THESIS
2
2001

This is to certify that the

dissertation entitled

PART I: DETECTION AND SEQUENCING OF OLIGOSACCHARIDES AND PEPTIDES.
PART II: CHROMATOGRAPHY AND DETECTION OF SULPHOQUINOVOSE AND SUSPECTED METABOLITES FROM RHIZOBIUM MELILOTI bv. 1021

presented by

James J. Bradford

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Chemistry

R. I. Hollingsworth
Major professor

Date 19 May, 2000

MSU is an Affirmative Action/Equal Opportunity Institution

O-12771

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

11/00 c:/CIRC/DateDue.p65-p.14

Part I: Detection and Sequencing of Oligosaccharides and Peptides. Part II: Chromatography and Detection of Sulphoquinavose and Suspected Metabolites From Rhizobium meliloti bv. 1021.

By

James J. Bradford

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

2000

De

glycosyla

sophistic

These s

glycoforr

required

A

presence

All redu

The che

detectio

reaction

derivati

T

and sulf

The atte

far have

ABSTRACT

By

James J. Bradford

Detailed analysis of carbohydrate structures associated with glycosylation remains a major analytical challenge that requires sophisticated instrumentation and a variety of sample preparation strategies. These samples are often available only in limited quantities of each glycoform. Therefore, sensitive techniques that are highly reproducible are required for the structure determination of oligosaccharides.

All sugars are polyalcohols. Modifications to these sugars include the presence of an amine group, a carboxylic acid group, and/or an acetal group. All reducing sugars also are in equilibrium either as a ketone or aldehyde. The chemistry of each of these groups can be taken advantage of to improve detection and structural determination through derivatization. These reactions provide for analysis specificity and, through the formation of derivatives, detectability and structural information can be obtained.

The techniques used for the separation and detection of phosphorilated and sulfated saccharides are similar to the techniques used for saccharides. The attempts to resolve the biosynthesis of Sulphoquinovosediacylglycerol so far have proven fruitless. Only recently has it become feasible to rigorously

test the

biosynth

and the

test the different chemical and genetic approaches. Work on determining the biosynthetic pathway has fallen into two areas: the sulfoglycolytic pathway, and the sugar-nucleotide pathway.

For Sean and Kaitlyn
Live Long and Prosper[1]

I

none of

research

strongly

approach

and his

M

discussi

chemistr

I hadn't

theoreti

Biochem

and Ga

experie

directio

them.

I

deserve

people

that co

those s

ACKNOWLEDGMENTS

I have had many teachers and advisors that have influenced me, but none of them compare to my mentor Dr. Hollingsworth. His approach to research is similar to his approach to mentoring, very hands on. He is strongly focused and very intent on his approach to science. I know that my approach to mentoring will be derived from my experiences within his lab and his interactions with me.

My fellow lab mates past and present, with which I have had many discussions, certainly deserve mention. Jie and Gang opened my eyes to the chemistry occurring within the laboratory. Our discussions revealed a world I hadn't envisioned. Guanfei and Hussein showed me the practical and theoretical chemistry that can be achieved through personal diligence. Biochemical discussions with Carol and Atima, and conversations with them and Gabriela nearly always resulted in some fun and a good learning experience. Ben and Rob gave me a view of biochemistry from a different direction than the one I had been trained in, it was a pleasure working with them.

I have had many teachers that influenced me. The two that most deserve mention are Mr. Michael Crowl and Dr. Richard Lees. These two people showed me how to teach and gave me an appreciation for instructors that could treat students as people. Thanks to Mr. Crowl, I was able to find those students that didn't know how to interact with instructors and get

them to.

teaching

student

think ab

F

support

me to su

little ha

instance

helped

frustrat

possible

them to. Thanks to Dr. Lees I knew how to treat them as equals while still teaching. To them I owe the most rewarding complaint I ever received from a student "He was a tough grader. He never gave direct answers, he made us think about the answers."

Finally, I must thank my parents, friends, wife, and children. The support from them has been overwhelming. My parents have always wanted me to succeed at any project that I had started, and knew when to push a little harder for me to complete it. My friends have kept me going, in some instances literally, from Bob Hazekamp, to Preston Davis and Ken Roth, they helped me retain some form of sanity by listening to me when I was frustrated. My two beautiful children have affected me more than I thought possible, they bring joy to my life.

TABLE OF CONTENTS

LIST of TABLES and GRAPHS

LIST OF FIGURES

2.9
Spectr
pyrene
extrac
V. In
in spe
2.10
Spectr
pyren
extrac
increa
voltag
2.11
Spectr
pyren
extrac
of 37
respe
2.12
Resul
cone. T
Spect
voltag
increa
2.13
Spect
deriva
applie
respe
2.14
Spect
pyren
extra
Spect
has th
2.15
Spect
pyren
extra
V. In
in spe

LIST OF ABBREVIATIONS

2-AA	2-Aminobenzoic acid
2-AP or α-AP	α-Aminopyridine
2-D	Two Dimensional
ABEE	Aminobenzoic Ethyl Ester
CAD	Collisionally Activated Dissociation
CCD	Charge Coupled Device
CE	Capillary Electrophoresis
CID	Collision Induced Dissociation
D.P.	Degree of Polymerization
ECD	Electron Capture Detector
EI	Electron Ionization
ESI	Electrospray Ionization
FAB	Fast Atom Bombardment
FACE	Fluorophore-Assisted Carbohydrate Electrophoresis
FID	Flame Ionization Detector
GAG	Glucoseaminoglycans
GC	Gas Chromatography
HPAEC	High pH Anion Exchange Chromatography
HPLC	High Pressure Liquid Chromatography
ISAA	Ion-Suppression Amine-Absorption
LIF	LASER Induced Fluorescence
m/z	mass to charge ratio
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NMR	Nuclear Magnetic Resonance
ODS	Octadecylsilyl
PA	Pyridylamination
PAD	Pulsed Amperometric Detection
PAGE	Polyacrylamide Gel Electrophoresis
PAGEFS	PAGE Fluorophore-labeled Saccharides
PGPR	Plant Growth Promoting Rhizobacteria
PMA	Pyrenemethylamine
PMP	1-Phenyl-3-Methyl-5-Pyrazolone
PNGase F	Protein N-glyconase F
RIC	Reconstructed Ion Chromatogram
RP-HPLC	Reverse Phase HPLC
SFD	Saccharide Fluorophore Derivative
SQDG	Sulphoquinovosyldiacylglycerol
TIC	Total Ion Current
UV	UltraViolet
vis	Visible

PART I:
DETECTION AND SEQUENCING OF OLIGOSACCHARIDES AND PEPTIDES.

weight
They a
structu
biolog
transl
major
the e
glyco
and
glyco
func
play
mod
adhe
and
carb
stru
unde

Chapter 1

CARBOHYDRATE ANALYSIS
LITERATURE REVIEW

Introduction: Carbohydrates may comprise as much as 10% of the weight of some membranes, usually in the form of glycolipid or glycoprotein.[1] They are widely distributed in nature and the determination of their exact structure is often an important feature in the study of their functional or biological properties. Glycosylation is therefore a prevalent post-translational modification of proteins and these glycoproteins constitute a major segment of protein profiles in most eukaryotes.[2] It is also known that the extent of glycosylation is not static, the carbohydrate components on both glycolipids and glycoproteins are known to change during cell development and differentiation. These alterations in the carbohydrate moieties of glycoproteins affect their solubility, stability, and most of all their functionality. The glycan compliment, of glycoconjugates, has been shown to play important biological roles these include: protection from proteases, modulation of activity, molecular recognition, immunogenicity, neuronal cell adhesion, sperm and egg fertilization, homing of lymphocytes, and bacterial and viral interactions.[1] A precise description of the role played by carbohydrates in cellular processes is frequently compromised by their structural complexity and, as in many areas of biological science, understanding will require the pursuit of greater molecular detail. This

requir

charac

order

are ad

saccha

any o

any

Struc

hydr

fragr

linke

(hyd

Rece

the

adap

NMI

whil

Gel

comp

requ

requires more information than other biopolymers. Typically, in order to characterize a polymer it is required to know the monomer composition, the order of appearance and the size of the polymer. For oligosaccharides there are additional requirements such as: the configuration and ring form of the saccharide residues, the position and character of any substituent groups on any of the monomers, the positions of the inter-residue linkages including any branching points, and the configuration of the glycosidic linkages.[3] Structural elucidation of the carbohydrate moieties typically involves specific hydrolysis of the oligosaccharide followed by detailed analysis of the resulting fragments. Methods that have been previously used for characterizing the N-linked oligosaccharides attached to glycoproteins employ chemical (hydrazinolysis) or enzymatic (PNGase-F) release from the polypeptide chain. Recently, the increasing use of instrumental approaches has been used for the determination of composition and sequence of oligosaccharides. Some are adaptations of generally applicable analytical methods (e.g., two-dimensional NMR spectroscopy, reverse phase HPLC, GC-MS, FAB-MS, and ESI-MS) while others have been developed specifically for oligosaccharides (e.g., Bio-Gel P4 column chromatography, HPAEC, lectin blotting). The most comprehensive structural information is currently provided by NMR, but this requires large pure quantities (> 500 nmol) of sample.[4]

Carbohydrate Chemistry: All sugars are polyalcohols. Modifications to these sugars include the presence of an amine group (glucosamine, galactosamine, etc.), a carboxylic acid group (e.g., sialic acid), and/or an acetal group (N-acetylglucosamine, N-acetylgalactosamine, etc). All reducing sugars also are in equilibrium either as a ketone or aldehyde (Figure 1 and Appendix A and B). The chemistry of each of these groups can be taken advantage of to improve detection and structural determination through derivatization. These reactions provide for analysis specificity and, through the formation of derivatives, detectability can be improved.

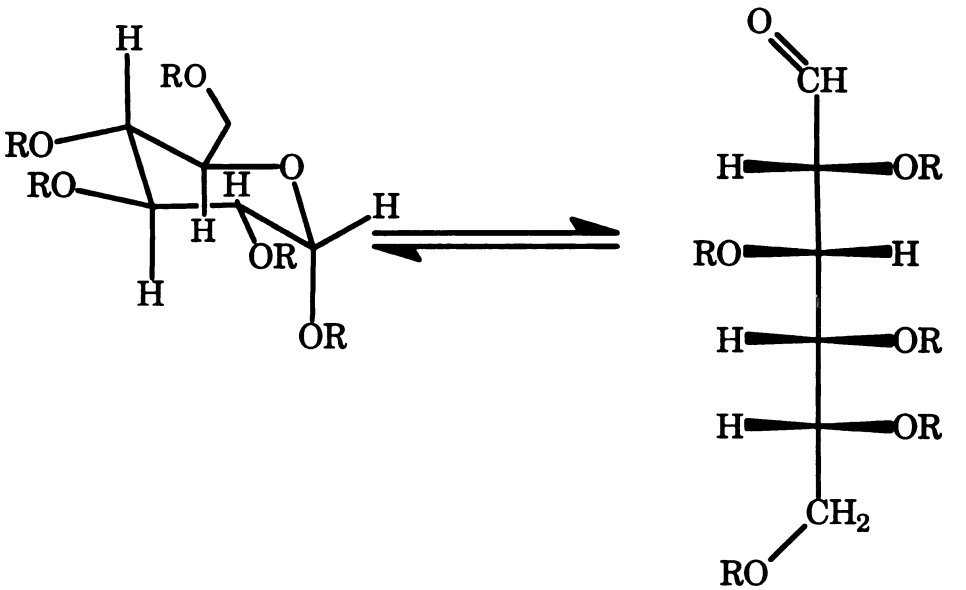


Figure 1.1: Isomerization of linear and pyranose form of glucose. For derivatization purposes, the linear form, containing the aldehyde, is required.

HPLC
for GC
GC an
detecta
and st
deriva
improv
chroma
applica
in the
deriva
than o
could b
20-30
with 7
al.7 an
uronic
the re
induce
only o
which

Derivatization of carboxylic acids has been employed for GC, MS, and HPLC analysis. Although monocarboxylic acids may be sufficiently volatile for GC, multifunctional carboxylic acids require derivatization for successful GC analysis. Various derivatives have been prepared which yield specific detectability in GC. Derivatization is also of value in improving volatility and stability of carboxylic acid compounds for MS analysis. For HLPC, derivatives have been used to confer UV absorption or fluorescence to improve detectability. The determination of carboxylic acids by chromatography continues to be a very widely used technique and is applicable to some modified sugars. Several reports deal with improvements in the sensitivity of the detection primarily by the used of fluorescent derivatives.[5] Recently, 1-pyrenyldiazomethane was found to be more stable than other diazomethane derivatizing agents and resulted in derivatives that could be separated by reversed-phase chromatography. Detection limits were 20-30 fmol[6]. In addition carboxylated carbohydrates have been derivatized with 7-aminonaphthalene-1,3-disulfonic acid as a fluorescent tag. El Rassi et al.[7] analyzed glucosaminoglycans (GAGs) by first digesting the bipolymer of uronic acid and hexosamine residues into disaccharides and then derivatizing the resulting fractions for separation by capillary electrophoresis and laser induced fluorescence (LIF). In this instance, the resulting disaccharide has only one acid group. For the purpose of quantitating large polysaccharides, which may have varying numbers of carboxylic acid groups on labile residues,

this approach would not give the necessary information to determine the degree of polymerization and therefore the concentration of each component. Since there may be a varying number of carboxylic acid groups, the concentration of the components in each chromatographic band would be uninterpretable since the number of derivatizing reagents present may vary from peak to peak; therefore, quantitation becomes impossible.

Derivatizing alcohols to improve detectability is more difficult than derivatizing carboxylic acids. Alcohols do not undergo S_N2 reactions very efficiently, since the -OH functionality is a poor leaving group. In addition, most alcohols are insufficiently acidic or basic to form stable ions by modification of solution pH. Therefore the most useful and general routes to produce derivatives of alcohols make use of reactions in which the oxygen acts as a nucleophile[8]. The derivatives of alcohols fall primarily into the three classes of alkyl, acyl, and silyl. Alcohols are insufficiently acidic (pK_a 16-19) to undergo alkylation by the common alkylation reagents but may be alkylated by use of very strong bases and alkyl halides or by diazomethane with suitable catalysis. Acyl derivatives of alcohols are most commonly prepared with acid anhydrides, acid chlorides, or reactive amide reagents. The classical technique for preparing acyl derivatives involves acylation with acetic anhydride, which may be base (frequently pyridine) catalyzed or acid catalyzed. The perfluoroacyl derivatives are more volatile than the corresponding nonfluorinated derivatives and may be used to impart

electron-capturing ability.[9] As with carboxylic acids, the use of fluorescent derivatizing agents enhances the sensitivity of the detection of alcohols in HPLC analyses. Anthracene-9-carbonyl chloride has been used as a derivatizing agent in the determination of short-chain alcohols.[10] For the purpose of quantitation, in this instance, it would be necessary to have a first hand knowledge of the number of –OH groups present on the saccharide since varying numbers of alcohols can be present on oligosaccharides. For this reason, the best approach to quantitation and detection utilizes the chemistry of a functional group that is represented only once on a reducing sugar.

Many of the derivatives that have been used for GC, GC-MS, MS, and HPLC analysis of aldehydes and ketones stem from the classical carbonyl derivatives such as hydrazones and oximes.[11] Whereas classically these derivatives had to be isolated, purified, and analyzed, modern chromatographic and spectrometric techniques can be applied to the derivatives without prior isolation. In the past, GC has been used for phenylhydrazones and dinitrophenyl hydrazones. The dinitrophenyl-hydrazones can be detected with ECD to enhance detection sensitivity by 2-3 orders of magnitude over that obtained with FID. The dinitrophenyl-hydrazones and the 5-dimethylamino-1-naphthalenesolfonyl hydrazones have been found useful for chromogenic derivatization for HPLC analysis. Other substituted hydrazones are also used as analytical derivatives, and the basic theme still offers potential for development of new reagents.

Carbohydrate Analysis:

Derivatization Techniques: The determination of carbohydrates begins with the knowledge of the monosaccharide composition, and the linkages between the individual residues. There are numerous derivatizing reagents that have been utilized for this purpose. The first application of GC to carbohydrates was reported in 1958 and described efforts to separate the fully methylated methyl glycosides of some monosaccharides.[12] The turning point toward widespread application of GC to carbohydrates was the publication in 1963 by Sweeley et al.[13] of the use of trimethylsilyl derivatives for GC analysis of carbohydrates including mono- through tetrasaccharides. In addition, alkyl derivatives of carbohydrates have been utilized in conjunction with acid hydrolysis of glycoside bonds and further derivatization.[13] Utilizing GC/MS it is then possible to determine the linkage positions between the residues by interpreting the fragmentation patterns associated with intramolecular bond cleavages. The fragmentation patterns associated with many of the derivatizing agents for GC/EIMS have been reviewed.[14] More recently, HPLC methods have been utilized to determine the monosaccharide composition. Kim[15] utilized reductive amination to determine the monosaccharide content of glycoproteins. His derivatization used p-aminobenzoic ethyl ester (ABEE) as the chromophore and RP-HPLC for separation. Fu and O'Neill[16] utilized a procedure developed by Honda and

coworkers[17] to derivatize monosaccharides with 1-phenyl-3-methyl-5-pyrazolone (PMP) and RP-HPLC to separate the resulting material. Another method developed by Sato et al.[18] used reductive amination with 2-aminobenzoic acid (2-AA) to derivatize monosaccharides that were then separated using capillary electrophoresis (CE) and detected using laser induced fluorescence (LIF). The monosaccharide compositions of oligosaccharides and glycoproteins were found to be highly accurate with similar or improved sensitivities to those assays previously developed.

These high resolution separation techniques, coupled to detectors that have extremely high sensitivity are required. This is due to the diversity in the oligosaccharide structures, combine this with the multiple sites of glycosylation that may occur, results in an impressive array of glycoforms, separating these glycoforms requires high resolution separations. These different structures can originate from glycoprotein heterogeneity that can be defined as the varying carbohydrate core branching patterns and composition associated with a unique protein/peptide glycosylation position. Each core glycostructure can exhibit further structural modifications to a given carbohydrate moiety, thus greatly increasing the structural diversity or microheterogeneity. The heterogeneity and microheterogeneity have been shown to vary between cell types and stages of development. Due to these variations in the structures from cell to cell and time to time, it is not

possib
glycar
protei
that
prepa
weigh
these
multi
and
Addit
each
that a
oligos
detect
data.
neces
peaks
used
of flu
assis

possible for a monosaccharide analysis to give an accurate picture of the glycan.[19]

Detailed analysis of carbohydrate structures associated with unique protein/peptide glycosylation position(s) remains a major analytical challenge that requires sophisticated instrumentation and a variety of sample preparation strategies.[20] For some glycans, a determination of molecular weight alone would be adequate to identify structure and ES profiling of these samples is direct and informative. However, it is common to find multiple structures (glycomers) that are consistent with an established mass, and alternative strategies are necessary to resolve these glycoforms.[21] Additionally, these samples are often available only in limited quantities of each glycoform obtained from the sample. Therefore, sensitive techniques that are highly reproducible are required for the structure determination of oligosaccharides. The goals of these derivatizations are to improve detectability of the saccharides and ease the interpretation of the resulting data. Many approaches have been taken to improve the separations necessary to remove ambiguities inherent in overlapping chromatographic peaks.

Polyacrylamide Gel Electrophoresis: One technique that is being used to determine saccharide structures is polyacrylamide gel electrophoresis of fluorophore-labeled saccharides[22] (PAGEFS) also known as fluorophore-assisted carbohydrate electrophoresis (FACE). In this method of glycan

analysis, the saccharides with hemiacetal functions are covalently labeled with a fluorophore and the derivatives are separated by polyacrylamide gel electrophoresis (PAGE). The saccharide-fluorophore derivatives can be subjected to PAGE in high-density gels yielding high-resolution separations. Detection of the bands on the gels was obtained in some instances by illuminating the slab gel on a UV transilluminator and photographing the resulting fluorescence to give high-quality negatives which could be scanned densitometrically. Over the range of 10 to 500 pmol, of saccharide per band, the response was nonlinear, but repeatable. The lower limit of detection using this technique was about 1 pmol.[3] Alternatively, gels are imaged digitally using a camera system based on a cooled charge coupled device (CCD). The quantitative potential of the system was shown to be good in that a linear response was obtained over the range 12 to 500 pmol per band. Repeatability was similar to that obtained with the photographic system. The sensitivity of the CCD camera was about 0.2 pmol per band. Digestions are carried out using glycosidases that will give an indication as to the sequence, linkage type, and, ultimately, structure of the saccharide, and can be performed either before or after the saccharides have been labeled with a fluorophore. In the enzymatic structural analyses described above, the glycan structural features are deduced by interpretation of the changes that occur after specific glycosidase treatments.

Column Chromatography: Column chromatography has long been one of the preferred techniques for separating mono and oligosaccharides. High-pH anion-exchange chromatography (HPAEC) with a pulsed amperometric detector (PAD) is a powerful tool for analysis of oligosaccharides from glycoproteins. There are also many protocols for the separation and analysis of neutral and acidic oligosaccharides[23-26] and monosaccharides.[23,27-29] Among the more recent variations on these protocols, Hayase et al.[30] used HPAEC-PAD for O-linked saccharides. They found that reduced desialylated O-linked oligosaccharides, especially those with short chains, are barely retained on a HPAE column because they lack the negative charge of either a sialic acid or an acidic anomeric hydroxyl group. In their work, the desialylated oligosaccharide-alditols were de-N-acetylated and then N-succinylated before HPAEC separation. The succinylation introduces a negative charge on an amino group of each amino sugar residue and causes the oligosaccharides to be retained much longer on a HPAE column. HPAEC was also used by Clarke[31] to separate peptidoglycan hydrolysates from several species of Proteeae. The single analysis afforded base line separation of glucosamine, the basic, and neutral, amino acids in an isocratic period and the peptide portion during a gradient segment.

Ion Exchange Chromatography: HPAEC-PAD is very sensitive, capable of very high resolution, and can separate very similar anionic oligosaccharide structures. However, while excellent for analytical purposes

HPAEC-PAD is less suitable for preparative work since sodium hydroxide can cause epimerization and degradation of sugars.[32] Also, the use of nonvolatile buffers requires an additional step for their removal. For buffers such as sodium acetate the nonvolatile cations can be removed using cation-exchange resin and then the volatile anions can be evaporated. However, this would require the use of large amounts of cation-exchange resin. Buffers with nonvolatile anions such as phosphate can be removed by gel filtration but this is time consuming.[33,34] The use of anion exchange columns is not applicable to mass spectrometry without extensive desalting procedures. Also, gradients cannot be used with refractive index detection which limits the applicability.

Several studies have shown that bonded-phase columns bearing amine functional groups can be used to separate oligosaccharides by HPLC. Ion-exchange HPLC[35] fractionates oligosaccharides into groups with different charges, but these remain heterogeneous with respect to carbohydrate content and linkage. Additionally, amine adsorption HPLC[36] cannot proceed unless the anionic fraction is removed. A different analytical procedure, ion-suppression amine-adsorption HPLC (ISAA-HPLC)[37] can fractionate neutral and anionic oligosaccharides, but does not adequately separate large neutral oligosaccharides or highly sialylated small oligosaccharides. By modifying the conditions used for amine adsorption and ion-exchange HPLC, Kondo et al.[36] established methods for the separation of complex mixtures of neutral

and

carbo

using

deriv

with

are

amin

reduc

chron

appli

at pH

corre

a UV

been

polyr

et al

resid

pyrid

with

oligo

rever

and sialylated PA-oligosaccharides based upon differences in net carbohydrate content (size) and the number of sialic acid moieties (charge), using an amine-bearing silica column.

Reductive Amination: To study closely related oligosaccharides derived from glycoconjugates, it is essential to be able to fractionate them with good resolution. A pyridylamination method in which oligosaccharides are fluorescently labeled with α-aminopyridine (α-AP), by reductive amination, has been developed by Hase et al.[38] Since its introduction reductive amination has been an important approach to placing a chromophore onto reducing sugars, and the approach has seen many applications. Initially, the specimens were analyzed by paper electrophoresis at pH 5.0. Each sugar derivative giving a single spot in addition to one corresponding to excess α-aminopyridine when the paper was scanned under a UV lamp. Since then, many primary amines with large chromophores have been utilized to obtain information on oligosaccharide degree of polymerization, and to quantitate the amount of saccharide present. Kondo et al.[39] reported an improved pyridylamation procedure in which sialic acid residues are not released during the reaction. The combination of pyridylamination with HPLC allows sensitive analysis of oligosaccharides, with a detection limit at the femtomole level of pyridylamino (PA)-oligosaccharides. PA-oligosaccharides have been analyzed by HPLC on reversed-phase and normal phase columns, and the elution positions of many

PA-oligosaccharides have been plotted two-dimensionally by Hase et al.[40] They separated pyridylaminated oligosaccharides by two-dimensional sugar mapping, a combination of size-fractionation and reversed-phase HPLC. This technique allows each sugar residue to make its own contribution to the elution time (partial elution time), and therefore the elution time of a PA-oligosaccharide is calculated as the sum of all contributing partial elution times. Utilizing this approach mono, di, and trisaccharides were obtained by partial acid hydrolysis and compared to authentic pyridylaminated standards. Under these conditions, mannose and glucose derivatives could not be separated in this system, although they were well separated by TSK gel sugar AXI column.[41] Hase et al. and Tomiya, et al.[42] have compiled an extensive 2-D map for neutral PA-oligosaccharides. This same group has also extended the use of this 2-D map to include GalNAc-containing N-linked oligosaccharides.[43] To utilize this map, the structure of unknown oligosaccharides are estimated by comparing the elution position (expressed in glucose units) of an unknown sample with those of standard oligosaccharides analyzed by both reversed- and normal-phase chromatography. The resulting coordinates are plotted as a two-dimensional map. The structure of unknowns is then characterized by comparing the elution position with those of the standard oligosaccharides on the map. In this study, the 2-D positions of the GalNAc-containing oligosaccharides did not correspond to any of those of the known GalNAc-free oligosaccharides.

After
chan
to th
degre
used
inclu
et al.
octyl
has
attac
2HCl
deriv
anth
al.[46]
struc
capil
comb
comb
mixt
cons

After β-N-acetylhexosaminidase digestion, however, their elution positions changed and coincided with known positions on the original map.

Derivatization with a chromophore of high, known, molar absorptivity to the reducing saccharide is key to quantitating saccharides, based on the degree of polymerization. Reverse phase HPLC is one of the most widely used techniques for separating oligosaccharides. Samples of this technique include the assays developed by Hase, as already mentioned. Hollingsworth, et al.,[44] also utilized reverse phase HPLC to separate N,N-(2,4-dinitrophenyl) octylamine derivatized oligosaccharides for FAB-MS analysis. Anumula[45] has developed an HPLC assay for sialic acids where a label is specifically attached, the fluorescent tag utilized in this instance is o-phenylenediamine 2HCl, and this yields a stable fluorescent quinoxaline derivative. Other derivatizing reagents investigated include 2-amino benzamide and 2-anthranilic acid, both were utilized to analyze N- and O-glycans by Bigge et al.[46]

Microbore HPLC: The characterization of the gross covalent structure of glycoproteins has been revolutionized with the advent of capillary and microbore high performance liquid chromatography (HPLC) combined with electrospray ionization mass spectrometry (ESI-MS). Using a combination of proteolytic enzymatic digestion and separation of the digest mixture by HPLC/ESI-MS, a glycoprotein mass map can be obtained consisting of both peptides and N- and O- linked glycopeptides.[47-50] This

methodology allows the nature and extent of site heterogeneity, including sialylation, of intact native glycopeptides. The location of those components bearing glycosylation in such digests is usually difficult because of inadequate chromatographic resolution. However, as recently reported by Carr et al.[49], collisional activation of electrospray generated ions, fragmentation occurring either in the source or within the second quadrupole of a triple quadrupole mass analyzer, induces dissociation reactions preferentially within the carbohydrate moieties themselves. The fragment ions are specific to the oligosaccharide structure attached to the peptide backbone.[51,52] Hence, monitoring of these, lower mass, carbohydrate ions throughout the chromatogram permits detection of glycosylated species. Burlingame[53] reported the use of this technique utilizing an ethanol: propanol: water: formic acid solvent system for the reversed-phase separation and electrospray mass spectrometric detection of tryptic peptides and glycopeptides from bovine fetuin. However, because of the complexity of carbohydrate heterogeneity, which is often present within and among sites in a glycoprotein, it is not possible in most cases to completely resolve the individual glycoforms using only one method. Depending on the degree of glycosylation, several glycoforms from the same amino acid sequence usually coelute, making interpretation of such mass spectra difficult. In addition, when several components (whether peptides and glycopeptides or glycopeptide glycoforms) coelute, possible problems with suppression and/or

usabl

meas

some

ident

infor

that

neces

seque

seque

fairly

gene

ultim

impo

deter

activ

(i.e.,

the

need

dime

obta

usable dynamic range may arise, especially when attempting to detect and measure nonstoichiometric quantities of glycopeptides by LC/ESIMS. For some glycans, a determination of molecular weight alone is sufficient to identify structure and ES profiling of these samples is direct, rapid, and informative. However, it is common to find multiple structures (glycomers) that are consistent with an established mass, and alternative strategies are necessary to resolve these glycoforms.

Peptide Analysis: Traditionally, short stretches of amino acid sequences have been used to search databases to locate protein or gene sequences, or to identify proteins of similar sequence. This constitutes a fairly unique address for a protein and can lead to successful searches. It is generally accepted that the molecular structure of a globular protein is ultimately determined by its amino acid sequence, within the constraints imposed by its chemical environment, and that the molecular structure determines its biological function.[54] The determination of the specificity and activity of a protein depends on obtaining the correct sequence of amino acids (i.e., primary structure) that comprise the polypeptide chain. This includes the knowledge of the presence of specific post-translational modifications needed for the proper folding of the oligopeptide chain to a precise three-dimensional structure. The amino acid sequence information is usually obtained by N-terminal analysis using Edman degradation. Conventional

sequencing strategies typically employ chemical reagents to remove one amino acid at a time from the amino terminus followed by isolation and analysis of the released amino acid derivative.[55,56] Mass spectrometry offers the advantage of speed and sensitivity in protein primary structure determination and also has a unique and important role in the determination of structural changes introduced into proteins by post-translational modifications. Both chemical degradation and peptide mapping approaches require the use of fairly homogeneous samples to avoid ambiguity in assigning the amino acid sequence and establishing the origin of the set of measured peptide ions, respectively. These approaches are not well suited to the study of biological problems where complex mixtures of peptides may be isolated.

Innovations such as electrospray ionization have led to improved methods for interfacing liquid separation techniques to mass spectrometers. Combining microcolumn liquid chromatography (micro-bore HPLC) with tandem mass spectrometry improves the ability to manipulate, for sequence analysis, small quantities of peptides contained in complex mixtures. Ions of a single mass-to-charge ratio can be selected for collision-induced dissociation (CID) even though that peptide may be introduced into the mass spectrometer as part of a mixture of peptides.[57] Under CID conditions, peptides fragment to create patterns characteristic of a specific amino acid sequence. These patterns are reproducible and, in general, predicable. But

even still, interpretation of a low energy CID spectrum for an unknown sequence proceeds by identifying a consecutive series of fragment ions whose differences correspond to residue masses for amino acids. This ion series can correspond to either the type –b or –y ion series or a combination of both (see next section for a description of the nomenclature). Interpretation of limited m/z range, low-resolution tandem mass spectra from multiply charged precursors can present difficulties due to the possibility of multiply charged product ions and the lack of unambiguous charge-state information. Even so, ESI/MS with instruments usable over a relatively low m/z range (<2000) have been shown to be attractive for molecular weight measurements of large biomolecules.[58,59] Electrospray interfaces serve as the basis for a near ideal interface for liquid chromatography and a range of separation formats. The potential of these methods is further enhanced by tandem mass spectrometry (MS/MS), which, ideally, permits efficient dissociation of the precursor molecular ion into product ions that are structurally distinctive to be useful for fingerprinting.

Electrospray techniques have found application in accurate measurements of molecular weights of peptides and proteins by exploiting the phenomenon of multiple protonation, which yields a coherent series of progressively more highly protonated species $(M + nH)^{n+}$. Smith and coworkers have published an excellent review.[60] For larger molecules, the more highly charged ions are more susceptible to dissociation.[61] Because

collision energy is proportional to the number of charges at a given m/z, it is reasonable to assume that the greater number of charges is a primary reason for an increased CAD efficiency. In addition, efficiency for MS/MS of multiply charged ions may be enhanced due to electrostatic repulsive forces.[62] The relatively facile fragmentation of extremely large proteins in their multiply-protonated forms doubtless reflects, in part, the destabilizing influence of the Coulombic repulsions within the multiply charged species.[63]

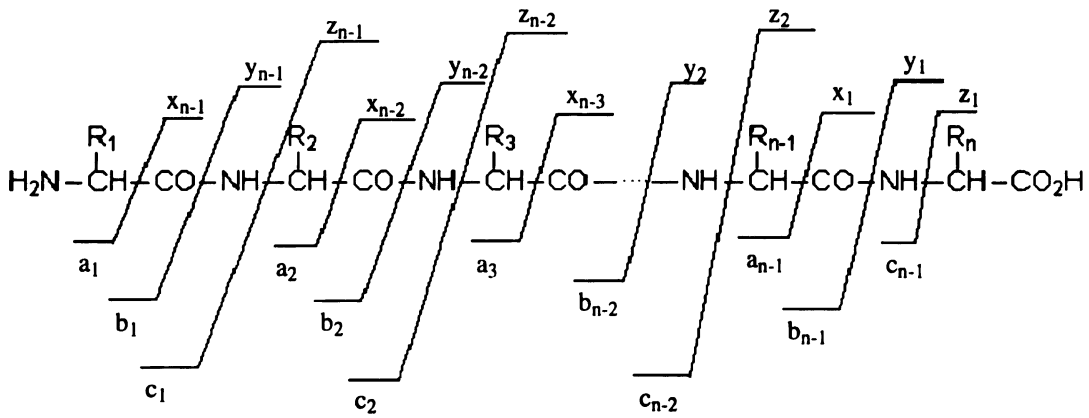


Figure 1.2: Peptide fragmentation nomenclature, typically for low energy CID only b and y type fragments are observed.

Peptide Nomenclature: Compared to the complexity of the saccharide fragmentation nomenclature, the system used for peptides is simple, particularly for low energy CID. The full series of fragments is indicated in Figure 1.4. Not only are fragments that account for backbone bonds breaking included, but also for side chains, just as in saccharide

frag

typi

the

addi

the

grou

bond

for

the

fragmentation. Fortunately, for low energy CID, the only fragments that are typically present are the -b and -y series. These are the ions that account for the sequence of the peptide and form from breaking the peptide bonds. Any additional fragmentation that may occur typically only relays information on the side chains. The fragmentations that arise from the side chains, the "R" groups, typically do not occur, nor do the fragmentations that arise from bonds breaking other than the amide linkage. These fragments are reserved for high energy CID, which may occur on magnetic sector instruments due the additional energy imparted during acceleration.

