Abstract

Abstract of thesis entitled

"Development of A Cost-effective Image Analysis System and Its Applications in Electrophoretic Separations"

submitted by CHAN Cheok-man, Andy

for the degree of Master of Philosophy in Chemistry

at The Hong Kong Polytechnic University in October. 1999.

Owing to the rapid advancement in microelectronics. image analysis is now widely applied to the fields of sciences. industry and medical treatment. In this research. image analysis was applied to study and to enhance separation techniques in chemical and biochemical studies.

Two-dimensional crossed immunoelectrophoresis (2D-CIEP) is a technique widely used for studying composition of protein mixtures in biological and clinical studies. A low cost image analysis system with the use of an optical flat-bed scanner and an IBM-compatible PC was set up in this work for capturing 2D-CIEP patterns. A computer package CIEPEASY was developed for modification and analysis of the acquired image to determine important peak parameters such as the migration

distance, peak height and peak area of the constituting components for both qualitative and quantitative studies. In this approach, more composition information of the standard and sample gel patterns could be extracted from the proposed image analysis system. In addition, the running cost of data analysis is low compared to conventional methods. The time required for data collection and interpretation of 2D-CIEP images was shortened significantly and the results obtained have a higher accuracy than those obtained by using conventional methods. Moreover, a linear relationship between the peak area and the amount of antigen present in a sample was confirmed accurately and reported for the first time in the literature. The developed system was applied to study abnormal 2D-CIEP patterns of prothrombin in patients with lupus anticoagulant. It was found that patients with a previous history of thrombosis showed significantly reduced peak heights. Further investigations are required to verify whether this difference is an indication of an increased of thrombosis.

Deoxyribonucleic acid (DNA) fingerprinting for personal identification is a powerful tool for criminal investigation for judgement in forensic science. In this research, the proposed image analysis system with the developed software package DNAEASY were utilized to enhance DNA fingerprinting analysis in silver stained electrophoretograms. Matching of the DNA alleles of the suspect and the reference standard is able to perform automatically. Intensities of alleles can be acquired for improving the judgement of stutter bands and sex identification function was provided. By comparing with the conventional vistual inspection method, more objective judgements can be obtained.

Development of A Cost-effective Image Analysis System and Its Applications in Electrophoretic Separations

by

CHAN Cheok-man, Andy

MASTER OF PHILOSOPHY

in

CHEMISTRY

at

The Hong Kong Polytechnic University

1999

Declaration

This is to declare that this work has been done by the author in the Department of Applied Biology and Chemical Technology of The Hong Kong Polytechnic University and this thesis has not been submitted to this or other institution for any academic qualification.

CHAN Cheok-man, Andy

Date: 7 October, 1999

Acknowledgment

I would like to thank the Research Committee of The Hong Kong Polytechnic University for providing a research fund (Grant No. V286) for support of this project. I also wish to express my gratitude to all those who helped in various ways during the course of this work. I would like to thank Prof. F.T. Chau, my supervisor, for his critical advice and supervision of this project as well as his comments on the final draft of this thesis and the publications. I am indebted to Dr. S.C.L. Lo, my co-supervisor, for contributing 2D-CIEP gel plates and invaluable comments on the project. Thanks are also given to Dr. A.K.M. Leung for his technical advice and assistance.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2D-CIEP	Two-dimensional crossed immunoelectrophoresis
AT-III	Antithrombin-III
BMP	Windows Bitmap Format
bp	Base pair
CCD	Coupled charged device
CD-ROM	Compact disk read-only-memory
CPU	Central Processing Unit
CT	X-ray computer tomography
DAF	DNA amplification fingerprinting
DIB	Windows Device Independent Bitmap Format
DIC	Differential interference contrast
DNA	Deoxyribonucleic acid
dpi	Dot per inch
FWHM	Full-width-at half-maximum
GCMS	Gas chromatography mass spectrometry
GIF	CompuServe Graphics Interchange Format
HMWK	High molecular weight kininogen
JPG	Joint Photographic Experts Group Format
LA	Lupus anticoagulant
LCMS	Liquid chromatography mass spectrometry
MRI	Magnetic resonance imaging

PC	Personal Computer
PCR	Polymerase chain reaction
PCR-STR	Polymerase chain reactions-based polymorphisms in short tandem repeat
PCX	ZSOFT Bitmap Format
PKK	Prekallikrein
PLG	Plasminogen
PT	Prothrombin
PT-2D-CIEP	Prothrombin 2D-CIEP
%RSD	Percentage of relative standard deviations
SAR	Synthetic aperture radar
SIMS	Secondary ion mass spectrometry
SLE	Systemic Lupus Erythematous
SMM	Slow-moving material
STR	Short tandem repeat
TBE	Tris-Borate-EDTA
TGA	True Vision Targa Format
TIFF	Tagged Image File Format
TLC	Thin layer chromatography
vWF	Von-Willebrand's factor

CHAPTER 1

Introduction

1.1 Background

Images acquired from different sources contain a large amount of essential information. Image analysis is simply a process to extract useful information from pictures (Glasbey and Horgan, 1995). Over the twenty years, the rapid development of microelectronics and microcomputer systems facilitates the use of image analysis techniques in different fields of science. In the past, both qualitative and quantitative image information generally obtained by human visual inspections and measurements which are repetitive tasks and time consuming. So, when the speed and storage capacity of a computer were greatly improved, image analysis applications of industrial, military, medical, chemical and biochemical areas are developed rapidly. In the change of the image analysis being carried out by human to computer, analysis time is shortened dramatically and human errors are minimized.

Scientists have developed applications of image analysis in document and map data analysis, medical image analysis, satellite image understanding, robot vision and fingerprint analysis (Kasturi and Trivedi, 1990). Moreover, it was also used to study the detection and analysis of human motion (Tsukiyama and Shirai, 1985). Much interest exists in applying some of the newer artificial intelligence techniques to pattern recognition (Freeman, 1986). In the textile technology, image analysis has been successfully applied in measuring fiber crimp, cotton maturity, fiber cross section shape and fiber damage. Application of imaging techniques to evaluating carpet appearance and its change with wear started in the early 1980's and is getting prosperous recently (Xu, 1994).

In the biochemical and biological science, image analysis has been employed, for instance, in the following areas: (1) to count and measure algal cells in an microscope image (Martin and Fallowfield, 1989); (2) to obtain numbers and sizes of muscle fibers in a cross section image for the development of a drug that enhances muscle development (Maltin *et al.*, 1989); (3) to study porosity and pore-size distribution within a sample of soil aggregates (Darbyshire *et al.*, 1989; Glasbey *et al.*, 1991); (4) to measure the fiber diameters automatically in an image of cashmere goat fibers in support of a goat breeding programme (Russel, 1991); (5) to understand the spatial structure of the fungal hyphae in relation to their environment (Ritz and Crawford, 1990); (6) to summarize the statistics on size and shape for making comparisons between different fish species (Strachan *et al.*, 1990) and so on.

In the separation chemistry, image analysis has been applied to the most sophisticated analytical instruments such as gas chromatography coupled to mass spectrometry (GC/MS) and liquid chromatography coupled to mass spectrometry (LC/MS). These involve the analysis of the chromatograms and matching of mass spectra library. Moreover, it has also been utilized in thin layer chromatography (TLC) for qualitative and quantitative analysis. A modern TLC scanning system such as the Camag TLC Scanner II with Labdata Station AT and CATS software (Fried and Sherma, 1994) can acquire TLC images, and perform calibration and quantitation automatically. For the application of secondary ion mass spectrometry (SIMS) in surface chemistry, image analysis technique is adopted to denoise the SIMS images (Leung *et al.*, 1998). For example, some researchers reported an application of a new image analysis technique, wavelet transform, for denoising SIMS images in recent year (Hutter *et al.*, 1996; Nikolov *et al.*, 1996).

1.2 Introduction to the Research Work

In this research, image analysis methods were developed and applied to the study of gel images as obtained from two-dimensional crossed immunoelectrophoresis (2D-CIEP) and DNA electrophoresis. In the past, the electrophoretic pattern was mainly analyzed by visual inspection. Nowadays, some of the instruments developed by laboratory and commercial companies are available. However, the price for such instruments is still expensive. In addition, there is a lack of image analysis system for 2D-CIEP and DNA electrophoresis specifically. Owing to the wide ranges of electrophoresis applications, image patterns of electrophoretograms obtained from different electrophoresis applications differ quite a lot and the information that has to be extracted from the these images is not the same. Commercial available image analysis software and systems are usually not able to suit specific single chemical and biochemical application. Development of automatically analytical system for these applications is still in demand.

Within this research work, specific image analysis system was developed to extract important information such as peak heights, peak areas, peak profiles and the migration distances from 2D-CIEP gel pattern. Such information can be used to determine the abnormality of patient protein in blood plasma. By applying this technique, more information that is essential can be obtained with higher accuracy and shorter time. While in the DNA electrophoresis, image analysis was employed in the field of pattern recognition for forensic purposes. This is known as DNA fingerprinting. It involves the comparison of the gel patterns from the suspect with

reference patterns. The degree of matching is usually subject to analyst determination. By the use of our developed image analysis system, the judgment will be much more objective and scientific.

In the above mentioned applications, gel images obtained from experiments were acquired and converted into digital form with the use of an optical scanner. Image analysis was then carried out on the digitized images by the specific developed software. Information was extracted, reported and archived by the software. Details will be discussed in the following chapters.

In Chapter 2, a brief introduction of image analysis will be given. This details the comparison between human and computer interpretation of images, the operations procedure of image analysis by computer and the hardware requirements for an image analysis system. Chapter 3 will describe the development of an image analysis technique to study 2D-CIEP gel patterns. Peak parameters were extracted by the developed software CIEPEASY. A linear relationship was found between the peak area and the amount of antigen present in a sample and is reported for the first time in the literature. With the help of the image analysis system and CIEPEASY, they were also applied to study abnormal 2D-CIEP patterns of prothrombin in patients with lupus anticoagulant. It was found that patients with a previous history of thrombosis had significantly reduced peak heights statistically. This will be described in Chapter 4. Chapter 5 will discuss the application of image analysis system in DNA fingerprinting. A software called DNAEASY was developed in this work to match the DNA alleles of the suspect and the reference. More objective

judgements can be achieved by this technique. Chapter 6 will give a brief conclusion of this research.

In conclusion, image analysis is the framework of this research work and thesis. Novel applications based on image analysis were studied, developed and applied to analyze images acquired from chemical and biochemical systems.

CHAPTER 2

Introduction to Image Analysis

2.1 Human and Computer Interpretation of Images

Visual imagery is one of the most important sensory inputs to the human perceptual system (Kasturi and Trivedi, 1990). When comparing the ability of vision system with computer, the human vision system is superb, particularly at qualitative tasks, for example to recognize a coarse picture even though it is not clear. However, for extracting quantitative information from images such as measuring a distance or counting an area within a region in an image, computers can do better than humans and can reduce the tedious aspects of image interpretation (Glasbey and Horgan, 1995).

Computers have been recognized as an effective and efficient tool for processing and analysis of high volumes of data since their development in the middle of this century (Kasturi and Trivedi, 1990). They are widely employed for a variety of applications such as image segmentation, object recognition, motion characterization, three dimensional feature analysis, and many other tasks associated with the broad area of scene understanding. In these applications, large volume of digital image data is produced and has to be manipulated. Therefore, computer offers an ideal platform for image analysis and measurement because it can perform simple and repeatable operations in very short time (Wojnar, 1999). Therefore, image analysis techniques are widely utilized in industrial, medical and scientific aspects of the society.

2.2 Image Analysis by Computer

Image analysis is simply the extraction of information from pictures (Glasbey and Horgan, 1995) and it is a process of image data transformation, leading to some actions, decisions or conclusion (Wojnar, 1999).

In the development of image analysis for a specific application, one generally has to employ image processing and pattern recognition techniques. Image processing refers to the manipulation of digitally encoded images and to convert them from a given form to another form that is more useful. The purpose of the conversion maybe to achieve a geometric transformation, to enhance the image so that its visual understanding by human is made easier, or to extract features of interest that will facilitate subsequence computer analysis. Pattern recognition is the ability to perceive structure in sensor-derived data, whether by humans, animals or computers. It is a fundamental activity that is intimately intertwined with our concept of intelligence. The sensor-derived data can be of any form that includes visual, tactile, sonic and so on (Freeman, 1986).

Image processing and pattern recognition are two closely related computer application areas, which together define the field of machine vision. There is an analogy here between a machine vision system and human vision – with computer image processing being analogous to the process that takes place in the human eye and the optic nerve, while pattern recognition representing more the visual perception activities that take place in the human brain (Freeman, 1986). Both image

processing and pattern recognition, however, also have their own distinct application areas. Image processing may be used purely to enhance an image, with no intent at subsequent machine interpretation. Similarly, pattern recognition has an application range that reaches far beyond visual images. It includes the interpretation of speech, seismic data, radio waves, and various other sensor derived signals not normally regarded as imagery (Freeman, 1986).

2.3 Basic Computer Hardware Components for Image Analysis

For an image analysis system, it is essential to have the following basic hardware components (Glasbey and Horgan, 1995).

2.3.1 A Central Processing Unit (CPU)

The central processing unit is the 'brain' of the computer. It is the main factor in determining the speed which a system works. Most computers have only one CPU. Computers with more than one are also built, and designed so that computing tasks can be shared out and performed on different CPUs. This is parallel computing. Many image analysis tasks involve repeating the same operation many times, and are well suited to parallel implementation.

2.3.2 A Display Monitor

The monitor is used to display the digital image for image analysis. Different monitors will differ in the number of pixels and the number of colors or greylevels that can be displayed. A monitor with an adequate number of pixels and number of color should be chosen.

2.3.3 Image Storage Device

For many applications, it will be necessary to store digital images so that they can be accessed on the computer in the future without the need to capture the image again. Computers usually store information on a hard disk which can be accessed quickly. Images for which rapid access is unnecessary can be stored on some other peripheral device. These include diskettes, cassettes, tapes of various sorts and various optical storage media. These are the commonly used backup archive media. Conventional diskettes can store only a little bit of information each, but are easy to use, convenient for transferring information between computers. Modern diskette available now can store up to 1000 mega bytes. Compact disk read-only-memory (CD-ROM) is an inexpensive way of storing a large amount of data or images, although it can be written only once.

In deciding what storage is to be used, the amount of storage needed should be considered. Digital images occupy a lot of space. If the images are to be $n \times n$ pixels in size, with b bytes per pixel, then each image will occupy at least bn^2 bytes.

For example, a 512 × 512 pixel image of 1-byte pixels will occupy 0.25 Mb. If many images are to be stored, consideration should be given with image compression techniques (Jain, 1989). These are ways of storing images in smaller amounts of storage space. They make use of the property that adjacent pixels tend to be similar, or identical, in value, and so image detail can be stored without the need to record each individual pixel. Compression has the disadvantage that images take longer to access. Some image analysis programs will offer compression as an option. Alternatively, stand-alone programs can be employed to compress images files as stored on the computer. They can then be decompressed before use.

2.3.4 Image Capture Devices

Any image, seen in the microscope or taken by any type of camera or other device has to be transformed into digital form that can be processed by computer. This process is called image acquisition (Wojnar and Majorek, 1994).

The image capture device is an equipment that provides the link between the real world and the digital image in a computer. It can be any machine that records a signal from the object under study and converts it to a set of pixel values. Such devices may be roughly divided into three types: electronic cameras, scanners and other image devices (Glasbey and Horgan, 1995).

Electronic cameras like the familiar video camera, which produces, as an electronic signal, images of the scene that it is focused at. Usually this is an analog

signal, but an analogue-to-digital converter will turn it into a digital signal, which may be arranged to form an image. This conversion is often done in an image frame store with the computer. A still video camera operates very like a photographic camera, but can store images in digital form, which can later be transferred to a computer.

