

## INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
800-521-0600

UMI<sup>®</sup>



UNIVERSITY OF ALBERTA

**High Sensitivity Protein Sequencing using Fluorescein Isothiocyanate for  
Derivatization on a Miniaturized Sequencer using Capillary Electrophoresis with  
Laser Induced Fluorescence Detection for Product Identification.**

BY

**Ian D. Ireland**



A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment  
of the requirements for the degree of **Doctor of Philosophy**.

DEPARTMENT OF CHEMISTRY

EDMONTON, ALBERTA,

CANADA

SPRING 2000



National Library  
of Canada

Acquisitions and  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada

Acquisitions et  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*

*Our file* *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-59975-2

**Canada**

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Ian D. Ireland


TITLE OF THESIS: High Sensitivity Protein Sequencing using  
Fluorescein Isothiocyanate for Derivatization on a  
Miniaturized Sequencer using Capillary  
Electrophoresis with Laser Induced Fluorescence  
Detection for Product Identification.

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE  
GRANTED: 2000

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights associated with the copyright in this thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written consent.

  
217 Thorncliff Place  
Edmonton, Alberta, Canada  
T5T 0A2

Date Oct 22, 1999

UNIVERSITY OF ALBERTA  
FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommended to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **High Sensitivity Protein Sequencing using Fluorescein Isothiocyanate for Derivatization on a Miniaturized Sequencer using Capillary Electrophoresis with Laser Induced Fluorescence Detection for Product Identification** submitted by **Ian D. Ireland** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.



Dr. N.J. Dovichi



Dr. F.F. Cantwell



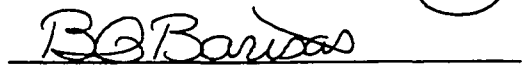
Dr. L. Li



Dr. D.L.J. Clive



Dr. G.D. Armstrong



Dr. B.G. Barisas

Date 22 Oct 99

*For my Parents*

## Abstract

Primary sequence analysis is an important step in determining the structure-function relationship for a given protein or peptide. By far the most popular sequencing method in use today is the Edman degradation. A number of automated Edman sequencers are commercially available. They all incorporate high performance liquid chromatography (HPLC) with ultraviolet (UV) absorbance detection for identification of the phenylthiohydantoin (PTH) amino acid sequencing products. This detection system is suitable for the analysis of proteins at the low picomole level. However, many biologically active proteins can only be isolated at the mid to low femtomole level. This thesis describes the development of a sequencer that has the potential to successfully analyse these rare proteins.

The new sequencer incorporates three major modifications from current instruments. First, sequencing products are identified by capillary electrophoresis with laser induced fluorescence (CE-LIF) detection. This approach allows for unambiguous identification of the fluorescein thiohydantoin (FTH) derivatives of the eighteen coded amino acids studied thus far. Furthermore, mass limits of detection are in the low zeptomole ( $1 \text{ zmol} = 1 \times 10^{-21} \text{ mols}$ ) range for all components in the separation. This sensitivity is seven to eight orders of magnitude better than that of HPLC with UV absorbance detection. The second modification involves a change from phenylisothiocyanate (PITC) to fluorescein isothiocyanate (FITC) as the primary derivatizing agent in the Edman degradation. It is this change that results in formation of the FTH-amino acid sequencing products that can be detected with such exquisite



sensitivity by CE-LIF. Finally, the sequencer has been miniaturized to make it more compatible with the volume requirements of CE. Reagents are delivered by syringe pumps to a reaction chamber that is a fraction of the size of those found on commercial sequencers.

This thesis presents data describing the CE-LIF separation of 18 FTH-amino acids. It then describes the syringe pump sequencer and the modified Edman degradation chemistry. It finishes by presenting sequencing data for five cycles of  $\beta$ -lactoglobulin at the 25 pmol, 10 pmol and 1 pmol levels. The data shows that this proof of principle instrument already has a sensitivity comparable to that of state of the art Edman sequencers.

## Preface

### 16 GENERATIONS OF SCIENCE

<b>PERSON</b>	<b>SCHOOL</b>	<b>DATE</b>
Ian D. Ireland	Alberta	1999
Norman J. Dovichi	Utah	1980
Joel M. Harris	Purdue	1976
Fred E. Lytle	M.I.T.	1968
David M. Hercules	M.I.T.	1957
Lockhart B. Rogers	Princeton	1942
Earle C. Caley	Ohio State	1928
Charles W. Faulk	Leipzig	1890
Wilhelm Ostwald	Dorpat	1878
Carl Schmidt	Giessen	1844
Justus von Liebig	Erlangen	1822
Joseph L. Gay-Lussac	Paris	1800
Claude L. Berthollet	Paris	1778
Jean B.M. Bucquet	Paris	1770
Antoine L. Lavoisier	Paris	1763
Guillaume F. Rouelle	Paris Apothecary School	1725

### Norther Lights Laser Lab

Ian Ireland – 1999  
Rong Jiang – 1998  
Devanand Pinto – 1998  
Sue Bay – 1998  
Karl Voss – 1998  
Darcy Hager – 1997  
H. John Crabtree – 1996  
Yanni Zhang – 1996  
Daniel Figeys – 1995  
Heather Starke – 1994  
JianYing Zhao – 1994  
JianZhong Zhang – 1994  
Karen Waldron – 1993  
DaYong Chen – 1993  
Robert McLaren – 1991  
YungFong Cheng – 1988  
Darryl Bornhop – 1987  
Dean Burgi – 1987  
Tomas Nolan – 1986  
Wayne Weimer – 1985  
Fahima Zarrin - 1985

## Acknowledgments

Obtaining a Ph.D. in any field involves a huge commitment in both time and energy. And, as with any such endeavor, its successful completion requires support, both practical and emotional, from a large number of people. In the next few paragraphs, I would like to express my appreciation to those people who helped me to reach such a meaningful personal milestone.

First, I would like to express my heartfelt thanks to Norm Dovichi, my supervisor. Norm, when I made the decision to join the Northern Lights Laser Lab, I was attracted not only by the wonderful science being done in the group but also by your genuine interest in and concern for the people that you supervise. I know that if I had to do it over again (heaven forbid), I would unhesitatingly make the same choice that I made almost seven (!) years ago.

I would also like to thank all of the other members of the NLLL with whom I had the pleasure of working. There always seemed to be someone in the group who had the expertise to help with any problem that arose. In particular, I would like to thank Xing-Fang Li and Darren Lewis for all of the work they did on the design and construction of the sequencer and Min Chen and Annika Renborg for their help with the FTH-amino acid separation.

There is another group in the Chemistry Department who played a large part in the success of this project. These people work in the machine, glassblowing and electronics shops. They were responsible for making the various sequencer components. Their input went beyond simply constructing the components. They often suggested design modifications that led to significant improvements to the finished instrument.

Some of the people who helped me the most while I was at the University of Alberta were not directly involved in the development of the protein sequencer. Instead, they provided the encouragement and emotional support outside of the lab that kept my spirits up during the bad times and made the good times that much better.

Larissa Powell came to Edmonton with me. For more than five years we made a home together and shared our hopes and dreams. Larissa, things didn't turn out the way we planned, but I know that you will be successful, both at work and at home.

I am grateful to Lorraine and Roland Zimmermann (and the boys) for their unwavering support. They gave me a place to go on those days when I needed to forget my worries and just escape for a while. Lorraine, it was a pleasure working with you and I'm glad that you are back doing something you enjoy.

I am also deeply indebted to the McDermott family – Mark, Christie, Jordan and Quentin. You opened up your home to me (not to mention your fridge and your liquor cabinet) and made me feel like I was part of the family. For that I will always be grateful. I cannot adequately express how much your understanding and encouragement have meant to me.

Finally, I come to my parents and my sister, Vikki – my family. You have always been there for me and you know me the best. It would take pages to describe all that you have done for me. Instead, I will content myself with two heartfelt thoughts: thank you and I love you.

Ian Ireland, Ph.D.

## Table of Contents

1	Introduction.....	1
1.1	Introduction.....	1
1.2	Capillary Electrophoresis .....	2
1.3	Laser Induced Fluorescence Detection.....	21
1.4	Protein Sequencing.....	34
1.5	References.....	56
	Appendix A: L- $\alpha$ -amino acid structures.....	62
2	Identification of Fluorescein Thiohydantoin Amino Acids by Capillary Electrophoresis with Laser Induced Fluorescence Detection as a Potential Method for Improving Sensitivity in Protein Sequencing .....	63
1.1	Introduction.....	63
1.2	Experimental.....	65
1.2.1	Fluorescein Thiohydantoin Amino Acid Synthesis.....	65
1.2.2	Capillary Electrophoresis with Laser Induced Fluorescence Detection .....	66
1.3	Results and Discussion.....	67
1.4	Conclusions.....	90
1.5	References.....	91
3	Development of a Syringe Pump Based Protein Sequencer .....	94
3.1	Introduction.....	94
3.2	Experimental.....	96
3.2.1	Reaction Chamber Design.....	96
3.2.2	Temperature Control Unit.....	97
3.2.3	Reagent Delivery System.....	98
3.3	Results and Discussion.....	98
3.3.1	Reaction Chamber Design.....	99
3.3.2	Temperature Control Unit.....	107
3.3.3	Reagent Delivery System.....	109
3.4	Conclusions.....	116

3.5	References.....	116
4	Evaluation of Fluorescein Isothiocyanate as a Reagent for Modified Edman Protein Sequencing .....	119
4.1	Introduction.....	119
4.2	Experimental.....	121
4.2.1	Materials and Reagents.....	121
4.2.2	Conversion .....	122
4.2.3	Extraction .....	123
4.2.4	Cleavage.....	124
4.2.5	Wash .....	124
4.2.6	Coupling.....	126
4.2.7	Protein Attachment to Solid Supports .....	128
4.3	Results and Discussion.....	129
4.3.1	Conversion .....	129
4.3.2	Extraction .....	134
4.3.3	Cleavage.....	136
4.3.4	Wash .....	139
4.3.5	Coupling.....	141
4.3.6	Protein Attachment to Solid Supports .....	143
4.4	Conclusions.....	147
4.5	References.....	147
5	Fluorescein Isothiocyanate as a Coupling Reagent for High Sensitivity Protein Sequencing in a Syringe Pump Based Instrument with Product Identification by Capillary Electrophoresis with Laser Induced Fluorescence Detection Identification. ....	150
5.1	Introduction.....	150
5.2	Experimental.....	153
5.2.1	Protein Sequencing .....	153
5.2.2	Sample Analysis .....	156
5.3	Results and Discussion.....	156
5.4	Conclusions.....	175
5.5	References.....	176

6	Conclusions and Future Work.....	179
6.1	Introduction.....	179
6.2	Discussion.....	181
6.3	References.....	186

## List of Tables

1.1	Effect of repetitive yield on the overall yield at different cycles. ....	49
2.1	Identification capabilities of the combined separations from Figure 2.7.....	84
2.2	Separation efficiencies for the 18-FTH-amino acids analysed by CE-LIF.....	86
2.3	Limits of detection ( $3\sigma$ ) for 18 FTH-amino acids by CE-LIF. ....	89
4.1	Steps involved in an Edman protein sequencing cycle.....	120
4.2	Procedure used to evaluate wash efficiency.....	125
4.3	Sequencing protocol for Cycle 1 of the peptide standard. ....	127
4.4	FTH-amino acid peak heights at different TFA concentrations. ....	131
4.5	FTH-amino acid peak heights at different conversion times. ....	133
4.6	Scintillation counting measurements from three [ $^{35}\text{S}$ ]-Met labelled DITC-GFM's.....	146
5.1	Typical sequencing protocol.....	155
5.2	Yield for five cycles from 25 pmol of $\beta$ -lactoglobulin. ....	159
5.3	Yield for four cycles from 10 pmol of $\beta$ -lactoglobulin.....	166
5.4	Comparison of individual repetitive yields from 25 pmol and 10 pmol of $\beta$ -lactoglobulin. ....	168



## List of Figures

1.1	Schematic of a capillary electrophoresis instrument. ....	5
1.2	Organization of the electrical double layer in a fused silica capillary. ....	7
1.3	Effect of EOF on separations by CZE. ....	10
1.4	Electronic transition energy level diagram. ....	22
1.5	Schematic of a basic fluorimeter. ....	27
1.6	Schematic of an off column CE-LIF detector using a sheath flow cuvette. ....	30
1.7	Picture of the interior of a sheath flow cuvette. ....	32
1.8	A typical CE-LIF instrument. ....	33
1.9	General approach for the sequential Edman degradation of a protein. ....	39
1.10	Flowchart of the steps in the Edman degradation. ....	41
1.11	Reaction scheme for the Edman degradation. ....	43
1.12	$\beta$ -Elimination of PTH-serine during the Edman degradation. ....	45
1.13	Schematic of an automated protein sequencer. ....	53
1.14	Flowchart of the double coupling Edman chemistry. ....	55
2.1	Comparison of the structures of FTH- Isoleucine and FTH-Leucine. ....	68
2.2	Effect of SDS concentration on the separation of 18 FTH-amino acids. ....	71
2.3	Migration time versus SDS concentration for 18 FTH-amino acids. ....	73
2.4	Effect of magnesium ion concentration on the separation of 18 FTH-amino acids. ....	76
2.5	Effect of pH on the separation of 18 FTH-amino acids. ....	79
2.6	Migration time vs. running buffer pH for 18 FTH-amino acids. ....	81
2.7	Combined separations that allow for the unambiguous identification of 18 FTH-amino acids. ....	83
3.1	Schematic of a capillary based reaction chamber. ....	100
3.2	Capillary based reaction chamber. ....	102
3.3	Schematic of the glass reaction chamber. ....	103
3.4	Glass reaction chamber. ....	104
3.5	Schematic of the Teflon® reaction chamber. ....	105
3.6	Picture of the Teflon® reaction chamber. ....	106
3.7	Schematic of the thermoelectric heating device. ....	108
3.8	Picture of the thermoelectric heating device. ....	109

3.9	Syringe pump schematic. ....	111
3.10	Syringe pump picture. ....	112
3.11	Schematic of the syringe pump sequencer. ....	114
3.12	Picture of the syringe pump sequencer. ....	115
4.1	Effect of TFA concentration on the conversion efficiency for five FTH-amino acids. ....	130
4.2	Effect of reaction time on the conversion efficiency of five FTH-amino acids. ....	132
4.3	Extraction of selected FTH-amino acids by anhydrous TFA. ....	135
4.4	Cleavage of the amino-terminal residue from FITC-canine albumin using anhydrous TFA. ....	137
4.5	Successive TFA cleavage steps on FITC-canine albumin. ....	138
4.6	Evaluation of the wash efficiency of 0.1% cyclohexylamine in methanol. ....	140
4.7	Cycle 1 from a peptide sequencing standard using FITC for coupling. ....	142
4.8	Glass beads after coupling and wash. ....	144
4.9	Comparison of an AP-GFM and a bare GFM after FITC coupling. ....	145
5.1	Electropherograms of the first five cycles from 25 pmol of $\beta$ -lactoglobulin. ....	158
5.2	Yield vs. Cycle Number for cycles 2-5 from 25 pmol of $\beta$ -lactoglobulin. ....	161
5.3	Electropherograms of cycles 2-5 from 10 pmol of $\beta$ -lactoglobulin. ....	165
5.4	Yield vs. Cycle Number for cycles 2-5 from 10 pmol of $\beta$ -lactoglobulin. ....	167
5.5	Yield vs. Cycle Number for cycles 2-5 from 10 pmol of $\beta$ -lactoglobulin. ....	171
5.6	Electropherograms of cycles 2-5 from 1 pmol of $\beta$ -lactoglobulin. ....	173
6.1	Schematic of an air tight valve block for the syringe pump sequencer. ....	182
6.2	Reaction chamber schematic. ....	184

## Symbols and Abbreviations

### The 20 Common Amino Acids:

A	Ala	alanine
R	Arg	arginine
N	Asn	asparagine
D	Asp	aspartic acid
C	Cys	cysteine
E	Glu	glutamic acid
Q	Gln	glutamine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
L	Leu	leucine
K	Lys	lysine
M	Met	methionine
F	Phe	phenylalanine
P	Pro	proline
S	Ser	serine
T	Thr	threonine
W	Trp	tryptophan
Y	Tyr	tyrosine
V	Val	valine

AP-GFM	Aminopropyltriethoxy glass fibre mat
ATZ	Anilinothiazolinone
CE	Capillary electrophoresis
CEC	Capillary electrochromatography
CE-LIF	Capillary electrophoresis with laser induced fluorescence
CE-TOAD	Capillary electrophoresis with thermo-optical absorbance detection
CGE	Capillary gel electrophoresis
CIEF	Capillary isoelectric focussing
CITP	Capillary isotachopheresis
CMC	Critical micelle concentration
CZE	Capillary zone electrophoresis
DITC	<i>p</i> -Phenylenediisothiocyanate
DITC-GFM	Diisothiocyanate glass fibre mat
DMF	Dimethylformamide

DNA	Deoxyribonucleic acid
DPTU	Diphenylthiourea
DPU	Diphenylurea
EOF	Electroosmotic flow
GFM	Glass fibre mat
HPLC	High performance liquid chromatography
FITC	Fluorescein isothiocyanate
FTC	Fluorescein thiocarbamyl
FTH	Fluorescein thiohydantoin
FTZ	Fluorescein thiazolinone
id	Inner diameter
IHP	Inner Helmholtz plane
LIF	Laser induced fluorescence
LOD	Limit(s) of detection
MECC	Micellar electrokinetic capillary chromatography
NMM	N-methylmorpholine
od	Outer diameter
OHP	Outer Helmholtz plane
PITC	Phenylisothiocyanate
PMT	Photomultiplier tube
PTC	Phenylthiocarbamyl
PTH	Phenylthiohydantoin
PVDF	Polyvinylidene difluoride
RY	Repetitive yield
SDS	Sodium dodecyl sulfate
TC	Thiocarbamyl
TED	Thermoelectric device
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
TOAD	Thermooptical absorbance detection
UV	Ultraviolet

## Chapter 1

### Introduction

#### 1.1 Introduction

The role of proteins and peptides in the life processes of all organisms cannot be overstated. If DNA is the blueprint of life, then proteins and peptides are its tools and materials. As such, determining how proteins function, and what causes them to malfunction, is of primary importance in many fields of research. And the first step in determining how a protein functions is to determine how it is put together, i.e. its amino acid sequence. This thesis describes the development of a protein sequencer that has the potential to improve the overall sensitivity of the sequencing process to the femtomole level.

Current state of the art protein sequencers can provide structural information from as little as a few picomoles of sample (1-3). This level of sensitivity has made them extremely useful tools for protein scientists. However, there are many proteins and peptides that are only weakly expressed and cannot be isolated even at the low picomole level. A sequencer that could provide structural information on low femtomole amounts of a sample would open up entirely new fields of research (4). Right now, the limit to sequencer sensitivity is related to the high performance liquid chromatography (HPLC) with ultraviolet (UV) absorbance detection technology used to identify the sequencing products (5). So any improvement in sensitivity will require an improvement in the sequencer detection system. Chapter 2 of this work describes a detection system for

protein sequencing products based on the use of capillary electrophoresis with laser induced fluorescence detection (CE-LIF). This technique shows a six order of magnitude improvement in sensitivity over the HPLC with UV absorbance detection systems currently in use.

There are two main obstacles associated with using CE-LIF for protein sequencing. First, the volume requirements of CE are approximately five orders of magnitude less than the volume requirements of HPLC. This disparity makes it impossible to couple a CE-LIF instrument to a sequencer designed to be used with an HPLC. Any improvement in sensitivity associated with the use of CE-LIF would be lost due to the extremely small percentage of each sample that would be analysed. For this reason, a new sequencer with volume requirements that are more compatible with CE had to be built. The construction of this instrument is described in Chapter 3. The second obstacle to using CE-LIF for sequencing is the fact that LIF detection requires fluorescent analytes and the products from the standard sequencing chemistry are not appreciably fluorescent. In Chapter 4, the optimization of a modified sequencing chemistry that generates fluorescent products is described. Preliminary sequencing results using this chemistry are then presented in Chapter 5.

## **1.2 Capillary Electrophoresis**

Electrophoresis involves the movement of charged species in solution due to the application of an external electric field. It is used extensively in the biological sciences as a means to identify and purify species of interest, often from very complex mixtures.

The utility of electrophoresis as a separation technique stems from the difference in migration rate for analytes that differ in either charge or size. A mixture of such analytes exposed to an electric field will become stratified according to their individual migration rates. The analyte or analytes of interest can then be identified and collected for further study.

The most popular electrophoresis method in use today incorporates a gel matrix as a solid support in which separation of the analytes of interest can take place. Unfortunately, there are some intrinsic properties associated with gel electrophoresis that serve to limit its utility. First, gel supports have a low surface area to volume ratio. Consequently they cannot effectively dissipate the Joule heat generated as ions move in an electric field. For this reason, applied field strengths must be kept low to minimize the bandbroadening that results when a gel overheats. Low field strengths also lead to low analyte velocities, which significantly extend analysis times. Another drawback of gel electrophoresis is that it is not as suitable for quantitative analysis as chromatographic techniques. For this reason, classical gel electrophoresis methods are used primarily for preparative work and for analyses where simple identification of the analyte is sufficient. Quantitative work has traditionally remained the forte of HPLC.

In 1967 Hjerten demonstrated that electrophoresis could be performed in a capillary (6). His work has led to an explosion of research into capillary electrophoresis (CE) (7). The advent of CE solved the problems that limited the utility of gel electrophoresis methods. The small inner diameter capillaries typically used in CE have a very high surface area to volume ratio. As such, they are much better able to dissipate the Joule heat generated during electrophoresis. Better heat dissipation means that higher

field strengths can be used without degrading the separation due to excessive bandbroadening. And higher field strengths lead to decreased analysis times and increased sample throughput. It is also much easier to perform quantitative analyses in a capillary than on a gel. There are a number of different detectors that can be used to identify analyte bands as they migrate through a capillary (8-11). All of these detectors generate a quantifiable signal that can be directly related to the concentration of the analyte. These characteristics make it possible to use CE for a large number of applications that had previously been restricted to HPLC.

Figure 1.1 shows the basic design of a CE instrument. A buffer filled fused silica capillary, typically with a length of between 20 and 100 cm and an inner diameter between 10 and 100  $\mu\text{m}$ , acts as a bridge between two reservoirs. A high voltage is applied at one reservoir while the other is held at ground. Analytes migrate through the capillary from the high voltage to low voltage end with velocities dictated by their individual electrophoretic mobilities. A detector is set at or near the low voltage end of the capillary to monitor the signal generated as the analyte bands pass. A plot of detector signal versus time produces an electropherogram with peaks corresponding to the migration time for each analyte.



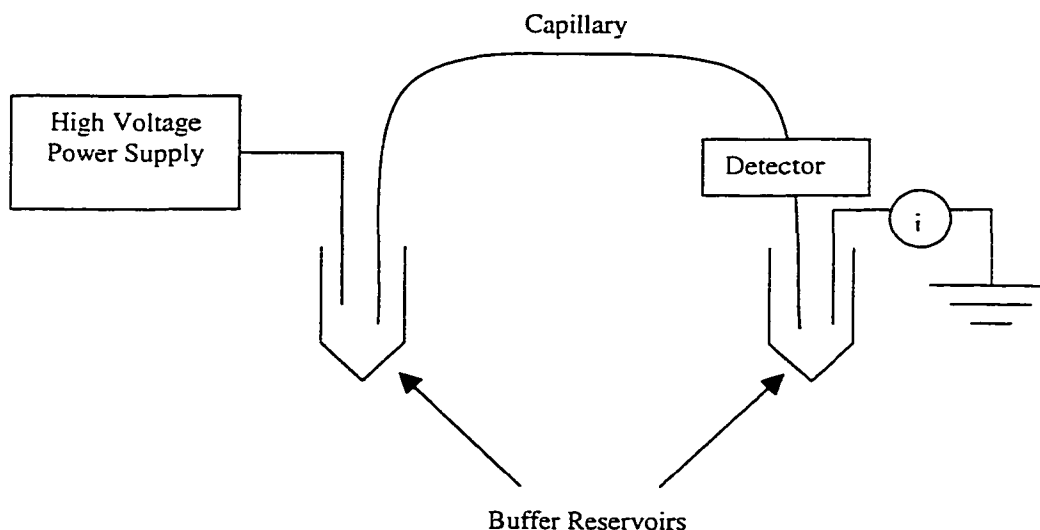


Figure 1.1 Schematic of a capillary electrophoresis instrument.

As with HPLC, a number of different modes of CE can be used to tackle a wide range of analytical problems. They include capillary zone electrophoresis (CZE), micellar electrokinetic capillary chromatography (MECC), capillary gel electrophoresis (CGE), capillary isotachopheresis (CITP), capillary isoelectric focussing (CIEF) and capillary electrochromatography (CEC). These methods are all based on the application of an electric field to a buffer filled capillary; however, each one incorporates a unique modification that results in a unique mechanism of separation. As such, each method is best suited to a different type of application. A detailed discussion of each mode of CE is beyond the scope of this thesis. Instead, this section will focus on CZE, which is at the heart of all of the other applications, and MECC, which is the mode of separation used for the work discussed in later chapters.

In CZE, analytes migrate through an open fused silica capillary according to their individual linear velocities. The electrophoretic velocity,  $v_{el}$ , of an analyte is given by the equation

$$v_{el} = \mu_{el}E \quad (1.1)$$

where  $E$  is the electric field strength and  $\mu_{el}$  is the electrophoretic mobility of the analyte; an intrinsic property that is related to its charge, effective size in solution and the solution viscosity. Electrophoretic mobility increases with increasing analyte charge and decreases with increasing analyte size and solution viscosity.

The dependence of  $\mu_{el}$  on the charge of the analyte means that cations and anions migrate in opposite directions in an applied electric field while neutral species do not move at all. Therefore, if electrophoretic mobility was the only factor that determined the linear velocity a given analyte, then a given CZE separation would be limited to the analysis of either positively charged or negatively charged species. Fortunately, the use of fused silica capillaries in CZE leads to a second factor that influences the overall migration rate of an analyte: electroosmotic flow.

Electroosmotic flow (EOF) describes the movement of bulk electrolyte through a fused silica capillary from the positive end to the negative end. It arises because there is an excess of mobile cations in the electrolyte solution. The presence of excess cations results in a higher net migration of ions toward the anode. The ions carry their hydration shell with them so net migration of the bulk solution toward the anode results. The excess cations come from the ionizable silanol groups on the inner surface of the

capillary. Deprotonation of these silanols creates a stationary negative charge on the capillary wall, which induces the formation of an electrical double layer at the capillary-solution interface. Figure 1.2 shows how this double layer is organized.

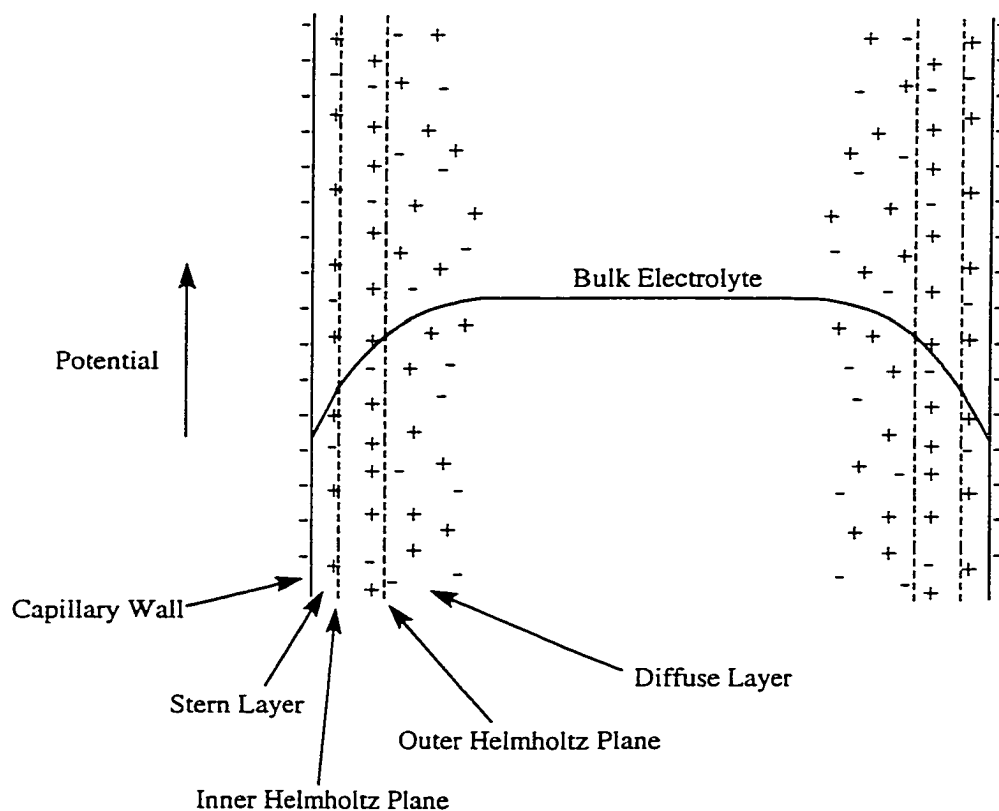


Figure 1.2 Organization of the electrical double layer in a fused silica capillary. The size of the double layer has been greatly exaggerated in this schematic.

The inner, or Stern, layer is composed of stationary ions that have been stripped of their waters of hydration. It extends from the capillary wall to the Inner Helmholtz Plane (IHP) at the locus of the electrical centres of the adsorbed ions. A second layer,

composed of mobile, solvated ions can then form in the region immediately adjacent to the stationary ions. This layer extends from the IHP to the Outer Helmholtz Plane (OHP) at the electrical centres of the solvated ions. Cations inside the OHP only partially counter the negatively charged silanols at the capillary wall. The remaining negative charge is balanced by an excess of solvated positive ions in the diffuse layer between the OHP and the bulk solution. The formation of the electrical double layer produces a potential gradient within the capillary. The shape of the gradient is also shown in Figure 1.2. Potential increases linearly from the capillary wall to the IHP. The gradient then decays exponentially from the IHP through the diffuse layer to the bulk solution.

The electroosmotic linear velocity can be expressed in terms of the solution dielectric constant,  $\epsilon$ , and viscosity,  $\eta$ , the applied electric field strength,  $E$ , and the zeta potential,  $\zeta$  measured at the IHP

$$v_{eo} = \frac{\epsilon\zeta}{4\pi\eta} E \quad (1.2)$$

The viscosity, dielectric constant and zeta potential are all constant for a given electrolyte solution so they can be combined to give the electroosmotic coefficient of mobility,  $\mu_{eo}$

$$\mu_{eo} = \frac{\epsilon\zeta}{4\pi\eta} \quad (1.3)$$

Substitution of equation 1.3 into equation 1.2 gives

$$v_{eo} = \mu_{eo} E \quad (1.4)$$

The overall linear velocity,  $v_t$ , of a given analyte can then be expressed as a combination of its electrophoretic velocity and the EOF velocity of the bulk solution

$$v_t = v_{el} + v_{eo} \quad (1.5)$$

Equation 1.5 can be expressed in terms of mobility coefficients and electric field strength

$$v_t = (\mu_{el} + \mu_{eo}) E \quad (1.6)$$

Electric field strength is a function of applied voltage,  $V$ , and capillary length,  $L$ , so equation 1.6 can be expanded to the form

$$v_t = (\mu_{el} + \mu_{eo}) \frac{V}{L} \quad (1.7)$$

The influence of EOF on separations in CZE is twofold. First, EOF is typically much stronger than a given analyte's electrophoretic mobility. As a result all analytes, regardless of size or charge, are effectively pumped from the high voltage to the low voltage end of the separation capillary. This electrophoretic pump makes it possible to separate cations and anions from each other and from neutral species in one analysis. However, it is important to note that neutral species, with no electrophoretic mobility,

cannot be separated from each other. They will all simply travel at the EOF velocity and reach the detector simultaneously. The effect of EOF on a multicomponent separation is illustrated in Figure 1.3.

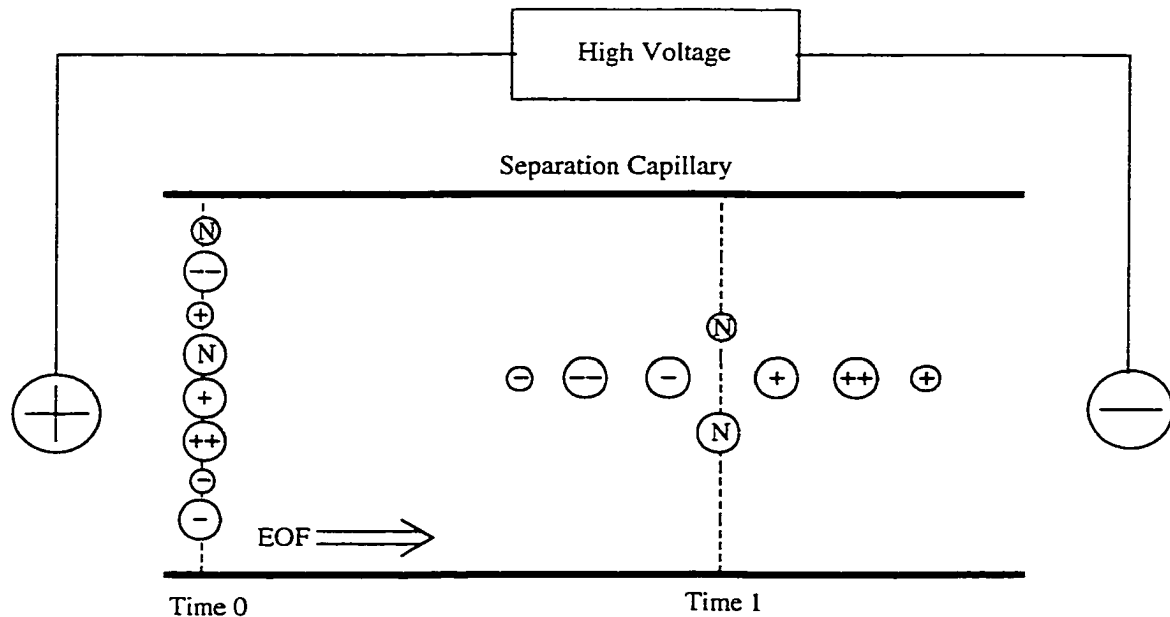


Figure 1.3 Effect of EOF on separations by CZE.

At Time 0, the sample has just been injected and all of the analytes are mixed in the injection plug. A high voltage is then applied across the capillary producing an EOF that begins to pump all of the analyte species through the capillary. At Time 1, the analytes have all migrated toward the negative electrode but they have also separated according to their individual electrophoretic velocities. The neutral species, with  $v_{el} = 0$ , have all moved with the EOF velocity and have not been separated from each other. The electrophoretic velocity of the cations has added to the EOF so they have moved ahead of

the neutrals, with the smaller and more highly charged species being the farthest in front. In contrast, the anions have fallen behind the neutrals because their electrophoretic velocities partially counter the EOF. The end result is a separation in which the cations reach the detector first, followed by a mixture of all of the neutral species and finally the anionic species.

The second influence of EOF is related to the flat flow profile that it produces across the capillary. The EOF velocity is influenced by the potential gradient between the IHP and the diffuse layer. Ions at the IHP are not free to move so the EOF is zero at this boundary. As the potential increases across the OHP and the diffuse layer, so does the EOF. It reaches a maximum velocity at the diffuse layer-bulk solution interface (12). This velocity is then maintained across the rest of the bulk solution. Because the diffuse layer only extends a few nanometers into the solution, EOF is essentially constant across the entire capillary. The constant velocity results in a flat flow profile that contrasts sharply with the pressure induced parabolic flow profile seen with HPLC (13).

Plug flow and the lack of a stationary phase combine to make many of the factors that contribute to bandbroadening in HPLC insignificant in CZE. In fact, the only significant contribution to bandbroadening in CZE is longitudinal diffusion. As a result, separation efficiency in CZE is typically much higher than in HPLC. Efficiencies in excess of 100,000 theoretical plates are common while efficiencies of 3,000,000 (14) and 10,000,000 (15) have been reported.

The effect of longitudinal diffusion on analyte peak shape in CZE can be described by the Stokes-Einstein equation

$$\sigma^2 = 2Dt_m \quad (1.8)$$

where  $\sigma^2$  is the peak variance, a measure of bandbroadening,  $D$  is the analyte diffusion coefficient and  $t_m$  is the analyte migration time. Migration time can be expressed in terms of capillary length,  $L$ , and analyte linear velocity,  $v_t$

$$t_m = \frac{L}{v_t} \quad (1.9)$$

Combining equation 1.9 and equation 1.7 gives an expression for migration time in terms of mobility

$$t_m = \frac{L^2}{(\mu_{el} + \mu_{eo})V} \quad (1.10)$$

which can be substituted into equation 1.8 to give the peak variance in terms of mobility

$$\sigma^2 = \frac{2DL^2}{(\mu_{el} + \mu_{eo})V} \quad (1.11)$$

Equation 1.11 illustrates how peak variance is related to the three experimental parameters that can be modified in CZE: capillary length, applied voltage and electroosmotic mobility.



The relationship between separation efficiency, expressed as the number of theoretical plates,  $N$ , and peak variance can be determined using the rate theory of column bandbroadening. In rate theory, the theoretical plate height,  $H$ , of an analyte peak as it exits the column is defined as the variance of that peak divided by the column length

$$H = \frac{\sigma^2}{L} \quad (1.12)$$

Plate height can also be defined as the column length divided by the number of theoretical plates

$$H = \frac{L}{N} \quad (1.13)$$

Combining equation 1.12 with equation 1.13 gives an expression for efficiency with respect to peak variance

$$N = \frac{L^2}{\sigma^2} \quad (1.14)$$

Substituting equation 1.11 into equation 1.14 then illustrates how column efficiency in CZE is related to the experimental parameters that can be controlled by the analyst

$$N = \frac{L^2}{\left( \frac{2DL^2}{(\mu_{ei} + \mu_{eo})V} \right)} = \frac{(\mu_{ei} + \mu_{eo})V}{2D} \quad (1.15)$$

A number of conclusions can be drawn from equation 1.15. First, separation efficiency is independent of capillary length. This result contrasts sharply with HPLC where efficiency is directly related to the column length. Second, efficiency increases with increasing EOF. However, it will be shown later in this section that increasing the EOF can adversely affect resolution, so this approach can actually be counterproductive. And finally, efficiency increases linearly with voltage. This relationship suggests that the best way to improve efficiency in CZE is to use the highest possible applied voltage. In practice this approach is limited by the increased Joule heating that results from an increase in voltage (16). Failure to dissipate the Joule heat generated during electrophoresis leads to the formation of a temperature gradient in the running buffer. Regions of the running buffer that are at different temperatures will have different viscosities. Because buffer viscosity affects linear velocity, analyte molecules in the different regions will migrate at different rates and bandbroadening will result. Furthermore, increased Joule heating can lead to degassing of the running buffer, which causes an interruption in the electric circuit and halts electrophoresis.

The experimental determination of separation efficiency is carried out using peak variance expressed in time units,  $\tau^2$ , not in distance units as described above. The relationship between the time domain and the distance domain is given in equation 1.16

$$\sigma = \tau v_t \quad (1.16)$$

where  $\sigma$  is the standard deviation of the analyte band in cm,  $\tau$  is the standard deviation in seconds and  $v_t$  is the linear velocity of the analyte in cm/s. Thus, the relationship between the two variances is

$$\sigma^2 = \tau^2 v_t^2 \quad (1.17)$$

or by substituting for  $v_t$  from equation 1.9

$$\sigma^2 = \frac{\tau^2 L^2}{t_m^2} \quad (1.18)$$

Substituting equation 1.18 into equation 1.14 gives the relationship between peak variance in the time domain and efficiency

$$N = \frac{L^2}{\left( \frac{\tau^2 L^2}{t_m^2} \right)} = \frac{t_m^2}{\tau^2} \quad (1.19)$$

For a Gaussian peak, the variance can be determined by measuring the baseline peak width,  $W$ , or alternatively, the peak width at half-maximum,  $W_{0.5}$

$$\tau^2 = \frac{W^2}{16} = \frac{W_{0.5}^2}{5.54} \quad (1.20)$$

Substituting equation 1.20 into equation 1.19 gives

$$N = 16 \left( \frac{t_m}{W} \right)^2 = 5.54 \left( \frac{t_m}{W_{0.5}} \right)^2 \quad (1.21)$$

an equation for separation efficiency in terms of an analyte's migration time and its peak width. Both of these parameters can be easily determined from an electropherogram, thereby making efficiency calculations straightforward.

It was stated earlier in this section that increasing the EOF can have a detrimental effect on separations in CZE even though it leads to improved efficiencies. The reason why increasing EOF is not necessarily desirable can be understood by looking at the factors that effect resolution in CZE. Resolution,  $R$ , is a measure of how well adjacent peaks in a separation are separated from each other. It can be expressed empirically by the equation

$$R = \frac{t_{m,b} - t_{m,a}}{\frac{1}{2}(W_a + W_b)} \quad (1.22)$$

where  $t_{m,a}$  and  $t_{m,b}$  are the migration times and  $W_a$  and  $W_b$  are the baseline peak widths of the two analytes. Resolution can also be expressed in terms of analyte mobilities (16)

$$R = \frac{N^{0.5}}{4} \left( \frac{\mu_b - \mu_a}{\bar{\mu} + \mu_{eo}} \right) \quad (1.23)$$

where  $\mu_b - \mu_a$  is the difference in electrophoretic mobility between the two analytes and  $\bar{\mu}$  is their average electrophoretic mobility. Combining equation 1.15 and equation 1.23 gives

$$R = \frac{1}{4} \left( \frac{(\bar{\mu} + \mu_{eo})V}{2D} \right)^{0.5} \left( \frac{\mu_b - \mu_a}{\bar{\mu} + \mu_{eo}} \right) = 0.177(\mu_b - \mu_a) \left( \frac{V}{D(\bar{\mu} + \mu_{eo})} \right)^{0.5} \quad (1.24)$$

Equation 1.24 shows that resolution improves as the difference in mobility between the analytes increases. It also shows that when the EOF travels in the same direction as the analytes, it causes a decrease in resolution. In effect, the analytes will have less time to separate before being swept out of the capillary. If the EOF opposes the mean analyte mobility, resolution will improve as  $\mu_{eo}$  approaches  $-\bar{\mu}$ . This approach is useful for the separation of species having very similar electrophoretic mobilities, although it does lead to long analysis times (16, 17).

The equation for resolution in CZE (1.24) also illustrates the one drawback of this technique: there has to be a finite difference between the electrophoretic mobilities of two analytes, i.e.  $\mu_b - \mu_a \neq 0$ , before any resolution can be achieved. For all neutral species however,  $\mu_b = \mu_a = 0$  and resolution goes to zero. This problem was overcome by Terabe in 1984 through the addition of a surfactant to the running buffer (18). Surfactants, like sodium dodecyl sulfate (SDS), are ionic molecules that also have a

significant hydrophobic chain. At low concentrations the surfactant molecules remain individually solvated in solution. However, when they are present in solution above the critical micelle concentration (CMC), they form aggregates, or micelles of between 30 and 100 molecules. The addition of micelles to the running buffer in CE has become known as micellar electrokinetic capillary chromatography (MECC). It is a popular method for the analysis of small molecules because it allows for the separation of not just anions and cations but also neutral species by CE.