HC

for s

here

appl

clea

nstr

om

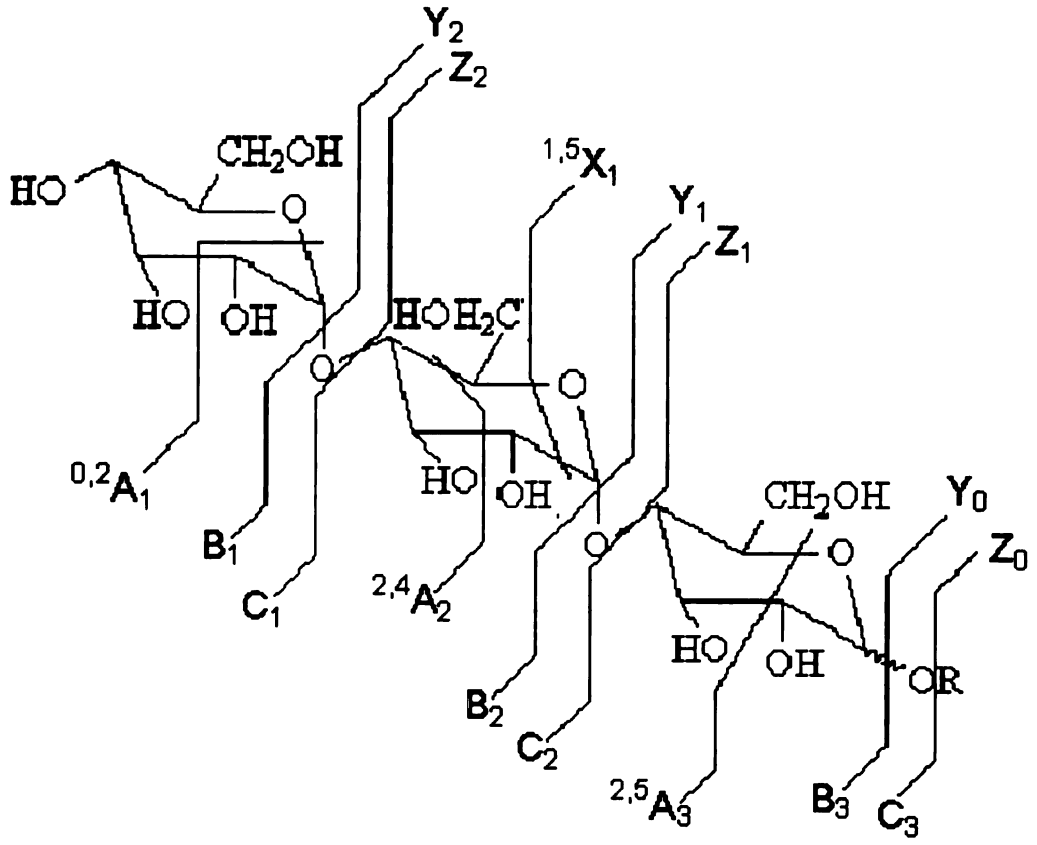


Figure 1.3: Fragmentation nomenclature for linear saccharides as developed by Domon and Costello.[64]

Carbohydrate Nomenclature: A modification of the nomenclature for saccharides that was developed by Domon and Costello[64] will be used here. Their system, developed for FAB ionization and high energy CID is applicable here since the fragmentations are sufficiently similar. Some ring cleavages observed at the high potentials utilized with magnetic sector instruments (high energy CID) are not observed, but are included for completeness.

The types of fragmentations that occur in glycoconjugates are more complicated than that of other large biomolecules, such as peptides and nucleic acids. Whereas with these other biomolecules the polymer of interest is linear, saccharides have numerous possible branch, or connection, points and all of these must be taken into account when revealing and reporting the structure. Therefore, it is necessary to define all connectivities, however determined. This system expands upon the nomenclature used for peptides. Here, A_i, B_i, and C_i labels are used to designate fragments containing a terminal (non-reducing end) sugar unit, whereas X_j, Y_j, and Z_j represent ions still containing the aglycone (or the reducing sugar unit). Subscripts indicate the position relative to the termini analogous to the system used in peptides, and superscripts indicate cleavages within carbohydrate rings (Figure 1.2).

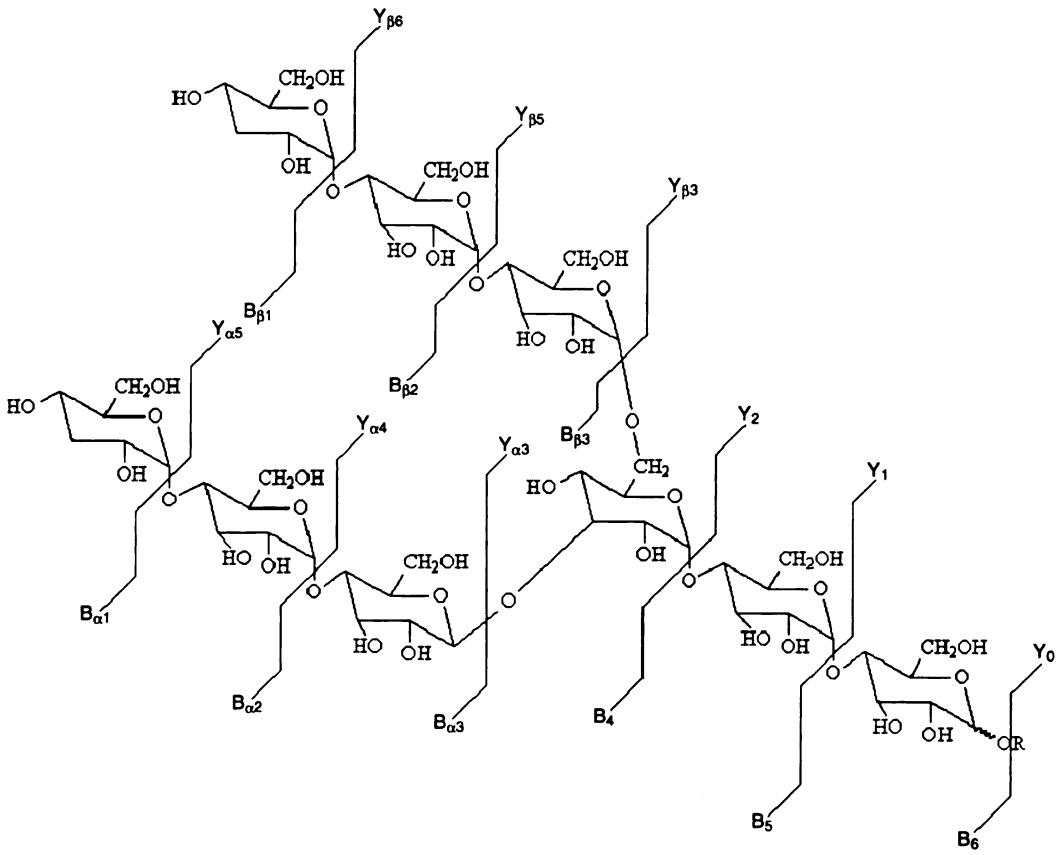


Figure 1.4: Nomenclature for branched saccharides as developed by Domon and Costello.[64]

As previously stated, oligosaccharides are unique for biomolecules in that their three dimensional structure includes branching. To describe this branching in the fragmentation nomenclature the convention introduced has been extended. The "core" of the oligosaccharide is labeled as previously introduced, the branches are labeled with the Greek letters α, β, γ, ..., in order of decreasing molecular weight (Figure 1.3). In the instances where two fragments are obtained from the same sugar, the Greek letters used to

represent each branch are given together (e.g., $Y_{\alpha\beta}$). Primes are used to indicate branch points on saccharides, other than the core region, the number of primes increases in the order of decreasing molecular weight of each sub-branch.

Chapter 2

STRUCTURE DETERMINATION OF REDUCING SACCHARIDES

Introduction: The rapid expansion of knowledge in the field of molecular biology during the third quarter of the twentieth century did not include carbohydrates. This exclusion is due to a preoccupation with proteins and nucleic acids. But, as contemporary researchers try to bring chemical reasoning to physiological function, carbohydrates and their conjugates are found as crucial participants in cellular processes.[65] A precise description of the role played by carbohydrates in cellular processes is frequently compromised by their structural complexity, and as in many areas of biological science, understanding will require the pursuit of greater molecular detail. The pursuit of this detail has its cost though; the analysis of carbohydrates is impeded by the lack of outstanding structurally detectable features. There are no groups that fluoresce or absorb well, nor groups that are easily ionized.

From a structural standpoint, oligosaccharides are ideally suited for events requiring molecular specificity. Their polyhedric character provides a platform with numerous functional groups for modification and interaction, while monomers within an oligomer can participate in multiple linkage and branching patterns, creating an array of structural possibilities from a

limited set of monosaccharides.[66] Traditional methods of carbohydrate analysis rely on chromatographic and electrophoretic techniques.[67-69] Since most sugars do not contain chromophoric moieties, the methods of direct detection of sugars are thus limited. This problem is compounded since the purification of N-linked, or O-linked, oligosaccharides from glycoproteins is problematic, a single glycoprotein often contains numerous types of closely related oligosaccharide structures.[70] Furthermore, it may be necessary to prepare milligram quantities of individual oligosaccharide glycomers in order to perform complete structural characterization, by NMR, and to assess bioactivity.

Although sugars do absorb at 185 nm,[71] the detection sensitivity is poor due to their low molar absorptivities.[72] Underivatized oligosaccharides have been purified using high-pH anion exchange chromatography (HPAEC).[73] The major advantage of this approach is that the isolated oligosaccharides retain their reducing end and thereby can be derivatized for further analysis. In this light it has been suggested that no single technique can be used to determine a complete saccharide structure. That is, the problems inherent in determining the degree of polymerization and the linkage positions cannot be determined by only one technique. Typically multiple stages of separation, and/or detection are involved in the analysis of oligosaccharides.

HPLC has played an important role in carbohydrate analysis. Several methods for quantification of monosaccharides without derivatization have been described.[74-77] Refractive index detectors are used for detecting underivatized sugars. However they are insensitive, nonspecific, and impossible to employ under gradient conditions. HPAEC-PAD has been accomplished with high sensitivity and selectivity,[78-82] but PAD has the limitation of possible electrochemical modification of the detected species. Borate anion-exchange chromatography with suitable postcolumn detection is reliable and easily implemented, but requires long analysis times and is relatively insensitive due to peak broadening.[83]

Utilizing a combination of techniques thus appears required for structure determination of saccharides. To this end a combination of HPLC with absorbance detection and ESI-MS has been employed in the work reported here to obtain not only information on the degree of polymerization, but also information on the linkage positions for numerous saccharides. Derivatization melds carbohydrate structural analysis with UV/Vis, or fluorescence, and is required for three reasons: First, detection sensitivity is increased, second, it enables quantitation, based on Beer's law, and third, with charge localization on the derivatizing reagent, structural features are more evident. Therefore, our procedure calls for the saccharides to be derivatized first, for ease in quantitating the individual species present. The HPLC separation is used to separate the individual saccharides to obtain

information on the degree of polymerization. Then, once separated, a UV/Vis detector set at the absorption maximum of the derivatizing reagent allows quantitation based on Beer's Law. The fractions obtained from the separation are then introduced into an electrospray mass spectrometer and, partially, fragmented. Typically, tandem analyzers are used for these fragmentations so that ions may be separated in complex mixtures and individually studied by CID. This dramatically increases the structural information available and introduces another level of separation, based on mass and charge. But, low energy CID is also available in electrospray by utilizing in-source fragmentation and a single quadruple mass spectrometer. Since the bonds that are fragmented by low-energy collisions, typically connect the monosaccharide residues, the resulting fragments provide monomer sequence and branching information. This allows the linkage position of the species to be determined and also confirms the molecular weight and, therefore, the degree of polymerization. The data obtained with this procedure are partially redundant, in order to decrease the chances of misinterpreting the degree of polymerization.

Materials: Pyrenemethylamine was purchased from Sigma, benzylamine, analine, and 2-aminopyridined were purchased from Aldrich. The maltodextrin ($n \leq 15$) was obtained from Synthon. Most sugars (arabinose, fructose, fucose, gentiobiose, glucose, lyxose, mannose, melibiose,

rhamnose, ribose turanose, and xylose) came from Sigma. Sophorose and rhamnose were also from Sigma. Galactose came from Aldrich, while lactose was obtained from Kodak.

Procedures: The reducing saccharide(s) to be derivatized are added to a one-dram vial, with two equivalents of derivatizing reagent and two equivalents of sodium cyanoborohydride. To this mixture, two hundred microlitres of water: methanol solution (1:1) with 5% acetic acid is added. The vial is capped with a screw cap, having a Teflon liner, and heated at 65 °C for one hour. Subsequently the solution is dried by evaporation using an N-evap. The sample is then ready for further analysis.

Most saccharides were purified by passing them through a column (6 x 1.5 inches) of silica with a mobile phase of isopropanol: ammonium hydroxide: water (6:4:1). These fractions were collected for further analysis. Aliquots of the samples were then spotted onto TLC plates, and the chromatograms were developed in a solvent of isopropanol: methanol: ammonium hydroxide: water (1:1:1:1). The resulting zones were visualized by observing the fluorescence of the derivatizing reagent under a UV lamp. Additionally, orcinol was used to visualize the saccharides. The fractions that were found to be fluorescent and to contain saccharides were then analyzed by mass spectrometry.

Reversed-phase HPLC analysis was accomplished using a Waters 600 (Milford, MA) multisolvent delivery system equipped with a Kratos (Ramsey,

NJ) Spectroflow 783 programmable absorbance detector set to monitor the absorbance maximum for each derivatizing reagent (240 nm for PMA). The separation of derivatized saccharides is achieved in either eighty or forty minutes on a Beckman (San Ramon, CA) Ultrasphere ODS column (4.6 mm x 25 cm with a mean particle diameter of 5 μm). The first separation has the elution performed at a flow rate of 0.4 ml min^{-1} at 30° C using a gradient. For maximum resolution of all saccharides, a mobile phase that initially consisted of 10% methanol: water, with 0.1% TFA was utilized, this was increased linearly to 100% methanol over seventy minutes after a ten minute hold. Alternatively, for increased resolution of mono through tetrasaccharides, the mobile phase was instead modified with 0.1 % HCl initially not only did this increase resolution for the lower degrees of polymerization, but also moved the bands further away from the derivatizing reagent and decreased the retention times.

Mass spectrometry was performed on a Platform (Fisons, MA) single quadrupole mass spectrometer equipped with an electrospray interface and operated in the positive ion mode. Solutions of analytes were obtained by dissolving approximately 1 μg of sample in 1 ml of methanol: water (1:1), this was diluted further as needed. Samples were introduced into the ESI source with a Sage 361 syringe infusion pump at a flow rate of 20 μl min^{-1}. The filter electrode was heated to 100 °C, the nebulization gas flow rate was 18 l hr^{-1},

and

optim

and the drying gas flow rate was 350 l hr^{-1}, all other parameters were optimized for maximum sensitivity while maintaining unit resolution.

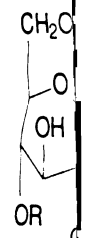
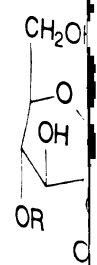
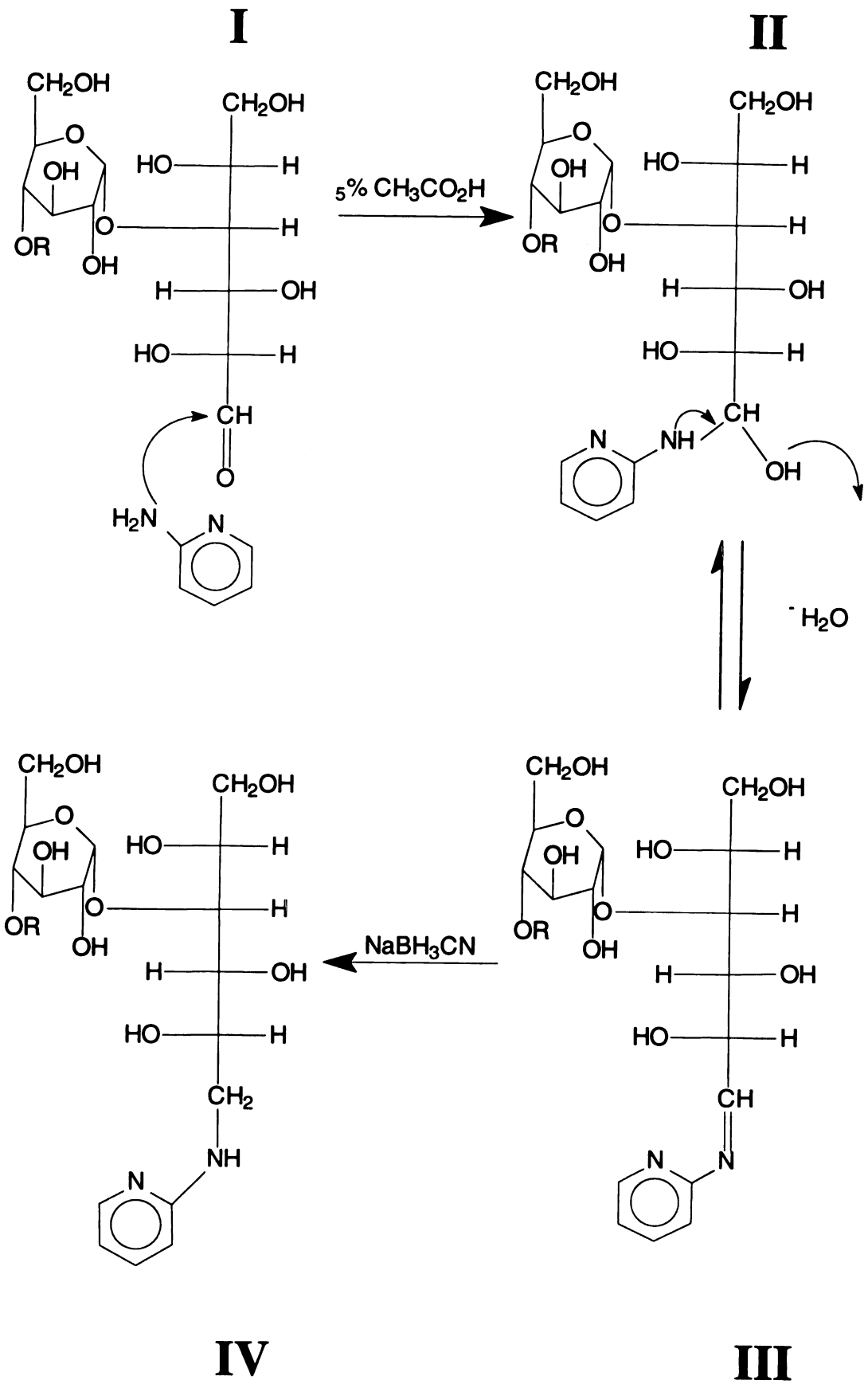


Figure 2.1: Scheme for reductive amination of saccharides with a derivatizing reagent. This example utilizes α-aminopyridine. For most experiments pyrenemethylamine was used.




I
II
III
IV
5% CH3CO2H
- H2O
NaBH3CN

0'00

Figure 2
with PM
increasin

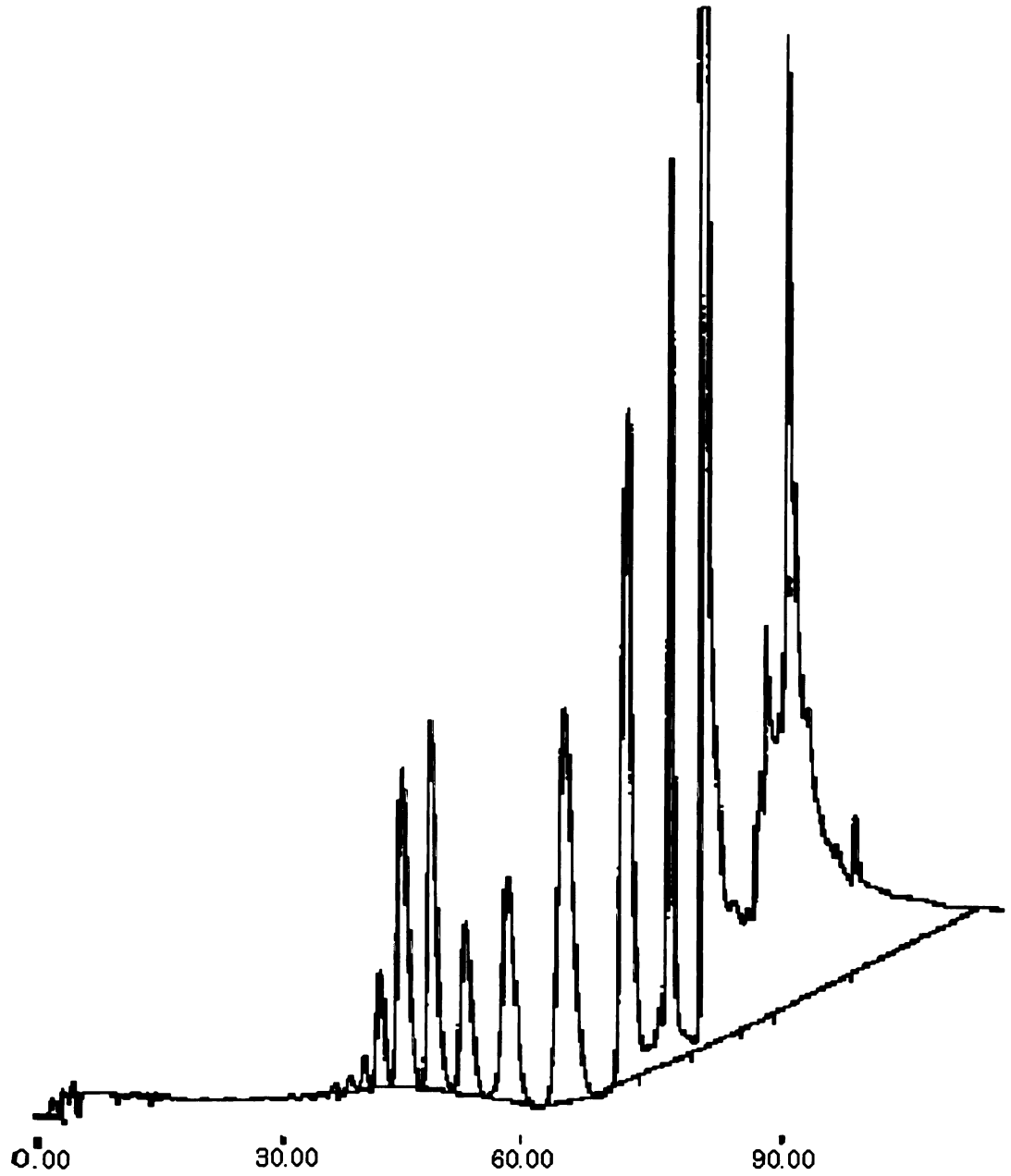


Figure 2.2: Maltodextrin separated on an ODS column after derivatization with PMA. The degree of polymerization of the maltodextrin decreases with increasing retention time.

R

performe

principle

aldose or

amino gr

water fro

is revers

reduction

cyanobor

Th

HPLC us

obtained

from one

detected

chromato

undertak

presented

experime

disacchar

in retenti

from tab

monosacc

Results and Discussion: Reductive amination (Figure 2.1) was performed using a modification of the procedure outlined by Hase et al.[35] The principle of the method is as follows: Saccharides (I) which are either an aldose or a ketose are reacted with a fluorophore which contains a primary amino group to yield a carbinolamine, or Schiff base (II). With the loss of water from the carbinolamine, an imonium ion is formed (III). This reaction is reversible and to obtain the final product, the imine (III) is trapped by reduction to a secondary amino derivative (IV) using sodium cyanoborohydride, a mild reducing agent.

The saccharide-fluorophore derivatives (SFDs) were separated by HPLC using a reversed phase, ODS column, where baseline separation is obtained for α1-4 linked maltodextrins having a degree of polymerization from one to fifteen (Figure 2.2). The fluorophore enables the derivatives to be detected with femtomole sensitivity using absorbance detection.[84] The chromatographic analysis of saccharides other than Glucα1→4Glu was also undertaken with this separation. The results for various saccharides are presented in graph 2.1 and table 2.1. The first obvious result from these experiments is that the retention times overlap for the mono and disaccharides while, individually, the saccharides maintain a small deviation in retention time (graph 2.1, error bars and the relative standard deviations from table 2.1). For the purpose of this study, it was found that the monosaccharide retention times overlap with the disaccharides, and their

configura

deoxy-m

typically

to be exp

the colur

the HPL

piece of i

plays onl

Hence, t

polymeri

seen wh

polymeri

Fo

attempt

separatio

these we

This may

content.

have bee

utilizing

oligosacc

apply to

configuration is much more critical than for the disaccharides. When a deoxy-monosaccharide is chromatographed, the retention time is longer, typically a barely resolved shoulder on the derivatizing reagent band. This is to be expected due to the increased interaction with the stationary phase of the column. Therefore retention time presents a problem for utilizing only the HPLC separation to determine the degree of polymerization. Another piece of information that is shown in this chart is that the linkage position plays only a small role in determining the retention time for disaccharides. Hence, the retention time is useful for approximating the degree of polymerization determination, but not the linkage position. It remains to be seen whether this finding applies to saccharides with larger degrees of polymerization when the PMA derivatizing reagent is used.

For lower degrees of polymerization, a separate assay was developed to attempt to improve resolution. By modifying the mobile phase, the separation between the mono through tetra saccharides was increased, and these were further separated from excess derivatizing reagent (Figure 2.3). This may allow for more of an unambiguous determination of monosaccharide content. But, other obstacles to strictly chromatographic determinations have been encountered. Hase et al.[41] and Tomiya et al.[42] suggest a method of utilizing glucose standards to set a "glucose unit" designation for oligosaccharides. Although, according to Tomiya, this does not successfully apply to saccharides containing sialic acid residues, their response to this

problem

labile sa

branched

Since it

the infor

detector

Figure
(D.P. ≤ 1

problem is to remove the sialic acid residues before analysis since these are labile saccharide units. This approach allowed them to categorize the branched saccharides correctly based upon Hase's glucose unit designation. Since it is obvious that a single method of detection could not provide all of the information required, the approach we have taken is to add a second detector that is mass sensitive.

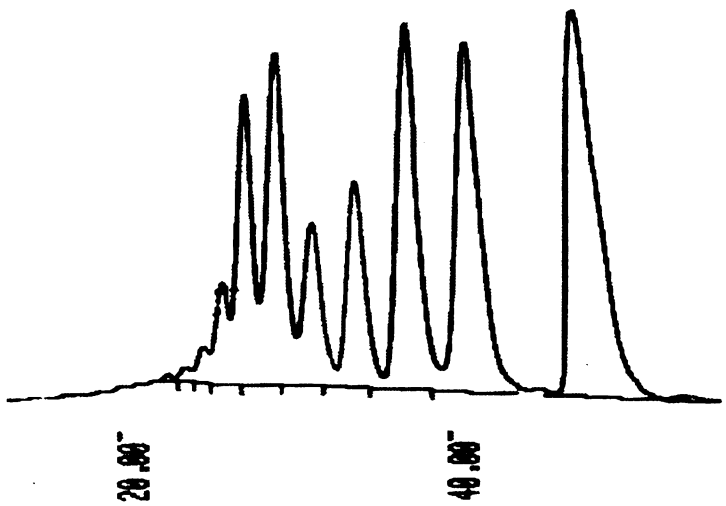
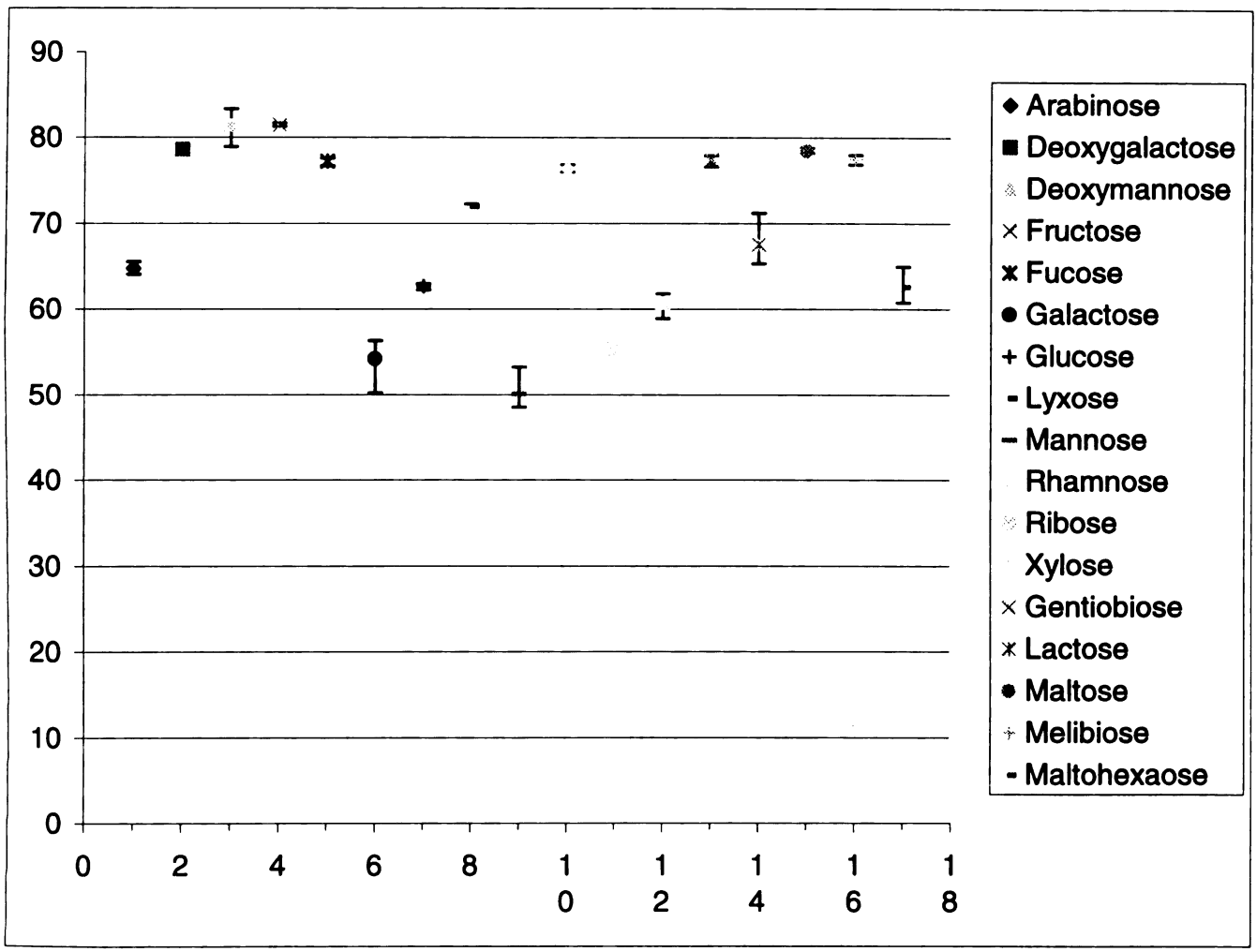


Figure 2.3: Alternative separation for maltodextrins derivatized with PMA (D.P. ≤ 15) on an ODS column using a methanol gradient.



Graph 2.1: Relationship between the average retention time over three consecutive injections on an ODS column verses the degree of polymerization of the saccharide. Error bars are included to indicate the range of retention times.

Sugar	D.P.	Mean	$S.D._{mean}$	R.S.D.
Arabinose	1	64.8	0.4	0.01
Deoxygalactose	1	78.6	0.2	0.00
Deoxymannose	1	81.4	1.3	0.02
Fructose	1	81.4	0.1	0.00
Fucose	1	77.2	0.3	0.00
Galactose	1	54.2	2.0	0.04
Gentiobiose	2	77.4	0.4	0.01
Glucose	1	62.6	0.2	0.00
Lactose	2	67.5	1.8	0.03
Lyxose	1	72.0	0.1	0.00
Maltohexaose	4	62.6	1.2	0.02
Maltose	2	78.4	0.2	0.00
Mannose	1	50.1	1.6	0.03
Rhamnose	1	76.3	0.3	0.00
Ribose	1	55.4	0.3	0.00
Xylose	1	60.7	0.9	0.02
Melibiose	2	77.5	0.3	0.00
Gentiobiose	2	80.5	0.6	0.01

Table 2.1: Analysis of the retention times of saccharides of varying degrees of polymerization.

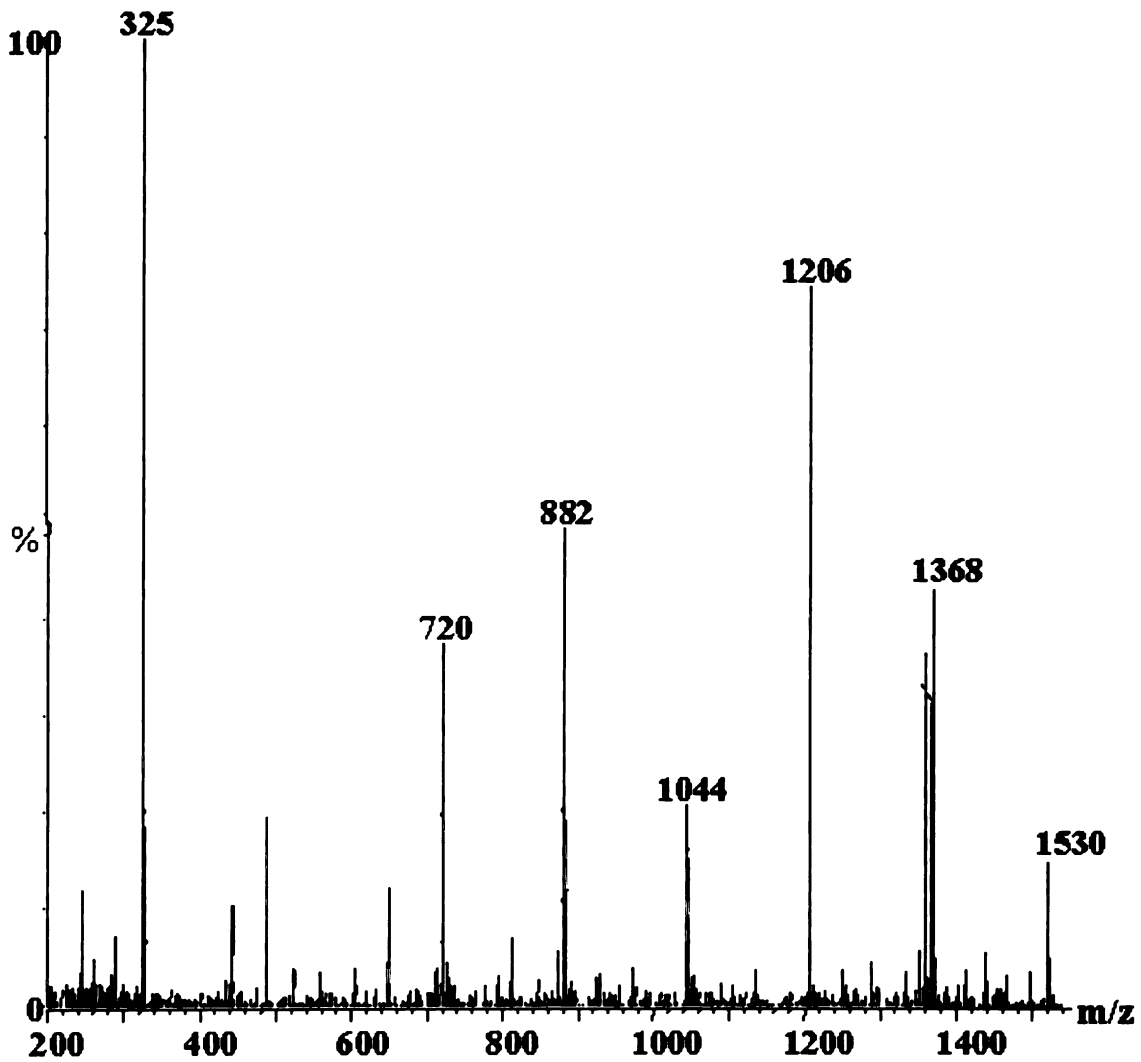


Figure 2.4: Electrospray mass spectrum of pyrenemethylamine derivatized maltodextrin (D.P. ≤ 15).

