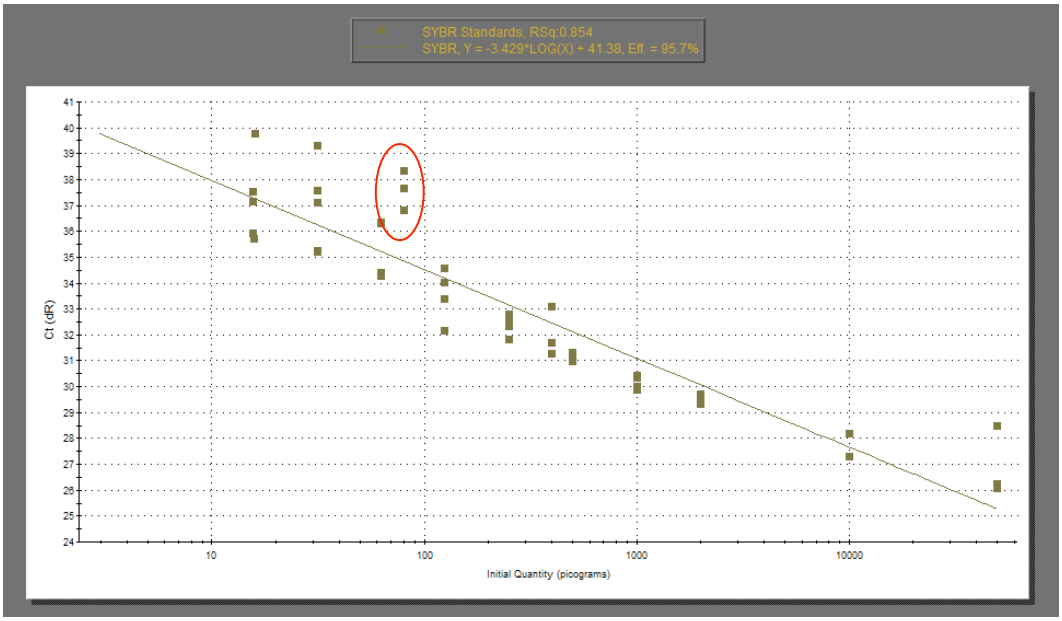


Figure 2.9. The calibration curve from meta-analysis of four experiments. Note the increase in variability as the initial quantity of DNA decreases.

C_q values for 0.080 ng of initial DNA template in the fourth calibration curve were conspicuously higher than the other initial DNA templates (circled in Figure 2.9). This had a direct bearing on the calibration curves as it invariably decreased the linearity of the curve. Moreover, the spread of the C_q values for this locus was higher (most relative standard deviations above two percent) than both ALDOA and GAPDH, of which the majority of relative standard deviation was less than one percent. Due to these reasons, optimisation of RPPH1 was not carried on further.

2.3.6 GAPDH optimisation

As with RPPH1, the GAPDH assay was first investigated for optimal primer concentrations using a primer matrix of 0.05-0.15 μM for both forward and reverse primers at 60°C annealing. The results indicated that the combination of 0.15 μM forward primer and 0.10 μM reverse primer had the lowest C_q value of 26.25. The second best C_q value of 26.48 was seen with forward and reverse primer concentrations of 0.15 μM. A single melting peak was observed in all samples at approximately 84.2°C (Figure 2.10).

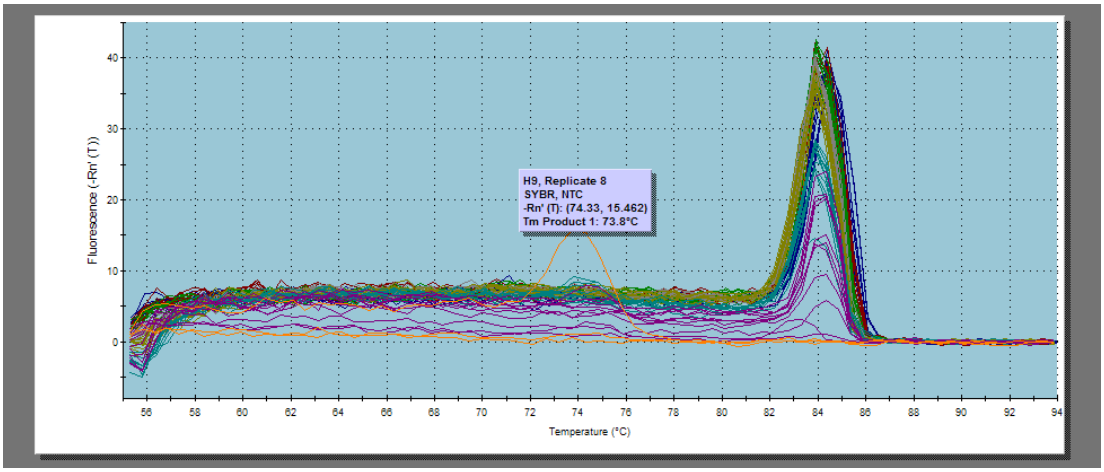


Figure 2.10. The dissociation curve of 11 repeats of serially diluted DNA and 3 NTCs amplified using the GAPDH assay. Note the single peak at 73.8°C.

Serial dilutions from 50 ng to 0.003 ng were amplified 11 times to generate a calibration curve. Three NTCs were also included. The mean, standard deviation and relative standard deviation are listed in Table 2.4 (page 53). The spread of the C_q values were narrow in all dilutions except at the lowest dilution of 0.003 ng (Figure 2.11). The amplification efficiency was 101.2% with an r^2 value of 0.981. One out of three NTCs showed a distinct melting peak at 73.8°C (Figure 2.10) and gave a starting DNA template quantity of 0.003 ng of DNA. A slightly smaller peak was seen at the same location for one of the 11 replicates of 0.003 ng DNA standard.

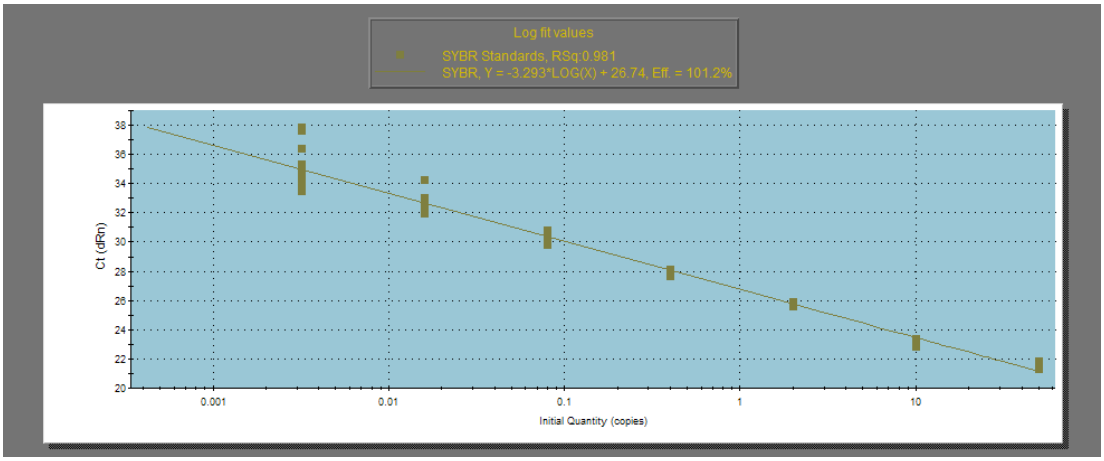


Figure 2.11. Calibration curve of 11 replicate dilutions amplified using the GAPDH primer set.

Further investigation into the cause of the anomalous melting peak was carried out by amplifying eight NTCs. Two of eight had minor amplification products, as shown by C_q values of 37.58 and 36.69 with melting peaks at 74.58°C. Contamination was unlikely because these secondary peaks seemed to occur

randomly. Had this been a result of contamination, the same melting peaks would be seen in all NTCs and at the same melting temperature with the reactions containing DNA (~84°C).

A newer, improved kit – Brilliant® II SYBR® Green Kit – was tried as an attempt to eliminate this non-specific product, as decreased non-specific products and low C_q values were claimed by the manufacturer. Due to the differences in the chemistries, the GAPDH assay was re-optimised. A five-by-five primer concentration matrix with 0.10-0.30 μM of both forward and reverse primers in 0.05 μM increments was carried out. The combination with the lowest C_q was achieved with 0.30 μM forward primer and 0.25 μM reverse primer with an annealing temperature of 64°C. However, the non-specific melting peak was not eliminated (data not shown).

The primary problems encountered in these optimisation experiments were the presence of non-specific products (TERT and GAPDH), low amplification efficiency (ALDOA) and low repeatability (RPPH1). The GAPDH assay exhibited the least variation with a relative standard deviation of less than one percent for 0.080 ng of initial DNA template and less than two percent even at a very low template amount of 0.016 ng (Table 2.4 on page 53). The higher sensitivity of GAPDH could be attributed to its imperfect matches to its abundant pseudogenes [158]. Other true multi-copy targets, on the other hand, were shown to have higher sensitivities – 0.006 ng with Plexor® HY [154] and 0.001 ng with Nicklas and Buel's *Alu* [148]. This made these assays advantageous over single-copy quantification loci like TERT, ALDOA and RPPH1 for quantifying extremely low levels of nuclear DNA.

In general, designing primers that are more specific and with no secondary structure could be the best solution. The non-specific products seen with both the TERT and GAPDH assays were most likely the result of either primer-dimer or truncated primers binding to each other rather than non-specific binding to the template DNA. Secondary peaks were only observed in the NTCs, suggesting that the primers had preference for binding to template DNA rather than to each other. Primers were synthesized using desalted purification; hence, it was possible that there were "contaminating" truncated primers inadvertently mixed in the solution and these primers could subsequently bind non-specifically during the qPCR experiments [182]. Desalting may not be the best purification option to use with SYBR® green,

an intercalating dye which detects all double-stranded DNA. Further purifications by HPLC or PAGE could be performed to eradicate truncated primers.

2.3.7 Simulated casework samples: GAPDH vs. Plexor® HY

2.3.7.1 Quantification results

Twenty simulated casework and five buccal swab samples were quantified using the GAPDH assay and Plexor® HY (Table 2.5). The casework sample experiments gave an efficiency of 99.7% and r^2 value of 0.985 with the GAPDH assay and an efficiency of 99.7% and r^2 value of 0.992 with the Plexor® HY kit. NTCs did not have any detectable DNA or melting peak in either experiment, indicating the absence of non-specific products. Sample S7-1 had the lowest DNA concentration, reported by the Plexor® HY kit at 0.013 ng/μL and by the GAPDH assay at 0.008 ng/μL. The sample with the highest DNA concentration was S7-3, with 6.48 ng/μL and 2.74 ng/μL reported with the Plexor® HY kit and the GAPDH assay, respectively.

Table 2.5. Quantification results for simulated casework (S) and buccal reference (B) samples using Plexor® HY kit and GAPDH assay. The ratio of Plexor® HY to GAPDH is also shown. DNA quantity is shown in ng/μL. Normality is p-value from Shapiro-Wilk test (N = 25).

Sample	HY	GAPDH	Ratio	Sample	HY	GAPDH	Ratio
S1-1	0.569	0.106	5.36	S7-1	0.013	0.008	1.56
S1-2	0.407	0.245	1.66	S7-2	0.442	0.160	2.77
S2-1	0.090	0.053	1.71	S7-3	6.480	2.740	2.37
S2-2	0.539	0.265	2.04	S7-4	1.050	0.901	1.17
S3-1	2.520	0.410	6.15	S7-A	0.874	0.385	2.27
S3-2	2.930	3.260	0.90	S7-B	0.231	0.100	2.31
S6-1	0.058	0.030	1.92	S7-C	0.097	0.026	3.72
S6-2	0.090	0.025	3.54	B1-M	1.350	0.895	1.51
S6-3	0.202	0.073	2.78	B2-F	0.990	1.890	0.52
S6-4	0.634	0.351	1.81	B3-M	0.585	0.385	1.52
S6-A	0.280	0.372	0.75	B4-M	1.950	1.510	1.29
S6-B	0.207	0.210	0.98	B5-M	2.200	2.020	1.09
S6-C	4.540	1.210	3.75				
Median	*0.569*	*0.351*	*1.81*	*Normality*	*<0.001*	*<0.001*	*0.004*

The ratio of Plexor® HY to GAPDH quantification results were calculated (Table 2.5). A ratio of one indicates that the results from both assays were the same. In 21 of 25 samples, the ratios were greater than one, meaning that the Plexor® HY results

were higher than the GAPDH results. Only in four samples (S3-2, S6-A, S6-B and B2-F) were the results of GAPDH higher than Plexor® HY (ratio < 1). The lowest ratio observed was 0.52 in sample B2-F and the highest was 6.15 in sample S3-1. A median of 1.81 was computed, indicating that the quantification results of Plexor® HY were almost twice of the GAPDH results. Since the samples were related and the data were not normally distributed, the non-parametric Wilcoxon Signed Rank Test was applied to statistically compare the two quantification methods [146]. The null hypothesis was "*the median of differences between GAPDH and Plexor® HY results equals 0*" and the alternative hypothesis was "*the median of differences does not equal 0*" (two-sided). The null hypothesis was rejected, indicating a significant difference between the two methods ($p = 0.002$, $W = 278$, $N = 25$).

The disparity between the two assays could be caused by a number of reasons. The product size of the Plexor® HY autosomal target (RNU2) is 99 bp while the product size of the GAPDH assay is 147 bp, approximately 1.5 times longer. Swango *et al.* [150] reported that in 23 casework samples, the ratio of nuCSF (a 67 bp product) to nuTH01 (170-190 bp) ranged from 0.7 to 13.9 with a median at 1.4. The results reported in this study were similar, ranging from 0.52 to 6.15 with median at 1.81. The ratios of the buccal samples were not as high as those of the simulated casework samples, possibly due to some inhibitors co-extracted with the casework samples. As the Brilliant® kit (used for GAPDH) was shown to be more affected by PCR inhibitors than the Plexor® HY kit (see Section 6.3.3.3), the lower quantification results of the GAPDH assay could inflate the ratios between the two kits in inhibited samples.

2.3.7.2 Accuracy of quantification

Due to the difference in the two quantification methods, further investigation was carried out to determine the accuracy of the two methods. It is known that allele peak heights in an STR profile are dependent on the initial DNA quantity added to the PCR reactions. Two mini-STR kits, MF and S5, were used to amplify serial dilutions of 9947A DNA standard and the twenty casework samples simultaneously (See Chapter 3 for details). Calibration curves were constructed using DNA standards from 1 ng to 0.016 ng amplified in triplicate (Figure 2.12). The peak heights were

averaged across the loci in each kit and a linear regression was fitted to the data, correlating average peak height to initial DNA quantity:

$$Q_{mf} = \frac{H_{mf} + 6.7831}{2.5436} (r^2 = 0.977)$$

Equation 2.1. Linear regression equation for the MF kit.

$$Q_{s5} = \frac{H_{s5} - 20.572}{2.1773} (r^2 = 0.961)$$

Equation 2.2. Linear regression equation for the S5 kit.

, where Q_{mf} = initial DNA quantity (in pg) for the MF kit, H_{mf} = average peak height per allele for the MF kit, Q_{s5} = initial DNA quantity (in pg) for the S5 kit, and H_{s5} = average peak height per allele for the S5 kit.

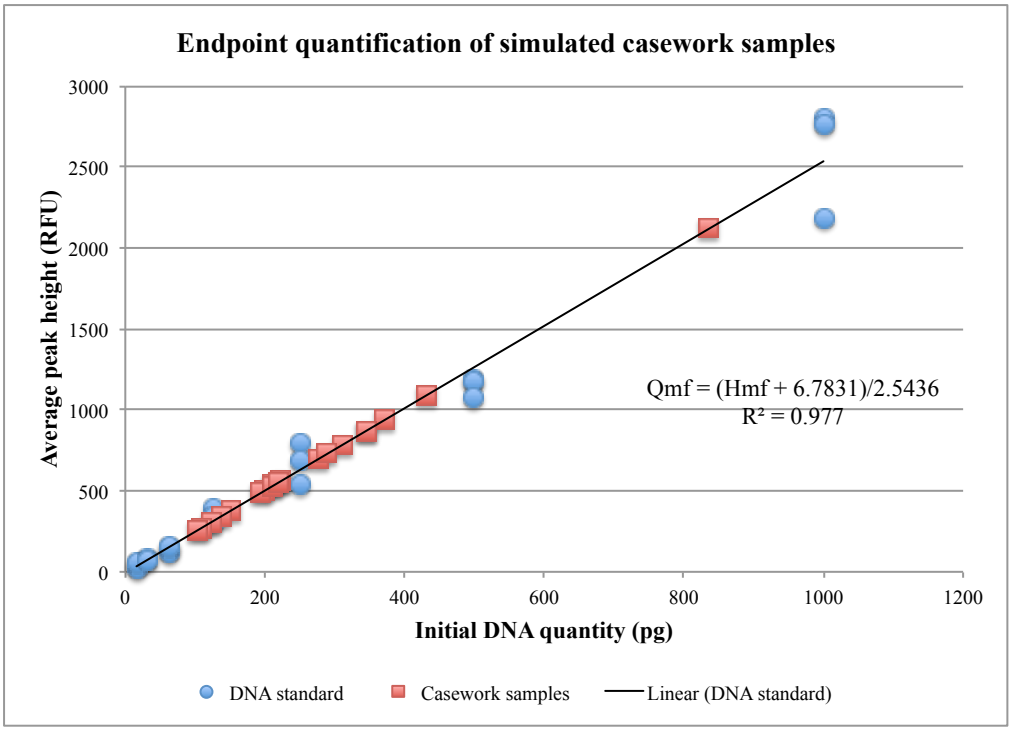


Figure 2.12. Calibration curve of MF peak heights (averaged over loci) to initial DNA quantity generated from triplicate amplification of 9947A serial dilutions (blue circles). Casework samples were quantified based on their peak heights (red squares).

The peak heights of the casework samples were then plotted onto the regression line and the initial DNA quantity of the casework samples were interpolated (Table 2.6). These interpolated initial DNA quantities were compared to the quantification results of the GAPDH assay and the Plexor® HY kit using the Spearman's ρ. The null hypothesis was "*the true correlation coefficient (ρ) is equal to 0*" and the alternative hypothesis was "*ρ is not equal to 0*" (two-sided). Mixtures were excluded from the

comparisons (bolded in Table 2.6) as the calibration curves were constructed from a single-source DNA profile and the average peak height per allele would be skewed by the alleles of the minor profile.

Table 2.6. The initial DNA quantity (in ng) of casework samples interpolated from calibration curves (Q_{mf} and Q_{s5}) and the reported initial DNA quantity based on the GADPH assay and Plexor® HY kit. Bold denotes mixed sample.

Sample	Q_{mf}	Q_{s5}	GAPDH	HY	Sample	Q_{mf}	Q_{s5}	GAPDH	HY
S1-1	0.201	0.268	0.134	0.569	S6-A	0.836	1.277	0.779	0.560
S1-2	0.277	0.337	0.311	0.407	S6-B	0.430	0.586	0.487	0.414
S2-1	0.222	0.319	0.331	0.448	S6-C	0.211	0.202	0.248	0.908
S2-2	**0.151**	**0.193**	**0.293**	**0.539**	S7-1	0.109	0.133	0.057	0.66
S3-1	0.310	0.326	0.097	0.504	**S7-2**	**0.219**	**0.302**	**0.282**	**0.663**
S3-2	**0.224**	**0.276**	**0.620**	**0.586**	**S7-3**	**0.103**	**0.099**	**0.570**	**1.296**
S6-1	0.213	0.252	0.194	0.290	S7-4	0.371	0.392	0.500	0.525
S6-2	0.138	0.216	0.164	0.448	S7-A	0.346	0.623	0.370	0.874
S6-3	**0.123**	**0.162**	**0.189**	**0.505**	S7-B	0.289	0.487	0.239	0.462
S6-4	**0.195**	**0.261**	**0.408**	**0.634**	S7-C	0.345	0.486	0.158	0.486

The other 14 samples were used in the statistical analysis. The correlation coefficient between GAPDH and Q_{mf} was 0.705 ($p = 0.005$) and between GAPDH and Q_{s5} was 0.644 ($p = 0.013$), indicating that the null hypothesis could be rejected. In contrast, the correlation coefficient between Plexor® HY and Q_{mf} was 0.293 ($p = 0.310$) and between Plexor® HY and Q_{s5} was 0.282 ($p = 0.329$), meaning that the null hypothesis could not be rejected. These statistical results showed that there existed a correlation between the initial DNA quantity interpolated from the MF and S5 peak height data and the quantification result of the GAPDH assay, suggesting that the assay better reflected the actual initial DNA quantity than the Plexor® HY kit. In contrast, the Plexor® HY could find its use in situations where a higher sensitivity is required.

The method described here was based on end-point quantification. As discussed earlier (Section 1.7.1), end-point analysis is not as accurate as real-time analysis due to non-linearity at the later cycles of PCR, but endpoint analysis methods, e.g. BodeQuant [183], were once used to quantify human DNA in a forensic laboratory [184]. Due to the lack of a reliable qPCR DNA standard such as the NIST SRM2372

[185] when the study was carried out, the method used in this experiment at least provided an approximation of the initial DNA quantity added to the MF and S5 PCR reactions, which was then used to compare to the quantification results of the GAPDH assay and Plexor® HY kit.

3 Comparisons of next-generation and mini-STR kits

3.1 Introduction

3.1.1 Next-generation and mini-STR kits

In 1999, the European Network of Forensic Science Institutes (ENFSI) DNA Working Group and European DNA Profiling Group (EDNAP) published the "European Standard Set" of seven STR loci [186], which was then extended by five loci in 2006 [187]. The original seven loci were D3S1358, D8S1179, D18S51, D21S11, FGA, TH01 and vWA with additional ones being D1S1656, D10S1248, D12S391, D14S1434 and D22S1045. Following these recommendations, commercial manufacturers have worked closely with the forensic DNA community to develop two next-generation kits – AmpF*l*STR® Next Generation Multiplex™ (Applied Biosystems) and PowerPlex® ESX/ESI (Promega Corporation) (Table 3.1). In addition, the National Institute of Standards and Technology (NIST) has been leading the research in next-generation multiplexes, with a 26plex autosomal STR multiplex being the latest innovation to aid identification in difficult samples and situations in which mass kinship analysis must be carried out (e.g. mass disasters) [188]. The benefits of small amplicon size in obtaining information from sub-optimal forensic samples have been widely demonstrated [34, 35, 46, 47, 189, 190].

Table 3.1 The markers of each next-generation and mini-STR kit compared to two current standard kits (SGM+ and PP-16 = PowerPlex® 16) and two recommendations (COD = CODIS and ENF = ENFSI). Tick mark (✓) indicates inclusion in the set. Plus (+) is optional inclusion. Chr = chromosome, MF = ABI Minifiler™, S5 = PowerPlex® S5, ESS = Mentype ESS, NGM = ABI NGM™ and ESX = PowerPlex® ESX/ESI.

Chr	Marker	MF	S5	ESS	NGM	ESX	SGM+	PP-16	COD	ENF
1	D1S1656			✓	✓	✓				✓
2	D2S1338	✓			✓	✓	✓			
	TPOX							✓	✓	
	D2S441			✓	✓	✓				✓
3	D3S1358				✓	✓	✓	✓	✓	✓
4	FGA	✓	✓		✓	✓	✓	✓	✓	✓
5	D5S818							✓	✓	
	CSF1PO	✓						✓	✓	
6	SE33				+	+				
7	D7S820	✓						✓	✓	
8	D8S1179		✓		✓	✓	✓	✓	✓	✓
10	D10S1248			✓	✓	✓				✓
11	TH01		✓		✓	✓	✓	✓	✓	✓
12	vWA				✓	✓	✓	✓	✓	✓
	D12S391			✓	✓	✓				✓
13	D13S317	✓						✓	✓	
15	PENTA E							✓		
16	D16S539	✓			✓	✓	✓	✓	✓	
18	D18S51	✓	✓		✓	✓	✓	✓	✓	✓
19	D19S433				✓	✓	✓			
21	D21S11	✓			✓	✓	✓	✓	✓	✓
	PENTA D							✓		
22	D22S1045			✓	✓	✓				✓
X/Y	AMEL	✓	✓	✓	✓	✓	✓	✓	✓	

There are variations in the number of loci, dyes used and amplicon sizes between the different kits (Table 3.2). The kits that genotype the highest number of STR loci are NGM and ESX/ESI, which amplify 15 STR loci plus the sex typing marker amelogenin. The three systems were designed to incorporate all the recommended ENFSI loci using a five-dye detection system. The only difference between the ESX and ESI kits is the placement of the amplicons based on size and fluorescent tags. ESX has the five new ESS loci as mini-STRs (shorter than 185 bp), while ESI has the current loci as mini-STRs (seven out of ten loci shorter than 230 bp). These kits have a discrimination power of one in a quintillion (10^{18}).

Table 3.2 The number of loci, the number of fluorescent dyes used, the size standard and the longest amplicon (in allelic ladder) of each mini-STR kit with PP-16 as a reference.

Kit	No. Loci + AMEL	Detection	Size Standard	Max Amplicon Size
PP-16	15+1	4 dyes	CXR 600	474 bp
MF	8+1	5 dyes	LIZ 500	260 bp
S5	4+1	3 dyes	ROX 600	260 bp
ESS	5+1	4 dyes	ROX 550	225 bp
NGM	15+1	5 dyes	LIZ 500	352 bp
ESX	15+1	5 dyes	LIZ 500	410 bp
ESI	15+1	5 dyes	LIZ 500	383 bp

The three mini-STR kits, AmpF*l*STR® MiniFiler™, PowerPlex® S5 and Mentype® ESS Pentaplex (Biotype® AG), all target less than eight STR loci. Consequently, the discrimination powers of these kits are much lower, e.g. the power of discrimination for the S5 kit is less than 1 in 190,000. Although this is not as high as the ESX/ESI kits, these three kits benefit from having all amplicons less than 260 bp, making them ideal for degraded samples. The S5 kit, genotyping five loci, is marketed as a screening method. Information gathered from this prior screening can then be used to determine the best downstream analysis method to proceed with. On the other hand, the MF kit, genotyping eight loci, is more appropriate as a supplementary analysis for challenging and difficult samples.

All kits claim sensitivity down to sub-optimal range for STR amplification (<0.10 ng) and the ability to overcome common PCR inhibitors such as hematin and humic acid. For instance, the S5 kit is capable of consistently providing full profiles for samples containing 0.05 ng of DNA and is extremely resistant to inhibitors [191]. Andrade *et al.* [192] compared the MF kit to the Identifiler® kit on casework samples, including blood, saliva, bone, tooth and hair. The MF kit was able to obtain more complete profiles and was suggested that it could also be used to cross check samples for false homozygous alleles and artefact peaks [192].

The ESI and MF kits have been developmentally validated for use according to the FBI/National Standards and SWGDAM guidelines [193, 194]. Concordance studies of MF with Identifiler® by NIST resulted in a high confidence for use with existing databases (99.7% concordance) [195]. Validation and concordance studies like these

ensure that the newly developed kits are ready for use with real casework samples by demonstrating that they are robust, reliable, and reproducible [196]. Additionally, internal validations must be carried out by local laboratories to verify that the conditions used can work effectively in-house [197].

3.1.2 LCN analysis

Low copy number (LCN) analysis was first proposed in 2000 by Gill *et al.* [198], in which an increased PCR cycle number of 34 cycles was used instead of the manufacturer's recommendation of 28 cycles. Although an increased sensitivity can be obtained, there are many caveats associated with using this technique [199]. The most common problems tied to LCN DNA analysis include increased stutters, allelic and locus dropout, allelic dropin, and complexity in interpretation of results [7, 198]. Other techniques, such as whole genome amplification [200] and post-PCR purification [201, 202], have been tried in order to circumvent the problems associated with increased cycles, but they also suffer from the same analytical problems.

Furthermore, there is much confusion about the term LCN analysis among reporting scientists, judicial personnel, and the media whether the term refers to a specific technique (34 cycles), specific interpretation criteria, and/or the stochastic effects observed [203]. Due to these issues, the UK court questioned the reliability, reproducibility, and lack of validation of LCN methods in the recent Omagh trial [204], sparking a review of the LCN methods [205]. The forensic science community has responded feverishly to these issues [206, 207]. A wider term called "low template DNA" (LT-DNA) has been proposed when referring to samples with a low amount of DNA, in which a stochastic effect is expected and seen, regardless of the method used to generate STR profiles [205, 207].

Recent attempts have been made to clarify ambiguous terms and to introduce statistics-based interpretation criteria [207-209]. Gill and Buckleton [210] have suggested that a single, universal interpretation rule be accepted and used without referring to the term LCN. However, a few forensic scientists believe "general acceptance" has not been reached and implementation of the proposed statistical frameworks is not widely carried out [211]. Currently, there is no indication of the debate abating [212-216].

3.1.3 Simulated casework samples

Simulated casework samples can be used to test the ability of the kits to amplify a variety of routine casework samples. These samples can be buccal swabs and dried body fluid stains, including blood, saliva, semen and vaginal secretions on various substrates. Proficiency-testing services, such as the Collaborative Testing Services Inc. (CTS) and the German DNA Profiling Group (GEDNAP), produce simulated casework samples and ship them out to any laboratory that is interested in obtaining certification in reporting DNA tests.

CTS is an internationally recognized inter-laboratory testing service. One of the services available is DNA profiling from dried body fluid stains. The CTS samples can contain a mixture of up to two persons.

GEDNAP was established in the 1980s and is now a part of EDNAP (European DNA Profiling Group). With the aim of standardising, maintaining and improving the quality of forensic medicine and forensic science services to the general public, GEDNAP has been carrying out blind trials with laboratories for over 15 years [217]. The GEDNAP simulated casework samples are composed of three reference samples (person A, B and C) and four stains (stain 1 to 4) each. Each stain could contain a mixture from up to three persons, and the stain type is considered typical in casework samples. Some samples can also contain a rare allele.

3.1.4 Aims

Prototype kits obtained from three manufacturers (Applied Biosystems, Promega Corporation and Biotype) will be assessed using various validation protocols and will form parts of the internal validation for the laboratory. Two mini-STR kits (S5 and MF) will be evaluated in terms of sensitivity and the effect of increased cycle. Both kits and three additional kits – ESS, PowerPlex® EP01 (ESX prototype) and PowerPlex® EP02 (ESI prototype) – were also used to amplify different types of simulated casework samples. The aim is to help forensic scientists make an informed decision on which kit to adopt and subsequently internally validate for their own laboratory uses. The results obtained will demonstrate the current standards in forensic STR typing.

3.2 Materials and methods

3.2.1 Sensitivity study

Three DNA sources – ABI 007 Control DNA (Applied Biosystems, CA, USA), CAMBIO DNA and Promega 9947A Control DNA (Promega Corporations, WI, USA) – were used for the sensitivity study. Initial DNA template of 1, 0.5, 0.25, 0.125, 0.063, 0.031 and 0.016 ng were used.

3.2.2 Increased cycle study

The three lowest concentrations of CAMBIO DNA – 0.063, 0.031 and 0.016 ng – were amplified with the S5 and MF kits using a 34 cycle PCR instead of the recommended 30 cycles, to investigate the effects of the increased cycle on sub-optimal DNA amounts [198, 205].

3.2.3 Simulated casework study

3.2.3.1 Simulated casework samples

Simulated casework samples obtained from CTS Tests 06/07/08 (coded "S1-X to S3-X") and GEDNAP Trial 36/37 (coded "S6-X and S7-X") were used to test the ability of the next-generation and mini-STR kits to amplify a variety of routine casework samples (Table 3.3). All 20 samples were amplified with ESS, MF and S5, and 12 samples were amplified with EP01 (ESX prototype) and EP02 (ESI prototype) (Table 3.3).

Table 3.3 Casework-type samples tested. Mix = mixture

Sample	Mix	Description	Kit
S1-1	No	Bloodstain on red flower print cloth	MF, S5, ESS
S1-2	No	Bloodstain on leopard print cloth	MF, S5, ESS
S2-1	No	Bloodstain on blue plaid fabric	MF, S5, ESS
S2-2	Yes	Blood and semen stain on patterned fabric	MF, S5, ESS
S3-1	No	Bloodstain on orange material	All
S3-2	Yes	Blood and semen stain on brown material	All
S6-1	No	Bloodstain on denim	All
S6-2	No	Saliva stain on envelope	All
S6-3	Yes	Mixed semen and saliva stain on cotton wool	All
S6-4	Yes	Bloodstain on cotton wool	All
S6-A	No	Bloodstain on cotton wool	All
S6-B	No	Bloodstain on cotton wool	MF, S5, ESS
S6-C	No	Bloodstain on cotton wool	MF, S5, ESS
S7-1	No	Bloodstain on wallpaper material	All
S7-2	Yes	Saliva mixture on a cotton swab	All
S7-3	Yes	Bloodstain on a cotton swab	All
S7-4	No	Semen stain on a bed sheet	All
S7-A	No	Bloodstain on cotton wool	All
S7-B	No	Bloodstain on cotton wool	MF, S5, ESS
S7-C	No	Bloodstain on cotton wool	MF, S5, ESS

3.2.3.2 DNA extraction

All samples were DNA extracted using a QIAGEN Micro-kit (QIAGEN, CA, USA) following the standard protocol for stain extraction. A negative extraction control (unused cotton swab) was included. All samples were eluted in sterile double distilled water (ddH_2O) at a final volume of 100 μL.

3.2.3.3 DNA quantification

Plexor® HY kit was used to quantify and determine the presence of inhibitors in the extracted samples. The reaction set-up, cycling conditions, and data analysis were

carried out as detailed in Section 2.2.4. Supplied Plexor® HY Genomic DNA Standard was diluted five-fold serially from 50 ng/μL to 3.2 pg/μL and used as a standard for calibration curves.

3.2.4 DNA amplification

All reactions were set up in triplicate and a negative control was included with each dilution set for the sensitivity and increased cycle studies. 10 μL DNA template was added to each PCR reaction.

For the simulated casework study, an optimal amount of DNA template (~0.5 ng) was added to each reaction based on prior quantification results (see Section 3.2.3.3) whenever possible and 5 μL of DNA template was added for samples that fell below the optimal range. All samples were amplified in duplicate, and a duplicate of both positive control and negative control were included with each batch.

All reactions were set up according to manufacturers' protocol and made up to a total volume of 25 μL with amplification-grade water.

3.2.4.1 MiniFiler™

Each PCR reaction was made of 10 μL Master Mix, 5 μL Primer Pair Mix and a chosen amount of DNA template. The cycling parameters used were an initial denaturation at 95°C for 11 minutes, then 30 cycles of 94°C for 20 seconds, 59°C for two minutes, and 72°C for one minute, followed by a final extension step at 60°C for 45 minutes.

3.2.4.2 PowerPlex® S5

Each PCR reaction was made of 5 μL 5X S5 Master Mix, 2.5 μL 10X S5 Primer Pair Mix, 7.5 μL ddH_2O, and a chosen amount of DNA template. The cycling parameters used were an initial denaturation at 96°C for two minutes, then 30 cycles of 94°C for 30 seconds, 60°C for two minutes, and 72°C for 90 seconds, followed by a final extension step at 60°C for 45 minutes.

3.2.4.3 Mentype® ESS

Each PCR reaction was made of 5 μL Reaction Mix A, 2.5 μL Primer Mix, 0.4 μL *Taq* (2.5 U/μL), and a chosen amount of template DNA. The cycling parameters

used were 94°C for four minutes, then 30 cycles of 94°C for 30 seconds, 60°C for two minutes, and 72°C for 75 seconds, followed by a final extension at 68°C for 60 minutes.

3.2.4.4 PowerPlex® EP01

Each PCR reaction was made of 5 μL PowerPlex® EP01 5X Master Mix, 2.5 μL PowerPlex® EP01 10X Primer Pair Mix, and a chosen amount of DNA template. The cycling parameters used were 96°C for two minutes, followed by 30 cycles of 94°C for 30 seconds, 60°C for two minutes, and 72°C for 90 seconds, then a final extension at 60°C for 45 minutes.

3.2.4.5 PowerPlex® EP02

Each PCR reaction was made of 5 μL PowerPlex® EP02 5X Master Mix, 2.5 μL PowerPlex® EP02 10X Primer Pair Mix, and a chosen amount of DNA template. The cycling parameters used were 96°C for two minutes, followed by 30 cycles of 94°C for 30 seconds, 59°C for two minutes, and 72°C for 90 seconds, then a final extension at 60°C for 45 minutes.

3.2.5 Detection

All MF, S5, EP01 and EP02 samples were analysed using the ABI 3100-*Avant* CE instrument according to manufacturer's instructions. For MF, a loading cocktail was made by adding 0.3 μL LIZ 500, 8.7 μL HiDi™ formamide and either 1 μL amplified product or MF allelic ladder to a 96-well plate. For S5, a loading cocktail was made by adding 0.5 μL ILS 600 and 9.5 μL HiDi™ formamide and either 1 μL amplified product or S5 allelic ladder to a 96-well plate. For EP01 and EP02, a loading cocktail was made by adding 1.5 μL ILS 500 and 10 μL HiDi™ formamide and either 1 μL amplified product or EP01 or EP02 allelic ladder to a 96-well plate.

All samples were denatured at 95°C for three minutes and immediately snap-cooled on ice for three minutes. The injection parameters for all samples were five seconds at three kV.

The ESS samples were analysed using the ABI 310 CE instrument. 1 μL amplified product or allelic ladder was added to 12 μL HiDi™ formamide and 0.5 μL ROX

550. The mixture was then denatured at 95°C for three minutes and snap-cooled for three minutes. The injection time was five seconds and the voltage was 15 kV. Each sample was run at 60 kV for 24 minutes.

3.2.6 Data analysis

Raw data were analysed and genotyped with Genemapper® ID 3.2.1 (Applied Biosystems, CA, USA) using bins and panels for each kit provided by the manufacturers. The allele detection threshold was set to 50 RFUs for all dyes. Relevant information, such as allele designations and peak heights (in RFUs) were exported to Microsoft® Excel for further manipulation. Statistical tests were carried out using JMP® 7.0.1 and PASW Statistics 18.0.

DNA profiles from duplicates of simulated casework samples were genotyped using the consensus method, whereby only alleles that were repeated in both EPGs were included in the "consensus profile". The profiles in other studies were genotyped separately for each replicate (no "consensus"). The alleles were then used to calculate a percentage profile in order for the different kits with different numbers of loci to be compared directly with the same scale. The percentage profiles were calculated based on the number of alleles observed compared to the maximum number of alleles that could be seen in a full profile. For example, a ten-allele profile genotyped using the S5 kit (five loci) would give a percentage profile of 100%. A homozygous locus with peak height above 150 RFUs was counted as two alleles. Below 150 RFUs, a homozygous locus was scored as a single allele with possible allele dropout (fail – F).

Mean peak heights per allele were computed as follows, the peak heights of the two alleles for a heterozygous locus were summed and divided by two and the peak height of a single allele for a homozygous locus was divided by two.