All of the surfactant molecules in a micelle are aligned with their charges at the surface and the hydrophobic chains in the interior. This orientation results in the formation of a discrete hydrophobic phase in the separation capillary. Analyte molecules will then partition between this phase and the hydrophilic bulk solution according to their own hydrophobicity. This differential partitioning is the same separation mechanism that is employed in reverse phase HPLC. However, partitioning of analytes between the micelles and the running buffer does not, on its own, result in separation. There must also be a difference in migration velocity between the two phases. Fortunately, the charged outer surface of the micelle gives it a significant electrophoretic velocity. Thus, the linear velocity of an analyte can be expressed as a combination of the electrophoretic velocity of the micelles,  $v_{mc}$ , and the EOF velocity (19)

$$v_t = N_{aq} v_{eo} + N_{mc} v_{mc} \quad (1.25)$$

where  $N_{aq}$  is the ratio of analyte molecules in the running buffer and  $N_{mc}$  is the ratio of analyte molecules in the micelles. These ratios determine the relative contribution of

each term in equation 1.25 to the overall analyte linear velocity. They can be expressed in terms of the number of analyte molecules in the aqueous phase,  $n_{aq}$ , and the number of molecules in the micellar phase,  $n_{mc}$

$$N_{aq} = \frac{n_{aq}}{n_{aq} + n_{mc}} \quad N_{mc} = \frac{n_{mc}}{n_{aq} + n_{mc}} \quad (1.26)$$

Equation 1.25 then becomes

$$v_t = \left( \frac{n_{aq}}{n_{aq} + n_{mc}} \right) v_{eo} + \left( \frac{n_{mc}}{n_{aq} + n_{mc}} \right) v_{mc} \quad (1.27)$$

In the limit of  $v_{mc} = 0$ , the second term goes to zero and equation 1.27 reverts to the linear velocity expression for an analyte in classical chromatography. Thus, MECC can be thought of as voltage pumped chromatography in which the micelles act as a pseudostationary phase. Other concepts from classical chromatography can also be used to describe analyte migration in MECC. The capacity factor,  $k$ , can be expressed as the ratio of analyte molecules in the micelles to analyte molecules in the running buffer

$$k = \frac{n_{mc}}{n_{aq}} \quad (1.28)$$

Substituting equation 1.28 into equation 1.27 gives

$$v_t = \left( \frac{1}{1+k} \right) v_{eo} + \left( \frac{k}{1+k} \right) v_{mc} \quad (1.29)$$

As with CZE, the linear velocities in the above equation can all be expressed in terms of the capillary length and migration time

$$v_t = \frac{L}{t_m} \quad v_{eo} = \frac{L}{t_{eo}} \quad v_{mc} = \frac{L}{t_{mc}} \quad (1.30)$$

where  $t_m$  is the analyte migration time,  $t_{eo}$  is the migration time of an unretained neutral molecule and  $t_{mc}$  is the migration time of the micelles. Substituting these equations into equation 1.29 and solving for  $t_m$  gives an expression for the migration time of a neutral species in MECC.

$$t_m = \frac{(1+k)t_{eo}}{1 + \left( \frac{t_{eo}}{t_{mc}} \right) k} \quad (1.31)$$

Combining equation 1.31 with 1.21 and 1.22 then gives the expression for the resolution of neutral analytes in MECC (19)

$$R = \left( \frac{N^{0.5}}{4} \right) \left( \frac{\alpha - 1}{\alpha} \right) \left( \frac{k_b}{k_b + 1} \right) \left( \frac{1 - \frac{t_{eo}}{t_{mc}}}{1 + \left( \frac{t_{eo}}{t_{mc}} \right) k_a} \right) \quad (1.32)$$



where  $\alpha$  is the selectivity factor defined as  $k_b/k_a$ . The resolution equation in MECC is very similar to that of standard chromatography. The only difference is the extra term related to the micelle migration time.

In recent years the number of applications of MECC has expanded dramatically. The ability of MECC to separate both ionic and neutral species has made it a valuable analytical tool in areas ranging from environmental, forensic and clinical analysis (20-25) to the analysis of foods and pharmaceuticals (26-30). These assays involved the separation and identification of both neutral and ionic species using a number of different surfactants under different running conditions. Chapter 2 of this thesis presents a separation of 18 Fluorescein thiohydantion amino acids by MECC. This separation is then used in Chapter 4 and Chapter 5 to evaluate the effectiveness of using fluorescein isothiocyanate as a reagent for high sensitivity protein sequencing.

### **1.3 Laser Induced Fluorescence Detection**

The relaxation of a molecule from its excited singlet electronic state back to its ground state through the emission of a photon of light is called fluorescence. The emitted light can be collected to give both qualitative and quantitative information about the sample. Fluorometric methods, although limited to a rather small set of suitable compounds, have found uses in almost all areas of the natural sciences. Their popularity results from the high sensitivity, wide dynamic range and selectivity associated with fluorescence detection. Chapter 2 of this thesis describes how laser induced fluorescence

detection was used in conjunction with capillary electrophoresis to identify protein sequencing products at the low zeptomole ( $1 \text{ zmol} = 1 \times 10^{-21}$  moles) level.

The various processes involved in generating fluorescence are illustrated in the electronic transition energy level diagram below (Figure 1.4).

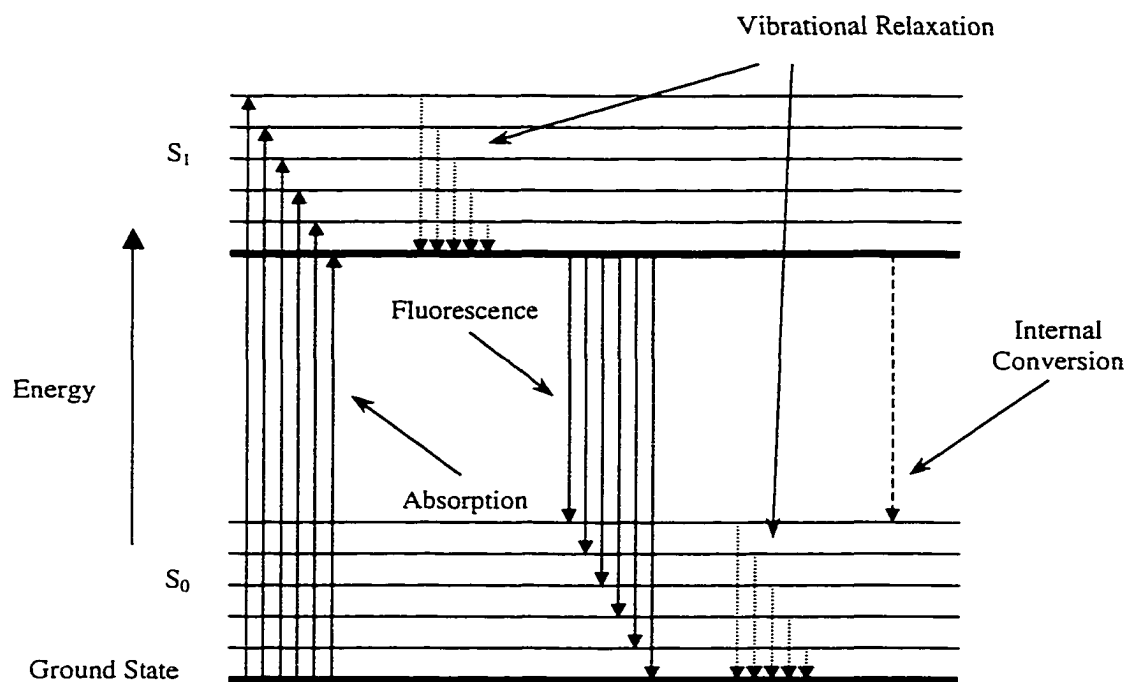


Figure 1.4 Electronic transition energy level diagram.

First, the molecules absorb light having the same energy as the transition from the ground state,  $S_0$ , to one of the vibrational levels of the first excited electronic state,  $S_1$ . The molecules rapidly relax from the excited vibrational levels to the ground vibrational level of the excited state. Fluorescence then occurs when the molecules relax back to one of the excited vibrational levels of the ground electronic state by emitting an appropriate wavelength of light. Finally, vibrational relaxation returns the molecules to the ground vibrational level of the ground electronic state.

The loss of energy to vibrational relaxation after both excitation and emission leads to an emission transition that has a lower energy than the original excitation transition. This phenomenon, called the Stokes shift, is seen empirically as a shift in the emission spectrum to higher wavelengths than the excitation spectrum. The difference in wavelength between excitation and emission is beneficial because it makes it easier to distinguish the fluorescent signal from the overwhelming signal of the excitation source.

There are a number of deactivation processes that compete with fluorescence. They include both internal and external conversion, intersystem crossing, phosphorescence and photodecomposition. Internal conversion describes any intramolecular process that leads to non-radiative relaxation of the molecule. For example, overlap between the vibrational energy levels of the excited and ground electronic states can lead to the transition from a vibrational level of the excited electronic state to an equally energetic excited vibrational level of the ground state. Vibrational relaxation can then return the molecule to the ground state without the emission of any light. Internal conversion is typically quite fast. In fact, for most species it is the predominate means by which relaxation of the excited molecules occurs.

External conversion involves transfer of the absorbed energy to solvent molecules, which can then release the energy as heat. Intersystem crossing occurs when the spin state of an excited electron is flipped. The change in spin is accompanied by a transition from the excited singlet state to an excited triplet state. Relaxation from the excited triplet state to the ground singlet state can then occur by internal conversion, external conversion or by phosphorescence. As with fluorescence, phosphorescence involves radiative relaxation of the molecule back to the ground state. However, phosphorescence specifically describes the transition from an excited triplet state to the ground singlet state. Absorption of energy by a molecule can also lead to photodecomposition. This process, also known as photobleaching, does not involve a transition back to the ground state, but rather results in destruction of the molecule.

The extent that each of the above mechanisms contributes to the overall relaxation process depends on the relative rate at which each occurs. This relative rate is expressed as quantum efficiency,  $\Phi$ , which is defined as the rate of the decay process in question divided by the sum of the rates of all of the decay processes. For example, the quantum efficiency of fluorescence,  $\Phi_f$ , can be expressed as

$$\Phi_f = \frac{k_f}{k_f + k_{ic} + k_{ec} + k_i + k_p + k_{pd}} \quad (1.33)$$

where  $k_f$  is the rate constant for fluorescence and the other  $k$  values are the rate constants of the other deactivation processes.

For most compounds, the rate of the non-radiative deactivation processes, especially internal conversion, is much higher than that of fluorescence. Consequently

the vast majority of molecular species have a vanishingly small fluorescent quantum efficiency and are unsuited to fluorometric methods. Those species that do fluoresce are typically aromatic compounds having a low energy  $\pi$  to  $\pi^*$  transition. This type of transition has a higher molar absorptivity and hence, a shorter lifetime than that of other electronic transitions. As a result, the rate of fluorescence is enhanced. Aromatic compounds also tend to have a fairly rigid structure due to the extensive  $\pi$  bonded ring system. The rigid structure leads to a decrease in the rate of internal conversion. The combination of increased rate of fluorescence and decreased rate of internal conversion results in an increased fluorescent quantum efficiency.

Fluorescent signal can be collected and used for concentration determination of a good fluorophore. Of course, quantitative fluorescence detection requires a linear relationship between the concentration of the analyte and the fluorescent signal that it generates. Fluorescence,  $F$ , can be related to the power of the incident light according to the equation:

$$F = k'(P_0 - P) \quad (1.34)$$

where  $k'$  is a constant related to the quantum yield of the analyte,  $P_0$  is the power of the beam incident to the sample and  $P$  is the power of the beam after it has traveled a distance  $b$  through the sample. Hence,  $(P_0 - P)$  represents the amount of radiant energy absorbed by the sample. Absorbance can be related to analyte concentration according to Beer's law

$$\frac{P}{P_0} = 10^{-\epsilon bc} \quad (1.35)$$

where  $\epsilon$  is the molar absorptivity of the analyte,  $b$  is the absorbance pathlength and  $c$  is the analyte concentration. Combining equations 1.34 and 1.35 and solving for  $F$  gives

$$F = k'P_0(1 - 10^{-\epsilon bc}) \quad (1.36)$$

Equation 1.36 describes a Taylor series and can be expanded to the form

$$F = k'P_0 \left[ 2.3\epsilon bc - \frac{(2.3\epsilon bc)^2}{2!} + \frac{(2.3\epsilon bc)^3}{3!} \dots \right] \quad (1.37)$$

For conditions where  $\epsilon bc < 0.05$ , the contribution of the higher order terms of the expansion becomes negligible and the equation can be simplified to

$$F = 2.3k'P_0\epsilon bc \quad (1.38)$$

and fluorescence varies linearly with analyte concentration. The other significant implication of equation 1.39 is that fluorescent signal is also directly related to the intensity of the incident radiation. Therefore, sensitivity can be improved by simply increasing the radiant power of the excitation source. This feature of fluorescence

detection results in limits of detection for a good fluorophore that are many orders of magnitude better than those of UV absorbance detection.

Instruments for fluorescent detection, both quantitative and qualitative, all have a similar basic design. This design is illustrated in Figure 1.5.

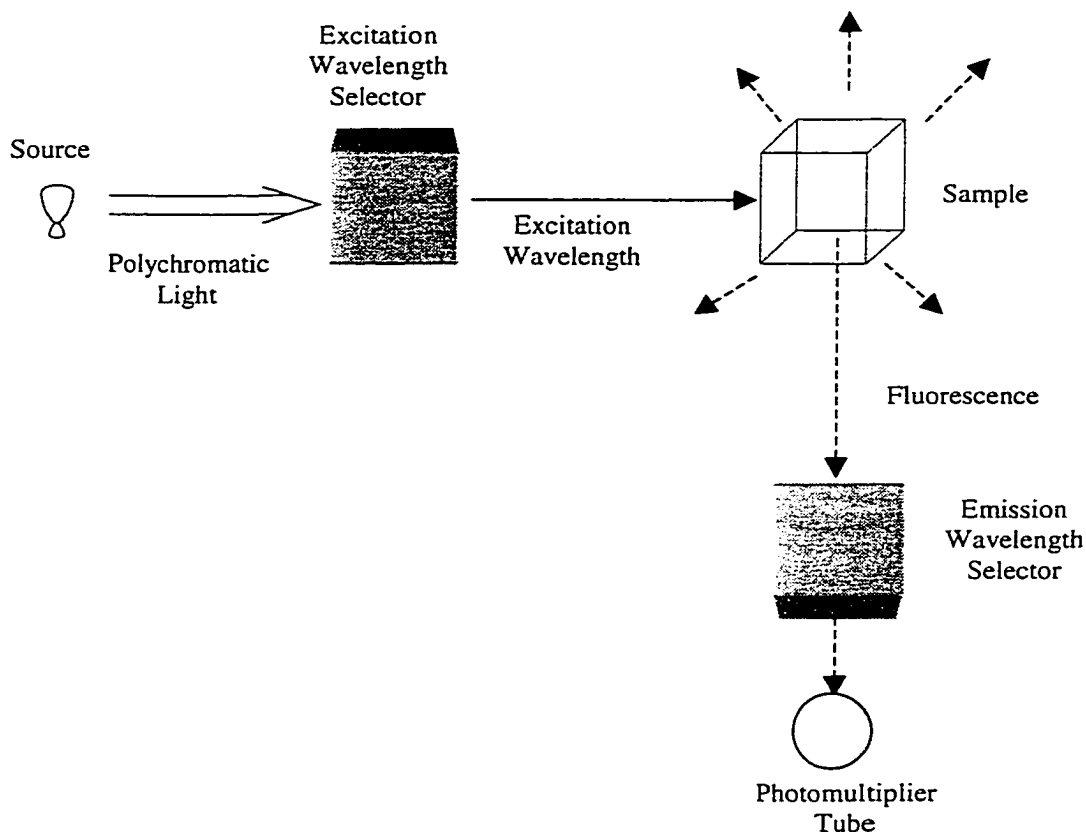


Figure 1.5 Schematic of a basic fluorimeter.

A monochromator or an optical filter is used to select an appropriate excitation wavelength from the polychromatic source, most commonly a xenon arc lamp. Other wavelengths are prevented from reaching the sample because they would interfere with

detection of the fluorescent signal. The excitation beam travels to the sample cell where a small percentage is absorbed by the analyte. A significant amount of the excitation beam is also scattered by the walls of the sample cell or by the sample solution itself. In order to minimize the noise generated by this scattering, fluorescence is collected at 90° to the incident excitation beam. Noise is further reduced by using a second monochromator or optical filter set so that only light at the desired emission wavelength passes through to the detector. Detection is performed using a photomultiplier tube (PMT), which produces an electronic signal proportional to the amount of light that reaches it. The entire analysis is performed in a light shielded housing to limit background noise from room lighting.

While most fluorimeters use a continuous light source to excite the sample, it is also possible to use light from a laser. Using a laser for excitation has a number of implications for fluorescence detection. First, no excitation filter or monochromator is necessary because the laser emits a coherent beam of monochromatic radiation. However, detection flexibility is reduced because only those molecules that absorb significantly at the wavelength emitted by the laser can be analysed. Second, lasers are often more powerful than a continuous light source. Because fluorescence is directly related to the power of the excitation beam, improved sensitivity results from laser induced fluorescence (LIF). A laser beam can also be focussed into a very small spot using a standard microscope objective. This particular aspect of LIF detection makes it ideal for use with CE. The typical inner diameter of a separation capillary is between 10  $\mu\text{m}$  and 100  $\mu\text{m}$ . As a result, detection volumes range from the low nanolitre to sub-picolitre level. Careful positioning of the focussed laser beam allows for efficient



excitation of analyte molecules even at these low sample volumes (31-34). Detection of the fluorescent signal can then be carried out using a second microscope objective positioned orthogonally to the incident laser beam.

The simplest design of an LIF detector for CE would involve on-column excitation of the analyte and collection of the resulting fluorescence. However, this approach is somewhat problematic because both the incident laser beam and the fluorescent signal undergo a significant amount of scattering as they pass through the curved capillary walls. The resulting increase in noise and decrease in analyte signal adversely affects detector sensitivity. This problem can be reduced by carrying out detection off-column in a sheath flow cuvette. Figure 1.6 shows a basic schematic for an off-column LIF detector.

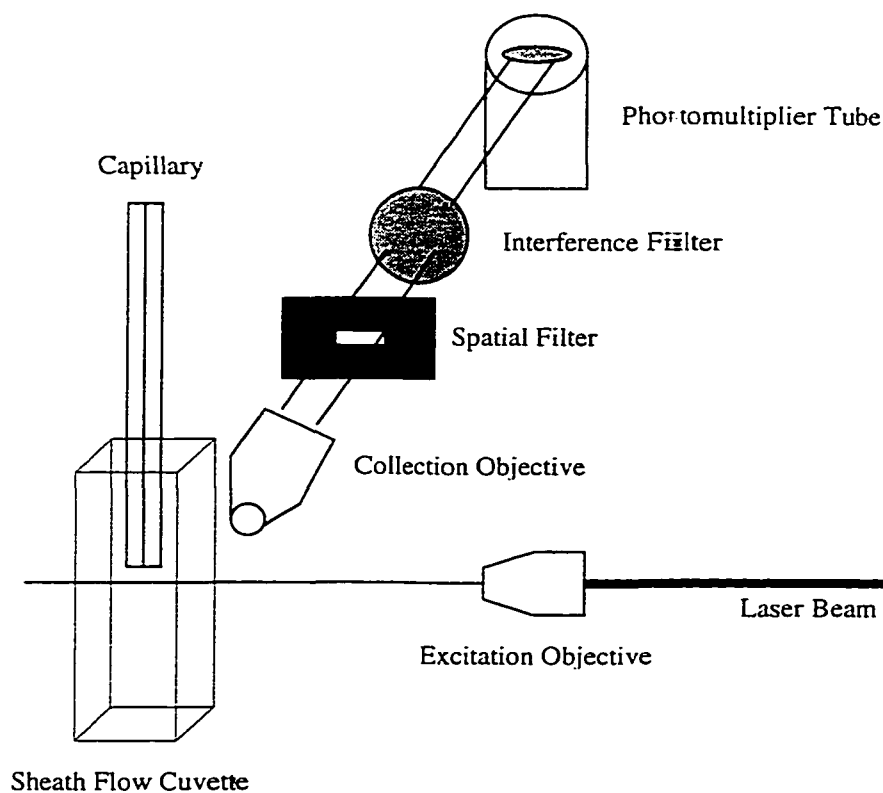


Figure 1.6 Schematic of an off column CE-LIF detector using a sheath flow cuvette.

The cuvette is a rectangular piece of quartz or glass containing a rectangular inner channel. Its smooth, flat walls allow for transmission of a perpendicular laser beam with a minimum of scatter. The outlet of the separation capillary is inserted into the cuvette with the end sitting just above the path of the laser beam. Sheath buffer, having the same composition as the running buffer, is then allowed to flow through the top of the cuvette, past the end of the capillary and out the bottom of the cuvette to waste.

The purpose of the sheath buffer is twofold. First, it provides the electrical contact that completes the electrophoretic circuit, with the metal cuvette holder acting as the electrode. Second, it carries the eluent away from the end of the capillary. In so doing, it prevents buildup of analyte in the detection region during the analysis. It also serves to maintain the integrity of the eluent stream for a short distance past the end of the capillary. In a static system, the eluent would fan out immediately after exiting the capillary. However in the cuvette, the sheath buffer is constantly flowing past the capillary. And when it reaches the end of the capillary, it pushes in toward the centre of the cuvette channel. Consequently, the tendency of the eluent stream to spread out as it exits the capillary is counteracted. It is this hydrodynamic focussing that allows for off-column LIF detection. The laser beam can be positioned so that it intersects the focussed eluent stream at a point just below the end of the capillary. Because the beam only travels through the smooth cuvette walls, scatter is minimized and sensitivity is enhanced. Figure 1.7 shows a picture of the interior of a sheath flow cuvette. The separation capillary is sitting slightly off centre in the channel. The green spot below the capillary tip is the fluorescent signal resulting from excitation of a  $1 \times 10^{-9}$  M fluorescein solution being electrophoretically pumped through the capillary. Although the spot is wider than the inner diameter of the capillary, it shows that the eluent stream remains cohesive enough to allow for off column detection.

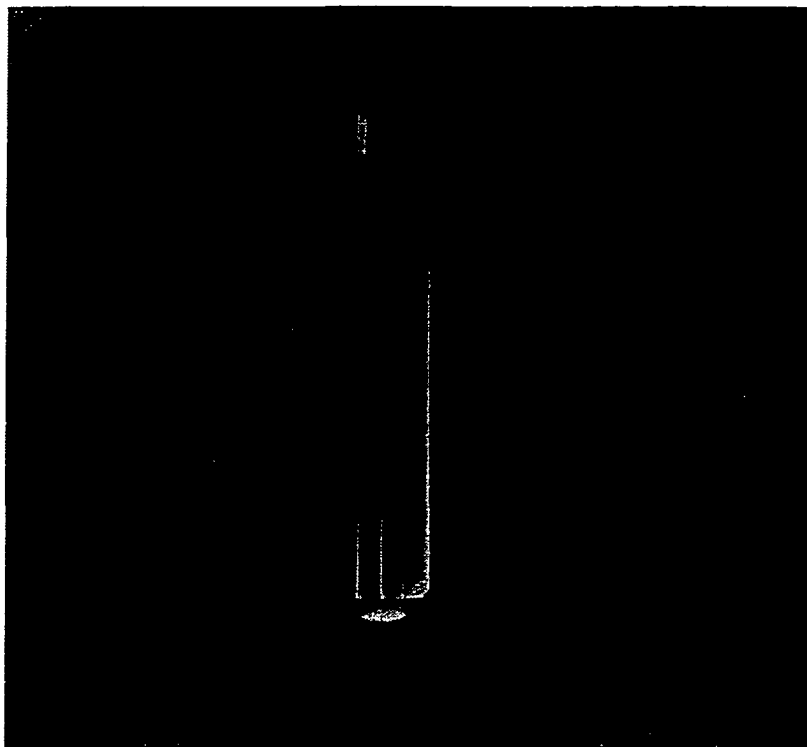


Figure 1.7 Picture of the interior of a sheath flow cuvette.

The sensitivity of CE-LIF is truly extraordinary. Unlike UV absorbance, LIF has no path length dependence. Therefore, the small detection volume of CE has no adverse effect on the analyte signal. In fact, the combination of small sample volume and high signal level leads to unparalleled mass limits of detection (LODs). This group routinely obtains mass LODs in the low attomole ( $1 \times 10^{-18}$  moles) to zeptomole ( $1 \times 10^{-21}$  moles) range, while both yoctomole ( $1 \times 10^{-24}$  moles) limits and even single molecule detection have been reported (35-40). A picture of a typical CE-LIF instrument is shown in Figure 1.8.

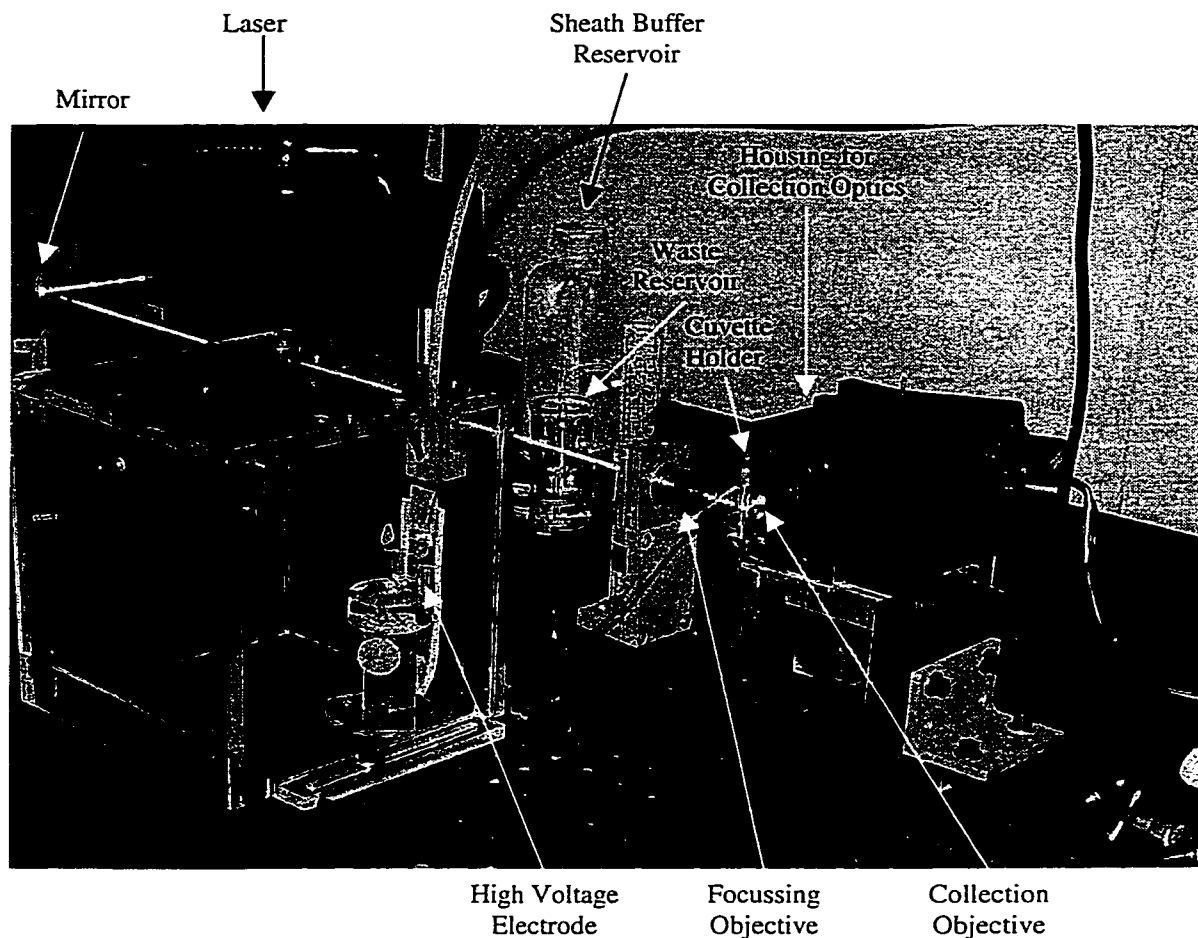


Figure 1.8 A typical CE-LIF instrument.

A mirror is used to direct the laser beam to the focussing objective. The beam then travels to the cuvette where it excites the analyte. Fluorescence is collected by the microscope objective mounted in the housing behind the cuvette holder. This housing also holds a spatial filter, an interference filter and a photomultiplier tube (PMT). Linear translation stages and adjustable mirror mounts are used to align the various components. The Plexiglas box in the foreground of the picture houses the injection end of the

capillary and isolates the high voltage electrode. The plastic squeeze bottle in the background acts as the sheath buffer reservoir; while the glass apparatus immediately in front of the squeeze bottle acts as a constant level waste reservoir.

Much of the work described in this thesis was performed using a CE-LIF instrument with the same basic design as the one described above. Chapter 2 describes the development of a CE-LIF separation for 18 different fluorescein thiohydantion amino acid derivatives. In Chapter 4, CE-LIF is used to evaluate how fluorescein isothiocyanate affects each step of the standard protein sequencing chemistry. And in Chapter 5 it is used to analyze fluorescent products generated from a novel protein sequencer.

#### **1.4 Protein Sequencing**

The successful determination of the amino acid sequence of insulin by Frederick Sanger in 1953 unequivocally demonstrated that proteins had a definite chemical structure. This work, for which Sanger received his first Nobel Prize in 1955, paved the way for the incredible explosion in protein research that has occurred over the last half century. And to this day, protein sequence analysis retains its importance as a means for understanding the structure-function relationship of biologically active proteins.

Proteins are a class of biopolymers responsible for regulating many of the biological functions in all living organisms. They are composed of linear chains of amino acids linked by the formation of a peptide bond between of the amine group of one amino acid to the carboxylic acid group of another. All proteins are made up from the set of twenty L- $\alpha$ -amino acids specified by the genetic code (see Appendix A). The specific

sequence of amino acid residues for a given protein is known as its primary structure. It is this structure that determines the three dimensional shape that gives a protein its unique function. Therefore, one of the first steps to understanding how a protein works, or, in the case of many diseases, why it doesn't work, is to determine its amino acid sequence.

Over the past fifty years, a number of different techniques have been developed to elucidate the primary structure of proteins and peptides. These methods can be divided into four main categories, each with its own strengths and weaknesses

1. Sequential degradation from the amino-terminal residue.
2. Sequential degradation from the carboxy-terminal residue.
3. DNA sequencing of the gene that codes for the protein.
4. Fragmentation and analysis of the protein by mass spectrometry.

Sequential degradation of a protein from its amino-terminus (N-terminus) was the first successful sequencing approach (41). Although the chemistry has changed over the years the general approach remains the same: the N-terminal amino acid is derivatised, isolated from the rest of the protein and analysed. The entire procedure is then repeated on the truncated protein. Because the protein of interest is analysed directly, the resulting sequence information is unambiguous. The lack of ambiguity in the sequencing results is a major advantage for this technique. The drawbacks of this technique are primarily related to the significant analysis time required to isolate and identify each residue and the lack of sensitivity associated with residue identification.

Protein sequencing from the carboxy-terminus (C-terminus) (42-46) follows the same general logic as sequencing from the N-terminus. As a result, the two approaches typically share the same advantages and suffer from the same disadvantages. However, the chemistry of C-terminal sequencing has not proven to be as reliable as that of N-terminal sequencing. Therefore, the everyday use of C-terminal sequencing has not, as yet, become commonplace in protein research.

Sequencing the gene that codes for a given protein is an elegant way to determine its primary structure. The advantage of this approach comes from the speed and sensitivity of DNA sequencing. Once the DNA sequence is known, the amino acid sequence can be determined from the codons that make up the gene (47). The disadvantage to this approach is that the location of the appropriate gene must be determined. Finding the gene requires an oligonucleotide probe that is specific to the protein of interest (3). However, creating the probe requires prior knowledge of at least part of the protein's primary structure. As well, many proteins undergo posttranslational modifications that cannot be inferred from the DNA sequence. Such sites can only be found by sequencing the protein itself.

Direct mass analysis of proteins is a relatively new, but rapidly expanding field (48, 49). It takes advantage of the high analysis speed and sensitivity that are hallmarks of mass spectrometry. However, this approach is currently most useful for identifying a protein by comparing its mass spectrum to the spectra contained in an appropriate database (50, 51). De novo sequencing by mass spectrometry has not become popular due to the difficulty in interpreting the complex spectra that result even from small peptides (52).



Gene reading, C-terminal sequencing and direct mass analysis are all valuable tools for protein structure research, but they each have a drawback that currently limits their use for de novo sequencing. Elucidation of the complete structure of novel proteins and peptides remains the forte of N-terminal degradation methods. Therefore, the remainder of this chapter will focus on the chemistry, instrumentation and current limitations associated with N-terminal protein sequence analysis.

The approach Sanger used to sequence insulin involved coupling the N-terminal amino acid with 2,4-dinitrofluorobenzene (41). The derivatised amino acid was identified colorimetrically after hydrolysis of the protein. It took him almost 10 years to successfully sequence the entire 51 amino acid insulin chain using this method. The reason the process took so long is because the protein had to be completely hydrolysed after each coupling step. As a result, determination of each successive residue required the preparation of fresh sample that had been cleaved in such a way that it was one residue shorter than the last sample to be analysed. Only then could the next residue be successfully labelled and analysed. Most of the 10 years that it took to sequence insulin were spent not on the sequencing reaction itself, but rather on the chemistry associated with selectively cleaving and isolating the peptide fragment required for successive cycles. The extreme amount of labour and the large amount of sample required to successfully sequence even a small protein using Sanger's approach, make it unsuitable as a method for high throughput protein sequencing.

In 1950, Per Edman proposed an alternate method for protein sequencing based on the use of isothiocyanate chemistry (53). This method had the advantage that the derivatised N-terminal amino acid could be selectively cleaved without destroying the

rest of the protein. The truncated protein still retained an active N-terminal site, so the entire process could be repeated. Each successive cycle would lead to the identification of the next residue in the protein. This process is portrayed in Figure 1.9.

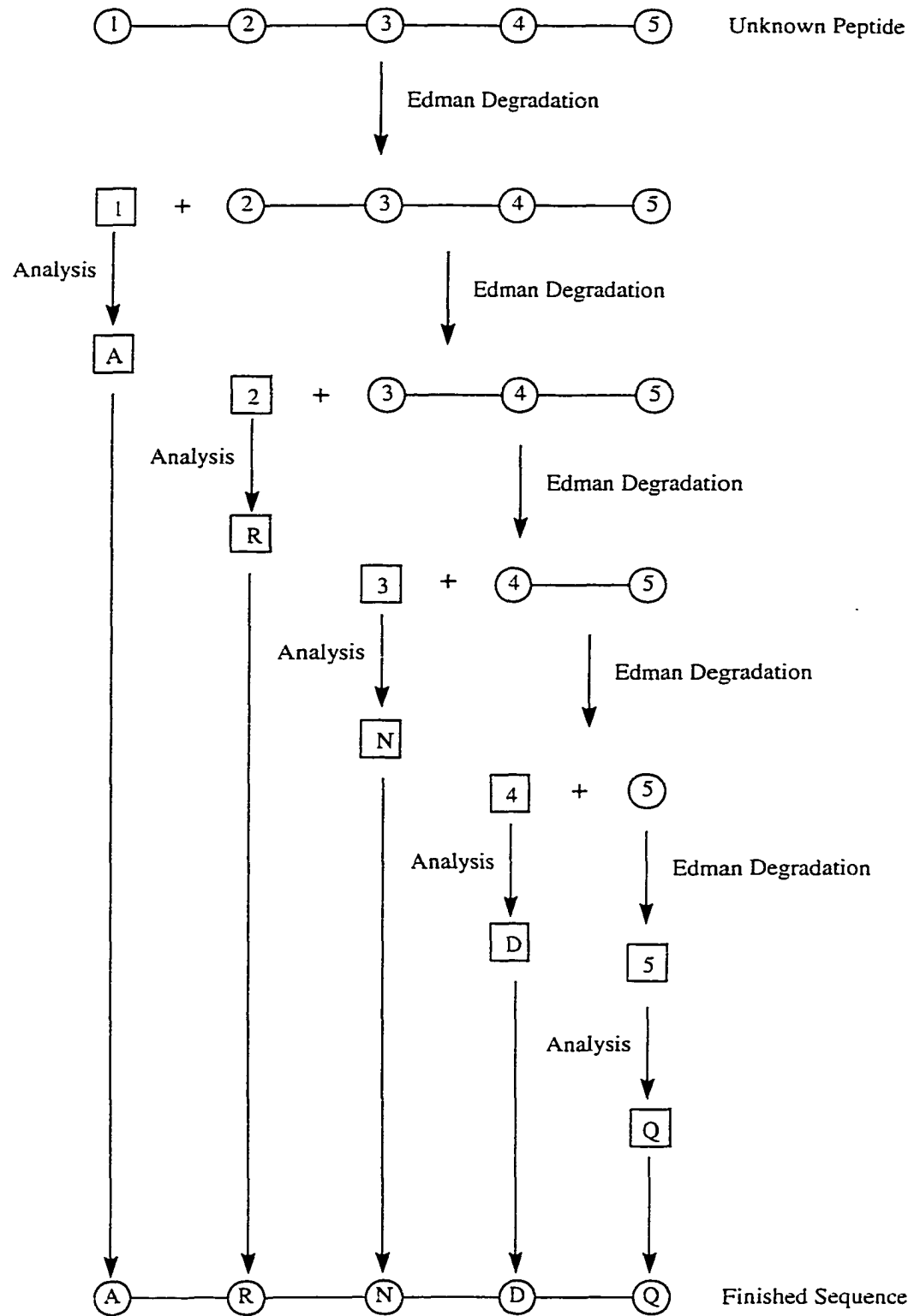


Figure 1.9 General approach for the sequential Edman degradation of a protein.

The Edman degradation is not a simple one step process. It actually consists of six discrete steps that must all work properly for the overall process to be successful. The first step involves coupling between the N-terminal amino acid of the protein and phenylisothiocyanate (PITC). When coupling is complete, excess PITC and other reaction byproducts are washed away to prevent them from interfering with later product identification. After washing, the derivatized amino acid is cleaved from the phenylthiocarbamyl (PTC) protein. It is then extracted and converted to a more stable form. Analysis of this product gives the identity of the original N-terminal residue. The entire process is then repeated on the truncated protein and the new N-terminal residue is identified. Figure 1.10 shows a flowchart of the steps involved in the Edman degradation.

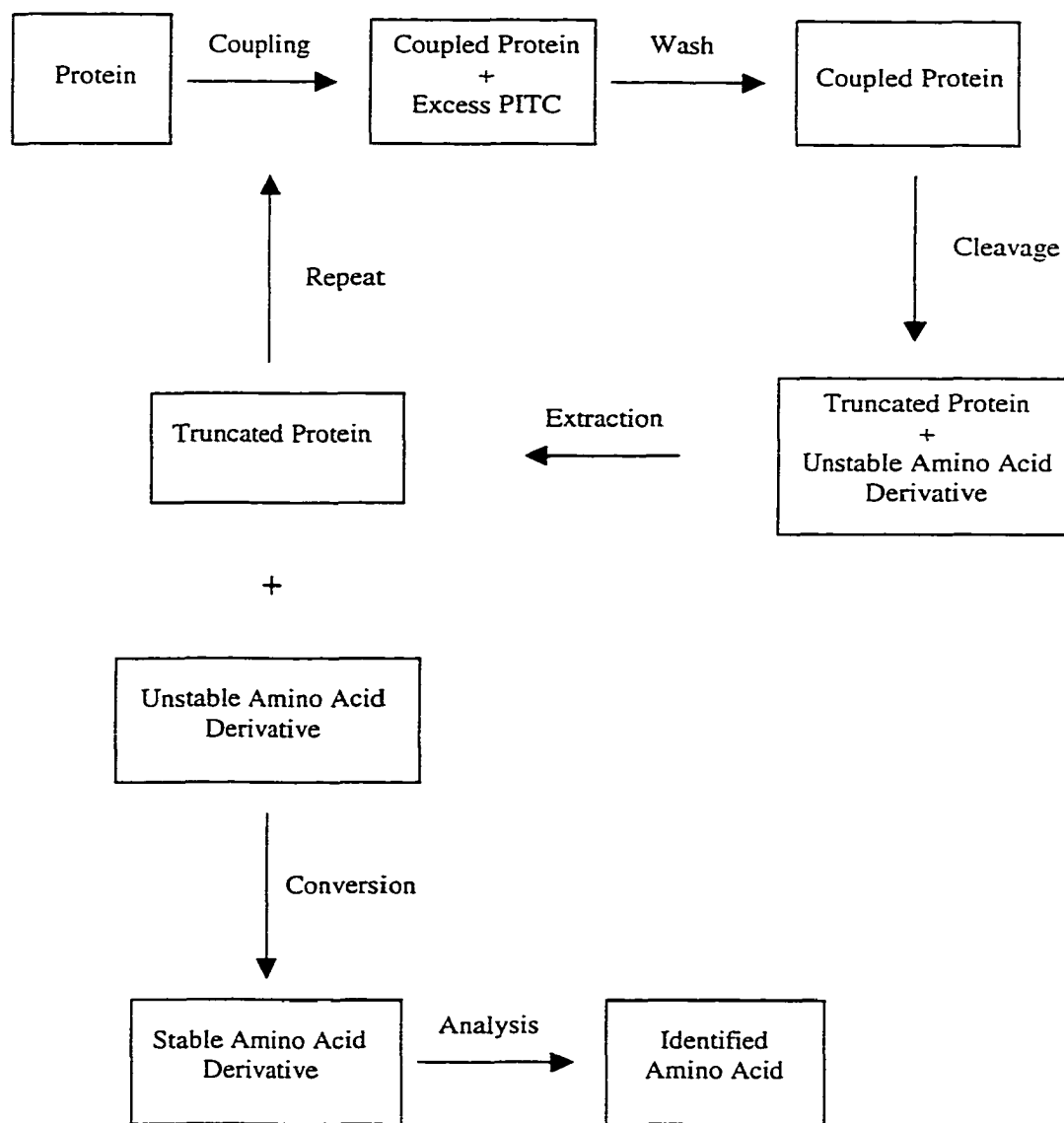


Figure 1.10 Flowchart of the steps in the Edman degradation.

The utility of the Edman degradation arises because it allows researchers to sequence multiple residues from a single sample. This ability is a direct result of the

unique reactivity of isothiocyanates with the N-terminal amino acid of a protein. To wit, the coupled residue can be cleaved from the remainder of the protein under conditions that do not result in destruction of the rest of the peptide backbone. Furthermore, the derivatised residue retains the singular feature of each amino acid, its unique side chain, while the truncated protein retains an active N-terminal site. These points are illustrated in the Edman reaction scheme shown in Figure 1.11.

Coupling involves nucleophilic attack on the isothiocyanate by the N-terminal primary amine of the protein under basic conditions. The cleavage reaction is then performed by treating the phenylthiocarbamyl (PTC) protein formed during coupling with anhydrous acid. This process results in cyclization of the derivatised residue through the sulfur from the isothiocyanate and the amide carbon. The resulting unstable intermediate decomposes at the peptide bond to produce an anilinothiazolinone (ATZ) amino acid and the truncated protein, which retains a primary amine functionality at the new N-terminus. Hydrolysis of any other sites on the peptide backbone does not occur because there is no water present during the cleavage reaction. The ATZ-amino acids formed during cleavage tend to slowly isomerize to the more stable phenylthiohydantoin (PTH) form. To prevent the presence of multiple products from each cycle, this conversion is driven to completion with aqueous acid. The PTH-amino acid is then analysed to give the identity of the original N-terminal residue while the truncated protein is taken through another cycle.

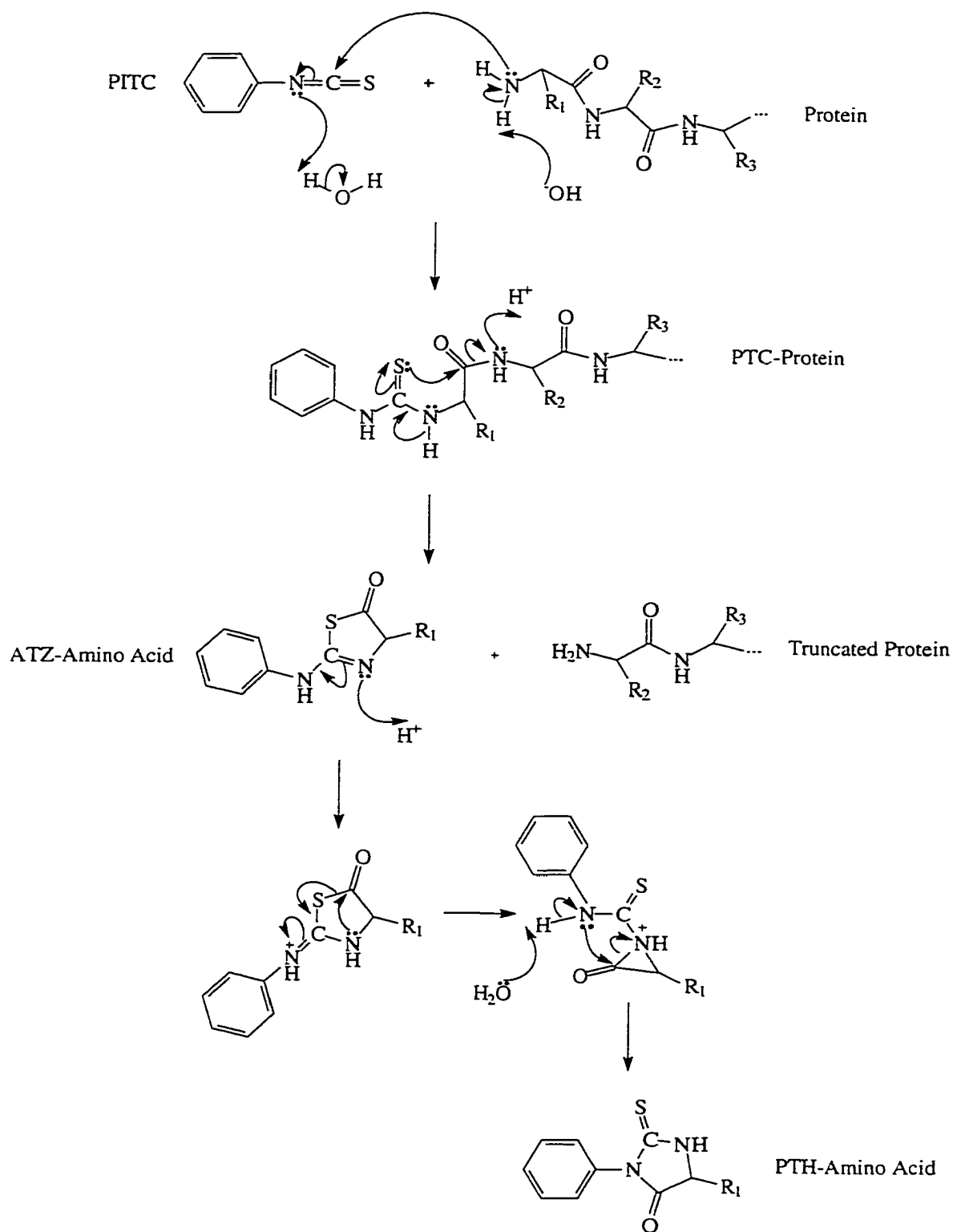


Figure 1.11 Reaction scheme for the Edman degradation.