By utilizing mass spectrometry to obtain mass information from the bands obtained from the HPLC separation the degree of polymerization can be determined with pmol sensitivity up to a degree of polymerization of ten. Figure 2.4 is two hundred pmol of the derivatized maltodextrin injected, shown are $DP_N \leq 8$. Sensitivity for this stage of structure determination is in

the tens of picomole range (e.g., figure 2.5 for the derivatized disaccharide maltose). By obtaining confirming results by mass spectrometry, not only can the degree of polymerization be confirmed, but also any modifications to the saccharide(s) can be detected due to changes in mass. Therefore, even if, for example, a deoxy-sugar or acetylated sugar were present, the mass of the resulting ion would indicate its presence. But, this will not give us all of the information that we seek. Information on branching, linkage positions, or where modifications to the sugars occur is not present, only indications that they exist. It is therefore necessary to fragment the saccharides to obtain ions that indicate this information.

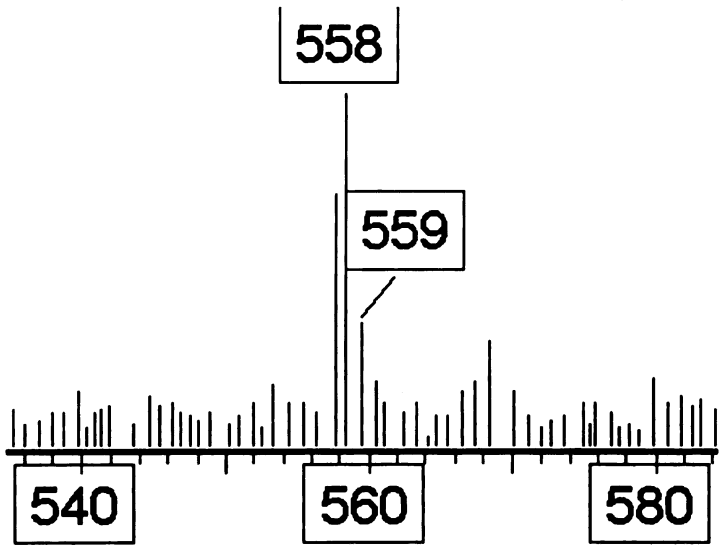


Figure 2.5: Two pmol μl^{-1} of disaccharide directly injected into the mass spectrometer. Total volume infused, 10 µl.

Utilizing a "soft" ionization method typically yields only molecular weight information. In order to obtain structural information, fragmentation of the ions is required and collisionally induced dissociation (CID) is used. This technique, for low energy collisions, typically is performed in the center quadrupole of a triple quadrupole mass spectrometer where a collision gas has been introduced. However, for the purposes of these experiments, a single quadrupole mass spectrometer was used. In-source fragmentation, a.k.a. cone fragmentation, relies on imparting additional energy to the ionized molecules to fragment them prior to any mass analysis. There are two ways in which this can be done. Conventional in-source fragmentation relies on manipulating the potential between the extraction cone and skimmer cone (Figure 2.6). Several groups have observed this.[85-87] Additionally, in-source fragmentation can be obtained, with similar results, by increasing the flow rate of the nebulization gas.

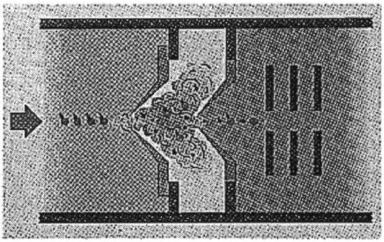
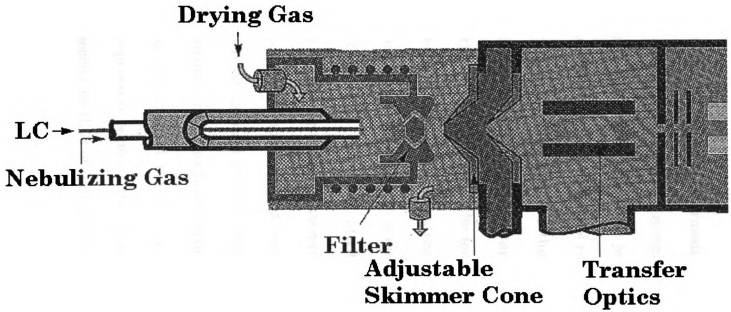


Figure 2.6: Schematic of the intermediate pressure region in an electrospray mass spectrometer .

Figure 2.7: Configuration of Fisons Platform electrospray mass spectrometer.


Drying Gas
LC
Nebulizing Gas
Filter
Adjustable
Skimmer Cone
Transfer
Optics

By adjusting the potential between the extraction cone and the skimmer cone in the region of intermediate vacuum (Figures 2.6 and 2.7), [88] where the gas expands through the sampling cone, or capillary, into the lower pressure region, forming a supersonic jet, fragmentation can occur. At the pressures typically encountered in this region, a shock wave is formed due to collisions with background gas, and there is considerable scattering of the expanding beam. Here, the potentials applied between the skimmer and the cone/capillary serve both to focus the ion beam through the skimmer orifice and to accelerate ions through this intermediate pressure region. Raising the cone potential has been shown to improve signal-to-noise ratios for some species in ESI which has been attributed to improved ion focusing.[89] Acceleration of ions through the intermediate-pressure region has also been found to cause collision-induced dissociation by increasing the energy of collisions with background gas, by imparting excess energy to the ions to induce fragmentations. This phenomenon has been attributed to the effects of collisional activation of the ions as they are accelerated through the intermediate pressure region of the ion source .

The second method of inducing fragmentation involves increasing the flow rate of the nebulization gas. In the ion-spray configuration, an additional flow of gas concurrent with the mobile phase flow is present (Figure 2.7). For the configuration of the Fisons Platform mass spectrometer,

the flow rate for obtaining pseudomolecular ions is 18 l hr^{-1}. Increasing this flow increases the rate of evaporation of the droplets and also increases the fragmentation present in the resulting mass spectrum. The data obtained are similar to those for conventional in source fragmentation (Figure 2.8 (maltose)). Since the "curtain" gas flow rate is more than ten times higher than the flow rate of the nebulization gas, the fragmentations are likely due to the increased evaporation of the droplets. This allows for more collisions prior to the ions passing through the extraction cone into the region of intermediate vacuum.

The fragmentations described here are sufficient to obtain linkage position as well as to indicate where modifications are present in the saccharide. From observing the differences in fragmentation as the potential is increased, differences in linkage position become evident, although there are similarities among all of the data. In particular, as the potential is increased, the intensity of adduct ions (Na^+ and K^+) also increases (Figures 2.8 - 2.16). These ions offer no increased structural information, they only decrease the intensity of the ions that do contain structural information. A potential needs to be chosen that is a trade off between the intensity of adduct ion(s) and the intensity of the fragmentation ions. This can provide a generalized procedure for sample analysis, even though increased sensitivity for the fragment ions could be obtained by optimizing the potential for each saccharide present. Optimization assumes prior knowledge of the sample

composition, which in instances of unknowns is not present. The potential that was chosen was 55 V.

Once the amount of energy required to induce fragmentations was investigated, another series of experiments was carried out that focused on disaccharides with varying linkage positions and configurations, and monosaccharides with varying configurations. This was pursued in order to determine differences in the fragmentation that would indicate the presence, or absence, of specific linkage configurations (Figures 2.9 – 2.16). In all of the fragmentations in this data set, the linkage position of the two saccharide systems was observed by a unique fragment ion. In the case of gentiobiose (Gluβ1→6Glu), the linkage position was indicated by a cleavage on the derivatized saccharide between carbons five and six, with the charge being retained on the derivative end of the molecule. In addition, a fragment ion is observed corresponding to $^{3,5}X_1$ (two bonds being broken within one ring) with the charge again being retained on the derivative side of the molecule. The fragment ions obtained for lactose (Galβ1→4Glu) (Figure 2.11 and 2.12) and maltose (Gluβ1→4Glu) are nearly identical, just as is the situation for EIMS.[90] Therefore, strictly utilizing any one form of mass spectrometry for the purpose of monosaccharide determination is likely impossible. The linkage position is well indicated with a characteristic cleavage, an ion indicating that the bond between carbons three and four was broken. In this instance the species that imparts the charge is sodium (Na^+), and the charge

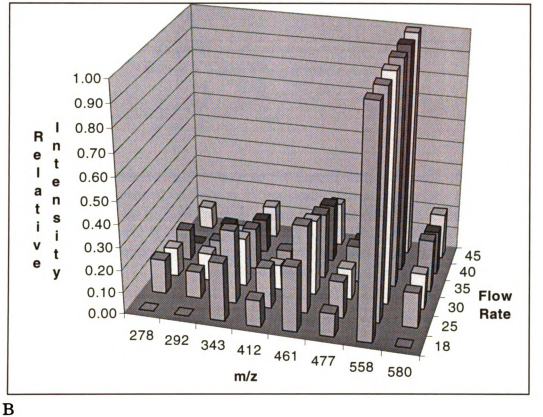
is retained on the saccharide portion of the SFD (m/z 278). As with gentiobiose (Gluβ1→6Glu), a ring fragmentation is also present yielding a fragment ion, again with Na^+ imparting the charge. Only, in this instance, the derivative part of the molecule retains the charge, and yields a $^{1,5}X_1$ fragment. The similarities expected between the spectra of gentiobiose and melibiose (Galα1→6Glu) are obvious upon inspection of their structures (Figures 9 and 10 and 13 and 14). Both disaccharides have a one to six linkage: gentiobiose has a beta linkage with glucose as both monomers, while melibiose has an alpha linkage with galactose linked to glucose. Both structures fragment between carbons five and six, leaving the charge on the derivatized saccharide, and both have a ring cleavage at $^{3,5}X_1$. The expanded region in figure 14 C gives an indication of the only difference. An additional bond is broken in melibiose, Z_1. This bond is more labile, possibly due to a decreased pi overlap in the glycosidic bond, from the differences in linkage type (i.e., alpha vs. beta linkage).

The final disaccharide observed was sophorose. This disaccharide consists of two glucose monomers with a beta linkage at the two position (i.e., Gluβ1→2Glu). In this instance, the bonds are very labile and, with increased potential applied to the extraction cone, the amount of secondary fragmentation increases. These additional fragmentations only lead to more difficult interpretation, and do not readily give more structural information. From the spectrum with the potential on the extraction cone at forty-five

volts, the relevant structural information is obtained. Again the fragmentation of interest occurs near the glycosidic linkage, and gives an ion characteristic of this position, m/z 246. Also a fragment can be seen in this spectrum involving the saccharide ring, which is a $^{1,5}X_1$ fragment.

Figure 2.8: Fragmentation of PMA derivatized maltose. The first set of data utilizes increased nebulization gas flow rate to fragment (Figure A). The second set of data utilize conventional in source fragmentation, increased potential on the extraction cone, to fragment derivatized maltose (Figure B).

A


Relative Intensity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
278 292 343 412 461 477 558 580
m/z
45
40
35
30
25
18
Flow Rate

B

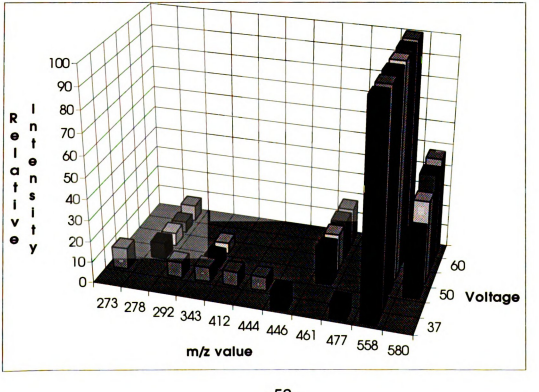
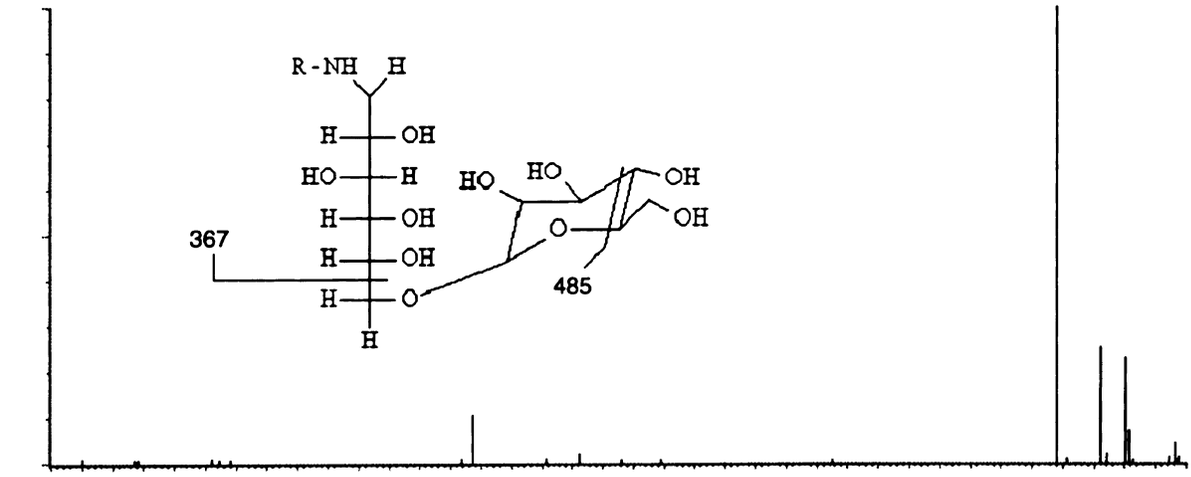
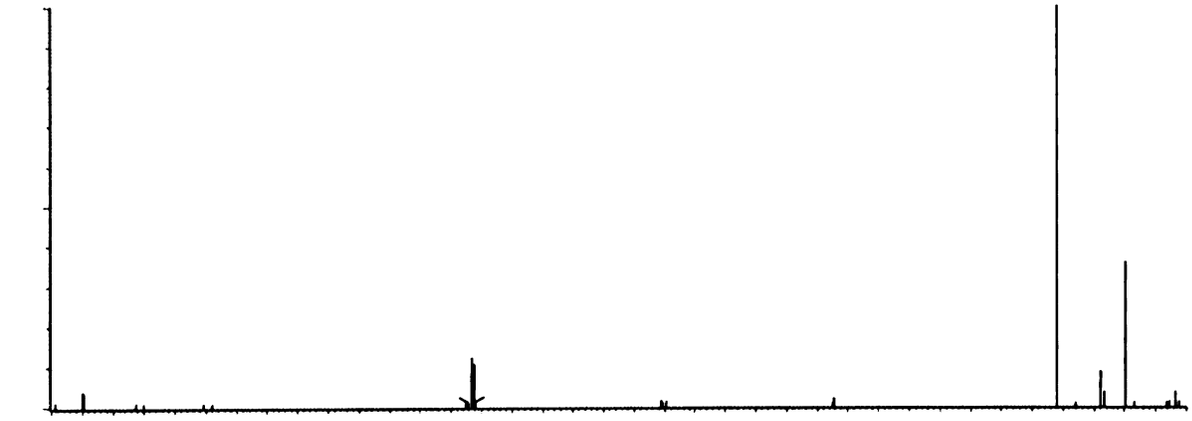

Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
273 278 292 343 412 444 446 461 477 558 580
m/z value
60
50
37
Voltage

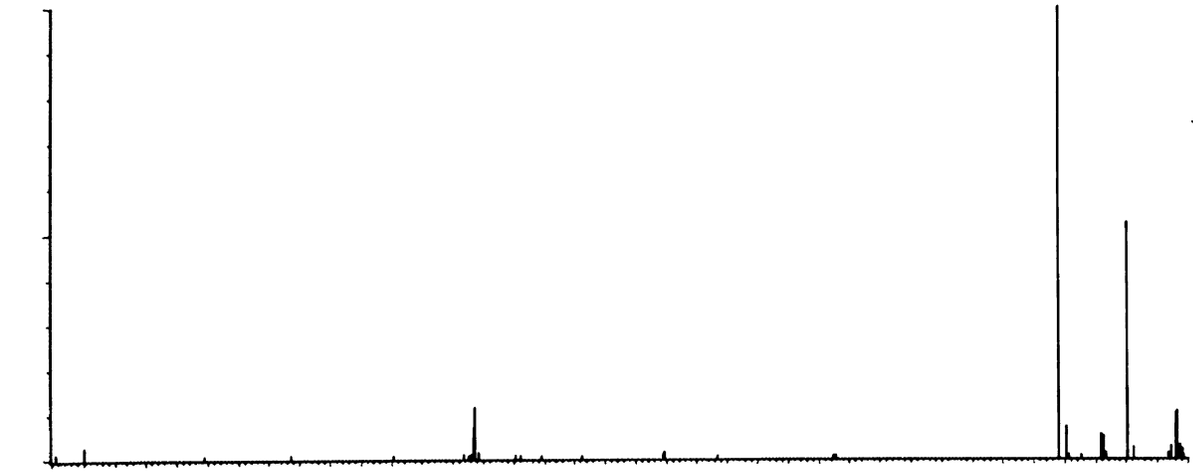
Figure 2.9: Spectra from experiments in which gentiobiose, derivatized with pyrenemethylamine, is subjected to increasing voltage applied to the extraction cone. Spectrum A is table with the normal operating voltage of 37 V. In spectrum B the voltage applied to the cone is increased to 45 V, while in spectrum C the voltage applied is further increased to 50 V.



A

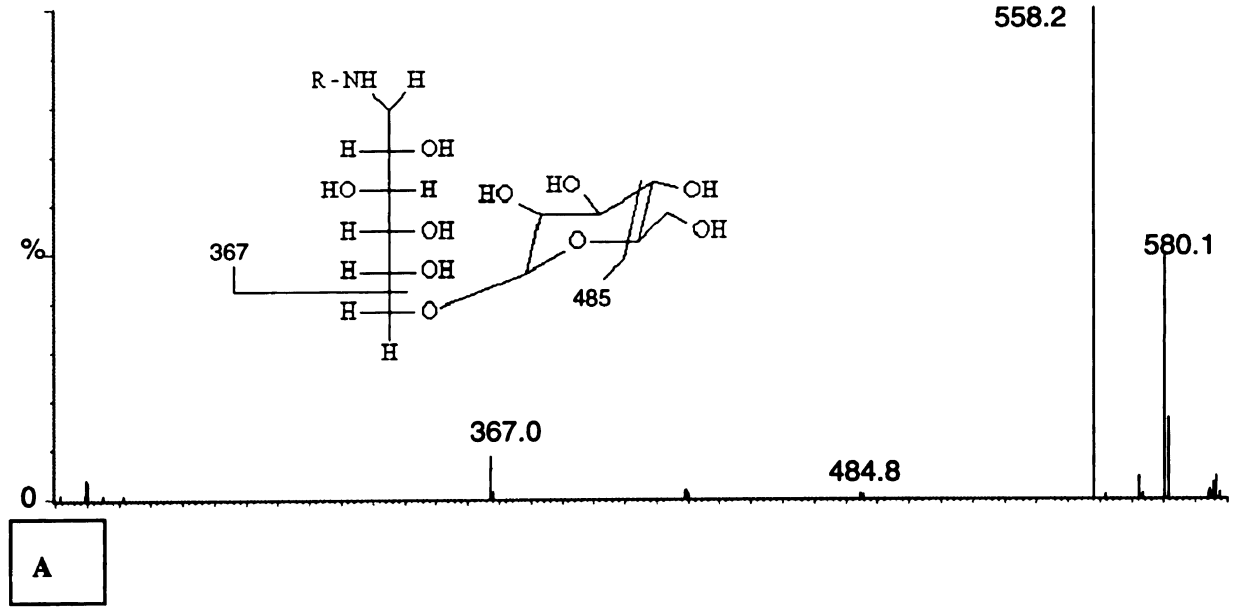


B

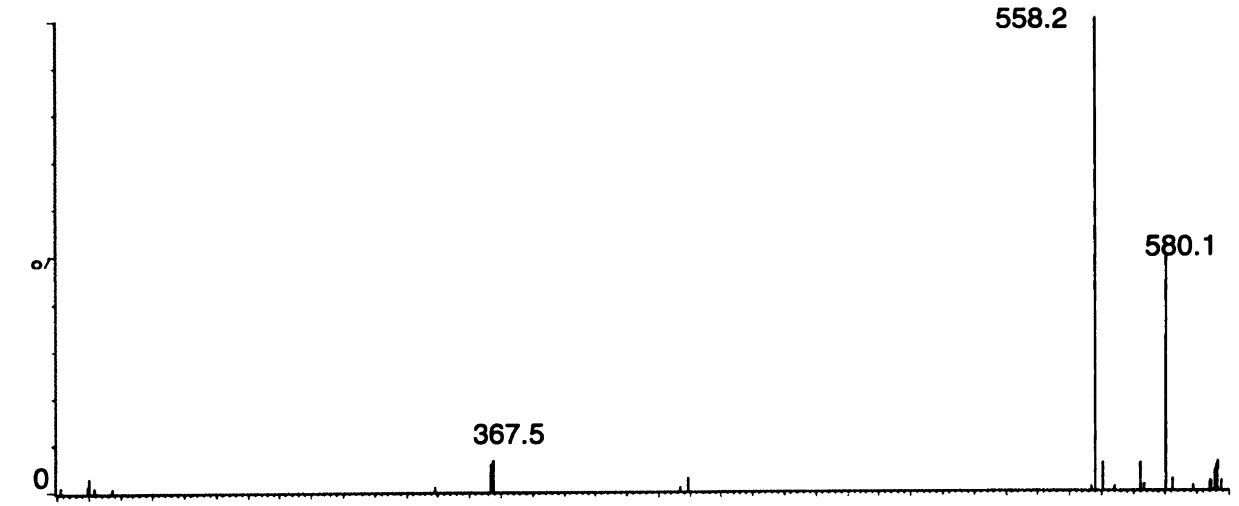


C

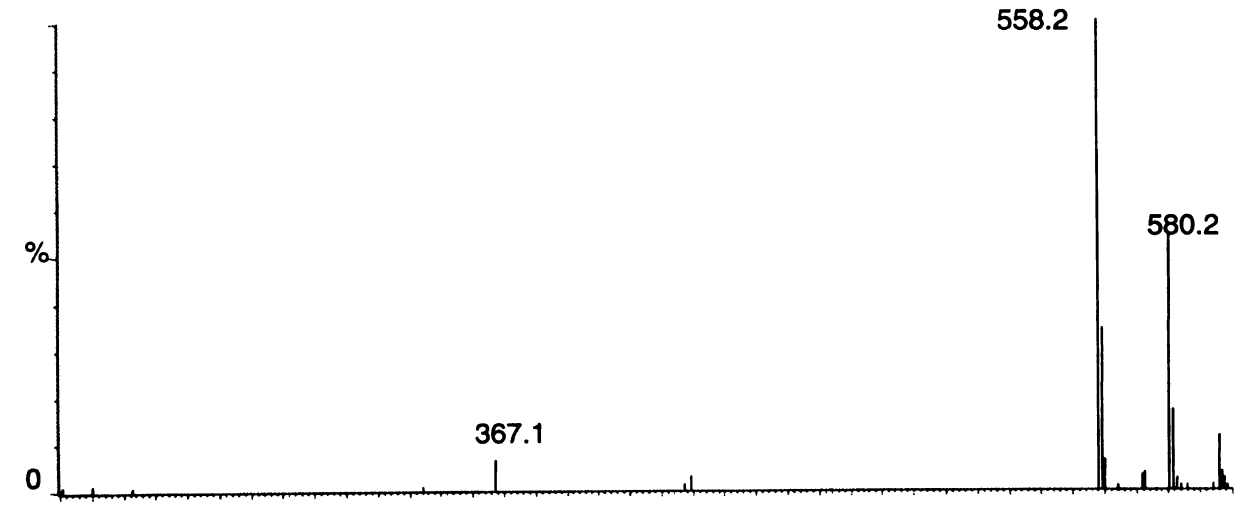
Figure 2.10: Spectra from experiments in which gentiobiose derivatized with pyrenemethylamine is subjected to increasing voltages applied to the extraction cone. Spectrum A has the operating voltage at 55 V. Spectrum B increases the voltage applied to the cone to 60 V, and spectrum C has the voltage applied increased to 65 V.


558.2
R-NH H
H OH
HO H
H OH
H OH
H O
H
HO HO OH
O OH
485
367
%
580.1
367.0
484.8
0

A

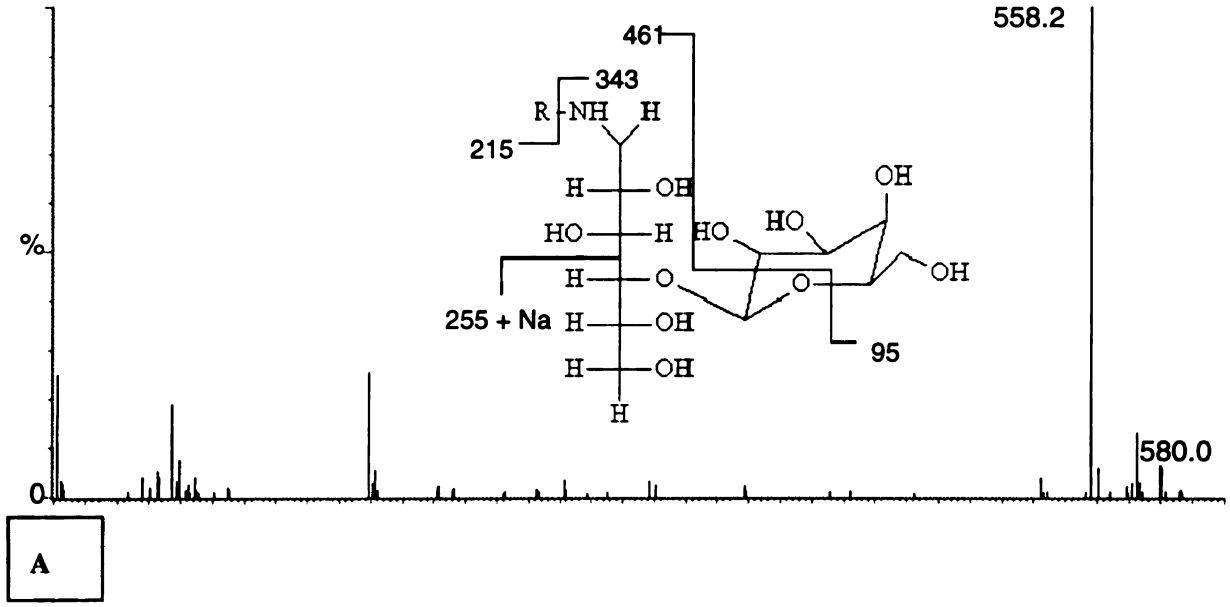

558.2
%
580.1
367.5
0

B

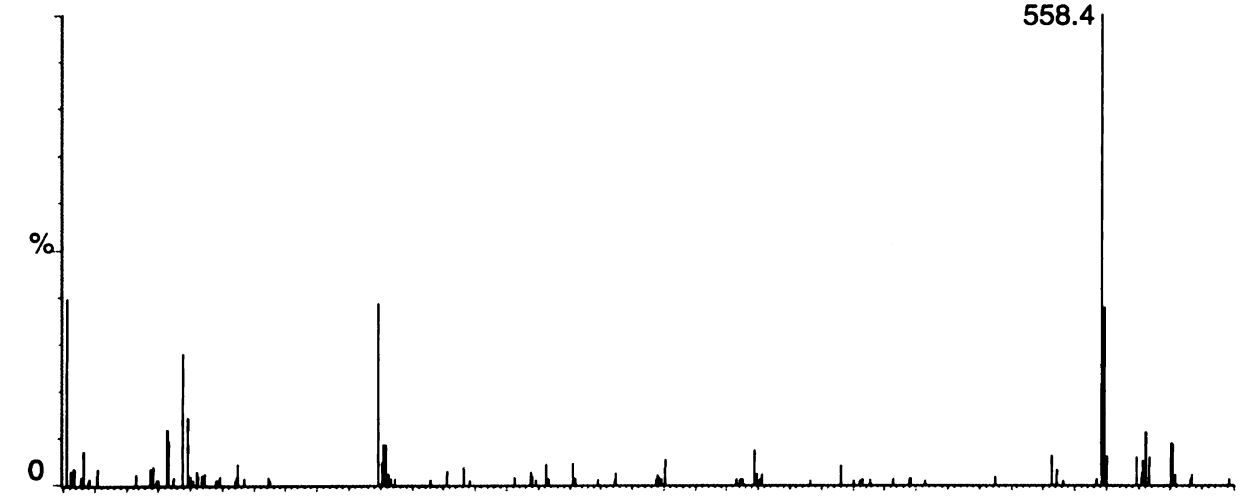

558.2
%
580.2
367.1
0

C

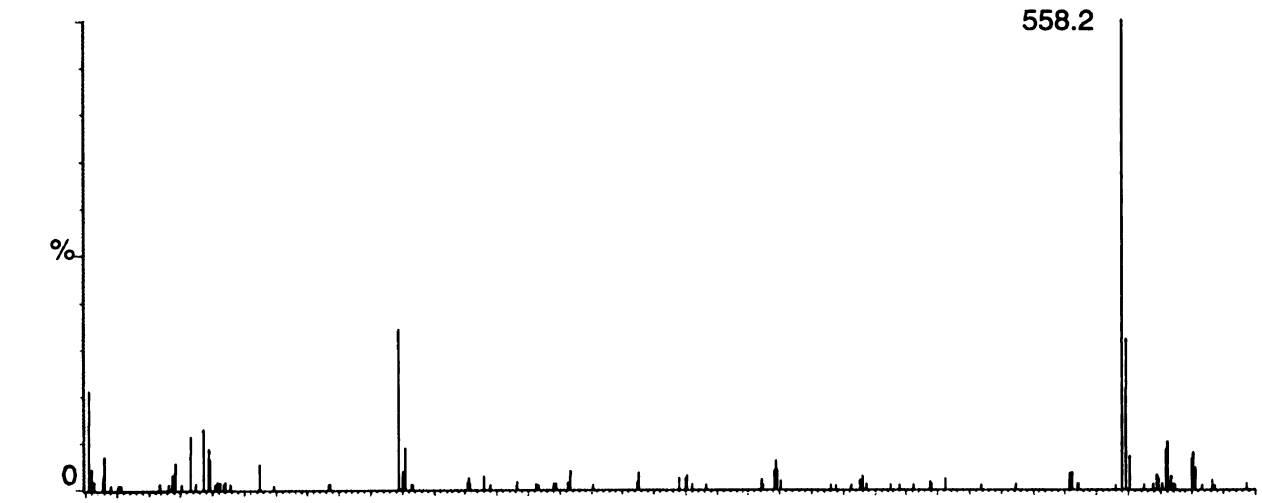
Figure 2.11: Spectra depicting experimental results where lactose, derivatized with pyrenemethylamine, is subject to increasing potentials applied to the extraction cone. The results shown here are for the normal operating voltage of 37 V, 45 V, and 50 V applied to the cone for spectra A, B, and C, respectively.


