IS-T--1803

DNA Typing by Capillary Electrophoresis

by

Zhang, Nanyan

MS Thesis submitted to Iowa State University

Ames Laboratory, U.S. DOE
Iowa State University
Ames, Iowa 50011

Date Transmitted: October 8, 1997

PREPARED FOR THE U.S. DEPARTMENT OF ENERGY
UNDER CONTRACT NO. W-7405-Eng-82.

DISCLAIMER

Table of Contents

GENERAL INTRODUCTION

Thesis Organization

This dissertation begins with a general introduction of the background and theory. Recent research and progress are also provided in the literature. One research paper follows the general introduction with their literature cited. General conclusions summarize the work. Finally, a list of cited references for the general introduction concludes this dissertation.

Background

Characterization, or "typing," of blood, serum, and other body fluids has been used for forensic and clinical purposes for more than 50 years.[1] In the last decade, methods have become available for deoxyribonucleic acid (DNA) typing, that is, for showing distinguishing differences in the genetic material itself. Developments in genetics have accumulated large amount of knowledge about human genome. The mordern molecular biology has built a lot of advanced biotecnologies. These knowledge and advances have provided a sufficient number of regions of DNA which can be chracterized. DNA typing is becoming a powerful tool in lots applications such as

absolute personal identification in forensic testing, in cell line authorization, tissue transplantation and disease diagnosis, etc.

Genetic Basis of DNA Typing

The human genome consists of DNA molecules in the form of a double helix in which the two strands of the DNA duplex are held together by weak hydrogen bonds. Each strand has a linear backbone of residues of deoxyribose (a 5-carbon sugar) which are linked by covalent phosphodiester bonds. Covalently attached to carbon atom number 1' of each sugar residue is a nitrogenous base, either a pyrimidine [cytosine(C) or thymine (T)], or a purine [adenine (A) or guanine (G); see Figure 1[2]]. A sugar with an attached base and phosphate group therefore constitutes the basic repeat unit of a DNA strand, a nucleotide. Genetic information is encoded by the sequence of bases in the DNA strands. Hydrogen bonding occurs between laterally opposed bases, base pairs, of the two strands of a DNA duplex according to Watson-Crick rules: A to T and C to G. The total number of nucleotides is about 3 billion in human genome. Genes are segments of the DNA molecules. Much of the DNA, the part that separates genes from one another, is noncoding.

A human has 23 pairs chromosomes. In each pair of chromosomes, one is inherited from mother , another from father. Each of the chromosomes is composed of a

