1992

Novel detection schemes and automated image analysis algorithms for planar chromatography and gel electrophoresis

Lance Bryant Koutny
Iowa State University

Order Number 9302006

Novel detection schemes and automated image analysis algorithms for planar chromatography and gel electrophoresis

Koutny, Lance Bryant, Ph.D.

Iowa State University, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Novel detection schemes and automated image analysis algorithms for planar chromatography and gel electrophoresis

by

Lance Bryant Koutny

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Chemistry
Major: Analytical Chemistry

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1992

TABLE OF CONTENTS

GENERAL INTRODUCTION.....1
Introduction to Charge Coupled Devices (CCDs).....2
Introduction to Personal Computers in the Lab.....37

PAPER 1. LASER-BASED INDIRECT FLUOROMETRIC DETECTION AND QUANTIFICATION IN THIN-LAYER CHROMATOGRAPHY.....43

INTRODUCTION.....44

EXPERIMENTAL SECTION.....47
Apparatus.....47
Chromatography.....50

RESULTS AND DISCUSSION.....51

CONCLUSION.....62

REFERENCES.....63

PAPER 2. ON-LINE DETECTION OF DNA IN GEL ELECTROPHORESIS BY ULTRAVIOLET ABSORPTION UTILIZING A CHARGE COUPLED DEVICE IMAGING SYSTEM.....66

INTRODUCTION.....67

EXPERIMENTAL SECTION.....70
Restriction Digest.....70
Gel Electrophoresis.....70
Detection.....71

RESULTS AND DISCUSSION.....75

CONCLUSION.....91

REFERENCES.....93

PAPER 3. ON-LINE DETECTION OF PROTEINS IN GEL ELECTROPHORESIS BY NATIVE ULTRAVIOLET ABSORPTION AND FLUORESCENCE UTILIZING A CHARGE COUPLED DEVICE IMAGING SYSTEM.....95

INTRODUCTION.....96

EXPERIMENTAL SECTION.....98

Gel Electrophoresis.....98
Detection.....99

RESULTS AND DISCUSSION.....103
General Considerations.....103
Native Fluorescence.....106
Ultraviolet Absorption.....111
Coomassie Blue Staining.....114
Detection Response.....114

CONCLUSION.....123

REFERENCES.....125

PAPER 4. AUTOMATED IMAGE ANALYSIS FOR DISTORTION COMPENSATION IN SEQUENCING GEL ELECTROPHORESIS.....127

INTRODUCTION.....128

EXPERIMENTAL SECTION.....133
Electrophoresis.....133
Image Acquisition.....134
Image Analysis.....136

RESULTS AND DISCUSSION.....141

REFERENCES.....162

PAPER 5. AN EXPERT SYSTEM FOR DATA ACQUISITION TO ACHIEVE CONSTANT SIGNAL-TO-NOISE: APPLICATION TO IMAGING OF DNA SEQUENCING SLAB GELS.....164

INTRODUCTION.....165

EXPERIMENTAL SECTION.....170
Preliminary Feature Recognition.....170
Specific Feature Recognition.....171
Decision Making.....173
Signal Accumulation.....174
Termination.....175
Test System.....175

RESULTS AND DISCUSSION.....182

REFERENCES.....195

GENERAL SUMMARY.....197

ADDITIONAL REFERENCES.....199

ACKNOWLEDGEMENTS.....202

GENERAL INTRODUCTION

The past 20 years have seen an explosion in the capabilities and availability of electronic semi-conductor devices which were used in all of the projects in this dissertation. The most publicized advance has been the continuing development of more powerful and accessible stand-alone computers commonly termed "personal computers." Another revolution in electronics has occurred in the field of video imaging, as evidenced by the recent advances made in the consumer camcorder industry. These devices are possible because of a solid-state imaging chip known as the "CCD." The remainder of this general introduction will be devoted to both of these technologies. The body of this dissertation will consist of five papers that have been published or submitted to Analytical Chemistry or Applied Spectroscopy.

Paper 1 (1) demonstrates computer control of a laser scanning imaging system for thin-layer chromatography. Papers 2 (2) and 3 (3) describe the application of CCD imaging to slab gel electrophoresis. All three of these papers also demonstrate enhanced performance of chemically non-intrusive detection schemes.

Paper 4 (4) is an example of automated image analysis directed at correcting signal distortions in slab gel electrophoresis. Paper 5 (5) is a computerized expert system

that demonstrates the modern personal computer's ability to interactively control experimental data acquisition, in this case, two-dimensional imaging.

Introduction to Charge Coupled Devices (CCDs)

In 1988, Photometrics Ltd. of Tucson, Arizona, introduced the first "scientific-grade cooled slow-scan charge coupled device imaging system," opening up numerous new possibilities in the field of analytical imaging. The first systems were a collection of many advances that had been made over the previous 30 years that were combined for the first time to obtain the ultimate levels of performance for the already well developed sensor known as the "charge coupled device" (CCD) (6).

History of solid-state imagers

The predecessors of true silicon-based imagers were light sensitive transistors and diodes created as a by-product of the extensive research being performed on the first solid-state devices in the early 1960s. Because of their high cost and poor performance, these initial devices were mostly considered an interesting oddity.

During the 1970s, small-scale integration followed by large-scale integration, and finally very large-scale

integration of solid state circuits allowed the assembly of small silicon chips with two-dimensional arrays of circuits. In the late 1970s and early 1980s, this technology was applied to create solid state imagers targeted for the consumer market for television cameras. These sensors worked quite well for the intended purpose, but were not of high enough quality for highly accurate and quantitative imaging.

By the mid 1980s, General Electric and several other manufacturers recognized the need for quantitative imaging in many scientific fields and created the first generation of truly high performance sensor arrays. These solid-state sensors fall into three categories: the charge coupled device, the charge injection device, and photodiode arrays. In the mid-1980s the CCD, for reasons of performance and availability, was chosen to be the sensor in modern imaging devices (7).

Principles of operation

Charge transfer devices are solid-state multichannel detectors that electronically integrate photon signal information in a manner analogous to the chemical based process in photographic film. Each individual element in the detector array is composed of several conductive electrodes overlying an insulating layer that forms a series of metal oxide semiconductor (MOS) capacitors. The insulator separates

the electrodes from a doped silicon region where the photo-generated charge is stored (8).

Charge coupled device structure and operation

Figure 1 shows a schematic diagram of a CCD. The serial register is simply a one-dimensional CCD which is used during readout. The majority of the area consists of the imaging area which is called the parallel register. This register consists of electron conducting channels which are vertical in this schematic. Permanent barriers referred to as channel stops prevent charge from migrating from one channel to the next. Perpendicular to this are the electrodes, or gates, which are used to collect and localize the photogenerated charge during integration, and, with proper potential switching, are used to shift the charge during readout. Figure 2 demonstrates the principle of charge collection in a single pixel during integration.

The absorption of a photon causes the formation of a hole-electron pair. The hole is held in the potential well formed by the negatively biased gates, and thus does not travel in the channel. Combining this with the channel stops localizes the charge in the pixel. A typical pixel is 20 x 20 μm with a capacity of 20,000 to 300,000 electrons. When "full well capacity" was exceeded on early CCD models the charge

Figure 1. Schematic drawing of the basic configuration of a typical full-frame imaging CCD.

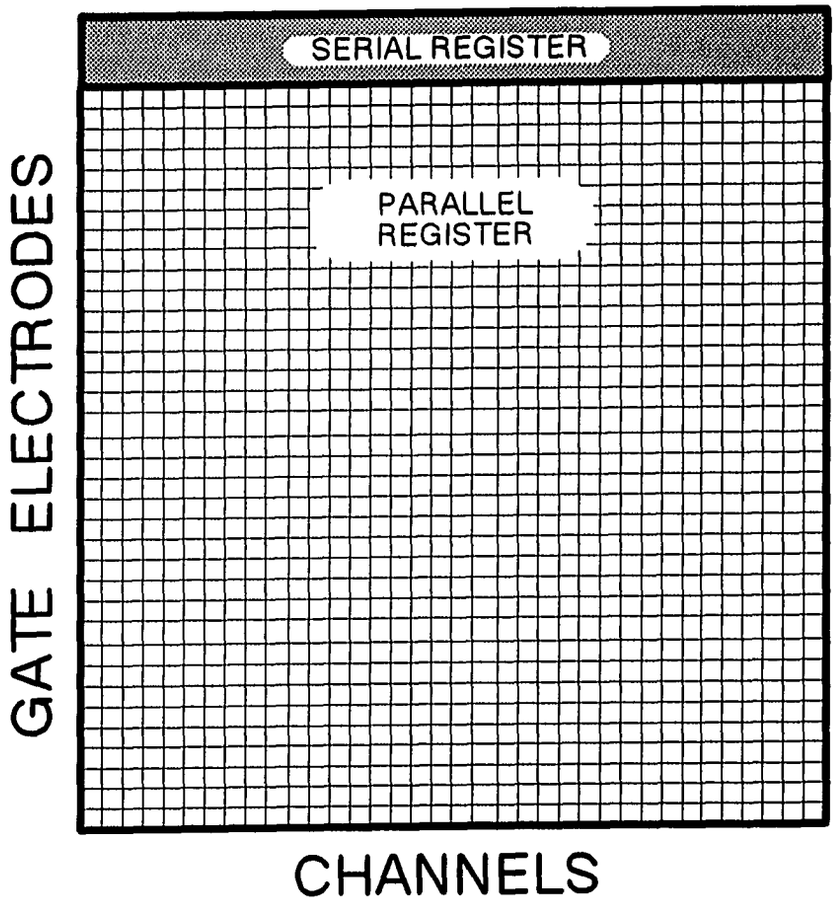
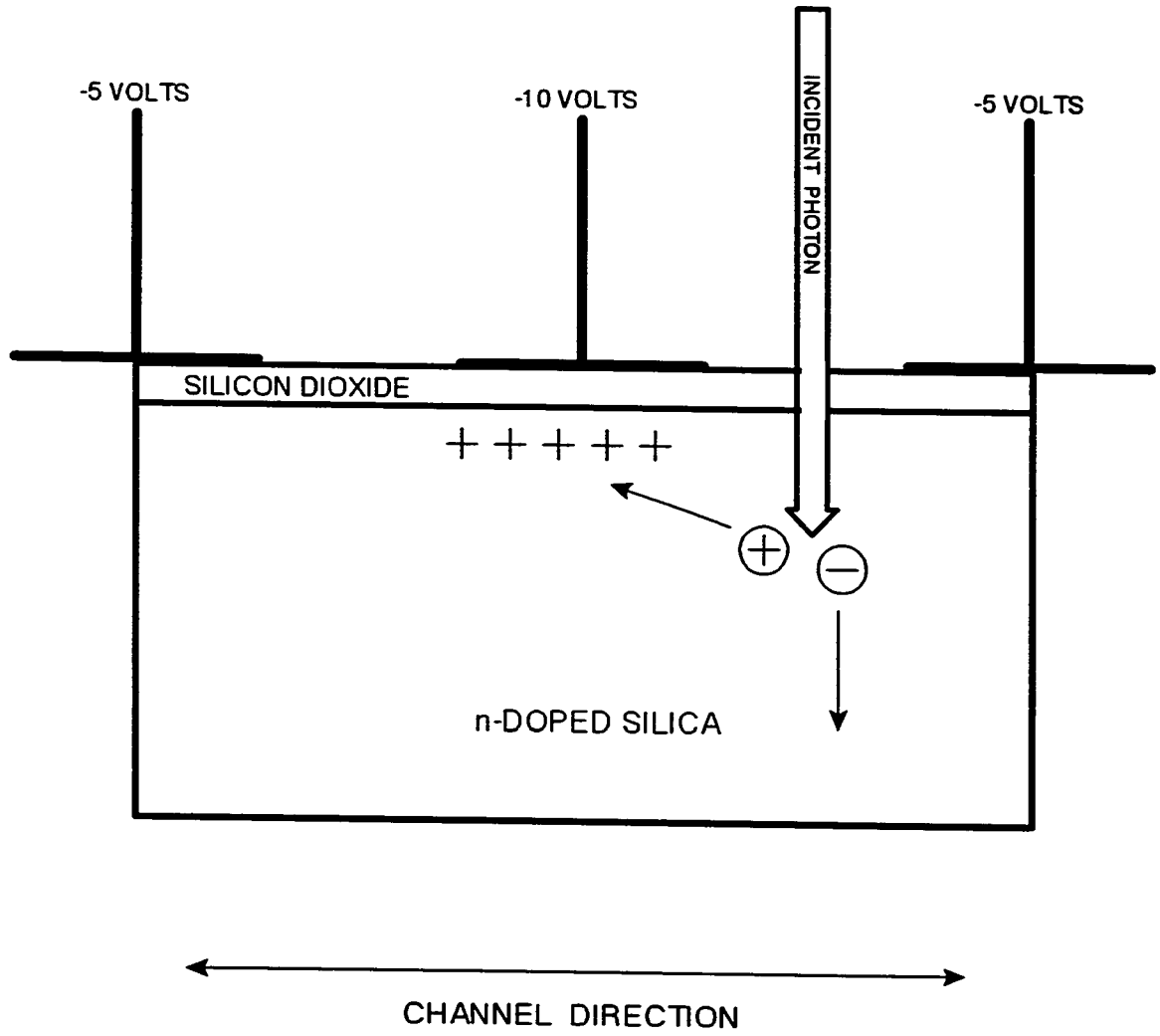

SERIAL REGISTER
PARALLEL
REGISTER
GATE ELECTRODES
CHANNELS

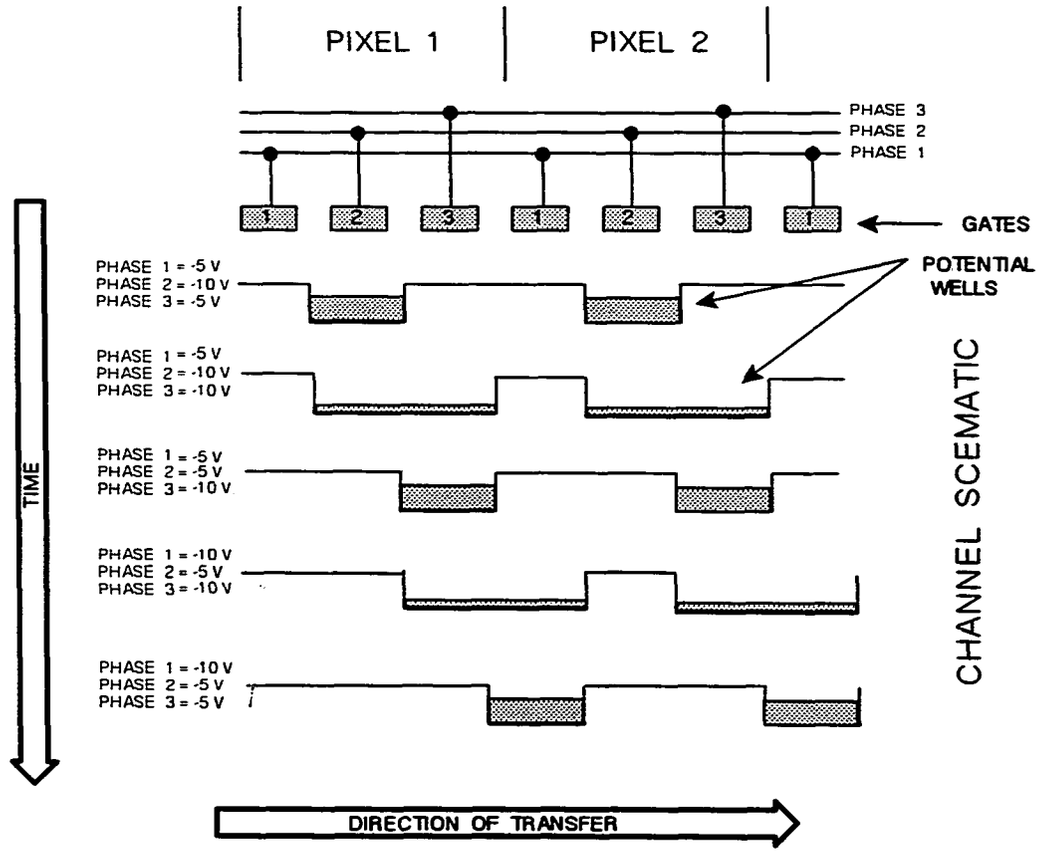
Figure 2. Schematic representation of a single pixel in an n-doped negatively biased CCD during signal integration.


-5 VOLTS
-10 VOLTS
INCIDENT PHOTON
-5 VOLTS
SILICON DIOXIDE
+ + + + +
n-DOPED SILICA
CHANNEL DIRECTION

spilled over into neighboring pixels. This condition was termed blooming, since extremely bright pixels appeared to expand, or bloom, into the surrounding image area. Most CCDs used since 1989 have special antiblooming architectures that provide a shunt for this excess charge. This feature is quite useful when looking for very weak optical signals of interest that are located close to extremely large signals since it allows long integration times without the large signal blooming out into the pixels containing the smaller signal.

Once the desired charge has been collected, readout must occur via a process known as charge transfer. Figure 3 is a schematic diagram of a section of a conductive channel in a three-phase CCD. By switching the voltage on the gates as indicated by the values for each given time, the potential wells move up the channels toward the serial register. With every six voltage switches, the potential wells and their associated charge packets move one pixel closer to the serial register. The gate switching is clocked such that all channels shift at the same time. This is the reason that all of the channels, as one unit, are called the parallel register. During a one pixel shift the charge packets at the top of the parallel register are transferred into the serial register shown in figure 1. At this point, parallel charge transferring is temporarily halted, and charge transfer occurs in the serial register toward the output amplifier in a manner

Figure 3. Schematic representation of the charge transfer process in two pixels contained in one channel of a 3-phase CCD.


PIXEL 1
PIXEL 2
PHASE 3
PHASE 2
PHASE 1
1
2
3
1
2
3
1
GATES
POTENTIAL WELLS
PHASE 1 = -5 V
PHASE 2 = -10 V
PHASE 3 = -5 V
PHASE 1 = -5 V
PHASE 2 = -10 V
PHASE 3 = -10 V
PHASE 1 = -5 V
PHASE 2 = -5 V
PHASE 3 = -10 V
PHASE 1 = -10 V
PHASE 2 = -5 V
PHASE 3 = -10 V
PHASE 1 = -10 V
PHASE 2 = -5 V
PHASE 3 = -5 V
TIME
DIRECTION OF TRANSFER
CHANNEL SCEMATIC

identical to that described for the parallel register. When an entire pixel of information has been transferred to the output amplifier, an amplified and measurable signal is produced, digitized, and stored in memory. Once the entire serial register has been read, another row of pixels is transferred from the parallel to the serial register and read out. This process continues until the entire image has been read.

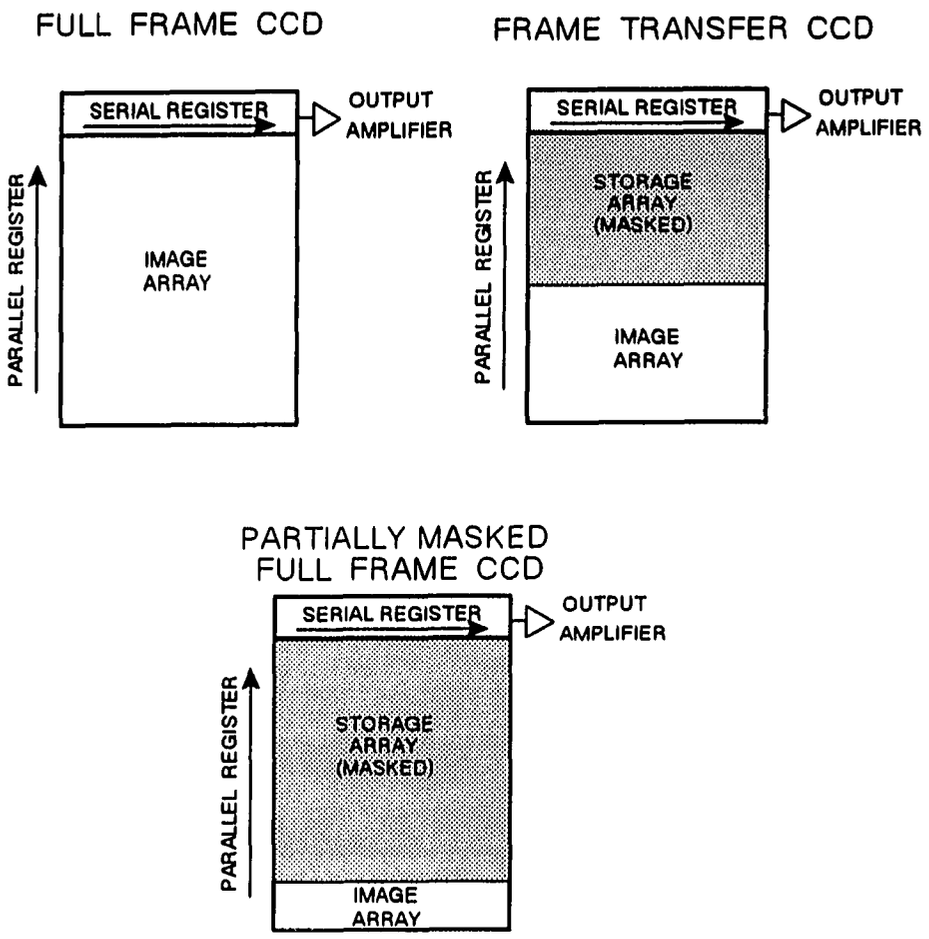
One major operating principle that must be considered is the readout rate of the pixel information. In traditional video applications the entire CCD had to be read out in the dead time between frames, which could not exceed one-thirtieth of one second. Thus, a typical 384 x 576 pixel CCD had to be clocked at 10 to 20 MHz. Unfortunately, read noise is directly related to readout rate, and at these high rates CCD performance suffers greatly in terms of reduced dynamic range and photometric accuracy. To reduce this noise to near zero, readout rates ranging from 50-250 kHz are used (7).

CCD architectures

There are three CCD architectures available, each with its own advantages and intended use. They are shown schematically in figure 4.

The most common type, and the one used in the following studies, is the full frame CCD. A full frame CCD has a single

Figure 4. Schematic diagrams of the three commonly available CCD formats.


FULL FRAME CCD
SERIAL REGISTER
OUTPUT AMPLIFIER
PARALLEL REGISTER
IMAGE ARRAY
FRAME TRANSFER CCD
SERIAL REGISTER
OUTPUT AMPLIFIER
PARALLEL REGISTER
STORAGE ARRAY (MASKED)
IMAGE ARRAY
PARTIALLY MASKED FULL FRAME CCD
SERIAL REGISTER
OUTPUT AMPLIFIER
PARALLEL REGISTER
STORAGE ARRAY (MASKED)
IMAGE ARRAY

parallel register used for photon collection and charge transfer. The most common mode of operation is the snapshot mode. A camera lens with an electronic shutter is placed in front of the CCD to collect light and focus an image onto the CCD. The set up is identical to a typical 35 mm camera, except that the photographic film has been replaced with the CCD. The shutter is closed prior to image acquisition, and any residual charge is cleared from the CCD by executing several rapid readouts. The shutter is then opened, and charge is allowed to collect. When the desired amount of integration has occurred, the shutter is closed and normal slow-scan readout occurs.

The major disadvantage of this type of CCD is that the shutter must be closed during readout to avoid blurring. Since the readout rate is limited by the slow speed of the serial register, any image accumulated during this time would be smeared during vertical charge transfer. If one is not concerned about missing a few seconds worth of signal, this type is the logical choice since the full frame CCD is the least expensive of the three.

The second type of CCD is the frame transfer device. This CCD is composed of one serial and two parallel registers. The lower parallel register functions as the imaging array, while the upper parallel register is covered with an opaque mask that prevents image collection. This upper register is

called the storage array. These CCDs can operate in a manner that eliminates the shutter and minimizes vertical blurring. First, an image is accumulated on the imaging array. When adequate signal has been accumulated, the image is quickly transferred via the parallel clocks to the masked storage array, and a new image begins accumulating in the image array. During this image accumulation, the image under the mask can be transferred to the serial register and read out at the normal slow scan rates. This design has the distinct advantage of being able to operate without a shutter, and, therefore, no events are missed. Exposure time, however, cannot be faster than storage array readout time or the series of images will pile up on each other. Smearing is only a problem if events happen quite quickly, since transferring the image to the storage array occurs in a parallel manner. Parallel shifts can occur at faster rates than are used in the serial register where readout must occur, and typically range from 10 to 20 MHz. In a typical frame transfer CCD with an image array of 500 x 500 pixels and a transfer rate of 10 MHz, the 500 shifts occur in on 50 microseconds.

The last type of CCD available is the partially masked full frame CCD. In this design, a large fraction of the upper part of the parallel register is masked, with a small area left unmasked at the bottom for image integration. The system operates by collecting a series of images by alternately

integrating and shifting the parallel register. When the masked area is full, the entire CCD is read out as in the full frame design. The result is an image composed of a series of images the size of the unmasked area stacked on top of each other. The advantage with this design is that a series of images can be collected quickly (serial readout does not limit exposure time), and the time between images can be rather short. Time resolution is limited only by the parallel transfer time. If a 1000 x 1000 pixel CCD with a 90% mask is used, the image area will be 100 x 1000 and 10 images can be acquired. The parallel transfer time between images would be only 10 microseconds if calculated in the same way as for the frame transfer device. The obvious disadvantages of this CCD are the limited imaging area and an extremely limited number of consecutive images one can acquire (9).

Modern CCD performance and specifications

With the introduction of cooled slow-scan CCD imaging systems, a new level of performance was reached for numerous imaging and detection methods (10-18). A basic understanding of the underlying principles and technologies is necessary so that one can take full advantage of the technology while avoiding the few drawbacks inherent to these systems.

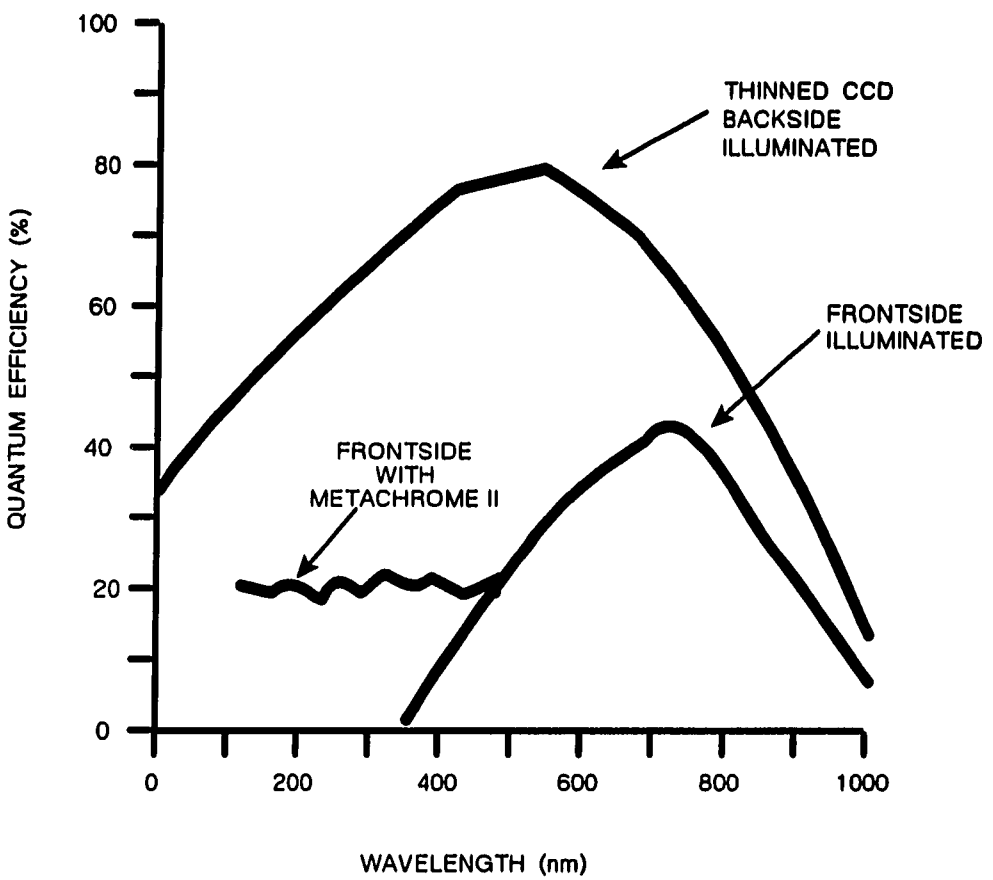
Sensor size and format CCD chips are available in an amazing variety of shapes and sizes as well as pixel

dimensions. The number of detector elements varies from one to 16 million, and overall photoactive areas from 0.2 mm^2 to over 3000 mm^2. The vast majority, however, are similar in size to the chips originally used for video camera applications, which is about 13 x 17 mm. Typical pixels are 20 x 20 μm (19). Such format specifications are important to consider when designing the imaging optics or evaluating sensor response.

