

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: **DNA Typing Using High Performance Liquid Chromatography**

Author(s): **James E. Girard ; Joseph Devaney ; Michael Marino**

Document No.: **196175**

Date Received: **September 05, 2002**

Award Number: **1999-IJ-CX-0033**

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

PROPERTY OF
National Criminal Justice Reference Service (NCJRS)
Box 6000
Rockville, MD 20849-6000

COPY

196175
c/l

DNA Typing using

High Performance Liquid Chromatography

Final report

NII Award No. 1999-IJ-CX-0033

James E. Girard
Joseph Devaney
Michael Marino

— 12/31/01

department of chemistry
american university
4400 massachusetts ave., nw
washington, dc 20016

RECEIVED

2002 MAY 22 P 1:13

FBI/DOJ
OFFICE OF THE
COMPTROLLER
CONTROL DESK

Abstract

Human identification through DNA analysis has undergone tremendous changes since the first criminal conviction in the United States using DNA evidence in the Florida v. Andrews case of 1987.

DNA is commonly used as forensic evidence to link suspects to crimes, exclude falsely accused suspects, reveal serial crimes, distinguish copycat crimes, identify the remains of a victim, and reconstruct accidents. In fact, DNA analysis is used in approximately 10,000 new criminal investigations annually. With the advent and use of the polymerase chain reaction (PCR) as a means to increase the copy number of specific DNA sequences, the use of DNA analysis as evidentiary material will continue to increase. Currently, gel electrophoresis is the DNA analysis method most commonly used. However, gel electrophoresis is a slow technique that typically takes more than two hours to complete, and once the electrophoresis is complete, the results can take days to process.

In this thesis, high performance liquid chromatography (HPLC) is used as a rapid DNA sizing/typing method. The chromatographic conditions for the separation of dsDNA were optimized. Using the optimal conditions for the separation of dsDNA, PCR

Continued
next page

products were sized using restriction enzyme fragments. In addition, a typing method for PCR products from the HUMTHO1 locus was accomplished. Finally, PCR products from different loci were multiplexed and typed using HPLC.

1. INTRODUCTION

Background and Statement of Purpose

In recent years there has been a tremendous increase in the growth of technologies being used to study genetic information. More than 30 genomes have been fully sequenced, and there are many more that will soon be completed. The completion of genomes and increase in genetic information that has been obtained has led to a push for functional genomics, and further developments of new methods for genomic data analysis. The goal of functional genomics is to understand how genes assign specific roles to different cells and to determine how they work together in the body (1). Analysis of gene expression microarrays allows us to study the expression level of thousands of genes simultaneously in a variety of tissues. Bioinformatics has led to better presentation and organization of the data, and software development has progressed so that the information that is obtained from microarrays is easier to read and analyze. With the relatively new technologies being used, it is possible to examine normal and diseased tissue samples, and identify all of the known genes on the arrays that differ in expression between them. Abnormal gene expression can reveal genes that affect particular diseases, therefore presenting new targets for drugs in different biochemical pathways, or areas involving extracellular components and membrane interactions.

In this study, Affymetrix GeneChip® probe arrays have been used to identify gene expression level changes greater than three-fold between experiments. The expression arrays use approximately 16-20 pairs of specific oligonucleotide probes to

seek out each transcript (2). Each probe array contains probes that represent a number of reference and control genes. The human probe array set consists of more than 60,000 genes and ESTs (expressed sequence tags) that include approximately 12,000 full length genes, the majority with known function. The 60,000 genes and ESTs are divided among five different chips (HG-U95A through E). The probe arrays are constructed by synthesizing the oligonucleotide probes onto a glass wafer. The square glass wafer is mounted in a plastic cartridge so that the oligonucleotides lie on the inner glass surface. A chamber is molded into the plastic casing directly under the glass to house buffer for hybridization and washing. The outer glass surface is kept clean to achieve the best results. Each 25 base pair oligonucleotide probe on the glass surface is unique to a particular gene. The 16-20 probe pair sets measure the gene expression level of each of the sequences in the human genome chip set using one perfect sequence match (PM) and a sequence mismatch set (MM) containing one different nucleotide (4).

Samples are processed from total RNA to fragmented cRNA, hybridized onto probe arrays and the expression levels measured. Data mining and analysis are key elements in fully understanding the meaning of the gene expression data for a given disease. Gene signatures or fingerprints are constructed for particular samples sets, and then compared for similarities and differences. This takes all of the known genes or ESTs that are detected or not detected on the chip and assigns which ones are present or absent in each sample set. Based upon the gene expression differentials between two sets, a fold change analysis is performed on the gene signature differential to see if the differences are statistically significant. The genes that are significantly different (fold change >3 and $p<0.05$) can then be viewed on their respective biochemical pathway maps and their functions analyzed.

For this project, I compared the differences in gene expression between two samples sets consisting of ten normal liver tissues and ten malignant liver tissues. The sample sets are matched, meaning that one normal tissue is taken from the same patient as one malignant, only in a location safely outside the area of the tumor. During the surgical procedure, normal and malignant sections of tissue are identified and removed. Expression levels among the normal and diseased cells may be more easily distinguished using the matched sets. The diseased tissues consist of hepatocellular carcinoma, metastatic adenocarcinoma, metastatic carcinoma, or a carcinoid tumor. The gene expression data was examined in

the citric acid (or TCA) cycle (consisting of 36 known genes), fatty acid metabolism (97 genes), glycolysis/gluconeogenesis (86 genes), and oxidative phosphorylation (243 genes), pathways associated with energy production. The purpose of this experiment was to determine which known genes are affected by the diseased tissues in the four pathways mentioned above. By looking at their placement in the four metabolic pathways selected, the genes discovered are examined to see how they are involved in the production of ATP (adenosine tri-phosphate). Liver tissue samples were chosen because of the liver's metabolic role in providing energy for the brain, muscle, and other peripheral organs. Also, because most compounds that absorb through the intestine pass through the liver, the liver is able to regulate many metabolite levels in the blood, an extremely important function.

Materials and Methods

Acquisition of Samples

The samples selected for this study were acquired following a strict set of guidelines for handling and preserving the tissues. The guidelines were previously agreed upon between the tissue accrual site and the receiving company, Gene Logic, Inc. The tissue samples collected sufficiently weighed greater than 500 mg, and were frozen in liquid nitrogen within 30 minutes of excision. Samples are shipped in dry ice to ensure specimen integrity. A pathologist from the tissue accrual site performs an in depth examination on all tissue samples and makes an evaluation. The pathology report, clinical summary and patient history are shipped with each sample, though honoring patient confidentiality. Gene Logic, Inc.'s own pathologist does a histological examination on slides created from each tissue as a quality control on the site of origin and diagnosis of the tissue being studied, and to evaluate how well the tissue has been preserved.

Preparation of Samples

The preparation of samples closely followed the protocol that the Affymetrix GeneChip® Expression Analysis Manual (Santa Clara, CA) outlines. Frozen tissue samples were pulverized into a fine powder using the Spex Certiprep 6800 Freezer Mill (Metuchen, NJ). TRIzol (Life Technologies, Rockville, MD) is then used to extract the total RNA from the crushed tissues. The tissue weight can range from 300-600 mg, and the total RNA yield for each sample was 200-500 µg. Double-stranded cDNA was created from 1-5 µg of mRNA using the Super Script Choice system (Life Technologies). A T7-(dT₂₄)

oligonucleotide served as a primer for the first strand cDNA synthesis. Phenol/Chloroform is then used to extract the cDNA, where it is then precipitated to a final concentration of 1 µg/µl using ethanol. Ambion's T7 MegaScript In Vitro Transcription Kit (Austin, TX) is used to synthesize cRNA from 2 µg of cDNA. The cRNA is biotin labeled using the nucleotides Bio-11-CTP and Bio-16-UTP (Enzo Diagnostics, Inc., Farmingdale, NY). After incubating at 37 °C for 6 hours, the labeled cRNA was cleaned up following the RNeasy Mini kit protocol (Qiagen). A fragmentation buffer consisting of 40 mM Tris-Acetate, at pH 8.1, 100 mM KOAc, and 30 mM MgOAc was then added to the cRNA at 94 °C for 35 minutes to fragment the cRNA. The Affymetrix protocol states that 55 µg of the fragmented cRNA be hybridized onto the 60K set of chips for 24 hours at 60 rpm in a hybridization oven set at 45 °C. The chips are then washed with a stringent and non-stringent buffer, and stained with Streptavidin Phycoerythrin (SAPE) (Molecular Probes, Eugene, OR) in the Affymetrix fluidics stations. SAPE was added twice with an anti-streptavidin biotinylated antibody (Vector Laboratories, Burlingame, CA) staining step in between to amplify the signal. Hybridization onto the probe arrays was detected using fluorometric scanning (Hewlett Packard Gene Array Scanner, Hewlett Packard Corporation, Palo Alto, CA) with an argon ion laser. The microarray images created from scanning were evaluated for quality control, and the data analyzed using Affymetrix GeneChip® software (version 3.3), and the Experimental Data Mining Tool (EDMT) software (version 1.1).

Data Analysis

The GeneChip Human Genome U95 array set representing ~60,000 genes or ESTs was used in the experiment. The expression levels of approximately 12,000 full length human genes are measured on the HG-U95A array. The other four arrays (HG-U95B, C, D and E) collectively survey the expression of over 50,000 human genes and ESTs. The Affymetrix GeneChip® 3.3 software subtracts the mismatch (MM) fluorescent intensities from the perfect match (PM) fluorescent intensities for each of the probe sets to account for the variability in hybridization among probe pairs and other factors that could affect the intensities.

The set of genes that are commonly present or commonly absent in N-1 samples of the sample set, where N represents the number of samples in the set, is used in determining whether there are detectable

genes in the set. In this case, N=10. The expression levels measured are defined as being detectable or not detectable in the normal and diseased tissue sets. Once this is determined, the over- and under-expressed genes for a disease can be seen (5).

1. INTRODUCTION

In a 1953 issue of the journal *Nature*, James Watson and Francis Crick proposed the structure of deoxyribonucleic acid (DNA) to be two polynucleotide chains (or strands) that spiral around an imaginary axis to form a double helix. The chains consist of nucleotides that are broken down into three parts consisting of a nitrogenous base joined to a pentose sugar bonded to a phosphate group. The nitrogenous bases are composed of two families: pyrimidines and purines. Pyrimidines (cytosine [C], thymine [T], and in ribonucleic acid (RNA) uracil [U]) are characterized by a six-membered ring composed of carbon and nitrogen. The second family is the purines (adenine [A], guanine [G]) that are composed of a five-membered ring fused to a pyrimidine type ring with both rings composed of nitrogen and carbon.

As stated above, the DNA double helix is comprised of two chains held together by hydrogen bonds between pairs of bases. A base pair (bp) can either be AT or GC; the AT base pair is held together by two hydrogen bonds and the GC base pair is held together by 3 hydrogen bonds. Therefore, the GC base pair is more stable and will denature (melt) at a higher temperature than the AT base pair.

DNA provides templates called genes for the flow of genetic information in cells. Genes are composed of introns and exons and can be on the order of 1000 base pairs (bps) in length. Introns and exons are transcribed into ribonucleic acid (RNA); however, introns are cut out of newly synthesized RNA molecules. Thus, processed RNA molecules are continuous exons that code for amino acids.

An amino acid is synthesized from a code given by a group of three RNA bases. The three base codes are entitled a codon. Sixty-four codons have been deciphered, of which sixty-one correspond to particular amino acids and three code for protein chain termination. Degeneracy exists because many amino acids are coded for by more than one triplet. Because of this information, the sequence of the human genome is an important source of information. Deciphering the sequence of the human genome can have implications ranging from forensics to molecular biology. Therefore, many methods exist to profile DNA. The following is an exploration into the field of DNA profiling.

DNA Profiling

Review of Methods of DNA Fingerprinting

One of the first methods to profile DNA was called restriction enzyme fragment length polymorphism (RFLP)/restriction digest fingerprinting.¹⁻⁴ In humans, RFLP loci with as many as 80 different alleles have been reported.⁵ These types of loci are termed variable number of tandem repeats (VNTR) consisting of sets of tandemly repeated nucleotide core sequences. The length of the core repeats can vary from 11 to 60 base pairs (bps). Conserved endonuclease restriction sites⁶ flank the repetitive region. Therefore, the length of the restriction fragment is proportional to the number of core repeat units. To do RFLP analysis, DNA is first extracted from human body fluid or tissue. Then the extracted DNA is digested with a restriction enzyme, subjected to electrophoresis, southern blotted⁷, and probed sequentially with several single locus probes.^{8,9} The restriction enzyme chosen as standard for digestion is Hae III.

The vast majority of individuals are heterozygous at loci commonly used in forensic identification, and thus two RFLP bands are present. These two bands that are present represent the Hae III restriction fragments that contain the VNTR region (D2S44) complementary to the probe. Although this method is well defined and highly informative, problems exist including: the amount of time for analysis¹⁰, the effects of degraded DNA, and the amount of template DNA needed for adequate results (20-100 ng). In addition, with the enzyme Hae III, partial digestion of the template DNA leading to extraneous bands on the gel or analysis format^{8,9} can pose a problem with explanation of the results. Extraneous bands can interfere with analysis of multiple donors.

Because RFLP suffers from the problems listed above, a new method for DNA analysis was needed. In fact, in some criminal cases, the starting sample is relatively small in nature (single hair follicle¹¹, 1 µl drop of blood), and can be exposed to influences like ultraviolet light, humidity, decay or storage in paraffin/denaturing conditions.¹² These conditions of storage can lead to degraded DNA, meaning that a technique is needed that examines only small parts of the genome. Fortunately, small amounts of short-length DNA can be amplified using polymerase chain reaction (PCR). PCR works by using synthetic oligonucleotide primers that flank the region to be amplified. Each primer is complementary to a different DNA strand. The target duplex DNA (template DNA) is denatured by heat and annealed to the primers that are in excess to prevent the two strands of the target DNA from re-

annealing with each other. The two primers are then extended on the separated target DNA template strands using a heat stable DNA polymerase and dNTPs that exist in the PCR mix. This results in a two-fold amplification. A second round of denaturation using heat, annealing, and primer extension would result in a four-fold amplification. The number of copies of template DNA using PCR will be 2^n , where n is the number of amplification cycles. With a method to amplify/recopy DNA, other fingerprinting methods, such as short tandem repeats that use minimal amounts of template DNA, were examined.

The genetic uniqueness of individuals is a central tenet of human biology. One method to examine the uniqueness of an individual is the use of short tandem repeat sequence variants (STR).¹³⁻²¹ Short tandem repeats are tandemly repeated units of sequence ranging between 2 and 7 bps that have high levels of heterozygosity, distinguishable alleles that are polymorphic based on the number of repeat units, and are amplified using PCR amplification. The human genome has an abundance of these sequence occurrences, averaging one tri- or tetra-nucleotide unit every 15 kbp. There have been over two thousand STR systems mapped out in the human genome, and it is estimated that thousands more remain to be discovered.²²

STRs are useful for disease diagnostics (Fragile X, Huntington's disease, Spinobulbar muscular atrophy), paternity testing, physical and genetic mapping, monitoring bone marrow transplants, and forensics. The use of STRs in forensics and parentage assessment was the focus of this study. Forensics and parentage assessment use STR systems because of the flow of genetic information from parent to offspring. The human genome of an offspring carries two copies of each chromosome, one copy from each parent. Chromosomes contain genes that have the DNA sequence to code for and build proteins. An exon is the region of DNA sequence in a gene that codes for a protein. The sequence that exists between genes or part of a gene that does not code for a protein is called an intron. The locations or particular positions of a gene, the intronic sequence of a gene, and the intronic sequence between genes are termed loci. A single location in the human genome is termed a locus. When a locus on a chromosome differs at the sequence level, it is called an allele. STRs are made up of different numbers of repeating units or alleles.

One example of an STR is the locus HUMTHO1. This particular locus is a tetrameric (4-bp repeat) system that will be analyzed using HPLC. Tetrameric repeat systems are used in the forensic

community because PCR amplification of these systems yields fewer artifacts, such as extra band production and repeat slippage or stuttering.⁹⁵⁻⁹⁷ The HUMTHO1 locus is located in intron 1 of the human tyrosine gene with a chromosomal position of 11p15-15.5.⁵⁵ The alleles present in the locus are represented in table 1, along with the bp size of each allele, and the repeat structure. The repeating unit of HUMTHO1 is (AATG). The alleles are numbered according to the number of tandem repeats present in the PCR-amplified products. If a sample has 9 AATG repeat motifs on one chromosome and 10 AATG repeat motifs on the other chromosome, the sample is designated a HUMTHO1 (9, 10). However, not all alleles in the HUMTHO1 locus carry just 4 bp repeats (AATG). Some alleles have base pair deletions (i.e., 9.3 allele in table 1). The allele type is the number of 4 bp repeats. An additional 0.1 is added for each bp less than 4 sandwiched within the string of repeats. A HUMTHO1 9.3 sample has 9 of the 4 bp repeats. An extra ATG exists between the sixth and seventh repeats, accounting for the 0.3 designation in the type (0.1 for each bp of the ATG).

All STR loci are parentally inherited. One allele from each parent is passed on to the offspring. A simple

example is a father that has the type HUMTHO1 (6, 9) and a mother that has the type HUMTHO1 (7, 8). The offspring can have the following HUMTHO1 types (6, 7; 6, 8; 9, 7; 9, 8). If the typing does not match one of these pairings, the offspring cannot be assigned to that set of parents.

Table 1 HUMTHO1 Locus

Allele Repeat #	Base Pair Size	Repeat Structure
3	146	[AATG] ₃
5	154	[AATG] ₅
6	158	[AATG] ₆
7	162	[AATG] ₇
8	166	[AATG] ₈
8.3	169	[AATG] ₅ ATG[AATG] ₃
9	170	[AATG] ₉
9.3	173	[AATG] ₆ ATG[AATG] ₃
10	174	[AATG] ₁₀
10.3	177	[AATG] ₆ ATG[AATG] ₄
11	178	[AATG] ₁₁
12	182	[AATG] ₁₂
13.3	189	[AATG][AACT][AATG] ₈ ATG[AATG] ₃

Another use of STRs is in forensic cases. If blood, skin, dandruff, or any other biological material is found at a crime scene, it can be analyzed using STRs and matched with the suspect's blood type. With this use of STRs, population databases have been set up for different STR systems.²⁴ The allele frequency for HUMTHO1 STR system has been analyzed for the United States populations (table 2).

Table 2 HUMTHO1 allele frequencies in United States populations
(Adapted from ref. 24)

Allele	African American (N = 198)	Caucasian (N = 197)	Hispanic (N = 206)
5	0.000	0.000	0.002
6	0.111	0.221	0.228
7	0.434	0.175	0.340
8	0.189	0.127	0.083
9	0.146	0.168	0.100

9.3	0.106	0.308	0.245
10	0.013	0.003	0.002

Heterozygosity is a way of measuring the genetic variation that is useful in establishing the diversity that exists in a population. In the case of STR loci, the heterozygosity is the appearance of different alleles for a locus of an individual. The observed heterozygosities for the locus HUMTHO1 in different populations are 70.2% for the African American sample, 72.1% for the Caucasian sample, and 77.2% heterozygosities for the Hispanic sample.²⁴

With the population data and the heterzygosities of the HUMTHO1 locus at acceptable levels for use in paternity testing and forensic analysis, this locus will be used for typing with HPLC as the separation mode.

Along with the STR systems for identification, a reliable method for sexing of samples has been accomplished.¹⁰² Amplification of the X-Y homologous amelogenin gene with a single primer pair is used for a sex test because it generates different length products from the X and Y chromosomes (table 3). This locus is co-amplified with STR loci making it a powerful tool in forensic science because, along with the identification power of STRs, the sex of the offender can be known.

Table 3 Amelogenin Locus

Allele	Base Pair Size
X chromosome	106
Y chromosome	112

Methods Used for STR Sizing/Typing

The first method used for analysis of STR's was slab gel electrophoresis.^{23, 24} This method is still in use today due to low cost of instrumentation and the fact that multiple samples can be analyzed at one time. However, gel electrophoresis is inefficient because the length of analysis time is long (typically two to three hours), the sample throughput is low, and the sample handling is labor-intensive. In addition, gel or electric field inhomogeneity can result in inconsistent migration of sample zones among lanes or even in a single lane. Therefore, other methods of electrophoresis were explored.

One method that is expected to replace slab gel electrophoresis is capillary electrophoresis (CE).²⁵

^{26, 27} However, CE suffers from run-to-run variation caused by differing buffer viscosities, analysis voltages, and temperature. Therefore, precise sizing measurements are difficult. In order to overcome these problems and to calibrate migration times, internal standards need to be added to the sample.^{28, 29} However, many of the internal standards that have been used were of a different base composition than the STR loci of interest, and unfortunately, electrophoretic mobility can be affected by base composition.³⁰

Another CE approach has been the co-injection of a STR ladder and STR PCR products each labeled with a different fluorescent dye. These experiments required a CE system with multi-wavelength detection capabilities.^{22, 25, 26, 31-34} It is necessary to choose fluorescent labels carefully to eliminate mobility shifts due to the dyes themselves. In practice, it is difficult to produce a perfect mobility match due to the different

chemical structures of the fluorescent labels. Due to the aforementioned problems, the method of HPLC will be developed and applied to STR analysis.

Liquid Chromatographic Methods of DNA Separation

High performance liquid chromatography (HPLC) has become a well-established technique for the analysis of biopolymers. Different chromatographic methods are used for the separation of single- and double-stranded DNA, including mixed-mode, size-exclusion, affinity, ion-exchange, reverse-phase, and ion-pairing reverse-phase (IPRP).

Mixed-mode chromatography³⁵⁻³⁷ was the first chromatographic method for nucleic acids. The separation is based on ionic and hydrophobic interactions. The resin consisted of a charged reverse-phase matrix with a quaternary ammonium derivative absorbed on a non-porous polymer support. One example was the modification of a reversed-phase support with a tetraalkylammonium salt to induce ionic interactions with the DNA.³⁵ This method suffers from an influence of AT content, base composition, and DNA conformation on the retention time of dsDNA.^{37,38} Also, this method allowed for bleeding of the stationary phase at low salt concentrations and required a long separation time for DNA. Therefore, this method would not be useful for sizing PCR products in a minimal amount of time. The idea that the size of dsDNA could be used for a separation was explored in a size-exclusion type of chromatography.

Size-exclusion or gel-permeation chromatography has been used for the separation of restriction fragments.³⁹ This type of chromatography relies on the penetration of the separating analyte into the pores of the stationary phase. The dsDNA is partitioned between the mobile phase and the stationary phase. The stationary phase is a mixture of neutral and hydrophilic porous particles composed of organic compounds, or silica coated with hydrophilic surfaces. The penetration into the pores is dependent on the shape of the dsDNA. This type of separation needs smaller fragments of dsDNA that can interact with the pores in stationary phase. In some cases, 200 bp is the upper limit of separation making this mode of chromatography not useful for STR systems. In addition, this type of chromatography suffers from poor resolution and separation times on the order of hours. In order to improve resolution, the interaction between strands of DNA was explored for a chromatographic separation.

Affinity chromatography has been investigated for the separation of ssDNA.⁴⁰ This method takes advantage of the base pairing that exists in duplex DNA. One strand of the target duplex DNA is attached

to the chromatographic support, thus the column highly specific. Lowering the salt concentration elutes the attached, duplexed DNA strand. The other method uses elevated temperature to dissociate the DNA duplex. One example is the separation of poly(A)_{12, 14, 16, 18} oligonucleotides using a stationary phase composed of amino linked thymidylic acids (dT)₁₈. One problem that exists with this method is the requirement that the column be packed specifically for the ssDNA of interest. This technique has not been used to separate dsDNA. The ssDNA requires long analysis times (>60 min) which is not useful for the separation of STR alleles since the STR products would be larger in bp size.

Anion exchange (AE) chromatography is a method that is rapid and reproducible for separation of ss- and ds-DNA. First, AE was shown to be a rapid technique for the separation of single-stranded oligonucleotides with single base resolution.⁴¹ AE chromatography has been utilized to separate dsDNA, but these separations show sequence dependent retention behavior of the dsDNA fragments.⁴²⁻⁴⁵ High Adenine and Thymine (AT) content of some dsDNA fragments cause tighter binding to the stationary phase than is expected based on the molecular size alone.³⁹ Therefore, two DNA fragments with the same number of base pairs will elute differently based on AT content. The fragment with the higher AT content will elute slower. To eliminate separation based on AT content, a new method of dsDNA separation called ion-pair reverse-phase (IPRP) HPLC has been used.

IPRP HPLC is composed of two different forms of chromatography, ion-pairing and reverse phase. Ion-pair (IP) chromatography uses a mobile phase consisting of a aqueous buffer with an added counter-ion of opposite charge to the sample molecule.⁹⁸ In the case of an ion pair separation of negatively charged DNA, the counter-ion would be positively charged. The counter-ion that will be used in this thesis is triethylammonium acetate (TEAA). The DNA negative ion forms ion pairs with the positive TEAA ion creating a net charge of zero. The neutral ion-pair can interact with the stationary phase for separation based on the size of the molecule. IP is used extensively with reverse-phase chromatography.