The reaction scheme shown in Figure 1.11 is applicable to all twenty of the coded amino acids. But as with any such diverse group, some variability in the effectiveness of the overall process is inevitable. For instance, proline, a secondary amine, and glycine are not as amenable to the cleavage reaction as the rest of the amino acids. If the cleavage reaction for these residues does not go to completion during the glycine or proline cycle, then the unreacted PTC-proteins bearing these residues will be cleaved and converted in the following cycle. They will then generate an extra peak during the analysis of that cycle. This carryover, or lag, continues on through the rest of the sequencing run. It is a significant problem because it can lead to difficulty in residue identification in later cycles when the absolute yield is no longer very high. When the N-terminal residue is known to be proline or glycine, lag can be minimised by increasing the cleavage time. This approach is not taken in every cycle because it can lead to an increase in non-specific hydrolysis of the peptide backbone due to the presence of trace amounts of water. Backbone cleavage leads to the formation of new N-terminal residues that will begin to be sequenced in the following cycle. As these secondary sites increase, they produce increasing background signal during analysis, which limits sequence read lengths.

Another problem encountered with the Edman method is  $\beta$ -elimination of the alcohol and thiol side chains of serine, threonine and cysteine. The mechanism of this reaction for PTH-serine is shown in Figure 1.12 below.



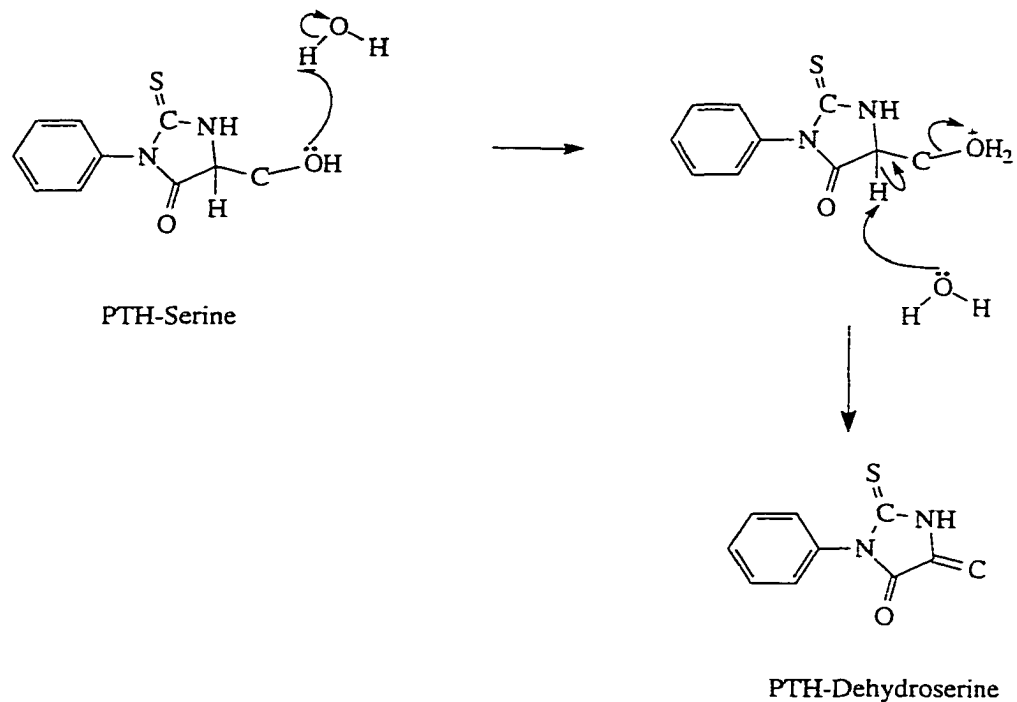


Figure 1.12  $\beta$ -Elimination of PTH-serine during the Edman degradation.

The degree of  $\beta$ -elimination varies for the three amino acids. Threonine is the least reactive followed by serine with cysteine being the most reactive. In fact, while both PTH-threonine and PTH-serine can be detected no PTH-cysteine survives the conversion process. Furthermore, PTH-cysteine and PTH-serine share the same elimination product. The lack of a unique cysteine sequencing product makes identification of this residue difficult. As a result, cysteine residues are typically modified prior to sequencing to a form that is stable to the Edman chemistry. A number of different approaches have been used to do this including conversion to pyridethyl-cysteine (54), carboxymethyl-cysteine and cysteic acid (55).

There is one other problem that can interfere with the Edman chemistry: blockage of the N-terminal amine by another functional group. Without a free amine group at the

N-terminus, the coupling reaction will not work and no sequence information will be obtainable from the sample. There are many examples of naturally occurring proteins that have a blocked N-terminus (56, 57). Any protein that is N-terminally blocked must be sequenced by another method. This problem is more severe for small peptides than for proteins. Proteins are too big to be directly sequenced from N-terminus to C-terminus. Instead, they are subjected to selective internal cleavage of the peptide backbone (58). This process results in the formation of a number of smaller fragments that can be individually purified and sequenced in their entirety. All of the internal cleavage products have a reactive N-terminus, so the only fragment that would need to be sequenced by alternate methods is the one containing the original blocked N-terminal residue. Smaller peptides may not have a site that is amenable to internal cleavage. As such, the entire blocked peptide would have to be sequenced by one of the other methods discussed above.

Blockage of the protein can also occur during the Edman chemistry. N-terminal glutamine residues are prone to cyclization to pyroglutamic acid. However the most problematic form of N-terminal blockage results when oxygen present during coupling causes desulfurization of the PTC-peptide (59). Without the thione present to act as a nucleophile, the cleavage reaction will not occur. This kind of protein blockage reduces the yield of PTH-amino acid from each cycle, which results in a corresponding decrease in repetitive yield for the overall sequencing process. Even a decrease in repetitive yield of only one or two percent can have a profound effect on the number of cycles that can be performed on a single sample. The effect of repetitive yield on sequencing will be discussed in more detail later in this section. Blockage of the sequencing sample during

coupling is controlled by minimizing the amount of oxygen and by adding small amounts of a reducing agent such as dithiothreitol.

Sequencer performance is evaluated using two standard figures of merit. The first is simply the initial yield, IY, expressed as a percentage of the theoretical yield

$$IY = \frac{Y_1}{Y_t} \times 100 \quad (1.40)$$

where  $Y_1$  is the PTH-amino acid yield in cycle 1 and  $Y_t$  is the theoretical yield based on the amount of protein loaded onto the sequencer. Values of IY can be quite variable due to differences in sample loading techniques and in the amount of washout from sample to sample. For this reason, initial yield is best used to compare sequencing runs of the same protein.

The second figure of merit for protein sequencing is the repetitive yield, RY, which gives a measure of reaction completeness in each cycle. It can be determined using the yields from two cycles and the equation

$$RY = \left( \frac{Y_j}{Y_i} \right)^{\frac{1}{(j-i)}} \quad i < j \quad (1.41)$$

where  $Y_j$  and  $Y_i$  are the yields in Cycle  $j$  and Cycle  $i$  respectively. Although a calculation of the repetitive yield between two cycles is useful, the overall repetitive yield is a much more meaningful number. In order to calculate the overall repetitive yield, it is first

necessary to express equation 1.41 into a form that can be conveniently graphed. The first step in this process is to eliminate the exponential term

$$\log RY = \frac{1}{(j-i)} (\log Y_j - \log Y_i) \quad (1.42)$$

Setting  $i = 0$  and rearranging then gives

$$j(\log RY) = \log Y_j - \log Y_i \quad (1.43)$$

or

$$\log Y_j = (\log RY)j + \log Y_i \quad (1.44)$$

Equation 1.44 has the form  $y = mx + b$ , so a plot of the log of the yield in each cycle versus the cycle number will have a slope that is equal to the log of the overall repetitive yield.

The relevance of repetitive yield in protein sequencing can be better understood by considering a hypothetical sequencing run at three different repetitive yields. Table 1.1 shows how the overall yield for the sequencing process changes as the repetitive yield changes (58).

Table 1.1 Effect of repetitive yield on the overall yield at different cycles.

Repetitive Yield	Overall Yield (%)			
	Cycle 10	Cycle 20	Cycle 50	Cycle 100
99%	90	82	60	37
95%	60	36	8	0.6
90%	35	12	0.5	0.003

If one assumes that proper sequence assignment becomes difficult below an overall yield of 35%, then a repetitive yield of 99% will allow for the analysis of 100 cycles from a given sample. In contrast, a repetitive yield of 90% will give a sequence read length of only 10 cycles. Of course, long read lengths are desirable because they make the most use of a given sample. Table 1 clearly indicates the necessity of having as high a repetitive yield as possible.

Manual sequencing using the Edman degradation is a very labour intensive process. Analysing even a small protein in this manner can take years to complete. However, the repetitive nature of the sequencing chemistry lends itself quite well to automation. An automated sequencer has a number of advantages over manual sequencing. First, throughput is increased because the instrument can be operated on a 24 hour basis. It is further increased because automation typically leads to a decrease in cycle times. Second, the sequencing steps can be performed in a closed system. Isolating the sample and all of the reagents minimizes contamination, which results in increased yields for the PTH-amino acids and an increased repetitive yield for the overall process. In turn, increased yields serve to improve sequencing sensitivity and to increase the

sequence read length for a given sample. Third, automated instruments can be miniaturized. Decreasing the size of the sequencer leads to a decrease in reagent volumes. Decreasing the reagent volumes also reduces contaminants, resulting in cleaner analyses and improved sensitivity.

The first automated protein sequencer, the Sequenator, was introduced by Edman in 1967 (60). The coupling, wash, cleavage and extraction steps were all carried out in a spinning cup. Reagents were delivered by pressurizing the various reservoirs with nitrogen and opening the appropriate valve for a fixed period of time. The extracted ATZ-amino acids were sent to a fraction collector where they were held until they could be manually converted to PTH-amino acids and analysed by thin layer chromatography (TLC). The spinning cup was used to increase reaction rates by producing a thin film of protein with a very large surface area. Because the protein was not bound in any way to the cup, non-polar wash and extraction solvents were used to minimize sample washout.

In 1971 Laursen introduced the solid-phase sequencer (61). It was primarily intended for the analysis of small peptides that were susceptible to washout on the spinning cup sequencer (58). The main modification incorporated by the solid-phase sequencer involved covalent attachment of the protein to a solid support prior to analysis. This approach had the advantage that none of the bound protein would be lost during the wash or extraction steps. It also led to faster cycles with cleaner products because less time was required for sample drying and stronger wash reagents could be used (62). As with the Sequenator, the solid phase sequencer was only automated up to the extraction step. The ATZ-amino acids were still collected on a fractionator for manual conversion and analysis by TLC. This instrument did not achieve broad acceptance due to the

inadequately developed coupling chemistry (58, 62). However, there are now coupling methods available that can be used to reliably attach a protein through either its amine sites (63) or its carboxyl sites (64, 65).

Even with the advances in coupling chemistry, solid-phase sequencing has not seen a significant growth in popularity. Instead, most sequencing today is performed on a gas-liquid phase instrument (4). This type of sequencer incorporates gas phase delivery of both the coupling base and the cleavage acid with liquid phase delivery of all of the other reagents. Gas phase delivery of the most polar reagents prevents sample washout during coupling and cleavage. Furthermore, the sample is typically adsorbed to either a polyvinylidene difluoride (PVDF) membrane or to a glass fibre mat (GFM) incorporating a polycationic carrier such as polybrene. Most proteins and peptides adsorb strongly to these matrices, so there is minimal sample loss with the fairly hydrophobic reagents used during wash and extraction.

Modern sequencers also no longer employ TLC for identification of the PTH-amino acid sequencing products. Instead they are analysed by HPLC with UV absorbance detection. The retention time of the individual PTH-amino acid generated in each cycle is used to confirm its identity. As well, analysis by HPLC allows for quantitation of the PTH-amino acids in each cycle. This information can be used to more reliably identify each residue, especially in the presence of background peaks in later cycles resulting from the gradual increase in lag and non-specific cleavage of the peptide backbone during analysis. In this way, sequence read lengths can be extended beyond those obtainable using TLC.

There is another significant advantage of using HPLC instead of TLC to separate the PTH-amino acids: each sample can be automatically analysed after conversion. Combined with on-line conversion, it allows for automation of the entire sequencing process. The sample is applied, the instrument is programmed to perform the desired number of cycles and the analysis is initiated. A block diagram of an automated protein sequencer is shown in Figure 1.13.

A suitable solid support carrying the protein sample is inserted into the reaction chamber. Coupling, wash, cleavage and extraction reagents are then delivered through valve block A according to a preprogrammed sequence of steps. The extracted ATZ-amino acid is carried from the reaction chamber to the conversion chamber; all other reagents go to waste. A stream of argon is used to dry the extraction solvent. Aqueous acid is then delivered to the chamber through valve block C and conversion occurs. The resulting PTH-amino acid is delivered to an HPLC system where it is analysed.



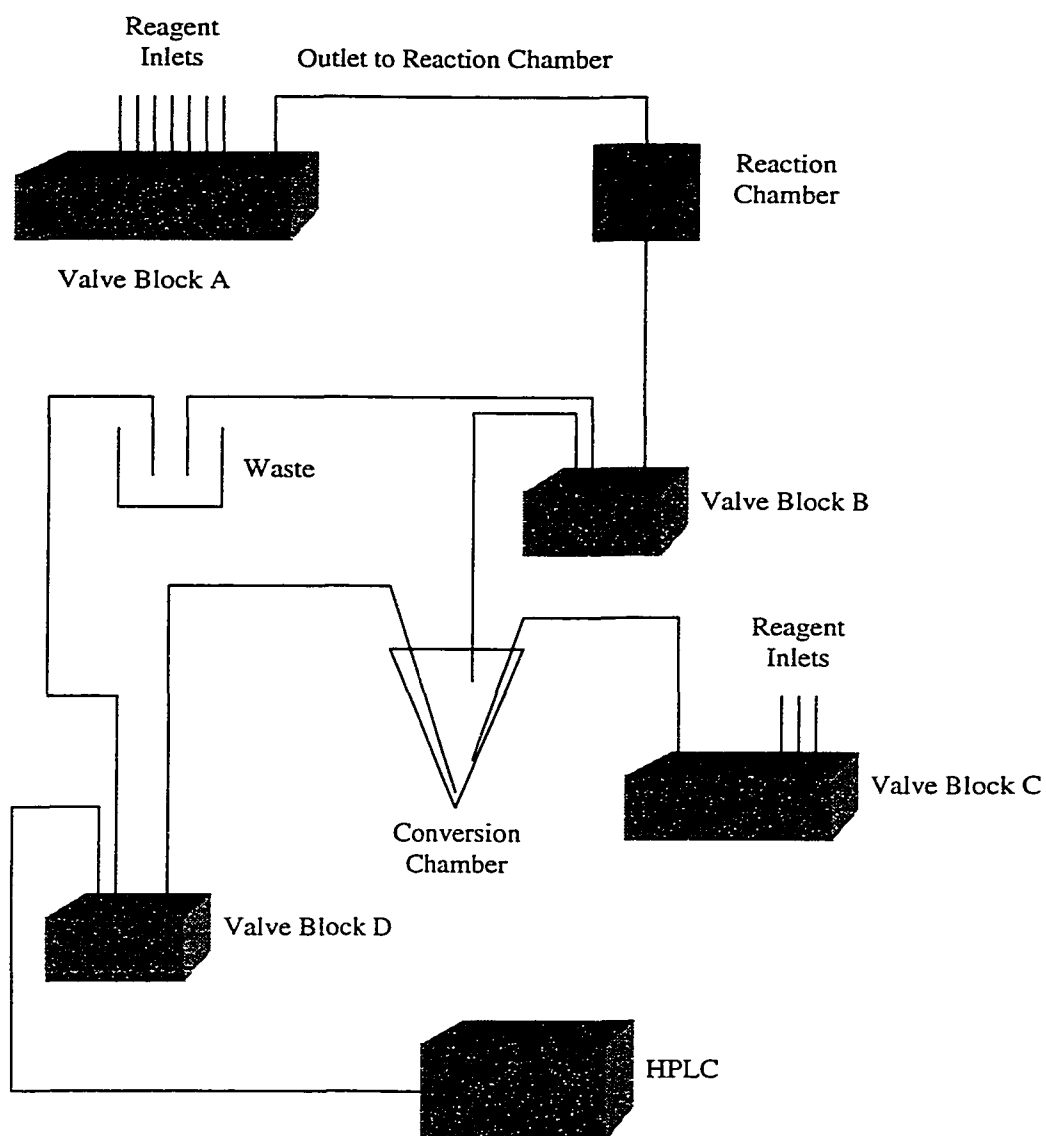


Figure 1.13 Schematic of an automated protein sequencer.

Automated sequencers have been a boon for protein researchers. They can provide sequence information on samples down to the low picomole level (1, 4).

Unfortunately many biologically active proteins cannot be isolated even at this level (2). Analysis of these hard to isolate samples requires an improvement in sequencer sensitivity. Presently, sensitivity is limited not by the sequencing chemistry, but rather by the HPLC with UV detection systems used to analyse the PTH-amino acids. Further improvements in this detection methodology are likely to result in only marginal improvements in sensitivity. Therefore, sub-picomole sequencing requires an alternate detection technology. A number of different detection systems have been suggested including mass spectrometry (66) thermo-optical absorbance (9, 67, 68) and fluorescence detection (10, 36, 69-79). The most promising approach involves the use of CE-LIF to detect fluorescein thiohydantoin (FTH) amino acid sequencing products (33, 36, 69, 70). The FTH-amino acids are produced when fluorescein isothiocyanate (FITC) is used for coupling instead of PITC. This group has previously demonstrated zeptomole limits of detection for FTH-amino acids analysed by CE-LIF (33, 36, 69).

High sensitivity protein sequencing using CE-LIF cannot be accomplished by simply replacing PITC with FITC in the Edman chemistry. The more extensive  $\pi$ -system of FITC decreases the electrophilicity of the isothiocyanate moiety and hence, makes it less reactive than PITC. A decrease in reactivity leads to a decrease in repetitive yield, which severely limits the utility of FITC coupling. Fortunately, this problem can be overcome by performing a second coupling using PITC to scavenge any N-terminal sites that didn't react with FITC. The cleavage and conversion reactions are identical for both reagents so no other extra steps are required. This point is illustrated in the double coupling flowchart below (Figure 1.14). Although a mixture of the FTH and PTH-amino

acid is produced in each cycle, only the FTH-amino acid is identified because PTH-amino acids are not fluorescent.

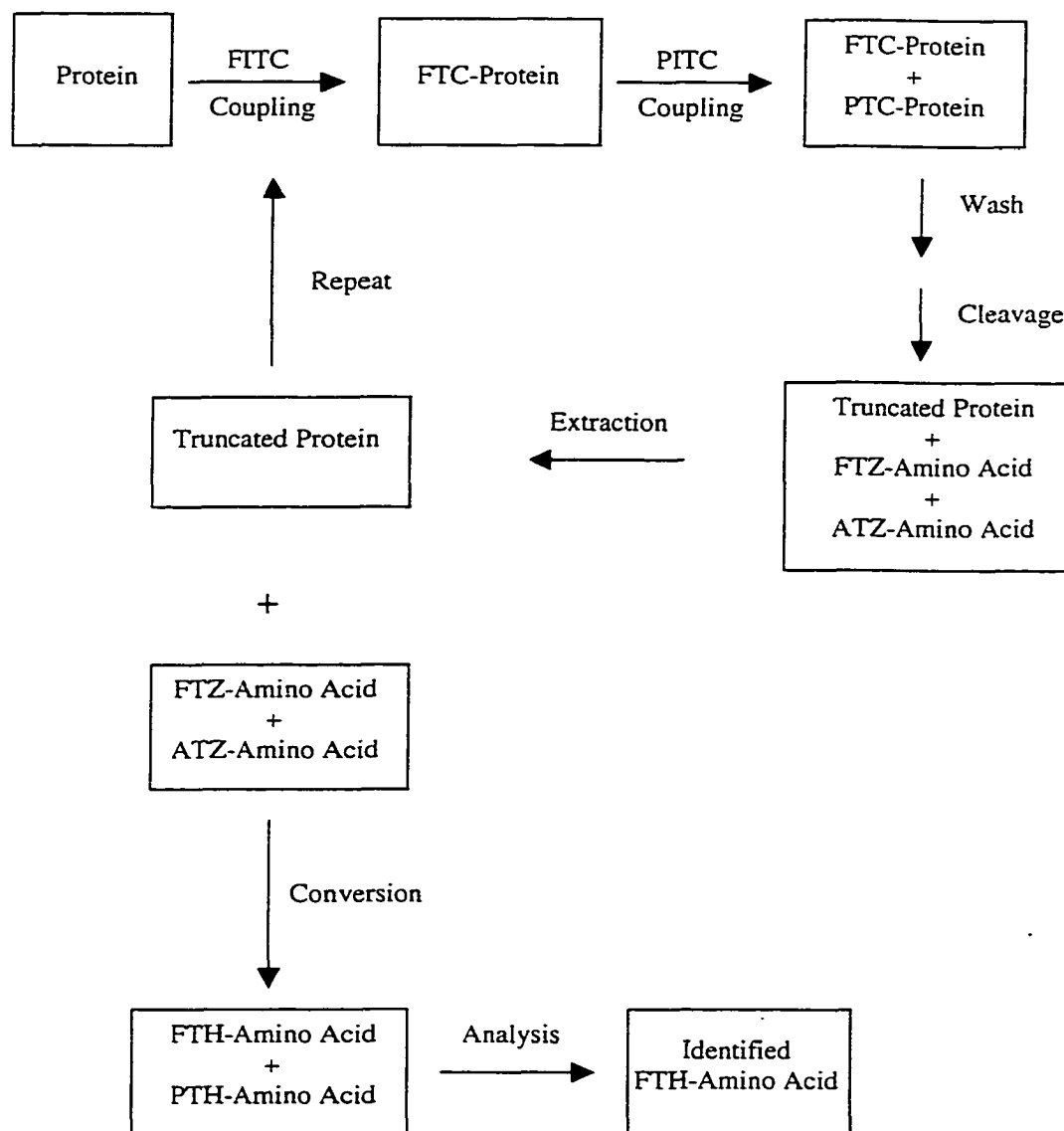


Figure 1.14 Flowchart of the double coupling Edman chemistry.

Even using the double coupling approach described above, it is unlikely that a standard automated sequencer fitted with a CE-LIF detection system would show much of an improvement in sensitivity. All modern sequencers have been designed for use with HPLC. The volume requirements of HPLC are four to five orders of magnitude greater than those of CE. Therefore, the sensitivity advantage of CE-LIF would be lost because only a tiny fraction of each sample would be analysed. This problem can only be overcome by building a sequencer that is more compatible with the volume requirements of CE. The design of such an instrument is described in Chapter 3 of this thesis. Chapter 4 then explores the effects of the double coupling chemistry with the miniaturized sequencer on each step of the sequencing process. And finally, preliminary sequencing data is presented in Chapter 5 that demonstrates how the miniaturized sequencer with CE-LIF detection of FTH-amino acids can be used for sequencing at the low picomole level. This proof of principle instrument shows great promise for one day being able to sequence proteins and peptides at the low femtomole level.

## 1.5 References

1. R. Aebersold, *Nature* **343**, 291 (1990).
2. S. Kent, et al., *BioTechniques* **5**, 314 (1987).
3. L. M. Smith, *Anal. Chem.* **60**, 381A (1988).
4. R. M. Hewick, M. W. Hunkapiller, L. E. Hood, W. J. Dreyer, *J. Biol. Chem.* **256**, 7990 (1981).
5. J. Calaycay, M. Rusnak, J. E. Shively, *Anal. Biochem.* **192**, 23 (1991).

6. S. Hjerten, *Chromatogr. Rev.* **9**, 122 (1967).
7. W. G. Kuhr, C. A. Monnig, *Anal. Chem.* **64**, 389R (1992).
8. G. J. M. Bruin, et al., *J. Chromatogr.* **559**, 163 (1991).
9. K. C. Waldron, N. J. Dovichi, *Anal. Chem.* **64**, 1396 (1992).
10. D. Chen, K. Adelhelm, X. L. Cheng, N. J. Dovichi, *Analyst* **119**, 349 (1994).
11. C. S. Liu, et al., *Electrophoresis* **19**, 3183 (1998).
12. C. S. Lee, in *Handbook of Capillary Electrophoresis* J. P. Landers, Ed. (CRC Press, New York, 1997) pp. 717.
13. D. McManigill, S. A. Swedberg, in *Techniques in Protein Chemistry* T. E. Hugli, Ed. (Academic Press, San Diego, 1989) pp. 468.
14. R. D. Smith, J. A. Olivares, N. T. Nguyen, H. R. Udseth, *Anal. Chem.* **60**, 436 (1988).
15. A. Guttman, A. S. Cohen, D. N. Heiger, B. L. Karger, *Anal. Chem.* **62**, 137 (1990).
16. J. W. Jorgenson, K. D. Lukacs, *Anal. Chem.* **53**, 1298 (1981).
17. K. K.-C. Yeung, C. A. Lucy, *Anal. Chem.* **70**, 3286 (1998).
18. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, *Anal. Chem.* **56**, 111 (1984).
19. S. T. Terabe, K. Otsuka, T. Ando, *Anal. Chem.* **57**, 834 (1985).
20. W. Kleibohmer, K. Cammann, J. Robert, E. Mussenbrock, *J. Chromatogr.* **638**, 349 (1993).
21. S. Takeda, S. Wakida, M. Yamane, A. Kawahara, K. Higashi, *J. Chromatogr.* **653**, 109 (1993).

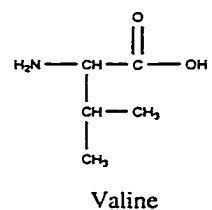
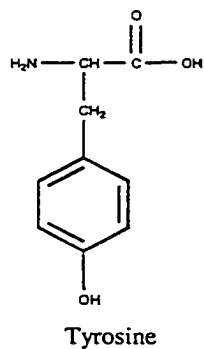
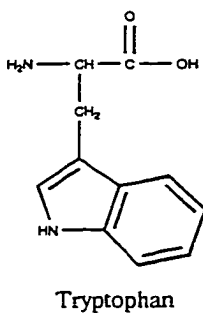
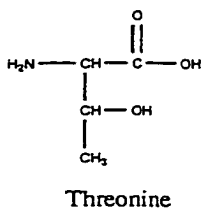
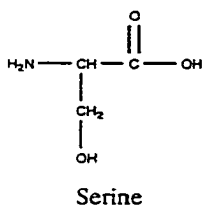
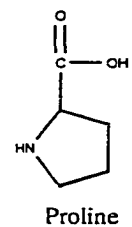
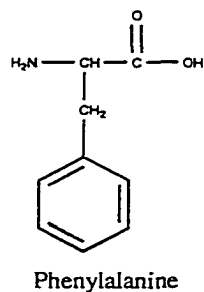
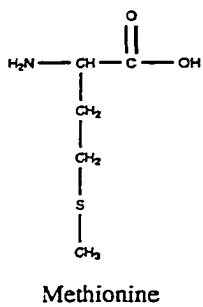
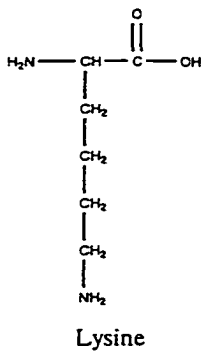
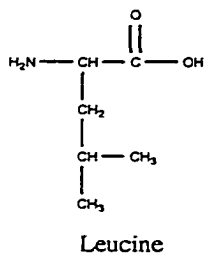
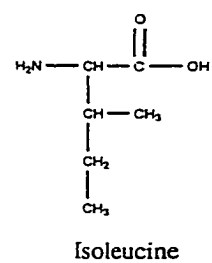
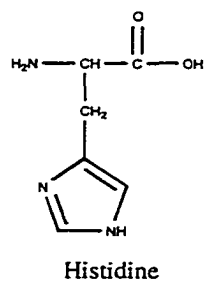
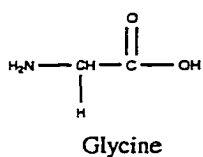
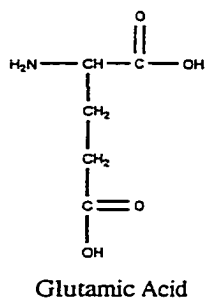
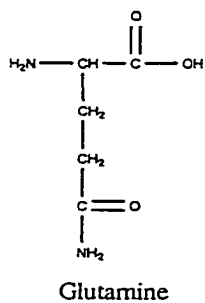
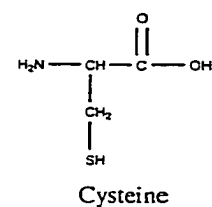
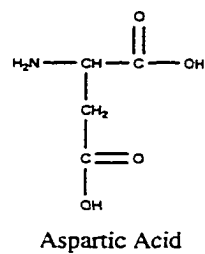
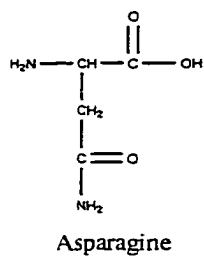
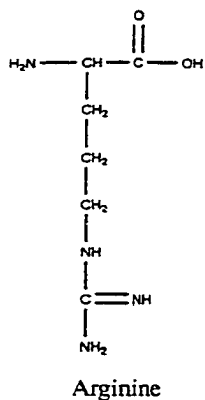
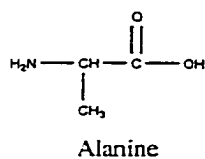
22. R. Weinberger, I. S. Lurie, *Anal. Chem.* **63**, 823 (1991).
23. P. Wernley, W. Thormann, *Anal. Chem.* **63**, 2878 (1991).
24. V. C. Trennery, J. Robertson, R. J. Wells, *Electrophoresis* **15**, 103 (1994).
25. A. J. Ji, et al., *J. Chromatogr.* **669**, 15 (1995).
26. D. E. Burton, L. L. Powell, X. Xi, *J. Microcolumn Separations* **6**, 19 (1994).
27. F. Ferres, M. A. Blazquez, M. I. Gil, F. A. Tomas-Barberan, *J. Chromatogr.* **669**, 268 (1994).
28. J. Vindevogel, R. Szucs, P. Sandra, L. C. Verhagen, *High Resolution Chromatogr.* **14**, 584 (1991).
29. S. Boonkerd, M. Lauwers, M. R. Detaevernier, Y. Michotte, *J. Chromatogr.* **695**, 97 (1995).
30. S. Croubels, W. Baeyens, C. Dewaele, C. V. Van Peteghem, *J. Chromatogr.* **673**, 267 (1994).
31. T. G. Nolan, W. A. Weimer, N. J. Dovichi, *Anal. Chem.* **56**, 1704 (1984).
32. F. Zarrin, N. J. Dovichi, *Anal. Chem.* **57**, 2690 (1985).
33. Y. Cheng, N. J. Dovichi, *Science* **242**, 562 (1988).
34. N. J. Dovichi, J. C. Martin, J. H. Jett, R. A. Keller, *Science* **219**, 845 (1983).
35. D. Chen, N. J. Dovichi, *Anal. Chem.* **68**, 690 (1996).
36. S. Wu, N. J. Dovichi, *J. Chromatogr.* **480**, 141 (1989).
37. D. B. Craig, J. C. Y. Wong, N. J. Dovichi, *Anal. Chem.* **68**, 697 (1996).
38. J. Y. Zhao, N. J. Dovichi, O. Hindsgaul, S. Gosselin, M. M. Palcic, *Glycobiology* **4**, 239 (1994).
39. D. Chen, N. J. Dovichi, *J. Chromatogr.* **657**, 265 (1994).

40. J. Y. Zhao, P. Diedrich, Y. Zhang, O. Hindsgaul, N. J. Dovichi, *J. Chromatogr.* **657**, 307 (1994).
41. F. Sanger, *Biochem. J.* **39**, 507 (1945).
42. H. Tschesche, in *Methods in Enzymology* C. H. W. Hirs, S. N. Timasheff, Eds. (Academic Press Inc., New York, 1977), vol. 47, pp. 73.
43. R. Hayashi, in *Methods in Enzymology* C. H. W. Hirs, S. N. Timasheff, Eds. (Academic Press Inc., New York, 1977), vol. 47, pp. 84.
44. K. Takamoto, M. Kamo, K. Satake, A. Tsugita, in *Methods in Protein Structure Analysis* M. Z. Atassi, E. Appella, Eds. (Plenum Press, New York, 1995) pp. 89.
45. V. L. Boyd, M. Bozzini, G. Zon, R. L. Noble, R. J. Mattaliano, *Anal. Biochem.* **206**, 344 (1992).
46. C. G. Miller, J. M. Bailey, D. H. Hawke, S. Early, J. Tso, in *Methods in Protein Structure Analysis* M. Z. Atassi, E. Appella, Eds. (Plenum Press, New York, 1995) pp. 119.
47. C. K. Mathews, K. E. vanHolde, *Biochemistry* (The Benjamin/Cummings Pub. Co., Redwood City, CA, 1990).
48. M. Mann, G. Talbo, *Current Opinion in Biotechnology* **7**, 11 (1996).
49. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, *Science* **246**, 64 (1989).
50. A. Shevchenko, M. Wilm, M. Mann, *J. Prot. Chem.* **15**, 481 (1997).
51. J. R. I. Yates, A. L. McCormack, D. Scheiltz, E. Carmack, A. Link *J. Prot. Chem.* **16**, 495 (1997).

52. F. W. McLafferty, E. K. Fridriksson, D. M. Horn, M. A. Lewis, R. A. Zubarev, *Science* **284**, 1289 (1999).
53. P. Edman, *Acta Chem. Scand.* **4**, 283 (1950).
54. M. Friedman, L. H. Krull, J. F. Cavins, *J. Biol. Chem.* **245**, 3868 (1970).
55. G. E. Tarr, *Methods Enzymol.* **47**, 335 (1977).
56. J. M. Wilkinson, E. M. Press, R. R. Porter, *Biochemical J.* **100**, 303 (1966).
57. K. Narita, *Biochim. Biophys. Acta.* **28**, 184 (1958).
58. L. R. Croft, *Introduction to Protein Sequence Analysis* (John Wiley & Sons, Toronto, 1980).
59. G. E. Tarr, *Anal. Biochem.* **63**, 361 (1975).
60. P. Edman, G. Begg, *Eur. J. Biochem.* **1**, 80 (1967).
61. R. A. Laursen, *Eur. J. Biochem.* **20**, 89 (1971).
62. J. B. C. Findlay, D. J. C. Pappin, J. N. Keen, in *Protein Sequencing A Practical Approach* J. B. C. Findlay, M. J. Geisow, Eds. (IRL Press, Oxford, 1989) pp. 69.
63. E. Wachter, W. Machleidt, H. Hofner, J. Otto, *FEBS Lett.* **35**, 97 (1973).
64. B. Wittmann-Liebold, M. Kimura, in *Modern Methods in Protein Chemistry* H. Tschesche, Ed. (de Gruyter, New York, 1984) pp. 229.
65. M. D. Davison, J. B. C. Findlay, *Biochem. J.* **236**, 389 (1986).
66. A. Ducret, E. J. Bures, R. Aebersold, *J. Prot. Chem.* **16**, 323 (1997).
67. N. J. Dovichi, K. C. Waldron, M. Chen, I. Ireland, *J. Prot. Chem.* **13**, 470 (1994).
68. M. Chen, K. C. Waldron, Y. Zhao, N. J. Dovichi, *Electrophoresis* **15**, 1290 (1994).
69. I. D. Ireland, et al., *J. Prot. Chem.* **16**, 491 (1997).



70. K. Muramoto, K. Nokihara, A. Ueda, H. Kamiya, *Biosci. Biotech. Biochem.* **58**, 300 (1994).
71. V. Farnsworth, K. Steinberg, *Anal. Biochem.* **215**, 200 (1993).
72. V. Farnsworth, K. Steinberg, *Anal. Biochem.* **215**, 190 (1993).
73. O. Imakyure, M. Kai, T. Mitsui, H. Nohta, Y. Ohkura, *Anal. Sci.* **9**, 647 (1993).
74. O. Imakyure, M. Kai, Y. Ohkura, *Anal. Chim. Acta.* **291**, 197 (1994).
75. H. Miyano, T. Nakajima, K. Imai, *Biomed. Chromatogr.* **2**, 139 (1987).
76. S. W. Jin, G. X. Chen, Z. Palacz, B. Wittmann-Liebold, *FEBS Lett.* **198**, 150 (1986).
77. H. Matsunaga, et al., *Anal. Chem.* **67**, 4276 (1995).
78. Z. Palacz, J. Salnikow, S.-W. Jin, B. Wittmann-Liebold, *FEBS Lett.* **176**, 365 (1984).
79. H. Hirano, B. Wittmann-Liebold, *Biol. Chem.* **367**, 1259 (1986).

Appendix A: Structures of the 20 coded L- $\alpha$ -amino acids.

## Chapter 2

### **Identification of Fluorescein Thiohydantoin Amino Acids by Capillary Electrophoresis with Laser Induced Fluorescence Detection as a Potential Method for Improving Sensitivity in Protein Sequencing**

#### **2.1 Introduction**

Protein sequence analysis describes any procedure that is used to determine the primary structure, or the connectivity of the amino acids, of a protein sample. Although there are a number of methods that can be used to sequence proteins (1-4), the most common approach is to use the isothiocyanate chemistry developed by Per Edman almost fifty years ago (2). The Edman chemistry has remained popular because it results in cleavage of the labelled N-terminal amino acid from the protein and it can be repeated on the truncated sample. By analysing the individual product generated in each successive cycle, it is possible to piece together the primary structure of the protein being studied.

Commercial protein sequencers based on the Edman method rely on HPLC with UV absorbance detection to identify the phenylthiohydantoin (PTH) amino acid products of sequencing. Using this detection technology, modern sequencers can provide primary sequence information on low picomole amounts of proteins and peptides (5, 6). They are unable to analyse smaller amounts due to the inherent sensitivity limitations associated with UV absorbance detection (7). Although this level of sensitivity has made sequencing instruments valuable tools for protein characterization, it is still inadequate for the primary structure analysis of a significant number of biologically active and

clinically useful proteins and peptides. An instrument that could provide sequence information on samples at the low femtomole level would open up new areas of research for protein scientists (8).

Improving sequencer sensitivity to the femtomole level will require an improvement in the detection technology used to analyse the sequencing products. While a number of alternative detection methods have been investigated (9-12), they have not, as yet, resulted in a significant decrease in the amount of protein needed for sequencing sensitivity. This chapter will explore the use of capillary electrophoresis with laser induced fluorescence detection (CE-LIF), one of the most promising alternatives to HPLC with UV absorbance detection, for the high sensitivity analysis of protein sequencing products. This group has previously achieved zeptomole ( $1 \text{ zmol} = 1 \times 10^{-21} \text{ mol}$ ) and yoctomole ( $1 \text{ ymol} = 1 \times 10^{-24} \text{ mol}$ ) limits of detection (LOD's) for a number of different fluorescent products using CE-LIF (13-16). The ability to analyse protein sequencing products at this level would represent a six order of magnitude improvement over current detection methods and would bring the goal of femtomole sequencing much closer to fruition.

Although CE-LIF is much more sensitive than HPLC with UV absorbance detection, it is not compatible with standard Edman chemistry (13). The PTH-amino acids that result do not fluoresce and hence, they are invisible to the LIF detector. The Edman chemistry must be altered so that fluorescent sequencing products are generated. A number of different reagents that would generate fluorescent sequencing products have been investigated (17-23). The majority of these reagents are modified isothiocyanates meant to replace phenylisothiocyanate (PITC) in the coupling step of the Edman

degradation. The most thoroughly investigated of these modified coupling reagents is fluorescein isothiocyanate (FITC) (13, 24, 25). It has been shown to work with both manual and automated protein sequencing however the detection systems employed were not sensitive enough to result in improved sequencing sensitivity (11, 26). This chapter will present a separation scheme for the fluorescein thiohydantoin (FTH) amino acid products generated from FITC coupling to eighteen of the twenty coded amino acids. The ability to unambiguously identify these potential sequencing products with the sensitivity afforded by CE-LIF is the first step in developing a sequencer that can successfully sequence proteins and peptides at the low femtomole level.

## 2.2 Experimental

### 2.2.1 Fluorescein Thiohydantoin Amino Acid Synthesis

Fluorescein thiohydantoin amino acids were prepared according to the method of Wu and Dovichi (27). Individual amino acids (Sigma Chemical, St. Louis, MO) were dissolved in water to a concentration of 0.01 M. The addition of several drops of 12.5% trimethylamine in water (Applied Biosystems, Foster City, CA) was necessary to completely dissolve all of the tyrosine. Fluorescein isothiocyanate (Molecular Probes, Eugene, OR) was dissolved in HPLC grade acetone (Sigma Chemical, St. Louis, MO) to a concentration of 1 mM and stored at 4°C. To derivatise the amino acids, 20 µL of amino acid solution was mixed with 20 µL of 1 mM FITC and 120 µL of pH 9.1 carbonate buffer in a 1.5 mL micro-centrifuge tube. The coupling reaction was allowed

to proceed in the dark for 4-8 hr. When coupling was complete, 120  $\mu\text{L}$  of anhydrous trifluoroacetic acid (TFA) (Sigma Chemical, St. Louis, MO) was added to each tube. The conversion reaction was allowed to proceed for 12 hr in the dark. The FTH-amino acids were then dried on a vacuum concentrator, redissolved in HPLC grade acetonitrile (Sigma Chemical, St. Louis, MO) and stored at 4°C until needed.

### 2.2.2 Capillary Electrophoresis with Laser Induced Fluorescence Detection

The capillary electrophoresis instrument has been described in detail elsewhere (28). An argon ion laser (Uniphase, San Jose, CA) was used to provide a 25 mW excitation beam at  $\lambda = 488 \text{ nm}$ . The beam was focused with a 10x microscope objective (Melles Griot, Nepean, ON) to a point approximately 100  $\mu\text{m}$  below the end of the separation capillary (Polymicro Technologies, Phoenix, AZ). The capillary was fixed inside a sheath flow cuvette that was housed in a locally constructed holder. Post-column fluorescence was collected at a right angle to the incident laser beam using a 50x, 0.60 numerical aperture microscope objective (Melles Griot, Nepean, ON) and imaged onto an iris adjusted to block scattered laser light. The image then passed through a 525 nm DF 40 nm interference filter (Omega Optical, Brattleboro, VT) to a photomultiplier tube (Hamamatsu, San Jose, CA) operated at 1000 V. The photomultiplier tube output was conditioned with a 0.1 s resistor-capacitor low-pass filter and recorded on a Macintosh Quadra computer.

Separation of the FTH-amino acids was carried out in buffers containing varying amounts of analytical grade sodium dihydrogen phosphate, sodium tetraborate, sodium

dodecyl sulphate and magnesium acetate. The separation capillaries used in this work were 60 cm long with an outer diameter of 150  $\mu\text{m}$  and an inner diameter of 30  $\mu\text{m}$ . The FTH-amino acids were analysed at a nominal concentration of  $2 \times 10^{-9}$  M. Samples were electrokinetically injected onto the capillary for 5 s at 25 V/cm and separated at 300 V/cm.

### **2.3 Results and Discussion**

Developing an electrophoretic separation for the FTH-derivatives of the coded amino acids is not a trivial exercise. The difficulty arises from the fact that the relative difference in structure between the various FTH-amino acids is quite small. A comparison of FTH-leucine and FTH-isoleucine (Figure 2.1) illustrates this point.