Spectral response Typical CCDs without modification have appreciable quantum efficiencies from 400 to 950 nm with a maximum of near 50% at 650 nm, as show in figure 5. Many approaches have been taken to improve quantum efficiencies and wavelength range. Light normally enters the CCD through the gates of the parallel register. These gates are made of very thin polysilicon and are essentially transparent at longer wavelengths but become quite opaque at wavelengths shorter than 400 nm (7).

The obvious solution to detection of ultraviolet radiation is to illuminate from the non-gate side, or backside illumination. To maintain the spatial integrity of the image, the charged pairs must be created within 10 μm of the gate structures. Thus, to implement backside illumination, the CCD must be thinned to 10 μm or less. This is achieved by carefully etching away the n-doped silica with acid and typically adds at least $10,000 to the cost of a CCD (20).

Figure 5. Quantum efficiency versus wavelength for frontside and backside illuminated CCDs.


100
80
60
40
20
0
QUANTUM EFFICIENCY (%)
THINNED CCD
BACKSIDE
ILLUMINATED
FRONTSIDE
ILLUMINATED
FRONTSIDE
WITH
METACHROME II
0
200
400
600
800
1000
WAVELENGTH (nm)

Back thinned CCDs have significant response from the soft x-ray to near-IR regions of the spectrum as shown in figure 5, and have the added advantage of higher quantum efficiencies at all wavelengths. An alternate, and much less expensive, method of achieving ultra-violet response is to coat the sensor with a wavelength conversion phosphor (21). Several techniques have been used to create these layers including thin fluorescent plastic films applied directly to the sensor (22-24) and vacuum deposition of thin film organic phosphors (25). Thin fluorescent plastic films do not work below 225 nm due to absorption by the polymer, and they tend to crack and peel with repeated temperature cycling. These films were abandoned early on as a UV enhancement method. Vapor deposited phosphors do not experience the above effects, and appear to show very little degradation or bleaching over the useful life of a CCD exposed to normal light levels. Originally the polycyclic aromatic hydrocarbon (PAH) coronene was used, but a poor match in its absorption spectra and the CCDs response curve resulted in near zero quantum efficiency at 400 nm. Around the time cooled CCD systems were introduced, a similar proprietary PAH was developed that eliminated this problem (26). The compound, known as MetaChrome II, is coated to a depth of 4000 angstroms, absorbs below 450 nm, and emits at 550 nm. Since the wavelength conversion occurs extremely close to the gates, no image

resolution is lost due to the random direction of the fluorescence process. The quantum efficiency curve for a MetaChrome II coated CCD is also shown in figure 5.

Full well capacity Although CCDs are integrating devices, there is an upper limit to the amount of signal that can be stored before readout must occur. Since each pixel can be considered to be a potential well where charge is stored during integration, this limit is termed the "full well capacity" or FWC, and is usually measured in electrons. Values for CCDs with normal sized pixels typically range from 10,000 to 500,000 electrons (7). This value is important for signal to noise considerations, as will be demonstrated later. A common mistake made when evaluating various CCDs is failure to compensate for pixel dimensions when comparing FWC performance. If, for example, one is evaluating two similarly constructed CCDs that are both 2 cm square with one being 500 x 500 pixels and the other 1000 x 1000 pixels, one might find that the first has a FWC of 400,000 electrons and the latter only 100,000 electrons, and assume the first to have better performance. It is proposed here that a more accurate measure of FWC performance is the pixel full well capacity divided by the pixel area. Thus, in the above example, both CCDs would show the same performance, as would be expected.

This comparison can be validated by examining a hypothetical experiment. Suppose both CCDs were placed in

identical imaging systems with identical images focused onto each sensor. Each pixel in the 1000 x 1000 pixel CCD would, on the average, only sense one-fourth as much light as its counterpart in the 500 x 500 pixel CCD, and would thus accumulate charge at one-fourth the rate. Thus, each CCD would reach the full well capacity after integrating for exactly the same amount of time. The original FWC values are still important though, especially if binning to be performed.

Binning Binning is a special mode of operation available on most cooled CCD systems that is useful at low light levels, and is controlled by clocking the registers in a slightly different manner than normal. The end result of binning is that the collected charge from more than one pixel is accumulated in the output node before readout occurs. Binning in the dimension parallel to the channels is accomplished by transferring more than one row of pixels into the serial register prior to serial readout. Binning in the other dimension is accomplished by transferring the desired number of pixels into the readout node before readout occurs. The advantages of binning are that readout occurs less often, introducing less read noise, and images are read more quickly since reading the node is time consuming in slow-scan devices. The major disadvantage of binning is that spatial resolution is lost. Binning is not applicable to high light level imaging since the total charge from the group of pixels being

binned cannot exceed the pixel full well capacity (7).

Noise considerations in CCDs Noise is defined as any undesirable signal components that interfere with the signal being measured, with the ultimate figure of merit being the ratio of the intensities of the signal to that of the noise, or signal to noise ratio. To ultimately discuss this ratio, a full understanding of the three sources of noise in the modern cooled CCDs is necessary.

Photonic noise The first type of noise inherent to these types of imaging systems is photonic noise, or shot noise. Although it is not strictly due to the sensor itself, it is significant because of the limited full well capacity. Shot noise is due to the fact that light consists of discreet packets of energy or photons. For a steady light source, the total number of photons emitted over any set of identical time periods varies according to a poisson distribution (27). In a similar manner, a CCD integrating signal will see the same distribution of intensities from pixel to pixel for a uniform light field. According to Poisson statistics, the magnitude of the shot noise is equal to the square root of the number of photons detected, and is thus quite a significant fraction of the measured signal with the relatively small average number of photon counts.

Dark current and dark current noise Dark current in CCDs is defined as the spontaneous formation of charged

couples, and is a direct result of the operating temperature of the chip itself and is usually reported as electron per unit time. In normal video equipment not employing cooling, this dark current is quite high and results in three very undesirable situations (28, 29).

First, integration time is severely restricted. The dark current in typical non-cooled CCDs would reach the full well capacity in less than one second, thus very little of the well is available for signal as exposures approach this limit. Secondly, slow scan readout is not possible since it typically requires one second to read out the CCD and the signal would be lost as the dark current exceeded the FWC. Lastly, having a dark current implies that there is a noise associated with it. Since the dark current is the result of randomly occurring events, the pixel to pixel noise associated with it follows the principles of Shottky (or shot) noise (27), and is equal the square root of the average number of charged pairs created per pixel during the integration and readout time. Dark current noise is the dominant factor in low light level imaging with traditional video equipment. Dark current is reduced via cooling by a factor of two for each 6°C decrease in temperature and is reduced to approximately one electron per pixel per second at -50°C (30). Imaging system manufacturers have achieved these temperatures by bonding the CCD to a thermo-electric cooling system and dissipating the

heat via a liquid recirculating unit. The dark current can be reduced further by cooling the chip via direct circulation of liquid nitrogen to the stage to which the CCD is bonded. Operating temperatures for these systems are typically -120°C with dark currents of approximately 0.005 electrons per pixel per second. Dark currents in this level are insignificant in all but the most demanding imaging situations. A fairly recent development that has further reduced dark currents is the introduction of CCDs operating in a multi-pinned phase (or MPP) mode (7). This technology has produced thermo-electrically cooled CCDs with liquid nitrogen CCD performance levels, and liquid nitrogen cooled CCDs with dark currents that cannot be measured. One major drawback of these systems is that as much as 75 percent of the full well capacity must be sacrificed. The noise from the dark current is calculated in the same manner as photonic noise since the detector cannot distinguish between it and incoming radiation.

Pre-amplifier noise and slow-scan readout Pre-amplifier noise is generated during pixel readout by the electronics in the on-chip output amplifier on the serial register. The ultimate source of this noise is due to the limitations of modern electronics and is discussed detail elsewhere (31). The operating parameter that affects the level of this noise the most is the readout rate, and this is commonly referred to as readout noise. The level of this

noise reaches a minimum of 5 to 10 electrons at a readout rate of 50 to 250 kHz depending on the system. Cooled CCDs are thus read at these rates and referred to as slow-scan devices. These slow readout rates result in very long full frame readout rates. In a 1024 x 1024 pixel full frame CCD operating at 200 KHz, full readout takes five seconds, which may limit the applicability of full frame CCDs.

Signal to noise in CCDs All signals and noises in CCDs are normally quoted in electrons, since that is ultimately the unit of detection. Assuming that the above mentioned noise sources are the only ones, or that they greatly outweigh any systematic contributions from the source of the signal, the total noise (NT) is defined as follows:

$$NT = NP + (NES + NED)^{1/2} \quad (1)$$

where NP equals the pre-amplification noise, NES is the number of electrons from signal and NED is the number of electrons from the dark current. The final signal to noise is therefore:

$$SNR = NES/NT \quad (2)$$

From a quick examination of these equations it is clear that there are three distinct regions where one noise source will dominate.

The pre-amplifier noise, which is typically ten electrons, will dominate when the signal is less than 100 electrons, and dark current is negligible. Such cases are

unusual, and better suited to other detectors such as microchannel plate intensified CCDs.

The integrated dark current will limit the SNR whenever it is equal to or greater than the signal itself. Such cases are common in thermo-electrically cooled CCD systems with incoming light fluxes at or below the dark current. These applications are more suited for liquid nitrogen cooled CCDs or the newer MPP mode thermo-electric cooled CCDs.

The most common case in CCD imaging is when the SNR is limited by shot noise, and this is most clearly demonstrated in the case of imaging absorption type signals such as those discussed later in Paper 2. For a single frame with a signal near the FWC, with insignificant dark current, the SNR will be approximately the square root of the FWC. For a typical CCD with a FWC of 250,000, the SNR will be around 500.

The SNR can be enhanced by acquiring multiple frames and averaging them. This will give nearly the same results as integrating for the total of the individual integrating periods since read noise is fairly insignificant, and dark current is cumulative. The SNR will increase as the square root of the number of frames read, with the only disadvantage being the excessive time spent reading frames, unless one has a frame transfer device. This frame averaging technique does not work well with non-cooled fast-scan CCD-based video imaging systems since read noise is the major source of noise

in this case (7).

Linear response and pixel calibration Since CCDs are solid-state devices with no gain mechanism, the response of any given pixel is quite linear throughout its full well capacity (32). Due to inconsistencies in individual pixel size and gate structures, each pixel in the sensor displays its own unique linear response. Depending on the grade of CCD chosen, this nonuniform pixel response can vary from 1% to nearly 10%. This would obviously be a major limitation in many applications, particularly those with high backgrounds. In these cases, pixel calibration is necessary, and is often referred to as flat-fielding. Traditional flat-fielding techniques are based on producing a calibration image by acquiring an evenly illuminated image at a level near the full well capacity, and normalized by dividing each pixel intensity by the average image intensity (7). To eliminate any contribution due to the dark current, another image is acquired over a similar time period and subtracted from the calibration image prior to normalization. Subsequent images are then multiplied by this calibration image to compensate for the varying pixel response.

A more sophisticated and powerful method of flat-fielding is to include non-detector non-uniformities in this calibration image. This is accomplished by designing the total system such that an image can be acquired that contains

all of the elements of the final image except the desired signal. Thus, the calibration image will correct for such things as non-uniform illumination of the scene, non-uniform imaging optics, and any systematic noise from the image scene, as well as non-uniform pixel response. This technique was utilized in Papers 2 and 3 with gel electrophoresis imaging. The effect of flat-fielding is easily demonstrated in figures 6 to 8. Figure 6 is the CCD image of a gel prior to electrophoresis. In the image, the starting wells are located along the left hand side. The direction of electrophoresis will be across the image. Many non-uniformities can be seen including uneven illumination an distortions due to the optical qualities of the gel itself. Figure 7 is the same gel after electrophoresis. The dark bands represent the signal from the separated DNA in this case. The same non-uniformities can be seen on this image as the last. Figure 8 is the flat fielded image created from the first two. The non-uniformities have been removed by calibration resulting in a significant enhancement of the features of interest.

Conclusion

The CCD has been called the ideal detector (33), and the specifications quoted by the manufactures and achieved in the lab support this to some extent. CCDs do have limitations, and care must be used during experimental design to exploit

Figure 6. The CCD image of a gel prior to electrophoresis.

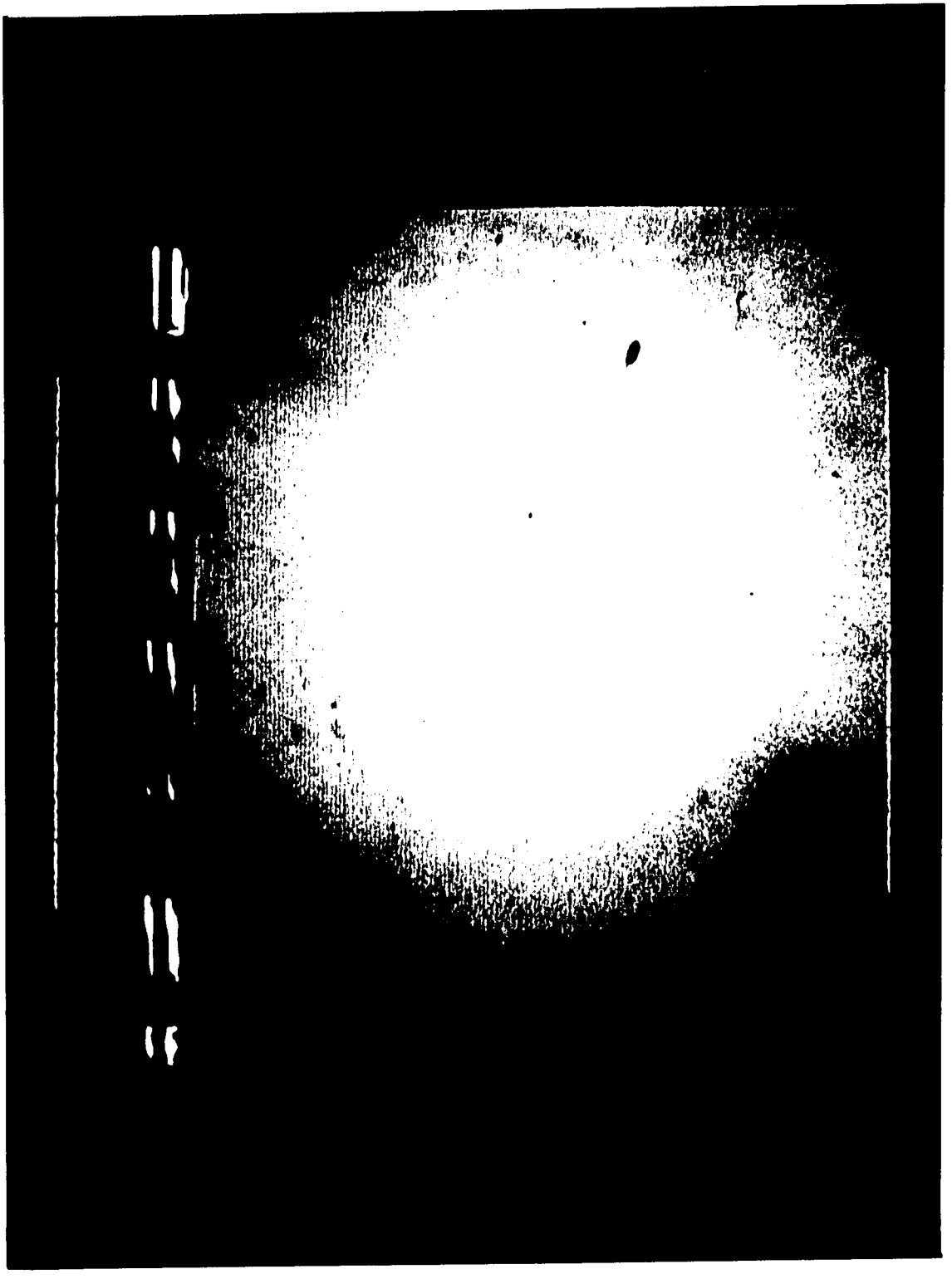


Figure 7. Image of gel in figure 6 after electrophoresis.

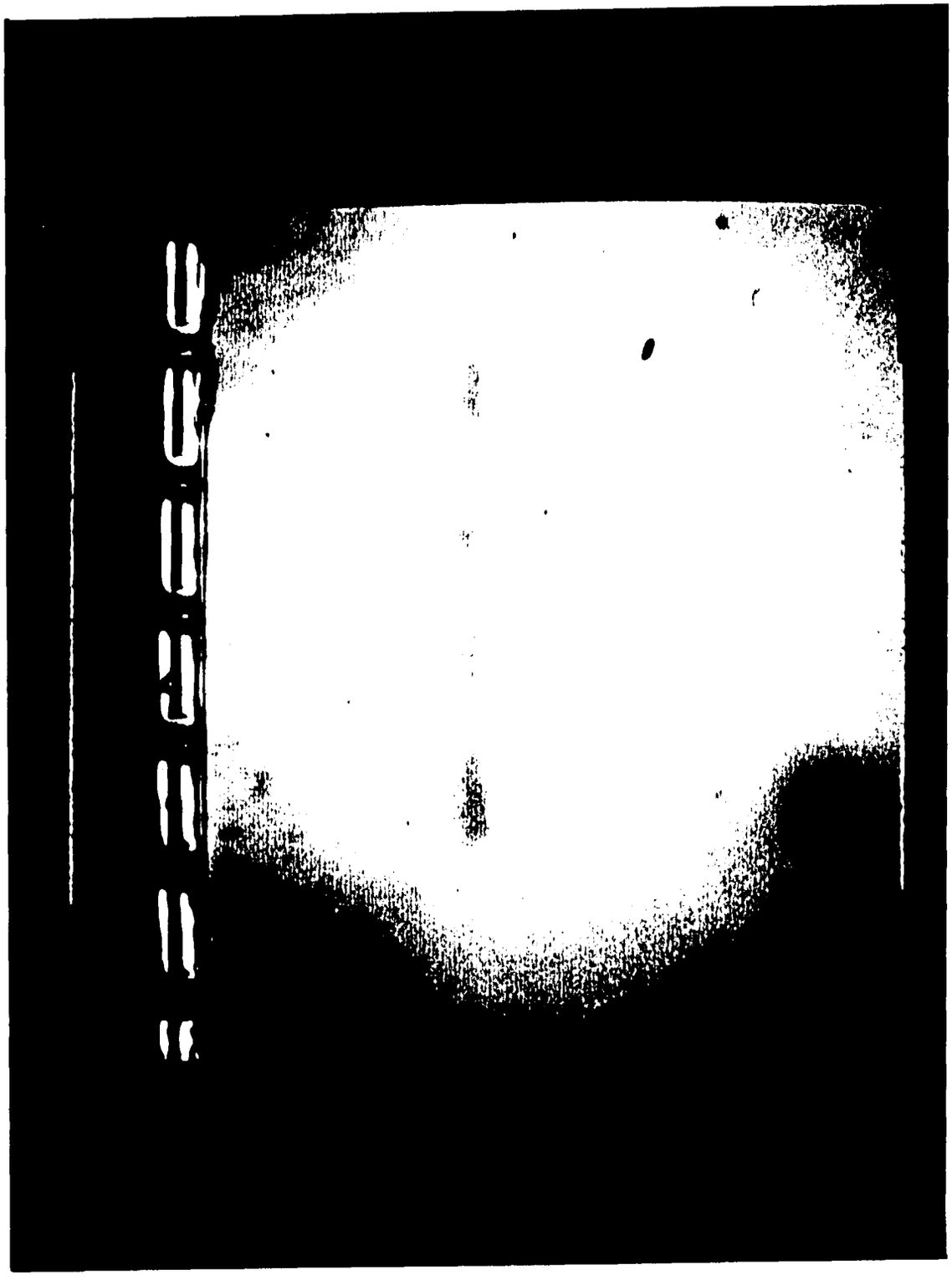
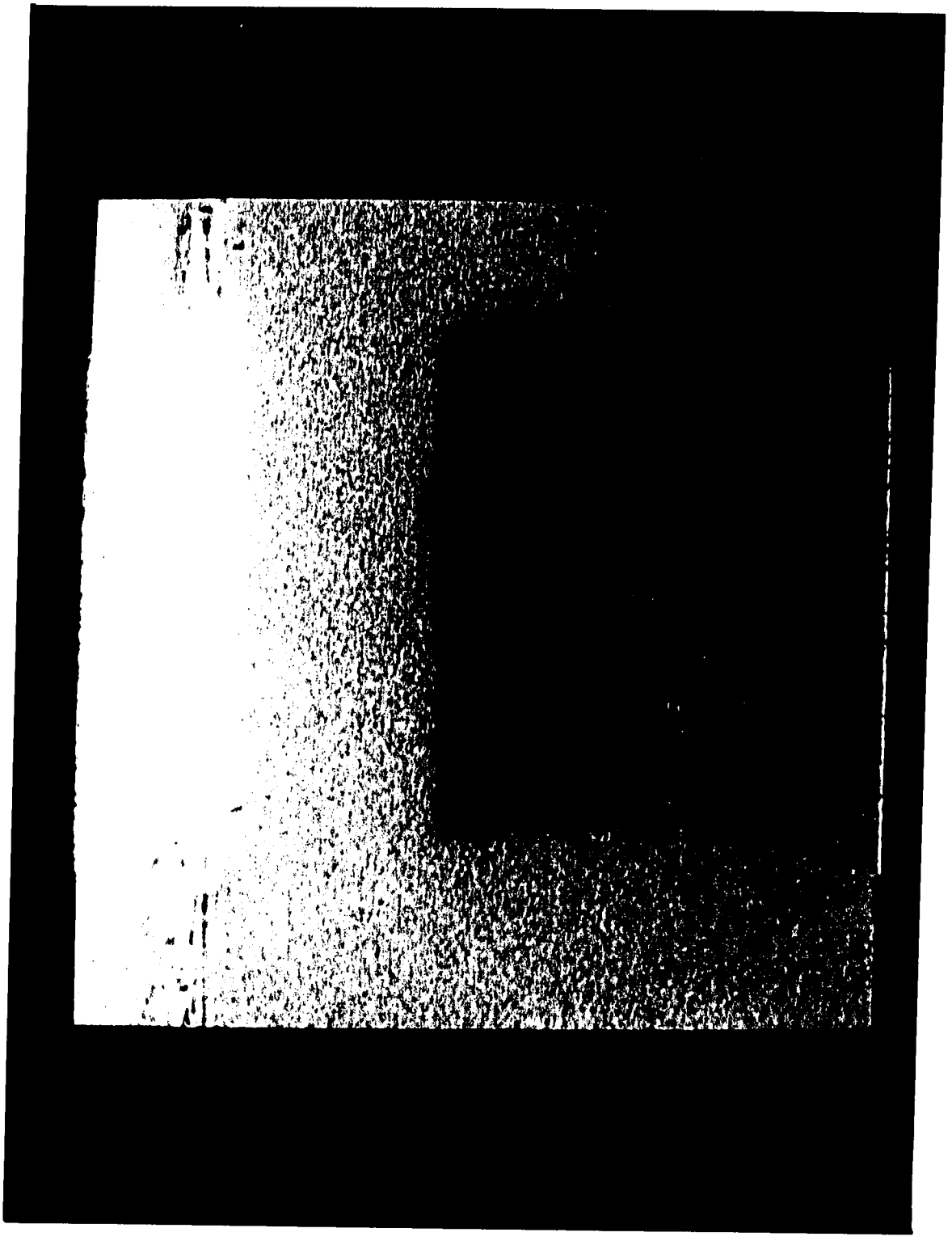


Figure 8. Flat-fielded image created by ratioing figures 6 and 7.



the nearly ideal aspects of performance, while avoiding some of the serious pitfalls. In the experimental systems we have applied the CCD to, its performance has not been a serious concern, except for one case in which there was a serious loss of signal during the readout period of our full frame CCD due to the slow-scan requirement. Although the CCD imaging system was adequate for this system, it was far from ideal.

Introduction to Personal Computers in the Lab

Although computers have been utilized by some analytical chemists for decades, it was the introduction and explosive growth of the personal computer (PC) in the 1980s that brought powerful computational techniques to the typical laboratory. By the late 1980s, the PC had become an indispensable tool in many disciplines. Today, PCs are involved in nearly every aspect of analytical research projects ranging from routine tasks such as data handling and automation of experimentation to novel approaches to analytical problems.

Data handling

The most obvious use of a computer is for data handling, which includes manipulation, visualization, and storage. Data entered into a computer can be manipulated via many methods. The most basic method is through simple computer

programs written for very specific tasks. These programs typically input the data set, perform the desired calculations based on the users input, and output the resulting data to a specified data file. A relatively sophisticated example of this type of post-run processing is demonstrated in paper 4.

Input output hardware

Perhaps the most dramatic development in PCs is the continuing emergence of easy to use yet powerful data input/output systems, and much of the work in this dissertation relied upon the abilities of these systems. These devices allow one to control many aspects of an experiment in an intelligent manner as well as collect the data generated, with the ultimate application being the intelligent real-time control of experiments based on the data being acquired from the experimental system. Their abilities allow for construction of expert systems without extensive knowledge of electronics or computer hardware.

System features

There are many different hardware systems available including several industry standards such as the GPIB IEEE-488. These boards are typically used for very specific interfacing applications. Although the performance in these systems is quite high, they are not very user friendly or

versatile and will not be discussed here. Truly versatile high performance has only been developed in systems that are not yet standardized, and vary greatly depending on manufacturer. The systems generally consist of an interface board that plugs into the parallel bus of the PC. Attached to this is usually an interface board where the incoming and outgoing signal wires are connected. The specifications below are typical of the systems manufactured by Data Translation Inc. and Keithley Instruments, but apply to most other manufacturers as well. The typical interface board is capable of four forms of communication with the computer or connected devices. These are analog input and output, and digital input and output. Analog input and output are of the most interest to analytical applications since the signals are generated by natural processes which are inherently analog. These boards allow for the interfacing of such signals (usually in a voltage format) to the powerful digital environment of today's laboratory computers. No analog signal can be precisely represented by a set of digital values, since the number of possible digital values is limited by the number of bits while analog values can have any value. This problem is known as digital noise and is manifested in unit step features on the resulting digital signal. This can be minimized by increasing the number of bits and assuring that the analog and digital ranges coincide. Typical high performance analog to digital

converters operate with 12 to 24 bits. The maximum digitizing frequency ranges from 60 Hz to a few MHz, with the upper range being limited by the clock speed of the host computer's bus. Boards typically have between 2 and sixteen analog channels that can be controlled at the same time. The voltage range that the digital values represent is often selectable by the operator. This assures that the analog signal being digitized utilizes the maximum number of bits. For example, it would be inappropriate to choose a digital range of 0 to 8 Volts for an analog signal that only varies from 0 to 1 Volts. In a 16 bit board case, the range of digital values would be the integers from 0 to 65535 ($2^{16} - 1$) but only values from 0 to 6554 would be used. This utilizes only the thirteen lowest bits, with the highest three being wasted. This ultimately results in the voltage steps between adjacent digital values being 8 times larger (2^3) than in the ideal case.

Digital input and output are simply specialized cases of coordinated analog input or output from several channels. The processes are simplified by the fact that only two analog voltage signals are required; low (0 volts) and high (+5 volts). Digital input and output typically occur on groups of eight channels resulting in 8-bit transfer of digital data. By coordinating more than one eight channel group, higher bit digital communication is possible. This digital communication is necessary to allow one to interface with the growing array

of analytical instruments and support equipment available with external digital logic control.

Perhaps the most important component of these input/output systems is the software package that creates the programming environments in which these systems are controlled. During the 1980's system manufacturers developed libraries of low level programmed sub-routines that could be easily called from high level computer languages, such as Fortran and Pascal. By the late 1980's, software packages appeared that eliminated the need for high level programming. These packages typically employ some sort of user interface, graphical or not, that allows the user to enter certain values into specified fields to cause the execution of an input or output operation. These values can be stored in a method file to execute sequences of operations. Using either high level programming or a user interface, it became quite easy for analytical chemists to begin to develop sophisticated data acquisition and control methods.

Conclusion

The modern PC with D/A and A/D board is perhaps the most versatile and flexible instrument available to the analytical researcher. The algorithm described in Paper 5 could easily be implemented using one of these systems, and demonstrates the type of control these systems are capable of exerting.

With the increasing speed and power of PCs and the sophistication of modern algorithms, the computer system can actually make real time decisions about control of experimental parameters based on data already collected.

PAPER 1

LASER-BASED INDIRECT FLUOROMETRIC DETECTION AND QUANTIFICATION IN THIN-LAYER CHROMATOGRAPHY

Reprinted with permission from Analytical Chemistry, 1989, 61, 1931.

INTRODUCTION

Since thin-layer chromatography (TLC) was described by Kirchner and his group (1), many improvements in all aspects of operation have been made, including high quality, high performance, and multimodal TLC plates, accurate and precise spotting techniques, instrumentalized development devices, and sophisticated detectors (2-6). These improvements have transformed TLC into a modern separation technique.