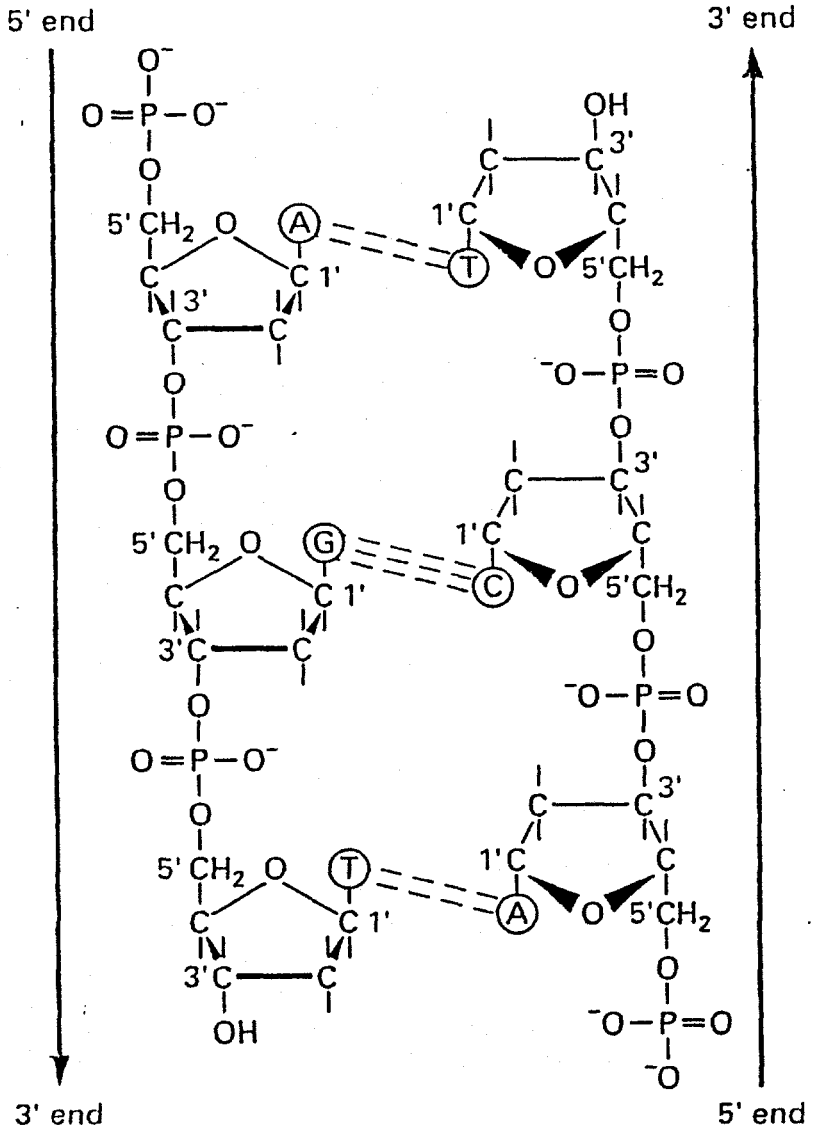

Figure 1. Structure of Double-stranded DNA

A = Adenine, C = Cytosine, G = Guanine, T = Thymine

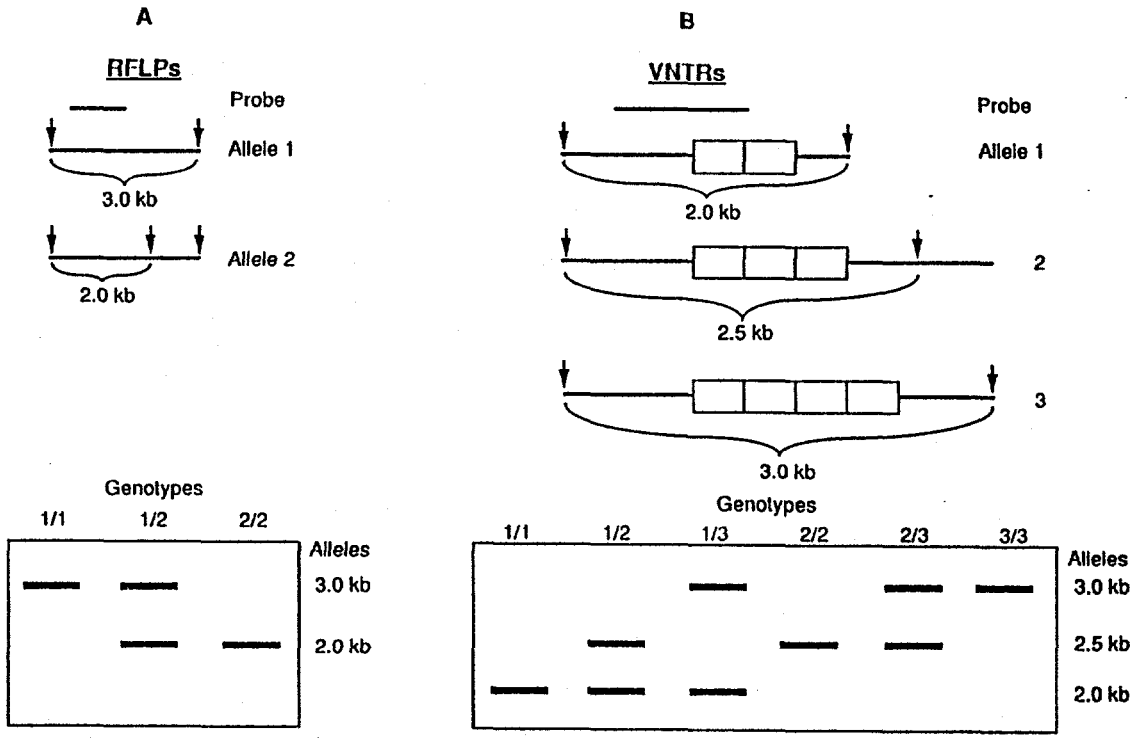

Figure 2. Two types of RFLPs Structure of alleles in chromosome is diagrammed at top; arrows indicate sites of cutting by enzyme; lengths of fragments demonstrated by probe. (short line above) are given. Electrophoretic patterns are diagrammed below. A. Diallelic RFLP system resulting from single nucleotide change. Electrophoretic patterns are those of three genotypes: homozygotes for either allele 1 or allele 2 and 1/2 heterozygote. B. Multiallelic VNTR system. With three alleles as diagrammed, three are six possible genotypes as demonstrated by electrophoresis.

long DNA molecule constructed as a double helix. A pair of DNA sequences, which are located at identical positions on each pair of chromosomes (except sex chromosomes X and Y) are referred to as alleles. An individual is said to be homozygous or heterozygous at a specific locus if the two alleles at that locus are, respectively, identical or different sequences.

DNA technology has revealed variations in the genome. In noncoding regions of DNA, it is estimated that at least one nucleotide per 300-1,000, on the average, varies between two people.[3] The nucleotide difference might change the recognition site for a particular site-specific endonuclease (restriction enzyme). Some regions of DNA contain repetitive units, multiple identical strings of nucleotides arranged in tandem. In VNTRs (variable number of tandem repeats), the number of repetitions of a sequence can vary from person to person. The repeating unit can be as small as a dinucleotides-e.g., the (TG)n polymorphism-or as large as 30, or even more nucleotides. Tandem repeats are not limited to noncoding segaments of DNA, although they are found less frequently in coding segaments.The two main types of variation : single-nucleotide differences and VNTRs , are both potentially recognizable by change in the lengths of fragments that result when DNA is cut with a restriction enzyme. The principle of restriction fragment lenght polymorphism is shown in Figure 2.[1]

Technological Basis of DNA Typing

The tools of DNA typing include restriction enzymes, electrophresis, probes, and the polymerase chain reaction.