Heterozygous peak balance (Hb_x) is the ratio of balance between two alleles of a heterozygous locus. Hb_x was calculated by dividing the smaller peak height by the larger peak height, resulting in a value that was always less than or equal to one. A value close to one meant that the two alleles were balanced.

Stutter is an artefact that is one tandem repeat less than the actual STR allele. Stutter proportion (S_p) was calculated by dividing the peak height of the stutter by the peak

height of its corresponding allele. Stutter thresholds were calculated by multiplying the standard deviation by three and then adding to the mean S_p.

The peak height data of EP01 and EP02 were used to investigate the effect of produce size on peak heights. Raw peak height data were transformed by taking log10 to convert the data to an approximate normal distribution. Size differences were calculated by subtracting the mean allele sizes of EP02 from those of EP01, i.e. negative values indicated that EP01 alleles were shorter and vice versa.

3.3 Results and discussion

3.3.1 Sensitivity study

3.3.1.1 Percentage profiles

The percentage profiles of three DNA sources typed with the S5 and MF kits were calculated from triplicate amplifications (Table 3.4). No allele was detected with any negative control. The two control DNA samples provided by the manufacturers (ABI 007 and 9947A) did not give a full profile with the S5 kit even at the recommended DNA template quantity of 0.5 to 1 ng. Only CAMBIO DNA produced full profiles using the recommended DNA starting template.

Table 3.4. Percentage profiles of three DNA sources amplified with S5, S5 excluding the TH01 locus (No-TH) and MF kits.

	ABI 007			CAMBIO			9947A		
DNA (ng)	***S5***	***No-TH***	***MF***	***S5***	***No-TH***	***MF***	***S5***	***No-TH***	***MF***
1.000	90	100	100	100	100	100	77	100	100
0.500	77	100	100	100	100	100	87	100	98
0.250	83	100	98	100	100	100	73	100	98
0.125	63	88	85	93	100	100	77	100	91
0.063	60	79	72	97	100	100	73	86	60
0.031	23	33	46	90	100	100	33	46	39
0.016	10	13	33	70	79	74	27	29	33

This problem was suspected to be the result of reduced fluorescent signals at the TH01 locus for the S5 kit, causing the locus to fail to be detected in most samples (Figure 3.1). For instance, only one of 21 amplifications of ABI 007 had no allelic dropout at the TH01 locus. This had been previously observed in a study conducted with SGM+ in this laboratory [218] and with one lot of SGM+ at the Forensic Science Service, UK (pers. comm. L. Welch). In order to overcome this aberration, data from the TH01 locus was removed from all calculations and the new data renamed as "No-TH" (Table 3.4). It was apparent that the new data gave higher percentage profiles than the original S5 data. As a consequence, the TH01 locus was excluded from all calculations in this study. The percentage profiles of "No-TH" were comparable to those of the MF kit at every amount of initial DNA template for the three different sources of DNA.

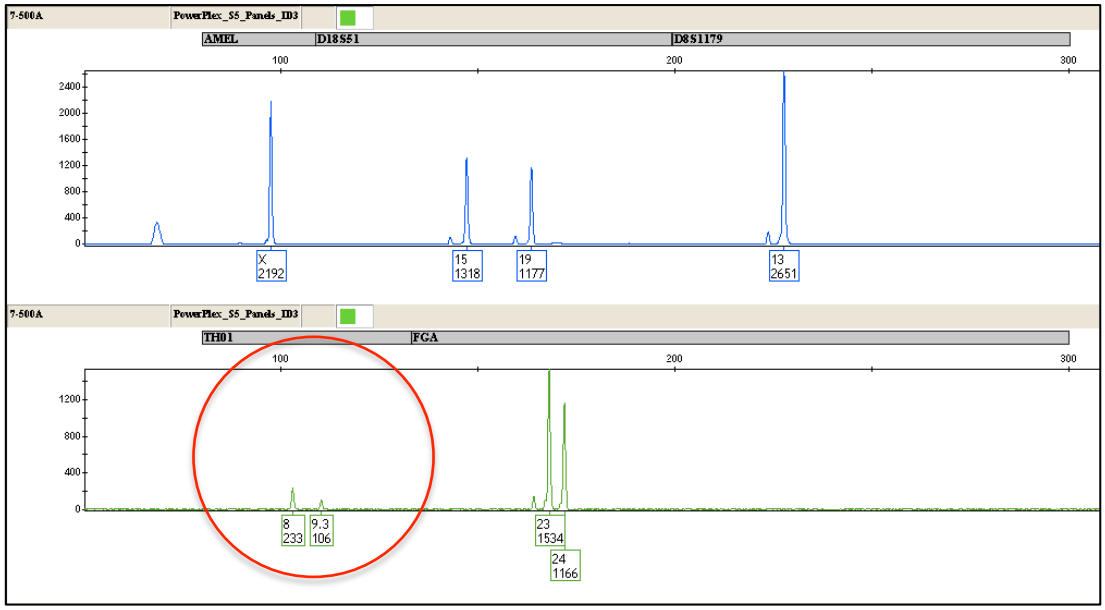


Figure 3.1. EPG of 9947A amplified with S5. Note the peak height of allele 8 and 9.3 (<250 RFUs) for TH01 (circled) while other alleles are above 1000 RFUs.

With CAMBIO DNA, full profiles were observed down to 0.031 ng initial DNA template for both kits. The other two DNA sources gave a lower than expected percentage profile, with the MF kit only able to produce a full profile at 0.5 ng (ABI 007) and 1 ng (9947A). This observation contradicted Mulero *et al.* [194] and Gehrig and Teyssier [197], who demonstrated that 0.063 and 0.040 ng DNA consistently yielded 100% profiles, respectively. Furthermore, the S5 kit only showed 100% profiles down to 0.25 ng (ABI 007) and 0.125 ng (9947A) of DNA, which was again inconsistent with the manufacturer's marketing claim of full profiles down to 0.05 ng DNA [191]. These disparities raised doubts concerning the accuracy of DNA concentration given by the manufacturers. This issue had been examined by NIST, who stated that there are variations in commercially assigned quantity in DNA standards and that the purported quantity is unreliable [184].

3.3.1.2 Peak heights

All dilution series exhibited decreasing mean peak heights with decreasing DNA templates (Table 3.5). In 15 of 21 reactions pooled from three sources of DNA, the S5 kit displayed higher mean peak heights (four reactions each of ABI 007 and 9947A and seven reactions of CAMBIO DNA). The highest difference of 2299 RFUs was seen at 0.5 ng CAMBIO DNA. In the other six reactions, MF produced higher mean peak heights with the highest difference of 532 RFUs at 1 ng ABI 007.

Both kits were able to give mean peak heights of over 150-200 RFUs for initial DNA template amount over the suboptimal range (>0.1 ng). Peak heights in these regions are sufficient for calling the two alleles in a homozygous locus, depending on the dropout threshold of each laboratory [203]. The mean peak heights of both kits at 0.063 ng CAMBIO DNA were above 400 RFUs. Based on an estimate by Tvedebrink *et al.* [209], a conservative probability of allelic dropout for this template amount was less than 0.0005 for S5 and 0.001 for MF, meaning that an analyst could be quite confident that no dropout had occurred.

Table 3.5. Mean peak height and standard deviation of three DNA sources amplified with S5 excluding the TH01 locus (No-TH) and MF kits.

	ABI 007		CAMBIO		9947A	
DNA (ng)	***No-TH***	***MF***	***No-TH***	***MF***	***No-TH***	***MF***
1.000	1426 ± 338	1958 ± 228	6975 ± 300	5493 ± 509	924 ± 91	1036 ± 50
0.500	934 ± 132	609 ± 215	4712 ± 285	2413 ± 767	514 ± 30	460 ± 25
0.250	531 ± 72	410 ± 103	2327 ± 275	1240 ± 290	297 ± 92	314 ± 42
0.125	217 ± 102	201 ± 6	1380 ± 146	836 ± 63	208 ± 59	190 ± 10
0.063	158 ± 26	107 ± 18	695 ± 113	437 ± 23	93 ± 26	96 ± 5
0.031	75 ± 12	94 ± 7	294 ± 48	174 ± 50	67 ± 23	63 ± 7
0.013	55 ± 49	73 ± 13	173 ± 57	129 ± 3	67 ± 24	57 ± 6

The widely different peak heights from different DNA sources further supported the idea that the manufacturers had inaccurately reported the control DNA, as observed by Kline *et al.* [184]. The peak heights of CAMBIO DNA were much higher than those obtained from the other two sources of DNA. For example, the mean peak height seen with 0.5 ng CAMBIO DNA was approximately 2000 RFUs while only 500-1000 RFUs were observed with the same amount of ABI 007 and 9947A. The peak heights of CAMBIO DNA were slightly higher than those reported by Mulero *et al.* [194], who obtained peak heights between 1500 and 2000 RFUs for 0.5 ng DNA. Given the results of percentage profiles and peak heights, it was likely that the assumed concentrations of all three commercial DNA sources were somewhat inaccurate, with ABI 007 and 9947A having lower concentrations and CAMBIO DNA having a higher concentration than that given by the manufacturers.

As ABI 007 and 9947A are intended to be used only as a PCR positive control and not for quantification, some inaccuracies are tolerated as long as a full profile is obtained by using the manufacturer's recommended amount. This again highlights the importance of internal validation, as claims of high sensitivity by manufacturers could be inaccurate if they have been based on serial dilutions of their own DNA standards. Caution must be exercised when comparing results from different laboratories. It was also possible that a dilution error led to these inaccuracies. Nonetheless, comparisons between S5 and MF were valid because they were only made within each DNA source.

3.3.2 Increased cycle study

3.3.2.1 Percentage profiles

Three lowest concentrations of DNA used in the sensitivity study were amplified using the manufacturer's recommendation of 30 cycles and 34 cycles. These concentrations corresponded to approximately the amount of DNA found in ten, five and three diploid cells, in which stochastic effects were expected and from which increased cycles had shown improvements in the DNA profiles obtained [198, 219, 220]. Both S5 and MF performed well with low concentrations of DNA without additional PCR cycles (Table 3.6).

At the manufacturers' recommendation of 30 cycles and 0.016 ng DNA, over 70% percentage profile was seen for both kits. Almost full profiles were obtainable at this level of template DNA when the PCR cycles were increased to 34. At the two other initial template amounts, there was almost no difference in percentage profiles between the two PCR cycles. With the only discernible increase in percentage profiles at 0.016 ng DNA, it was concluded that both kits were sensitive enough to amplify at least five cells' worth of DNA (0.031 ng) without the need for an increase in the optimal number of PCR cycles.

Table 3.6. Percentage profiles of the two mini-STR kits from samples amplified with 30 and 34 cycles PCR. Number in parentheses indicates the number of allelic dropin from three amplifications.

	30 cycles		34 cycles	
DNA (ng)	***S5***	***MF***	***S5***	***MF***
0.063	100 (4)	100 (0)	100 (4)	100 (7)
0.031	93 (4)	100 (0)	100 (5)	100 (3)
0.016	70 (4)	74 (1)	97 (6)	94 (4)

3.3.2.2 Allelic dropin

Allelic dropin refers to spurious additional alleles seen in a DNA profile [221]. Only one or two alleles are seen per profile, as opposed to gross contamination, in which extra alleles are seen for most loci. The sources of these spurious alleles could be from the manufacturing process of plastic-wares or DNA transferred locally via dust particles [198, 222].

The total allelic dropin observed from a triplicate of three DNA amounts were counted. At 30 cycles PCR, the number of dropin for the S5 kit was 12, which was greater than the single dropin observed for the MF kit. Increasing the number of PCR cycles resulted in an increase in dropin for both kits – 15 alleles for S5 and 14 alleles for MF. The MF kit was more affected by allelic dropin with increased cycle. Nonetheless, all alleles observed were never repeated in triplicate amplification of the same DNA template (see Appendix Table 6); therefore, the dropin alleles would not be reported if the "consensus profile" method were used [198, 210].

This increase was expected as it had been demonstrated before by Gill *et al.* [198], who observed 30 alleles in 30 replicates of negative control. In contrast, Petricevic *et al.* [206] only saw 13 alleles in 97 negative controls using the same technique. The discrepancy between the two studies were most likely due to the increase in stringency and improvement in decontamination procedures over the ten years between the two publications.

3.3.2.3 Peak heights

Increasing the number of cycle to 34 resulted in an approximately tenfold increase in mean peak heights of those obtained with the recommended 30 cycles (Table 3.7). For example, the lowest average seen with 0.016 ng of initial template amplified

using 30 cycles with the MF kit (129 RFU) was increased to 1761 RFUs with the extra four cycles. This gain in peak heights was comparable for both kits.

Table 3.7. Mean peak height and standard deviation of samples amplified with 30 and 34 cycles PCR.

	30 cycles		34 cycles	
DNA (ng)	***S5***	***MF***	***S5***	***MF***
0.063	586 ± 102	437 ± 23	5708 ± 271	4692 ± 178
0.031	271 ± 37	174 ± 50	3292 ± 790	2788 ± 152
0.016	166 ± 55	129 ± 3	2058 ± 1215	1761 ± 545

3.3.2.4 Heterozygous peak balance (Hb_x)

A mixed result in terms of changes in Hb_x was obtained (see Appendix Table 4 and Appendix Table 5). While a decrease in Hb_x was expected, the data obtained here did not support this. Five of 12 loci (42%) pooled from three DNA template amounts showed a decrease in Hb_x for the S5 kit, while 15 of 27 loci (56%) showed a decrease for the MF kit.

The S5 kit had mean Hb_x of 0.70 and the MF kit had a mean Hb_x of 0.61 when 34 PCR cycles were used. 78% of Hb_x for S5 was over 0.5, the National recommendations of the Technical UK DNA working group [203]. In contrast, only 68% of Hb_x for MF was over this threshold (Figure 3.2).

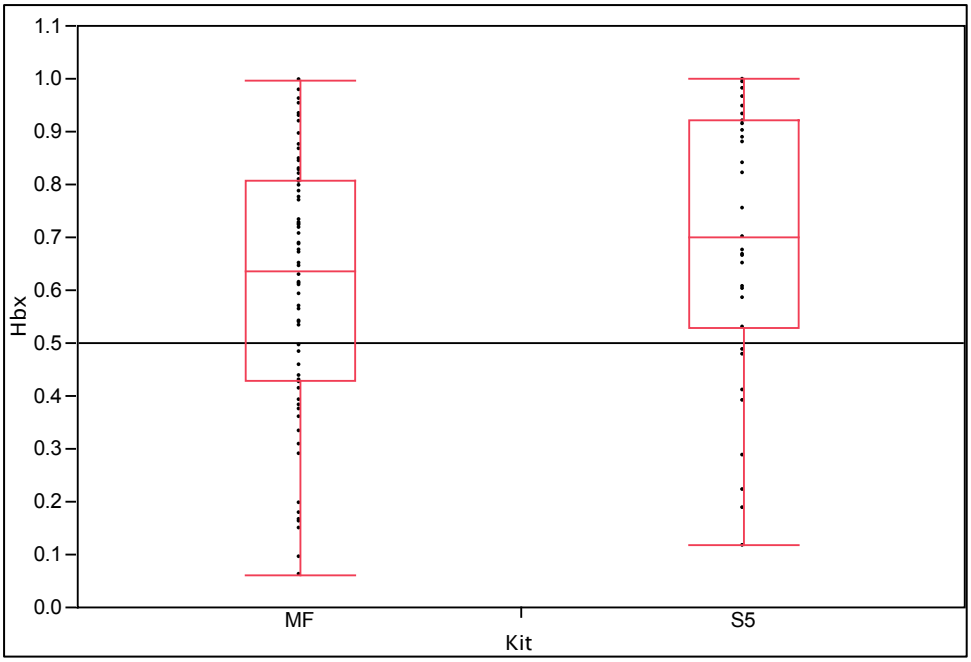


Figure 3.2. Boxplot of Hb_x pooled from triplicate amplification (34 cycles) of three initial DNA quantities (0.063, 0.031 and 0.013 ng) using the MF (N=78, mean = 0.61) and S5 (N=35, mean = 0.70) kits. Reference line shown at 0.5.

Both kits could yield full profiles and acceptable peak heights (>150 RFUs) with low quantities of DNA using the manufacturers' recommended protocols. In addition, increased PCR cycle showed no difference in terms of Hb_x while there was a significant increase in allelic dropin for all DNA quantities. It was clearly demonstrated that there was no benefit in adding extra PCR cycles to the optimal 30 cycles for both S5 and MF at 0.063 and 0.031 ng of DNA, while the ten-fold increase in peak heights and additional alleles gained for 0.016 ng should be an acceptable compromise for the increased dropin.

3.3.3 Simulated casework study

3.3.3.1 Quantification

Plexor® HY was used to simultaneously quantify the autosomal and Y targets of the simulated casework samples (Table 3.8). Internal PCR control (IPC) data indicated that all simulated casework samples used for this study were free from PCR inhibitors. The analysis of melt-curves for all samples did not show any non-specific products or primer-dimers.

Table 3.8. Autosomal and Y-target quantification results (in ng/μL) for each casework sample. The ratio of the two quantities is also given in [Auto]/[Y]. Template added indicates how much DNA (in ng) was added to PCR reactions.

Sample	Autosomal	Y	[Auto]/[Y]	Template added
S1-1	0.569	0.345	1.6	0.569
S1-2	0.407	-	-	0.407
S2-1	0.090	-	-	0.448
S2-2	0.539	0.042	**12.9**	0.539
S3-1	2.520	2.310	1.1	0.504
S3-2	2.930	4.250	0.7	0.586
S7-1	0.013	-	-	0.066
S7-2	0.442	0.344	1.3	0.663
S7-3	6.480	3.010	**2.2**	1.296
S7-4	1.050	1.630	0.6	0.525
S7-A	0.874	-	-	0.874
S7-B	0.231	0.188	1.2	0.462
S7-C	0.097	0.127	0.8	0.486
S6-1	0.058	0.058	1	0.290
S6-2	0.090	-	-	0.448
S6-3	0.202	0.143	1.4	0.505
S6-4	0.634	0.732	0.9	0.634
S6-A	0.280	0.527	0.5	0.560
S6-B	0.207	0.450	0.5	0.414
S6-C	4.540	-	-	0.908

Six simulated casework samples were obtained from female donors, while the other fourteen were either from a male donor or were mixtures with a male contributor, indicated by the presence of detectable Y-target. Most samples had an autosomal to Y ratio of about 1, meaning that there were equal parts of autosomal and Y DNA in the DNA extract. S2-2 and S7-3, however, displayed imbalanced autosomal to Y ratios of 12.9:1 and 2.2:1, respectively (bolded in Table 3.8). These contrasting ratios signified a male/female mixture with the male being a minor contributor. Other autosomal to Y ratios were within the normal variation (0.4 to 2.0) observed in the

94% of the population [121]. This meant that they were either from only male donors (both single and mixed) or had an almost equal contribution from male and female donors.

The lowest DNA concentration was observed in sample S7-1 with only 0.013 ng/μL, while the highest concentration was 6.480 ng/μL in sample S7-3. An optimal template amount (~0.5 ng) was added to the PCR reactions for most samples, except S6-1 (0.290 ng), S7-1 (0.066 ng) and S7-3 (1.296 ng).

3.3.3.2 Percentage profiles

The reference DNA profiles for casework samples were obtained using PP-16. Only loci in common with PP-16 were used for the calculation of percentage profiles (Table 3.1). The overlapping loci for the kits were as follows:

- S5 – FGA, D8S1179, TH01, D18S51, AMEL
- MF – FGA, CSF1PO, D7S820, D13S317, D16S539, D18S51, D21S11 and AMEL
- EP01/EP02 – D3S1358, FGA, D8S1179, TH01, vWA, D16S539, D18S51, D21S11 and AMEL
- ESS – None

Alleles were called using the "consensus profile" method from duplicate amplification. Both S5 and MF gave full profiles in 18 of 20 samples (90%), followed by EP01 with full profiles in 9 of 12 samples (75%). EP02 had the lowest number of full profiles in 8 of 12 samples (67%). Some samples had all alleles present in one replicate but because they were not seen in both replicates, they were not scored in the "consensus profile" (starred in Table 3.9).

Table 3.9. Percentage profiles (%) of casework samples typed with S5, MF, EP01 and EP02. Scoring was based on alleles in common with reference profiles. Asterisk (*) indicates that all missing alleles were detected in one amplification but not scored in the consensus profile.

Sample	S5	MF	EP01	EP02	Sample	S5	MF	EP01	EP02
S1-1	100	100	-	-	**S1-2**	100	100	-	-
S2-1	100	100	-	-	**S2-2**	*95**	*75**	-	-
S3-1	100	100	100	100	**S3-2**	100	*97*	*97*	*94*
S7-1	100	100	100	*94**	**S6-1**	100	100	100	100
S7-2	100	100	*88*	*85*	**S6-2**	100	100	100	100
S7-3	100	100	100	100	**S6-3**	*75**	100	*72*	*75*
S7-4	100	100	100	100	**S6-4**	100	100	100	100
S7-A	100	100	100	100	**S6-A**	100	100	100	100
S7-B	100	100	-	-	**S6-B**	100	100	-	-
S7-C	100	100	-	-	**S6-C**	100	100	-	-

Sample S7-1, from which the initial DNA template added to each PCR reaction was only 0.066 ng, gave a full profile for all kits except for EP02. This demonstrated that even with a sub-optimal amount of DNA, a full profile could still be obtained with the kits. Full profiles were also obtained with all kits from a three-person mixture (S7-3).

MF correctly typed two rare alleles, allele 28.3 at D21S11 in sample S7-C (Figure 3.3) and allele 16 at D16S539 in sample S6-4. The allele frequency was approximately 1 in 19000 and 1 in 12000, respectively. The EP01 and EP02 kits also correctly reported this rare allele 16 at D16S539. Unfortunately, these two loci were not included in the S5 and the ESS kit and thus a comparison was not possible.

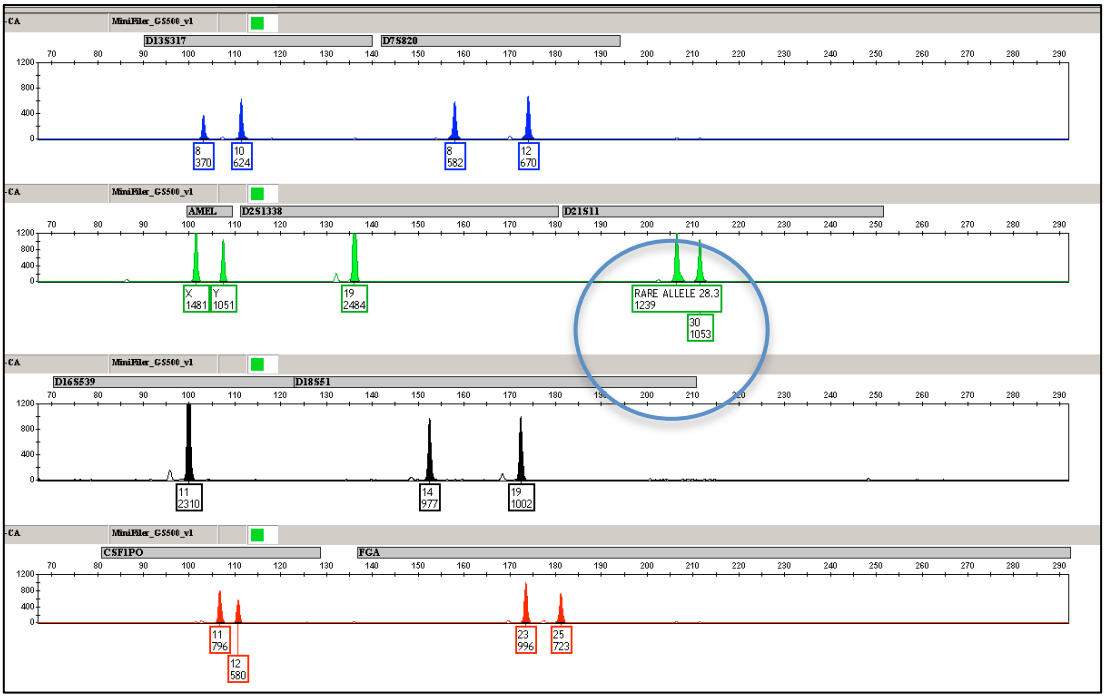


Figure 3.3. EPG of S7-C typed with MF. Rare allele 28.3 was observed at D21S11 (blue circle).

Sample S6-3, which was a mixed semen and saliva stain, was the most difficult to amplify as shown by the failure of all kits except MF to score a full profile. This was most likely due to this sample being a low-level mixture (Figure 3.4). It was probable that the major contributor's DNA was preferentially amplified and hence the minor profile was either not amplified or masked. Although the total DNA template added for S6-3 was 0.505 ng, it was not possible to tell how much DNA was from the major and minor contributors. More DNA could have been added to bring up signal from the minor contributor, as only 5 μL was used as 10 μL had been previously demonstrated to yield full profiles with PP-16. The maximum amount allowed for S5, EP01, and EP02 was 17.5 μL and this should be more than sufficient to generate a full profile in any kit. In most cases the profiles obtained were indistinguishable from reference profiles of PP-16 obtained earlier with twice the amount of DNA added.

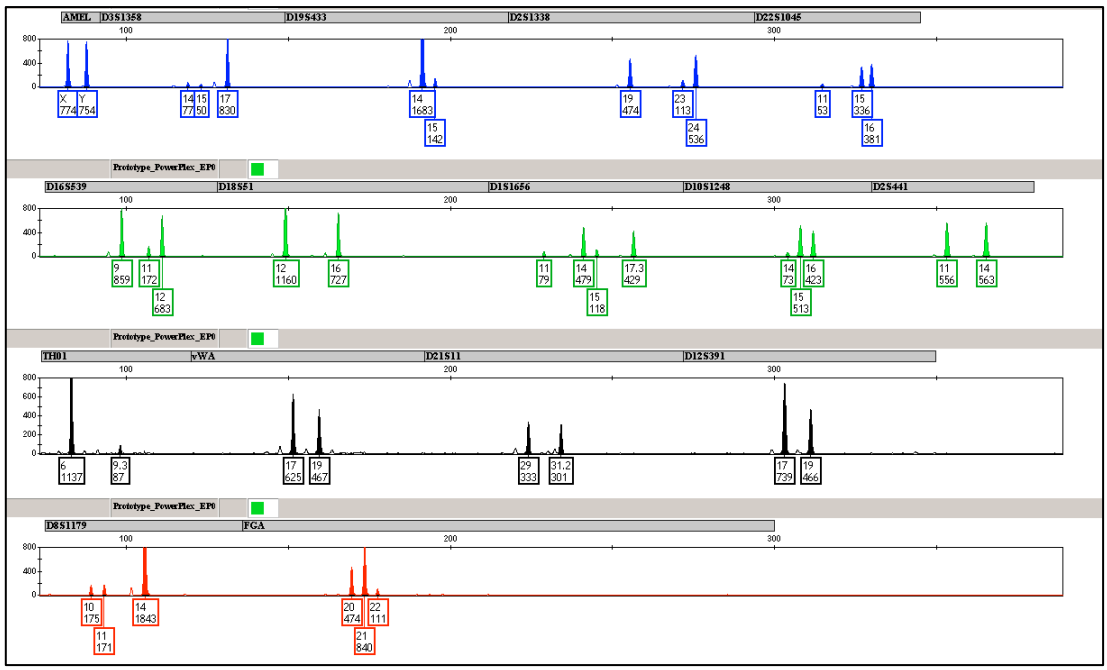


Figure 3.4. EPG of sample S6-3 (low-level mixture) typed with EP02. Note the low-level mixture at most loci.

From the percentage profiles, it seemed that the S5 and MF kits performed better, but these two kits detected a smaller number of loci (4 and 8 STRs, respectively) while the EP01 and EP02 kits were designed to amplify 15 STRs each. The random match probability (P_m) – the probability of finding an identical genotype from an unrelated individual in the population – was lower for the S5 and MF kits. P_m is calculated by multiplying the genotype frequency across all loci. The genotype frequency for a heterozygous locus is $2pq$ and for a homozygous locus is $p^2 + p(1-p)\theta$, where p and q = allele frequency and θ = correction factor for population substructure (NRC II Recommendation 4.1). A conservative value of 0.01 for θ is used for a large population and 0.03 for a smaller or more inbred population.

OmniPop version 200.1 (http://www.cstl.nist.gov/strbase/populationdata.htm), an Excel macro developed by Brian Burritt, was used to calculate the P_m of S7-1, in which a 100% percentage profile was seen with all kits except EP02, based on the allele frequencies for US Caucasian [223]. The P_m of S5, MF and EP01 were 1.79 x 10^5, 3.49 x 10^7 and 4.40 x 10^9, respectively. EP02, with 94% percentage profile, had a P_m of 1.11 x 10^9, which was higher than both MF and S5.

3.3.3.3 Peak heights

Only single-source profiles were chosen for this analysis (S3-1, S6-1, S6-2, S6-A, S7-1, S7-4 and S7-A) (Figure 3.5). Except for S6-1 and S7-1, the amount of DNA template added was optimal (~0.5 ng).

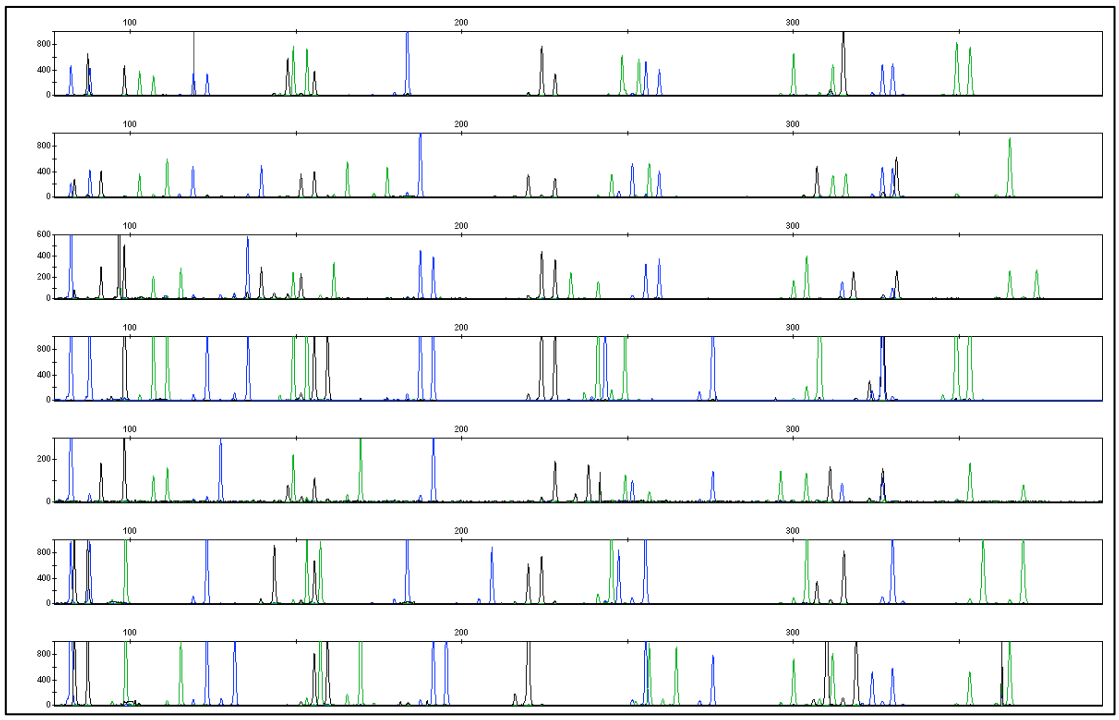


Figure 3.5. EPGs of single-source simulated casework samples typed with EP02. Top to bottom: S3-1, S6-1, S6-2, S6-A, S7-1, S7-4 and S7-A. All y-axes were from 0 to 1000 RFU, except for S6-2 (3rd panel - 600 RFU) and S7-1 (5th panel - 300 RFU).

The highest mean peak height was seen with the ESS kit (1456 RFU), followed by the S5 kit (1020 RFU). EP01 had the third highest mean peak height (976 RFU). The lowest mean peak heights were seen with the MF (930 RFU) and the EP02 kit (687 RFU). As expected, the standard deviations for these were widespread due to the different qualities of the seven casework samples (Figure 3.6).

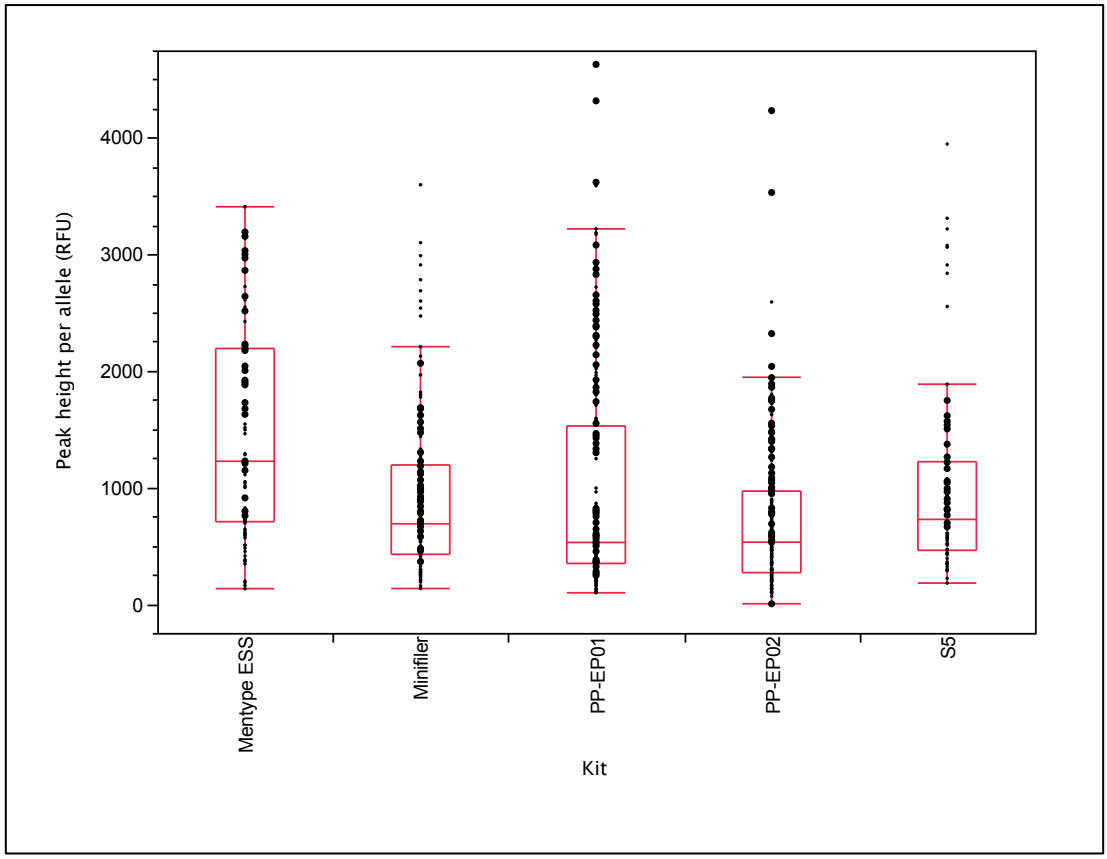


Figure 3.6. Boxplot of mean peak heights per allele (in RFU) categorized by each kit. N = 90, 126, 224, 224, and 70 respectively.

The data were not normally distributed and the variances were not equal (data not shown). A non-parametric test for related samples (Friedman's test) was used to compare the peak heights of each kit. The null hypothesis *"the distribution of mean peak heights is the same across categories of kits"* could be rejected ($p = 0.008$, Chi-squared = 13.829, DF = 4, N = 7 samples for each kit). Pairwise comparisons confirmed that the peak heights obtained with the ESS kit were significantly different from EP02 ($p = 0.004$). All other pairs were not significantly different ($\alpha = 0.05$). Based on peak heights, the ESS kit excelled in typing casework samples. This could be because the amplicon sizes of the ESS kit were the shortest of all the kits (Table 3.2), which agreed with the results from Butler *et al.* and Dixon *et al.* using 15 year-old bloodstains and artificially degraded DNA respectively [17, 35]. The two studies observed increased allelic dropout and reduced peak heights from larger loci in STR kits. Internal validation of the MF kit by the Institut de Médecine Légale (Geneve, Switzerland) using heat degraded buccal swabs also demonstrated that the kit suffered from lowered peak heights and allelic dropout at larger loci [197].

3.3.3.4 Heterozygous peak balance (Hb_x)

All kits had a mean Hb_x of over 0.74 (Figure 3.7). The null hypothesis *"the distribution of Hb_x is the same across categories of kits"* could not be rejected (Friedman's test, p = 0.225, Chi-squared = 5.667, DF = 4, N = 7 samples for each kit). This indicated that all kits were comparable in terms of Hb_x when genotyping a single-source DNA sample.

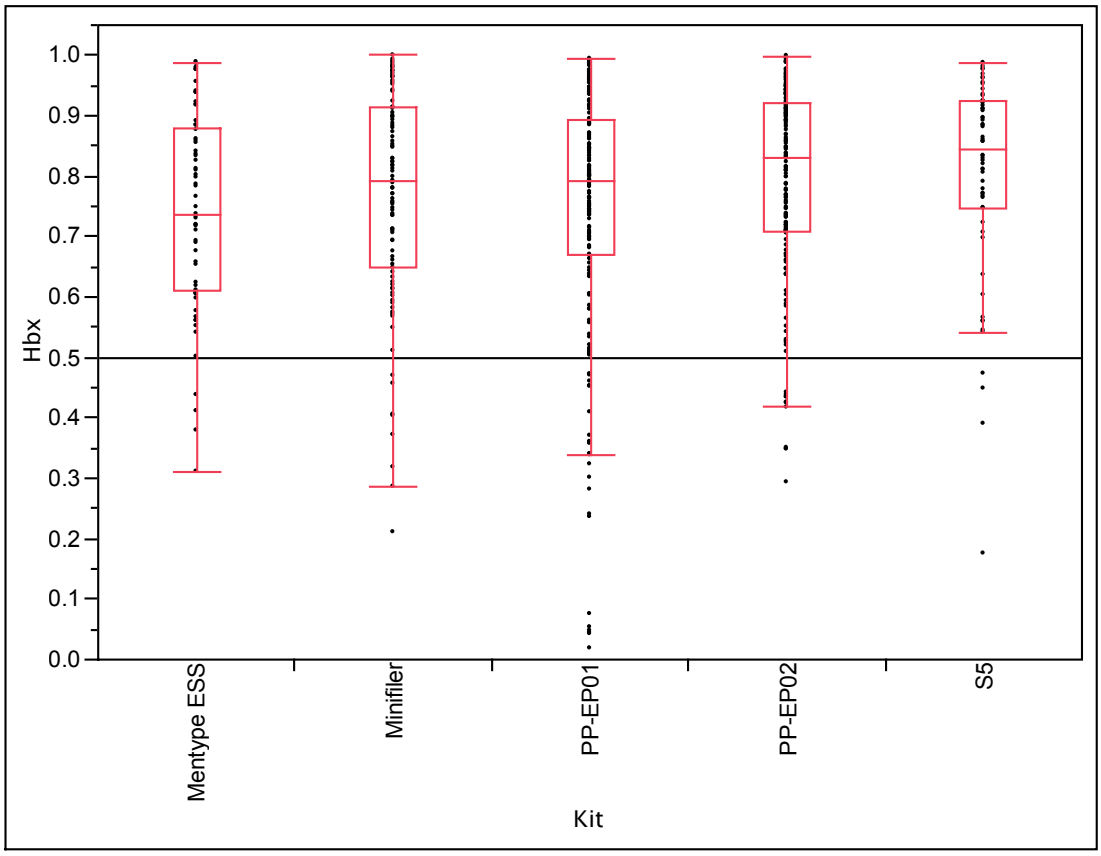


Figure 3.7. Boxplot of heterozygous peak balance (Hb_x) categorized by each kit. Mean = 0.74, 0.77, 0.74, 0.79, and 0.80, respectively. N = 65, 100, 189, 171, and 60, respectively. Reference line is shown at 0.5.