Detection and quantification on TLC plates are important considerations. The simplest detection method is based on the human eye aided by a vast array of selective spray reagents, or by the use of plates impregnated with fluorescent indicators to allow components to be detected by absorption at the excitation or emission wavelength. For quantitative analysis, traditionally the plates were scanned by mechanically driven densitometers (7). This is usually a slow procedure. Recently, various methods have been suggested for improving the detection limit, speed, and scope of application. These include photoacoustic spectrometry (8-10), flame ionization (11), photothermal deflection (12, 13), laser fluorometry (14, 15), mass spectrometry (16-19), and computer-aided video densitometry (5, 20, 21). All these detection methods respond to a specific property of the analyte. If the analyte does not possess that specific property, difficulties

in detection will arise. There is thus a need for a sensitive universal detection scheme for TLC.

Recently, several indirect detection methods for liquid chromatography (LC) have been demonstrated (22-26). The detection mechanisms have been clearly identified (27, 28). Briefly, the detector responds to some physical property of the chromatographic eluent. A constant background signal is then maintained when the analytes are absent. When the analytes elute, displacement of the eluent causes a change in the background signal. So, analytes can be detected indirectly. Following the same principles, the indirect fluorometric detection of anions (29), cations (30), and nonelectrolytes (31) in TLC has also been achieved. However, the results so far have not been optimized. The visual observation of the TLC plates (which are illuminated by a UV lamp) is obviously not sensitive compared to instrumental detection methods. Also, visual observation cannot be reliably used for quantitative measurements. In this paper, we describe a laser-based indirect fluorometric detection scheme for TLC with high sensitivity (picograms) and wide linear dynamic range (more than 2 orders). Reliable scanning is easily performed because the TLC plates and detector are fixed while the excitation laser beam is moved. In this system, a microcomputer controls the X-Y scanning and at the same time collects the data. This allows a total data

acquisition time of 35 s for a data array of 256 × 64. There is also an improvement in the signal-to-noise ratio (S/N) resulting from data averaging.

EXPERIMENTAL SECTION

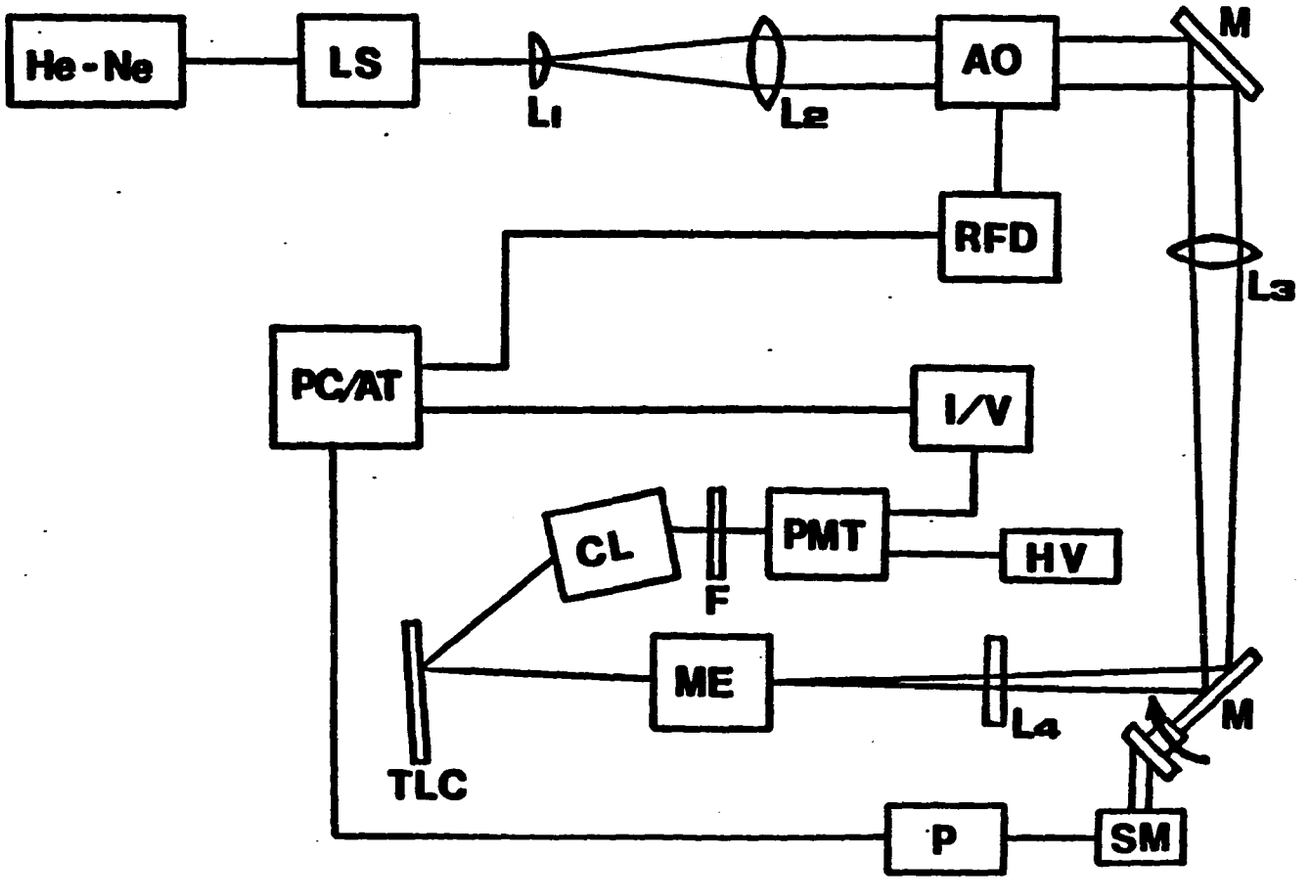
Apparatus

The experimental setup for two-dimensional TLC scanning in laser-based indirect fluorometric detection is shown in Figure 1. A He-Ne laser (Uniphase, Manteca, CA) was used as an excitation light source at 633 nm at a power of 8 mW. In order to maintain a constant laser power, a laser power stabilizer (Cambridge Research and Instrumentation, Cambridge, MA, LA 100) was used. An acoustooptic modulator (Andersen Laboratories, Inc., Bloomfield, CT), which was controlled by a radio frequency (rf) driver, was used to deflect the laser beam. The change in the frequency of rf input to the acoustooptic device causes deflection of the first-order laser beam (32,33).

The acoustooptic modulator and a rotating mirror combine to scan the laser beam in the horizontal and vertical directions, respectively. To obtain the optimum spatial resolution, we introduced a cylindrical beam expander with L_1 and L_2 and used long focal length lenses L_3 and L_4. The focused image onto a microscope eyepiece (12× magnification) was enlarged to an area 40 × 50 mm (horizontal × vertical) with a laser spot size of about 1.5 mm on the TLC plate.

The fluorescence signal was collected by a camera lens

Figure 1. Schematic diagram of laser-based indirect fluorometric detector for thin-layer chromatography. NeNe = Helium-Neon laser, LS = laser stabilizer, AO = acousto-optic modulator, M = mirror, RFD = radio frequency driver, PC/AT = personal computer/AT compatible, I/V = current-voltage converter, PMT = photomultiplier tube, HB = high voltage, CL = camera lens, ME = microscope eyepiece, SM = stepping motor, P = power for stepping motor, F = cut-off filter, TLC = thin-layer chromatographic plate, L_1 = 25 mm f.l. cylindrical lens, L_2 = 400 mm f.l. spherical lens, L_3 = 1000 mm f.l. spherical lens, L_4 = 250 mm f.l. cylindrical lens.


He-Ne
LS
L1
L2
AO
M
RFD
L3
PC/AT
I/V
CL
F
PMT
HV
ME
L4
M
TLC
P
SM

(28-105 mm, *f*2.8-*f*3.8, Vivitar Corp., Santa Monica, CA), passed through a cutoff filter (to remove scattered 633 nm light), and directed into a R928 photomultiplier tube (PMT) (Hamamatsu, Middlesex, NJ) operated at 850 V. The output of the PMT was converted into voltage via a resistor and was fed into a data acquisition system consisting of an analog to digital I/O interface (Data Translation, Marlborough, MA, DT 2827) and a microcomputer (IBM, Boca Raton, FL, PC/AT). This system was also used to control the rf output and the stepping motor in order to synchronously scan the laser beam. After each horizontal measurement, the laser beam was moved vertically to the next horizontal line by the stepping motor. For our experiment, a data array of 256 × 64 points (horizontal × vertical) was acquired from the entire scan area. Each point was averaged 256 times. Data acquisition took 35s. Six additional minutes was needed for data averaging and reduction. The final data array was 64 × 64 points. The data were then normalized to minimize the fluctuations in the background signal.

Chromatography

A K_6 silica gel plate (Whatman, Clifton, NJ) was pretreated with 2 × 10^{-6}M Nile Blue A perchlorate in methanol for 20 min. The plate was then dried with a heat gun. A 0.1-

1.0μL methanol solution of test sample containing crocein orange G and orange G (Aldrich, Milwaukee, WI) was spotted by a microsyringe (Hamilton, Reno, NV). The thin-layer plate was developed to a distance of about 40 mm from the origin in a developing solution containing 2-butanol/acetone/water (75:15:10 (v/v)). After drying, the TLC plate was placed in the apparatus for measurements. For comparison with the detection limit obtained visually, the TLC plate was put under a sodium lamp.

RESULTS AND DISCUSSION

In indirect fluorometric detection, the limit of detection can be estimated from the following equation (24, 25):

$$C_{lim} = C_F/DR \cdot TR \quad (1)$$

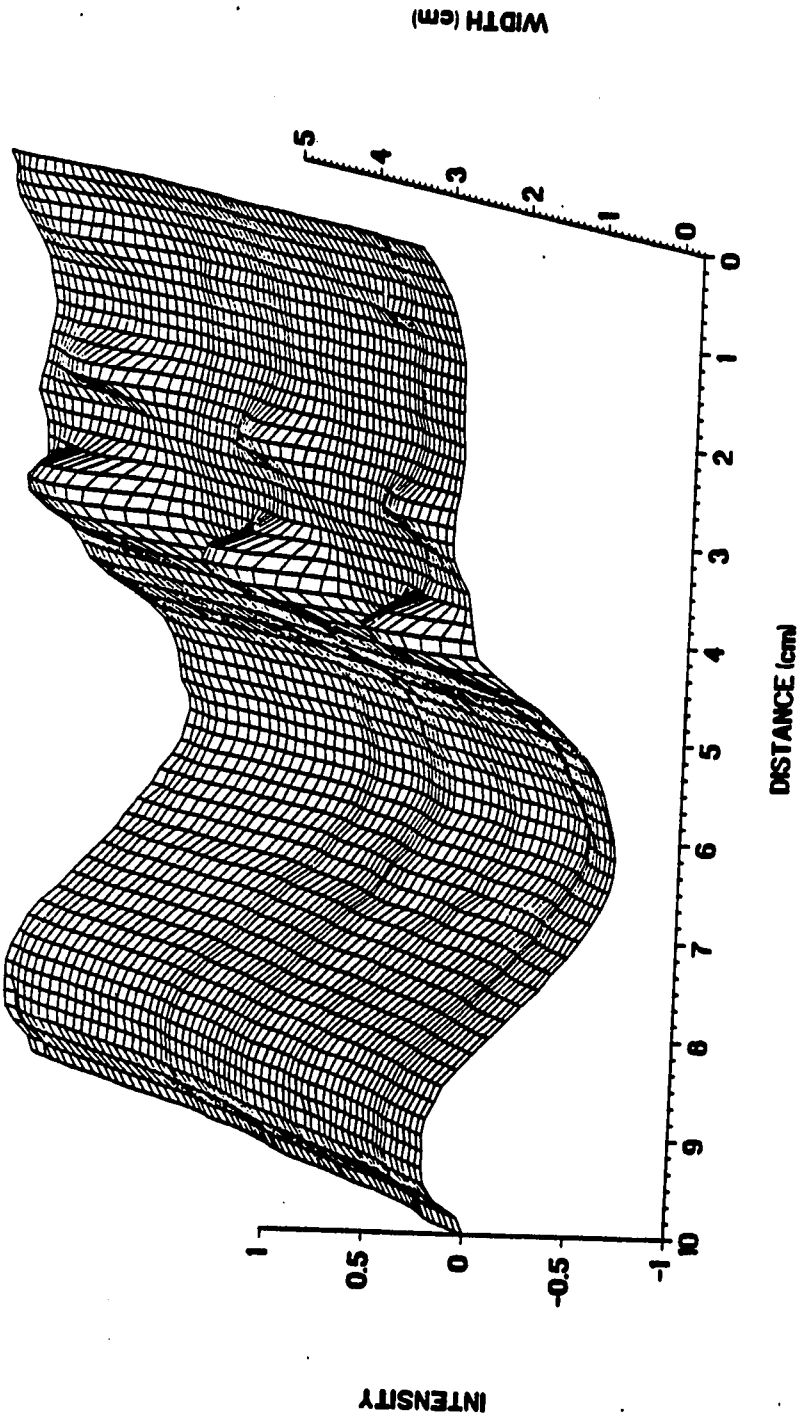
where C_{lim} is the molar concentration of analytes at the detection limit, C_F is the molar concentration of the fluorophor, DR is the dynamic reserve (which is defined as the ratio of the background signal to the noise level), and TR is the transfer ratio (which is defined as the number of fluorescing molecules transferred by one analyte molecule). From eq 1, it is easy to see that if one wants to decrease the detection limit, a more stable background signal (large DR) and a lower concentration of the fluorophor (smaller C_F are needed. Even though a large transfer ratio can improve the detection limit, it is generally fixed by the stationary phase, the developing solvent system, and the analyte. In our previous papers (29, 31), even though we demonstrated indirect fluorometric detection in TLC, the detection limit was not optimum. The human eye has a low contrast, which lowers the DR. In this work, we can easily stabilize the background signal (increase DR) and use lower concentrations of the fluorophor in the pretreating solution (smaller C_F) due to the high sensitivity of phototubes. These conditions help to

improve the detection limit.

In this work, we chose Nile Blue as the fluorophor because it can be excited by a He-Ne laser. This will greatly enhance cost and convenience considerations in an eventual instrument. Since detection is by indirect fluorescence, the choice of the fluorophor is quite arbitrary. The analytes were chosen so that they do not absorb at the He-Ne laser wavelength or the fluorescence wavelength, to emphasize the indirect detection mode. Before the TLC plate is developed, the analyte spots do not show any signal. This confirms that displacement of Nile Blue on the plate is responsible for the indirect fluorometric response. The colored analytes also allow us to compare indirect fluorescence with conventional visualization of the spots on TLC plates.

Figure 2 shows a two-dimensional chromatogram of the indirect fluorometric detection of orange G and crocein orange G. The response is dependent on the analyte because TR is a function of R_f (25).It can be seen from the figure that 80 pg of each analyte can be detected. However, the conditions are not perfect for quantification because of fluctuations in the background fluorescence signal which are caused by variations in the intensity of the laser spot with the angle of deflection and inconsistencies in fluorophor concentration due to the development. To compensate for these fluctuations, the signal can be normalized based on the background levels in a

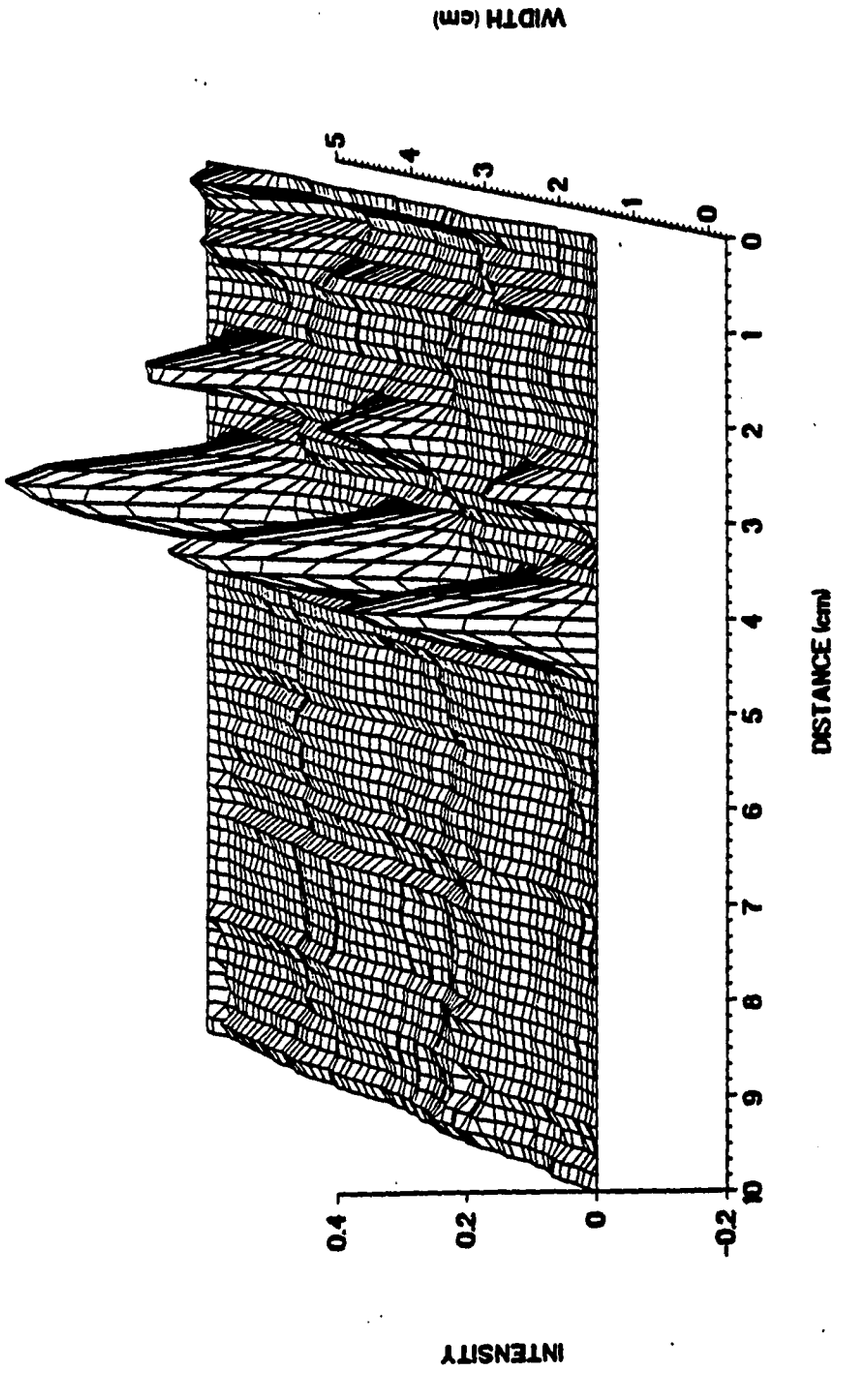
Figure 2. Two-dimensional chromatogram by indirect fluorometric detection of orange G and crocein orange G without data normalization. The left series of peaks are crocein orange G, the series of peaks are orange G. The amount spotted are 80, 120, and 150 pg counting from bottom to top for each sample.


WIDTH (cm)
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
10
DISTANCE (cm)
1
0.5
0
-0.5
-1
INTENSITY

lane in between the sample lanes. The normalized data are shown in Figure 3. The data in Figure 2 and Figure 3 have been inverted to facilitate viewing. Thus, "zero" corresponds to background fluorescence, and "1" would correspond to no fluorescence. Figure 3 shows that the signal-to-noise ratio is readily improved by normalization. The detection limit of crocein orange G is approximately 6 pg. This detection limit is over 100 times lower than when the eye is used to view the same dark spots under a sodium lamp (visual indirect fluorometry). This is also 1000 times lower than those limits reported earlier (31) that were obtained with a near-UV fluorophor, about 1000 times better than using the eye to look at the colored spots in room light (indirect densitometry) without Nile Blue pretreatment, and 100 times better than photoacoustic detection of the same two analytes (8, 9). It should be noted that the detection limit depends on the square root of the number of data points averaged. It is thus possible to reduce the total scan time to 1 s and still achieve detectability in the 40-pg level.

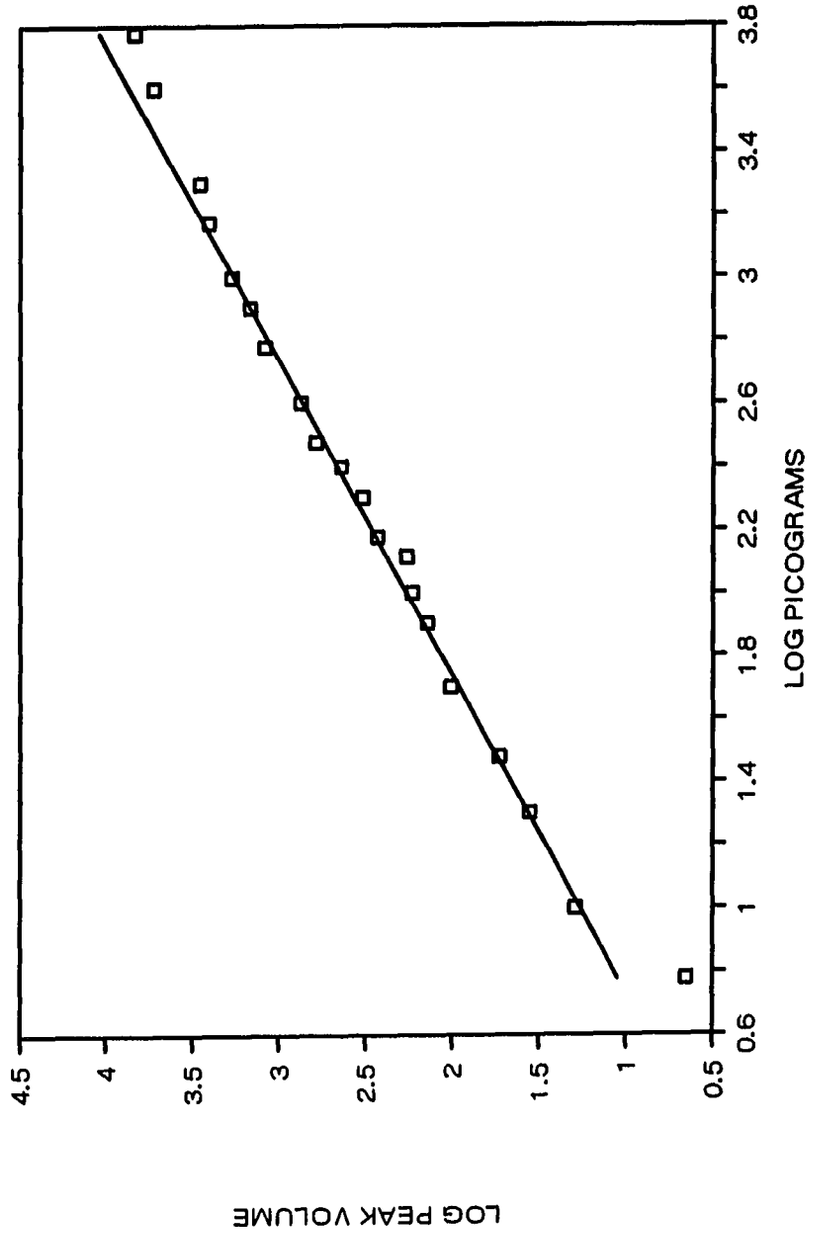
Because of the optical medium, a nonlinear response is usually observed for direct densitometry measurements (34, 35). This can be corrected by applying the Kubelka-Munk function. If the detection mode is based on fluorescence or absorbance, for example, higher concentrations of analytes will lead to nonlinear calibration plots due to self-

Figure 3. Two-dimensional chromatogram of indirect fluorometric detection of orange G and crocein orange G with data normalization. The peak orders and amounts are the same as Figure 2.


WIDTH (cm)
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
10
DISTANCE (cm)
0.4
0.2
0
-0.2
INTENSITY

absorption, self-quenching, and detector saturation. In indirect fluorometric detection, the response is not based on the analyte itself. A linear response is therefore easier to obtain. Figure 4 shows a calibration curve for crocein orange G. A line with unit slope is shown for comparison. Based on the linear regression criteria described by Johnson (36), linearity was observed over 2 orders of magnitude with relative deviations of approximately 10%. Note that the calibration curve was based on the peak volume of the TLC spots in Figure 3. This properly accounts for bank broadening due to development. The peak volume can be estimated by determining the peak area under a single scan through the peak raised to the $^{3}/_{2}$ power. Due mostly to variations in the concentrations of the fluorophor from plate to plate, and partly due to variations in light collection geometry, it was necessary to include a standard spot on each TLC plate to allow normalization among different plates. If the concentration of the analyte is too high (beyond 3.2 in Figure 4), the calibration plot begins to show nonlinearity. This happens because the TR in eq 1 will change at higher concentrations. The signal also falls off at the low concentration end, probably because of difficulties in integration and in background normalization for small signals. However, the linear range achieved is large enough for this approach to be broadly applicable.

Figure 4. Analytical calibration curve for indirect fluorometric detection based on peak volumes. The test sample is crocein orange G. A line with unit slope is drawn for comparison.


LOG PEAK VOLUME
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0.6
1
1.4
1.8
2.2
2.6
3
3.4
3.8
LOG PICOGRAMS

CONCLUSION

We have developed a laser-based indirect fluorometric detection scheme for TLC. The method is universal, is very sensitive, and has a large linear dynamic range. The laser and the optics used should allow the construction of a rugged, fully automated instrument eventually. At present, commercial TLC plates have to be pretreated with the fluorophor to allow indirect detection. It is conceivable that this step can be incorporated into the manufacturing process. The entire process of TLC separations is then substantially simplified; i.e. the tedious and unreliable derivatization steps can be avoided completely.

REFERENCES

1. Kirchner, J. G.; Miller, J. M.; Keller, G. J. Anal. Chem. **1951**, 23, 420-425.
2. Donovan, J.; Gould, M.; Majors, R. E. LC·GC **1987**, 5, 1024-1028.
3. Jaenchen, D. E.; Issaq, H. J. J. Liq. Chromatogr. **1988**, 11, 1941-1965.
4. Jaenchen, D. E. Am. Lab. **1988**, 3, 66-73.
5. Dallas, F. A. A.; Read, H.; Ruane, R. J.; Wilson, I. D. Recent Advances in Thin-Layer Chromatography, Plenum: New York, 1988.
6. Touchstone, J. C. J. Chromatogr. Sci. **1988**, 26, 645-649.
7. Touchstone, J. C.; Sherma, J. Densitometry in Thin-Layer Chromatography: Practice and Applications, Wiley: New York, 1979.
8. Kawazumi, H.; Yeung, E. S. Appl. Spectrosc. **1988**, 42, 1228-1231.
9. Kawazumi, H.; Yeung, E. S. Appl. Spectrosc. **1989**, 43, 249-253.
10. Imaeda, K.; Ohsawa, K.; Uchiyama, K.; Nakamura, S.; Tokieda, T. Anal. Sci. **1986**, 2, 9-13.
11. Ranny, M. Thin-Layer Chromatography with Flame Ionization Detection, Reidel: Boston, 1987.
12. Chen, I. I.; Morris, M.D. Anal. Chem. **1984**, 56, 19-21.

13. Chen, I. I.; Morris, M.D. Anal. Chem. **1984**, 56, 1674-1677.

14. Huff, P.B.; Sepaniak, M. J. Anal. Chem. **1983**, 55, 1992-1994.

15. Belenkii, B. G.; Gankina, E. S. J. Chromatogr. **1986**, 365, 315-320.

16. Chang, T. T.; Andrawes, F.; Kaiser, R. E. (Ed.) Proceedings of the Third International Symposium on Instrumental HPTLC, Institute for Chromatography, Bad Burkheim, W. Germany, 1985, 427-433.

17. Chang, T. T.; Lay, J. O.; Francel, R. J. Anal. Chem. **1984**, 56, 109-111.

18. Ramaley, L.; Vaughn, M.-A.; Jamieson, W. D. Anal. Chem. **1985**, 57, 353-358.

19. Kraft, R.; Otto, A; Zoepfl, H. J.; Etzold, Gerhard Biomed. Environ. Mass Spectrom. **1987**, 14, 1-4.

20. Ford-Holevinski, T. S.; Radin, N. S. Anal. Biochem. **1985**, 150, 359-363.

21. Belchamber, R. M.; Read, H.; Roberts, J. D. M. J. Chromatogr. **1987**, 395, 47-53.

22. Bobbitt, D. R.; Yeung, E. S. Anal. Chem. **1984**, 56, 1577-1581.

23. Mho, S. I.; Yeung, E. S. Anal. Chem. **1985**, 57, 2253-2256.

24. Takeuchi, T.; Yeung, E. S. J. Chromatogr. **1986**, 370, 83-

92.