A scanner is a good and relatively cheap solution. It is the best solution if the images for analysis are available in the form of photographs or hard copies of some other type. Most of modern scanners are flatbed scanners and work like photocopiers. They are widely available in desktop publishing. Hand-held scanners, which are cheaper, produce lower quality images. However, scanners work relatively slowly, so they are not a good solution for massive data input. They can often produce images with higher spatial resolution and more pixels than cameras. But, they are not suitable for freezing an image at a chosen moment for a changing scene.

A CCD (coupled charged device) video camera is currently one of the most frequently used device in computer image analysis. It is a smart device, offering good sensitivity and speed, being standard equipment on many optical microscopes. However, some problems may arise if we want to analyze color or higher resolution images because appropriate cameras and adapters are very expensive. Besides, the adapters require special software to be coupled to the computer.

Contemporary electron microscopes are equipped with modules for storage of digital images. Therefore, there is no problem to analyze the images from electron microscope using computer. Many other types of imaging equipment, particularly

medical imaging equipment such as differential interference contrast (DIC) microscope (Martin and Fallowfield, 1989), magnetic resonance imaging (MRI) (Fowler *et al.*, 1990), X-ray computer tomography (CT) (Simm, 1992), ultrasound imaging (Simm, 1992), landsat thematic mapper (TM) imaging (Glasbey and Horgan, 1995), synthetic aperture radar (SAR) imaging (Glasbey and Horgan, 1995) has been developed to capture images from different sources.

After image acquisition, an image is converted to a digital form. In image analysis, it is always necessary to save the digital images for further manipulation or analysis. Digital images may be stored in form of computer files with several different formats. These differ in how the pixel values and extra information are arranged. Commonly used image formats include Tagged Image File Format (TIFF), True Vision Targa Format (TGA), Windows Bitmap Format (BMP), CompuServe Graphics Interchange Format (GIF), Windows Device Independent Bitmap Format (DIB), ZSOFT Bitmap Format (PCX) and Joint Photographic Experts Group Format (JPG) (Media Architect, 1996). If necessary, programs that convert between different formats may be used.

2.3.5 Printing

In image analysis, a printer or other device to produce permanent copies of images is usually necessary. Most printers can produce some forms of image printout, but the quality of the result can be very variable. Printers designed mainly for printing text usually give very poor reproductions of images, and are only suitable

for very limited purposes. High resolution color printers are useful and available nowadays. For normal desktop printers, printer resolution can now reach 1440 × 1440 dpi (dot per inch). They are usually adequate for general image analysis purposes. Other devices such as plotters and slide-writers can also be used for special applications.

2.4 Operations of Image Analysis

2.4.1 Digital Image

To a computer, any image is a mosaic of very small areas, called pixels, filled with a single greylevel or digitally defined color (Wojnar, 1999). The brightness in each pixel is governed by a digital number. So the entire digital image is represented by a long list of digital numbers: each number represents the intensity of a pixel, and the number's location in the list represents the pixel's location in the image. When the number of pixels is large enough and the size of the pixels is sufficiently small, they are not perceived individually (Kuni, 1988). Thousands of pixels, touching each other and locating within a grid, give us an illusion of realistic, smooth picture. This pixel nature of computerized images allows us to store them and transform them as matrices of numbers. This is the basis for computer-aided image analysis. The number of pixels in an image is directly related to the spatial resolution of the image (Kuni, 1988). An image that shows much fine detail has a large number of small pixels. Since images typically contain a large number of pixels, it is convenient to keep the storage space as small as possible for each pixel.

2.4.2 Image Analysis Steps

One way of describing image analysis is to recognize distinct steps that follow each other logically, namely display, filters, geometric transformation, segmentation algorithm, mathematical morphology and measurement (Hall, 1979; Rosenfeld and Kak, 1982; Glasbey and Horgan, 1995).

2.4.2.1 Display

Display of an array of pixel values as a picture on a computer screen is the first stage in analyzing a digital image. Digital images can be divided into three different types of binary, greyscale and color image (Glasbey and Horgan, 1995).

A binary image is the most straight forward type to display. Pixels take only value of 0 and 1. The display simply needs to distinguish between these two levels. A natural way to do this is to show one value as black and the other as white. This is known as the "black and white" image. Any computer monitor is capable to display this kind of image.

Another kind of image is greyscale image. Greyscale display is to use the pixel values to specify the brightness with which a pixel is illuminated on a computer screen. Grey images are usually described by 256 greylevels. This corresponds to 8 bit per pixel as $256 = 2^8$. In this representation 0 equals black and 255 denotes whites. 256 greylevels are quite sufficient for most applications as humans can

distinguish approximately only 30 to 40 gray levels. In some applications, however, other depths of image data are used: 2 (binary images), 12, 16, or 32 bits per pixel (Wojnar, 1999).

Color images are most commonly stored as RGB (Red, Green, Blue) images. In fact, each of the RGB channels is a single grey image. Analysis of color images can be broken down into analyses of individual grey components being put together at the end to produce the final color image (Wojnar, 1999). This is referred to as the RGB system. Most technological use of color, from photography to television screens and computer monitors employs the RGB system to produce different colors. The screen of a color monitor consists of many small light-emitting dots of three types – blue, green and red. The brightness of each of the three types of dots determines what color is seen in a particular part of the screen. If each component can be displayed in 256 different intensities then there are 256^3 (= 16,777,216) different possible colors that a pixel can have. This is known as the true color.

To observe the final detail of an image, we may not be able to examine this detail easily in a standard display on a monitor. We would like to enlarge a part of the image. This is known as zooming. We cannot usually enlarge the pixels on the monitor, and zooming must consist of allocating more than one pixel on the monitor for each pixel in the image. The simplest form of zooming is pixel replication. It is equivalent to holding a magnifying glass in front of the screen. It is not always necessary to replicate pixel values in the computer memory. Some computer hardware has the ability to replicate pixels on the screen by making use of the way that the display control electronics access the computer memory. On other occasions,

we may wish to reduce the size of an image. This may be for display purposes, because our original image has too many pixels to be accommodated on the monitor. Alternatively, we may wish to reduce the size of an image so that it can be processed in less computer time, or occupy less storage space. One can form the reduced image by selecting a subset of pixels from the original image. This is known as pixel sampling. For example, if we take every third pixel in every third row of the original image, we shall have reduced the image size by a factor of nine. Another method is known as block averaging. One may define each pixel in the reduced image to be the average of a three-by-three block of pixels in the original image. This method reduces the loss of information in the process (Glasbey and Horgan, 1995).

2.4.2.2 Filters

Most images are affected to some extent by noise, that is, unexplained variation in data: disturbances in image intensity that are either uninterpretable or not of interest. Real images often contain some artifacts induced during specimen preparation, like scratches, smearing, relief, pull-outs, comet tails or lapping tracks. Image analysis can be simplified if this kind of noise is filtered out prior to any quantitative analysis. Image filters may be used to emphasize edges-boundaries between objects or parts of objects in images. Filters provide an aid to visual interpretation of images and can also be used as a precursor to further digital processing (Glasbey and Horgan, 1995; Wojnar, 1999).

Filters are neighbor-type operations. In other words, the pixel value after filtering is a function of its own value and the greylevels of its neighbors. Usually,

filters return values that are weighed means of neighboring pixels. Filters of various types are also among the most frequently used tools for image treatment. Common filters used in image analysis are brightness, contrast, sharpening, blurring and softening, smoothing, edges, maximum and minimum filters (Glasbey and Horgan, 1995; Media Architect, 1996; Wojnar, 1999). The duty of an image analyst is to decide the appropriate filters in order to enhance the efficiency of the noise removal.

2.4.2.3 Geometric Transformation

When an image is obtained by a capture device and loaded into the computer, it is usually necessary to adjust its alignment and position in the image analysis software. It can be accomplished generally by the function of rotation, scaling, and curvilinear or warped mapping (Freeman, 1986). This is most important when different images have to be overlaid together for measurement and comparison.

2.4.2.4 Segmentation Algorithm

Segmentation involves dividing an image up into regions, which correspond to different objects or parts of objects, by classifying all pixels (Glasbey and Horgan, 1995). Every pixel in an image is allocated to one of number of these categories. A good segmentation is typically one in which pixels in the same category have similar greyscale and form a connected region or neighboring pixels that are in different categories have dissimilar greylevel values.

Segmentation is often the critical step in image analysis. There are three general approaches to segmentation, termed thresholding, edge-based methods and region-based methods (Glasbey and Horgan, 1995). In thresholding, pixels are allocated to categories according to the greylevel range. In edge-based segmentation, an edge filter is applied to the image, pixels are classified as edge or non-edge depending on the filter output, and pixels that are not separated by an edge are allocated to the same category. Region-base segmentation algorithms operate iteratively by grouping together pixels that are neighbors and have similar greylevel values, and splitting groups of pixels that are dissimilar in value. Despite all of these common approaches to segmentation, an image analyst may usually have to design his/her own algorithm in order to extract more useful information from a specific image type.

2.4.2.5 Mathematical Morphology

Mathematical morphology is a theory that provides a number of useful tools for image analysis (Russ, 1995). Morphology is an approach to image analysis that is based on the assumption that an image consists of structures that may be handled by set theory. Due to their complexity, morphological operators are implemented only in advanced packages. On the other hand, mathematical morphology enables detection of various features in an image in a way somewhat similar to human intuition. As a consequence, application of morphological operators enables detection of features not available with other analysis methods. Morphological transformations such as hit, miss, erosion, dilation, opening and closing are used quite commonly in image analysis (Wojnar, 1999).

2.4.2.6 Measurement

Measurement, the extraction of quantitative information from images, is the final stage in the analysis. The objective may simply be to count the number of objects in a scene, compute their areas, or measure distances between objects such as computing the cross-sectional areas of the muscles. These measures of size can be obtained simply by counting the number of pixels in the muscle regions. Three general categories of measurement are size, shape and boundary. Measurements are usually taken from the images output from segmentation algorithms, which have possibly also been processed using morphological operators. In some application, measurements can be obtained directly from the original image (Glasbey and Horgan, 1995).

2.5 Conclusion

Recently, image analysis has been applied to different field of science rapidly. Using image analysis techniques instead of the human manual operations, analysis time is greatly reduced. Accuracy and reliability of the results obtained are higher. Moreover, with the improvement of the peripheral devices for computer and the development of new image processing methods, the image analysis techniques will become a powerful tool for scientific research in chemistry and biochemistry.

CHAPTER 3

Development of Image Analysis Method to Two-Dimensional Crossed Immunoelectrophoresis Study

3.1 Introduction

Electrophoresis as an analytical and preparative tool began with Tiselius in 1937 (Rilbe, 1983) and is defined as the migration of a charged particle in an electric field (Dunbar, 1987). The technique is usually used to separate proteins in aqueous mixtures. To characterize proteins and other antigenic substances, immunoelectrophoresis, an application of the electrophoresis technique, was developed by using both electrophoretic migration mobilities of these components in a matrix (usually agarose) and their immunological properties. There are many variations of this technique such as rocket immunoelectrophoresis, crossed immunoelectrophoresis, line immunoelectrophoresis and tandem-crossed immunoelectrophoresis (Dunbar, 1987). Yet, they are all based on the electrophoretic migration of antigens in an antibody-containing gel and the specific immunoprecipitation of the antigens by using the corresponding precipitating antibodies (Dunn, 1993). However, when applying these methods to gel electrophoresis, they are usually unable to resolve individual components from complex mixtures. Two-dimensional crossed immunoelectrophoresis (2D-CIEP) was then developed for this purpose with the resolving power increased by several orders of magnitude compared to the one-dimensional methods as mentioned above (Dunn and Burghes, 1986). This technique is widely used to detect the presence of abnormal protein in a protein mixture (Baker *et al.*, 1999). Abnormal binding of a protein to a cofactor could also be detected by the 2D-CIEP system (Michiels and Hamulyak, 1998; Nasher and Ciznar, 1998; Shirotani *et al.*, 1999).

In the past, electrophoretic pattern was mainly analyzed by visual inspection. Nowadays, several types of instrument are available for studying electrophoretic pattern. These include densitometers, laser densitometers, television cameras, coupled charged device (CCD) array scanners, radioisotope imagers and photostimulable phosphor-imaging systems and hardwares (Dunn, 1993). Several laboratories and commercial companies have also developed computer systems for automating quantitative analysis of 2D gel images (Sondergaard *et al.*,1987; Dunn, 1993). However, the price for such instruments is still high. In this work, a low cost image analysis system consisting of an IBM compatible-PC and a flatbed scanner was set up for capturing the 2D-CIEP pattern. In addition, a software called CIEPEASY for Windows was developed for manipulating 2D-CIEP images. In this way, the image of a 2D-CIEP profile can be acquired faster and easier in a cheaper way by using the proposed image capture system compared with other imaging instruments. The profile can also be digitized under high resolution to hold useful information of the 2D-CIEP gel for archiving purposes. Furthermore, our software CIEPEASY was designed specifically for analyzing 2D-CIEP patterns while other software packages are usually designed for general 2D electrophoretograms. Hence, CIEPEASY can be used to determine important peak parameters of peak area, position and height of a 2D-CIEP peak automatically within a short time interval. The results thus obtained are more accurate than those deduced from the conventional methods (Andrews, 1986). Besides, quantitative analysis on the 2D-CIEP pattern can also be performed.

3.2 Methods of Investigation

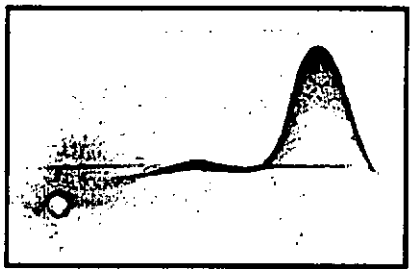
3.2.1 Principle of Two-Dimensional Crossed Immunoelectrophoresis

When any charged molecule is placed in an electric field, a force is exerted on it. The magnitude of the driving force depends on both the strength of the electrical field as well as the charge of that molecule. Since proteins possess charges as a result of amino acid residues in their backbones, they may migrate in the presence of an electric field at any pH. At high pH (e.g. pH 8.6 as used in the first dimension of 2D-CIEP), proteins will possess a net negative charge and hence move towards the anode. The rate of migration of a protein in this electric field also depends on the charge density, that is, the charge to mass ratio. The greater the value of this ratio, the faster the molecule will move. So, different proteins can be separated (Dunbar, 1987). In the 2D-CIEP experiment, the electrophoretic process is carried out in two different steps with the direction of the second electrophoresis step being at right angles to that of the first one (Melvin, 1987). In the first step, agarose without antibody is used while agarose with antibody to the antigen being tested is employed in the second step (Bøg-Hansen, 1990).

The first 2D-CIEP-separation method was developed in middle 1970s for both qualitative and quantitative purposes (Rickwood *et al.*, 1990). Since then, it has been widely applied to examine the purity of protein preparation and to analyze the composition of protein mixtures. Moreover, with the use of selected antibodies, the technique may be employed to investigate presence of the corresponding antigens and

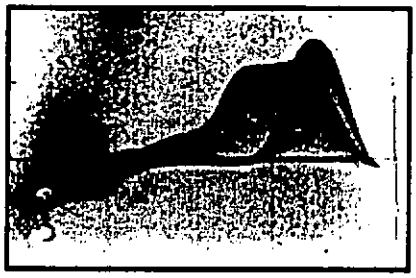
to check the identity or degree of differences between two antigens (Lo *et al.*, 1990; Michiels and Hamulyak, 1998; Nasher and Ciznar, 1998; Shirotani *et al.*, 1999). In this application, sample proteins were first applied to the wells cut in the agarose gel and then separated electrophoretically in that agarose gel. Afterwards, these separated proteins were electrophoresed at right angles to the direction of the first development in another layer of agarose containing specific antibodies to the antigens being tested. With an equivalent amount of antibodies, each protein will give rise to a precipitate or a peak. Figures 3.1a and 3.1b (Lo *et al.*, 1990) show the 2D-CIEP precipitin profiles of plasma samples from a normal control and a patient, respectively. Usually, peak parameters such as migration distance, peak height and peak area of a 2D-CIEP peak are measured manually, for example, by visual inspection. However, in some cases, due to the abnormal appearance of the 2D-CIEP gel (e.g. Figure 3.1b), it is not easy to identify the peak profile or parameters correctly by visual inspection. Hence, the accuracy of the results obtained is not high. Furthermore, other parameters such as the peak position and the shape of the peak profile that may contain important information cannot be deduced accurately. To illustrate this, the 2D-CIEP profile of a plasma protein called anti-thrombin III from a normal person in the presence of heparin was shown in Figure 3.1a. It consists of only one peak with its shape like a Gaussian curve. However, the profile of a patient (Figure 3.1b) is composed of two overlapping peaks. Figure 3.1c depicts a combination of Figure 3.1a and 3.1b together. It can be seen that the two profiles are different. This information is important for clinical diagnosis. Presence of a mutant protein that lacks a binding epitope has been described using this technique (Shirotani *et al.*, 1999). Besides, Bøg-Hansen (Bøg-Hansen, 1990)

Figure 3.1 The AT-III-CIEP pattern of a plasma sample of (a) a normal person, and (b) a patient (no. 1) and (c) a combination of Figures 3.1a and 3.1b. Sodium heparin (15 units/mL) was added to the agarose in the first dimension.