3.3.3.5 Stutter proportion (S_p)

The five kits had mean S_p of less than 0.10 (Figure 3.8), with the ESS kit having the highest mean S_p. Only 10 of 444 alleles (2.3%) had stutters of over 15%, ranging from 0.15 (D18S51/S5) to 0.26 (FGA/EP02). No significance difference between the kits was observed (Friedman's test, p = 0.075, Chi-squared = 8.480, DF = 4, N = 5 samples for each kit due to no stutter observed for S6-2 with MF kit and S7-1 for S5 and EP02 kits).

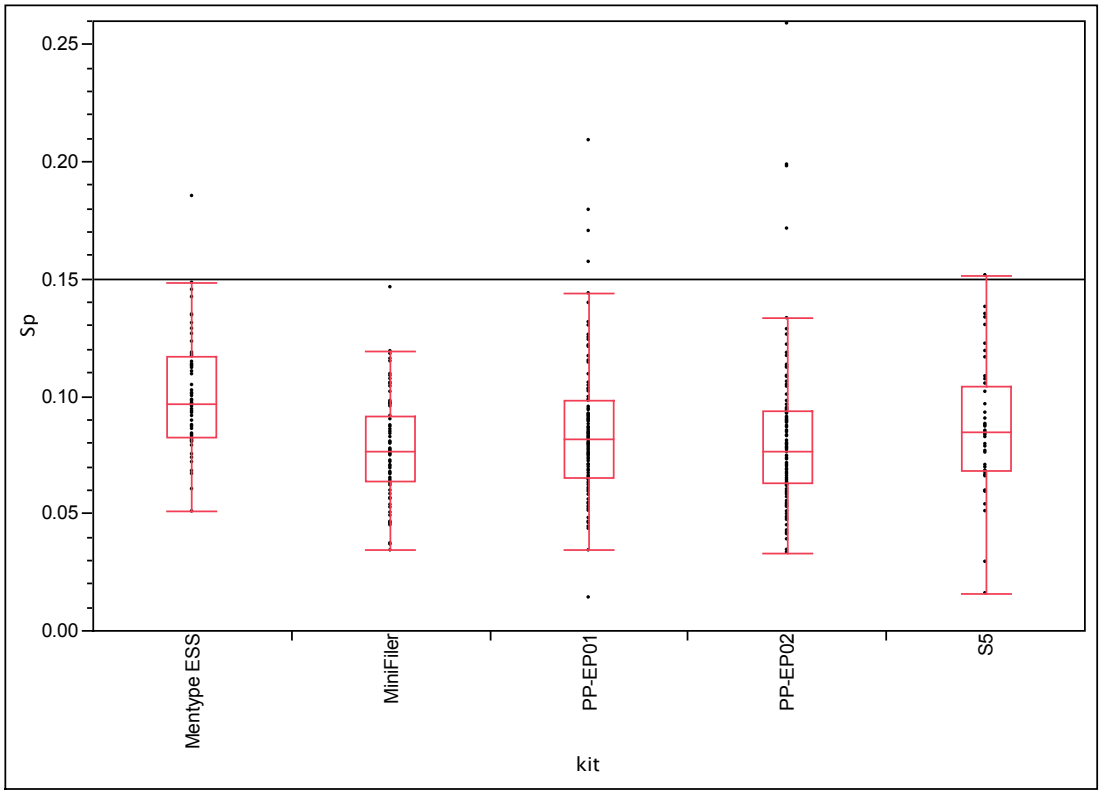


Figure 3.8. Boxplot of stutter proportion (S_p) categorized by each kit. Mean = 0.101, 0.078, 0.085, 0.082 and 0.087, respectively. N = 70, 84, 126, 120, and 44, respectively. Reference line at 0.15 is shown.

Wide variations in S_p were observed for all loci (Table 3.10 – data pooled from five kits). Stutter thresholds were calculated according to ABI recommendation that three standard deviations be added to the mean S_p to calculate the interpretation threshold for a locus [1]. Assuming a normal distribution of S_p, this meant that 99.73% of stutters would not exceed the threshold.

Table 3.10. The mean S_p and standard deviation for each locus. Threshold is three times the standard deviation added to the mean S_p. PP-16 and SGM+ are recommended threshold values from Greenspoon *et al.* [33] and SGM+ Manual [224].

Locus	N	Mean	Std Dev	Threshold	PP-16	SGM+
CSF1PO	7	0.07	0.011	0.10	0.11	-
D10S1248	32	0.09	0.018	0.15	-	-
D12S391	38	0.10	0.026	0.18	-	-
D13S317	8	0.06	0.013	0.09	0.09	-
D16S539	25	0.07	0.027	0.15	0.12	0.13
D18S51	39	0.08	0.026	0.16	0.13	0.16
D19S433	19	0.07	0.037	0.18	-	0.17
D1S1656	39	0.10	0.019	0.15	-	-
D21S11	20	0.08	0.017	0.13	0.12	0.13
D22S1045	32	0.10	0.029	0.19	-	-
D2S1338	40	0.09	0.024	0.16	-	0.15
D2S441	21	0.07	0.025	0.15	-	-
D3S1358	22	0.09	0.015	0.13	0.10	0.11
D7S820	4	0.05	0.013	0.09	0.09	-
D8S1179	27	0.07	0.017	0.13	0.08	0.12
FGA	48	0.09	0.033	0.19	0.09	0.11
TH01	7	0.06	0.056	0.23	0.05	0.06
VWA	16	0.11	0.038	0.22	0.14	0.11

The results were comparable to those reported for PP-16 by Greenspoon *et al.* [33] and SGM+ panel data for Genemapper® ID 3.2.1. Most thresholds were higher than those of the two commercial kits, due to the higher standard deviations of the next-generation and mini-STR kits. This could be the result of a combined effect of the smaller sample size and multiple kits used in this study. For instance, the number of stutters observed here ranged from 4 to 48 for each locus while the range of PP-16 stutters were from 49 to 232. Moreover, since this study used stutter data compiled from five different kits, the variations were expected to be high, although it is notable that more than 97% all of the stutters were below the guidelines for stutter interpretation [203].

3.3.3.6 Effect of short amplicons

EP01 and EP02 incorporated the same loci but the product sizes and dyes had been deliberately shifted between the two kits by the manufacturer in order to have different loci as mini-STRs [225]. This gave a unique opportunity to confirm the finding that shorter amplicons survive degradation better than longer amplicons [17,

35, 149, 150], as direct comparisons could be made using the same loci while the chemistries of the kits being compared were very similar. Moreover, it could serve as an indicator of the degree of degradation of a sample, if there was a difference in survival rates between a short and a long amplicon (mini-STR vs. conventional STR) in these two kits.

Simulated casework samples were used to determine if there was a pattern in the differences in peak heights when the two kits were compared, as it was known that preferential amplification of short, low molecular weight loci occurred in degraded samples [1]. In other words, it was expected that higher peak heights should be observed when a locus was a mini-STR and that loci that were very different in size between the kits should show a significant difference in peak heights. The loci in which EP01 were much shorter were D1S1656, D12S391, D10S1248, D22S1045 and D2S441 (91-252 bp) (top rows of Table 3.11). The loci in which EP02 were much shorter were FGA, D8S1179, D18S51 and D16S539 (117-190 bp) (bottom rows of Table 3.11). The other loci were approximately the same size (up to 78 bp difference).

Table 3.11. Paired t-test results between EP01 and EP02 peak heights by locus. N = number allele peaks, size diff = mean size difference of the alleles, t = t-statistics, 95% CI = 95% confidence interval and *p* = p-value of paired t-test. Bolded p-values indicates significance at alpha of 0.05.

	Locus	N	Size diff (bp)	t	95% CI		*p*
<-----EP02 shorter / EP01 shorter ----->	D2S441	24	-252	1.96	0.98	2.00	0.063
	D22S1045	23	-224	8.67	2.77	5.27	**0.000**
	D10S1248	23	-202	2.13	1.01	1.73	**0.045**
	D12S391	22	-158	-1.65	0.62	1.06	0.114
	D1S1656	24	-91	2.29	1.03	1.67	**0.031**
	D2S1338	26	-28	2.13	1.01	1.46	**0.043**
	D21S11	24	-5	0.53	0.79	1.50	0.601
	vWA	26	0	2.83	1.06	1.43	**0.009**
	AMEL	20	0	2.98	1.11	1.77	**0.008**
	D3S1358	22	0	5.21	1.36	2.05	**0.000**
	D19S433	20	35	2.59	1.06	1.66	**0.018**
	TH01	24	78	2.33	1.02	1.43	**0.029**
	FGA	26	117	0.78	0.90	1.27	0.443
	D8S1179	24	126	-3.11	0.61	0.90	**0.005**
	D18S51	26	150	-3.60	0.58	0.86	**0.001**
	D16S539	25	190	0.71	0.88	1.29	0.485

Most loci had transformed peak heights that were normally or approximately normally distributed, as evaluated by the Shapiro-Wilk test and estimation of density plots (data not shown). A paired t-test with unequal variance was used to compare the transformed peak heights of the two kits by loci. The p-values and 95% confidence intervals were also calculated (Table 3.11). With log-transformed data, the confidence interval indicated the ratio of peak heights of EP01 to EP02 [226], i.e. an interval that included one meant that there was no significant difference in the peak heights between the kits. It was obvious that the significant differences seen were not dependent on allele size alone as both loci with almost no difference in size and ones with large differences showed statistically significant differences between the two kits. Moreover, the two loci with the most extreme size differences – D2S441 and D16S539 – did not exhibit a significant difference in peak heights.

This lack of a distinctive pattern of preferential amplification could be attributed to the quality of the simulated casework samples used. EPGs of these samples did not show conspicuous degradation patterns expected in degraded samples (Figure 3.5). The result here confirmed that the simulated casework samples were of high quality. It would be interesting to carry out this comparison using known, artificially degraded samples.

4 *In silico* evaluation of nucleosome protection theory of forensic STRs

4.1 Introduction

Recent advancements in forensic DNA analysis have focused on improving analysis techniques, such as pyrosequencing [227], increased PCR cycles [198], post-PCR purification [201] and mini-STR designs [35]. These improvements have proved to be successful in obtaining better DNA profiles with degraded DNA samples often found in mass disasters and samples exposed to the environment. However, the intrinsic structural properties of DNA that might prevent its degradation have been overlooked. Using these structural properties as a guide, forensic scientists might be able to choose the loci that can better withstand degradation and hence obtain more information from a degraded sample.

The binding of octameric histone cores to 147 bp of DNA in order to form a nucleosome is a complex, multifactorial process that limits the interactions of DNA with other proteins. The formation and location of these nucleosomes is known to depend on the following factors: dinucleotide periodicity, base stacking, GC content, and chromatin remodelers [51, 56-64, 66, 68]. It has been shown that certain properties, such as low deformation energy [228] and periodicity repeats of GG/CC dinucleotides [59], favour nucleosome formation. They are called "nucleosome positioning signals" (NPS) [61] (Section 1.5.3).

Dixon *et al.* [16, 17] suggested that a nucleosome could offer protection to the 147 bp of DNA bound to it from the attacks of endonucleases, which would freely digest DNA at exposed sites (Figure 1.8 on page 18). This is possible because the structural changes of DNA to form nucleosomes exclude them from being accessed by protein [229]. Nucleosomes usually occupy the promoter regions of strictly regulated genes, preventing access to transcription-factor binding sites and consequent binding of enzymes [86]. On the other hand, the promoter regions of constitutive genes are free of nucleosomes, permitting unregulated access by transcription factors [80]. An *in silico* whole human genome annotation for nucleosome exclusion regions also showed that regions free of nucleosomes correlated well with DNase I hypersensitive sites [230]. From this information, an inference can be made that DNA bound in the nucleosomes could be protected against DNases.

This *in silico* study was carried out to evaluate the "nucleosome forming potentials" (NFPs) (how likely it is for a certain sequence of DNA to be bound by nucleosomes) of 60 forensically important markers (58 STRs plus amelogenin X and Y). After analysis of the software available, two NPSs – DNA bendability based on known stiff sequences and dinucleotide base stacking – were analysed by two freely available tools, NXSensor [93] and nuScore [94], respectively.

NXSensor searches for three sequences that are known to be rigid and therefore resist bending into a nucleosome. These sequences, when located near each other, could indicate a nucleosome-free region of DNA [93]. A modified version of NXSensor has been shown *in silico* to achieve good correlations with regions lacking nucleosomes [230]. NuScore works by determining the energy needed to bend a sequence of DNA. This deformation energy is calculated based on the specific arrangements of dinucleotides and their interactions, a phenomenon called dinucleotide stacking. The six possible interactions between the neighbouring two bases are tilt and shift (x-axis), roll and slide (y-axis), and twist and rise (z-axis) [231] (Figure 4.1). Locations of minimal deformation energy have been shown to correspond well to empirically determined locations of nucleosome dyads, the centre of the nucleosome [94].

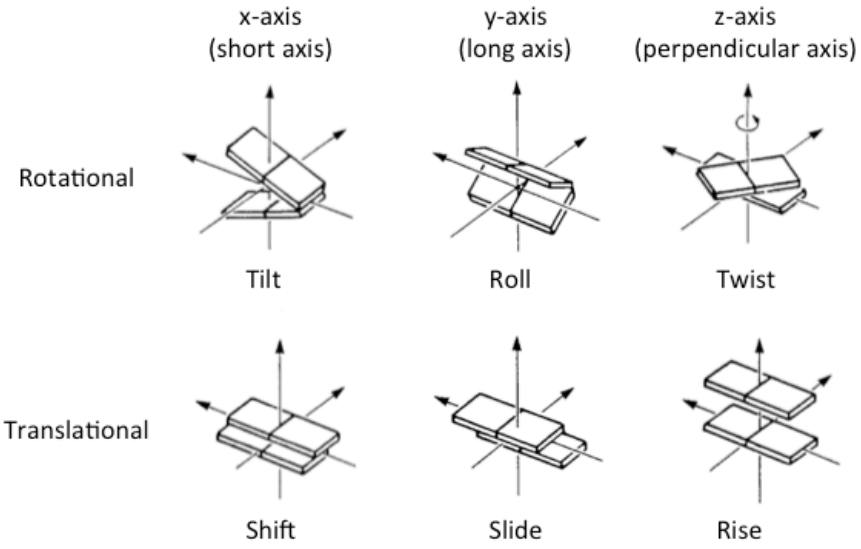


Figure 4.1. Six possible dinucleotide interactions on three axes. Adapted from Dickerson [232].

4.1.1 Aims

It is hypothesized that some forensically important STR loci may be more protected by nucleosomes than other loci. Determining which loci are protected could allow them to be incorporated into future forensic identification kits, resulting in a higher discrimination power for certain degraded sample types (saliva, bone, and decomposed remains) than with current profiling methods. The result of this *in silico* evaluation will be used for further experiments to empirically determine the protection that nucleosomes confer on DNA in degraded forensic samples.

4.2 Materials and methods

4.2.1 Selecting markers and obtaining base sequences

Fifty-eight STR markers and amelogenin X and Y, totalling 60 sequences (Table 4.1), were selected based on past use and current recommendations by the forensic community. Sequences were obtained from the NCBI Human Genome Map (http://www.ncbi.nlm.nih.gov/genome/guide/human/). These sequences were centre-aligned at the tandem repeat units and truncated for 200 bp at both the 5' and 3' end, yielding a sequence of 400 bases.

4.2.2 NXSensor mechanisms and parameters

The algorithm of Nucleosome eXclusion Sensor (NXSensor version 1.3.1) (http://www.sfu.ca/~ibajic/NXSensor/) reads an input sequence for three known nucleosome-free sequences: 10 bases of poly-A, 10 bases of poly-T, and a combination of Gs and Cs ($A_{\geq 10}$, $T_{\geq 10}$, or $[(G/C)_3N_2]_{\geq 3}$). If any of these sequences are found, the program outputs the sequence in FASTA format and highlights the nucleosome-free region.

All 60 sequences were evaluated and accessibility scores were given as a measure of how accessible the input sequence was to DNA-binding proteins. The score was calculated using the following formula:

$$A = \frac{L_o}{L_i - L_a}$$

Equation 4.1. Accessibility score of NXSensor.

where A = accessibility score; L_o = the total length (in bp) of open contiguous segment that are longer than the minimum length of open segment; L_i = the length of input sequence; and L_a = the total length of ambiguous segments. An accessibility score of zero indicated the whole input sequence contains no sequence that inhibits nucleosome formation while a score of one indicated the whole sequence is stiff and hinders the formation of a nucleosome. A marker was deemed to have a high NFP if the accessibility score was close to zero and a low NFP close to one.

Table 4.1. The 58 STR markers plus amelogenin X and Y, their GenBank accession number and chromosomal position in the latest human GRCh37 assembly.

Locus name	Genbank	Chromosomal position	Locus name	Genbank	Chromosomal position
CD4	M86525	Chr 12: 6.897 Mb	D2S441	AC079112	Chr 2: 68.239 Mb
CSF1PO	X14720	Chr 5: 149.455 Mb	D3S1358	AC099539	Chr 3: 45.582 Mb
D10S1248	AL391869	Chr 10: 131.093 Mb	D3S1545	L16413	Chr 3: 161.673 Mb
D10S1435	AL354747	Chr 10: 2.243 Mb	D3S3053	AC069259	Chr 3: 171.751 Mb
D11S4463	AP002806	Chr 11: 130.872 Mb	D3S4529	AC117452	Chr 3: 85.852 Mb
D12ATA63	AC009771	Chr 12: 108.322 Mb	D4S2364	AC022317	Chr 4: 93.517 Mb
D12S391	G08921	Chr 12: 12.450 Mb	D4S2366	G08339	Chr 4: 6.485 Mb
D13S317	AL353628.7	Chr 13: 82.692 Mb	D4S2408	AC110763	Chr 4: 31.304 Mb
D14S1434	AL121612	Chr 14: 95.308 Mb	D5S2500	AC008791	Chr 5: 58.699 Mb
D16S539	AC024591	Chr 16: 86.386 Mb	D5S818	AC008512	Chr 5: 123.111 Mb
D17S1301	AC016888	Chr 17: 72.681 Mb	D6S1017	AL035588	Chr 6: 41.677 Mb
D17S974	AC034303	Chr 17: 10.519 Mb	D6S474	AL357514	Chr 6: 112.879 Mb
D18S51	AP001534	Chr 18: 60.949 Mb	D7S820	AC004848	Chr 7: 83.789 Mb
D18S853	AP005130	Chr 18: 3.990 Mb	D8S1115	AC090739	Chr 8: 42.536 Mb
D19S433	AC008507	Chr 19: 30.416 Mb	D8S1179	AF216671	Chr 8: 125.907 Mb
D1GATA113	Z97987	Chr 1: 7.443 Mb	D9S1122	AL161789	Chr 9: 79.689 Mb
D1S1171	AF017307	Chr 1: 201.917 Mb	D9S2157	AL162417	Chr 9: 136.035 Mb
D1S1627	AC093119	Chr 1: 106.964 Mb	F13A1	M21986	Chr 6: 6.321 Mb
D1S1656	G07802	Chr 1: 230.905 Mb	FES	X06292	Chr 15: 91.432 Mb
D1S1677	AL513307	Chr 1: 163.560 Mb	FGA	M64982	Chr 4: 155.509 Mb
D20S1082	AL158015	Chr 20: 53.866 Mb	HPRTB	M26434	Chr X: 133.615 Mb
D20S161	L16405	Chr 20: 16.621 Mb	LPL	D83550	Chr 8: 19.815 Mb
D20S438	L29933	Chr 20: 38.051 Mb	Penta D	AP001752	Chr 21: 45.056 Mb
D20S482	AL121781	Chr 20: 4.506 Mb	Penta E	AC027004	Chr 15: 97.374 Mb
D21S11	AP000433	Chr 21: 20.554 Mb	SE33	V00481	Chr 6: 88.987 Mb
D21S1437	G08082	Chr 21: 21.646 Mb	TH01	D00269	Chr 11: 2.192 Mb
D221045	AL022314	Chr 22: 37.536 Mb	TPOX	M68651	Chr 2: 1.493 Mb
D2S1242	L17825	Chr 2: 221.217 Mb	vWA	M25858	Chr 12: 6.093 Mb
D2S1338	G08202	Chr 2: 218.879 Mb	Amelogenin X	M55418	Chr X: 11.311 Mb
D2S1776	AC009475	Chr 2: 169.145 Mb	Amelogenin Y	M55419	Chr Y: 6.736 Mb

The default settings used were:

- 147 bp window size
- Minimum length of open segments = 10
- Minimum number of exclusion sequences considered significant = 1.

A 147 bp window size was chosen as this was the same length of DNA in a nucleosome. A minimum length of open segment (not bound by nucleosome) of 10 bp was chosen as a sequence shorter than this between two nucleosome-bound regions was unlikely to be effectively accessed by a protein. The last option of minimum exclusion sequence of one meant that the presence of one known nucleosome-free sequence was sufficient to exclude a nucleosome formation within the window size.

4.2.3 NuScore mechanisms and parameters

NuScore (http://compbio.med.harvard.edu/nuScore/) was used to evaluate the DNA deformation energy based on dinucleotide stacking properties – tilt, shift, roll, slide, twist and rise. Randomized sequences were generated 100 times with the same dinucleotide content as the input sequence. The program options selected were:

- Template 2cv5 (human)
- Best of two orientations
- 164 bp window size.

The human template 2cv5 was chosen as the DNA sequences evaluated in this study were from the human genome. The threading template used for imposing the nucleosome structure on the input sequence was asymmetrical and the underlying nucleosome structure of the DNA sequences were not known; hence, best of two orientations was selected. A 164 bp window size was chosen as recommended by the program and was used to determine how many neighbouring positions of the position being evaluated should be used for calculating the nucleosome positioning score (NPScore).

Aside from NPScore, DNA deformation energy was also used in this study. The DNA deformation energy measures the amount of energy required to impose the structure of the nucleosome bend onto the input sequence whilst the NPScore shows the significance in deviation of the deformation energy at one point from its

neighbouring positions. Supplementary materials from Tolstorukov *et al.* [94] used an NPScore threshold of less than or equal to -2 to indicate a possible nucleosome dyad location, and the same threshold was applied in this study. A more stringent threshold of -3 was evaluated as well.

4.2.4 Comparisons of original base sequences with random arrangements

All 60 sequences were compared with random sequences of the same dinucleotide content generated by the nuScore software to determine if the positioning of the nucleosome dyad is dependent upon the specific arrangements of dinucleotides and their interactions. The number of locations with an NPScore more negative than two thresholds (-2 and -3) were counted and compared statistically.

A marker was deemed to have a high NFP when the locations of NPScore crossing the threshold (NPScore ≤ -2) were high and vice versa. Statistical comparisons were performed using PASW Statistics 18.0.

4.3 Results and discussion

4.3.1 STR sequences

Two hundred base pairs upstream and downstream from the centre of the repeat units were used for the sequence analysis described in this study. This total of four hundred base pairs was chosen because the four hundred base pairs unit wholly encompasses the repeat motifs of the markers as well as the possible primer binding sites flanking the motifs. Moreover, the largest loci of the widely-used commercial kits do not generally go beyond this size, albeit with a few exceptions, such as Penta E (size range 379 to 474 bp) and FGA (size range 322 to 444 bp) of PowerPlex® 16 [233]. In addition, most STRs in current use do not have repeat units (without the flanking regions) that exceed 147 bp, which is equivalent to 36-37 repeat units for a tetranucleotide STR. Amplicon lengths of loci in commercial kits such as SGM+ extend greater than the repeat unit size because of the flanking regions selected for convenient primer design and multiplexing. Using mini-STR primers as shown in [190] as an example, the actual amplicon length of the markers can be reduced to less than the nucleosomal protection size of 147 bp.

Given that STRs have varying number of repeats depending on each individual and that the methods used centre-aligned the sequence, the flanking regions will change accordingly with each allele. This could have an effect on the NFPs. However, centre-aligning the sequences was deemed important because, in theory, the closer the nucleosome dyad is to the centre of the repeat units, the higher the chance that the primer binding sites and the repeat units would remain intact after DNA degradation (due to nucleosome protection) and that successful PCR amplification could occur.

4.3.2 NXSensor

An overall median of 0 (no inhibition sequence) and a standard deviation of 0.019 indicated that the majority (81.7%) of markers tested had high NFP. Consequently, they were more probable to associate with histones to form nucleosomes. Eleven of 60 markers (18.3%) contained short nucleotide sequences that were deemed “stiff” and were less probable to exist as nucleosomes. Their accessibility scores ranged from 0.028 to 0.098 (maximum score = 1) (Figure 4.2).

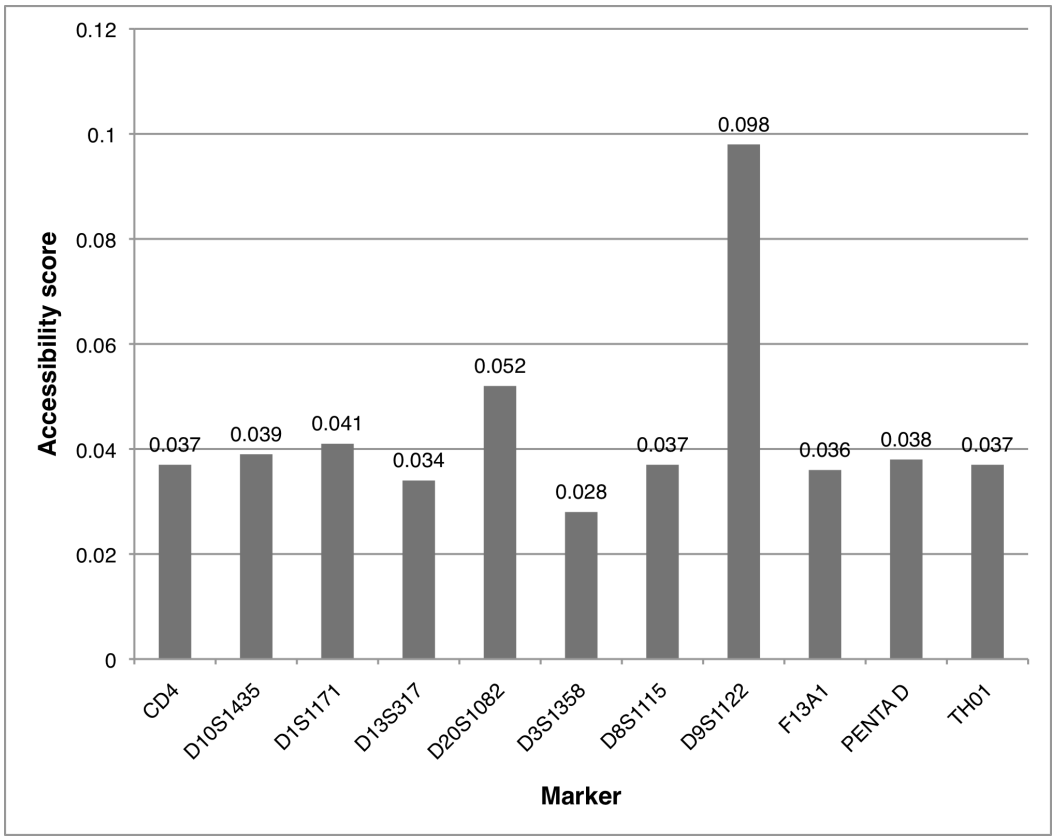


Figure 4.2. Accessibility scores (calculated using Equation 4.1) of 11 STR loci with nucleosome exclusion sequences.

Based on the NXSensor results, the markers were divided into two groups – ones with exclusion sequences and ones without. The validity of nucleosome protection conferred upon the markers could be empirically determined by designing primers for the 11 (particularly D9S1122) with the highest accessibility scores (Figure 4.2), and then compare their performance on artificially degraded DNA and casework samples with a marker that had no exclusion sequence, e.g. vWA. Potentially, if the hypothesis proposed in this study was correct, the 11 loci with accessibility scores of more than zero should exhibit properties associated with degraded DNA and/or LT-DNA amplification [198] while these effects should be dampened with the other 49 loci. However, as the accessibility scores of these 11 loci were extremely low (<0.10), the effect of stiff sequences hindering nucleosome formation should be almost negligible, i.e. the loci were not sufficiently discriminated based on the accessibility scores. Hence, NXSensor would not be further used for the empirical study of nucleosome protection in Chapter 6.

4.3.3 NuScore

NuScore outputted the results into a graphical format (e.g NPScore profile of D18S51 – Figure 4.3).

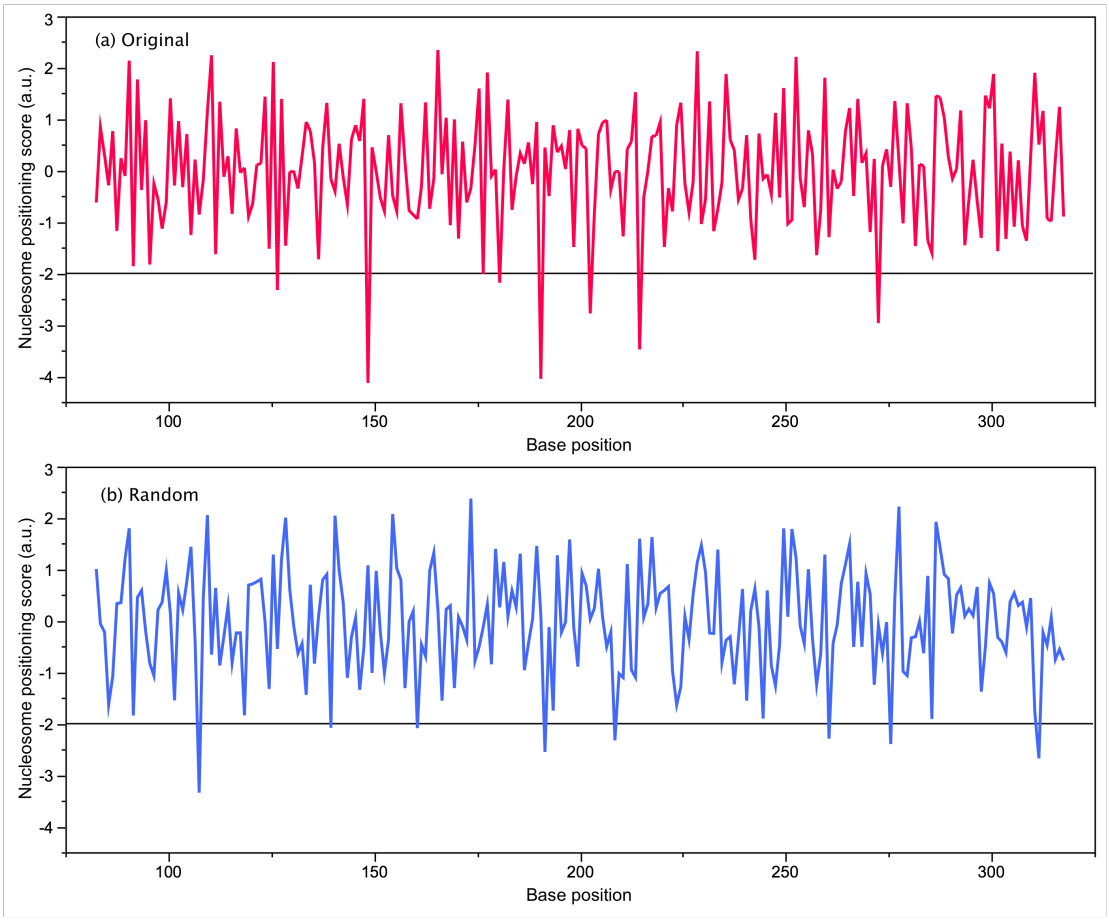


Figure 4.3. Nucleosome positioning score profile of D18S51 (a.u. = arbitrary units) with reference line at -2: (a) original base sequences and (b) random dinucleotide profile of the same composition.

The alternating high-low score seen in the figure was typical of every sequence. The minima signified locations where there was potential for a nucleosome dyad to exist. In this profile, a reference line is shown at -2, as suggested by Tolstorukov *et al.* [94]. Values below -2 and -3 were counted for each STR locus. All loci displayed at least one possible location for a nucleosome dyad (threshold of -2) in the 400 bases input (Table 4.2). The medians of possible nucleosome dyad locations were 7 and 1 for the threshold of -2 and -3 respectively, with standard deviations at 2.380 and 0.851. The markers with the highest number of potential dyad locations (12) were D21S11 and D10S1435, meaning that these two loci were the most likely to be bound to nucleosomes.

Table 4.2. Number of possible locations for a nucleosome dyad with the threshold of -2 and -3 for the 60 markers tested. Original (Ori) and random (Ran) indicate the arrangement of the base sequences. Original arrangement is found in a human genome and random is the same dinucleotides arbitrarily rearranged. The table is arranged in descending order of possible locations for Ori in threshold -2.

Marker	Threshold -2		Threshold -3		Marker	Threshold -2		Threshold -3	
	Ori	Ran	Ori	Ran		Ori	Ran	Ori	Ran
D10S1435	12	4	0	2	**D19S433**	7	8	1	2
D21S11	12	5	1	0	**D1S1677**	7	6	1	1
CD4	11	8	1	1	**D4S2408**	7	12	0	2
D5S2500	11	8	1	2	**D5S818**	7	7	0	0
F13A1	11	4	1	0	**D8S1115**	7	5	0	1
D12ATA63	10	7	0	1	**D12S391**	6	7	1	1
D1S1627	10	7	2	1	**D14S1434**	6	5	2	1
D1S1656	10	5	2	1	**D17S1301**	6	6	2	0
D2S1338	10	4	0	0	**D1S1171**	6	6	1	1
D7S820	10	5	1	1	**D21S1437**	6	7	0	1
PENTA E	10	8	1	1	**D2S441**	6	4	0	1
CSF1PO	9	1	0	0	**D3S3053**	6	4	0	0
D3S1545	9	7	2	0	**D6S474**	6	1	1	0
D4S2364	9	4	2	0	**D8S1179**	6	2	0	0
FGA	9	4	2	1	**PENTA D**	6	8	0	1
VWA	9	5	0	2	**TH01**	6	6	2	0
AMEL_Y	8	9	2	0	**D2S438**	5	3	1	0
D18S51	8	8	3	1	**D4S2366**	5	8	0	1
D1GATA113	8	6	0	1	**D9S1122**	5	5	0	0
D20S1082	8	7	0	1	**D9S2157**	5	11	0	0
D20S482	8	7	0	0	**HPRTB**	5	2	0	1
D22S1045	8	6	3	0	**TPOX**	5	7	0	0
D3S1358	8	4	2	0	**D17S974**	4	5	1	0
D6S1017	8	6	1	1	**D20S161**	4	7	1	1
FES	8	3	1	0	**D2S1776**	4	7	0	1
LPL	8	4	1	1	**D3S4529**	4	8	0	1
AMEL_X	7	7	0	1	**D13S317**	3	3	0	1
D10S1248	7	9	0	1	**D2S1424**	3	8	1	1
D11S4463	7	6	1	1	**D18S853**	2	6	0	0
D16S539	7	5	1	0	**SE33**	2	7	0	0
Median	7	6	1	1	*StDev*	2.380	2.184	0.851	0.629

Within the central 100 bp, the markers were divided into three groups based on their scores. Group A comprised 27 markers with scores from zero to two, i.e. there were two or less positions in the central 100 bp that crossed the threshold of -2. Group B comprised 28 markers with scores between three and five, inclusive, and group C comprised five markers whose scores were six or above (Table 4.3).

Table 4.3. 60 markers divided into three groups according to the number of positions crossing the threshold of -2 at the central 100 bases (sorted in alphabetical order).

Group A (0-2)		Group B (3-5)		Group C (6-8)
AMEL_X	D2S1776	AMEL_Y	D3S1545	D18S51
D10S1248	D2S438	CD4	D4S2364	D21S11
D11S4463	D2S441	CSF1PO	D4S2408	D22S1045
D12ATA63	D3S3053	D10S1435	D5S818	D5S2500
D12S391	D3S4529	D14S1434	D6S1017	F13A1
D13S317	D4S2366	D17S974	D7S820	
D16S539	D6S474	D19S433	D8S1115	
D17S1301	D9S1122	D1GATA113	D8S1179	
D18S853	D9S2157	D1S1627	FES	
D1S1171	HPRTB	D1S1656	FGA	
D20S1082	SE33	D1S1677	LPL	
D20S161	TH01	D20S482	PENTA D	
D21S1437	TPOX	D2S1338	PENTA E	
D2S1424		D3S1358	VWA	

NuScore results showing at least one potential nucleosome dyad location (NPScore ≤ -2) within the 400 bp of each marker were expected. Nucleosomes serve to facilitate compacting of the chromatin for higher order structure [64] and, as a nucleosome binds approximately 147 bp of DNA, at least one dyad should be seen in a sequence as long as 400 bp. The core 100 bp was more important as discussed in Section 4.3.1 and was used to categorize the STR loci. The markers in group A and group C could be targeted for further empirical comparisons for evidence that may validate the protective capabilities of nucleosomes on STR loci.

4.3.4 Comparisons of original base sequences with random arrangements

A set of 100 random sequences with the same dinucleotide composition was generated by the nuScore program for each marker. The number of possible dyad locations for both arrangements (original and random) and both thresholds (-2 and -3) were computed (Table 4.2). Comparing the scores from the original configuration to the random sequences revealed the effect of dinucleotide stacking on NFPs. For

instance, using D18S51 as an input sequence, a random sequence profile was generated and displayed for direct comparison with the original profile (Figure 4.3), in which differences in the positions of the maxima and the minima were conspicuous. These positional differences suggested that the NFP of this input sequence depended on the arrangement of the bases and not the dinucleotide content.