25. Takeuchi, T.; Yeung, E. S. J. Chromatogr. **1986**, 366, 145-152.

26. Pfeffer, W.; Takeuchi, T.; Yeung, E. S. Chromatographia **1987**, 24, 12-126.

27. Ishii, D.; Takeuchi, T. J. Liq. Chromatogr. **1988**, 11, 1865-1874.

28. Yeung, E. S. Accounts Chem. Res. **1989**, in press.

29. Ma, Y.; Yeung, E. S. Anal. Chem. **1988**, 60, 722-724.

30. Ma, Y.; Yeung, E. S. Mikrochim. Acta. **1988 III**, 327-332.

31. Ma, Y.; Yeung, E. S. J. Chromatogr. **1988**, 455, 382-390.

32. Lekavich, J. Lasers and Applications **1986**, 4, 59-64.

33. Young, M. Optics and Lasers, Springer-Verlag: Berlin, 1986, p. 193.

34. Ebel, S.; Glaser, E. J. High Resolut. Chromatogr. Chromatogr. Commun. **1979**, 2, 36-38.

35. Costanzo, S. J.; Cardone, M. J. J. Liq. Chromatogr. **1984**, 7, 2711-2718.

36. Johnson, D. C. Anal. Chim. Acta. **1988**, 204, 1-5.

PAPER 2

ON-LINE DETECTION OF DNA IN GEL ELECTROPHORESIS BY ULTRAVIOLET ABSORPTION UTILIZING A CHARGE COUPLED DEVICE IMAGING SYSTEM

Reprinted with permission from Analytical Chemistry, **1991**, 63, 746.

INTRODUCTION

Gel electrophoresis is a powerful technique for the separation of high molecular weight biomolecules such as proteins and nucleic acids (1). In particular, agarose gel electrophoresis is widely used in the separation of DNA fragments in restriction analysis (2-4). The method most commonly used for visualizing the DNA bands in the gel matrix is fluorescence staining with ethidium bromide (5), which is very sensitive and usually performed following electrophoresis. However, the staining and destaining of gels may take an hour to complete. Another way to visualize the bands is to include ethidium bromide in the running buffer (6). This procedure is simple and allows the monitoring of the electrophoresis during the run, but incorporating the dye in the running buffer will change the mobility of the DNA and degrade detectability because of the larger background signal. Since ethidium bromide is mutagenic, it is inconvenient to dispose of large amounts of contaminated buffer solution (7). Finally, recovery of DNA fragments from the stained complex is inconvenient and time consuming. For radio-labeled DNA samples, one can use autoradiography. Sensitive detection is possible by using exposure times of hours or days (8). Naturally, this requires strict laboratory practice because of the handling of radio-isotopes.

The basic units of DNA, the nucleotides, all show moderate absorption around 260-270 nm. This implies that they can be detected by absorption at 254 nm (Hg lamp). Two such methods, ultraviolet shadowing (9) and ultraviolet imaging (10) have been used for detecting nucleic acids. In these methods, the DNA bands absorb uv light, leaving an under-exposed area on a fluorescent or photographic background that corresponds to the location of the DNA fragments in a gel. Two major disadvantages of the manner in which these methods were implemented were that real-time monitoring of the electrophoresis was not possible, and detectability was poor. The detection scheme reported here is also based on the uv absorption process, however, the distinct advantages of our method are simplicity, good detectability, reliable quantitation, and the capability of real-time, in situ monitoring.

Charge coupled devices (CCD) and charge transfer devices (CTD) have recently made a substantial impact on spectrochemical measurements (11,12). Commercial versions offer low level light sensitivity that is competitive with photomultiplier tubes and dynamic range equivalent to 16-bits over 500 x 500 pixels. CCD has been used for fluorescence detection in slab gels (13,14), but not for absorption. In this work, the UV light is converted to visible light after interacting with the DNA fragments, so that a standard camera

lens can be used to image the slab gel onto a CCD. The ability to detect low light levels is desirable to speed up data acquisition time, to allow the use of a commercial lamp, and to avoid damage to the DNA fragments during imaging.

EXPERIMENTAL SECTION

Restriction Digest

Lambda DNA and restriction endonuclease Hind III were purchased from BRL (Gaithersburg, MD). Restriction digests were carried out using conditions suggested by the manufacturer except that excess Hind III was used to ensure a complete digestion. Dilution of the Hind III digests were prepared with either the running buffer or digestion buffer. All loading samples contained 5% Ficoll solution.

Gel Electrophoresis

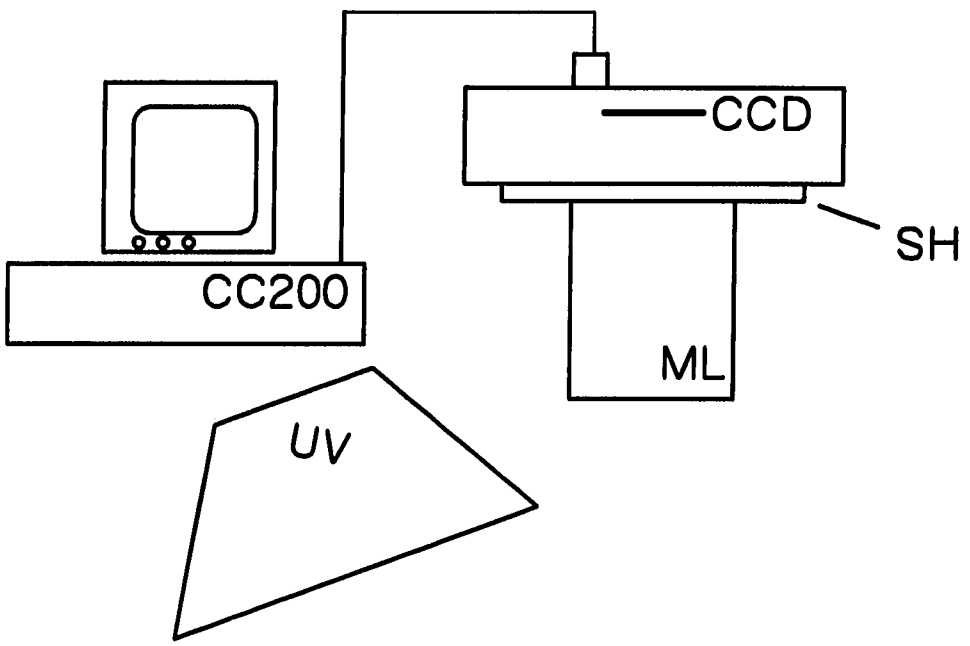
The running buffer contained 20 mM Tris, 1 mM EDTA, and was titrated to pH 8.1 using 85% phosphoric acid; all chemicals were analytical grade and were purchased from Sigma (St. Louis, MO). Ultrapure grade agarose was purchased from Bio-Rad (Richmond, CA). Red fluorescent plexiglass was obtained from Plexiform (Iowa City, IA) and was cut into 70 mm x 100 mm x 3 mm plates. The back side of the plexiglass plate was spray-painted flat black. Agarose gel of 0.67% was cast as an open gel as usual on the fluorescent plate, or, in most cases, as a sandwich gel between the fluorescent plexiglass and a 70 mm x 80 mm x 1.5 mm quartz plate. All gels used were

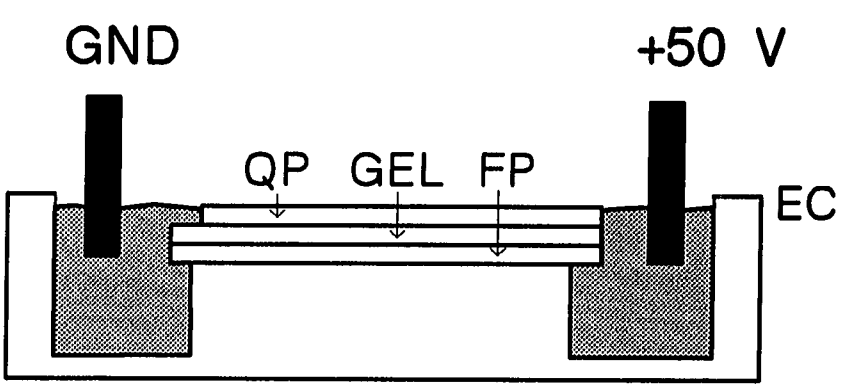
approximately 2 mm in thickness. Sample wells were formed by using either a 1 mm x 2.6 mm or 1 mm x 5.5 mm comb. Typically, 2-4 μL of sample was loaded. Electrophoresis was carried out horizontally in a Mini-Sub cell (Bio-Rad) at a constant voltage of 50 volts. The level of the running buffer was adjusted to touch both ends of the gel only, instead of covering the whole gel surface as in the conventional submerged system. In conventional submerged gel systems, particles may deposit on the surface of the gel while it is being cast, or flow on top of the running buffer covering the gel. This will create artifacts on the detected signal by reflection and scattering. The sandwich gel running system described above will eliminate such problems. In addition, a sandwich gel can prevent the gel from dehydrating. We have also applied this detection scheme to a conventional vertical running gel system. The results were similar to those reported below for the horizontal system. In separate studies, we found that cooling of the gel is not a problem even for applied fields up to 10 V/cm and currents of 100 mA. Additional cooling can be readily implemented by running coolant around the quartz plates.

Detection

Figure 1 shows a schematic of the detection setup. A

Figure 1. Experimental set-up. CCD = charge coupled device, SH = shutter, ML = macro lens, UV = ultraviolet lamp, EC = electrophoresis cell, QP = quartz plate, FP = fluorescent plate, GEL = agarose gel, CC200 = data system for camera.


CCD
SH
CC200
ML
UV


GND
+50 V
QP
GEL
FP
EC

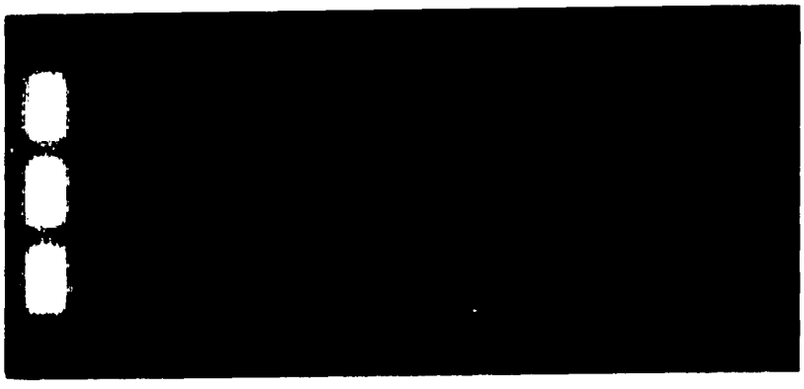
scientific grade cooled CCD digital camera system (Photometrics Ltd., Tucson, AZ) was used to image the gels during electrophoresis. The system was equipped with a thermoelectrically cooled Thomson TH7882 384 x 576 pixel CCD. The camera head, with a 90 mm f/2.5 tele-macro lens (Tamron Co.), was securely mounted 55 cm above the gel. This particular lens has flat-field focusing properties at low aperture settings, thus assuring high light throughput without loss of image sharpness. The gels were illuminated by a 30 Watt hand-held uv lamp (Cole-Parmer) operating at 254 nm, which was securely mounted 45 cm above the gel. Appropriate filters were placed on the lamp and camera to avoid detecting the 254 nm light and other wavelengths emitted by the lamp. With this arrangement, exposure times were typically 5 seconds or less. Since the CCD is an integrating device, it was not necessary to use frame averaging to enhance the images of the relatively low light level scenes. Using a Photometrics 200 series camera, images can be processed and viewed as the gel is being run, and later transferred to a laboratory computer (IBM PC/AT) for storage and further processing. The video images presented in this paper were produced using a SEM-IPS Image Analysis System running IBAS Version 2.0 software (Zeiss Kontron). To enhance contrast, an appropriate linear grey scale was chosen using the IBAS software. The images were printed using a Sony UP-5000 video printer.

RESULTS AND DISCUSSION

Since the reduction in fluorescence signal due to absorption by the DNA bands was small, from 1 to 10 percent in this study, it was necessary to correct the image for variations in individual pixel gain, non-uniformities in illumination, and signal distortion due to the gel itself. This technique is commonly called flat-fielding. This was accomplished by placing the loaded gel in the detection chamber, illuminating it with the UV lamp, acquiring this blank image and storing it digitally before the running voltage was applied. Then, at desired times throughout the electrophoresis, images were acquired and flat-fielded by simply dividing the image by the blank, pixel by pixel, and multiplying the resulting ratios by a constant. These calculations were accomplished on 576 x 384 pixel images in only a few seconds. The resulting image could then be viewed immediately.

Hind III restriction digest of lambda DNA (15) generates six major fragments ranging in size from 2 to 23 kb. A flat-fielded image acquired with the CCD camera of a typical electrophoretic run is shown in Figure 2. The total amount of DNA loaded was 0.25 μg. Where DNA is present in the gel, a portion of the uv light that would have excited the fluorescent background is absorbed, resulting in dark bands.

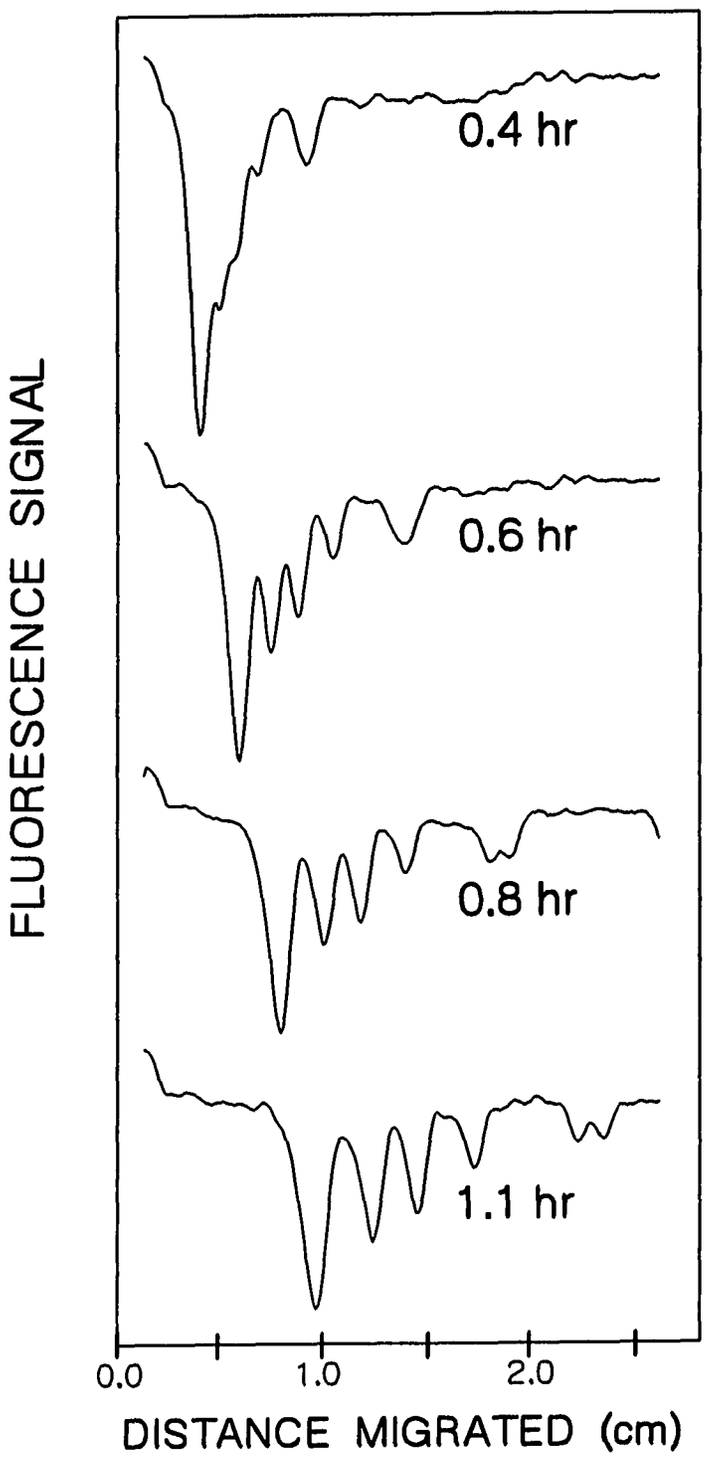
Figure 2. Flat-fielded image of typical electrophoresis run at completion. 0.25 μg DNA per lane. Electrophoresis direction is from left to right.



The loading wells at the top of each lane show up as bright bands. These are the locations where all the DNA resided before electrophoresis. When the fragments move away from the wells, the absorbance becomes zero once again. The flat-fielding process therefore produces bright bands at these locations. Figure 3 is the plot of the fluorescent intensity versus location of bands at different times during electrophoresis. To improve the signal to noise ratio, the columns of pixels that make up the lane, typically 20 to 30, were averaged to one column, and then smoothed with an 11-point Savitzky-Golay smoothing routine (16). The six fragments were reasonably well separated from each other after approximately one hour. Plots of the locations of the individual bands as a function of time gave linear relationships as expected. However, these plots can provide important insights into other separation modes, such as pulsed-field gel electrophoresis (17).

Accurate quantitation of DNA in a gel is not necessary in most restriction analysis. The location of the restriction fragments in a gel rather than their amounts are of most concern. Still, it may be useful to estimate the amount of DNA in a band. Since the molar absorptivities of the DNA fragments are proportional to their molecular weight, and the pathlength and volume of the gel imaged by any given pixel are constant, it can easily be shown, using the Beer-Lambert law,

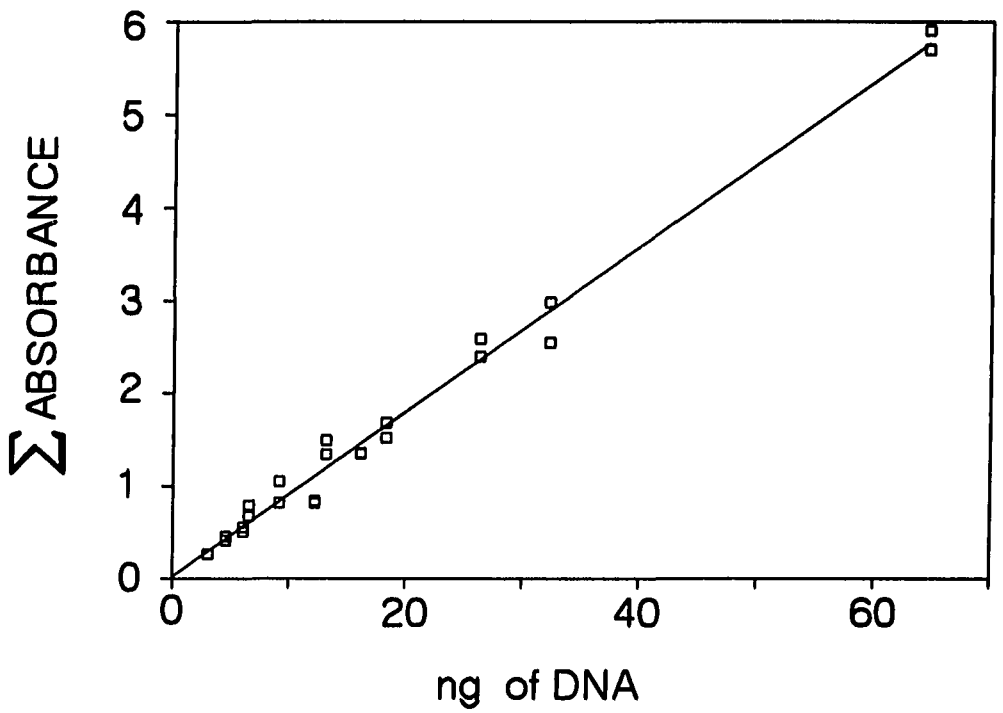
Figure 3. Fluorescent signal versus position in gel at various times during electrophoresis.


FLUORESCENCE SIGNAL
0.4 hr
0.6 hr
0.8 hr
1.1 hr
0.0
1.0
2.0
DISTANCE MIGRATED (cm)

that the absorbance signal from a pixel is proportional to the amount (weight) of DNA in the area imaged by the pixel. Thus, the summation of the absorbance values for the pixels in a band should be proportional to the amount of DNA in that band. Figure 4 is a plot of the summation of absorbance values of the pixels that make up a given band versus the weight of DNA in that band. A total of 23 bands spanning the 4 largest fragments from one gel are included here. This linear relationship is clearly shown by Figure 4. It is interesting to note that Fig. 4 allows an independent determination of the size of the DNA fragment. On a relative basis, since all fragments come from the same DNA, the ratio of the integrated absorbances for each band should be identical to the ratio of the fragment sizes. If the initial molar amount of DNA, the exact pixel size, and gel thickness are determined, one can then use the molar absorptivity of the average base to estimate the number of bases in each band and thus the number of bases per mole of the fragment in question. This of course depends on the assumption that complete digestion is performed. An independent determination of the fragment size as suggested here will help to decipher the complicated migration patterns in e.g. pulsed-field electrophoresis (17). This means that reliable quantitation can add substantially to the information content in gel electrophoresis.

With this absorption detection system, we obtained a

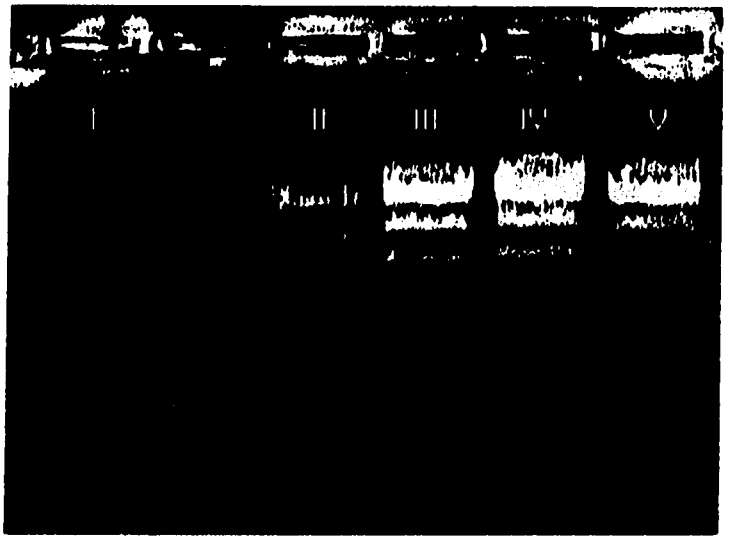
Figure 4. Integrated absorbance signal vs. weight of DNA in each band, and least-square best-fit line.


6
5
4
3
2
1
0
0
20
40
60
ΣABSORBANCE
ng of DNA

detection limit of approximately 5 ng of double-stranded DNA per band. This value is about 5 times higher than typical values using ethidium bromide stain and standard visualization, but it is at least 25 and 65 times lower than those obtained previously using shadowing and imaging respectively (9,10). The improvement is a direct result of the higher contrast possible in a CCD system (vs. photographic plates) and the incorporation of flat-fielding. The amount of DNA used here is typical, so this detection scheme is ready for implementation in molecular biology. If however even better detection limits are desired, we found that frame averaging can be employed. For example, averaging 9 frames will improve the signal-to-noise ratio by 3X since our system is shot-noise limited. This obviously increases the data acquisition time proportionately.

As mentioned above, DNA bands are usually detected with ethidium bromide staining after electrophoresis. Inclusion of ethidium bromide in the running buffer is also common since this will allow electrophoresis to be monitored during the run. However, inclusion of the dye will change the conformation and charge distribution of DNA, resulting in the change of their mobility. Figure 5 is the flat-field image after a run in which unstained (I) and ethidium bromide stained (II to V) samples were applied to the same gel. The total amount of DNA in each lane was 0.4 μg. The

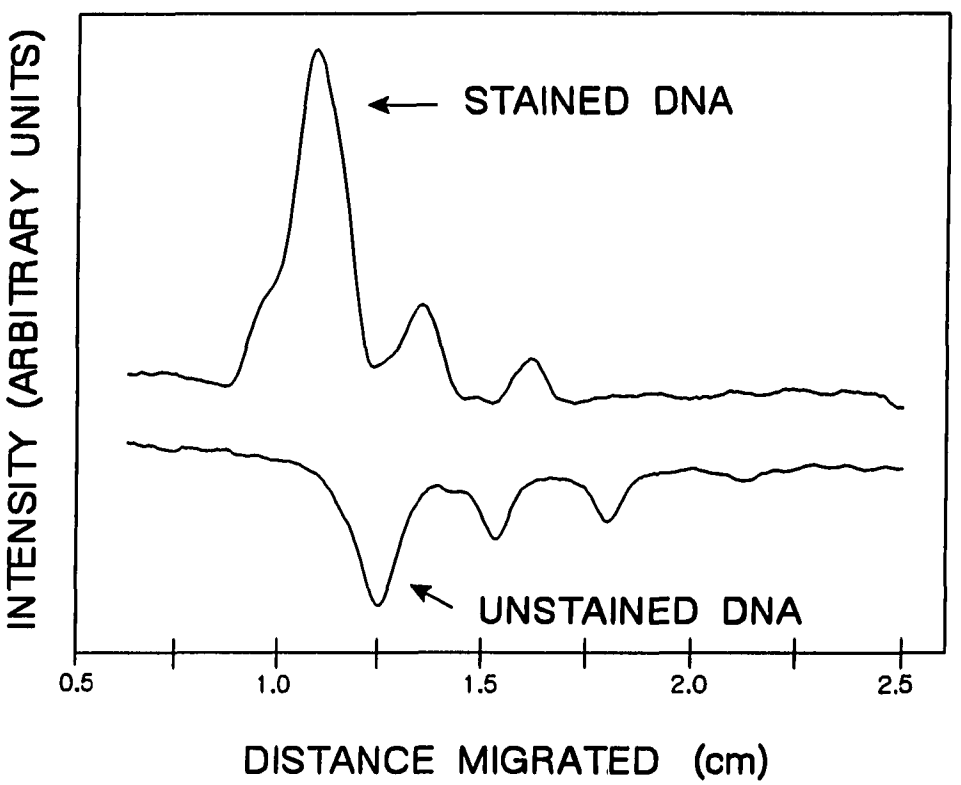
Figure 5. Flat-field image of a gel with both unstained and stained DNA fragments. The amount of ethidium bromide in each sample wsa 0, 2.5, 13, 37, 150 μg/mL in bands I-V. Electrophoresis direction is from top to bottom.


I
II
III
IV
V

concentration of ethidium bromide in the stained samples ranged from 2.5 to 150 μg/mL and was mixed with the DNA before loading (18). The dark bands in lane I, and elsewhere in the image, are the uv absorption bands, while the bright bands corresponded to the fluorescing dye-stained DNA. The fluorescent bands in lane II were fuzzy and weak, which was probably due to a lack of a sufficient amount of dye to fully complex all of the DNA. At high dye concentrations, all of the DNA formed complex with the dye, resulting in intense fluorescent bands. This was the case in lanes III to V. It is interesting to note that as running times became longer, ethidium bromide dissociates from the DNA complex. Thus, after a certain time the fluorescent bands became weaker and finally converted to uv absorption bands. Figure 6 is a plot of intensity versus location of the stained and unstained bands in Figure 5. The positive and negative peaks correspond to the stained and unstained DNA respectively. It is clear that the unstained DNA fragments have a higher mobility than the stained ones. The change in mobility is about 10-15%, as expected (19). This highlights the importance of being able to visualize the bands without chemical treatment.

One of the essential parts of this detection method is the fluorescent background. We used red fluorescent plexiglass throughout the experiment. However, other types of fluorescent materials can also be used. The stronger the

Figure 6. Intensity versus location of bands for stained and unstained DNA on same gel.


INTENSITY (ARBITRARY UNITS)
STAINED DNA
UNSTAINED DNA
0.5
1.0
1.5
2.0
2.5
DISTANCE MIGRATED (cm)

material fluoresces, the shorter will be the exposure time for acquisition of the image. For example, when we used a quartz plate with orange fluorescent paint sprayed on it as the fluorescent background, the required exposure time was only 0.5 second. Although it may not be as useful, we obtained similar results by using a running buffer containing a non-interacting fluorescent dye. This technique has the disadvantages of reducing the theoretical percentage decrease in fluorescence by a factor of two, and introducing unnecessary material into the gel.

CONCLUSION

The CCD camera system proved to be a convenient means of acquiring the desired images. Because of the integrating ability and sensitivity of the camera system, images could be acquired quickly and easily. The CCD's low noise characteristics coupled with the flat fielding technique allowed for low detection limits. Compared to the previous shadowing and imaging methods, the detection scheme reported here has better detectability and allows electrophoresis to be monitored during the run. The latter is especially useful in cases where long running times are required, such as pulsed-field electrophoresis for the separation of chromosomal DNA (17). For example, utilizing on-line monitoring, one could change operating conditions during a run, such as switching times, to achieve the desired separation. Since this detection method does not use ethidium bromide, no hazardous waste disposal is necessary. Furthermore, recovery of native DNA from a gel would be simpler since no DNA-dye complex would be formed. This would be especially valuable in preparative scale applications, or for collecting the fragments for further mapping or sequencing. Quantitative results are obtained here, which can be used as an independent measure of the fragment size. Finally, the absorption method is useful for the detection of small DNA fragments, such as oligomers,

which cannot be stained with ethidium bromide.