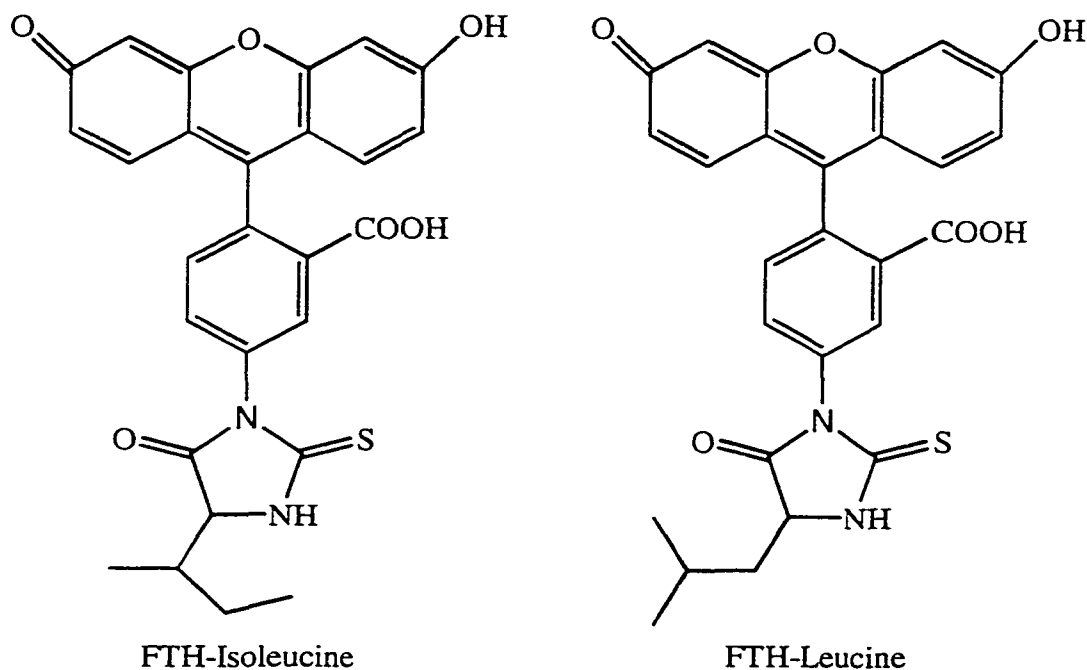


Figure 2.1 Comparison of the structures of FTH-Isoleucine and FTH-Leucine.

The two molecules differ only in their amino acid side chains while sharing a large fluorescein thiohydantoin moiety. As a result, their migration characteristics are expected to be quite similar. This same argument applies to all of the eighteen FTH-amino acid derivatives studied in this work. The FTH-derivatives of the other two coded amino acids, lysine and cysteine, were excluded not because they could not be separated but rather because of complications associated with the Edman chemistry.

The lysyl side chain contains a primary amine that is also quite reactive with isothiocyanates. This secondary reactive site does not pose a problem for the standard Edman chemistry because only one derivatizing reagent, PITC, is used and the reaction at each of the two sites essentially goes to completion. As a result, only one product is



observed when a lysine residue is sequenced. In contrast, the decreased reactivity of FITC necessitates the use of PITC as a scavenger. Combining two reactive sites and two derivatizing reagents leads to four possible doubly labelled sequencing products, three of which fluoresce. Any lysine standard should, therefore, be composed of an appropriate mixture of the three expected fluorescent products. The method used to label the amino acids in this work would not produce such a mixture. Excess amino acid was used to prepare the standards to minimize the amount of unreacted FITC and other byproduct peaks that could overload the detector and interfere with the analysis. Under these conditions, the predominant products from lysine labelling would instead, be singly labelled at either the  $\alpha$  or  $\epsilon$  amine site.

Cysteine decomposes under the conditions used for Edman sequencing. As such, it is typically identified after modification of the side chain thiol to a more stable form. A number of modifications have been successfully employed to produce stable cysteine derivatives (29, 30). Identification of the FTH-cysteine derivative that is best suited to the overall FTH-amino acid separation, while important, is not a necessary requirement for demonstrating that CE-LIF can be used to identify protein sequencing products. For this reason, no cysteine derivatives were included in this work.

Because of the similarity in structure between all of the FTH-amino acids, it is not possible to separate them simply using capillary zone electrophoresis (CZE) (10, 27, 31). This chapter explores the use of micellar electrokinetic capillary chromatography (MECC) along with modifications to the electroosmotic flow rate and careful control of buffer pH to improve separation to the point where the FTH-derivatives of eighteen of the twenty coded amino acids can be unambiguously identified.

MECC is a very useful electrophoretic technique for the analysis of complex mixtures. It employs a surfactant that produces micelles in the running buffer. Analyte molecules can then partition between the micelles and the running buffer as they migrate through the separation capillary. By combining the separation characteristics of CZE with the partition effects of chromatography, it is often possible to separate analytes having very similar electrophoretic mobilities.

The most common surfactant used in MECC is sodium dodecyl sulfate (SDS). In this work the effect of SDS on the separation of FTH-amino acids was studied. Figure 2.2 shows the electropherograms obtained for the analysis of eighteen coded amino acids at different SDS running buffer concentrations. Each FTH-amino acid peak is labelled with the single letter code corresponding to its original amino acid. The two serine peaks result from partial dehydration of its side chain alcohol during the Edman chemistry. There is an extra peak in all of the electropherograms that has the same migration time as FITC and has therefore been labelled as such.

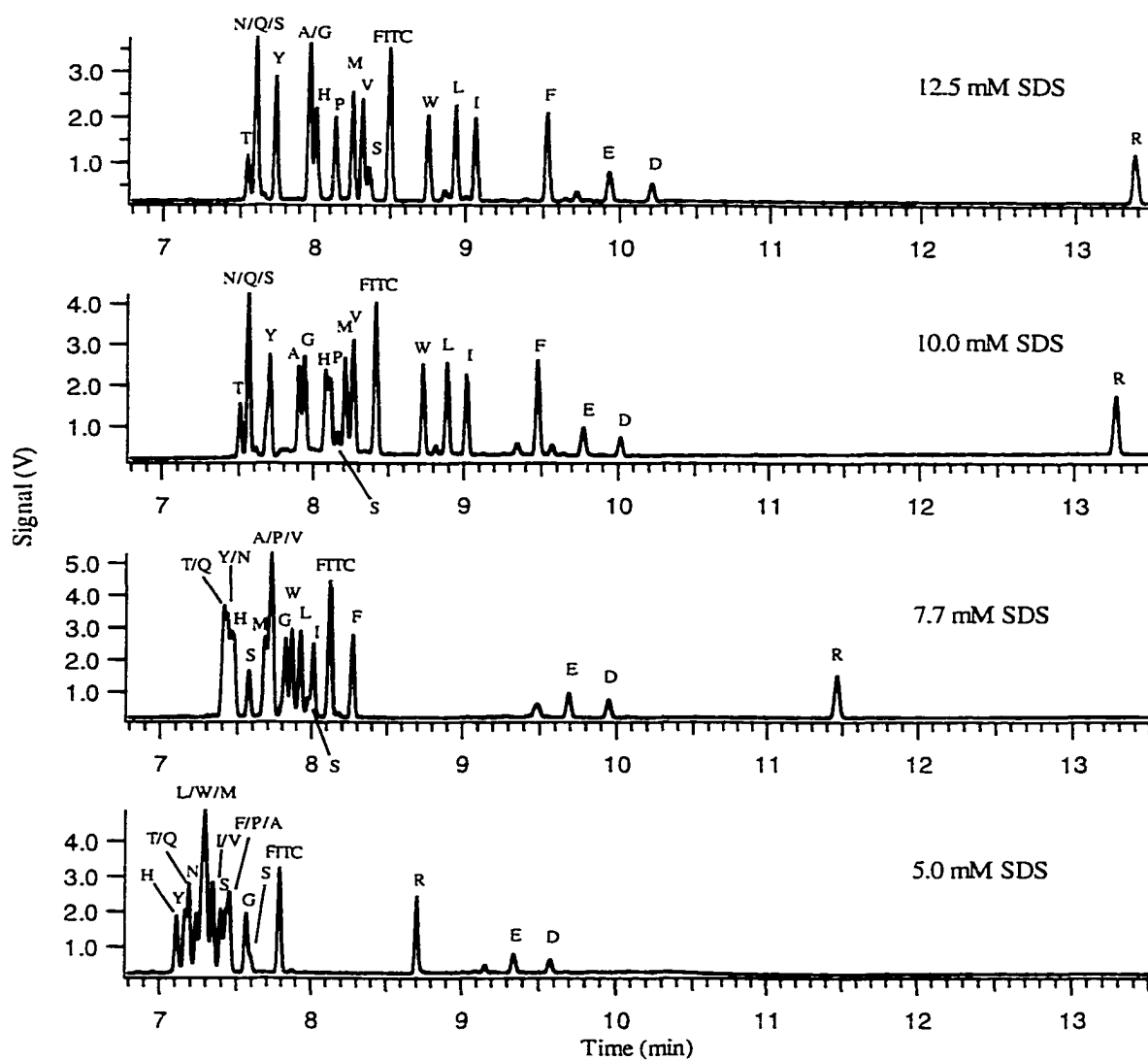


Figure 2.2 Effect of SDS concentration on the separation of 18 FTH-amino acids. The separations were performed in 10 mM phosphate buffer at pH 7.0 with SDS concentrations of 5.0 mM, 7.7 mM, 10.0 mM and 12.5 mM.

Increasing the SDS concentration in the running buffer served to spread out the FTH-amino acid peaks. The total analysis time increased from just over 9.5 min with 5.0 mM SDS in the running buffer to 13.5 min when the SDS concentration was increased to 12.5 mM. Of more importance however, was the effect that increasing the SDS

concentration had on the front of the separation. In the 5.0 mM SDS running buffer there were seventeen peaks in a 0.7 min migration window (7.1 min to 7.8 min). Even with the increased efficiency associated with CE it would not be possible to adequately resolve all of these peaks in such a small window. In contrast, the migration window for the same seventeen peaks increased to 2.1 min (7.5 min to 9.6 min) in the 12.5 mM running buffer. Not surprisingly, this three-fold increase in the migration window was accompanied by a significant improvement in resolution among the FTH-amino acids.

By far the largest increase in migration time with increasing SDS concentration occurred for the FTH-arginine peak. Such a large and singular increase indicates that its interaction with SDS is much stronger than that of any of the other FTH-amino acids. This result is somewhat surprising as partitioning into SDS micelles is more favorable for hydrophobic compounds and arginine has a polar guanidino group on its side chain. As such, it is unlikely that FTH-arginine would partition preferentially in the micellar phase. Instead, it is more likely that the interaction is due to ion pairing between the guanidino side chain, which is protonated in the pH 7.0 running buffer, and the negatively charged dodecyl sulfate anion.

The rest of the FTH-amino acids behaved generally as expected. The more hydrophobic derivatives showed a greater degree of interaction with the micelles than the hydrophilic ones did. This trend can be seen in Figure 2.3, which gives a plot of migration time against SDS concentration for each FTH-amino acid.

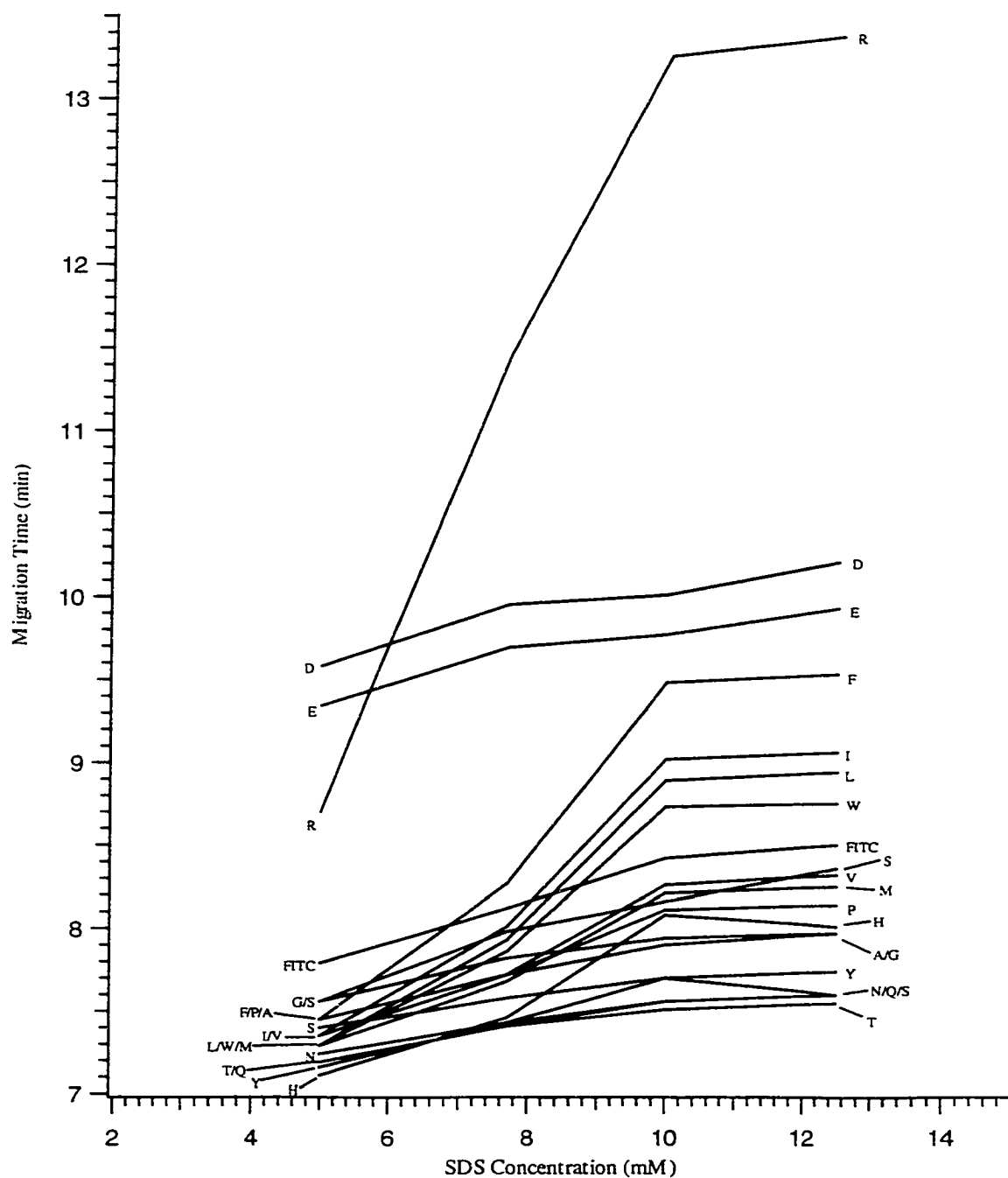


Figure 2.3 Migration time versus SDS concentration for 18 FTH-amino acids. The separations were performed in 10 mM phosphate buffer at pH 7.0 with SDS concentrations of 5.0 mM, 7.7 mM, 10.0 mM and 12.5 mM.

Some general trends become apparent upon inspection of the plot in Figure 2.3. First, migration times increased as the SDS concentration increased. However, this trend was much less pronounced above an SDS concentration of 10.0 mM. And in fact, the earlier migrating FTH-serine species and FTH-histidine actually showed a decrease in migration time when the SDS concentration was increased from 10.0 mM to 12.5 mM.

The second trend was that, as a group, the FTH-amino acids with hydrophobic side chains (A, V, W, L, I and F) showed a larger increase in migration time with increasing SDS concentration than those with hydrophilic side chains. Furthermore, within the group, the degree to which each analyte interacted with the micelles increased with increasing side chain hydrophobicity. This trend can be seen most clearly by comparing FTH-alanine, with a methyl side chain, to FTH-phenylalanine, with a phenyl side chain. FTH-alanine showed only a small increase in migration time while the more hydrophobic FTH-phenylalanine showed an increase in migration time that was second only to that of FTH-arginine, which, as described earlier, has an entirely different mechanism of interaction with SDS.

Although the addition of SDS significantly enhanced the separation of the FTH-amino acids, it did not, on its own, effect separation of all the components. It shifted the more hydrophobic species away from the hydrophilic species but it did not help to separate the hydrophilic species from each other. In fact, when the SDS concentration rose above 7.7 mM, FTH-asparagine, FTH-glutamine coalesced into one peak. The comigration of these two components at higher SDS concentrations is problematic because their structures differ by only a single methylene group. This similarity in structure would be expected to make them one of the more difficult analyte pairs to

separate. As such, it was decided that further improvement in the separation would be attempted by using a running buffer containing 7.7 mM SDS and investigating other avenues for enhancing the resolution of the FTH-amino acids.

One way to improve resolution in CE is to decrease the electroosmotic flow rate (EOF) without changing the applied voltage (32). Because the voltage is fixed, the electrophoretic mobilities of the various analytes will not be affected. Therefore, when the EOF is lowered, analytes will be swept through the separation capillary more slowly with the result that small differences in individual electrophoretic mobilities will be magnified. In effect, the different analyte bands will be given more time to separate.

The most straightforward way to reduce EOF is to lower the pH of the running buffer so that the silanol groups on the capillary wall remain protonated (33). Unfortunately, this approach is not suitable for the FTH-amino acid separation. The running buffer pH would have to be less than 4.5 to significantly reduce EOF, but the fluorescence of fluorescein and its derivatives decreases significantly below pH 7 (34). However, it has recently been shown that the addition of metal cations such as  $Zn^{2+}$ ,  $Cu^{2+}$  and  $Mg^{2+}$  to the running buffer can also significantly reduce EOF (35). The metal cations reduce the change in potential across the diffuse layer between the outer Helmholtz plane and the bulk solution. The electroosmotic velocity is directly related to this zeta potential, so as it decreases so does the EOF. In this work, magnesium acetate was used to decrease the EOF. Figure 2.4 shows the effect on the FTH-amino acid separation of adding 2.0 mM magnesium acetate to the running buffer. Higher concentrations were not explored because they led to precipitate formation in the separation capillary.

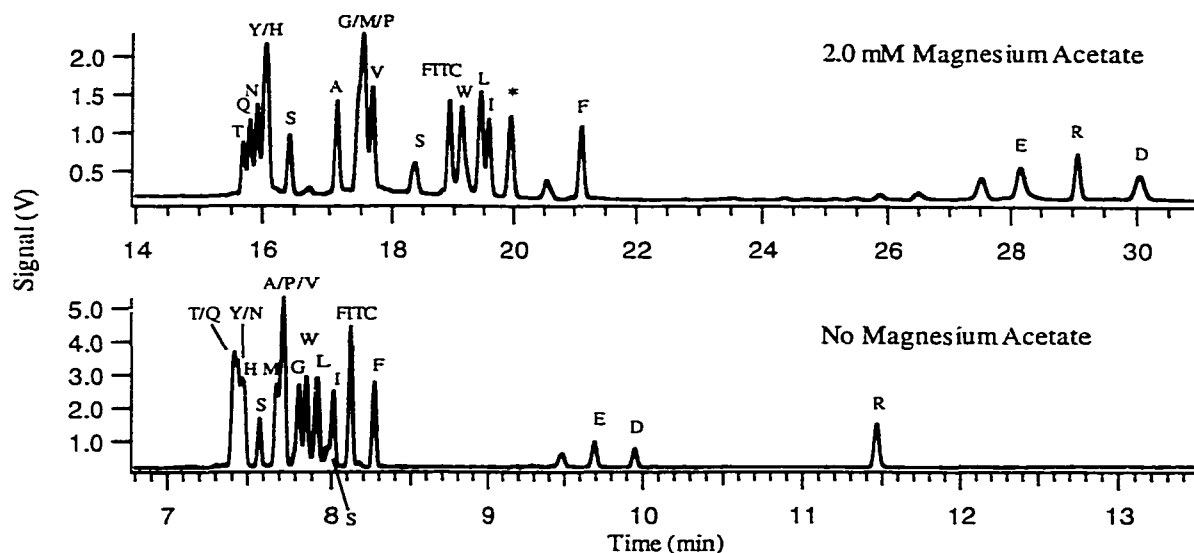


Figure 2.4 Effect of magnesium ion concentration on the separation of 18 FTH-amino acids. The separations were performed in 10 mM phosphate/7.7 mM SDS buffer at pH 7.0 with a magnesium acetate concentrations of 0.0 mM and 2.0 mM respectively.

As expected, the addition of magnesium ions to the running buffer significantly reduced the EOF in the separation capillary. This effect can be seen in the increase in analysis time from 11.5 min to just over 30 min. The decreased EOF also had the desired effect on the FTH-amino acid separation: analyte peaks are more spread out and the resolution at the very front of the separation has improved. There is also an extra peak in the second separation (\*). This peak disappeared when fresh FITC was used, indicating that it is a FITC decomposition product and is not related to the FTH-amino acids. However, it may be produced under the harsher conditions used for protein sequencing, so its effect on the separation cannot be ignored.



More surprisingly, adding magnesium acetate to the running buffer also had an effect on the migration order of the FTH-amino acids. The most striking example of this effect is the way FTH-arginine was overtaken by FTH-aspartic acid. Other changes in migration order involved movement of the more hydrophobic FTH-amino acids to longer migration times relative to the more hydrophilic ones. In this respect it seems that the addition of magnesium acetate affected how the FTH-amino acids interacted with SDS. Those FTH-amino acids that had a significant interaction with the micelles prior to the addition of magnesium acetate saw an increase in their partition coefficients. In contrast, the interaction between FTH-arginine and SDS was greatly decreased when magnesium acetate was added. This difference in behavior for FTH-arginine is further evidence that its mechanism of interaction with SDS differs from that of the other FTH-amino acids.

The change in migration order results from the fact that there is more than one mechanism of separation at work during the analysis. If the separation mechanism was based solely on the different electrophoretic mobilities of the analytes, as in capillary zone electrophoresis, then reducing the EOF would improve resolution without altering the migration order. However, the addition of SDS introduces two other separation mechanisms, partitioning of analytes between the micelles and the running buffer and, in the case of FTH-arginine, ion pairing with the SDS. As such, the migration time of each component derives from the particular balance that exists between the different separation mechanisms that are acting on it. Therefore, the effect of any buffer additive on the overall separation will depend on how it alters the balance between the different separation mechanisms acting on each species.

As with the addition of SDS to the running buffer, the addition of magnesium acetate improved the separation of the FTH-amino acids but it did not lead to the complete resolution all of the components. Further improvements in resolution require a third method for altering analyte migration characteristics. One way to do this is to change the pH of the running buffer. This approach has been used successfully for the separation of PTH-amino acids by capillary electrophoresis with thermo-optical absorbance detection (36, 37). A change in pH can lead to changes in one or both of the electrophoretic mobility and the partition coefficient for a given analyte. In this respect, it represents a method for modifying the FTH-amino acid separation that is complementary to the addition of SDS and magnesium acetate to the running buffer.

As stated earlier, the fluorescent signal generated by the fluorescein moiety begins to decrease as the pH drops below 7 (34). It was therefore decided that the effect of pH on the FTH-amino acid separation would be evaluated between pH 6.6 and pH 7.3. Figure 2.5 shows four electropherograms for the FTH-amino acid separation using running buffers in this pH range.

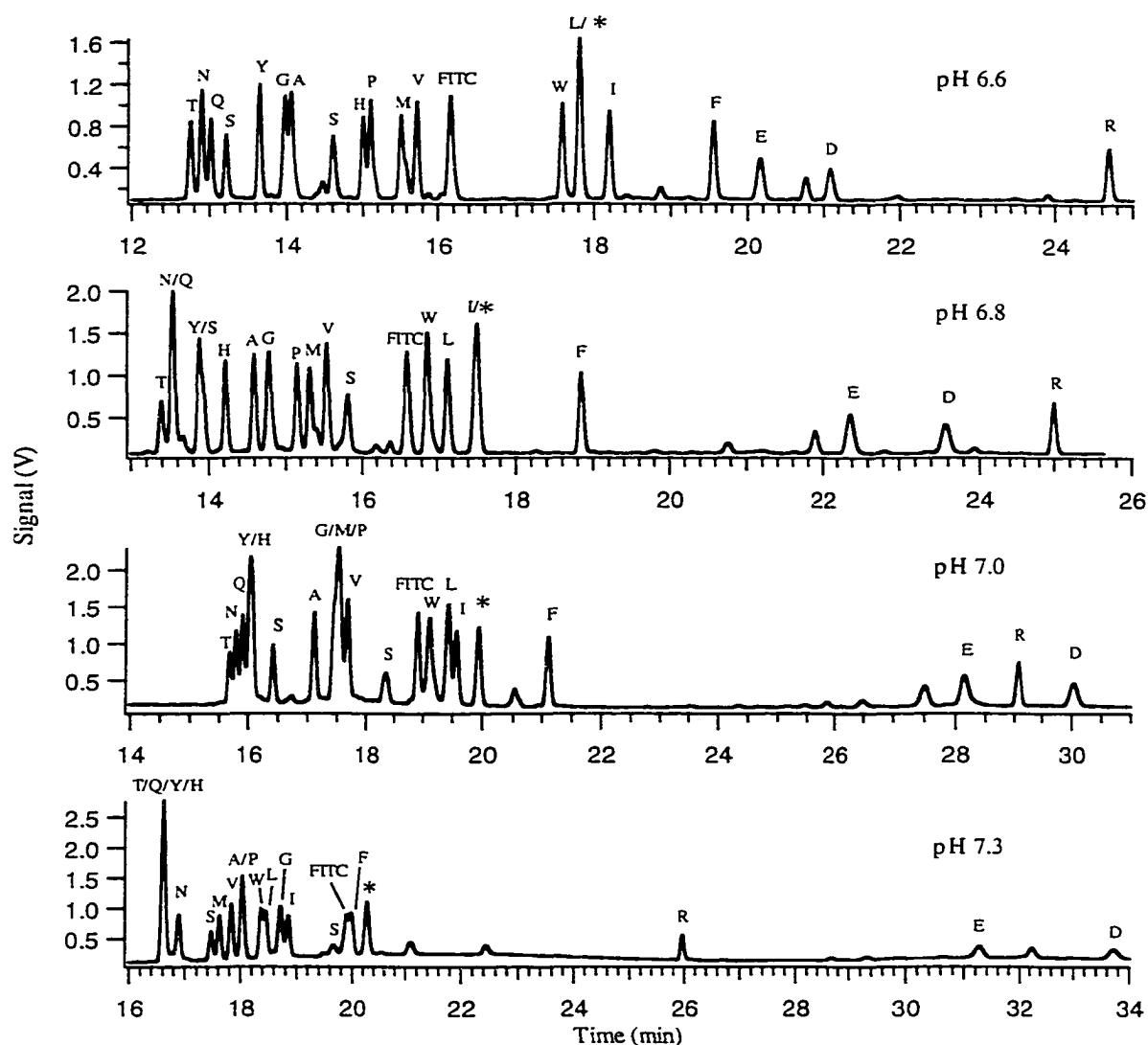


Figure 2.5 Effect of pH on the separation of 18 FTH-amino acids. The separations were performed in 10 mM phosphate/7.7 mM SDS/2.0 mM magnesium acetate buffer at pH 7.3, pH 7.0, pH 6.8 and pH 6.6 respectively.

The data in Figure 2.5 indicates that the running buffer pH had a strong effect on the FTH-amino acid separation. This effect is probably due to a change in the distribution of charged species for each of the FTH-amino acids. The fluorescein moiety has two ionizable protons, one from the carboxylic acid functional group and one from

the phenol. The pKa of the carboxylic acid is approximately 3.5, so it is completely deprotonated at the pH of the running buffers used here. The phenol however, has a pKa of approximately 6.5 (34). Therefore, there will be a combination of both the phenol and the phenolate present for each of the FTH-amino acids in the pH range studied here. Due to the change in charge, the two forms will have different electrophoretic mobilities and will probably have different partition coefficients with the SDS micelles as well. As the pH drops from 7.3 to 6.6, the ratio of the alcohol to the phenolate form will increase. Such a change will alter the migration characteristics of each of the FTH-amino acids. The magnitude of the change will depend on exactly how each component's electrophoretic mobility and partition coefficient is affected. Figure 2.6 shows how the migration times of the various FTH-amino acids varied with running buffer pH.

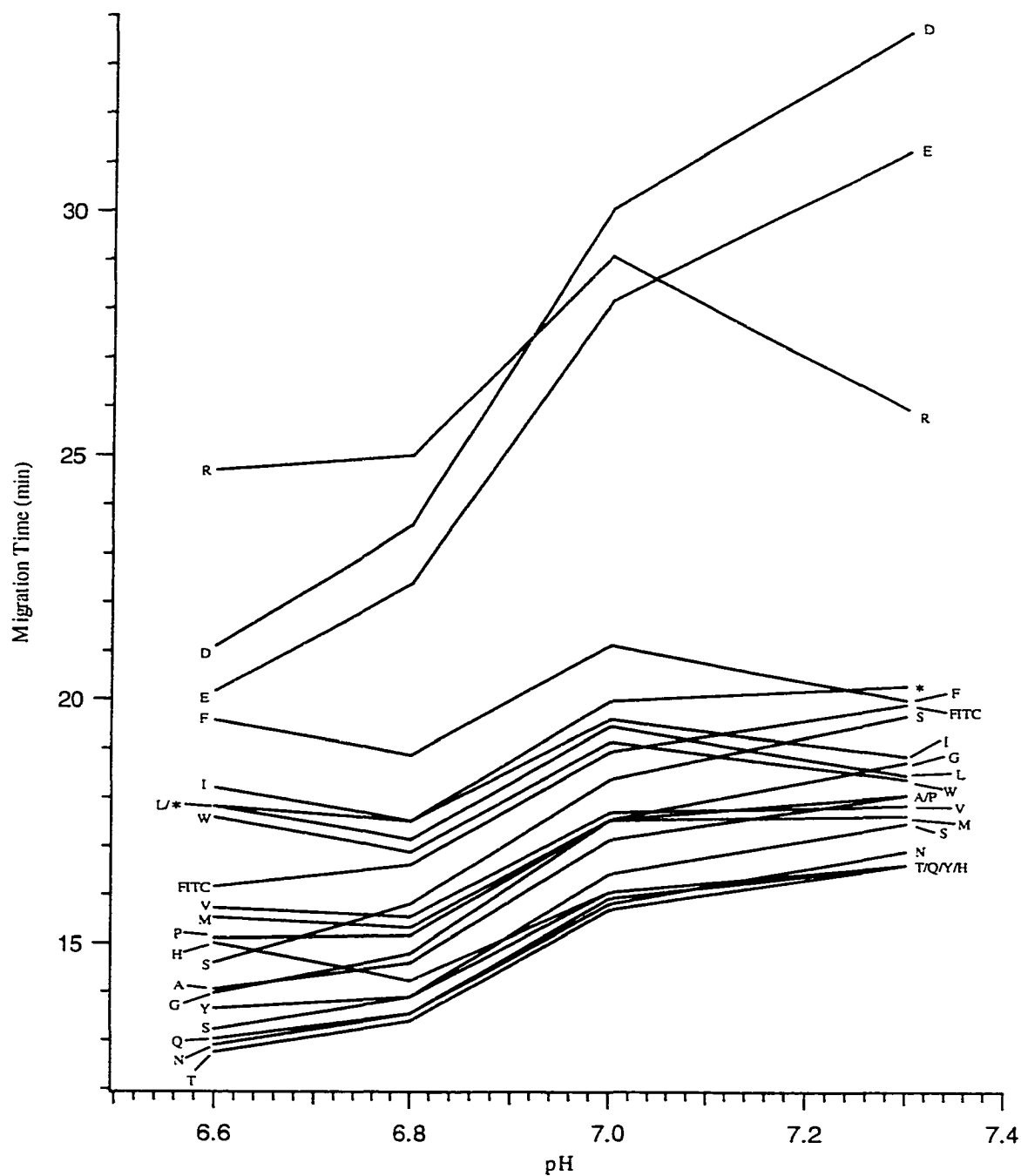


Figure 2.6 Migration time vs. running buffer pH for 18 FTH-amino acids. The separations were performed in 10 mM phosphate/7.7 mM SDS/2.0 mM magnesium acetate buffer at pH 7.3, pH 7.0, pH 6.8 and pH 6.6 respectively.

Lowering the running buffer pH served to spread out the FTH-amino acid peaks. For the most part this effect was due to a decrease in migration time of the more hydrophilic FTH-amino acids, while the more hydrophobic FTH-amino acids showed very little change in migration time. Protonating the phenolate on the fluorescein moiety lead to an increased partition coefficient for hydrophobic species with the micelles. This increase is not surprising because the loss of a negative charge would reduce the electrostatic repulsion between the FTH-amino acids and the dodecyl sulfate anions. The partition coefficient for the more polar FTH-amino acids remained low even after protonation of the phenolate so these species were affected more by the decrease in electrophoretic mobility associated with loss of the negative charge. Because their electrophoretic mobilities opposed the EOF, this decrease led to a decrease in migration time. This effect was most striking with FTH-aspartate and FTH-glutamate. Both of these species have acidic side chains that would make partitioning into the micelles unfavorable even with the loss of the negatively charged phenolate. As such, the decrease in electrophoretic mobility could not be countered through increased interaction with the micelles and their migration times decreased dramatically.

The results in Figure 2.6 show that there is no one running buffer pH that allows for complete separation of all of the FTH-amino acids studied. However, a comparison of the analyses performed at pH 6.6 and pH 6.8 shows that combining the information from both separations allows for the unambiguous identification of all of the components. This point is illustrated in Figure 2.7. Using the complementary separations eliminates any uncertainty in FTH-amino acid identification.

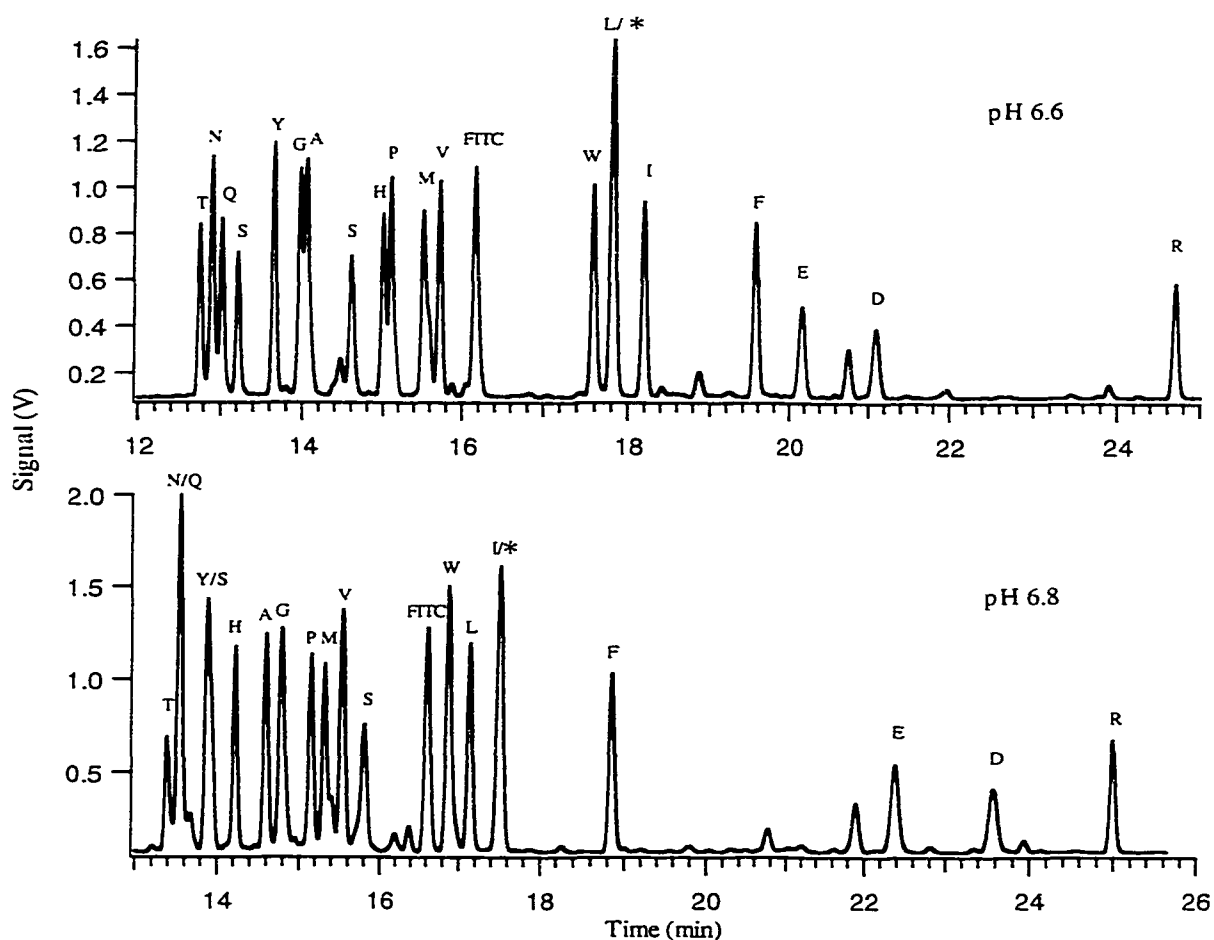


Figure 2.7 Combined separations that allow for the unambiguous identification of 18 FTH-amino acids. The separations were performed in 10 mM phosphate/7.7 mM SDS/2.0 mM magnesium acetate running buffer at pH 6.8 and pH 6.6 respectively.

Although neither separation gives adequate resolution of all the components, any overlap in one of the separations is not present in the other. For example, FTH-asparagine and FTH-glutamine comigrated at pH 6.8 but they were resolved at pH 6.6. Conversely, FTH-alanine and FTH-glycine were baseline resolved at pH 6.8 but they were poorly resolved at pH 6.6. The FITC byproduct peak (\*) interfered with FTH-

leucine in one separation and with FTH-isoleucine in the other. Only by using both separations could the identity of either product be confirmed. Other examples of how the two separations compliment each other include FTH-tyrosine and FTH-serine which are only separated at pH 6.6 and FTH-histidine and FTH-proline which are only baseline resolved at pH 6.8. These results are summarized in Table 2.1.

Table 2.1 Identification capabilities of the combined separations from Figure 2.7.

Amino Acid	Peak Resolved (yes/no)		
	pH 6.8	pH 6.6	Combined
FTH-Ala	yes	no	yes
FTH-Arg	yes	yes	yes
FTH-Asn	no	yes	yes
FTH-Asp	yes	yes	yes
FTH-Gln	no	yes	yes
FTH-Glu	yes	yes	yes
FTH-Gly	yes	no	yes
FTH-His	yes	no	yes
FTH-Ile	yes	yes	yes
FTH-Leu	yes	yes	yes
FTH-Met	yes	yes	yes
FTH-Phe	yes	yes	yes
FTH-Pro	yes	no	yes
FTH-Ser 1	no	yes	yes
FTH-Ser 2	yes	yes	yes
FTH-Thr	yes	yes	yes
FTH-Trp	yes	yes	yes
FTH-Tyr	no	yes	yes
FTH-Val	yes	yes	yes

The table illustrates how performing separations at both pH 6.6 and pH 6.8 makes it possible to identify all of the FTH-amino acids studied. However, the time associated



with analyzing each protein sequencing sample twice would be a serious drawback for any instrument using this detection system. The added time would include not only the second analysis but also the time required to change the running buffers and equilibrate the capillary. Fortunately, the unique design of the CE-LIF system makes it possible to analyse multiple samples simultaneously with only minor modifications to the detection apparatus. In fact, this group has developed CE-LIF instruments that can analyse up to 96 samples simultaneously (38). Using one of these multiple capillary instruments would eliminate the extra time associated with performing two analyses.

One of the figures of merit for any separation is the efficiency of the system. One of the advantages of CE is that it typically results in much higher efficiencies than HPLC. The improved efficiency derives from the plug flow profile that is typical of CE as opposed to the parabolic flow profile that is typical of HPLC (39). The advantage afforded by high efficiency separations is in part due to the decreased peak width that results. It is because of the high efficiency associated with CE that the FTH-asparagine peak at 12.9 min. can be adequately resolved from the FTH-threonine peak at 12.8 min. and the FTH-glutamine peak at 13.0 min in the pH 6.6 separation. Table 2.2 gives the efficiency for each of the 18 FTH-amino acids from the pH 6.6 separation except where otherwise noted.

Table 2.2 Separation efficiencies for the 18-FTH-amino acids analysed by CE-LIF.

Amino Acid	Efficiency
FTH-Ala *	$2.6 \times 10^5$
FTH-Arg	$7.0 \times 10^5$
FTH-Asn	$2.6 \times 10^5$
FTH-Asp	$1.2 \times 10^5$
FTH-Gln	$2.8 \times 10^5$
FTH-Glu	$2.2 \times 10^5$
FTH-Gly *	$2.0 \times 10^5$
FTH-His *	$3.1 \times 10^5$
FTH-Ile	$4.1 \times 10^5$
FTH-Leu *	$3.7 \times 10^5$
FTH-Met	$3.1 \times 10^5$
FTH-Phe	$4.1 \times 10^5$
FTH-Pro *	$2.7 \times 10^5$
FTH-Ser 1	$3.0 \times 10^5$
FTH-Ser 2	$1.8 \times 10^5$
FTH-Thr	$2.4 \times 10^5$
FTH-Trp	$3.1 \times 10^5$
FTH-Tyr	$3.2 \times 10^5$
FTH-Val	$2.7 \times 10^5$

\* Efficiency calculated from the pH 6.8 separation.

Separation efficiency, N, was determined using the equation

$$N = 5.54 \left( \frac{t_m}{w_{0.5}} \right)^2 \quad (2.1)$$

where  $t_m$  is the migration time for each component and  $w_{0.5}$  is the peak width at half maximum.

Separation efficiencies in excess of 100 000 are typical for CE. In this respect the results obtained for the FTH-amino acid separation are not surprising. There is however one surprisingly high value. FTH-arginine had an efficiency of 700 000 theoretical plates. The reason for this high efficiency is unclear but it is probably related to its strong interaction with SDS. In general, the other FTH-amino acids that interacted significantly with SDS also had higher efficiencies than those species that had less interaction with the surfactant. This trend can be most clearly seen in the almost two-fold higher efficiency of FTH-phenylalanine over FTH-glutamate even though the migration times for the two peaks differ by only 0.6 min.

The other, more important, figure of merit for the FTH-amino acid separation is the limit of detection attainable for each analyte. It is this value that will determine the utility of the FTH-amino acids and the CE-LIF system for protein sequencing. The LOD can be calculated using the equation

$$\text{LOD} = 3\sigma_b \left( \frac{C_s}{h} \right) \quad (2.2)$$

where  $\sigma_b$  is the standard deviation in the baseline signal,  $C_s$  is the nominal concentration of the FTH-amino acid and  $h$  is its peak height. The mass LOD can then be determined using the injection volume,  $\text{Vol}_i$ , which can be calculated for electrokinetic injection using the equation (40)

$$\text{Vol}_i = \frac{(\mu_{ep} + \mu_{eo})\pi^2 V_i t_i}{L} \quad (2.3)$$

where  $\mu_{ep}$  is the electrophoretic coefficient of mobility of the analyte,  $\mu_{eo}$  is the electroosmotic coefficient of mobility,  $r$  is the capillary radius,  $V_i$  is the injection voltage,  $t_i$  is the injection time and  $L$  is the capillary length. Equation 2.3 simplifies to the form

$$\text{Vol}_i = \frac{\pi r^2 L V_i t_i}{V_m t_m} \quad (2.4)$$

where  $V_m$  is the running voltage and  $t_m$  is the migration time of the analyte. This equation accounts for the fact that electrokinetic injection induces a bias based on the mobility of each analyte. An analyte with a higher mobility will be preferentially injected onto the capillary. Table 2.3 gives both the concentration LOD and the mass LOD for each FTH-amino acid based on a nominal concentration of  $2 \times 10^{-9}$  M for each component in the analysis mixture.

Table 2.3 Limits of detection ( $3\sigma$ ) for 18 FTH-amino acids by CE-LIF.

Amino Acid	Concentration LOD ( $M \times 10^{-12}$ )	Injection Volume (nL)	Mass LOD (moles $\times 10^{-21}$ )
FTH-Ala	5.0	0.81	4.0
FTH-Arg	9.6	0.47	4.5
FTH-Asn	7.8	0.91	7.1
FTH-Asp	17	0.50	8.4
FTH-Gln	10	0.90	9.4
FTH-Glu	12	0.53	6.5
FTH-Gly	4.8	0.80	3.8
FTH-His	5.2	0.83	4.4
FTH-Ile	9.6	0.65	6.2
FTH-Leu	5.2	0.69	3.4
FTH-Met	5.7	0.77	4.4
FTH-Phe	6.0	0.62	3.7
FTH-Pro	5.4	0.78	4.2
FTH-Ser	13	0.89	12
FTH-Thr	9.3	0.88	8.2
FTH-Trp	4.0	0.70	2.8
FTH-Tyr	7.3	0.86	6.3
FTH-Val	4.4	0.76	3.4

The limits of detection obtained for the FTH-amino acids using CE-LIF are exquisitely low. They represent a six order of magnitude improvement over those attainable using HPLC with UV absorbance detection (41). These results illustrate quite clearly the potential for improvement in protein sequencing sensitivity that exists through the use of CE-LIF for identification of the FTH-amino acids.