If the NFPs of each STR depended on the nucleotide content, a strong correlation between the NFPs before (original) and after re-shuffling of bases (random) was expected. The Spearman's ρ was used to determine the correlation between the NFPs. The correlation between the possible dyad locations of the original and random configurations at threshold of -2 gave a p-value of 0.532 ($\rho = -0.082$, $N = 60$), indicating no correlation between the two. When the threshold was set to -3, no correlation was again observed between the two configurations ($p = 0.345$, $\rho = -0.124$, $N = 60$).

The change in the locations of the minima and the lack of correlation between the NFPs suggested that both the numbers of possible dyad locations and the positions they might take up depended on the arrangement of the bases and not the dinucleotide content. However, due to the low number of positions crossing the threshold at -3 as a result of high stringency (Table 4.2), threshold -2 was chosen for categorization of STR loci (Table 4.3).

4.3.5 Comparison of NXSensor and nuScore

Comparison of accessibility scores from NXSensor to threshold -2 and -3 from nuScore scores for each STR locus yielded correlation coefficients of 0.049 ($p = 0.712$) and -0.092 ($p = 0.483$) respectively, using Spearman's ρ. This observation revealed that there was no correlation of accessibility score to the number of possible dyad locations.

The lack of correlation between the two programs may be due to their basic design. The NXSensor program searches for sequences that are inhibitory to nucleosome binding based on known strong inhibitory signals [230] while the nuScore program evaluates the dinucleotide stacking properties of the input sequence. Nucleosome binding and attraction is a multifactorial event, with commonly known variables being dinucleotide periodicity and stacking, GC content and chromatin remodelers [59, 61-63, 66, 69, 230]. Since this attraction depends on more than one single factor,

a significant difference might not be observed when only one or two factors are considered, as in this study. Furthermore, these signals only indicate the probability of finding a nucleosome at a given location and not exact positions determined by high-resolution mapping.

Also, all of the loci evaluated in these studies, with the exception of amelogenin, were located within an intron or intergenic region. Recent studies utilizing advanced methods, including high-resolution nucleosome mapping of multiple eukaryotes, revealed that non-coding regions had lower nucleosome occupancy (density) than coding regions [86, 234-240]. The low occupancy in intergenic regions is due to the bias of the promoter region towards being nucleosome-free and a high occupancy in genes directly downstream from transcription start sites [86]. Additionally, variation in the length of linker DNA exists between coding and non-coding regions. The longer linker lengths found in non-coding regions lowered the nucleosome occupancy [239]. The reason that coding regions have such high nucleosome density is still unknown, but has been suggested to be due to increasing residence time for higher fidelity with enzymes and/or higher nucleosome reassembly due to the act of transcription itself [80]. This means that non-coding regions might actually degrade faster than coding regions if nucleosomes protect bound DNA from degradative enzymes. A low occupancy in non-coding regions makes the specific locations of nucleosomes more important. If a nucleosome dyad is positioned within the repeat units of an STR, the possible protection against degradation can extend to both directions and cover the repeat units as well as primer binding sites.

In addition to evaluating forensically relevant loci, other areas in the human genome could be analysed and their NFPs determined. This way, areas with extreme NFPs can be located and primers can be designed to amplify them in degraded samples. By using these extreme loci, detection of nucleosome protection is facilitated as even a small effect can be observed. This could serve as a pilot experiment before progressing to STR loci. Nonetheless, detection of nucleosome protection in these other loci does not guarantee that STR loci are affected or protected in the same way, as nucleosome positioning is highly dynamic.

Other nucleosome prediction programs are also available online for research use. The most recent program released in December 2009 is “FineStr”

(http://www.cs.bgu.ac.il/~nucleom/) [95], which is based on the universal nucleosome positioning pattern of *Caenorhabditis elegans* [71]. Another noteworthy program based on discriminant analysis of dinucleotide frequency is called "Recon" (http://wwwmgs.bionet.nsc.ru/mgs/programs/recon/) [241]. It was not used in this study because the human dinucleotide periodicity pattern is based on GC dinucleotides and not the AT dinucleotides used by "Recon" [71]. Given the relatively fast nature of the field, the two programs that are freely available and based on most up-to-date data when the study was carried out were used.

5 Using the Taguchi method for rapid qPCR optimisation

5.1 Introduction

In this study, attempts at optimisation of qPCR assays have been carried out using a combination of the factorial method (matrix of forward primer and reverse primer concentrations) and trial-and-error (see Chapter 2). A more systematic approach is deemed necessary, as many new assays needed optimisation before they can be compared in order to evaluate the protective capabilities of nucleosomes. Using the factorial method to determine the optimal conditions requires a lot of resources, as it involves testing all the levels of all factors against one another. For instance, previous optimisations experiments (e.g. ALDOA – Section 2.3.4 on page 52) used the factorial method (3x3 matrix) to optimise primer concentrations, in which nine reactions were needed to look at three levels of forward primer and three levels of reverse primer. If one included three levels of annealing temperature in the factorial method, 27 reactions would be needed (3x3x3 matrix). An alternative approach to optimisation widely used in engineering is the Taguchi method, which is used to reduce the time and effort for optimisation processes [242]. This method has been previously applied with a hydrolysis probe [243] but has never been demonstrated with SYBR® Green I dye and qPCR [244].

The Taguchi method can be used when the objective of the experiment is "larger-better", "smaller-better" or "on-target-better" [245]. It has been used to optimise PCRs using the "larger-better" equation [246-248]. In contrast to PCR which uses endpoint measurements, qPCRs are based on real-time analysis where measurements are taken during each cycle (see Section 1.7.1). The cycle number at which fluorescence significantly differs from the background noise is called the quantification cycle (C_q) [130], which is inversely proportional to the log of the initial quantity of DNA [131]. An optimised qPCR assay will also have the lowest C_q value possible. Therefore, qPCRs requires the "smaller-better" equation (Equation 5.1) [245].

$$\eta = -10 \log_{10} \left[\frac{1}{r} \left(\sum_{i=1}^{r} y^2 \right) \right]$$

Equation 5.1. Smaller-better signal-to-noise ratio equation.

where η = signal-to-noise ratio, r = number of repeats in an experiment, y = response (C_q). This equation gives a signal-to-noise ratio (η) that is negative, with values close to zero indicating better conditions.

5.1.1 Aims

The Taguchi method will be employed in the optimisation of three qPCR assays – CD4, D1S1627, and RPPH1 – and the optimal conditions obtained will be compared with ones determined by the factorial method to ascertain whether the Taguchi method could be adapted to qPCR experiments. The three loci are used as a pilot study and if successfully optimised, the Taguchi method will be further used to optimize assays that will be used for nucleosome protection study. The benefits of using this method will also be explored, such as percent contribution (P_C) of each factor and performance prediction of untested levels.

5.2 Materials and methods

5.2.1 Reaction set-up

Three primer sets were designed to bind to a part of the CD4 gene, D1S1627 STR locus, and ribonuclease P RNA component H1 (RPPH1) gene (Table 5.1). Each qPCR reaction included 12.5 μL Brilliant® II SYBR® Green Low ROX Master Mix, 1 ng DNA template (CAMBIO) and variable amounts of primer depending on the experiment being investigated (Table 5.2). The final reaction volume of 25 μL was made up with amplification-grade water.

Table 5.1. Forward and reverse primer sequences and lengths for CD4, D1S1627, and RPPH1.

Primer	Sequence (5'-3')	Length
CD4 Forward	TGGAGTCGCAAGCTGAACTA	20
CD4 Reverse	CAGAGTGAGAACCTGTCTTGAAAA	24
D1S1627 Forward	CATGAGGTTTGCAAATACTATCTTAAC	27
D1S1627 Reverse	TTTTAATTTTCTCCAAATCTCCA	23
RPPH1 Forward	CATCTCCTGCCCAGTCTGA	19
RPPH1 Reverse	GTCACTCCACTCCCATGTCC	20

Table 5.2. Three factors and three levels for each assay optimised in this study.

Factor	CD4 (Level 1, 2, 3)	D1S1627 (Level 1, 2, 3)	RPPH1 (Level 1, 2, 3)
Forward primer (μM)	0.20, 0.40, 0.60	0.20, 0.40, 0.60	0.20, 0.40, 0.60
Reverse primer (μM)	0.20, 0.40, 0.60	0.20, 0.40, 0.60	0.20, 0.40, 0.60
Annealing temperature (°C)	55, 60, 65	55, 60, 62	62, 66, 64

qPCR was run on a Stratagene MX3005P using the following parameters: an initial denaturation of 95°C for 10 minutes followed by 35 cycles of 95°C denaturation for 30 seconds, desired annealing temperature (Table 5.2) for 60 seconds, and 72°C extension for 30 seconds. Dissociation curve analysis was completed by holding at

95°C for 60 seconds then ramping the temperature up from 55°C to 95°C. C_q values were determined using the MxPro™ software version 4.10. All reactions were carried out in duplicate with at least one NTC.

5.2.2 Taguchi method

The chosen response variable was C_q, of which a low C_q value was better. Three factors at three levels were chosen for optimisation (Table 5.2) and hence the L_9 orthogonal array was selected (Table 5.3). All calculations were done by inputting the formulae (Equation 5.1 to Equation 5.4) into a Microsoft® Excel spreadsheet. Using the experimentally determined C_q values, the signal-to-noise ratio (η) of each experiment (experiment 1 to 9) was calculated using Equation 5.1 (page 108).

Table 5.3 The modified L_9 orthogonal array. Numbers 1, 2 and 3 indicate the levels tested for each experiment.

Experiment	Forward primer	Reverse primer	Annealing temperature
1	1	1	1
2	1	2	2
3	1	3	3
4	2	1	2
5	2	2	3
6	2	3	1
7	3	1	3
8	3	2	1
9	3	3	2

The percent contribution (P_C) of each factor to the total variation observed in the qPCR experiment was calculated using the following equation [245]:

$$P_x = \frac{SS_x - V_e \times \upsilon_x}{SS_T}$$

Equation 5.2. Percent contribution.

where SS_x = the sum-of-squares of the signal-to-noise ratio of factor x, V_e = the variance of error, υ_x = the degree of freedom of factor x, and SS_T = the total sum-of-squares.

The following equation was used to predict the signal-to-noise ratio of a reaction carried out with optimal conditions [245]:

$$\eta_{\text{opt}} = \eta_m + \sum_{i=1}^{f} (\overline{\eta_i} - \eta_m)$$

Equation 5.3. Predicted optimal signal-to-noise ratio.

where η_m = the overall mean of signal-to-noise ratio, f = the number of factors, η_i = the mean of the signal-to-noise ratios at the optimal level of each factor i.

The 95% confidence interval of prediction (CI), in which the signal-to-noise ratio of the confirmation experiment should fall given that the prediction model is suitable, was then calculated using the follow equation [245]:

$$CI = \sqrt{F_{(1,\upsilon_e)} \times V_e (\frac{1}{n_{\text{eff}}} + \frac{1}{n_{\text{conf}}})}$$

Equation 5.4. 95% confidence interval prediction

where $F_{(1,ve)}$ = the F-value with the first degree of freedom equal to 1 and the second degree of freedom equal to the degree of freedom of error υ_e, V_e = the variance of error, n_{eff} = the effective sample size determined by N/(1+υ), where N = the total number of experiments and υ = degree of freedom of all factors combined, and n_{conf} = the number of confirmatory tests conducted.

A regression analysis was applied to the data obtained from the orthogonal array by plotting the average signal-to-noise ratio against the levels of each factor and fitting a quadratic curve onto the plotted points.

5.2.3 Factorial method

Three factors at all levels (3x3x3 matrix) were crossed to determine optimal conditions. The same reaction set-up for the Taguchi method was used for the factorial method experiments. The combination of levels that gave the lowest average C_q value from duplicate runs was taken to be the optimal condition.

5.3 Results and discussion

The η for all Taguchi experiments (see Appendix Table 7), the mean η for each level of each factor, and the P_C of each factor (Table 5.4) were calculated. The highest mean η among the three levels was determined as the optimal level. The P_C reflected the amount of variation in the C_q values that was accounted for by each factor. The strongest contributor in all assays was annealing temperature, which was also the main contributor to another PCR-based technique [247].

Table 5.4. The average signal-to-noise ratios of each level of each factor. The highest ratios are shown in bold and the optimal levels are shown in the "Optimal" column. Percent contribution is shown in the "P_C" column.

Assay		Level				
	Factor	**1**	**2**	**3**	**Optimal**	**P_C (%)**
CD4	**Forward**	-29.208	**-29.126**	-29.159	0.40 µM	2.38
	Reverse	-29.273	-29.114	**-29.106**	0.60 µM	13.59
	Temperature	-29.396	-29.162	**-28.935**	65°C	82.74
	Error					1.29
D1S1627	**Forward**	-29.575	**-29.517**	-29.536	0.40 µM	-5.08
	Reverse	-29.807	-29.454	**-29.368**	0.60 µM	39.39
	Temperature	-29.819	-29.432	**-29.378**	62°C	42.46
	Error					23.22
RPPH1	**Forward**	-28.753	**-28.649**	-28.656	0.40 µM	21.41
	Reverse	-28.700	-28.712	**-28.647**	0.60 µM	5.94
	Temperature	-28.742	-28.741	**-28.576**	64°C	62.65
	Error					10.01

The P_C of error is important and should be less than 15 percent [245]. The P_C of error for D1S1627 was 23.22 percent, indicating other significant factors that were not tested in our study, such as annealing time, were contributing to the total variation. Moreover, the P_C of forward primer concentration of D1S1627 was negative (-5.08

percent), indicating that the variation in C_q values due to the forward primer are smaller than those caused by untested factors.

The η from the orthogonal experiments were used to calculate the predicted η_{opt} (Equation 5.3) and its 95% CI (Equation 5.4) that should be obtained when the optimal conditions were used (Table 5.5). A confirmatory experiment was run using the optimal conditions and the observed η from this experiment was compared to the predicted η_{opt}.

Table 5.5. Confirmatory test conditions, predicted η_{opt} based and observed η. "Yes" and "No" in the Within CI (95%) row indicate whether the observed ratio is within the prediction interval.

	CD4	D1S1627	RPPH1
Forward (μM)	0.40	0.40	0.40
Reverse (μM)	0.60	0.60	0.60
Temperature (°C)	65	62	64
η_{opt} predicted	-28.84 ±0.121	-29.18 ±0.703	-28.50 ±0.158
η observed	-28.81	-29.14	-28.78
Prediction error	-0.03	-0.04	0.28
Within CI (95%)	Yes	Yes	No

Both CD4 and D1S1627 had observed η within the confidence interval, meaning the model was accurate, while the observed η of RPPH1 lay outside the 95% CI. This indicated that the three-factors model for RPPH1 was not suitable.

Since the P_C of the reverse primer for RPPH1 was the lowest of the three factors (Table 5.4 on page 113), we created two new models: one with annealing temperature as the only factor, and one with both forward primer and annealing temperature as factors. However, the observed η of the confirmation experiment did not fit in these models (data not shown). Inspection of the dissociation curves of RPPH1 revealed primer-dimers in some NTCs (Figure 5.1), which would have affected the C_q values. As SYBR® Green I binds non-specifically to any double-stranded DNA, such primer-dimers would have brought forward the exponential gain phase of fluorescence, resulting in a false "low C_q". This again highlighted the

importance of rigorous primer design and purification when using a non-specific quantifier such as SYBR® Green I.

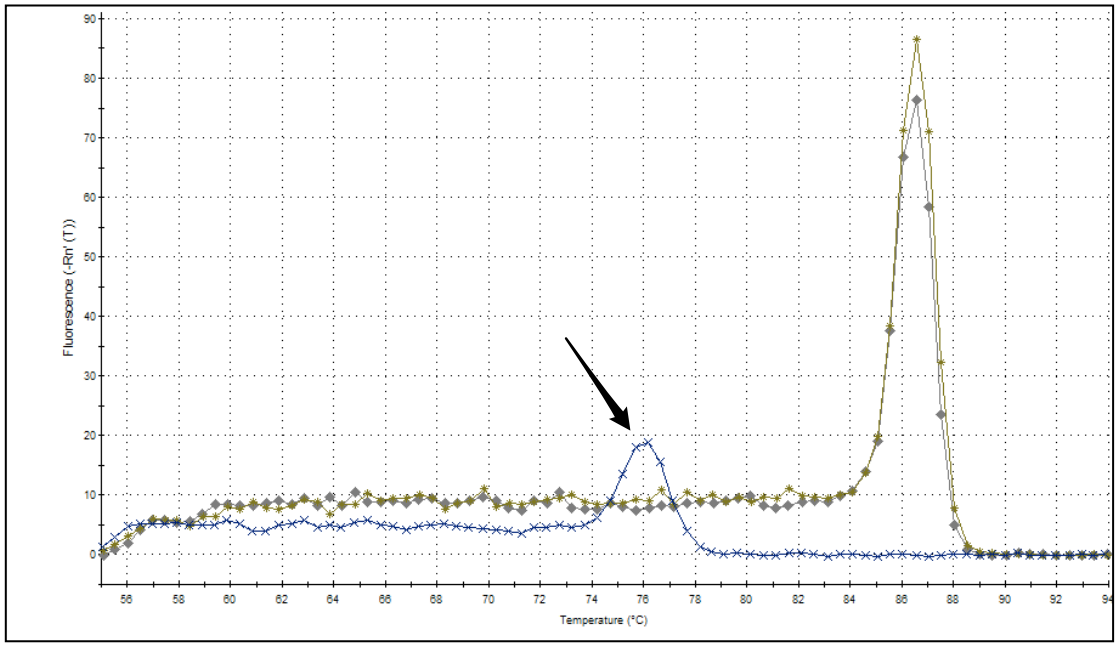


Figure 5.1. Dissociation curve of RPPH1. The yellow and grey lines represent 1 ng of human DNA and the blue line is the NTC. The arrow points to a non-specific product with a dissociation peak at approximately 76°C.

The same optimal conditions were obtained using the factorial method and the Taguchi method for all three loci with only two exceptions: the forward primer concentration of D1S1627 and the reverse primer concentration of RPPH1 (Table 5.6). However, the differences in C_q values for the two optimal conditions (factorial vs Taguchi) were only 0.22 for D1S1627 and 0.08 for RPPH1, which was considered minimal. Interestingly, these two factors coincided with very low P_Cs shown by the Taguchi method (Table 5.4). It was possible that due to the weak contributions, any concentration between 0.40 and 0.60 μM would not alter the C_q values significantly.

Table 5.6. Optimal conditions as determined by the factorial method, the Taguchi method and regression analysis following the Taguchi method. Bold indicates a difference in the factorial and Taguchi method.

Assay	Factor	Factorial	Taguchi	Regression
CD4	Forward (μM)	0.40	0.40	0.44
	Reverse (μM)	0.60	0.60	0.51
	Temperature (°C)	65	65	65
D1S1627	Forward (μM)	**0.60**	**0.40**	0.45
	Reverse (μM)	0.60	0.60	0.56
	Temperature (°C)	62	62	62
RPPH1	Forward (μM)	0.40	0.40	0.49
	Reverse (μM)	**0.40**	**0.60**	0.60
	Temperature (°C)	64	64	64

The η of untested level could be predicted using regression analysis [248]. The η were plotted against the levels of each factor and a quadratic regression curve was fitted to the points [247], e.g. Figure 5.2. These curves could be used to identify levels that should be tested if further optimisation was needed. In Figure 5.2, the highest point in the quadratic curve was approximately at 0.50 μM, which meant that although 0.50 μM was not tested, it should be included in subsequent optimisation experiments. The highest point found in the curve of each factor was determined (Table 5.6). The RPPH1 assay was also included because although the prediction model did not fit, the same optimal condition from both the factorial and Taguchi method was observed.

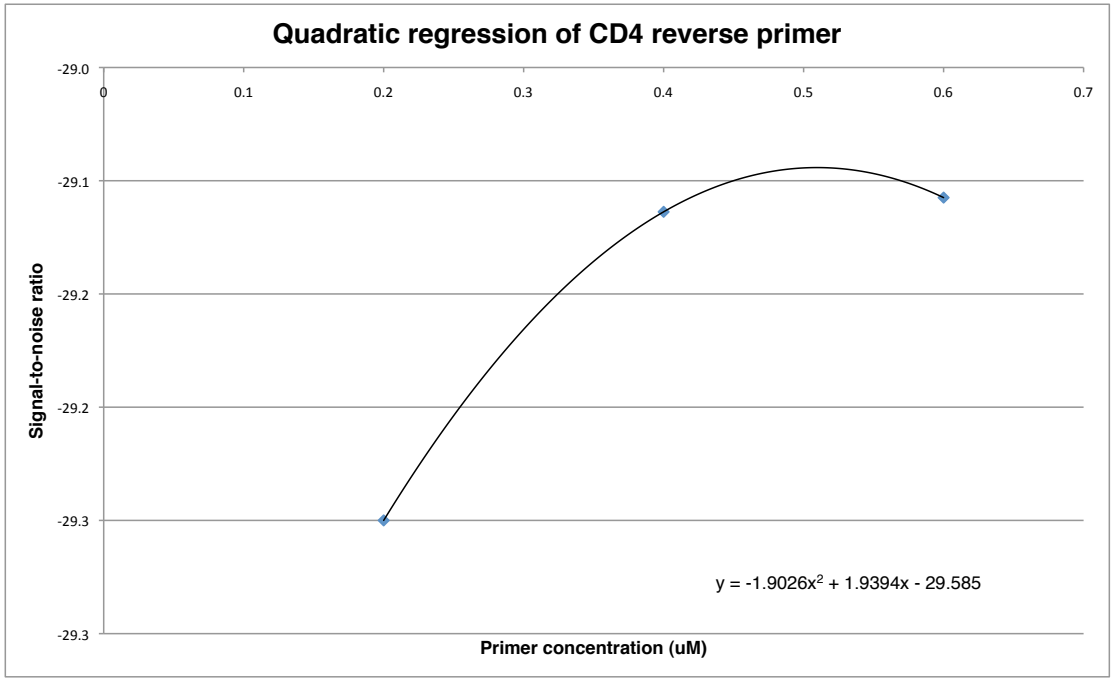


Figure 5.2. A quadratic regression curve for the signal-to-noise ratio of the three levels of the reverse primer concentration of CD4. The highest signal-to-noise ratio (optimal condition) is observed at 0.51 µM

Like any design-of-experiment method, one needs to bear in mind the limitations and weaknesses of the Taguchi method, such as the deficiency in handling interactions between factors and handling noise [249]. Compared to other quality improvement techniques such as Six Sigma, the Taguchi method fails to explore the systems but is useful in finding optimal settings [250]. Other useful experimental designs that could be beneficial to assay optimisation include fractional factorial design and factorial design with blocking. Nonetheless, the Taguchi method raises awareness in experimental design, choosing controllable and uncontrollable factors, and a systematic approach in problem solving [244].

27 experiments were conducted to optimise each assay using the factorial method whereas only nine experiments were needed using the Taguchi method. Optimizing more factors and levels could increase the savings exponentially, e.g. investigating four factors at four levels each would save 240 reactions (256 reactions vs. 16 reactions). The data obtained showed that the Taguchi method was useful for determining which factor had more influence on a qPCR assay by looking at percent contribution. In forensic science, novel methods are often created and optimised to fit specific needs of casework, and thus the Taguchi method can be applied to save

valuable time and resources. The Taguchi method will be applied to optimise newly designed assays for the determination of the protective capabilities of nucleosomes in Chapter 6.

6 An investigation into the protective capabilities of nucleosome on forensic STRs

6.1 Introduction

6.1.1 Saliva and degradation of DNA

Saliva is a heterogeneous fluid composed mainly of water, electrolytes, mucus, antibacterial compounds and various enzymes. On average, one mL of saliva contains 4.3 x 10^5 epithelial cells and 1.694 x 10^7 bacteria cells [11]. These bacterial cells provide additional enzymes such as micrococcal nucleases [18] that aid the degradation of DNA. Although there are variations in the species and number of bacteria in an individual's mouth [18], the combined effects of their enzymes must contribute toward DNA degradation. Other enzymes, such as peroxidases and lysozymes, that degrade DNA can also be found in saliva [17]. 100 μL saliva contains roughly the same amount of DNA as a buccal swab [14], and it has been shown to be a reliable source of DNA in multiple forms (liquid, pellet, and frozen) [13, 15].

Degradation of DNA in saliva has been shown to occur faster than DNA degradation in blood and semen samples [16, 17]. In tissues containing more enzymes with increased activity, such as liver and muscle, degradation of DNA is also faster [112]. In pig carcasses, the majority of DNA degradation has been to shown to occur during the first 24 hours in blood and porcine skeletal muscles [112].

Standardized methods of DNA degradation for validation experiments in forensic science use DNase I, MNase, sonication or a combination thereof [101, 121, 149]. The size of DNA fragments from these degradation methods is roughly the same as post-necrotic DNA fragments (~200 bp) and those found in forensic casework samples [16, 34, 35, 38, 101]. It has been suggested that the degradation of DNA in forensic stains is neither apoptosis nor necrosis. Apoptosis is excluded because fragmentation continues even after all ATPs have been exhausted, while the random degradation process via necrosis is not observed [112]. To date, the process of DNA degradation in forensic stains is not well characterized [92].

Degradation of saliva stains can be carried out using the "incubation method" to simulate "a time-course series of degraded stains in their *natural state*", as stated by

Dixon *et al.* [17]. This technique has been shown to provide rapid degradation of DNA in a natural state, unlike other artificial degradation methods such as DNase I, sonication, and MNase [17]. Saliva stains are subjected to human body temperature in a closed-tube, high humidity environment. The temperature of 37°C is optimal for enzymes and should speed up the degradation of DNA. The increased humidity of a closed-tube incubation should also degrade DNA through hydrolysis [107].

6.1.2 Nucleosome positioning and protection

Nucleosome positions have been shown to have preference for certain base sequences and could be partially predicted using computer models [93-95]. Interactions of many factors, e.g. dinucleotide stacking and DNA methylation, dictate the occupancy of nucleosomes at different locations [51, 56-64, 68, 71, 82, 84, 251]. According to Dixon *et al.* [16], the binding of DNA and histone into a nucleosome could limit the access of the DNA to endonucleases. NuScore, one of the more recent programs, was used in Chapter 4 to classify 58 past and present forensic STR loci into three groups – "low", "medium" and "high" – based on the probability of the existence of a nucleosome dyad. STR loci can be subjected to further *in vitro* experimentations to determine whether the occupation of nucleosomes actually helps protect DNA in degraded samples. This can be achieved by designing mini-STR assays to amplify the randomly selected loci from each nuScore group. These assays can be tested on degraded samples and the success rates of each group can be compared. A locus in the "high" nuScore group should perform better than a locus in the "low" nuScore group, given that nuScore accurately predicts nucleosome occupation and that nucleosome occupation prevents nucleases from attacking the bound DNA.

6.1.3 Primer design and optimisation

Primer design and optimisation is important for the comparison of results from multiple primer sets. Softwares, such as Primer3Plus [252] and OligoPerfect™ Designer (http://tools.invitrogen.com/content.cfm?pageid=9716), are freely available and are widely used for designing primers. Fifty-eight STR loci are considered as candidates for this study [253], including some loci which have published mini-STR primer sets [188].

Using mini-STR products should allow for the differences in quantification of degraded samples, as it has been shown that smaller fragments of DNA survive after degradation [35]. The Taguchi method (previously described in Chapter 5) can be applied to optimise these new primer sets [245]. High molecular weight, intact DNA samples from male placental extract must be used during the optimisation process in order to obtain optimal PCR conditions. For qPCR experiments, construction of calibration curves must be performed at least in duplicate for each primer set to assess the amplification efficiency and sensitivity of the assay. A more detailed discussion of calibration curves, amplification efficiency, and r^2 values can be found in Chapter 2.

6.1.4 Hematin

The PCR process is sensitive to inhibitors that are present in and co-extracted with forensic samples. For instance, DNA extracted from blood and faeces can contain haemoglobin and bilirubin, both of which are PCR inhibitors [254]. Different buffers and polymerases have different capabilities in overcoming inhibitors [254, 255]; therefore different commercial PCR and qPCR kits have different thresholds when inhibitors are present. As a result, the quantification result of one assay might differ from another assay due to the different chemistries used.

Hematin is routinely used to simulate the inhibitory effect of heme in bloodstains [121, 122]. It is a compound formed from the oxidation of FeII (ferrous) to FeIII (ferric) [256] and inhibits *Taq* by binding to the DNA template as well as quenching fluorescent dyes [257]. Addition of bovine serum albumin (BSA) counteracts the effect of these compounds by binding to them, thereby overcoming inhibitions in PCR reactions [258]. Increasing the inhibitor concentration in a qPCR results in a delayed quantification cycle (C_q) [121, 122].

6.1.5 Aims

It is hypothesized that nucleosomes could offer protection to the bound DNA by offering protection against degradation by enzymes. If one could empirically determine whether STR loci with a higher probability of binding are less likely to be degraded, then choosing and incorporating these loci into a “high nucleosome

binding" multiplex could mean that a complete, or better partial, DNA profile could be obtained from degraded forensic samples.

Degraded saliva samples will be prepared by subjecting saliva stains to degradation using the incubation method, and degradation of DNA in these samples will be examined. Two experiments will be set up to investigate the incubation method and a final one will be used for the investigation of nucleosome protection.

Mini-STR primers for use with real-time qPCR detection will be designed and optimised using intact DNA samples and SYBR® Green I dye. 14 randomly chosen loci will be used to amplify degraded saliva samples to determine if there is any difference in the quantification result between different groups of nucleosome protection as determined by nuScore results.

Confirmation of this result will be carried out by amplifying a range of simulated casework samples with five selected primer sets. Statistical methods will be employed to compare the differences between nuScore groups and amplification of STR loci in both degraded saliva samples and simulated casework samples. The correlation between nuScore and survivability of STR loci in degraded samples will be investigated to evaluate the protection of STR loci by nucleosomes.

6.2 Material and method

6.2.1 Degradation of saliva samples

6.2.1.1 Sample collection

Two donors, one male (KF) and one female (YS), donated 20 mL saliva each on three separate occasions. For each sample to be degraded, a 0.25 mm^2 piece of cotton (UV cross-linking for 20 mins) was put into a 0.2 mL PCR tube. 25 µL amplification-grade water was added to each tube, followed by 10 µL of saliva. Tubes were incubated at 37°C and at each designated time-point:

- Experiment "day" = 0, 1, 2, 3, 4, 5, 6, 7, 14, and 28 days (n = 30);
- Experiment "week" = 0, 7, 14, 28, 56, and 70 days (n = 18);
- Experiment "final" = 0, 1, 2, 3, 4, 5, 6, 7, 14, and 28 days (n = 10).

Three 0.2 mL PCR tubes from each donor were taken out of the incubator and immediately frozen to suspend degradation for the "day" and "week" experiments. One tube for each donor was used for the "final" experiment. The "day" and "week" experiments were used to explore the result of the incubation method while the "final" experiment was used for the investigation of nucleosome protection.

6.2.1.2 DNA extraction

All "day" and "week" samples underwent DNA extraction using the QIAGEN QIAcube® with the "forensic casework samples" protocol of the QIAamp® DNA Investigator Kit with a final elution volume of 60 µL. The three samples at each time point were then pooled together to form a single sample to average out the variation in different stains.

The "final" samples were similarly extracted but with an increased elution volume of 100 µL. The 0-day samples of this set were further diluted five-fold and twenty-fold for donor KF and YS, respectively, to obtain a concentration that was optimal for the DNA genotyping and quantification kits used.

6.2.1.3 Quantification

The "day" and "week" samples were quantified in duplicate, according to the manufacturer's protocol, using a Plexor® HY kit with a Stratagene MX3005P, as

detailed in Section 2.2.4. Raw data were exported to the Plexor® Data Analysis Software version 1.5.4.18, which was used to determine the concentrations of DNA in the degraded samples.

6.2.1.4 SGM+ DNA amplification

Saliva samples from the "final" experiment were amplified in duplicate, including a PCR positive and a PCR negative, with SGM+, using the manufacturer's recommended protocol. The samples were amplified in a 25 µL reaction volume. Each tube contained 9.55 µL PCR Reaction Mix, 5 µL of primers, 0.45 µL Ampli*Taq* Gold® DNA polymerase, optimal amount of DNA template, and amplification-grade water. The cycling parameters were: an initial denaturation at 95°C for 11 minutes followed by 28 cycles of 94°C for 60 seconds, 59°C for 60 seconds, and 72°C for 60 seconds, and a final extension at 60°C for 45 minutes.

6.2.1.5 DNA detection and genotyping

PCR products were prepared for detection on an ABI 3130 CE instrument by preparing a master mix of 15 µL of HiDi-Formamide and 0.5 µL of GeneScan® 500 ROX. One µL of PCR product or allelic ladder was added. Each tube was denatured at 95°C for three minutes and then snap-cooled on ice for three minutes. Each sample was injected for 16 seconds at 1.2 kV and run for 30 minutes at 60°C and 15.0 kV. The polymer used was POP-7 and the capillary length was 36 cm. Handling of raw data and allele calling was made using Genemapper® 3.2.1, as detailed in Section 3.2.5.

6.2.2 STR primers

6.2.2.1 Primer selection

23 loci with suitable primer sets and a desired product size of less than 147 bp (except for D21S11 and D18S51) were subjected to randomization for further experimentation (Table 6.1). The candidates in each group were randomized separately. Five loci were randomly chosen from each group (Table 6.2); all four loci from group C were used.

Table 6.1. NuScore groups based on Thanakiatkrai and Welch [253]. The nuScore of each primer is indicated in parentheses (X). Asterisks (*) mark a failed optimisation resulting in a new randomly chosen locus. Bold indicates the final chosen loci.

Group A	Group B	Group C
D10S1248 (0)	**D1GATA113 (3)**	**D18S51 (6)**
D20S1082 (0)	D1S1627 (3)	**D22S1045 (6)**
D1S1171* (1)	D14S1434 (3)	**D5S2500 (6)**
D12ATA63* (2)	**D8S1179 (3)**	**D21S11 (8)**
D16S539* (2)	D1S1677 (4)	
D2S441 (2)	**FES (4)**	
D9S2157 (2)	**CD4 (5)**	
TH01 (2)	CSF1PO (5)	
TPOX (2)	D10S1435 (5)	
	D4S2364 (5)	

Table 6.2. The STR loci selected for optimisation. The chromosomal position was based on the human GRCh37 (Feb 2009). The repeat motif was based on the sequences from Genbank top strand.

Locus	Accession no.	Chromosomal Position	Location	Min allele	Max allele	Repeat Motif
D10S1248	AL391869	10: 131.093 Mb	10q26.3	8	19	GGAA
D20S1082	AL158015	20: 53.866 Mb	20q13.2	8	17	ATA
D9S2157	AL162417	9: 136.035 Mb	9q34.2	7	19	ATA
TH01	D00269	11: 2.192 Mb	11p15.5	3	14	TCAT
TPOX	M68651	2: 1.493 Mb	2p25.3	4	16	AATG
D1GATA113	Z97987	1: 7.443 Mb	1p36.23	7	13	GATA
D8S1179	AF216671	8: 125.907 Mb	8q24.13	7	20	TCTR
FES	X06292	15: 91.432 Mb	15q25-qter	7	15	ATTT
CD4	M86525	12: 6.897 Mb	12p12-pter	4	15	TTTTC
D4S2364	AC022317	4: 93.517 Mb	4q22.3	7	11	GRAT
D18S51	AP001534	18: 60.949 Mb	18q21.33	7	40	AGAA
D22S1045	AL022314	22: 37.536 Mb	22q12.3	8	19	ATT
D5S2500	AC008791	5: 58.699 Mb	5q11.2	14	24	GRYW
D21S11	AP000433	21: 20.554 Mb	21q21.1	12	41.2	TCTR

Table 6.3. The forward primer, reverse primer, minimum product size, and maximum product size of each candidate locus. "L" = primer length.

Locus	Forward primer (5'-3')	L	Reverse primer (5'-3')	L	Min size	Max size
D10S1248	TTAATGAATTGAACAAATGAGTGAG	25	CAACTCTGGTTGTATTGTCTTCAT	24	82	126
D20S1082	ACATGTATCCCAGAACTTAAAGTAAAC	27	CAGAAGGGAAAATTGAAGCTG	21	76	103
D9S2157	CAAAGCGAGACTCTGTCTCAA	21	AAAATGCTATCCTCTTTGGTATAAAT	26	75	111
TH01	GGCCTGTTCCTCCCTTATTT	20	CACAGGGAACACAGACTCCA	20	61	105
TPOX	GAACAGGCACTTAGGGAACC	20	TCCTTGTCAGCGTTTATTTGC	21	69	117
D1GATA113	TTCTTAGCCTAGATAGATACTTGCTTC	27	TCAACCTTTGAGGCTATAGGAA	22	83	107
D8S1179	TTTTTGTATTTCATGTGTACATTCG	25	TCCTGTAGATTATTTTCACTGTGG	24	83	135
FES	TTTAGGAGACAAGGATAGCAGTTC	24	CCTGGCGAAAGAATGAGACT	20	80	112
CD4	TGGAGTCGCAAGCTGAACTA	20	CAGAGTGAGAACCTGTCTTGAAAA	24	75	130
D4S2364	CTAGGAGATCATGTGGGTATGATT	24	CAGTGAATAAATGAACGAATGGA	23	70	86
D18S51	TGAGTGACAAATTGAGACCTTG	22	TTGCTACTATTTCTTTTCTTTTTCTCT	27	98	230
D22S1045	ATTTTCCCCGATGATAGTAGTCT	23	CGAATGTATGATTGGCAATATTTTT	25	78	111
D5S2500	CTGTTGGTACATAATAGGTAGGTAGGT	27	TCGTGGGCCCCATAAATC	18	86	126
D21S11	TGAGTCAATTCCCCAAGTGAA	21	CCAGAGACAGACTAATAGGAGGTAGA	26	131	249

6.2.2.2 Reaction optimisation

Optimisations were performed in duplicate for each STR primer set using the Taguchi method and the L_9 orthogonal array (see Chapter 5). The three factors optimised were forward primer concentration, reverse primer concentration and annealing temperature.

The levels of primer concentration tested were 0.20, 0.40 and 0.60 μM, based on the recommendation for the Brilliant® II SYBR® Green Low ROX Master Mix kit [179] (Agilent Technologies, CA, USA). Annealing temperatures ranged from 55°C to 65°C, depending on the melting temperatures of the primers (see Appendix Table 8). Calibration curves were constructed in duplicate with intact DNA control samples to confirm the optimal conditions. The optimal conditions were used with both the degraded saliva samples and simulated casework samples.

6.2.2.3 Amplification of saliva samples

Using the determined optimal conditions (Table 6.7 on page 138), ten saliva samples from each donor ("final" experiment) were amplified in duplicate for each locus. A total reaction volume of 12.5 μL, consisting of 6.25 μL 2X Brilliant® II SYBR® Green Low ROX Master Mix, 1 μL DNA (sample or control), the optimal amount of forward and reverse primers (Table 6.7 on page 138), and Milli-Q water (Millipore, MA, USA), was used. Two negative controls and a positive control (SRM-A) were included in each plate.