Reverse-phase (RP) chromatography is the most widely used type of separation.⁹⁸ RP is used to separate neutral molecules in solution of the basis of hydrophobicity.⁹⁸ The stationary phase is non-polar while the mobile phase is polar. The column used in this thesis is of the reverse-phase type. Alkylated poly(styrene-divinylbenzene) (PS-DVB-C₁₈) is the packing in the DNAsep™ reverse-phase

HPLC column. This type of column has two main theories that explain the separation mechanism, solvophobic and partitioning. In the solvophobic theory, the stationary phase is thought to behave more like a solid than a liquid, and retention is considered to be related primarily to the hydrophobic interactions between the solutes and the stationary phase (solvophobic effects).⁹⁸ Because of the solvophobic effects, the solute binds to the surface of the stationary phase, thereby reducing the surface area of the analyte exposed to the mobile phase. Adsorption increases as the surface tension of the mobile phase increases. Hence, solutes are retained more because of solvophobic interactions with the mobile phase than through specific interactions with the stationary phase.

The partitioning model of retention states that the stationary phase is an important part of the retention process for an analyte. The solute (or analyte) is embedded in the chains of carbons in the stationary phase rather than adsorbed on the surface of the stationary phase as in the solvophobic theory.⁹⁸ Therefore, the solute is considered to be partitioned between the mobile phase and the stationary phase. Both mechanisms are believed to work together for the separation of analytes using RP chromatography.⁹⁸

IPRP chromatography is a result of a combination of the methods of IP and RP. This type of chromatography is useful for the analysis of negatively charged DNA. The TEAA molecule in the mobile phase is the IP agent, and the RP column is alkylated poly(styrene-divinylbenzene) (PS-DVB-C₁₈). There are three mechanisms or models proposed for the IPRP separation of molecules.

The ion-pair model⁹⁹ postulates that the IP reagent contains bulky organic substituents that cause the ion-pair to be hydrophobic in character and adsorb onto the hydrocarbon stationary phase. The dynamic ion-exchange model¹⁰⁰ proposes that the unpaired organic counter-ion absorbs to the surface of the non-polar stationary phase, and is present in the mobile phase. Thus, a dynamic equilibrium forms between the IP agent in the mobile phase and the IP agent absorbed onto the stationary phase. This type of interaction causes the column to behave as an ion-exchanger, and the sample ions are separated based on conventional ion-exchange. The ion-interaction model^{79, 101} is an intermediate between the two previous models, and proposes the formation of an electrical double layer at the stationary phase surface. As in the previous model, it is suggested that a dynamic equilibrium occurs between the IP agent absorbed onto the stationary phase and the IP in the mobile phase. However, this model proposes that the primary layer of charge is attracted to a second layer of loosely held ions of the opposite charge.⁹⁸ Transfer of solutes through the double layer to the stationary phase is both a function of electrostatic effects, and the solvophobic effects responsible for the retention in RP chromatography.

With the unique properties of an IPRP separation, this method was applied to dsDNA in the form of restriction enzyme fragments that are smaller than 600-bp. Unfortunately, this separation took more than three hours and was proven impractical for analytical or preparative purposes.^{45, 58}

The IPRP separations of DNA were later investigated by Huber et al.⁴⁷⁻⁵¹ The results of this investigation provided a new column packing material and a different mobile phase that allowed for a more

rapid separation that took less than 25 minutes for dsDNA (up to 2176 bp). This separation was based on DNA fragment size, not AT content.⁴⁶⁻⁵¹ The IPRP columns (50 × 4.6 mm inner diameter) are packed with alkylated poly(styrene-divinylbenzene) (PS-DVB-C₁₈) and are commercially available as DNAsep™ columns (Sarasep Inc., Santa Clara, CA). Due to the rapid separations using IPRP, this method will be explored for dsDNA sizing and typing.

Materials and Methods

Chemicals

HPLC grade acetonitrile (ACN) (EM Science, Gibbstown, NJ) and triethylammonium acetate (TEAA) (Applied Biosystems, Foster City, CA) were used to constitute the mobile phase.

Preparation of the Mobile Phase

The mobile phase consists of 0.1 M TEAA (Solvent A) and 0.1 M TEAA-25% ACN (Solvent B). In order to keep the concentration of TEAA constant and unaffected by volume contraction during the mixture of organic solvents with water, the mobile phase was prepared as follows: for Solvent B, 50 ml of the 2 M TEAA stock solution was added to 250 ml of ACN, and this was diluted to 1000 ml with the addition of deionized water.

DNA Extraction

Whole blood samples were collected in 5-ml tubes containing the anticoagulant EDTA. DNA was extracted from 300 ml of whole blood using the Puregene DNA Isolation Kit (Gentra Systems, Inc., Minneapolis, MN) following the manufacturer's protocol.

STR Amplification

The yields of purified genomic DNA isolated from nucleated cells fell within the range of 5-15 µg for 300 µL of whole blood. The concentrations were determined by UV spectroscopy using a spectrophotometer. DNA amplification was performed as described in the Perkin-Elmer ABI PRISM STR protocol (Perkin-Elmer, Foster City, CA). The amplified tetrameric STR locus will be HUMTHO1. The other locus of interest is the amelogenin locus. These loci will be amplified using primers from the Forensic Service of the British Home Office for use in forensic casework. The information for the HUMTHO1 locus is in table 1 and the amelogenin locus is in table 3. Table 4 is a description of a 2-bp repeating unit that occurs in human mitochondrial DNA that will be co-analyzed with HUMTHO1 alleles and amelogenin. The primers used for the PCR amplification of all loci used in this study are in table 5.

Table 4 mtDNA dinucleotide repeat

Repeat #	Base Pair Size	Repeat Structure
6	137	(CA) ₆

5	135	$(CA)_5$
4	133	$(CA)_4$

Table 5 Primers to be used for PCR amplification of all loci

Locus	Primer pairs
HUMTHO1	5'-CCCTGGGCTCTGTAAAGAATAGTG-3' 5'-ATCAGAGCTAAACTGGGAAGCTG-3'
Amelogenin	5'-GTGGGCTGAAAAGCTCCGATTAT-3'
mt dinucleotide repeat	5'-GTGATTCCCATTGGCCTGTTCCCTC-3' 5'-CTCCCATACTACTAATCTCA-3' 5'-TTGAGGAGGTAAAGCTACATA-3'

Instrumentation

Two HPLC systems were used for analysis. The first HPLC system is a Dionex DX-300, non-metallic, polyetherketone (PEEK) system with a variable wavelength detector set at 260 nm, and a DNAsep™ column. The column oven temperature will be maintained at the optimal temperature using (Model 505, Scientific Systems, Inc., State College, PA) with a pre-column filter (0.2 µm PEEK filter, Upchurch Scientific, Oak Harbor, WA), a 50-µl sample loop, and a Dionex AS3500 autosampler with the capacity to handle 105 samples. The Dionex HPLC system uses Peaknet 4.30 for data handling.

The second system used is a Transgenomic non-metallic PEEK system HPLC with a variable wavelength detector set at 260 nm, a DNAsep™ column, and an autosampler with the capacity to handle 96 samples. The Transgenomic system uses Hitachi model D-7000 chromatography data station software for data analysis.

Statement of Dissertation

Liquid chromatography is a method that is useful for the separation of dsDNA. However, this method has not been optimized for the separation of STR systems. The parameters that affect the liquid chromatographic separation of dsDNA with a bp size between 100 and 400 were examined and optimized (chapter 2). Once optimization of the separation parameters was completed, HPLC WAS used for the sizing/typing of the HUMTHO1 locus (chapter 3). Finally, the HUMTHO1 locus, a mtDNA repeat, and the amelogenin locus were multiplexed using HPLC (chapter 3). In addition, the sizing method developed can be used for other PCR products.

2. OPTIMIZATION OF SEPARATION PARAMETERS USING LIQUID CHROMATOGRAPHY FOR DSDNA WITH BASEPAIR SIZES BETWEEN 51-458

Discussion

Effects of Column Temperature

The effect of column temperature in reverse-phase (RP) liquid chromatography has been studied both theoretically and experimentally.⁶⁰⁻⁶² The results of these studies can be summarized to show that column temperature has three main effects. First, elevated column temperatures will usually reduce the analysis time because of the exothermic enthalpy changes associated with the transfer of solutes from the mobile to stationary phases that dominate the retention process in most chromatographic

systems.^{61,63} Second, a change in temperature can have a pronounced effect on the efficiency of a separation because an increase in temperature will reduce viscosity of the solvent and increase diffusion rates of the solute, having an enhanced effect on the mass transfer rate.^{61,64} Li and Carr found that increasing temperature could increase column efficiency by 30% with a RP type of separation.⁶⁵ Third, the change in retention with temperature is often different for various analytes, and temperature can affect the selectivity of a separation.⁶⁶⁻⁶⁸ Chloupek et al. found that the selectivity is often improved at elevated temperatures.⁶⁹

In the development of a liquid chromatographic method, the column temperature can be an important tool for large solutes.⁶⁵ It is thought that for large solutes, the

temperature can affect retention, efficiency, and selectivity. The backpressure caused by using smaller particles and high flow rates for ultrafast separation can be reduced by higher column temperatures. In addition, the miniaturization of column dimensions will enable faster thermal equilibration, creating new opportunities for selectivity and retention control based on temperature programming.⁶⁴

Column temperature is an important parameter for dsDNA separation using liquid chromatography because the column temperature has been shown to affect the resolution between fragments and an affect the physical properties of a DNA fragment.⁵⁰ Using alkylated non-porous poly(styrene-divinylbenzene) (PS/DVB-C₁₈) that compose the DNasep™ column, Huber et al. found 50°C to be the optimum column temperature for the separation of dsDNA.⁴⁸ However, the temperature of the column was only explored in 10°C increments. In Oefner's study, a temperature of 60°C decreased the resolution for dsDNA fragments, but the effect of 1°C increments from 50°C to 60°C on separation efficiencies of dsDNA was never explored.

At the optimal column temperature for the separation of dsDNA, the fragments should not undergo partial denaturation because this would affect the sizing experiments by reducing the accuracy since the separation would be based on sequence and not the size of the DNA fragments. The partial denaturation of the dsDNA helix on the column produces a change in retention time. The retention time for a partially denatured dsDNA fragment decreases as the charge density of a formally helical structure is disturbed. This produces a bubble formation in the dsDNA helix limiting the number of charges accessible for electrostatic interaction between the DNA and the double layer of the stationary phase. The charge density of the formally helical region decreases from two negative charges per base pair to one negative charge per base in the single stranded region of the bubble. Incorporation of these properties of dsDNA into a method enables the examination of a dsDNA separation for the partial denaturation and melting of a fragment. The effects of column temperature on the separation of dsDNA in the form of restriction enzyme cleaved plasmid DNA (RE ladders) and PCR-amplified DNA were examined. The plasmid DNA fragments were used as sizing markers throughout the experiments to size PCR-amplified DNA (target DNA) between 100 to 400 bp in length. The sizing markers need to be both smaller than the target DNA and larger than the target DNA because the markers will be used to draw a line of regression that will include the bp sizes of the target DNA and thus acquire a bp size for the target DNA.

The plasmids treated with the types of restriction enzymes (RE) used in this study cleave with blunt ends. This means that the enzyme cleaves off excess bases and leaves no bases that overhang the duplexed DNA. The plasmids treated with Hae III and Msp I cleave due to the interaction of the phosphate group on the 5' end. Therefore, the fragments from the pUC18 plasmid and the pBR322 plasmid contain a phosphate group on the 5' of both strands (parallel and anti-parallel). The RE fragments have two more negative charges than the PCR-amplified alleles since PCR does not add a phosphate group on the 5' end. This might affect the sizing assay. The size and bp compositions of the fragments from the RE ladders used in this study are listed in tables 6, 7, and 8.

The PCR-amplified DNA samples used for the temperature study are from the locus HUMTH01. The amplification conditions are located in the materials and methods in chapter 1 along with a detailed description of each allele.

Table 6 Fragment sizes and AT content for pUC18 cleaved with Hae III

Fragment size (bp)	%AT
80	41.0
102	45.0
174	41.4
257	49.0
267	49.4
298	40.9
434	42.2
458	57.4
587	56.9

Table 7 Fragment sizes and AT content for pUC18 cleaved with Msp I

Fragment size (bp)	%AT
67	46.3
89	53.9
110	50.0
147	39.5
190	49.5
242	51.2
353	41.6
404	56.6
489	43.6
501	56.5

Table 8 Fragment sizes and AT content for pBR322 cleaved with Hae III

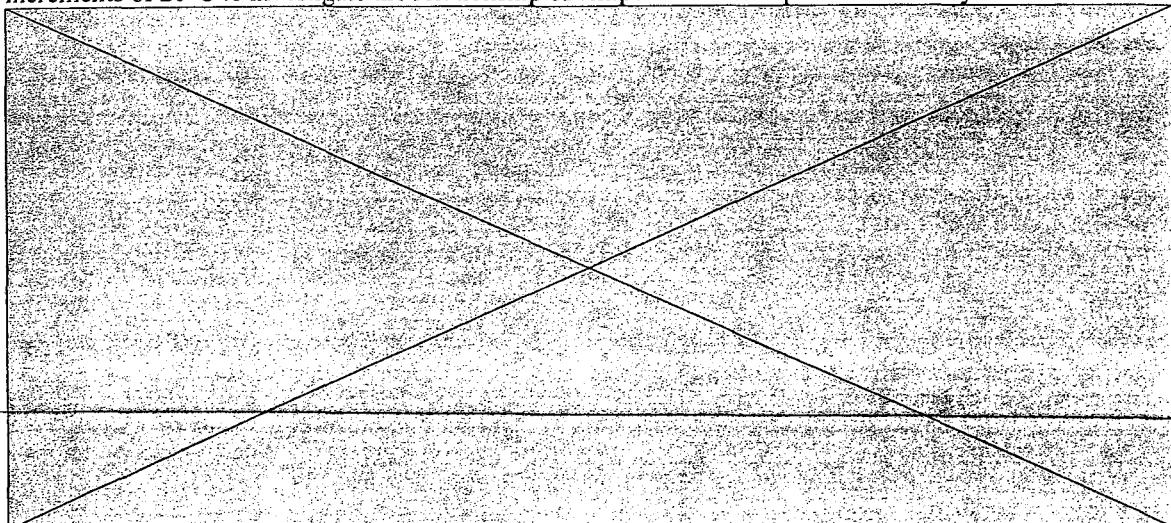
Fragment size (bp)	%AT
51	29.0
57	38.6
64	32.8
80	41.0
89	39.0
104	35.6
123	43.9
124	34.7
184	40.8
192	52.6
213	38.5
234	40.0
267	49.0
434	57.4

458

42.2

For the analysis of RE fragments and PCR-amplified dsDNA, the chromatographic conditions were held constant at a flow rate of 0.9 ml/min and a gradient where the percentage of acetonitrile (ACN) was increased 2.0% per minute (2% ACN per min). The flow rate and gradient were held constant. The column temperature was the only chromatographic condition changed during the analyses.

One of the first effects of raising the temperature of the column is that the dsDNA fragment is retained longer in the stationary phase (figure 1). These results are contrary the observations of Li and Carr who studied raising the column temperature for the separation of alkylbenzenes.⁷⁰ Based on their study, the analytes' retention times became shorter as the column temperature was raised. However, Li and Carr used increments of 20°C to investigate the relationship of temperature and separation and alkylbenzenes and



DNA are different types of solutes. With an increase in temperature of the column, the dsDNA begins to straighten due to the defolding and derotating of the double helix.⁴⁸ The defolding and elimination of any secondary structure in the DNA helix exposes more phosphate groups on the DNA backbone for ion-pairing formation with the triethylammoniumacetate (TEAA) ion. This allows stronger interaction with the alkylated polymer beads, giving a longer retention time. More acetonitrile will be required to move the dsDNA from the stationary phase into the mobile phase at higher temperatures.

Figure 1 Retention time for a 173-bp HUMTHO1 product with increasing column temperature

The HUMTHO1 PCR product increases in retention time (RT) from 50°C to 57°C, but the RT decreases once the column temperature reaches 57°C. Table 9 represents the change in retention time for each degree change in column temperature. The temperature change from 50°C to 51°C is the most substantial change in RT. This is because the dsDNA is becoming more rod-like, and is exposing more phosphate groups for interaction with the stationary phase.⁵⁰ As can be seen in table 9, the increase in temperature reaches a plateau where an increase in temperature does not affect the retention time substantially. The temperature change from 53°C to 54°C causes a larger change in retention time than the changes in retention time that occur from 54°C to 55°C, 55°C to 56°C and 56°C to 57°C. It is possible that the dsDNA helix for the HUMTHO1 (9.3, 9.3) sample has defolded and derotated to reach a maximum linear state at the temperature of 54°C. Therefore, any increases in temperature are not going to produce any further ion-pairing sites for the TEAA to interact with on the DNA. Once the column temperature reaches 57°C, the dsDNA begins to denature thus, reducing the RT.

Table 9 Change in column temperature and the effect on the change in retention time of a HUMTHO1 (9.3, 9.3)

Temperature 1 (T_1)	Retention time for T_1 (R_1)	Temperature 2 (T_2)	Retention time for T_2 (R_2)	$\Delta (R_2 - R_1)$ (min)
50	13.33	51	13.49	0.16
51	13.49	52	13.58	0.09
52	13.58	53	13.63	0.05
53	13.63	54	13.69	0.06
54	13.69	55	13.72	0.03
55	13.72	56	13.75	0.03
56	13.75	57	13.76	0.01
57	13.76	58	13.74	-0.02
58	13.74	59	13.57	-0.17

To explore any correlation between the sequence of the dsDNA fragment and the increase in retention time due to a rise in the temperature of the column, dsDNA fragments of different bp sizes and different sequences were analyzed. The results were plotted as retention time versus temperature (figure 2). As can be seen in figure 2, all of the RE fragments are retained longer as the column temperature is increased. Since

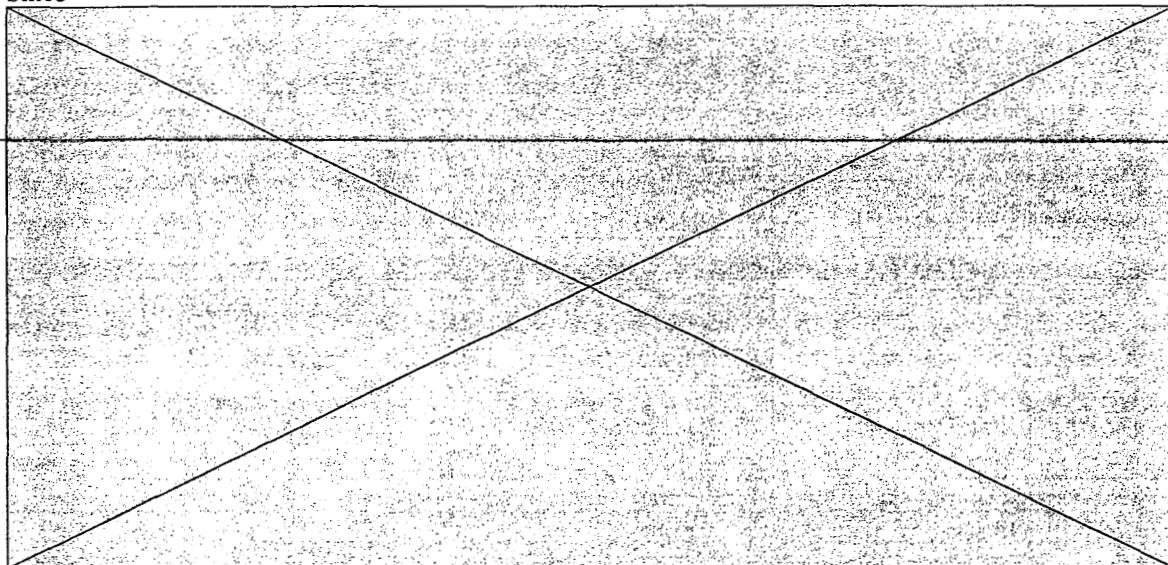
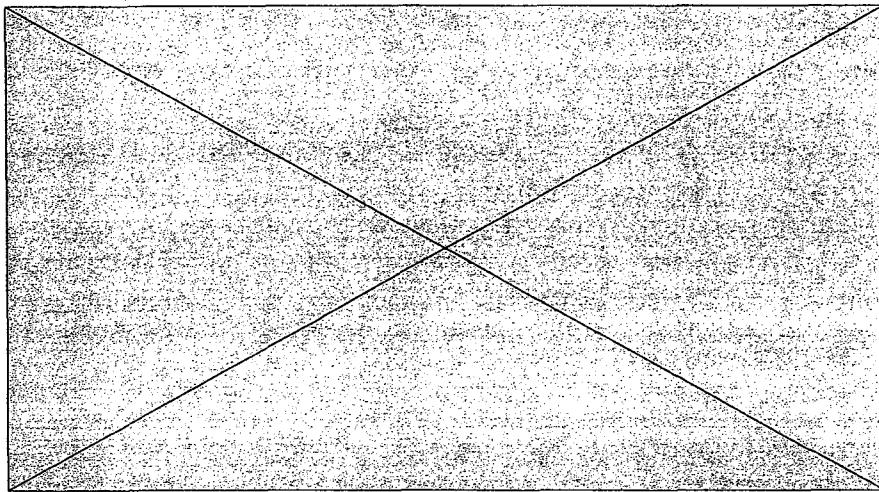


Figure 2 Retention time of RE fragments with increasing column temperature

secondary structure in the dsDNA is eliminated, the phosphate groups of each bp can have more interactions with the TEAA ion. This methodology also indicates the temperature that dsDNA begins to partially denature. As the dsDNA begins to denature, there is a decrease in retention time because there are not as many charges accessible for electrostatic interaction between the DNA and the double layer of the stationary phase. Although all the fragments examined had different numbers of bps and different percentages of GC, no partial denaturation of any of the markers took place. Therefore, an optimum temperature for dsDNA separation can be found that would allow for all possible fragments to be in a non-denatured state and have the greatest interaction with the stationary phase. Hence, any unwinding of the dsDNA helix for any marker would not compromise sizing accuracy and typing.



Upon examination of the chromatograms, the 458-bp peak from the RE ladder pBR322 cut with Hae III showed a decrease in retention time at a temperature of 55°C (figure 3). The panel marked A shows five peaks from the RE ladder pBR322 cut with Hae III at a column temperature of 54°C. The panel marked B shows the same five fragments separated at 55°C, only the 458-bp fragment co-elutes with the 434-bp fragment. The decrease in retention time of the 458-bp fragment is caused by the partial denaturation of the helix. Because of the partial denaturation of this fragment at a temperature of 55°C, the sizing integrity while using the RE ladders would be compromised at this temperature. Therefore, 54°C is the upper limit for sizing when using the RE ladders. The optimum temperature for analysis of the dsDNA fragments analyzed in this study lies below 55°C.

Figure 3 Separation of RE fragments at the column temperatures of 54°C and 55°C
Panel A is the separation of dsDNA fragments from the pBR322-Hae III at a column
temperature of 54°C. Panel B is the same separation at a column temperature of 55°C. At
55°C, the 458-bp fragment partially melts and co-elutes with the 438-bp fragment. The
separation conditions for both chromatograms were: flow rate of 0.9 ml/min and 2%
increase in ACN per min.

To find the optimal temperature for separation of dsDNA, the RE ladders consisting of pBR322 cleaved with Hae III, pUC18 cleaved with Hae III, and pUC18 cleaved with Msp I were analyzed using column temperatures starting at 50°C, and were repeatedly analyzed as the column temperature was raised by 1° increments to 55°C. The flow rate of the mobile phase was kept constant at 0.9 ml/min, and the rise in percentage acetonitrile (%ACN) per minute was kept at two percent ACN per min. These conditions were used for all the column temperature experiments. The resolution for fragments in the size range of STR alleles (51-458 bp) was calculated using the equation:

$$(1) \quad R_s = 2(t_2 - t_1) / W_1 + W_2$$

where R_s is resolution, t_n is retention time of a component, and W_n is peak width of a component. Since tetranucleotide STR fragments differ in length by divisions of 4 bp (except in the cases of deletions of parts of repeats), fragments that had differing lengths similar to the actual STR products were examined for resolution calculations. The following fragment pairs were used for the resolution calculations: 51-57 ($\Delta = 6$ bp), 57-64 ($\Delta = 7$ bp), 184-192 ($\Delta = 8$ bp), 213-234 ($\Delta = 21$ bp), 257-267 ($\Delta = 10$ bp), 267-298 ($\Delta = 31$ bp), and 434-458 ($\Delta = 24$ bp).

The results of these experiments can be seen in figures 4 and 5. Figure 4 shows the resolution for RE fragments with a bp size < 64, and figure 5 shows the resolution for RE fragments with a bp size > 184. There is a slight increase in resolution for all fragment pairs with the increasing column temperature. However, once 55°C is reached, the resolution decreases for all fragment pairs, and the resolution can not be calculated for the 434-458 bp fragments because the two fragments elute as one peak. The increase in column temperature creates an environment for the dsDNA fragment to unwind and allow for maximum interaction with the ion-pairing agent, TEAA. An increase in ion pairing agent interaction with the dsDNA helix requires the exposure of the charged phosphate groups that are a part of the sugar backbone of DNA. The exposure of phosphate groups for interaction with the ion-pairing agent means that the molecule is derotating and becoming rod-like in nature. The increase in retention time and increase in resolution with increasing column temperature counters the fact that as a DNA molecule increases in size, the dsDNA becomes more randomly coiled in free solution.⁷¹ However, in the chromatographic column, the dsDNA is not in free solution. The DNA flows through the narrow channels of a packing material. This causes stretching of the DNA molecules by shear forces.⁷¹ Therefore, the effect of the increased ion-pairing interaction with the exposed phosphate groups and the removal of secondary structure in the dsDNA helix contributes to increased resolution for RE ladder fragments.