Restriction Fragment Length Polymorphism (RFLP)----Southern Blotting

In this RFLP approach shown in Figure 3, DNA is subjected to controlled fragmentation with restriction enzymes that cut double-stranded DNA at sequence-specific positions. The long DNA molecules are thereby reduced to a reproducible set of short pieces called restriction fragments (RFs), which are usually several hundred to several thousand basepairs long. Many hundreds of thousands of fragments producecd by digestion of human DNA with a single restriction enzyme; each fragment has a distinct sequence and length. For analysis of RFs to demonstrate RFLPs, the fragments are separated electrophoretically on the basis of size. Electrophoresis, typically performed on agarose or acrylamide gels, results in large fragments at one end and small fragments at the other. The fragments are denatured (i.e., rendered single-stranded), neutralized, and transferred from the gel to a nylon membrane, to which they are fixed; this facilitates detection of specific RFLPs and VNTRs. RFLPs that are defined by specific sequences are detected by hybridization with a probe, a short segament of single-stranded DNA tagged with a group such as radioactive phosphorus, that is used to detect a particular complementary DNA sequence. The nylon membrane is placed in a bath that contains the

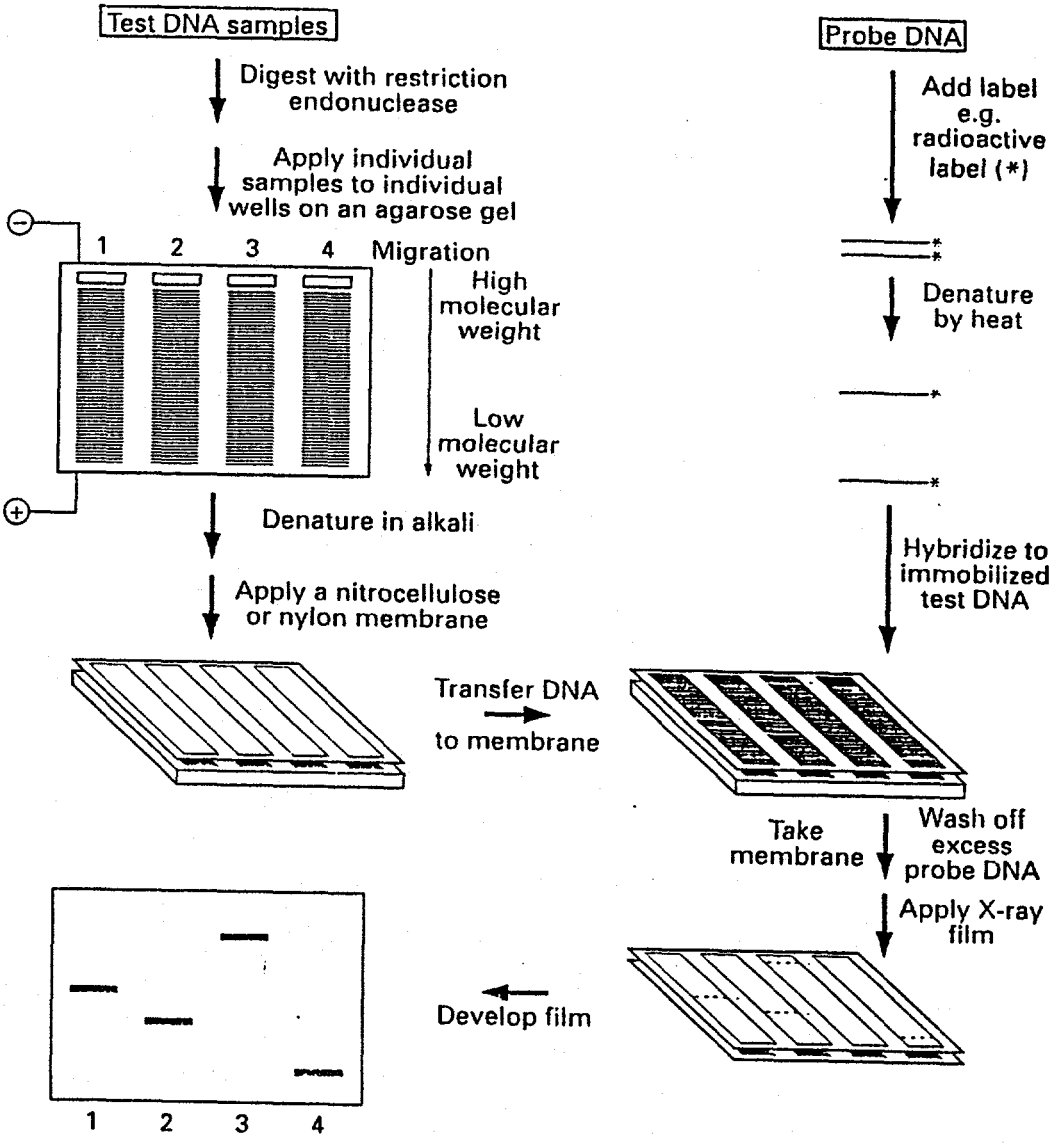

Figure 3. Southern blot hybridization

probe, and the probe hybridizes to the target denatured RF. Nonspecifically bound probe is washed off. The pattern of probe binding is visualized with autoradiography.