Calibration curves were constructed from five concentrations of a serial five-fold dilution of a commercial male placental DNA extract (CAMBIO). The DNA concentrations used were 10, 2, 0.4, 0.08, 0.016 ng/μL. Each concentration was amplified in duplicate.

The thermal cycling condition for all loci were: initial denaturation at 95°C for 10 minutes, followed by 35 cycles of 95°C for 30 seconds, 60-65°C for 1 minute and 72°C for 30 seconds. A dissociation curve analysis was carried out by denaturation at 95°C for 1 minute, annealing at 55°C for 30 seconds, and continuous fluorescence

monitoring during the ramping up of temperature to 95°C then holding the temperature for 30 seconds.

6.2.2.4 Amplification of simulated casework samples

Five primer sets (CD4, D4S2364, D20S1082, D10S1248, and FES) were selected based on the results from the saliva samples for further work with simulated casework samples obtained from GEDNAP Trial 40 and 41 (coded "0-X" and "1-X") (see Chapter 3 for a discussion on simulated casework samples and Appendix Table 10 for descriptions). All samples were amplified in duplicate, including two negative controls and two positive controls (SRM-A) (see Section 6.2.3.2 on explanation of SRM-A). Thermal cycling and data collection was carried out as in Section 6.2.2.3.

6.2.2.5 Inhibition by hematin

Reaction set-up and thermal cycling for Plexor® HY was carried out as in Section 2.2.4. Porcine hematin powder (Sigma-Aldrich) was resuspended in 0.1 M sodium hydroxide and then diluted with Milli-Q water. Each reaction was made of 6.25 μL Brilliant® II SYBR® Green Low ROX Master Mix, 2.1 ng SRM-A DNA, 0.20 μM D18S51 forward primer, 0.40 μM D18S51 reverse primer, Milli-Q water, and 0 - 30 μM hematin (in 5 μM interval). Thermal cycling and data collection was carried out as in Section 6.2.2.3.

6.2.3 Data analysis

6.2.3.1 Real-time qPCR data collection

Data from qPCR runs on the Stratagene Mx3005P were collected using Stratagene MxPro™ qPCR Software Version 4.10. For all experiments, replicates were "treated individually" with the following algorithm enhancements selected: amplification-based threshold, adaptive baseline and moving average. Quantification cycles (C_q values), r^2 values and reaction efficiencies were all automatically determined by the software. ROX was used to normalise the SYBR® Green I signals. Raw data were exported to Microsoft® Excel for further manipulation. Reactions that did not cross the amplification threshold ("no C_q") were given the value of 0 ng/μL for the calculation of average concentration from a duplicate in order to prevent

overestimation of DNA concentrations in degraded saliva samples that fall in the stochastic region of PCR [198]. The average concentration was used because replicates of a sample could not be assumed to be independent.

6.2.3.2 Normalisation

Some run-to-run variations due to unequal amplification efficiencies were expected from using different primer sets and from running experiments on different days [154]. Normalisation should minimize the differences due to different primer sets and reaction efficiencies. The concentrations of a 1 in 25 dilution of SRM-A (NIST, MD, USA) (the positive control) was used to normalise the results of both degraded saliva and simulated casework samples quantified in the same plate. SRM-A is the A-component of the NIST Standard Reference Material 2372 specifically prepared for qPCR standardisation from buffy coat white blood cells of an anonymous male. It is certified to contain 52.4 ng of DNA per μL [185]. The dilution should have a concentration of 2.1 ng/μL.

The quantification results of the STR primer sets were normalised as follows. A "normaliser" was first obtained by dividing 2.1 ng, which was the exact amount added to the reaction well, by the quantification result of SRM-A in each plate (Table 6.4). The "normaliser" was then multiplied with the quantification result of the sample wells in the same plate, thereby normalising the DNA concentrations in these wells by a factor derived from a known sample.

Table 6.4. The concentration of SRM-A (singly quantified) in ng/μL for all loci and the normaliser (2.1/SRM-A) used to correct for inter-run variation. RSD% is relative standard deviation.

Locus	SRM-A	Normaliser	Locus	SRM-A	Normaliser
D10S1248	2.29	0.92	FES	1.43	1.47
D20S1082	2.44	0.86	CD4	1.94	1.08
D9S2157	1.61	1.3	D4S2364	1.92	1.09
TH01	1.87	1.12	D18S51	3.23	0.65
TPOX	1.91	1.1	D22S1045	3.19	0.66
D1GATA113	2.93	0.72	D5S2500	2.94	0.71
D8S1179	1.79	1.17	D21S11	2.88	0.73
Mean	*2.31*	*0.97*	*Std Dev*	*0.615*	*0.259*
Median	*2.12*	*1*	*RSD%*	*27*	*27*

6.2.3.3 Statistical analysis

All statistical data analyses and graphs were done using Microsoft® Excel, PASW Statistics 18.0 and GraphPad Prism® 5.0 (GraphPad Software, CA, USA). All tables are arranged by ascending nuScore unless otherwise stated. The level of significance (α) for any hypothesis test was 0.05. Statistical tests were previously discussed in Section 1.8. Normality was determined by the Shapiro-Wilk test and homogeneity of variance was assessed using the Fligner-Killeen test [145].

The Mann-Whitney *U* test [146] was used to compare the distributions in DNA concentration between the two donors in the "day" and "week" degradation experiments. The Friedman's test [146] was used to compare dependent samples receiving different treatments. The concentration of each sample across the different loci was ranked, and the distribution of the ranks of each locus was then compared. If a p-value of less than 0.05 was obtained, pairwise comparisons were carried out to determine which loci differed significantly. Determination of correlation between the nuScore and Friedman's test ranks were done using the Spearman's ρ [146]. A value close to one or negative one indicated a strong correlation and a value close to zero indicated weak correlation. Both nonparametric methods were used because they could effectively deal with small sample sizes.

6.3 Results and discussion

6.3.1 Degradation of saliva samples ("day" and "week")

6.3.1.1 DNA concentration and rate of degradation

The highest concentration of DNA was found in the 0 day sample of YS (28.5 ng/µL) (Table 6.5). The YS samples, both in the "day" and "week" experiments, had a higher DNA quant at day 0 than the KF samples. A sudden drop in DNA concentration was observed between day zero and day one for both donors, suggesting that there was a rapid degradation of DNA within the first twenty-four hours. The DNA concentrations in both experiments seemed to stabilize after seven days of degradation (Figure 6.1).

Table 6.5. Quantification results in ng/µL (means and standard deviations) of degraded saliva samples (duplicate analysis at each time-point) by donor and experiment. A "-" indicates no sample tested.

Time (days)	KF				YS			
	Day		Week		Day		Week	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0	6.695	0.431	3.466	0.221	28.500	3.536	16.86	0.865
1	1.025	0.021	-	-	1.855	0.007	-	-
2	0.394	0.004	-	-	1.103	0.166	-	-
3	0.365	0.037	-	-	0.588	0.144	-	-
4	0.291	0.005	-	-	1.235	0.163	-	-
5	0.259	0.010	-	-	0.351	0.025	-	-
6	0.365	0.083	-	-	0.654	0.017	-	-
7	0.242	0.034	0.517	0.063	0.120	0.022	0.050	0.006
14	0.320	0.068	0.648	0.008	0.098	0.004	0.040	0.005
28	0.145	0.006	0.398	0.074	0.126	0.018	0.025	0.004
56	-	-	0.231	0.035	-	-	0.011	0.002
70	-	-	0.244	0.003	-	-	0.012	0.001

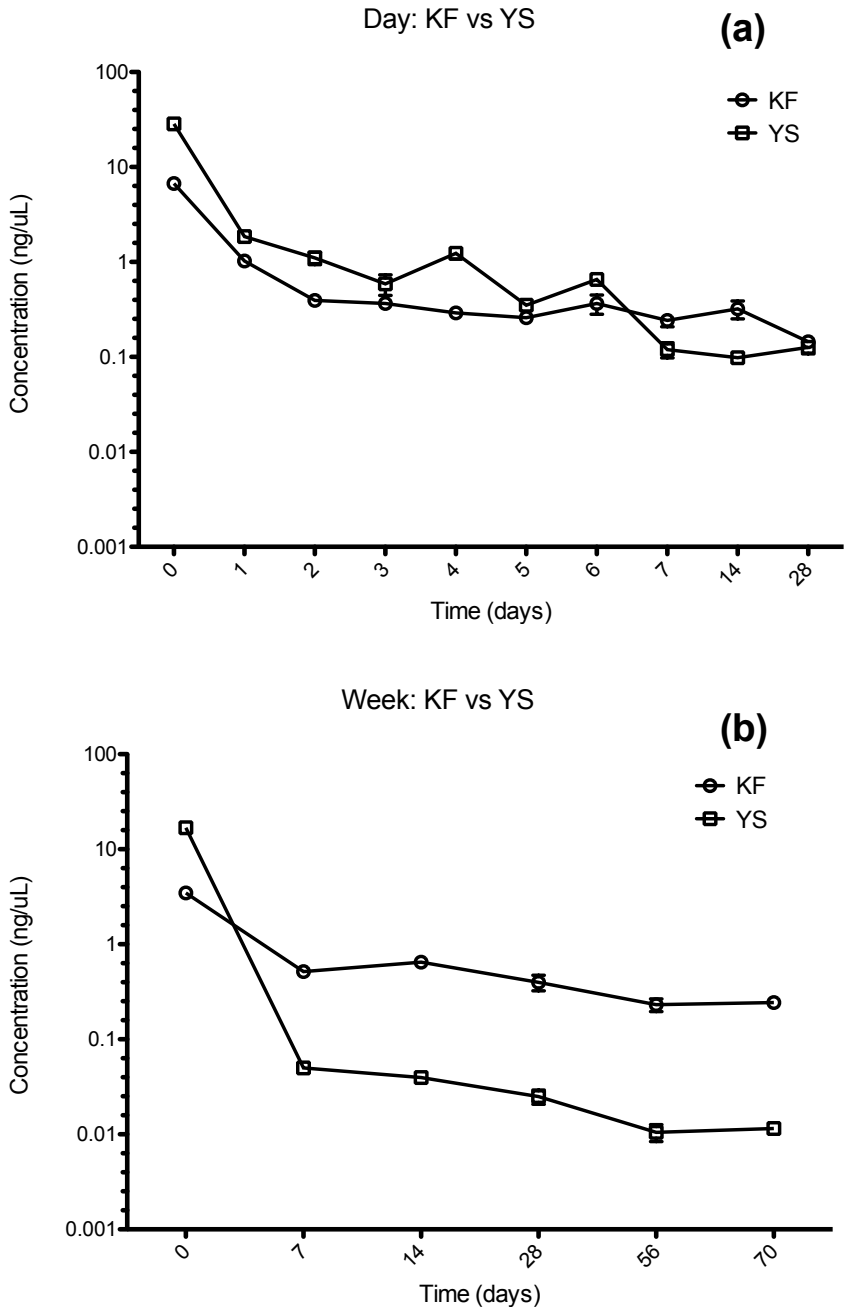


Figure 6.1. DNA concentrations by time of the day (a) and week (b) experiment. KF samples are shown as circles and YS samples are shown as squares. Both graphs have error bars showing one standard deviation (n=2). DNA concentration is shown in log 10.

The rate of degradation seemed to vary considerably between donors and experiments. At day 28, the YS-week sample had only 0.040 ng of DNA per μL, while the YS-day sample had approximately three times that concentration (0.126 ng/μL). Even after 70 days, the KF-week sample had 0.244 ng/μL. Some time-points, e.g. day four and day six of the day experiment, also showed an increase

("jump") in DNA concentration when compared to the prior time-point (Figure 6.1). Another interesting point to note was the crossing of the two donor lines at approximately day seven. YS samples, even with a much higher starting DNA concentration, appeared to degrade faster than the KF samples.

This "jump" had been observed before by Dixon *et al.* [17], and could be attributed to the unequal amount of bacteria, the randomness of degradation and the unequal starting amount of DNA in the saliva that had been deposited on the cotton fabric. Despite the best attempts at mixing the saliva thoroughly, the heterogeneous nature of saliva might have prevented an equal number of the donor cells and bacterial cells in each 10 µL saliva sample. These variations could have led to different degradation rates and consequently a seemingly erratic increase in DNA concentration at later time-points was seen.

Using the same experimental set-up, Dixon *et al.* [17] degraded two reference saliva samples. A similar log-curve degradation pattern for both samples was seen, based on SGM+ profile success rates. They reported 19.0 and 15.8 ng/µL of DNA in the two "week 0" samples, comparable to the range of 3.47 to 28.50 ng/µL of DNA found in the day 0 and week 0 samples in this study. After two weeks of degradation only 0.10 ng/µL and 0.06 ng/µL of DNA was detected in the experiment by Dixon *et al*. Again, this result was comparable to the YS samples in this study, but the KF samples had a much higher concentration of DNA at that time-point. This difference was attributed to the difference in product sizes. The qPCR method based on Nicklas and Buel used in their study targeted a 124 bp *Alu* product, while the Plexor® HY used in this study targeted the 99 bp Human RNU2 locus. Shorter products have been shown to give a higher estimate of DNA in degraded samples [120, 149, 154].

6.3.1.2 Establishing degradation

In degraded DNA samples, gel electrophoresis can be carried out to ascertain the samples are degraded, as degraded DNA exhibits a distinctive smear pattern [1]. However, it was not used in this study because the amount of DNA in each degraded saliva sample was lower than the sensitivity of ethidium bromide staining (10-20 ng required per band) [141]. Even at day 1, loading 10 µL of the degraded sample onto the gel would only equal to 10-20 ng for the whole lane, which was insufficient for

visualisation on a gel. SYBR® Green staining, with a better sensitivity of 0.06 ng per band, could have been used but no facility was available for this technique [141].

The samples were deemed degraded by the decreasing concentrations of DNA (Table 6.5) and the increasing ratios of the concentrations of the autosomal target to the Y target of Plexor® HY (Table 6.6). The autosomal target was 99 bp long and the Y target was 133 bp long [154]. Products of different length had been used to evaluate the degree of degradation in DNA samples [149, 150] (see Section 2.3.7.1). Day 0 samples for both experiments (KF only – male) had ratios close to one (Table 6.6), while the samples from other days had much higher ratios, indicating that the smaller autosomal target survived better than the longer Y target, which was expected in degraded samples [1]. Moreover, the incubation method used here brought about degradation of DNA in prior studies, as shown by SGM+ EPGs and linear regression relating allelic dropout to product size [17, 109]. In both studies, it was shown that allelic dropouts in incubated saliva samples increased progressively with incubation time.

Table 6.6. The ratio of the concentrations of autosomal to Y targets for KF samples.

Day experiment		Week experiment	
Time (days)	**[Auto]/[Y]**	**Time (days)**	**[Auto]/[Y]**
0	0.75	0	1.10
1	2.55	7	2.10
2	2.40	14	2.55
3	1.80	28	3.15
4	2.15	56	3.85
5	2.20	70	2.70
6	2.15		
7	3.05		
14	2.75		
28	2.35		

The incubation method was thought to be a more “natural” way to obtain degraded DNA, as the histones were not intentionally removed prior to degradation. Thus, it was possible for the DNA bound in nucleosomes to be protected from degradation

enzymes, oxidation and hydrolysis. The conventional way of artificially degrading DNA for experiments involves subjecting purified, extracted DNA to DNase I, MNase, sonication or a combination of these methods. Degradation in these "artificial" ways would only cut up the pieces randomly, albeit with a slight sequence preference depending on the enzyme used. Furthermore, degradation of DNA has been shown to be faster in a hydrated (high humidity) state, as was used in this experiment, than in a dried state [108]. The term "natural" here was used to contrast with the "artificial" ways of degradation. It was not used to mimic real-world degradation in forensic stains, as these stains are dry and not subjected to 100% humidity as in this method.

The degradation pattern from this incubation method was also observed in studies employing DNase I as the degrading agent. A rapid decrease in DNA concentration was seen in the early minutes of degradation. Swango *et al.* [149] found that by 2.5 minutes, only half of the starting amount of DNA remained. In addition, Timken *et al.* [151] also observed a rapid degradation rate in the first five minutes, followed by a more stable decrease after fifteen minutes of degradation with DNase I. The incubation method was comparable to the artificial, enzyme-based method in terms of the rate of degradation. This result also agreed with Johnson and Ferris [112], who demonstrated that a faster degradation rate was observed in the first 24 hours post-mortem. Although the end result of both degradation methods was the same, the incubation method should give degraded samples that could be used for further investigation of nucleosome protection during degradation.

6.3.1.3 Statistical comparisons

Statistical analysis was used to determine whether the two donors differed in their DNA concentrations in the various degradation samples. Non-parametric tests were used, as the assumptions of randomness, normality, and homogeneity of variance could not be fulfilled. The Mann-Whitney *U* test [146] was used to compare the distributions in DNA concentration between the two donors. The null hypothesis was *"the distributions of DNA concentration were the same across the two donors"* while the alternative hypothesis was *"the distributions of DNA concentration were different"*. The test revealed that the null hypothesis could not be rejected, indicating that the two donors did not differ significantly ($p = 0.423$, $N = 32$, Mann-Whitney *U*

= 106.0). Thus, using the information from both day and week experiments, it was concluded that the distributions of the two donors did not differ significantly. Analyzing the two experiments separately revealed that the two donors still did not differ significantly in either the week experiment ($p = 0.065$, $N = 12$, Mann-Whitney $U = 6.0$) or the day experiment ($p = 0.579$, $N = 20$, Mann-Whitney $U = 58.0$).

The Mann-Whitney U test was used to determine if there was any difference in the DNA concentrations of the saliva samples collected during the two different experiments. Only samples with the same time-points in both experiments were used (0, 7, 14, and 28 days). The null hypothesis for this test was *"the distributions of the DNA concentration were the same across the two experiments (day and week)"* while the alternative was *"the distributions of DNA concentration were different"*. This test indicated that the saliva samples did not differ in the two experiments ($p = 0.959$, $N = 16$, Mann-Whitney $U = 31$).

Saliva was used in this study to obtain a degradation series, as saliva had been shown to degrade faster than blood, due to the presence of various enzymes that aid degradation process as well as countless bacteria [17]. The bacteria in the human mouth produce an additional degradation enzyme called micrococcal nuclease, which cleaves DNA non-specifically. Moreover, saliva could be collected non-invasively, which was an advantage over the collection of blood samples.

The degradation method used in this study was adapted from [17], and it is worth noting that this was only the second study known in the forensic literature to use this method of *in vivo* degradation in contrast to artificial degradation of DNA using nucleases [101, 150].

6.3.2 Primer design and optimisation

Primer sets were optimised using the Taguchi method and the L_9 orthogonal array (see Chapter 5). Optimisation failed for: D16S539 due to presence of primer dimers; D1S1171 due to high percentage errors (34.38%); and D12ATA63 due to low amplification efficiency and r^2 in the calibration curves. All other primers were successfully optimised with acceptable percentage errors of less than 15%, except for D9S2157 (Table 6.7). Optimal conditions were used for the comparison of degraded saliva samples and simulated casework samples (Sections 6.3.3 and 6.3.3.2).

Table 6.7. The optimal conditions (forward primer concentration, reverse primer concentration, and annealing temperature) of the 14 STR primer sets. The percentage error (PE) marked with an asterisk (*) indicated a percent over the recommended value of 15%.

Locus	Forward (μM)	Reverse (μM)	Temp (°C)	PE(%)	r^2	Eff
D10S1248	0.40	0.20	62	14.71	.993	100.8
D20S1082	0.40	0.60	62	2.18	.996	96.1
D9S2157	0.40	0.40	60	17.76*	.998	99.2
TH01	0.60	0.40	65	10.82	.998	99.1
TPOX	0.60	0.60	65	9.68	.999	92.6
D1GATA113	0.40	0.40	62	5.16	.998	106.8
D8S1179	0.60	0.60	62	14.33	.995	97.1
FES	0.40	0.40	65	10.49	.992	103.8
D4S2364	0.40	0.60	62	4.49	.996	99.7
CD4	0.40	0.60	65	1.29	.997	99.6
D18S51	0.20	0.60	60	7.80	.993	99.5
D22S1045	0.40	0.60	60	8.12	.997	97.4
D5S2500	0.40	0.60	60	8.66	.999	99.4
D21S11	0.60	0.20	60	8.82	.997	103.7

6.3.3 Nucleosome protection

6.3.3.1 Saliva samples

6.3.3.1.1 Quantification results

Degraded saliva samples were quantified using 14 primer sets randomly selected to represent the three nuScore groups (see Appendix Table 9). Further investigation into the data revealed that most of the degraded saliva samples had been contaminated in the laboratory with SGM+ allelic ladder, leading to elimination of four loci that overlapped SGM+ (see Appendix A.9). All further analyses were carried out with the remaining ten loci.

The highest concentration of DNA in any sample was found with TPOX at 394.46 pg/μL (YS0), corresponding to almost 8 ng/μL in the original, undiluted sample. A trend of decreasing DNA concentration with increasing incubation time was

observed for all loci (Figure 6.2 and Figure 6.3), as previously demonstrated using Plexor® HY in Section 6.3.1.1. Every degraded saliva sample had less than 200 pg of DNA per µL (see Appendix Table 9) and most samples had DNA concentrations corresponding to only a single or a few cells. Nonetheless, quantification using the ten primer sets was possible up to 7 days of degradation. Further from that time-point (14 and 28 days), some samples did not cross the amplification threshold for both replicates, resulting in a "no C_q". One of ten loci and four of ten loci had "no C_q" at both replicates for KF and YS, respectively. The DNA concentrations of all loci were not from a normally distributed population ($p < 0.001$), but the variances were homogeneous ($p = 0.122$).

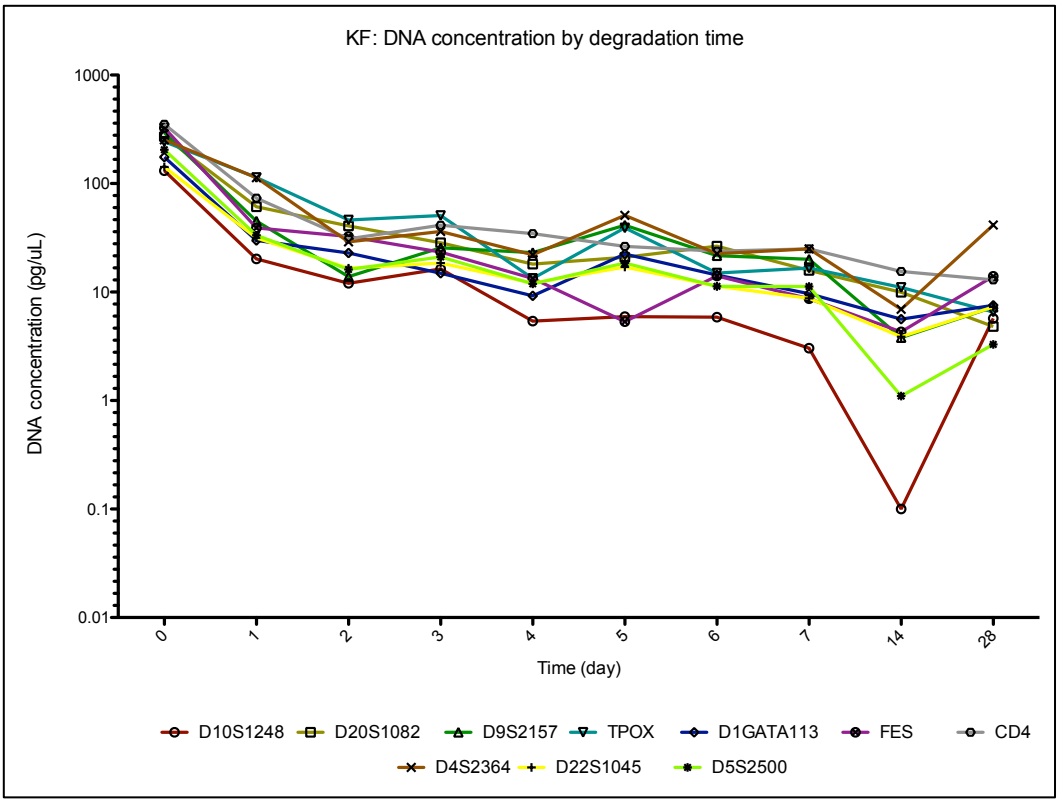


Figure 6.2. DNA concentration of KF saliva samples plotted by degradation time in days. Each line represents different STR loci. "No C_q" was given a value of 0.1 pg/µL.

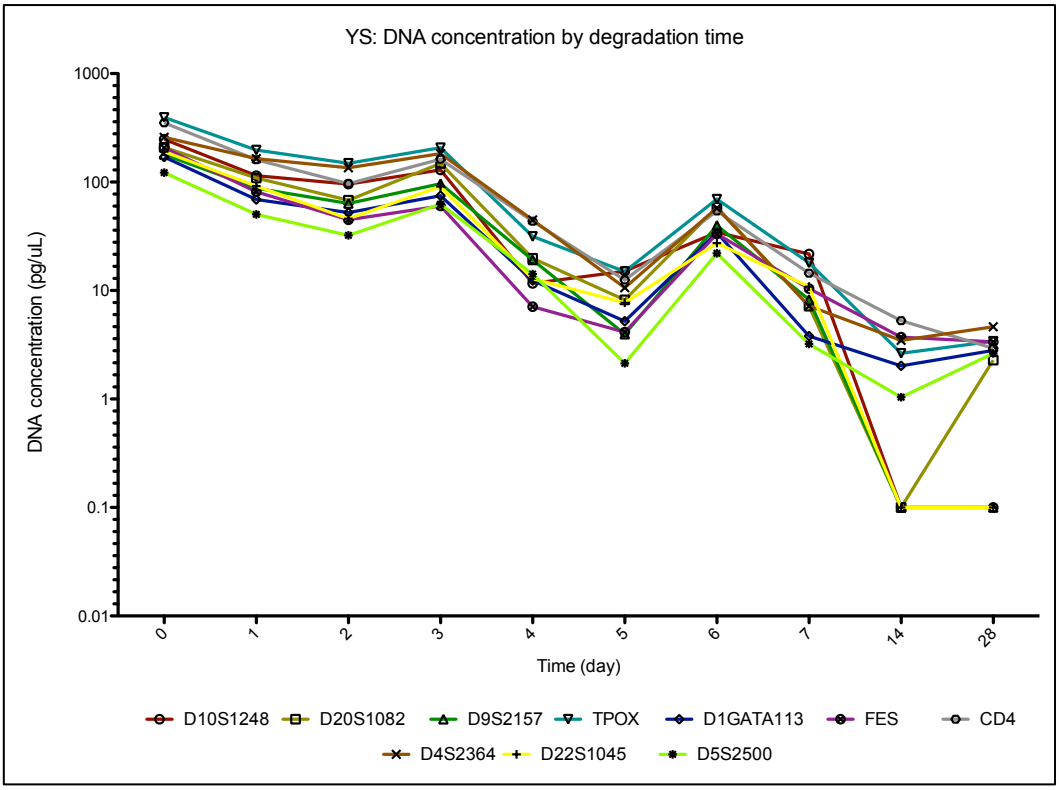


Figure 6.3. DNA concentration of YS saliva samples plotted by degradation time in days. Each line represents different STR loci. "No C_q" was given a value of 0.1 pg/µL.

6.3.3.1.2 Comparisons between nucleosome loci

The three nuScore groups [253] were compared to determine whether the group of loci with high nuScore survived degradation better than the other two groups (Figure 6.4). All groups exhibited a similar pattern of degradation at all days except for day 14.

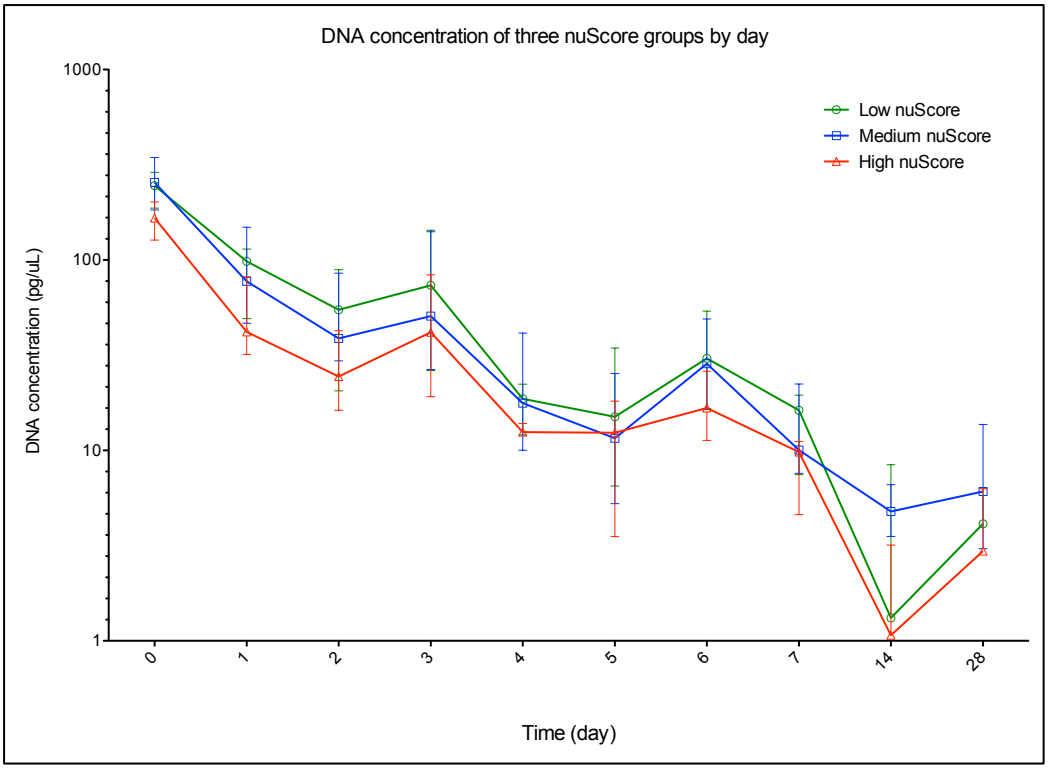


Figure 6.4. Median DNA concentration of three nuScore groups plotted by degradation time in days (four loci each for low and medium nuScore groups and two loci for high nuScore group). Error bars represent interquartile range.

The median DNA concentrations of low nuScore and medium nuScore groups were consistently higher than those of the high nuScore group up to seven days of degradation. This was surprising, as the high nuScore group was hypothesized to have the highest protective capabilities against degradation. The smaller sample size in the high nuScore group could have affected the result, as additional loci could have shifted the median (two loci compared to four loci for the other two groups).

Further investigations were performed to see if there was any significant difference in DNA concentrations among the ten loci using the Friedman's test, which ranked the concentration of each sample across the ten loci (Table 6.8). For example, the highest concentration of sample KF0 was seen with CD4 (351.38 pg/μL), and so this would be given the highest rank of ten. The next highest concentration was with FES (326.27 pg/μL) and so this would be ranked 9. This ranking was done for all samples and the distributions of the ranks were then compared. The null hypothesis was "*the distributions of DNA concentration of the ten loci are the same*". The alternative hypothesis was "*the distribution of each locus are not the same*". The null hypothesis was rejected ($p < 0.001$, Chi-squared = 93.161, DF = 9, N = 20 for each group).

Table 6.8. NuScore and mean ranks of the STR primers. Ranks when removing day 0 (non-degraded) samples are also shown.

Locus	nuScore	Rank	Rank w/o day 0
D10S1248	0	3.62	3.58
D20S1082	0	6.22	6.19
D9S2157	2	5.18	5.14
TPOX	2	8.35	8.44
D1GATA113	3	3.85	4.00
FES	4	4.70	4.44
CD4	5	8.45	8.33
D4S2364	5	8.20	8.33
D22S1045	6	3.68	3.75
D5S2500	6	2.75	2.78

It was clearly seen that the average ranks of CD4, TPOX and D4S2364 were the three highest, while the lowest average rank of 2.75 was seen with D5S2500. Unlike Timken *et al.* [151], the day 0 samples in this experiment were included in the Friedman's test due to its high robustness against outliers. Removing day 0 samples did not notably alter the ranks (Chi-squared = 84.538, DF = 9, N = 18 for each group) (Table 6.8). 18 of 45 pairwise comparisons gave significant differences, all of which included the three highest-ranked loci (CD4, D4S2364, and TPOX) except for one pair (D20S1082-D5S2500) (Table 6.9).

Table 6.9. Eighteen significantly different pairwise comparisons.

Locus 1	Locus 2	P-value	Locus 1	Locus 2	P-value
CD4	D10S1248	<0.001	D4S2364	D22S1045	<0.001
CD4	D1GATA113	<0.001	D4S2364	D5S2500	<0.001
CD4	D22S1045	<0.001	D4S2364	FES	0.012
CD4	D5S2500	<0.001	TPOX	D10S1248	<0.001
CD4	D9S2157	0.028	TPOX	D1GATA113	<0.001
CD4	FES	0.004	TPOX	D22S1045	<0.001
D20S1082	D5S2500	0.013	TPOX	D5S2500	<0.001
D4S2364	D10S1248	<0.001	TPOX	D9S2157	0.041
D4S2364	D1GATA113	<0.001	TPOX	FES	0.006

No correlation between the nuScore groups and the rank of DNA concentrations of the primers was observed using Spearman's ρ ($p = 0.647$, correlation coefficient = -0.166, $N = 10$). This confirmed the result from the comparison of three nuScore groups. Clearly, although the quantification results were statistically significant for some pairwise comparisons, a locus with a higher nuScore was not more successful in amplifying degraded DNA.

Using the same three primer sets in this study, Paez [259] degraded commercial DNA samples with DNase I. She observed that D22S1045 (group 3) performed worse than the other0 two loci from groups 1 and 2. This was not surprising, as commercial DNA extracts were purified before degradation and thus no nucleosome protection would have been in place. In contrast, the degraded saliva samples in this study should have been protected by nucleosome and so group 3 was expected to perform better.

Shorter PCR products tended to survive the degradation process better than longer products [149, 150]. In this study, the ratio of the longest product to the shortest product possible was 1.88 (CD4 to TPOX) (Figure 6.5). Knowing the exact sizes of the PCR products would allow a statistical examination of the effect of size differences and could have been done using a chip-based electrophoretic system such as the Bioanalyzer (Agilent Technologies). A model taking into accounts both factors

– size and nuScore – could be constructed with this additional information to assess the contribution of each factor to the variation in quantification results.

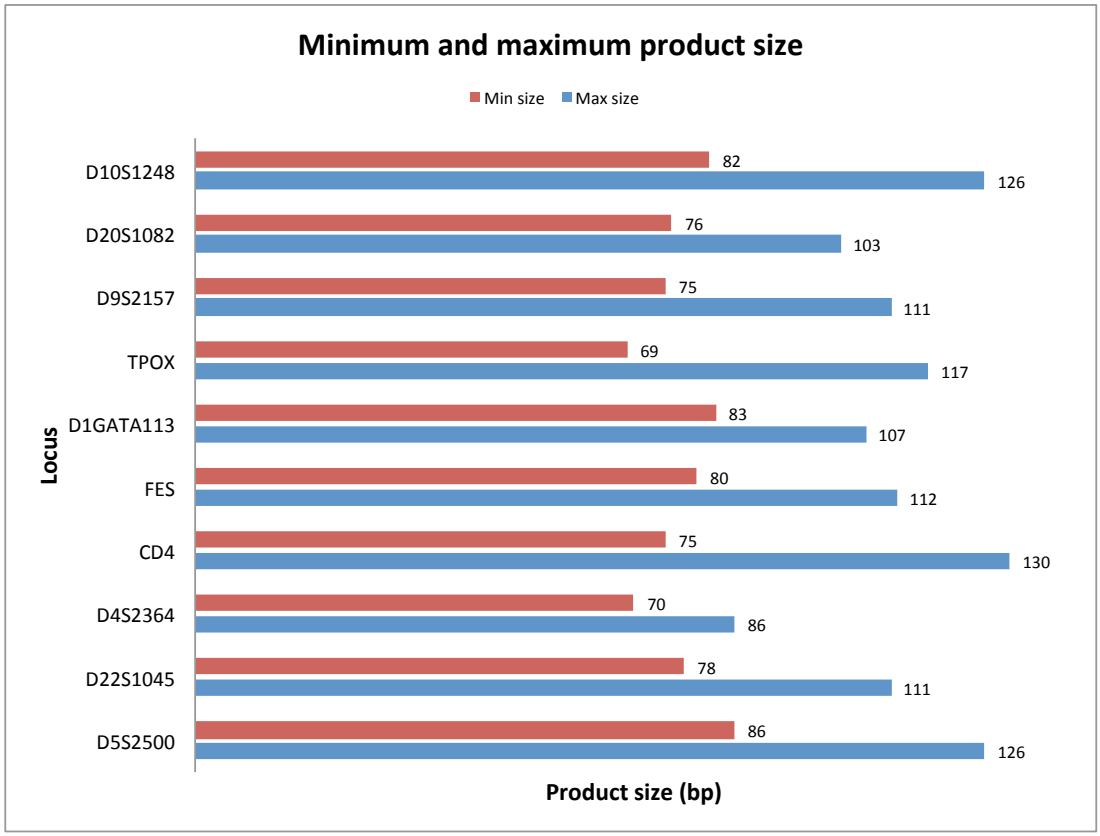


Figure 6.5. The minimum and maximum product sizes of all nucleosome loci based on primer binding sites and Genbank GRCh37 (Feb 2009) sequences.

Additionally, a size-stratification strategy, in which one forward primer and multiple reverse primers are used to amplify products of different lengths – could be utilised to directly quantify the effect of product size on survivability. Alternatively, it is also possible to look for evidence of nucleosome protection if an access to a large dataset of EPGs is available. As NFPs of STR loci differ, combing a large dataset to detect for the effect of nucleosome protection is possible; however, other factors, such as product size, that are known to affect peak heights and number of allelic dropouts in degraded samples must be kept in mind by accounting for their influences in a statistical model. As such, a direct experimental approach that was used in this study is preferred, as many factors, including product sizes, can be controlled to a certain degree. Loci from different groups can also be randomly selected to prevent any bias in data analysis.

Significant differences between certain loci but not between nuScore groups could have been caused by a number of factors. The different amplification efficiency between loci, even after optimisation and normalisation, could have been the cause. Nonetheless, the r^2-values and amplification efficiencies were all within the recommended range and thus should not drastically affect the results (Table 6.10).

Table 6.10. R-squared value and amplification efficiency of each primer set for degraded saliva samples.

Locus	r^2	Efficiency (%)	Locus	r^2	Efficiency (%)
D10S1248	.993	100.8	FES	.992	103.8
D20S1082	.996	96.1	D4S2364	.996	99.7
D9S2157	.998	99.2	CD4	.997	99.6
TPOX	.999	92.6	D22S1045	.997	97.4
D1GATA113	.998	106.8	D5S2500	.999	99.4

Secondly, nuScore might not be a true indicator of nucleosome occupation. The prediction by nuScore was only based on one known property of nucleosome positioning [94]. This meant that a high nuScore might not equal actual nucleosome occupation. Hence, there was no correlation between nuScore and ranks of loci. Other programs, such as FineSTR [95], could be used in conjunction to evaluate the nucleosome forming potentials of STR loci. Better models for nucleosome prediction are constantly being made [71].