558.2
461
343
R-NH H
215
H OH
HO H
H O
255 + Na H OH
H OH
H
HO
HO
OH
OH
O
95
%
0
580.0

A


558.4
%
0

B


558.2
%
0

C

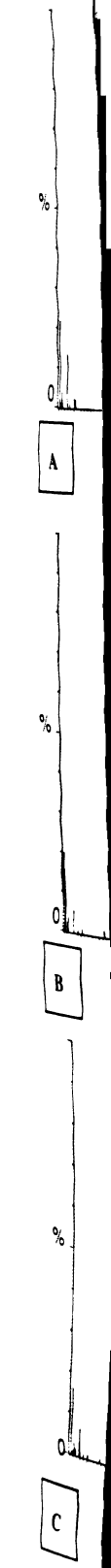
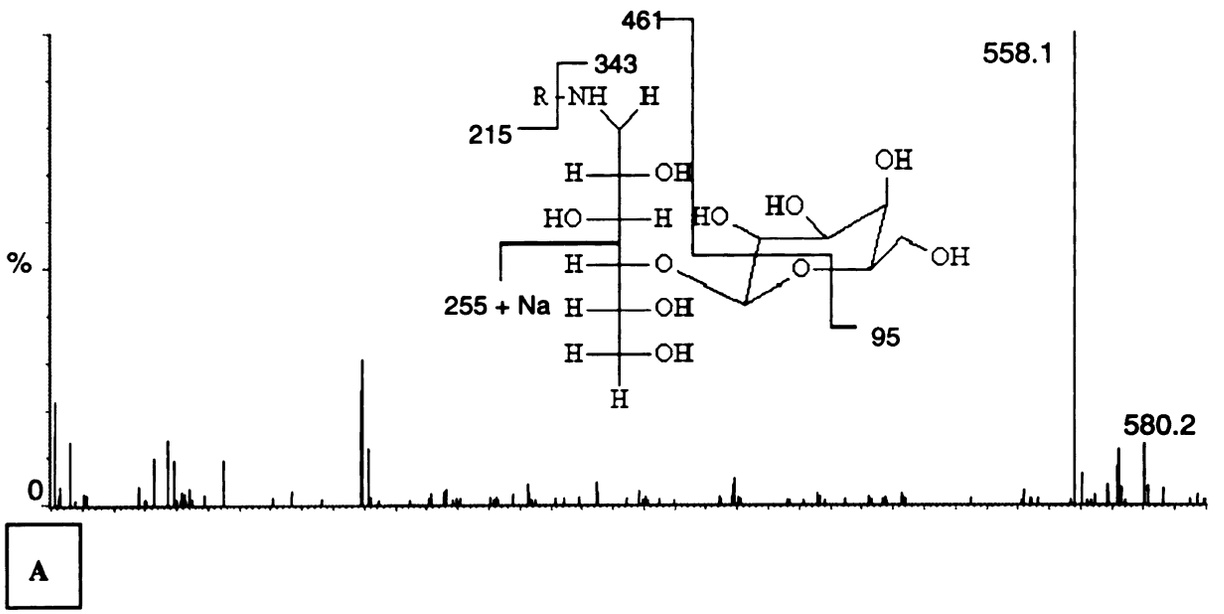
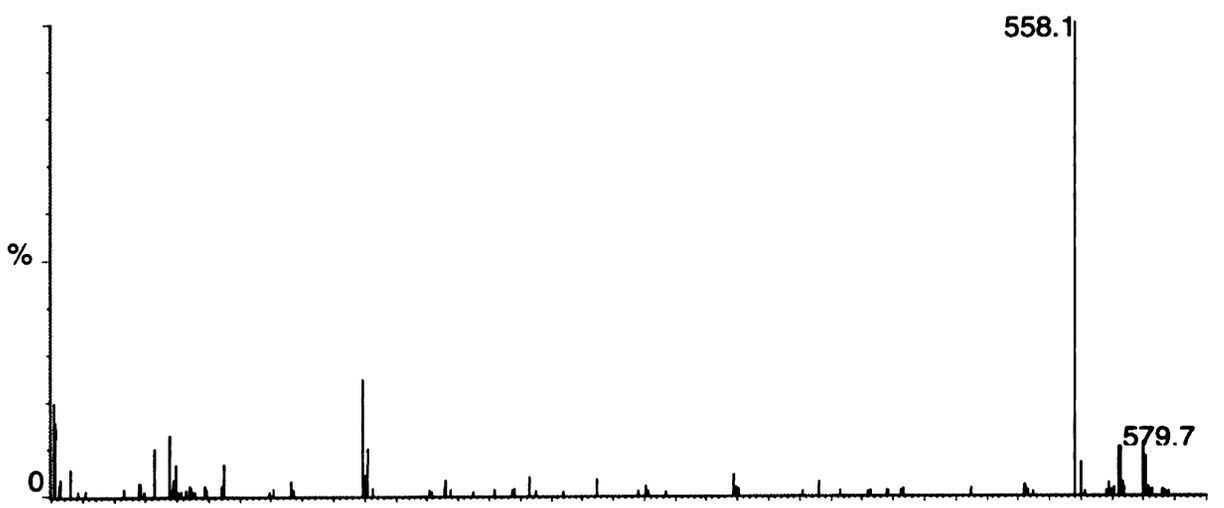


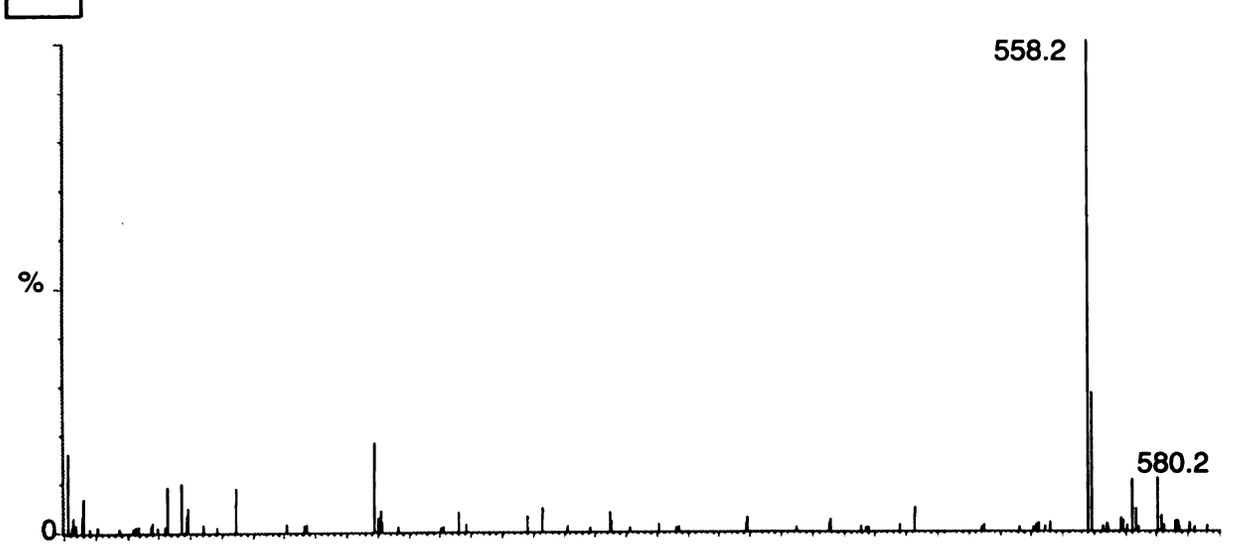
Figure 2.12: Results from experiments in which increasing voltages are applied to the cone. The saccharide derivatized with PMA for these spectra is lactose. Spectrum A has the operating voltage set to 55 V. Spectrum B increases the voltage applied to the cone to 60 V, and spectrum C has the voltage applied increased to 65 V.


461
343
R-NH H
215
H OH
HO H
H O
255 + Na H OH
H OH
H
HO
HO
OH
OH
O
95
558.1
580.2
%
0

A


558.1
579.7
%
0

B


558.2
580.2
%
0

C

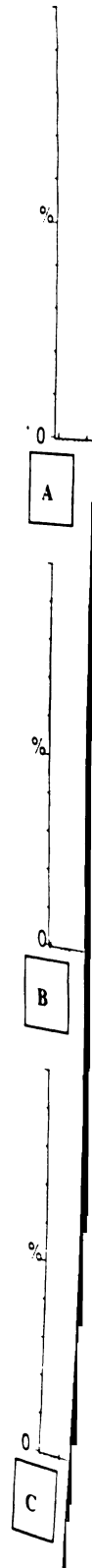
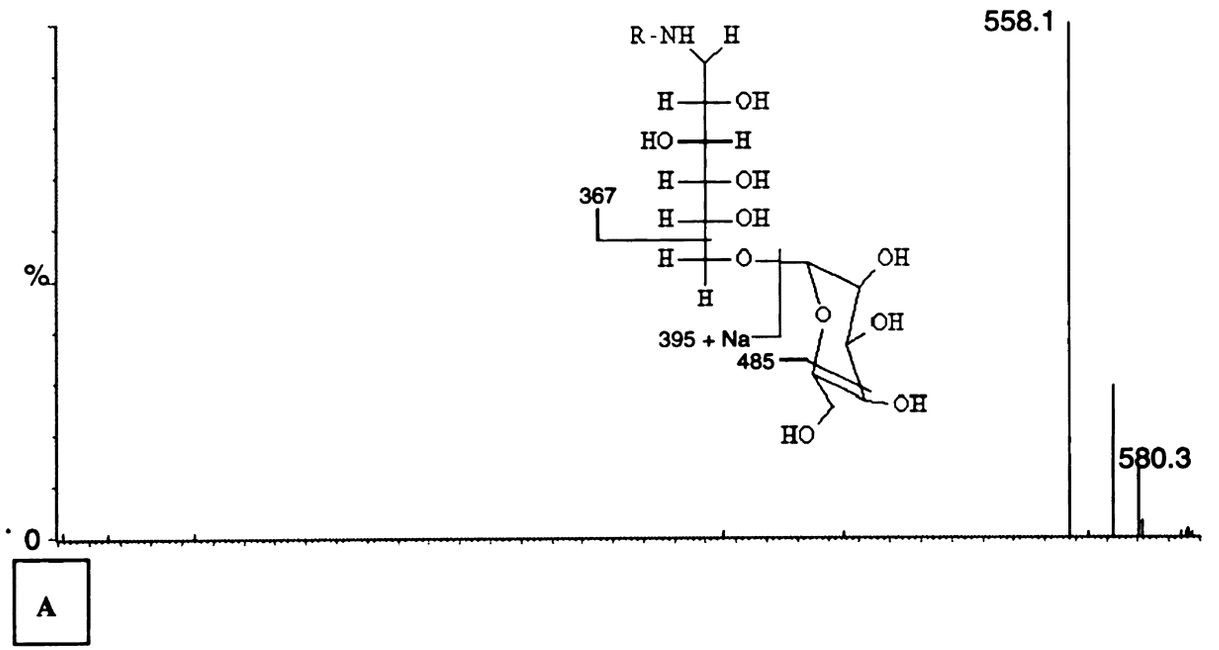
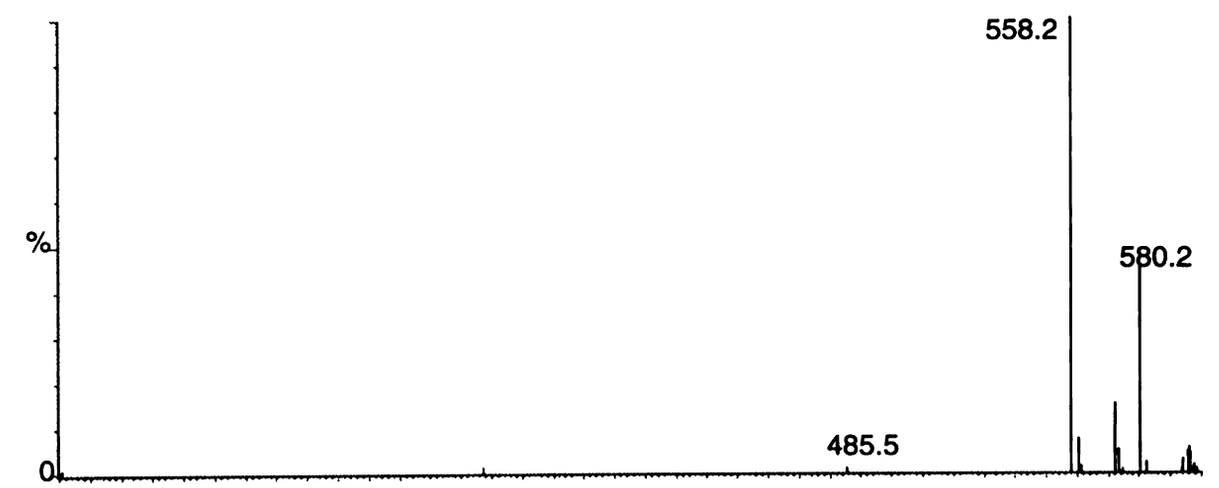


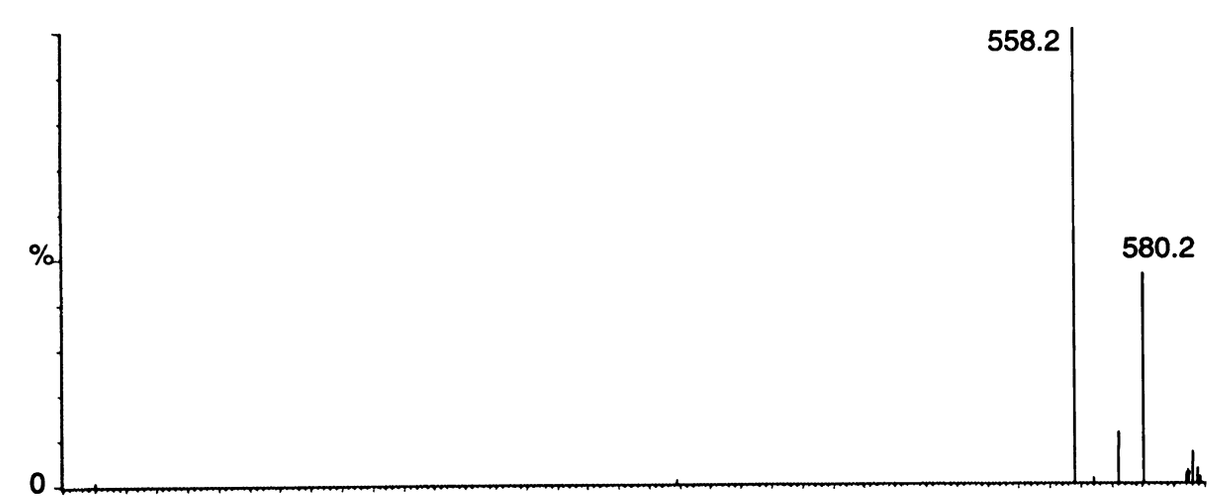
Figure 2.13: Spectrum A, B, and C show the experimental results where melibiose, derivatized with pyrenemethylamine, is subjected to increasing voltages applied to the extraction cone. The voltages applied are 37, 45, and 50 V, respectively.


558.1
R-NH H
H OH
HO H
H OH
H OH
H O
H
367
395 + Na
485
OH
OH
O
OH
HO
580.3
%
0

A

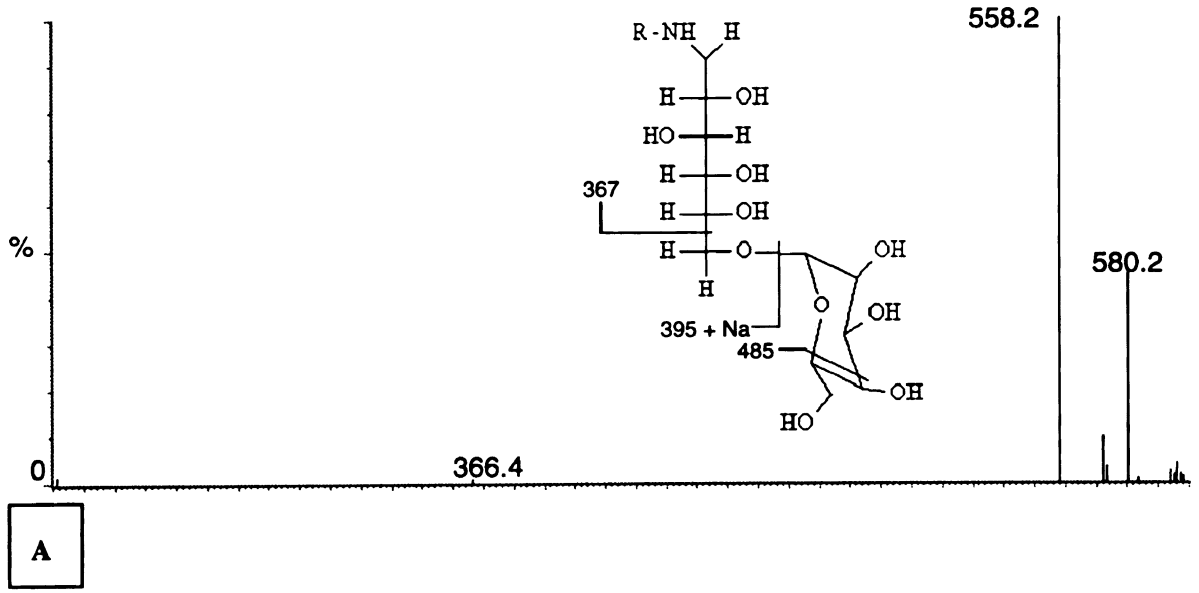

558.2
580.2
485.5
%
0

B

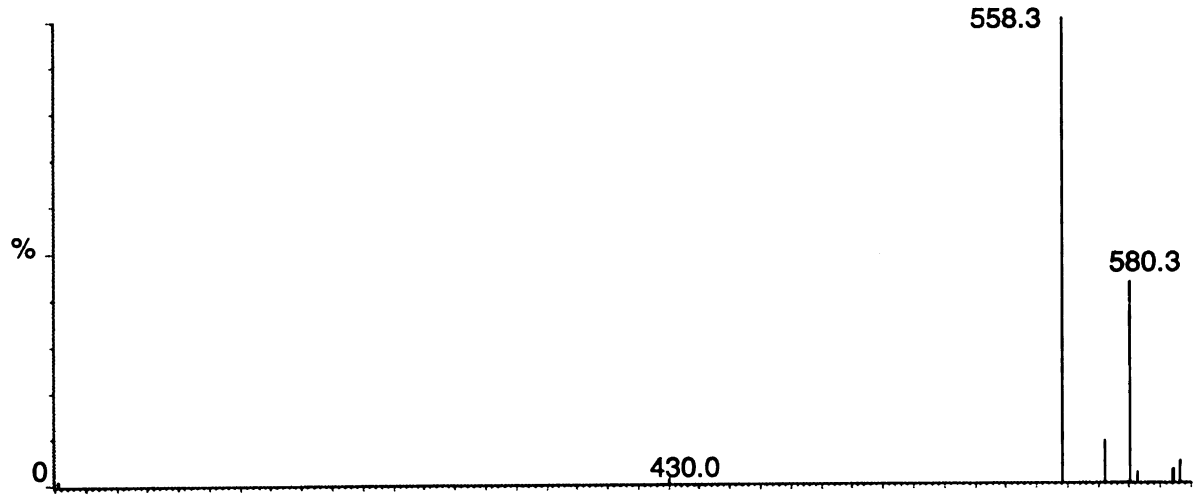

558.2
580.2
%
0

C

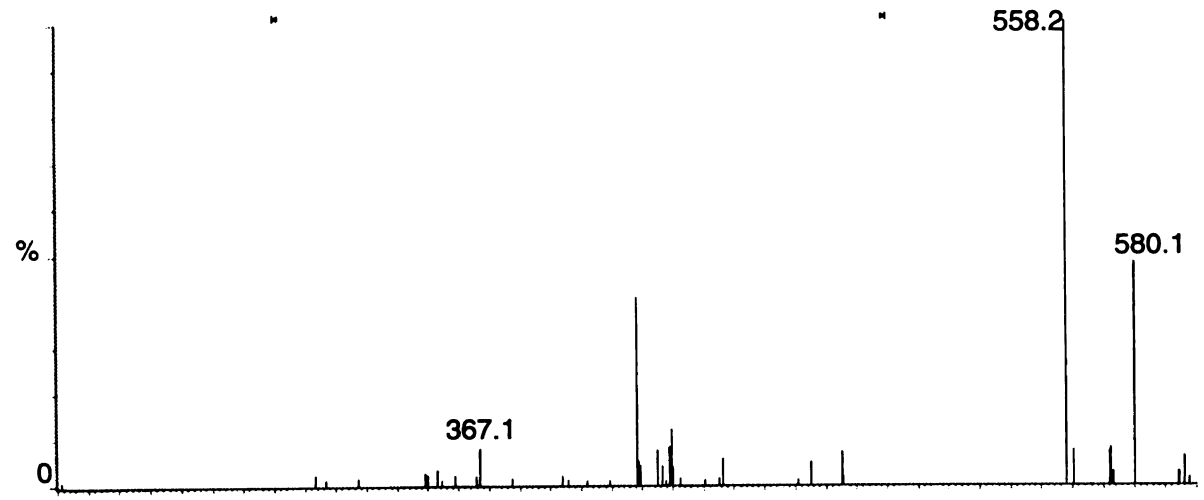
Figure 2.14: Spectra from experiments in which melibiose, derivatized with pyrenemethylamine, is subjected to increasing potentials applied to the extraction cone. Spectrum A has the operating potential set to 55 V. Spectrum B increases the voltage applied to the cone to 60 V, and spectrum C has the potential applied increased to 65 V.


R-NH H
H OH
HO H
H OH
H OH
H O
H
OH
OH
OH
HO
O
367
395 + Na
485
558.2
580.2
%
0
366.4

A

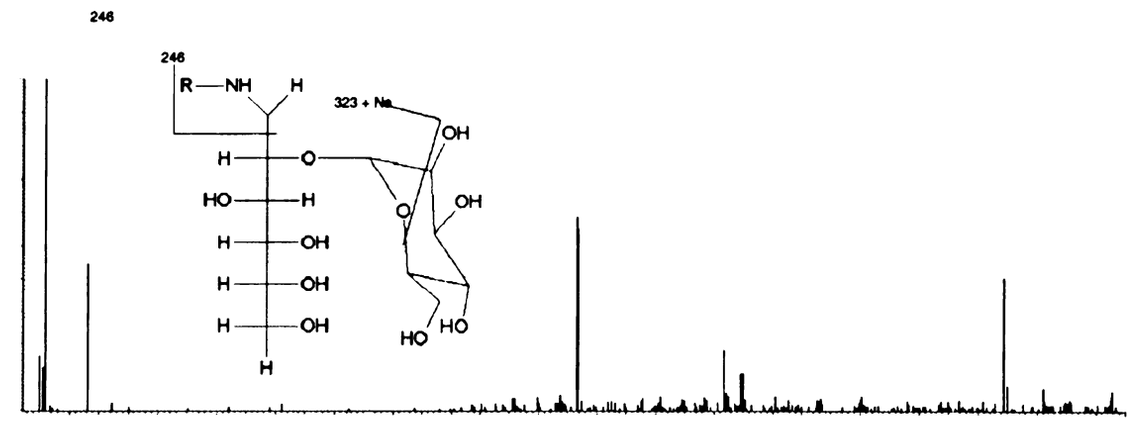

558.3
580.3
%
0
430.0

B

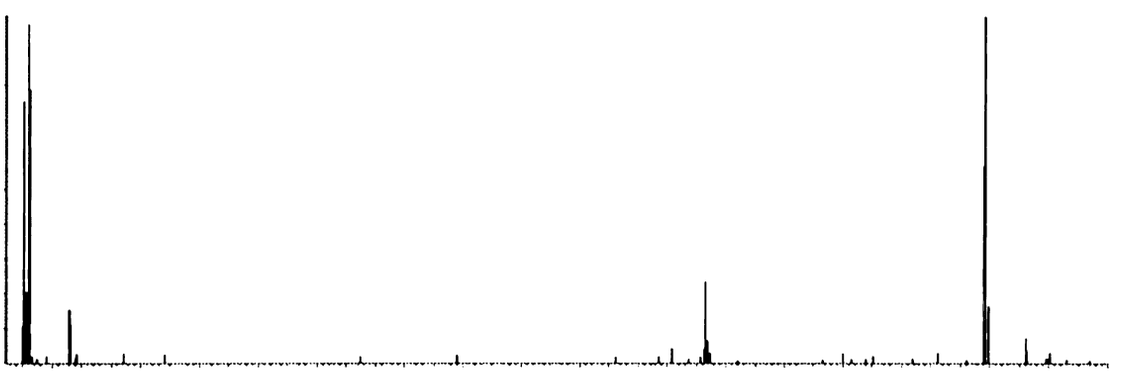

558.2
580.1
%
0
367.1

C

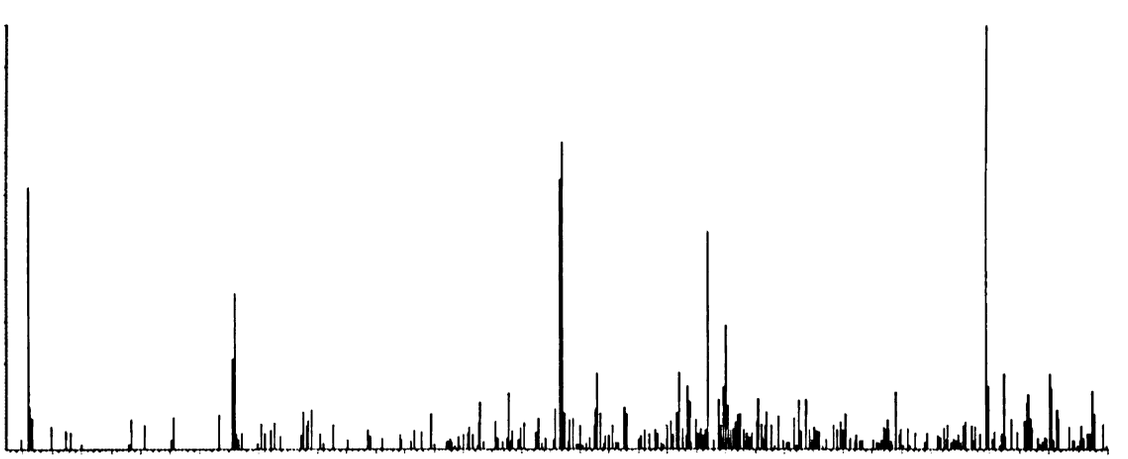
Figure 2.15: Spectra from experiments in which sophorose, derivatized with pyrenemethylamine, is subjected to increasing voltage applied to the extraction cone. Spectrum A is table with the normal operating voltage of 37 V. In spectrum B the voltage applied to the cone is increased to 45 V, while in spectrum C the voltage applied is further increased to 50 V.



A

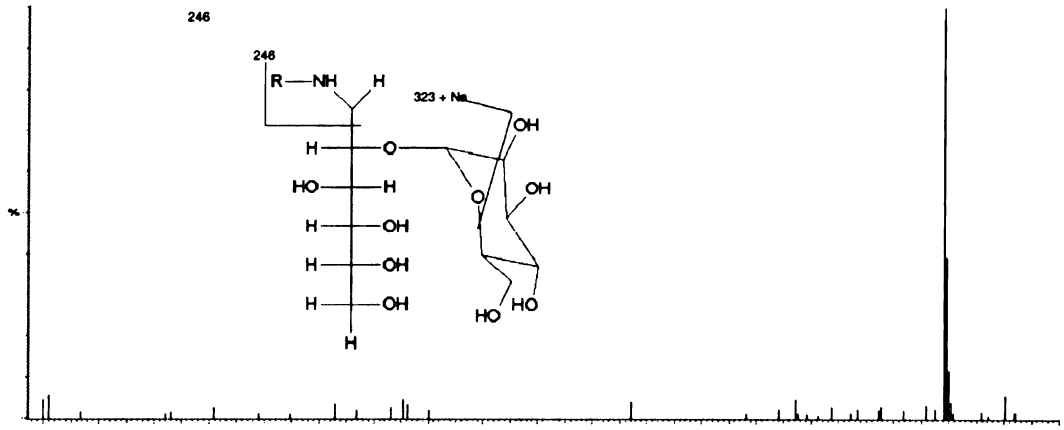


B

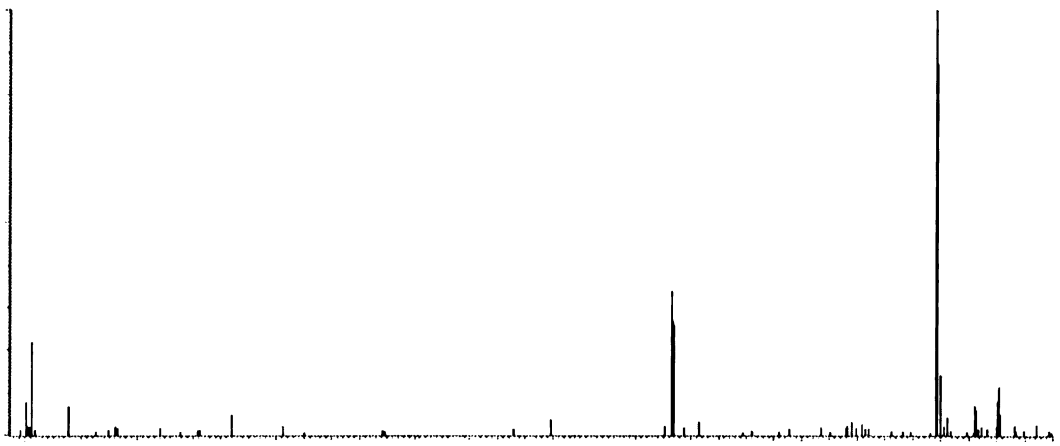


C

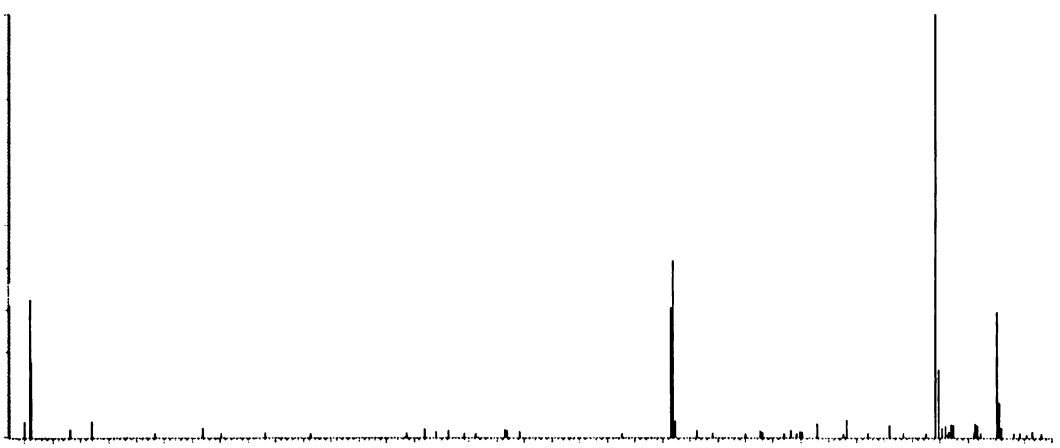
Figure 2.16: Spectra depicting experimental results where sophorose, derivatized with pyrenemethylamine, is subjected to increasing voltages applied to the extraction cone. Spectrum A has the operating potential set to 55 V. Spectrum B increases the voltage applied to the cone to 60 V, and spectrum C has the potential applied increased to 65 V.



A



B



C

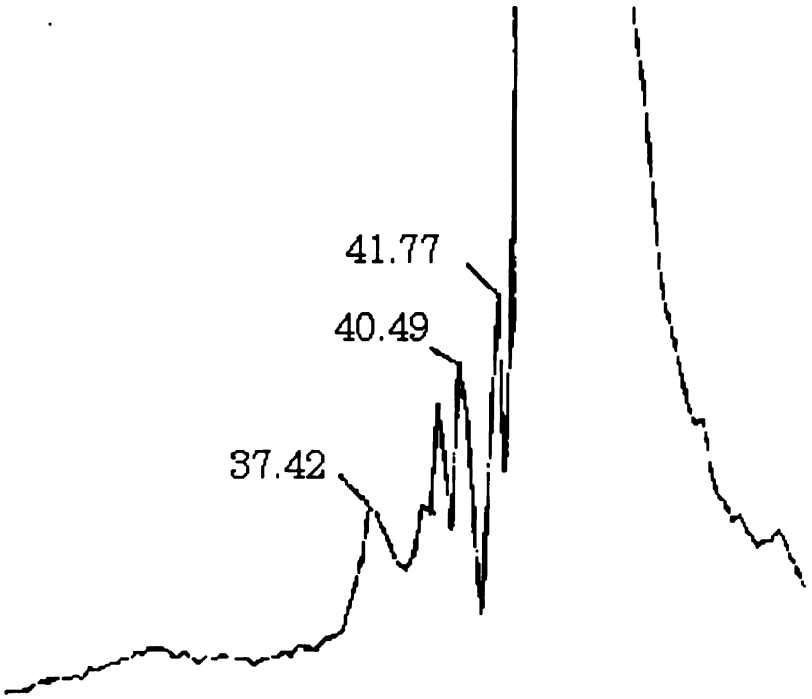


Figure 2.17: Separation of maltodextrin (n ≤ 15) by microbore HPLC. There is overlap of the various degrees of polymerization. The first eluting band at 37.42 minutes contains degrees of polymerization from ten to eight, the second eluting band at 40 minutes contains degrees of polymerization from nine to seven.

Conclusion: This assay has many strong points, but there is always room for improvement. The separation of the maltodextrin (n ≤ 15) is baseline, and gives a good indication of the degree of polymerization. The absorbance properties of the PMA allow for detection of the saccharides on the femtomole level, and the lower limit of detection can be improved further by using fluorescence detection for the HPLC separation in the first stage. Even though for monosaccharides the retention appears to be very dependent on the saccharide, this in itself can be used as a diagnostic tool to obtain more information from the separation. Since, the degree of polymerization is confirmed using MS, the retention time is also a diagnostic tool.

Unfortu

2.5, the

through

m/z ≥ 1

the ser

assay,

using

microb

of frac

initial

the re

conver

suffici

(Figur

first e

ten t

polym

overla

ions i

resul

num

Unfortunately, the MS limits the sensitivity of the assay. As shown in figure 2.5, the detection limit is two pmol μl^{-1}, this appies to the disaccharide through the decasaccharide. This instrument's sensitivity begins to fall off at $m/z \geq 1000$ and declines even more rapidly beyond m/z 1500. In this instance, the sensitivity is only limited by technology. This does not invalidate the assay, or the techniques.

Improvements to the assay are needed. The separation was developed using a conventional 25 cm x 4 mm ODS HPLC column. By utilizing microbore HPLC (μ-bore HPLC), the assay can be improved, and the problem of fraction collection, with the concomitant sample loss, is removed. The initial steps toward a separation have already begun (Figure 2.17). Although the resolution is poor, for the same degrees of polymerization as the conventional separation, many of the oligosaccharides are separated, and it is sufficient to indicate another instrumental shortcoming. The chromatogram (Figure 2.17) is the reconstructed ion chromatogram (RIC) from the MS. The first eluting band at 37.42 minutes contains degrees of polymerization from ten to eight, the second eluting band at 40 minutes contains degrees of polymerization from nine to seven. This could be an indication that there is overlap between the two bands, but the intensities of the pseudomolecular ions increase for all of the ions in the second eluting band. Included with this result are the following: first, that there are apparent differences in the number average degree of polymerization obtained from the chromatographic

separa

mixtur

eleven

picture

throug

these

with

groups

higher

increa

Altho

(n ≤ 1

larger

separations (Figures 2.3 and 2.4) and the mass spectrum of the unseparated mixture (Figure 2.4). Second, that the higher degrees of polymerization, eleven through fifteen, are not detected. With this information a more clear picture is obtained. All indications are that the linear polymer, upon passing through the extraction cone, is being sheared into smaller pieces and that these fragments are being detected. This does not appear to be associated with the skimmer cone potential since it has been observed by other groups[94,102,105] that increasing the potential on the extraction cone can reveal higher molecular weight species. But what has been shown here, is that increasing the cone potential on oligosaccharides increases fragmentation. Although this is a benefit for structure determination of smaller saccharides ($n \leq 10$), it is not beneficial for the detection, or structure determination, of larger saccharides.

PR

glycop

Hence

(e.g., a

of glyc

to an

isolat

struct

of spe

for s

struct

varia

postu

size

are o

facto

accu

Chapter 3

PRIMARY AND SECONDARY SEQUENCE DETERMINATION FOR PEPTIDES UTILIZING A SINGLE QUADRAPOLE MASS SPECTROMETER

Introduction: Understanding the glycobiology (structure-function) of glycoproteins requires the structural determination of the entire molecule. Hence structure of the protein as well as any post-translational modifications (e.g., any attached sugar chains, which often occur at any of the possible sites of glycosylation) must be determined. These glycoforms invariably give rise to an impressive array of structures. Individual glycopeptide forms may be isolated using modern methods of separation, although, profiling the structure in the absence of workup may best represent a natural distribution of species. Attempts at isolation and derivatization introduce opportunities for sample disproportionation, especially for samples with contrasting structural types. However, for protein glycoconjugates the assumption is that variations in the glycan do not affect the ionization of the glycopeptide, an postulation that is more plausible as the peptide size increases relative to the size of the carbohydrate.[91] Even so, higher molecular weight glycoproteins are often difficult to directly ionize by electrospray. This reflects a number of factors, such as high molecular weights leading to high-charge states and the accumulation of adducts that may challenge the resolution limits of the mass

spectrometer. For these materials, ES-MS may provide a spectrum that appears as a large unresolved "hump" (Figure 3.1) as opposed to nonglycosylated, or deglycosylated, species of similar molecular weight (Figure 3.2). The presence of adducts (e.g., multiple glycoforms) can limit the effectiveness of HPLC/MS interfacing by decreasing the number of glycopeptides that can be directly ionized. Alternatively, the peptide sequence in addition to the glycan distribution may be assessed separately after deglycosylation. Additional possibilities, in the study of large biological polymers, are offered by collisional activation of multiply charged ions created by electrospray ionization. Multiple charging allows higher molecular weight precursors to be collisionally activated and enhances fragmentation in these heavier species. To date, most reports of collisional activation of selected multiply charged peptides have involved triple quadrupole instruments and multiple low-energy collisions, typically with an inert collision gas in the second quadrupole.[92-94] Peptide precursors with mixed charge states have been fragmented by low-energy collisions in the ESI source.[81-96] High-energy collisional activation has been examined with doubly charged ions produced by fast-atom bombardment (FAB).[97-99]

Figure 3.1: Two hundred pmol μl^{-1} of fetuin characterizing the poorer sensitivity of a glycosylated peptide due to a lack of single, distinct charge series. The presence of a mixture, due to differences in glycosylation, decreases sensitivity. For these materials, ES-MS may provide a spectrum that appears as a large unresolved "hump" as opposed to nonglycosylated, or deglycosylated, species of similar molecular weight.