(a)

(b)



(c)

suggested that the area enclosed by the precipitate is proportional to the amount of antigen to be investigated. Traditionally, the quantitative measurement is achieved by setting the area equal to the product of the peak height and its width at half height (Andrews, 1986). However, if the peak is asymmetric, the peak area obtained will not be so accurate. As a result, new methods have to be developed for analyzing 2D-CIEP profiles in a more scientific manner.

3.2.2 Image Processing of Two-Dimensional Crossed Immunoelectrophoresis Profiles

In order to extract peak parameters from the 2D-CIEP profiles automatically or semi-automatically, the 2D-CIEP pattern must be digitized first. In this investigation, an optical flatbed scanner coupled with a PC was used to achieve this purpose. After image acquisition, the digitized 2D-CIEP pattern is then converted into the grayscale with 256 greylevels and stored in a common graphical format such as BMP, PCX, or TIFF. In terms of image processing terminology, a grayscale image is represented by the intensities of light for all the pixels of the image on a continuous scale (Wegner, 1992). An absolutely black pixel has a greylevel of 0 while an absolutely white pixel has a greylevel of 255. Parameters such as migration distance, peak height, peak area, peak profile, well center and others of the 2D-CIEP profile are extracted from the digitized image. Details are given as follows.

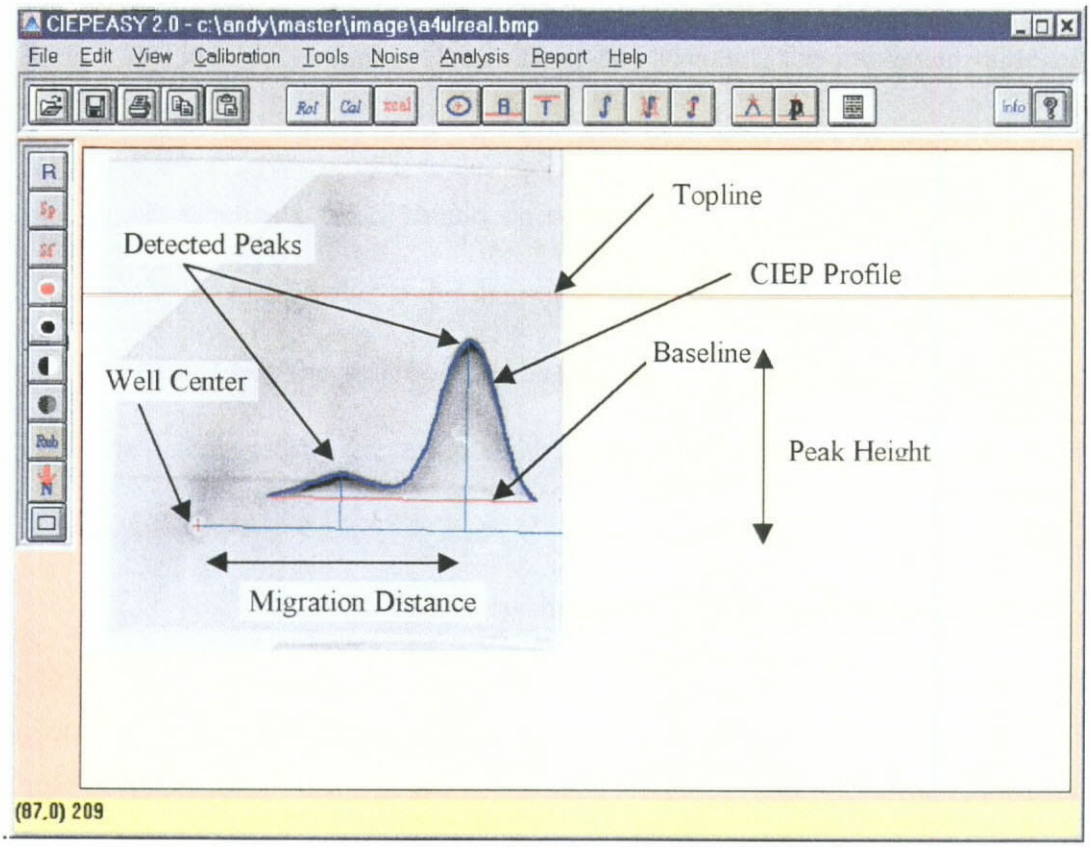
3.2.2.1 Reduction of Background Noise

As shown in Figure 3.1b, unwanted spots or noise are usually observed in the antithrombin-III (AT-III) 2D-CIEP pattern. The noise may be contributed by diffusion of the sample in the gel or from staining of the gel in the final stage (Bøg-Hansen, 1990). Sometimes, its intensity is comparable to that of the peak signal and leads to inaccurate results. Three different methods were used in this investigation to reduce the noise. In the first method, the greylevel value of a certain pixel within the image is simply assigned to be 255 (white color) manually. One might also define the greylevel values for all the pixels within a sub-region to have values of 255. The last method involves the use of a threshold to remove noise. Any pixel within the image that has a greylevel value greater than the threshold value will be converted to 255.

3.2.2.2 Method for Determination of Well Boundary and Well Center

In Figure 3.1, a circular pattern is observed on the lower left corner. It originates from sample loading during the experiment and is called sampling well (Bøg-Hansen, 1990). Within this region, the position of the well center is a very important quantity for determining other peak parameters such as the migration distance (Figure 3.2) in the AT-III 2D-CIEP pattern. In order to determine the well center, the well boundary has to be defined first. As the region within the well should be clear and the greylevel of the constituting pixels have values usually near 255, the well boundary can be identified by setting the noise threshold value above the

Figure 3.2 Screen layout of the main menu of the software CIEPEASY together with a 2D-CIEP pattern.



greylevel value of the noise around the well. In general, pixels outside the well boundary have lower greylevel value than that find insides the sample well. In this investigation, the threshold value is assigned to have the minimum greylevel value around the boundary region. Once this quantity is set, the greylevel value of the pixels within the defined region for the well will be compared with this value to distinguish whether a pixel should be within or outside the well boundary. If the greylevel value of a pixel is larger than the threshold, then the pixel is within the well or vice versa. Once the well boundary is defined, the x- or y-coordinate of the well center is determined by averaging all the x- or y-coordinates of the pixel within the well as follows:

$$x-\text{coordinate} = \frac{\text{Sum of all the } x\text{ - coordinates of pixels within the well}}{\text{Total number of pixels within the well}} \qquad (3.1)$$

and

$$y-\text{coordinate} = \frac{\text{Sum of all the } y\text{ - coordinates of pixels within the well}}{\text{Total number of pixels within the well}}. \qquad (3.2)$$

3.2.2.3 Method for Determination of 2D-CIEP Peak Parameters

Several parameters are required to quantify the presence of a particular protein in the 2D-CIEP pattern. That includes migration distance, peak height and baseline (Figure 3.2). The migration distance for a protein is defined as separation between the well center and the position of the peak maximum in the horizontal direction. The baseline of a 2D-CIEP image represents the separation between gel layers with and without antibodies. In general, the upper part of the gel contains antibodies and the lower part does not. The baseline has to be assigned before the

determination of peak height and area. Besides, a topline (Figure 3.2) should also be assigned to facilitate the determination of the overall peak profile.

The boundary of the 2D-CIEP profile is determined in this investigation by scanning the greylevel value of each pixel between the baseline and topline. Two methods have been implemented in this work for such purpose. In the first method, scanning of pixel intensities is performed vertically upwards from the baseline to the topline and then from left to right. Generally, a pixel in the peak boundary of a 2D-CIEP profile has a sudden change in the greylevel from a higher value to a lower one in the upward scanning. A pixel with the maximum change is assumed to be that representing the boundary. In this manner, all the boundary of the 2D-CIEP profile is determined (Figure 3.2). The change is from a higher greylevel value to a lower one. All the pixels on the boundary thus determined are joined together to form the 2D-CIEP profile (Figure 3.2). However, in some cases, the boundary cannot be detected successfully by this algorithm owing to the poor quality of the 2D-CIEP profile or the presence of background noise. In order to improve the accuracy, various filter functions were introduced to assist the boundary determination. Such filters utilize different combination of threshold values to either enhance the boundary or suppressing the background noise to locate the boundary position.

Once the 2D-CIEP profile is determined, the positions of all local peak maxima can be found out and the peak heights (Figure 3.2) can be obtained by measuring the perpendicular distance, in terms of pixel, from the peak maximum to the baseline. Shoulder peak is often observed in the 2D-CIEP image with overlapping with other peaks (Figure 3.1b). The peak with the highest rate of change

of slope on the peak boundary will be identified as the shoulder peak position. The peak height and migration distance of the shoulder peak can also determined using similar approach as mentioned before. The area of the region between the baseline and the 2D-CIEP profile is regarded as peak area. The peak area of the 2D-CIEP profile is proportional to the amount of antigen or protein applied in the sample well (Svendsen, 1979) and is determined by counting the total number of pixels within the region between the peak boundary and the baseline.

3.3 Experimental

3.3.1 Software Implementation

2D-CIEP images analysis was carried out by the software CIEPEASY developed in the present work. CIEPEASY can help the user to determine the required parameters such as migration distance, 2D-CIEP profile, peak height and peak area automatically for quantitative analysis. The program was coded in Microsoft® Visual Basic Version 3.0, Professional Edition (Microsoft Corporation, WA, US) under Microsoft® Windows™ 3.1 environment on a PC compatible with a 133MHz Pentium® processor. ImageKnife/VBX™ 2.0 (Media Architects, Inc., OR, US) was also embedded in the program for graphic file manipulation.

CIEPEASY provides a user-friendly graphical interface for 2D-CIEP gel pattern analysis (Figure 3.2). All the functions have their own pull down menus and icons. Major functions and options available are listed in Table 3.1. 2D-CIEP

electrophoretic pattern can be imported to CIEPEASY in BMP, TIFF, GIF, PCX, JPEG, DIB or TGA graphic format. General image processing tools such as contrast, brightness, sharpness and softness adjustments are also available in CIEPEASY. As the image may not be acquired in an absolutely vertical or horizontal position, this will affect the accuracy of the analysis. A rotation function is provided to the user to adjust the image position circularly on the screen to minimize this type of error. Background noise and noise incurred during the experiment can be reduced by the methods mentioned previously. Different threshold values can be assigned by the user to facilitate the determination of the 2D-CIEP profile. After the user defines the location of the sample well, baseline and topline, peak profile and other peak parameters will be determined automatically using the methods as described before. The peak height and migration distance may be reported in pixel unit or in centimeter (cm) while the peak area in the total number of pixel or in cm^2. With regard to data reporting and archiving, images processed by CIEPEASY can be saved in the form of PCX, TIFF, TGA, JPEG, DIB or BMP graphic files. A hard copy of the image can also be provided upon request. The "Copy" and "Paste" functions available in CIEPEASY provide image exchange capability to other software packages such as Microsoft® Word or other image processing softwares. Moreover, information extracted from the image under study may be exported to Microsoft® Write for further data manipulation and storage.

Table 3.1 Major functions available in CIEPEASY

Options provided	Functions available
Image Processing	Adjustment of image size
	Adjustment of zooming ratio
	Adjustment of sharpness and softness
	Adjustment of contrast and brightness
	Conversion of a color image to that in grayscale
	Rotation of an image
Noise Reduction	Rubber erasing
	Noise bounding and cutting
	Noise level exclusion
Image Analysis	Determination CIEP profile
	Adjustment of shape of the CIEP profile
	Determination of peak area
	Determination of peak height and migration distance
	Determination shoulder peak
Reporting and Archives	Saving processed image
	Exporting of report to Microsoft® Write for further manipulation and archives.

Images of 2D-CIEP gel patterns as studied in this work were captured by a HP ScanJet IIc flatbed scanner (Hewlett-Packard Company, US) at a resolution of 200 dot per inch (dpi). This resolution level was found to be good enough for the present investigation. Although higher resolution can be used, more memory and storage space is required. All digitized images were stored using the BMP graphic format with 256 greyscales. No image post-processing steps such as contrast and brightness adjustment were applied on the scanned images. All scanned 2D-CIEP gel images were then imported and analyzed by CIEPEASY.

3.3.2 2D-CIEP Experiment

CIEPEASY was applied in this work to study the effects of different sample sizes in the anti-thrombin III (AT-III) CIEP with the presence of 15 units/mL of heparin in the first dimension agarose as described previously (Lo *et al.*, 1990). Briefly, 1.65 mL of 0.9% molten agarose in barbitone buffer (0.0375 M Na-barbitone, 0.015 M barbitone, pH 8.6) containing 15 units/mL Na-heparin was poured onto an area of 1.5 cm × 5 cm on the hydrophobic side of a gelbound (Bio-Rad Laboratories Inc., US). After solidification, a sample well was cut on the left hand side of the agarose. Plasma sample and 3 µL of a blue marker (0.5% Evan's blue and 5% bovine serum albumin in barbitone buffer) was put into the well. The first dimension electrophoresis was applied from left (cathode) to right (anode) at 1.5 mA/cm until the blue marker reached 1 cm of the right hand side of the gel. Subsequently, 3.9 mL 0.9% agarose containing AT-III antibodies (Calbiochem-Behring, Australia) was poured onto the remaining 3.5 cm × 5 cm of the gelbond. After solidification, the gel was turned at right angle and the second dimension

electrophoresis was applied from the side of plain agarose gel (cathode) to the side of the antibodies containing gel (anode) at 1 mA/cm for 16 hours. Prothrombin 2D-CIEP was also performed in a similar fashion. The only difference is that no heparin was added in the first dimension agarose. Prothrombin antibodies were purchased from Dakopatts, US.

Precipitin patterns in the gel were visualized by Coomassie blue staining. Briefly, the agarose gel was washed in 0.9% saline for 4 hours. After a brief dipping in distilled water, the agarose gel was squashed between several folds of filter paper and then blow-dried. Finally, the gel was stained in staining solution (0.25% Coomassie brilliant blue, 45.4% methanol and 9.2% glacial acetic acid) for 30 minutes. Precipitin patterns were visualized after destaining in destaining solution (45.4% methanol and 9.2% glacial acetic acid) for about 1 hour or until prominent precipitin patterns with clear background can be seen. Gels were prepared and provided by the co-supervisor Dr. S.C.L. Lo from the Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University.