Finally, protection by nucleosomes might not exist in degraded saliva stains. Loci with a high nuScore might be occupied by a nucleosome, but during cell death, most or all histones could be removed by lysosomal proteases [92]. In apoptosis, a cell death pathway, chromosomal DNA is degraded by endonucleases to form oligomers of about 180 bp; while in necrosis, lysosomal proteases help to remove histones, followed by a random degradation of DNA by endo- and exonucleases [92]. After a certain period of degradation, there would be no difference between a nucleosome-bound locus and a nucleosome-free locus. To date, the mechanism for degradation of DNA in forensic stains has not been characterized [92]. The lack of difference between the groups supports the idea that the degradation process in saliva stains is mainly due to necrosis.

Instead of using the incubation method, MNase digestion could have been applied directly to fresh saliva samples to obtain mono-nucleosomes followed by DNA extraction using an adapted protocol from nucleosome mapping studies [260]. Then the mono-nucleosome samples can be amplified using the STR primer sets to determine whether some loci are more likely to be found as nucleosomes. As MNase preferentially cuts linker DNA [62, 114], loci found within nucleosomes should give a higher quantification result. This information could then be correlated to NFP data from a bioinformatics program like nuScore and would serve as a proof-of-concept.

6.3.3.2 Simulated casework samples

To determine whether the results obtained with a known degradation series could be further confirmed with other sample types, five STR loci were chosen for further testing using simulated casework samples: CD4 and D4S2364 (highest DNA concentrations in saliva samples); D10S1248 and FES (lowest DNA concentrations); D20S1082 (middle DNA concentrations). Seventeen simulated casework samples were amplified with these five STR primer sets (see Appendix Table 10). The quantification result of the FES locus was noticeably lower than the other loci, while CD4 and D4S2364 were the loci with highest medians (Figure 6.6). The results were not normally distributed (all $p < 0.05$) and the variances were not homogeneous ($p < 0.001$, Chi-squared = 24.5598, DF = 4).

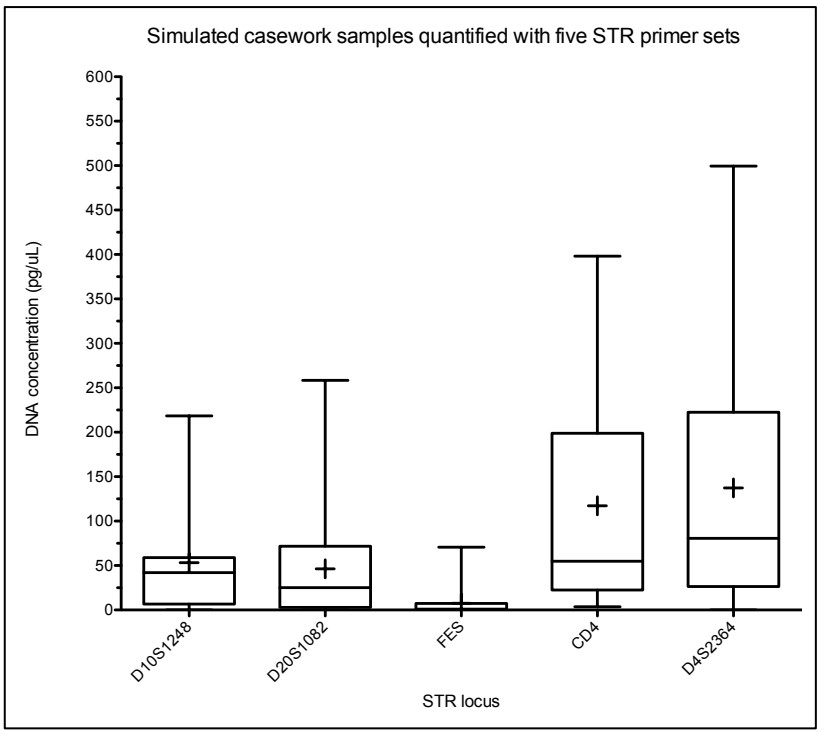


Figure 6.6. Boxplot showing DNA concentrations of simulated casework samples (N = 17) by locus. Plus (+) indicates mean DNA concentration.

The differences in the results of each locus were statistically examined using the Friedman's test. The null hypothesis was "*the distributions of the five loci are the same*" while the alternative hypothesis was "*the distributions among them are different*". The null hypothesis was rejected ($p < 0.001$, Chi-squared = 38.514, DF = 4, N = 17 for each group). Ranking the loci based on their performances exhibited a similar pattern to the degraded saliva samples, with CD4 and D4S2364 having a higher ranking, while D20S1082, which was one of the middle-ranked loci, had a slightly lower rank than D10S1248 (Table 6.11).

Table 6.11. Friedman's two-way ANOVA result from the simulated casework samples. N = 17 for each locus. Pairwise comparisons that were significant are shown with corresponding p-values. R^2 and efficiency (%) of each locus are also shown.

Locus	nuScore	Rank	r^2	Eff	Significant pairs (p-value)	
D10S1248	0	2.85	.998	95.9	-	
D20S1082	0	2.41	.998	100.2	-	
FES	4	1.44	.993	103.8	-	
CD4	5	4.35	.996	103.6	D20S1082 (0.003)	FES (<0.001)
D4S2364	5	3.94	.995	105.4	D20S1082 (0.048)	FES (<0.001)

The results from amplifying simulated casework samples with five selected STR primer sets confirmed that nuScore was not a good indicator of survivability of an STR locus. Although both D10S1248 and D20S1082 had a nuScore of zero, both were ranked higher than FES, which had a nuScore of four. As previously discussed, the difference in quantification results observed could be due to many reasons, such as nuScore not being a good predictor of nucleosome occupation or removal of histones by proteases.

6.3.3.3 Inhibition of Brilliant® II SYBR® Green Low ROX kit

The simulated casework samples were also quantified using the Plexor® HY kit (see Appendix Table 10). The ratio of the autosomal target of Plexor® HY to the average DNA concentration of the five STR primer sets was 13.8±23.5, suggesting a difference between the two kits by about one order of magnitude. It was hypothesized that this difference was due to PCR inhibition.

Hematin was used to determine the capability of the Brilliant® II SYBR® Green Low ROX kit to withstand inhibition. The only reactions that crossed the amplification threshold were the replicate that contained no hematin (0 μM). Their reported DNA concentrations were 2.37 ng/μL and 2.16 ng/μL. All the other spiked reactions did not amplify at all, indicating complete inhibition of PCR, even at a low concentration of 5 μM. This suggested that the Brilliant® II SYBR® Green Low ROX kit was very sensitive to inhibition by hematin.

Two samples (0-A and 0-B) were selected for duplicate amplification with SGM+ by adding an optimal template amount (1 to 2 ng) as determined by the two different quantification kits. Full profiles were obtained from reactions in which the optimal amount was based on Plexor® HY quantification while no profiles were observed using the STR primer quantifications (Figure 6.7). The colour of both DNA extracts was yellowish, suggesting the presence of heme-compounds. Since the simulated casework samples were extracted using the Chelex procedure, adding too large a bloodstain could result in some inhibition [1]. This could be the reason why the reported DNA concentrations of the STR primer sets were much lower than those of Plexor® HY.

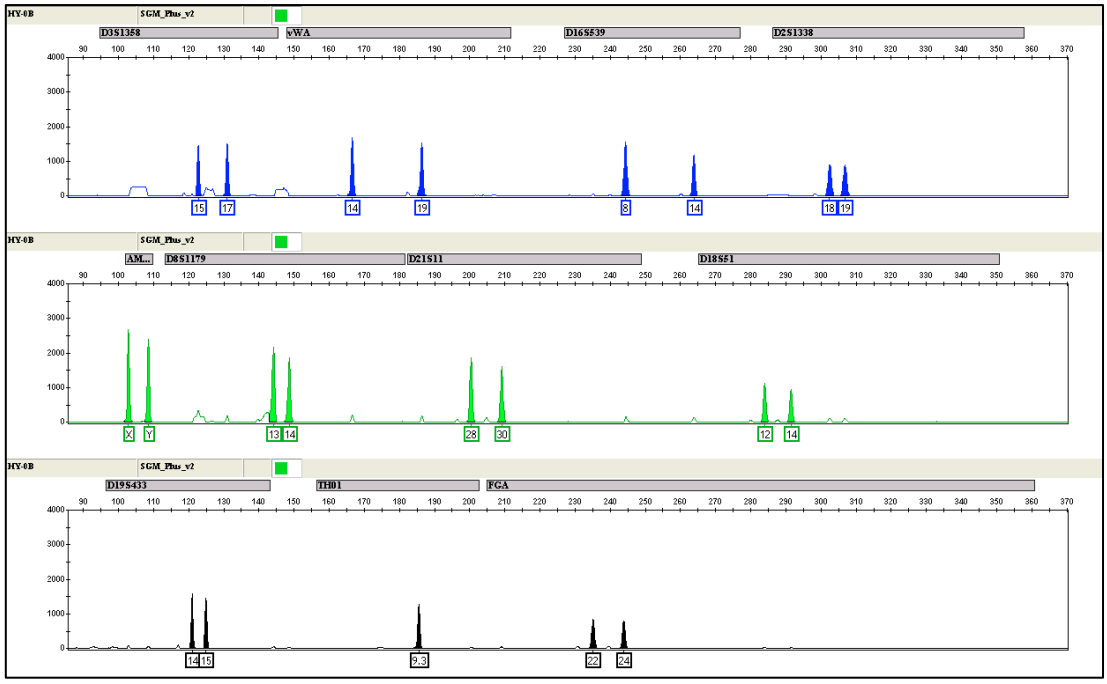


Figure 6.7. EPG of sample 0-B amplified with SGM+ based on HY quantification result.

Plexor® HY contained an internal positive control (IPC) that functioned as a detector for PCR inhibition. In the absence of any inhibition, these IPC should have C_q values of approximately 22-23. However, most samples had no inhibition based on their IPC results, although samples 0-1, 0-2M, 0-3, 0-4, 1-A, 1-1M and 1-3 had delayed C_q values of about 1 cycle (data not shown).

Different kits, with their different buffers, primers and polymerases, have different thresholds for PCR inhibitors. In a study by Krenke [121], Plexor® HY could withstand up to 25 uM of hematin, at which point a delay of more than 2 C_q values in the IPC (2 ng) were observed, whereas Quantifiler® Duo showed a significant decrease in IPC concentration at 10 µM of hematin, followed by complete inhibition at 15 µM [122]. In contrast, Swango *et al.* [150] saw a delayed C_q at 45 µM of hematin. This means that the casework samples could have contained inhibitors which were at a level only detrimental to the Brilliant® II SYBR® Green amplifications.

Another factor that could have contributed to the difference in quantification result was the time elapsed between the two quantification methods. The Plexor® HY experiment was carried out approximately four weeks before the STR primer sets. Since the Chelex resins were prepared in water, some degradation of DNA was

expected after four weeks of storage at 4°C, due to acid hydrolysis and lack of chelation by EDTA (Chelex beads removed prior to storage) [261].

7 General discussion and conclusion

The process of forensic identification has come a long way since its inception in 1985, when Alec Jeffreys developed and published the technique of "DNA fingerprinting" [4, 5, 262]. The first case that this technique was used for was a first in many ways. Two homicides – Lynda Mann and Dawn Ashworth – took place in 1983 and 1986 in Narborough, UK. Richard Buckland, who confessed to the murder of Dawn Ashworth, denied murdering Lynda Mann. However, DNA fingerprinting showed that the semen samples taken from the two crime scenes were from the same person, and the profile did not match Richard Buckland [263]. This made him the first person to be exonerated using DNA evidence [263]. The case also initiated the first DNA "mass screen" in 1987, in which 4000 men from the area were subjected to DNA profiling over the course of six months. Colin Pitchfork, who evaded the screening by paying Ian Kelly to give a sample in his place, raised police suspicion and was arrested in September of the same year. He was finally sentenced to 30 years of imprisonment in January 1988 [1, 7, 8].

Over twenty years since that first case, the current technology in use in forensic DNA profiling is based on STR analysis. The first national DNA database was set up in the UK in 1995, and currently holds the DNA profiles of over 5.4 million individuals [264]. In 2008/09 alone, there were 36,727 crimes with a scene-to-database match [265]. The largest database in the world is CODIS (US), which as of last November contained over 9 million offender profiles, and has assisted over 120,000 investigations to date [266]. In addition to establishing scene-to-scene and scene-to-suspect links, DNA can be used for intelligence purposes. A familial match, in which DNA from a crime scene is closely matched to a DNA profile in the database, can lead investigators to a possible blood relative of the perpetrator and eventually to the perpetrator [267]. This technique was first used in the UK in the case of Craig Harman, who never had a prior conviction, and thus his DNA was not in the database. A close match (16 out of 20 alleles) to his relative's profile in the database was found, and subsequently led to Mr. Harman's arrest and confession [268]. Other intelligence leads from DNA that are not based on STRs include accurate predictions of hair colour [269] and eye colour [270, 271], but the two technologies have not found their way into mainstream use.

In a casework environment, the DNA Advisory Board Quality Assurance Standards for Forensic DNA Testing Laboratories mandates that all forensic DNA samples are quantified before STR genotyping can take place [272]. The current standard procedure for quantification is qPCR-based, which is favoured because it is straightforward, species-specific and automatable. It can also be multiplexed, and has a high dynamic range [122]. Accurate forensic quantification is very important, because STR genotyping kits have a narrow optimal input range of 0.5 to 2.0 ng [154]. When the input is lower than the optimal range, a partial profile could result, while higher than optimal range could produce off-scale data [122]. Phenomena, such as allelic dropout and dropin, lower heterozygous balance and increased stutters, are common outside the optimal range, and complicate profile and mixture interpretation [198].

Moreover, accurate quantification empowers a forensic scientist to wisely choose a downstream protocol. An appropriate method can be chosen and carried out depending on the sample quality and quantity. For instance, in the case of a mixture from a female victim and a male perpetrator, knowing the quantity of male DNA in the sample helps with choosing whether Y-STR or conventional STR is best for the case. If there is a small amount of male DNA, Y-STR is given a higher weight when making the decision. If there is a lot of male DNA, conventional STR can be used, and in the case of very low or undetectable male DNA, there might be no need to genotype the sample. In cold cases or cases where samples are very limited, using a small amount of DNA extract, e.g. 2 μL, for quantification provides a forensic scientist with knowledge to make an informed decision. Choosing the most suitable method for DNA profiling undoubtedly increases the chance of obtaining a good STR profile and consequently a higher evidential value.

Quantification of forensic samples by qPCR has evolved over the years, and efforts have been made using both multi-copy and single-copy loci, including mitochondrial DNA [147], *Alu* elements [148], human RNU2 multi-copy locus [121], STR loci [149], and other single-copy loci [122]. The multi-copy loci have higher sensitivity than single-copy loci [121], e.g. mitochondrial DNA makes a highly sensitive assay, as there are over 1,000 copies per cell [273]. As a result, a smaller amount of sample extract is needed for the quantification assay, and the majority of the extract is left

for downstream analyses. However, the variation in copy number, e.g. five to 20 copies in the human RNU2 target in Plexor® HY [121, 154], between individuals and DNA standards can result in an overestimation or underestimation of the DNA concentration of the unknown sample. In contrast, single-copy loci provide an accurate prediction of the result of subsequent STR genotyping, as the copy number is identical to an STR locus [154]. The ratio of autosomal DNA to Y DNA in a sample is also more accurate, providing a better estimate of female to male DNA in a mixture [154]. Nonetheless, due to the variation in length of different alleles, using an STR locus itself as a target in a quantification assay could affect the amplification efficiency and the quantification result might not be accurate [122].

Both multi-copy and single-copy loci were explored as candidates for accurate quantification of forensic samples. Optimisation was necessary before any newly designed assay can be used to obtain reliable results. The optimisation strategy used in these experiments was time-consuming due to the trial-and-error approach. Preliminary trials with *Alu* elements and single-copy loci (TERT, ALDOA and RPPH1) failed due to numerous reasons, such as low r^2, too low or too high efficiency, and the presence of secondary structure. Since the assays were detected by SYBR® Green I, the fluorescence gain from amplification of non-specific products and secondary structures interfered with the detection of the intended target. This called for a more stringent primer design, and additional optimisation can be pursued, such as increasing the concentration of magnesium chloride to increase assay specificity. A hydrolysis probe-based method can also be tried in order to only detect desired PCR product. Better purification methods, e.g. HPLC, eliminate truncated primers from the synthesis process [182]. It was suspected that the truncated primers present in the *Alu* study could mis-prime with other subfamilies of *Alu* and result in some amplification of non-specific products. GAPDH, a housekeeping gene with 62 pseudogenes [158], was successfully optimised, but a decision was made to use Plexor® HY to quantify casework and degraded DNA samples due to it being a multiplex kit that can simultaneously yield useful information regarding sample inhibition and autosomal to Y DNA ratio. The kit had also been validated and shown to work well with degraded samples [121].

Conventional STR kits experience a phenomenon called "preferential amplification" in degraded samples [274, 275]. Preferential amplification results in an unbalanced profile, with smaller loci being successfully genotyped and larger loci failing to amplify. This happens because the longer loci have a higher probability of being degraded by nucleases and other damaging processes [106, 107]. Wiegand and Kleiber [38] first demonstrated the benefits of shortening PCR products by shifting the primers to bind as close as possible to the repeat units. Since then, many loci have been explored and shortened to "mini-STRs". They have been used multiple times in real cases, such as genotyping a burnt cigarette stub from a fire scene [276], degraded bone samples in various stages of decomposition [47], 14-year-old bloodstains [35], decomposed and charred femur and tibia [46], and skeletal fragments from the World Trade Center incident [277].

From this pioneering work and through recommendations by professional forensic bodies, e.g. SWGDAM and ENFSI, manufacturers of STR kits such as Applied Biosystems and Promega Corporation began investing in and developing new kits. The first commercial mini-STR kits available were AmpF*l*STR® MiniFiler™, PowerPlex® S5, and Mentype® ESS, followed by the next-generation PowerPlex® ESX/ESI and AmpF*l*STR® NGM™ kits. Developmental validations have to be carried out by the manufacturers as well as internal validations by laboratories looking to adopt the kits [1]. The sample types required for validation of any kit include, but are not limited to, standards, mixtures, simulated casework samples and non-human samples. Other factors that form parts of a validation procedure are consistency, precision, stutters, heterozygous balance and inhibition studies [1]. Casework samples, whether simulated or adjudicated, are integral to the validation of these new kits, because they determine whether the kit under investigation is robust enough for real forensic samples. The SWGDAM validation guidelines [196] stress the importance of using casework samples: "*The ability to obtain reliable results should be evaluated using samples that are representative of those typically encountered by the testing laboratory"*.

Five next-generation and mini-STR kits were evaluated as a part of this study. Comparing different STR, next-generation and mini-STR kits is complicated by the fact that the number of loci and the targeted loci are different. Even at the same

locus, the primers have been designed differently and so the amplicons are of different lengths. Compounded by the fact that the other components included in the buffer, such magnesium chloride, BSA, and dNTPs, differ for these kits, a conclusive decision about which one to use is difficult to make. Nonetheless, all kits performed equally well in all respects; thus, any laboratory looking to adopt one of the kits should consider the intended purpose of the kit rather than the performance. For instance, if extra discrimination power is required, the NGM™, ESX and ESI kits should be the first choices. The NGM™ kit has a probability of identity (PI) value of 2.21×10^{-19}, which is lower than the MF kit (8.21×10^{-11}) and SGM+ kit (2.99×10^{-13}) [43]. Nonetheless, the MF kit works well as an adjunct kit for degraded samples, as the loci in the MF kit are mini-STRs of the largest loci in the SGM+ and Identifiler® kits plus additional CODIS loci. The ESS kit, incorporating all of the new ENFSI recommended loci [187], is also intended as an adjunct kit. In the case of cross-border data sharing in Europe, the NGM™, ESX, and ESI kits are most appropriate because they all amplify the same loci, all of which overlap with both CODIS and ENFSI recommendations. In contrast, the S5 kit, which is the least discriminatory but at the same time the cheapest, is perfect for screening samples to exclude individuals in casework scenarios or in preliminary mass screenings or a mass disaster.

As with all techniques, there are limitations to mini-STRs. Redesigning primers to mini-STRs makes the product sizes overlap. This means that at the current detection limit of five dyes, approximately no more than eight to ten loci can be mini-STRs in a kit, restricting the discrimination power. Additionally, some loci cannot be made into mini-STRs due to their extensive allele range. Mobility modifiers have been used by Applied Biosystems to spread out the loci [32, 197], but they are not available to other manufacturers due to their proprietary nature and patent restriction. Promega have tried to overcome this limitation by shuffling the mini-STR and STR loci in their ESX and ESI kits, e.g. a locus is a mini-STR size in one kit and a conventional STR size in the other kit. Hence, degraded samples can be typed once with each kit to obtain the maximum number of alleles possible [225]. Another problem with new kits is non-concordance, in which moving the primers closer to the repeat units results in a different allele to the one typed with conventional STR

primers [278]. This may happen if one set of primers spans a primer-binding site mutation which will result in a null allele; or, a deletion in the flanking region included with conventional STR primers will not be amplified with mini-STR primers binding closer to the repeat unit, resulting in a different allele. Concordance studies have to be carried out to ascertain that the new kits will show the same allele and further investigations should be made in the case of non-concordance as to why the alleles are different [195, 279]. One way of overcoming non-concordance in the case of primer-binding site mutations, is by including a degenerate primer to bind over the mutation site [278].

Using mini-STRs is one way to obtain an STR profile from degraded samples. Other techniques currently used or under investigation include pyrosequencing [227], increased PCR cycles [198], post-PCR purification [201], whole genome amplification [200] and mass spectrometry [280]. Nevertheless, an intrinsic property of DNA sequence that could protect STR loci from degradation has been overlooked. It has been suggested that nucleosomes could protect the DNA bases inside from degradative enzymes and processes [16, 49], therefore, choosing loci that are more likely to be protected could increase the chances of obtaining a full profile from degraded samples. This protection is partly due to the structural changes imposed by the curving of DNA around the octameric histone core, making it difficult for enzymes to access the underlying bases and binding sites [64]. MNase, a nuclease produced by *Staphylococcus aureus* found on human skin, illustrates how degradation of DNA could be protected by nucleosomes. MNase progressively digests linker (non-nucleosomal) DNA and then the more exposed ends of nucleosomal DNA, leaving a protected, core unit of approximately 125 bp [49]. In constitutive genes, the promoter and enhancer regions are devoid of nucleosomes, as constant access to the DNA is needed for transcription; whereas in regulated genes, these regions are often blocked by nucleosomes [72, 83-86].

Prediction of nucleosome position is complicated, because nucleosomes are positioned differently *in vitro* and *in vivo*. Nucleosome positioning and nucleosome occupancy is also a matter of intense debate, especially regarding the properties that determine these locations [65, 71, 72, 87, 88, 90, 251, 281, 282]. *In vitro* positioning of nucleosomes has been shown to largely depend on DNA sequence, with certain

sequences and dinucleotide combinations requiring less energy to bend than others [283-285]. This property is termed "deformation energy", and sequences that are easily bent are favoured by nucleosomes [62]. Evidence for the existence of periodic dinucleotide occurrences guiding nucleosome locations are also abundant [63, 67, 68, 82, 83, 282, 286].

In vivo nucleosome positioning is highly dynamic, with external factors such as chromatin remodelers, histone acetylations and DNA methylations reshaping the nucleosome landscape. Chromatin remodelling has been shown to override sequence-dependent nucleosome positions [72]. Acetylation reduces histone-DNA interactions, possibly facilitating sliding and eviction of histones [74], as well as interfering with higher order chromatin structures [55]. Furthermore, methylation moves nucleosomes away from methylated sites [65, 77]. Alternative to the *in vitro* sequence-dependent model, *in vivo* nucleosome positions might follow a barrier model, in which the first nucleosome downstream from a promoter region is well positioned. This well positioned +1 nucleosome imposes a constraint on the positions of subsequent nucleosomes with increasing fuzziness further away from it [78].

The deformation energies of STR loci were evaluated using two computer programs – NXSensor and nuScore. While NXSensor searches for sequences that are known to repel nucleosomes due to their stiffness [93], nuScore threads the STR sequence and evaluates the deformation energies of dinucleotides [94]. A locus with low deformation energy in multiple locations was deemed to have high NFP. It was hypothesized that loci with higher NFP could withstand degradation better than the ones with lower NFP due to the protection conferred by nucleosomes. Assessment of NFPs of forensically important STRs could be beneficial as nucleosome-bound DNA could be less likely to degrade due to being in the "bound configuration". The majority of the markers evaluated have no nucleosome inhibitory sequence. The range of possible dyad locations observed in this study suggests that the probability of a sequence to be bound varies from locus to locus. Selecting a high-NFP locus for a multiplex could increase the chance of obtaining a balanced profile with fewer allelic dropout. Using nuScore, the STR loci were categorized into three groups – high, medium and low NFP [253].

In order to evaluate whether there is a difference in terms of the survival rates of the loci between the three groups, a "natural" degradation method that preserves the chromatin structure must be used. Saliva samples were obtained from two donors and two time periods. They were used to explore a degradation protocol called the incubation method [17]. Saliva samples were successfully degraded by incubation at 37°C with 100% humidity. The temperature was maintained at 37°C for optimal degradative enzyme activity from both human endogenous nucleases and bacterial nucleases. Initial degradation was rapid, followed by a more gradual degradation rate. Although the DNA concentrations of the two donors differed, statistical analysis showed that the difference was not significant. Similarly, the two time periods were also shown not to differ significantly. These two comparisons confirmed that this degradation method provided a replicable way of degrading saliva samples.

Although one might argue that MNase is a better candidate for this study as the enzyme preferentially cuts linker DNA to yield intact nucleosomes [114], it was believed that this method could create a bias for nucleosome protection that might not exist in actual degradation of forensic samples. Another bias that could have been introduced with the use of MNase was preferential cutting of loci that contained CATA and CTA [115]. The conventional way of obtaining degraded samples is treating purified DNA with DNase I [101, 122, 193], but it was considered inappropriate for this study. This was because purified DNA had their histones removed during the process and thus no information would be gained regarding the protection by nucleosomes. Nonetheless, the rate and pattern of degradation from both the incubation method and DNase method seemed very similar [149, 151].

New mini-STR primers, chosen to amplify the loci from the three NFP groups, were designed to quantify the degraded saliva samples. These primers had to be optimised before any result could be considered valid. The factorial method utilized earlier to optimise primers for accurate quantification was inefficient and tedious because many reactions were needed. Other methods of optimisation, such as fractional factorial design and Taguchi method [245], that could help to save resources were considered. The Taguchi method, an optimisation method with roots in engineering, has been applied a number of times in biosciences [244], but never tried with a qPCR assay using SYBR® Green I dye. Employing the orthogonal array, it proved to be

quick, systematic and informative. Not only did it achieve the same results as the factorial method, additional information, such as percent contribution and prediction of untested levels with regression, were obtained [245, 247]. There are caveats associated with the method, including not being able to handle interactions between factors elegantly and lacking exploratory aspects [249, 287]. After successfully optimizing three SYBR® Green I assays, the Taguchi method was chosen as the method of choice for the optimisation of the new mini-STR primers for the investigation of nucleosome protection.

Degraded saliva samples were quantified with 14 newly designed STR primer sets, which had been optimised using the Taguchi method. The three nuScore groups were expected to show different survivability due to their differences in NFPs [253]. It was hypothesized that nuScore would predict the survivability of an STR locus in DNA degraded "naturally". It was expected that an STR locus with a high nuScore (indicating multiple locations in the DNA sequence where the nucleosome dyad could exist) would have a higher probability of being bound in a nucleosome. The association of DNA to histones in a nucleosome should limit the access of endo- and exonucleases to the bases within them. Consequently, they would survive better than unbound, exposed DNA. However, no correlation was observed between survivability of STR loci and their NFPs in degraded saliva samples, which could be explained by four reasons:

- Nucleosome protection might only work in living cells and not in hydrated, degraded samples or dead bodies. Following cell death and depletion of ATP, most or all histones could by removed during necrosis by lysosomal proteases, and subsequently DNA would be degraded by endo- and exonucleases [92].
- The way that NFP was scored might be unsuitable. A high NFP from nuScore was defined as having many possible dyad locations in the input sequence. It was hypothesized that having multiple locations indicated a higher probability of being a nucleosome. If, however, *in vivo* positioning of nucleosomes truly follows the barrier model discussed above, multiple locations in an input sequence could be interpreted that the exact location of nucleosome is "fuzzy" (not well positioned) [78]. Thus defining a high NFP

in this way would be inappropriate, as multiple locations would have no association with the probability of the sequence being bound to histones.

- The nuScore NFP was based on deformation energy, which could have easily been overridden by chromatin remodelers while the cells were alive [62, 65]. NuScore, as previously discussed in Chapter 4, was only based on one factor – dinucleotide stacking properties of the input sequence. Other factors were known to influence nucleosome occupation, with the common ones being GC content, chromatin remodelers, DNA methylation, and histone variants. The nucleosome positions in a newly synthesized DNA strand might be based on deformation energy determined by base sequence, then other processes such as chromatin remodelling and DNA methylation shift the nucleosomes to other locations. As the cell is dying the final locations of nucleosomes are maintained until cell death and beyond, because ATP-dependent enzymes would have stopped moving them.
- Additional uncertainty in the data could have been introduced by the differences in product size. It was assumed that this should not significantly affect the result, since all products were mini-STRs and roughly of similar sizes, but this effect was not quantified due to the lack of size information. Differences in amplification success of a short target to a long target has been documented [120, 149], e.g. in highly degraded samples from DNase degradation, the quantity of a 67 bp target was about five to eight times higher than a 170-190 bp target [120]. It would have been better if the qPCR products were sized and accounted for, probably in a statistical model.

From the results of this experiment, it can be concluded that nucleosome protection does not exist for "naturally" degraded saliva samples, given that nuScore-based NFP accurately represents the probability of finding a nucleosome. In order for a more general conclusion to be drawn, other nucleosome prediction programs must be tried, as well as other sample types and degradation techniques. More programs must be evaluated because different programs have been written using different algorithms. Researchers are always improving the models and datasets are constantly updated [71]. A new program called FineSTR has been claimed to be able to resolve nucleosome positions to a single-base resolution, with algorithms based on the

previously published 10.4 base periodicity and *Caenorhabditis elegans* "bendability matrix" [70, 95]. This algorithm is completely different to NXSensor and nuScore used here and should provide more insights into the issue of nucleosome protection. It was not used to evaluate the STR loci in this study, but given that the quantification had been done, it was possible to re-evaluate the STR loci with a new scoring system.

Further evidence can be sought by increasing the sample size, sample types, the number of primers and the number of donors. In this study only one type of sample – incubated saliva stain – was used in the degradation series. Obviously, other sample types such as bloodstains and semen stains should be evaluated, as some differences were observed between the degraded saliva samples and simulated casework samples that were mostly made of blood and semen stains. A different degradation pattern could be revealed. Coupling that with an increase in the number of donors and primers, there will be an increase in confidence in the result and more general conclusions can be drawn.

Forensic DNA and structural molecular biology are both rapidly moving fields. This nature dictates that hypotheses must be debated and updated in light of new information, not unlike the judicial process, in which the odds of innocence and guilt are constantly evolving through presentation of new evidence. Advancements in computer modelling and instrumentations increase the possibility of revealing and understanding even the tiniest change in DNA molecules and chromatin structure. Although the data obtained from this study does not support the hypothesis of nucleosome protection, it provides additional knowledge to both fields. At the very least, one now knows that nucleosome forming potentials of forensic STRs do not play a role in preventing DNA degradation in saliva samples. As STRs are firmly established in the forensic community, alternative methods that are STR-based and further expand the understanding of the mechanisms of DNA degradation, such as the one carried out here, are worth investigating. To date, this study is the first to directly investigate the possibility of using knowledge of nucleosome positions to understand and help to prevent DNA degradation.

References

1. Butler, J.M., *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*. 2005, Elsevier Academic Press: London, UK.
2. Watson, J. and F. Crick, *Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid.* American Journal of Psychiatry, 2003. **160**(4): p. 623-624.
3. Klintschar, M., U. Immela, M. Kleiber and P. Wiegand, *Physical location and linked genes of common forensic STR markers.* International Congress Series, 2006. **1288**: p. 801-803.
4. Jeffreys, A., V. Wilson and S. Thein, *Hypervariable 'minisatellite' regions in human DNA.* Nature, 1985. **314**(6006): p. 67-73.
5. Gill, P., A. Jeffreys and D. Werrett, *Forensic application of DNA 'fingerprints'.* Nature, 1985. **318**(6046): p. 577-579.
6. Jeffreys, A., J. Brookfield and R. Semeonoff, *Positive identification of an immigration test-case using human DNA fingerprints.* Nature, 1985. **317**(6040): p. 818-819.
7. Jobling, M.A. and P. Gill, *Encoded evidence: DNA in forensic analysis.* Nature Reviews Genetics, 2004. **5**(10): p. 739-751.
8. Saad, R., *Discovery, development, and current applications of DNA identity testing.* Proceedings (Baylor University. Medical Center), 2005. **18**(2): p. 130.
9. Ginsburg, J. and A. Costoff, *Gastrointestinal Physiology: Gastrointestinal secretions*, in *Essentials of Human Physiology*, T. M. Nosek, Editor. 2000, Gold Standard Multimedia Inc.
10. Kaplan, M.D. and B.J. Baum, *The functions of saliva.* Dysphagia, 1993. **8**(3): p. 225-229.
11. Dawes, C., *Estimates, from salivary analyses, of the turnover time of the oral mucosal epithelium in humans and the number of bacteria in an edentulous mouth.* Archives of Oral Biology, 2003. **48**(5): p. 329-336.
12. Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell*. 2002, Garland Science: New York, USA.

13. Quinque, D., R. Kittler, M. Kayser, M. Stoneking and I. Nasidze, *Evaluation of saliva as a source of human DNA for population and association studies.* Analytical Biochemistry, 2006. **353**(2): p. 272-277.
14. Lee, H.C. and C. Ladd, *Preservation and collection of biological evidence.* Croatian Medical Journal, 2001. **42**(3): p. 225-228.
15. Ng, D.P.K., D. Koh, S.G.L. Choo, V. Ng and Q. Fu, *Effect of storage conditions on the extraction of PCR-quality genomic DNA from saliva.* Clinica Chimica Acta, 2004. **343**(1-2): p. 191-194.
16. Dixon, L.A., C.M. Murray, E.J. Archer, A.E. Dobbins, P. Koumi and P. Gill, *Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes.* Forensic Science International, 2005. **154**(1): p. 62-77.
17. Dixon, L.A., A.E. Dobbins, H.K. Pulker, J.M. Butler, P.M. Vallone, M.D. Coble, W. Parson, B. Berger, P. Grubwieser, H.S. Mogensen, N. Morling, K. Nielsen, J.J. Sanchez, E. Petkovski, A. Carracedo, P. Sanchez-Diz, E. Ramos-Luis, M. Brion, J.A. Irwin, R.S. Just, O. Loreille, T.J. Parsons, D. Syndercombe-Court, H. Schmitter, B. Stradmann-Bellinghausen, K. Bender and P. Gill, *Analysis of artificially degraded DNA using STRs and SNPs--results of a collaborative European (EDNAP) exercise.* Forensic Science International, 2006. **164**(1): p. 33-44.
18. Richardson, R.L. and M. Jones, *A bacteriologic census of human saliva.* Journal of Dental Research, 1958. **37**(4): p. 697-709.
19. Wren, J.D., E. Forgacs, J.W. Fondon, A. Pertsemlidis, S.Y. Cheng, T. Gallardo, R.S. Williams, R.V. Shohet, J.D. Minna and H.R. Garner, *Repeat polymorphisms within gene regions: phenotypic and evolutionary implications.* American Journal of Human Genetics, 2000. **67**(2): p. 345-356.
20. Clayton, T., J. Whitaker and C. Maguire, *Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci.* Forensic Science International, 1995. **76**(1): p. 7-15.
21. Hagelberg, E., B. Sykes and R. Hedges, *Ancient bone DNA amplified.* Nature, 1989. **342**(6249): p. 485.
22. Higuchi, R., C. Von Beroldingen, G. Sensabaugh and H. Erlich, *DNA typing from single hairs.* Nature, 1988. **332**(6164): p. 543-546.

23. Hochmeister, M., R. Dirnhofer, U. Borer, B. Budowle, J. Jung and C. Comey, *PCR-based typing of DNA extracted from cigarette butts.* International Journal of Legal Medicine, 1991. **104**(4): p. 229-233.
24. Pizzamiglio, M., A. Mameli, G. Maugeri and L. Garofano, *Identifying the culprit from LCN DNA obtained from saliva and sweat traces linked to two different robberies and use of a database.* International Congress Series, 2004. **1261**: p. 443-445.
25. Pizzamiglio, M., A. Mameli, D. My and L. Garofano, *Forensic identification of a murderer by LCN DNA collected from the inside of the victim's car.* International Congress Series, 2004. **1261**: p. 437-439.
26. Pizzamiglio, M., A. Marino, A. Coli, T. Floris and L. Garofano, *The use of mini-STRs on degraded DNA samples.* International Congress Series, 2006. **1288**: p. 498-500.
27. Schulz, M.M. and W. Reichert, *Archived or directly swabbed latent fingerprints as a DNA source for STR typing.* Forensic Science International, 2002. **127**(1-2): p. 128-130.
28. Tanaka, M., T. Yoshimoto, H. Nozawa, H. Ohtaki, Y. Kato, K. Sato, T. Yamamoto, K. Tamaki and Y. Katsumata, *Usefulness of a toothbrush as a source of evidential DNA for typing.* Journal of Forensic Sciences, 2000. **45**(3): p. 674-676.
29. Wickenheiser, R., C. Macmillan and C. Challoner, *Case of identification of severely burned human remains via paternity testing with PCR DNA typing.* Canadian Society of Forensic Science Journal, 1999. **32**(1): p. 15-24.
30. Mullis, K., F. Ferré and R. Gibbs, *PCR: The Polymerase Chain Reaction*. 1994, Birkhäuser: Pennsylvania, USA.
31. Cotton, E., R. Allsop, J. Guest, R. Frazier, P. Koumi, I. Callow, A. Seager and R. Sparkes, *Validation of the AmpFℓSTR® SGM Plus™ system for use in forensic casework.* Forensic Science International, 2000. **112**(2-3): p. 151-161.
32. Collins, P.J., L.K. Hennessy, C.S. Leibelt, R.K. Roby, D.J. Reeder and P.A. Foxall, *Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFℓSTR®*

Identifiler® PCR Amplification Kit. Journal of Forensic Sciences, 2004. **49**(6): p. 1265-1277.