REFERENCES

1. Andrews, A. T. Electrophoresis: Theory, Techniques, and Biochemical and Clinical Applications, Oxford University Press: New York, 1988.
2. Johnson, P. H.; Grossman, L. I. Biochemistry **1977**, 16, 4217-4225.
3. Harris-Warrick, R. M.; Elkana, Y.; Ehrlich, S. D.; Lederberg, J. Proc. Nat. Acad. Sci. USA **1975**, 72, 2207-2211.
4. Polsky, F.; Edgell, M. H.; Siedman, J. G.; Leder, P. Anal. Biochem. **1978**, 87, 397-410.
5. Aaij, C.; Borst, P. Biochim. Biophys. Acta **1972**, 269, 192-200.
6. Sharp, P. A.; Sugden, B.; Sambrook, J. Biochemistry **1973**, 12, 3055-3063.
7. Brunk, C. F.; Simpson, L. Anal. Biochem. **1977**, 82, 455-462.
8. O'Farrell, P. H. J. Biol. Chem. **1975**, 250, 4007-4021.
9. Hassur, S. M.; Whitlock, H. W. Anal. Biochem. **1974**, 59, 162-164.
10. Clarke, P.; Lin, H.; Wilcox, G. Anal. Biochem. **1982**, 124, 88-91.
11. Sweedler, J. V.; Bilhorn, R. B.; Epperson, P. M.; Sims, G. R.; Denton, M. B. Anal. Chem. **1988**, 60, 282A-291A.

12. Epperson, P. M.; Sweedler, J. V.; Bilhorn, R. B.; Sims, G. R.; Denton, M. B. Anal. Chem. **1988**, 60, 327A-335A.

13. Sutherland, J. C.; Lin, B.; Monteleone, D. C.; Mugavero, J.; Sutherland, B. M.; Trunk, J. Anal. Biochem. **1987**, 163, 446-457.

14. Jackson, P.; Urwin, V. E.; Mackey, C. D. Electrophoresis **1988**, 9(7), 330-339.

15. Murray, K.; Murray, N. E. J. Mol. Biol. **1975**, 98, 551-564.

16. Savitzky, A.; Golay, J. E. Anal. Chem. **1964**, 36, 1627-1639.

17. Schwartz, D. C.; Cantor, C. R. Cell **1984**, 37, 67-75.

18. Angermuller, S. A.; Sayavedra-Soto, L. A. Biotechniques **1990**, 8, 36-37.

19. Fritsch, E. F.; Maniatis, T. Molecular Cloning: A Laboratory Manual, 2nd Ed., page 6.15. Cold Spring Harbor Laboratory Press: New York, 1989.

PAPER 3

ON-LINE DETECTION OF PROTEINS IN GEL ELECTROPHORESIS BY ULTRAVIOLET ABSORPTION AND BY NATIVE FLUORESCENCE UTILIZING A CHARGE COUPLED DEVICE IMAGING SYSTEM

Reprinted with permission from Analytical Chemistry, in press.

INTRODUCTION

Slab gel electrophoresis is the most common technique for the separation of high molecular weight biomolecules such as proteins (1). Acrylamide gels, as described by Laemmli (2), are generally the matrices of choice for the separation of SDS-denatured proteins via electrophoresis. Agarose gels, similar to those used for nucleic acids, are also useful for the separation of proteins but have not been widely applied (3). Agarose gels are advantageous for many reasons including: simplicity of gel casting, easy sample recovery, and nontoxicity to both the experimenter and the proteins. In the past, agarose was not used because of its poor resolving power at molecular weights below 40,000 Daltons. New agarose gel systems are available that will resolve proteins ranging from 20,000 to 200,000 Daltons with or without SDS denaturing (4). In this study, agarose gel was chosen for its optical qualities and ability to be cast in an open system that can be imaged as the experiment is running.

The method most commonly used for visualization of the protein bands in the gel matrix is staining with coomassie brilliant blue. This process requires several hours and can only be done post-run. It is desirable to be able to monitor the electrophoresis during the run. There is one system available capable of on-line detection that is based on Schlieren optics. That system works quite well for acrylamide

gels (5). It is the goal of this paper to demonstrate two additional on-line detection schemes and compare them to traditional coomassie staining.

Of the twenty common amino acids that are the basic units of proteins, tryptophan, phenylalanine, and tyrosine display significant absorption in the ultraviolet region between 250 and 300 nm (6). This property is the basis of detection of protein in most capillary electrophoresis experiments (7). Detection of DNA via absorption of 254 nm ultraviolet radiation has been demonstrated in agarose slab gels in a method called ultraviolet shadowing (8), which we recently improved upon by utilizing a modern charge-coupled device (CCD) imaging system (9). This shadowing system should also be capable of on-line detection of proteins in their native form in an agarose matrix, as we will attempt to demonstrate here.

Proteins that contain tryptophan or tyrosine groups are known to fluoresce between 300 and 400 nm when excited with ultraviolet radiation within the absorption band of the three absorbing amino acids (10). This native fluorescence has been exploited in capillary electrophoresis detection (11,12), but never in slab gels. This native fluorescence will be the basis of the other on-line visualization technique presented here.

EXPERIMENTAL SECTION

Gel Electrophoresis

The running buffer contained only 5 mM sodium phosphate (dibasic) titrated to pH 10.2 using 1 M sodium hydroxide. Both chemicals were analytical grade and were purchased from Sigma (St. Louis, MO). Ultrapure grade agarose was purchased from Bio-Rad (Richmond, CA). The protein standards used, conalbumin type V and chicken egg albumin grade V, were also purchased from Sigma. Both proteins were prepared in solution at 1 μg per μL concentrations in the running buffer. The human serum was derived by centrifugation from whole blood obtained from a healthy adult male, and was diluted by a factor of 20 in the running buffer.

Agarose gels of 1.5 percent weight-to-volume in running buffer were cast in one of two sandwich-like manners. For native fluorescence detection, the gel was cast on a clear plexiglass tray and covered with a 70 mm × 80 mm × 1.5 mm non-fluorescent quartz plate (Quartz Scientific). For ultraviolet absorption detection, the plexiglass tray was replaced with one made from red fluorescent plexiglass acquired from Plexiform (Iowa City, IA). In both cases, the gels were 3 mm thick, and the backs of the trays were spray-painted flat black. Sample wells were formed with a 1 mm × 5.5 mm comb and were approximately 2 mm deep. In all cases, 5 μL of sample

were loaded in each well.

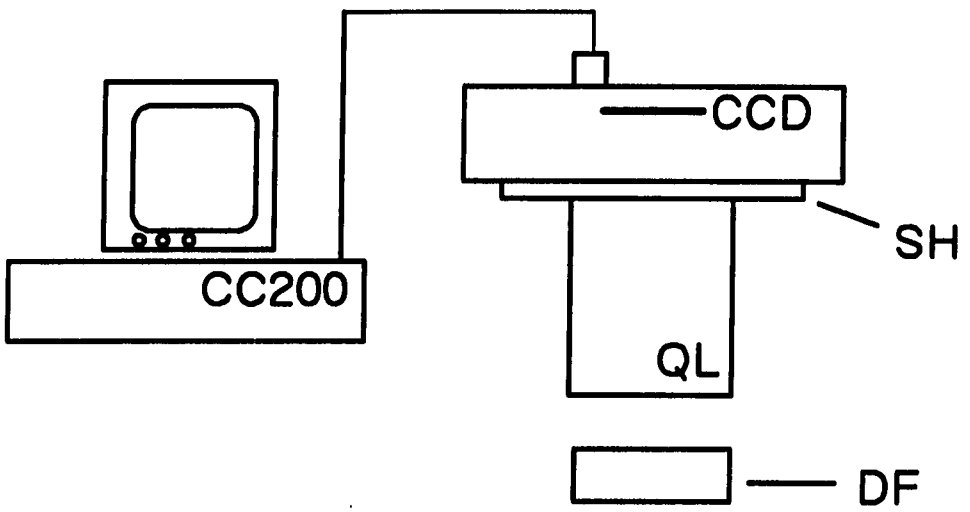
Electrophoresis was carried out horizontally in a mini-sub cell (Bio-Rad) at a constant 50 volts. The current measured was 20 to 25 mA. The running buffer level was adjusted to touch both ends of the gel without running on top of the gel sandwich. This was done to leave good optical access to the gel through the quartz plate.

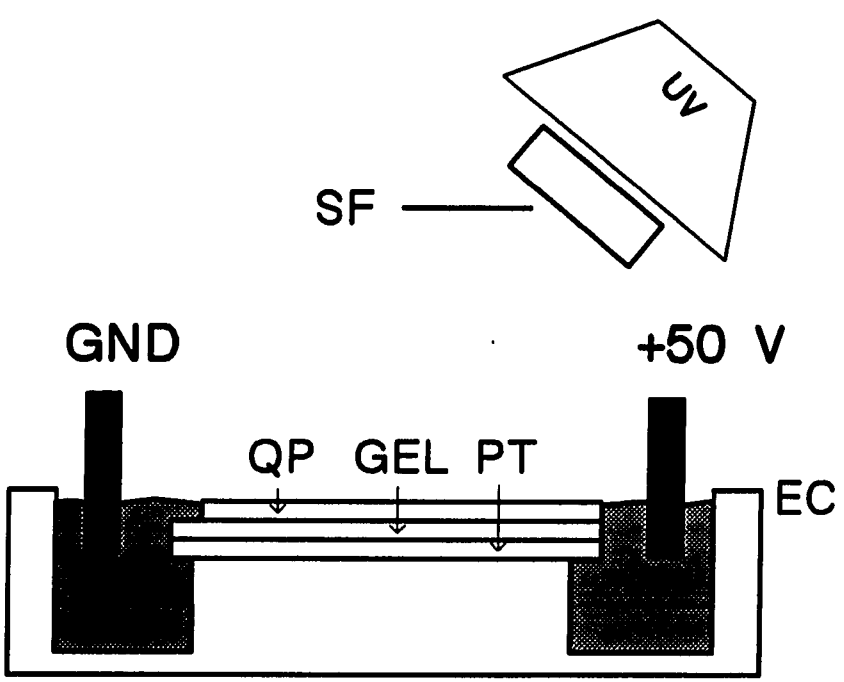
For the staining detection method, gels were run using on-line detection, removed after completion of electrophoresis, and placed in a staining solution containing 40 percent methanol, 10 percent glacial acetic acid, and 0.5 percent coomassie brilliant blue (Sigma) for four hours. Destaining, using the above solution minus the coomassie brilliant blue, required 36 hours due to the thickness of the gels. The destained gel was then placed on a clear plexiglass tray that had been spray-painted white for detection.

Detection

Figure 1 shows a schematic of the detection set-up used for all three detection modes. A scientific grade cooled CCD digital camera system (PhotoMetrics Ltd, model CC200, Tucson, AZ) was used to image the gels during electrophoresis. The system was equipped with a thermoelectrically cooled (Thomson TH7882) 384 × 576 pixel CCD. The camera head, with a 105 mm f/4.5, UV Nikkor quartz camera lens (Nikon, Japan), was

Figure 1. Experimental set-up. CCD = charge-coupled device camera, SH = shutter, QL = quartz lens, DF = detection filter, UV = ultraviolet lamp, SF = source filter, EC = electrophoresis cell, QP = quartz plate, GEL = agarose gel, PT = plexiglass tray, CC200 = data system for camera.


CCD
SH
CC200
QL
DF


UV
SF
GND
+50 V
QP GEL PT
EC

securely mounted 55 cm above the gel. For native fluorescence, a 30 watt hand-held UV lamp, operating at 254 nm, was mounted 15 cm above the gel. A 2" × 2" 10 nm bandpass interference filter centered at 254 nm with 13 percent transmittance (Melles Griot), was placed on the UV lamp, with the remainder of the light emitting area masked off. To isolate the detector from the excitation source, a 305 nm glass cutoff filter (Melles Griot) was placed in front of the camera lens. For UV absorption detection, the interference filter was removed, and a 550 nm glass long-pass filter was placed in front of the camera lens. During detection of stained gels, the UV lamp was removed and diffuse room light was used for illumination.

RESULTS AND DISCUSSION

General Considerations

To correct the images for variations in individual pixel gain, non-uniformities in illumination, and distortions due to the gel itself, the CCD imaging system was operated in a mode commonly called flat-fielding. This was accomplished by placing the loaded gel in the detection chamber, illuminating it with the ultraviolet lamp, acquiring a blank image, and storing that digitally prior to application of the running voltage. Then, at desired intervals throughout electrophoresis, images were acquired and flat-fielded by simply dividing the current image by the blank, pixel by pixel, and multiplying the result by a constant factor. These calculations require only a few seconds for our 576 × 384 pixel images. The resulting images can then be displayed immediately, and are also stored digitally for later analysis and printing. Flat-fielding was not applied to the staining technique since the gel must be handled, and it would have been nearly impossible to place the gel in the exact position it was at prior to electrophoresis.

A major concern in both on-line detection schemes was fluctuations in background signals due to changes in the optical properties, such as fluorescence and scattering, of the gel during electrophoresis. Such fluctuations are not

corrected by flat-fielding, and, if large enough, tend to obscure the signal bands. Acrylamide gels were found to have very unstable fluorescent and scattering backgrounds that were possibly due to artifacts of the polymerization reaction and the fact that the gels did not adhere well to the quartz coverplate. Even with extensive pre-running of these gels in the horizontal sandwiched configuration, a steady background could not be achieved. Agarose gels have a higher background signal level due to scattering from the more opaque gel, but it is a constant background which is easily normalized during flat-fielding. One problem with our sandwiched agarose gels, however, is that this background changes as the gel heats up to near its melting point (typically 36°C), and becomes less opaque. During electrophoresis, gels experience Joule heating that is directly related to the electrical current being passed through the gel. To minimize this heating, all gels were run with a 5 mM buffer instead of the 500 mM buffers typically used. This reduced the current from over 100 mA to near 20 mA for all runs, and gel heating was no longer a serious problem. An alternative approach that would allow for normal buffer strengths would be to design a system incorporating back-side gel cooling while maintaining front-side optical access.

Many components of the on-line imaging system had to be carefully selected due to the fact that both absorption and fluorescence occur in the ultraviolet. Current CCDs, as

manufactured, have an insignificant quantum efficiency below 400 nm. The CCD in our system was coated with MetaChrome II, a proprietary chromophore that is vacuum deposited on the CCD sensor when the imaging system is built (13). With this coating, a quantum efficiency of 20 percent is attained for the entire ultraviolet region. To prevent unnecessary background signal, the plate covering the gel was constructed from Suprasil (Heraeus Amersil) fused silica, which does not fluoresce when irradiated with UV light. The optimum excitation and absorption wavelength of proteins is near 280 nm, but there is no convenient and intense broad area illumination source available at this wavelength. Instead, a low-pressure mercury lamp emitting at 254 nm was used. Although this is quite far from the maximum absorption wavelength, enough light was absorbed to demonstrate both on-line techniques. For the UV absorption technique, the use of a more appropriate wavelength would result in larger absorbance signals, thereby increasing the signal-to-noise ratio. For native fluorescence, a wavelength closer to the absorption maximum would shorten the amount of time required for signal accumulation, and thus reduce the amount of background accumulated, again resulting in an increased signal-to-noise ratio. The last component of the imaging system subject to the ultraviolet requirement was the camera lens. Normal flint glass lenses do not pass light below 350 nm, and are thus not applicable in the native fluorescence

scheme. The lens required, which was made with quartz and lithium fluoride optical elements, was quite expensive and had to be custom built.

Native Fluorescence

Figure 2 is the flat-fielded image of protein native fluorescence detected after 70 minutes of electrophoresis. The image was accumulated for 200 seconds, while electrophoresis was suspended, and flat-fielded with a 200 second blank frame as described earlier. The wells are located at the top of the image, and electrophoresis proceeded vertically down the imaged area. Counting from the left, lanes 2 and 4 were loaded with the two protein standards, and lanes 1, 3, and 5 were loaded with diluted human blood serum. Previously (12), we have found that the fluorescence efficiencies of proteins are substantially higher at pH 10 than at pH 7. This explains our choice of the running buffer to optimize detectability.

Figure 3 presents the information from lanes 2 and 3 in a graphical manner. The plots were generated by averaging the columns of pixels that compose a lane into a one-dimensional chromatogram-like data set versus distance travelled in the gel. For the protein standard mixture, the first band, located 2.5 cm from the starting well, represents conalbumin with a molecular weight of 68,000 Daltons, while the second

Figure 2. Flat-fielded image of protein native fluorescence in a slab gel after 70 minutes of electrophoresis.

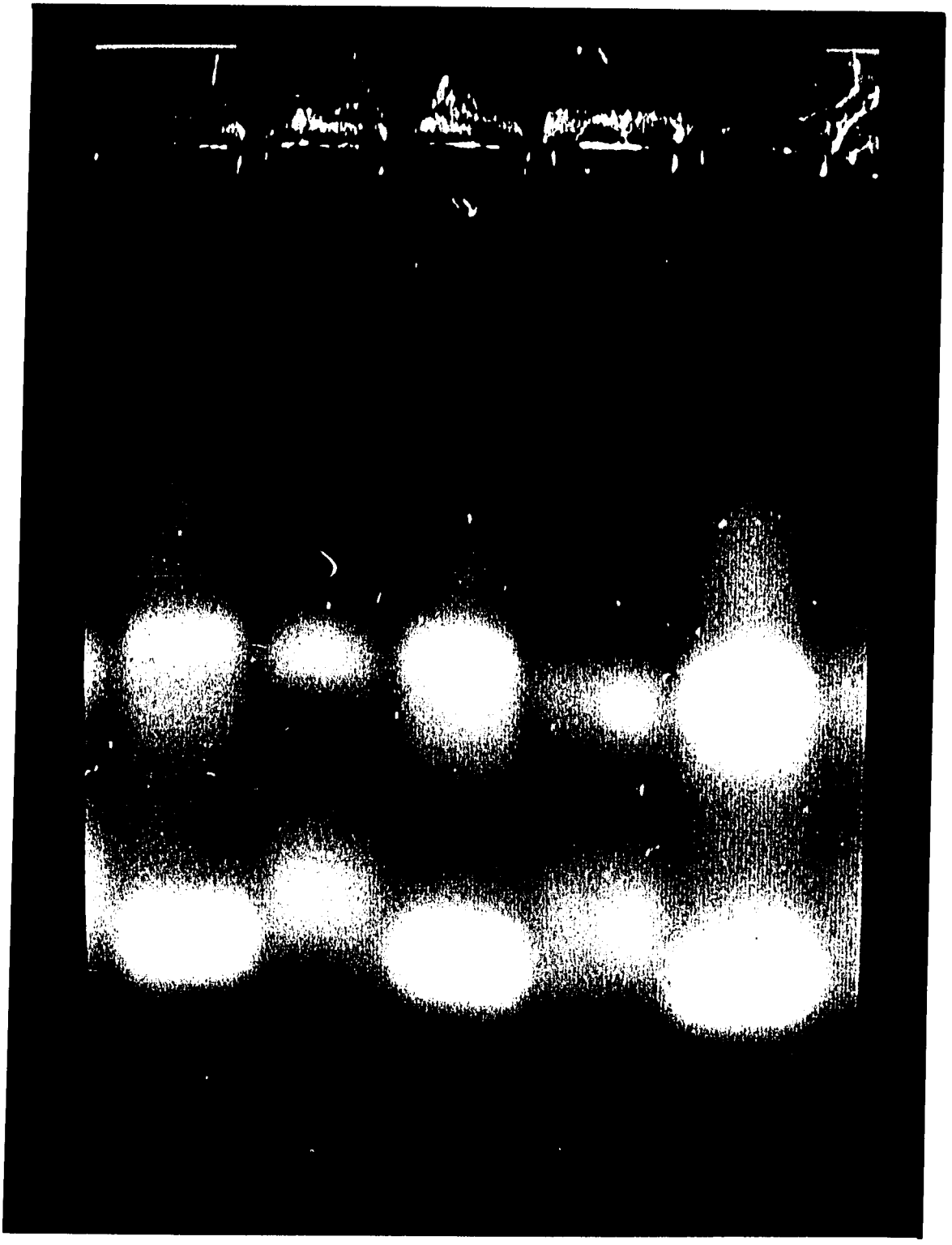
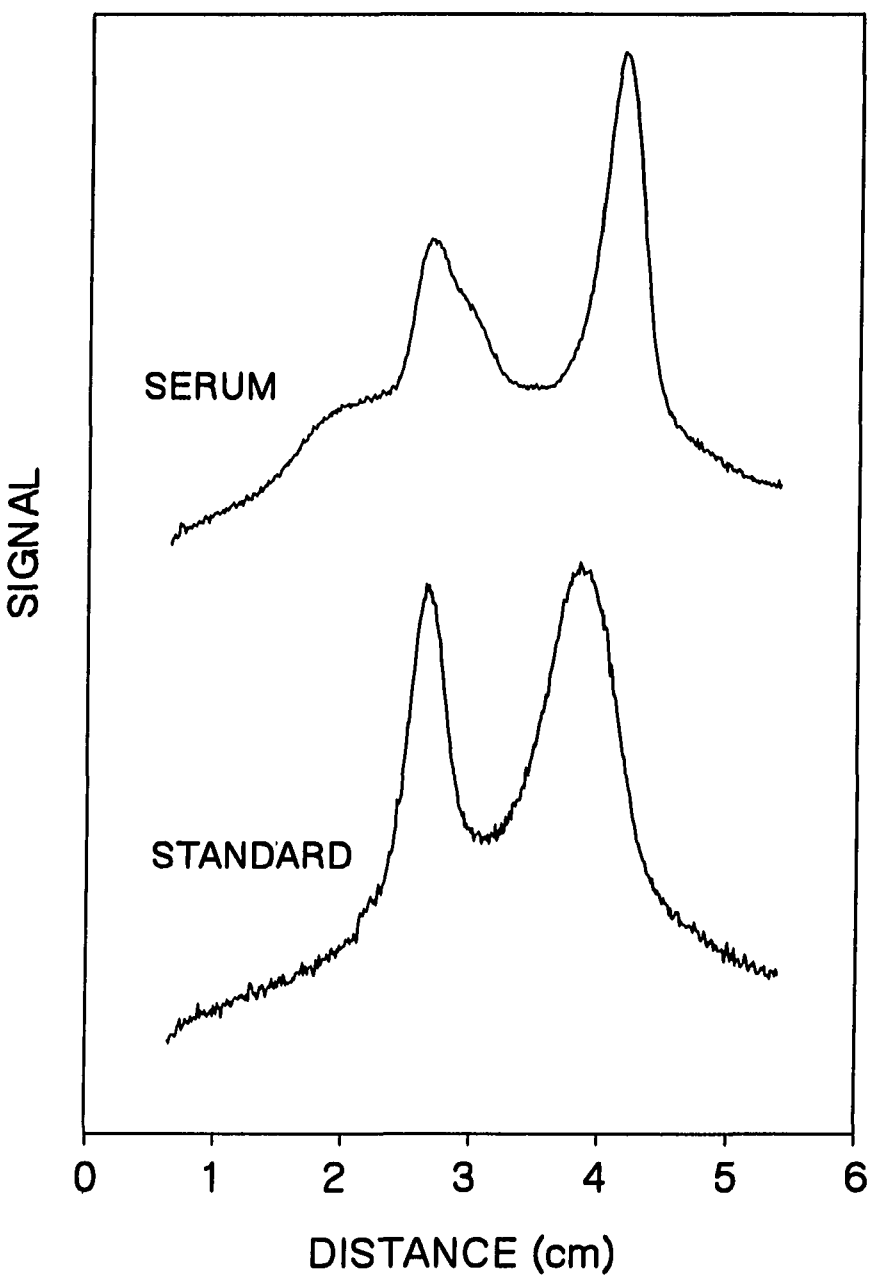


Figure 3. One-dimensional plots of separate lanes containing a standard protein mixture and human serum with native fluorescence detection.

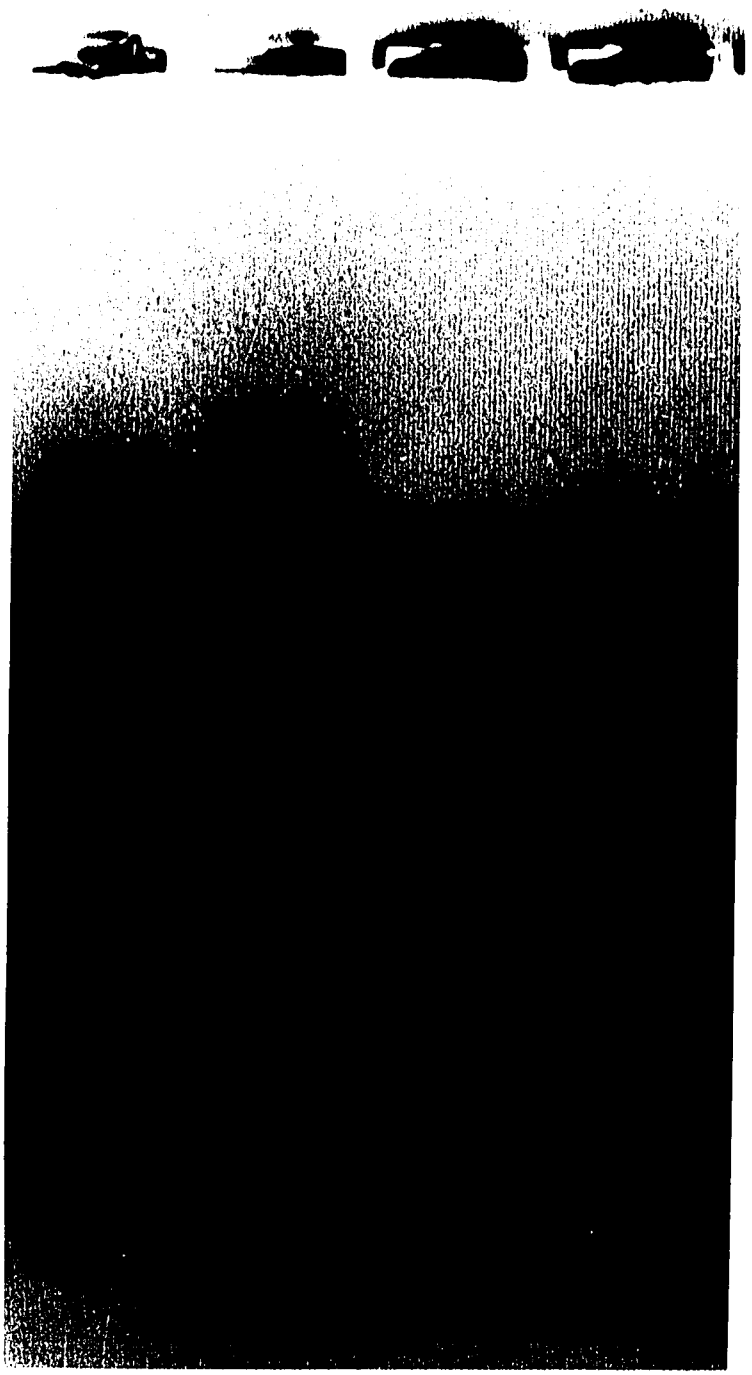

SERUM
STANDARD
SIGNAL
0
1
2
3
4
5
6
DISTANCE (cm)

band, at 3.75 cm, is chicken egg albumin at 45,000 Daltons. In the plot for human blood serum, the unresolved features from 1.5 to 3.0 cm are the four major globulin fractions while the well resolved band at 4.0 cm is human albumin (14). These separations are typical of slab gel electrophoresis at typical loadings, in agarose or in acrylamide gels. It can be seen that detectability is more than adequate for these applications. Exposure times under a minute are clearly sufficient, particularly if more intense lamps are used.

Ultraviolet Absorption

Figure 4 is the flat-fielded image of an ultraviolet absorption gel image after 70 minutes of electrophoresis. As before (9,15), the CCD monitors the fluorescence of the plexiglass in the visible region rather than the source intensity at 254 nm. This avoids interference from UV fluorescence from the gel, the buffer, or the optical components, and allows a standard CCD chip to be used. Thus, we expect UV absorption to be applicable to polyacrylamide gels as well. Since absorbance measurements using CCDs are inherently shot-noise limited due to the limited integrating capacity of the individual sensor elements (15), frame averaging was used prior to flat-fielding. In this case, 16 frames, with an integration period of three seconds each, were acquired, averaged, and flat-fielded with a similarly acquired

Figure 4. Flat-fielded image of protein ultraviolet absorption in a slab gel after 70 minutes of electrophoresis.



blank image taken prior to electrophoresis. The orientation of the gel is the same as before with lanes 1 and 2 being the standard protein mixture and lanes 3 and 4 being human serum. A one-dimensional plot of lanes 2 and 3 is shown in Figure 5. The same features identified before are clearly visible here, except that the relative intensities are different. This shows that ultraviolet absorption is a viable real-time technique for monitoring protein gel electrophoresis.

Coomassie Blue Staining

To allow a fair evaluation of the two on-line techniques, a coomassie blue stained gel image was run and is shown in Figure 6. This is actually the same gel that is shown in Figure 2 after staining and destaining as described earlier. Figure 7 shows the one-dimensional plots of lanes 2 and 3, and, once again, the same features occur with different relative intensities. It is important to note that Figure 6 provides an exact one-to-one correspondence to Figure 2, i.e. the detection techniques described here can directly replace standard visualization methods.

Detection Response

The difference in the relative protein band intensities from method to method is easily explained by examining the

Figure 5. One-dimensional plots of separate lanes containing a standard protein mixture and human serum with ultraviolet absorption detection.