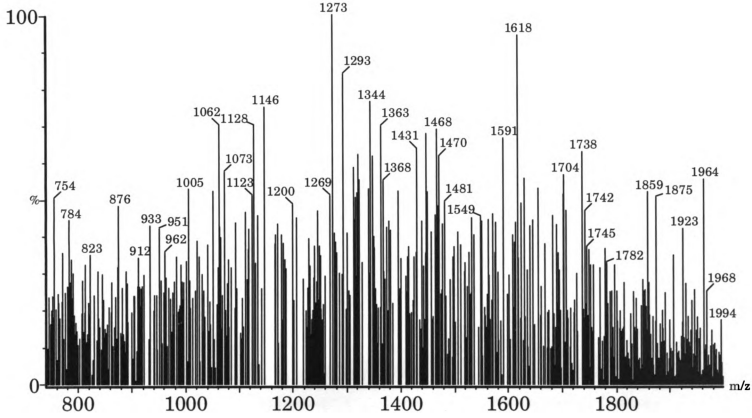
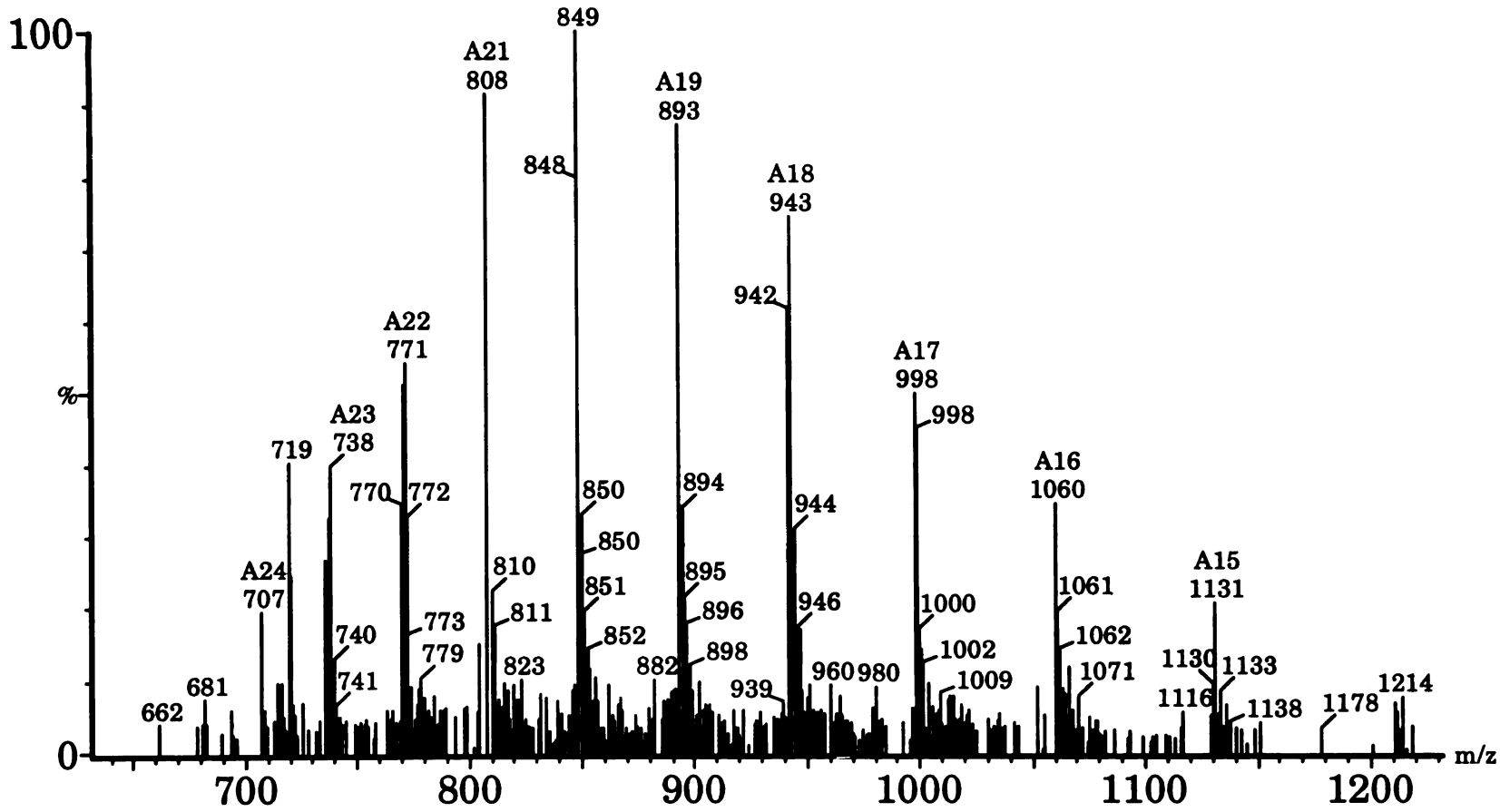

100
%
0
754
784
823
876
912
933
951
962
1005
1062
1128
1073
1123
1146
1200
1269
1273
1293
1344
1363
1368
1431
1468
1470
1481
1549
1591
1618
1704
1738
1742
1745
1782
1859
1875
1923
1964
1968
1994
800
1000
1200
1400
1600
1800
m/z

Figure 3.2: Two hundred fmol μl^{-1} of horse heart myoglobin, HHM, showing an envelope of multiple charge states.


100
%
0
700
800
900
1000
1100
1200
m/z
A21
808
849
A19
893
848
A18
943
942
A22
771
A17
998
998
A23
738
719
770
772
850
894
944
A16
1060
A24
707
850
810
895
A15
1131
851
946
1061
811
896
1000
740
773
852
1062
779
898
1002
823
882
1130
1133
741
960
980
1009
1071
662
681
939
1116
1138
1178
1214

One characteristic of ESI instrumentation is the large pressure differential that is necessary for combining atmospheric pressure ionization with the low pressures required for mass analysis and ion detection. The pressure differential is usually achieved either with a sampling cone/skimmer or a capillary/skimmer arrangement. In either configuration as the gas expands through the sampling cone or capillary into the lower pressure region it forms a supersonic jet. At the pressures typically encountered in this region, a shock wave is formed due to collisions with background gas, and there is considerable scattering of the expanding beam. The position of the skimmer in the source relative to the cone or capillary is extremely important for ensuring minimal ion losses in this region.[100] Potentials applied between the skimmer and the cone/capillary serve both to focus the ion beam through the skimmer orifice and to accelerate ions through this intermediate-pressure region. Raising the cone potential has been shown to improve signal-to-noise ratios in the ESI analysis of pesticides in water and also for the analysis of some large peptides. This has been attributed to improved ion focusing.[101] Acceleration of ions through the intermediate-pressure region has also been found to cause collision-induced dissociation by increasing the energy of collisions with background gas.[102,103]

Experimental: The in source fragmentation mass spectra were obtained on a Platform (Fisons, MA) single quadrupole mass spectrometer under control of a PC data system operating MassLynx. The fragmentations of the peptides occurred under conditions where the potential difference between the extraction cone and skimmer cone were increased to obtain maximum fragmentation. Typically, the potential was set to 55 V.

All peptide samples were obtained from the Michigan State MacroMolecular Structure Facility. They were purified by HPLC utilizing a mobile phase of acetonitrile and water on an ODS column.

Solutions of analytes were obtained by dissolving approximately 1 μg of sample in 1 ml of acetonitrile: water (1:1). These were introduced into the ESI source with a Sage syringe infusion pump at a flow rate of 20 μl min^{-1}. The filter electrode was heated to 100° C, the nebulization gas flow rate was 18 l hr^{-1}, and the drying gas flow rate was 150 l hr^{-1}.

Collisional activations occurred with positively charged ions. The potential on the adjustable extraction cone to accomplish this was adjusted from thirty-seven to sixty-five volts in approximately five volt increments for selected peptides. The conditions were dependent upon whether molecular weight or sequence information was needed. Differences in charge state were taken into account only for interpretation of sequence and/or pseudo-molecular weight calculations. No attempts were made to influence the charge state of any species.[104,105] Data acquisition occurred at 100 u s^{-1} and

the instrument was set to observe the mass range from 100 to 1500 u. Resolution was adjusted such that for singly charged species the valley was 50% for unit resolution across the mass range. All other parameters were adjusted for optimum performance and sensitivity.

Results and Discussion: For the twenty-one peptides utilized the average molecular mass was 1414 ± 174 u; the average degree of polymerization was 13 ± 1. For these experiments, all pseudomolecular ions were first confirmed, by setting the potential on the cone to thirty-seven volts, optimizing all other parameters, and observing the, typically, doubly charged pseudomolecular ion (Figure 3.3). Species of higher charge were usually of considerably smaller abundance and, in most instances, the singly protonated species was also present, but at much lower intensity. Once the species of interest was confirmed, the extraction cone potential was raised to fifty-five volts (Figure 3.4). The choice of this potential was not arbitrary. The optimum potential was obtained by observing the fragmentation pattern upon increasing the potential in five volt increments (Figure 3.5) for a couple of representative peptides. As was the case with the previously discussed saccharides, increasing the potential on the skimmer cone decreases the total ion current without a concomitant increase in signal-to-noise ratio obtained using a triple quadrupole. Although this alters the lower level of detection for the fragment ions, the sequence remains interpretable and comparable to

the data that can be obtained from ion trap and triple quadrupole instruments. Once optimum settings were determined, the potential was not changed for each species since it was desirable to find a standard set of values to apply to unknown peptides. It was determined, again, that 55 V was an optimum setting that allowed reasonable fragmentation with minimal amounts of adduct ions.

The ion currents of individual oligomers were measured over a range of sampling cone potentials (Figure 3.5). The results are analogous to the results for the peptide in figure 3.4, perhaps suggesting that there may be a common effect dependent on both the mass and the charge of an ion rather than either the charge or the mass alone. There is, however, a difference, in regards to the exact dependency of ion current upon cone potential between the multiply charged ions of the single mass. From an inspection of the optimum extraction cone potentials for various ions in different spectra, Shiel et al.[106] found that the optimum cone potential varied as approximately m/z (i.e., linear dependence) in the case of multiply charged ions of the single mass, but as approximately $(m/z)^{1/2}$ in the case of singly charged ions of multiple masses. The increase in peptide bond cleavage in multiply charged species is more likely from increased charge repulsion.

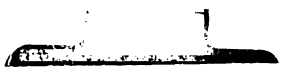


Figure 3.3: Example of a peptide (QKSQRSQAENPV) analyzed by the Fisons Platform mass spectrometer with the potential set to 37 V to obtain molecular weight information.

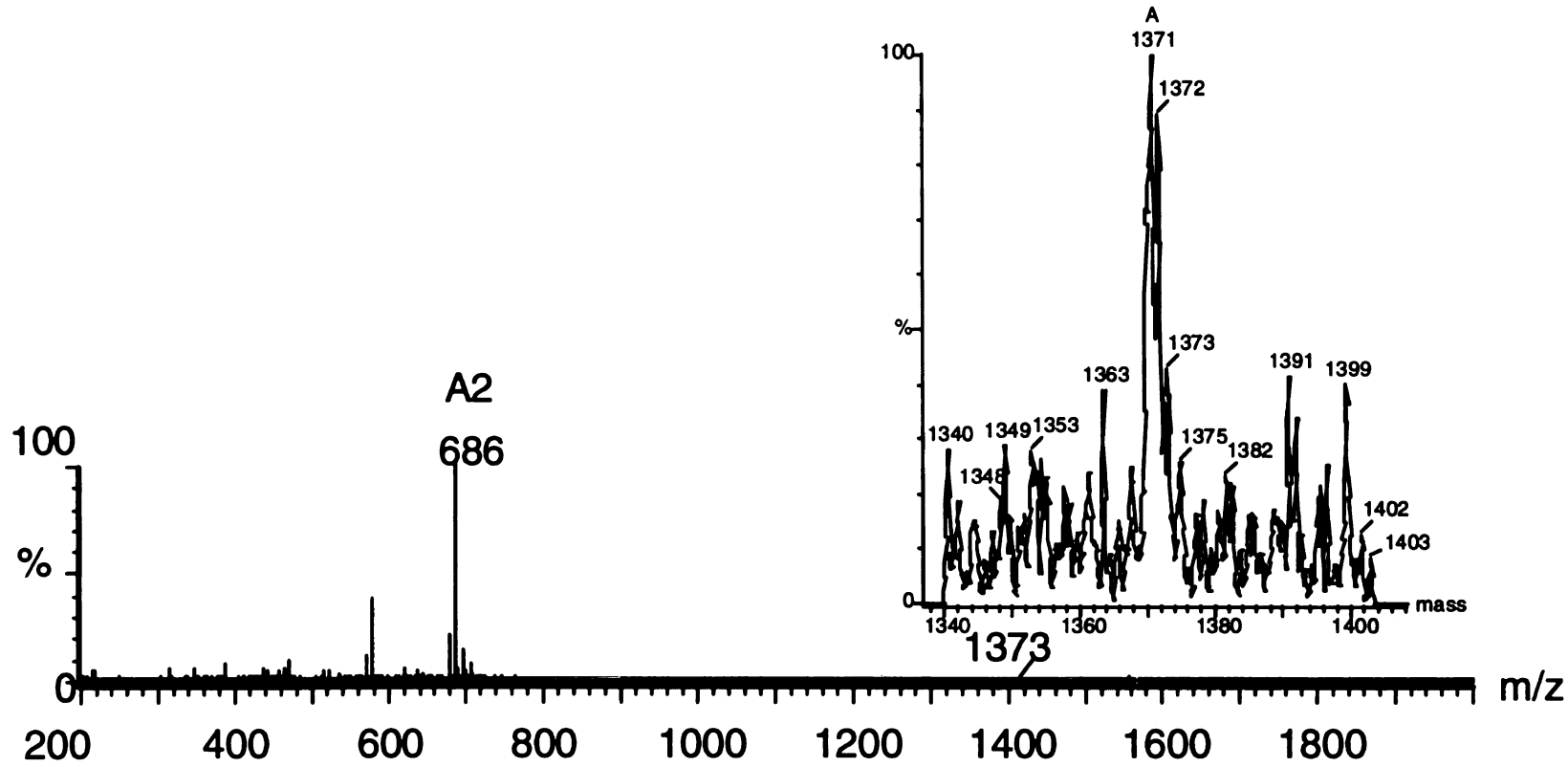

A2
686
100
%
0
200
400
600
800
1000
1200
1400
1600
1800
m/z
A
1371
1372
1373
1363
1391
1399
1340
1349
1353
1348
1375
1382
1402
1403
mass
1340
1360
1380
1400
1373

Figure 3.4: The same peptide as figure 3.3 fragmented using in-source fragmentation (cone potential set to 55 V). Nearly the entire sequence of the peptide is obtained, and the linkages not cleaved can be inferred from the data.

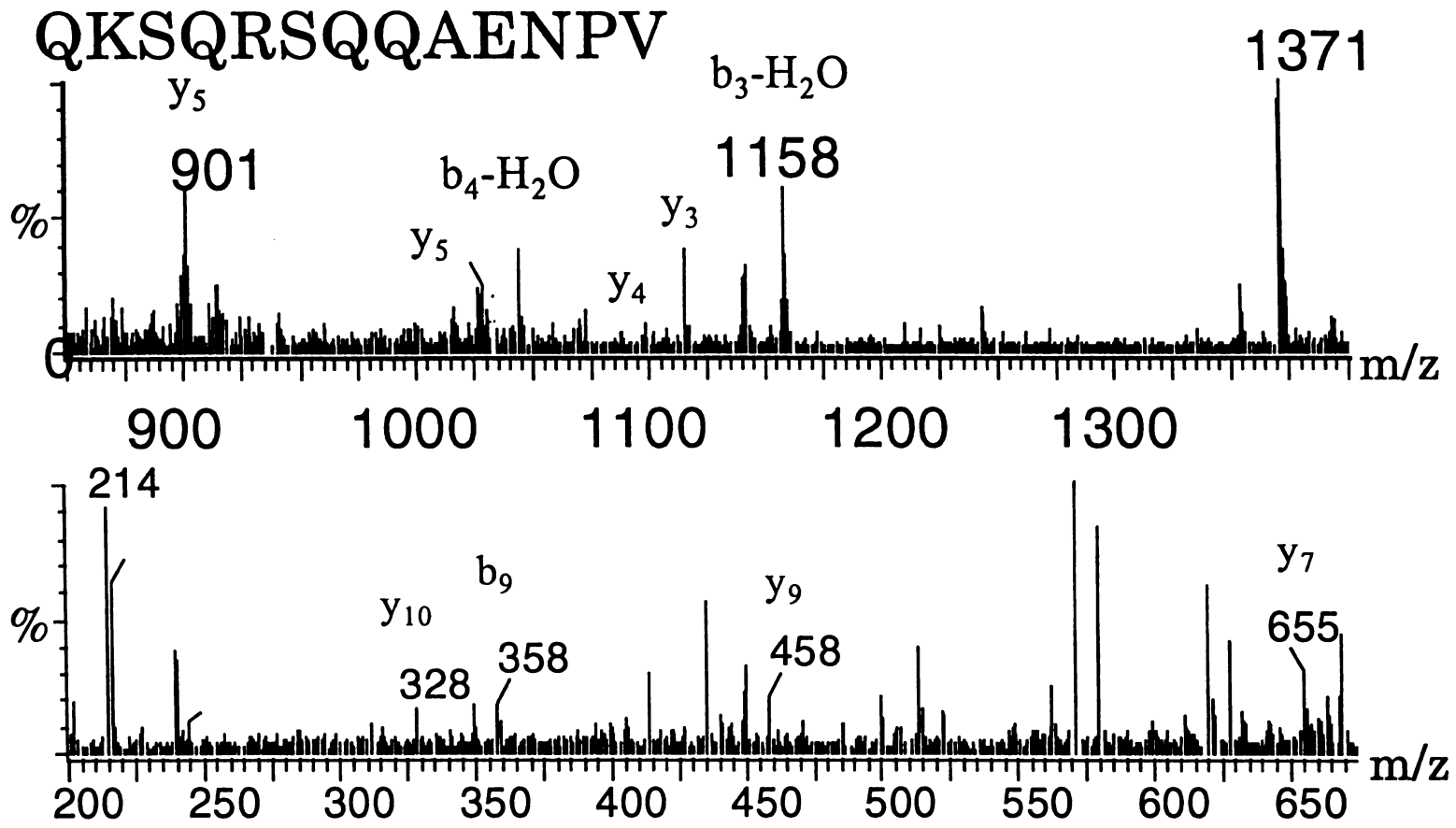

QKSQRSQQAENPV
y5
901
b3-H2O
1158
1371
b4-H2O
y3
y5
y4
%
0
m/z
900
1000
1100
1200
1300
214
b9
y9
y7
y10
655
458
358
328
%
m/z
200
250
300
350
400
450
500
550
600
650

Figure 3.5: Ion current obtained for two peptides (QKSQRAQAENPV, solid line, and QKSQRSQDENPV, dashed line) obtained under conditions of increasing cone potential.

V(cone) vs. TIC

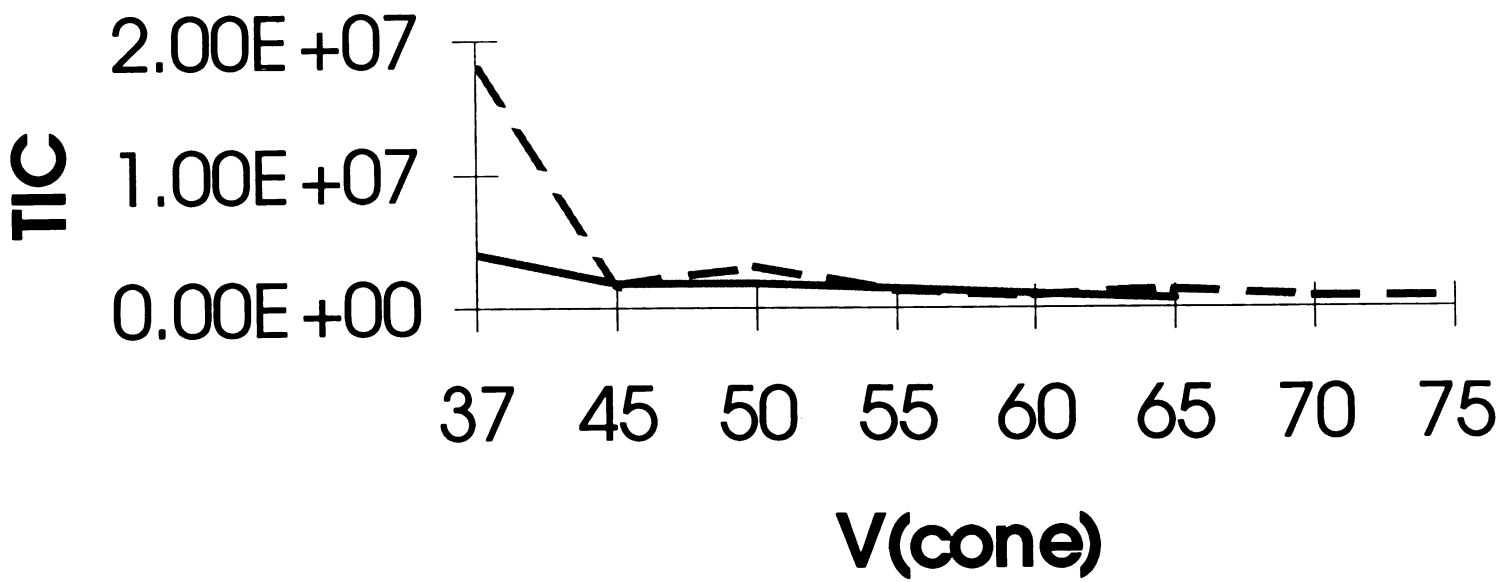


Table 3.1: Data from the in-source fragmentation of twenty one peptides. The type of fragment and the peptide bond that fragmented are indicated.

		a	l	p	f	w	g
Mol. Wt.	Sequence						
1481	ARTTHAGSLPQKSQ	Y9					y8
1516	ARTTHYGSLPAKSQ	B11	b9				
1557	ARTTHYGSLPQKAQ		b9	y5			
1547	ARTTHYGSLAQKSQ	B10	b9				b7
1531	ARTTHYGSAPQKSQ	B9		y5			b7
1587	ARTTHYASLPQKSQ			y5			
1371	QKSQRSQAENPV						
1370	QKSQRSQAENPV	Y5		y2			
1413	QKSQRSQDENPV			y2			
1413	QKSQRSQDENPV			y2			
1413	QKSQRSQDENPV			y2			
1371	QKSQRSQAENPV	Y5		y2			
1437	GAGLSLSRFSWGAE		y9				b12
1437	GRGLSLSAFSWGAE	b8,b13	y9,y11				b12,y12
1480	GRGLSASRFSWGAE		b4,y11		y6	b11,y4	b3,b12,y3,y12
1446	GRGLSLSRASWGAE	B13					b12
1480	GRGASLSRFSWGAE	B13	b6			b11	b12
1506	GRGLSLARFSWGAE	Y8	b6,y11		y6		y12
1413	GRGLSLSRFSWGGE		b4,y9		y6	y4	y2,y3
1506	GRGLALSRFSWGAE	Y2			y6		
1240	AAAAHYGSLPAAAA	B3,b4,b11,b12,b13,y3,y4,y11,y12,y13	b9	b10,y5			y8
1489	PSQGKGRGLSLSRF		y4				y7,y9,y11

Table 3.2: Continuation of the data from the in-source fragmentation of twenty one peptides. The type of fragment and the peptide bond that fragmented are indicated.

		s	t	n	q	y	h	k	r	d	e
Mol. Wt.	SEQUENCE										
1481	ARTTHAGSLPQKSQ	b13,y7	b3,y11,y12		b11		b5		b2		
1516	ARTTHYGSLPAKSQ	y2	b3,b4				b5		b2		
1557	ARTTHYGSLPQKAQ	b8	b3,b4,y7		y4		b5		b2		
1547	ARTTHYGSLAQKSQ		b4				b5		y13		
1531	ARTTHYGSAPQKSQ	b13,y7	b4,y11		b11,y4		b5		b2		
1587	ARTTHYASLPQKSQ	b8	b4			y9			b2		
1371	QKSQRSQAENPV										
1370	QKSQRSQAENPV	b6,y7		y3	b4,y6,y9			b2	b5,y8		b9,y4
1413	QKSQRSQDENPV	y10		b10	b4,y9				y8	b8	b9,y4
1413	QKSQRSQDENPV	y10		b10				y11			y4
1413	QKSQRSQDENPV	b6		b10	y9				y8		b9
1371	QKSQRSQAENPV	y7		y3	y6,y9				b5,y8		b9
1437	GAGLSLSRFSWGAE	y8,y10									
1437	GRGLSLSAFSWGAE	y8,y10									
1480	GRGLSASRFSWGAE	B5,b7,b10,y8,y10							y13		
1446	GRGLSLSRASWGAE								y13		
1480	GRGASLSRFSWGAE	b5,b10,y8							b2,b8		
1506	GRGLSLARFSWGAE	b5,y10							y7,y13		
1413	GRGLSLSRFSWGGE	b5,b7,b10							y13		
1506	GRGLALSRFSWGAE								b2,y7		
1240	AAAAHYGSLPAAAA	b8					y10				
1489	PSQGKGRGLSLSRF	b2,y5,y13			b3,y12			b5,y10			

From tables 3.1 and 3.2, the application of in-source fragmentation to peptides was very successful. This learning set consisted of three families of peptides, in which the sequences were similar, and two peptides in which the sequences were divergent. In no instance is the complete sequence obtained. However, for most peptides, a reasonable partial sequence is present. Those amino acids not identified in the sequence ladder typically can be inferred from the missing masses. The first set of peptides, artthy??s?p?k?p (rows one through six), has only internal differences. In nearly all instances, one of the fragments is at alanine (4/6), leucine (3/5), glycine (3/5), serine (8/11), threonine (9/12), glutamine (3/6), histidine (5/6), or arginine (6/6). For this series, no fragments were observed where peptide bond cleavage occurred at lysine. This does not lead to the conclusion that no fragments will occur at lysine. The next set of data, in tables 3.1 and 3.2 (rows seven through twelve), includes two fragments at lysine in five instances. This data set of five in the list contains only two sequences, and is intended to show consistency. In most instances, the data are identical or complimentary (i.e., either the –b, or –y series is present). The only discrepancies occur for lysine and aspartic acid (1/3). But, even with these differences the sequences of the peptides are similar and consistent. In the next set of data, g?g????????g?e (rows thirteen through nineteen), similar fragmentations to the first data set are obtained. With alanine present as an internal residue, fragmentations at

this residue are very likely, the peptide fragments at alanine six out of thirteen occurrences, the peptide bond associated with leucine (8/14) also resulted in ions that were characteristic. Similar to the first data set, peptide bonds at glycine (10/25), serine (12/21) and arginine (9/14) also fragmented to give characteristic ions. Among the fragment ions that are not present, in most instances, are those where the terminal amino acid is lost. The y_1 and b_1 ions are not included in this assessment, since they were typically below the lowest m/z value acquired. Only in sample twenty is the terminal fragment present. The large number of alanine residues nearly guaranties the presence of one of the terminal fragments. The final peptide in this study was very cooperative in its fragmentation, since there are ions representative of all but three residues. This was expected, due the number glycine and serine residues present. This data set cannot be exhaustive, since the number of possible combinations of amino acids with a degree of polymerization of thirteen is 1.9×10^{22}. But, this set of data has shown what was intended. The use of in-source fragmentation is consistent, accurate, and reliable for determining the primary and secondary structure of peptides. By utilizing sets of peptides that are similar, I have shown that the results are sufficient to differentiate between them.

Conclusion: The application of a single quadrupole mass spectrometer to sequencing peptides is a relatively new approach. Triple

quadrupoles, where the central quadrupole contains a collision gas to induce fragmentation, have been utilized since their introduction for structural characterization of molecules. With the introduction of LC interfaces, the usefulness of mass spectrometry was extended to those molecules that are not sufficiently volatile for GC or that decomposed upon heating so a direct insertion probe could not be used. The concomitant increase in signal-to-noise ratio that is an added benefit for triple quadrupoles, is absent in this single quadrupole technique. That is not to say that the technique is severely limited. Sample concentration requirements are greater for in-source fragmentation and the spectrum tend to have more information, due to secondary fragmentation. Also, being unable to choose an ion in the first quadrupole to fragment are limitations. The information for sequencing is present, nonetheless, and with improved software, and sufficient experience, sequencing of peptides is possible on single quadrupole instruments.

Finally, the argument could be made that current sequencing technology[107] has surpassed the need for mass spectrometry. It is true that current automated sequencers have similar, and in some instances better, sensitivity than quadrupole mass analyzers. These sequencers are not able to obtain information on the post-translational modifications that may be present in a protein or peptide. Glycoconjugates are only one example of modifications that cannot be interpreted automatically. In these instances, it

is necessary to use complementary techniques (e.g., in-source fragmentation or CID) to obtain the sequence of a peptide.

Part II:
Chromatography and Detection of Sulphoquinavose and Suspected Metabolites From Rhizobium meliloti bv. 1021.

Chapter 4

SULPHOQUINOVOSE
LITERATURE REVIEW

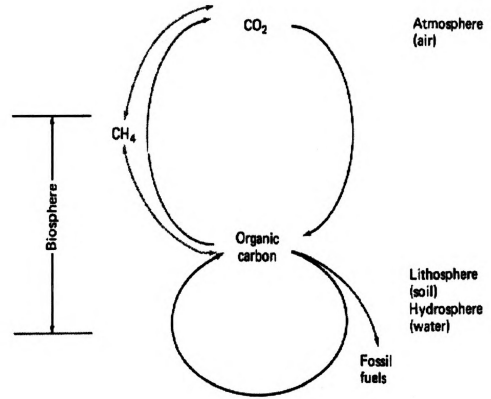
Introduction: Life on this planet is dependent on a series of metabolic cycles. Among these are the carbon, nitrogen (Figure 4.1), and sulfur cycles (Figure 4.2). A cycle contains no beginning, and therefore, no end. But, in order to describe them, you must start somewhere. The carbon cycle, it can be said, begins in plants, which combine carbon dioxide from the atmosphere with water to make plant tissues such as leaves, stems, and fruits. Animals eat the plants and convert the plant tissues into animal tissues. The cycle restarts when the animals die and their decaying tissues are eaten by soil organisms, a process that releases carbon dioxide.[108] Proteins make up a large portion of animal tissues and nitrogen is an essential element of all proteins. The availability of nitrogen in forms that plants can use is a basic determinant of the fertility of soils; the role of soil organisms in facilitating the nitrogen cycle is therefore of great importance. Nitrogen is necessary for the biosynthesis of amino acids and the purine and pyrimidine bases, the nitrogenous building blocks of proteins and nucleic acids. When a plant or animal dies, soil organisms break up the complex proteins, polypeptides, and nucleic acids in their structures and produce

ammonium ions, nitrates, and nitrites that plants can then use to build their structural tissues.[109] Soil organisms also participate in the sulfur cycle mostly by breaking up the naturally abundant sulfur compounds in the soil so that this vital element is made available to plants. The smell of rotten eggs so common in swamps and marshes is due to the hydrogen sulfide produced by these microorganisms.[109]

Microbial biogeochemical cycling activities and interpopulation relationships have a major influence on agricultural practices. As in other areas of economic microbiology, some microbial activities are beneficial and others are detrimental, and it is the controlled balance of microbial activities that is important in determining agricultural success. Fertilizers and pesticides have become an integral part of modern agricultural practice because crop yields depend on maintaining high levels of soil fertility and limiting destruction due to infectious diseases and pest populations. The environmental fate and usefulness of these chemical fertilizers and pesticides are greatly influenced by the metabolic activities of microorganisms in soil.[110] In some cases the microbial populations themselves provide essential inputs of fixed forms of nitrogen to soil and to effective control of pest populations. Both bacteria and blue-green algae can fix nitrogen directly from the atmosphere, but this is less vital to plant development than the symbiotic relationship between the bacteria genus Rhizobium and leguminous plants and certain trees and shrubs.[109] This bacterium, in return for secretions from

their host that encourage bacterial growth and multiplication, fix nitrogen in nodules of the host plant's roots, providing nitrogen to the plant in a usable form. Scientists concerned with plant mineral nutrition have studied biological nitrogen fixation for more than one hundred years.[111] Among the most urgent problems that face Mankind are providing sufficient nourishment to the constantly growing World, reducing environmental pollution and conserving energy. The intense interest in biological nitrogen fixation shown by scientists from divers fields of research is obviously connected with the fact that understanding the main steps of this process will help people cope with these global problems.[112]

Figure 4.1: The carbon (A) and nitrogen (B) cycles: the chemical forms and key processes in the biogeochemical cycling of carbon and nitrogen, respectively. Bacteria mediate all the critical steps of nitrogen fixation, nitrification, and denitrification.