An interesting result of the temperature study for RE fragments was the difference in resolution based on the fragment sizes. Table 10 shows the difference in resolution for different size fragments at a column temperature of 54°C. The fragments that are < 64 bp have the highest resolution values. For the fragments > 184 bp, the resolution decreases. The fragments that differ by seven bp (57, 64) have a resolution that is 2.4 times greater than the fragments that differ by 3 times as many bp (213, 234). Therefore, fragment size is going to have an impact on the amount of separation between peaks that can be attained.

With these facts in mind, all STR systems that are going to be analyzed using HPLC should be the smallest size possible. This should be an easy task because all the STR systems to be analyzed in this study have primer pairs that will allow all sizing of fragment whose size is below 200 bp.

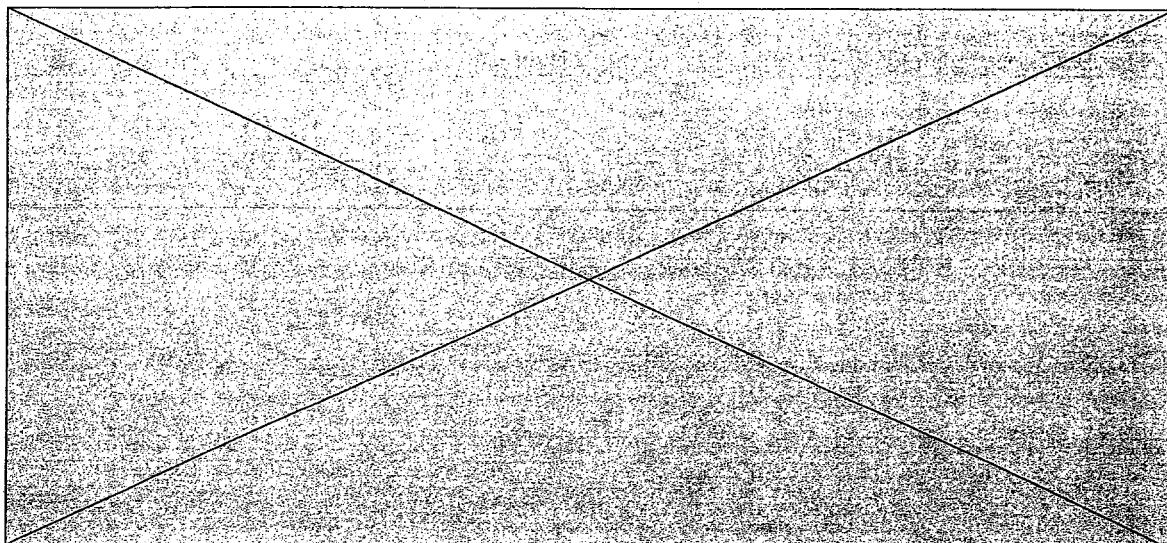


Figure 4 Resolution for RE fragments ($< 64\text{-bp}$) with increasing column temperature

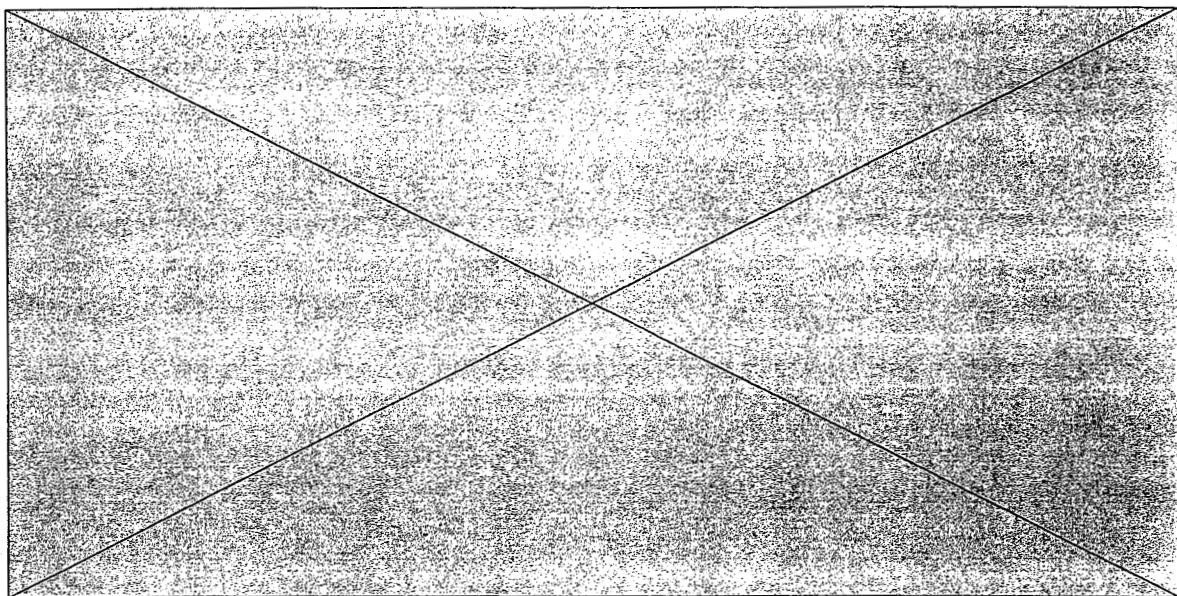


Figure 5 Resolution for RE fragments ($> 184\text{-bp}$) with increasing column temperature

Table 10 Resolution for different RE fragment sizes at a column temperature of 54°C

Fragment sizes (bp)	Δ BP	Resolution at 54°C
51, 57	6	7.61
57, 64	7	9.07
184, 192	8	1.57
213, 234	21	3.76
257, 267	10	1.94
267, 298	31	4.22
434, 458	34	1.59

The theoretical number of plates (N) for RE fragments was examined while increasing the column temperature. The formula for N is

$$(2) \quad N = 16 \times (RT / W)^2$$

where RT is the retention time of the peak of interest, and W is the peak width obtained by drawing tangents to each side of the peak and calculating the distance between the two points where the tangents cross the baseline. Increases in N take place for all fragment pairs as the temperature increases (figure 6 and figure 7). Since the formula for N contains retention time and peak width, changes in both will affect N. The retention time of the fragments increases, and the effect coupled with the shrinking peak widths, the value of N increases.

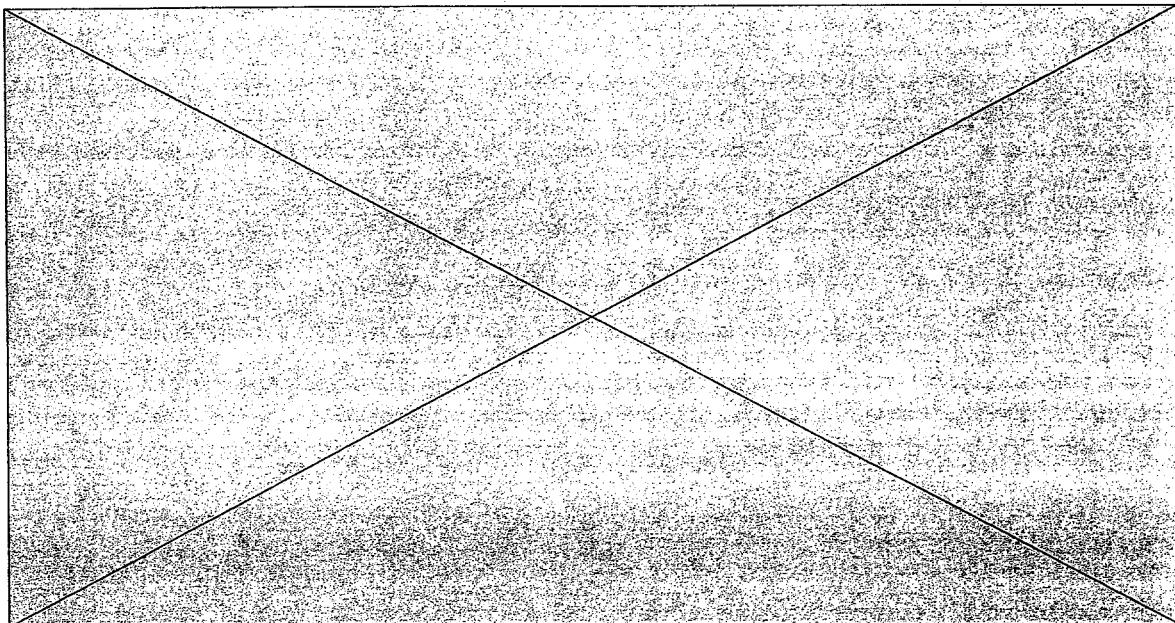


Figure 6 Theoretical number of plates for RE fragments (< 147-bp) with increases in column temperature

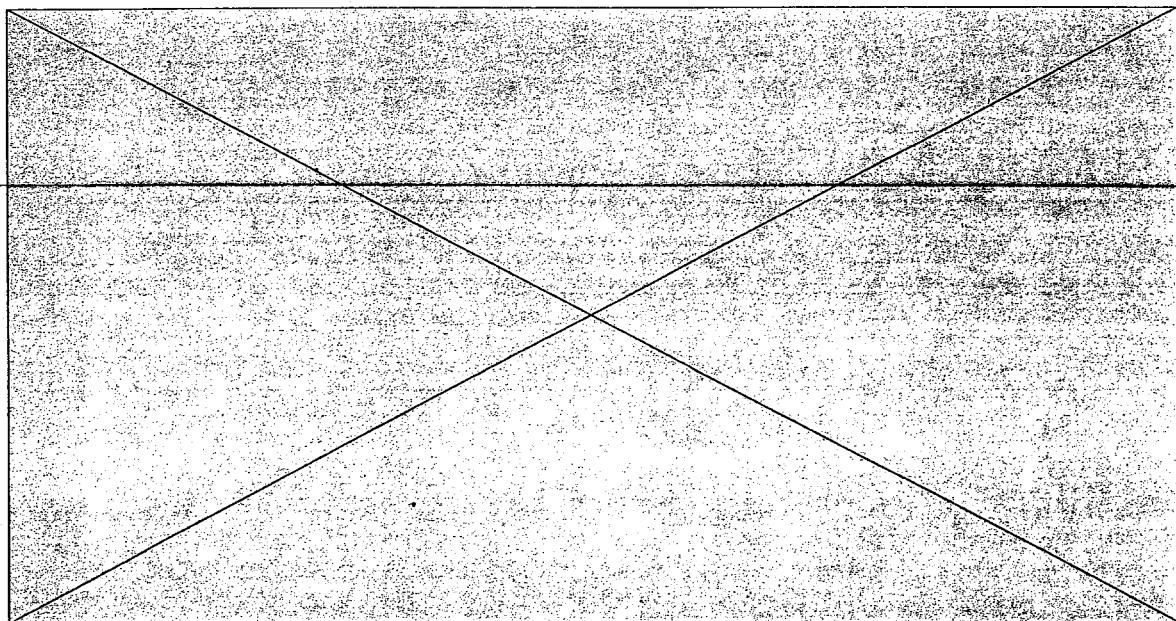


Figure 7 Theoretical number of plates for RE fragments (> 184-bp) with increases in column temperature

The purpose of this study is to develop a HPLC method to size/type PCR-amplified STR fragments. Therefore, the effects of column temperature needed to be examined on actual PCR samples (table 11). All of the samples used for this part of the study had been previously typed using gel electrophoresis and sequenced to assure proper allele designation.

Table 11 STR alleles used for column temperature, flow rate, and gradient optimization studies

Alleles	Allele sizes (bp)	Δ bp
6, 9	158, 170	12
8, 10	166, 174	8
9, 10	170, 174	4

One of the first effects of raising column temperature on the separation of dsDNA has already been discussed. The retention time is longer for PCR-amplified dsDNA with increasing column temperature (figure 8). There did not appear to be any melting of the fragments as signified by shortening of retention time until they were subjected to a column temperature of 57°C.

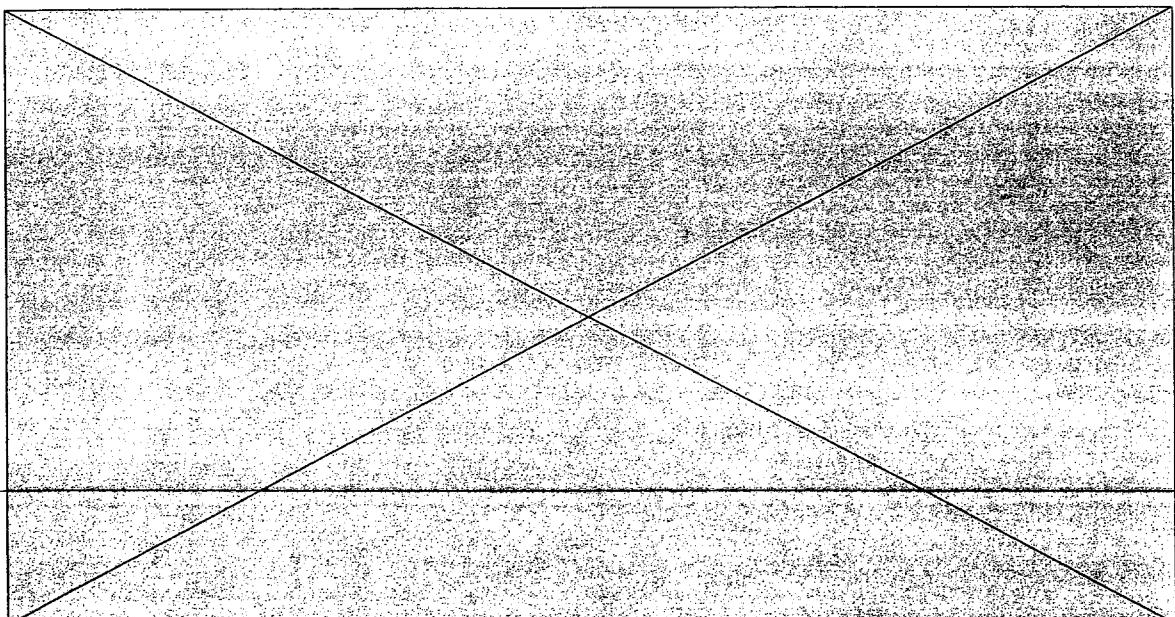
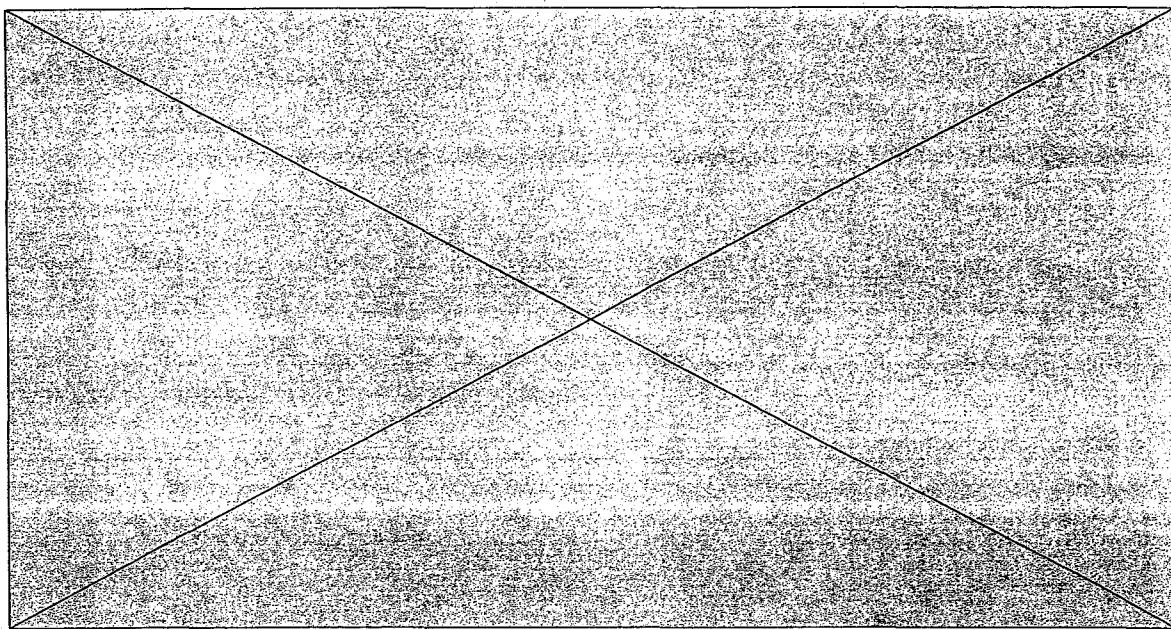


Figure 8 Retention time for HUMTHO1 PCR products with increasing column temperature

The effect of column temperature on the resolution for PCR-amplified HUMTHO1 is of great interest for the sizing/typing assay. The smallest repeat unit for all the STR systems in this study is four bp. However, in the HUMTHO1 locus there is a deletion (9.3 allele) that can occur reducing the number of bp between two alleles to one bp (the difference between the 9.3 and 10 alleles). With this in mind, the resolution for the HUMTHO1 sample (9, 10) should be greater than 1.5 because this signifies baseline separation. The resolution would need to be that high in order not to have co-elution of any samples with the allele types of 9, 9.3, or 10.

The changing of the column temperature and its effect on resolution for HUMTHO1 products are shown in figure 9. The resolution for all samples increased until reaching a maximum at 54°C. At 55°C, the resolution dropped a significant amount. The resolution between the (9, 10) sample was 1.41 at a column temperature of 54°C. The maximum resolution needed for this sample was 1.5, although, with adjustments in flow rate and steepness of gradient, a resolution of 1.5 should be attainable. These results agree with the effect of column temperature on the resolution



for RE fragments.

Figure 9 Resolution for HUMTHO1 PCR products with increasing column temperature

The theoretical number of plates (N) for HUMTHO1 products was examined while increasing the column temperature (figure 10). The formula used to calculate N can be found in equation 2. As the temperature increased from 50°C to 51°C, the number of theoretical plates stayed the same or increased for all fragment pairs. Then, from 52°C to 53°C, the value of N decreased until 53°C where N reversed its descent and began to increase until it reached at maximum at 54°C. These data correlate with the resolution which also reaches a maximum at 54°C. A balance between the peak width and the retention time is reached at 54°C, creating a high number of theoretical plates for the allele peaks. The peak widths for the HUMTHO1 (9.3, 9.3) fragment can be seen in table 12. Even with the increases in retention time, the peak width decreases, lowering the value of N. The peak width at 54°C is the lowest, leading to a high value for N. In addition, the peak width measurements display lower values at lower column temperatures. This is the reason that the N values are high at 50°C.

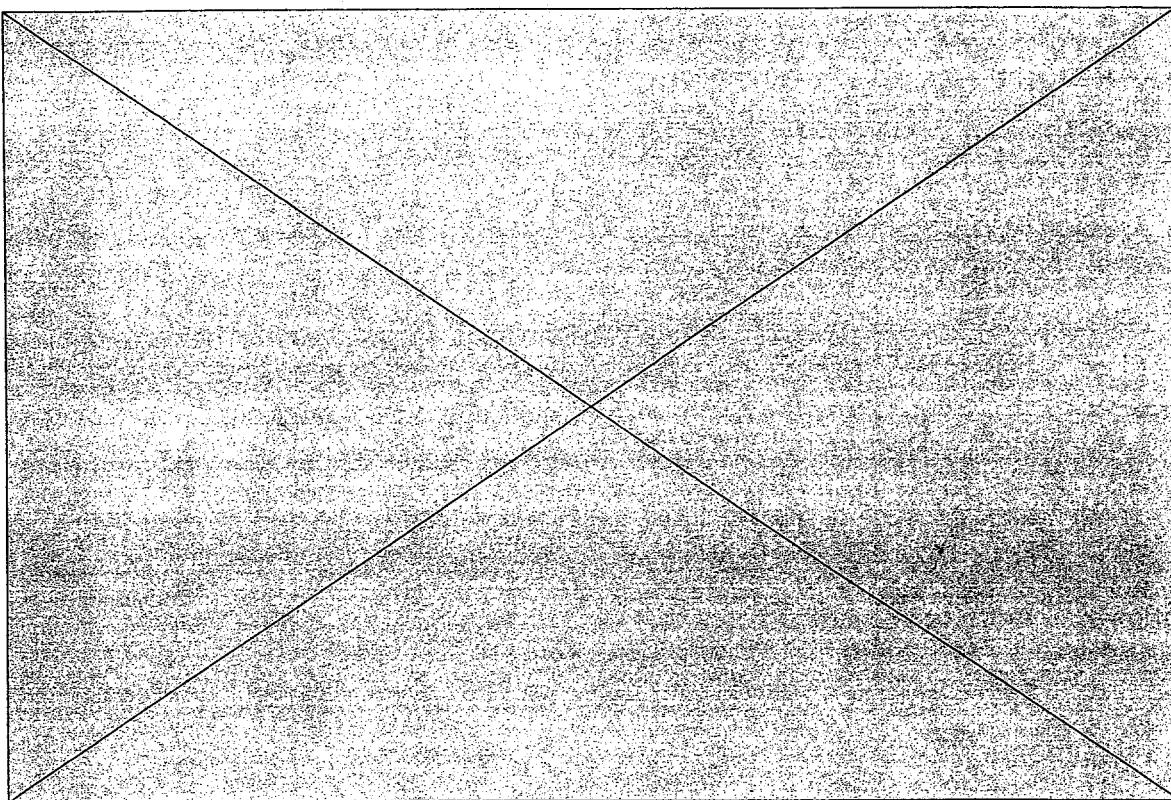


Figure 10 Number of theoretical plates for HUMTHO1 PCR products with increasing column temperature

Table 12 Peak widths for HUMTHO1 allele (9.3, 9.3) with increasing column temperature

Column Temperature (°C)	Peak Width (min)
50	0.131
51	0.129
52	0.116
53	0.118
54	0.117
55	0.133

The selectivity factor (α) was calculated for the HUMTHO1 products with increases in column

temperature . The α -value is calculated using the following equation

$$(3) \quad \alpha = k'_{\text{2}} / k'_{\text{1}}$$

where the k'_{2} is the capacity factor for the peak of interest and k'_{1} is the capacity factor for peak adjacent to the peak of interest. The α -value is a term that describes the separation of two adjacent peaks and the capacity factor is a direct measure of the strength of the interaction of the sample with the packing material. The α -value does not take into account the effects of peak width and band broadening. However, if one of the two alleles were undergoing any partial melting leading to a retention time shift, the α -value would be affected. Therefore, the α -value is useful as a method to check for any partial melting of the allele peaks.

The allele pairs from the resolution study were used for examining the effects of column temperature on the α -values (figure 11). The α -values remain constant for all pairs of alleles until the column temperature reaches 58°C. From 58°C to 59°C, the 9 and 10 alleles co-elute as one peak, and the α -value becomes zero. This is followed by the 6/9 and 8/10 alleles co-eluting at a column temperature of 60°C with α -values of zero. All the results for both the RE fragments and the HUMTHO1 alleles point to a column temperature of 54°C as the optimum column temperature for the sizing and typing experiments. At this column temperature, no partial melting of any of the RE fragments that will be used as markers took place. In addition, the HUMTHO1 alleles did not melt at a column temperature of 54°C. With a column temperature found for the analysis of STR products, the next step is to find the proper flow rate for optimum analysis.