33. Greenspoon, S., J. Ban, L. Pablo, C. Crouse, F. Kist, C. Tomsey, A. Glessner, L. Mihalacki, T. Long, B. Heidebrecht, C. Braunstein, D. Freeman, C. Soberalski, B. Nathan, A. Amin, E. Douglas and J. Schumm, *Validation and implementation of the PowerPlex 16 BIO System STR multiplex for forensic casework.* Journal of Forensic Sciences, 2004. **49**(1): p. 71-80.
34. Coble, M. and J. Butler, *Characterization of new miniSTR loci to aid analysis of degraded DNA.* Journal of Forensic Sciences, 2005. **50**(1): p. 43-53.
35. Butler, J.M., Y. Shen and B.R. Mccord, *The development of reduced size STR amplicons as tools for analysis of degraded DNA.* Journal of Forensic Sciences, 2003. **48**(5): p. 1054-1064.
36. Opel, K., D. Chung, J. Drabek, J. Butler and B. Mccord, *Developmental validation of reduced-size STR Miniplex primer sets.* Journal of Forensic Sciences, 2007. **52**(6): p. 1263-1271.
37. Tsukada, K., K. Takayanagi, H. Asamura, M. Ota and H. Fukushima, *Multiplex short tandem repeat typing in degraded samples using newly designed primers for the TH01, TPOX, CSF1PO, and vWA loci.* Legal Medicine, 2002. **4**(4): p. 239-245.
38. Wiegand, P. and M. Kleiber, *Less is more--length reduction of STR amplicons using redesigned primers.* International Journal of Legal Medicine, 2001. **114**(4-5): p. 285-287.
39. Watterson, J., V. Blackman and D. Bagby, *Considerations for the analysis of forensic samples following extended exposure to the environment.* The Forensic Examiner, 2006. p. 19-25.
40. Hellmann, A., U. Rohleder, H. Schmitter and M. Wittig, *STR typing of human telogen hairs--a new approach.* International Journal of Legal Medicine, 2001. **114**(4-5): p. 269-273.
41. Grossman, P.D., W. Bloch, E. Brinson, C.C. Chang, F.A. Eggerding, S. Fung, D.M. Iovannisci, S. Woo, E.S. Winn-Deen and D.A. Iovannisci, *High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay*

and sequence-coded separation. Nucleic Acids Research, 1994. **22**(21): p. 4527-4534.

42. *AmpFℓSTR® MiniFiler™ PCR amplification kit user guide*. 2007, Applied Biosystems: Foster City, California.
43. *The NGM™ Kit: The Perfect Union of Data Quality and Data Sharing*. 2009. [Accessed 10 Jan 2011]; Available from: http://marketing.appliedbiosystems.com/images/Product_Microsites/NGM/downloads/PCRAmplificationKit_SS_v2.pdf.
44. Alonso, A., P. Martin, C. Albarrán, P. Garcia, L. Fernandez De Simon, M. Jesús Iturralde, A. Fernández-Rodriguez, I. Atienza, J. Capilla, J. García-Hirschfeld, P. Martinez, G. Vallejo, O. García, E. García, P. Real, D. Alvarez, A. León and M. Sancho, *Challenges of DNA profiling in mass disaster investigations.* Croatian Medical Journal, 2005. **46**(4): p. 540-548.
45. Biesecker, L., J. Bailey-Wilson, J. Ballantyne, H. Baum, F. Bieber, C. Brenner, B. Budowle, J. Butler, G. Carmody, P. Conneally, B. Duceman, A. Eisenberg, L. Forman, K. Kidd, B. Leclair, S. Niezgoda, T. Parsons, E. Pugh, R. Shaler, S. Sherry, A. Sozer and A. Walsh, *Epidemiology. DNA identifications after the 9/11 World Trade Center attack.* Science, 2005. **310**(5751): p. 1122-1123.
46. Fondevila, M., C. Phillips, N. Naveran, L. Fernandez, M. Cerezo, A. Salas, A. Carracedo and M.V. Lareu, *Case report: Identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur.* Forensic Science International: Genetics, 2008. **2**(3): p. 212-218.
47. Opel, K.L., D.T. Chung, J. Drábek, N.E. Tatarek, L.M. Jantz and B.R. Mccord, *The application of miniplex primer sets in the analysis of degraded DNA from human skeletal remains.* Journal of Forensic Sciences, 2006. **51**(2): p. 351-356.
48. Brown, T.A., *Genomes*. 2002, Wiley-Liss: Oxford, UK.
49. Read, C.M., J.P. Baldwin and C. Crane-Robinson, *Structure of subnucleosomal particles. Tetrameric (H3/H4)2 146 base pair DNA and*

hexameric (H3/H4)2(H2A/H2B)1 146 base pair DNA complexes. Biochemistry, 1985. **24**(16): p. 4435-4450.

50. Richmond, T.J. and C.A. Davey, *The structure of DNA in the nucleosome core.* Nature, 2003. **423**(6936): p. 145-150.
51. Ercan, S. and J. Lieb, *New evidence that DNA encodes its packaging.* Nature Genetics, 2006. **38**(10): p. 1104-1105.
52. Hartl, D.L. and E.W. Jones, *Essential genetics : a genomics perspective*. 2006, Jones and Bartlett Publishers: Boston, USA.
53. Schones, D.E., K. Cui, S. Cuddapah, T.-Y. Roh, A. Barski, Z. Wang, G. Wei and K. Zhao, *Dynamic regulation of nucleosome positioning in the human genome.* Cell, 2008. **132**(5): p. 887-898.
54. Lusser, A. and J.T. Kadonaga, *Chromatin remodeling by ATP-dependent molecular machines.* Bioessays, 2003. **25**(12): p. 1192-1200.
55. Corona, D.F.V., C.R. Clapier, P.B. Becker and J.W. Tamkun, *Modulation of ISWI function by site-specific histone acetylation.* EMBO Reports, 2002. **3**(3): p. 242-247.
56. Ioshikhes, I., I. Albert, S. Zanton and B. Pugh, *Nucleosome positions predicted through comparative genomics.* Nature Genetics, 2006. **38**(10): p. 1210-1215.
57. Ioshikhes, I., A. Bolshoy, K. Derenshteyn, M. Borodovsky and E.N. Trifonov, *Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences.* Journal of Molecular Biology, 1996. **262**(2): p. 129-139.
58. Kiyama, R. and E. Trifonov, *What positions nucleosomes?-A model.* FEBS Letters, 2002. **523**(1-3): p. 7-11.
59. Kogan, S.B., M. Kato, R. Kiyama and E.N. Trifonov, *Sequence structure of human nucleosome DNA.* Journal of Biomolecular Structure and Dynamics, 2006. **24**(1): p. 43-48.
60. Lowary, P.T. and J. Widom, *New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning.* Journal of Molecular Biology, 1998. **276**(1): p. 19-42.

61. Peckham, H.E., R.E. Thurman, Y. Fu, J.A. Stamatoyannopoulos, W.S. Noble, K. Struhl and Z. Weng, *Nucleosome positioning signals in genomic DNA.* Genome Research, 2007. **17**(8): p. 1170-1177.
62. Rando, O.J. and K. Ahmad, *Rules and regulation in the primary structure of chromatin.* Current Opinion in Cell Biology, 2007. **19**(3): p. 250-256.
63. Segal, E., Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I.K. Moore, J.-P.Z. Wang and J. Widom, *A genomic code for nucleosome positioning.* Nature, 2006. **442**(7104): p. 772-778.
64. Kornberg, R.D. and Y. Lorch, *Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome.* Cell, 1999. **98**(3): p. 285-294.
65. Segal, E. and J. Widom, *What controls nucleosome positions?* Trends in Genetics, 2009. **25**(8): p. 335-343.
66. Yuan, G.-C., Y.-J. Liu, M.F. Dion, M.D. Slack, L.F. Wu, S.J. Altschuler and O.J. Rando, *Genome-scale identification of nucleosome positions in S. cerevisiae.* Science, 2005. **309**(5734): p. 626-630.
67. Salih, F., B. Salih and E.N. Trifonov, *Sequence structure of hidden 10.4-base repeat in the nucleosomes of C. elegans.* Journal of Biomolecular Structure and Dynamics, 2008. **26**(3): p. 273-282.
68. Bettecken, T. and E. Trifonov, *Repertoires of the nucleosome-positioning dinucleotides.* PLoS ONE, 2009. **4**(11): p. e7654.
69. Trifonov, E., *Curved DNA*. CRC Critical Reviews in Biochemistry, 1985. **19**(2): p. 89-106.
70. Gabdank, I., D. Barash and E. Trifonov, *Nucleosome DNA bendability matrix (C. elegans).* Journal of Biomolecular Structure and Dynamics, 2009. **26**(4): p. 403-411.
71. Trifonov, E., *Nucleosome positioning by sequence, state of the art and apparent finale.* Journal of Biomolecular Structure and Dynamics, 2010. **27**(6): p. 741-746.
72. Zhang, Y., Z. Moqtaderi, B.P. Rattner, G. Euskirchen, M. Snyder, J.T. Kadonaga, X.S. Liu and K. Struhl, *Intrinsic histone-DNA interactions are not*

the major determinant of nucleosome positions in vivo. Nature Structural and Molecular Biology, 2009. **16**(8): p. 847-852.

73. Whitehouse, I. and T. Tsukiyama, *Antagonistic forces that position nucleosomes in vivo.* Nature Structural and Molecular Biology, 2006. **13**(7): p. 633-640.
74. Manohar, M., A.M. Mooney, J.A. North, R.J. Nakkula, J.W. Picking, A. Edon, R. Fishel, M.G. Poirier and J.J. Ottesen, *Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding.* Journal of Biological Chemistry, 2009. **284**(35): p. 23312-23321.
75. Arya, G. and T. Schlick, *A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments.* The Journal of Physical Chemistry A, 2009. **113**(16): p. 4045-4059.
76. Narlikar, G.J., H.-Y. Fan and R.E. Kingston, *Cooperation between complexes that regulate chromatin structure and transcription.* Cell, 2002. **108**(4): p. 475-487.
77. Nathan, D. and D.M. Crothers, *Bending and flexibility of methylated and unmethylated EcoRI DNA.* Journal of Molecular Biology, 2002. **316**(1): p. 7-17.
78. Mavrich, T.N., I.P. Ioshikhes, B.J. Venters, C. Jiang, L.P. Tomsho, J. Qi, S.C. Schuster, I. Albert and B.F. Pugh, *A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome.* Genome Research, 2008. **18**(7): p. 1073-1083.
79. Koerber, R.T., H.S. Rhee, C. Jiang and B.F. Pugh, *Interaction of transcriptional regulators with specific nucleosomes across the Saccharomyces genome.* Molecular Cell, 2009. **35**(6): p. 889-902.
80. Arya, G., A. Maitra and S. Grigoryev, *A structural perspective on the where, how, why, and what of nucleosome positioning.* Journal of Biomolecular Structure and Dynamics, 2010. **27**(6): p. 803-820.
81. Mirny, L.A., *Nucleosome-mediated cooperativity between transcription factors.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(52): p. 22534-22539.

82. Cui, F. and V. Zhurkin, *Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro.* Journal of Biomolecular Structure and Dynamics, 2010. **27**(6): p. 821-841.

83. Hall, M.A., A. Shundrovsky, L. Bai, R.M. Fulbright, J.T. Lis and M.D. Wang, *High-resolution dynamic mapping of histone-DNA interactions in a nucleosome.* Nature Structural and Molecular Biology, 2009. **16**(2): p. 124-129.

84. Yuan, G.-C. and J.S. Liu, *Genomic sequence is highly predictive of local nucleosome depletion.* PLoS Computational Biology, 2008. **4**(1): p. e13.

85. Valouev, A., J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, H. Peckham, K. Zeng, J.A. Malek, G. Costa, K. Mckernan, A. Sidow, A. Fire and S.M. Johnson, *A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning.* Genome Research, 2008. **18**(7): p. 1051-1063.

86. Lee, W., D. Tillo, N. Bray, R.H. Morse, R.W. Davis, T.R. Hughes and C. Nislow, *A high-resolution atlas of nucleosome occupancy in yeast.* Nature Genetics, 2007. **39**(10): p. 1235-1244.

87. Zhang, Y., Z. Moqtaderi, B.P. Rattner, G. Euskirchen, M. Snyder, J.T. Kadonaga, X.S. Liu and K. Struhl, *Evidence against a genomic code for nucleosome positioning.* Nature Structural and Molecular Biology, 2010. **17**(8): p. 920-923.

88. Kaplan, N., I. Moore, Y. Fondufe-Mittendorf, A.J. Gossett, D. Tillo, Y. Field, T.R. Hughes, J.D. Lieb, J. Widom and E. Segal, *Nucleosome sequence preferences influence in vivo nucleosome organization.* Nature Structural and Molecular Biology, 2010. **17**(8): p. 918-920.

89. Kaplan, N., I.K. Moore, Y. Fondufe-Mittendorf, A.J. Gossett, D. Tillo, Y. Field, E.M. Leproust, T.R. Hughes, J.D. Lieb, J. Widom and E. Segal, *The DNA-encoded nucleosome organization of a eukaryotic genome.* Nature, 2009. **458**(7236): p. 362-366.

90. Pugh, B.F., *A preoccupied position on nucleosomes.* Nature Structural and Molecular Biology, 2010. **17**(8): p. 923.

91. Vinogradov, A.E., *Noncoding DNA, isochores and gene expression: nucleosome formation potential.* Nucleic Acids Research, 2005. **33**(2): p. 559-563.
92. Alaeddini, R., S. Walsh and A. Abbas, *Forensic implications of genetic analyses from degraded DNA--a review.* Forensic Science International: Genetics, 2010. **4**(3): p. 148-157.
93. Luykx, P., I.V. Bajić and S. Khuri, *NXSensor web tool for evaluating DNA for nucleosome exclusion sequences and accessibility to binding factors.* Nucleic Acids Research, 2006. **34**(Web Server issue): p. W560-565.
94. Tolstorukov, M.Y., V. Choudhary, W.K. Olson, V.B. Zhurkin and P.J. Park, *nuScore: a web-interface for nucleosome positioning predictions.* Bioinformatics, 2008. **24**(12): p. 1456-1458.
95. Gabdank, I., D. Barash and E. Trifonov, *FineStr : a web server for single-base-resolution nucleosome positioning.* Bioinformatics, 2010. **26**(6): p. 845-846.
96. Fiers, W., R. Beyaert, W. Declercq and P. Vandenabeele, *More than one way to die: apoptosis, necrosis and reactive oxygen damage.* Oncogene, 1999. **18**(54): p. 7719-7730.
97. Edinger, A. and C. Thompson, *Death by design: apoptosis, necrosis and autophagy.* Current Opinion in Cell Biology, 2004. **16**(6): p. 663-669.
98. Lamkanfi, M., W. Declercq, M. Kalai, X. Saelens and P. Vandenabeele, *Alice in caspase land. A phylogenetic analysis of caspases from worm to man.* Cell Death and Differentiation, 2002. **9**(4): p. 358-361.
99. Nagata, S., *DNA degradation in development and programmed cell death.* Annual Review of Immunology, 2005. **23**: p. 853-875.
100. Williamson, R., *Properties of rapidly labelled deoxyribonucleic acid fragments isolated from the cytoplasm of primary cultures of embryonic mouse liver cells.* Journal of Molecular Biology, 1970. **51**(1): p. 157-168.
101. Bender, K., M. Farfán and P. Schneider, *Preparation of degraded human DNA under controlled conditions.* Forensic Science International, 2004. **139**(2-3): p. 135-140.

102. Eguchi, Y., S. Shimizu and Y. Tsujimoto, *Intracellular ATP levels determine cell death fate by apoptosis or necrosis.* Cancer Research, 1997. **57**(10): p. 1835-1840.

103. Hofreiter, M., D. Serre, H.N. Poinar, M. Kuch and S. Pääbo, *Ancient DNA.* Nature Reviews Genetics, 2001. **2**(5): p. 353-359.

104. Antheunisse, J., *Decomposition of nucleic acids and some of their degradation products by microorganisms.* Antonie van Leeuwenhoek, 1972. **38**(3): p. 311-327.

105. Corry, J., *A review. Possible sources of ethanol ante- and post-mortem: its relationship to the biochemistry and microbiology of decomposition.* Journal of Applied Bacteriology, 1978. **44**(1): p. 1-56.

106. Poinar, H.N., *The top 10 list: criteria of authenticity for DNA from ancient and forensic samples.* International Congress Series, 2003. **1239**: p. 575-579.

107. Lindahl, T., *Instability and decay of the primary structure of DNA.* Nature, 1993. **362**(6422): p. 709-715.

108. Marrone, A. and J. Ballantyne, *Hydrolysis of DNA and its molecular components in the dry state.* Forensic Science International: Genetics, 2010. **4**(3): p. 168-177.

109. Dixon, L.A., *An investigation into the use of single nucleotide polymorphisms for forensic identification purposes (Thesis)* in School of Biosciences, 2006, University of Wales: Cardiff.

110. Hall, A. and J. Ballantyne, *Characterization of UVC-induced DNA damage in bloodstains: forensic implications.* Analytical and Bioanalytical Chemistry, 2004. **380**(1): p. 72-83.

111. Goffin, C. and W.G. Verly, *Interstrand DNA crosslinks due to AP (apurinic/apyrimidinic) sites.* FEBS Letters, 1983. **161**(1): p. 140-144.

112. Johnson, L.A. and J.A. Ferris, *Analysis of postmortem DNA degradation by single-cell gel electrophoresis.* Forensic Science International, 2002. **126**(1): p. 43-47.

113. Heins, J., J. Suriano, H. Taniuchi and C. Anfinsen, *Characterization of a nuclease produced by Staphylococcus aureus.* Journal of Biological Chemistry, 1967. **242**(5): p. 1016.

114. Telford, D. and B. Stewart, *Micrococcal nuclease: its specificity and use for chromatin analysis.* International Journal of Biochemistry and Cell Biology, 1989. **21**(2): p. 127-137.

115. Clark, D., *Nucleosome positioning, nucleosome spacing and the nucleosome code.* Journal of Biomolecular Structure and Dynamics, 2010. **27**(6): p. 781-793.

116. Hörz, W. and W. Altenburger, *Sequence specific cleavage of DNA by micrococcal nuclease.* Nucleic Acids Research, 1981. **9**(12): p. 2643-2658.

117. Vanecko, S. and M. Laskowski, *Studies of the specificity of deoxyribonuclease I. III. Hydrolysis of chains carrying a monoesterified phosphate on carbon 5'.* Journal of Biological Chemistry, 1961. **236**: p. 3312-3316.

118. Douvas, A. and P. Price, *Some effects of calcium and magnesium ions on the activity of bovine pancreatic deoxyribonuclease A.* Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis, 1975. **395**(2): p. 201-212.

119. Junowicz, E. and J. Spencer, *Studies on bovine pancreatic deoxyribonuclease A: I. General properties and activation with different bivalent metals.* Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis, 1973. **312**(1): p. 72-84.

120. Hudlow, W.R., M.D. Chong, K.L. Swango, M.D. Timken and M.R. Buoncristiani, *A quadruplex real-time qPCR assay for the simultaneous assessment of total human DNA, human male DNA, DNA degradation and the presence of PCR inhibitors in forensic samples: a diagnostic tool for STR typing.* Forensic Science International: Genetics, 2008. **2**(2): p. 108-125.

121. Krenke, B.E., N. Nassif, C.J. Sprecher, C. Knox, M. Schwandt and D.R. Storts, *Developmental validation of a real-time PCR assay for the simultaneous quantification of total human and male DNA.* Forensic Science International: Genetics, 2008. **3**(1): p. 14-21.

122. Barbisin, M., R. Fang, C.E. O'shea, L.M. Calandro, M.R. Furtado and J.G. Shewale, *Developmental validation of the Quantifiler® Duo DNA quantification kit for simultaneous quantification of total human and human*

male DNA and detection of PCR inhibitors in biological samples. Journal of Forensic Sciences, 2009. **54**(2): p. 305-319.

123. Fleming, R.I. and S. Harbison, *The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids.* Forensic Science International: Genetics, 2010. **4**(4): p. 244-256.
124. Heid, C., J. Stevens, K. Livak and P. Williams, *Real time quantitative PCR.* Genome Research, 1996. **6**(10): p. 986-994.
125. Wittwer, C.T., M.G. Herrmann, A.A. Moss and R.P. Rasmussen, *Continuous fluorescence monitoring of rapid cycle DNA amplification.* Biotechniques, 1997. **22**(1): p. 130-131, 134-138.
126. Vanguilder, H., K. Vrana and W. Freeman, *Twenty-five years of quantitative PCR for gene expression analysis.* Biotechniques, 2008. **44**(5): p. 619-626.
127. *Brilliant® SYBR® Green QPCR Master Mix: Instruction Manual*. 2007, Stratagene: La Jolla, California.
128. Yuryev, A., *PCR Primer Design*. 2007, Humana Press: New Jersey.
129. *Introduction to Quantitative PCR: Methods and Application Guide*. 2004, Stratagene: La Jolla, California, USA.
130. Bustin, S.A., V. Benes, J.A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M.W. Pfaffl, G.L. Shipley, J. Vandesompele and C.T. Wittwer, *The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments.* Clinical Chemistry, 2009. **55**(4): p. 611-622.
131. Higuchi, R., C. Fockler, G. Dollinger and R. Watson, *Kinetic PCR analysis: real-time monitoring of DNA amplification reactions.* Nature Biotechnology, 1993. **11**(9): p. 1026-1030.
132. Wong, M. and J. Medrano, *Real-time PCR for mRNA quantitation.* Biotechniques, 2005. **39**(1): p. 75-85.
133. Nolan, T., R.E. Hands, W. Ogunkolade and S.A. Bustin, *SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations.* Analytical Biochemistry, 2006. **351**(2): p. 308-310.
134. Zipper, H., H. Brunner, J. Bernhagen and F. Vitzthum, *Investigations on DNA intercalation and surface binding by SYBR Green I, its structure*

determination and methodological implications. Nucleic Acids Research, 2004. **32**(12): p. e103.

135. Giglio, S., P.T. Monis and C.P. Saint, *Demonstration of preferential binding of SYBR Green I to specific DNA fragments in real-time multiplex PCR.* Nucleic Acids Research, 2003. **31**(22): p. e136.
136. Freeman, W.M., S.J. Walker and K.E. Vrana, *Quantitative RT-PCR: pitfalls and potential.* Biotechniques, 1999. **26**(1): p. 112-125.
137. Morrison, T., J. Weis and C. Wittwer, *Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification.* Biotechniques, 1998. **24**(6): p. 954-958, 960, 962.
138. Wilhelm, J. and A. Pingoud, *Real-time polymerase chain reaction.* Chembiochem : a European journal of chemical biology, 2003. **4**(11): p. 1120-1128.
139. Förster, T., *Zwischenmolekulare energiewanderung und fluoreszenz.* Annalen der Physik, 1948. **437**(1-2): p. 55-75.
140. Stryer, L. and R. Haugland, *Energy transfer: a spectroscopic ruler.* Proceedings of the National Academy of Sciences of the United States of America, 1967. **58**(2): p. 719-726.
141. Haugland, R.P., *The Handbook: A Guide to Fluorescent Probes and Labeling Technologies*. 2005, Molecular Probes Inc.
142. Tyagi, S. and F.R. Kramer, *Molecular beacons: probes that fluoresce upon hybridization.* Nature Biotechnology, 1996. **14**(3): p. 303-308.
143. Marras, S.a.E., S. Tyagi and F.R. Kramer, *Real-time assays with molecular beacons and other fluorescent nucleic acid hybridization probes.* Clinica Chimica Acta, 2006. **363**(1-2): p. 48-60.
144. Shapiro, S. and M. Wilk, *An analysis of variance test for normality (complete samples).* Biometrika, 1965. **52**(3-4): p. 591.
145. Conover, W.J., M.E. Johnson and M.M. Johnson, *A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data.* Technometrics, 1981. **23**(4): p. 351-361.
146. Hollander, M. and D.A. Wolfe, *Nonparametric statistical methods*. 1999, Wiley-Interscience: New York, USA.

147. Alonso, A., P. Martín, C. Albarrán, P. García, O. García, L.F. De Simón, J. García-Hirschfeld, M. Sancho, C. De La Rúa and J. Fernández-Piqueras, *Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies.* Forensic Science International, 2004. **139**(2-3): p. 141-149.

148. Nicklas, J.A. and E. Buel, *Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples.* Journal of Forensic Sciences, 2003. **48**(5): p. 936-944.

149. Swango, K.L., W.R. Hudlow, M.D. Timken and M.R. Buoncristiani, *Developmental validation of a multiplex qPCR assay for assessing the quantity and quality of nuclear DNA in forensic samples.* Forensic Science International, 2007. **170**(1): p. 35-45.

150. Swango, K.L., M.D. Timken, M.D. Chong and M.R. Buoncristiani, *A quantitative PCR assay for the assessment of DNA degradation in forensic samples.* Forensic Science International, 2006. **158**(1): p. 14-26.

151. Timken, M., K. Swango, C. Orrego and M. Buoncristiani, *A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: implications for quantifying DNA in degraded samples.* Journal of Forensic Sciences, 2005. **50**(5): p. 1044-1060.

152. Tringali, G., A. Barbaro, E. Insirello, P. Cormaci and A. Roccazzello, *Rapid and efficacious real-time quantitative PCR assay for quantitation of human DNA in forensic samples.* Forensic Science International, 2004. **146**: p. S177-S181.

153. Walker, J., G. Kilroy, J. Xing, J. Shewale, S. Sinha and M. Batzer, *Human DNA quantitation using Alu element-based polymerase chain reaction.* Analytical Biochemistry, 2003. **315**(1): p. 122-128.

154. Lasalle, H.E., G. Duncan and B. Mccord, *An analysis of single and multi-copy methods for DNA quantitation by real-time polymerase chain reaction*. Forensic Science International: Genetics, 2010. **In Press** (Corrected Proof). doi: 10.1016/j.fsigen.2010.03.002.

155. Mighell, A.J., A.F. Markham and P.A. Robinson, *Alu sequences.* FEBS Letters, 1997. **417**(1): p. 1-5.

156. Schmid, C., *Alu: structure, origin, evolution, significance and function of one-tenth of human DNA.* Progress in Nucleic Acid Research and Molecular Biology, 1996. **53**: p. 283-319.

157. Butte, A.J., V.J. Dzau and S.B. Glueck, *Further defining housekeeping, or "maintenance," genes Focus on "A compendium of gene expression in normal human tissues".* Physiological Genomics, 2001. **7**(2): p. 95-96.

158. Liu, Y.-J., D. Zheng, S. Balasubramanian, N. Carriero, E. Khurana, R. Robilotto and M.B. Gerstein, *Comprehensive analysis of the pseudogenes of glycolytic enzymes in vertebrates: the anomalously high number of GAPDH pseudogenes highlights a recent burst of retrotrans-positional activity.* BMC Genomics, 2009. **10**(1): p. 480.

159. Eisenberg, E. and E.Y. Levanon, *Human housekeeping genes are compact.* Trends in Genetics, 2003. **19**(7): p. 362-365.

160. Rozen, S. and H. Skaletsky, *Primer3 on the WWW for general users and for biologist programmers* in *Bioinformatics: Methods and Protocols (Methods in Molecular Biology).* 2000, Humana Press*:* New Jersey. p. 365-386.

161. Nakamura, T., G. Morin, K. Chapman, S. Weinrich, W. Andrews, J. Lingner, C. Harley and T. Cech, *Telomerase catalytic subunit homologs from fission yeast and human.* Science, 1997. **277**(5328): p. 955.

162. Zavlaris, M., K. Angelopoulou, I. Vlemmas and N. Papaioannou, *Telomerase reverse transcriptase (TERT) expression in canine mammary tissues: a specific marker for malignancy?* Anticancer Research, 2009. **29**(1): p. 319-325.

163. Greider, C., *Telomere length regulation.* Annual Review of Biochemistry, 1996. **65**: p. 337-365.

164. Allsopp, R.C., H. Vaziri, C. Patterson, S. Goldstein, E.V. Younglai, A.B. Futcher, C.W. Greider and C.B. Harley, *Telomere length predicts replicative capacity of human fibroblasts.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(21): p. 10114-10118.

165. Kim, N.W., M.A. Piatyszek, K.R. Prowse, C.B. Harley, M.D. West, P.L. Ho, G.M. Coviello, W.E. Wright, S.L. Weinrich and J.W. Shay, *Specific*

association of human telomerase activity with immortal cells and cancer. Science, 1994. **266**(5193): p. 2011-2015.

166. Lorentzen, E., B. Siebers, R. Hensel and E. Pohl, *Structure, function and evolution of the Archaeal class I fructose-1,6-bisphosphate aldolase.* Biochemical Society Transactions, 2004. **32**(Pt 2): p. 259-263.
167. Taguchi, K. and Y. Takagi, *[Aldolase]*. Rinsho Byori. The Japanese Journal of Clinical Pathology, 2001. **Supplement 116**: p. 117-124.
168. Baer, M., T.W. Nilsen, C. Costigan and S. Altman, *Structure and transcription of a human gene for H1 RNA, the RNA component of human RNase P.* Nucleic Acids Research, 1990. **18**(1): p. 97-103.
169. Altman, S., L. Kirsebom and S. Talbot, *Recent studies of ribonuclease P.* FASEB Journal, 1993. **7**(1): p. 7-14.
170. Sirover, M.A., *New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase.* Biochimica et Biophysica Acta, 1999. **1432**(2): p. 159-184.
171. Sirover, M.A., *New nuclear functions of the glycolytic protein, glyceraldehyde-3-phosphate dehydrogenase, in mammalian cells.* Journal of Cellular Biochemistry, 2005. **95**(1): p. 45-52.
172. Krynetski, E.Y., N.F. Krynetskaia, A.E. Gallo, K.G. Murti and W.E. Evans, *A novel protein complex distinct from mismatch repair binds thioguanylated DNA.* Molecular Pharmacology, 2001. **59**(2): p. 367-374.
173. Sundararaj, K., R. Wood, S. Ponnusamy, A. Salas, Z. Szulc, A. Bielawska, L. Obeid, Y. Hannun and B. Ogretmen, *Rapid shortening of telomere length in response to ceramide involves the inhibition of telomere binding activity of nuclear glyceraldehyde-3-phosphate dehydrogenase.* Journal of Biological Chemistry, 2004. **279**(7): p. 6152-6162.
174. Nakagawa, T., Y. Hirano, A. Inomata, S. Yokota, K. Miyachi, M. Kaneda, M. Umeda, K. Furukawa, S. Omata and T. Horigome, *Participation of a fusogenic protein, glyceraldehyde-3-phosphate dehydrogenase, in nuclear membrane assembly.* Journal of Biological Chemistry, 2003. **278**(22): p. 20395-20404.

175. Bailey, A.D., T. Pavelitz and A.M. Weiner, *The microsatellite sequence (CT)n x (GA)n promotes stable chromosomal integration of large tandem arrays of functional human U2 small nuclear RNA genes.* Molecular and Cellular Biology, 1998. **18**(4): p. 2262-2271.

176. Hammarstrom, K., G. Westin, C. Bark, J. Zabielski and U. Petterson, *Genes and pseudogenes for human U2 RNA. Implications for the mechanism of pseudogene formation.* Journal of Molecular Biology, 1984. **179**(2): p. 157-169.

177. Greenspoon, S.A., R.C. Bohr, A. Pittock, J. Grubb, K.M. Horsman and J.D. Ban, *Predicting the success of STR typing using the Plexor® HY System and extraction of mock cases using the Differex™ System* in 18th International Symposium on Human Identification, Promega Corporation: Hollywood, CA, USA.

178. Mcginnis, S. and T. Madden, *BLAST: at the core of a powerful and diverse set of sequence analysis tools.* Nucleic Acids Research, 2004. **32**(Web Server issue): p. W20-25.

179. *Manual: Brilliant II SYBR® Green QPCR Master Mix with ROX*. 2009, Stratagene.

180. Li, H., X. Cui and N. Arnheim, *Direct electrophoretic detection of the allelic state of single DNA molecules in human sperm by using the polymerase chain reaction.* Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(12): p. 4580-4584.

181. Rychlik, W., B. White and B. White, *Selection of Primers for Polymerase Chain Reaction* in *PCR Protocols: Current Methods and Applications.* 1993, p. 31-40.

182. Invitrogen, *Oligo Purity Selection Guide*. [Accessed 4 January 2011]; Available from: http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Nucleic-Acid-Amplification-and-Expression-Profiling/Oligonucleotide-Design/Oligo-Ordering-Details/Oligo-Purity-Selection-Guide.html.

183. Fox, J.C., C.A. Cave and J.W. Schumm, *Development, characterization, and validation of a sensitive primate-specific quantification assay for forensic analysis.* Biotechniques, 2003. **34**(2): p. 314-318, 320, 322.

184. Kline, M.C., D.L. Duewer, J.W. Redman and J.M. Butler, *Results from the NIST 2004 DNA Quantitation Study.* Journal of Forensic Sciences, 2005. **50**(3): p. 570-578.

185. Kline, M.C., D.L. Duewer, J.C. Travis, M.V. Smith, J.W. Redman, P.M. Vallone, A.E. Decker and Butler, *Production and certification of NIST Standard Reference Material 2372 Human DNA Quantitation Standard.* Analytical and Bioanalytical Chemistry, 2009. **394**(4): p. 1183-1192.

186. Schneider, P., *Expansion of the European Standard Set of DNA Database Loci—the Current Situation.* Profiles in DNA, 2009. **12**(1): p. 6-7.

187. Gill, P., L. Fereday, N. Morling and P.M. Schneider, *New multiplexes for Europe-amendments and clarification of strategic development.* Forensic Science International, 2006. **163**(1-2): p. 155-157.

188. Hill, C., J. Butler and P. Vallone, *A 26plex autosomal STR assay to aid human identity testing.* Journal of Forensic Sciences, 2009. **54**(5): p. 1008-1015.

189. Asamura, H., S. Fujimori, M. Ota and H. Fukushima, *MiniSTR multiplex systems based on non-CODIS loci for analysis of degraded DNA samples.* Forensic Science International, 2007. **173**(1): p. 7-15.

190. Hill, C., M. Kline, M. Coble and J. Butler, *Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples.* Journal of Forensic Sciences, 2008. **53**(1): p. 73-80.

191. *Technical Manual: PowerPlex® S5 System*. 2008, Promega Corporation: Wisconsin, USA.

192. Andrade, L., A. Bento, A. Serra, M. Carvalho, J. Gamero, C. Oliveira, L. Batista, V. Lopes, F. Balsa and F. Corte-Real, *AmpFℓSTR® MiniFiler™ PCR amplification kit: The new miniSTR multiplex kit.* Forensic Science International: Genetics Supplement Series, 2008. **1**(1): p. 89-91.

193. Tucker, V.C., A.J. Hopwood, C.J. Sprecher, R.S. Mclaren, D.R. Rabbach, M.G. Ensenberger, J.M. Thompson and D.R. Storts, *Developmental*

validation of the PowerPlex® ESI 16 and PowerPlex® ESI 17 Systems: STR multiplexes for the new European standard. Forensic Science International: Genetics, 2010. **In Press** (Corrected Proof). doi: 10.1016/j.fsigen.2010.09.004.

194. Mulero, J.J., C.W. Chang, R.E. Lagacé, D.Y. Wang, J.L. Bas, T.P. Mcmahon and L.K. Hennessy, *Development and validation of the AmpFℓSTR® MiniFiler™ PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA.* Journal of Forensic Sciences, 2008. **53**(4): p. 838-852.
195. Hill, C., M. Kline, J. Mulero, R. Lagace, C. Chang, L. Hennessy and J. Butler, *Concordance study between the AmpFℓSTR® MiniFiler™ PCR amplification kit and conventional STR typing kits.* Journal of Forensic Sciences, 2007. **52**(4): p. 870-873.
196. *Revised Validation Guidelines.* Forensic Science Communications, 2004. **6**(3).
197. Gehrig, C. and A. Teyssier, *Internal validation of the AmpFℓSTR® MiniFiler™ kit.* Forensic Science International: Genetics Supplement Series, 2008. **1**(1): p. 115-117.
198. Gill, P., J. Whitaker, C. Flaxman, N. Brown and J. Buckleton, *An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA.* Forensic Science International, 2000. **112**(1): p. 17-40.
199. Gill, P., *Application of low copy number DNA profiling.* Croatian Medical Journal, 2001. **42**(3): p. 229-232.
200. Ballantyne, K., R. Van Oorschot and R. Mitchell, *Comparison of two whole genome amplification methods for STR genotyping of LCN and degraded DNA samples.* Forensic Science International, 2007. **166**(1): p. 35-41.
201. Smith, P.J. and J. Ballantyne, *Simplified low-copy-number DNA analysis by post-PCR purification.* Journal of Forensic Sciences, 2007. **52**(4): p. 820-829.
202. Forster, L., J. Thomson and S. Kutranov, *Direct comparison of post-28-cycle PCR purification and modified capillary electrophoresis methods with the 34-cycle "low copy number" (LCN) method for analysis of trace forensic*

DNA samples. Forensic Science International: Genetics, 2008. **2**(4): p. 318-328.

203. Gill, P., R. Brown, M. Fairley, L. Lee, M. Smyth, N. Simpson, B. Irwin, J. Dunlop, M. Greenhalgh, K. Way, E. Westacott, S. Ferguson, L. Ford, T. Clayton and J. Guiness, *National recommendations of the Technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes.* Forensic Science International: Genetics, 2008. **2**(1): p. 76-82.
204. *DNA test halted after Omagh case*. 2007. [Accessed 22 Dec 2010]; Available from: http://news.bbc.co.uk/1/hi/uk/7156051.stm.
205. Caddy, B., G.R. Taylor and A.M. Linacre, *A review of the science of low template DNA analysis*. 2008. [Accessed 12 January 2011]; Available from: http://www.bioforensics.com/conference10/Workshop/Caddy_Report.pdf.
206. Petricevic, S., J. Whitaker, J. Buckleton, S. Vintiner, J. Patel, P. Simon, H. Ferraby, W. Hermiz and A. Russell, *Validation and development of interpretation guidelines for low copy number (LCN) DNA profiling in New Zealand using the AmpFlSTR SGM Plus multiplex.* Forensic Science International: Genetics, 2010. **4**(5): p. 305-310.
207. Buckleton, J., *Validation issues around DNA typing of low level DNA.* Forensic Science International: Genetics, 2009. **3**(4): p. 255-260.
208. Balding, D. and J. Buckleton, *Interpreting low template DNA profiles.* Forensic Science International: Genetics, 2009. **4**(1): p. 1-10.
209. Tvedebrink, T., P. Eriksen, H. Mogensen and N. Morling, *Estimating the probability of allelic drop-out of STR alleles in forensic genetics.* Forensic Science International: Genetics, 2009. **3**(4): p. 222-226.
210. Gill, P. and J. Buckleton, *A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number.* Forensic Science International: Genetics, 2010. **4**(4): p. 221-227.
211. Budowle, B., A. Eisenberg and A. Van Daal, *Low copy number typing has yet to achieve "general acceptance".* Forensic Science International: Genetics Supplement Series, 2009. **2**(1): p. 551-552.
212. Buckleton, J. and P. Gill, *Further Comment on "Low copy number typing has yet to achieve "general acceptance"" by Budowle, B., et al, 2009. Forensic*

Sci. Int. Genetics: Supplement Series 2, 551-552. Forensic Science International: Genetics, 2011. **5**(1): p. 7-11.