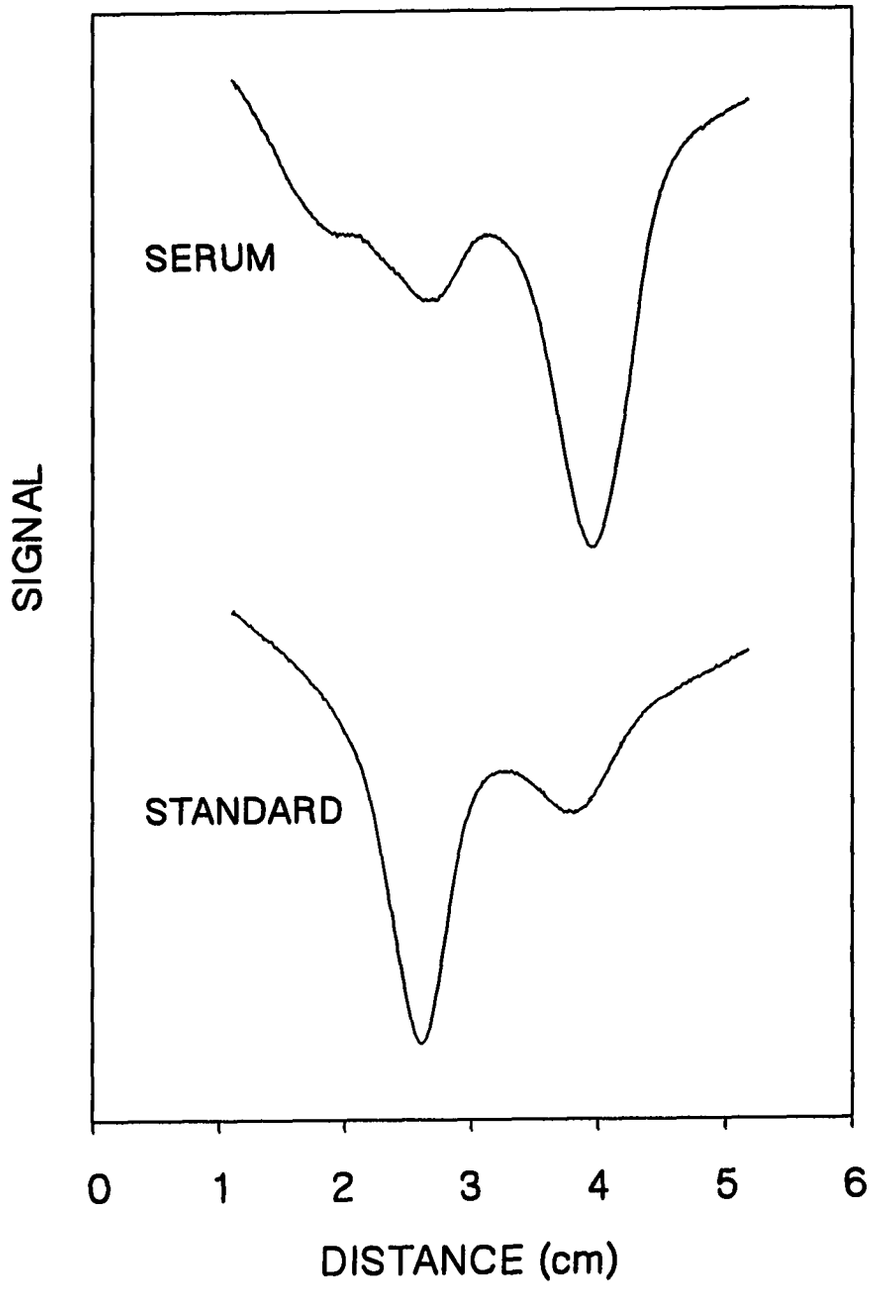

SERUM
STANDARD
SIGNAL
0
1
2
3
4
5
6
DISTANCE (cm)

Figure 6. Image of coomassie blue stained protein gel after 70 minutes of electrophoresis.

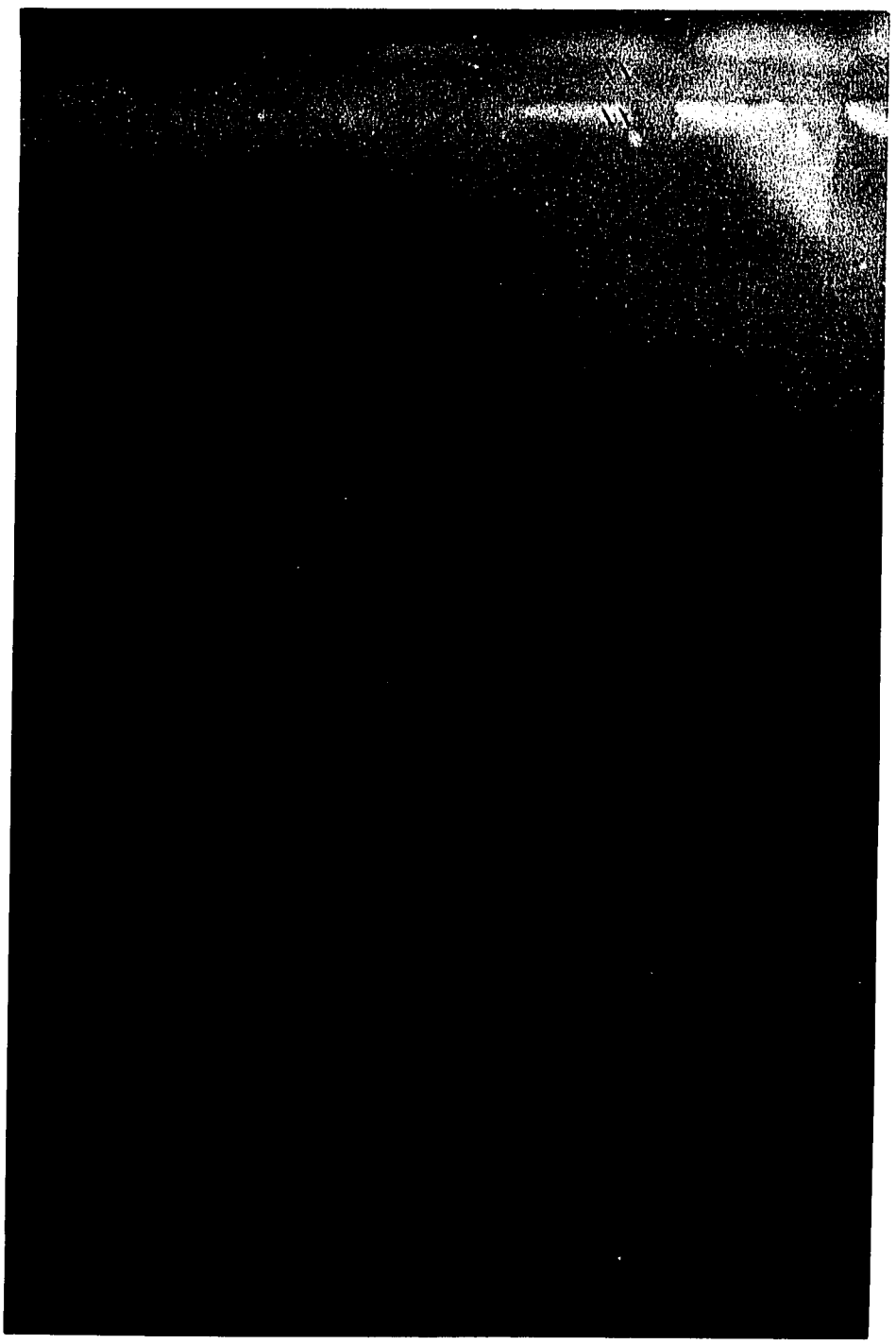
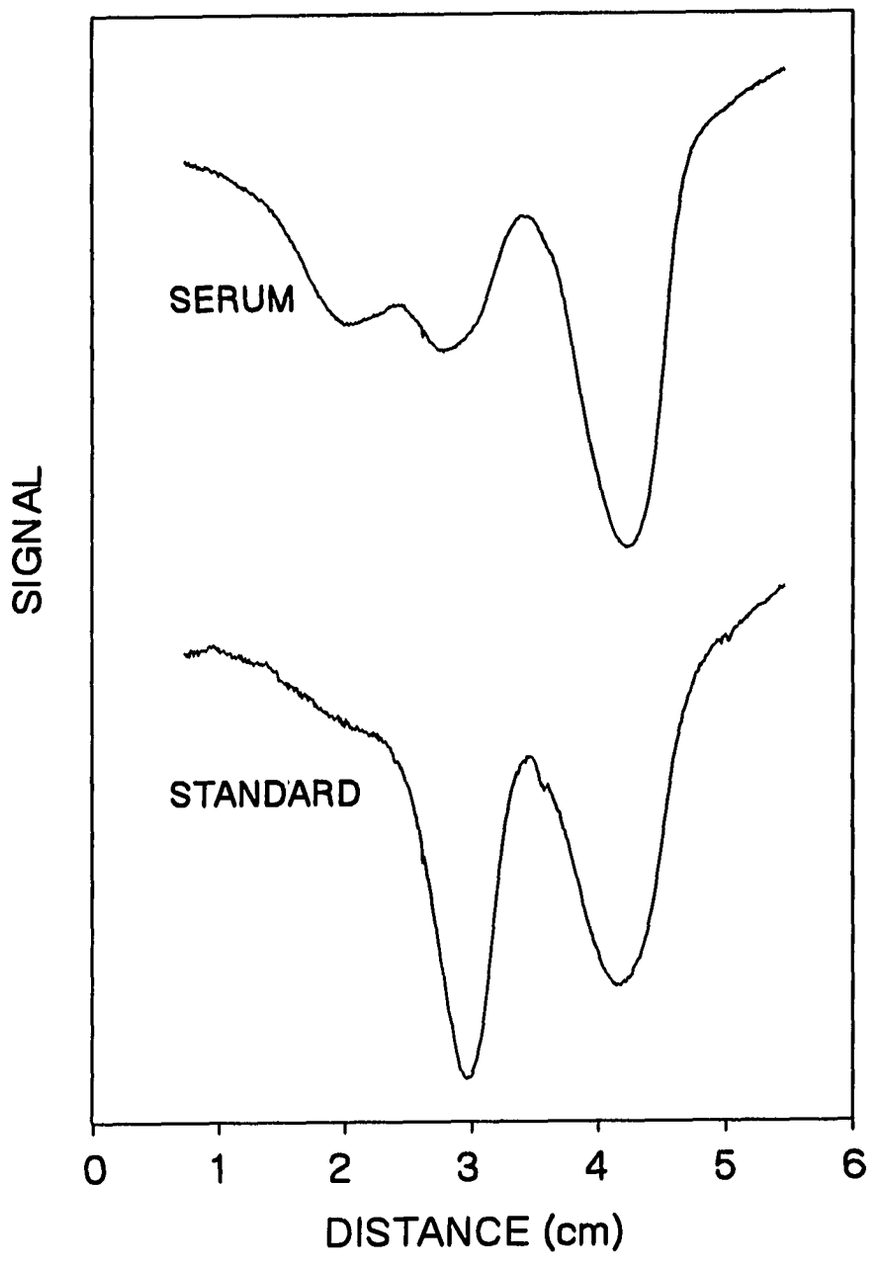


Figure 7. One-dimensional plots of separate slab gel lanes containing a standard protein mixture and human serum after coomassie blue staining. The run is the same one displayed in Figures 2 and 3.


SERUM
STANDARD
SIGNAL
0
1
2
3
4
5
6
DISTANCE (cm)

signal source in each case. In the traditional staining techniques, quantitative information is not available since the signal intensities are not proportional to the amount of protein present in the band due to inconsistencies in the staining process (16).

In the direct fluorescence scheme, a protein's response is determined by many factors, of which the most important is the percentage of tryptophan residues (10). Since not all proteins and peptides contain aromatic amino acids, this detection scheme is not universal. This actually may be advantageous when attempting to detect a small amount of a fluorescing protein in the presence of a large amount of a similarly sized non-fluorescing protein. It should be feasible to some day disperse the fluorescence spectrally, to provide identification based on tryptophan/tyrosine ratios or to study protein conformations during electrophoresis.

The ultraviolet absorption technique is nearly universal since it only requires that a protein contains any of the three absorbing amino acids. The varying response of proteins depends on a combination of percent absorbers and total mass, and is demonstrated clearly in Figure 5. In the standard mixture, both proteins are present at the same total mass as loaded, but the second band (chicken egg albumin) has a lower percentage of absorbers than the first (conalbumin) and thus displays a smaller signal.

Unlike traditional staining methods, both on-line

techniques will give quantitative information. Although the signal obtained for different proteins at the same loaded mass varies, any individual protein will give a predictable signal based on concentration and its optical properties. This is assuming that the run conditions are identical so that spectral shifts due to conformations and the local environments are reproduced. The acquisition of an entire image as opposed to single-point densitometry also allows integration of the intensity (signal) over all pixels within a band to further improve quantitation (17).

CONCLUSION

The CCD camera system is well suited for the on-line detection schemes presented in this paper because of its low noise characteristics and high sensitivity. The ability of the imaging system to perform flat-fielding was essential to both techniques as optical non-uniformities in the gels would have seriously degraded detection. As performed in this study, both on-line techniques appear to be competitive in terms of detectability with conventional staining methods. It should be noted, however, that both on-line techniques could benefit from a more intense light source and a better spectral match with the absorption bands, as well as longer signal integration times.

Both on-line methods are compatible with traditional separation and detection methods. Other than switching to agarose gels and running at lower buffer strengths, both of which are desirable in many cases, the methods impose no restrictions on the operating parameters for slab gel electrophoresis. Since neither the gel nor the proteins are altered, traditional methods such as staining can be performed after on-line detection, as shown in Figures 2 and 6, if desired. The on-line methods also have the advantage of simplifying protein recovery since no chromophores or radioactive labels must be removed after electrophoresis. In fact, if proteins were to be prestained for absorption or for

fluorescence detection, it is likely that the separation will be compromised due to heterogeneity created by derivatization at multiple sites. Finally, the most attractive feature associated with on-line detection is the ability to manipulate the separation interactively during electrophoresis (15).

REFERENCES

1. Andrews, A. T. Electrophoresis: Theory, Techniques, and Biochemical and Clinical Applications, Oxford University Press: New York, 1988.
2. Laemmli, U. K. Nature **1970**, 227, 680-685.
3. Moos, M. Jr.; Nguyen, N. Y.; Lin, T.-Y. J. Biol. Chem. **1988**, 263, 6005-6008.
4. FMC ProSieve Gel System: The Agarose System for Electrophoretic Separation of Proteins, FMC BioProducts: Rockland, 1992.
5. Takagi, T.; Kubota, H. Electrophoresis **1990**, 11, 361-366.
6. Teale, F. W. J. Biochem. J. **1960**, 76, 381-388.
7. Jorgenson, J. W.; Wahlbroehl, Y. J. Chromatogr. **1984**, 315, 135-143.
8. Hassur, S. M.; Whitlock, H. W. Anal. Biochem. **1974**, 59, 162-164.
9. Chan, K. C.; Koutny, L. B.; Yeung, E. S. Anal. Chem. **1991**, 63, 746-750.
10. Konev, S. V. Fluorescence and Phosphorescence of Proteins and Nucleic Acids, Plenum Press: New York, 1967.
11. Swaile, D. F.; Sepaniak, M. J. J. Liq. Chromatogr., **1991**, 14, 869-893.
12. Lee, T. T.; Yeung, E. S. J. Chromatogr. **1992**, 595, 319-325.
13. Sims, G. R.; Griffin, F. Proceedings Of SPIE-The

International Society for Optical Engineering **1989**, 1071, 31-42.

14. Putnam, F. W. The Plasma Proteins, Vol. I, Academic Press: New York, 1960, p. 56.
15. McGregor, D. A.; Yeung, E. S. Anal. Chem. **1992**, 64, 1-6.
16. Yamamoto, H.; Nakatani, M.; Shinya, K.; Kim, B.-H.; Kakuno, T. Anal. Biochem. **1990**, 191, 58-64.
17. Ma, Y.; Koutny, L. B.; Yeung, E. S. Anal. Chem. **1989**, 61, 1931-1933.

PAPER 4

AUTOMATED IMAGE ANALYSIS FOR DISTORTION COMPENSATION IN SEQUENCING GEL ELECTROPHORESIS

Reprinted with permission from Applied Spectroscopy, **1992**, 46, 136.

INTRODUCTION

With the increasing momentum and ambitions of the Human Genome project, it has become clear that computers must play a major role at all levels of analysis, including the direct interpretation of data derived from slab gel electrophoresis; a task that is traditionally performed manually at great cost in terms of human labor and is extremely time consuming. Many hardware systems and computer algorithms have been developed to automate this task (1-9). Unfortunately most of these do not automatically correct for distortions in a gel due to heating, distorted electrical fields, and non-uniform salt concentrations, and those that do are quite complex and prone to error. Others do not entirely remove human judgement from the analysis. This can result in serious misinterpretation of data due to inappropriately determined band locations.

There are several steps towards producing useful information from DNA sequencing gels or determining fragment sizes in restriction fragment mapping gels with the aid of a computer. The raw image must first be digitized. Many approaches have been used, including single-point densitometers (9) and two-dimensional image detectors (3). The specific digitizing system used is not pertinent to this investigation, neither is the specific visualization mode, which can be reading from a photographic image of fluores-cence, chemiluminescence, radiography, or direct monitoring of

the gel (10). In any case, an array of intensities is recorded for the pixels in the image. We use here a charge coupled device (CCD) camera to collect the information efficiently, and use 14-bit analog-to-digital conversion to increase accuracy in the intensity values.

Next, the electrophoresis direction and the width of the lanes (pixel columns) must be defined. Although alignment of the imager with the lasers is straightforward when gels are properly run, one cannot assume the lanes are straight and are of uniform width. Detection of the space in between lanes is fairly easy (6) so that wandering of the tracks due to heating and gel inhomogeneities can be corrected for. In this work, we assume that the tracks have been lined up properly and that each lane has a constant width, i.e. the pixel columns follow exactly the electrophoresis direction.

The most important correction to be made is straightening of the bands within each lane. Distortions such as tilt, smiling, frowning and smudging seriously compromise resolution in sequencing gels and introduce reading errors. The general idea is to use one column as the standard and progressively shift the other columns, one at a time, to line up rows that belong to the same band. This is far superior to simply reading the rows as they are and then using deconvolution techniques to enhance the information content. The reason is that gel distortions have distinct physical basis behind them, and arbitrary mathematical functions used for deconvolution

(4) cannot properly reflect these intricacies. One of the approaches toward intralane band straightening involves the calculation of a correlation function between adjacent columns (3,5,6). Presumably, maximizing the correlation function by shifting the pixels in each column will result in the proper alignment. Maximization can be performed one region at a time along the lane to allow variable shifts along each column, and correlations for more than one column at a time within the same lane can be used. The major difficulty with a simple correlation scheme is that bands often do not show the same width (number of row pixels) across the lane. Edge effects in loading the gel and concentration-dependent smearing of the bands are some of the possible reasons for variable band widths. It is easy to see that there will be ambiguity in the required shifts for correcting such distortions. A solution to this is to first determine the peak maxima of the stronger bands along each column, and then perform correlation based on these maxima alone (5). The weaker bands in the lane are then shifted based on interpolation.

Correlation-based matching is deficient in three respects. It does not take advantage of the fact that band distortions are generally slowly varying functions for physical reasons (heating, concentration, gel distortion, loading), especially if the runs are performed with care. One should take into consideration one band at a time rather than pairs of adjacent columns. For example, one dust particle or an imperfection in

the film localized in one column can unnecessarily cause shifting errors over a significant area around it. Correlation schemes also do not address interlane corrections. Since there is space in-between lanes, it is inappropriate to simply correlate the edges of the successive lanes. Furthermore, the lanes are likely to contain different information (e.g. different terminators, or standard versus unknown fragments) which one is trying to place on the same linear scale. Finally, there are no clear rules that can guide the optimal shift that should be applied to successive bands. A constant shift within a given range (rows) is easy to implement but may only be valid for a very limited range. If variable shifts are used for neighboring bands, the results will be easily influenced by localized noise in the data set.

In this work, we have built upon the band straightening approach based on peak recognition and time-shifting of adjacent columns to match the peaks (5). Instead of simply relying on the peak maxima, we detect the centroids of the major peaks. This has the advantage of using all the information from a band rather than a single data point to enhance noise immunity. In fact, centroids are preferred indicators of positions in chromatographic data analysis (11). Instead of using arithmetic interpolation, we fit the centroids to a third degree polynomial across the columns in the band. This smooths out the contribution of outliers due to local noise in the data. Our approach allows for

fractional pixel shifts where there are only a few pixels per band. Finally, the polynomial fits are extrapolated across the lanes, taking into account the gap between lanes. This way, adjacent lanes are shifted so that sequence determination can be made with improved accuracy.

EXPERIMENTAL SECTION

Electrophoresis

Sequencing gel electrophoresis was performed on acrylamide gel using plasmid DNA under standard operating conditions. Detection was accomplished by radiolabeling followed by a standard autoradiogram procedure. The conventional slab gel electrophoresis was performed on an in-house pulsed field gel electrophoresis (PFGE) setup, which is described in detail elsewhere (12). A restriction digest of Lambda DNA with endonuclease Hind III (BRL, Gaithersburg, MD) was carried out using conditions suggested by the manufacturer, and 130 nanogram total DNA was loaded into each 5 μl well. Electrophoresis was performed using a 0.6 percent agarose gel for one hour with a ten millimolar THAM/phosphoric acid/EDTA buffer at 6 V/cm and a 0.9 second switching time. A running buffer of 0.5 μg/ml ethidium bromide was used for implementation of fluorescence detection. To ensure that the data exhibited distortion, the gel was run at a higher voltage than normal, and was covered with a quartz plate. A higher voltage produced excess heat in the gel, and the quartz plate reduced heat transfer efficiency to the buffer. It should be noted that both of these conditions are actually desirable. A higher voltage results in faster run times, and a cover plate on the gel allows for better optical access to the gel,

resulting in lower noise when optical imaging is to be employed. The net result was a temperature gradient in the gel, with the center of the gel having the highest temperature, and the edges being the coolest. Experimentally, this resulted in increasing mobilities for DNA fragments in the central lanes of the gel.

Image Acquisition

A scientific-grade cooled charge coupled device (CCD) digital camera system (Photometrics Ltd., Tucson, AZ) was chosen for this purpose. The system is equipped with a thermoelectrically cooled Thomson TH7882 384 x 576 pixel CCD. Although there are many systems available for digitizing gel information, CCD systems have many advantages which have been discussed in detail elsewhere (13-15). An image of the sequencing gel used in this study was obtained by placing an autoradiogram on a light box and mounting the CC200 camera head, equipped with a 90 mm flat-field lens (Tamron, Tokyo), 50 cm above the light box. To reduce noise, both random and systematic, frame averaging and flat-fielding techniques were applied. Noise reduction is beneficial when feature recognition methods are to be applied. Frame averaging is simply the acquiring of multiple images followed by the averaging of the pixel values from each image. Although CCDs are integrating devices, there is an upper limit to the

intensity that can be digitized by any given pixel. The accuracy of this value is shot-noise limited, and frame averaging reduces noise as the square root of the number of frames averaged. Experience has shown that averaging sixteen frames is adequate for imaging gels. Flat fielding is a technique used to reduce systematic noise or variations in intensity not related to the signal, such as irregular illumination and non-uniform pixel response. This is accomplished by acquiring a frame-averaged image of the light box setup without the autoradiogram in place, followed by the acquisition of a frame-averaged image of the autoradiogram. Flat-fielding is then accomplished by simply dividing the gel image by the blank pixel by pixel. CCD exposure time was 1.5 seconds per frame, and frame averages and flat-fielding were carried out by a dedicated microprocessor in the CC200 system automatically, resulting in a total data acquisition time of 7 minutes.

The image of the restriction digest DNA gel was acquired by directly imaging the gel using fluorescent staining with ethidium bromide. The same CC200 camera head was mounted 50 cm above the PFGE setup as described earlier. This setup was chosen because direct optical access to the gel is present. Camera operation was the same as above except that the blank image was that of the gel before electrophoresis was begun. Thus, when flat-fielding was performed, any fluctuations in optical signal due solely to inconsistencies in the gel were

normalized out. Once acquired, the images could be displayed immediately on the high-resolution video monitor of the CC200 system and transferred in binary format to a host personal computer via a dedicated IEEE interface (National Instruments, Austin, TX), for further processing.

Image Analysis

Feature recognition

During feature recognition the computer must first obtain information describing the general features of the gel image, such as outer boundaries, orientation of gel and lanes, and lane boundaries. Outer boundary definition is performed by the user, and is a trivial task. Orientation of the gel is fixed such that the "starting wells" are situated in a horizontal manner, and the direction of electrophoresis is vertical from top to bottom of the image. For this work, lane boundary information is also defined by the user. It is relatively easy to automate this step by searching the image data array for groups of pixel columns that are closely related, as reported by Elder (8), thus reducing human intervention requirements.

Once the boundaries are set, specific feature recognition can begin. In this case, the DNA bands in each lane are the specific features to be identified and characterized. For each lane, bands are ultimately described in a three-

dimensional array consisting of band number (top to bottom), pixel column number (left to right), and pixel row number (top to bottom) of the centroid of intensities in the band. First, rough peak location ranges are determined by setting a threshold based on the total range of the intensities in each pixel column of interest, followed by a search for continuous groups of pixels that exceed that threshold. The threshold should be set high enough to include only the strongest features. These will be least affected by noise. Computation will also be minimized. Exact peak locations are then calculated as the weighted average (or centroid) of each continuous pixel range. This is a fairly standard approach in analyzing chromatographic data. Although not as sophisticated as other techniques, this method is easily implemented, is not computationally intensive, and provides adequate results. This is however superior to simply using the peak maximum, which is affected more by noise and by asymmetry in the bands.

At this point, data validation is performed. It is imperative that each pixel column in a given lane report the same set of peaks. To verify this, we use the assumption that DNA bands are continuous and relatively smooth features. The peak locations that are believed to compose a band only need to be checked for continuity of these locations across the pixel columns that compose a given lane. If a new location is found to be discontinuous (beyond a few pixels) from adjacent columns, then that location must be a spurious peak and can be

rejected. The band number for all subsequent peaks in that pixel column must then be reduced by one to correct the numbering scheme. If a peak is expected based on peaks in adjacent columns but not found, the omission can be corrected by inserting an appropriate entry into the data array. Since peak locations vary little from pixel column to pixel column and since these locations are eventually going to be fitted to a polynomial, it is acceptable to simply use the peak location from the previous pixel column for the overlooked peak location. In this case, the band number for the current and all subsequent peaks in that pixel column must be increased by one to correct the numbering scheme.

Image manipulation utilizing feature analysis

Image manipulation, based on feature analysis, can begin once the features (major centroids) have been identified. Intralane straightening can be accomplished quite easily using the information obtained during feature recognition. To accomplish this in any given lane, the first pixel column is chosen as the standard and polynomial regressions are performed to obtain functions that best fit the peak locations in each pixel column to the peak locations in the standard. Third order polynomial regression is used because acceptable results are obtained with minimal computational effort. Higher-order polynomials are not used to minimize contributions from noise. Intralane straightening is then

accomplished by applying the functions obtained to each peak location in the given lane. Other pixels are shifted by interpolating the parameters of these functions between the peaks. Since most new pixel locations are fractional, the intensity of the pixel is distributed proportionally between the two nearest integer positions. Since all of the intensity information is retained, no quantitative information is lost. It should also be noted that since the functions are applied to all pixels, weak bands not detected during feature recognition will also be straightened.

Interlane straightening is a more difficult problem because there are no features between lanes to allow analysis of the distortions. Therefore one must rely on extrapolation of the functions determined for the lane adjacent to the intralane area. The distortions in a given lane are described by the vertical shifts in pixel locations at regular intervals throughout the pixel column for each pixel column in the lane. A second-order polynomial regression is then performed for the shifts in location at each interval versus pixel column position, and the resulting functions are extrapolated to the first pixel column of the next lane. It was found that second-order regression provided the best results. Higher order regressions fail because the positions being extrapolated to lie outside of the range of positions used in the regression, and these functions tend to become erratic and "explode" beyond the regression range. Once all of the

shifts at regular intervals are found for the first pixel column in the next lane, a polynomial regression can be used to obtain a function to normalize this column to the first pixel column of the last lane in a manner similar to that described for intralane straightening.

In actual use, the above two methods are implemented as follows. Starting with the leftmost lane, intralane straightening is performed, and an interlane correction function is obtained for the next lane. In the remaining lanes, intralane straightening is applied, the next interlane straightening function is determined, and all previous interlane straightening functions are applied to all pixel columns in the current lane.

The algorithm was coded in Turbo Basic (Borland, Scotts Valley, CA), a compilable form of BASIC, on a 80386 based IBM compatible personal computer operating at 25 Mhz. Total analysis time for a typical four lane gel is approximately one minute. This could be reduced further by optimizing the program and switching to a computer more suited to computationally intense image analysis.