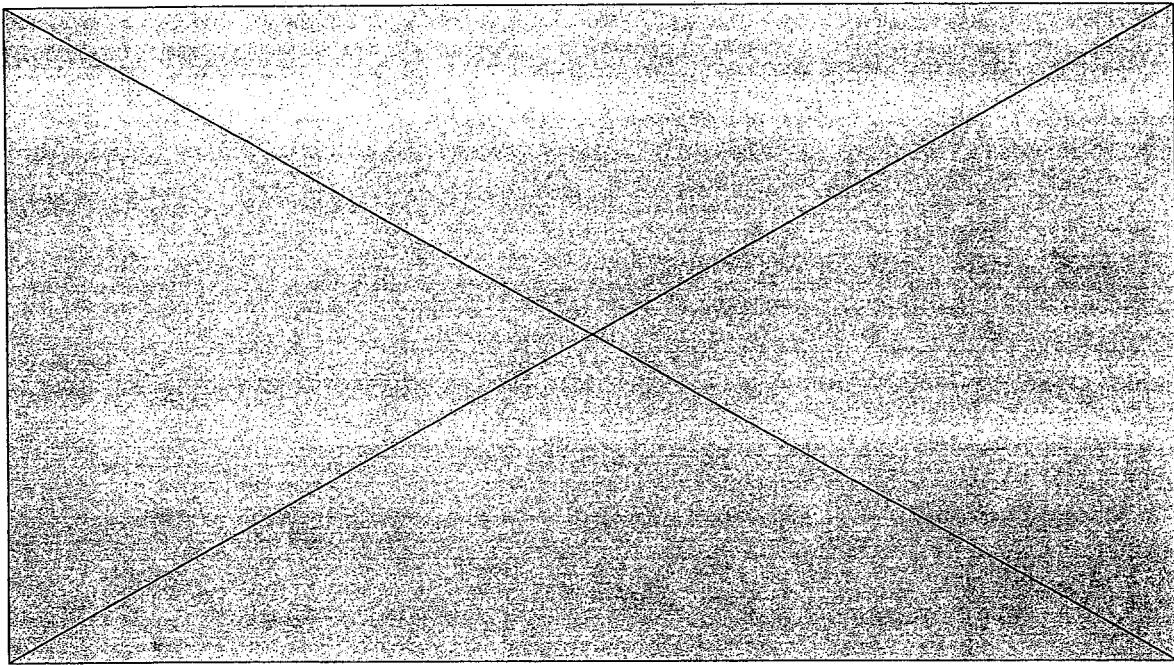


Figure 11 Selectivity factor for HUMTHO1 PCR products with increasing column temperature

Effects of Flow Rate

In liquid chromatographic separations, mobile phase flow results from the generation of a pressure gradient along the stationary phase. Because of this, the column head pressure is always greater than atmospheric pressure, and varies with the volumetric flow rate, mobile phase composition, and temperature. The pressure (or flow rate) is not commonly considered an important chromatographic parameter because the small changes in viscosity of the solute and the changes in the density of liquid induced by pressure over the modest pressure range (< 350 bar) employed in liquid chromatography are not thought to have any effect on the separation of analytes.^{72, 73}

Theoretical studies have considered the role of pressure-induced changes in the viscosity and density of the solvent as well as the unified approach to solute retention that includes the variable of pressure.^{74, 75} In both of these cases, the effect of pressure on solute retention was predicted to be negligible under reverse-phase conditions. However, two experimental studies demonstrate that pressure can affect solute retention when the mobile phase is polar.⁷⁶⁻⁷⁸ Thus, even though the bulk properties of polar mobile phases are not significantly affected under typical pressure conditions as predicted by Martin et al. and Martire and Boehm, chromatographic retention is disturbed possibly through shifts in the solute ionization state and solute equilibrium between mobile phase and stationary phase.^{74, 75}

The chromatography used in this study has a mobile phase that contains polar compounds. Consequently, flow rate should effect the separation of dsDNA. From earlier work dealing with the separation of RE fragments, an increase in flow rate exerted a pronounced increase in resolution between a 162 bp and 210-bp fragments.⁴⁸ However, STR fragments can differ by as little as 4 bp. Therefore, the effect of flow rate on the separation of fragments, in the range of bp sizes 51-458-bp that differ in length by as little as 4-bp needed, to be examined.

The purpose of the first part of this study was to analyze RE ladders at flow rates (ml/min) of 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 and 1.1 to see the affect on retention time, resolution, and N. The temperature for all analyses was held constant at 54°C based on the column temperature study described earlier in this chapter. The percentage increase of ACN per minute was two (2% ACN per min), and was linear in nature. The DNA fragments analyzed were 51-458-bp in length to match the possible sizes of STR loci. In addition,

the resolution between fragments differing by as little as six bp was examined with increases in flow rate.

One of the first effects of increasing flow rate is a reduction in the retention time for RE fragments (figure 12). The DNA fragments move rapidly with the faster moving solvent. Therefore, it is not surprising that the retention time decreased as the flow rate was increased.

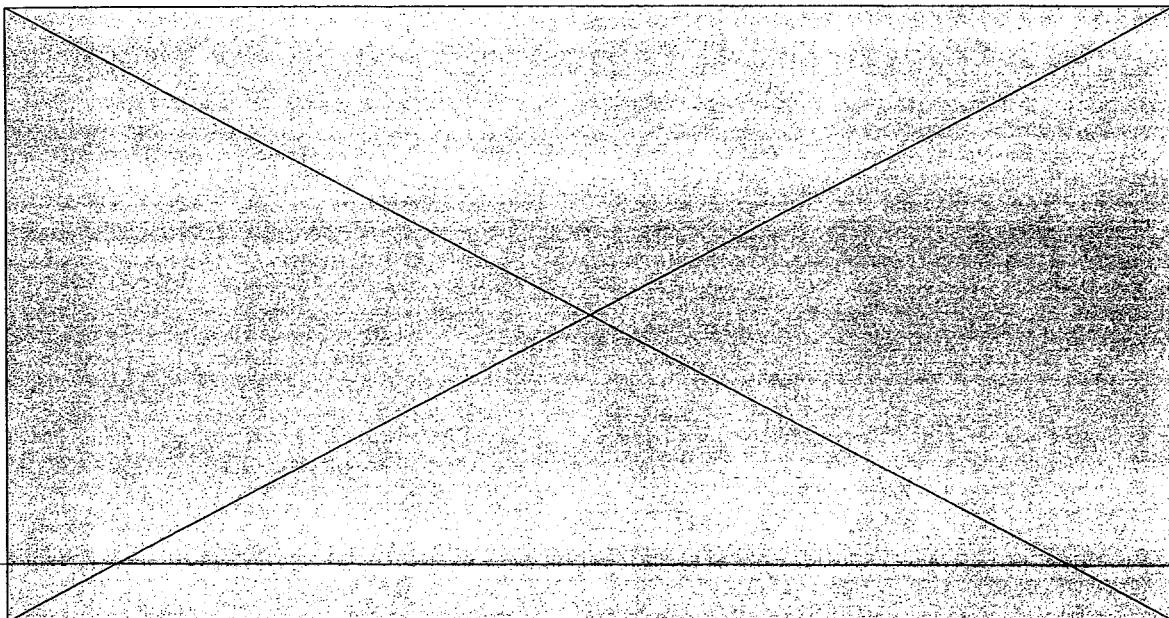


Figure 12 Retention time for RE fragments with increasing flow rate

Hirabayashi and Kasai¹⁰³ showed that DNA separations are affected by flow rate because the higher the flow rate, the more extended the dsDNA fragment. Problems will arise in this size based separation because of reduction in interaction of the DNA analyte with the stationary phase.⁷¹ This is confirmed by a paper that measured resolution at different flow rates (0.3 – 1.0 ml/min) for the separation of dsDNA fragments. Lower flow rates resulted in better resolution for fragments less than 100 bp in size.⁷¹ The results we obtained when increasing flow rate confirmed the ideas of Hirabayashi and Kasai.¹⁰³ The higher the flow rates, the fewer interactions with the stationary phase,⁷¹ and this leads to a reduction in resolution. Since the dsDNA will not interact with the stationary phase based on its size, but rather the interaction will be a purely physical interaction that is reduced in time by increases in flow rate. Logically, the lower the flow rate, the more each dsDNA fragment will interact with the stationary phase, causing the

separation to be size based. Therefore, the lower the flow rate the, better the separation (increase in resolution). However, the total time for the analysis will be increased with lower flow rates.

To check the effects of flow rate on resolution, RE fragment pairs with the following bp sizes were analyzed: 51-57 ($\Delta = 6$ bp), 57-64 ($\Delta = 7$ bp), 184-192 ($\Delta = 8$ bp), 213-234 ($\Delta = 21$ bp), 257-267 ($\Delta = 10$ bp), 267-298 ($\Delta = 31$ bp), and 434-458 ($\Delta = 24$ bp). The resolution was calculated using equation 1.

The results for RE fragments

< 64 bp are shown in figure 13, and the results for RE fragments > 184 bp are shown in figure 14. As can be seen in the figures, 0.8 ml/min is the flow rate that gave the best resolution for all pairs of fragments.

Table 13 shows the resolution obtained at 0.8 ml/min for all fragment pairs. Again, there is a large difference in resolution based on the fragment sizes. This data further confirms the need for all alleles to be sized to be the smallest bp length possible.

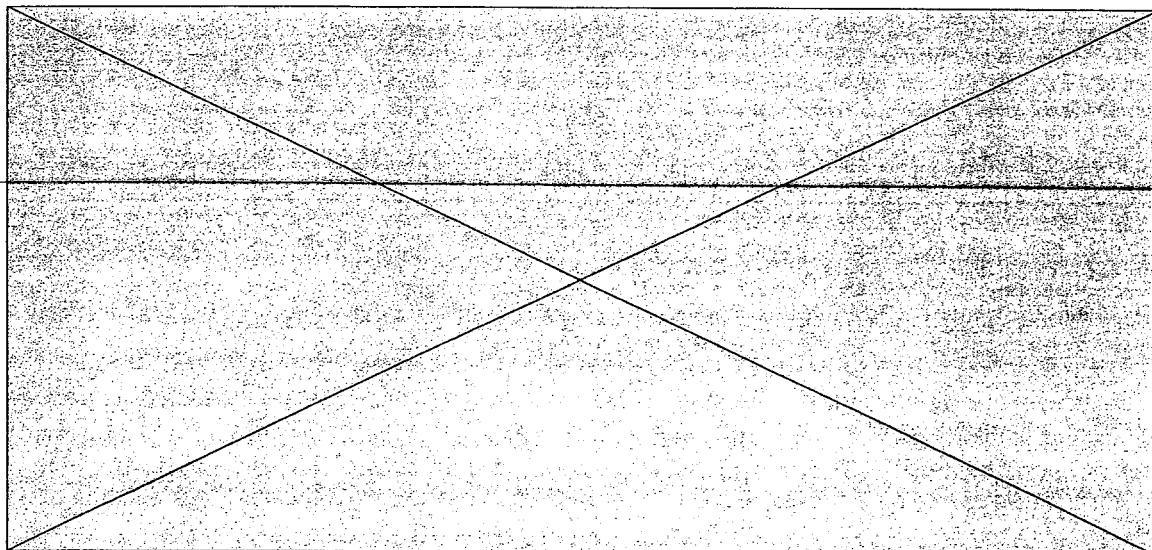


Figure 13 Resolution for RE fragments (< 64-bp) with increasing flow rate

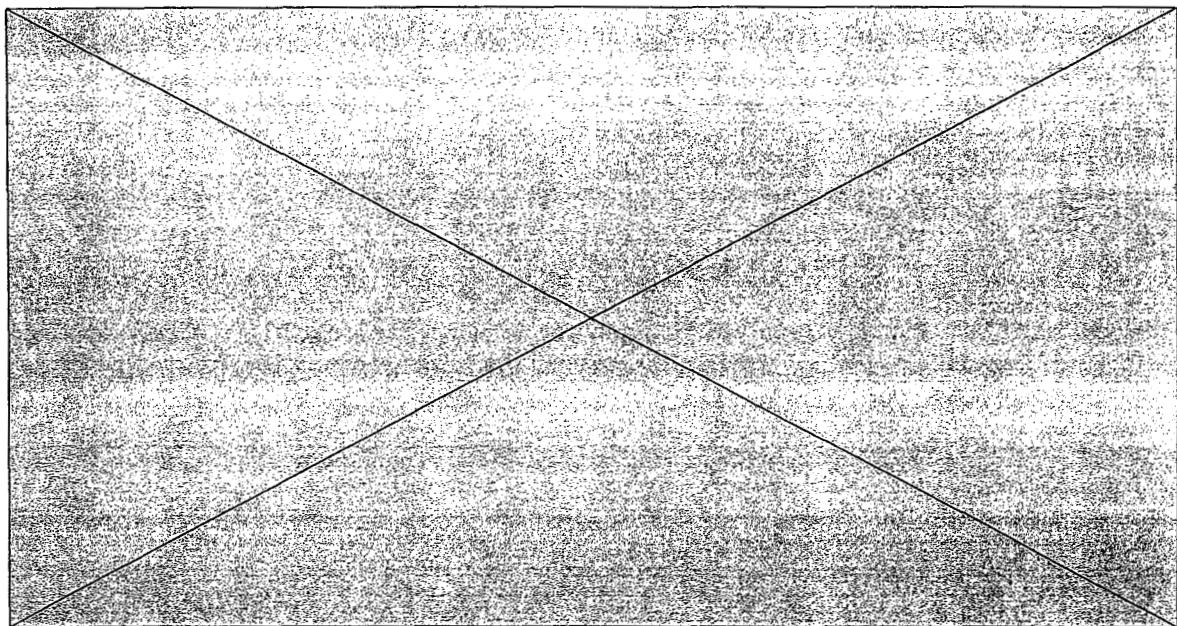


Figure 14 Resolution for RE fragments (>184-bp) with increasing flow rate

Table 13 Resolution for RE fragments at a flow rate of 0.8 ml/min

Fragment sizes (bp)	Δ BP	Resolution at flow rate of 0.8 ml/min
51, 57	6	10.93
57, 64	7	8.93
184, 192	8	1.98
213, 234	21	4.23
257, 267	10	2.01
267, 298	31	4.44

434, 458	34	1.73
----------	----	------

The effect of flow rate on the theoretical number of plates (N) for RE fragments was examined (figure 15). As the flow rate of 0.5 ml/min increases to a flow rate of 0.7 ml/min, the retention time decreases (shown in figure 12). The retention time at 0.5 ml/min is part of the reason that the value of N is high. However, the figure shows that N has a second maximum at a flow rate of 0.8 ml/min. This is in agreement with the study of the effect of flow rate on resolution. Following the maximum N value at 0.8 ml/min, N drops steadily for all fragments except the for the 458-bp fragment. The 458-bp N value reaches a third maximum at a flow rate value of 1.1 ml/min, but this is the only fragment that has a third maximum.

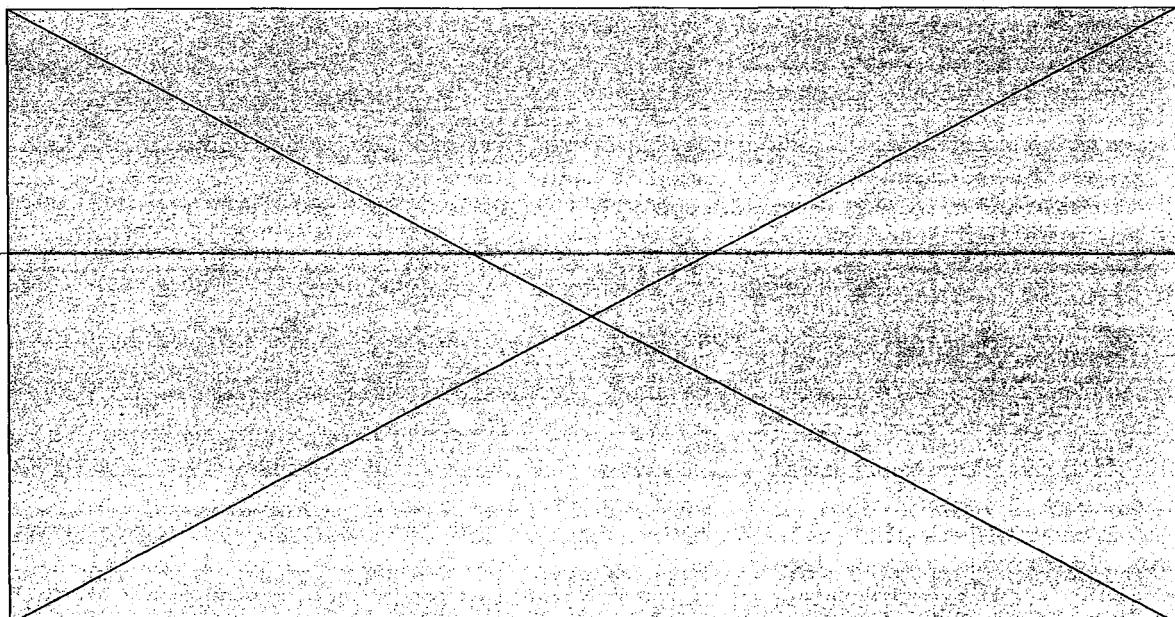


Figure 15 Theoretical number of plates for RE fragments with increasing flow rate

The results of the resolution and N value studies demonstrated that the flow rate for the RE fragments that gave the best results was 0.8 ml/min. All studies for the effects of flow rate changes were repeated on STR PCR-amplified fragments.

For STR fragments, the flow rate was examined from 0.5 ml/min to 1.2 ml/min in increasing steps of 0.1 ml/min. The conditions of separation included a column temperature of 54°C and an increase in the

percentage of ACN per minute of two (2% ACN per min). HUMTHO1 alleles were used for the flow rate study. These PCR amplified samples are in table 11.

The first effect on the separation of the HUMTHO1 alleles was a decrease in retention time as the flow rate increased (figure 16). This agrees with the results attained for the RE fragments.

Figure 17 is the result of resolution calculations using equation 1 for the HUMTHO1 PCR products with increases in the flow rate. From the figure, the resolution for all alleles reaches a maximum at a flow rate of 0.8 ml/min. For the HUMTHO1 (9, 10) alleles, a resolution of 1.73 was attained. This resolution (≥ 1.5), according to the literature, allows for a degree overlap of less than 1%, and mixtures with concentrations as high as 10:1 could easily be distinguished.⁷⁹ This is very important when ladder spiking experiments will be done. It is possible that the mixture for the typing assay will be even greater than 10:1.

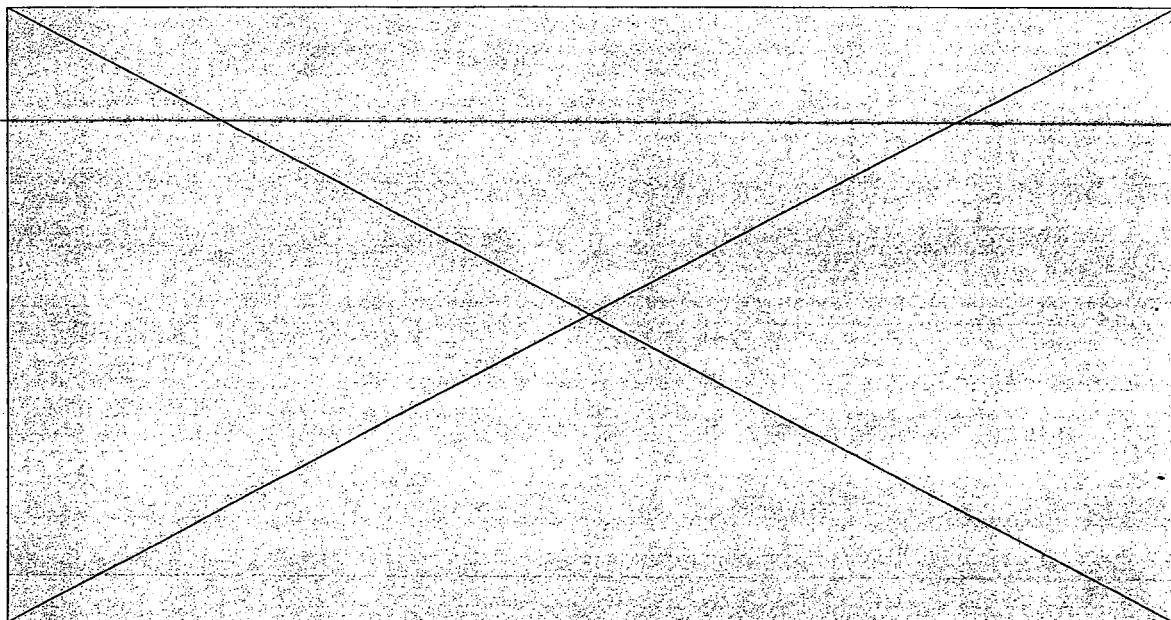


Figure 16 Retention time for HUMTHO1 PCR products with increasing flow rate

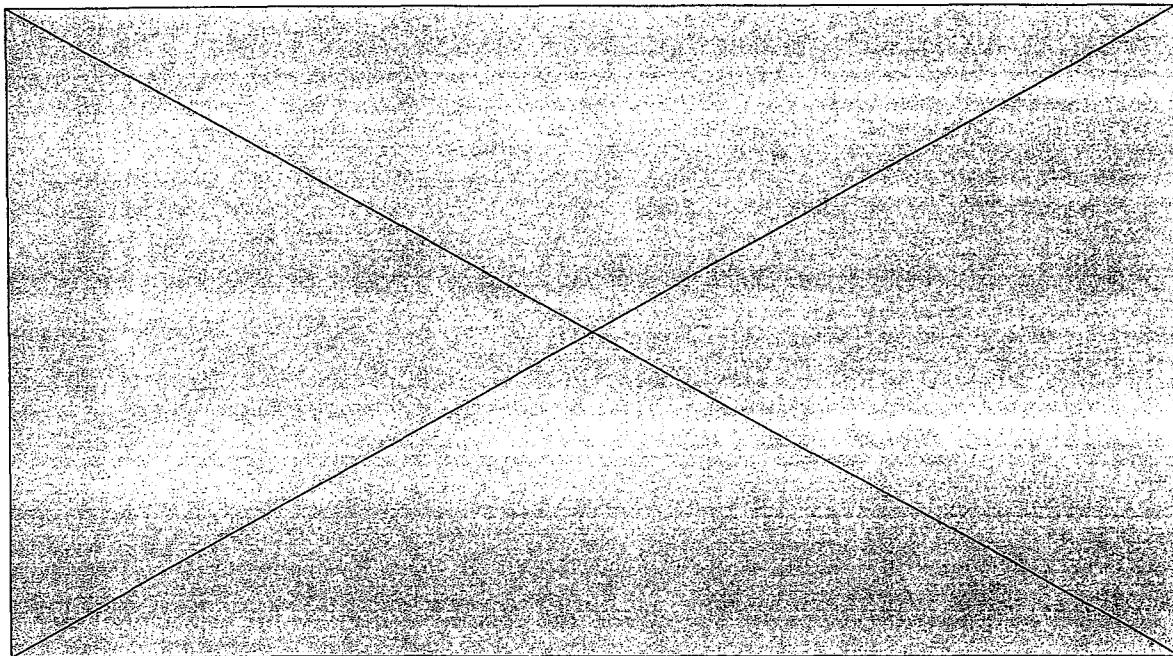


Figure 17 Resolution for HUMTHO1 PCR products with increasing flow rate

The effect of flow rate on the theoretical number of plates (N) for HUMTHO1 alleles is shown in figure 18. In the calculation of N (equation 3), the peak width and retention time are in the equation. For the HUMTHO1 PCR products, the retention time is greatly affected by the flow rate as displayed by figure 16. From the retention time for all fragments at 0.5 ml/min to the retention time of all fragments at 1.2 ml/min, there is an average time difference of 6.18 minutes. This retention time shift will affect the calculation of N . The peak width (min) that is reported in table 14 for the HUMTHO1 allele 8 (166-bp) is used to calculate N and it has a low value at a flow rate of 0.8 ml/min. This, coupled with the longer retention time of the 166-bp sample at a flow rate of 0.8 ml/min, is the reason for N reaching a maximum at this flow rate.

Table 14 Effect of flow rate (ml/min) on the peak width for the HUMTHO1 allele 8

Flow rate (ml/min)	Peak Width (min)
0.5	0.254
0.6	0.209
0.7	0.157
0.8	0.160

0.9	0.164
1.0	0.171
1.1	0.165
1.2	0.169

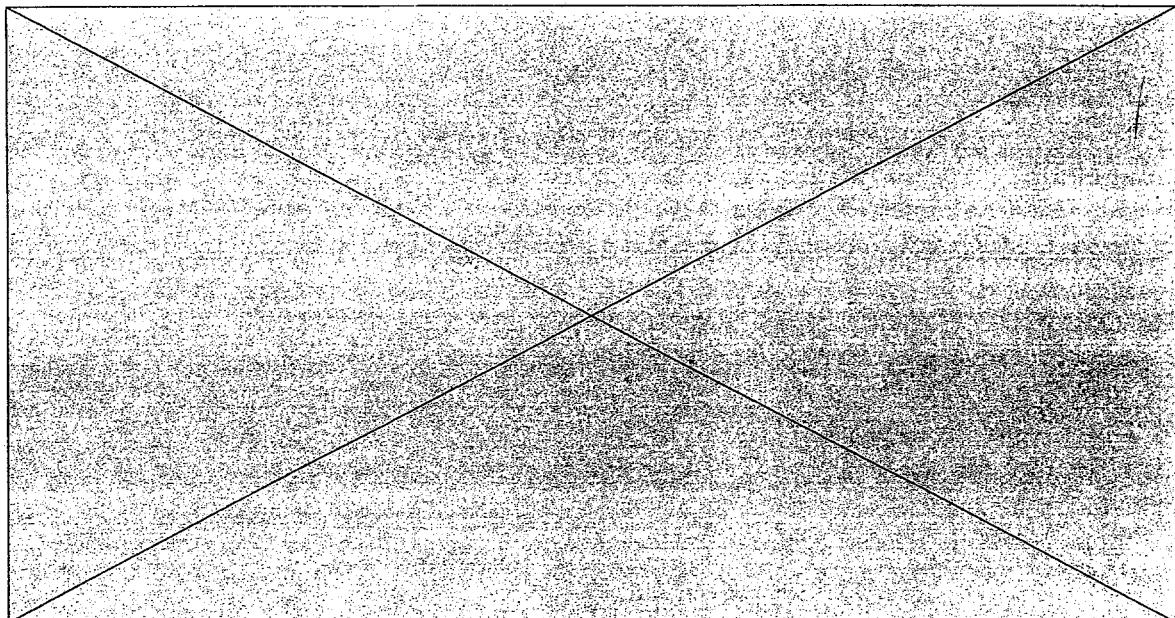


Figure 18 Theoretical number of plates (N) for HUMTHO1 PCR products with increasing flow rate

The flow rate for the analysis of STR PCR products found a maximum for N at a flow rate of 0.8 ml/min. The flow rate, coupled with the column temperature, was then used to find the optimal percentage of ACN per minute for analysis of dsDNA using the DNAsep™ column.

Effects of the Slope of the Acetonitrile Gradient

The term reverse-phase liquid (RP) chromatography is derived from the fact that the mobile phase is polar, and the stationary phase is non-polar. Typically, RP chromatography employs a hydrophobic phase covalently bound to a silica gel or a polymeric support and a hydrophilic mobile phase such as a mixture of water and acetonitrile for separation of solutes.