It is important to note that the extreme mass sensitivity of CE-LIF derives in part from the small sample volumes that are analysed. In this respect, CE-LIF is incompatible with current commercially available sequencers. These instruments have been designed with the volume requirements of HPLC in mind. As such, simply replacing the HPLC system with a CE-LIF system would result in a volume mismatch that would effectively

preclude any improvement in sensitivity. A new, highly miniaturized sequencer designed with the volume requirements of CE-LIF in mind is required. Such an instrument is described in Chapter 3 of this thesis.

## 2.4 Conclusions

As an analytical method, CE-LIF provides unparalleled sensitivity. This sensitivity makes it an attractive candidate to replace the HPLC with UV absorbance detection system currently used to identify the PTH-amino acid products of protein sequencing. However, it is first necessary to produce a set of fluorescent sequencing products and then to develop a separation for these products by CE-LIF. This work describes such a separation for the FTH-amino acid products from 18 of the 20 coded amino acids. In order to identify all of the FTH-amino acids unambiguously, it was necessary to investigate a number of methods for improving resolution in CE. These methods included using MECC, modifying the EOF in the separation capillary and adjusting the running buffer pH. Even so, it was necessary to use two separations performed at different running buffer pH's to adequately resolve all 18 components. Fortunately, the CE-LIF system can be modified to allow both separations to be performed simultaneously.

The CE-LIF separation of the FTH-amino acids had limits of detection in the low zeptomole range for all of the analytes. It is seven to eight orders of magnitude more sensitive than the best PTH-amino acid separation by HPLC with UV absorbance detection. Such a dramatic improvement in sensitivity indicates that this approach could

well lead to a significant decrease in the amount of protein required for sequencing. Unfortunately, current sequencers are not compatible with the volume requirements of CE. The construction of a sequencer that can be used with the CE-LIF system should make femtomole protein sequencing a reality.

## 2.5 References

1. F. Sanger, *Biochem. J.* **39**, 507 (1945).
2. P. Edman, *Acta Chem. Scand.* **4**, 283 (1950).
3. A. Inglis, *Anal. Biochem.* **195**, 183 (1991).
4. K. Biemann, in *Protein Sequencing, A Practical Approach* J. B. C. Findlay, M. J. Geisow, Eds. (IRL Press, New York, 1989) pp. 99-118.
5. R. Aebersold, *Nature* **343**, 291 (1990).
6. S. Kent, et al., *BioTechniques* **5**, 314 (1987).
7. P. Tempst, C. Geromanos, H. Elicone, Erdjument-Bromage, *Methods: A Companion to Methods in Enzymology* **6**, 248 (1994).
8. L. M. Smith, *Anal. Chem.* **60**, 381A (1988).
9. X. Li, et al., *Talanta* **44**, 401 (1997).
10. Y. Cheng, N. J. Dovichi, *Science* **242**, 562 (1988).
11. K. Muramoto, K. Nokihara, A. Ueda, H. Kamiya, *Biosci. Biotech. Biochem.* **58**, 300 (1994).
12. A. Ducret, E. J. Bures, R. Aebersold, *J. Prot. Chem.* **16**, 323 (1997).
13. I. D. Ireland, et al., *J. Prot. Chem.* **16**, 491 (1997).

14. D. Chen, N. J. Dovichi, *J. Chromatogr.* **657**, 265 (1994).
15. D. Chen, N. J. Dovichi, *Anal. Chem.* **68**, 690 (1996).
16. J. Y. Zhao, N. J. Dovichi, O. Hindsgaul, S. Gosselin, M. M. Palcic, *Glycobiology* **4**, 239 (1994).
17. H. Matsunaga, et al., *Anal. Chem.* **67**, 4276 (1995).
18. V. Farnsworth, K. Steinberg, *Anal. Biochem.* **215**, 190 (1993).
19. A. Tsugita, K. Masaharu, C. S. Jone, N. Shikama, *J. Biochem.* **106**, 60 (1989).
20. O. Imakyure, M. Kai, Y. Ohkura, *Anal. Chim. Acta.* **291**, 197 (1994).
21. H. Hirano, B. Wittmann-Liebold, *Biol. Chem.* **367**, 1259 (1986).
22. S. W. Jin, G. X. Chen, Z. Palacz, B. Wittmann-Liebold, *FEBS Lett.* **198**, 150 (1986).
23. H. Miyano, T. Nakajima, K. Imai, *Biomed. Chromatogr.* **2**, 139 (1987).
24. K. Muramoto, H. Kawauchi, K. Tuzimura, *Agric. Biol. Chem.* **42**, 1559 (1978).
25. H. Maeda, H. Kawauchi, *Biochem. Biophys. Res. Commun.* **31**, 188 (1968).
26. H. Kawauchi, K. Muramoto, J. Ramachandran, *Int. J. Peptide Protein Res.* **12**, 318 (1978).
27. S. Wu, N. J. Dovichi, *J. Chromatogr.* **480**, 141 (1989).
28. E. A. Arriaga, Y. Zhang, N. J. Dovichi, *Anal. Chim. Acta* **299**, 319 (1995).
29. G. E. Tarr, *Methods Enzymol.* **47**, 335 (1977).
30. M. Friedman, L. H. Krull, J. F. Cavins, *J. Biol. Chem.* **245**, 3868 (1970).
31. K. C. Waldron, S. Wu, C. W. Earle, H. R. Harke, N. J. Dovichi, *Electrophoresis* **11**, 777 (1990).
32. J. W. Jorgenson, K. D. Lukacs, *Anal. Chem.* **53**, 1298 (1981).



33. R. M. McCormic, *Anal. Chem.* **60**, 2322 (1988).
34. H. Diehl, R. Markuszewski, *Talanta* **36**, 416 (1989).
35. J. H. Jumppanen, M.-L. Riekkola, *J. Microcol. Sep.* **6**, 595 (1994).
36. M. Chen, K. C. Waldron, Y. Zhao, N. J. Dovichi, *Electrophoresis* **15**, 1290 (1994).
37. K. C. Waldron, N. J. Dovichi, *Anal. Chem.* **64**, 1396 (1992).
38. J. Z. Zhang, N. J. Dovichi, Patent 5,567,294 (US, 1996).
39. D. McManigill, S. A. Swedberg, in *Techniques in Protein Chemistry* T. E. Hugli, Ed. (Academic Press, San Diego, 1989) pp. 468-478.
40. R. A. Wallingford, A. G. Ewing, *Adv. Chromatogr.* **29**, 1 (1989).
41. J. A. Loo, C. G. Edmonds, R. D. Smith, *Science* **248**, 201 (1990).

## Chapter 3

### Development of a Syringe Pump Based Protein Sequencer

#### 3.1 Introduction

In order to understand how a given protein or peptide functions it is important to have a knowledge of its primary structure. By far the most commonly used method for protein sequence analysis used today involves sequential degradation of the sample using isothiocyanate chemistry first developed by Edman in 1950 (1). This method involves coupling of the primary amine site on the amino (N) terminal amino acid of the protein with phenylisothiocyanate (PITC) under basic conditions. Excess PITC and other byproducts are then removed through a series of wash steps so that they will not interfere with the final identification of the labelled amino acid. The derivatised amino acid is then cleaved using anhydrous acid and extracted in organic solvents. The resulting anilinothiazolinone (ATZ) amino acid is converted to the more stable phenylthiohydantoin (PTH) form, which is then identified. The power of the Edman method comes from the fact that the protein is not destroyed during the cleavage reaction. Furthermore, it retains a primary amine moiety at the new N-terminus. It is therefore possible to repeat the entire procedure on the truncated protein. As each successive cycle of the Edman degradation is performed, another amino acid is identified and the primary structure of the protein is revealed.

The repetitive nature of the Edman chemistry lends itself very well to automation. In fact, Edman himself developed the first automated sequencer, the Sequenator, in 1967

(2). This instrument essentially reproduced the conditions used for manual sequencing from the initial coupling step through to extraction of the ATZ-amino acids. The Sequenator, along with similar instruments that followed shortly thereafter, required nanomole amounts of protein for sequencing (2-5). However, the difficulty in isolating many proteins at this scale has since driven a push to improve sequencing sensitivity that continues to this day. Currently, sequencers based on the Edman degradation can reliably analyse proteins and peptides at the low picomole level (6). This increase in sensitivity has resulted from advances including miniaturization and automation of the instrumentation (7-11), optimization of the sequencing chemistry (10, 12), improved protein immobilization (13-16) and enhanced detection sensitivity for the PTH-amino acids (9, 17, 18).

As good as current sequencers are, they still cannot analyse the large number of rare proteins and peptides that can only be isolated in femtomole amounts. Because many of these proteins perform important biological functions, an instrument that is sensitive enough to successfully sequence them would be extremely beneficial for protein researchers. A number of approaches to femtomole Edman sequencing have been investigated. They typically incorporate a modified detection system, often in conjunction with modifications to the degradation chemistry itself, that is much more sensitive than HPLC with UV absorbance (19-23). Chapter 2 of this thesis described one such approach involving the use of capillary electrophoresis with laser induced fluorescence (CE-LIF) detection of fluorescein thiohydantoin (FTH) amino acid sequencing products. Unfortunately, current sequencers are not compatible with CE-LIF. They were designed to be compatible with HPLC detection methods. Consequently, they

work best with sample volumes in the 10  $\mu\text{L}$  to 100+  $\mu\text{L}$  range. In contrast, CE sample volumes are typically 1 nL or less. Coupling a CE-LIF system to a commercial sequencer would therefore, lead to a volume mismatch that would undermine any sensitivity enhancement associated with the improved detector. This chapter describes a miniaturized protein sequencer that has been designed with the volume requirements of CE in mind. By significantly decreasing reagent and sample volumes, it should be possible to realize the sensitivity advantage of CE-LIF and successfully sequence proteins at the femtomole level.

## **3.2 Experimental**

### **3.2.1 Reaction Chamber Design**

Three different reaction chamber designs were evaluated. The first design incorporated polyimide coated fused silica capillaries of different inner and outer diameters (id, od). Up to seven pieces of 50  $\mu\text{m}$  id by 150  $\mu\text{m}$  od capillary tubing were partially inserted into a 700  $\mu\text{m}$  by 850  $\mu\text{m}$  reaction capillary. Epoxy resin was then used to hold the smaller capillaries in place. The free ends of the capillaries were connected to individual syringe pumps and, for argon, to a two way solenoid valve. The protein support sat in the reaction capillary immediately below the reagent delivery capillaries.

The second design was made from 1 mm by 4 mm glass tubing. Short glass side arms were added to a longer central piece of tubing. The side arms acted as inlets into which 100  $\mu\text{m}$  by 350  $\mu\text{m}$  capillaries were inserted. Each capillary was in turn connected

to a syringe pump. Argon was delivered through one end of the central tube. The protein support was inserted through the other end to a position below the last reagent inlet.

The third reaction chamber was made from a solid block of Teflon. A 1mm channel was machined through the centre of the block along its long axis. Six smaller diameter channels (350  $\mu\text{m}$ ) were then drilled so that they intersected the main channel at a 45° angle. Teflon coated fused silica capillaries (100  $\mu\text{m}$  by 350  $\mu\text{m}$ ) were inserted into the small channels. The sequencing reagents were delivered through these capillaries while argon was delivered through Teflon tubing inserted into the top of the central channel. The protein support sat in the main channel below the last reagent port.

### 3.2.2 Temperature Control Unit

The temperature control unit was made from two thermoelectric heating devices (Melcor, Trenton, NJ) mounted on opposite arms of a standard laboratory screw clamp. Copper blocks were attached to both sides of the thermoelectric devices to enhance heat transfer. Temperature control involved adjusting the voltage applied to the heating units using a variable voltage, variable current power supply (Kepco, Flushing, NY). A thermocouple (Cole Parmer, Concord, ON) was used to monitor the reaction chamber temperature.

### 3.2.3 Reagent Delivery System

The reagent delivery system, consisting of syringe pumps and the necessary control circuitry, was built in-house. The pumps were made using stepper motors (Eastern Air Devices, Dover, NH) mounted in pairs to an aluminum frame. A surplus computer power supply capable of delivering 30 amps of current was used to power the motors. Individual Hurst 220001 stepper motor drivers (Emerson, Princeton, IN), a square wave generator and a 5 V power supply controlled step rate and direction for each pump. The design and construction of the control circuitry has been described elsewhere (24). Gas-tight 250  $\mu$ L Hamilton syringes (Chromatographic Specialties, Brockville, ON) were used to deliver all liquid reagents. The syringes were connected to the delivery capillaries (Polymicro Technologies, Phoenix, AZ) with a 1 cm piece of 0.2 mm inner diameter Teflon tubing (LC Packings International, San Francisco, CA) flared at each end to accept the syringe tip and the capillary. Argon was delivered via a two-way solenoid valve (Lee Company, Westbrook, CT) to dry the reaction chamber between steps.

## 3.3 Results and Discussion

A sequencer that is compatible with CE-LIF must be able to work reliably with low to sub-microlitre reagent volumes. This requirement impacts on every aspect of the instrument design. First, the size of the reaction chamber must be significantly reduced to ensure uniform delivery of all reagents over the protein support. Decreasing the size of the reaction chamber then makes it necessary to decrease the size of the heating unit used

for temperature control during the coupling and cleavage reactions. As well, the reagent delivery system must be completely redesigned. The valves used in modern sequencers are too large to accurately and reproducibly deliver reagents at the low microlitre level. In fact, there are currently no valve systems small enough to work reliably at this level.

The protein sequencer described here addresses all of the issues discussed above. It is essentially a valveless instrument that uses syringe pumps to deliver the sequencing reagents to a highly miniaturized reaction chamber. A number of different reaction chambers were tested during instrument development. Each basic design is discussed along with its individual merits and drawbacks. The construction of a miniaturized heating unit made from thermoelectric devices is then described, followed by a description of the syringe pump based reagent delivery system. The entire sequencer is then presented in the configuration in which it will be used.

### 3.3.1 Reaction Chamber Design

In many ways, the reaction chamber is the heart of a protein sequencer. It is here where much of the degradation chemistry takes place. An effective reaction chamber must meet a number of requirements. It must allow for facile delivery of the sequencing reagents to the sample. However, it must do this in such a way that cross contamination is minimized. Furthermore, it has to allow for adequate removal of excess reagents and extraction of the cleaved residue. If even one of these requirements is not met, then the entire sequencing process will fail.

A number of reaction chamber designs were investigated for use with the syringe pump sequencer. The goal was to produce a miniaturized chamber that was more compatible with the volume requirements of CE without compromising any of the required functionality. This section will describe the various reaction chambers that were investigated and discuss their advantages and drawbacks.

The first reaction chamber was made entirely from varying sizes of fused silica capillary tubing. Reagents were delivered through a collection of small capillaries (50  $\mu\text{m}$  by 150  $\mu\text{m}$ ) into a larger capillary (700  $\mu\text{m}$  by 850  $\mu\text{m}$ ) where the protein support was held. The ends of up to seven delivery capillaries could be inserted into the reaction capillary. They were held in place using ordinary epoxy resin. A schematic of this reaction chamber is shown in Figure 3.1.

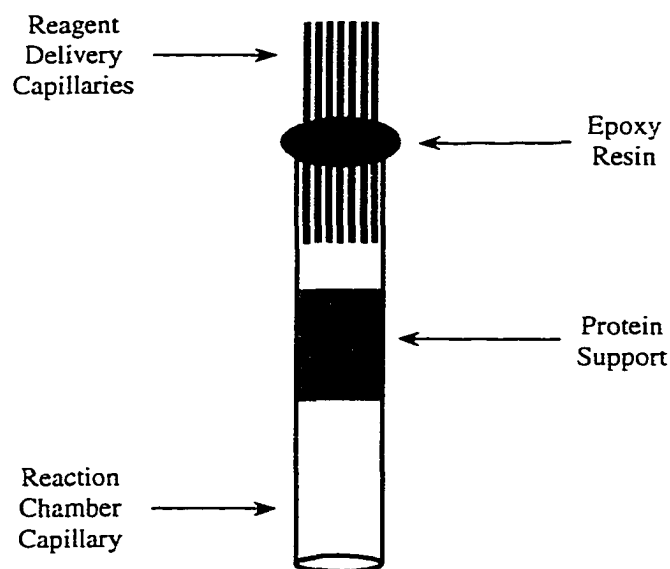


Figure 3.1 Schematic of a capillary based reaction chamber.



The rationale behind designing a capillary based reaction chamber was that it could be made quite easily and it would best match the volume requirements of CE. Unfortunately, the need to pack all of the delivery capillaries tightly together led to serious problems with reagents creeping up via capillary action toward the top of the reaction chamber during delivery. As a result, reagents would end up above the ends of the capillaries where they could not be removed efficiently. This problem led to high levels of byproducts in the sequencing samples. Furthermore, mixing of the basic coupling buffer and the cleavage acid interfered with the degradation chemistry by making it difficult to maintain the required pH during both reactions.

A number of different configurations of the capillary reaction chamber were investigated in an attempt to overcome the reagent mixing and wash problems. Adjustments were made to the number of delivery capillaries along with their relative positions both with respect to each other and the protein support. However, no configuration was found in which seepage of the reagents by capillary action didn't cause significant contamination. A typical capillary reaction chamber is pictured in Figure 3.2.

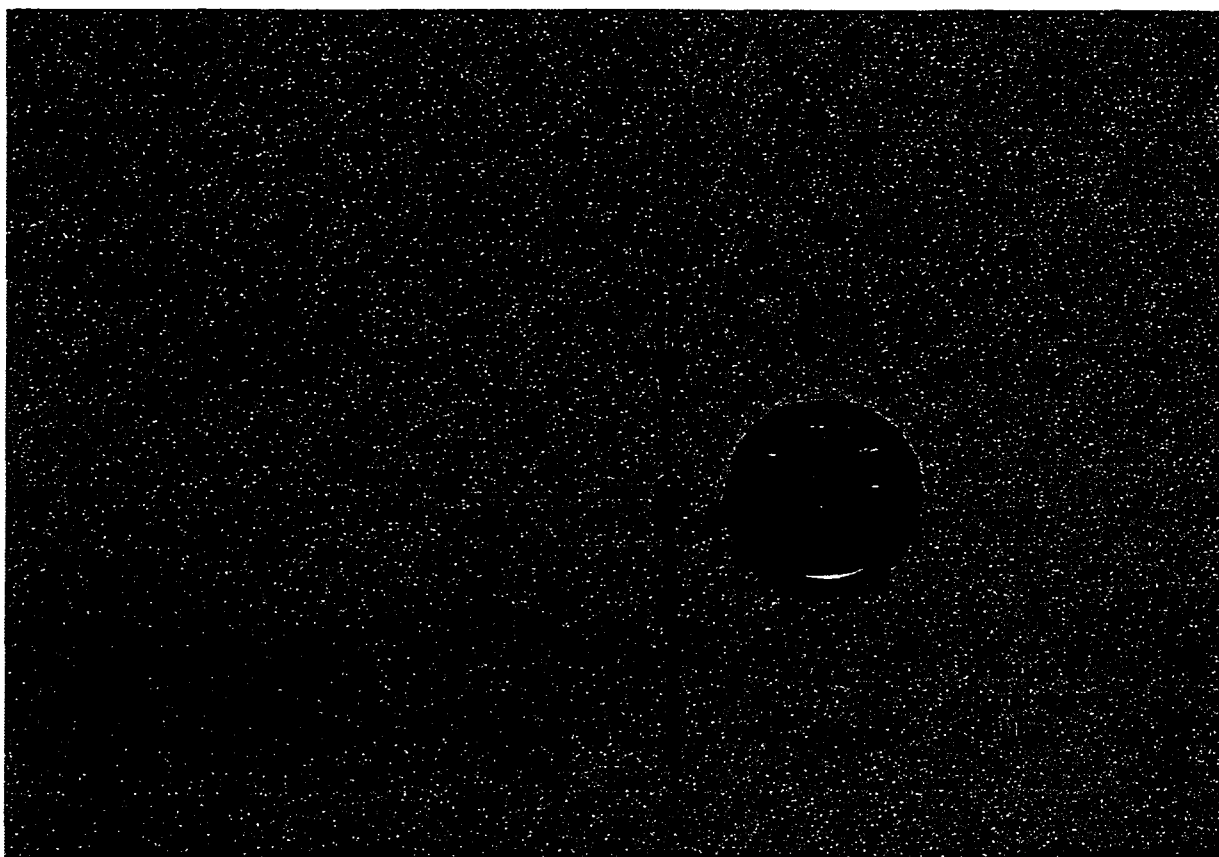


Figure 3.2 Capillary based reaction chamber.

The second reaction chamber design attempted to overcome the contamination problem by spatially separating the various reagent inlets. The chamber was made in house using 1 mm by 4 mm glass tubing. Individual reagent inlets were made by joining short pieces of the same sized tubing to the longer central piece using standard glassblowing techniques. Each inlet was tapered to a diameter of less than 350  $\mu\text{m}$  just before it joined the main channel. Reagent delivery capillaries (100  $\mu\text{m}$  by 350  $\mu\text{m}$ ) were then inserted as far as they would go into each inlet port. As long as the capillaries had

smooth, even ends, this type of connection would allow for low pressure reagent delivery with no leakage. Argon was delivered through Teflon tubing connected to the end of the central channel. A schematic of the glass reaction chamber is shown in Figure 3.3.

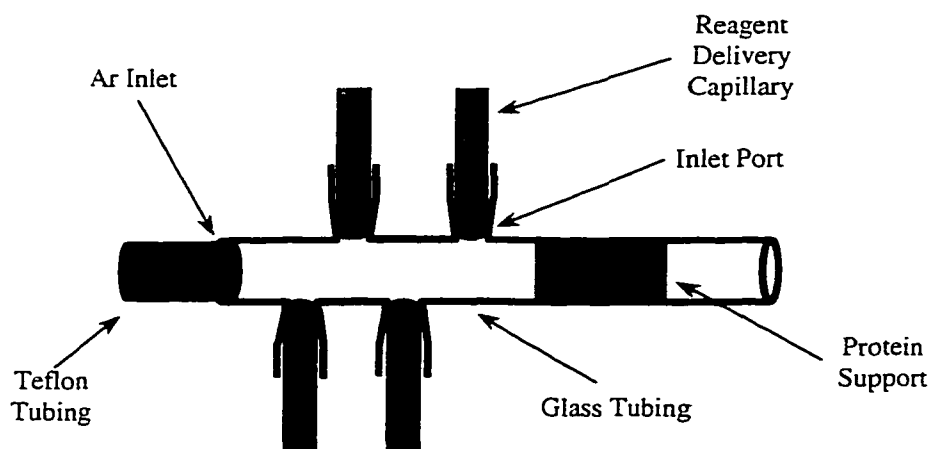


Figure 3.3 Schematic of the glass reaction chamber.

Although the glass reaction chamber did not suffer from the reagent creep problem of the capillary reaction chambers, it had a drawback of its own that led to significant contamination. The tapered inlet could not be positioned so that the capillaries would sit precisely at the edge of the main channel. As a result, a dead volume was produced between the end of the capillary and the main channel. A pool of reagent would collect in this region and mix with the other reagents as they were delivered. This mixing again led to high background and poor pH control during the degradation reactions. A picture of the glass reaction chamber is given in picture 3.4.

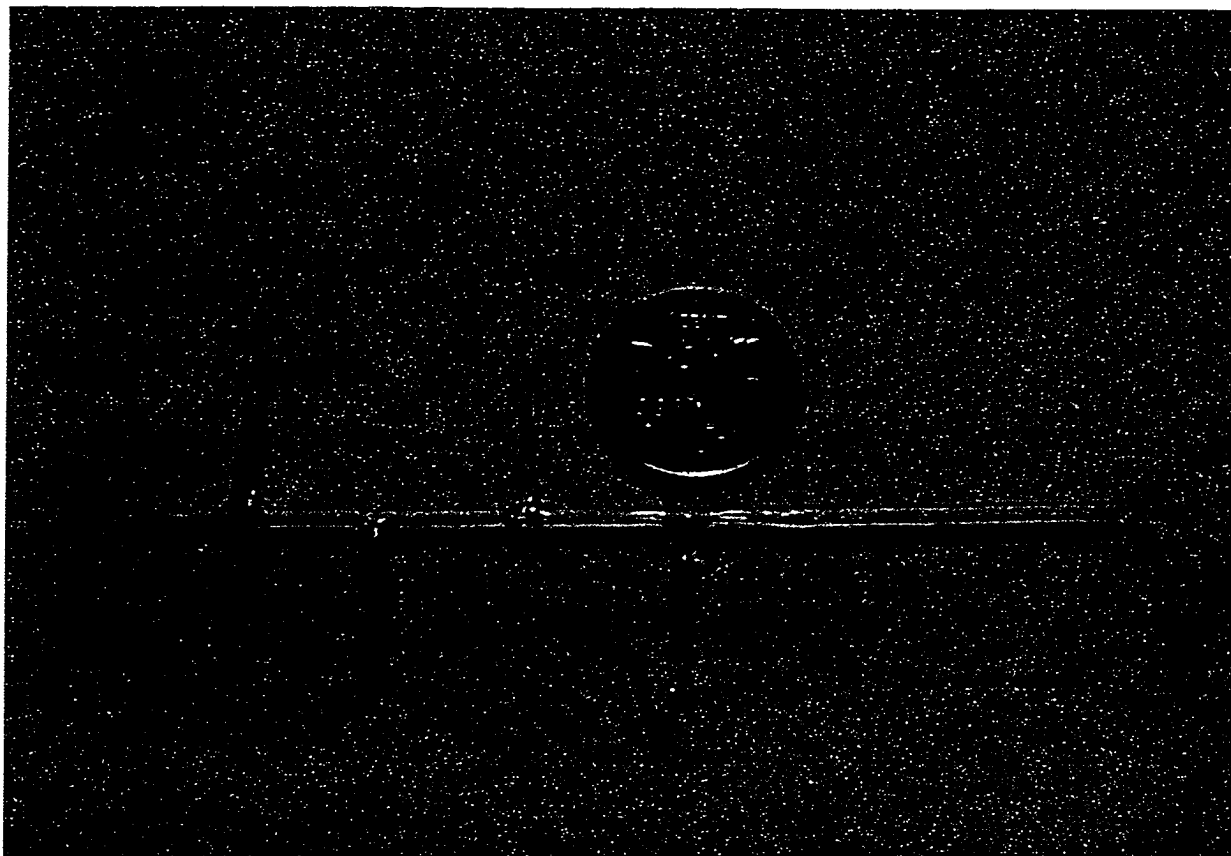


Figure 3.4 Glass reaction chamber.

The third reaction chamber was made from a single block of chemically inert Teflon. A 1 mm central channel was drilled through the block, while five 350  $\mu\text{m}$  channels were drilled so that they intersected it at a 45° angle. Teflon coated capillaries (100  $\mu\text{m}$  by 350  $\mu\text{m}$ ) were inserted into the smaller channels so that their ends were flush with the central channel. Sequencing reagents were delivered through these capillaries, while argon was delivered through a piece of 1 mm od Teflon tubing inserted into the top

of the central channel. The protein support was inserted through the bottom of the central channel to a position just below the last reagent inlet. A schematic and a picture of the Teflon reaction chamber are displayed in Figure 3.5 and Figure 3.6 respectively.

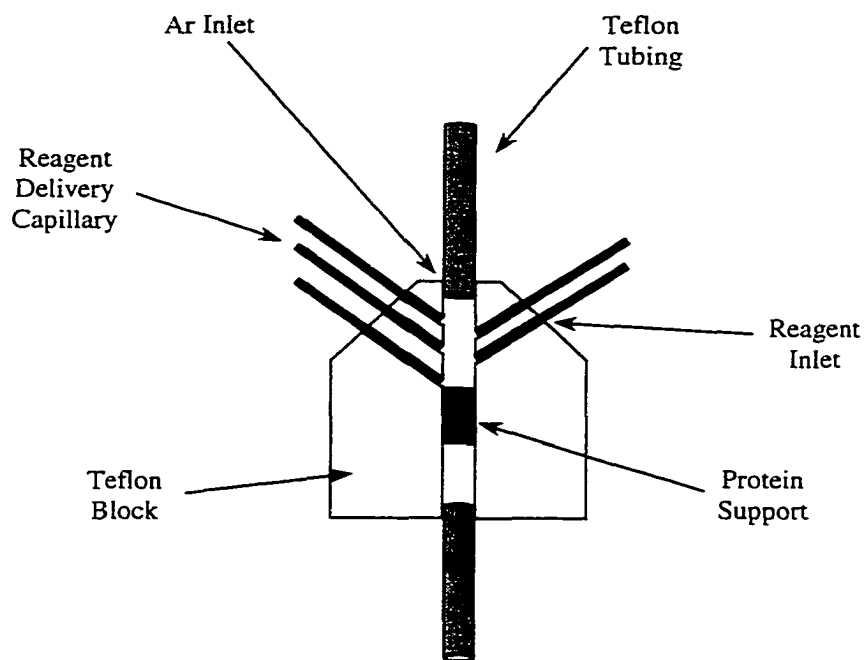


Figure 3.5 Schematic of the Teflon reaction chamber.

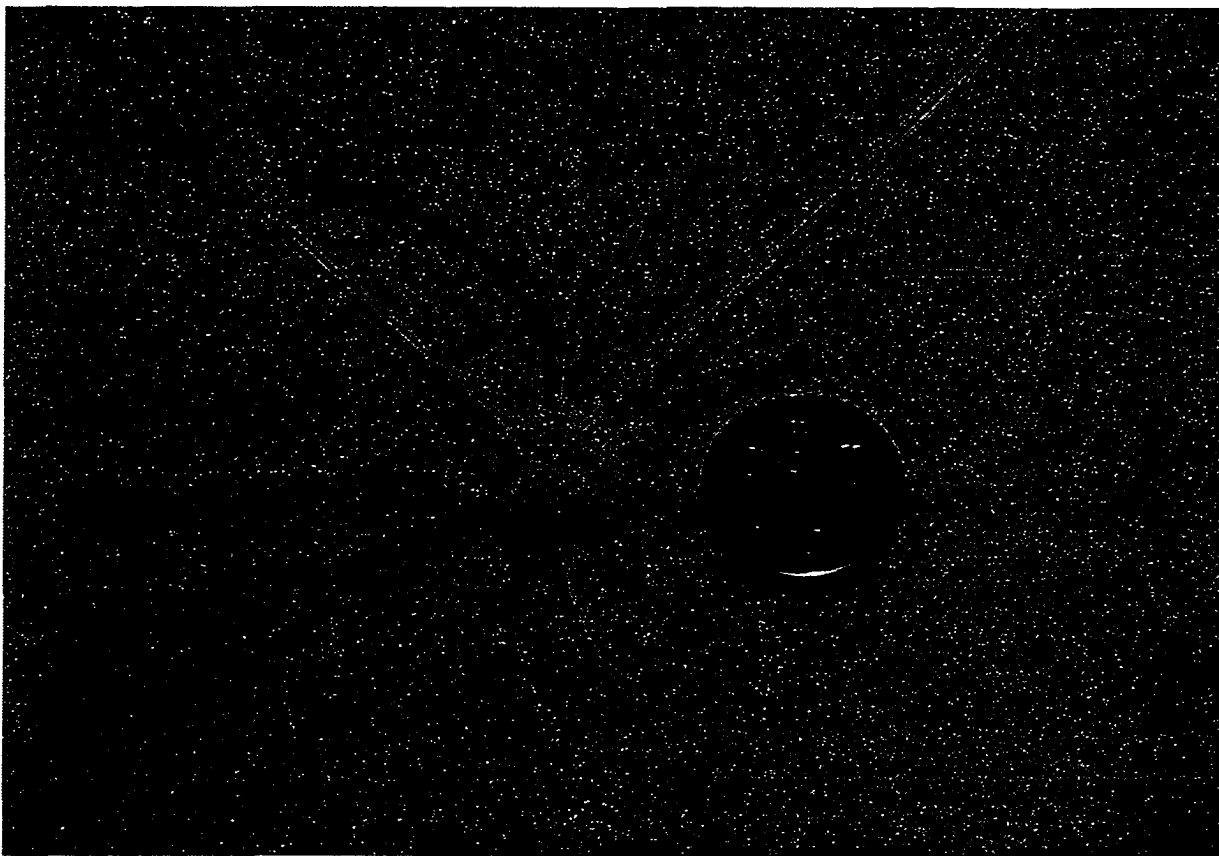


Figure 3.6 Picture of the Teflon reaction chamber.

The design of the Teflon reaction chamber eliminated the problems that plagued both the glass and the capillary reaction chambers. Because the ends of the reagent capillaries were flush with the central channel, there were no dead volume problems like there were with the glass reaction chamber. Furthermore, the tight fit of the delivery capillaries in the inlet channels prevented reagent seepage along the outside of the capillaries during delivery. Consequently, excess reagents could not get trapped between the capillaries and the channel walls from where they would slowly leak out and interfere

with both the degradation chemistry and sample cleanup. These two improvements allowed the Teflon unit to meet all of the requirements mentioned above for a functional reaction chamber. The open central channel made for easy delivery of the reagents to the protein support, while at the same time providing no dead spaces where they could build up. Reduced reagent buildup produced less contamination and more efficient sample cleanup and extraction. Because the Teflon reaction chamber was able to meet the standards set out above; it was used for all of the other instrument development work.

### 3.3.2 Temperature Control Unit

Modifying the heating unit proved to be the most straightforward step in developing the miniaturized sequencer. It was made from two thermoelectric devices (TED's) connected in series and mounted onto either side of a standard laboratory screw clamp. The TED's each had a  $1 \text{ cm}^2$  surface area, which matched the size of the various reaction chambers quite well. Copper plates were attached to either side of the TED's using epoxy resin. These plates served to enhance heat transfer and to provide uniform contact with the different reaction chambers. For the capillary and glass reaction chambers, uniform contact was achieved by milling an appropriately sized channel into the inside plates. The chambers were positioned in these channels and the two halves of the heating unit were clamped around them. The Teflon reaction chamber was simply tightened in place between the plates. A schematic of the heating unit is shown in Figure 3.7.

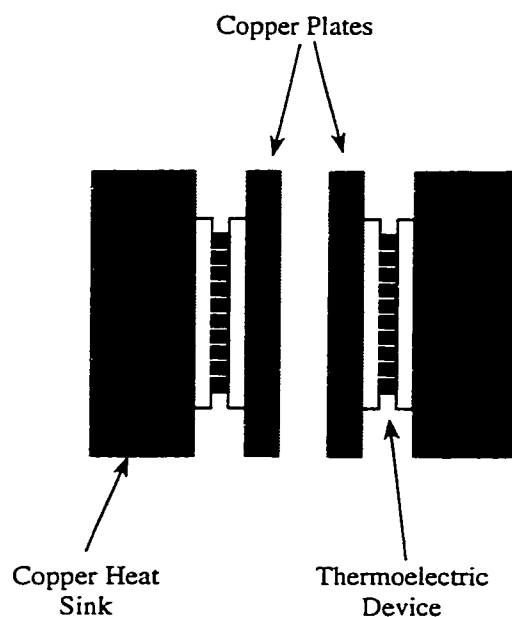


Figure 3.7 Schematic of the thermoelectric heating device.

A variable voltage, variable current power supply was used to control the heating unit. The applied voltage was increased until the desired reaction chamber temperature was reached. A thermocouple was used to measure this temperature. For both the glass and Teflon reaction chambers, it was simply inserted into the main channel where the sequencing sample would be positioned. The capillary reaction chamber however, did not have a large enough opening for the thermocouple. Therefore, the temperature was set by inserting the thermocouple between the TED's and adjusting the applied voltage accordingly. It was then assumed that the reaction chamber reached this same temperature during sequencing. The heating unit is pictured in Figure 3.8 holding the Teflon reaction chamber.



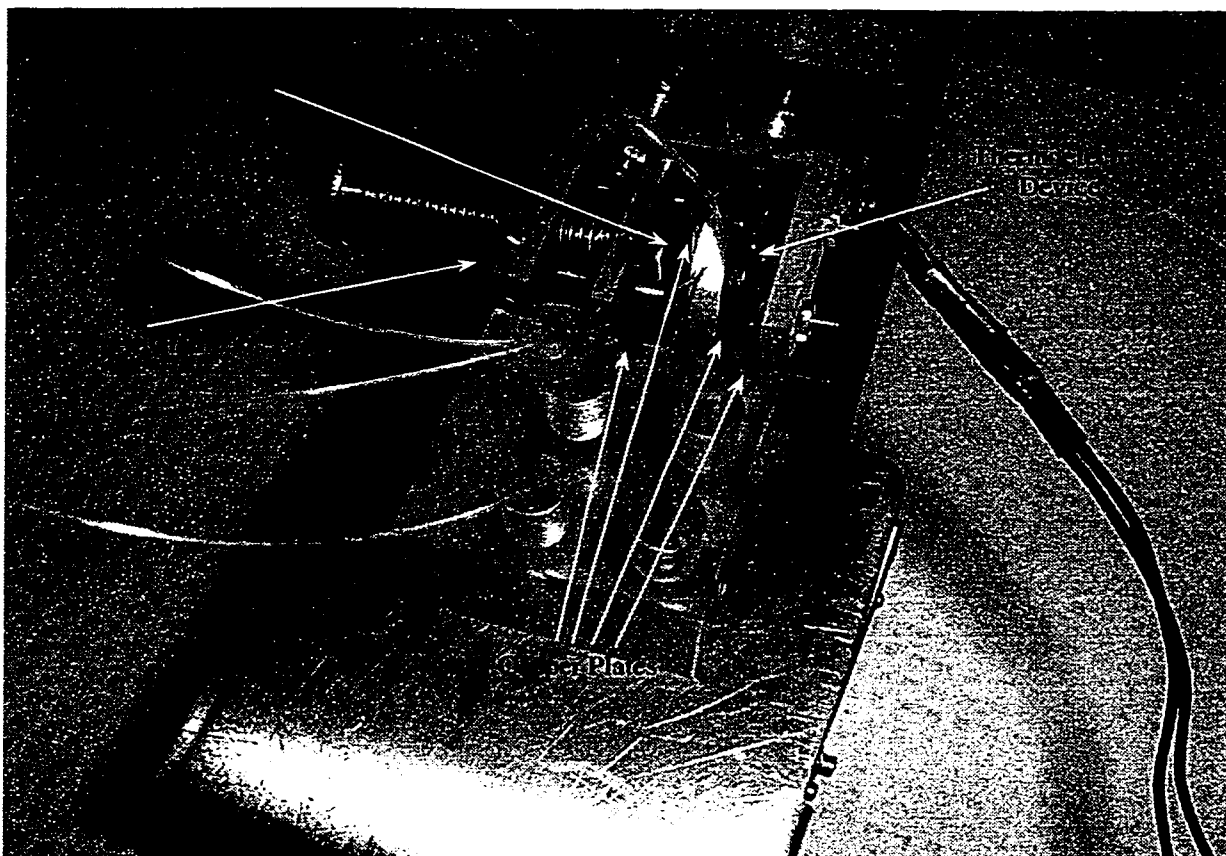


Figure 3.8 Picture of the thermoelectric heating device.

### 3.3.3 Reagent Delivery System

Most commercial protein sequencers use positive pressure displacement for delivery of the various sequencing reagents. Each reagent bottle is connected to an argon

manifold and an outlet valve. When the valve is opened, the pressure of the argon gas forces the reagent through the delivery line to the reaction chamber. The delivery volume is then a function of the argon pressure at the reagent bottle and the length of time that the valve is left open. While this system works well for sequencers coupled to HPLC analysis systems, it would not be compatible with the CE-LIF analysis system described in the previous chapter. The valve system required to control reagent deliveries cannot currently be miniaturized enough to handle the low to sub-microlitre reagent volumes required.

A syringe pump reagent delivery system provides an elegant alternative to the positive pressure displacement system. No valves are required to regulate the flow of the reagents. Instead, delivery volumes are a function of the step size of the pump, the number of steps taken and the size of the reagent syringe. The system would also be easy to miniaturize. For example, a ten-fold decrease in reagent volume can be achieved by leaving the pump settings constant and switching from a 250  $\mu\text{L}$  syringe to a 25  $\mu\text{L}$  syringe. An argon pressure delivery system cannot come close to matching this flexibility. As well, commercial stepper motors with step sizes of  $1.25 \times 10^{-4}$  inches/step are readily available. When combined with low volume syringes, these motors can easily deliver reagents at the sub-microlitre level. A set of six syringe pumps was built using stepper motors like the one described above. This section will describe the pumps and the means by which they were integrated with the reaction chamber and heating unit to produce a protein sequencer compatible with CE-LIF.

The syringe pumps were constructed in-house by mounting two stepper motors onto each of three aluminum housings. A front plate was attached to the end of each

stepper motor screw to force it to move linearly instead of simply rotating with the motor. The plates also pressed against the syringe plungers during reagent delivery. At the front of each housing was a syringe mount. It consisted of a block of aluminum with a V-shaped depression in the middle. A spring loaded clamp was attached to the top of the mount to keep the syringes from moving during reagent delivery or withdrawal. Figure 3.9 and Figure 3.10 shows a syringe pump schematic and an overhead picture of the six pumps respectively.

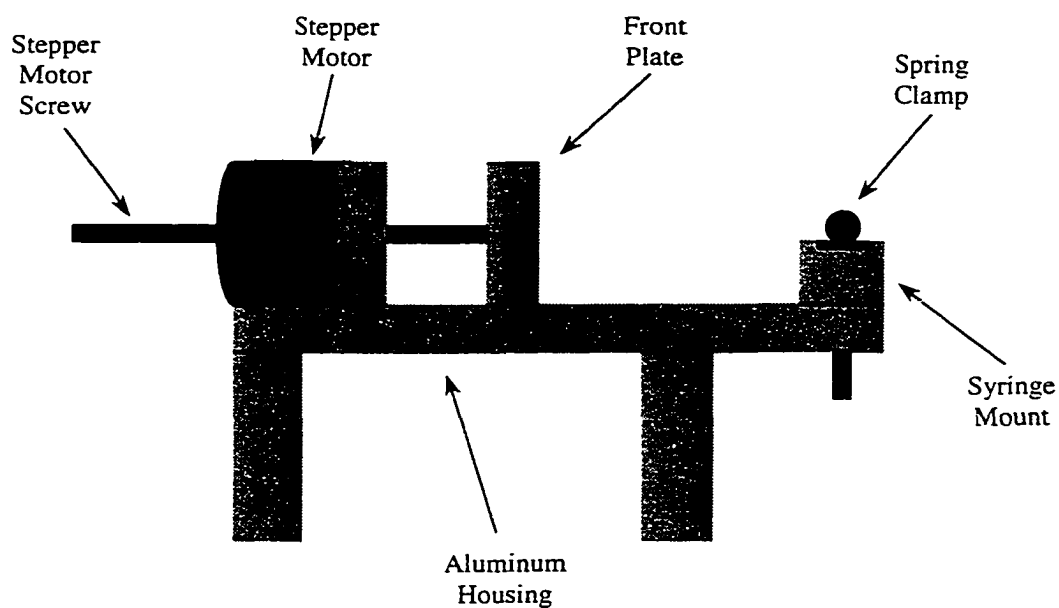


Figure 3.9 Syringe pump schematic.