RESULTS AND DISCUSSION

Figure 1 is an image of a typical sequencing gel that exhibits distortions so severe that normal sequencing routines based on band centroids result in an incorrect sequence. This image was chosen because the reliable sequence analysis was performed elsewhere on the gel where no distortion was perceivable. These other lanes can be used as a standard for later comparison. Figure 2, an expanded view of a portion of figure 1, demonstrates that normal sequencing routines based on band centroids would fail for this particular gel. From the standard, it is known that the sequence of this section is GTACACTAG. Figure 3 is a plot of intensity verses migration distance for the four lanes in figure 2, with each trace representing the average of the pixel columns that compose that lane. The centers of the peaks are equal to the centroids of the bands they are derived from. The peaks are numbered according to the correct sequence. Peaks 1 and 2 are ambiguous whereas both 3 and 4, and 5 and 6, are cases in which two base pairs would be reversed during sequence reading. The bands are broad and the intensities are low, as expected from the skewed positions among columns. There are also small bumps in the various lanes due to overlap between adjacent lanes.

Figure 4 shows the same data presented in Figure 1 after

Figure 1. Sequencing gel image with zero background and positive peak intensities.


GATC

Figure 2. Expanded view of sequencing gel image.

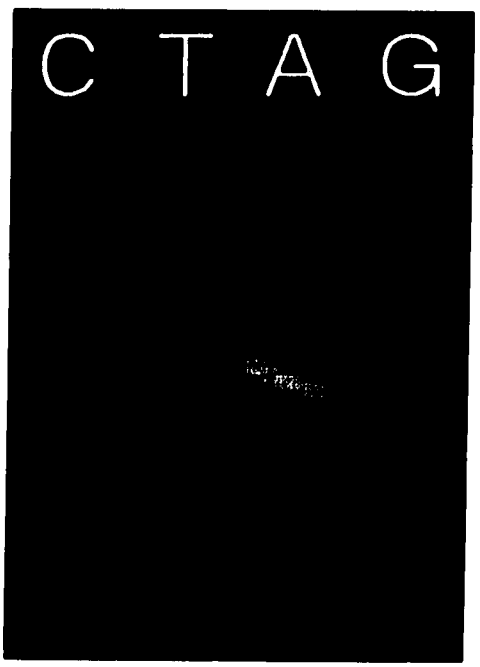

C T A G

Figure 3. Average pixel intensity versus position for each of four sequencing lanes. Actual base pair sequence is indicated.

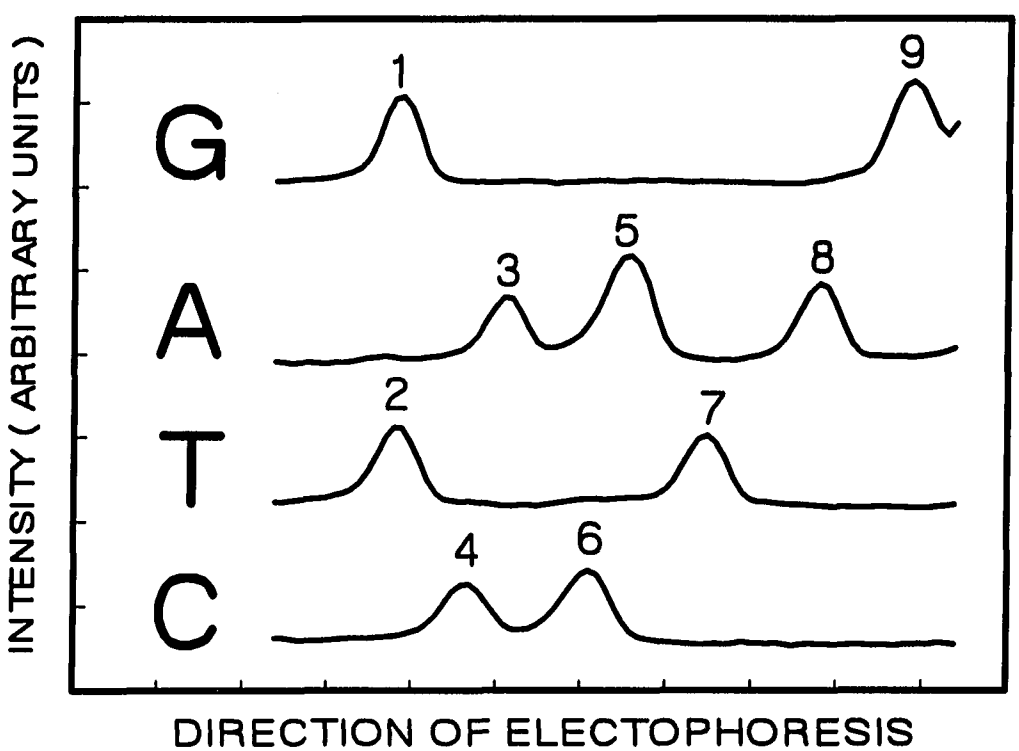

INTENSITY (ARBITRARY UNITS)
G
A
T
C
1
9
3
5
8
2
7
4
6
DIRECTION OF ELECTOPHORESIS

Figure 4. Sequencing gel image following analysis.

application of our band straightening algorithm. In this particular gel, the original lanes overlapped due to a distortion in the horizontal direction. Since the overlapping pixel columns were too complex, they were treated as interlane areas and blanked out to facilitate viewing. Quantitative information is lost in the final results, but is not needed since the goal is simply to read the sequence. Figure 5 is an expanded view of the same area shown in Figure 2, and Figure 6 is a plot similar to Figure 3 for the straightened data. It is clear from both of these that the sequence can be easily read by locating band centroids. Indeed, the correct sequence was read from Figure 6.

To better measure the performance of our system, a plot of distance between adjacent bands versus band number for the raw, straightened, and standard (undistorted portion of the same gel) data is presented in Figure 7. In the trace for the raw data, negative values correspond to positions where the sequence would be read incorrectly. The traces for the straightened and standard data are quite similar and correspond to sequencing data that is relatively consistently spaced, with the spacing increasing toward the bottom of the gel, which is what one would expect.

To further demonstrate the performance and usefulness of our technique, a set of standard DNA fragments was analyzed. Figure 8 shows an image of a gel in which severe distortion

Figure 5. Expanded view of post-analysis sequencing gel.

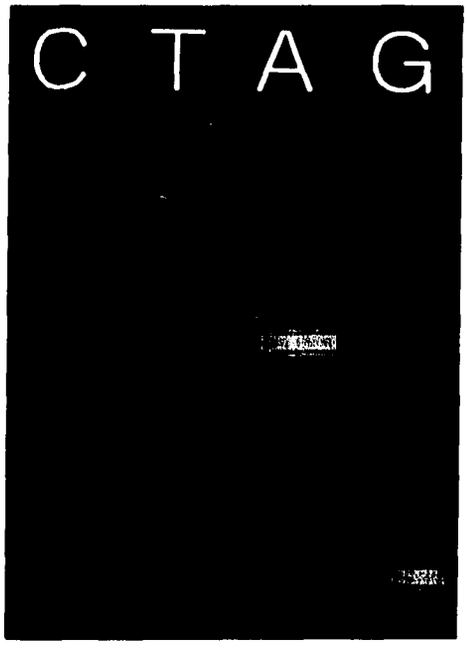

C T A G

Figure 6. Average pixel intensity versus position for each sequencing lane after analysis.

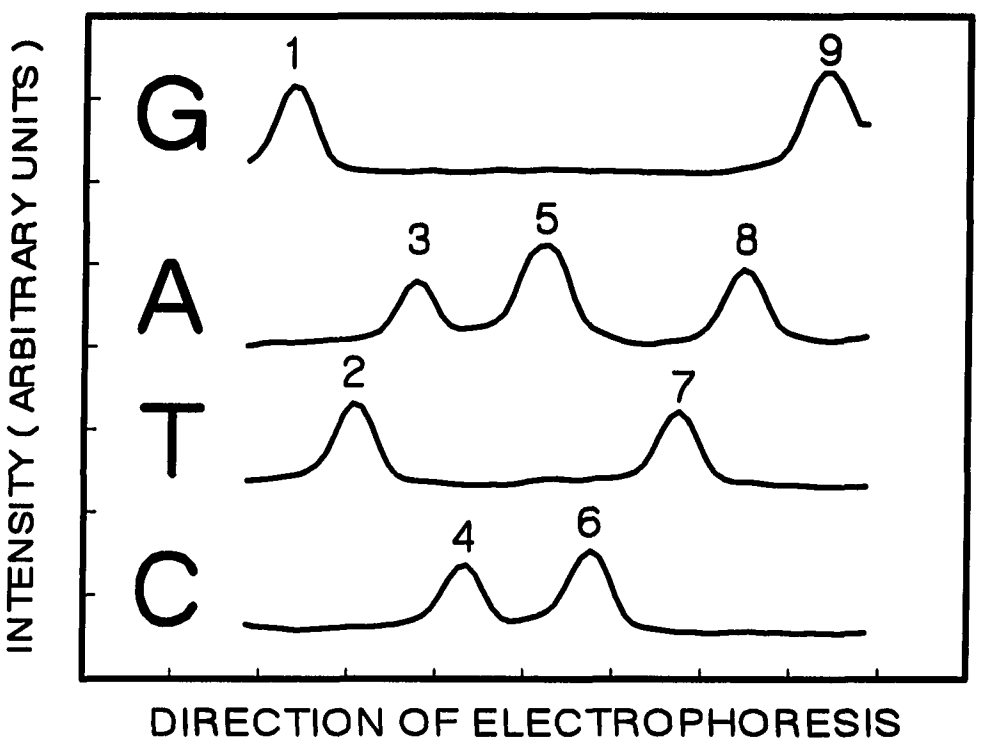

INTENSITY (ARBITRARY UNITS)
G
A
T
C
1
9
3
5
8
2
7
4
6
DIRECTION OF ELECTROPHORESIS

Figure 7. Spacing between adjacent bands for sequencing gel prior to and following analysis, as well as for a standard lane.

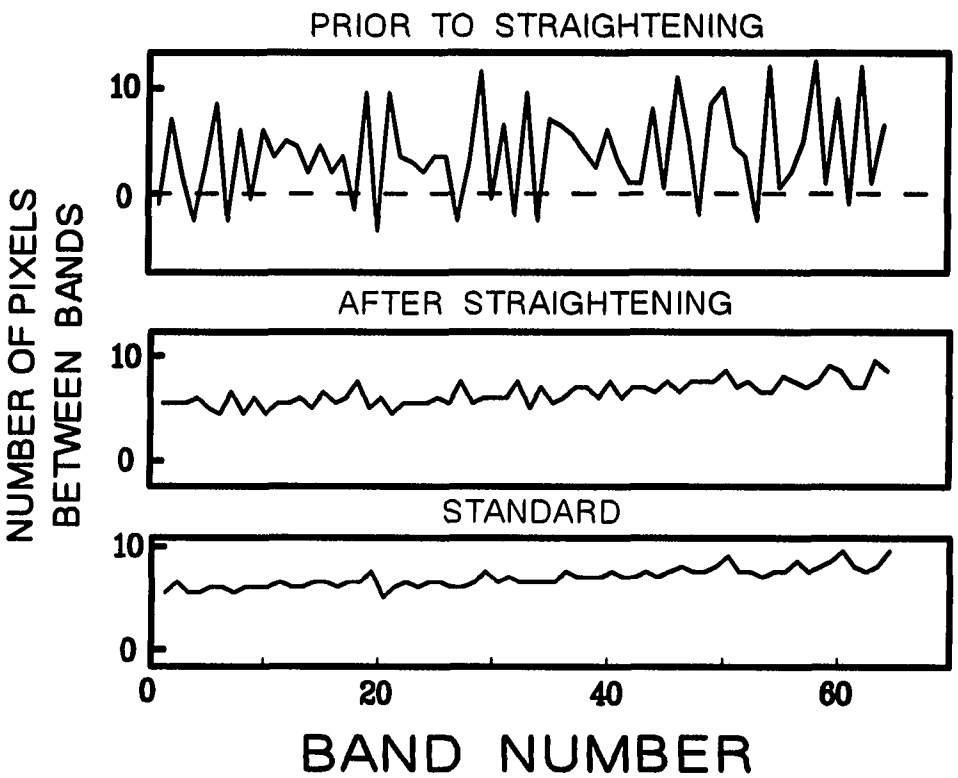
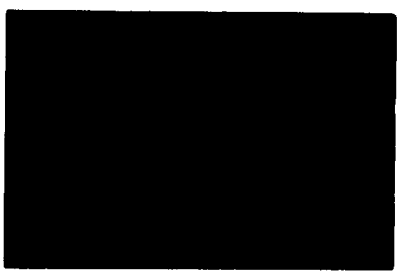
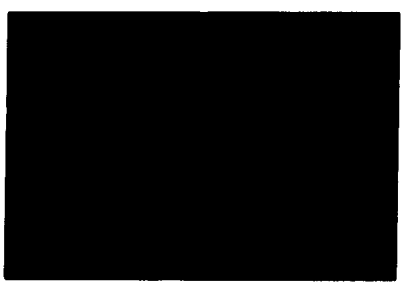

PRIOR TO STRAIGHTENING
10
0
AFTER STRAIGHTENING
10
0
STANDARD
10
0
0
20
40
60
NUMBER OF PIXELS BETWEEN BANDS
BAND NUMBER

Figure 8. Hind III restriction digest of Lambda DNA agarose gel electrophoresis with direct fluorescence imaging.



occurred across the four lanes in which identical Hind III digest samples were run. Since simple molecular size algorithms are based on migration distances, each lane would give substantially different results. Figure 9 shows the image after straightening. It is clear that the migration distance for each band is reproducible for all lanes after correction.

Figure 9. Agarose gel electrophoresis image of Figure 8 following analysis.



REFERENCES

1. Komaromy, M.; Goven, H. Nucleic Acid Res. **1984**, 12, 675.
2. Staden, R. Nucleic Acid Res. **1984**, 12, 499.
3. Elder, J. K.; Green, D. K.; Southern, E. M. Nucleic Acid Res. **1986**, 14, 417.
4. Taylor, P. J. Biomed. & Biophys. Methods **1987**, 14, 71.
5. Sanders, J. Z.; Petterson, A. A.; Hughes, P. J.; Connell, C. R.; Raff, M.; Menchen, S.; Hood, L. E.; Teplow, D. B. Electrophoresis **1989**, 11, 3.
6. Elder, J. K.; Southern, E. M. in: Bishop, M. J.; Rawlings, C. J. (Eds.) Nucleic Acid and Protein Sequence Analysis: A Practical Approach, IRL Press: Oxford, pp. 219-229 1987.
7. Ehrhardt, W.; Englisch, U.; Neuhorr, V. Electrophoresis **1989**, 10, 265.
8. Elder, J. K. Electrophoresis **1990**, 11, 440.
9. Galat, A. Electrophoresis **1989**, 10, 659.
10. Chan, K. C.; Koutny, L. B.; Yeung, E. S. Anal. Chem. **1991**, 63, 746.
11. Barber, W. E.; Carr, P. W. Anal. Chem. **1981**, 53, 1939.
12. Chu, G.; Vollrath, D.; Davis, D. W. Science **1986**, 234, 1582.
13. Freeman, S. E.; Larcom, L. L.; Thompson, B. D. Electrophoresis **1990**, 11, 425.
14. Boniszewski, Z. A. M.; Comley, J. S.; Huges, B.; Read, C.

A. Electrophoresis **1990**, 11, 432.

15. Sutherland, J. C.; Sutherland, B. M.; Emrick, A.; Monteleone, D. C.; Ribeiro, E. A.; Trunk, J.; Son, M.; Serwer, P.; Poddar, S. K.; Maniloff, J. BioTechniques **1991**, 10, 492.2

PAPER 5

AN EXPERT SYSTEM FOR DATA ACQUISITION TO ACHIEVE CONSTANT SIGNAL-TO-NOISE: APPLICATION TO IMAGING OF DNA SEQUENCING GELS

Reprinted with permission from Analytical Chemistry, in press.

INTRODUCTION

In chemical analysis, parallel or multiplexed data acquisition and data treatment is often the rule rather than the exception. For information that is inherently multidimensional, such a mode of operation is essential for handling the large volume of data. Even in the case of one-dimensional information, parallel processing allows substantial savings in the total measurement or computation time.

Traditionally, each data point in parallel measurements is treated with equal weight. The total data acquisition or data processing time is therefore evenly divided among all the data points. This allows processing in the absence of any knowledge about the data set of interest. In the end, some data points will acquire a higher signal-to-noise ratio (S/N) than others. Universally, however, the criterion for satisfactory measurement is to achieve a certain minimum S/N for each and every data point. It is in fact wasteful to consider a given data point further once the minimum S/N level has been achieved. The data acquisition or data processing time for that data point could be more beneficially used to enhance the S/N of other data points. There is therefore a need for expert systems that will allow data acquisition to achieve constant S/N rather than constant time (effort) for the individual data points.

Certain common measurement schemes are readily modified to provide constant S/N over the data set, although we were not aware of any literature precedents of such a mode of operation. In scanning spectrophotometry, very often the system is shot-noise (photon statistics) limited. A constant scan rate will result in data points with varying S/N levels. It should however be straightforward to integrate at each wavelength interval until a certain total count (photocurrent) is accumulated. One can then move on to the next wavelength interval and start integrating again. The scan will become asymmetric in time, but constant in S/N. It is obvious that the total scan time will be reduced. In emission spectrometry, for example, one can further impose a requirement that each wavelength interval will only be integrated for a certain maximum period. This will allow one to quickly skip over regions with little useful information. The gain in efficiency will depend on the amount of useful information versus the amount of useless information, and the range of data values within the set. The spectrometer will have to be controllable interactively, e.g. stepping motor drive, and a resident algorithm must be available to make decisions in real time.

Not all analytical measurements can directly take advantage of this concept. In chromatography, for example, where time is dictated by the separation and not the measurement process, interactive data acquisition is typically not possible. An

exception is in pulsed-field gel electrophoresis (1). In multiplexed spectroscopy such as Fourier transform variants, all data points must be acquired before a valid transform is allowed. The distribution of S/N in the resulting spectrum does not follow a simple dependence on the S/N of the individual points in the interferogram. Even if the mirrors can be asynchronously scanned, one would not necessarily want to implement constant S/N data acquisition.

With the ultimate goal of the Human Genome Initiative being to sequence all 3×10^9 base pairs in human DNA, it is clear that computers must be involved in all steps of the detection and interpretation of sequencing data. Traditional human methods simply cannot complete the task in the time frame desired, and human operation is severely prone to errors due to fatigue and misjudgment. One step in this automation process is imaging electrophoretic gels and storing the information in a format the computer can easily analyze later. Many approaches have been used (2-8), including single point densitometers (9,10) and two dimensional detectors (11-19). The algorithm described here is applicable to systems in which individual picture elements (pixels) can be interrogated and read in any order desired, with the goal being to significantly reduce the number of pixels that must be read and analyzed to absolutely identify and locate all of the signal bands. Bands in sequencing gels often vary in intensity and thus require varying degrees of interrogation by

the imaging device before they can be absolutely confirmed. Previous techniques (20) simply scan all areas repeatedly until the S/N is such that all bands are detectable, and thus much time and effort is wasted scanning areas that are more readily decipherable.

One detector that our technique will be especially well suited for is the charge injection device (CID) based imaging systems. These will be commercially available in 1993. These systems will be quite similar to the scientific-grade cooled charge coupled device (CCD) based imaging systems that first appeared in 1987, except that one will be able to read the pixel intensities in any pattern desired, even nondestructively (21). One limitation of these devices is that each detector element can only hold a certain amount of charge before it becomes saturated. This so-called full well capacity is typically less than 250,000 electrons. Thus, even though these are integrating devices, the data may need to be read out well before enough photons have been detected to assure an adequate S/N to positively identify the image features. This is especially true in absorption-based detection methods such as the ultraviolet absorption method for DNA demonstrated by these authors previously (12). Because the bands appear as very small decreases in the large background signal, the initial image that was read before detector saturation was always shot-noise limited. Although some bands were clearly detectable, other bands required as

many as 64 frames of images to be acquired and averaged before positive identification. Since the images were shot-noise limited, this provided a factor of eight improvement in detection limit.

Another limitation in CID imaging is the data acquisition rate. To assure an accurate reading of the charge in these elements, readout must be performed in a slow-scan mode. Typical readout rates are on the order of 200,000 elements per second. Since typical devices now contain 250,000 to 4 million picture elements, full image readout can take several seconds. Thus, a constant S/N mode of data acquisition for signal averaging will significantly reduce the amount of time spent and the amount of computation required.

EXPERIMENTAL SECTION

In any process that mimics human vision it may at first seem complex, but when properly broken down into smaller tasks using appropriate assumptions, vision algorithms need not be sophisticated or require state-of-the-art programming techniques or prohibitively expensive computer hardware. The present image analysis algorithm can be broken down into several simple steps as follows.

Preliminary Feature Recognition

Before the features of interest can be analyzed, the expert system must first identify the general features. The first step towards this goal is to acquire one full image and perform image analysis to describe and record these features. Defining the expected orientations of the gel features will greatly simplify this task. The algorithm will assume that the gel is oriented such that the starting wells are situated in a horizontal manner, and that the direction of electrophoresis is vertical. In the selected convention, pixel rows are horizontal and pixel columns are vertical. The only general features that need to be located are the lanes. At this point, it is assumed that the lanes are straight and of uniform width, or that lane wandering due to heating and gel inhomogeneities has been corrected for using an

appropriate algorithm (11). Lanes are easily identified and their positions recorded by simply searching for groups of pixel columns that have closely related sets of features, as reported by Elder (8). This can be accomplished at a low S/N level. After this step, the imaged area corresponding to the empty lane boundaries as well as the region preceding the "starting wells" can be omitted from further processing.

Specific Feature Recognition

Once these lane boundaries are defined, recognition of specific features can begin, which in this case are the DNA fragment bands inside of each lane. At this point, it is assumed that the bands are fairly straight and horizontal, or that an automated algorithm has been applied to straighten them, such as the one previously described by us (22). That algorithm would be ideal in this situation because it does not require that all bands be visible, and it requires no time-consuming human intervention. It is also assumed that the bands are correctly ordered, and do not significantly overlap in their pixel row locations. Again, by applying the algorithm previously described by us, one assures such a condition.

With straight lanes and bands, the information contained in the image is essentially four related one dimensional arrays, with each array being composed of the average intensities of

the pixel columns that make up each lane versus the corresponding row locations. Bands in each lane can then be identified with any peak finding algorithm. For this work a simple threshold method was implemented based on the intensity and statistics of the background noise. The noise is assumed to conform to a Gaussian or normal distribution, and its magnitude is determined by calculating the standard normal deviation, which is simply the root-mean-square (RMS) of the pixel intensities in the non-lane area (23). This is commonly referred to as the RMS noise. The RMS noise of the data in the one dimensional arrays will be this value divided by the square root of the number of columns averaged to produce it. The threshold is then set as this value multiplied by an appropriate factor. The magnitude of this factor will determine the probability of locating false signals. A search is then conducted for groups of data points in the one dimensional arrays that exceed this threshold. Any group is considered a band if the number of consecutive row locations exceeds the full width at half maximum (in pixels) expected for the data. This is easy for DNA sequencing gels. Band locations are then recorded to a file as a data pair consisting of the lane number and the centroid of the pixel row locations of the groups of data points identified as a band.

Decision Making

At this point, the expert system must use the information already gained to make decisions regarding the areas that should be read again, and those regions that can be skipped. Obviously, the pixel columns between lanes need not be probed anymore. The only information contained in these areas is the RMS noise, and it can be assumed that the noise will accumulate as the square root of the number of times an area is read, while the signal will accumulate proportional to this number (20). The next step is to decide which pixel rows no longer need to be probed. This is quite easily performed if one assumes that once a band is found, no other bands will be found in the same pixel row locations in the other three lanes. This is quite significant as it allows the algorithm to not only mark the area in which the band was found as inactive, but also three more equally sized areas in other lanes where it can be concluded that no bands reside. By the time most of the bands have been located, this assumption will have allowed the expert system to mark nearly all of the image as inactive versus only one-fourth without utilizing this assumption. A convenient method of programming this information is to create a Boolean data array with one entry for each row. The array is initially set to all true, which is defined to mean that all pixels in that row, which are also in lane areas, are active. When a band is located, a block of

row locations in the Boolean array centered at the band location and equal in number to the expected width of a band in pixels is marked false, or becomes inactive.

Signal Accumulation

Once the initial image has been fully analyzed, a reasonable approach must be found for continuation of the search. Instead of performing the above methods after each pass of the image, it is more prudent to do so only after the S/N has increased significantly for the bands not yet located. For simplicity a square root of two improvement in S/N was chosen. Thus, each pass must double the total number of previous reads of each active pixel before band location and decision making occurs again. Since noise accumulates at a rate of the square root of number of reads, the threshold must be increased by a factor of the square root of two each pass. The signal however will double each pass, resulting in the enhanced S/N that can be exploited to locate bands not found in previous passes. The advantage of choosing this square root of two improvement is easily demonstrated. If, for example, 64 reads are needed to locate the last band, the program will only make seven passes through the band location and decision making portions of the algorithm, which are the most complex. Although not every band will be located at the earliest possible time, computations will be reduced by a factor of

nine.

Termination

Since DNA sequencing data consists of a series of bands that are fairly evenly distributed in the direction of electrophoresis, it is quite simple for the expert system to determine when the reading process is complete. The distances between neighboring bands versus band number should be fairly monotonic as one moves down the gel image, and if this is not the case, the process should continue.

Test System

The expert system was implemented on a 80486-based IBM compatible personal computer (PC Plus, Ames, IA) running at 33 Mhz and operating under DOS 5.0 (MicroSoft). Programming was performed using Turbo BASIC (Borland), which is a compilable version of BASIC that incorporates many PASCAL structures. The 64 simulated images were stored on the hard disk, and pixel intensities were passed on to resident memory (simulated data acquisition) as necessary.

To demonstrate the technique, data was artificially created that would mimic what a detector would report if it were reading a DNA sequencing gel that followed all of the assumptions listed earlier. The data was created starting

with a noiseless image data array of 28 bands randomly distributed in 4 lanes. The dimensions of this image are 384 × 250 pixels and therefore contains 96,000 pixels. Each band is 13 by 52 pixels, and is Gaussian along the pixel column direction. Band intensities vary from 40 counts to 580 counts in steps of 20 counts. The full width at half maximum value for these bands is approximately 5 pixels. The maximum intensities of a band are constant over the middle 40 pixel columns, and decrease in a Gaussian manner on each of the ends which consist of six columns each. This is a good approximation of the results that occurs due to diffusional broadening as a signal band travels through a gel. Pixel resolution is appropriate to avoid blurring due to the detector resolution. The spacing between lanes is 5 pixels. This image was superimposed (added) upon 64 similarly sized images of artificial noise which were Gaussian in nature with an average intensity of zero and root-mean-square variation of 290. The noise in all 64 frames is independent of that in the other frames. Figure 1 is a visual representation of the image before noise was added. Figure 2 is the same with one of the 64 frames of noise added. The images were formatted so that they could be read into and displayed on a Photometrics CC200 Imaging System (Photometrics Ltd., Tucson, AZ). The expert system therefore reads and uses these frames as if they are real gel images obtained from a CCD camera in real time.

Figure 1. Test sequencing image prior to noise addition.