CO_2
Atmosphere
(air)
CH_4
Biosphere
Organic
carbon
Lithosphere
(soil)
Hydrosphere
(water)
Fossil
fuels

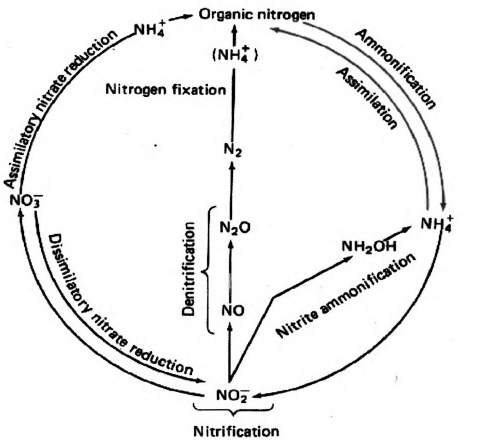
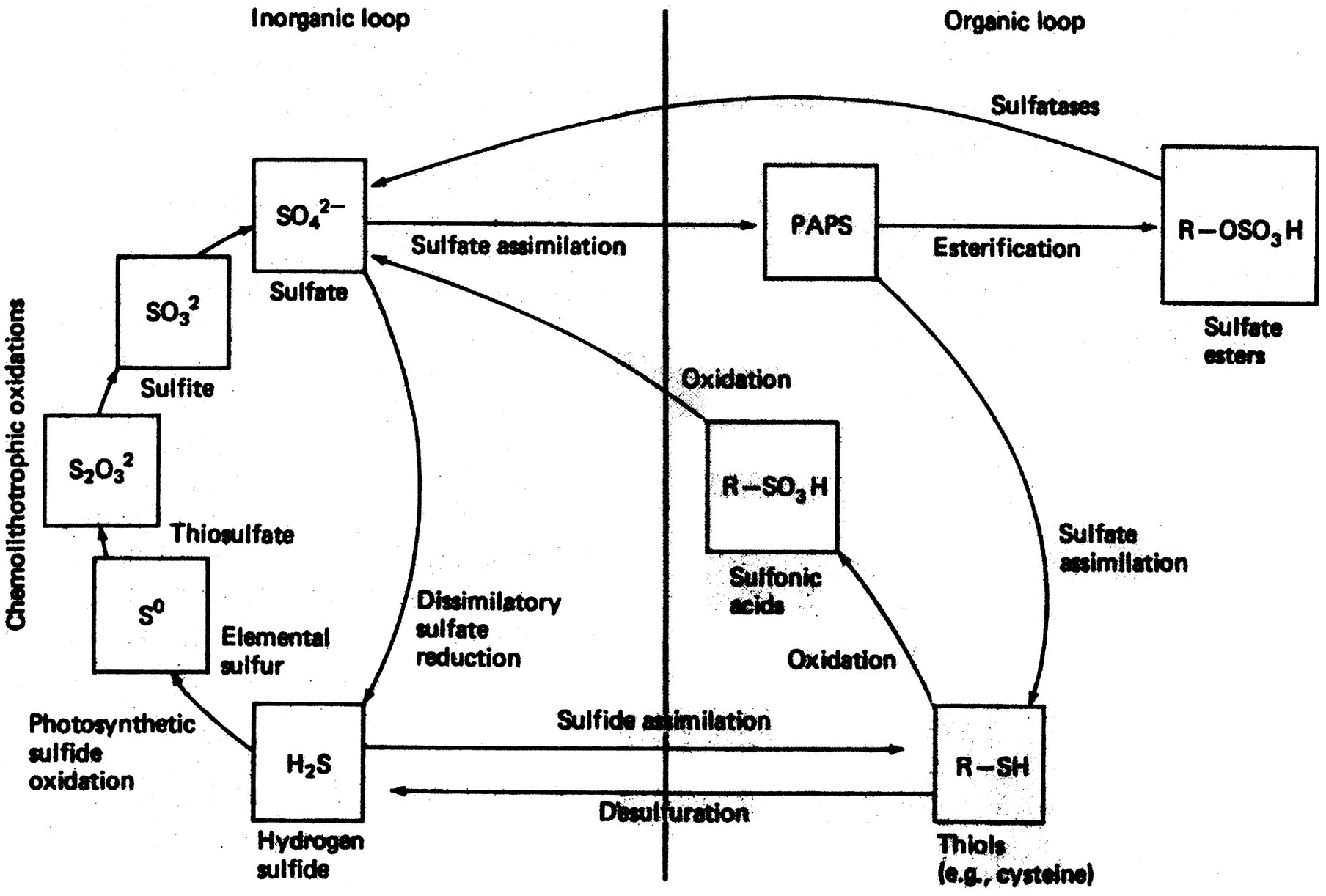

Organic nitrogen
NH_4^+
(NH_4^+)
Assimilatory nitrate reduction
Ammonification
Assimilation
Nitrogen fixation
N_2
NO_3^-
N_2O
Denitrification
NH_2OH
NH_4^+
Dissimilatory nitrate reduction
NO
Nitrite ammonification
NO_2^-
Nitrification

Figure 4.2: The sulfur cycle, showing various transformations of organic and inorganic compounds.


Inorganic loop
Organic loop
Sulfatases
SO_4^{2-}
Sulfate
Sulfate assimilation
PAPS
Esterification
$R-OSO_3H$
Sulfate esters
SO_3^2
Sulfite
Oxidation
$R-SO_3H$
Sulfonic acids
Sulfate assimilation
$S_2O_3^2$
Thiosulfate
Chemolithotrophic oxidations
S^0
Elemental sulfur
Dissimilatory sulfate reduction
Oxidation
Photosynthetic sulfide oxidation
Sulfide assimilation
H_2S
Hydrogen sulfide
Desulfuration
R–SH
Thiols (e.g., cysteine)

Bacterial Symbiosis: Certain species of nitrogen-fixing bacteria can coexist intimately, symbiotically, with legumes and other plants, providing the plants with necessary nitrogen. In this symbiotic association the bacteria become encased in nodules that grow in the roots of plants through which nitrogen that has been fixed by the resident bacteria is obtained by the plant. These organisms were originally described as endorhizosphere microorganisms. However, the microflora that colonize this specialized habitat inside healthy roots are now more commonly referred to as endophytes or internal root colonists.[113] This habitat has already been identified as an important reservoir for isolation of N_2-fixing plant growth-promoting rhizobacteria (PGPR) in some cereal roots. Cyanobacteria have developed similar relationships with various life forms, such as liverworts, hornworts, cycads, and at least one genus of flowering plants (Gunnera).[114] Their symbiotic relationship with fungi has earned its own designation - the coexistent species are called lichen. Rhizobia (now classified into three genera, Rhizobium, Bradyrhizobium, and Azorhizobium) are the soil bacteria that can elicit the formation of nitrogen-fixing nodules on the roots of selected species of the legume family.[109]

The legume-Rhizobia symbiosis is estimated to fix per year as much nitrogen as the fertilizer industry and is of great agronomic, economic, and ecological importance[108]. This efficient nitrogen-fixing association occurs as a

result of the formation of a new specialized organ on the plant root, the root nodule, which is induced by the prokaryotic symbiont on its specific plant partner.[115] The infection begins with the bacteria secreting a class of lipo-oligosaccharides in response to plant hormones, flavonoids, which trigger the early steps of legume nodule formation. The Rhizobial genes, the Nod genes, are essential for this process[116,117] and determine host recognition. These genes have been found to specify the synthesis of excreted lipo-oligosaccharide signals that are capable of eliciting many of the plant responses characteristic of the bacteria themselves.[108] Purified Nod factors, when applied to legume seedlings at concentrations as low as 10^{-12} M, stimulate the differentiation of epidermal cells into root hairs and also deform root hairs .[108] The structures of many of the products of the nod genes have been determined.[118-126] Uniformly, they consist of an oligosaccharide backbone of β-1,4-linked N-acetyl-D-glucosamine residues varying in length from three to five sugar units. A structurally varied fatty acid group is attached to the nitrogen group of the nonreducing amino sugar moiety. The presence of other substitutions is dependent on the species and strain (Appendix E and Figure 4.1). In response to the bacteria the plant initiates cell division, induction of specific changes in cell morphology, and triggering of a plant organogenic program leading to the formation of the nodule.[127] Within these root or stem structures, the bacteria are able to convert atmospheric nitrogen, under conditions of nitrogen limitation, into ammonia,

which is used by the plant as a nitrogen source. Initially it was speculated that attachment of Rhizobia to legume roots was involved in host specificity. The specificity of adherence by compatible Rhizobia species was proposed to be mediated by specific binding of particular polysaccharide structures present on the bacterial cell surface to host plant lectins.[128,129]. The lectin recognition hypothesis has been validated in one case, although many studies fail to demonstrate any degree of host specificity at the attachment step.[130,131]

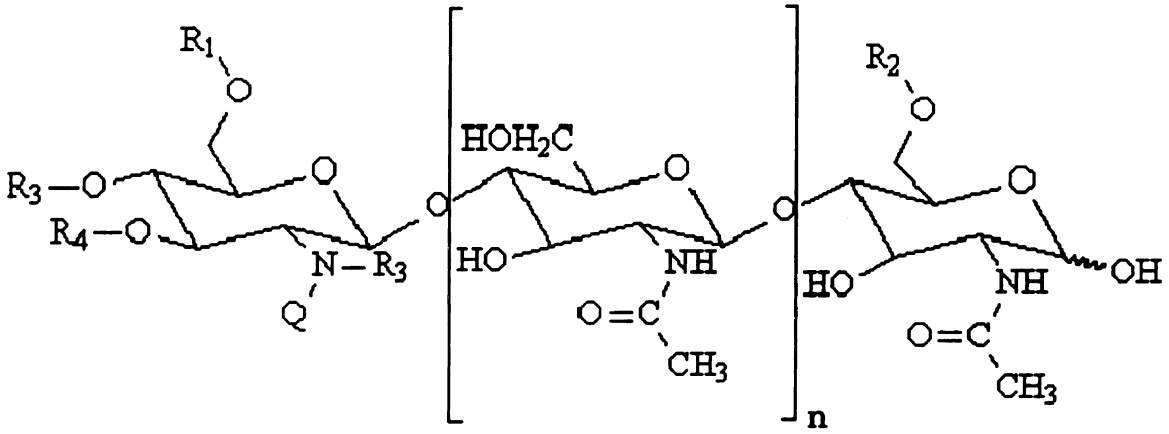


Figure 4.3: The general structure of nodulation factors.

Biosynthesis of Bacterial Factors: Since sulfolipids make up a major contribution to the sulfur cycle, knowledge of their metabolism is important. It is generally accepted that the sulfolipid sulfoquinovosyl-diacylglycerol (SQDG) is present in all-higher plants, mosses, ferns, algae, and in several species of Rhizobia.[23, 131] In spite of its importance, attempts to elucidate the biosynthesis of SQDG have so far proved fruitless. Only recently has it become feasible to rigorously test the different chemical and

genetic approaches. Work on determining the synthesis has fallen into two areas: the sulfoglycolytic pathway, and the sugar-nucleotide pathway that utilizes UDP-glucose and/or UDP-galactose (Figure 4.2 and 5.1). The former scheme is thought to involve sulfonated analogs of the intermediates of glycolysis (Appendix D). Here sulfolactaldehyde and dihydroxyacetone phosphate condense, via an aldolase reaction, to form 6-sulpho-6-deoxy-D-fructose-1-phosphate, which would be a direct precursor of sulfoquinovose. In the alternative to the sulfoglycolytic pathway, the sugar-nucleotide pathway, sulfur is introduced at the level of hexose, either glucose or galactose, and not at the triose as predicted by the sulfoglycolytic pathway.

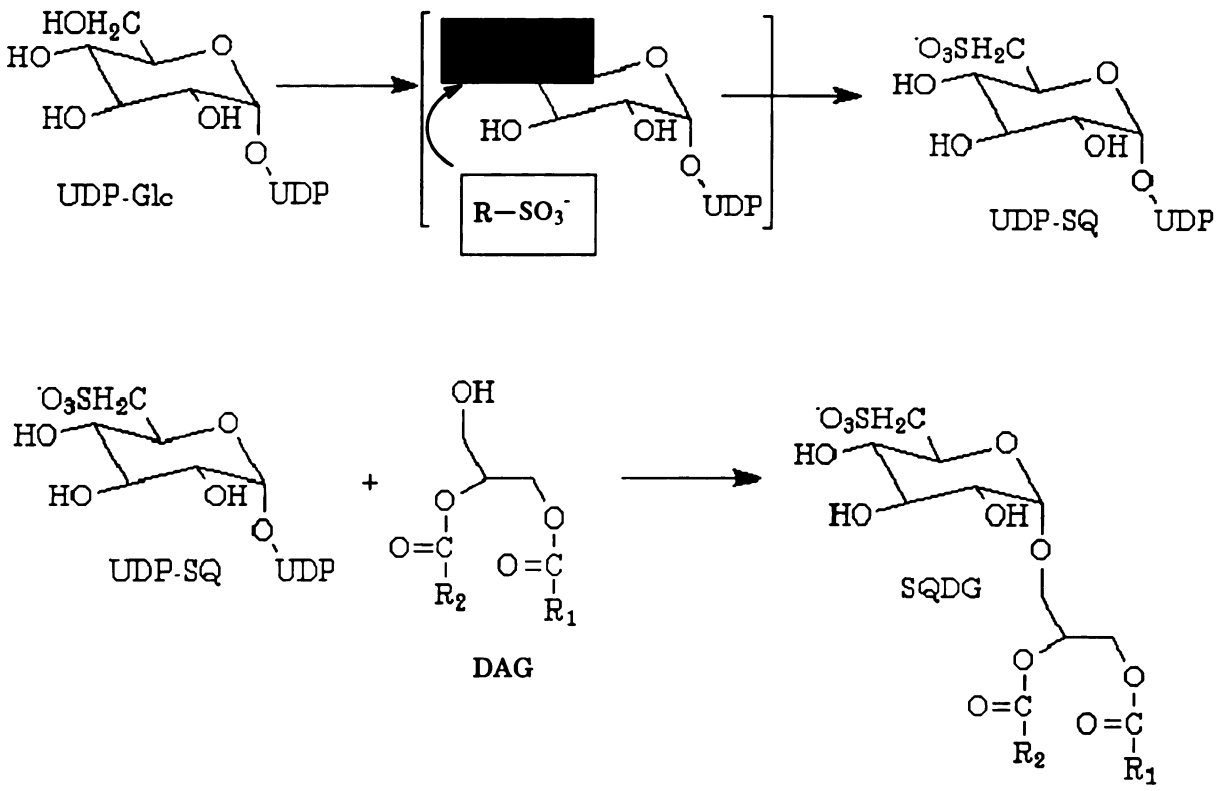


Figure 4.4: Schematic for the sugar-nucleotide pathway of sulfoquinovose synthesis.[131]

Despite nearly forty years of efforts, our current knowledge about the reactions leading up to the formation of the sulfoquinovose (SQ) head group still falls short of the claim that the pathway of SQDG biosynthesis has been solved. The complexity of this classic problem of plant biochemistry becomes apparent if one considers that three major pathways contribute to the biosynthesis of SQDG: carbohydrate metabolism, sulfur assimilation, and fatty acid biosynthesis. Of these three metabolic processes, carbohydrate metabolism has been most intensely studied, and many ideas regarding the

possible pathway for SQDG biosynthesis have been derived by analogy to the common reactions of gluconeogenesis and glycolysis.[132]

There are many approaches that are used, singly or in combination, to work out the chemical details of a metabolic pathway. The first and most direct is to study the pathway in vitro, in a cell-free extract of a tissue capable of catalyzing the overall metabolic processes in question. Another important approach to elucidation of metabolic pathways is the study of genetic mutations in which a given enzyme fails to be synthesized in active form. Such a defect, if not lethal, may result in the accumulation and excretion of the substrate of the defective enzyme. Another powerful method for establishing the general outline of a metabolic pathway is to use an isotopic form of an element to label a given metabolite and then follow the label through its cycle.

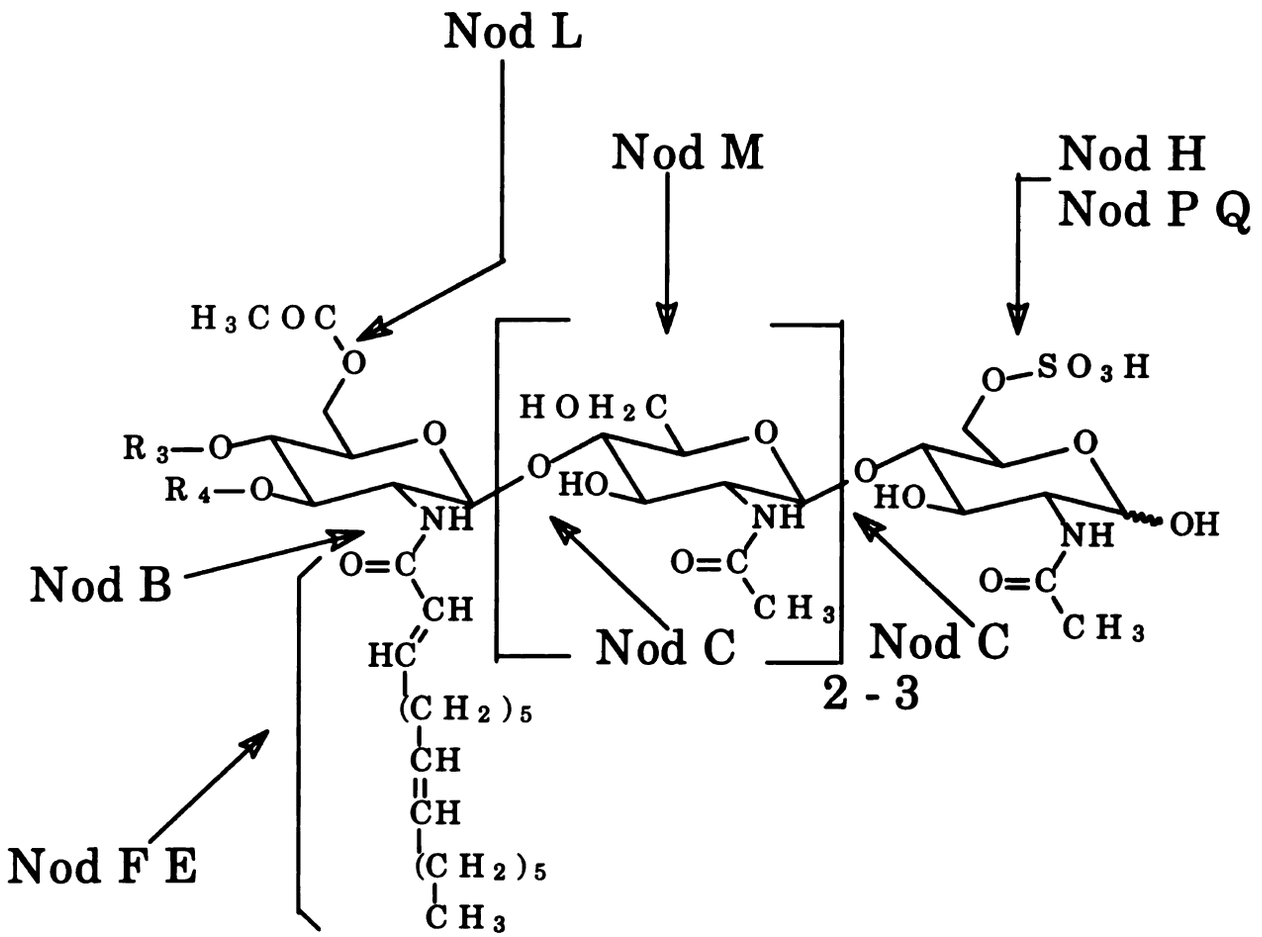


Figure 4.5: Schematic of sulfoquinovose with the nod genes indicated that are responsible for the formation of the various bonds.

Genetic Studies: In the study of the formation of Rhizobial head groups, a number of bacterial genes have been identified that are required for the formation of the nodulation factors. These nodulation genes (designated nod, nol, and noe) of the symbiotic soil bacteria Rhizobia, may be divided into two classes. One class comprises genes which, when mutated, completely abolish nodulation on all legumes. Since genes in this group share significant sequence homology and can be complemented between different Rhizobial genera/species, they are often called "common." Nod ABCIJ and nod D are the best known examples.[131-134] Genes of the second class are necessary for the interaction with certain, but not all, legumes. Expression of these genes permits nodulation of additional hosts, and for this reason, they have been termed host specificity of nodulation (hsn) genes. By definition, they are unique to one or a few Rhizobia.[108,109,135,136] In most cases, mutations cannot be fully complemented by the introduction of the corresponding genes from other Rhizobia. Mostly mutations result in alteration or extension of the host range. In 1990, Lerouge et al.[137] showed that the products of the nod genes are N-acylated-D-glucosamine. Numerous other investigators have confirmed these findings.[138,139] Since these substances are the products of the nod genes, they are called Nod factors. Principal differences among the Nod factors of the various Rhizobia concern the length of the core molecule as well

as the substitutions to both the reducing and nonreducing residues. Presumably, the hsn genes are responsible for these substitutions.

Separation and Detection of SQ, SQDG, and Metabolites: Techniques for the separation and detection of phosphorilated and sulfated saccharides are similar to the techniques used for saccharides. Many of those that have been used include TLC on silica plates.[140-144] These assays used mobile phases that were mixtures of organic and aqueous phases, typically a first dimension would contain chloroform: acetone: methanol: acetic acid: water, in various proportions. Joyard et al. used this as a first dimension and isopropyl ether: acetic acid as the second dimension for the detection of polar lipids. A deviation on this theme was used by Olsthoorn et al. where butanol: ethanol: water was used, several reverse phase TLC assays, ODS coated TLC plates, using acetonitrile: water were also developed for lipooligosaccharides and some exclusively for SQDG.[141,145,146] In all instances a radioactive nuclide, ^{13}C, ^{35}S or both, was detected by autoradiography.

Attempts to determine the correct biosynthetic pathway for SQ and SQDG have also utilized ion exchange chromatography. Lee et al.[147] using a PA 100 column with a mobile phase of 75mM KH_2PO_4 at a flow rate of 1ml min^{-1} with detection at 254 nm looked for methods to detect UDP-galactopyranose and UDP-galactofuranose. Also attempting to determine UDP-sugars were Hull and Montgomery.[148] Their assay used a PAX-100

column and two three step gradients all consisting of water, 200 mM NaOH, and 50 % isopropanol, conductometric detection was used. The goal was to separate UDP-sugars, nucleotides, and sugar phosphates. A related assay for sugar phosphates in biological extracts was developed by Smits et al.[149] Yeast extracts were analyzed by solid-phase extraction and anion-exchange chromatography with a gradient of sodium acetate, from 100 mM to 500 mM and pulsed amperometric detection.

In addition to HPAEC, reverse phase HPLC has been used for separating SQ, SQDG, its precursors, and in the search for metabolites. Sanjuan et al.[150], to prove that a 2-O-methylfucose moiety is present in the lipo-oligosaccharide nodulation signal of *B. japonicum*, used a acetonitrile: water gradient at a flow rate of 1 ml min^{-1} to prove the purity of the nod factor prior to MS and NMR analysis. Synthetic standards have also been chromatographed by reverse phase HPLC. Rossak et al.[151] used these standards to confirm the retention time of UDP-sulfoquinovose. By inactivating sqd D they caused an accumulation of UDP-sulfoquinovose and with the use of standards, confirmed that they co-chromatographed.

Normal phase HPLC has also been used for the separation of sulfated glycans. Karamanos et al.[152] separated 24 variously sulfated galactosaminoglycan- and hyaluronan-derived disaccharides on an Econosphere NH_2 column. In order to obtain separation for all of these species, three separate mobile phases were needed. For non-sulfated Δ-

disaccharides, constituents on the column were eluted with 5 mM sodium dihydrogen orthophosphate buffered with orthophosphoric acid to pH 2.55, with a flow rate of 1.1 ml min^{-1}. Monosulfated Δ-disaccharide's required a mobile phase of 50 mM sodium dihydrogen orthophosphate buffered with orthophosphoric acid to pH 2.50 with the flow rate adjusted to 0.7 ml min^{-1}. Finally, for over-sulfated species 50 mM sodium sulfate 10 mM sodium acetate buffered with acetic acid to pH 5.0 with a flow rate of 1.5 ml min^{-1} was required.

HPLC is not capable of the high resolution of capillary electrophoreses. This technique has been used to separate neutral, charged, and highly charged oligosaccharides. Using open tubular capillary columns, Novotny, et al.[153-155] have been able to obtain separations for all of these species.

Isolation of SQ and Metabolites: Reverse phase HPLC is not strictly for analysis and purity conformation. Shultze, et al.[156] used preparative reverse phase HPLC to purify sulfated lipooligosaccharides that exhibited different degrees of plant host specificity from *R. meliloti*. Their preparative scale (47 x 300 mm) separation used a mobile phase of 20 % acetonitrile: water with a flow rate of 120 ml min^{-1}. Once purified, a family of sulfated lipooligosaccharides was obtained that exhibited different degrees of host plant specificity. A modification of this procedure was used by Poupot, et al.[157,158] Their separation used a semi-preparative reverse phase HPLC

column (250 x 7.5 mm) and a linear gradient of 20 % acetonitrile: water to 100 % acetonitrile, at a flow rate of 2 ml min^{-1}. With these purified nod factors, they were able to determined that *R. tropici* nod factors are sulfated or nonsulfated chitopentasaccharides containing an N-methyl-N-acyl-glucosaminyl terminus, and that the nod factor sulfation is limited by the amount of sulfate. A similar separation was used by Mergaert, et al.[159] to determine that modifications in a D-arabinosyl residue, an N-methyl and a carbamoyl group are present in the nod factors of a strain of *A. caulinodans*.

In addition to reverse phase HPLC, column chromatography using silica gel has also been used to purify SQ, SQDG, and other nod factors. Sanjuan, et al.[123,149] and Carlson et al.[160] used silica gel 60 columns with a 60 % acetonitrile: water gradient for the purification of nod factors from *B. japonicum*. Both used this procedure to determine the structures and/or biological activities of lipooligosaccharide nodulation signals.

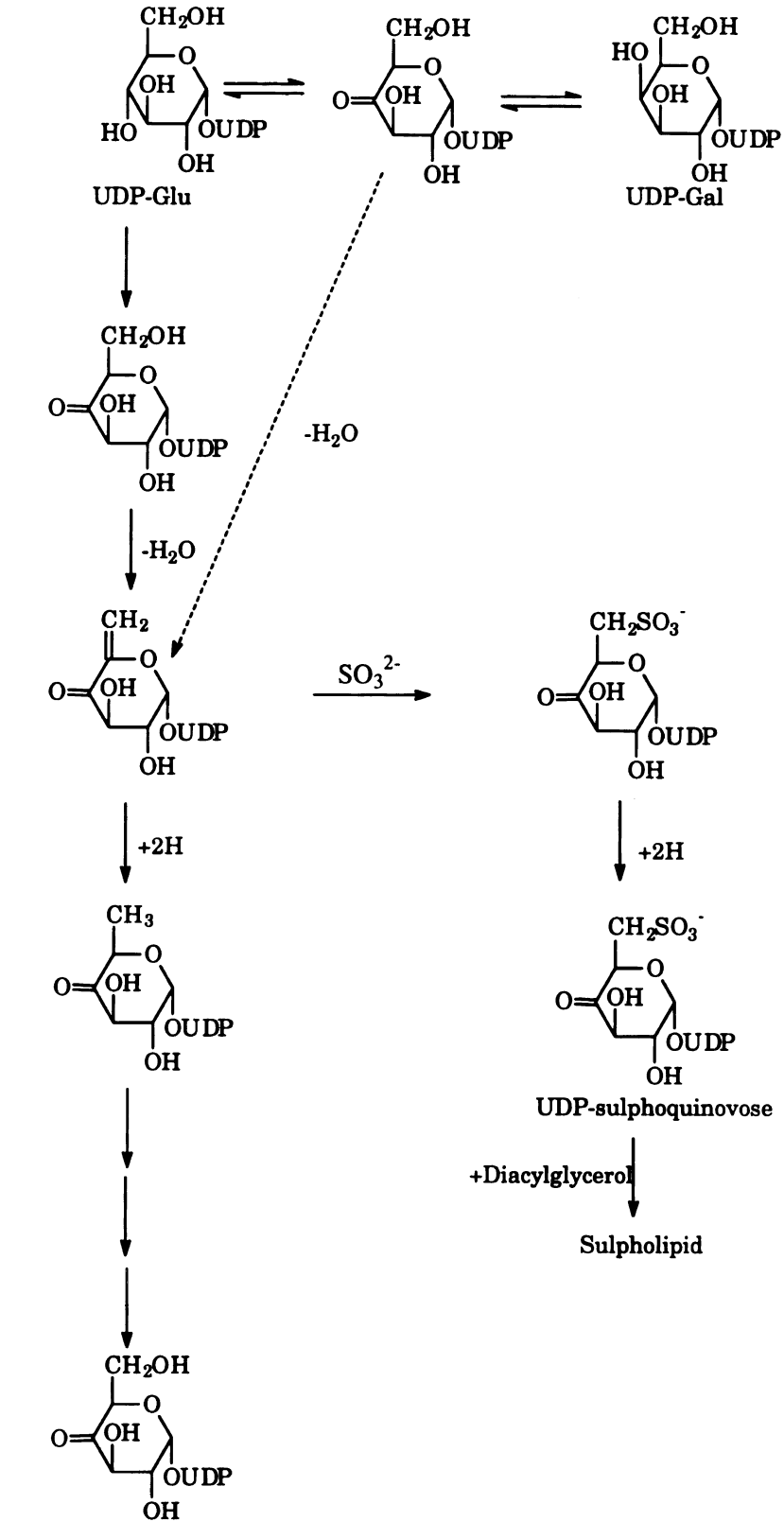
Chapter 5

SYNTHESIS OF SYNTHETIC STANDARDS OF SUSPECTED SULPHOQUINOVOSE METABOLITES

Introduction: The study of metabolism has as its goal the complete, detailed description of the reactions by which the constituents of the cell are built up and broken down (intermediary metabolism) and of the mechanisms by which the chemical energy derived from the degradative reactions is conserved to be utilized in synthetic processes (energy metabolism or bioenergetics). In metabolism each consecutive step in the pathway brings about a small, specific chemical change, usually the removal, transfer, or addition of a specific atom, molecule, or functional group.[161] As previously mentioned, there are several techniques for determining a biochemical pathway, other methods are available, but must be used in conjunction with the primary procedures.

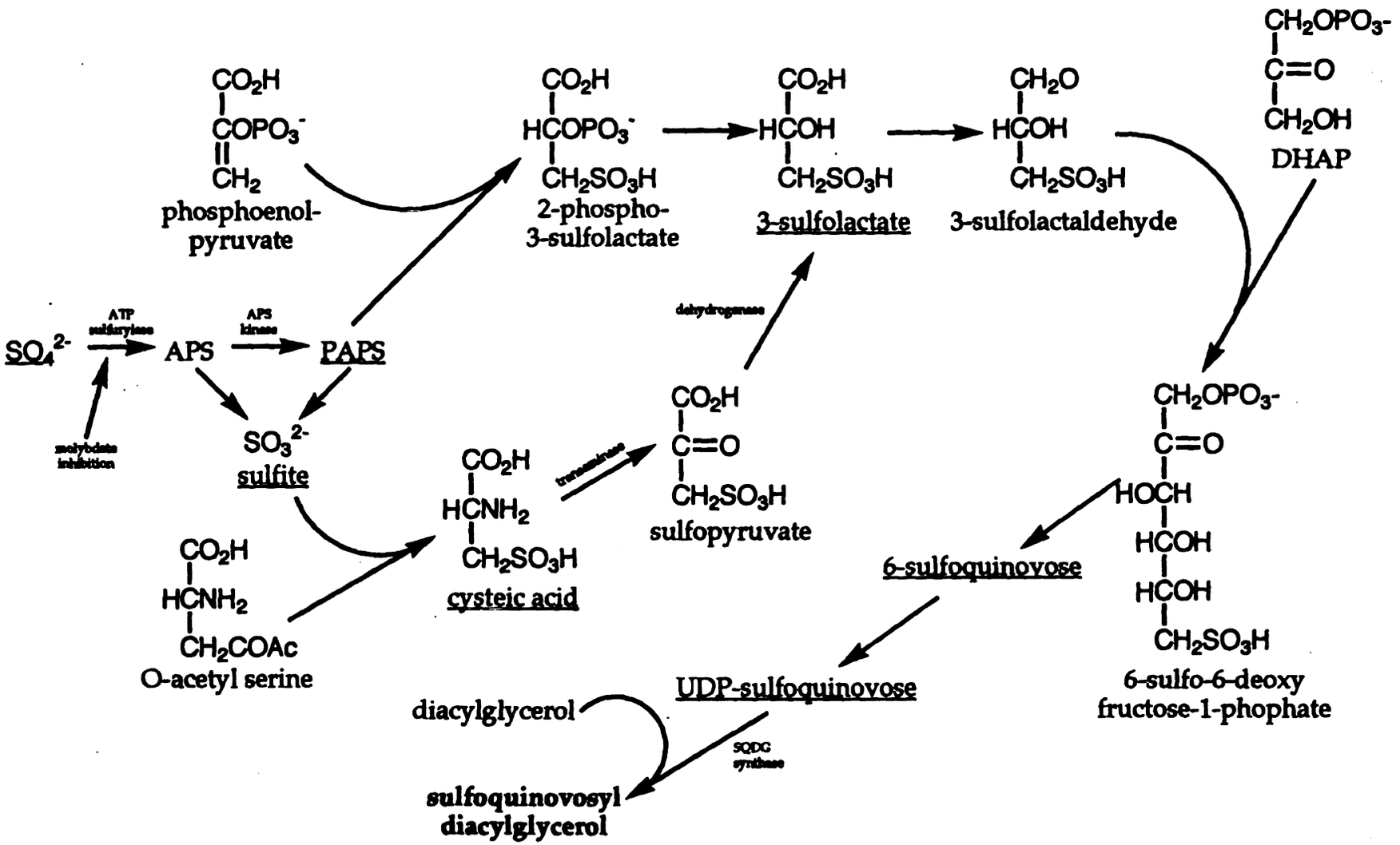
In order to discover the biological pathway for formation of critical biomolecules for Rhizobia symbiosis, the use of standard molecules can be invaluable. In the instance of SQ, there are currently two proposed pathways, one involves UDP-glucose (Figure 5.1[131]), and the other utilizes a modified glycolysis pathway.

Figure 5.1: One proposed pathway for a nucleotide sugar to form sulfolipids starting with either UDP-glucose or UDP-galactose.



Pathway for synthesis of
6-deoxy sugars

Figure 5.2: An alternative route to the formation of sulfolipids using the gluconeogenesis pathway.


phosphoenol-pyruvate
2-phospho-3-sulfolactate
3-sulfolactate
3-sulfolactaldehyde
DHAP
ATP sulfurylase
APS kinase
APS
PAPS
molybdate inhibition
sulfite
dehydrogenase
sulfopyruvate
cysteic acid
O-acetyl serine
6-sulfoquinovose
6-sulfo-6-deoxy fructose-1-phophate
UDP-sulfoquinovose
diacylglycerol
SQDG synthase
sulfoquinovosyl diacylglycerol

Materials and Methods: The 3-Bromo-acetaldehyde-dimethyl acetal was obtained from Aldrich. The sodium sulfite used in all reactions was purchased from Sigma. The methyl α-D-glucopyranoside was obtained from Sigma. Carbon tetrabromide, triphenylphosphine (TPP), and sodium metaperiodate were all acquired from Aldrich. The derivatizing reagents, α-aminopyridine, p-nitrobenzylamine, and pyrenemethylamine were all obtained from Aldrich. All solvents were analytical grade from Baker.

Synthesis: The formation of the sulfonic acid of acetaldehyde dimethylacetal occurred by dissolving 2.11 g of bromoacetaldehyde dimethylacetal in 20 ml of a water: methanol solution (4:1). To this solution 8 g of sodium sulfite ($NaSO_4$) was added, and the reaction, shown in Figure 5.3A, occurred in one hour at 20 °C. The mother liquor from this reaction was evaporated and then dissolved in 18 ml of water and 2 ml of 12 M (concentrated) HCl. The mixture was left at room temperature overnight, for twelve hours, then concentrated to a syrupy solution to form the aldehyde, as shown in Figure 5.3B. This product was dissolved in 50 ml of water and calcium chloride (3.33 g) was added, to this solution. Sodium cyanide (1.50 g), dissolved in 20 ml of water, was then introduced. The mixture was stirred at room temperature for 36 hours and calcium hydroxide (2.22 g) was then added. Stirring was continued for a further 6 hours and then the mixture

was filtered and concentrated to form the sulfonic acid of lactic acid, as shown in Figure 5.3C.