RP liquid chromatography is the most popular mode of chromatography for the analytical separations used in biological and biomedical science applications.⁷⁹ The popularity exists in part because

an aqueous mobile phase is used, and therefore aqueous samples can be injected directly without pre-treatment. Another reason for the popularity is the wide range of polarity that analytes can be separated by this technique.

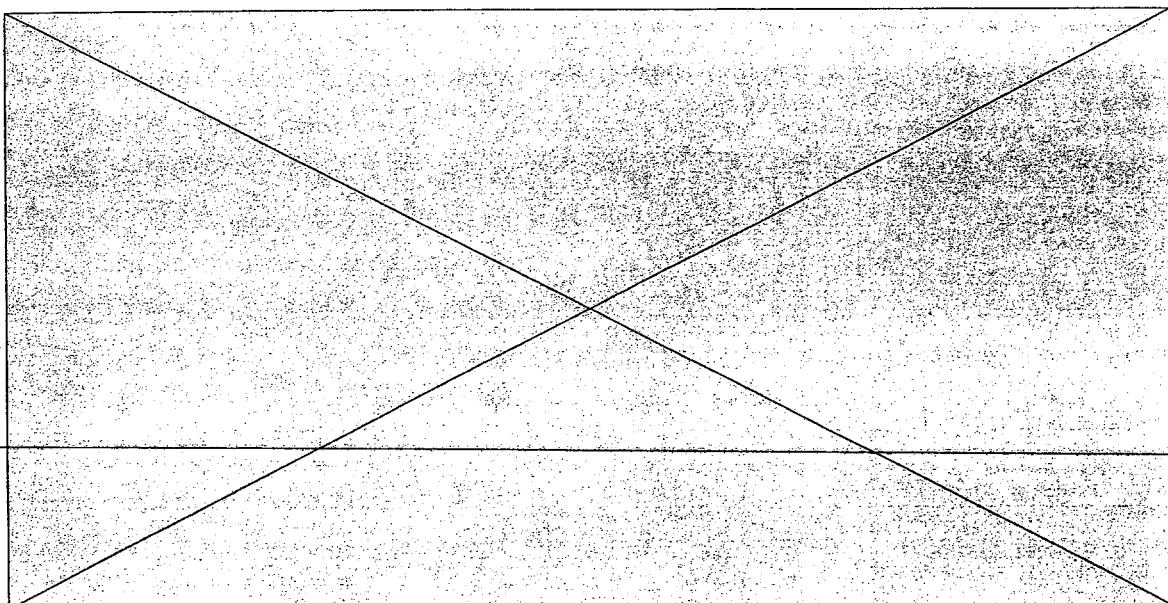
The reason for the retention of analytes on the column in RPLC is the hydrophobic effect.⁸⁰ The retention by the stationary phase arises from the repulsion of non-polar (hydrophobic) regions of the solute molecules by the water molecules in the mobile phase. Hydrophobic interactions are mediated by the specific solute-solvent interactions in the mobile phase and the stationary phase. The retention in RP is a result of the interaction of the solute with the mobile phase because molecules of the organic modifier added to the mobile phase saturate the stationary phase. For example, strong modifiers, such as acetonitrile, will saturate the surface of the stationary phase at concentrations of less than 10%.

Using RP with an ion-pairing agent such as TEAA has become a popular method for controlling the retention of ionic compounds. This type of chromatography has two mechanisms proposed for the retention of solutes. The first mechanism proposed is that the ion-pairs are formed in the mobile phase between the ion-pairing agent and the solute, and are distributed between the mobile phase and the stationary phase.⁷⁹ The second mechanism is a dynamic ion-exchange model that takes place with a monolayer of ion-pairing agent absorbed onto the surface of the stationary phase, and the solute interacting with the absorbed ion-pairing agent.⁷⁹ Both mechanisms have mathematical expressions that are similar and have been proven by modeling the relationship between retention (k') and the concentration of the ion-pairing agent.

The optimization of IPRP liquid chromatography involves the relationship of the effects of pH, organic modifier type, organic modifier concentration, and the effects of the type of IP agent. For the separation of dsDNA using the DNAsep™ column, the pH, organic modifier type, concentration of organic modifier, and the IP agent have been optimized. However, the effect of the organic modifier concentration per minute has not been examined. The organic modifier concentration can control the separation based on the mechanisms proposed for IPRP separation.

For the analysis of dsDNA, the rate of increase in % ACN per minute was changed, while the same chromatographic conditions were used. The flow rate was 0.8 ml/min, the column temperature was

54°C, and the gradient conditions were linear. The changes in % ACN per min were 0.5, 1.0, 1.5, 2.0, and 3.0. The first effect of changing the % ACN per minute is a change in retention time (figure 19). The time dsDNA spends on the column is increased with a decrease in % ACN per min. The average change in retention time from 3.0 % ACN per min to 0.5% per min was increased to 18.55 min. Adjusting the starting % ACN in the mobile phase could shorten the time of analysis. However, this would influence the theoretical number of plates (N) for analysis because band broadening would take place at the lower flow rates.



To examine the effects of changing the % ACN per min on resolution (R_s), the HUMTHO1 6, 9 alleles and the 184-192 bp (Δ bp = 8) RE fragments were analyzed using the conditions discussed above (figure 20). The R_s increased as the % ACN per min decreased. This correlates well with the difference in retention time as the % ACN per min changes. At the lower values of % ACN per min, the dsDNA fragments have more time to interact with the stationary phase and allow for maximum IP agent interaction. However, time becomes a factor. The analysis at 0.5 % ACN per min gave the highest resolution values but the total separation time was over 25 min, whereas with 1.0 % ACN per min, the total analysis took just over 15 min. With this in mind, the 1.0 % ACN per min will be used for all dsDNA analysis.

Figure 19 Retention time for HUMTHO1 PCR products and RE fragments with increasing % ACN per minute

In addition, the theoretical number of plates (N) for the HUMTHO1 products and the RE fragments were compared using the differing % ACN per min. Figure 21 is comparison of N with increasing % ACN per min. N increases with increases in the % ACN per min. The higher the % of ACN per min, the less band broadening and the sharper the peak. In addition, N is calculated using the retention time (equation 2). The retention time for the 0.5% ACN per min is ~ 25 min, leading to a small value for N, while the retention time for the 3.0% ACN per minute is > 10 min.

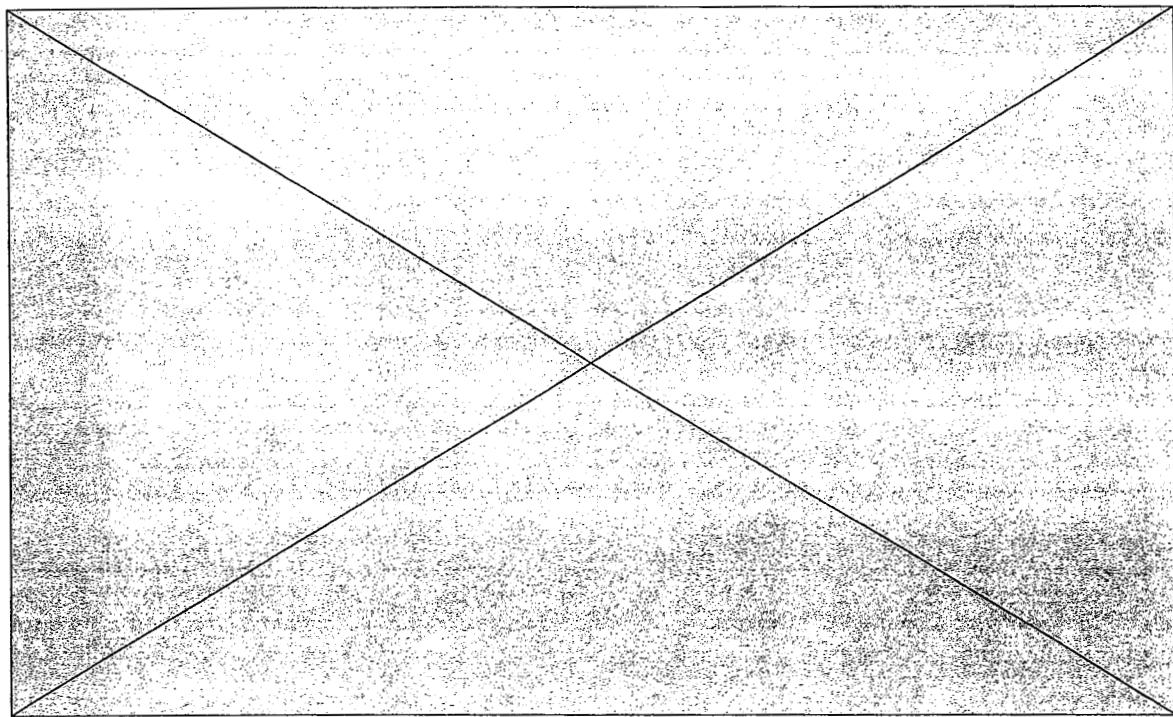


Figure 20 Resolution for HUMTHO1 PCR product and RE fragments with increasing % ACN per minute

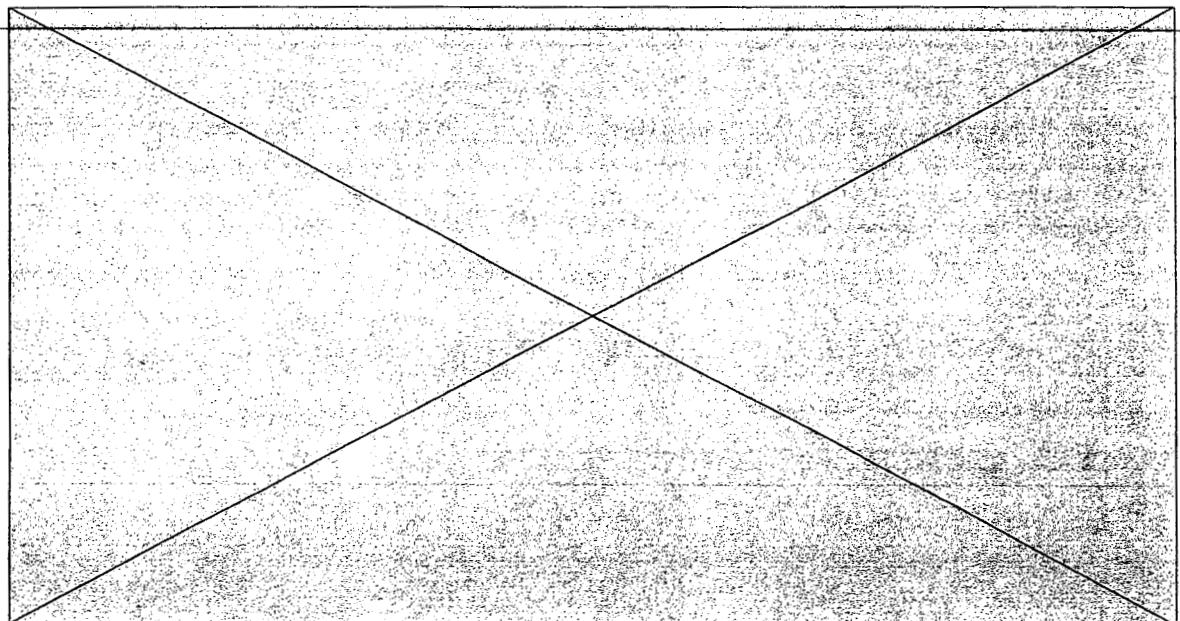


Figure 21 Theoretical number of plates for HUMTHO1 PCR products and RE fragments with increasing % ACN per minute

Conclusion

The optimal conditions for the analysis of PCR-amplified and RE fragments were a column temperature of 54°C, a flow rate of 0.8 ml/min, and a rate of increase in % ACN of 1.0 per minute.

The actual starting and ending points for a linear gradient are calculated based on the size of dsDNA that is to be separated.

3. Sizing and typing of PCR-amplified dsDNA using HPLC

Discussion

Separation Conditions

The optimal conditions for the analysis of PCR-amplified and RE fragments were a column temperature of 54°C, a flow rate of 0.8 ml/min, and a % ACN of 1.0 per minute. The starting and ending points for the linear gradient are calculated based on the size of dsDNA that is to be separated.

Sizing of the HUMTHO1 Locus Using RE Markers

Linear Least Squares Analysis of Four RE Markers to Size HUMTHO1 Alleles

To measure the length (bp) of a DNA fragment in gel electrophoresis, a relationship is established between the mobility of an unknown fragment and the mobility of fragments of known length. With gel electrophoresis, problems arise in the fact that gel inhomogeneities exist. In addition, temperature and voltage gradients across the gel affect the electrophoretic mobility of dsDNA fragments, and the sequence of dsDNA affects the electrophoretic mobility. However, the idea of using the mobility of dsDNA

fragments of known length to size unknown fragments has become a popular method to size dsDNA.

The dsDNA fragment bp length affects the elution time in HPLC. Therefore, a sizing standard can be applied to HPLC separation for the purpose of sizing dsDNA.

The first method examined for sizing of PCR products was the use of RE fragments as markers. The fragment size and the sequence of the marker (fragment) could have an effect on the sizing accuracy. Therefore, different sized markers with different sequences were tested for the accuracy of sizing dsDNA HUMTHO1 PCR products of known bp length. The markers that allowed for the highest accuracy in sizing would be used to size a group of unknown samples.

The method for the sizing of PCR products using RE fragments begins by using the retention times of the RE fragment to draw a line of regression. With this line of regression, a retention time of an unknown peak could be interpolated with the equation of the line between the standards to find the number of bps in the unknown peak. In the case of the HUMTHO1 locus, all alleles for the United States population fall between 154-bp and 174-bp. Therefore, the markers used for the sizing of the HUMTHO1 locus should be < 154-bp and > 174-bp. The RE ladders used for the experiments in chapter 2 contain fragments that meet this requirement.

All RE fragments used for the sizing study were HPLC purified using the Dionex DX 500 system.

The fragments were collected and re-analyzed to assure that the correct fragment was isolated. The fragments were used in their collected states.

In the first set of experiments, fragments from the plasmid pBR322 cleaved with the RE Hae III were used as size markers. The fragments had bp sizes and %AT of 123 (43.9%), 124 (34.7%), 184 (40.8%) and 192 (52.6%), respectively. The fragments were chosen from one RE-cleaved ladder and were close to the unknown peaks in bp size as recommended in Huber et al.⁴⁹ and Elder and Southern.⁸⁷

The results of the sizing experiments using the above markers are listed in table 15 and an example chromatogram is shown in figure 22. The table contains data consisting of three analyses for five different alleles for the HUMTHO1 locus. The retention times of the RE fragments (y variable) were plotted against the number of bps for each RE fragment (x variable). Least squares analysis of the RE fragment bp size and retention time produced an equation that allowed the calculation of bp size from

retention times of PCR products. The bp size generated for the HUMTHO1 alleles was on the average 6.40 bp larger than the actual bp size of the allele. Upon examination of the regression equation, the adjusted R^2 value was 0.99 for all the regression lines for the four fragments, meaning that the least squares equation eliminates 99% of the error in predicting the dependent variable. These results show that the AT content of a fragment does not affect the separation. If the prediction equation was affected by AT content, it would not eliminate that much of the error in predicting the dependent variable. In addition, the R_s for the 123 and 124 bp fragments is 1.29, and the R_s for the 184 and 192 fragments is 2.48, meaning that the size of the fragment controls the R_s and the size of the fragments affects the sizing of the alleles.

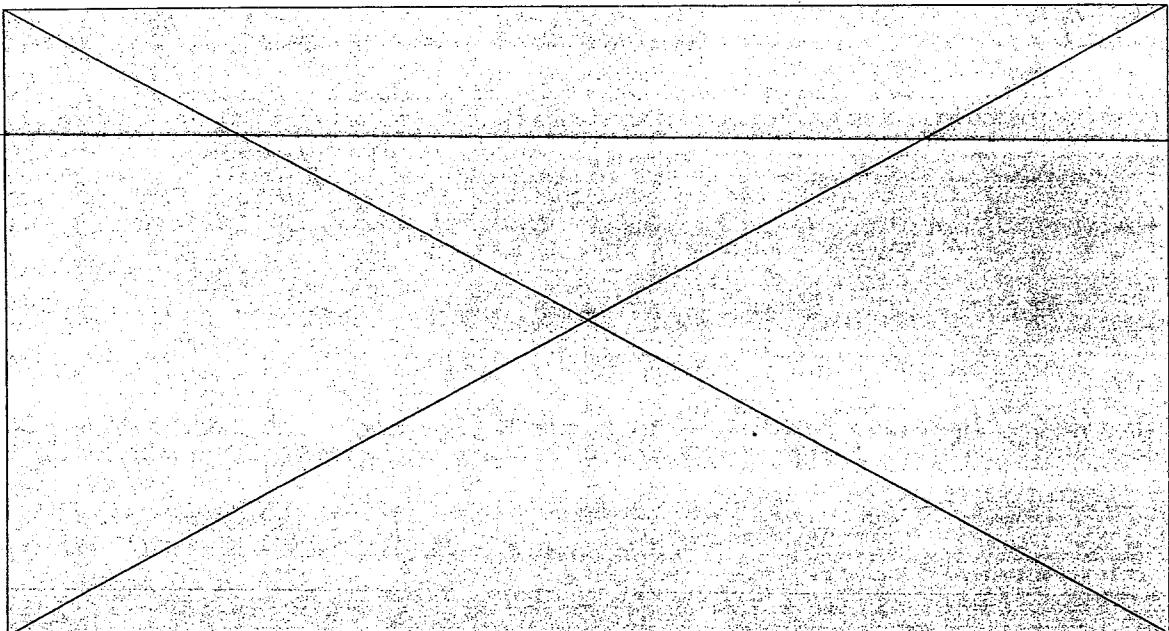


Figure 22 Separation of HUMTHO1 alleles with markers from pBR322-Hae III

The markers are 123, 124, 184, and 192-bp in length. The sample is a HUMTHO1 (9, 10) with the bp sizes calculated using least squares fit. The heteroduplex is a result of the PCR reaction. The separation conditions are as follows: column temperature of 54°C, an increase in ACN of 1%, and a flow rate of 0.8ml/min. The gradient started at 43% B and ended at 57% B.

Table 15 Sizing of HUMTHO1 alleles using RE fragments of bp size 123, 124, 184, and 192

HUMTHO1 allele	Allele bp size	Size using RE fragments	Δ bp from allele size and sizing method
6	158	164.87	6.87
7	162	169.09	7.09
8	166	172.78	6.78
9	170	176.00	6.00
10	174	179.27	5.27

To investigate the difference in the size of the RE fragments and the difference in the sizing results for the HUMTHO1 alleles, fragments from the pUC18-Msp I were used as sizing markers. The fragments used for size markers had bp sizes and %AT of 110 (50.0%), 147 (39.5%), 190 (49.5%) and 242 (51.32%), respectively. Each allele was sized three times using the markers. The results of these experiments are listed in table 16 and an example chromatogram is shown in figure 23.

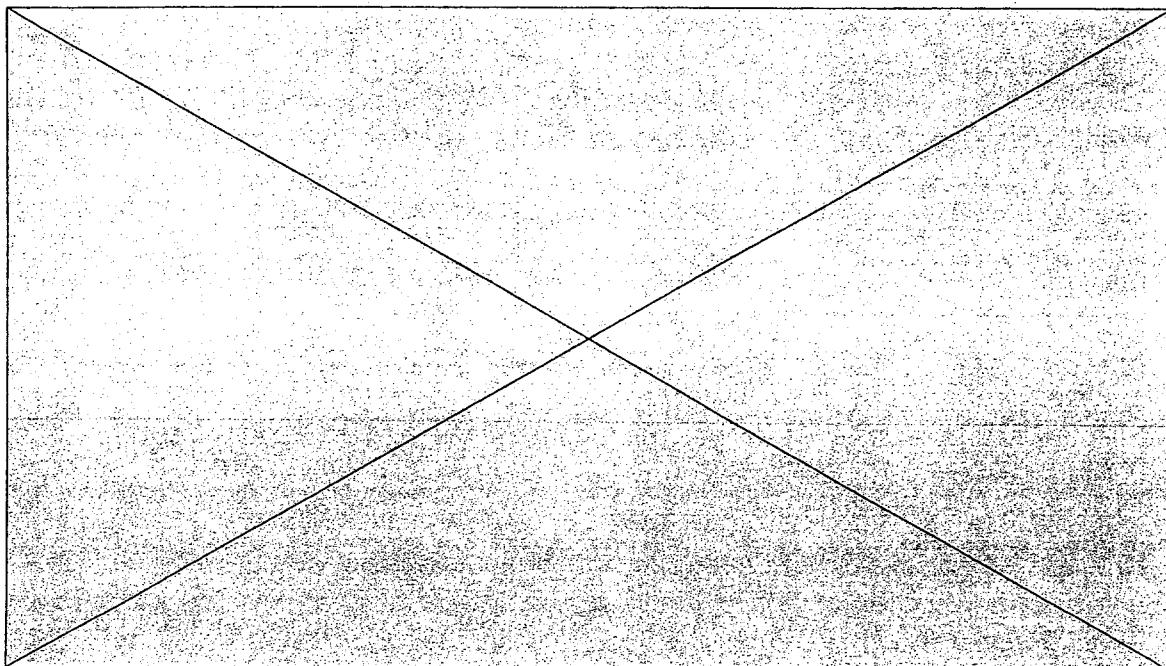


Figure 23 Separation of HUMTHO1 allele with markers from pUC18-Msp I

The markers are 110, 147, 190, and 242 -bp in length. The sample is a HUMTHO1 (9.3, 9.3) with the bp sizes calculated using least squares fit. The separation conditions are as follows: column

temperature of 54°C, an increase in ACN of 1%, and a flow rate of 0.8ml/min. The gradient started at 40% B and ended at 60% B.

The sizing values were on the average 12.2 bp larger than the actual HUMTHO1 allele bp size. The adjusted R² value for the analyses was, on the average, 0.94, meaning that 94% of the error is reduced through the use of the least squares equation to predict the dependent variable. Even with the high adjusted R² value, the calculated bp size for the HUMTHO1 alleles is larger by an average of 12.2 bp. This demonstrates that the size of the RE fragments used in the sizing experiments affects the sizing accuracy for the HUMTHO1 alleles. The difference in calculated bp size for HUMTHO1 alleles is closer in value (Δ bp = 6.40) for the four fragments from the pBR322 cleaved with the RE fragment Hae III than the calculated bp size (Δ bp = 12.22) for the HUMTHO1 alleles sized with the four fragments from the pUC18 plasmid cleaved with the RE Msp I. The fragments from the pBR322 plasmid are closer in bp size to the alleles than the fragments from the pUC18 plasmid. To obtain accurate sizing data, the markers must be as close to the PCR products as possible.

Table 16 Sizing of HUMTHO1 alleles using RE fragments of bp size 110, 147, 190, and 242

HUMTHO1 allele	Allele bp size	Size using RE fragments	Δ bp from allele size and sizing method
6	158	170.61	12.61
8	166	177.32	11.32
9	170	182.55	12.55
9.3	173	184.80	11.80
10	174	186.83	12.83

The use of four markers to generate a line of regression did not produce results that are high in accuracy. However, the results produced could be used with a correction factor in an equation that calculates the bp size of the PCR products. Other methods for sizing the alleles of STR products need to be examined.

Local Sizing of HUMTHO1 Alleles Using RE Markers

Another sizing method that uses four markers to calculate bp size was explored.⁴⁹ This method is called local dsDNA sizing.⁴⁹ The method uses dsDNA fragments that are close in bp size to the unknown fragments as markers. Using two fragments below the PCR product and one fragment above the PCR product creates a linear regression curve. The bp size is determined for the PCR product from the curve. Another curve is created using one fragment below and two fragments above the PCR product, and a second value is obtained for the bp size of the PCR product. The two size values for the PCR product are averaged to determine the bp length.

This sizing method was used to calculated bp sizes for HUMTHO1 PCR products using the 123, 124, 184, and 192-bp fragments from the pBR322 plasmid cut with the Hae III RE. A local sizing method was used with the first set of markers consisting of fragments with bp sizes of 123, 124, and 192, and a second set of markers consisting of fragments with bp sizes of 124, 184, and 192. Both of the calculated sizes using the two sets of markers were averaged to find the local sizing result. The results are located in table 17. Each allele was analyzed three times to generate an average calculated bp size. The sizing method generated bp sizes that averaged 5.82 bp larger than the actual allele size. This method produced bp sizes that were closer to the allele size than the method of using four markers to produce a linear regression curve and the utilizing of least squares analysis to generate a bp size for the PCR product. The Δ bp improved because the sizing method used markers that were closer in bp size to the unknown allele peaks. For example, the first least square analysis used the 123, 124, and 184-bp markers, and the second least squares analysis used the 124, 184, and 192 bp markers. Therefore, the sizes of the markers in each analysis were close in bp size to the unknown PCR product. With the size of the markers close to the size of the unknown peak, the calculated bp size was more accurate than using the four markers alone to generate a least squares fit.