3.4 Results and Discussion

Figure 3.3 shows the 2D-CIEP pattern for different sample sizes of heparin in the AT-III-CIEP study. Columns A and B are, respectively, the 2D-CIEP images from the plasma samples of a normal person and a patient (no. 1) with the corresponding sample sizes listed in column C. The AT-III-CIEP abnormality of plasma from the patient is found to be dose dependent (Figure 3.3). An abnormal AT-

III-CIEP precipitin arc which was indicated by arrow 3 in Figure 3.3 was observed with a sample size larger than 5 μL when comparing with the 2D-CIEP patterns of the normal person. However, when the sample size was decreased to 3 μL, normal AT-III-CIEP precipitin arc that was indicated by arrows 1 and 2 was observed. The results as obtained from analyzing the 2D-CIEP patterns of plasma samples by using CIEPEASY are listed in Tables 3.2 to 3.4. Arrows in Figure 3.3 indicate the peaks being studied. By applying CIEPEASY to study both normal person and patient, a linear relationship was found between the peak area and the sample volume with the correlation coefficient greater than or equal to 0.97 (see Table 3.2) although some of the 2D-CIEP patterns of the patient are abnormal. However, this is not true if the peak height instead of peak area was utilized (Table 3.3). The calibration curve based on the peak area can be employed to determine the amount of antigen present in human plasma. To our best knowledge, this is the first time in the literature reporting a linear relationship between peak area and the sample size in 2D-CIEP study. To test this hypothesis further, two more sets of CIEP experiments with increasing sample sizes were performed. Although with different concentrations of antibodies, both additional experiments use anti-prothrombin antibodies as the precipitating antibodies. These antibodies were chosen because the precipitin patterns formed are different when compared to that of AT-III antibodies (Figure 3.4). In the first experiment, plasma samples from 2 patients (nos. 2 and 3) and normal pooled plasma were used. In the second experiment, one patient and another normal plasma samples were used. Results of the first experiment were given

Figure 3.3 Effects of different sample sizes in the AT-III-CIEP study with the presence of 15 units/mL of heparin in the first dimension agarose. Image on the left (column A) and the right (column B) hand sides are, respectively, the 2D-CIEP images from the plasma samples of a normal person and a patient with the sample sizes listed in column C.

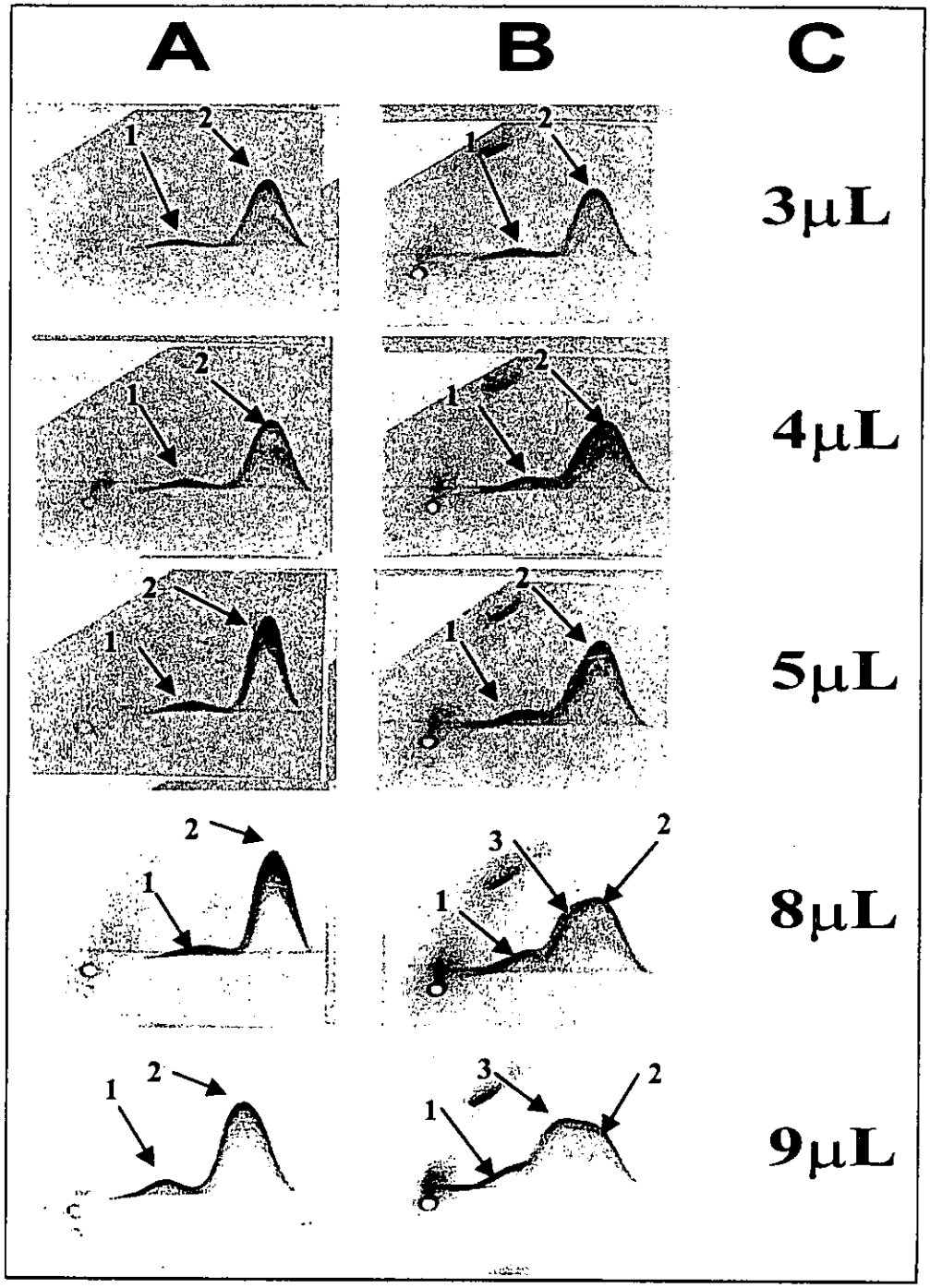


Table 3.2 Areas of the AT-III-CIEP peaks with different sample sizes as shown in Figure 3.3.

Plasma sample size (μL)	Area under the AT-III-CIEP profile / cm^2	
	Normal	Patient
3	0.92	0.98
4	1.05	1.25
5	1.10	1.48
8	1.35	1.73
9	1.53	1.88
Correlation coefficient	0.99	0.97

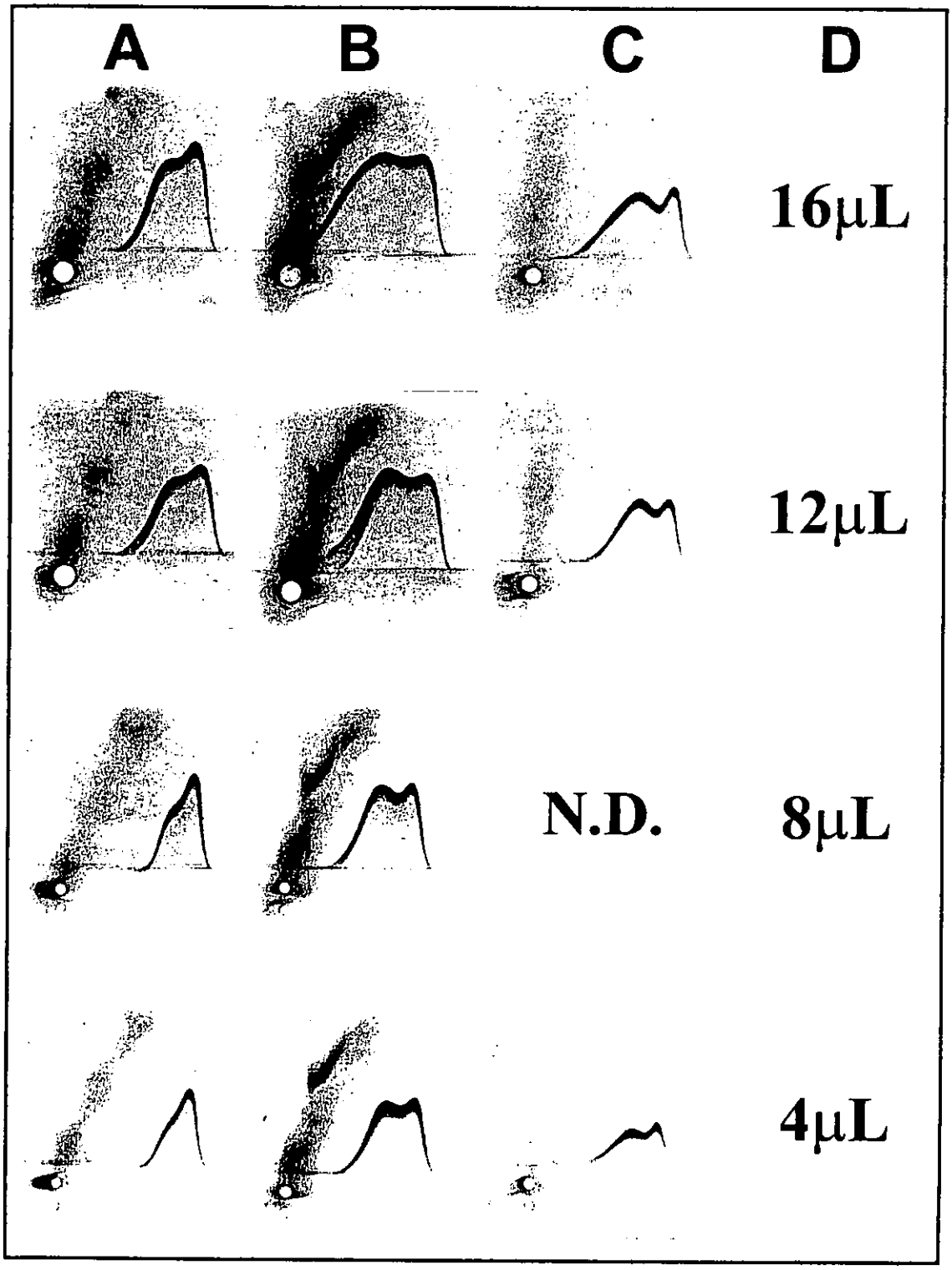
Table 3.3 Peak heights of the AT-III-CIEP peaks with different sample sizes as shown in Figure 3.3.

Plasma sample size (μL)	Peak height / cm				
	Normal		Patient		
	Peak 1	Peak 2	Peak 1	Peak 2	Peak 3
3	0.17	1.24	0.18	1.26	N/A
4	0.19	1.32	0.25	1.30	N/A
5	0.19	1.79	0.32	1.59	N/A
8	0.20	1.92	0.39	1.44	1.21
9	0.32	1.70	0.48	1.26	1.35
Correlation coefficient	0.79	0.77	0.98	0.04	N/A

Table 3.4 Migration distances of the AT-III-CIEP peaks with different sample sizes as shown in Figure 3.3.

Plasma sample size (μL)	Migration distance / cm				
	Normal		Patient		
	Peak 1	Peak 2	Peak 1	Peak 2	Peak 3
3	1.24	2.57	1.38	2.51	N/A
4	1.26	2.49	1.26	2.41	N/A
5	1.45	2.59	1.32	2.44	N/A
8	1.55	2.64	1.21	2.27	1.80
9	1.27	2.49	1.22	2.30	1.88
Percentage of relative standard derivation (%RSD)	10.22	2.56	5.60	4.18	N/A

Figure 3.4 Effects of different sample sizes in prothrombin-CIEP. Images in column A are prothrombin-CIEP images of a normal plasma sample while that in columns B and C are CIEP images from 2 patients (nos. 2 and 3). Column D represents the sample sizes.



in Table 3.5. Very high correlation was obtained between the peak areas and the amounts of plasma samples used. Results of the second experiment were very similar with correlation coefficients around 0.99. Hence to avoid repetition, the results were not shown. Taken overall, our results indicate strongly that with the use of our proposed image analysis technique, the 2D-CIEP method can be extended for quantitative study. It should be mentioned that the accuracy of the areas as listed in Table 3.5 is not known. It is because to our knowledge there is no report on the true values of this kind in the literature.

It should be pointed out that peak areas of plasma samples from patient no. 1 as seen in the AT-III-CIEP study were found to be larger than those of normal person. This implies that the amount of antigen in the plasma of the patient is higher than that of the normal control. Chromogenic enzyme assay of AT-III activity of the patient plasma using commercially available reagent S2238 (KabiVitrum / Ortho Diagnostic System, US) and performed as previously described (Shah *et al.*, 1992) confirmed that this patient had an elevated level of 130% when compared to a pooled normal control of 100% (unpublished data).

The migration distances of different 2D-CIEP peaks of the constituting components of a plasma sample provide important information about the characteristics of the antigen. These quantities and their percentages of relative standard deviations (%RSD) of AT-III 2D-CIEP are listed in Table 3.4. It is obvious that some of the gel patterns of the patient contained three peaks with a few of them being present as shoulder peaks. As the pore size of the agarose matrix is relatively larger than that of the plasma proteins, these proteins are separated mainly by their

differences in their overall molecular electrical charge densities. The additional shoulder peak observed might represent the presence of abnormal variants of that protein with different molecular electrical charge densities. As the electrical charge density of the abnormal antigen decreased, the antigen would migrate slower. The migration distances of different antigens on the gel in the first electrophoresis would be different. This gives rise to various peaks on the 2D-CIEP gel. The overall results from this work support the abnormality of the patient's antigen in blood plasma. The %RSD values of the migration distance as given in Table 3.4 are generally low which indicate that the results obtained in this work are precise. For clinical diagnosis, by using the proposed image analysis technique, one can actually locate precisely the migration distance of a specific antigen. From a study on a control group of normal persons, an indicator can be set up based on the migration distances which can then be used for comparing samples from patients as a clinical reference.

The advantage of our method cannot be obtained by using the conventional methods such as visual inspection, counting squares, using ruler for measurement or other approximation methods like the product of peak height and full-width-at half-maximum (FWHM) (Andrews, 1986). A study on the precision of the results obtained in this work was also performed. The AT-III-CIEP gel pattern as shown in Figure 3.1a was analyzed by both the conventional methods and ours. For the former ones, migration distance and peak height were measured by ruler whereas peak area was measured by utilizing graph paper and the FHWM approximation method. The results thus obtained are listed in Tables 3.6 and 3.7. It can be seen that both methods give comparable results for peak parameters. Replicate analyses on the same

Table 3.5 Areas of prothrombin 2D-CIEP peaks with different sample sizes as shown in Figure 3.4. Prothrombin antibodies at a concentration of 0.4% were used in the second dimension.

Plasma sample size (μL)	Area under the prothrombin-CIEP profile / cm^2		
	Normal	Patient #2	Patient #3
4	1.19	2.0	0.97
8	1.55	2.42	1.73
12	2.36	3.97	N.D.
16	2.88	4.61	2.45
Correlation coefficient	0.99	0.98	0.98

Table 3.6 AT-III-CIEP analysis results as obtained by the conventional methods (see text).

Parameters	Analysis Method		
	Ruler	Graph paper	Approximation (FWHM)
Peak Height / cm	1.8	N/A	N/A
Migration Distance / cm	2.5	N/A	N/A
Peak Area / cm^2	N/A	1.03	1.04

Table 3.7 Results of analyzing the AT-III-CIEP in Figure 3.3 by CIEPEASY (see text).

Parameters	Replicates					Average	%RSD
	1	2	3	4	5		
Peak Height (cm)	1.78	1.79	1.78	1.79	1.78	1.78	0.31
Migration Distance (cm)	2.58	2.60	2.59	2.59	2.54	2.58	0.91
Peak Area (cm^2)	1.08	1.13	1.11	1.11	1.10	1.11	1.64

normal person's AT-III-CIEP pattern (5 μL sample) as shown in Figure 3.3 by using CIEPEASY were also performed to test its precision. Low %RSD (see Table 3.6) indicates that CIEPEASY can obtain reproducible results. Similar results were obtained with prothrombin-CIEP precipitin peaks using plasma samples from control, normal, pooled plasma and 2 other patients. It is obvious that the proposed image analysis method is superior to the conventional ones using ruler, graph paper and FWHM approximation in determining peak parameters (Andrews, 1986). Normally in measuring a distance by utilizing ruler, an absolute error of ±1mm is anticipated. For a 2D-CIEP-gel image scanning with a resolution of 200 dpi, the absolute error incurred by our method in measuring the peak height or migration distance is only ±2 pixels, or ±0.254mm. Analytical error will be further reduced when a higher resolution image is used. But the image processing time will be increased as the image size becomes larger. More serious error may incur in determining other parameters such as the well center by visual inspection and peak area by the counting square methods using graph paper. CIEPEASY can help the user to save more time in determining peak parameters of a 2D-CIEP pattern and minimizing personal errors.