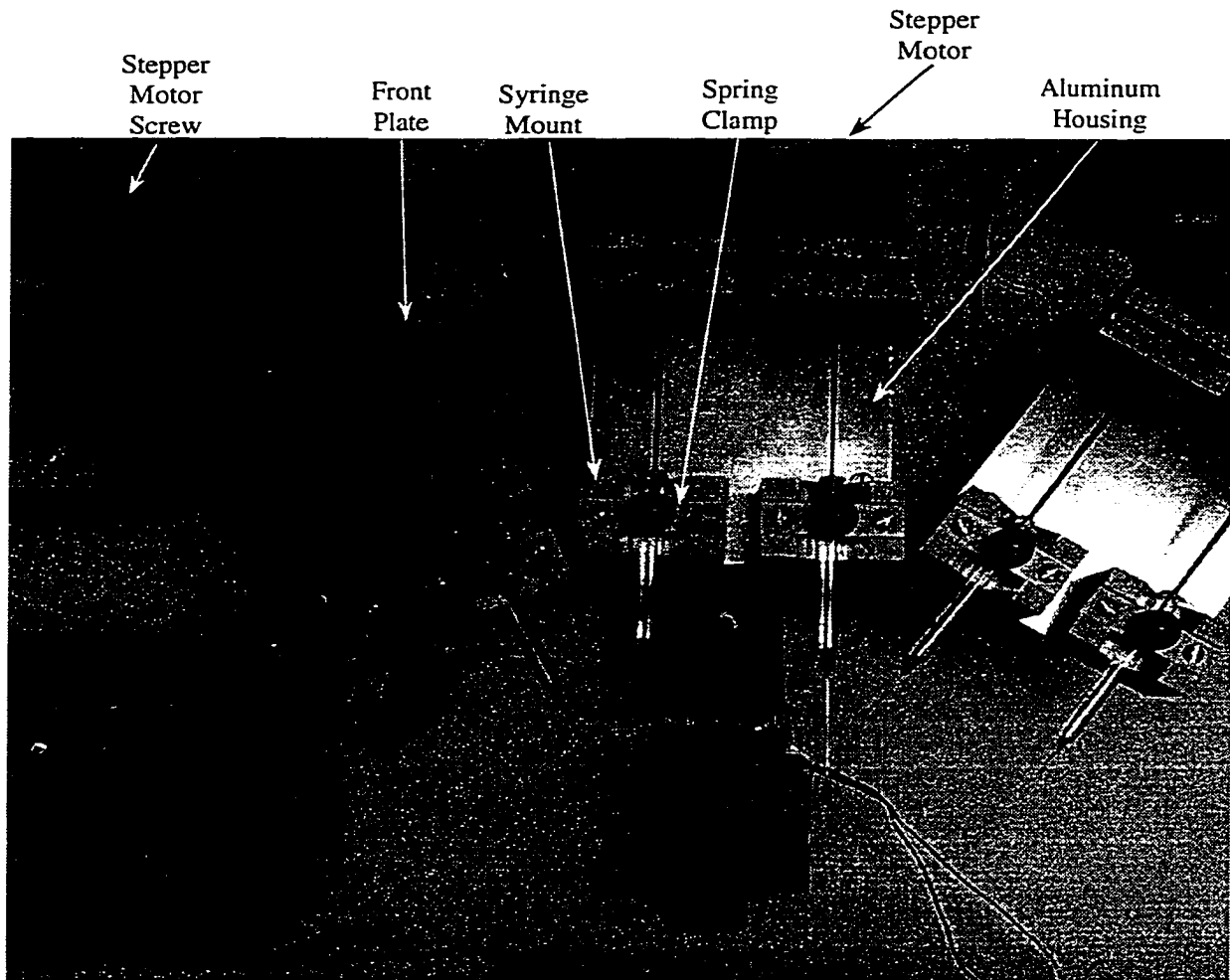


Figure 3.10 Syringe pump picture.

A spare computer power supply capable of delivering 30 amps of current was used to power all six motors. The high current capacity was necessary because the motors drew in excess of 2 amps each. The motors were controlled by individual Hurst 220001 stepper motor drivers. These integrated circuits controlled not only the step sequence of the motors but also the step speed and direction. The step speed was set using a square wave generator and the step direction was set using a standard 5 V power

supply. A single circuit board was used for all six stepper motor drivers. The design and construction of this board has been described elsewhere (24).

The schematic in Figure 3.11 shows how the various components of the miniaturized sequencer were combined to make the complete instrument. The syringes and the reagent capillaries were connected using short lengths of 200  $\mu\text{m}$  id Teflon tubing. Both ends of the tubing were flared so that they would fit over the tip of the syringe and the end of the capillary respectively. Argon, the only reagent not delivered by syringe pump, was delivered to the reaction chamber from a gas manifold via a two-way solenoid valve. The reaction chamber was held in the heating unit simply by tightening the screw clamp (not shown in the schematic). The temperature of the heating unit was maintained by the variable voltage power supply. The signals from the pulse generator, the 5V power supply and the syringe pump power supply were transferred through the individual stepper motor control circuits to the syringe pumps. A picture of the syringe pump sequencer is shown in Figure 3.12.

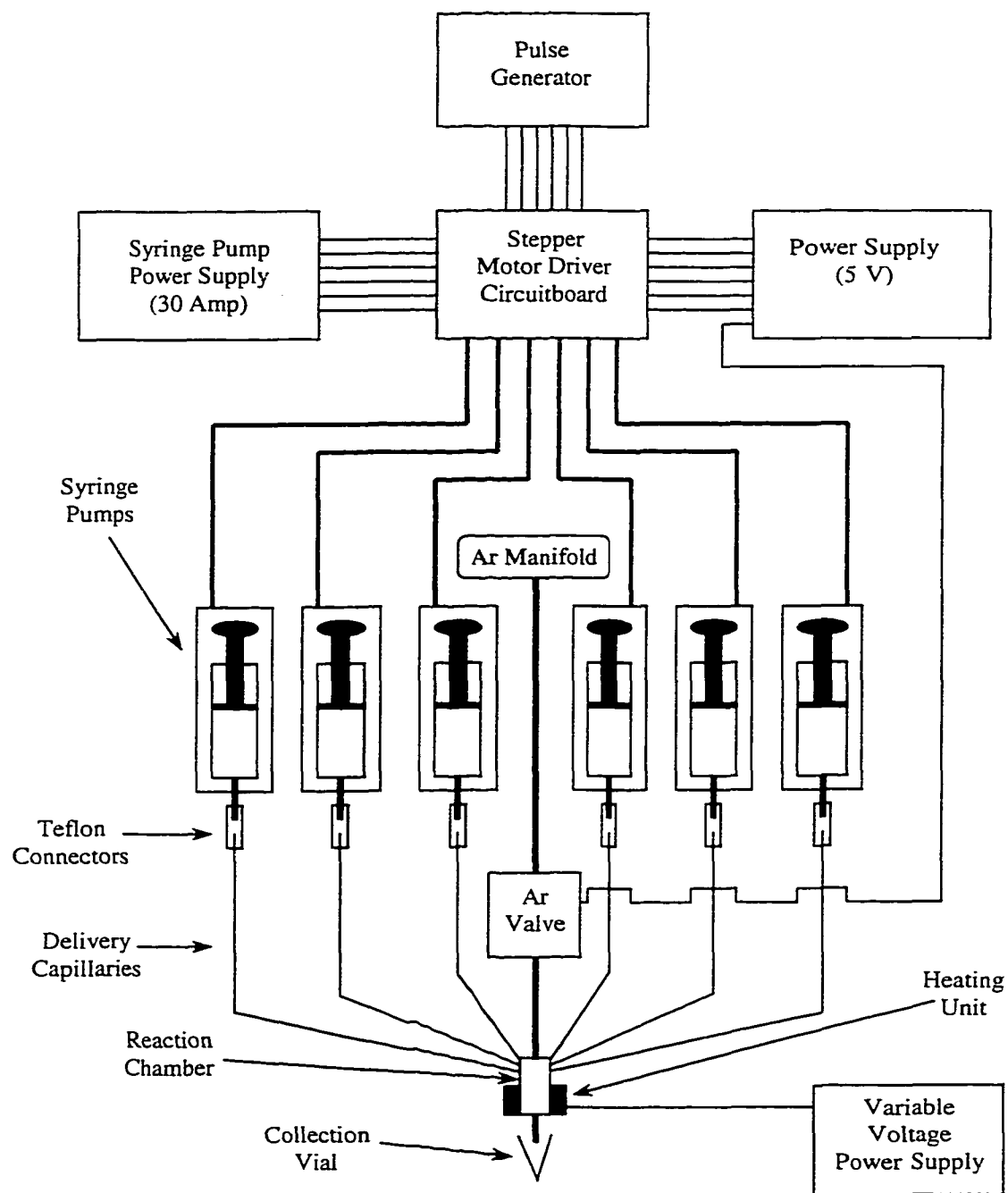


Figure 3.11 Schematic of the syringe pump sequencer.

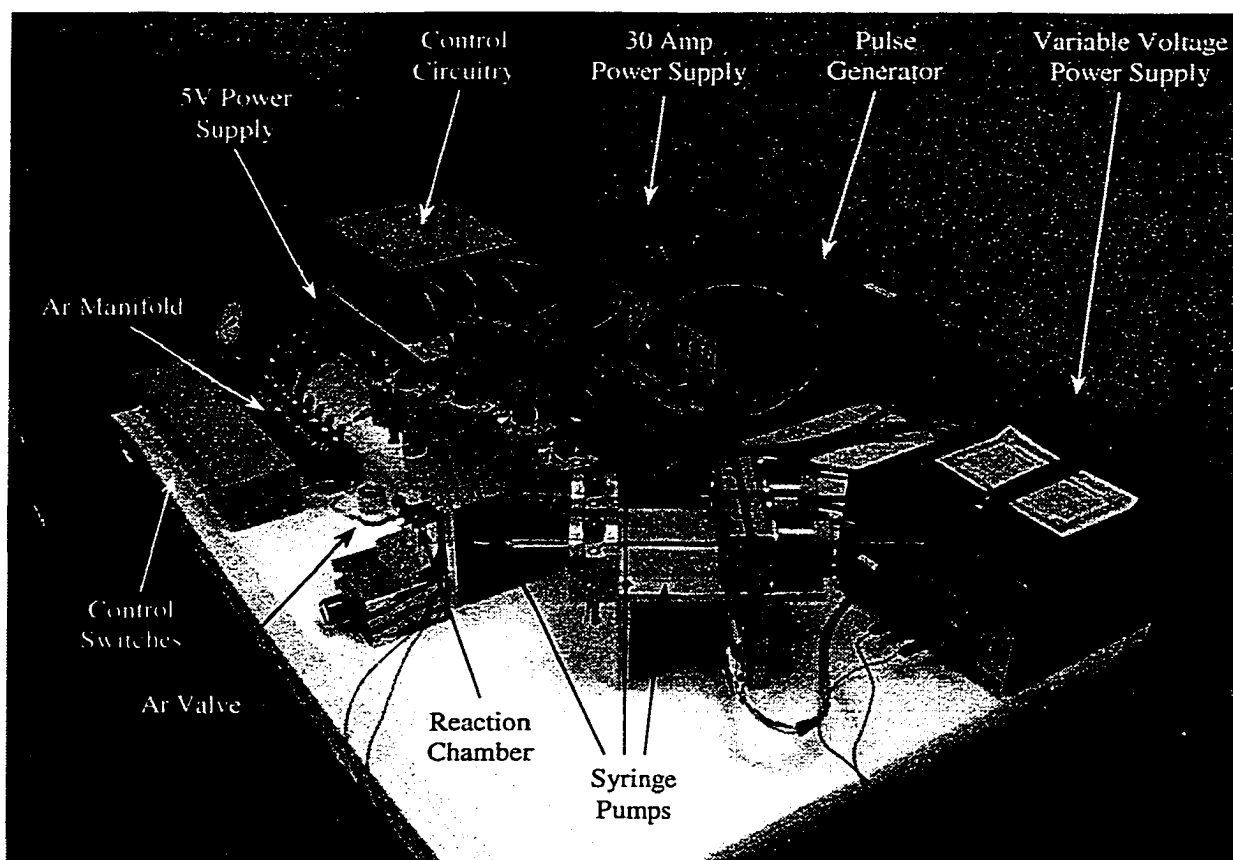


Figure 3.12 Picture of the syringe pump sequencer.

The syringe pump sequencer described here has the potential to be the first sequencer compatible with the volume constraints of CE-LIF. Even the current, first generation instrument is capable of delivering sub-microlitre volumes of the different sequencing reagents. The stepper motors used to make the syringe pumps have a minimum step size of  $1.25 \times 10^{-4}$  inches, while a typical syringe has a barrel length of 3 inches. It therefore takes 24000 steps to deliver the entire contents of a full syringe.

Thus the volume delivered per step for a 25  $\mu\text{L}$  syringe is 1.0 nL. Of course it is unrealistic to expect this kind of performance from the sequencer but it nicely illustrates why the syringe pump design is so promising.

### 3.4 Conclusions

Development of a protein sequencer that is compatible with CE-LIF requires the miniaturization of a number of components including the reaction chamber, heating unit and reagent delivery system. Combining these components produces a sequencer that can perform all of the Edman degradation steps up to extraction of the ATZ-amino acids. Conversion to the PTH-amino acids must be performed off-line due to the lack of appropriate valves for quantitative and selective delivery of the extraction solvent to an appropriate reaction flask. In this respect the syringe pump sequencer is still not yet up to the standards of commercial instruments. However, it is only a first generation instrument designed to prove that Edman-type sequencing can be performed at a volume scale that is compatible with CE-LIF. That question is addressed for the syringe pump sequencer in the next two chapters this thesis.

### 3.5 References

1. P. Edman, *Acta Chem. Scand.* **4**, 283 (1950).
2. P. Edman, G. Begg, *Eur. J. Biochem.* **1**, 80 (1967).
3. T. Hase, K. Wada, M. Ohmiya, H. Matsubata, *J. Biochem.* **80**, 993 (1976).



4. G. Martinez, C. Kopeyan, H. Schweitz, M. Lazdunski, *FEBS Lett.* **84**, 247 (1977).
5. W. T. Butler, J. E. Finch, E. J. Miller, *Biochemistry* **16**, 4981 (1977).
6. R. Aebersold, *Nature* **343**, 291 (1990).
7. B. Wittmann-Liebold, H. Graffunder, H. Kohls, *Anal. Biochem.* **75**, 621 (1976).
8. B. Wittmann-Liebold, *FEBS Lett.* **36**, 247 (1973).
9. W. Machleidt, H. Hofner, in *Methods in Peptide and Protein Sequence Analysis* C. Birr, Ed. (Elsevier/North Holland Biomedical Press, Amsterdam, 1980) pp. 35.
10. R. M. Hewick, M. W. Hunkapiller, L. E. Hood, W. J. Dreyer, *J. Biol. Chem.* **256**, 7990 (1981).
11. J. Calaycay, M. Rusnak, J. E. Shively, *Anal. Biochem.* **192**, 23 (1991).
12. R. A. Laursen, *Eur. J. Biochem.* **20**, 89 (1971).
13. B. Wittmann-Liebold, M. Kimura, in *Modern Methods in Protein Chemistry* H. Tschesche, Ed. (de Gruyter, New York, 1984) pp. 229.
14. M. D. Davison, J. B. C. Findlay, *Biochem. J.* **236**, 389 (1986).
15. E. Wachter, W. Machleidt, H. Hofner, J. Otto, *FEBS Lett.* **35**, 97 (1973).
16. R. Aebersold, G. D. Pipes, R. E. Wettenhall, H. Nika, L. E. Hood, *Anal. Biochem.* **187**, 56 (1990).
17. C. L. Zimmerman, E. Apella, J. J. Pisano, *Anal. Biochem.* **77**, 569 (1977).
18. D. Hawke, P. M. Yuan, J. E. Shively, *Anal. Biochem.* **120**, 302 (1982).
19. M. Chen, K. C. Waldron, Y. Zhao, N. J. Dovichi, *Electrophoresis* **15**, 1290 (1994).
20. K. Muramoto, K. Nokihara, A. Ueda, H. Kamiya, *Biosci. Biotech. Biochem.* **58**, 300 (1994).

21. V. Farnsworth, K. Steinberg, *Anal. Biochem.* **215**, 190 (1993).
22. A. Ducret, E. J. Bures, R. Aebersold, *J. Prot. Chem.* **16**, 323 (1997).
23. I. D. Ireland, et al., *J. Prot. Chem.* **16**, 491 (1997).
24. D. F. Lewis, PhD Thesis, University of Alberta (1999), in preparation.

## Chapter 4

# Evaluation of Fluorescein Isothiocyanate as a Reagent for Modified Edman Protein Sequencing

### 4.1 Introduction

Protein sequence analysis involves a series of chemical reactions that result in cleavage of the derivatized N-terminal amino acid from the protein of interest. The derivatized residue is then identified, typically by HPLC with UV absorbance detection. Repeated cycles of this procedure allow for the identification of successive residues from the protein. In this way its primary sequence is determined.

By far the most commonly used approach to protein sequencing today is the Edman method. In the Edman method, phenylisothiocyanate (PITC) is coupled to the N-terminal residue of the protein under basic conditions. The coupled residue is then cleaved under acidic conditions, converted to a more stable isomer and identified. All four of these processes must be carried out successfully in order for a protein sequencing to work. In addition, there are other steps that, although they need not be considered to understand the basic premise of the Edman method, are still necessary for the method to be practically applicable. These extra steps include protein immobilization, removal of excess labelling PITC and extraction of the anilinothiazolinone (ATZ) amino acid cleavage products. Table 4.1 shows all of the steps involved in a sequencing cycle and the order in which they are performed.

Table 4.1 Steps involved in an Edman protein sequencing cycle.

Step	Description
1	Immobilization of the protein on a solid support.
2	Coupling of the N-terminal amino acid with phenylisothiocyanate.
3	Removal of excess PITC and other byproducts.
4	Cleavage of the derivatized N-terminal amino acid.
5	Extraction of the derivatized amino acid.
6	Conversion of the derivatized amino acid to a more stable form.
7	Identification of the derivatized amino acid.

There are currently a number of instruments on the market that use the Edman method to sequence proteins. These instruments can provide sequence information on low picomole amounts of proteins and peptides (1, 2). However, there are many biologically active proteins and peptides that cannot be easily isolated even at this level (2, 3). Consequently, a great deal of effort has been put into improving the sensitivity associated with protein sequencing (1, 4-15). Much of this work has involved modifications to the Edman chemistry to result in sequencing products that are amenable to high sensitivity analysis (8, 16-24). One such modification involves replacing PITC with fluorescein isothiocyanate (FITC) (6, 25-30).

Chapter 2 of this work describes a method for the analysis of the fluorescein thiohydantoin (FTH) amino acids generated when FITC is used for coupling. The limit of detection (LOD) for the FTH-amino acids is in the zeptomole range for all of the analytes. This LOD represents a six order of magnitude improvement on the sensitivity

of the UV absorbance detection system used with the phenylthiohydantoin (PTH) amino acids generated from the standard Edman chemistry. As such, the use of FITC shows considerable promise for improving sequencing sensitivity.

Changing coupling reagents is a significant modification of the Edman method. It could lead to problems with any one of the seven sequencing steps identified in Table 4.1. As such, it is important to determine which steps, if any, are adversely affected by the use of FITC for coupling. Because the success of any one step depends directly on the success of all of the preceding steps, it is necessary to evaluate the effects of FITC use by starting at the final step and working backwards. Otherwise it would be impossible to determine exactly from where any problems arose. Chapter 2 of this work describes how the various FTH-amino acids can be identified using capillary electrophoresis with laser induced fluorescence detection (CE-LIF). This chapter will describe how each of the remaining six steps in its turn is affected by the incorporation of FITC for coupling.

## **4.2 Experimental**

### **4.2.1 Materials and Reagents**

The FTH-amino acids were identified using the CE-LIF system described in Chapter 2. The separation was carried out in buffers containing varying amounts of sodium dihydrogen phosphate, sodium tetraborate, sodium dodecyl sulphate and magnesium acetate (Sigma Chemical, St. Louis, MO). Polyimide coated fused silica

capillary was purchased from Polymicro Technologies (Phoenix, AZ). The separation conditions used for each step are described in more detail below.

The instrument used to evaluate the various sequencing steps is described in Chapter 3. Coupling buffer (NMM, 5% N-methylmorpholine in 70/30 methanol:water with 0.1% cyclohexylamine) and wash solvent (0.1% cyclohexylamine in methanol) were purchased from Millipore Canada (Mississauga, ON). FITC was purchased from Molecular Probes (Eugene, OR), and was dissolved in HPLC grade acetone (Sigma Chemical, St. Louis, MO) to give a  $1 \times 10^{-3}$  M working solution. This solution was stored at 4°C prior to use. 5% phenylisothiocyanate in heptane and  $\beta$ -lactoglobulin were purchased from Applied Biosystems (Foster City, CA). Anhydrous trifluoroacetic acid (TFA) was purchased from Sigma Chemical (St. Louis, MO). Teflon coated fused silica capillary was purchased from Polymicro Technologies (Phoenix, AZ).

#### 4.2.2 Conversion

A representative set of amino acids consisting of arginine, asparagine, isoleucine, methionine, and proline (Sigma Chemical, St. Louis, MO) was used for the conversion study. Individual fluorescein thiazolinone (FTZ) amino acids were prepared by adding 100  $\mu$ L of  $1 \times 10^{-3}$  M FITC in acetone to 500  $\mu$ L of a  $1 \times 10^{-2}$  M solution of each amino acid in pH 9 carbonate buffer. The reaction was allowed to proceed at room temperature for 4 hr. The coupling buffer was then removed using a C-18 sep-pack. Each sample was applied to the cartridge, washed with water, eluted with methanol and dried on a vacuum concentrator. The dried FTZ-amino acids were reconstituted in 500  $\mu$ L of

methanol prior to use. A mixture of the FTZ-amino acids was created by aliquotting 25  $\mu\text{L}$  of each analyte into a vial and drying off the methanol on a vacuum concentrator. Conversion was carried out in 500  $\mu\text{L}$  of 5% to 25% TFA for 10 min or 25% TFA for 5 min to 40 min. The FTH-amino acids were dried on a vacuum concentrator and stored at 4°C prior to analysis. Analysis was carried out on the CE-LIF system described in Chapter 2. A pH 7.0, 10 mM phosphate/10 mM SDS running buffer was used with a 300 V/cm electric field strength. Injection was at 25 V/cm for 5 s onto a 60 cm long 50  $\mu\text{m}$  id capillary. Samples were reconstituted for analysis in 500  $\mu\text{L}$  of running buffer containing 1% acetonitrile.

#### 4.2.3 Extraction

An FTZ-amino acid mixture was prepared as described for the conversion study and reconstituted in 100  $\mu\text{L}$  of methanol. A 20  $\mu\text{L}$  aliquot of the mixture was applied to a phenylene diisothiocyanate derivatized glass fibre mat (DITC-GFM). After the mat dried, a strip was cut that would fit into the central channel of the reaction chamber described in Chapter 3. Extraction consisted of consecutive deliveries of 4  $\mu\text{L}$  of TFA to the reaction chamber. Each volume of TFA was collected into a separate vial containing 20  $\mu\text{L}$  of distilled, deionized water. Conversion was carried out at 65°C for 10 min. The converted samples were dried on a vacuum concentrator and stored at 4°C prior to analysis. Analysis was performed as described for the conversion study. The first extraction sample was analysed at a 50 times greater dilution than the other samples to prevent detector overload.

#### 4.2.4 Cleavage

The cleavage reaction was evaluated using FITC labelled canine albumin (Molecular Probes, Eugene, OR). A 0.67 mM solution of FITC-canine albumin was prepared in pH 9.0 10mM dibasic sodium phosphate/1% SDS coupling buffer. 30  $\mu$ L of this solution was applied to a DITC-GFM. The mat was heated at 55°C for 1hr. It was then washed three times with 1 mL volumes of coupling buffer followed by three washes with deionized water, methanol and acetone. The coupled mat was viewed on a fluorescence microscope to determine if any protein was present. A strip of the canine albumin-GFM was cut off and inserted into the reaction chamber.

Cleavage consisted of a 3  $\mu$ L delivery of neat TFA followed by a 5 min hold step. The cleaved FTZ-amino acid was extracted with three 4  $\mu$ L TFA deliveries. The extract was collected into 48  $\mu$ L of deionized water and converted at 65°C for 10 min. The extraction and conversion procedures were repeated four times. After conversion the samples were dried on a vacuum concentrator and stored at 4°C prior to analysis. The samples were analysed according to the procedure described for the conversion study.

#### 4.2.5 Wash

The efficiency of FITC removal was evaluated using 0.1% cyclohexylamine in methanol as the wash solvent. A strip of DITC-GFM was loaded into the reaction chamber and subjected to the procedure outlined in Table 4.2. This procedure is a



modification of the method used by Li *et al* for coupling using the standard Edman chemistry (31).

Table 4.2 Procedure used to evaluate wash efficiency.

Step	Reagent Delivered/Function	Time (s)
1	NMM	15
2	FITC	15
3	Hold	300
4	Ar, 5 psi	10
5	FITC	15
6	NMM	15
7	Hold	300
8	Ar	30
9	Wash	240
10	Ar	10
11	Wash	90
12	Ar	10
13	Repeat Steps 9-12 three more times	---

All reagents were delivered at a flow rate of 8  $\mu\text{L}/\text{min}$ . The wash solvent from each 90 s delivery was collected, diluted to 500  $\mu\text{L}$  in running buffer and analysed according to the procedure described in the conversion study.

#### 4.2.6 Coupling

A peptide sequencing standard (Sigma Chemical, St. Louis, MO) was covalently attached to a DITC-GFM according to the method described in section 4.2.4 of this chapter. The sequence of the standard is given below:

Tyr - Ala - Glu - Gly - Asp - Val - His - Ala - Thr - Ser - Lys - Pro - Ala - Arg - Arg

A single sequencing cycle was performed according to the protocol in Table 4.3. All reagents were delivered at a rate of 8 $\mu$ L/min. The TFA was collected into 36  $\mu$ L of distilled water. Conversion was carried out at 65°C for 10 min. The converted sample was dried on a vacuum concentrator. The dried sample was reconstituted in 500  $\mu$ L of running buffer containing 1% acetonitrile. Analysis was carried out in a 60 cm long, 30  $\mu$ m inner diameter fused silica capillary. The running buffer was composed of 10 mM phosphate/3.8 mM borate/7.7 mM SDS/2.0 mM magnesium acetate at pH 7.03. The sample was injected at 50 V/cm for 5 s and analysed at 300 V/cm.

Table 4.3 Sequencing protocol for Cycle 1 of the peptide standard.

Step	Reagent Delivered/Function	Time (s)
1	NMM	15
2	FITC	15
3	Hold	300
4	Ar, 5 psi	10
5	FITC	15
6	NMM	15
7	Hold	300
8	Ar	10
9	Repeat Steps 5-7	---
10	Ar	30
11	Wash	720
12	Ar	30
13	Wash	720
14	Ar	180
15	TFA	23
16	Hold	300
17	TFA	30
18	Ar	3
19	Repeat Steps 17-18 twice more	---

#### 4.2.7 Protein Attachment to Solid Supports

Derivatised glass fiber mats were prepared according to the method of Aebersold *et al* (32). The underivatised glass fibre mats (Applied Biosystems, Foster City, CA) were washed in anhydrous TFA for 2 hr, then dried under Ar for 5 hr. Next, they were immersed in a solution of 2% (v/v) 3-aminopropyltriethoxysilane (Sigma Chemical, St. Louis, MO) in 95% acetone:water for 2 min. The silanized mats were washed ten times with 3 mL volumes of acetone for 5 min each time. The mats were then dried under Ar. The silane linkage was cured by heating the mats at 110°C for 45 min. Efficiency of silanization was monitored by coupling FITC with a derivatised GFM in pH 9 carbonate buffer. Coupling was carried out at 58°C for 45 min followed by five washes with 3 mL of carbonate buffer, five washes with 3 mL of distilled water and five washes with 3 mL of acetone. The coupled GFM was compared to an unsilanized GFM that had undergone the same coupling protocol using a fluorescence microscope.

Twenty 9 mm diameter aminopropyltriethoxy-GFM's (AP-GFM's) were coupled with 10 mL of 0.3 M phenylenediisothiocyanate (DITC) (Sigma Chemical, St. Louis, MO) in 10% (v/v) pyridine/DMF. Coupling was carried out for 2 hr with mixing. Coupled GFM's were washed with four 5 mL volumes of methanol and four 5 mL volumes of acetone, dried under Ar and stored at 4°C prior to use. Efficiency of DITC coupling was monitored with [<sup>35</sup>S]-methionine ([<sup>35</sup>S]-Met) (Amersham, Oakville, ON). Each mat was soaked with 1.6 μL of 30 mM [<sup>35</sup>S]-Met and 20 μL of 10 mM of dibasic sodium phosphate:1% SDS coupling buffer at pH 9.0. Coupling proceeded at 55°C for 1 hr. Coupled DITC-GFM's were washed with 2×1 mL coupling buffer,

4×1 mL deionized water, 3×1 mL methanol and 3×1 mL methanol. Quantitation of bound [<sup>35</sup>S]-Met was carried out on a scintillation counter.

### 4.3 Results and Discussion

#### 4.3.1 Conversion

Conversion of the FTZ-amino acids into their FTH-amino acid counterparts involves the same reaction as the conversion of ATZ-amino acids to PTH-amino acids. However, substituting the fluorescein moiety for the phenyl moiety may lead to a change in the rate or efficiency of the reaction. In this work, both the %TFA and the reaction time were varied to determine what effect, if any, they had on conversion efficiency. In order to simplify the interpretation of the data, a set of five amino acids – arginine, asparagine, isoleucine, methionine and proline – was used to evaluate the conversion reaction. The first four were chosen because their side chains are, in turn, basic, acidic, non-polar, and polar. Proline was chosen because it is the least reactive amino acid with the Edman chemistry.

The effect of TFA on conversion was evaluated using solutions of 5%, 10% and 25% acid in distilled water. Figure 4.1 shows the results obtained for the different TFA concentrations.

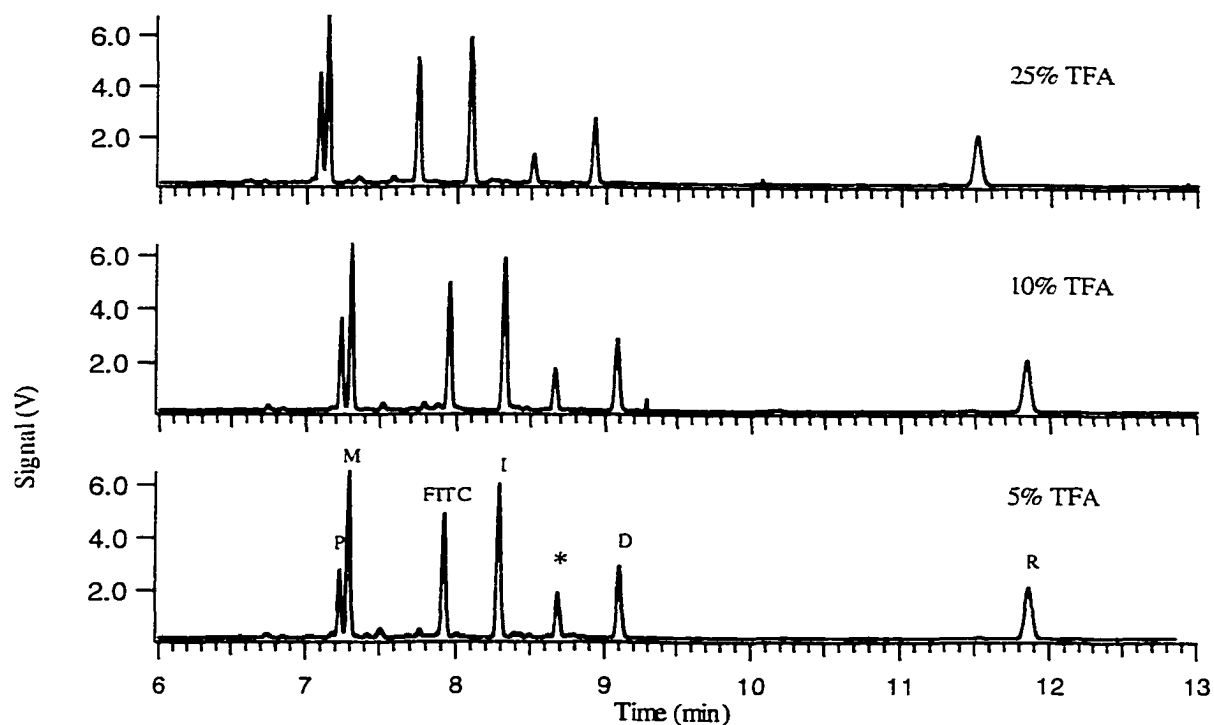


Figure 4.1 Effect of TFA concentration on the conversion efficiency for five FTH-amino acids.

All of the electropherograms look quite similar. Only the FTH-proline peak appears to change significantly with %TFA. This observation is borne out by the comparison of peak heights shown in Table 4.4.

Table 4.4 FTH-amino acid peak heights at different TFA concentrations.

TFA Conc. (%)	Peak Height (V)				
	FTH-Pro	FTH-Met	FTH-Ile	FTH-Asp	FTH-Arg
5	2.6	6.3	5.8	2.7	1.9
10	3.5	6.2	5.7	2.7	1.9
25	4.4	6.6	5.7	2.6	2.0

The FTH-proline peak height increased with increasing TFA concentration. Conversion of the other four FTZ-amino acids did not appear to be sensitive to changes in TFA concentration.

There is one unknown peak (\*) in the three electropherograms shown in Figure 4.1. Its peak height decreased as the TFA concentration increased. This fact, combined with the corresponding increase in the FTH-proline peak, suggests that the unknown peak results from unconverted FTZ-proline.

The effect of reaction time was studied by allowing the conversion reaction to proceed for 5 min, 10 min, 20 min and 40 min in 25% TFA in distilled water. The electropherograms obtained for the various conversion times are shown in Figure 4.2.

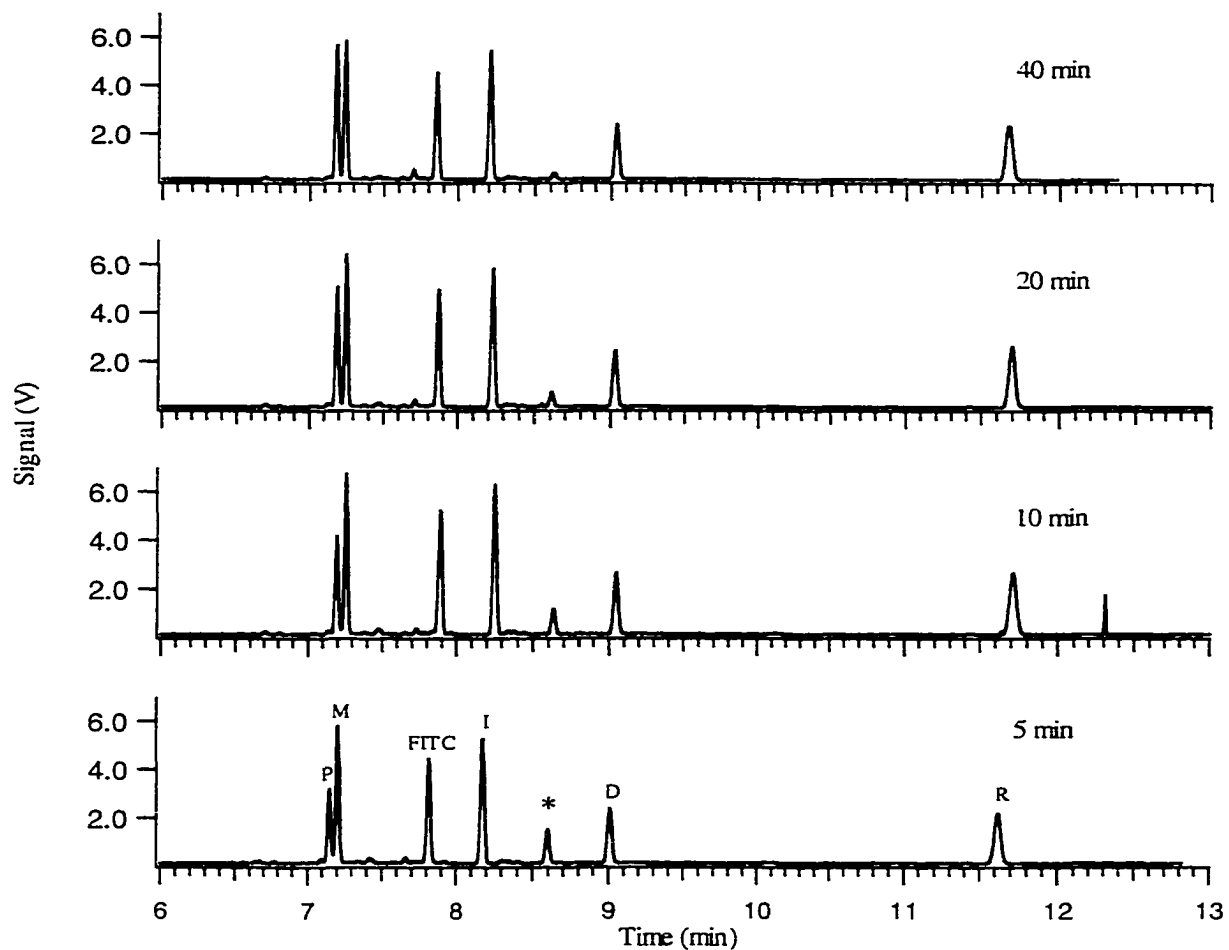


Figure 4.2 Effect of reaction time on the conversion efficiency of five FTH-amino acids.

Again, all of the electropherograms are quite similar. The only FTH-amino acid peak that shows a significant change is FTH-proline. Its signal increased with increasing conversion time. The unknown peak was also still present but it decreased steadily as the reaction time increased. This correlation again suggests that the unknown peak is FTZ-



proline. The effect of conversion time on the peak heights of the FTH-amino acids is summarized in Table 4.5.

Table 4.5 FTH-amino acid peak heights at different conversion times.

Conv. Time (min)	Peak Height (V)				
	FTH-Pro	FTH-Met	FTH-Ile	FTH-Asp	FTH-Arg
5	3.0	5.6	5.1	2.2	2.0
10	4.0	6.6	6.1	2.5	2.5
20	4.9	6.2	5.6	2.3	2.5
40	5.5	5.7	5.3	2.2	2.2

The height of the FTH-proline peak increased by 2.5 V from the 5 min reaction to the 40 min reaction. This trend is not surprising because proline is the least reactive amino acid with the standard Edman chemistry. The other four FTH-amino acids all followed a similar trend. Peak height increased when the conversion time was increased from 5 min to 10 min. It then started to decrease as the conversion time was further increased.

The above results indicate that switching from FITC to PITC in the sequencing chemistry does not adversely affect the conversion reaction. The results further indicate that conversion should be carried out in 25% TFA for 10 min. With the exception of proline, these conditions allowed for complete reaction of all of the amino acids studied.

It is expected that the other coded amino acids will behave in much the same manner as the test group.

#### 4.3.2 Extraction

Extraction efficiency was evaluated using the same representative set of amino acids used in the conversion study and neat TFA as the extraction solvent. Because the syringe pump based sequencer is designed to be used with covalently bound proteins, highly polar solvents can be used without fear of sample washout. The use of polar solvents is necessary because FITC is more polar than PITC. Less polar solvents would not be as effective in extracting the FTZ-amino acids. TFA was chosen for extraction because it is also used for the cleavage reaction, which immediately precedes extraction, and in the conversion reaction which follows it. The results of the extraction study are shown in Figure 4.3. Three successive extractions were performed on a single piece of DITC-GFM to which the FTZ-amino acids had been added. The first extraction sample was analysed at a 50 times greater dilution than the other two samples to prevent overloading of the LIF detection system.

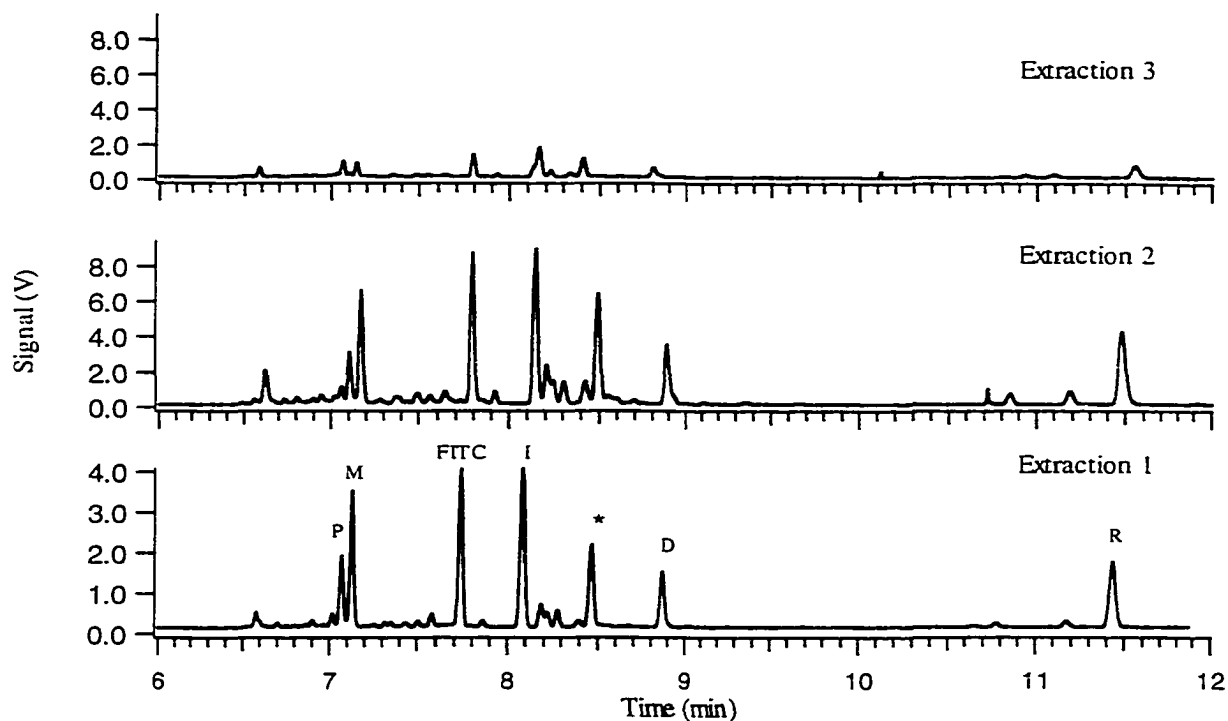


Figure 4.3 Extraction of selected FTH-amino acids by anhydrous TFA. Each extraction consisted of consecutive deliveries of 4  $\mu\text{L}$  of TFA to the reaction chamber. Extraction 1 was analysed at a 50-fold greater dilution than Extraction 2 and Extraction 3.

The results clearly indicate that neat TFA is an excellent solvent for extraction of the FTZ-amino acids. The individual extractions consisted of only 4  $\mu\text{L}$  of TFA. Even with such a small volume of solvent, the first extraction removed greater than 95% of the total FTZ-amino acids. And after the second extraction greater than 99% of the FTZ-amino acids were removed. It is apparent that TFA will work quite well for extraction of the cleavage products generated from the use of FITC in the Edman chemistry.

### 4.3.3 Cleavage

The cleavage reaction, which liberates the derivatized N-terminal amino acid from the protein sample, is the same regardless of which isothiocyanate is used for coupling. However, the rate of cleavage may be affected by the choice of label. As such, a reaction time suitable for the complete cleavage of the phenylthiocarbamyl (PTC) amino acids produced through PITC coupling may not result in complete cleavage of the fluorescein thiocarbamyl (FTC) amino acids produced through FITC coupling. And any FTC-protein molecules that do not undergo cleavage in a given cycle are still available for cleavage in successive cycles. This delayed cleavage, or lag, causes spurious peaks to appear in the electropherograms of later sequencing cycles and makes data interpretation more difficult. Lag also decreases repetitive yield, which limits the number of cycles that can be performed on a given sample. It is therefore necessary to determine how long it takes for the cleavage reaction to go to completion.

The cleavage reaction was evaluated using FITC labelled canine albumin covalently attached to a DITC-GFM. Prior labelling of the protein did not interfere with coupling to the DITC-GFM because on average, ten lysine residues per molecule were labelled while canine albumin actually has thirty lysine residues per molecule (33, 34). Attachment to the mat could still occur through one or more of the unlabelled lysine residues.

The success of the cleavage reaction was evaluated by analysing the converted extract for the presence of FTH-glutamic acid. Although FITC coupling to the canine albumin was not complete, enough of the N-terminal glutamic acid sites should have

been labelled to generate a discernable signal after cleavage and conversion. The results of the cleavage study are shown in Figure 4.4 and Figure 4.5.

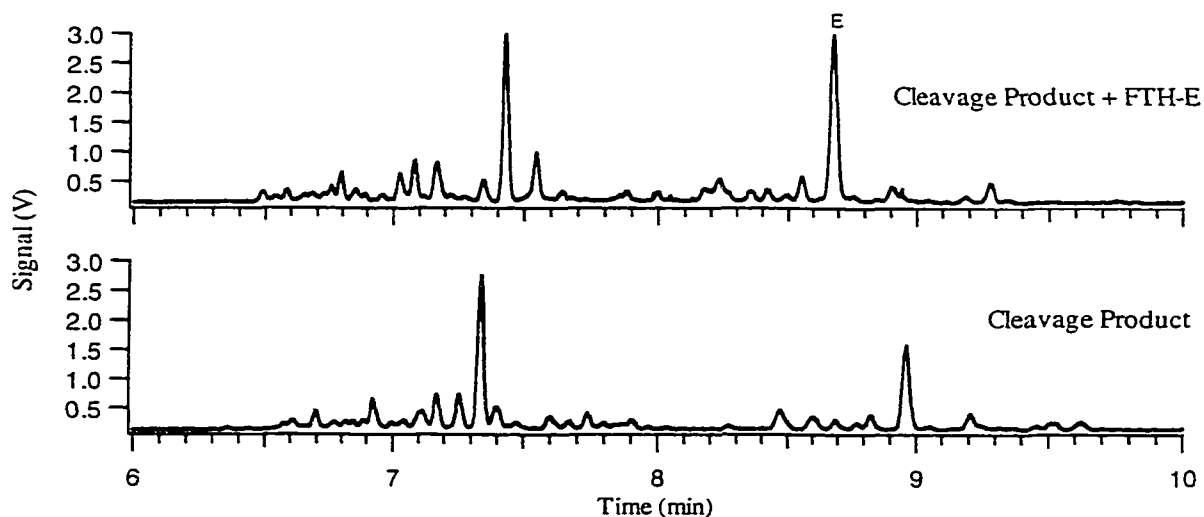


Figure 4.4 Cleavage of the amino-terminal residue from FITC-canine albumin using anhydrous TFA. Cleavage consisted of a 3  $\mu\text{L}$  delivery of neat TFA followed by a 5 min hold step. The cleaved FTZ-amino acid was extracted with three 4  $\mu\text{L}$  TFA deliveries.