The complete process: DNA digestion, electrophoresis, membrane transfer, and hybridization, usually needs several days to complete. These procedures are routinely used in molecular biology, biochemisrty, genetics, and clinical DNA diagnosis. RFLP analysis with single-locus probes is usually designed to result in a simple pattern of one or two RFLP bands, depending on whether the person is homozygous or heterozygous, respectively. The range of variation shown in the patterns from different persons depends on how many different alleles exist at the particular target locus, e.g., how many different tandem repeats are in population as a whole.[1]

Polymerase Chain Reaction (PCR) for Amplifying DNA

Techniques for analyzing DNA are changing rapidly. One key technique introduced about a decade ago is the polymerase chain reaction, which allows a million or more copies of a short region of DNA to be easily made. For DNA typing, one amplifies a geneticaly informative sequence, usually 100-2,000 nucleotides long, and detect the genotype in the amplified product. Because large quantity of pure material is made by PCR, DNA typing can rely on various detection methods other than using radio active substance. The variability of PCR technique itself and the technical improvement of salb gel electrophoresis affords more options of analysis methods. Meanwhile the rapid

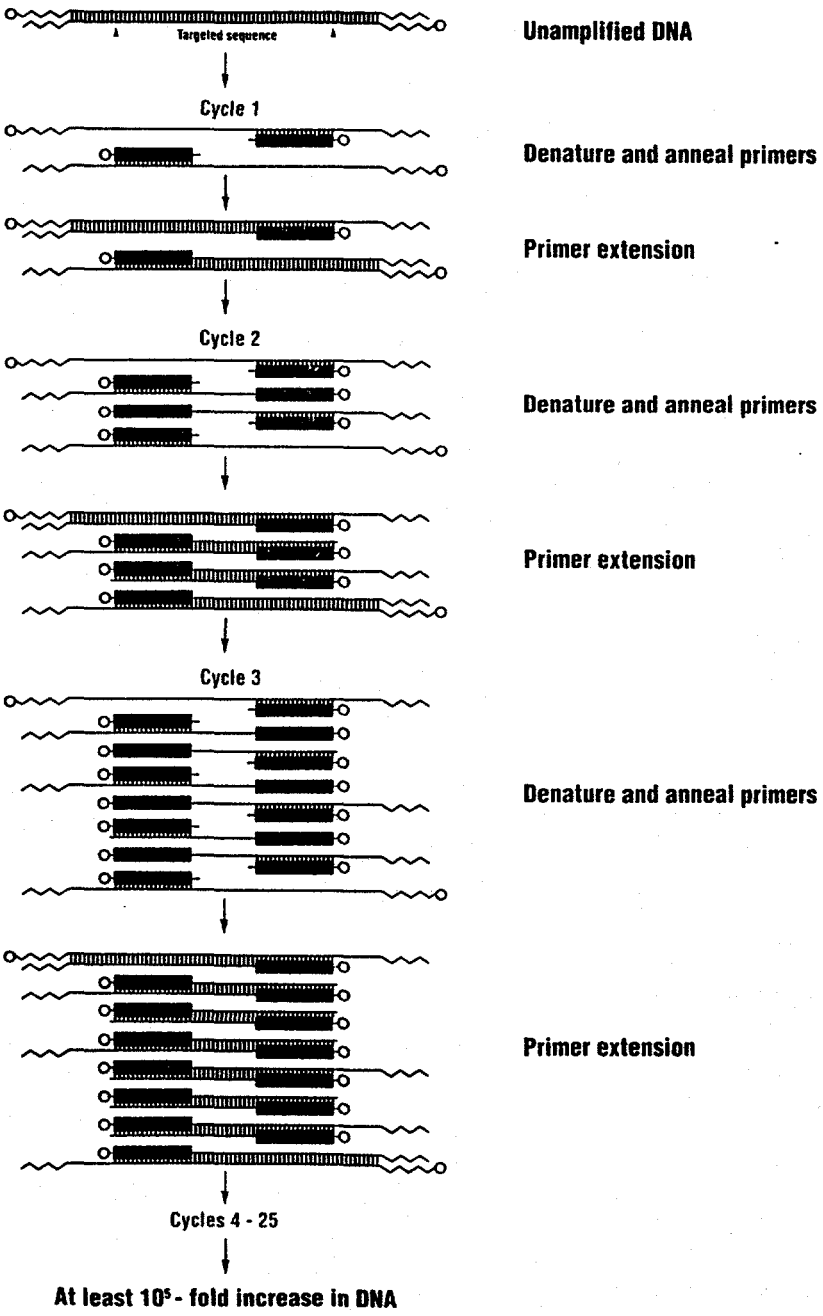

Figure 4. Polymerase Chain Reaction (PCR)

progress of capillary electrophoresis (CE) provides a powerful tool to genetic analysis. DNA typing becomes faster and more sensitive based on CE. Another important advantage of PCR technique is speed and sensitivity. It permits the use of very small sample of tissue or body fluids-theoretically even a single nucleated cell. It gets rid of the intensive labor needed by RFLP.

The PCR process (Figure 4) is simple; indeed, it is analogous to the process by which cells replicate their DNA.[4-5] Two short oligonucleotides are hybridized to the opposite strands of a target DNA segament in positions flanking the sequence region to be amplified. The two oligonucleotides serve as primers for an enzyme-mediated replication of the target sequence. The PCR amplification process itself consists of a three-step cycle:

1. The double-srtranded template DNA is dissociated into single strands by incubation at high temperature, typically 94°C.