3.5 Conclusion

In conclusion, an image analysis system with a PC and an optical flatbed scanner and a software package CIEPEASY was developed to analyze 2D-CIEP pattern. From the results of AT-III-CIEP and prothrombin-CIEP studies, a linear relationship between the peak area and the sample size was observed the first time by using this new method. Besides, the results also indicated that CIEPEASY was

superior to the conventional method in term of the time for image analysis as well as accuracy and precision of results obtained.

CHAPTER 4

Application of Image Analysis Technique to Study Abnormal Two-Dimensional Crossed Immunoelectrophoresis Patterns of Prothrombin in Patients with Lupus Anticoagulant

4.1 Introduction

Lupus anticoagulant (LA) was first described in a patient with Systemic Lupus Erythematous (SLE) in 1952 (Conley and Hartmann, 1952). Many other associations with this entity have been reported (Schleider *et al.*, 1976; Ruggieri *et al.*, 1985; Cohen *et al.*, 1989) and thrombosis appears to be the most common (Bowie *et al.*, 1963). The finding of patients with the LA or anticardiolipin antibodies can help identify those patients more prone to thrombotic episodes. However, at present, evidence that these antibodies that cause the thromboses are lacking.

Although mechanism of thrombosis in patients with the LA is still uncertain, abnormalities in terms and levels and functions of procoagulants/fibrinolytic factors found in plasma from patients with the LA have sometimes been inferred as the cause of thrombosis. With a two-dimensional crossed immunoelectrophoretic (2D-CIEP) technique, abnormal precipitin arcs associated with the presence of slow-moving material (SMM) were seen in patients with antibodies (Edson *et al.*, 1984). Importantly, these antibodies commonly occurred in patients with quantitatively normal prothrombin levels as measured in clotting studies as well as rocket immunoelectrophoretic assays. This abnormality would therefore remain undetected without the use of 2D-CIEP techniques.

In this work, prothrombin (PT), factors IX, X as well as von-Willebrand's factor (vWf), plasminogen (PLG), prekallikrein (PKK) and high molecular weight

kininogen (HMWK) in plasma samples from 21 previously reported patients with the LA were analysed by 2D-CIEP techniques (Lo *et al.*, 1990). It was aimed to correlate any abnormality found in the various pro-coagulation and/or fibrinolytic factors with a previous history of thrombosis.

An image analysis system with an IBM PC compatible coupled with an optical scanner was established for capturing 2D-CIEP patterns. A computer package CIEPEASY (Chan *et al.*, 1999) as developed in the previous chapter was utilized to analyze the 2D-CIEP electrophoretograms of these patients. CIEPEASY can be used for modification and analysis of the 2D-CIEP patterns acquired to determine the peak parameters such as the migration distance, peak height and area under the precipitin ares for both qualitative and quantitative studies. In this approach, the time required for data collection and interpretation of 2D-CIEP gel images was shortened significantly and the results thus obtained have a higher accuracy than those by using conventional methods (Andrew, 1986). More information could be extracted from the proposed image analysis system in a scientific and statistical presentation.

4.2 Experimental

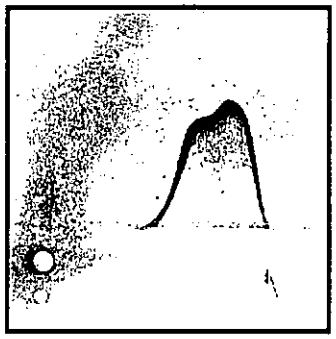
PT, factors IX, X, vWF, PLG, PKK and HMWK in plasma samples from two groups of previously reported patients (n = 21) with the LA were analyzed by 2D-CIEP techniques (Lo *et al.*, 1990). The 2D-CIEP experimental procedure can be found in Chapter 3 and is not repeated in here. All 2D-CIEP

immunoelectrophoretograms were provided by Dr. S.C.L. Lo, Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University. One groups of patients had a previous history of thrombosis (thrombotic group. number of patient = 13) and the other group without (non-thrombotic group. number of patient = 8). In the experiment, normal persons 2D-CIEP immunoelectrophoretograms were used as control. To quantitatively compare subtle changes seen on the 2D-CIEP immunoelectrophoretograms. image analysis was applied to these patterns by CIEPEASY. All the 2D-CIEP electrophoretograms were scanned using an optical flat-bed scanner HP ScanJet IIc (Hewlett Packard Company. US) at a resolution of 200 dpi (dot per inch) and analysed by CIEPEASY in this study. With the help of CIEPEASY program. peak heights. migration distances. and areas under the precipitin arcs of the electrophoretograms were measured and compared.

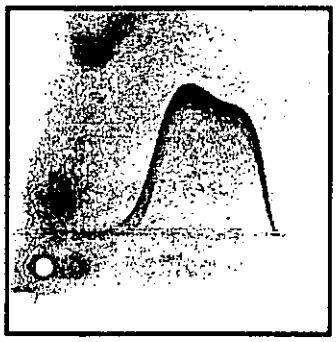
4.3 Results

Out of the 21 LA patients in this study. 10 patients had one or more abnormal 2D-CIEP patterns among the different proteins analysed. But these 10 patients all had abnormal prothrombin 2D-CIEP precipitin arcs and six of them had previous thrombotic episodes. Figures 4.1a and 4.1b show the selected prothrombin 2D-CIEP electrophoretograms of the control and the abnormal patient respectively.

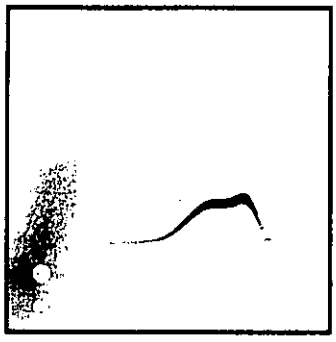
Figure 4.1 Selected prothrombin 2D-CIEP electrophoretograms of (a) a control sample, (b) an abnormal patient and (c) a thrombotic patient indicating a significant reduced anionic peak.



(a)



(b)



(c)

In the 2D-CIEP investigation of PT, factors IX, X, vWf, PLG, PKK and HMWK, patterns of the precipitin arcs from all 11 control samples were highly consistent. For all the normal 2D-CIEP electrophoretograms and with the use of the proposed image analysis system, two times the standard deviations of the peak heights had never been greater than 0.5 cm while that of the migration distance had never been greater that 0.22 cm and that of the area had never been greater than 0.64 cm^2.

In all the 2D-CIEP eletrophoretograms analysed, except for prothrombin 2D-CIEP electrophoretograms, there are no significant differences among the control, thrombotic and non-thrombotic groups in terms of peak heights, migration distances and areas under the precipitin arcs. As for the image analysis of prothrombin 2D-CIEP (PT-2D-CIEP) electrophoretograms, the results are summarized in Table 4.1. When comparing different groups, it was found that patients with a previous history of thrombosis had significantly reduced peak heights for the more anionic peak than that of the control (*t*-test, $p = 0.007$). This is not seen in the non-thrombotic group ($p = 0.17$). Figure 4.1c shows a selected thrombotic patient's prothrombin 2D-CIEP electrophoretograms which indicate a significant reduced anionic peak.

4.4 Discussion

Out of the 21 LA patients in this study, 10 of them have demonstrated an abnormal prothrombin 2D-CIEP, with all 10 showing the presence of slow-moving materials associated with normal antigenic levels of prothrombin. These results are

Table 4.1 Results of image analysis of prothrombin 2D-CIEP electrophoretograms by using CIEPEASY.

	More Anionic Peak		Less Anionic Shoulder		
	Peak Height (cm)	Migration Distance (cm)	Peak Height (cm)	Migration Distance (cm)	Area (mm^2)
Control (Mean ± 2 S. D.)	1.78 ± 0.50	2.48 ± 0.22	1.41 ± 0.48	1.95 ± 0.22	1.79 ± 0.64
LA patients with a thrombotic history (Mean ± 2 S. D.)	1.25 ± 0.82	2.49 ± 0.38	1.19 ± 0.90	2.07 ± 0.48	1.53 ± 1.36
LA patients without a thrombotic history (Mean ± 2 S. D.)	1.50 ± 1.03	2.54 ± 0.16	1.32 ± 0.87	2.07 ± 0.17	1.64 ± 1.14

consistent with those reported previously (Edson *et al.*,1984. Fleck *et al.*,1988). suggesting the presence of a non-neutralising antibody to prothrombin in these patients.

When the PT-2D-CIEP electrophoretograms were scanned and analysed by CIEPEASY, it was found that the more anionic peaks of samples from patients of the thrombotic group had a statistically reduced peak height when compared with normal ($p = 0.007$). This phenomenon is not observed in the non-thrombotic group. Whether this reduced peak height represents a decreased amount of more anionic prothrombin sub-species in LA patients with a thrombotic history remains unclear although Murakami and coworkers suggested that anti-prothrombin antibodies reduced prothrombin activities in a patient with LA (Murakami *et al.*, 1996). Possibilities also exist that autoantibodies to prothrombin increased the clearance rate of prothrombin. Further studies are required to verify whether such a difference in peak height represent an increased risk of thrombosis.

4.5 Conclusion

Without the use of CIEPEASY, it was difficult to compare 2D-CIEP patterns between normal persons and patients quantitatively. Moreover, the accuracy of the analysis was greatly improved when comparing the measurement of peak parameters like peak heights, migration distance and peak areas by using ruler or graph paper. Also, the time needed for the analysis of electrophoretograms was greatly reduced. In

clinical application, image analysis also make the medical judgement more objective. Human error is also minimized by transferring the tedious repetitive pattern measurements from labor work to computer.

CHAPTER 5

Development of Image Analysis Method to Enhance DNA Fingerprinting Analysis in Forensic Science

5.1 Introduction

In board definition, forensic science is a science used for the purpose of the law. Thus, any branch of science used in the resolution of legal disputes is in the barrier of forensic science. This definition covers criminal prosecutions in the widest sense, including consumer and environmental protection, and health and safety at work, as well as civil proceedings such as breach of contract and negligence. However, in general usage the term is applied more narrowly to the use of science in the investigation of crime by the police and courts as evidence in resolving the issue in an subsequent trial that includes the investigation of offences such as murder, violent assault, robbery, arson, breaking and entering, fraud, motoring offences illicit drugs, and poisoning (Cobb, 1998).

Deoxyribonucleic acid (DNA) fingerprinting for personal identification is a powerful tool for criminal investigation and justice. Characterization, or "typing" of DNA for purposes of criminal investigation can be thought of as an extension of the forensic typing of blood that has been common for more than 50 years (National Research Council, 1992). DNA fingerprinting can be an essential adjunct to personal identification in forensic science. The first suspect to have been convicted largely on the basis of DNA analysis of blood samples was sentenced at the Crown Court at Leicester on 22 January, 1988. Since this case, DNA technology has become commonplace in forensic laboratories around the world and has been instrumental in establishing both guilt and innocence in many court cases (Watson,1998).

DNA, the active substance of the genes, carries the coded messages of heredity in every living thing: animals, plants, bacteria, and other microorganisms. In humans, the code-carrying DNA occurs in all cells that have a nucleus, including white blood cells, sperm, cells surrounding hair roots, and cells in saliva. These would be the cells of greatest interest and sampling target at the crime scene. (National Research Council, 1992). Each individual, except identical twins, possesses a unique genetic signature that can be visualized using recombinant DNA techniques. The signatures resemble the bar codes on supermarket products. These DNA fingerprints or profiles, have broad applications in forensic analysis, paternity testing, diagnostic medicine, plant and animal sciences, wildlife poaching and homicide, rape, and assault cases (Lorne, 1990).

The use of short tandem repeat (STR) analysis of polymorphic loci is the commonly used method for individual identification in DNA analysis. Coupled with the application of polymerase chain reaction (PCR) technology, the system has enabled the development of rapid, sensitive analysis even on highly degraded DNA (Sparkes *et al.*, 1996). After gel electrophoresis on the processed DNA fragments, fluorescent labeling or silver staining is chosen as a visualization and detection method for the gel patterns. For fluorescent labeling, it has an excellent resolution and selectivity. The laser detector gives high accuracy and reproducibility result. However, the equipment is much more expensive and the time require for analysis is longer than using the silver staining method (Micheli and Bova, 1997). In contrast, silver staining detects nucleic acids with high sensitivity and avoiding fluorophore or radioisotopic labeling. Silver staining of complex nucleic acid profiles that separated in polyacrylamide gels provides high band resolution. These gels permit the

retrospective recovery and examination of nucleic acids, constitute experimental records, avoid the costs of photography, and allow easy densitometrical scanning of DNA profiles for data analysis (Kirby, 1990). However, selectivity and reproducibility are not as good as fluorescent labeling.

For the DNA electrophoretogram analysis after sliver staining, allele band matching can be arduous if it is done by the eye and hand. Traditionally, bands are sized and matched directly on gels, autoradiographic or photographic films or photocopies on transparency overlays. The presence, absence and intensity of a band in particular location is noted relative to the weight standards, usually with the help of rulers or alignment devices and a light box. Allele matching by eye and hand is demanding and may suffer from error and bias from investigator (Micheli and Bova, 1997). However, DNA fingerprinting requires a precise and objective matching rule for declaring whether two samples match (National Research Council, 1992).

Nowadays, image analysis for DNA fingerprinting can now be automated with appropriate hardware and software tools. The image can be recorded via a high resolution video camera, scanning device, or phosphorimage analyzer. The image is edited with appropriate software to reduce background, to perform band alignment, and band matching. Automated image analysis has increased the level of accuracy with which fingerprints are analysed and has simplified enormously the task (Gill *et al.*, 1991). Some of the computer software have been developed for analysing sliver stained images, such as "Kodak Digital Science 1D Image Analysis Software" (Eastman Kodak Company, 1997). However, it has a number of limitations. For each image analysis, only one single molecular weight standard (STR locus) can be

analysed at one time and it does not provide the function of sex identification. So more rapid image analysis system and software with more functions for the analysis of DNA silver stained electrophoretograms should be developed.

In this work, a low cost image analysis system consisting of an IBM compatible-PC and a flatbed scanner was set up for capturing the DNA gel pattern. In addition, a software package called DNAEASY for Windows was developed for manipulating DNA gel images. Comparing with other imaging instruments, the proposed image capture system is faster and cheaper and the DNA patterns can be digitized for archiving purposes. Furthermore, the software DNAEASY was designed specifically for analyzing DNA silver stained gel patterns. It can be used to determine the matching of allele band positions in sample with standard automatically within a short time interval. Three STR loci are able to be analyzed at the same time with sex identification provided. Moreover, the intensity of the allele bands are provided for the judgement of shutter bands. The results thus obtained are more accurate and objective than those deduced from the conventional methods.

5.2 Method of Investigation

5.2.1 Principles of DNA Fingerprinting Analysis

In DNA fingerprinting analysis, genomic DNA was isolated from samples such as bloodstains or buccal swabs from random individuals (Micheli and Bova, 1997). The DNA samples were quantified prior to amplification using the polymerase chain reactions-based polymorphisms in short tandem repeat (PCR-STR)

systems. Then, the amplified products were separated by gel electrophoresis and detected by silver staining method. Image analysis of the gel images was carried out after.