Figure 4.4 shows the electropherogram from cleavage of the N-terminal glutamic acid residue, and the electropherogram from the N-terminal residue spiked with an FTH-glutamic acid standard. There are two main peaks in each separation. The identity of the first peak is unknown but the second peak is the expected FTH-glutamic acid. This conclusion can be drawn from the fact that the amplitude of the second peak grew in the spiked sample and no new peak was observed. There is an unexplained shift in migration

time for the spiked sample, but it is related to the CE system and was not observed for any other samples analysed in this work.

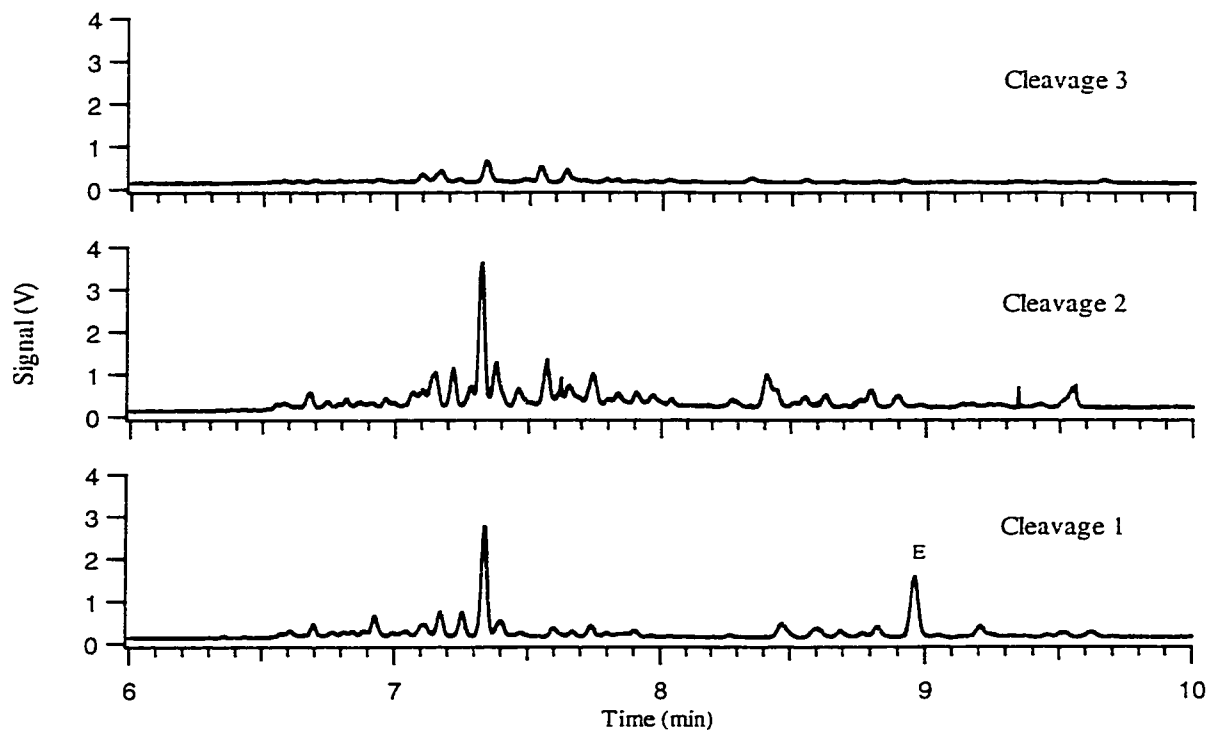


Figure 4.5 Successive TFA cleavage steps on FITC-canine albumin. Each step consisted of a 3  $\mu\text{L}$  delivery of neat TFA followed by a 5 min hold step. The cleaved FTZ-amino acid was extracted with three 4  $\mu\text{L}$  TFA deliveries.

The successive cleavage results in Figure 4.5 indicate that the reaction essentially went to completion with the first cleavage. There was, at best, only a small FTH-glutamic acid peak in the second cleavage and none in the third. Cleavage with liquid

TFA is therefore, an efficient way to generate identifiable sequencing products from FITC labelled protein samples.

#### 4.3.4 Wash

Because of the extreme sensitivity of the CE-LIF detection system, the removal of excess FITC and other byproducts from the reaction chamber after coupling is critical. In order to ensure complete coupling, a large excess of FITC is used. Therefore, the signal resulting from FITC and other fluorescent byproducts would mask the signal generated by the desired FTH-amino acids and severely limit the sensitivity of the syringe pump based sequencer. The more polar nature of FITC makes it necessary to use a strong solvent in order to remove it effectively. The solvent chosen for this work was 0.1% cyclohexylamine in methanol because it is currently used in other solid phase sequencers that use the standard Edman chemistry, and therefore can be obtained with a very low level of contaminants.

Efficiency of the cyclohexylamine/methanol wash solvent was evaluated by performing a coupling reaction on a DITC-GFM followed by successive deliveries of the wash solvent. After an initial 32  $\mu$ L wash step, successive 12  $\mu$ L washes were performed and collected. They were then analysed to determine if byproduct levels were decreasing. The results of the first four 12  $\mu$ L washes are shown in Figure 4.6.

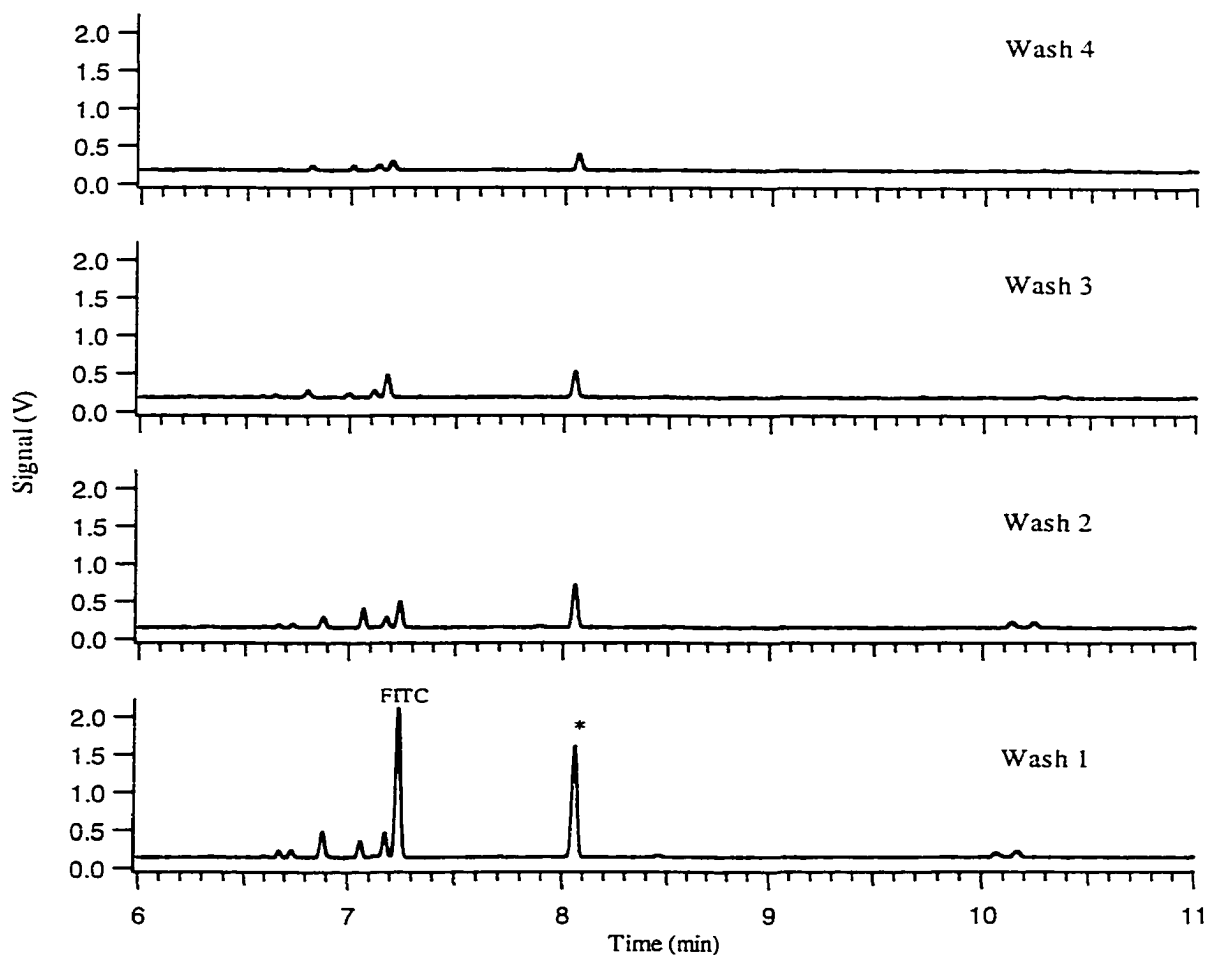


Figure 4.6 Evaluation of the wash efficiency of 0.1% cyclohexylamine in methanol. The wash solvent from each 12  $\mu$ L delivery was collected, diluted to 500  $\mu$ L in pH 7.0, 10 mM phosphate/10 mM SDS running buffer and analysed.

The results in Figure 4.6 show a significant decrease in byproduct peak signal in successive washes. In fact, the two main peaks in Wash 1 are almost completely eliminated by Wash 4. The first of these peaks migrates at the same time as FITC and has been labelled as such. The identity of the other main peak (\*) has not been determined but is presumably the fluorescent analog of the diphenylthiourea generated



when PITC is used for coupling. The decrease in these two peaks indicates that 0.1% cyclohexylamine in methanol is an effective wash solvent for FITC based protein sequencing.

#### 4.3.5 Coupling

Successful modified Edman sequencing requires efficient coupling of FITC with the protein sample. If the coupling reaction does not work, then the N-terminal amino acid will not be cleaved and no product will be generated. In this work, the coupling step was evaluated by performing a single sequencing cycle on a strip of FITC-GFM to which peptide sequencing standard had been covalently attached. The sequencing protocol was a modified version of the one previously used with the standard Edman chemistry and the syringe pump based sequencer (31). The only difference was the use of FITC for coupling instead of PITC. Coupling success was evaluated by analysing the converted extract for the presence of FTH-tyrosine, the expected first residue from the peptide standard. Figure 4.7 gives the results from the first cycle of the peptide standard.

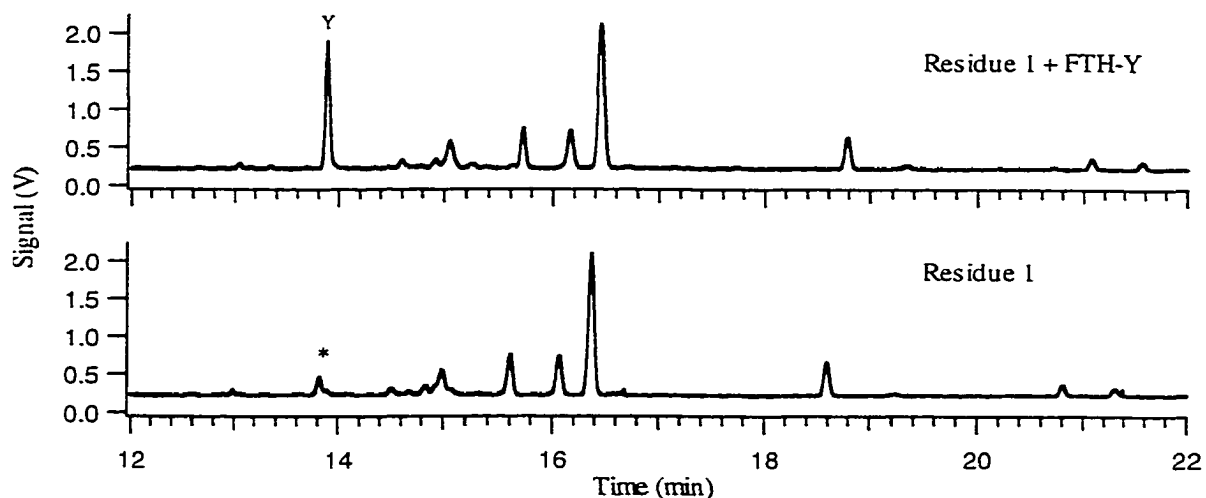


Figure 4.7 Cycle 1 from a peptide sequencing standard using FITC for coupling. The sample was prepared in 500  $\mu\text{L}$  of 10 mM phosphate/3.8 mM borate/7.7 mM SDS/2.0 mM magnesium acetate running buffer at pH 7.03.

The results of the first cycle indicate that a small amount of FTH-tyrosine was produced. There is a small peak in the lower electropherogram that increased considerably when the sample was spiked with an FTH-tyrosine standard. Although the peak is quite small in the unspiked sample, it still indicates that the coupling reaction worked. Furthermore, the sample was diluted to 500  $\mu\text{L}$  for analysis. Less than 1 nL of this sample was then injected onto the separation capillary. It is therefore not surprising that the peak is so small. It is also important to consider that the sequencing protocol has not yet been optimized for use with FITC. A low signal level is therefore not unexpected. The results obtained here do indicate that coupling with FITC is possible and it gives a useful starting point for optimizing the entire sequencing protocol; this work is described in Chapter 5 of this thesis.

#### 4.3.6 Protein Attachment to Solid Supports

In standard automated sequencers the protein sample is typically adsorbed to a solid support. It is held in place due to its stronger affinity for the support than for the wash and extraction reagents used in the sequencing chemistry. This approach works because PITC and the ATZ-amino acids are soluble in fairly nonpolar solvents that will not cause sample washout. When FITC is used for sequencing, more polar wash and extraction solvents are required to adequately remove excess reagent and collect the FTZ-amino acids. A consequence of using more polar solvents for sequencing is that adsorption of the protein to the solid support is no longer strong enough to prevent washout. Samples must be covalently bound to the support if FITC is to be used as a coupling reagent. There are commercially available solid supports for covalent protein attachment. They consist of PVDF membranes or glass beads chemically modified to have a functional group available to react with either primary amine or carboxylic acid sites on the protein. Unfortunately these supports are not amenable for use with FITC sequencing. The PVDF membranes dissolve in the acetone used to make the FITC coupling solution, while the pores in the glass beads make it impossible to adequately wash away excess FITC after coupling. Figure 4.8 shows a picture of FITC-glass beads after a FITC coupling and wash cycle. The picture was taken using a fluorescence microscope to illustrate the amount of fluorescent material present even after extensive washing. The extreme sensitivity of the CE-LIF detection system would be compromised because the samples would have to be run at a significant dilution so that the FITC and other fluorescent byproduct peaks would not swamp the detector.

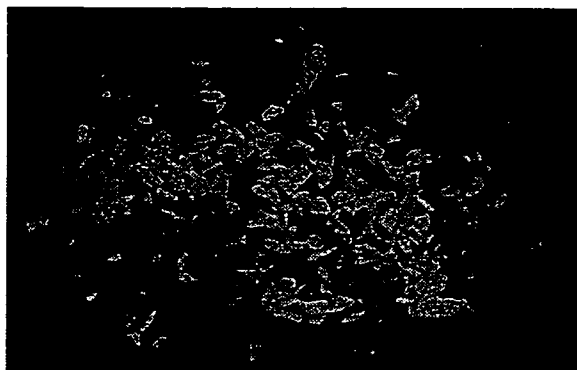


Figure 4.8 Glass beads after coupling and wash.

Glass fiber mats represent an alternative to glass beads as a solid support for covalent protein attachment. The mats are amenable to the same chemistry as the beads and their more open structure should allow for improved washing (32, 35). The mats used in this work were chemically modified to covalently bind proteins through both the N-terminus and any lysyl side chains. The first step in the derivatization process is the attachment of aminopropyltriethoxy silane to the GFM. After silanization, the free aminopropyl group is available to undergo further reaction with phenylenediisothiocyanate. This reaction leaves a free isothiocyanate that can couple with any primary amine sites on the protein sample to be sequenced.

The success of the silanization reaction was monitored by coupling the free amine groups with FITC and viewing the mat under a fluorescence microscope. Figure 4.9 shows a picture of a FITC coupled AP-GFM along with a blank consisting of a bare GFM that had undergone the same coupling and wash protocol.



Figure 4.9 Comparison of an AP-GFM and a bare GFM after FITC coupling.

The AP-GFM shows a diffuse fluorescent signal that is not present on the bare GFM. This diffuse signal results from fluorescent groups that are uniformly spread throughout the mat. The fact that it is only present in the AP-GFM image indicates that the silanization reaction did indeed work. The localized fluorescent spots in both images are due to small crystals of FITC that were not completely washed off.

The degree of DITC coupling to the AP-GFM was monitored using radioactive [ $^{35}\text{S}$ ]-Methionine. It is a small molecule that should react quantitatively with any free isothiocyanate sites on the DITC-GFM. Once the [ $^{35}\text{S}$ ]-Met is attached to the mat, it can be quantitatively analysed with a scintillation counter. The mat is immersed in scintillant that absorbs the energy from the  $\beta$ -particles emitted from the radioactive decay of the [ $^{35}\text{S}$ ]. The scintillant releases the energy through phosphorescence. The phosphorescence signal is measured to determine the amount of [ $^{35}\text{S}$ ]-Met and hence the number of active

sites for protein attachment on the DITC-GFM. Table 4.6 gives the results of scintillation counting measurements on three DITC-GFM's labelled with [<sup>35</sup>S]-Met.

Table 4.6 Scintillation counting measurements from three [<sup>35</sup>S]-Met labelled DITC-GFM's.

Sample	Raw Counts (CPM)	Corrected Counts (CPM)	[ <sup>35</sup> S]-Met Load (nmol)
Underivatised Mat	4969	---	0.0
Mat 1	1 489 093	1 484 084	2.48
Mat 2	1 206 940	1 201 971	2.01
Mat 3	1 160 660	1 155 691	1.93
		AVG	2.1
		S.D.	0.3

The average [<sup>35</sup>S]-Met load is 2.1 nmol. Only one strip of DITC-GFM can be loaded in the reaction chamber for sequencing. The strip corresponds to approximately 10% of the total mat. Therefore, approximately 200 pmol of active sites are available to bind protein samples for sequencing. Although this load capacity is much less than that of commercial protein sequencers, it should still be sufficient for the syringe pump based sequencer. This instrument, in combination with the CE-LIF detection system, is being developed to improve the sensitivity of protein sequencing. As such, a low sample capacity will not impede instrument function.

#### 4.4 Conclusions

Automated protein sequence analysis involves a number of discrete steps that must all be carried out successfully. If any one of the individual steps fails, then the entire process will fail. In the fifty years since Edman first developed his method of protein sequencing, all of the sequencing steps have been optimized for use with PITC as the coupling reagent. However, the sensitivity of PITC based sequencing appears to have reached its lower limit. Further improvements in sequencer sensitivity will require a switch to a more sensitive coupling reagent. The work described here explored the use of FITC as a replacement for PITC in the sequencing chemistry. Such a significant modification could have led to problems in any of the seven sequencing steps. As such, the effect of FITC was evaluated for each step, starting with the final step and working backwards through each succeeding step. It was found that every step could still be performed successfully when FITC was used for coupling.

The work described in this chapter paves the way for further studies aimed at bringing together all of steps that were evaluated here. These studies are the subject of the next chapter of this thesis. Chapter 5 describes how the individual steps were combined and optimized to allow for successful protein sequencing at the low picomole level using the syringe pump based sequencer.

## 4.5 References

1. R. Aebersold, *Nature* **343**, 291 (1990).
2. S. Kent, et al., *BioTechniques* **5**, 314 (1987).
3. L. M. Smith, *Anal. Chem.* **60**, 381A (1988).
4. G. E. Tarr, J. F. Beecher, D. J. McKean, *Anal. Biochem.* **84**, 622 (1978).
5. D. J. C. Pappin, J. Coull, H. Koester, *Anal. Biochem.* **187**, 10 (1990).
6. H. Maeda, N. Ishida, H. Kawauchi, K. Tsujimura, *J. Biochem.* **65**, 777 (1969).
7. J. Y. Chang, E. H. Creaser, K. W. Bently, *Biochem. J.* **153**, 607 (1976).
8. R. Aebersold, et al., *Protein Science* **1**, 494 (1992).
9. B. Wittmann-Liebold, H. Graffunder, H. Kohls, *Anal. Biochem.* **75**, 621 (1976).
10. R. M. Hewick, M. W. Hunkapiller, L. E. Hood, W. J. Dreyer, *J. Biol. Chem.* **256**, 7990 (1981).
11. R. A. Laursen, *Eur. J. Biochem.* **20**, 89 (1971).
12. J. Calaycay, M. Rusnak, J. E. Shively, *Anal. Biochem.* **192**, 23 (1991).
13. J. E. Walker, I. M. Fearnley, R. A. Blows, *Biochem. J.* **237**, 73 (1986).
14. C. L. Zimmerman, E. Apella, J. J. Pisano, *Anal. Biochem.* **77**, 569 (1977).
15. D. Hawke, P. M. Yuan, J. E. Shively, *Anal. Biochem.* **120**, 302 (1982).
16. E. J. Bures, et al., *Anal. Biochem.* **224**, 364 (1995).
17. V. Farnsworth, K. Steinberg, *Anal. Biochem.* **215**, 190 (1993).
18. H. Matsunaga, et al., *Anal. Biochem.* **67**, 4276 (1995).
19. A. Tsugita, K. Masaharu, C. S. Jone, N. Shikama, *J. Biochem.* **106**, 60 (1989).
20. O. Imakyure, M. Kai, T. Mitsui, H. Nohta, Y. Ohkura, *Anal. Sci.* **9**, 647 (1993).



21. O. Imakyure, M. Kai, Y. Ohkura, *Anal. Chim. Acta.* **291**, 197 (1994).
22. H. Hirano, B. Wittmann-Liebold, *Biol. Chem.* **367**, 1259 (1986).
23. S. W. Jin, G. X. Chen, Z. Palacz, B. Wittmann-Liebold, *FEBS Lett.* **198**, 150 (1986).
24. H. Miyano, T. Nakajima, K. Imai, *Biomed. Chromatogr.* **2**, 139 (1987).
25. Y. Cheng, N. J. Dovichi, *Science* **242**, 562 (1988).
26. K. Muramoto, K. Nokihara, A. Ueda, H. Kamiya, *Biosci. Biotech. Biochem.* **58**, 300 (1994).
27. K. Muramoto, H. Kawauchi, K. Tuzimura, *Agric. Biol. Chem.* **42**, 1559 (1978).
28. K. Muramoto, H. Kamiya, H. Kawauchi, *Anal. Biochem.* **141**, 446 (1984).
29. H. Kawauchi, K. Muramoto, J. Ramachandran, *Int. J. Peptide Protein Res.* **12**, 318 (1978).
30. H. Maeda, H. Kawauchi, *Biochem. Biophys. Res. Commun.* **31**, 188 (1968).
31. X. Li, et al., *Talanta* **44**, 401 (1997).
32. R. H. Aebersold, D. B. Teplow, L. E. Hood, S. B. Kent, *J. Biol. Chem.* **261**, 4229 (1986).
33. J. W. Dixon, B. Sarkar, *J. Biol. Chem.* **249**, 5872 (1974).
34. S. Spitzauer, et al., *J. Allergy Clin. Immunol.* **93**, 614 (1994).
35. W. Machleidt, E. Wachter, *Methods Enzymol.* **47**, 263 (1977).

## Chapter 5

### **Fluorescein Isothiocyanate as a Coupling Reagent for High Sensitivity Protein Sequencing in a Syringe Pump Based Instrument with Product Identification by Capillary Electrophoresis with Laser Induced Fluorescence Detection**

#### **5.1 Introduction**

The most common approach used today to determine the primary structure of proteins and peptides is based on a series of chemical steps developed by Pehr Edman almost fifty years ago (1). In the Edman method, phenylisothiocyanate (PITC) couples to the N-terminal amino acid of the protein being sequenced. Subsequent steps result in cleavage and extraction of the labelled residue followed by its conversion to a phenylthiohydantoin (PTH) amino acid. The PTH-amino acid is analysed to identify the original N-terminal amino acid. Successive cycles are used to establish the primary structure of the protein.

Although commercial instruments can generate sequence information from as little as a few picomoles of sample, there are a large number of biologically active proteins and peptides that cannot easily be isolated even at the picomole level (2, 3). A protein sequencer that could generate structural information on femtomole amounts of sample would open up new areas of research in biochemistry (3). Currently, sequencer sensitivity is limited by the HPLC with UV absorbance detection systems used to identify the PTH-amino acid sequencing products (4). A number of alternative methods for

product identification have been explored in an effort to overcome the limits of sensitivity associated with UV absorbance detection (5-14). This chapter describes a method for high sensitivity protein sequencing using capillary electrophoresis with laser induced fluorescence detection (CE-LIF) for product identification.

CE-LIF is one of the most sensitive methods currently available for the analysis of biomolecules. Mass limits of detection for CE-LIF are routinely in the low attomole ( $1 \times 10^{-18}$  moles) to zeptomole ( $1 \times 10^{-21}$  moles) range with yoctomole ( $1 \times 10^{-24}$  moles) LOD's reported in the literature (10, 15-19). This extreme mass sensitivity is due, in part, to the small detection volume that is probed in CE: typically less than 1 nL. As a result, the absolute amount of an analyte that is detected is small even at relatively high concentrations. For example, an analyte with a concentration LOD of  $1 \times 10^{-6}$  M would have a 1 fmol mass LOD in a 1 nL detection volume. Mass sensitivity is further enhanced by the extremely low concentration LOD's that can be obtained using laser induced fluorescence. Fluorescence signal is directly proportional to the power of the incident excitation source. The collimated beam from a laser directs a large amount of energy to the detection region; thereby increasing the signal generated as the analyte molecules pass through. Increasing the signal generated by an analyte leads to an improved concentration LOD. A good fluorophore can have a concentration LOD of less than  $1 \times 10^{-12}$  M. The combined effects of low detection volume and enhanced signal make CE-LIF an ideal candidate to improve the detection sensitivity associated with protein sequencing.

Standard sequencing instruments have been designed for use with HPLC detection systems. As such, the reagent and sample volumes have been optimised with

the volume requirements of HPLC in mind. Because the volume requirements of CE are much smaller than those of HPLC, it is not possible to simply attach a CE system to a standard sequencer. This group has developed a miniaturised sequencer that is more compatible with the volume requirements of CE (7). It allows for a significant reduction in both the reagent volumes used for sequencing and the sample volume used for analysis. The instrument was evaluated using the standard Edman chemistry and capillary electrophoresis with thermo-optical absorbance detection (CE-TOAD) to identify the PTH-amino acids. Much like commercial instruments, the sensitivity of the miniaturised sequencer was limited to the low picomole level by the detection system used for product identification. By switching to CE-LIF for product analysis, it should be possible to overcome this limitation.

Before any sensitivity enhancement can be realized using CE-LIF for protein sequencing, it is first necessary to generate products that are strongly fluorescent. This requirement necessitates a modification to the standard Edman chemistry. A number of different labelling reagents have been suggested as possible candidates to generate fluorescent sequencing products (12, 20-28). The most promising of these reagents is fluorescein isothiocyanate (FITC) (9-11, 29-34). It reacts with amino acids in the same manner as PITC but it generates fluorescein thiohydantoin (FTH) amino acids that are quite amenable to high sensitivity analysis by CE-LIF (9, 10). Using FITC for sequencing does have a problem however: it contains an extensive conjugated  $\pi$  system that reduces the reactivity of the isothiocyanate moiety. A decrease in reactivity leads to a decrease in repetitive yield, which, in turn, severely limits the length of sequence that can be determined for a given sample. This problem can be overcome by the continued

use of PITC during coupling, but as a scavenger instead of a primary reagent. It reacts with sites missed by the less reactive fluorescent label thereby minimizing lag. The rest of the sequencing chemistry remains the same so both the fluorescently labelled amino acids and the PITC labelled amino acids are cleaved. Combined cleavage minimizes protein loss during sequencing, which also serves to extend sequence reads. This double coupling approach has been demonstrated with both manual sequencing and automated instruments (11, 29-31). Again however, the detection methods used were not sensitive enough to improve the overall sequencer sensitivity. The advent of CE-LIF shows promise of overcoming this limitation.

## 5.2 Experimental

### 5.2.1 Protein Sequencing

The protein sequencer was built in house and has been described in Chapter 3. Coupling buffer (NMM, 5% N-methylmorpholine in 70/30 methanol:water with 0.1% cyclohexylamine) and wash solvent (0.1% cyclohexylamine in methanol) were purchased from Millipore Canada (Mississauga, ON). FITC was purchased from Molecular Probes (Eugene, Or). It was dissolved in HPLC grade acetone (Sigma Chemical, St. Louis, MO) to give a working solution with a concentration of  $1 \times 10^{-3}$  M. This solution was stored at 4°C prior to use. 5% phenylisothiocyanate in heptane and high purity  $\beta$ -lactoglobulin sequencing standard were purchased from Applied Biosystems (Foster City, CA).

Anhydrous trifluoroacetic acid (TFA) was purchased from Sigma Chemical (St. Louis, MO). Teflon coated fused silica capillary was purchased from Polymicro Technologies (Phoenix, AZ).

Table 5.1 shows a typical sequencing protocol including the reagent delivered in each step, the length of time for each step, and comments. Coupling buffer (NMM), FITC, PITC and neat TFA were delivered at 8  $\mu\text{L}/\text{min}$ . Wash 1 and Wash 2 (0.1% cyclohexylamine in methanol) were both delivered at 80  $\mu\text{L}/\text{min}$ . Wash 2 and neat TFA shared the same delivery capillary. The acid was only connected for the cleavage and extraction steps. Coupling buffer, FITC and PITC solutions were withdrawn from the delivery capillaries after coupling to prevent leakage into the reaction chamber during cleavage and extraction. The reaction chamber was held at 50°C throughout the sequencing protocol.  $\beta$ -lactoglobulin was covalently coupled to diisothiocyanate-glass fiber mats (DITC-GFM's) prior to sequencing according to the method described in Chapter 3. Protein loads of 25 pmol, 10 pmol and 1 pmol were obtained by adding 1  $\mu\text{L}$  of coupling solution containing the desired amount of protein to a fresh strip of DITC-GFM cut so it would fit into the reaction chamber. Each strip was inserted into the reaction chamber immediately prior to sequencing and removed before the next sample was analysed. Sequencing conditions were the same for all three samples with one exception: total coupling time was increased from 30 minutes/cycle for the 25 pmol sample to 45 minutes/cycle for both the 10 pmol and 1 pmol samples.

Table 5.1 Typical sequencing protocol.

Step	Reagent Delivered/Function	Time (s)	Comments
1	NMM	8	
2	Ar	1	
3	NMM	8	
4	FITC	15	
5	Hold	900	Couple
6	Ar	10	
7	Repeat Steps 3-6	933	Couple
8	Repeat Steps 3-6	933	Couple
9	Withdraw FITC	3	Reduce contamination
10	NMM	8	
11	PITC	15	
12	Hold	300	Scavenge
13	Ar	10	
14	Repeat Steps 10-13	333	Scavenge
15	Ar	30	
16	Withdraw NMM and PITC	3	Reduce contamination
17	Wash 1	187	Deliver full 250 $\mu$ L syringe volume
18	Wash 2	187	Deliver full 250 $\mu$ L syringe volume
19	Ar	60	Dry the reaction chamber Switch Wash 2 with TFA
20	TFA	12	Fill delivery capillary
21	TFA	23	
22	Hold	300	Cleave
23	Ar	1	Collect the TFA
24	TFA	30	Extract
25	Ar	1	Collect the TFA
26	TFA	30	Extract
27	Ar	30	Collect the TFA Switch TFA with Wash 2
28	Wash 2	187	Deliver full 250 $\mu$ L syringe volume
29	Wash 1	187	Deliver full 250 $\mu$ L syringe volume
30	Ar	300	Dry the reaction chamber

Conversion of the fluorescein thiazolinone (FTZ) amino acid intermediates was performed off line. They were extracted with neat TFA into a 200  $\mu$ L vial containing 50  $\mu$ L of distilled, deionized water. The solution was mixed, heated at 65°C for 10 minutes

and dried on a vacuum centrifuge. The converted FTH-amino acids were stored at 4°C until analysed.

### 5.2.2 Sample Analysis

Capillary electrophoresis was carried out as described in Chapter 2 of this thesis with some minor modifications. The separation capillary had an inner diameter of 30  $\mu\text{m}$ ; the sample was electrokinetically injected at 50 V/cm for 5 seconds; and the aqueous running buffer for all samples consisted of 15.0 mM sodium dihydrogen phosphate, 3.75 mM sodium tetraborate, 7.7 mM sodium dodecyl sulphate and 2.0 mM magnesium acetate at a pH of 6.86. Sequencing samples were dissolved in either 50 or 500  $\mu\text{L}$  of running buffer containing 1% acetonitrile prior to analysis. FTH-amino acid standards were prepared as described in Chapter 2. They were analysed at a nominal concentration of  $2 \times 10^{-9}$  M for each derivatised amino acid.

## 5.3 Results and Discussion

The performance of the syringe pump sequencer was evaluated by analysing successively smaller amounts of  $\beta$ -lactoglobulin covalently coupled to a DITC-GFM. Instrument performance was monitored with respect to both the efficiency of the sequencing chemistry and overall sensitivity using protein loads of 25 pmol, 10 pmol and 1 pmol respectively.



Sequencing results for the first five cycles from a 25 pmol load of  $\beta$ -lactoglobulin are shown in Figure 5.1. Samples were injected from 50  $\mu$ L of running buffer to prevent peaks from overloading the detector. Inspection of the electropherograms in Figure 5.1 shows that identification of the expected FTH-amino acid is straightforward in all five cycles. The expected sequence for the first five residues of  $\beta$ -lactoglobulin is given below:

L - I - V - T - Q

Even the FTH-leucine peak in cycle 1 is easy to identify. This residue, along with any lysyl residues, covalently couples to the DITC-GFM and should therefore give very little signal. The presence of such a strong FTH-leucine peak indicates that coupling of the protein to the DITC-GFM was not very efficient. There is an increase in signal in cycle 2 as protein molecules that were coupled via the N-terminal leucine become available for sequencing. Cycle 3 has a very prominent FTH-valine peak but it co-migrates with a byproduct peak. The relative contribution of both components to the overall peak is difficult to determine because of the variability in intensity of the byproducts from cycle to cycle. Cycle 4 and cycle 5 both show a marked decrease in signal compared to earlier cycles. However, identification of the FTH-amino acid peaks is still not difficult.

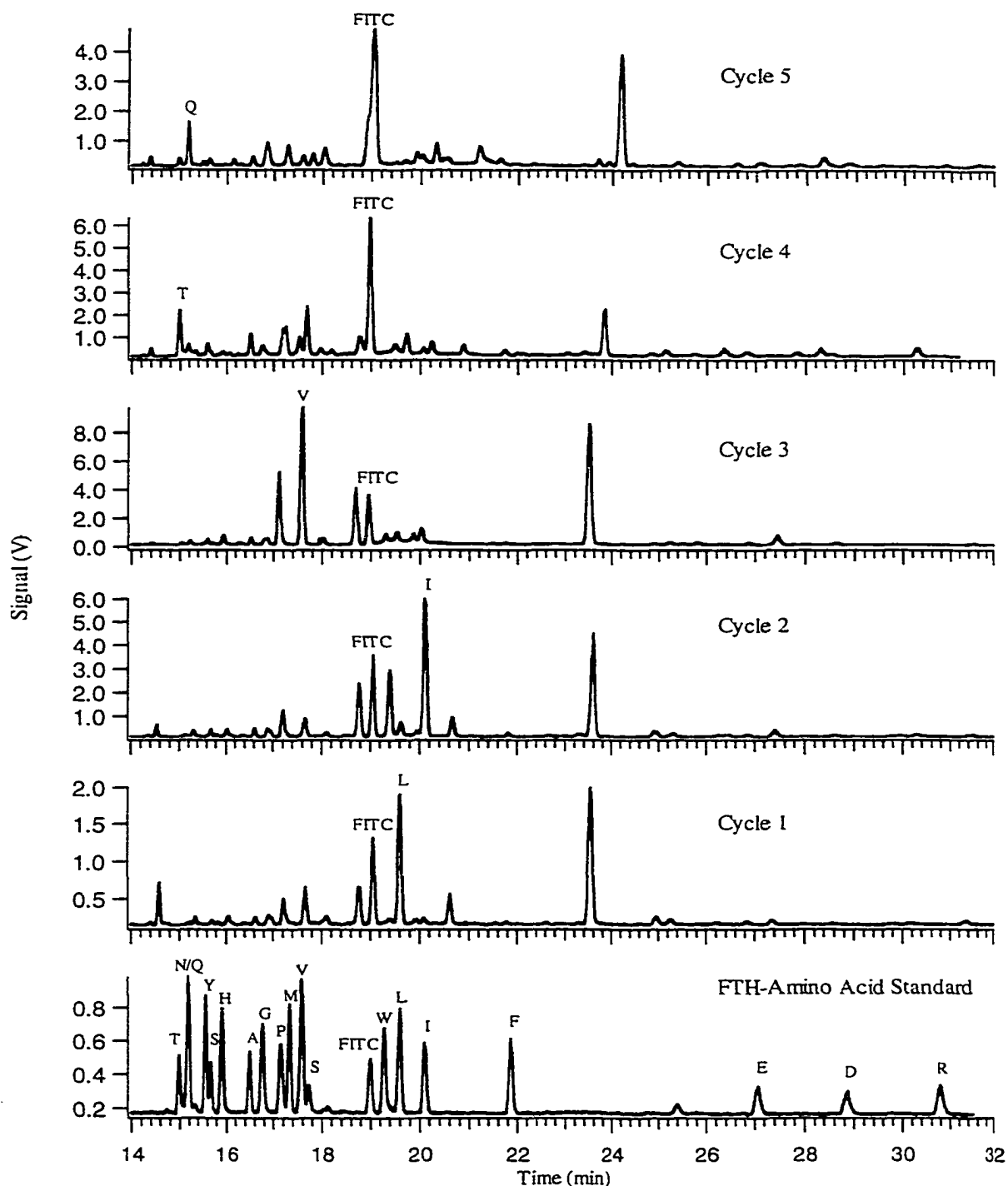


Figure 5.1 Electropherograms of the first five cycles from 25 pmol of  $\beta$ -lactoglobulin. Samples were reconstituted in 50  $\mu$ L of mobile phase containing 1% acetonitrile. The standard was run at a nominal concentration of  $2 \times 10^{-9}$  M for each FTH-amino acid.

Table 5.2 gives the calculated yields for the five cycles from 25 pmol of  $\beta$ -lactoglobulin. The yield for each cycle was calculated using the nominal concentration and peak height of the corresponding FTH-amino acid from the standard separation.

Table 5.2 Yield for five cycles from 25 pmol of  $\beta$ -lactoglobulin.

Cycle	Expected Residue	Yield (pmol)	Yield (%)
1	Leu	0.28	1.1
2	Ile	1.4	5.6
3	Val	1.2	4.8
4	Thr	0.90	3.6
5	Gln	0.39	1.6

The initial yield for a sequencing run measures the percentage of the applied sample that is actually available to be sequenced. It gives an indication of both the efficiency of protein binding and the effectiveness of the sequencing chemistry in generating measurable products. Because the  $\beta$ -lactoglobulin covalently coupled to the DITC-GFM through both the N-terminal amino acid and any lysine residues, the yields for these cycles will be substantially reduced. As such, the yield in the first cycle is not a true indication of the initial yield. The initial yield is more accurately determined from cycle 2 since any protein molecules coupled through the N-terminal amino acid become available for sequencing in the second cycle. Therefore, the initial yield from the

analysis of 25 pmol of  $\beta$ -lactoglobulin was 5.6%. This value is extremely low. Commercial solid supports typically give initial yields of between 30% and 70%.

There are a couple of factors that could explain the low initial yield obtained with the syringe pump based sequencer. First, double coupling chemistry produces a mixture of both FTH-amino acids and PTH-amino acids in each cycle. Only the FTH-amino acids are detected so the calculated initial yield does not include any PTH-isoleucine that was produced in cycle 2. The 5.6% initial yield may simply be indicating that the FITC coupling chemistry is not yet optimised. Second, a substantial amount of product may have been lost during the manual conversion of the FTZ-amino acid intermediates to the FTH-amino acids. This process involves a significant amount of sample handling and exposes the unstable intermediates to the atmosphere. Incorporating on-line conversion into the current sequencer design would reduce this type of sample loss.

Although the yields from all five cycles are quite low, they do serve to illustrate the potential sensitivity of the sequencer. Identification of FTH-leucine in cycle 1 was straightforward even though only 280 fmol of product was generated. Furthermore, the sample was injected from 50  $\mu$ L of running buffer. This volume could easily be decreased by as much as two orders of magnitude without having a serious impact on sample injection. With such a reduction in sample volume, 2.8 fmol of FTH-leucine would generate the same signal as 280 fmol did in Figure 5.1. This illustration provides a striking example of the potential of the sequencer. It also demonstrates that improvements in initial yield are not critical to improvements in overall sequencer sensitivity. If 2.8 fmol of product represented a 5.6% initial yield, then the protein load

would be only 50 fmol. The ability to sequence even this small an amount of protein would be a significant improvement over current instruments.

Of more importance to the potential utility of the syringe pump sequencer is the repetitive yield that can be attained. A high repetitive yield is necessary for extended sequence reads. The overall repetitive yield for a sequencing run can be determined from a log-linear plot of yield vs. cycle number. Figure 5.2 shows this plot for the analysis of 25 pmol of  $\beta$ -lactoglobulin. Again, because of loss of the N-terminal amino acid to coupling with the DITC-GFM, the yield from the first cycle was not included in the plot.

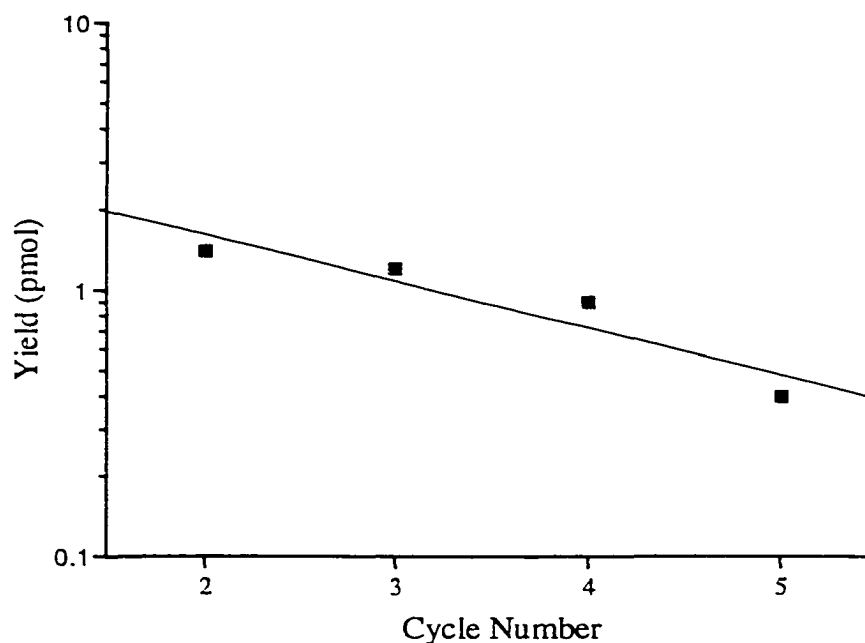


Figure 5.2 Yield vs. Cycle Number for cycles 2-5 from 25 pmol of  $\beta$ -lactoglobulin.