2. The temperature is lowered to allow the ologonucleotide primers to bind to their complementary sequences in the DNA that is to be amplified.

3. A DNA polymerase extends the primers from each of the two primer-binding sites across the region between them, with the target sequence as template.

Because the extension products of one primer bind the primer in successive cycles, there in principle a doubling the target sequence in each cycle. However, the efficiency of amplificatrion is not 100%, and the yield from a 30 cycle amplification is generally about 10^6-10^7 copies of the target sequence. The efficiency can be improved by amplifying

several different products in the same reaction mix; this is termed multiplex amplification. Sveral methods have been coupled with PCR for the detection of genetic variation in the amplified DNA. Some are listed in Table 1.[7-13]

Table 1. Some PCR-Based Systems for the Detection of Genetic Variation

Sequence-based detection systems
Allele-specific oligonucleotide (ASO)[7]
Allele-specific priming of PCR[8]
Oligonucleotide-ligation assay (OLA)[9]
Restriction-site-specific cleavage (Amp-FLPs)[10]
Denaturing gradient gel electrophoresis[11]
Chemical cleavage of mismatched heteroduplxes[12]
Length-variation systems
Simple insertions and deletions
VNTR polymorphisms[13]
Analysis of nucleotide sequences

Capillary Gel Electrophoresis for DNA Analysis

General Introduction of Capillary Electrophoresis

Capillary electrophoresis (CE) is a modern analytical technique which permits rapid and efficient separations of charged components present in small sample volumes. Separation are based on the differences in electrophoretic mobilities of ions in electrophoretic media inside small capillaries with 10 -200 μm inner diameter.[14] CE has showed a lot of applications in chemical, biological ,biomedical and pharmaceutical applications. CE offers clear advantages over slab-gel electrophoresis in terms of speed, ease of automation and quantitation. CE provides efficiencies up to two orders of magnitude greater than high-performance liquid chromatography (HPLC). The distinct capillary electroseparation methods include:

(A) Capillary zone electrophoresis (CZE)

(B) Capillary gel electrophoresis (CGE)

(C) Micellar electrokinetic capillary chromatography (MEKC or MECC)

(D) Capillary electrochromatography (CEC)

(E) Capillary isoelectric focusing (CIEF)

(F) Capillary isotachophoresis.

Capillary Gel Electrophoresis for DNA Separation and Detection

Review of Separation Matrices

The main separation mechanism in capillary gel electrophoresis (CGE) is based on differences in solute size as analytes migrate through the pores of the gel-filled column. Gels are potentially useful for electrophoretic separations mainly because they permit separation based on "molecular sieving". Futher more, they serve as anti-convective media, they minimize solute diffusion, which contribute to zone broadening, they prevent solute adsorption to the capillary walls and they help to eliminate electroosmosis. Capillary gel electrophoresis has bcome the norm for nucleic acids analysis since the first appearance of DNA analysis by CE in 1988[15,16]. Since the size to charge ratio is constant for nucleic acid in free solution; they can not be separated in CZE mode. High separation efficiency can be acheived by employing appropriate gel matrices for specific DNA sample.

Two types of gel matrices can be distinguished : 1) a relatively high-viscosity, crosslinked gel that is chemically anchored to the capillary wall ("chemical" gel), and 2) a relatively low-viscosity, polymer solution (physical gel). Table 2[17] summarizes the main differences between these gels.

The typical example of crosslinked gel is polyacrylamide. Karger and Cohen[40] have made significant contributions in demonstrating the extremely high separation

Table 2. Characteristics of Gel Matrices Used in CGE

Chemical Gels
Crosslinked and/or chemically linked to the capillary wall
Well-defined pore structure
Pore sizes cannot be varied after polymerization
Heat Sensitive
Particulates can damage thr gel matrix
Not replaceable; general high viscosity

Physical Gels
Not crosslinked or attached to the capillary wall
Entangled polymer networks of linear or branched hydrophilic
Dynamic pore structure
Pore size can be varied
Heat insensitive
Particulates can be easily removed
Gel is replaceable

efficiency of polyacrylamide gel filled capillary. The polymerization of the high viscosity, linear polyacrylamide is carried out within the capillary. Bifunctional reagents are necessary to provide linkages to both the capillary wall and the polymer gel matrix for the reason of stability. The polymerization mixture which composed of monomer, cross-linker, initiator and denaturing reagent are carefully filled into capillary. The

polymerization process needs relative long time and careful handling. The pore size of the polymer are determined by the T% and C%[14] When applications requiring low electric field (<300 V/cm), the bifunctional reagent may not be required. The crosslinked PA gel is usually used for high resolution single stranded oligonucleotides separation and DNA sequencing under denaturing condition 7-9 M urea.