PCR is a DNA amplification technique which allows a million or more copies of a short region of DNA to be easily made. It has been particularly useful in the field of forensic science, where many biological evidence samples contain either extremely small quantities of DNA, or degraded DNA. Because the quantity of DNA fragment is amplified and improved. DNA fingerprinting can rely on methods of detection that do not use radioactive substances. Furthermore, the technique of PCR amplification permits the use of very small amount of tissue or body fluids, theoretically even a single nucleated cell for sample preparation (National Research Council, 1992). In forensics, the quantity and quality of DNA in a sample can be important factors in deciding how to proceed with the analysis of that sample. Therefore, samples containing only a few nanograms of degraded DNA are more suitable for analysis by PCR methods (Walsh *et al.*, 1992).

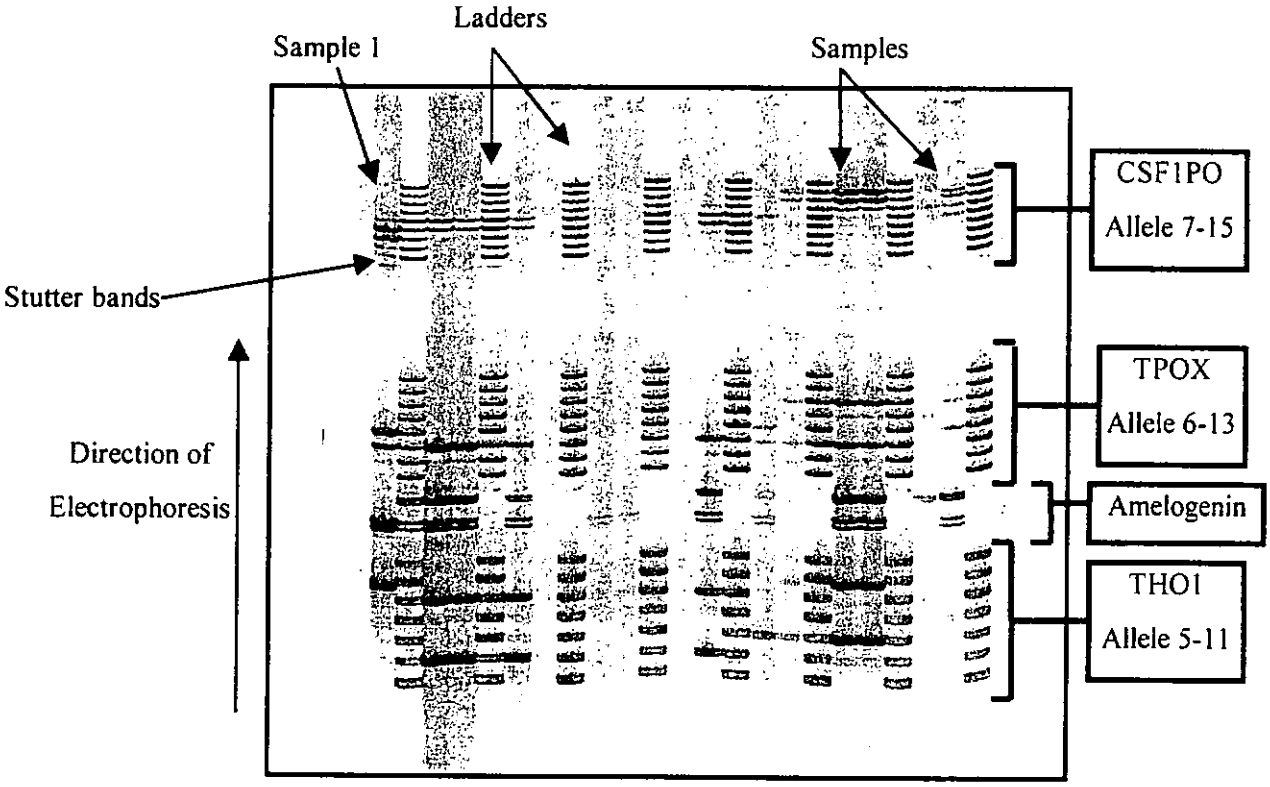
STR loci are a class of polymorphic markers which occur throughout the human genome and consist of simple tandemly repeated sequences of 1-6 base pair (bp) in length. DNA profiling based on PCR amplification of STRs (PCR-STR system) has the advantage of being more sensitive than conventional techniques. STR systems have been shown to be successful in typing material containing highly degraded DNA. Furthermore, the ability to resolve STR amplification products on polyacrylamide gels allows precise allele sizing (Kimption *et al.*, 1994). It also enhances the ability to distinguish discrete alleles and it can co-amplify three STR

loci in conjunction with the X-Y homologous gene Amelogenin for sex identification (Sparkes, 1996).

After DNA is processed by the PCR-STR system, DNA amplification fingerprinting (DAF) profiles are obtained. DAF profiles can be electrophoretically separated in polyester-backed polyacrylamide gels by using a vertical, open-faced horizontal or miniaturized electrophoretic unit. Separation of the DNA fragments which are produced by restriction enzyme cleavage is carried out by conventional sub-marine gel electrophoresis in agarous gels. The rate of the movement of the linear double-stranded molecules is directly proportional to the voltage applied and inversely proportional to the log of molecular weight (Kirby, 1990). After silver staining, clearly resolved DNA fingerprints are obtained (Figure 5.1).

Figure 5.1 shows a DNA fingerprinting gel image after sliver staining. The direction of the electrophoresis was from the bottom to the top of the gel. In this work, the gel were separated into four regions. THO1, TPOX and CSF1PO were the three STR loci. And the Amelogenin was the X-Y homologous gene for sex identification. They were developed from the bottom to the top of the lane. From the left to the right of the image, the gel was divided into twenty-four lanes. Eight of them were size-markers (ladders) which are standard lanes used for comparison with the sample individuals and monitoring the electrophoretic variations during the

Figure 5.1 A DNA fingerprinting gel image after sliver staining.



experiment. The other sixteen lanes were the sample lanes where indicating DNA samples obtained from suspects. One ladder was placed between two sample lanes. For the three loci in each ladder, the fragments of CSF1PO are the lightest whereas the fragments of TPO1 are the heaviest. The lightest fragment in the CSF1PO locus moved fastest to the highest of the gel and the heaviest fragment in the THO1 locus moved slowest and stayed at the lowest part of the gel. Each of them also contains different molecular fragments of various weights themselves. Therefore, all the fragments were separated after electrophoresis.

For different bands in each locus, they are known as alleles. Each allele is assigned a number to indicate its molecular weight. For locus CSF1PO, there are 9 alleles with number ranging from 7-15. The lightest allele from the top is named allele 15. Down to the bottom, the heaviest allele is named allele 7. In similar principles, alleles numbers of TPOX and THO1 with range 6-13 and 5-11 are assigned respectively. Sample alleles molecular weight can be known after matching the sample and ladder alleles. However, in the experiment, artifactual production of "stutter" bands are frequently observed (Stimson and Mertz, 1997). These shutter bands which have generally lower band intensities are observed in the positions of the sample alleles. From experts empirical point of view, if the intensity of an allele is lower than 30% of the maximum intensity of the STR loci, it can be considered as a stutter band and needed not be considered in the allele assignment. They may confuse the investigators in the judgement of the allele matching.

5.2.2 Image Capture and Image Processing of DNA Gel Images

In this work, an image analysis system with an IBM PC compatible and an optical scanner were utilized to capture DNA gel images. After image acquisition, images were imported into the developed software package, DNAEASY, to perform the allele matching automatically. The results listed in the analysis report include the allele number matched for various loci in each sample together with its band intensity provided and the sex type indication of the sample individual. The algorithm of detection are listed as follows.

5.2.2.1 Matching of the Alleles between Sample and Ladder Lanes

To match the sample alleles with ladders, the first thing to do is to locate all the positions of alleles in the three loci of the ladders. To start, user has to assign a detection region. This detection region should be set to just containing all the sample and ladder lanes. DNAEASY will then automatically separate each of the sample and ladder lanes (Figure 5.2). The boundary lines are in blue where the central lines of the lanes of the samples and ladders are in green and red respectively. The central lines are set right at the middle of the lanes. The greylevel values of the pixels located at the ladder central lines are then determined and obtained. A sample greylevel plot across the central of a ladder is shown in Figure 5.3. The y-axis is the intensity of pixels and x-axis is the vertical position of the pixels on the central line from the top to the bottom of the gel image. Due to the intensity change across

Figure 5.2 Screen layout of the DNAEASY

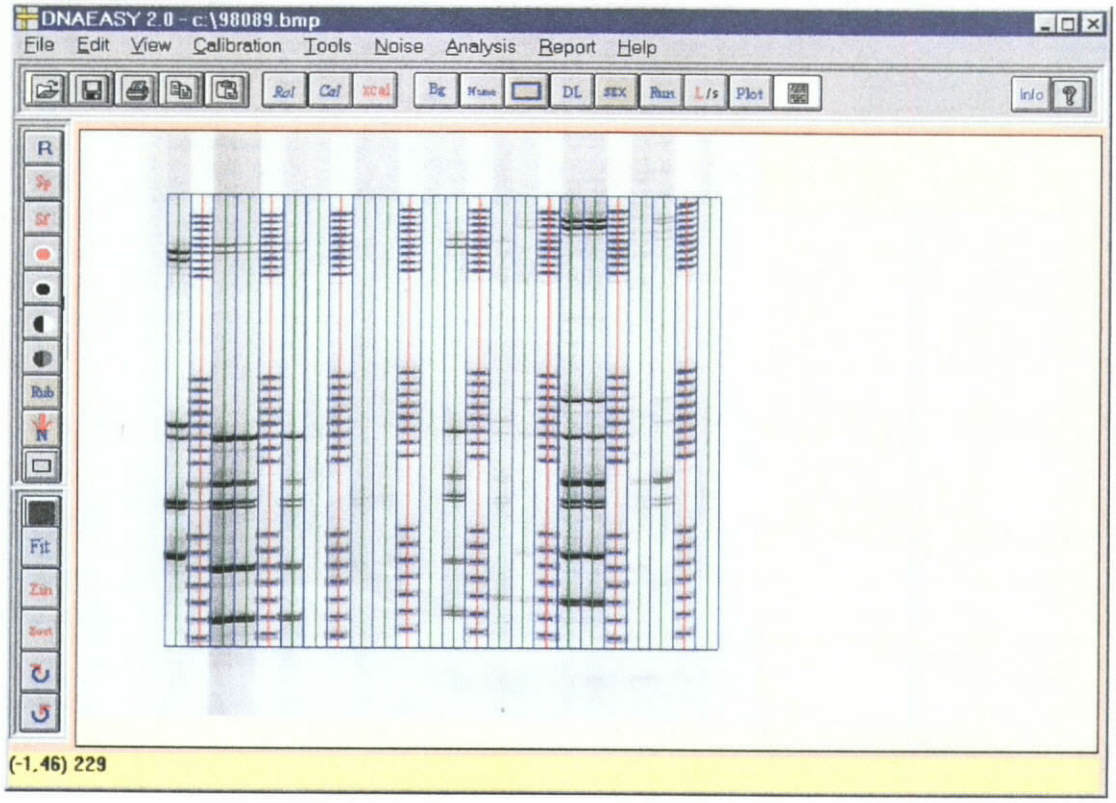
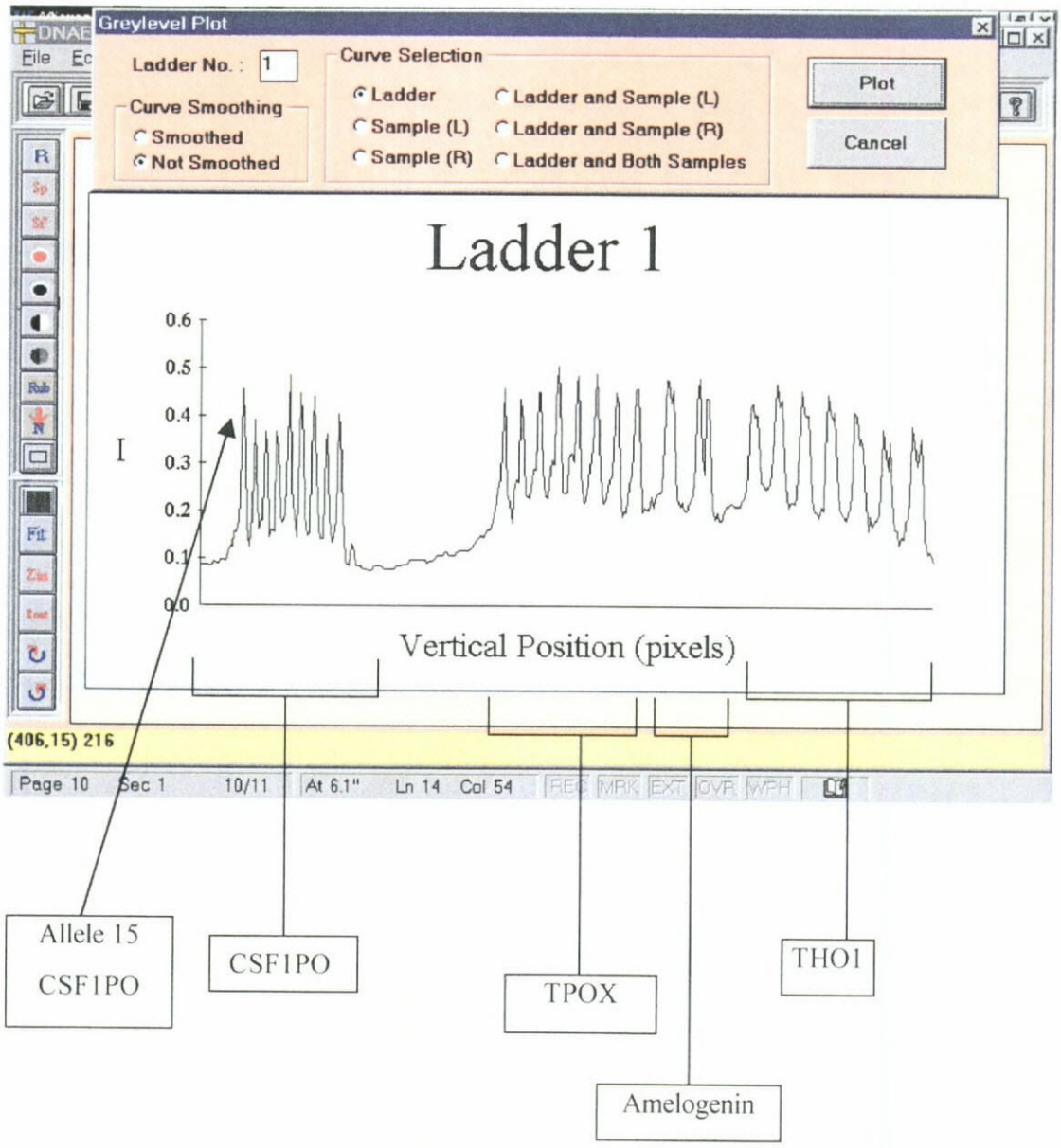