213. Budowle, B., A.J. Eisenberg and A. Van Daal, *Response to Comment on "Low copy number typing has yet to achieve "general acceptance"" (Budowle et al., 2009. Forensic Sci. Int. Genetics: Supplement Series 2, 551-552) by Theresa Caragine, Mechthild Prinz.* Forensic Science International: Genetics, 2011. **5**(1): p. 5-7.
214. Budowle, B. and A. Van Daal, *Comment on "A universal strategy to interpret DNA profiles that does not require a definition of low copy number" by Peter Gill and John Buckleton, 2010, Forensic Sci. Int. Genetics 4, 221-227.* Forensic Science International: Genetics, 2011. **5**(1): p. 15.
215. Budowle, B. and A. Van Daal, *Reply to Comments by Buckleton and Gill on "Low copy number typing has yet to achieve 'general acceptance"' by Budowle, B., et al., 2009. Forensic Sci. Int.: Genet. Suppl. Series 2, 551-552.* Forensic Science International: Genetics, 2011. **5**(1): p. 12-14.
216. Caragine, T. and M. Prinz, *Comment on "Low copy number typing has yet to achieve "general acceptance"" by Budowle, B., et al., 2009. Forensic Sci. Int. Genetics: Supplement Series 2, 551-552.* Forensic Science International: Genetics, 2011. **5**(1): p. 3-4.
217. Rand, S., M. Schürenkamp, C. Hohoff and B. Brinkmann, *The GEDNAP blind trial concept part II. Trends and developments.* International Journal of Legal Medicine, 2004. **118**(2): p. 83-89.
218. Thanakiatkrai, P., *Effects of increased PCR cycles and varying injection times on STR profile quality (Thesis)* in Pure and Applied Chemistry, 2007, University of Strathclyde: Glasgow.
219. Castella, V., N. Dimo-Simonin, C. Brandt-Casadevall and P. Mangin, *Consensus profiles and databasing of casework samples amplified with 34 PCR cycles: an empirical approach.* International Congress Series, 2004. **1261**: p. 532-534.
220. Kloosterman, A. and P. Kersbergen, *Efficacy and limits of genotyping low copy number DNA samples by multiplex PCR of STR loci.* International Congress Series, 2003. **1239**: p. 795-798.

221. Gill, P., C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir and D. Genetics, *DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures.* Forensic Science International, 2006. **160**(2-3): p. 90-101.

222. Gill, P., D. Rowlands, G. Tully, I. Bastisch, T. Staples and P. Scott, *Manufacturer contamination of disposable plastic-ware and other reagents--an agreed position statement by ENFSI, SWGDAM and BSAG.* Forensic Science International: Genetics, 2010. **4**(4): p. 269-270.

223. Butler, J.M., R. Schoske, P.M. Vallone, J.W. Redman and M.C. Kline, *Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations.* Journal of Forensic Sciences, 2003. **48**(4): p. 908-911.

224. *AmpFℓSTR® SGM Plus™ PCR Amplification Kit User's Manual*. 2006, Applied Biosystems: USA.

225. Sprecher, C.J., R.S. Mclaren, D.R. Rabbach, M.G. Ensenberger, P.M. Fulmer, B.E. Krenke, C. Corona, M.A. Scurria, T.D. Good, S.J. Dwight, M.G. Mcdougall and D.R. Storts, *The PowerPlex® ESX and ESI Systems: Meeting the new European standards.* Profiles in DNA, 2009. **12**(2).

226. Bland, J.M. and D.G. Altman, *The use of transformation when comparing two means.* BMJ (Clinical Research Ed.), 1996. **312**(7039): p. 1153.

227. Anna-Maria, D., E. Hanna and A. Marie, *Forensic analysis of autosomal STR markers using Pyrosequencing.* Forensic Science International: Genetics, 2010. **4**(2): p. 122-129.

228. Travers, A.A., *The structural basis of DNA flexibility.* Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 2004. **362**(1820): p. 1423-1438.

229. Polach, K.J. and J. Widom, *Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation.* Journal of Molecular Biology, 1995. **254**(2): p. 130-149.

230. Radwan, A., A. Younis, P. Luykx and S. Khuri, *Prediction and analysis of nucleosome exclusion regions in the human genome.* BMC Genomics, 2008. **9**: p. 186.

231. Dickerson, R., *Definitions and nomenclature of nucleic acid structure parameters.* Journal of Biomolecular Structure and Dynamics, 1989. **6**(4): p. 627-634.

232. Dickerson, R.E., *DNA bending: the prevalence of kinkiness and the virtues of normality.* Nucleic Acids Research, 1998. **26**(8): p. 1906-1926.

233. *Technical Manual: PowerPlex® 16 System*. 2008, Promega Corporation: Wisconsin, USA.

234. Chodavarapu, R.K., S. Feng, Y.V. Bernatavichute, P.-Y. Chen, H. Stroud, Y. Yu, J.A. Hetzel, F. Kuo, J. Kim, S.J. Cokus, D. Casero, M. Bernal, P. Huijser, A.T. Clark, U. Krämer, S.S. Merchant, X. Zhang, S.E. Jacobsen and M. Pellegrini, *Relationship between nucleosome positioning and DNA methylation.* Nature, 2010. **466**(7304): p. 388-392.

235. Kotekar, A.S., J.D. Weissman, A. Gegonne, H. Cohen and D.S. Singer, *Histone modifications, but not nucleosomal positioning, correlate with major histocompatibility complex class I promoter activity in different tissues in vivo.* Molecular and Cellular Biology, 2008. **28**(24): p. 7323-7336.

236. Lee, C.-K., Y. Shibata, B. Rao, B.D. Strahl and J.D. Lieb, *Evidence for nucleosome depletion at active regulatory regions genome-wide.* Nature Genetics, 2004. **36**(8): p. 900-905.

237. Spies, N., C.B. Nielsen, R.A. Padgett and C.B. Burge, *Biased chromatin signatures around polyadenylation sites and exons.* Molecular Cell, 2009. **36**(2): p. 245-254.

238. Ercan, S., Y. Lubling, E. Segal and J.D. Lieb, *High nucleosome occupancy is encoded at X-linked gene promoters in C. elegans*. Genome Research, 2010. **In Press** (Corrected Proof). doi: 10.1101/gr.115931.110.

239. Cohanim, A. and T. Haran, *The coexistence of the nucleosome positioning code with the genetic code on eukaryotic genomes.* Nucleic Acids Research, 2009. **37**(19): p. 6466-6476.

240. Nahkuri, S., R. Taft and J. Mattick, *Nucleosomes are preferentially positioned at exons in somatic and sperm cells.* Cell Cycle, 2009. **8**(20): p. 3420-3424.

241. Levitsky, V., *RECON: a program for prediction of nucleosome formation potential.* Nucleic Acids Research, 2004. **32**(Web Server issue): p. W346-349.

242. Taguchi, G., *Introduction to quality engineering : designing quality into products and processes*. 1986, Quality Resources: New York, USA.

243. Brunborg, I.M., T. Moldal and C.M. Jonassen, *Quantitation of porcine circovirus type 2 isolated from serum/plasma and tissue samples of healthy pigs and pigs with postweaning multisystemic wasting syndrome using a TaqMan-based real-time PCR.* Journal of Virological Methods, 2004. **122**(2): p. 171-178.

244. Rao, R.S., C.G. Kumar, R.S. Prakasham and P.J. Hobbs, *The Taguchi methodology as a statistical tool for biotechnological applications: a critical appraisal.* Biotechnology Journal, 2008. **3**(4): p. 510-523.

245. Ross, P.J., *Taguchi Techniques for Quality Engineering: Loss Function, Orthogonal Experiments, Parameter and Tolerance Design*. 1988, McGraw-Hill: New York, USA.

246. Ballantyne, K., R. Van Oorschot and R. Mitchell, *Reduce optimisation time and effort: Taguchi experimental design methods.* Forensic Science International: Genetics Supplement Series, 2008. **1**(1): p. 7-8.

247. Caetano-Anolles, G., *DAF optimization using Taguchi methods and the effect of thermal cycling parameters on DNA amplification.* Biotechniques, 1998. **25**(3): p. 472-480.

248. Cobb, B. and J. Clarkson, *A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods.* Nucleic Acids Research, 1994. **22**(18): p. 3801-3805.

249. Parks, J.M., *On stochastic optimization: Taguchi methods demystified; its limitations and fallacy clarified.* Probabilistic Engineering Mechanics, 2001. **16**(1): p. 87 - 101.

250. Mast, J.D., *A methodological comparison of three strategies for quality improvement.* International Journal of Quality & Reliability Management, 2004. **21**(2): p. 198-213.

251. Travers, A., E. Hiriart, M. Churcher, M. Caserta and E. Di Mauro, *The DNA sequence-dependence of nucleosome positioning in vivo and in vitro.* Journal of Biomolecular Structure and Dynamics, 2010. **27**(6): p. 713-724.

252. Untergasser, A., H. Nijveen, X. Rao, T. Bisseling, R. Geurts and J.a.M. Leunissen, *Primer3Plus, an enhanced web interface to Primer3.* Nucleic Acids Research, 2007. **35**: p. W71-W74.

253. Thanakiatkrai, P. and L. Welch, *Evaluation of nucleosome forming potentials (NFPs) of forensically important STRs*. Forensic Science International: Genetics, 2010. **In Press** (Corrected Proof). doi: 10.1016/j.fsigen.2010.05.002.

254. Rådström, P., R. Knutsson, P. Wolffs, M. Lövenklev and C. Löfström, *Pre-PCR processing: strategies to generate PCR-compatible samples.* Molecular Biotechnology, 2004. **26**(2): p. 133-146.

255. Huggett, J.F., T. Novak, J.A. Garson, C. Green, S.D. Morris-Jones, R.F. Miller and A. Zumla, *Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon.* BMC Research Notes, 2008. **1**: p. 70.

256. Akane, A., K. Matsubara, H. Nakamura, S. Takahashi and K. Kimura, *Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification.* Journal of Forensic Sciences, 1994. **39**(2): p. 362-372.

257. Opel, K., D. Chung and B. Mccord, *A study of PCR inhibition mechanisms using real time PCR.* Journal of Forensic Sciences, 2009. **55**(1): p. 25-33.

258. Kreader, C.A., *Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein.* Applied and Environmental Microbiology, 1996. **62**(3): p. 1102-1106.

259. Paez, M.S., *Analysis of degraded DNA using real-time qPCR (Thesis)* in Pure and Applied Chemistry, 2010, University of Strathclyde: Glasgow.

260. Shivaswamy, S., A. Bhinge, Y. Zhao, S. Jones, M. Hirst and V.R. Iyer, *Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation.* PLoS Biology, 2008. **6**(3): p. e65.

261. Kasper, Y. and C. Lenz, *Stable 8-year storage of DNA purified with the QIAamp DNA Blood Mini Kit*. 2004. Available from: http://www.qiagen.com/literature/qiagennews/weeklyarticle/04_03/e10/default.aspx.

262. Jeffreys, A., V. Wilson and S. Thein, *Individual-specific 'fingerprints' of human DNA.* Nature, 1985. **316**(6023): p. 76-79.

263. Myska, S., *Twenty years of DNA evidence*. 2006. [Accessed 19 Jan 2011]; Available from: http://news.bbc.co.uk/1/hi/uk/6031749.stm.

264. *NPIA: Statistics*. 2010. [Accessed 14 Jan 2011]; Available from: http://www.npia.police.uk/en/13338.htm.

265. *NPIA: Basic facts - FAQ*. 2010. [Accessed 14 Jan 2011]; Available from: http://www.npia.police.uk/en/13340.htm.

266. *FBI -- NDIS Statistics*. 2010. [Accessed 14 Jan 2011]; Available from: http://www.fbi.gov/about-us/lab/codis/ndis-statistics.

267. Gershaw, C.J., A.J. Schweighardt, L.C. Rourke and M.M. Wallace, *Forensic utilization of familial searches in DNA databases.* Forensic Science International: Genetics, 2011. **5**(1): p. 16-20.

268. Bhattacharya, S., *Killer convicted thanks to relative's DNA*. 2004. [Accessed 14 January 2011]; Available from: http://www.newscientist.com/article/dn4908-killer-convicted-thanks-to-relatives-dna.html.

269. Branicki, W., F. Liu, K. Duijn, J. Draus-Barini, E. Pośpiech, S. Walsh, T. Kupiec, A. Wojas-Pelc and M. Kayser, *Model-based prediction of human hair color using DNA variants*. Human Genetics, 2011. **In Press** (Corrected Proof). doi: 10.1007/s00439-010-0939-8.

270. Walsh, S., F. Liu, K.N. Ballantyne, M.V. Oven, O. Lao and M. Kayser, *IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information*. Forensic Science International:

Genetics, 2010. **In Press** (Corrected Proof). doi: 10.1016/j.fsigen.2010.02.004.

271. Walsh, S., A. Lindenbergh, S.B. Zuniga, T. Sijen, P.D. Knijff, M. Kayser and K.N. Ballantyne, *Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence*. Forensic Science International: Genetics, 2010. **In Press** (Corrected Proof). doi: 10.1016/j.fsigen.2010.09.008.

272. *Quality Assurance Standards for Forensic DNA Testing Laboratories*. 2009. [Accessed 12 Jan 2011]; Available from: http://www.fbi.gov/about-us/lab/codis/qas_testlabs.

273. Bogenhagen, D. and D.A. Clayton, *The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid.* Journal of Biological Chemistry, 1974. **249**(24): p. 7991-7995.

274. Chung, D.T., J. Drábek, K.L. Opel, J.M. Butler and B.R. Mccord, *A study on the effects of degradation and template concentration on the amplification efficiency of the STR Miniplex primer sets.* Journal of Forensic Sciences, 2004. **49**(4): p. 733-740.

275. Takahashi, M., Y. Kato, H. Mukoyama, H. Kanaya and S. Kamiyama, *Evaluation of five polymorphic microsatellite markers for typing DNA from decomposed human tissues--correlation between the size of the alleles and that of the template DNA.* Forensic Science International, 1997. **90**(1-2): p. 1-9.

276. Romano, C., E. Di Luise, D. Di Martino, S. Spitaleri and L. Saravo, *A novel approach for genotyping of LCN-DNA recovered from highly degraded samples.* International Congress Series, 2006. **1288**: p. 577-579.

277. Holland, M.M., C.A. Cave, C.A. Holland and T.W. Bille, *Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks.* Croatian Medical Journal, 2003. **44**(3): p. 264-272.

278. Drábek, J., D. Chung, J. Butler and B. Mccord, *Concordance study between Miniplex assays and a commercial STR typing kit.* Journal of Forensic Sciences, 2004. **49**(4): p. 859-860.

279. Alenizi, M.A., W. Goodwin, S. Hadi, H.H. Alenizi, K.A. Altamar and M.S. Alsikel, *Concordance between the AmpFℓSTR® MiniFiler™ and AmpFℓSTR® Identifiler® PCR amplification kits in the Kuwaiti population.* Journal of Forensic Sciences, 2009. **54**(2): p. 350-352.

280. Butler, J.M., *Fundamentals of forensic DNA typing*. 2010, Elsevier: Burlington, MA.

281. Stein, A., T. Takasuka and C. Collings, *Are nucleosome positions in vivo primarily determined by histone-DNA sequence preferences?* Nucleic Acids Research, 2009. **In Press** (Corrected Proof). doi: 10.1093/nar/gkp1043.

282. Chung, H.-R. and M. Vingron, *Sequence-dependent nucleosome positioning.* Journal of Molecular Biology, 2009. **386**(5): p. 1411-1422.

283. Xu, F. and W. Olson, *DNA architecture, deformability, and nucleosome positioning.* Journal of Biomolecular Structure and Dynamics, 2010. **27**(6): p. 725-739.

284. Fernandez, A.G. and J.N. Anderson, *Nucleosome positioning determinants.* Journal of Molecular Biology, 2007. **371**(3): p. 649-668.

285. Trifonov, E., *Base pair stacking in nucleosome DNA and bendability sequence pattern.* Journal of Theoretical Biology, 2010. **263**(3): p. 337-339.

286. Rapoport, A.E., Z.M. Frenkel and E.N. Trifonov, *Nucleosome Positioning Pattern Derived from Oligonucleotide Compositions of Genomic Sequences.* Journal of Biomolecular Structure and Dynamics, 2011. **28**(4): p. 567-574.

287. Nair, V., B. Abraham, J. Mackay, J. Nelder, G. Box, M. Phadke, R. Kacker, J. Sacks, W. Welch, T. Lorenzen, A. Shoemaker, K. Tsui, J. Lucas, S. Taguchi, R. Myers, G. Vining and C. Wu, *Taguchi's parameter design: a panel discussion.* Technometrics, 1992. **34**(2): p. 127-161.

A Appendices

A.1 Primer-BLAST result for *Alu* primer set with 100% match

>NR_037676.1 Homo sapiens chromosome 14 open reading frame 82 (C14orf82), non-coding RNA

```
Product length = 124
Forward primer        GTCAGGAGATCGAGACCATCCC  22
Template        1118  ......................  1139
Reverse primer        TCCTGCCTCAGCCTCCCAAG  20
Template        1241  ......................  1222
```

>NM_001163377.1 Homo sapiens glutaminyl-peptide cyclotransferase-like (QPCTL), transcript variant 2, mRNA

```
Product length = 424
Reverse primer        TCCTGCCTCAGCCTCCCAAG  20
Template        1901  ......................  1882
Reverse primer        TCCTGCCTCAGCCTCCCAAG  20
Template        1478  ......................  1497
```

>NT_167206.1 Homo sapiens chromosome Y genomic contig, GRCh37.p2 reference primary assembly

```
Product length = 573
Reverse primer         TCCTGCCTCAGCCTCCCAAG  20
Template        93732  ......................  93713
Reverse primer         TCCTGCCTCAGCCTCCCAAG  20
Template        93160  ......................  93179
```

>NM_001099677.1 Homo sapiens chromosome 8 open reading frame 79 (C8orf79), transcript variant 2, mRNA

```
Product length = 1275
Reverse primer        TCCTGCCTCAGCCTCCCAAG  20
Template        4949  ......................  4930
Reverse primer        TCCTGCCTCAGCCTCCCAAG  20
Template        3675  ......................  3694
```

A.2 Calibration curve construction of the *Alu* assay

Appendix Table 1. The mean C_q values and standard deviation from duplicates of serially diluted DNA used to construct the calibration curve.

Amount (ng)	Mean	SD
1.000	17.64	0.54
0.500	18.41	0.12
0.250	18.78	0.10
0.125	19.96	0.26
0.063	20.65	0.01
0.031	21.63	0.05
NTC	29.27	0.82

A.3 Genetic profiles for DNA standards used

Appendix Table 2. Genotype of each control DNA source for the S5 kit.

Locus	Amel	D18	D8	TH01	FGA
ABI 007	X,Y	12,15	12,13	7,9.3	24,26
CAMBIO	X,Y	15,16	12,13	9.3,9.3	20,21
9947A	X,X	15,19	13,13	8,9.3	23,24

Appendix Table 3. Genotype of each control DNA source for the MF kit.

Locus	D13	D7	Amel	D2	D21	D16	D18	CSF	FGA
ABI 007	11,11	7,12	X,Y	20,23	28,31	9,10	12,15	11,12	24,26
CAMBIO	12,14	8,10	X,Y	23,25	28,29	11,13	15,16	11,12	20,21
9947A	11,11	10,11	X,X	19,23	30,30	11,12	15,19	10,12	23,24

A.4 Increased cycle study

Appendix Table 4. Mean Hb_x of samples at varying levels (0.016, 0.031 and 0.063 ng) amplified with 30 cycles and 34 cycles with the S5 kit. "+" and "-" indicate if Hb_x of 30 cycles is higher ("+") or lower ("-") than the 34 cycles.

Template	Locus	30 cycles	34 cycles	+/-
0.016 ng	AMEL	0.43	0.26	+
	D18S51	0.71	0.70	+
	D8S1179	0.31	0.45	-
	FGA	0.37	0.82	-
0.031 ng	AMEL	0.64	0.40	+
	D18S51	0.62	0.69	-
	D8S1179	0.86	0.60	+
	FGA	0.77	0.79	-
0.063 ng	AMEL	0.82	0.94	-
	D18S51	0.78	0.79	-
	D8S1179	0.84	0.80	+
	FGA	0.82	0.88	-

Appendix Table 5. Mean Hb_x of samples at varying levels (0.016, 0.031 and 0.063 ng) amplified with 30 cycles and 34 cycles with the MF kit. "+" and "-" indicate if Hb_x of 30 cycles is higher ("+") or lower ("-") than the 34 cycles.

Template	Locus	30 cycles	34 cycles	+/-
0.016 ng	AMEL	0.50	0.34	+
	CSF1PO	0.31	0.57	-
	D13S317	0.26	0.32	-
	D16S539	0.22	0.27	-
	D18S51	0.66	0.45	+
	D21S11	0.59	0.33	+
	D2S1338	0.00	0.29	-
	D7S820	0.33	0.85	-
	FGA	0.19	0.66	-
0.031 ng	AMEL	0.58	0.63	-
	CSF1PO	0.80	0.63	+
	D13S317	0.71	0.58	+
	D16S539	0.59	0.54	+
	D18S51	0.67	0.70	-
	D21S11	0.81	0.53	+
	D2S1338	0.80	0.75	+
	D7S820	0.51	0.42	+
	FGA	0.62	0.59	+
0.063 ng	AMEL	0.84	0.66	+
	CSF1PO	0.63	0.63	+
	D13S317	0.71	0.84	-
	D16S539	0.72	0.62	+
	D18S51	0.76	0.73	+
	D21S11	0.65	0.77	-
	D2S1338	0.80	0.68	+
	D7S820	0.70	0.71	-
	FGA	0.75	0.77	-

Appendix Table 6. Allelic dropin in the S5 and MF kits. 30C and 34C refer to the number of cycles used. Reference is the profile of CAMBIO DNA.

Kit	S5					MF								
Sample	AMEL	D18S51	D8S1179	TH01	FGA	D13S317	D7S820	AMEL	D2S1338	D21S11	D16S539	D18S51	CSF1PO	FGA
30C-16pg	-	-	-	-	20.2	-	-	-	-	-	-	-	-	19.2
30C-16pg	-	9.2	-	-	18.2,19	-	-	-	-	-	-	-	-	-
30C-16pg	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30C-31pg	-	-	-	10	-	-	-	-	-	-	-	-	-	-
30C-31pg	-	-	-	9	20,21	-	-	-	-	-	-	-	-	-
30C-31pg	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30C-63pg	-	9.2,19	-	-	-	-	-	-	-	-	-	-	-	-
30C-63pg	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30C-63pg	-	-	-	-	26,30	-	-	-	-	-	-	-	-	-
34C-16pg	-	-	-	-	28,30	-	-	-	-	-	-	12	-	-
34C-16pg	-	-	-	4	-	7,8	-	-	-	-	-	-	-	-
34C-16pg	-	-	-	-	16,18,22.2	10	-	-	-	-	-	-	-	-
34C-31pg	-	-	-	7,7.3	29,30	-	-	-	-	-	-	-	-	-
34C-31pg	-	-	-	-	-	-	-	-	-	-	-	11,20	-	-
34C-31pg	-	-	-	6	-	-	-	-	-	-	-	-	-	29
34C-63pg	-	-	-	-	-	-	-	-	-	-	-	10	-	24
34C-63pg	-	20	-	-	24,28	-	-	-	-	-	-	7,18	-	-
34C-63pg	-	20.2	-	-	-	-	-	-	-	35	-	12.2	-	26
Reference	X,Y	15,16	12,13	9.3,9.3	21	12,14	8,10	X,Y	23,25	28,29	11,13	15,16	11,12	20,21

A.5 C_q values for the Taguchi method

Appendix Table 7. Quantification cycles of duplicate runs (C_q1 and C_q2) for all experiments and their corresponding signal-to-noise ratios (η). The overall mean signal-to-noise ratios are also given.

	CD4			D1S1627			RPPH1		
Exp	C_q1	C_q2	η	C_q1	C_q2	η	C_q1	C_q2	η
1	30.06	29.87	-29.532	31.88	31.67	-30.042	27.55	27.55	-28.802
2	28.68	28.76	-29.164	29.66	29.81	-29.465	27.70	27.63	-28.839
3	28.08	27.83	-28.929	28.93	28.87	-29.218	27.02	26.93	-28.619
4	29.09	28.86	-29.241	30.57	30.20	-29.653	27.46	27.21	-28.734
5	27.77	27.50	-28.829	28.78	28.83	-29.189	26.81	26.67	-28.543
6	29.34	29.06	-29.308	30.50	30.66	-29.709	27.09	27.18	-28.671
7	28.19	28.48	-29.047	30.64	30.64	-29.726	26.83	26.78	-28.564
8	29.23	29.44	-29.348	30.60	30.54	-29.706	27.34	27.45	-28.753
9	28.53	28.37	-29.082	28.67	28.85	-29.176	27.13	27.02	-28.651
Mean			-29.164			-29.543			-28.686

A.6 Levels of annealing temperature for STR primers

Appendix Table 8. The melting temperatures (°C) of the forward (Tm-F) and reverse (Tm-R) primer and the annealing temperatures (°C) tried with the Taguchi method (Ann1-3).

Locus	Tm-F	Tm-R	Ann1	Ann2	Ann3
D10S1248	60.8	60.9	56	60	62
D20S1082	60.5	62.6	57	60	62
D9S2157	62.3	60.1	57	60	62
TH01	63.0	64.0	55	60	65
TPOX	62.4	63.9	55	60	65
D1GATA113	60.4	61.5	57	60	62
D8S1179	61.4	60.8	57	60	62
FES	61.5	63.3	55	60	65
D4S2364	61.9	62.5	56	60	62
CD4	63.7	63.5	55	60	65
D18S51	61.7	60.0	55	60	65
D22S1045	60.6	63.4	55	60	65
D5S2500	60.4	65.6	57	60	62
D21S11	64.4	61.8	55	60	65

A.7 Quantification results for KF and YS degraded saliva samples

Appendix Table 9. The normalised concentration (in pg/μL) of all degraded saliva samples, arranged in increasing nuScore (see Table 6.1). The number after the sample name indicated the time (in days) when the sample was collected.

KF	D10S1248	D20S1082	D9S2157	TPOX	D1GATA113	FES	CD4	D4S2364	D22S1045	D5S2500
0	131.42	271.07	294.65	243.16	175.32	326.27	351.38	252.83	142.07	205.23
1	20.23	61.2	45.18	113.73	29.9	39.02	73.35	113.1	31.5	33.39
2	12.05	40.69	13.87	46.07	22.95	32.53	30.89	29.05	16.59	16.14
3	16.3	28.44	25.59	50.75	14.88	23.38	41.25	36.21	18.52	21.04
4	5.4	18.11	23.12	13.34	9.23	13.33	34.55	22.07	12.11	11.9
5	5.94	20.9	41.25	39.03	22.38	5.34	26.34	50.96	17.04	18.53
6	5.87	26.47	21.62	14.95	14.33	14.09	23.76	22.59	11.35	11.28
7	3.04	15.84	20.02	16.62	9.65	8.69	24.98	25.04	8.74	11.25
14	0	9.94	3.81	11.03	5.62	4.29	15.51	6.94	3.88	1.1
28	5.72	4.82	7.25	6.56	7.55	13.95	12.95	41.5	7.29	3.29
Mean	*20.60*	*49.75*	*49.64*	*55.52*	*31.18*	*48.09*	*63.50*	*60.03*	*26.91*	*33.32*
Median	*5.91*	*23.69*	*22.37*	*27.83*	*14.61*	*14.02*	*28.62*	*32.63*	*14.35*	*14.02*
Std. Dev	*39.43*	*79.43*	*87.08*	*73.25*	*51.24*	*98.40*	*102.56*	*73.66*	*41.18*	*61.09*

(continued on next page)

YS	D10S1248	D20S1082	D9S2157	TPOX	D1GATA113	FES	CD4	D4S2364	D22S1045	D5S2500
0	247.84	209.28	179.73	394.46	170.28	207.35	351.6	258.28	190.58	122.23
1	114.41	108.83	87.87	197.12	69.01	81.27	160.6	164.32	91.97	50.39
2	95.91	67.96	63.57	148.67	52.23	45	96.37	135.65	45.91	32.31
3	129.35	148.14	96.52	207.19	74.77	60.23	162.54	182.63	90.69	62.48
4	11.63	19.83	19.19	31.52	12.39	7.1	43.63	44.71	12.87	14.2
5	15.06	8.2	3.97	14.99	5.22	4.13	12.51	10.64	7.73	2.13
6	34.43	58.64	39.8	69.51	34.13	33.28	54.04	57.68	27.39	22.06
7	21.75	7.19	8.28	18.02	3.8	10.44	14.44	7.2	10.85	3.22
14	0	0	0	2.64	2.02	3.72	5.27	3.47	0	1.04
28	0	2.28	0	3.41	2.8	3.37	2.94	4.63	0	2.63
Mean	*67.04*	*63.04*	*49.89*	*108.75*	*42.67*	*45.59*	*90.39*	*86.92*	*47.80*	*31.27*
Median	*28.09*	*39.24*	*29.50*	*50.52*	*23.26*	*21.86*	*48.84*	*51.20*	*20.13*	*18.13*
Std. Dev	*79.95*	*71.82*	*58.24*	*128.07*	*52.96*	*63.05*	*109.87*	*91.55*	*60.88*	*38.53*

Combined	D10S1248	D20S1082	D9S2157	TPOX	D1GATA113	FES	CD4	D4S2364	D22S1045	D5S2500
Mean	43.82	56.39	49.76	82.14	36.92	46.84	76.95	73.48	37.35	32.29
Median	13.56	23.69	22.37	35.28	14.61	14.02	32.72	38.86	14.73	15.17
Std. Dev	65.81	74.02	72.10	105.15	51.06	80.44	104.36	82.04	51.71	49.72

A.8 Quantification result for simulated casework samples

Appendix Table 10. Quantification result (in pg/μL) of Plexor® HY and five selected primer sets for simulated casework samples obtained from GEDNAP (Trial 40 and 41 recoded as 0-X and 1-X). Each set has 3 single-source samples (A-C) and 4 unknown samples (1-4). (F) and (M) indicated differential extractions for female and male DNA. Autosomal target (HY-Auto) and Y target (HY-Y) are both shown. "-" = no Y target detected. Auto/nuPrimer is the ratio of HY-Auto to the average DNA concentration of the five STR primers. Normality is p-value from Shapiro-Wilk test.

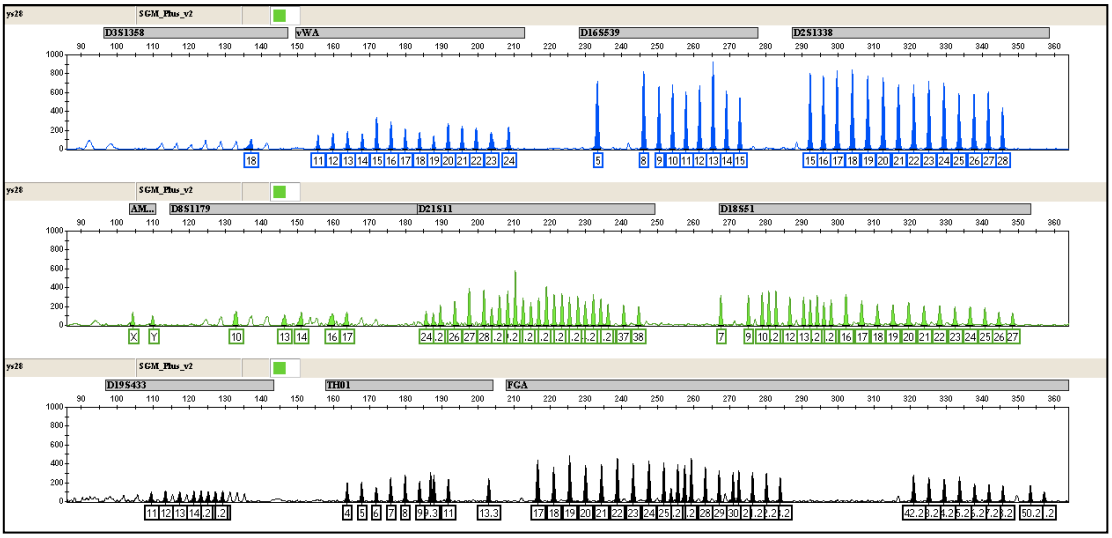
Sample	Description	HY-Auto	HY-Y	D10S1248	D20S1082	FES	CD4	D4S2364	Auto/nuPrimer
0-A	Bloodstain on cellulose swab	1930.00	-	181.48	115.75	1.72	240.24	439.19	9.9
0-B	Bloodstain on cotton swab	446.00	1470.00	49.59	8.24	0	185.19	171.18	5.4
0-C	Bloodstain on cellulose swab	1620.00	1450.00	13.43	69.12	0	110.44	121.45	25.8
0-1	Bloodstain mixture on cellulose swab	2890.00	652.00	150.56	67.89	7.95	398.13	366.82	14.6
0-2F	Semen stain on cellulose swab	76.50	79.40	64.98	25.18	21.7	46.39	40.12	1.9
0-2M	Semen stain on cellulose swab	2.44	1.27	4.19	0.71	2.87	8.92	0.69	0.7
0-3	Bloodstain mixture on cellulose swab	1580.00	261.00	52.82	74.24	0	212.49	273.67	12.9
0-4	Bloodstain on sugar cube	121.00	215.00	45.88	25.17	0.92	32.43	46.35	4.0
1-A	Bloodstain on cotton swab	280.00	561.00	8.7	0.61	0	33.24	80.58	11.4
1-B	Bloodstain on cellulose swab	331.00	1330.00	0	4.34	0	54.86	37.37	17.1
1-C	Bloodstain on cotton swab	4810.00	1480.00	16.9	27.03	0	111.46	82.21	101.2
1-1F	Bloodstain mixture on cellulose swab	464.00	123.00	44.25	93.54	6.67	130.7	116.21	5.9
1-1M	Bloodstain mixture on cellulose swab	10.00	0.71	0	1.6	0	4.52	0	8.2
1-2F	Saliva stain on nylon swab	57.50	NA	8.37	5.87	13.19	12.71	15.44	5.2
1-2M	Saliva stain on nylon swab	1.97	0.31	4.87	0	0	3.63	0	1.2
1-3	Bloodstain on cornflake	607.00	523.00	218.35	258.45	70.7	355.62	499.55	2.2
1-4	Bloodstain on tea bag	188.00	NA	42.07	8.48	1.4	52.25	44.22	6.3
	Mean	*906.79*	*581.91*	*53.32*	*46.25*	*7.48*	*117.25*	*137.36*	*13.8*
	Median	*331.00*	*392.00*	*42.07*	*25.17*	*0.92*	*54.86*	*80.58*	*6.3*
	Std. Dev	*1308.27*	*596.87*	*66.57*	*65.99*	*17.35*	*122.59*	*159.94*	*23.5*
	Normality	*<0.001*	*0.008*	*0.001*	*<0.001*	*<0.001*	*0.008*	*0.002*	

A.9 Contamination by allelic ladder

Due to unexpectedly high DNA concentrations for some samples amplified with loci that overlapped with the SGM+ kit, an investigation into the contamination of all degraded saliva samples was carried out. This investigation was carried out in two stages. First, SGM+ EPGs were used to detect contamination. Second, a real-time qPCR experiment was used to determine whether the allelic ladder contamination had interfered with the amplification of true DNA templates extracted from saliva in the STR primer sets experiment (see Chapter 6).

For SGM+, the reaction components were the same as Section 6.2.1.4. Separation and detection by capillary electrophoresis was carried out using the same protocols as detailed in Section 6.2.1.5. For the real-time qPCR experiment, all 14 primer sets were used to amplify a 1 in 20 dilution of the SGM+ allelic ladder in duplicate. One negative control and one positive control (SRM-A) were included for each primer set. The thermal cycling condition was the same as previous experiments (Section 6.2.2.3).

18 out of 20 samples were contaminated with SGM+ allelic ladder (Appendix Figure 1), and two were not contaminated (YS0 and YS7). Sample KF0 and YS6 (two of 18) were only contaminated with ladder peaks that were below the detection threshold of 50 RFUs. It was surmised that the reason that KF0 and YS0 had low-level contamination and no contamination, respectively, could be the dilution of extract (KF0 diluted 1:5 and YS0 diluted 1:20), which would have diluted down the contaminating ladder at the same time. On the other hand, the negative PCR control had no peaks (data not shown), leading to the conclusion that the contamination was in the sample extracts.



Appendix Figure 1. The EPG of sample YS28 amplified with the SGM+ kit. Note the presence of allelic ladder contamination in all loci.

In the qPCR experiment, all primer sets that overlapped with SGM+ had C_q values of less than 11.90 in wells containing allelic ladder compared to over 25 using SRM-A (Appendix Table 11). These extremely low C_q values indicated a very high amount of template, since 2.1 ng of the positive control had C_q values ranging from 24.58 to 32.70. The loci that did not overlap with SGM+ did not yield a C_q value when the allelic ladder was used as the DNA template.

Appendix Table 11. Quantification cycle (C_q) of positive control (SRM-A) and allelic ladder (Ladder) amplified using the primer sets for each locus. Bold **typeface indicated loci overlapping with SGM+. "-" = No C_q.**

Locus	SRM-A	Ladder	Locus	SRM-A	Ladder
D10S1248	26.15	-	FES	32.70	-
D20S1082	25.06	-	CD4	27.23	-
D9S2157	27.26	-	D4S2364	23.93	-
TH01	**27.15**	**8.91**	**D18S51**	**28.47**	**7.39**
TPOX	26.97	-	D22S1045	26.04	-
D1GATA113	24.58	-	D5S2500	24.81	-
D8S1179	**24.62**	**6.83**	**D21S11**	**25.69**	**11.9**

An SGM+ allelic ladder is a collection of many alleles taken from multiple sources, mixed together in varying amount to obtain balanced peak heights, and re-amplified using the primers used in the SGM+ kit [1]. The STR primer sets used to investigate the protective capability of nucleosomes were designed to be specific to only the intended target sites; hence there were no binding of these primers to the

contaminated allelic ladders in the degraded saliva samples for the loci that did not overlap with SGM+. Since there was no binding, no amplification took place and there was no increase in fluorescence. It was concluded that this contamination did not interfere with the amplification of the intended target of the primers that did not overlap with the loci in the ladder. For the loci that did overlap, the primer sets used in this study bound successfully “inside” the alleles in the allelic ladder, as they bound closer to the repeat motifs than the SGM+ primers. Data from the four overlapping loci were omitted from all calculations and comparisons.