The non-crosslinked, replaceable polymer networks have a dynamic pore structure and are more flexible. Polymer networks of variable viscosity can be made by carefully selecting the concentration and chain length of the linear polymers. The often used polymers are: hydroxyethylcellulose (HEC)[18-19], hydroxymethylcellulose (HMC)[20-22], hydroxypropylmethylcellulose (HPMC)[23-25], polyacryloylaminoethoxyethanol[26,27], ficoll-400[28], polyethyleneglycol[20,23], glucomannan[29], polyvinyl alcohol[30], poly(ethylene oxide) (PEO)[31] . A fresh gel can be used for every sample injection. Also higher temperature (50-70°C) and electric field can be used without damaging the gel as would be the case with chemical gels.

Column Coating and Poly(ethylene oxide) gel matrix.

Precoated capillary is desirable for capillary gel electrophoresis. The role of coating here is to suppress the elctroosmotic flow which if exists, will damage the gel matrix. Commercially available polysiloxane coated capillaries (e.g., DB-1 from J & W Scientific) can be used for this purpose, as well as others such as those coated with

polyvinyl alcohol[32] or polyacrylamide[33]. The most widely used coating method in research lab is γ-methacryloxy propyltrimethoxysilane and polyacrylamide developed by Hjerten[34]. Also dynamic coating with the cellulose additives[23] and PEO[35] was found to reduce the EOF in certain degree.

No column coating was found necessary when 1.5 % 8 000 000 and 1.4% 600 000 PEO gel was used for DNA sequencing separation, if the PEO gel was filled into the capillary after the capillary was flushed with 0.1M HCl first.[36] More understanding about the capillary treatment has been obtained since that for PEO gel. Methanol was also used to flush the bare capillary before gel was filled. No electroosmotic flow was observed during separation. Generally, if coating of capillary wall is necessary is needed to suppress the EOF and ensure the separation performance is determined by the viscosity, physical and chemical nature of the gel matrices and the pH of the buffer employed.

Resolution and Efficiency of Gel-Filled Colum

The following formula was used for the calculation of resolution:

$R = dt / 4\sigma_t$

where dt is the difference in time of elution between two consecutive peaks (differing by one nucleotide) and σ_t is the standard width of a single peak.

The number of theoretical plates, given by:

$N = (t / \sigma_t)^2$

t is the total time for a given species to elute.

It was found that capillary gel electrophoresis was capable of providing superior performance not only in terms of faster analysis time (3x), but also better resolution (2.4x), as well as higher separation efficiency (5.4x) than the conventional automated slab gel instrument.[14]

Sample Injection and Matrix Effect; Quantitation

In CE, the pressure injection mode is generally recommended for quantitative work: the composition of the sample plug introduced into the capillary is exactly that of the sample vial from which the injection took place. With replaceable gels, both injection modes can be used. Only electrokinetic injection can be employed with high viscosity crosslinked gel. Electrokinetic injection often yields more efficient peaks than does pressure injection. When electrokinetically injected from low ionic strength sample solutions, DNA fragments are effectively stacked against the viscous, polymer network medium. No sample bias should be expected for electrokinetic injection of DNA sample which has various fragments sizes, because DNA fragments essentially have the same mass-to-charge ratio in free solution, they should migrate into the capillary with the same mobility.

The matrix strongly effects the amount of DNA sample to be injected , so as to effect the separation performance, especially variable amount of salt are present.

Desalting is necessary sometimes to get good injection. The ultrafiltration procedure removes low MW sample constituents as demonstrated by Schwartz et al[21], resulting in efficient DNA peaks. However, loss of DNA due to adsorption on filter can occur[36]. So, purely for quantitation, pressure injection are preferable. An external standard with same matrix can be used for quantification .

Detection for Capillary Gel Electrophoresis

The requirements of the good detection are universal to all CE modes including CGE. To preserve the separation efficiency, the detection volume should be small enough (10% of peak volume)[37]. Other than small cell volume, a good detector should provide a high sensitivity, large dynamic range, and fast response.

Two major optical detection methods are used: UV absorbance and laser induced fluorescence detection. The detection limit with UV detection is around 0.5 μg/ml[14]. LIF is the most sensitive detection method so far. Figure 5 shows the schematic diagram of the LIF-CE detection. The very high sensitivity comes from the high intensity of laser and the efficient coupling to the capillary core[37]. Limit of detection was 0.01 amol (≅ 10^{-11} M), which corresponded to approximately 6000 molecules of fluorescently tagged DNA[14]. A number of detection schemes have been used for the detection of nucleic acids by fluorescence. On approach is based on the native DNA fluorescence in the low UV-region[38]. This allows analysis of the DNA molecule in its "natural "state. The most