Figure 5.3 Greylevel plot of a ladder across a central line

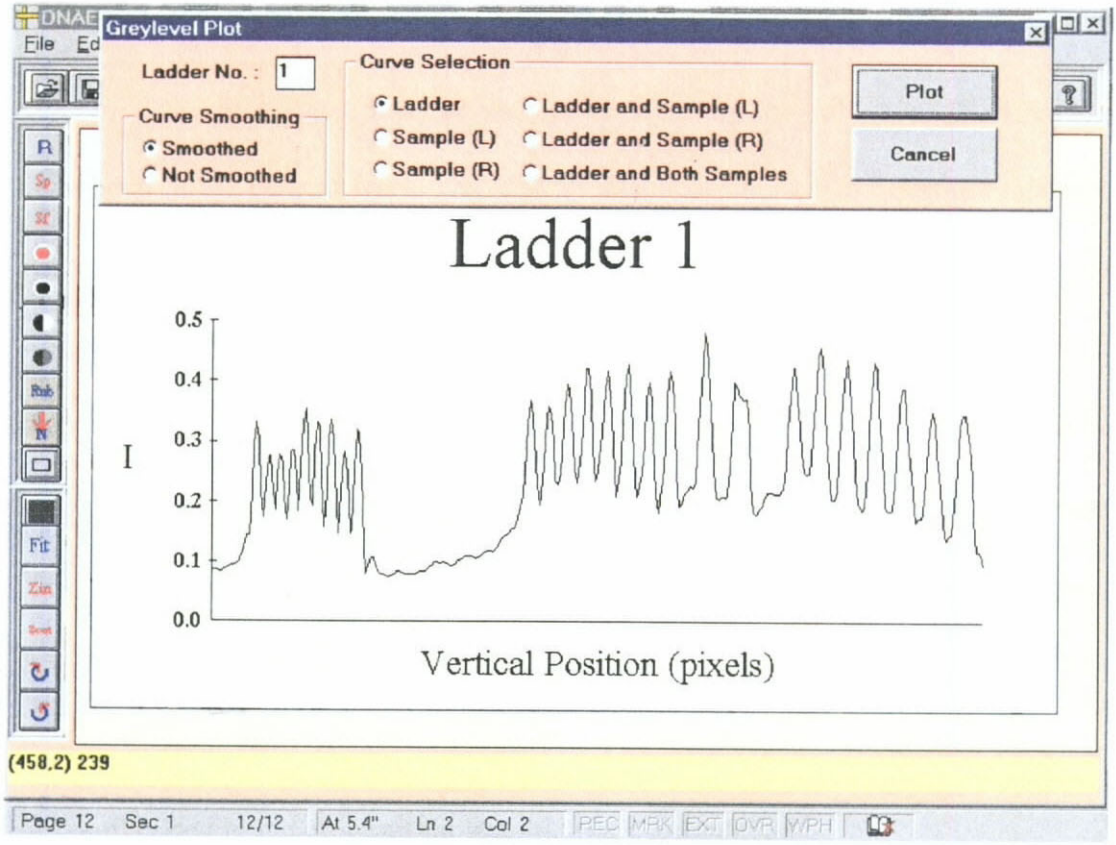


different alleles, there are many peaks found in the plot. And the positions of those peaks maxima are assigned to be positions correspond are the ladder alleles positions. There are some split peaks appearing in real situation as shown in Figure 5.3. These split peaks will certainly affect the correct detection of the allele positions. Data smoothing method such as Savitzky and Golay least squares procedures (Savitzky and Golay, 1964) was applied to smooth the curve. Figure 5.4 shows the result of a selected curve with smoothing. All the split peaks disappear with the proposed smoothing treatment. All the peak maxima and the alleles locations are then determined by the first derivatives of the curves and are indicated by blue lines (Figure 5.2). In the first derivative profile of the selected lane, the zero crossing point represents the location of the peak maximum. The sample lanes and alleles positions are detected similarly. Matching of detected allele bands in the sample lanes are completed by comparing the alleles positions with the standard ladder loci and the alleles numbers of the detected bands together with band intensities are assigned and reported respectively.

5.2.2.2 Determination of the Sex Type of a Sample Individual

The region named Amelogenin in Figure 5.1 is used for sex identification. Within this region, two bands can be identified. The upper and lower bands correspond to chromosome Y and X respectively. If DNAEASY detects two bands in this region, the sex type is male for that sample. On the other hand, if only one lower band X is detected, the sex type is female. The algorithm of determination of X and Y bands is similar to that mentioned in section 5.2.2.1.

Figure 5.4 Smoothed greylevel plot of a ladder across a central line



5.3 Experimental

5.3.1 Software Implementation

DNA gel images analysis was carried out by the software DNAEASY as developed in the present work. DNAEASY can help the user to determine matching between sample alleles and ladder alleles in various loci with known molecular weight automatically. Moreover, intensity of the matched alleles can be reported for the determination of the stutter bands. Also, DNAEASY is able to identify the sex type of the sample individuals. The program was coded in Microsoft® Visual Basic Version 3.0. Professional Edition (Microsoft Corporation, WA, US) under Microsoft® Windows™ 3.1 environment on a PC compatible with a 133MHz Pentium® processor. ImageKnife/VBX™ 2.0 (Media Architects, Inc., OR, US) was also embedded in the program for graphic file manipulation.

Same as CIEPEASY (Section 3.3.1), DNAEASY provides a user-friendly graphical interface for DNA gel pattern analysis (Figure 5.2). All the functions have their own pull down menus and icons. Major functions and options available are listed in Table 5.1. DNA electrophoretic pattern can be imported to DNAEASY in BMP, TIFF, GIF, PCX, JPEG, DIB or TGA graphic format. User may input sample names for individual identification. Image processing functions such as contrast, brightness, sharpness and softness adjustments are available. Rotation function is also provided for image alignment. Background noise incurred during the experiment can be reduced by noise reduction functions (Section 3.2.2.1). After the user defines the detection region by setting the boundary, matching of the sample and ladder

Table 5.1 Major functions available in DNAEASY

Options provided	Functions available
Image Processing	Adjustment of image size Adjustment of zooming ratio Adjustment of sharpness and softness Adjustment of contrast and brightness Conversion of a color image to that in grayscale Rotation of an image
Noise Reduction	Rubber erasing Noise bounding and cutting Noise level exclusion
Image Analysis	Determination of sample and ladder alleles positions with intensity Sex type identification of sample individuals Greylevel intensity plot for sample and ladder lanes
Reporting and Archives	Saving processed image Exporting of report to Microsoft® Write for further manipulation and archives.

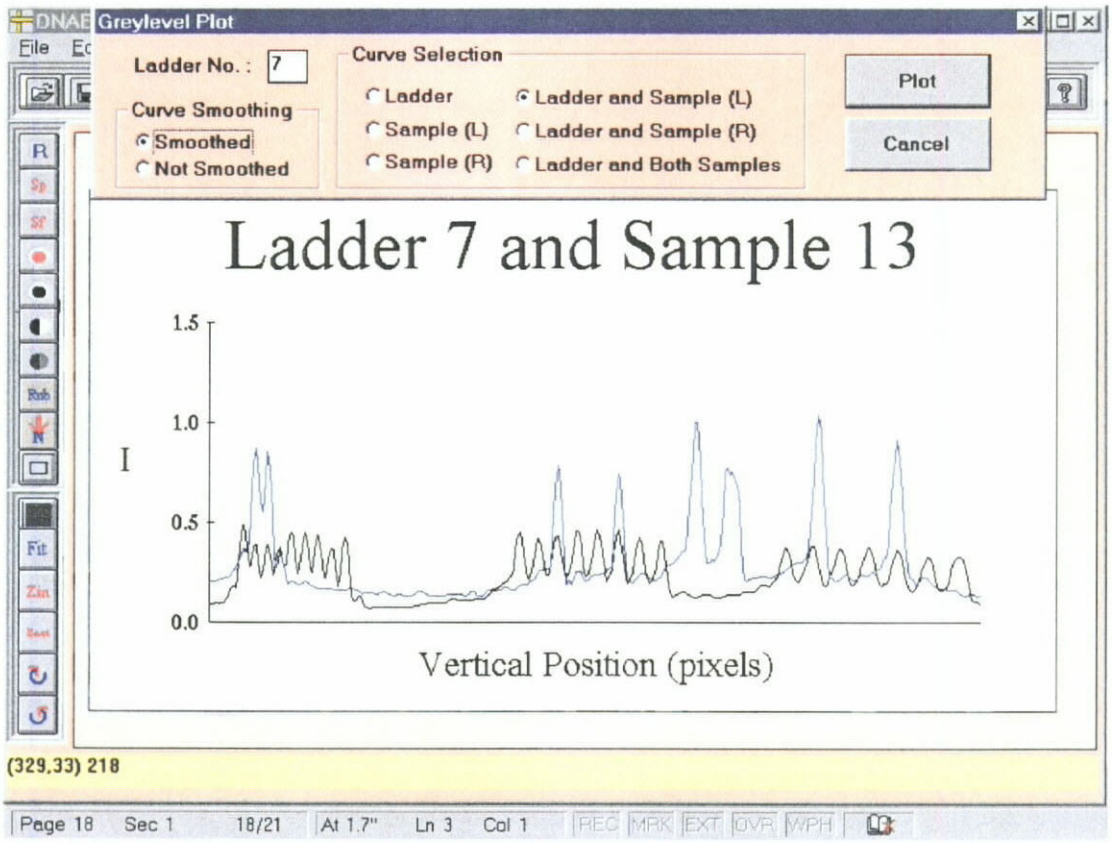
alleles, intensities of alleles and sex types will be determined automatically. Moreover, DNAEASY is able to show greylevel intensity plots across the central of lanes for both smoothed and non-smoothed curves (Figure 5.3 and 5.4). For visual determination of alleles matching, greylevel intensity curves of sample and ladder can be overlapped for comparison (Figure 5.5). With regard to data reporting and archiving, images processed by DNAEASY can be saved in the form of PCX, TIFF, TGA, JPEG, DIB or BMP graphic files. A hard copy of the image can also be provided from printing The "Copy" and "Paste" functions available in DNAEASY provide image exchange capability to other software packages such as Microsoft® Word or other image processing softwares. Moreover, information extracted from the image under study may be exported to Microsoft® Write for further data manipulation and storage.

In this investigation, images of DNA gel patterns as studied in this work were captured by a HP ScanJet IIc flatbed scanner (Hewlett-Packard Company, US) at a resolution of 75 dpi. All digitized images were stored using the BMP graphic format with 256 greyscales. No image post-processing steps such as contrast and brightness adjustment were applied on the scanned images. All scanned DNA fingerprinting gel images were then imported and analyzed by DNAEASY.

5.3.2 DNA Fingerprinting Experiment

The DNA gel were prepared and provided by Dr. Law, M.Y. and Dr. Chiu, H.C.T. from Forensic Science Division, Hong Kong Government Laboratory. The

Figure 5.5 Greylevel intensity plot shows overlapping curves of ladder and sample lanes. The black and blue lines represent the ladder and sample profiles respectively.



fingerprinting experimental procedures consist of three parts and are described in the following.

5.3.2.1 Genomic DNA Isolation and Quantiation

Genomic DNA were isolated from bloodstains or buccal swabs from random individuals using Chelex extraction procedure (Walsh *et al.*, 1991). DNA samples were quantified prior to amplification using the slot-blot hybridization method (Walsh *et al.*, 1992).

5.3.2.2 Amplification of Sample DNA

Co-amplification of the CSF1PO, TPOX, THO1 loci and Amelogenin was performed using the GenePrint™ STR System (Promega Corporation, Madison,WI). The PCR was carried out in 25μL reaction volumes containing 0.5-1.0 ng template DNA and 0.75 units of Taq DNA polymerase. The PCR reactions were placed into a PTC-100 thermal cycler (MJ Research Inc.,US). It involves denaturation at 96°C for 2 mins and then 10 cycles of denaturation at 94°C for 1 min; primer annealing at 64°C for 1 min; and primer extension at 70°C for 1.5 min. Then the PCR was continued for 20 cycles of denaturation at 90°C for 1 min; primer annealing at 64°C for 1 min; and primer extension at 70°C for 1.5 min.

5.3.2.3 Detection of Amplified Products

Amplified products were separated by electrophoresis through a 0.4 mm thick, 4 % denaturing polyacrylamide gel containing 0.5x TBE and 7M urea. 2.5 μL of each sample was mixed with 2.5 μL of the loading solution (10mM NaOH, 95% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol), denatured at 95°C for 3 minutes, and chilled on ice prior to loading 5 μL of the mixture. The gels were subjected to electrophoresis for 30 to 60 minutes prior to loading of the sample. Electrophoresis was performed at 40W with the SA 32gel apparatus (Gibco BRL, Bethesda, MD) in 0.5x TBE for about 1.5 hours at ambient temperature. The developed electrophoresis DNA patterns were then detected by the silver staining method (Bassam *et al.*, 1991).

5.4 Results and Discussion

The DNA gel shown in Figure 5.1 was analyzed by DNAEASY and the results of sample 1 was reported in Table 5.2 for the illustration of the software competence. By visual determination of Figure 5.1, there are 2 dark bands in CSF1SO locus, 2 dark bands in TPOX locus and 1 dark band in the TH01 locus. Moreover, there is a few gray bands appeared around the two dark bands of CSF1SO. These are the stutter bands (Figure 1). In the Amelogenin region, a lower band X was observed showing that the sex type of sample 1 is female. In Table 5.2, the alleles found that matched by DNAEASY in sample 1 were reported. Image

Table 5.2 Image Analysis Report of the DNA gel as shown in Figure 5.1

Image Analysis Report

Image information

File Path : c:\98089.bmp
Image Format : BMP
Color : 256 grayscales
Dimensions : 490 x 352 pixels
Resolution of image : x- 75 dpi , y- 75 dpi

Analysis Results

Sample 1

Loci :	Allele No.	(Relative Intensity in the same locus)
CSF1PO :	10.0	(100.0%)
CSF1PO :	9.0	(76.8%)
CSF1PO :	8.0	(4.4%)
TPOX :	9.0	(100.0%)
TPOX :	8.0	(69.0%)
THO1 :	9.3	(100.0%)
AMG- :	X	(100.0%)

information was also provided in the image analysis report. When comparing with the results, DNAEASY located all the dark alleles and reported sex type correctly. A grey band (allele 8) is also detected in CSF1SO. As the relative intensity is only 4.4% when comparing to the darkest alleles in the same locus, it is easy to judge that the allele is a stutter band coming from the artifacts during experiment.

An image analysis system was established with the use of an optical scanner for image acquisition and the software DNAEASY as developed in this work for image processing and analysis. Through which the efficiency on the determination of DNA fingerprinting on silver stained gel is enhanced. The system produces accurate matching within a shorter period of analysis time when compared with the human visual inspection. The analysis taken by DNAEASY on the gel in Figure 5.1 take only about 15 seconds for 16 samples. Moreover, when there is a large number of samples, repetitive analyses are tedious for visual determination. Human errors can be minimized by using DNAEASY. As objectiveness and accuracy are of great importance in the forensic science, image analysis by computer system is much more objective than human determination especially in the judgement of stutter bands. Furthermore, the proposed system provides an inexpensive way for DNA analysis and information such as images and reports can be saved for permanent records.

However, DNAEASY still has some limitations. The software supports only the loci CSF1PO, TPOX and THO1 and the sequence of lanes where ladder and sample are placed in electrophoresis is fixed. Improvements can be made to cope with these limitations in future development.

5.5 Conclusion

Image analysis technique was applied to enhance the DNA fingerprinting analysis of sliver stained gels. The system is easy to use and inexpensive. The results obtained are more accurate and objective and the analysis time is shortened when comparing with the conventional methods.

CHAPTER 6

Conclusion

Nowadays, image analysis is applied to different aspects of the society. Instead of using conventional methods mainly through visual and manual operation, image analysis usually enhance the extraction of essential information from images obtained from various applications. The analysis time is shortened significantly and human errors are minimized by avoiding repetitive manual measurements. Moreover, more information can be extracted and the results are more objective and reproducible. In this research, image analysis is the framework and associated methods were developed and applied to study two dimensional crossed immunoelectrophoretograms and DNA fingerprint electrophoretograms. A brief summary of the results are given in the following.

In Chapter 3, image analysis technique was applied to study 2D-CIEP gel patterns. Peak parameters such peak heights, migration distances, peak profile and peak areas were extracted by the developed image analysis system with software CIEPEASY automatically. A linear relationship was found between the peak area and the amount of antigen present in a sample and is reported for the first time in the literature. In Chapter 4, the system was applied to study abnormal 2D-CIEP patterns of prothrombin in patients with lupus anticoagulant. Patients with a previous history of thrombosis showed significantly reduced peak heights. In Chapter 5, the proposed image analysis system coupled with the self-developed software DNAEASY was applied to enhance DNA fingerprinting analysis, matching of the DNA alleles of the suspect and the standard automatically. Intensity of alleles and sex identification results are provided. More objective judgements can be obtained.