Table 17 Local sizing method applied to HUMTHO1 alleles using RE fragments of bp size 123, 124, 184, and 192

HUMTHO1 allele	Allele bp size	Size using RE fragments	Δ bp from allele size and sizing method
6	158	164.21	6.21
7	162	168.44	6.44
8	166	172.31	6.31

9	170	175.25	5.28
10	174	178.87	4.87

Since the proximity of the bp size of the markers to the unknown allele peaks is necessary to generate high accuracy in sizing, the local sizing method was performed using the fragments from the pUC18 plasmid cut with the RE Msp I. The markers were 110, 147, 190, and 242 bp, respectively. A local sizing method was used with the first set of markers consisting of fragments with bp sizes of 110, 147, and 190, and a second set of markers consisting of fragments with bp sizes of 147, 190, and 242. Using the two sets of markers, both of the calculated sizes were averaged to find the local sizing result. The results are located in table 18. The calculated bp size was larger than the true size of the allele by an average of 3.36 bp. The calculated bp size is closer to the true size using these four markers for local sizing rather than the markers from the pBR322 plasmid for local sizing. This set of markers has two fragments (147-bp and 190-bp) close to the allele sizes (154-bp to 174-bp) affecting the local sizing method.

Table 18 Local sizing method applied to HUMTHO1 alleles using RE fragments of bp size 110, 147, 190, and 242

HUMTHO1 allele	Allele bp size	Size using RE fragments	Δ bp from allele size and sizing method
6	158	160.71	2.71
8	166	169.74	3.74
9	170	173.55	3.55
9.3	173	175.95	2.95
10	174	177.84	3.84

To examine the effects of moving the markers closer to the allele size range, fragments from both RE-cleaved plasmid ladders were used to locally size HUMTHO1 alleles. The markers were the 124, 184, and 192 bp fragments from the pBR322 plasmid cleaved with Hae III, and a 147-bp fragment from the pUC18 plasmid cleaved with Msp I. The local method used the 124, 147, and 184 bp markers to find a least squares fit to calculate a bp size for the sample. This was averaged with the calculated bp size from

the 147, 184, and 192-bp markers. The results are listed in table 19 and an example chromatogram is in figure 24.

The average difference in bp that the calculated size and the actual bp size differ is +2.74. This sizing method used the markers that were closest in bp size to the HUMTHO1 products. The results are in agreement with the premise that the markers must be as close as possible to the sample that is to be sized. The HUMTHO1 products have allele bp sizes that are constantly making the selection of markers an easy task. The sequence of the marker does not affect the sizing, but the size of the marker affects the accuracy of the sizing method.

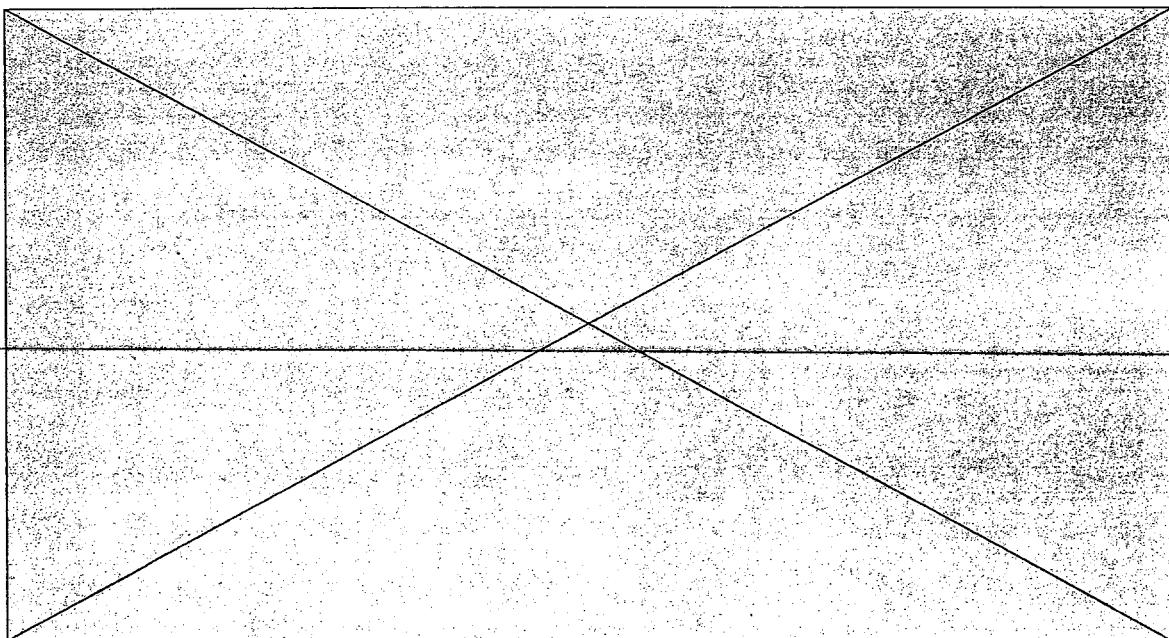


Figure 24 Local sizing method for HUMTHO1 alleles using RE markers

The markers used for the local sizing method are 124, 147, 184, and 192 -bp in length. The sample is a HUMTHO1 (8, 10) with the bp sizes calculated using the local sizing method. The separation conditions are as follows: column temperature of 54°C, an increase in ACN of 1%, and a flow rate of 0.8ml/min. The gradient started at 40% B and ended at 60% B.

Table 19 Local sizing method applied to HUMTHO1 alleles using RE fragments of bp size 124, 147, 184, and 192

HUMTHO1 allele	Allele bp size	Size using RE fragments	Δ bp from allele size and sizing method
6	158	161.60	3.80

8	166	169.55	3.55
9	170	172.50	2.50
9.3	173	174.65	1.65
10	174	176.18	2.18

The use of 4 size markers to size HUMTHO1 products using least squares analysis and using a local sizing method produced bp sizes for alleles that are constant. Therefore, the results from the four marker experiments can be used to size HUMTHO1 alleles as long as a calculated size is known for each allele. With a calculated size for each allele, a type can be called for each sample that represents the allele repeat number.

Linear Least Squares Analysis of Two RE Markers to Size HUMTHO1 Alleles

The sizing of PCR fragments has been accomplished using two internal markers or standards. Butler et al. used 150 and 300-bp fragments as markers to size a STR system by drawing a least squares fit line between the two markers.²⁸ The retention times of the alleles of the STR system were fit to the least squares line and a bp size was calculated. This same methodology was applied to HPLC to size HUMTHO1 alleles.

Based on the results of using four markers compared to using the three nearest markers (local sizing) to size the HUMTHO1 alleles, the sizing markers should be as close to the alleles as possible. With the markers that were close in bp size to the alleles, the calculated bp sizes were closer to the allele bp sizes as calculated by the primer sites on the intron sequence. The markers used for this study were within 40-bp of the HUMTHO1 alleles.

The markers to be used in this study are the 123, 124, 184, 192 and 213-bp fragments from the pBR322 plasmid cleaved with the Hae III RE, and the 147 and 190-bp fragments from the pUC18 plasmid cleaved with the Msp I RE. All markers were HPLC purified before use as sizing markers.

Markers of different sizes were combined and a HUMTHO1 (9.3, 9.3) homozygote sample was analyzed 3 times for each set of markers. The bp size of the HUMTHO1 (9.3, 9.3) allele is 173 bp. Table 20 contains all the data for this study. The RE markers always size the HUMTHO1 PCR product to have

more bp than the sample contains. This is possibly due to the extra base that is added with the use of *Termus aquaticus* (Taq) polymerase.⁸⁴ Taq polymerase has the ability to catalyze the non-templated addition of nucleotides to the 3' termini of blunt-ended DNA duplexes.^{85,86} Adenine is the base that is typically added after the extension step in PCR. The reason for this particular occurrence is unknown.⁸⁵ Therefore, the PCR-amplified DNA could have two extra bases attached as overhangs on the 3' termini of the PCR products.

The extra base addition is displayed in figure 25. The panel labeled A is the HUMTHO1 (9.3, 9.3) sample amplified with Taq gold and normal amplification conditions. The amplification conditions are described in chapter 1. The sample has an extra peak that is the result of the nontemplated addition of a nucleotide to the 3' end of the amplification product. In dealing with the Taq extra base addition, it has been suggested that an extra PCR step of 72°C for 45-min would add the extra bases on all of the PCR products. Panel B of figure 25 displays the PCR product with the 45-min extension at 72°C. The PCR product in panel B only has one peak, as opposed to the PCR product in panel A that has a mixture of PCR product without the extra base and PCR product with the extra base. The extra peak was eliminated with the use of a 45-min extension of 72°C because . Therefore, all subsequent PCR for the HUMTHO1 locus had an added 45-min extension to the PCR program at a temperature of 72°C.

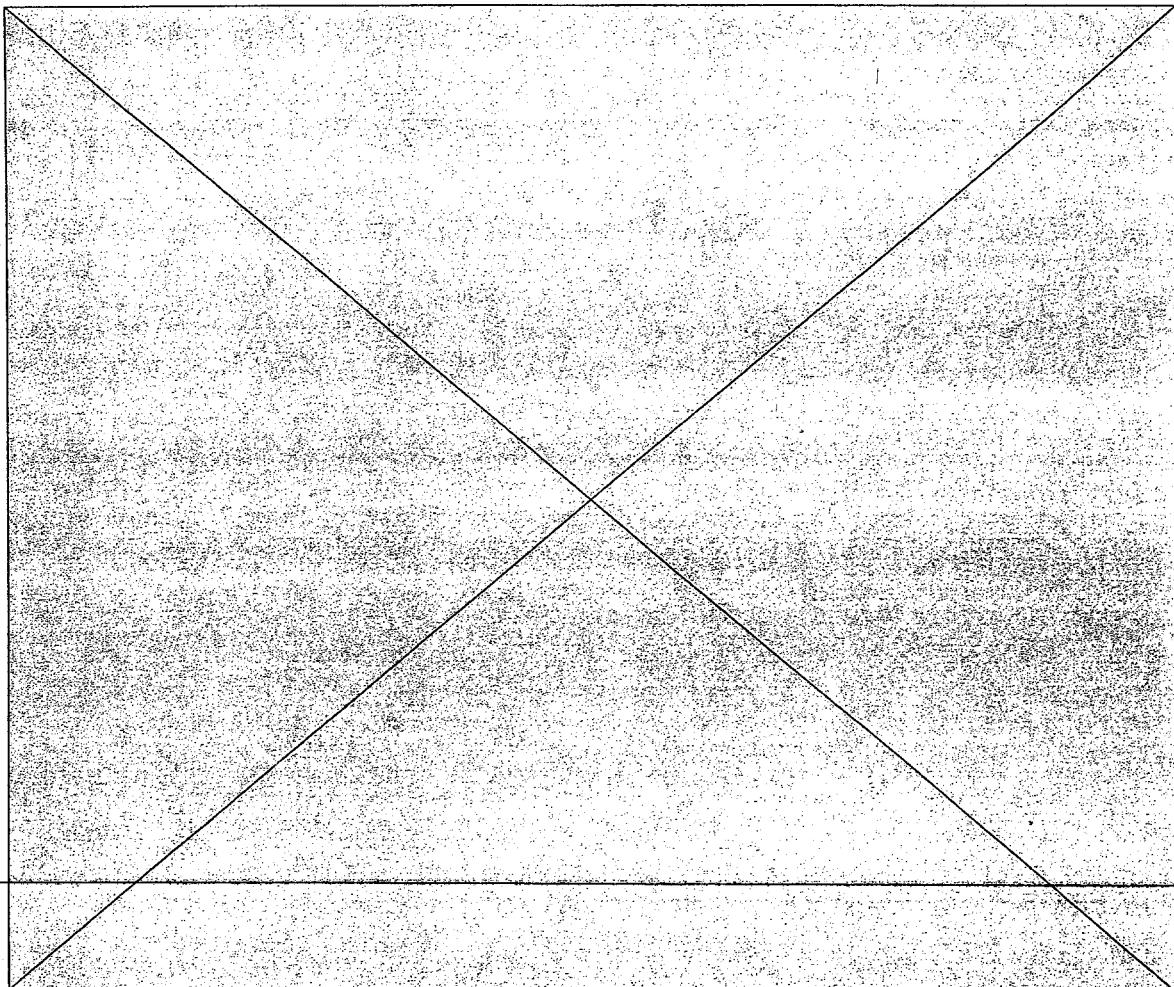


Figure 25 Comparision of amplification conditions for HUMTHO1 alleles with Taq gold

Panel A is an amplification of a HUMTHO1 (9.3, 9.3) with Taq gold. The peak labeled "+ base" is an extra base addition from the use of Taq. Panel B is the same sample with a 45-min annealing extension. There is no "+ base" peak. The separation conditions were listed in figure 23.

Table 20 Sizing method using two markers applied to HUMTHO1 allele

Markers (bp)	Allele bp size	Calculated bp size	Δ bp between sizing method and allele size
123, 213	173	183.17	10.17
124, 213	173	182.64	12.64
147, 213	173	180.74	7.74
123, 192	173	178.82	5.82
124, 192	173	178.54	5.52
147, 192	173	177.93	4.93
147, 190	173	175.95	3.30
147, 184	173	176.03	2.15

The markers that gave the bp size calculations closest to the actual allele size were the 147 and 190-bp markers and the 147 and 184-bp markers, as can be seen in table 20. Therefore, the 147 and 190-bp marker pairs were examined along with the 147 and 184-bp marker pairs in further detail.

The marker pair consisting of the 147 and 190 bp fragments was used to fit the HUMTHO1 samples using a least squares fit. The results of analyzing samples with these markers are listed in table 21 with figure 26 serving as an example chromatogram for the sizing assay. The table contains the results of the sizing experiments which used the two markers, the standard deviation for each sizing method, and the number of base pairs that the calculated size differed from the actual allele size.

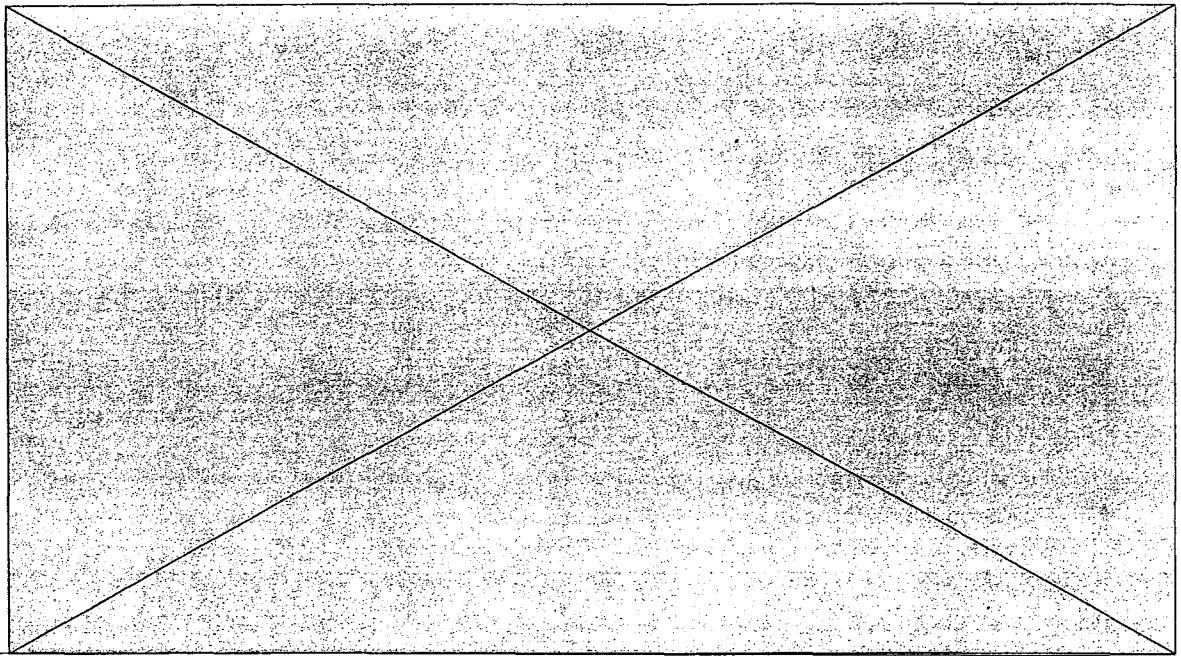


Figure 26 Sizing method for HUMTHO1 alleles using 147 and 190-bp markers
The markers used for the sizing method are 147 and 190-bp. The sample is a HUMTHO1 (6, 6) with the bp size calculated using the least squares fit from the markers. The separation conditions are as follows: column temperature of 54°C, an increase in ACN of 1%, and a flow rate of 0.8ml/min. The gradient started at 47% B and ended at 58% B.

Table 21 Sizing method applied to HUMTHO1 alleles using RE fragments of bp size 147 and 190 ($n = 5$ for each allele)

HUMTHO1 allele	Allele bp size	Size using RE fragments	Standard deviation of allele size	Δ bp between sizing method and allele size
6	158	161.02	0.48	3.02
7	162	165.05	0.63	3.05
8	166	170.27	0.66	4.27
9	170	173.58	0.37	3.58
9.3	173	176.41	0.97	3.41
10	174	177.77	0.39	3.77

The calculated size for each allele was on the average 3.51 bp larger than the size of the allele calculated from the intron sequence. However, the size of the allele calculated from the intronic sequence data does not include the extra base addition from the use of Taq polymerase. The extra base addition may account for some of the difference between the calculated bp size and the allele size. If the extra base addition from the Taq polymerase is added, the bp size of the allele would be increased from X to X + 1 to account for the extra base addition on each of the 3' termini. Even with the extra base addition calculated for each allele, the calculated size from the two markers is 2.51 bp larger than the allele size. In addition, each RE fragment is generated from plasmids that are treated with types of restriction enzymes that cleave to be blunt ended in nature. This means that the enzyme cleaves off excess bases and leaves no overhang. The plasmids treated with Hae III and Msp I cleave leaving a phosphate group on the 5' end. Therefore, the fragments contain a phosphate group on the 5' of both strands (parallel and anti-parallel). This causes an extra two negative charges to be present on all fragments cleaved with the respective enzymes. Therefore, the RE plasmids are not in the same form as the PCR products. The PCR products contain an extra base on the 3' termini with an extra phosphate group and the RE-cleaved fragments only have an extra phosphate group on the 5' termini. These differences between the PCR-amplified dsDNA and the RE-cleaved fragments account for a size difference and a charge difference for each dsDNA moiety.

The marker pair consisting of the 147 and 184-bp fragments was used to fit the HUMTHO1 samples using least squares analysis (table 22). The table contains the results of the sizing using the two markers, the standard deviation for each sizing method, and the number of base pairs each calculated size differed from the experimental size from the primers.

Table 22 Sizing method applied to HUMTHO1 alleles using RE fragments of bp size 147 and 184 (n = 5 for each allele)

HUMTHO1 allele	Allele bp size	Size using RE fragments	Standard deviation of allele size	Δ bp between sizing method and allele size
6	158	160.19	0.49	2.19
7	162	164.41	0.43	2.41

8	166	168.68	0.33	2.68
9	170	172.42	0.26	2.42
9.3	173	174.74	0.34	1.74
10	174	176.22	0.34	2.22

The calculated size for each allele was on the average 2.28 bp larger than the size of the allele calculated from the intron sequence and the primers. This size difference is smaller than the average bp difference for the 147-bp and 190-bp markers. The difference in accuracy of the sizing protocol using the 147-bp and 184-bp markers stems from the size of the markers. The 184-bp marker is closer to the size of the alleles than the 190-bp marker. When using four markers to size the dsDNA, it was found that the closer the bp size of the marker to the bp size of the target DNA allele, the more accurate the sizing assay.

The standard deviation for the sizing using the 147-bp and 184-bp markers was on the average 0.36. This standard deviation is better than the standard deviation for sizing using the 147-bp and 190-bp marker (0.58). Therefore, the 147-bp marker and the 184-bp marker will be used for the typing assay for HUMTHO1 alleles.

Sizing of HUMTHO1 PCR Products Using the 147-bp and 184-bp Markers

The size generated for alleles with the use of the two markers (147-bp and 184-bp) is larger than the size of the PCR product as calculated from the intronic sequence data and the primer pair used for the PCR amplification. To generate a standard for use in comparing unknown samples, a ladder was made of the HUMTHO1 6, 7, 8, 9, and 10 alleles (figure 27) and sized using the markers. Each ladder was analyzed twice and the results are listed in table 23. The use of the ladder is a method to generate a standard to use for the assignment of types for the unknown samples. The standard error for typing when using an Applied Biosystems (ABI) Prism 310 Genetic Analyzer is set to ± 0.5 nucleotide (nt) for all alleles. This method of typing is currently accepted for forensic casework in the United States and is used to analyze samples for both national and state convicted offender DNA databases. With the use of the ladder as a standard, a type can be given to alleles that fall within ± 0.5 nt of the allele sizes for the ladder. If a sample does not fall within ± 0.5 nt, the sample will be re-analyzed and re-sized. In addition, a

HUMTHO1 (9.3, 9.3) sample was analyzed and is included in the ladder sizes (figure 28).

Table 23 Average Base Pair Sizes of the HUMTHO1 allelic ladder to be used for the typing assay (n = 2)

HUMTHO1 allele	Cal. Size
6	160.11
7	164.47
8	168.56
9	172.32
9.3	174.55
10	176.40

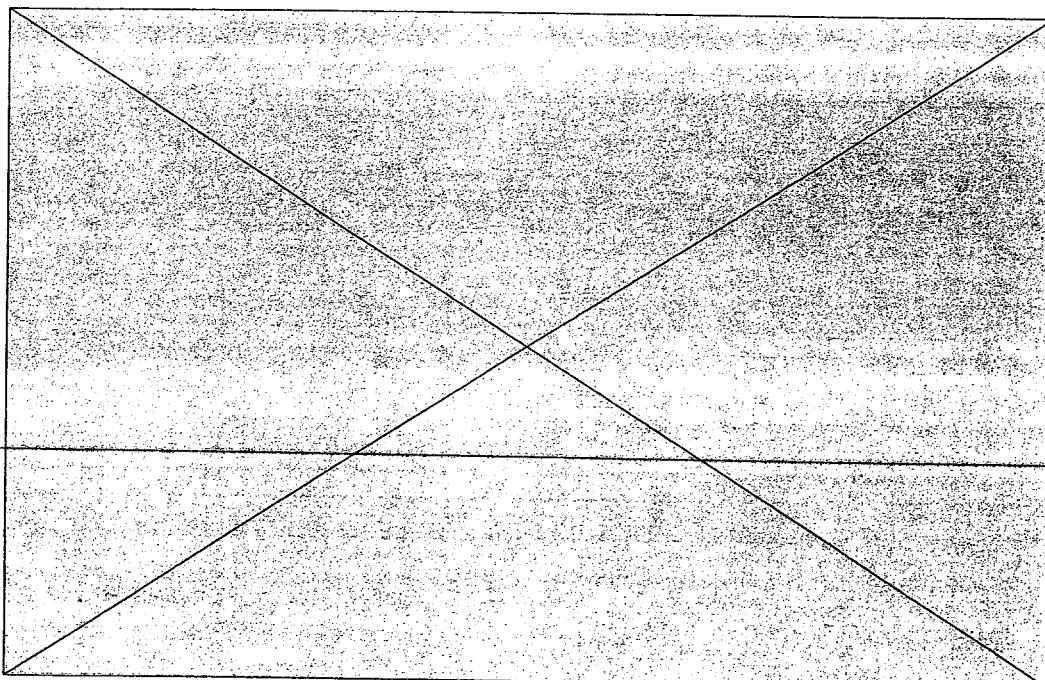


Figure 27 Sizing of HUMTHO1 allelic ladder using 147 and 184-bp markers
The ladder is composed of PCR products from the HUMTHO1 locus (alleles 6, 7, 8, 9, and 10). The calculated bp sizes are above the alleles. The markers are from the pUC18-Msp I and pBR322-Hae III ladders. The separation conditions are the same as in figure 26.

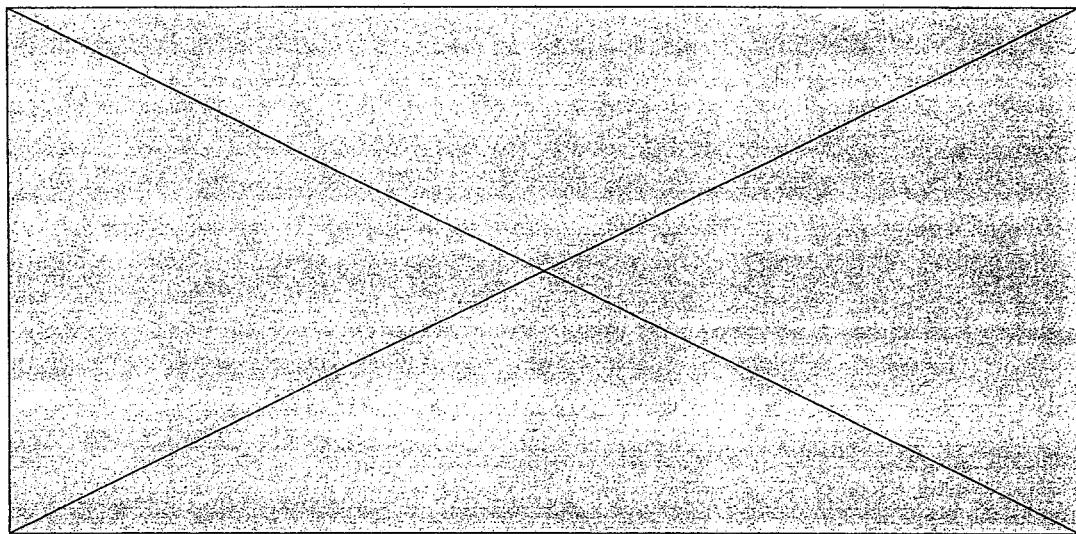


Figure 28 Sizing of HUMTHO1 9.3 allele using 147 and 184-bp markers

The calculated bp size is above the allele. The markers are from the pUC18-Msp I and pBR322-Hae III ladders. The separation conditions are the same as in figure 26.