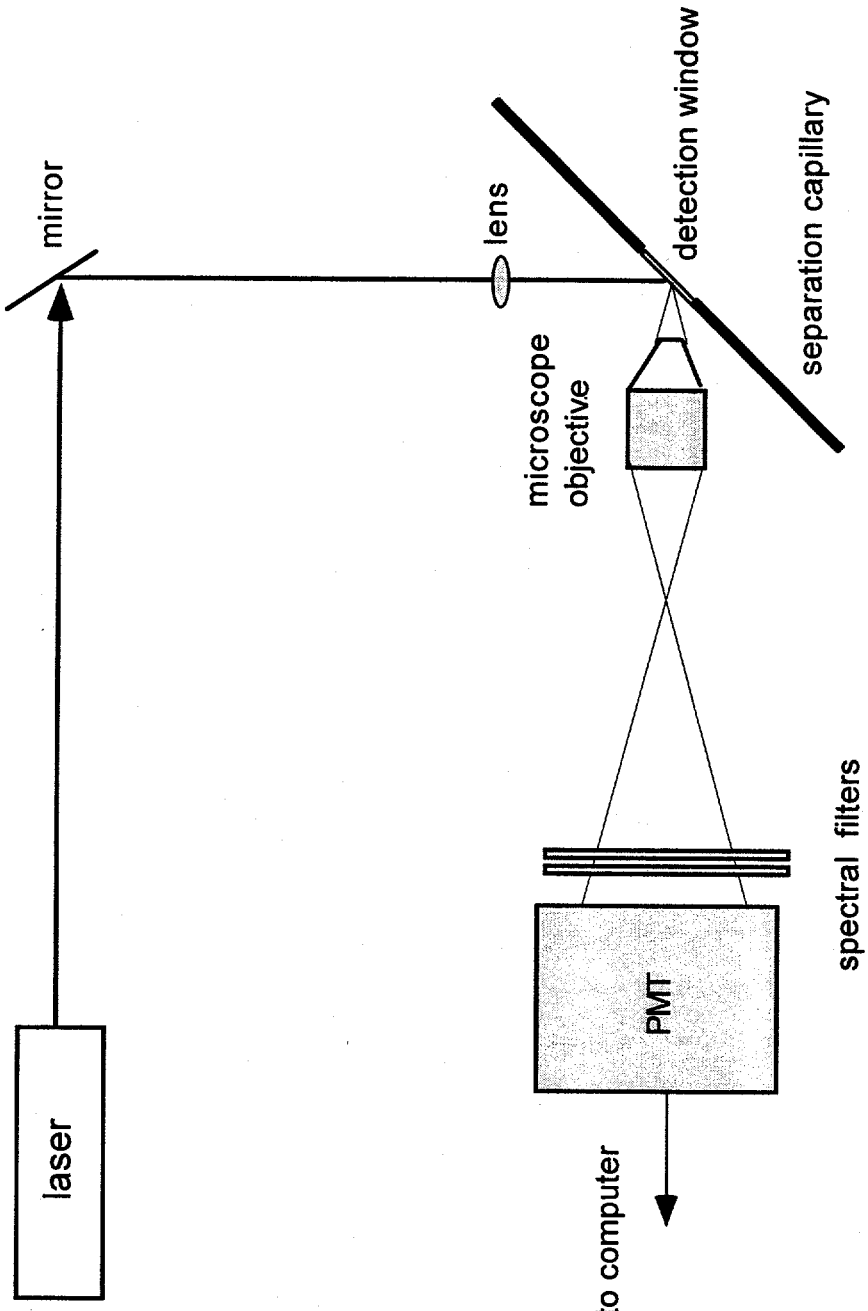

Figure 5. Schematic diagram for laser induced fluorescence detection

straight forward LIF detection scheme involves the use of fluorescent intercalators. The dye is added to the CE buffer and / or sample and specially interacts with sample dsDNA or RNA molecules. The dye insert between the base pairs of DNA, providing enhanced resolution. The DNA-dye complex fluoresces strongly when excited by appropriate laser. The final LIF detection scheme for DNA involves direct labeling of the analyte with a suitable fluophore. Fluorescently labeled probes and primers are used in many molecular biology applications involving hybridization and PCR. DNA primers and probes are usually synthesized with a fluorescent label attachment of a dye using commercial DNA labeling kit.

Pages 21-43 — Reprint - Removed for separate processing

GENERAL CONCLUSIONS

Capillary electrophoresis is becoming more and more important in nucleic acid analysis including DNA sequencing, typing and disease gene measurements. This work summarized the background of DNA typing. The recent development of capillary electrophoresis was also discussed. The second part of the thesis showed the priciple of DNA typing based on using the allellic ladder as the absolute standard ladder in capillary electrophoresis system. Future work will be focused on demonstrating DNA typing on multiplex loci and examples of disease diagnosis in the on-line format of PCR-CE. Also capillary array electrophoresis system should allow high throughput, fast speed DNA typing.

REFERENCES

1. Committee on DNA Technology in Forensic Science, *DNA Technology in Forensic Science,* National Academy Press, Washington, D.C., **1992**
2. Strach, T. *The Human Genome,* **1992**, Bios Scientific Publishers Limited
3. Longo, M., Berninger, M. Harley, J. *Gene*, **1990,** 93, 125
4. Erlich, H. *PCR Technology: principles and applications for DNA amplification*, New York, Stockton Press, **1989**
5. Rose, E. A. *Applications of the Polymerase Chain Reaction to Genome Analysis*,