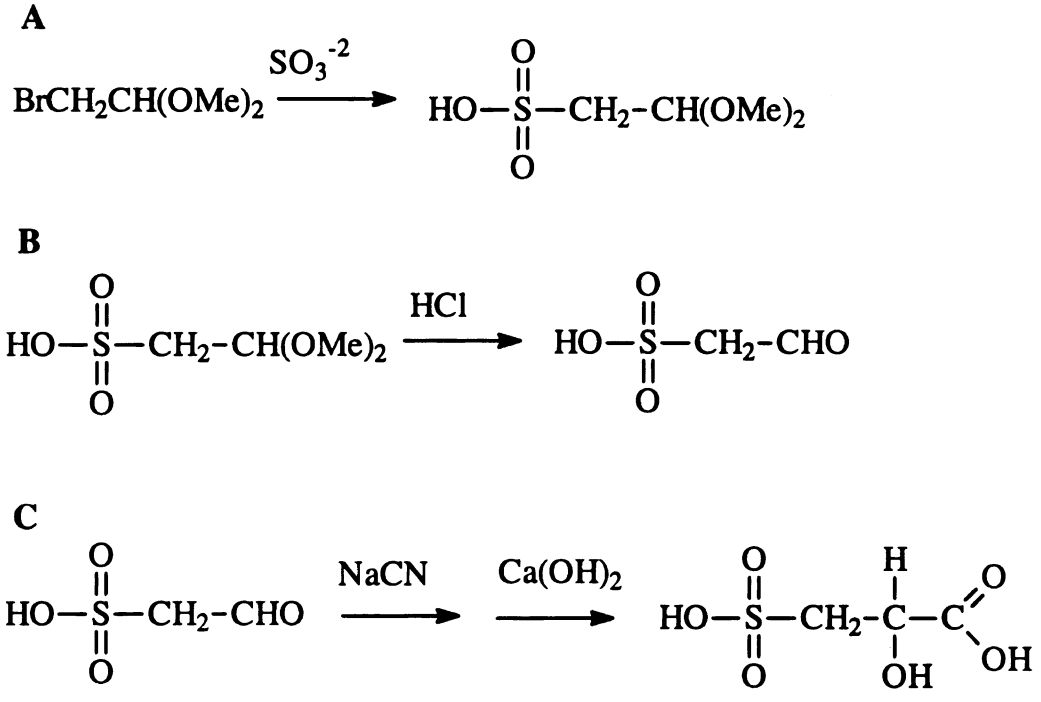


Figure 5.3: A-B. Formation of the sulfonic acid of acetaldehyde. C. Formation of the sulfonic acid of lactic acid.

The methyl glycoside of sulfoquinovose was synthesized starting with methyl α-D-glucopyranoside. The glycoside was suspended in dry pyridine (200 ml) and triphenylphosphine (TPP) (31.5 g, 0.12 mol) was added. The mixture was cooled to 10° C after which carbon tetrabromide (39.7 g) was introduced. The mixture was then heated to 100° C for 45 minutes, and the pyridine was removed on a rotary evaporator, the scheme is shown in Figure 5.4A. Water (500 ml) was added and the milky suspension was extracted 3 times (400 ml per extraction) with dichloromethane. The aqueous solution

was concentrated to a syrup which was taken up in water (150 ml) to which sodium sulfite (19 g) was added. The solution was heated on a steam bath for 18 hours and then concentrated to a syrup to which methanol (350 ml) was added. A white precipitate formed and the mixture was left at 10° C for 12 hours. It was then filtered and the white solid retained. An aliquot of the material (2.5 g) was dissolved in water (20 ml). Sodium metaperiodate (5.0 g) was then added, and the mixture was kept at 4° C for 12 hours in darkness. The reaction is shown in Figure 5.4B and is believed to take place via cyclic periodate intermediates.[162] This formed the sodium salt of propanal-2-ol-3-sulfonic acid.

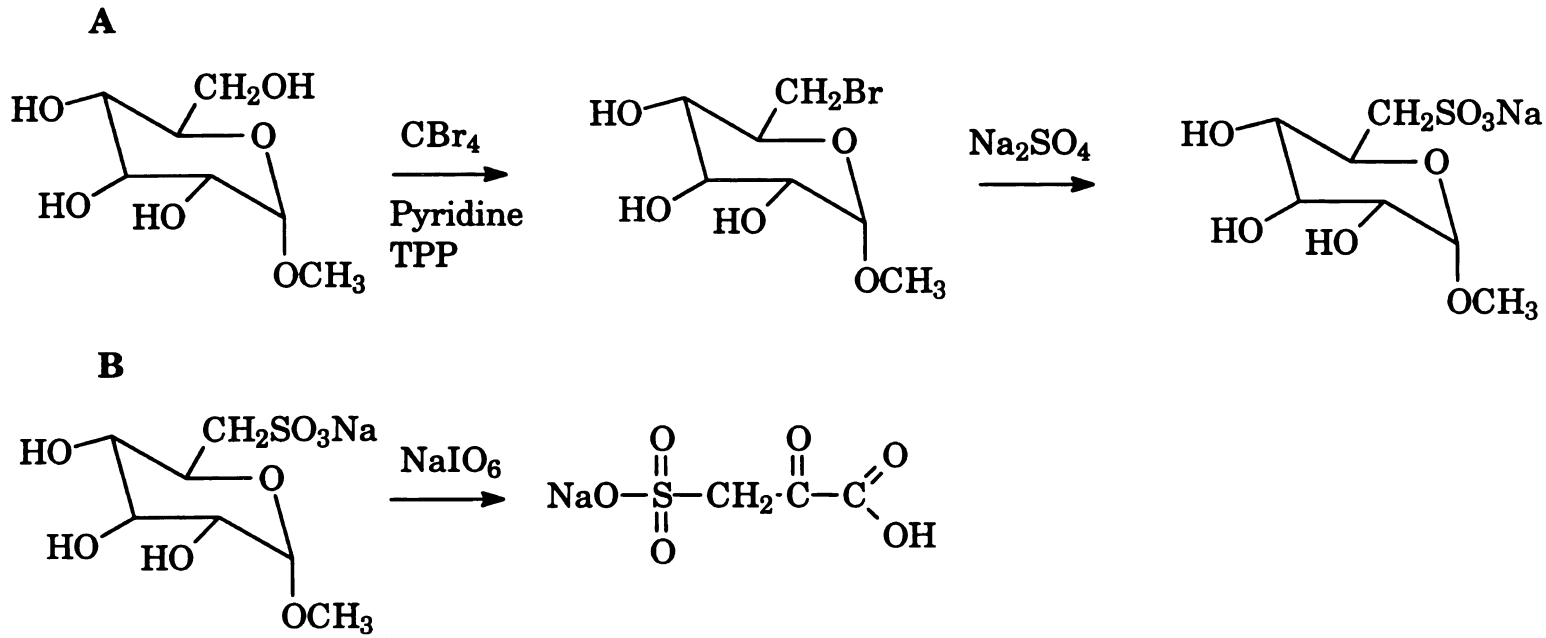


Figure 5.4: A. The formation of the methylglycoside of sulfoquinovose from methyl-glucopyranoside. B. Formation of the sulfonic acid of propanal-2-ol-3-sulfonic acid from the methylglycoside of sulfoquinovose by periodate oxidation.

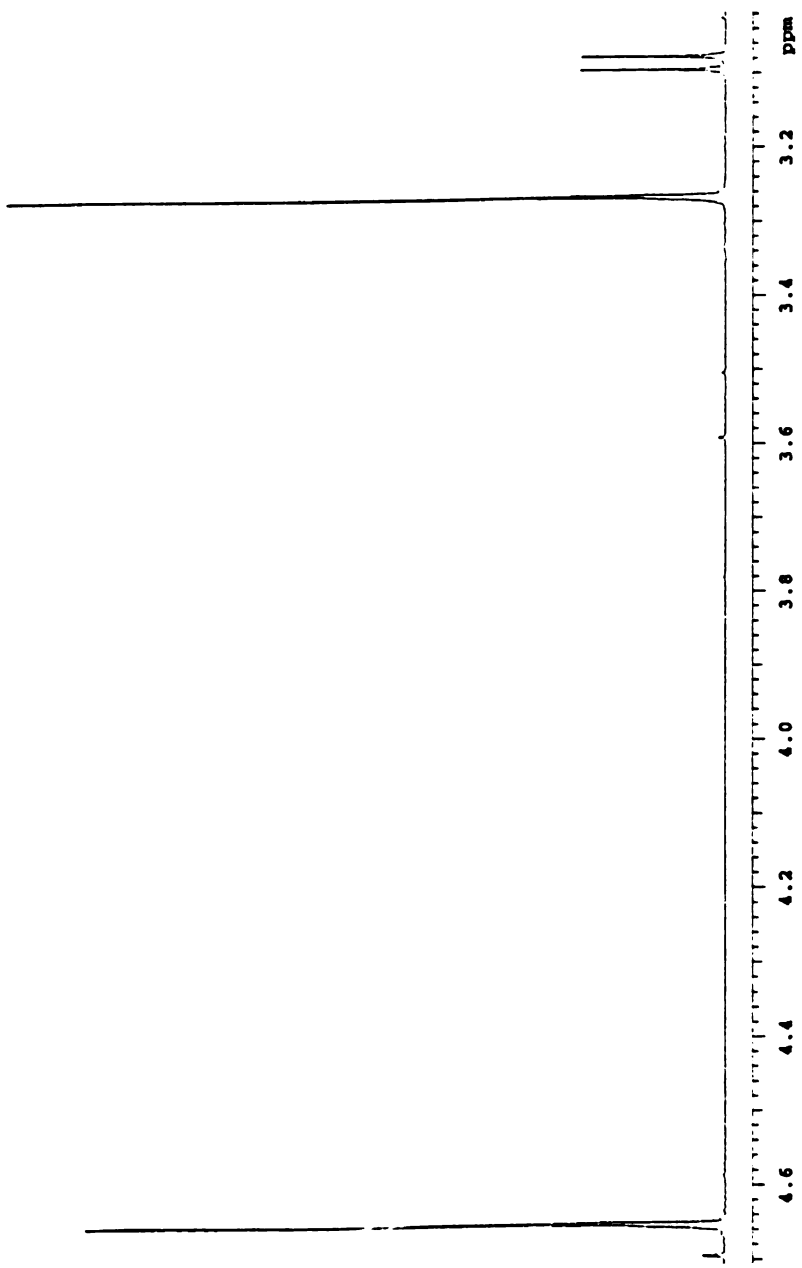
Separations: The HPLC separation was performed on saccharides derivatized by reductive amination with glucose as the initial test molecule. A separation was developed involving three derivatizing reagents: α-aminopyridine, p-nitrobenzylamine, and pyrenemethylamine. A baseline separation was obtained on a reverse phase, ODS, column with a gradient initially consisting of 10 % methanol modified with 0.1 % trifluoroacetic acid and adjusted linearly to 100 % methanol over 75 minutes. The initial flow rate was of 1 ml min^{-1} and decreased to 0.5 ml min^{-1} linearly over 20 minutes then remained at 0.5 ml min^{-1} for the remainder of the separation. Each derivatizing reagent/derivatized glucose pair was separated from the subsequent reagent and its derivatized glucose by ten to twenty minutes. All species were detected by a single wavelength UV detector operating at 240 nm. Confirmation of the elution order was obtained by collecting fractions, and acquiring a full spectrum on a scanning UV/Vis spectrophotometer. The individual derivatizing reagents were also injected and their retention times determined.

In addition to the reverse phase HPLC separation, since all molecules of interest are negatively charged due to the sulfonic acid group, an anion exchange assay was developed. For this separation a diethylaminoethyl (DEAE) column operating at a flow rate of 1.0 ml min^{-1} was used. The mobile phase consisted of 10 mM Na Borate modified with 2 % acetonitrile The pH was lowered to 4.4 using 1 M HCl. Complete separation of sulfoquinovose

from the sulfonic acid of acetaldehyde occurred in fifteen minutes. Refractive index detection was used to detect both species.

Since all of these species are negatively charged, the final set of separations again involved electrophoresis. A complete separation of all synthesized species was obtained by a combination of paper electrophoresis and TLC. The paper electrophoresis was carried out on Watman # 3mm chr paper that was 40 cm long. The mobile phase consisted of 0.5 M ammonium acetate that was saturated with $Ca(OH)_2$, and the electrophoresis occurred at 325 V for 1.5 hours. Bands were detected using silver nitrate to visualize the standards. The standards, and mixtures of the standards of interest, were removed from the paper by cutting and eluted into water. These eluted bands were lyophilized, and the resulting material dissolved in water: methanol (1:1) and spotted onto silica TLC plates. This dimension of separation occurred using a mobile phase of 1-propanol: methanol: ammonium hydroxide: water (1:1:1:1). To detect the resulting bands, the plates were developed with orcinol spray.

Figure 5.5: NMR first reaction product, the sulfonic acid of acetaldehyde dimethylacetal from the starting material of bromo-acetaldehyde-dimethylacetal.


4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
ppm

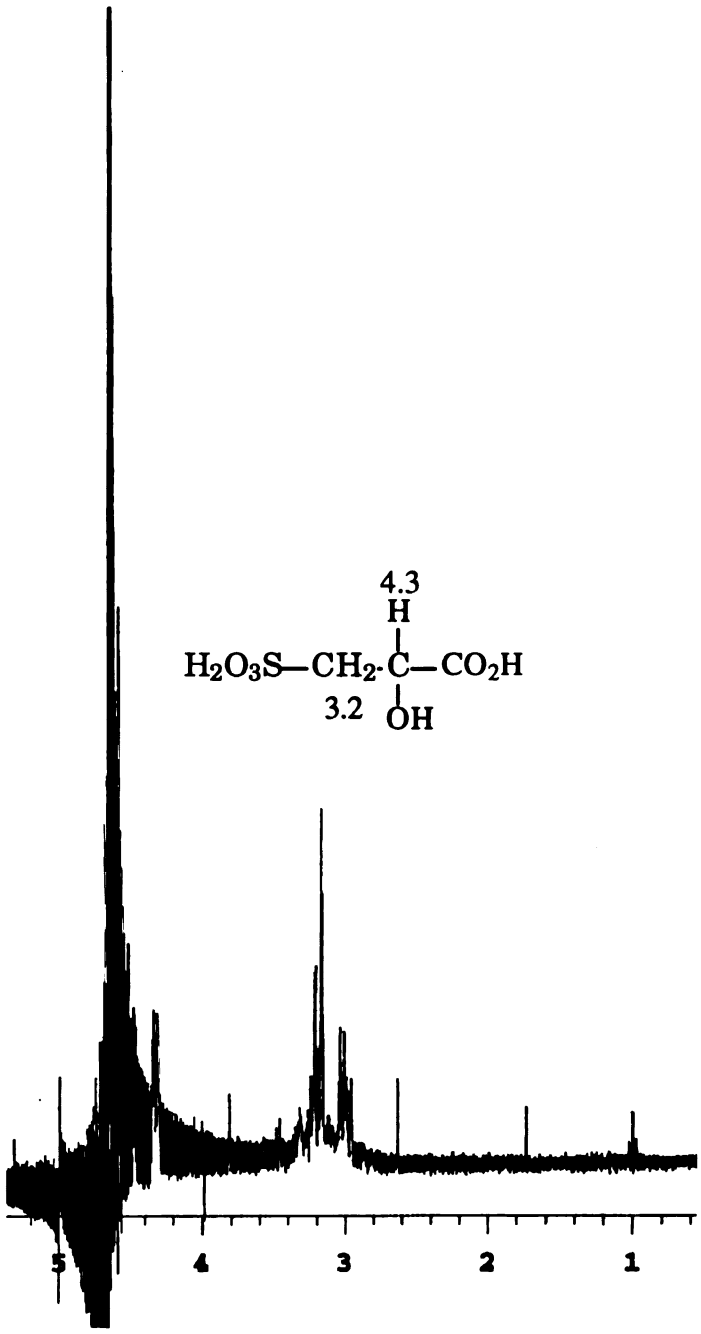


Figure 5.6: The final reaction product of the first series, the sulfonic acid of lactic acid, one of the suspected metabolites of sulfoquinovose. **Formation of the sulfonic acid of lactic acid after the reaction of the sulfonic acid of acetaldehyde with $^{-}$CN and $Ca(CO_3)_2$.**

Results and Discussion:

Organic Reactions: The sulfonic acid of acetaldehyde, a.k.a.: 2-oxo-ethanesulfonic acid, has previously been prepared. Roy and Mewling,[163] synthesized the sulfonic acid of acetaldehyde in several ways. In one instance, the method started with cysteine, which was reacted with barium chloride in aqueous sodium hydroxide. In addition Wang and Sayre.[164] used as starting reagent 2-morpholin-4-yl-ethanesulfonic acid and obtained the same product in 24 hours at room temperature at pH 8.

Since salts of sulfonic acids can be prepared by treatment of primary alkyl halides with sulfite ion[165], our synthetic method starts from 2-bromo-acetaldehyde-dimethyl acetal this by a Strecker type reaction.[166,167] It is clear from the resonances in figure 5.5B that the desired product was obtained. In particular the doublet at 3.13 ppm indicates the presence of a carbon with two protons next to a sulfonic acid group and a carbon with only one proton. In addition, the resonance at 4.75 ppm is a triplet indicating a proton on a carbon adjacent to a carbon with two protons. From the coupling of the resonances, these two carbons must be adjacent to one another. Its chemical shift is reasonable for a carbon next to two acetals. The large singlet gives evidence for the presence of the two acetals at 3.29 ppm, this did not deviate significantly from the starting material as shown in Figure 5.5A. Typically, the formation of a dehalosulfonation requires refluxing conditions.[165] In this instance, likely due to resonance stabilization from the oxygen atoms beta to

the halogen, this reaction occurred at 4° C and was exothermic. In order to remove the dimethyl acetal, the resulting sulfonic acid was subjected to acid hydrolysis. The alkoxy group, -OMe, is not a leaving group, and so must be converted to the conjugate acids before they are hydrolyzed by dilute HCl. The reaction proceeds by an S_N1 mechanism since the carbocation formed is stabilized by resonance through the unshared electron pair on the oxygen adjacent to the cationic center.[164] The proton NMR of the product is similar to the sulfonic acid acetaldehyde dimethylacetal, except for the loss of the resonances attributed to the acetals and the presence of an alcohol resonance at 3.27. This indicates that the aldehyde has been formed (figure 5.6).

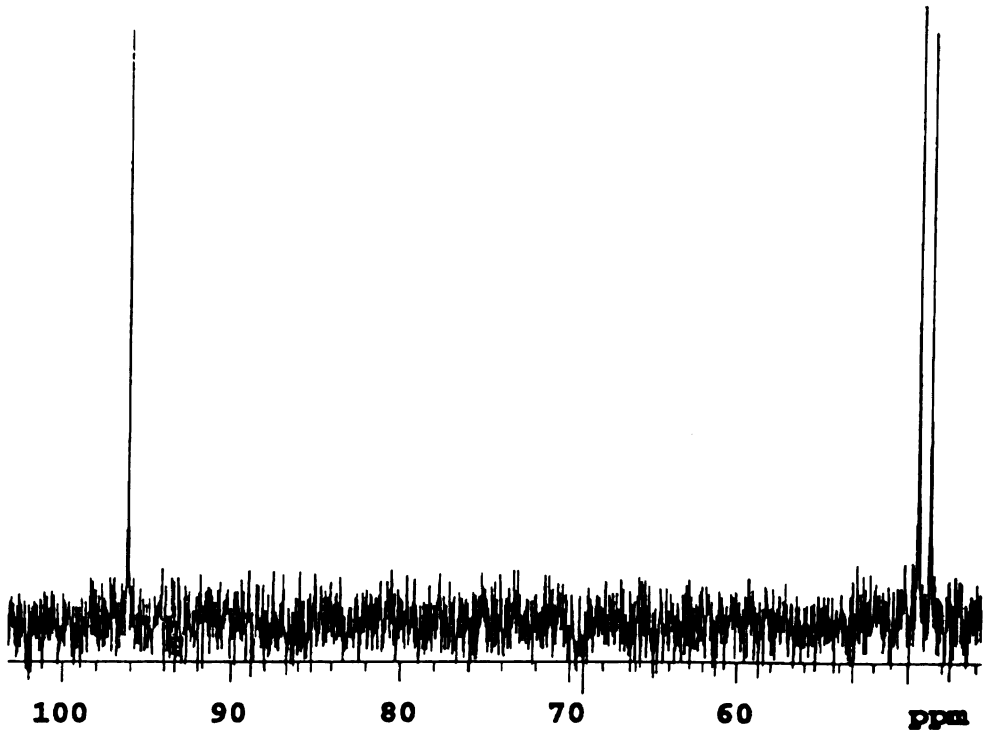


Figure 5.7: The ^{13}C spectrum from the sulfonic acid of acetaldehyde dimethylacetal.

In order to produce the sulfonic acid of lactic acid, the acetaldehyde sulfonic acid first must be converted to the cyanohydrin by addition of HCN. This reaction occurs very slowly when pure HCN is used, but rapidly when cyanide ion, CN^-, is added. This is because HCN is a weak acid ($pK_a = 9.1$) and therefore is not significantly dissociated. The cyanide anion, however, is strongly nucleophilic, and addition occurs by a nucleophilic addition pathway. Protonation of the anionic tetrahedral intermediate yields the stable tetrahedral cyanohydrin product plus regenerated cyanide ion[168]. Cyanohydrin formation is useful because of the further chemistry that can be carried out. These functional groups can be hydrolyzed to yield carboxylic acids, thus providing a method for transforming an aldehyde into a different functional group while lengthening the carbon chain by one unit.

Figure 5.8: NMR of the synthesized SQ-Me from the starting material of the glycopyranose methylglycoside.

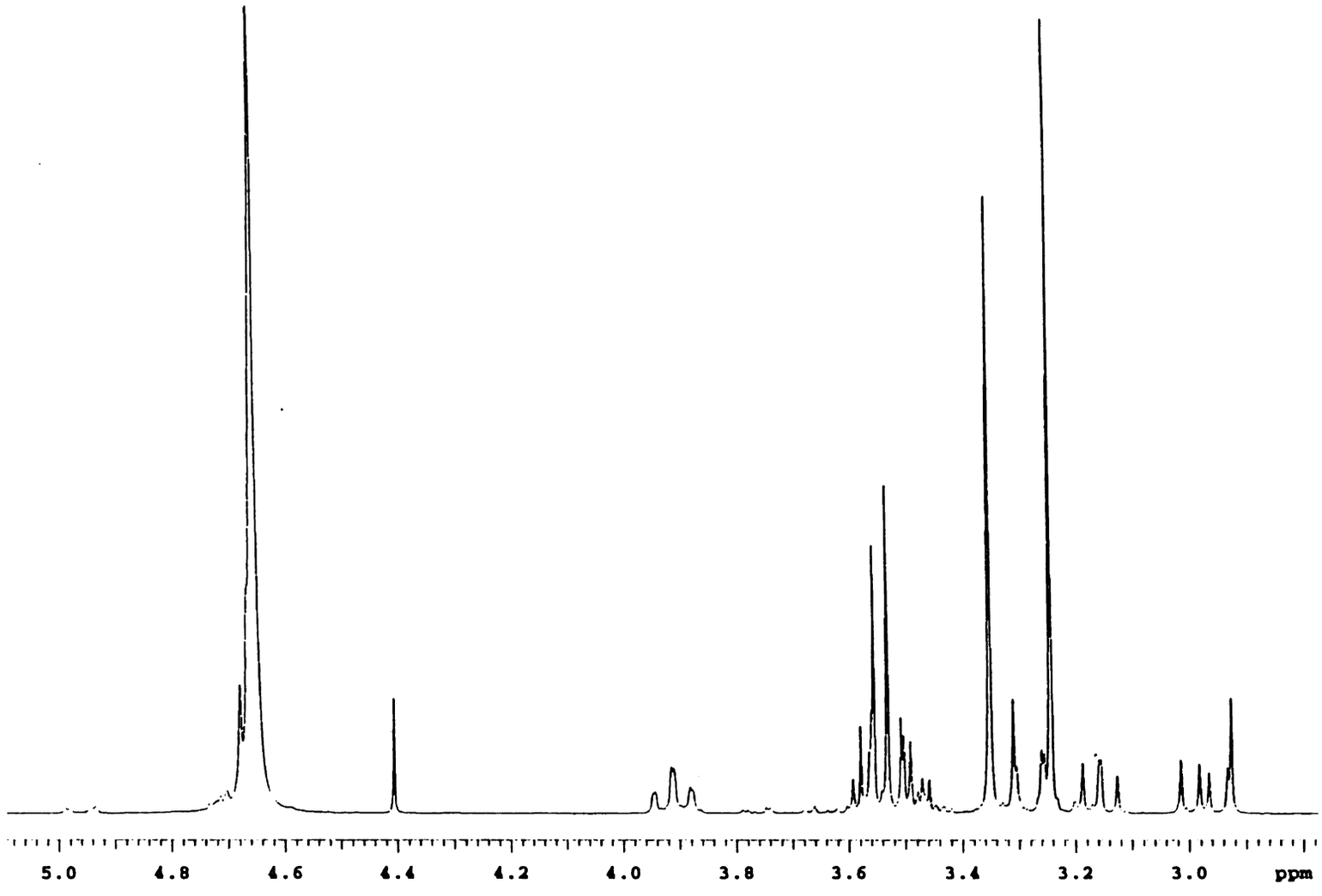
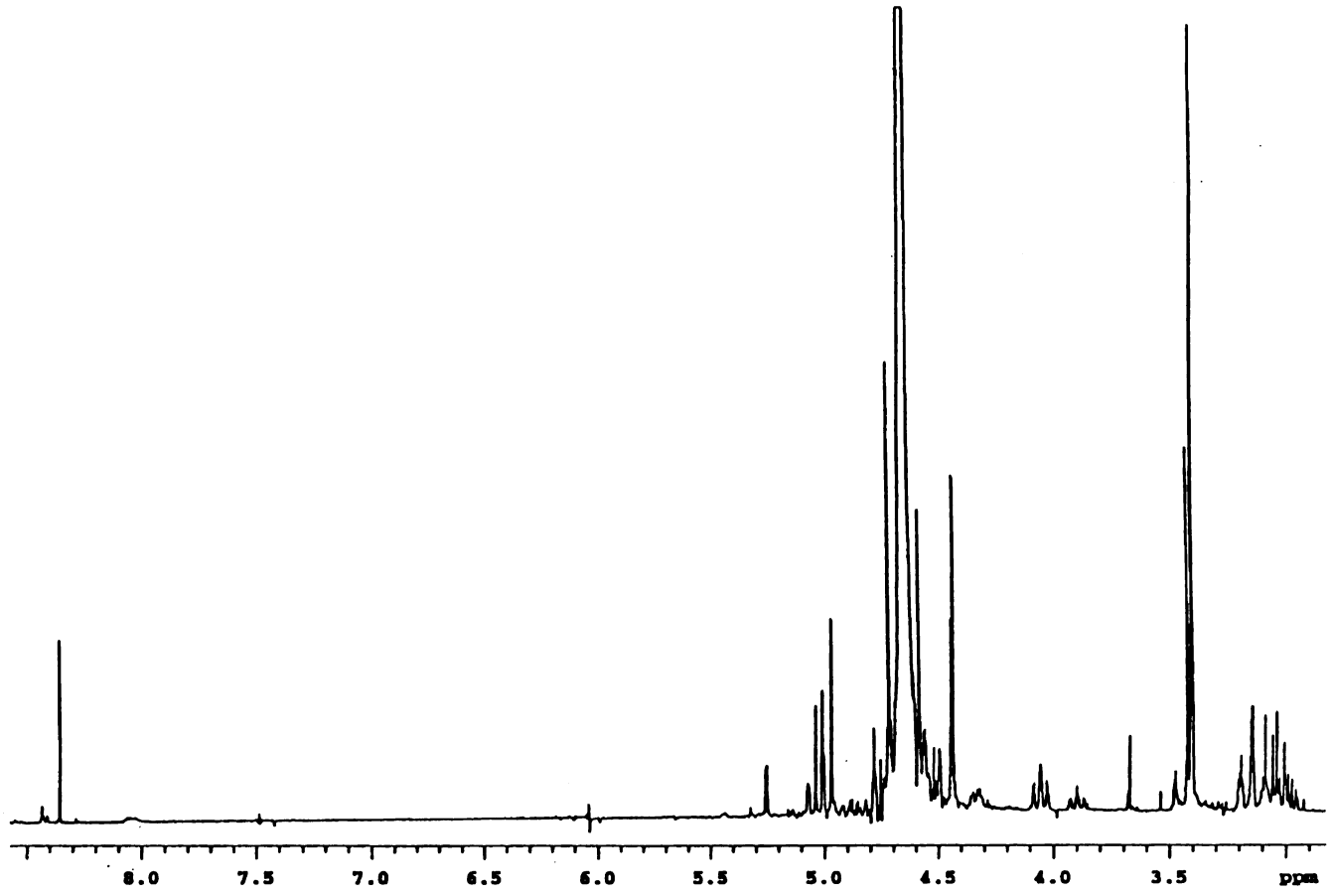
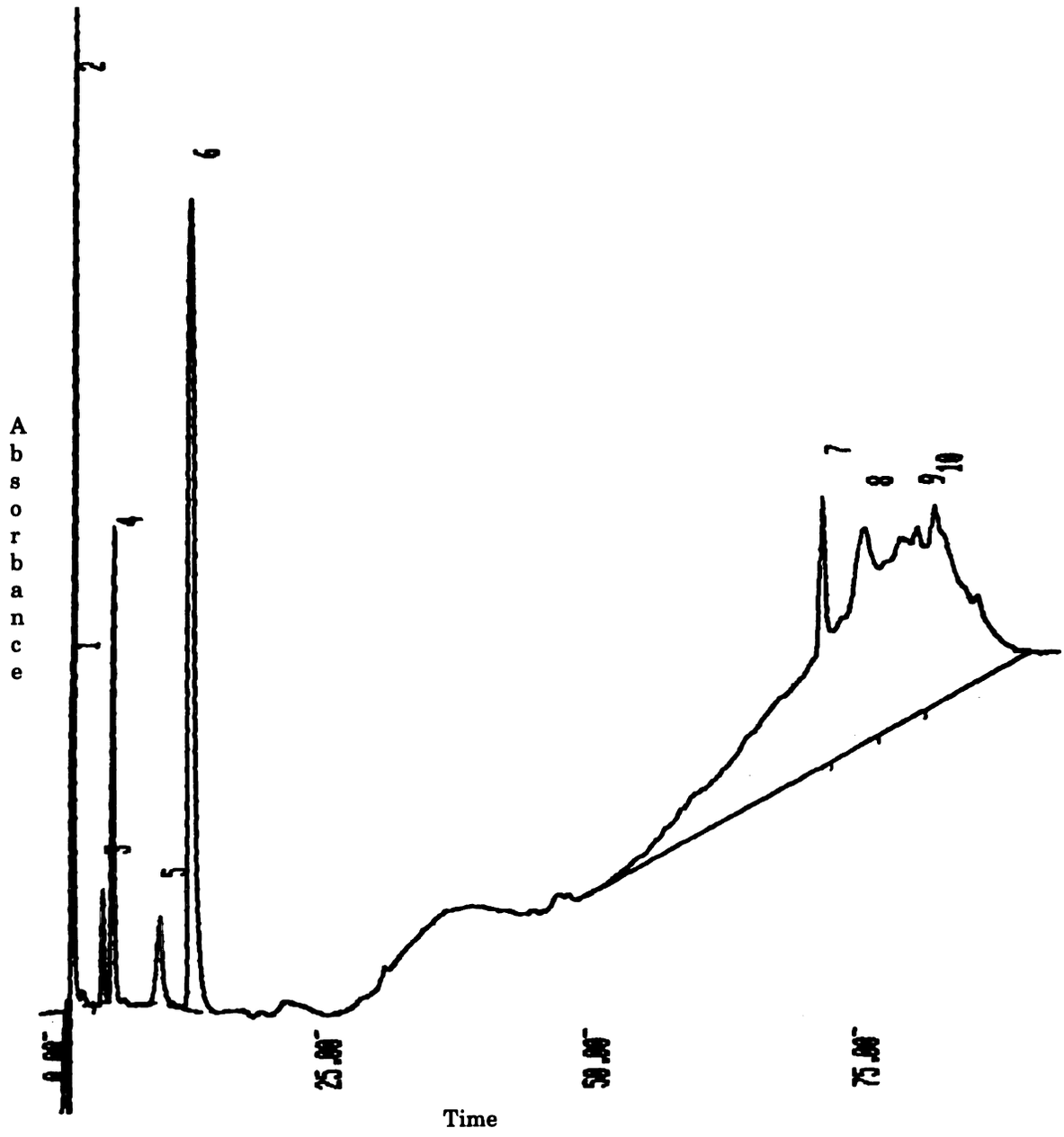

5.0
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
ppm

Figure 5.9: NMR of the reaction product of SQ-Me after periodate oxidation, the formation of the sulfonic acid of propanal-2-ol-3-sulfonic acid.


8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
ppm

The reaction, to form sulfoquinovose, proceeds in two steps, intermediate formation and intermediate decomposition and involves three components. Intermediate formation involves the formation of the TPP-Br complex from TPP and CBr_4. An ionic mechanism is responsible for the formation of the TPP-Br complex. It is initiated by the polarizing action of the permanent dipole of phosphane on the symmetrical but readily polarizable CBr_4, and it leads to heterolytic bond cleavage by a direct interaction between the phosphorus atom and the carbon bromine bond.[169,170] If the –OH groups available for the reaction are chiral intermediate decomposition occurs with extensive inversion of configuration,. The final component of this system is α-methylglucopyranose. Since this is the methyl-glycoside this system can no longer enter into the straight chain configuration (Figure 1.1), and the only alcohol group present that can undergo halogenation is on carbon six. All other hydroxyls are blocked from the reaction by the closed ring. The intermediate decomposition with the TPP-Br complex is thus a method of converting α-methylglucopyranose into 6-bromo-α-methylglucopyranose that can be achieved under very mild conditions. The sulfonato-de-halogenation reaction, the Strecker reaction, to form sulfoquinovose from 6-bromo-α-methylglucopyranose has previously been discussed.[165]

Figure 5.10: Reverse phase HPLC separation with absorbance detection at 240 nm of α-aminopyridine, p-nitrobenzylamine, and pyrenemethylamine after reductive amination using glucose as the test molecule.


Absorbance
Time
25.00
50.00
75.00
1
2
3
4
5
6
7
8
9
10

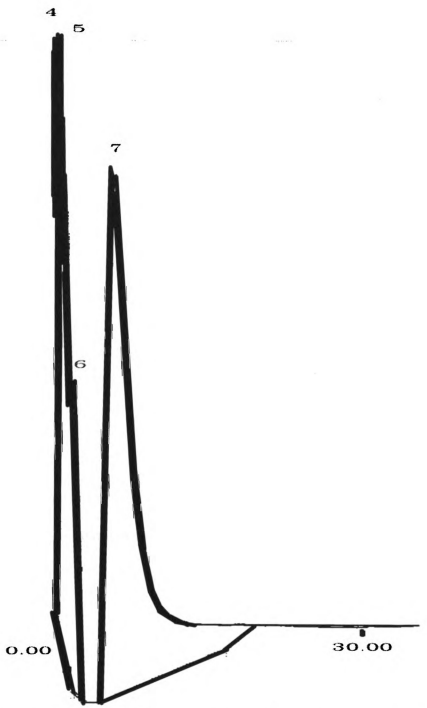


Figure 5.11: Anion exchange separation of SQ and the sulfonic acid of acetaldehyde.