The equation of the best fit line for the sequencing data is

$$\log Y_b = (\log RY)b + \log Y_a \quad (5.1)$$

where RY is the overall repetitive yield and  $Y_a$  is the theoretical yield of cycle 0. Hence the overall repetitive yield is determined from the inverse log of the best fit line for the sequencing data. Commercial instruments typically have repetitive yields of 95% or greater. If the repetitive yield drops much below this, the amount of protein lost in each cycle is too great to allow for an extended sequence read, especially for small protein loads that are close to the instrument's limit of sensitivity. The overall repetitive yield determined from the plot in Figure 5.2 was 67%. That is, one third of the protein available to be sequenced in each cycle is lost to subsequent cycles. So for a protein sample giving an initial yield of 1.4 pmol, the expected yield in the tenth cycle would be only 25 fmol. In contrast, the expected yield in the thirtieth cycle for the same sample with a 95% repetitive yield would be 300 fmol. An amount of product that could still be easily identified by CE-LIF.

It is important to determine the reasons why the repetitive yield was so low and how it can be improved. One possible cause for the low repetitive yield is incomplete coupling between the isothiocyanates and the protein. However, this would lead to a significant amount of lag in successive cycles. Inspection of the electropherograms in Figure 5.1 shows very little lag. Therefore, something else must be reacting with the N-terminal amino acid and blocking it from further sequencing.

Blockage occurs when a contaminant reacts with the protein during coupling. Protein molecules that react with the contaminant can no longer undergo cleavage and are lost to further sequencing. The most problematic contaminant in protein sequencing is

atmospheric oxygen. It reacts readily with the thiocarbamyl (TC) protein produced by both FITC and PITC coupling (35). Commercial sequencers are painstakingly designed to minimise the amount of oxygen contamination for this reason. The syringe pump based sequencer is not as well isolated from atmospheric oxygen as commercial instruments. Each reagent is manually drawn into its delivery syringe. The reagents are exposed to the atmosphere during this process and undoubtedly absorb some oxygen. Oxygen could also be diffusing directly into the reaction chamber during coupling. The chamber is made of Teflon, which is slightly permeable to oxygen. And the outlet from the reaction chamber is open for the entire sequencing cycle. Oxygen can diffuse into the reaction chamber through either of these routes. Furthermore, there are extended hold steps during coupling that would allow significant amounts of oxygen to reach the protein. The addition of an on-line conversion apparatus would help isolate the reaction chamber and eliminate this problem.

The results for the five cycles from 25 pmol of  $\beta$ -lactoglobulin clearly indicate that double coupling sequencing with a syringe pump based instrument is feasible. Identification of the expected residue was unambiguous in every cycle. Both the initial yield and repetitive yield were low but these problems are typical for a first generation instrument. They should improve as modifications are made to the reaction chamber and other components are brought on line.

Notwithstanding the low initial and repetitive yield, the signal from 25 pmol of  $\beta$ -lactoglobulin was still well above the limit of detection for the CE-LIF system. Therefore, the protein load was reduced to 10 pmol to further evaluate the potential of the sequencer. The sequencing protocol was modified in an attempt to improve coupling

efficiency. The FITC coupling time was increased from 30 minutes/cycle to 45 minutes/cycle. The first cycle was performed using PITC as the sole coupling reagent in order to minimize the coupling time. This modification would, in turn, minimize the amount of protein lost to oxygen contamination during a cycle in which most of the N-terminal residues were already covalently coupled to the DITC-GFM.

Sequencing results from cycles 2–5 for the 10 pmol sample are shown in Figure 5.3. Cycle 1 was not shown because no fluorescent products were generated. It is again possible to identify the expected FTH-amino acid in each cycle. In fact, cycle 2 and cycle 3 had to be injected from 500  $\mu\text{L}$  of running buffer to prevent the analyte signal from overloading the detector. The signal dropped significantly in cycle 4 and cycle 5 so those samples were both analysed from 50  $\mu\text{L}$  of running buffer.



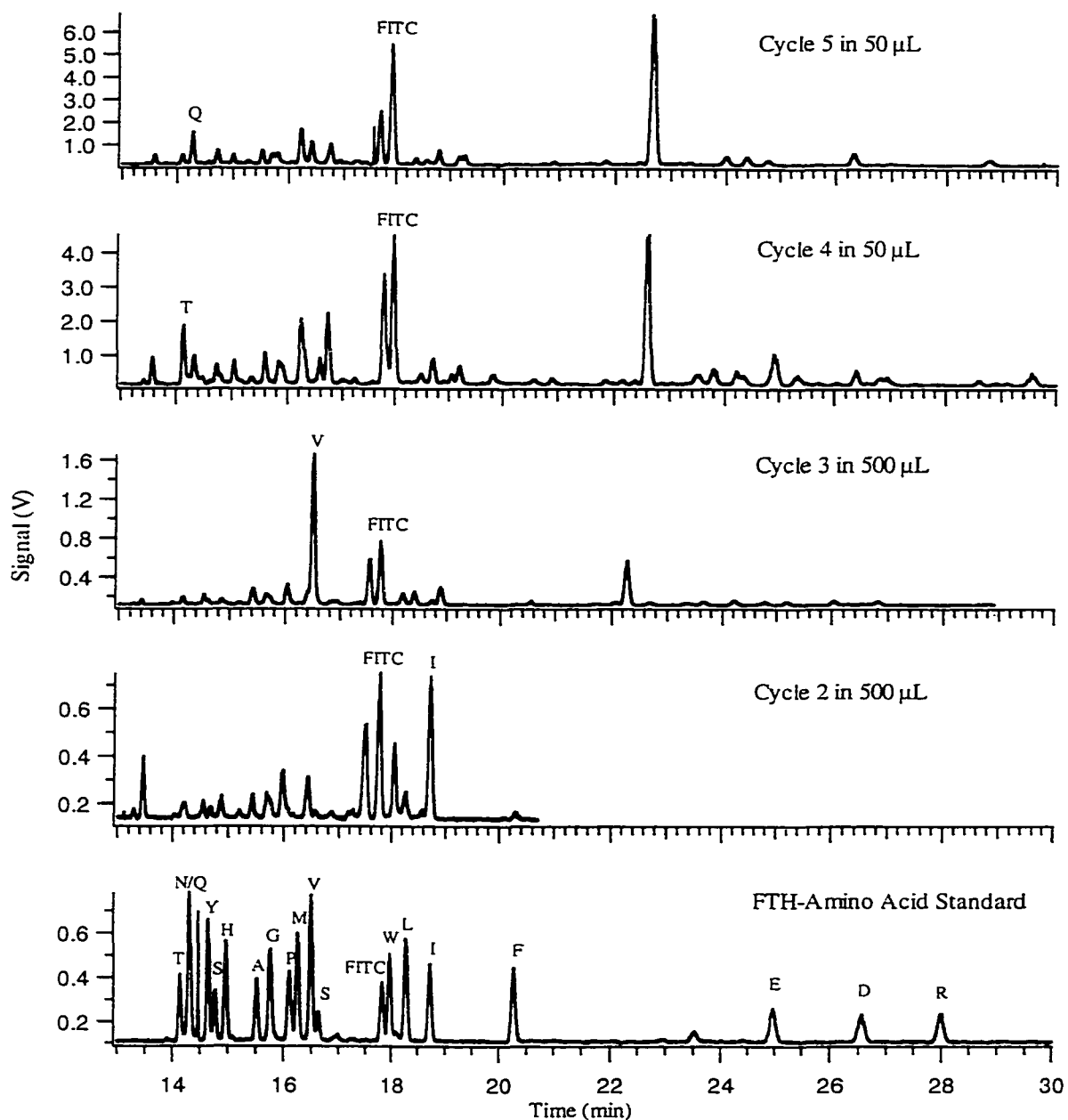


Figure 5.3 Electropherograms of cycles 2-5 from 10 pmol of  $\beta$ -lactoglobulin. Samples from cycles 2 and 3 were reconstituted in 500  $\mu$ L of running buffer containing 1% acetonitrile. Samples from cycles 4 and 5 were reconstituted in 50  $\mu$ L of running buffer containing 1% acetonitrile. The standard was run at a nominal concentration of  $2 \times 10^{-9}$  M for each FTH-amino acid.

The yield was again calculated for each cycle. The results are shown in Table 5.3.

Table 5.3 Yield for four cycles from 10 pmol of  $\beta$ -lactoglobulin.

Cycle	Expected Residue	Yield (pmol)	Yield (%)
2	Ile	1.8	18
3	Val	2.1	21
4	Thr	0.69	6.9
5	Gln	0.42	4.2

The initial yield increased to 18% for the 10 pmol sample from 5.6% for the 25 pmol sample. This 3-fold increase, accompanied as it is by the increased FITC coupling time, suggests that the double coupling chemistry is not yet optimized. Further optimization will require modifying the coupling temperature or the amount and concentration of FITC. Simply extending the coupling time further is not practical because it would result in extremely long cycle times. And the extreme sensitivity of the CE-LIF system means that even an 18% initial yield is good enough for low femtomole sequencing.

The overall repetitive yield was again calculated for the five cycles by plotting yield vs. cycle number. The plot is shown in Figure 5.4.

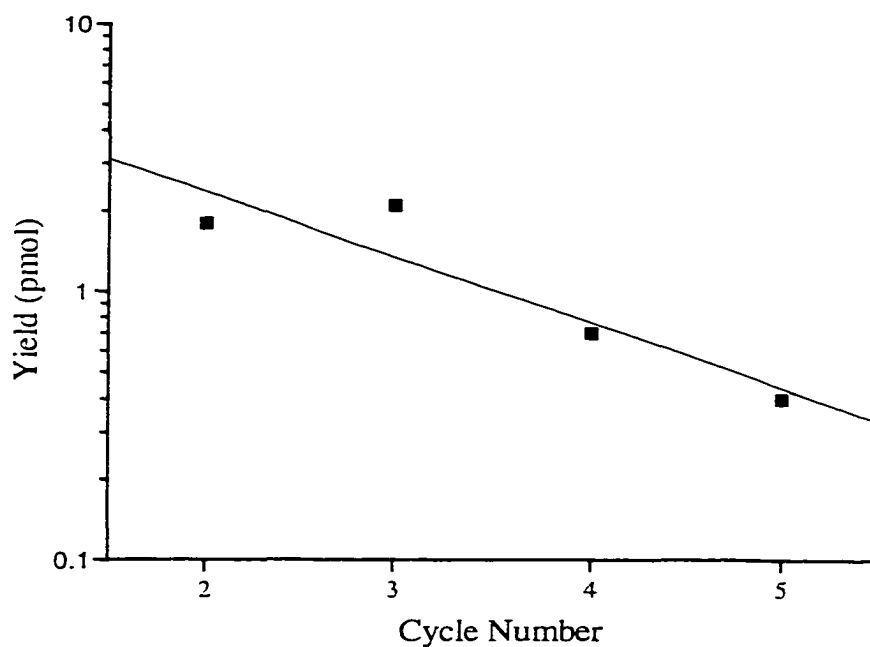


Figure 5.4 Yield vs. Cycle Number for cycles 2-5 from 10 pmol of  $\beta$ -lactoglobulin.

The overall repetitive yield for the 10 pmol sequencing run was only 57%. This value represents a 10% drop compared to the 25 pmol analysis. The only significant difference in the sequencing protocols for the two analyses was the increased FITC coupling time. It indicates that increasing the coupling time has an adverse effect on the overall repetitive yield. This correlation further suggests that oxygen contamination is limiting the repetitive yield. Longer coupling times allow more oxygen to diffuse into the reaction chamber and react with the TC-protein.

A comparison of individual repetitive yields for each cycle from repeat analyses of the same protein can provide information on the cycle to cycle reproducibility of the sequencing conditions. Although repetitive yields can vary considerably for different residues in different samples, they should be consistent for individual residues from

successive analyses of the same protein. Sample to sample differences for the same protein indicate that the sequencing conditions are not very reproducible. Individual repetitive yields are calculated using the equation

$$RY = \left( \frac{Y_b}{Y_a} \right)^{\frac{1}{b-a}} \times 100 \quad \text{for } a < b \quad (5.2)$$

where  $Y_a$  and  $Y_b$  are the yields from cycles a and b respectively and RY is the repetitive yield for cycle b. Table 5.4 gives the individual repetitive yields from cycles 3-5 for both the 10 pmol and 25 pmol  $\beta$ -lactoglobulin samples. Due to the loss of the N-terminal amino acid through coupling with the DITC-GFM, RY calculations were made using the yield from cycle 2 for  $Y_a$ .

Table 5.4 Comparison of individual repetitive yields from 25 pmol and 10 pmol of  $\beta$ -lactoglobulin.

Cycle	Expected Residue	25 pmol Repetitive Yield (%)	10 pmol Repetitive Yield (%)
3	Val	88	117
4	Thr	80	62
5	Gln	65	62

The individual repetitive yield results from 25 pmol of  $\beta$ -lactoglobulin are considerably different from those obtained from 10 pmol of  $\beta$ -lactoglobulin. The repetitive yield in cycle 3 increased from 88% to 117% respectively. This increase does not appear to be due to an increase in the unresolved byproduct peak. In fact, the relative contribution of the byproduct in cycle 3 from the 10 pmol sample appears to be smaller than in the corresponding cycle from the 25 pmol sample. Comparison of cycle 4 from both sequencing runs shows the opposite trend from cycle 3: the repetitive yield dropped from 80% to 62%. And in cycle 5 there was no appreciable difference between the two samples. Variations of this kind in repetitive yield indicate that the sequencing conditions are not reproducible from cycle to cycle.

There are a number of byproduct peaks in the electropherograms from both the 25 pmol and the 10 pmol samples. For the most part each peak is present in every cycle but with differing intensities. This variability is further indication that cycle to cycle sequencing conditions are not consistent. One of the byproduct peaks has the same migration time as FITC and has been labelled as such on all of the electropherograms. The other major peaks are probably fluorescent analogues of diphenylthiourea (DPTU) and diphenylurea (DPU), two common byproducts from the standard Edman chemistry. Contamination of the reaction chamber, both by oxygen and the sequencing reagents themselves, is the most likely explanation for the minor byproducts. Oxygen can react with the byproducts in much the same manner as it can with the coupled protein. Contamination by the sequencing reagents results because the syringe pump sequencer has no valves. There will always be some mixing of the reagents where the delivery

capillaries end. This type of contamination will always be problematic with the current reaction chamber design.

Many of the byproduct peaks do not co-migrate with any of the FTH-amino acid peaks so they do not interfere with sequence determination. However, there are a number of byproducts that migrate between 15 and 18 minutes. These peaks migrate in the same region as the majority of the FTH-amino acids. As such, interference between the byproducts and the FTH-amino acids is a problem, especially for FTH-valine and FTH-proline. The presence of these byproducts and the variable signal they generate could lead to errors in sequence analysis. Signal from a byproduct could be mistaken for signal from an FTH-amino acid or conversely, the signal from an FTH-amino acid could be buried by the signal from a byproduct. Both cases would lead to an incorrect sequence assignment for the cycle in question. This type of chemical noise will be more problematic as the protein load is decreased.

There is a byproduct peak in cycle 2 for both protein samples that is not present in any other cycle. In order to determine whether this peak is due to the protein sample or way related to the solid support, two cycles were run using a blank DITC-GFM. The results of the blank runs are shown in Figure 5.5.

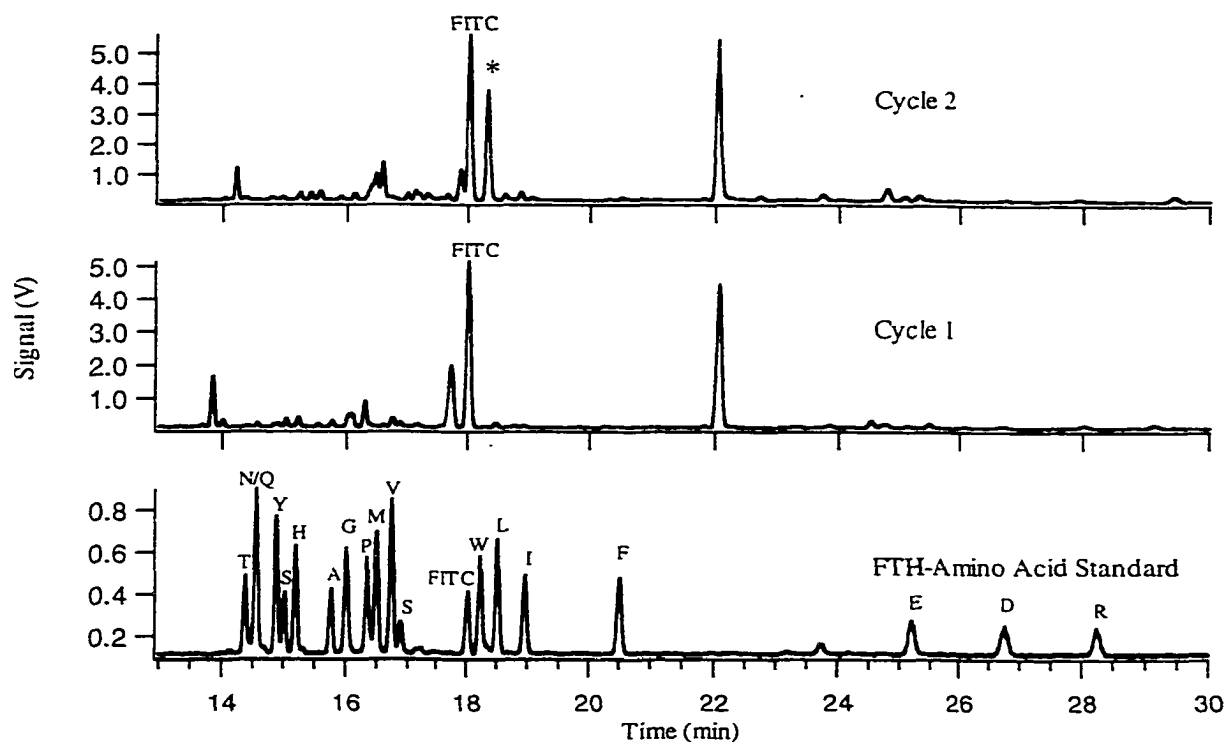


Figure 5.5 Electropherograms of two blank cycles run on an uncoupled DITC-GFM. Samples were reconstituted in 50  $\mu\text{L}$  of mobile phase containing 1% acetonitrile. The standard was run at a nominal concentration of  $2 \times 10^{-9}$  M for each FTH-amino acid.

Cycle 2 again contains a significant peak (\*) that is not present in cycle 1. The absence of protein and the fact that the extra peak only appears in the second cycle indicates that the peak results from a contaminant coupled to the DITC-GFM. If the byproduct was not covalently attached to the mat, then it would have been removed when the GFM's were washed or it would have generated a signal in the first cycle. Instead, it appears to be activated in the first cycle, then coupled with FITC and cleaved in the second cycle in much the same way as protein molecules coupled through the N-terminal

amino acid. The peak however, does not correspond to any of the common amino acids. In all cases it migrates between FTH-tryptophan and FTH-leucine. There is also no discernable secondary sequence in the five cycles run from 25 pmol of  $\beta$ -lactoglobulin. The combination of these two facts suggests that the contaminant is not another protein.

The signal generated for cycle 2 and cycle 3 from 10 pmol of  $\beta$ -lactoglobulin was well above the limit of detection for the CE-LIF. In fact, the sample had to be diluted by an order of magnitude to prevent detector overload. To further evaluate the sensitivity of the syringe pump sequencer, the protein load was decreased to 1 pmol. Figure 5.6 shows the electropherograms of cycles 2-5 from 1 pmol of  $\beta$ -lactoglobulin.



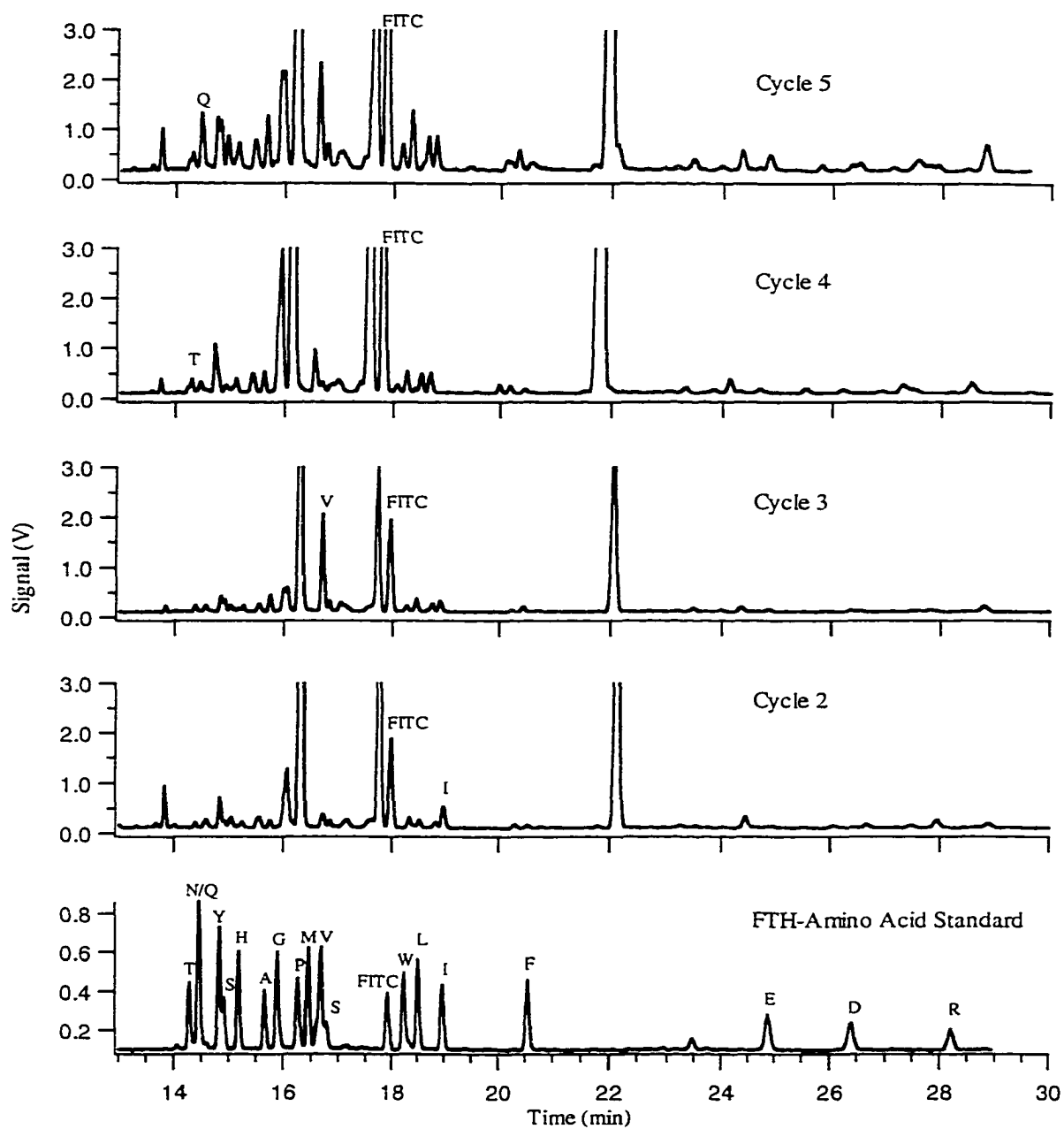


Figure 5.6 Electropherograms of cycles 2-5 from 1 pmol of  $\beta$ -lactoglobulin. Samples were reconstituted in 50  $\mu$ L of running buffer containing 1% acetonitrile. The standard was run at a nominal concentration of  $2 \times 10^{-9}$  M for each FTH-amino acid.

The signal from the FTH-amino acids is greatly reduced compared to the higher protein loads. The expected peaks can be identified in each cycle but they are not the main peaks in the electropherograms. Analysis of an unknown at this protein load would not be possible. Chemical noise would lead to significant errors in the sequence determination.

The sensitivity of the syringe pump sequencer is limited not by the sensitivity of the CE-LIF detection system but rather by chemical noise from byproducts generated during sequencing. The peaks generated by the FTH-amino acid products from 1 pmol of  $\beta$ -lactoglobulin were no larger than a number of the minor byproduct peaks. As a result, identification of the actual sequencing products from 1 pmol of an unknown protein would be extremely difficult. Improving the sensitivity of the sequencer will require a reduction in chemical noise so that the signal from the FTH-amino acids is significantly greater than the signal from the interfering byproducts. A reduction in chemical noise would also lead to a reduction in sample volume used for product analysis. The current 50  $\mu$ L sample volume is at least two orders of magnitude greater than that required for injection in CE. Reducing the sample volume would result in a corresponding increase in the signal generated by the FTH-amino acids. As such, modifications to the sequencer that reduce the chemical noise will result in lower sample volumes and lead to immediate improvements in instrument sensitivity. In this respect, the syringe pump instrument differs from commercial sequencers. They are limited by the lack of sensitivity associated with identification of the PTH-amino acids using HPLC with UV absorbance detection. Significant improvements in sensitivity for this type of detection system are unlikely so it will be difficult to decrease the protein requirements of any sequencer using

this approach. In contrast, a large decrease in chemical noise is expected as revisions and improvements are made to first generation instruments like the syringe pump sequencer. For example, the first automated sequencers based on single coupling and using manual conversion required nanomoles of protein, while current sequencers using the same chemistry can sequence proteins at the low picomole level. A similar progression for the sequencer described here is not unreasonable. Such a progression would result in a sequencer that could successfully analyse proteins at the low femtomole level.

#### **5.4 Conclusions**

High sensitivity sequencing using FITC for coupling and CE-LIF for product identification is a viable alternative to the standard PITC chemistry using HPLC with UV absorbance detection. The data presented here demonstrates that a first generation sequencer designed to be used with double coupling chemistry and CE-LIF detection can generate sequence information on low picomole amounts of protein. The current design suffers from low repetitive yield, which severely limits the number of cycles that can be run from a single sample. However, modifications to the reaction chamber and on-line conversion of the FTH-amino acids will result in improved repetitive yields and will allow for extended sequence reads.

Furthermore, the sensitivity of the sequencer is currently limited by chemical noise, not the CE-LIF detection system. In fact, the sensitivity of the CE-LIF system could be improved by at least two orders of magnitude by simply decreasing the sample volume used for injection. As improvements are made to the design of the sequencer and

conversion of the FTH-amino acids is brought on line, chemical noise will decrease significantly. It will then be possible to reduce the sample volume used to analyse the FTH-amino acids. The resulting decrease in sample volume will generate instant sensitivity improvements for the overall sequencing process. When these improvements are realized, the syringe pump sequencer based on the use of FITC for coupling and CE-LIF for product identification will be able to successfully sequence low femtomole amounts of proteins and peptides.

## 5.5 References

1. P. Edman, *Acta Chem. Scand.* **4**, 283 (1950).
2. S. Kent, et al., *BioTechniques* **5**, 314 (1987).
3. L. M. Smith, *Anal. Chem.* **60**, 381A (1988).
4. P. Tempst, C. Geromanos, H. Elicone, Erdjument-Bromage, *Methods: A Companion to Methods in Enzymology* **6**, 248 (1994).
5. K. C. Waldron, N. J. Dovichi, *Anal. Chem.* **64**, 1396 (1992).
6. K. C. Waldron, et al., *Talanta* **44**, 383 (1997).
7. X. Li, et al., *Talanta* **44**, 401 (1997).
8. M. Chen, K. C. Waldron, Y. Zhao, N. J. Dovichi, *Electrophoresis* **15**, 1290 (1994).
9. Y. Cheng, N. J. Dovichi, *Science* **242**, 562 (1988).
10. S. Wu, N. J. Dovichi, *J. Chromatogr.* **480**, 141 (1989).

11. K. Muramoto, K. Nokihara, A. Ueda, H. Kamiya, *Biosci. Biotech. Biochem.* **58**, 300 (1994).
12. V. Farnsworth, K. Steinberg, *Anal. Biochem.* **215**, 190 (1993).
13. R. Aebersold, et al., *Protein Science* **1**, 494 (1992).
14. E. J. Bures, et al., *Anal. Biochem.* **224**, 364 (1995).
15. D. Chen, N. J. Dovichi, *Anal. Chem.* **68**, 690 (1996).
16. D. B. Craig, J. C. Y. Wong, N. J. Dovichi, *Anal. Chem.* **68**, 697 (1996).
17. J. Y. Zhao, N. J. Dovichi, O. Hindsgaul, S. Gosselin, M. M. Palcic, *Glycobiology* **4**, 239 (1994).
18. D. Chen, N. J. Dovichi, *J. Chromatogr.* **657**, 265 (1994).
19. J. Y. Zhao, P. Diedrich, Y. Zhang, O. Hindsgaul, N. J. Dovichi, *J. Chromatogr.* **657**, 307 (1994).
20. H. Matsunaga, et al., *Anal. Chem.* **67**, 4276 (1995).
21. V. Farnsworth, K. Steinberg, *Anal. Biochem.* **215**, 200 (1993).
22. A. Tsugita, K. Masaharu, C. S. Jone, N. Shikama, *J. Biochem.* **106**, 60 (1989).
23. O. Imakyure, M. Kai, T. Mitsui, H. Nohta, Y. Ohkura, *Anal. Sci.* **9**, 647 (1993).
24. O. Imakyure, M. Kai, Y. Ohkura, *Anal. Chim. Acta.* **291**, 197 (1994).
25. H. Hirano, B. Wittmann-Liebold, *Biol. Chem.* **367**, 1259 (1986).
26. S. W. Jin, G. X. Chen, Z. Palacz, B. Wittmann-Liebold, *FEBS Lett.* **198**, 150 (1986).
27. H. Miyano, T. Nakajima, K. Imai, *Biomed. Chromatogr.* **2**, 139 (1987).
28. Z. Palacz, J. Salnikow, S.-W. Jin, B. Wittmann-Liebold, *FEBS Lett.* **176**, 365 (1984).

29. K. Muramoto, H. Kamiya, H. Kawauchi, *Anal. Biochem.* **141**, 446 (1984).
30. K. Muramoto, H. Kawauchi, K. Tuzimura, *Agric. Biol. Chem.* **42**, 1559 (1978).
31. H. Kawauchi, K. Muramoto, J. Ramachandran, *Int. J. Peptide Protein Res.* **12**, 318 (1978).
32. H. Maeda, H. Kawauchi, *Biochem. Biophys. Res. Commun.* **31**, 188 (1968).
33. H. Maeda, N. Ishida, H. Kawauchi, K. Tsujimura, *J. Biochem.* **65**, 777 (1969).
34. H. Kawauchi, K. Tuzimura, H. Maeda, N. Ishida, *J. Biochem.* **66**, 783 (1969).
35. L. R. Croft, *Introduction to Protein Sequence Analysis* (John Wiley & Sons, Toronto, 1980).

## Chapter 6

### Conclusions and Future Directions

#### 6.1 Introduction

The ability to analyse ever smaller amounts of proteins and peptides is becoming increasingly important in many different research disciplines. For this reason, a great deal of effort is currently being put forth both to develop new high sensitivity methods for protein analysis and to improve the sensitivity of methods that are already in use by protein researchers. This thesis describes the construction and evaluation of a unique, syringe pump based protein sequencer that promises to improve the sensitivity associated with primary structure analysis by two to three orders of magnitude.

Sensitivity enhancement with the syringe pump sequencer results from the use of capillary electrophoresis with laser induced fluorescence (CE-LIF) for analysis of the sequencing products. Chapter 2 of this work describes the optimization of a separation for the fluorescein thiohydantoin (FTH) amino acid derivatives of eighteen of the twenty coded amino acids. These FTH-amino acids are the expected sequencing products when fluorescein isothiocyanate (FITC) is used as the primary sequencing reagent. The limits of detection for the separation varied from a low of 3.4 zmol to a high of 12 zmol. This sensitivity is seven to eight orders of magnitude better than that obtained by HPLC with UV absorbance detection of the phenylthiohydantoin (PTH) amino acid products commonly generated during sequencing.

Commercial protein sequencers have been designed for use with HPLC. As a result, their volume requirements are not compatible with those of CE. Injection volumes in CE are typically no more than one or two nanolitres, so if CE-LIF is to be used for product identification, then a sequencer must be developed with this volume in mind. Chapter 3 describes a protein sequencer designed specifically to be used with CE-LIF detection. It incorporates syringe pumps for reagent delivery to a highly miniaturized reaction chamber.

The production of FTH-amino acid sequencing products requires a modification of the standard Edman chemistry whereby phenylisothiocyanate (PITC) is replaced by FITC. In Chapter 4 the effect of switching to FITC as the primary coupling reagent is evaluated for each step of the sequencing process. The results demonstrate that FITC can be used for sequencing with the syringe pump instrument.

Finally, Chapter 5 presents sequencing data using the syringe pump sequencer with FITC coupling and product identification by CE-LIF. The data clearly demonstrate that sequence information can be obtained from as little as one picomole of protein using this instrument. However, low repetitive yields severely limited the number of cycles that could be analysed from a given sample. Even so, the sequencer clearly showed that it has the potential to significantly improve the sensitivity associated with primary structure analysis of proteins and peptides.



## 6.2 Discussion

Although the syringe pump instrument was able to provide sequence information from 1 pmol of protein using the FITC coupling chemistry, the repetitive yield was too low and the chemical noise was too high to allow for the analysis of more than five residues. These problems must be addressed before the sequencer can be considered as a viable alternative to the standard Edman sequencers currently available. The development of a sealed reagent delivery system and reaction chamber to prevent contamination from atmospheric oxygen during sequencing should result in both an improved repetitive yield and a reduction in chemical noise.

One way to produce a sealed sequencer involves building a miniaturized valve block having the same basic design as one first proposed by Wittmann-Liebold (1). The miniaturized valve block can be constructed by endmilling a channel along one surface of a Teflon block and then drilling adjacent inlet holes for each reagent through the block. A 100  $\mu\text{m}$  deep channel with a width of 500  $\mu\text{m}$  would be compatible with the reagent volumes currently used for sequencing with the syringe pump sequencer. The inlet holes can be drilled to match the outer diameter of the reagent delivery capillaries. There will also have to be an outlet hole drilled directly through one end of the channel. The corresponding outlet capillary will be connected to a separate reaction chamber so that disassembly of the valve block will not be necessary whenever a sample is loaded. The reagent valve block can be sealed using a Teflon membrane and a second Teflon block. The membrane will be positioned between the two blocks, which will then be bolted together. The second block will have holes drilled through it that will match the position

of the inlet holes in the first block. These holes will allow for pressurization of the membrane with argon, which will in turn, seal the inlet holes on the other block. Reagent delivery can then be accomplished by releasing the pressure on the membrane over the desired reagent inlet hole followed by pumping the desired volume from the appropriate syringe pump. After delivery the membrane can be re-pressurized to prevent leakage. A schematic of the valve block is shown in Figure 6.1.

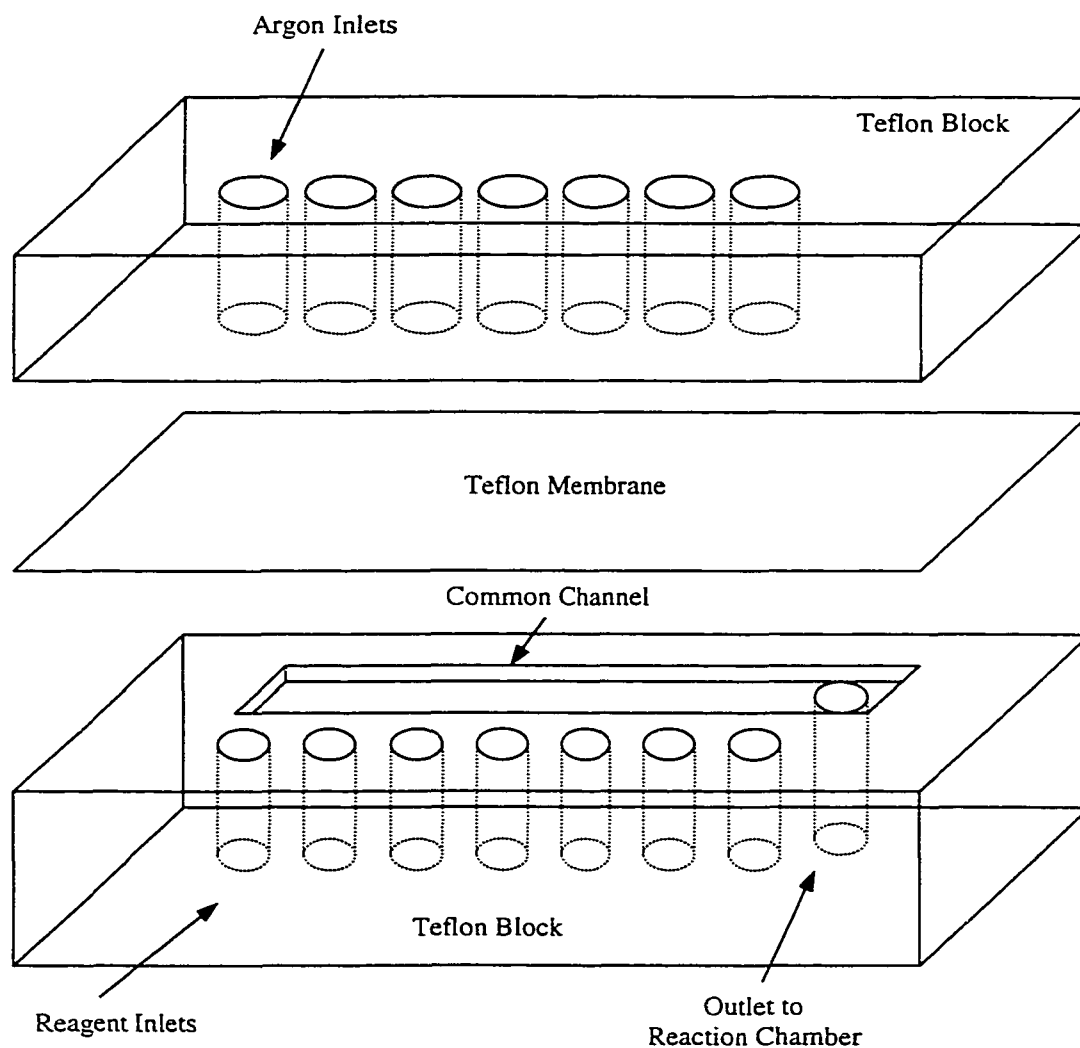


Figure 6.1 Schematic of an air tight valve block for the syringe pump sequencer.

A sealed reaction chamber can be constructed in a similar fashion to the valve block. A channel can again be endmilled into a Teflon block, but this time it will have to be widened in the centre so that it can hold a protein coupled glass fibre mat (GFM). Inlet and outlet holes will have to be drilled through the block intersecting the channel at each end. The outlet line will also have to be valved to prevent diffusion of oxygen into the reaction chamber during coupling and cleavage. This design will also make sample loading much easier and more reproducible than it is currently because it will be possible to accurately position the mat in the chamber instead of blindly inserting thin strips into a narrow channel. The reaction chamber can be sealed by tightening a smooth glass plate over the Teflon block. Using the glass plate will make it possible to monitor reagent delivery during sequencing and make troubleshooting easier. The reaction chamber can be heated through the Teflon block using a Feltier device positioned so that it sits under the GFM well. Figure 6.2 shows the basic design of this reaction chamber.

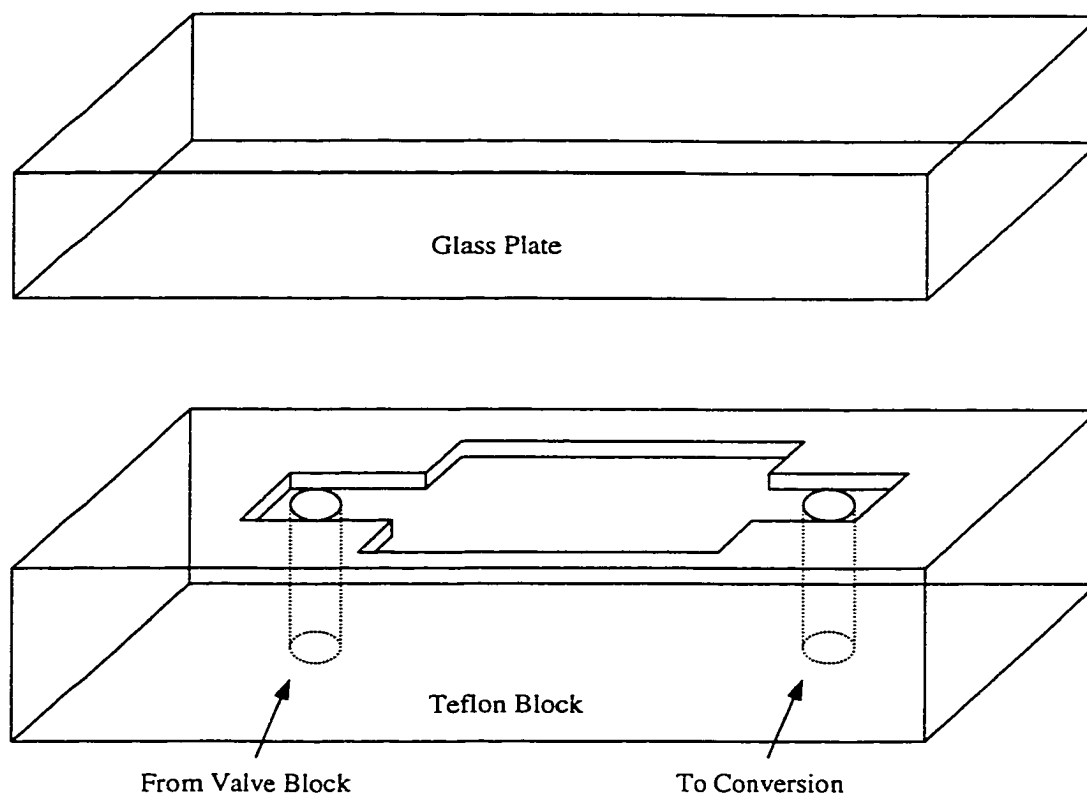


Figure 6.2 Reaction chamber schematic.

By using the valve block and reaction chamber above it should be possible to dramatically reduce contamination during coupling and cleavage. Such a reduction should lead to a significant improvement in repetitive yield. This improvement will result in the longer sequence reads required for the syringe pump sequencer with the FITC based chemistry to be useful for the analysis of unknown samples.

The sealed sequencer should also lead to a reduction in the size of the byproduct peaks in the sample electropherograms. Because it is this chemical noise that currently

limits the sensitivity of the syringe pump sequencer, any decrease will lead to an immediate improvement in sequencer sensitivity. All of the results presented in Chapter 5 were obtained using on a 50  $\mu\text{L}$  sample volume due to the size of the byproduct peaks in the separation. Even so, it was possible to sequence five residues from a 1 pmol sample of  $\beta$ -lactoglobulin. As such, a decrease in chemical noise that allows for a reduction in sample volume will make femtomole sequencing possible with the syringe pump sequencer. For example, if the sample volume can be decreased to 1  $\mu\text{L}$ , a volume that still allows for reproducible injection in CE, then it should be possible to sequence 20 fmol of protein. This sensitivity, in combination with increased sequence read lengths, will make the syringe pump sequencer a viable alternative to current sequencers for the high sensitivity analysis of proteins and peptides.

Although the modifications described above should make femtomole sequencing possible, the instrument will still require a significant amount of manual sample handling during the conversion and analysis steps. If these steps could also be automated, then the utility of the syringe pump sequencer would be further increased. The main obstacle to automated conversion and analysis will be the small sample volume required for CE. Significant miniaturization of the conversion chambers used by current sequencers will be necessary before the manipulation of 1  $\mu\text{L}$  samples will be possible. This work will benefit from current advances in the field of microfluidics. By using the same microlithography techniques employed in microchip fabrication, devices can be produced that are able to isolate and move and mix solutions at the low nanolitre level (2). These devices can also be used to carry out chemical reactions and analyse the resulting products (2, 3, 4). While these methods are still in the early stages of development, they

meet the exact requirements of an on-line conversion and analysis system for the syringe pump sequencer.

The protein sequencer described in this thesis incorporates major modifications with respect to existing instruments in three major areas: the detection methodology, the sequencing chemistry and the reagent delivery system. This first generation, proof of principle instrument was shown to have a sensitivity comparable to that of state of the art Edman sequencers. If it follows the trend established during the development of standard Edman sequencers, a further 100-fold to 1000-fold improvement in overall instrument sensitivity can reasonably be expected as the syringe pump sequencer is further refined and the FITC coupling chemistry is optimized. Such an improvement would make protein sequencing at the low femtomole level straightforward and would give researchers a valuable new tool for the study of weakly expressed proteins and peptides.

### 6.3 References

1. B. Wittmann-Liebold, in *Modern Methods of Protein Chemistry* H. Tschesche, Ed. (Walter de Gruyter, Berlin, 1983) pp. 229.
2. M. A. Burns, et al., *Science* **282**, 484-487 (1998).
3. A. G. Hadd, D. E. Raymond, J. W. Halliwell, S. C. Jacobsen, J. M. Ramsey, *Anal. Chem.* **69**, 3407 (1997).
4. A. T. Woolley, K. Lao, A. N. Glazer, R. A. Mathies, *Anal. Chem.* **70**, 684 (1998).