FASEB J., **1991**, 5, 46-54

6. Arnheim N. , Levenson, C.H. *Polymerase Chain Reaction*, Chem Eng News, **1990** 68, 36-47
7. Conner, B.J., Reyes, A.A., Morin, C., Itakura, K. Teplitz, R.L., Wallace, R.B. *Proc. Natl. Acad. Sci. USA* , **1983**, 80, 278-282
8. Chehab, F.F., Kan, Y.W. *Proc. Natl. Acad. Sci. USA* , **1989**, 86, 9178-9182
9. Nickerson, D.A. Kaiser, R., Lappins, Steward, J., Hood, L., Landegren, U. *Proc. Natl. Acad. Sci. USA* , **1990,** 87, 8923-8927
10. Kasai, K., Nakamura, Y., White, R. *J. For. Sci.*,**1990**, 35, 1196-1200
11. Myers, R.M., Maniatis, T., Lerman, L.S. *Methods Enzymol.*, **1987**, 155, 501-527
12. Cotton, RGH, Rodriques, N.R., Campbell, K.D. *Proc. Natl. Acad. Sci. USA*, **1988,** 85, 4397-4401
13. Nakamura, Y., Lepport, M.,*Science,* **1987**, 235, 1616-1622
14. Li, S.F.Y. *Capillary Electrophoresis*, **1992**, Elsevier Science Publishers B.V.
15. Cohen, A.S., Najarian, P.R., Paulus, A., Guttman, A., Smith, J.A., Karger, B.L. *Proc. Natl. Acad. Sci. USA,* **1988,** 85, 9660
16. Kasper, T.T., Melera, M., Gozel, P., Brownleo, R.G. *J. Chromatogr.* **1988**, 458, 303
17. Schwartz, H., Guttman, A. *Separation of DNA by Capillary Electrophoresis*, Beckman, 1995
18. Grossman, P.D., Soane, D.S. *J Chromatogr.* **1991**, 559, 257
19. Nathakarnkitkool, S., Oefer, P., Bartsch, G., Chin, M.A., Bonn, G.K.

Electrophoresis ,**1992**, 13, 18

20. Zhu, M.D., Hansen, D.L., Burd, S., Gannon, F. *J. Chromatogr.*, **1989**, 480, 311

21. MacCrehan, W.A., Rasmussen, H.T., Northrop, D.M. *J. Liq. Chrom.*, **1992**, 15, 1063

22. Kim, Y., Morris, M.D. *Anal. Chem.* **1994** ,66, 1168

23. Schwartz, H. E., Ulfelder, K.J., Chen, F.-T.A., Pentoney, S.L., Jr. *J. Cap. Elec.* **1994,** 1, 36

24. Ulfelder, K., Schwartz, H.E., Hall, J.M., Sunzeri, F.J. *Anal. Biochem.* **1992,** 200, 260

25. Baba, Y., Ishimaru, N., Samata, K., Tsuhako, M. *J. Chromatogr.*, **1993**, 558, 273

26. Chiari, M., Nesi, M., Righetti, P.G. *Electrophoresis*, **1994**,15, 616

27. Nesi, M., Righetti, P.G., Patrosso, M.C., Ferlini, A., Chiari, M. *Electrophoresis,* **1994**, 15, 644

28. Righetti, P.G., Chiari, M. *Electrophoresis*, **1991**,12,55

29. Izumi, T., Yamaguchi, M., Yoneda, K., Okuyama, T., Shinoda, T. *J. Chromatogr.* A **1993**,652, 41

30. Kleemis, M.H., Gilges, M., Schomburg, G. *Electrophoreisis,* **1993**, 14,515

31. Chang, H.T., Yeung, E.S. *J. Chromatogr ., B* **1995**, 669, 113-123

32. Schomburg, G., *Capillary Electrophresis*: Theory and Practice Camillieri, P.(Ed). Boca Raton, CRC press, **1993**

33. Strege, M., Lagu, A., *Anal. Chem.* **1991,** 63, 1233

34. S. Hjerten, *J. Chromatogr.* **1985**, 347, 191

35. Iki, N.; Yeung, E.S., *Anal. Chem.* **1995,** 67, 1913-1919

36. Fung, E.N.; Yeung, E.S. *Anal. Chem.*, **1995**, 67, 1913-1919

37. Butler, J.M., McCord, B.R., Jung, J.M., Allen, R.O. *BioTechniques*, **1994**,17, 1062

38. E.S. Yeung, *Adv. Chromatogr.*, **1995**, 35, 1

39. McGregor, D.A., Yeung, E.S. *J. Chromatogr. A,* **1994**, 680, 491

40. Cohen, A.S., Najarian, D.R., Paulus, A., Guttman, A., Smith, J. A., Karger, B.L. *Proc.Natl. Acad. Sci. USA* **1988**, 85, 9660

ACKNOWLEDGMENTS

My greatest gratitude goes to my major professor, Dr. Edward S. Yeung. I thank him for providing me an exciting environment for doing research. I have learned not only chemistry itself, but also the way toward a successful scientist with high self motivation and confidence. My experience here will benefit through my whole life.

I would like to thank Dr. Robert S. Houk and Dr. Donald J. Graves for being on my committee and reviewing the entire thesis.

Thanks also goes to all the members in Dr. Yeung's group. I thank them for the wonderful time we spent together.

I should thank my husband Qing Tang. Without his encouragement, I can not do these only.

Nothing in words can express my gratitude to my father, Jian Zhang; mother, Yue-e Jiang . Their dearest love follows me everywhere at any time although I live far away from them.

This work was performed at Ames Laboratory under Contract No. W-7405-ENG-82 with the U.S. Department of energy. The United States government has assigned the DOE Report number IS-T 1803 to this thesis.