In conclusion, image analysis was applied into different fields of chemical and biochemical systems for information extraction to enhance the performance of the existing methods. Image analysis is anticipated to be adopted more widely in the other fields of chemistry and biochemistry in the future.

References:

Andrew, A.T. *Electrophoresis: Theory, Techniques & Biochemical and Clinical Applications*, 2nd Ed., Clarenden, Oxford, pp.205-240 (1986)

Bøg-Hansen T.C. "Two-dimensional Gel Electrophoresis". In Hames, B.D. and Rickwood, D., eds., *Gel Electrophoresis of Proteins: A Practical Approach*. 2nd Ed., IRL Press, Oxford, pp.287-300 (1990)

Baker, D.C., Robbe, S.L., Jacobson, L., Manco-Johnson, M.J., Holler, L. and Lefkowitz, J. "Hereditary Deficiency of Vitamin-K-Dependent Coagulation Factors in Rambouillet Sheep". *Blood Coagulation Fibrinolysis*, Vol. 10, pp.75-80 (1999)

Bassam, B.J., Anolles, G and Gresshoff, P.M. "Fast and Sensitive Silver Staining of DNA in Polyacrylamide Gels". *Analytical Biochemistry*, Vol. 196, pp.80-83 (1991)

Bowie, E.J.W., Thompson, J.H., Pascuzzi, C.A. and Owen, C.A. "Thrombosis in Systemic Lupus Erythematosus Despite Circulating Anticoagulants". *Journal of Laboratory and Clinical Medicine*, Vol. 62, pp.416-430 (1963)

Chan, A.C.M., Lo, S.C.L. and Chau, F.T. "Application of Image Analysis Techniques to Two-dimensional Crossed Immunoelectrophoresis Study". *Computers and Chemistry*, in press (1999)

Cobb, P. "Forensic Science". In White, P., ed., *Crime Scene to Court - The Essentials of Forensic Science*. The Royal Society of Chemistry, Cambridge, pp.1-4 (1998)

Cohen, H., Mackie, I.J., Anagnostopoulos, N., Savage, G.F. and Machin, S.J. "Lupus Anticoagulant, Anticardiolipin Antigbodies, and Human Immunodeficiency Virus in Haemophilia". *Journal of Clinical Pathology*, Vol. 42, pp.629-633 (1989)

Conley, C.L. and Hartmann, R.C. "A Hemorrhagic Disorder Caused by Circulating Anticoagulant in Patients with Disseminated Lupus Erythematosus". *Journal of Clinical Investigation*, Vol. 31, pp.621-622 (1952)

Darbyshire, J.F., Griffiths, B.S., Davidson, M.S. and McHardy, W.J. "Ciliate Distribution amongst Soil Aggregates". *Revue d'Ecologie et de Biologie du Sol*, Vol. 26, pp.47-56 (1989)

Dunbar, B.S. *Two-Dimensional Electrophoresis And Immunological Techniques*. Plenum Press, New York, pp.1-23, 123-159 (1987)

Dunn, M.J. and Burghes, A.H.M. "High Resolution Two-dimensional Polyacrylamide-gel Electrophoresis". In Dunn M. J., ed., *Gel Electrophoresis of Proteins*. Wright, Bristol, pp.203-261 (1986)

Dunn, M.J. *Gel Electrophoresis: Proteins*. βios Scientific Publishers, Oxford, pp.7-11, 129-138, 163-165 (1993)

Eastman Kodak Company. *Kodak Digital Science™ 1D Image Analysis Software*. New Haven (1997)

Edson, J.R., Vogt, J.M. and Hasegawa, D.K. "Abnormal Prothrombin-Crossed-Immunoelectrophoresis in Patients with Lupus Inhibitors". *Blood*, Vol. 64, pp.807-816 (1984)

Fleck, R.A., Rapaport, S.I. and Rao, L.V.M. "Anti-prothrombin Antibodies and the Lupus Anticoagulant". *Blood*, Vol. 72, pp.512-519 (1988)

Fowler, P.A., Casey, C.E., Cameron, G.G., Foster, M.A. and Knight, C.H. "Cyclic Changes in Composition and Volume of the Breast During the Menstrual Cycle, Measured by Magnetic Resonance Imaging". *British Journal of Obstetrics and Gynaecology*, Vol. 97, pp.595-602 (1990)

Freeman, H. "Image Processing and Pattern Recognition: A General Overview". In Cappellini, V and Marconi, R., eds., *Advances in Image Processing and Pattern Recognition.* Elsevier Science Publishers, Amsterdam, pp.3-6 (1986)

Fried, B and Sherma, J. *Thin-Layer Chormatography: Techniques and Applications.* Marcel Dekker, Inc., New York. 451pp (1994)

Gill. P. Evett. I.W., Woodroffe, S., Lygo, J.E., Millican, E., Webster. M. "Databases. Quality Control and Interpretation of DNA Profiling in the Home-Office-Forensic Science Service". *Electrophoresis.* Vol. 12. pp.204-209 (1991)

Glasbey, C.A., Horgan. G.W. and Darbyshire. J.F. "Image Analysis and Three-dimensional Modeling of Pores in Soil Aggregates". *Journal of Soil Science.* Vol. 42. pp.479-486 (1991)

Glasbey. C.A. and Horgan. G.W. *Image Analysis for the Biological Sciences.* John Wiley & Sons. New York. 218pp (1995)

Hall. E.L. *Computer Image Processing and Recognition.* Academic Press. New York. 584pp (1979)

Hewlett Packard Company. *Using the HP ScanJet IIc Scanner with Microsoft Windows.* US (1991)

Hutter. H., Brunner. C., Nikolov. S., Mittermayr. C.R. and Grasserbauer. M., "Imaging Surface Spectroscopy for Two- and Three-Dimensional Charaterization of Materials". *Fresenius' Journal of Analytical Chemistry.* Vol. 355. pp.585-590 (1996)

Jain. A.K. *Fundamentals of Digital Image Processing.* Prentice-Hall. Englewood Cliffs. NJ. 569pp (1989)

Kasturi. R and Trivedi. M.M. *Image Analysis Applications.* Marcel Dekker. New York. 436pp (1990)

Kimpton, C., Fisher, D., Watson, S., Adams, M., Urquhart, A., Lygo, J. and Gill, P. "Evaluation of an Automated DNA Profiling System Employing Multiplex Amplification of Four Tetrameric SRT Loci". *International Journal of Legal Medicine*, Vol. 106, pp.302-311 (1994)

Kirby, L.T. *DNA Fingerprinting: An Introduction*, Stockton, New York, 365pp (1990)

Kuni, C.C. *Introduction to Computers & Digital Processing in Medical Imaging*. Year Book Medical, London, 169pp (1988)

Leung, A.K.M., Chau, F.T. and Gao, J.B. "Wavelet Transform: A Method for Derivative Calculation in Analytical Chemistry". *Analytical Chemistry*, Vol. 70, pp.5222-5229 (1998)

Lo, S.C.L., Salem, H.H., Howard, M.A., Oldmeadow, M.J. and Firkin, B.G. "Studies of Natural Anticoagulant Proteins and Anticardiolipin Antibodies in Patients with the Lupus Anticoagulant". *British Journal of Haematology*, Vol. 76, pp.380-386 (1990)

Maltin, C.A., Hay, S.M. Delday, M.I., Lobley, G.E. and Reeds, P.J. "The Actions of the β-agonist Clenbuterol on Protein Metabolism in Innervated and Denervated Phasic Muscles". *Biochemistry Journal*, Vol. 261, pp.965-971 (1989)

Martin, N.J. and Fallowfield, H.J. "Computer Modeling of Algal Waste Treatment Systems". *Water Science and Technology*, Vol.21, pp.277-287 (1989)

Media Architects Inc. *ImageKnife/VBA™ Reference Manual Version 2.0*, United States, 217pp (1996)

Melvin M. *Electrophoresis*, John Wiley & Sons, Chichester, pp.82-86 (1987)

Micheli. M.R. and Bova, R. *Fingerprinting Methods Based on Arbitrarily Primed PCR*. Springer, Heidelberg , 439pp (1997).

Michiels. J.J. and Hamulyak, K. "Laboratory Diagnosis of Hereditary Thrombophilia". *Seminars in Thrombosis and Hemostasis*. Vol. 24. pp.309-320 (1998)

Morrison. M. *The Magic of Image Processing*. Sams Publishing. Carmel. 303pp (1993)

Murakami. F.. Iijima. K., Nakamura, K.. Ikawa. S.. Hayashibara. H. and Siraki. K. "Transitions of Each Inhibitor in A Patients with Lupus Anticoagulant and Anti-prothrombin Antibody". *Rinsho Byori*. Vol. 44. pp.692-696 (1996)

Nasher. K. and Ciznar. I. "Characterization of Shigella-Dysenteriae Type-1 Capsular Polysaccharide by Immunochemical Methods". *Folia Microbiologica*. Vol. 43. pp.707-712 (1998)

National Research Council (U.S.). Committee on DNA Technology in Forensic Science. *DNA Technology in Forensic Science*. National Academy Press. Washington. 185pp (1992)

Nikolov. S.G.. Hutter. H. and Grasserbauer. M.. "De-noising of SIMS Images via Wavlet Shrinkage". *Chemometrics and Intelligent Laboratory Systems*. Vol. 34. pp.263-273 (1996)

Rickwood. D.. Chambers. J.A.A. and Spragg. S.P. "Two-dimensional Gel Electrophoresis". In Hames. B.D. and Rickwood. D.. eds.. *Gel Electrophoresis of Proteins: A Practical Approach*. 2nd Ed.. IRL Press. Oxford. pp.217-272 (1990)

Rilbe. H. "Basic Theory of Electrophoresis: Definitoins. Terminology and Comparison of the Basic Techniques". In Simpson. C.F. and Wittaker. M.. eds.. *Electrophoretic Techniques*. Academic Press. London. pp.1-25 (1983)

Rosenfeld, A. and Kak, A.C. *Digital Picture Processing*. Academic Press, New York, 457pp (1982)

Ruggieri, A.P., Galbraith, R.M. and Lazarchick, J. "Common Variable Immune Deficiency Syndrome Associated with the Lupus Anitcoagulant". *Clinical and Experimental Rheumatology*, Vol. 3, pp.71-73 (1985)

Russ, J.C. *The Image Processing Handbook*. 2nd Ed., CRC Press, Boca Raton, Florida (1995)

Russel, A.J.F. "Cashmere Production – the Viable Alternative". *Outlook on Agriculture*, Vol.20, pp.39-43 (1991)

Savitzky, A. and Golay, M.J.E. "Smoothing and Differentiation of Data by Simplified Least Square Procedures". *Analytical Chemistry*, Vol. 36, pp.1627-1639 (1964)

Schleider, M.A., Nachman, R.L., Jaffe, E.A. and Coleman, M. "A Clinical Study of the Lupus Anticoagulant". *Blood*, Vol. 48, pp.499-509 (1976)

Shah, J.K., Mitchell, L.G., Paes, B., Ofosu, F.A., Schmidt, B. and Andrew, M. "Thrombin Inhibition Is Impaired in Plasma of Sick Neonates". *Pediatric Research*, Vol. 31, pp.391-395 (1992)

Shirotani, H., Tokunaga, F. and Koide, T. "Cellular and Functional-Characterization of 3 Recombinant Antithrombin Mutants that Caused Pleiotropic Effect-Type Deficiency". *Journal of Biochemistry*, Vol. 125, pp.253-262 (1999)

Simm, G. "Selection for Lean Meat Production in Sheep". In Speedy, A.W. eds., *Recent Advances in Sheep and Goat Research*, CAB International, pp.193-215 (1992)

Sondergaard. I., Conradsen, K., Hagerup, M. and Poulsen. L.K. "Image-Processing and Pattern-Recognition Algorithms for Evaluation of Crossed Immunoelectrophoretic Patterns – (Crossed Radioimmunoelectrophoresis Analysis Manager Cream)". *Analytical Biochemistry*, Vol. 165, pp.384-391 (1987)

Sparkes. R., Kimpton. C., Gilbard, S., Carne. P., Andersen. J., Oldroyd. N., Thomas. D., Urquhart. A. and Gill, P. "The validation of a 7-locus multiplex STR test for use in forensic casework". *International Journal of Legal Medicine*. Vol. 109. pp.195-204 (1996)

Stimson. P.G. and Mertiz. C.A. *Forensic Dentistry*. CRC Press LLC. Boca Raton. 301pp (1997)

Strachan. N.J.C., Nesvadba. P. and Allen. A.R. "Fish Species Recognition by Shape Analysis of Images". *Pattern Recognition*. Vol. 5. pp.539-544 (1990)

Svendsen. P.J. "Quanitiative Immunoelectrophoresis". In Deyl. Z., Everaerts. F.M., Prusik. Z. and Svendsen. P.J., eds., *Electrophoresis A Survey of Techniques and Applications Part A: Techniques*. Elsevier Scientific Publishing Company. Amsterdam. pp.133-165 (1979)

Tsukiyama. T. and Shirai. Y. "Detection of the Movements of Persons from a Sparse Sequence of TV Images". *Journal of Pattern Recognition*. Vol.18. pp.207-213 (1985)

Walsh. P.S., Metzgar. D.A. and Higchi. R. "Chelex-100 as a Medium for Simple Extraction of DNA for PCR-Based Typing from Forensic Material". *Biotechniques*. Vol. 1. pp.91-98 (1991)

Walsh. P.S., Varario. J. and Reynods. R. "A Rapid Chemiluminescent Method for Quantitation of Human DNA". *Nucleic Acids Research*. Vol. 20. pp.5061-5065 (1992)

Watson, N. "The Analysis of Body Fluids". In White. P., ed., *Crime Scene to Court – The Essentials of Forensic Science*, The Royal Society of Chemistry. Cambridge, pp.289-320 (1998)

Wegner, T. *Image Lab*, Waite Group Press, Corte Madera, pp.53-58 (1992)

Wojnar. L. and Majorek. M. *Komputerowa Analiza Obrazu*. Fotobit-Design. Cracow (1994)

Wojnar. L. *Image Analysis Applications in Materials Engineering*. CRC. New York. 245pp (1999)

Xu. B. *Assessment of Carpet Appearance by Image Analysis*. University Microfilms International. Michigan. 154pp (1994)

List of Publications and Conference/Symposium Presentation

Journal Articles

1. Chan, C.M., Lo, S.C.L., Chau, F.T. and Leung, A.K.M. "Application of Image Analysis Technique to Two-dimensional Crossed Immunoelectrophoresis Study". *Computers and Chemistry*, in press (1999)

2. Lo, S.C.L., Chan, A.C.M., Kwok, F.S.L. and Chau, F.T. "Abnormal 2-Dimensional Cross Immunoelectrophoretic Patterns of Prothrombin in Patients with the Lupus Anticoagulant", submitted for publication (1999)

3. Chan, T. P., Chan, C.M., Chau, F.T., Lo, S.C.L., Law, M.Y. and Chiu, C.T. "Application of Image Analysis Technique to Enhance Fingerprinting Analysis in Forensic Science", submitted for publication (1999)

Conference Presentation

1. Chan, C.M., Chau, F.T. and Lo, S.C.L. "Application of Image Analysis to Two-Dimensional Crossed Immunoelectrophoresis". A poster presented at the Fourth Symposium on Chemistry Postgraduates Research in Hong Kong, Hong Kong Baptist University, Hong Kong, April 19, 1997, A-23 (1997)

2. Chan, C.M., Lo, S.C.L. and Chau, F.T. "Application of Image Analysis Technique to Study Abnormal Two-dimensional Crossed Immunoelectrophoresic Patterns of Prothrombin in Symposium in Lupus Anticoagulant". A poster presented at the Fifth Chemistry Postgraduate Research Symposium in Hong Kong in conjunction with The First Annual Congress of The Hong Kong Chemical Society, The Hong Kong Polytechnic University, Hong Kong, April 25, 1998, P-17 (1998)