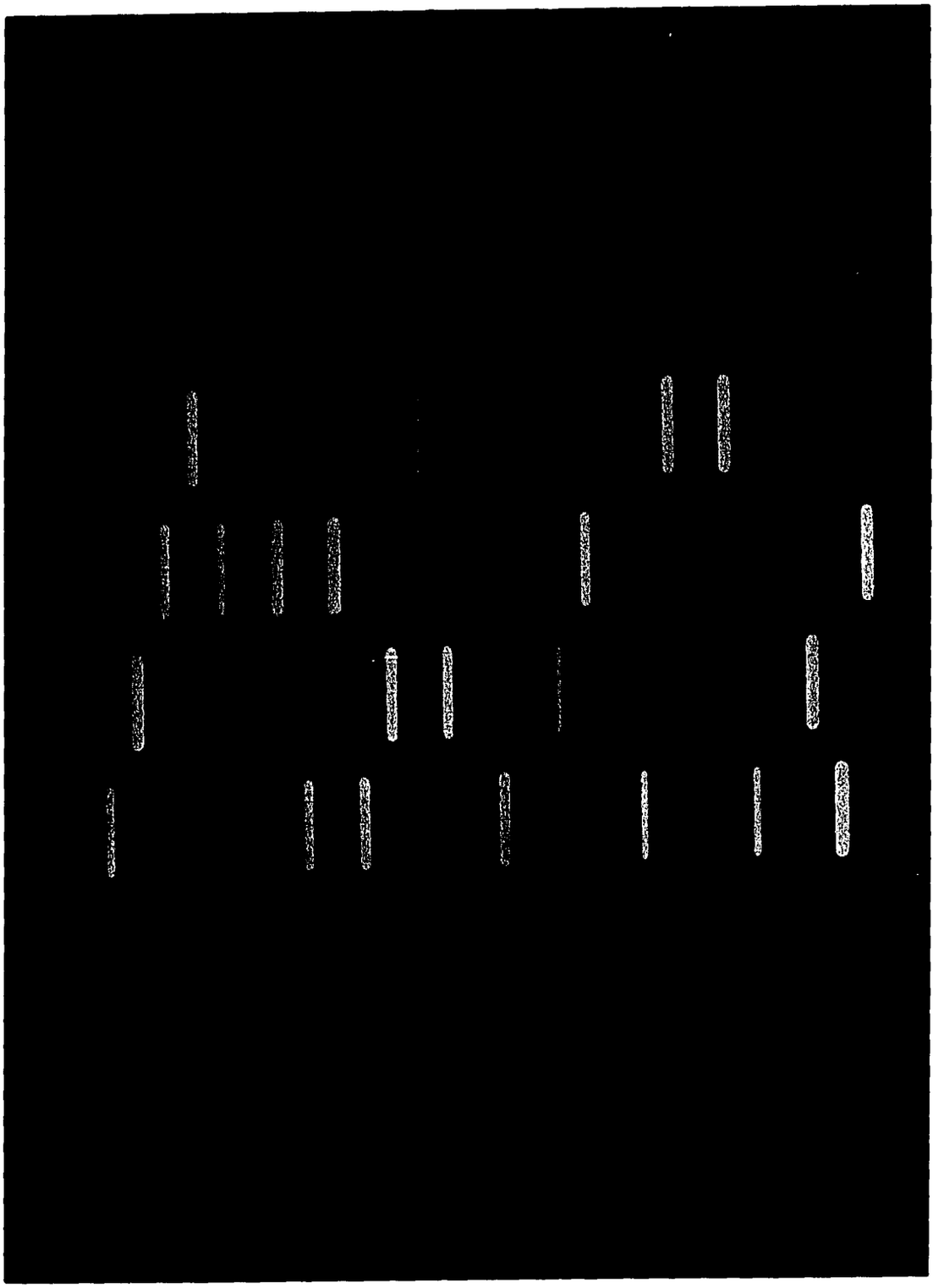


Figure 2. Test sequencing image superimposed on one frame of noise.

RESULTS AND DISCUSSION

Since the lanes are 52 pixels across, and the noise is random, the data stored in a one-dimensional array for a given lane as described earlier have a RMS noise of approximately 40 pixel counts after one reading. Thus, after accumulating one image, the bands have a S/N ranging from approximately 1 to 16. The one-dimensional representations of the lanes for the noiseless image is shown in Figure 3. Figure 4 shows the same representation for a single image with noise added. The S/N will increase by a factor of the square root of the number of frames accumulated until, after 64 frames, the range is 8 to 96, as shown in Figure 5. It is clear that some bands can be positively identified earlier than others. Figure 6 is a similar plot of the four one-dimensional arrays into which the signal and noise were selectively accumulated by the system. When any given band was detected, data acquisition and accumulation was terminated for the band's pixel row locations for each lane. Thus, the noise in other lanes at this position is representative of the noise used to determine the signal to noise ratio threshold. The weaker bands accumulate to a larger total signal because they must overcome a larger accumulated noise.

The expert system itself is quite simple yet elegant. By reducing the data as it is read, the computer program only

Figure 3. One-dimensional representation of lanes in noiseless image.

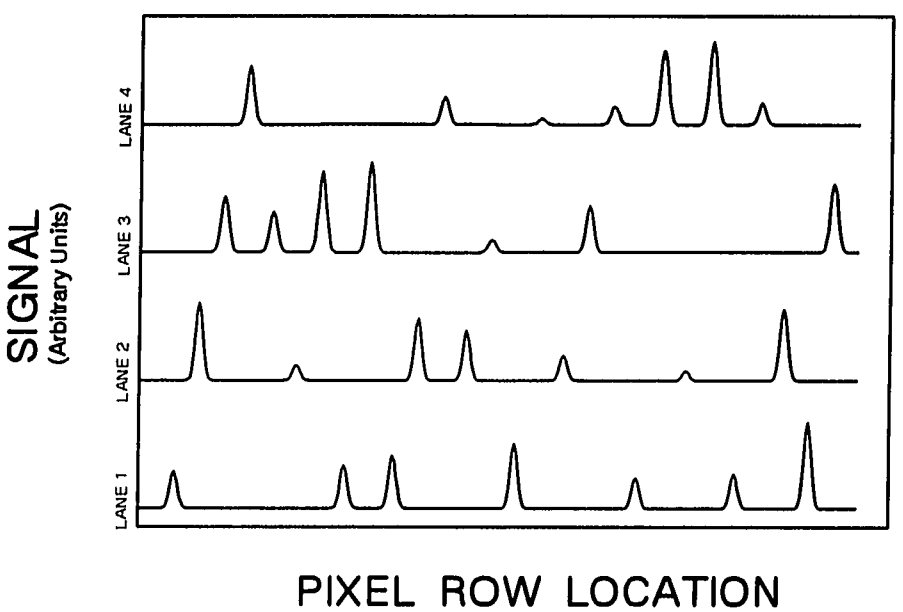

SIGNAL
(Arbitrary Units)
LANE 4
LANE 3
LANE 2
LANE 1
PIXEL ROW LOCATION

Figure 4. One-dimensional representation of lanes in image with noise.

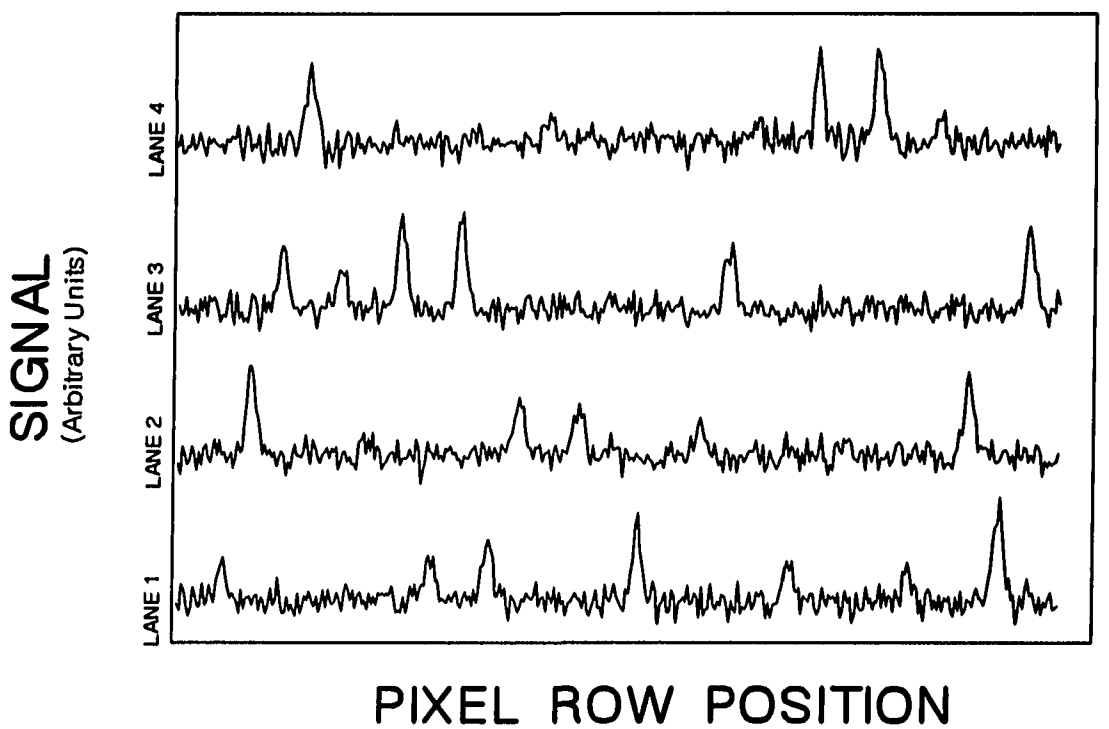

SIGNAL
(Arbitrary Units)
LANE 4
LANE 3
LANE 2
LANE 1
PIXEL ROW POSITION

Figure 5. One-dimensional representation of lanes in image after accumulating 64 frames.

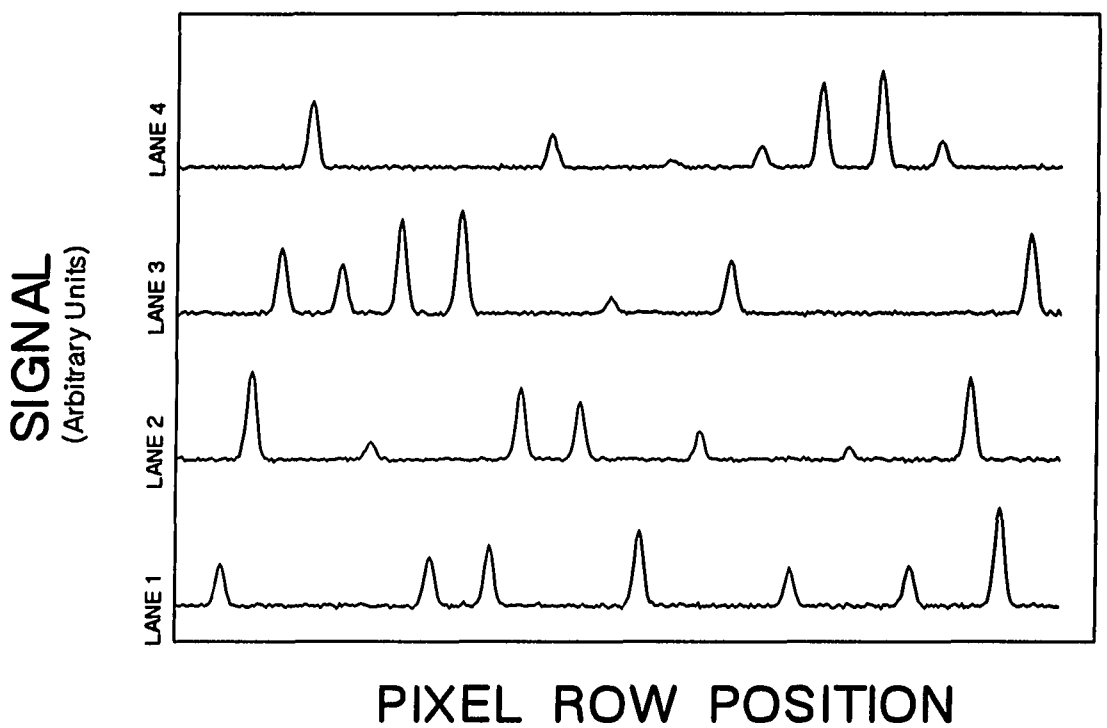
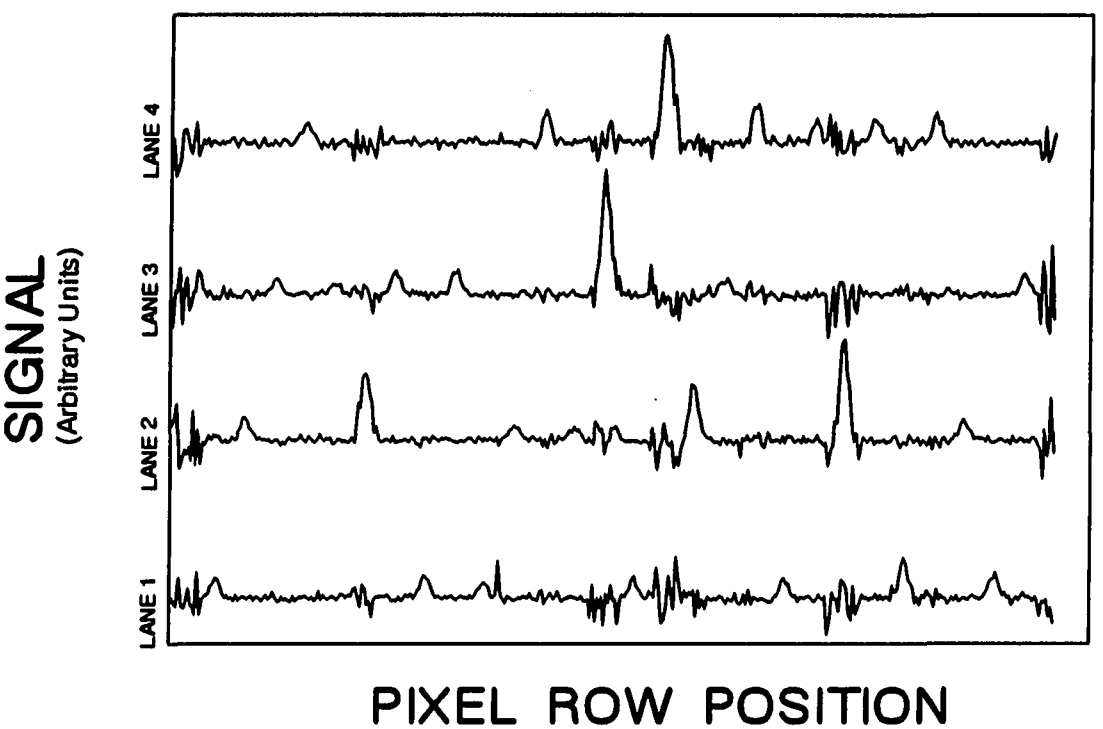

SIGNAL
(Arbitrary Units)
LANE 4
LANE 3
LANE 2
LANE 1
PIXEL ROW POSITION

Figure 6. One-dimensional representation of lanes with selective accumulation.


SIGNAL
(Arbitrary Units)
LANE 4
LANE 3
LANE 2
LANE 1
PIXEL ROW POSITION

requires a few variables such as lane locations, threshold parameters and five one-dimensional arrays, and contains no time-consuming multi-dimensional array calculations.

The threshold chosen for this study was three times the RMS noise, and all bands were located in seven passes with a maximum number of reads for any area of 64. Sixteen of the 28 bands were located in the first pass, and between one and three in each subsequent pass. Table 1 presents a summary of some of the important operating parameters and results for each pass. Bands were located when their cumulative signal to noise ratio ranged from 6.5 to 11.5. The lower value of this range is explained by the fact that five pixels must exceed the threshold which is equal to the noise. Since this is the full width half maximum of a band, and bandshapes are Gaussian, one would expect positive identification of a band when the maximum intensity exceeded two times the threshold, or six times the noise intensity. The upper value is due to the fact that feature recognition and decision making does not occur after each frame selectively accumulated, and may proceed up to two time that required for positive identification, which would result in a signal to noise ratio of 12.

The last row of Table 1 reports the cumulative percentage of the lane areas marked as inactive after each pass. These values can be used to calculate the most important figure of merit of the system, which is the factor by which data

Pass Number	1	2	3	4	5	6	7
Number of Frames Accumulated	1	2	4	8	16	32	64
Threshold (3 times signal to noise ratio)	120	170	240	340	480	679	960
Number of Bands Located	16	3	3	2	1	2	1
Signal to Noise Ratio of Bands Located	6.5 and up	7 to 10	7 to 11	8.5 to 11.5	10	8.5 to 11.5	8
Total percentage of lane areas marked inactive	54	64	74.5	81	84.5	91.5	95

Table 1. Some important operating parameters and results for each pass executed by the program on the sample data set.

acquisition is reduced. Only 806,944 pixels were read versus the 6,144,000 that would have been read if all 64 frames had been accumulated in their entirety, resulting in the factor of 7.6 reduction in data points read. Depending on the distribution of band intensities, this value could be significantly higher or lower, but will never drop below one, as it would be in the unlikely case that all bands were of the same intensity, or when all bands were located in one pass. This factor could be increased by performing analysis after each set of selective accumulations of pixel intensities at the expense of computational load.

Single point densitometers that allow random access would see the full effect of this reduced data acquisition factor, since the entire detection time is spent acquiring data. In CID detection schemes worthwhile benefit would only be seen when the signal is bright enough that much more time is spent reading the pixels than is spent integrating signal.

Another performance figure that must be mentioned is computational overhead time. Presently, while the computer is performing peak location and making decisions, data acquisition cannot proceed. This is another reason why a simple method with minimal variables and data arrays is advantageous. As written in Turbo BASIC and without careful optimization, the program required only 1.2 seconds to execute all but the data acquisition part of the algorithm on a 33-Mhz 80486-based computer operating under DOS 5.0. This value

could easily be significantly improved with more appropriate programming techniques or hardware.

In any automated DNA sequencing procedure, absolute accuracy is required. Since a statistical method is used in this method to determine the cutoff threshold used, precise error rates for the algorithm can be calculated. In Gaussian statistics, the probability of a given noise element exceeding three times the RMS value is one in 770 (23), and the chance of five points in a row exceeding this threshold is one in 2.7 × 10 (14). Since the data actually read contained approximately 800,000 possible combinations of five pixels in a row, and 28 bands were successfully located, the chances of a false peak in this example was 1 in every 924 million peaks found. This error rate is quite acceptable and can be increased or decreased in a predictable manner by altering the cutoff threshold.

Another important aspect of any automated image analysis process is the elimination or reduction of human interaction and judgement. This expert system does quite well in this respect. The only input required from the human operator is the expected dimensions of the bands, and this information is quite easily determined by observing past runs, and should remain constant for a given experimental setup. Beyond requiring this information, the algorithm executes completely independently.

REFERENCES

1. McGregor, D. A.; Yeung, E. S. Anal. Chem. **1992**, 64, 1-5.
2. Komaromy, M.; Goven, H. Nucleic Acid Res. **1984**, 12, 675.
3. Staden, R. Nucleic Acid Res. **1984**, 12, 499.
4. Taylor. P. J. Biomed. & Biophy. Methods **1987**, 14, 71.
5. Sanders, J. Z.; Petterson, A. A.; Hughes, P. J.; Connell, C. R.; Raff, M.; Menchen, S.; Hood, L. E.; Teplow, D. B. Electrophoresis **1989**, 11, 3.
6. Elder, J. K.; Southern, E. M. in: Bishop, M. J.; Rawlings, C. J. (Eds.) Nucleic Acid and Protein Sequence Analysis: A Practical Approach, IRL Press: Oxford, pp. 219-229, 1987.
7. Ehrhardt, W.; Englisch, U.; Neuhorr, V. Electrophoresis **1989**, 10, 265.
8. Elder, J. K. Electrophoresis **1990**, 11, 440.
9. Galat, A. Electrophoresis **1989**, 10, 659.
10. Autran, J. C.; Abbal, P. Electrophoresis **1988**, 9, 205.
11. Elder, J. K.; Green;, D. K.; Southern, E. M. Nucleic Acid Res. **1986**, 14, 417.
12. Chan, K. C.; Koutny, L. B.; Yeung, E. S. Anal. Chem. **1991**, 63, 746.
13. Freeman, S. E.; Larcom, L. L.; Thompson, B. D. Electrophoresis **1990**, 11, 425.

14. Boniszewski, Z. A. M.; Comley, J. S.; Hughes, B.; Read, C. A. Electrophoresis **1990**, 11, 432.

15. Sutherland, J. C.; Sutherland, B. M.; Emrick, A.; Monteleone, D. C.; Ribeiro, E. A.; Trunk, J.; Son, M.; Serwer, P.; Poddar, S. K.; Maniloff, J. BioTechniques **1991**, 10, 492.

16. Gray, A. J.; Beecher, D. E.; Olson, M. V. Nucleic Acid Res. **1984**, 12, 473.

17. Toda, R.; Fujita, T.; Ohashi, M. Electrophoresis **1984**, 5, 42.

18. Prehm, J.; Jungblut, P.; Klose, J. Electrophoresis **1987**, 8, 562.

19. Gray, A. J.; Beecher, D. E.; Olson, M. V. Nucleic Acid Res. **1984**, 12, 473.

20. Baxes, G. A. Digital Processing: A Practical Primer, Prentice Hall: New Jersey, p. 167, 1984.

21. Sweedler, J. V.; Bilhom, R. B.; Epperson, P. M.; Sims, G. R.; Denton, M. B. Anal. Chem. **1988**, 60, 282.

22. Koutny, L. B.; Yeung, E. S. Applied Spectroscopy **1992**, 46, 136.

23. Ingle, J. D.; Crouch, S. R. Spectrochemical Analysis, Prentice Hall: New Jersey, pp. 543-546, 1988.

GENERAL SUMMARY

The research described here is all directed toward the automation of the planar chromatography techniques of thin-layer chromatography and slab gel electrophoresis, which is necessary if the lofty goals of the human genome project are to be met.

An often neglected step toward automation is the simplification of sample preparation, which is often the most significant bottleneck in any analysis. Toward this goal, we have developed instrumental techniques to improve the performance of detection schemes such as indirect fluorescence, native fluorescence and ultraviolet shadowing. Adequate improvement has been demonstrated to allow for the elimination of time consuming sample pretreatment techniques such as chromophore or radioactive labeling in many situations. These schemes have the distinct advantage of relying upon the physical characteristics of the analyte itself, which allows for more accurate quantification. By eliminating chemically intrusive detection, sample recovery is also simplified. To achieve the enhanced detection performance demonstrated it was necessary to take advantage of the unprecidented noise characteristics of the new generation of imagers that CCDs represent.

Another aspect of automation addressed here is the automation of the data interpretation step in the analysis of

slag gels, since traditional human techniques are too slow and error-prone. In this type of post-run analysis, ultrafast computational results are not required, and the relatively simple algorithm presented in Paper 4 executed quite satisfactorily on a fairly inexpensive IBM PC/AT compatible computer.

The last paper in this dissertation presents an example of an expert system, which is one of the most powerful types of automation. The expert system described performs a decision making process that humans are incapable of, due to the computer's ability to accurately interpret incoming data and make decisions on a time scale impossible for humans.

Analytical chemists are in a position to make significant contributions in the process of automating processes such as slab gel electrophoresis due to their broad training in the areas of separation science, instrumentation, and data analysis. The ultimate completion of the automation task will, however, require communication and cooperation amongst a wide array of disciplines including molecular biology, computer science, and electrical engineering.

ADDITIONAL REFERENCES

1. Ma, Y.; Koutny, L. B.; Yeung, E. S. Anal. Chem. **1989**, 61, 1931.
2. Chan, K. C.; Koutny, L. B.; Yeung, E. S. Anal. Chem. **1991**, 63, 746.
3. Koutny, L. B.; Yeung, E. S. Submitted to Anal. Chem. May 1992.
4. Koutny, L. B.; Yeung, E. S. Applied Spectroscopy **1992**, 46, 136.
5. Koutny, L. B.; Yeung, E. S. Submitted to Anal. Chem. May 1992.
6. Denton, M. B.; Lewis, H. A.; Sims, G. R. Charge-Injection and Charge-Coupled Devices in Practical Chemical Analysis. In Multichannel Image Detectors; Talmi, Y., Ed.; ACS Symposium Series 236; American Chemical Society: Washington, D.C., **1983**; Vol. 2, pp. 133-154.
7. Charge-Coupled Devices for Quantitative Electronic Imaging. Photometrics Ltd.: Tucson, **1990**.
8. Michon, G. J.; Burke, H. K.; Brown, D. M. Proc. of a Symposium on "Charge Coupled Device Technology for Scientific Imaging Applications," **1975**, pp. 106-115.
9. User's Manual for CC200 Imaging System. Photometrics: Tucson, **1988**, Version 1.0.
10. Sutherland, J. C.; Sutherland, B. M.; Emrick, A.;

Monteleone, D. C.; Ribeiro, E. A.; Trunk, J.; Son, M.; Serwer, P.; Poddar, S. K.; Maniloff, J. BioTechniques **1991**, 10, 492.

11. Boniszewski, Z. A. M.; Comley, J. S.; Huges, B.; Read, C. A. Electrophoresis **1990**, 11, 432.
12. Campion, A.; Woodruff, W. H. Anal. Chem. **1987**, 59, 1299A.
13. Epperson, P. M.; Sweedler, J. V.; Bilhorn, R. B.; Sims, G. R.; Denton, M. B. Anal. Chem. **1988**, 60, 327A.
14. Freeman, S. E.; Larcom, L. L.; Thompson, B. D. Electrophoresis **1990**, 11, 425.
15. Janesick, J.; Blouke, M. Sky & Telescope, **1987**, 74, 238.
16. Cheng, Y.-F.; Piccard, R. D.; Vo-Dinh, T. Applied Spectroscopy **1990**, 44, 755.
17. Oldham, P. B. Analytical Instrumentation **1990**, 19, 49.
18. McGregor, D. A.; Yeung, E. S. Anal. Chem. **1992**, 64, 1.
19. Sweedler, J. V.; Bilhom, R. B.; Epperson, P. M.; Sims, G. R.; Denton, M. B. Anal. Chem. **1988**, 60, 282.
20. Burt, D. J. "Spectral Response Analysis for Back-Illuminated Devices," EEV Ltd.: London.
21. Sims, G. R.; Griffin, F. Proceedings of SPIE-The International Society for Optical Engineering **1989**, 1071, 31-42.
22. Viehmann, W. Proceedings of SPIE-The International Society for Optical Engineering **1979**, 196, 90.
23. Stuart, F. E.; Haas, A. K. Proceedings of SPIE-The

International Society for Optical Engineering **1982**, 331, 52.

24. Cullum, M.; Deiries, S.; D'Odorico, S.; Reiss, R. Astron. Astrophys. **1985**, 153, L1.
25. Blouke, M. M.; Cowens, M. W.; Hall, J. E.; Westphal, L. A.; Cristensen, A. B. Appl. Opt. **1980**, 19, 3318.
26. Inokuchi, H.; Harada, K.; Kondow, T. J. Opt. Soc. Am. **1964**, 54, 842.
27. Ingle, J. D.; Crouch, S. R. Spectrochemical Analysis, Prentice Hall: New Jersey, p. 141, 1988.
28. Freeman, S. E.; Thompson, B. D. Anal. Biochem. **1990**, 186, 222.
29. Keenan, T. P.; Krawetz, S. A. CABIOS **1988**, 4, 203.
30. "Ultra-Low Dark Current Inverted Mode Devices," EEV Ltd.: London.
31. Burt, D. J. "CCD Performance Limitations: Theory and Practice," EEV Ltd.: London.
32. Epperson, P. M. Optical Engineering **1987**, 26, 715.
33. Wang, Y.; McCreery, R. L. Anal. Chem. **1990**, 62, 2647.

ACKNOWLEDGEMENTS

This work was performed at Ames Laboratory, which is operated by Iowa State University for the U.S. Department of Energy under contract number W-7405-Eng-82.

With this out of the way, I can thank those responsible for steering me toward a career in analytical chemistry. As childhood friends and cohorts, Frank Vasquez and I discovered the wonderful concoctions that could be created with common household chemicals (sorry about the bathroom, Mom).

My formal training started in high school with the enthusiastic and knowledgeable presentation of chemistry by Mr. Gary Rieck. I would like to thank him for the encouragement to pursue chemistry in college.

My appreciation and thanks go to the entire chemistry department at Cornell College, who provided me with a most excellent undergraduate education and experience. A special thanks to Dr. William Deskin for deciding that I would work on an analytical project with Dr. J. Michael Ramsey during my science semester at Oak Ridge National Laboratory. I commend the Great Lakes College Association for funding such a worthwhile program.

Probably the most important decision so far in my career was the choosing of a graduate program. The strongest influences in this decision came from Dr. Ramsey and his group members at ORNL, who suggested that I attend ISU, and in

particularly recommended Dr. Yeung. I have now returned to ORNL as the "new and improved" Lance Koutny.

I must thank Dr. Yeung for his expert guidance and patience throughout my graduate studies. I feel fortunate to have had the opportunity to work for a person with his ability to conceive of and explore novel approaches to scientific challenges.

Everyone I have worked with in Dr. Yeung's research group has earned my respect and gratitude for all they have done for me both professionally and personally. Particular thanks go to Yinfa Ma for helping to orient me to the group and showing patience in our joint effort. King Chan and David McGregor deserve thanks for sharing their combined knowledge of biotechniques.

Thanks to Andy at PC+ for creating my faithful, digital sidekick, Bam Bam, a 386 PC-compatible computer (PC compatibles are the only REAL computers). Bam Bam woke up on time every morning, or night, and never complained despite the extent of the torture my research tried to inflict.

A special thanks goes to William Pfeffer for taking care of me when necessary, which was probably his fault to begin with, and for showing me that chemists didn't have to be boring people.

Parlez vous francais, Tommy?

David Arthur McGregor! If not for you I probably could have finished my graduate work in four, or fewer, years, but I

would not have learned to chili-dip (play excellent golf) or been reintroduced to the honorable sport of billiards. Thanks, and best of luck in the future, you'll need it.

Thank you to all of my family members for their support during my incredibly long academic career. Hang in there Val, you do get to the light at the end of the academic tunnel (and it's not a train).

Last, but definitely not least, I thank my wife and long-time companion, Tracy, for supporting me, financially and emotionally, for the last 12 years. Without her none of this would have been possible or worthwhile. I especially thank her for willingly leaving her job this spring to devote all her time to the awesome task of formatting and typesetting this dissertation within the stringent guidelines of the graduate thesis office; her ten fingers work much better than my two.