With this data, a table was calculated for each bp size ± 0.5 nt (table 24). This will be used to type the unknown samples. The table will allow for easy typing of the samples. Once an unknown sample's allele sizes are calculated, the calculated size is compared to table 24. The allele types are in table 24 have ranges that an unknown sample can be placed. This method is called binning, and is used by ABI for the genotyping.³⁴

Table 24 Allele bp sizes ± 0.5 nt to be used for typing the HUMTHO1 locus

HUMTHO1 allele (type)	Cal. Size	Cal. size range for each type
6	160.11	(159.61, 160.61)
7	164.47	(163.97, 165.97)
8	168.56	(168.06, 169.06)
9	172.32	(171.82, 172.82)
9.3	174.29	(173.79, 174.79)
10	176.40	(175.90, 176.90)

The samples in the next table were analyzed with a gradient of 47% eluent B to 58% eluent B. The samples were typed according to the above table. The samples were analyzed on an ABI 310 Genetic Analyzer using standard protocols from the Armed Forces DNA Identification Laboratory, and the type for each sample was given.

Using the typing method with the 147-bp marker and the 184-bp marker allowed for the typing of all samples correctly. This typing procedure allowed for the correct typing of all 9.3 alleles, but when injected with the ladder, the 9.3 allele co-elutes with the 10 allele. To allow the sizing and typing method to work, the 9.3 allele will be analyzed separately from the ladder to generate a standard size. The separation of the

9.3 and 10 allele calls for single bp resolution at 173-bp. Currently, single bp separation can be obtained with fragments with a length up to 124-bp. The conditions of separation do not allow for the single base resolution at the bp size of the HUMTHO1 PCR products.

This sizing method can be used for any STR loci with any markers. In the case of the HUMTHO1 locus, the best markers were the fragments with the bp sizes 147-bp and 184-bp because these markers gave relatively close bp approximations. However, any markers could be used for sizing as long as a ladder is analyzed first to get the bp sizes for each allele. With this method, there was no interference from components of the PCR mix with the size markers as there is in capillary electrophoresis (CE). In CE, the components of the PCR mix interfere with the size marker injection. Therefore, the method of using size markers composed of RE-cleaved plasmid DNA accurately sizes PCR products from STR loci with HPLC.

Table 25 Allele sizes and types calculated using the 147 and 184-bp markers.

Sample	BP Sizes	Allele Type according to table 24	Allele Type using ABI 310 Genetic Analyzer
1	171.90, 173.95	9, 9.3	9, 9.3
2	168.73, 176.30	8, 10	8, 10
3	172.08, 176.05	9, 10	9, 10
4	173.89	9.3, 9.3	9.3, 9.3
5	159.66	6, 6	6, 6
6	164.53, 172.40	7, 9	7, 9
7	174.52	9.3, 9.3	9.3, 9.3
8	160.58, 168.90	6, 8	6, 8
9	164.00, 172.16	7, 9	7, 9
10	160.28, 172.34	6, 9	6, 9
11	164.14, 172.16	7, 9	7, 9
12	160.46, 172.57	6, 9	6, 9
13	160.29, 168.65	6, 8	6, 8
14	171.86, 174.78	9, 9.3	9, 9.3
15	164.40, 172.40	7, 9	7, 9

Sizing of HUMTHO1 PCR products amplified with Pfu using 147-bp and 184-bp markers

As discussed previously, Taq polymerase has the ability to add an extra base on the 3' terminus of a PCR product. This may affect the sizing because there are extra charges that can ion pair with the TEAA and affect the HPLC separation. To determine if the Taq polymerase affected the sizing, a polymerase isolated from the hyperthermophilic marine archaebacterium, Pyrococcus furiosus (Pfu), was used for PCR

amplifications. Pfu does not add an extra base on the 3' terminus of PCR products. The polymerase Pfu has 3' → 5' exonuclease-dependent proofreading activity that causes a ~10-fold lower error rate than Taq.⁸⁸

In working with the Pfu polymerase, the amplification conditions change from that of Taq gold. Pfu has activity at room temperature and must be kept on ice. The results from the Pfu amplifications for sizing are listed in table 26. Each allele was sized twice.

Table 26 Sizing method applied to HUMTHO1 alleles amplified using Pfu polymerase using RE fragments of bp size 147 and 184

HUMTHO1 allele	Allele bp size	Size using RE fragments	Δ bp from allele size and sizing method
6	158	160.59	2.59
8	166	168.08	2.08
9	170	171.1	1.10
10	174	174.65	0.65

The alleles amplified with Pfu gave better results as the allele size increased. For example, the 6 allele had a size difference of 2.59-bp, while the 10 allele had a size difference of only 0.65-bp. There was not a standard change between the actual bp size and the calculated bp size. Therefore, difficulty would arise in using Pfu amplified dsDNA for sizing because there is no constant that could be added to the sizing equation.

In addition, the amplifications produced extra bands (figure 29). The amplification with Pfu is compared with the amplification with Taq gold for the same alleles in figure 30. The amplification with Taq gold does not produce the extra bands. The reason for the extra bands is not known.

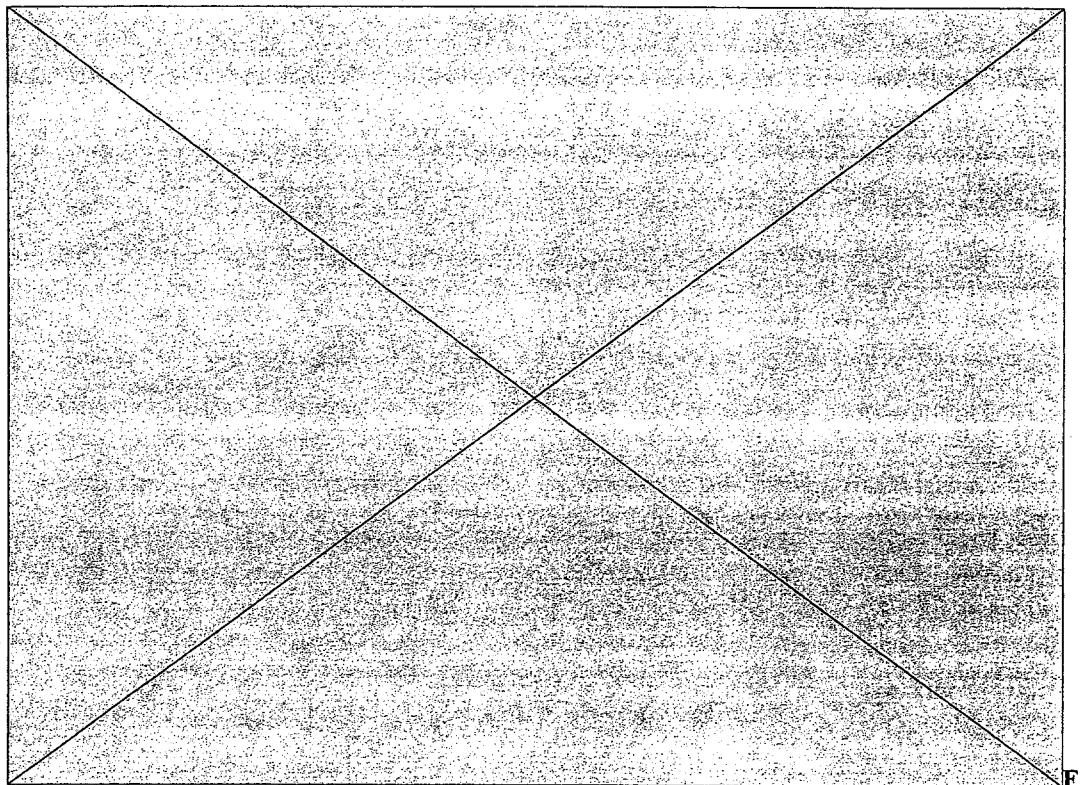


figure 29 HUMTHO1 alleles amplified with Pfu

The HUMTHO1 alleles pictured in panels A, B, and C were amplified with Pfu polymerase. In addition, the HUMTHO1 alleles in panels B and C have an additional peak along with the allele peaks.

When the alleles amplified with Pfu were analyzed with the 147-bp marker, The retention time of the marker changed when compared to the Taq amplified samples (table 27). The retention time for the 147-bp marker, when added to Pfu amplified dsDNA, is less than the retention time of the marker when added to Taq gold amplified dsDNA. Therefore, Taq was used for all sizing experiments because it causes none of the problems associated with Pfu.

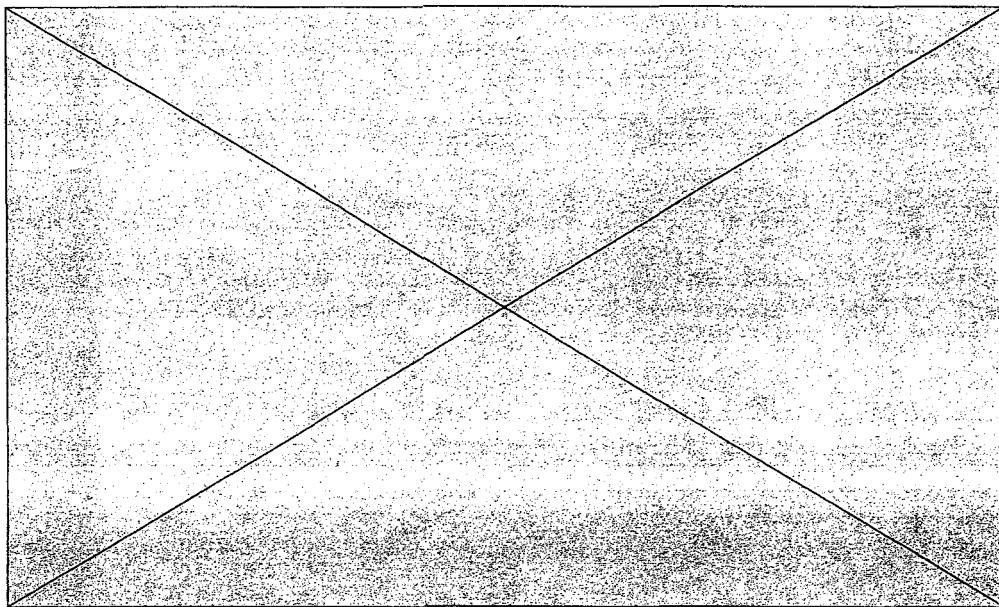


Figure 30

HUMTHO1 sample amplified with PFU and Taq

Panel A is a HUMTHO1 (7, 9) amplified with Pfu polymerase. The peaks marked with an (*) are extra peaks from the use of the Pfu polymerase. The 147-bp marker retention time changed from the retention time in Panel B that is the Taq gold amplified HUMTHO1 (7, 9) sample.

Table 27 Comparison of the retention time of 147-bp maker when mixed with Pfu amplified dsDNA and Taq gold amplified dsDNA

Analysis number	Retention time of 147-bp marker with Pfu amplified dsDNA	Retention time of 147-bp marker with Taq gold amplified dsDNA
1	10.55	10.97
2	10.64	10.98
3	10.55	10.99
5	10.58	10.99
6	10.51	11.00
7	10.57	11.01
8	10.59	10.99
9	10.55	10.98
10	10.56	10.99
Average (n=10)	10.57	10.99
Standard deviation	0.036	0.012

Normalization of HUMTHO1 PCR Samples

Another method that was explored to type alleles was normalization of the separation using a standard marker. This method was used to type D1S80 alleles with CE as a separation method.⁵³ Normalization with CE produced results with only 92.5 % accuracy. In some cases (4.0 %), an allele assignment was not possible because the allele normalized retention time was between two normalized retention times.

To normalize retention times for use as typing method for HPLC, the 147-bp marker was used as the peak that the HUMTHO1 alleles were normalized against. The normalized retention time (NRT) for each allelic peak was calculated using the following equation

$$(4) \quad NRT = \text{retention time for allele} / \text{retention time for 147 bp marker}$$

with the units of both retention times in min.

To calculate an NRT for the 5, 6, 7, 8, 9, 9.3, and 10 alleles in the HUMTHO1 locus, each allele was analyzed five times and equation 4 was used to calculate the NRT (table 28).

Table 28 NRT for the HUMTHO1 alleles using 147-bp marker
(n = 5 for each allele)

HUMTHO1 allele	NRT	Standard deviation of NRT
6	1.086	0.0018
7	1.115	0.0023
8	1.143	0.0032
9	1.167	0.0019
9.3	1.182	0.0026
10	1.192	0.0023

From the calculated means and standard deviation, a 99.7% confidence interval was calculated for each allele's NRT (table 29). This was used as the normalization standard table. As can be seen in table 30, the allele designated 9.3 and 10 have NRTs that overlap. Therefore, the 9.3 and 10 alleles will be grouped into one type (9.3/10).⁸⁹

Table 29 NRT confidence intervals for the HUMTHO1 alleles using a 147-bp marker

HUMTHO1 allele	NRT	Confidence Interval for NRT (99.7%)
6	1.086	(1.081, 1.091)
7	1.115	(1.108, 1.122)
8	1.143	(1.133, 1.152)
9	1.167	(1.162, 1.172)
9.3	1.182	(1.175, 1.189)
10	1.192	(1.185, 1.199)

With the NRT time confidence intervals from table 30, HUMTHO1 samples were typed using the normalization procedure (table 30). The NRT data agrees with the actual size data from the ABI 310 Genetic analyzer. The only drawback to the use of NRT is that the 9.3 and 10 allele are typed as one allele. Although this test could not be used for paternity testing but it could be useful as a quick method to screen samples. This method only requires one RE marker and does not require the mixing of any HPLC-purified markers.

Table 30 Typing of HUMTHO1 samples with normalization using table 29

Sample	NRT	Allele Type according to table 29	Allele Type using ABI 310 Genetic Analyzer
1	1.166, 1.187	9, 9.3/10	9, 9.3
2	1.183	9.3/10, 9.3/10	9.3, 9.3
3	1.145, 1.195	8, 9.3/10	8, 10
4	1.184	9.3/10, 9.3/10	9.3, 9.3
5	1.086	6, 6	6, 6
6	1.118, 1.171	7, 9	7, 9
7	1.184	9.3/10, 9.3/10	9.3, 9.3
8	1.087, 1.143	6, 8	6, 8
9	1.110, 1.164	7, 9	7, 9
10	1.087, 1.167	6, 9	6, 9
11	1.162, 1.183	9, 9.3/10	9, 9.3
12	1.165, 1.186	9, 9.3/10	9, 9.3
13	1.115, 1.168	7, 9	7, 9

It is possible to amplify a marker with the HUMTHO1 alleles in the PCR reaction. The only requirement is that a constant bp sized PCR marker is used that will co-amplify with the HUMTHO1 alleles.

For example, a PCR product could be amplified in the human mitochondrial (mt) genome to be used for this purpose. In mtDNA, the only change from individual to individual is in the sequence not size.

However, this is beyond the scope of this thesis.

Genetic Typing with the Use of an Allelic Ladder

A new approach to STR loci typing has been the use of a STR allelic ladder with the same bp composition as the unknown alleles from samples.^{27,29,30} The ladder and unknown sample are co-injected and electrophoresed together. The samples can be intercalated with a fluorescent dye. There are some problems with this type of analysis. The first problem is that the dyes used to make the ds- or ssDNA fluorescent are carcinogenic or expensive. In addition, some ladders may have different sequences than the actual samples; the differences can occur in the repeat units or in the base pairs that are on either side of the repeats. In fact, the locus VWA31A, chromosomal location 12p12-pter, is a 4 bp repeat system that has allele variants consisting of base changes (the repeating unit changes from a TCTG to a TCTA).⁹⁰

Capillary electrophoresis is a technique that can be influenced by the sequence of a dsDNA fragment.⁹¹

Therefore, if the STR system that is being analyzed contains sequence variation, the size of the allele or the type could be misidentified.

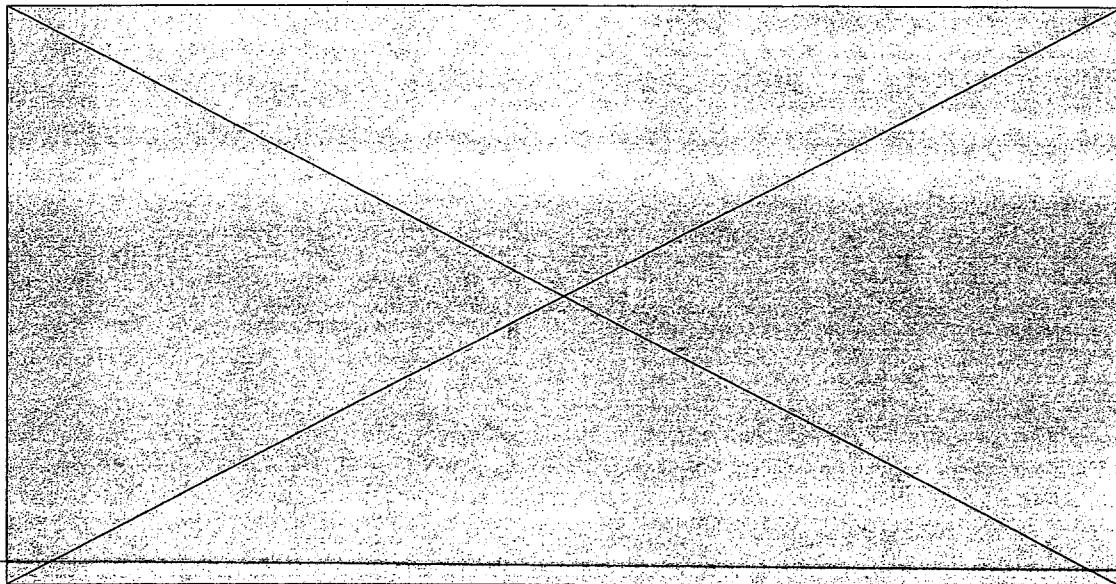


Figure 31 Chromatogram of the HUMTHO1 ladder for co-injection typing

The ladder is composed of HUMTHO1 PCR products. The separation conditions are the same as used in figure 26.

The column used for the separation of dsDNA in this study was unaffected by sequence of a dsDNA fragment but is influenced by the size of the dsDNA fragment. The STR system HUMTHO1 was analyzed for typing by using a ladder made of the 6, 7, 8, 9, and 10 alleles amplified from individuals. These alleles were used straight from the PCR reactions and were adjusted for concentration effects. The heteroduplexes^{25, 30, 54} produced from the PCR were reduced by dilution and did not affect the ladder. An HPLC chromatogram of the ladder can be seen in figure 31.

In all cases, the ladder was analyzed first, followed by the ladder with a sample added to it. Typically, the ladder consisted of 8.5 µl of a mixture of three samples with the HUMTHO1 types [(6, 6), (8, 10), (7, 9)]. The homozygote allele gives an amplification that is very concentrated and it must be

diluted more than the other samples that are mixed. Figures 32-35 are four samples that were typed using a ladder. The samples were all typed correctly. The only drawback is that the separation will not resolve the 9.3 and 10 allele.

Figure 32 in the figure contains a separation of a HUMTHO1 sample with the type (7, 9) co-injected with the HUMTHO1 allelic ladder. The heteroduplexes from the PCR reaction are retained less than the alleles. The next sample is a HUMTHO1 ladder spiked with a (9, 10) sample (figure 33). In this separation, the heteroduplexes are not visible due to the low concentration of the sample. This shows that a ladder can be used as long as the sample concentration is kept low. Figure 34 and 35 show the HUMTHO1 ladder spiked with a HUMTHO1 (8, 10) and a HUMTHO1 (6, 6), respectivley.

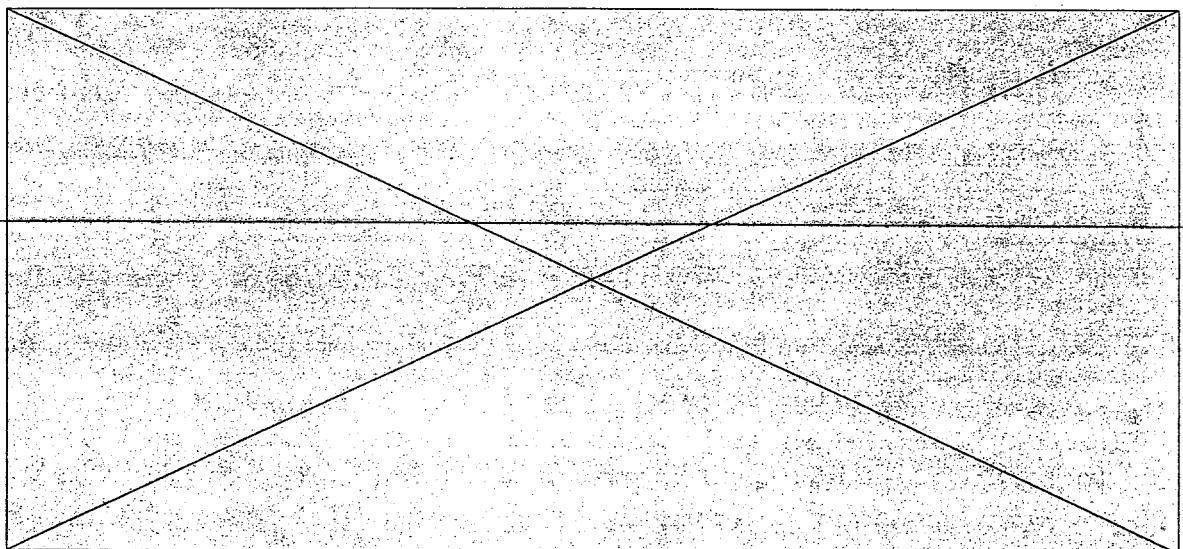


Figure 32 Typing of a HUMTHO1 (7, 9) by co-injection with an allelic ladder
The separation conditions are in figure 26.

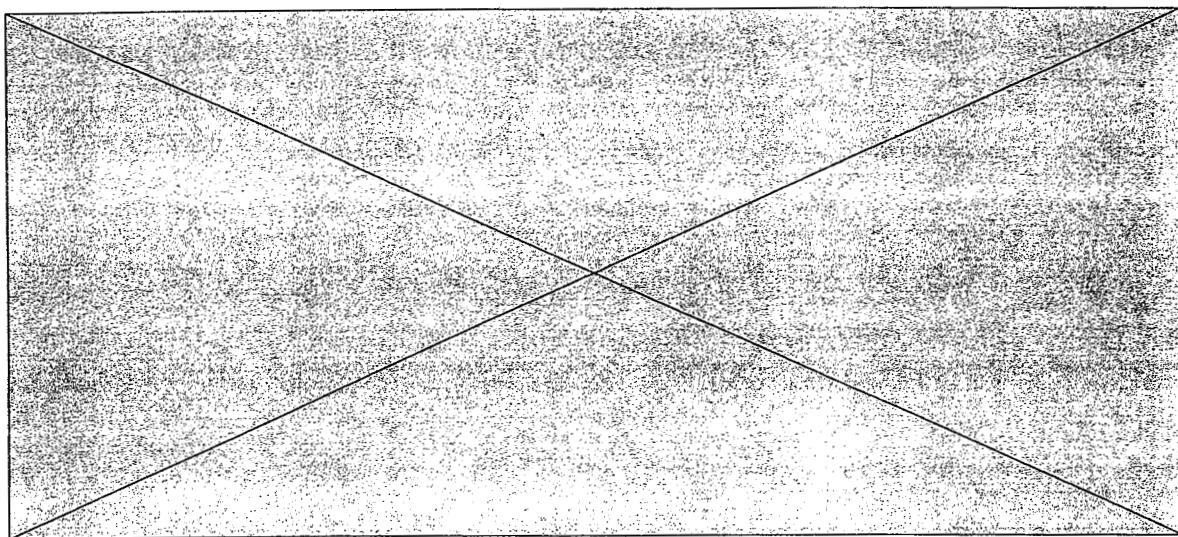


Figure 33 Typing of a HUMTHO1 (9, 10) by co-injection with an allelic ladder
The separation conditions are in figure 26.

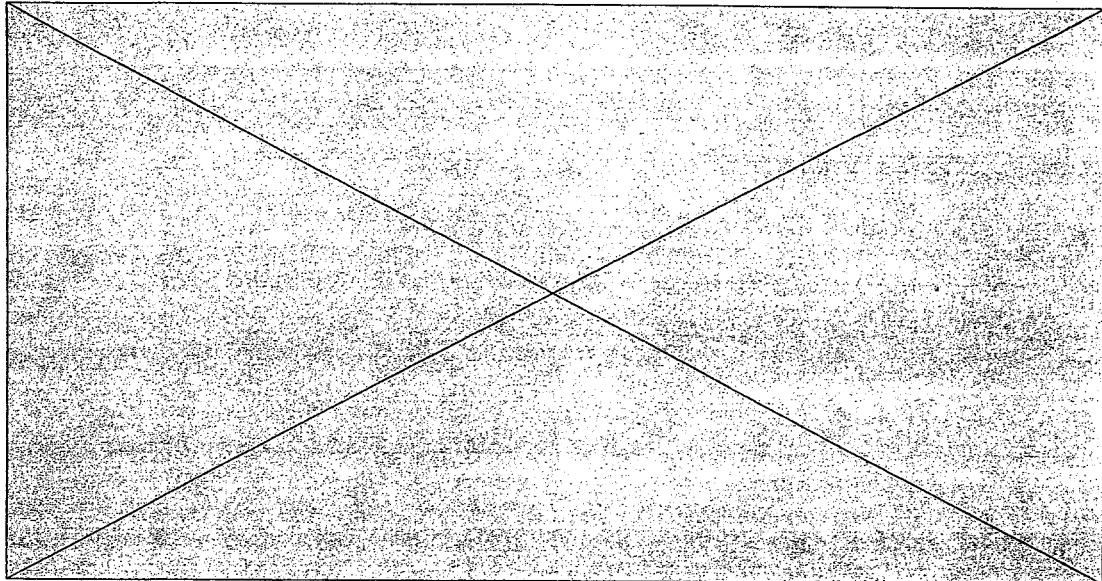


Figure 34 Typing of a HUMTHO1 (8, 10) by co-injection with an allelic ladder
The separation conditions are in figure 26.

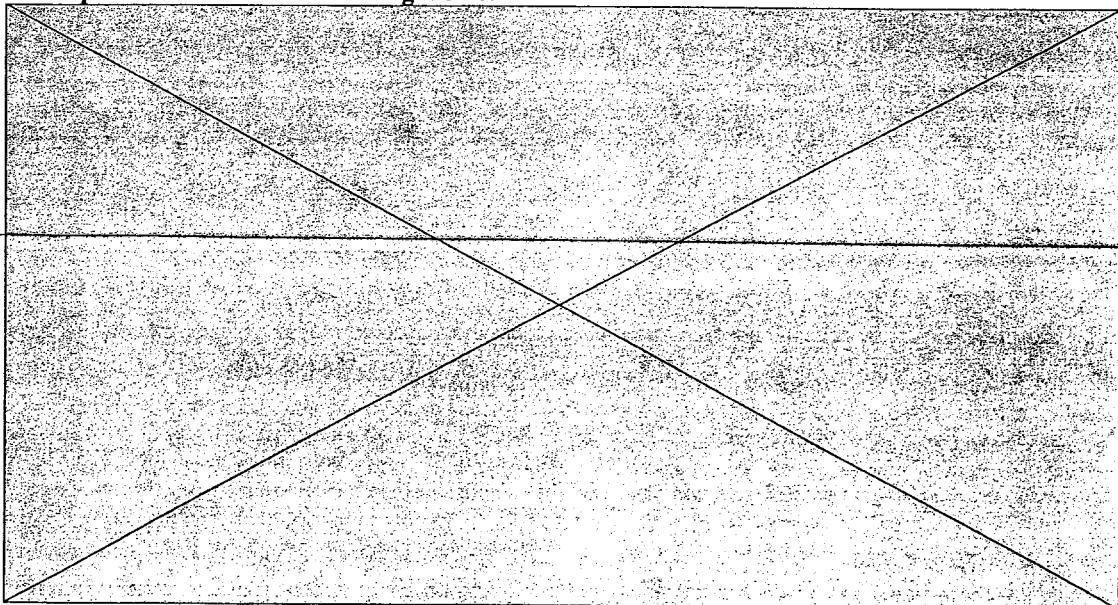


Figure 35 Typing of a HUMTHO1 (6, 6) by co-injection with an allelic ladder
The separation conditions are in figure 26.

Multiplexing of HUMTHO1, mtDNA Repeat, and Amelogenin

The HUMTHO1 alleles can be multiplexed with other loci to increase identification power. One such loci that can be analyzed with the HUMTHO1 locus is a short dinucleotide repeat polymorphism that occurs at position 514 in the D-loop region of the human mitochondrial genome.⁹² The number of repeats and the primers for PCR amplification of the mtDNA repeat can be found in chapter 1, tables 4 and 5. This

locus was sized using the 124-bp marker and the 184-bp marker and the results are listed in table 32. The results were confirmed by sequencing.

With the data from table 31, a table was calculated for each bp size \pm 0.5 nt (table 32). This will be used as the method to calculate a genotype for each sample. This table will allow for easy typing of the samples when multiplexed with HUMTHO1 samples.

Table 31 Sizing of the mtDNA dinucleotide repeat using RE fragments of bp size 124 and 184 (n = 3)

Number of mtDNA dinucleotide repeats	bp size of allele	Cal. size
4	133	140.80
5	135	143.05
6	137	145.50

Table 32 Allele bp sizes for mtDNA dinucleotide repeat \pm 0.5 nt for typing

Number of mtDNA dinucleotide repeats	Cal. Size	Cal. size range for each type
4	140.80	(140.30, 141.30)
5	143.05	(142.55, 143.55)
6	145.50	(145.00, 146.00)

The X-Y homologous amelogenin gene codes for an integral protein in mammalian tooth enamel development.⁹³ This gene has been used extensively for the sex-typing of forensic materials.⁹⁴

Amplification of a small portion of the amelogenin gene produces fragments that are distinct from the X and the Y chromosomes and consequently, information regarding the sex of the donor. These fragments have the sizes 106 and 112 bp, respectively.

The amelogenin gene PCR products were sized using 104-bp and 124-bp fragments as markers. The results of this study are listed in table 33.

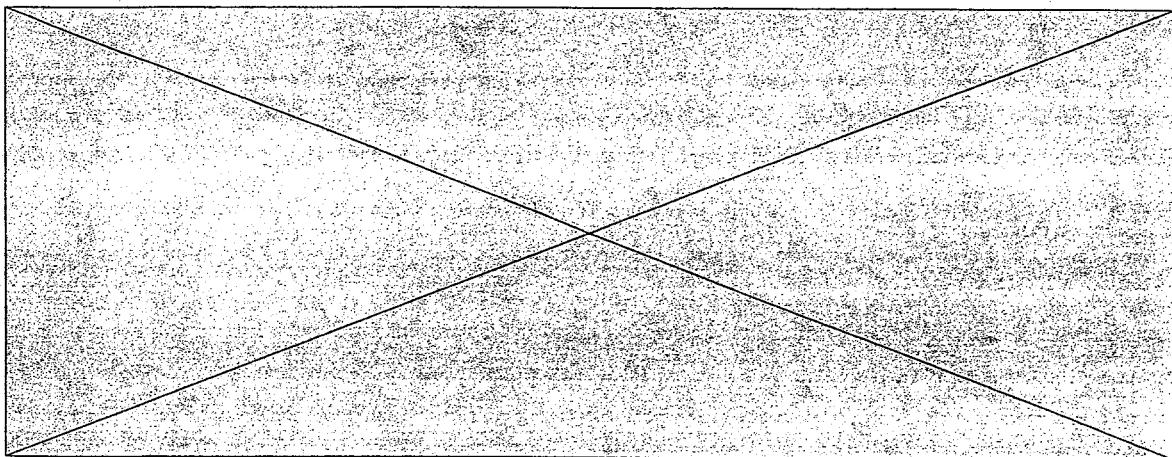
Table 33 Sizing of the amelogenin gene sex typing PCR products using RE fragments of bp size 104 and 124 (n = 3)

amelogenin gene type	bp size of allele	Cal. size
X	106	110.68
Y	112	116.25

With the data from table 33, a table was calculated for each bp size \pm 0.5 nt (table 34). This will be used as the method to calculate a genotype for each sample. This table will allow for easy typing of the samples when multiplexed with HUMTHO1 and mtDNA dinucleotide repeat samples.

Table 34 Allele bp sizes for amelogenin gene \pm 0.5 nt for typing

Amelogenin gene type	Cal. Size	Cal. size range for each type
X	110.68	(110.18, 111.18)
Y	116.25	(115.75, 116.75)



Using the information in tables 24, 32, and 34, PCR products from the amelogenin locus, the mtDNA dinucleotide repeat, and the locus HUMTHO1 were mixed and co-injected with the respective sizing markers for each locus. An example chromatogram is pictured in figure 36. The calculated bp results for five analysis are listed in table 35 and the allele bp sizes are located in table 36.

Figure 36 Sizing of the multiplex consisting of the amelogenin locus, the mtDNA dinucleotide repeat locus and the HUMTHO1 locus
The amelogenin locus sample sizes were calculated using the 104-bp and 124-bp markers. The +A peaks are from amplification with Taq gold. The mtDNA repeat calculated size was calculated using the 124-bp marker and the 184-bp markers. The HUMTHO1 sample calculated sizes are above the allele peaks. The conditions for the HPLC separation are present in figure 26.

Table 35 Allele sizes calculated for the multiplexed amelogenin, mtDNA repeat, and HUMTO1 loci using the 104, 124, 147, and 184-bp markers

Sample	BP Sizes for amelogenin (104 and 124-bp markers)	BP Sizes for mtDNA repeat (124 and 184-bp markers)	BP Sizes HUMTHO1 (147 and 184-bp markers)
1	110.72, 116.34	142.80	172.36, 174.43
2	110.67	143.11	169.03, 176.71
3	110.76	140.61	164.29, 172.60
4	110.91, 116.45	143.30	174.19
5	110.62, 116.32	142.93	160.55, 168.96
6	110.81, 116.46	143.11	160.52, 168.71

Using the information in tables 25, 32, and 34, the types for each locus were conferred (table 36).

Table 36 Types for the samples used in table 35

Sample	amelogenin type	mtDNA repeat type	HUMTHO1 type
1	X, Y	5	9, 9.3
2	X, X	5	8, 10
3	X, X	4	7, 9
4	X, Y	5	9.3, 9.3
5	X, Y	5	6, 8
6	X, Y	5	6, 8

The samples were all typed correctly according to the sizing results and sequence results from the ABI 310 Genetic Analyzer. Therefore, three different loci of different AT content were analyzed and typed correctly.

Conclusion

The separation of dsDNA on a HPLC with a DNAsep™ column was optimized with the use of RE fragments and PCR products with respect to column temperature, flow rate (ml/min), and % ACN per min. The optimal conditions for the highest resolution and fastest separation were a column temperature of 54°C, flow rate of 0.8 ml/min, and 1.0 % ACN per min. With these optimal conditions, the HUMTHO1 locus was separated and sized using RE fragments as markers. In addition, the amelogenin locus, a dinucleotide repeat in the human mt genome, and the HUMTHO1 locus were simultaneously sized and typed using RE fragments.

References

1. Giusti, A. M.; Budowle, B. *Applied and Theoretical Electrophoresis* 1995, 5, 89-98.
2. Budowle B.; Baechtel, F. S.; Comey, C. T.; Giusti, A. M.; Klevan; L. *Electrophoresis* 1995, 16, 1559-1567.
3. Budowle, B.; Giusti, A. M. *J. Forensic Sci.* 1995, 40, 236-238.
4. Gregory, S. G.; Howell, G. R.; Bentley, D. R. *Genome Res.* 1997, 7, 1162-1168.
5. Balazs, I.; Baird, M.; Clyne, M.; Meade, E. *Am. J. Hum. Genet.* 1989, 44, 182-190.
6. Marino M. A.; Turni, L. T.; Del Rio, S. A.; Williams, P. E. *J. Chromatogr.* 1994, 676, 185-189.
7. Southern, E. M. *J. Mol. Biol.* 1975, 98, 303-517.
8. Benzinger, E. A.; Emerek, E. A.; Grigsby, N. L.; Duewer, D. L.; Lovekamp M. L.; Deadman, H.; Thompson, J. L.; Sallee, P. J.; Riech, A. K. *J. Forensic Sci.* 1997, 5 ,850-863.
9. Duewer, D. L.; Benzinger, E. A. *J. Forensic Sci.* 1997, 5, 864-872.
10. Chakraborty R.; Kidd, K. K. *Science* 1991, 254, 1735-1739.

11. Vigilant, L.; Pennington, R.; Harpending H.; Kocher, T. D.; Wilson, A. C. *PNAS*, 1989, 86, 9350-9354.
12. Schmitt, C.; Benecke, M. *Electrophoresis* 1997, 18, 690-694.
13. Mills, K. A.; Even, D.; Murray, J. C. *Hum. Mol. Genet.* 1992, 1, 779.
14. Polymeropoulos, M. H.; Xiao, H.; Rath, D. S.; Merril, C. R. *Nucleic Acids Res.* 1991 , 19, 3753.
15. Frazier, R. R. E.; Millican, E. S.; Watson, S. K.; Oldroyd, N. J.; Sparkes, R. L.; Taylor, K. M.; Panchal, S.; Bark, L.; Kimpton, C. P.; Gill, P. D. *Electrophoresis* 1996, 17, 1550-1552.

16. Whitaker, J. P.; Clayton, T.M.; Urquhart, A. J.; Millican, E. S.; Downes, T. J.; Kimpton, C. P.; Gill, P. *BioTechniques* 1995, 18, 670-677.
17. Fregeau, C. J.; Fourney, R. M. *BioTechniques* 1993, 15, 100-113.
18. Corach, D.; Sala, A.; Penacino, G.; Sotelo, A. *Electrophoresis* 1995, 16, 1617-1623.
19. de Pancorbo, M. M.; Castro, A.; Alonso, S.; Fernandez-Fernandez, I.; Barbero, C.; Garcia-Orad, A.; Izaguirre, N.; Iriondo, M.; de la Rua, C. *Electrophoresis* 1995, 16, 1612-1616.
20. Sparkes, R.; Kimpton, C.; Watson, S.; Oldroyd, N.; Clayton, T.; Barnett, L.; Arnold, J.; Thompson, C.; Hale, R.; Chapman, J.; Urquhart, A.; Gill, P. *Int. J. Legal Med.* 1996, 109, 186-194.
21. Bullard, K. M.; Hietpas, P. B.; Ewing A. G. *Electrophoresis* 1998, 19, 71-75.
22. Wang, Y.; Wallin, J. M.; Ju, J.; Sensabaugh, G. F.; Mathies, R. A. *Electrophoresis* 1996, 17, 1485-1490.
23. Morin P. A.; Smith, D. G. *BioTechniques* 1995, 19, 223-8.
24. Budowle, B.; Smerick, J. B.; Keys, K. M.; Moretti, T. R. *J. Forensic Sci.* 1997, 42, 846-849.
25. Wang, Y.; Ju, J.; Carpenter, B. A.; Atherton, J. M.; Sensabaugh, G. F.; Mathies, R. A. *Anal. Chem.* 1995, 67, 1197-1203.
26. Woolley, A. T.; Sensabaugh, G. F.; Mathies, R. A. *Anal. Chem.* 1997, 69, 2181-2186.

27. Zhang, N.; Yeung, E. S. *Anal. Chem.* 1996, 68, 2927-2931.
28. Butler, J. M.; McCord, B. R.; Jung, J. M.; Lee, J. A.; Budowle, B.; Allen, R. O. *Electrophoresis* 1995, 16, 974-980.
29. Isenberg, A. R.; McCord, B. R.; Koons, B. W.; Budowle, B.; Allen, R. O. *Electrophoresis* 1996, 17, 1505-1511.
30. Zhang, N.; Yeung, E. S. *J. Chromatogr.* 1997, 768, 135-141.
31. Clark, S.; Mattheis, R. *Anal. Chem.* 1997, 69, 1355-1363.
32. Mansfield, E. S.; Robertson, J. M.; Vainer, M.; Isenberg, A. R.; Frazier, R. R.; Ferguson, K.; Chow, S.; Harris, D. W.; Barker, D. L.; Gill, P. D.; Budowle, B.; McCord, B. R. *Electrophoresis* 1998, 19, 101-107.
33. Isenberg, A. R.; Allen, R. O.; Keys, K. M.; Smerick, J. B.; Budowle, B.; McCord, B. R. *Electrophoresis* 1998, 19, 94-100.
34. Lazaruk, K.; Walsh, P. S.; Oaks, F.; Gilbert, D.; Rosenblum, B. B.; Menchen, S.; Scheibler, D.; Wenz, H. M.; Holt, C.; Wallin J. *Electrophoresis* 1998, 19, 86-93.
35. Bischoff, R.; McLaughlin L. W. *J. Chromatogr.* 1984, 296, 329-337.
36. Larson J. E.; Hardies, S. C.; Patient, R. K.; Wells, R. D. *J. Biol. Chem.* 1979, 254, 5535-5541.

37. Patient, R. K.; Hardies, S. C; Wells, R. D. *J. Biol. Chem.* **1979**, 254, 5542-5547.
38. Patient, R. K.; Hardies, S. C.; Larson J. E.; Inman R. B.; Maquat, L. E.; Wells, R. D. *J. Biol. Chem.* **1979**, 254, 5548-5554.
39. Kato, Y.; Yamasaki, A.; Onaka, A.; Kitamura, T.; Hashimoto, T.; Murotsu, T.; Fukushige, S.; Matsubara, K. *J. Chromatogr.* **1989**, 478, 264-268.
40. Goss, T. A.; Bard, M.; Jarrett, H. W. *J. Chromatogr.* **1990**, 508, 279-287.
41. Haupt, W.; Pingoud, A. *J. Chromatogr.* **1983**, 260, 419-427.
42. Sumita, C.; Baba, Y.; Hide, K.; Ishimaru, N.; Samata, K.; Tanaka, A.; Tsuhako M. *J. Chromatogr. A* **1994**, 661, 297-303.
43. Katz, E. D.; Dong, M. W. *BioTechniques* **1990**, 8, 546-555.
44. Katz, E. D.; Haff, L. A.; Elsteen, R. *J. Chromatogr.* **1990**, 512, 433-444.
45. Strege, M. A.; Lagu, A. *J. Chromatogr.* **1991**, 555, 109-124.
46. Eriksson, S.; Glad, G.; Pernemalm, P. A.; Westman, E. *J. Chromatogr.* **1986**, 359, 265-274.
47. Huber, C. G.; Oefner, P. J.; Bonn, G. K. *Anal. Biochem.* **1993**, 212, 351-358.
48. Huber, C. G.; Oefner, P. J.; Preuss, E.; Bonn, G. K. *Nucleic Acids Res.* **1993**, 21, 1061-1066.
49. Huber, C. G.; Oefner, P. J.; Bonn, G. K. *Anal. Chem.* **1995**, 67, 578-585.

50. Huber, C. G.; Berti, N. *Anal. Chem.* **1996**, 68, 2959-2965.
51. Oefner, P. J.; Huber, C. G.; Puchhammer-Stockl, E.; Umlauft, F.; Grunewald, K.; Bonn, G. K.; Kunz, C. *BioTechniques* **1994**, 16, 898-908.
52. Oefner, P. J.; Huber, C. G.; Umlauft, F.; Berti, G-N.; Stimpfl, E.; Bonn, G. K. *Anal. Biochem.* **1994**, 223, 39-46.
53. Mitchell, J. W.; Walsh, D. *Electrophoresis* **1998**, 19, 80-85.
54. Marino, M. A.; Devaney, J. M.; Smith, J. K.; Girard, J. E. *Electrophoresis* **1998**, 19, 108-118.
55. Puers, C.; Hammond, H. A.; Jin, L., Caskey, C. T.; Schumm, J. W. *Am. J. Hum. Genet.* **1993**, 53, 953-958.
56. Schumm, J. W.; Lins, A. M.; Micka, K. A.; Sprecher, C. J.; Rabbach, D.; Bacher, J. W. Proceedings from the First European Symposium on Human Identification (1996). **1997**, 90-104.
57. Roberston, J. M.; Squeglia, J. B.; Badger, C. A.; Juston, A. C.; Ballantyne, J. *Electrophoresis* **1995**, 16, 1568-1576.
58. Gill, P.; Evett, I. *Genetica* **1995**, 96, 69-87.
59. Usher, D. A. *Nucleic Acids Res.* **1979**, 6, 2289-2306.

60. Colin, H.; Diez-masa, J. C.; Guichon, G.; Czajkowska, T.; Mieddziak, I. *J. Chromatogr.* **1978**, 167, 41-65.
61. Antia, F.; Horváth, Cs. *J. Chromatogr.* **1988**, 435, 1-15.
62. Robbat, A.; Liu, T. Y. *J. Chromatogr.* **1990**, 513, 117-135.
63. Kenkelem, L. V.; Lewitus, A. J.; Kana, T.; Craft, N. E. *Mar. Ecol.: Prog. Ser.* **1994**, 114 (3), 303-313.
64. Chen, H.; Horváth, Cs. *J. Chromatogr.* **1995**, 705, 3-20.
65. Li, J.; Carr, P. W. *Anal. Chem.* **1997**, 69, 2202-2206.
66. Pirkle, W. H.; Burke, J. A. *J. Chromatogr.* **1991**, 557, 173-185.
67. Takagi, T.; Suzuki, T. *J. Chromatogr.* **1992**, 625, 163-168.
68. Labl, M.; Fang, S.; Hruby, K. J. *J. Chromatogr.* **1991**, 586, 145-148.
69. Chloupek, R. C.; Hancock, W. S.; Marchylo, B. A.; Kirkland, J. J.; Boyes, B. E.; Snyder, L. R. *J. Chromatogr.* **1994**, 686, 45-59.
70. Li, J.; Carr, P. W. *Anal. Chem.* **1997**, 69, 837-843.
71. Ksai, K. *J. Chromatogr.* **1993**, 618, 203-221.
72. Ringo, M. C.; Evans, C. E. *Anal. Chem.* **1997**, 69, 643-649.

73. Ringo, M. C.; Evans, C. E. *Anal. Chem.* **1997**, 69, 4964-4971.
74. Martin, M.; Blu, G.; Guiochon, G. *J. Chromatogr. Sci.* **1973**, 11, 641-654.
75. Martire, D. E.; Boehm, R. E. *J. Phys. Chem.* **1987**, 91, 2433-2446.
76. Tanaka, N.; Yoshimura, T.; Araki, M. *J. Chromatogr.* **1987**, 406, 247-256.
77. McGuffin, V. L.; Evans, C. E. *J. Microcolumn Sep.* **1991**, 3, 513-520.
78. McGuffin, V. L.; Evans, C. E.; Chen, S. H. *J. Microcolumn Sep.* **1993**, 5, 3-10.
79. Potter, R. L.; Lewis, R. V. *High Performance Liquid Chromatography Advances and Perspectives*, 4th ed.; Academic Press: Orlando, 1986; Chapter 1.
80. Molnar, I.; Horváth, Cs. *Clin. Chem.* **1976**, 22, 1497-1502.
81. Klintschar, M.; Crevenna, R. J. *Forensic Sci.* **1997**, 42, 907-910.
82. Lins, A. M.; Sprecher, C. J.; Puers, C.; Schumm, J. W. *BioTechniques* **1996**, 20, 223-228.
83. Beckman, J. S.; Weber, J. L. *Genomics* **1992**, 12, 627-631.
84. Clark, J. M. *Nucleic Acids Res.* **1988**, 16, 9677-9686.

85. Magnuson, V. L.; Ally, D. S.; Nylund, S. J.; Karanjawala, Z. E.; Rayman, J. B.; Knapp, J. I.; Lowe, A. L.; Ghosh, S.; Collins, F. S. *BioTechniques*, 1996, 21, 700-709.
86. Wenz, H.-M.; Roberston, J. M.; Menchen, S.; Oaks, F.; Demorest, D. M.; Scheibler, D.; Rosenblum, B. B.; Wike, C.; Gilbert, D. A.; Efcavitch, J. W. *Genome Res.* 1998, 8, 69-80.
87. Elder, J. K.; Southern, E. M. *Anal. Biochem.* 1983, 128, 227-331.
88. Cline, J.; Braman, J. C.; Hogrefe, H. H. *Nucleic Acids Res.* 1996, 24, 3546-3551.
89. Edwards, A.; Civitello, A.; Hammond, H.A.; Caskey, C.T. *Am. J. Hum. Genet.* 1991, 49, 746-756.
90. Brinkmann, B.; Sajantila, A.; Goedde, H.W.; Matsumoto, H.; Nishi, K.; Wiegand, P. *Eur. J. Hum. Genet.* 1996, 4, 175-182.
91. Wenz, H. M. *Nucleic Acids Res.* 1994, 22, 4002-4008.
92. Bodenteich, A.; Mitchell, L. G.; Polymeropoulos, M. H.; Merril, C. R. *Hum. Mol. Genet.* 1992, 1, 140.
93. Nakahori, Y.; Takenka, O.; Nakagome, Y. *Genomics* 1991, 9, 264-269.
94. LaFountain, M.; Schwartz, M.; Cormier, J.; Buel, E. *J. Forensic Sci.* 1998, 43, 1188-1194.
95. Levison, G.; Gutman, G. A. *Mol. Biol. Evol.* 1987, 4, 397-401.
96. Schlotterer, C.; Tautz, D. *Nucleic Acids Res.* 1992, 20, 211-215.

97. Sprecher, C. J., Puers, C.; Lins, A. M.; Schumm, J. W. *BioTechniques*, 1996, 20, 266-276.
98. Weston, A.; Brown, P. R. HPLC and CE Principles and Practices, 1st ed.; Academic Press: San Deigo, 1997.
99. Horvath, C.; Melander, W. R.; Molnar, I. J. *Chromatogr.* 1976, 125, 129-156.
100. Kissinger, P. T. *Anal. Chem.* 1977, 49, 883.
101. Bidlingmeyer, B. A. *J. Chromatogr. Sci.* 1977, 18, 525-539.
102. Sullivan, K. M.; Mannuci, A.; Kimpton, C. P. Gill, P. *BioTechniques*, 1993, 15, 636-641.
103. Hirabayashi, J.; Kasai, K. *Anal. Biochem.*, 1989, 178, 336-341.

PROPERTY OF
National Criminal Justice Reference Service (NCJRS)
Box 6000
Rockville, MD 20849-6000