Separations: As mentioned earlier, many HPLC methods have been developed for the separation of carbohydrates. These include separations that are specific for individual monomers, as well as oligomers, polymers, and charged species. Detectors for these separations have included refractive index, amperometry, and mass spectrometry. UV/Visible absorption detectors require derivatization to increase the sensitivity of the assay. Multiple derivatizations, to increase sensitivity as well as selectivity, have not been undertaken in the HPLC separation of carbohydrates and their metabolites. The test molecule in this study was glucose, an aldose that undergoes reductive amination. This technique was later extended to sulfoquinovose, one of the target molecules. The derivatizing reagents that were chosen are all primary amines with conjugated ring systems. They were α-aminopyridine, p-nitrobenzylamine, and pyrenemethylamine. It has been previously mentioned that saccharides and modified saccharides require more than one separation dimension in order to assure accurate and precise identification. Reductive amination has been discussed previously and will not be reintroduced here (p. 36 and Figure 2.1). The order of elution for this separation was confirmed by several techniques. First, derivatizing reagents were injected separately to compare retention times. Then to confirm that there were no strong interactions between the derivatizing reagents they were co-injected. Finally, fractions were collected and a full UV/Vis spectrum was obtained using a scanning spectrophotometer. Fractions were then

collected to determine the order of elution, whether the derivatized saccharide eluted first or the derivatizing reagent. These fractions were subjected to phenol/sulfuric acid to test for the presence of sugars. In addition, the retention time of the individual derivatizing reagents vs. the derivatized glucose were compared. These analyses determined that the derivitized saccharide was the first to elute in each pair of peaks. A typical chromatogram is shown in Figure 5.11, the peaks labeled 3 and 4 correspond to those for α-aminopyridine, 5 and 6 correspond to the derivatized glucose and excess derivatizing reagent of p-nitrobenzylamine, and the peaks labeled 7 and 8 correspond to those for pyrenemethylamine.

In addition to reverse phase HPLC, an ion exchange chromatographic assay was developed (Figure 5.12) since all of the species of interest are negatively charged. Anion-exchange chromatography of innately anionic carbohydrates have been performed previously.[171] Typically alkaline solutions are necessary for elution of carbohydrates in HPAEC. Sodium hydroxide is the base of choice at a concentration of 15-20 mM for monosaccharides.[146, 170] For the purpose of this separation sodium hydroxide could not be used since NaOH has been shown to cause rearrangements upon drying of the sample, this would be unacceptable since further analysis with NMR is required. The requirement for complete structural analysis with NMR also precluded the use of pulsed amperometric detection. Although sensitivity and selectivity is greater for amperometric detection compared to

refractive index detectors, the technique may be destructive due to redox reactions.

With the above limitations in mind, the following assay was developed. For the ions of interest for this assay a diethylaminoethyl (DEAE) column was used to separate the two most different molecules of interest. The mobile phase chosen to separate the molecules of interest consisted of 10 mM Na Borate modified with 2% acetonitrile at a pH of 4.4. This not only allowed for separation based on charge density, but also allowed for separation based on the size of the molecules. It is also advantageous that the separation was obtained using a borate buffer system, in that removal of buffer could be obtained by subjecting samples to dry ice prior to passing samples through a cation exchange resin. Choosing sulfoquinovose and the sulfonic acid of acetaldehyde a complete separation was obtained in fifteen minutes. A refractive index detector was used to detect eluting peaks. It was determined that the size of the molecules was critical for the separation, experiments that altered the pH were undertaken. As the pK_a of the sulfonic acid group ($pKa \approx 1.92$) was approached, the retention time of the two species should have increased, due to increased interaction with the stationary phase. The hydrogen ion concentration at $pH \geq 5$ is critical and the separation is altered with changes in pH as small as 0.2 units. Below pH 4.5 it is no longer a factor and there is little change in the separation to pH 2. In order to confirm the order of elution, the standards were injected individually. Also, fractions

were collected, and desalted, for NMR analysis. It was confirmed that the first eluting band was sulfoquinovose, or the methylglycoside, and the second eluting band was the sulfonic acid of acetaldehyde.

The final separations for the detection of sulfoquinovose and the suspected metabolites were paper electrophoresis and TLC. Paper electrophoresis was chosen, as was ion exchange chromatography, because all of the species of interest are negatively charged and any neutral or positively charged species would not interfere with the assay. Ammonium acetate was chosen as the buffer since it is volatile and, as repeatedly mentioned, a second dimension of separation would be required. In addition, calcium hydroxide was added. Although, $Ca(OH)_2$ is not volatile, it did not interfere with the second separation. Also, Ca^{2+} chelates well with the hydroxyls on sugars. The separation occurred in 1.5 hours at 325 V on a water cooled stage. Here two dimensions of separation are required. Since the separation is based on charge density, any phosphorylated or sulfated sugar would produce the same migration in electrophoresis. Also, considering the low molecular weight of the expected metabolites a complete separation in the first dimension is not expected. To develop the separation, electrophoresis was done on the standards individually and in a mixture. The migration distances for the sulfonic acid of acetaldehyde and the sulfonic acid of lactic acid were the same, 4 cm. For SQ and the sulfonic acid of propanal-2-ol-3-sulfonic acid, the migration distance was 9 cm. Once the separation was

complete the individual standard bands were visualized using silver nitrate solution. The bands were re-aligned with the lane containing a mixture of standards, and the individual bands were cut out, eluted into water, frozen, and finally lyophilized. The resulting solid was taken up in a water: methanol solution (1:1) and introduced onto TLC plates. Complete separation occurred with a mobile phase of 1-propanol: methanol: ammonium hydroxide: water (1:1:1:1). Individual bands were visualized using orcinol spray. The final R_f values of the individual standards were 0.31, 0.43, and 0.76 for the sulfonic acids of acetaldehyde, lactic acid, propanal-2-ol-3-sulfonic acid and for SQ, respectively.

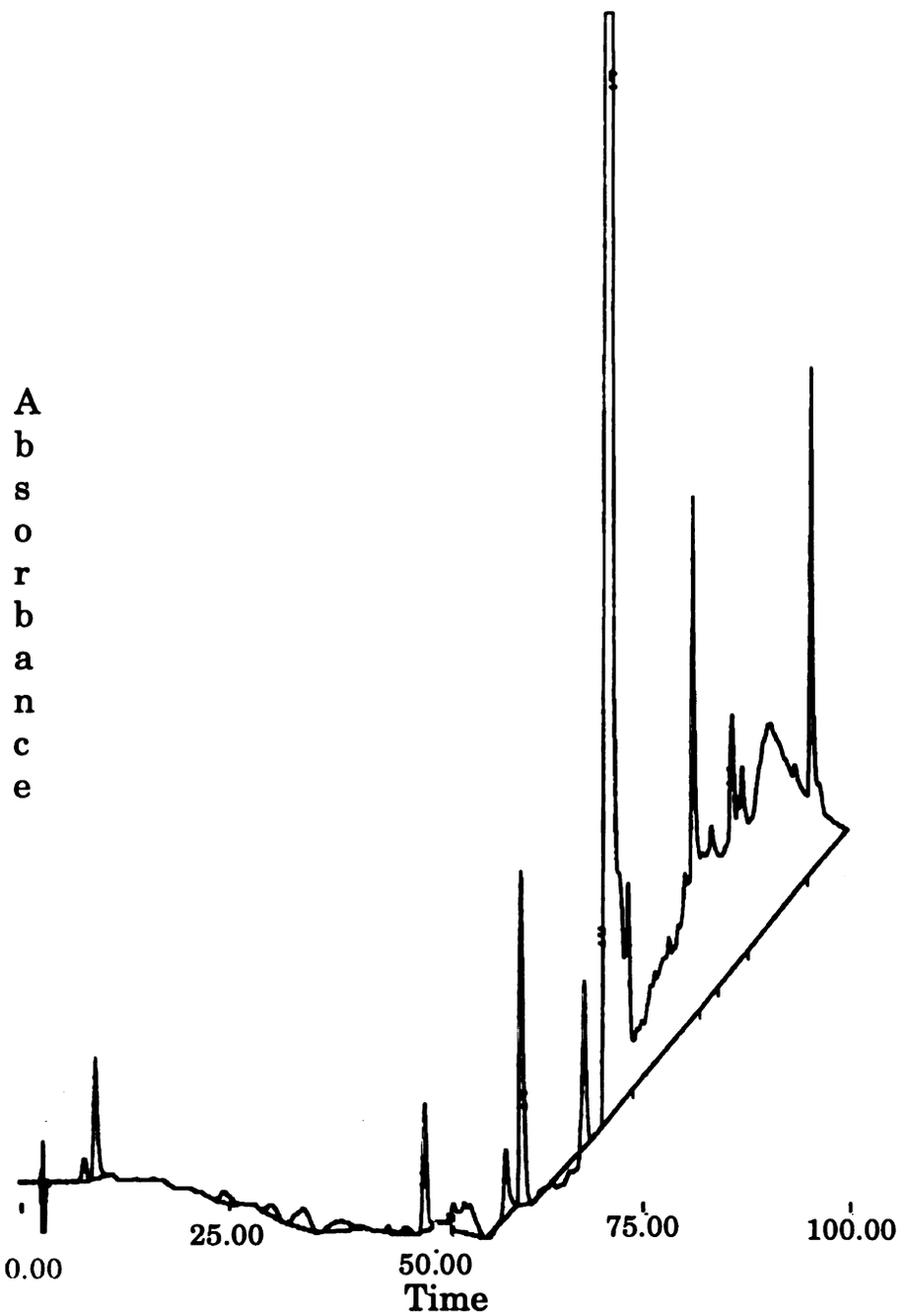


Figure 5.12: Separation of SQ derivatized with α-aminopyridine, p-nitrobenzylamine, and pyrenemethylamine.

Conclusion: Although the target molecules (standards) were all successfully synthesized, and each separation reached its individual goal, the final goal of obtaining a "gold standard" by which the pathway to determining sulfoquinovose biosynthesis was not achieved. The reverse phase HPLC assay is both sensitive and selective using a diode array detector, and was utilized in the development of the assay. The separation, although developed for glucose also works well for sulfoquinovose (Figure 5.13). The shortcoming is that not all of the suspected metabolites, from either proposed pathway, will undergo reductive amination. In addition, all other reducing sugars will undergo reductive amination, which complicates the interpretation of the chromatogram when attempting to identify sulfoquinovose and the detectable metabolites.

Ion exchange chromatography with refractive index detection is not as sensitive an assay as the reverse phase chromatography described above, however all positively charged and neutral species are removed from the sample in the void volume, and the negatively charged species are separated. The molecules of interest can be collected and structural analysis can be performed on these fractions after removing the salts. The shortcoming here is that the detector is not selective. Overlapping bands may go unnoticed. The option again is to include a second dimension of detection, and here NMR was chosen. Due to sample loading limitations of the column, multiple

injections to collect fractions are required to obtain sufficient material for NMR analysis.

Many two dimensional separations have been developed for TLC. Using paper electrophoresis as the first dimension allows for many neutral and positively charged species, such as other saccharides that are not of interest, to be removed. This allows for a second, or even third, dimension that is less complicated to interpret. By using preparative scale TLC, a sample that is possibly more pure may be obtained. Here, again, the difficulty is with sample loading limitations in the first dimension. Due to diffusive broadening of the bands, the initial spot must be small and concentration effects cause the migration of the sample to be altered. The standards would have to be spotted onto the paper at approximately the same concentrations as an unknown to assure that migration distances would correlate.

Acknowledgements: I would like to thank Dr. Hollingsworth for the synthesis of sulfoquinovose and the sulfonic acid of lactic acid. I would also like to thank him for his assistance in removal of the methyl acetals from the sulfonic acid of acetaldehyde dimethylacetal.

Chapter 6

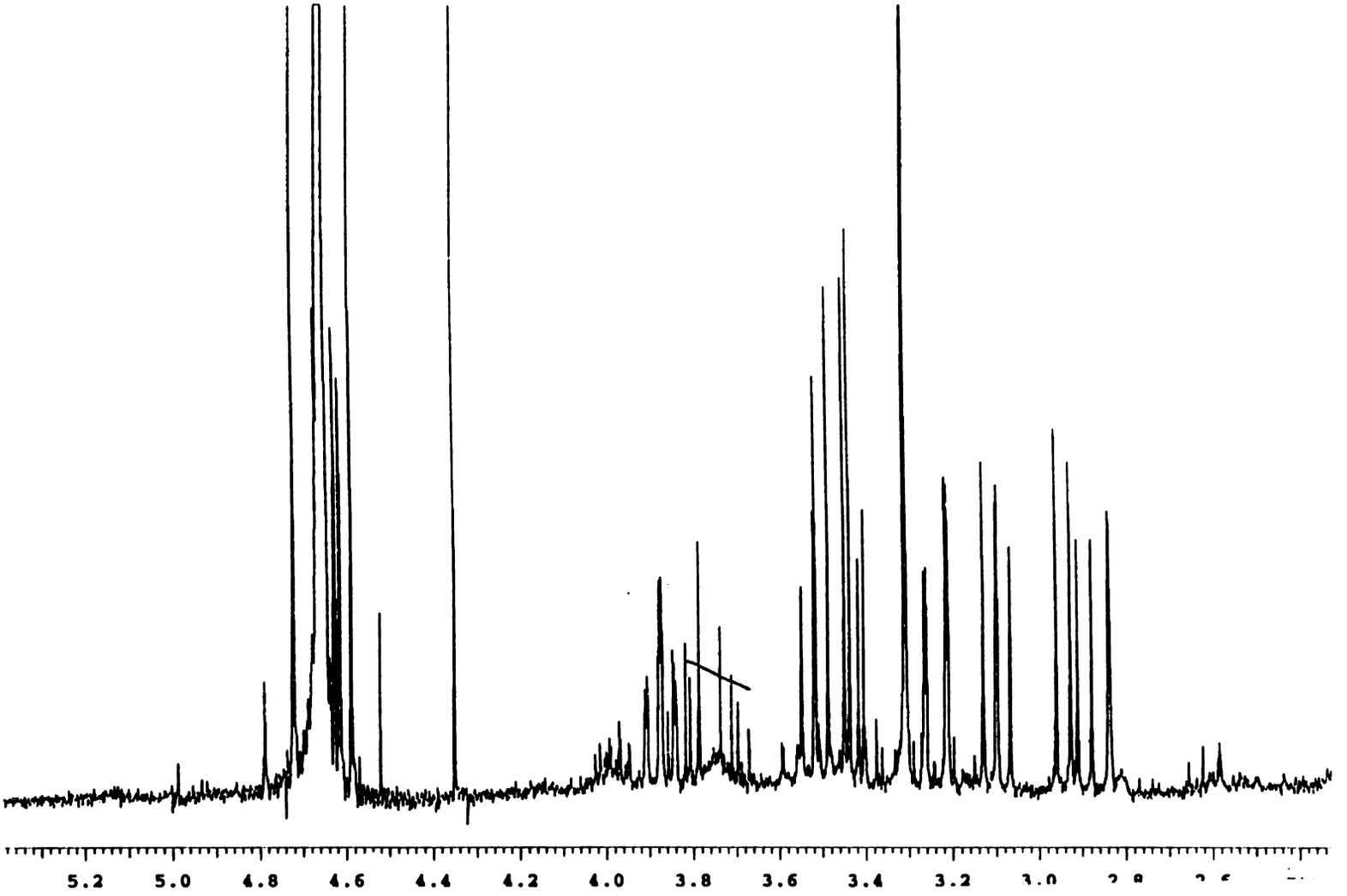
EXTRACTION AND STRUCTURE DETERMINATION OF SULFOQUINOVOSE

Introduction: One important class of biomolecules is the oligosaccharides. These can be present in nature as free carbohydrates or as glycoconjugates (e.g., glycoproteins and glycolipids). The challenge in the separation and analysis of these materials is that they do not contain strong chromophores for absorption in the UV or visible region. They are also highly heterogeneous in terms of their degree of polymerization, base-unit composition, and even molecular shape (i.e., functional group modifications). The bioanalysis of oligosaccharides has been demonstrated in ion-pairing chromatography by conductivity detection, in ion-exchange chromatography by pulsed amperometric detection, in supercritical fluid chromatography by mass spectrometry, and by direct measurement using matrix assisted laser desorption mass spectrometry and fast-atom bombardment mass spectrometry. Mass spectrometry alone, particularly soft ionization techniques, cannot be used for complete structure determination of molecules. For structure determination, carbon and proton NMR are invaluable, although the amounts of sample required and relative purity of the sample necessary for interpretation are greater than for mass spectral analysis.

Materials and Methods: R. trifolii 1021 was grown at 29° C in 4l shaken flasks containing 2l of modified Bergensen's, B III, media each.[172] The cultures were supplemented with either the methyl glycoside of sulfoquinovose, as the carbon and sulfur source or $H_2{}^{35}SO_3$ as the sulfur source. To acclimate the bacteria to sulfoquinovose, as the carbon and sulfur source, starter cultures were used. Test tubes containing 5 ml of sterile media were serially diluted to 10 ml. The bacteria in each tube were incubated until late stationary phase before being transferred to the next tube, thereby increasing the concentration of SQ the cells were exposed to. Once acclimated to SQ-Me, these cells were then grown in two litre batches until they were in late stationary phase and were then harvested by centrifugation at 7.5 krpm for 25 minutes on a Sorvall RC2-B centrifuge equipped with a GSA rotor. The supernatant was removed and the resulting pellet was extracted by stirring with a mixture of ethanol: water (1:1) for three hours at 4° C. Cells were then centrifuged again to obtain a pellet and the supernatant was passed through a cation exchange column. The column was then washed with a mobile phase of ethanol: water (1:1, 500 ml) and the eluent collected. All fractions were evaporated on a rotary evaporator and redissolved in a solution of propanol: methanol: ammonium hydroxide (3:1:1). This was passed through a silica column (6x1.5 inches), and 5 ml fractions were collected. These fractions were spotted onto silica TLC plates and

developed using a mobile phase of 1-propanol: methanol: ammonium hydroxide: water (1:1:1:1). Detection of the spots of interest was obtained by spraying with orcinol and heating, in the instance of the bacteria grown on SQ-Me, and by use of Autorad detection of radioactivity, for ^{35}S. Fractions with similar purity were then combined and NMR analysis performed.

Figure 6.1: First eluting fraction after sample was purified from R. 1021.


5.2
5.0
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2

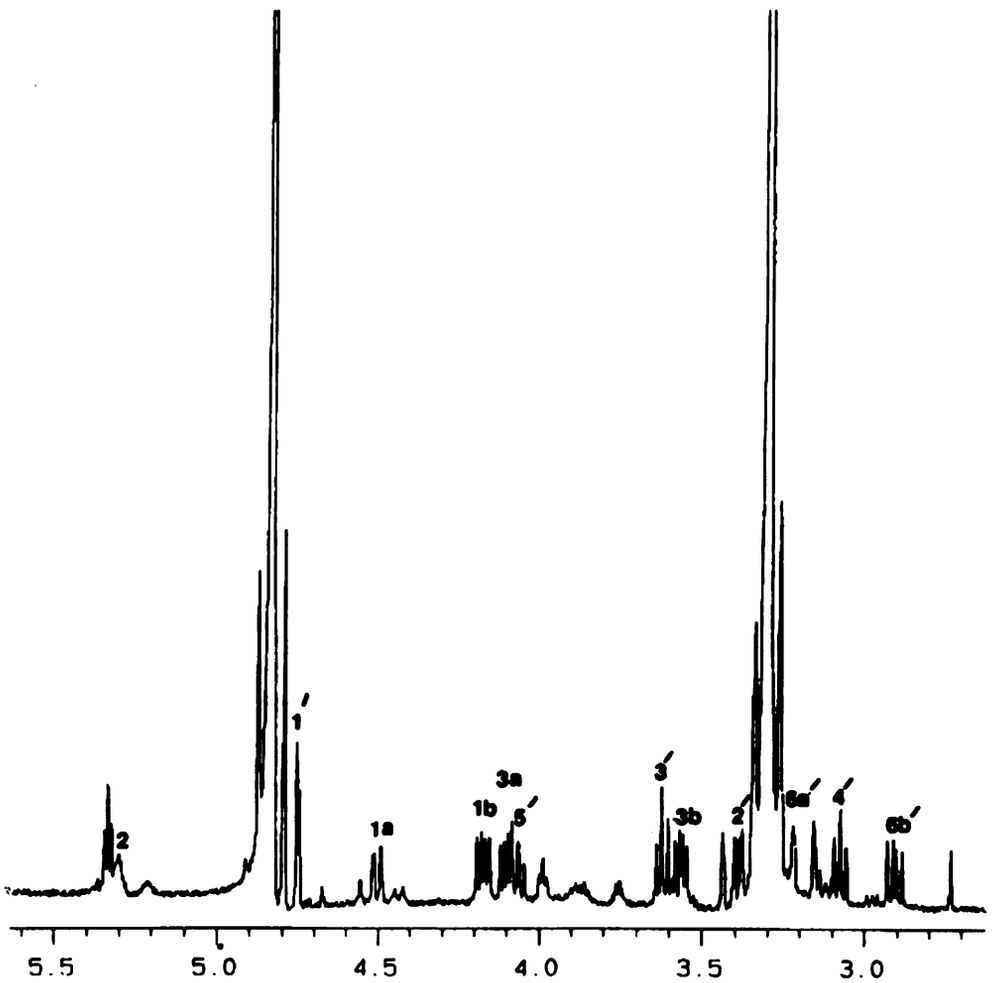


Figure 6.2: NMR of SQDG from the published work of Cedergren et al.

Results and Discussion: In order to harvest sufficient material for NMR analysis, none of the separations developed in the previous chapters was amenable to scale up. Once the bacteria had been lysed and centrifuged, to obtain a crude extract, which was passed through a cation exchange column. Most cations are removed, and the column does not retain neutral species or anions. An aliquot of the eluent was analyzed by NMR to determine whether molecules relevant to sulfoquinovose metabolism were

present in abundance. This was to be determined by comparing the crude extract to the synthetic standards. Further purification was then undertaken. The procedure used consisted of a silica flash column and an aqueous mobile phase of 1-propanol: methanol: ammonium hydroxide (3:1:1). This first dimension of separation resulted in several overlapping bands, determined by following the movement of ^{35}S labeled peaks that were obtained from a second set of incubations of R. 1021. By running standards alongside the unknown sample on the same TLC plates, many of the R_f values corresponded to those from the bacterial samples. In Figure 6.1 as compared to Figure 6.2, it was determined that sulfoquinovose is present as a major negatively charged component obtained from the biological samples and that the resonances from the NMR analysis of the column chromatography correspond well to those obtained previously from preparative TLC. In particular, all of the resonances corresponding to the carbohydrate protons of SQ are present, and at similar frequencies. From the data presented in Figures 6.1 and 6.2, the published spectrum from Cedergren et al.[173] of SQDG, as compared to Figure 5.9 from the synthesis of the methylpyranoside of sulfoquinovose, it is apparent that the synthesized material has been converted into sulfoquinovose, and, from the TLC retention factors, that the ^{35}S has been taken up by the bacteria and included in its metabolism.

Conclusion: After more than forty years, the question still remains as to the biosynthesis of sulfoquinovose. The answer lies in obtaining sufficient material to obtain 2-D NMR and definitive proof for the structure of all metabolites. In addition, the enzymes responsible must be isolated and the reactions proven in vitro, so that no questions remain. Glycolysis and gluconeogenesis are two of the most well known examples of biosynthetic pathways. If sulfoquinovose is synthesized using this pathway, perhaps many of the enzymes involved are also involved in the synthesis of SQ. General forms of all of these enzymes are available commercially.

The specific nutritional requirements for different microorganisms vary greatly. For example, heterotrophic microorganisms require organic matter, but autotrophic microorganisms do not. Some microorganisms are fastidious, having complex specific growth factor requirements, and therefore, these microorganisms are able to reproduce only under greatly restricted conditions. Other microorganisms are capable of reproducing under less restricted conditions. Some microorganisms have simple nutritional requirements and are able to reproduce by using a single carbon source as their growth substrate. Such organisms have the metabolic capability of interconnecting the various biochemical pathways so that a particular carbon compound can be transformed into the carbon backbones of all the required macromolecular constituents of those organisms. The fact that SQ and UDP-glucose can be used as substrates for growing Rhizobia is not surprising.

These organisms have been utilizing the carbon sources available for growth. The ability of Rhizobia to survive requires its adaptability and this extends to its sources for food.

The fact that overlapping bands were obtained in the separation of sulfoquinovose and its metabolites reinforces the notion that multiple separations are required to obtain pure samples for analysis. One of these bands was sulfoquinovose (Figure 6.1) another has been tentatively identified as a sugar, although from the TLC it has the same R_f as UDP-glucose (Atima), it is not likely this compound. Since many of the assays for SQ and metabolites utilize TLC solely for identification based on R_f and the comparison with standards, this data indicates that this practice is prone to errors. Final identification of SQ and its metabolites requires obtaining sufficient material for detailed structural analysis. This will require that larger batch sizes, likely more than ten times the volume used in this study. Following the metabolism using ^{35}S should be performed on a smaller scale, so that the elution of metabolites may be monitored. But for structural analysis of metabolites, the availability of larger quantities of material is crucial. This is especially true when the material, in the absence of mutants where intermediates build up, is present in only minute quantities. In addition cell free, in vitro, experiments need to be conducted. With the wealth of knowledge available on gluconeogenesis and glycolysis that have been accumulating for the past hundred 150 years, from Pastuer's initial

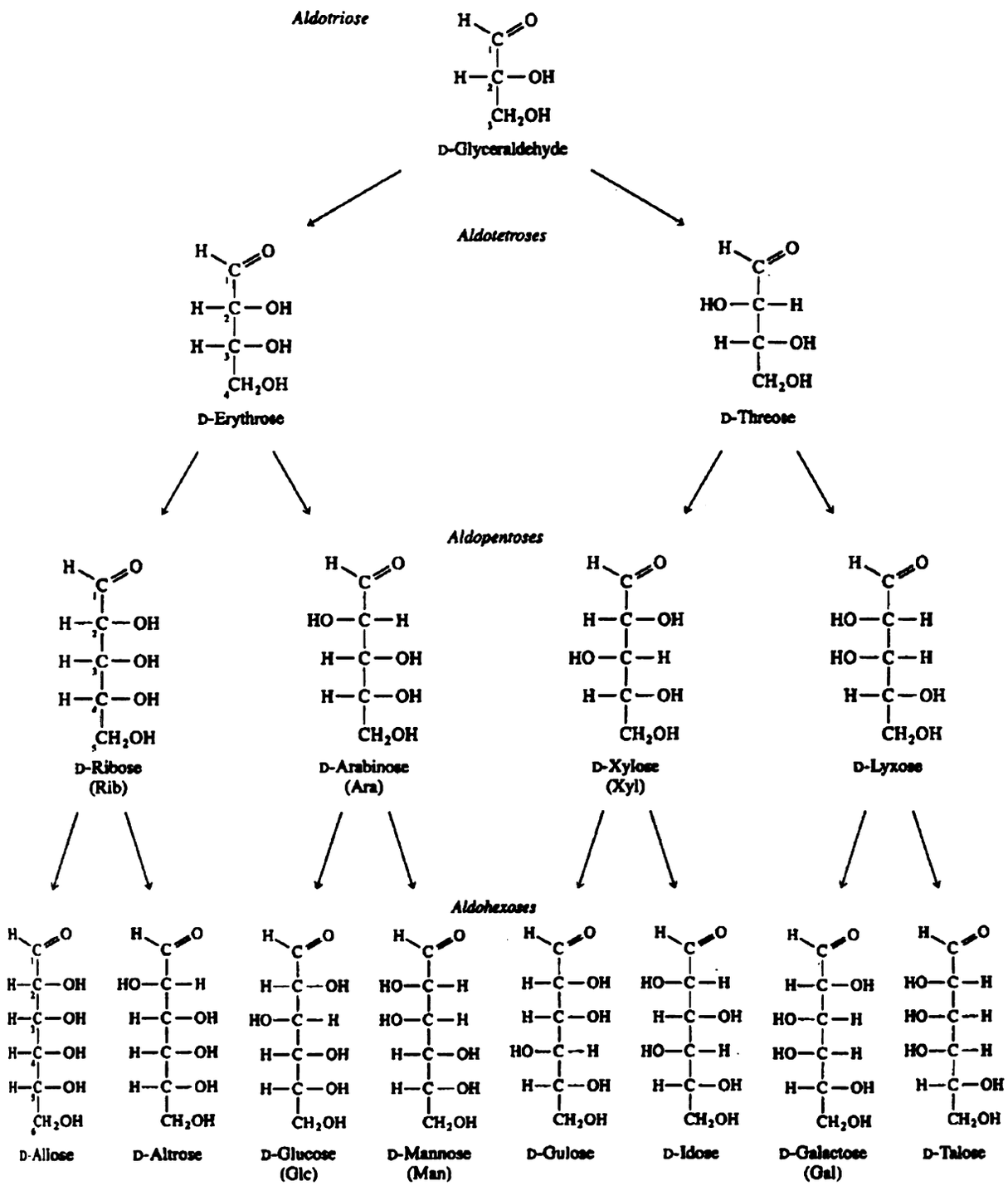
experiments in determining the mechanisms of fermentation to the present, if the suspected metabolites follow this pathway the in vitro experiments are a logical next step.

Acknowledgment: I would like to thank Dr. Hollingsworth for his help in acclimating the bacteria to sulfoquinovose. Atima Sharma's help in starting and maintaining the bacterial cultures was invaluable.

APPENDICES

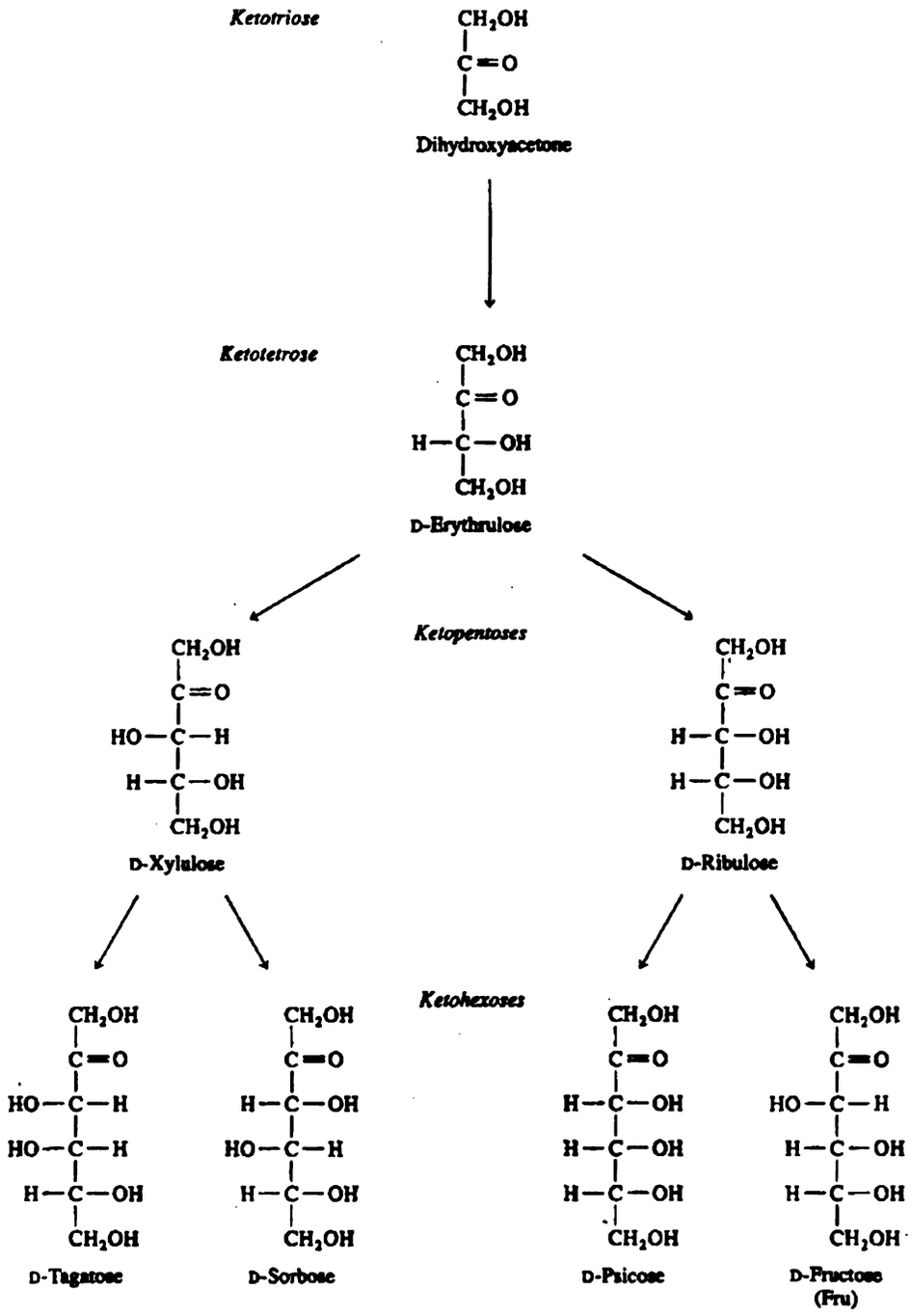
APPENDIX A

THE D-ALDOSES



APPENDIX B

THE D-KETOSES



APPENDIX C

AMINO ACIDS AND THEIR MASSES

Amino Acid Group	Amino Acid	3 (1) Letter Code	Elemental Composition	Mono-isotopic Mass	Aver. Mass
Hydrophobic (nonpolar) R Groups	Alanine	Ala (A)	C_3H_5NO	71.03711	71.08
	Valine	Val (V)	C_5H_9NO	99.06841	99.13
	Leucine	Leu (L)	$C_6H_{11}NO$	113.08406	113.16
	Proline	Pro (P)	C_5H_7NO	97.05276	97.12
	Phenylalanine	Phe (F)	C_9H_9NO	147.06841	147.18
	Tryptophan	Trp (W)	$C_{11}H_{10}N_2O$	186.07931	186.21
	Methionine	Met (M)	C_5H_9NOS	131.04049	131.20
Hydrophilic (polar) R groups	Glycine	Gly (G)	C_2H_3NO	57.02146	57.05
	Serine	Ser (S)	$C_3H_5NO_2$	87.03203	87.08
	Threonine	Thr (T)	$C_4H_7NO_2$	101.04768	101.11
	Cysteine	Cys (C)	C_3H_5NOS	103.00919	103.15
	Asparagine	Asn (N)	$C_6H_{12}N_4O$	156.10111	156.19
	Glutamine	Gln (Q)	$C_5H_8N_2O2$	128.05858	128.13
	Tyrosine	Tyr (Y)	$C_9H_9NO_2$	163.06333	163.18

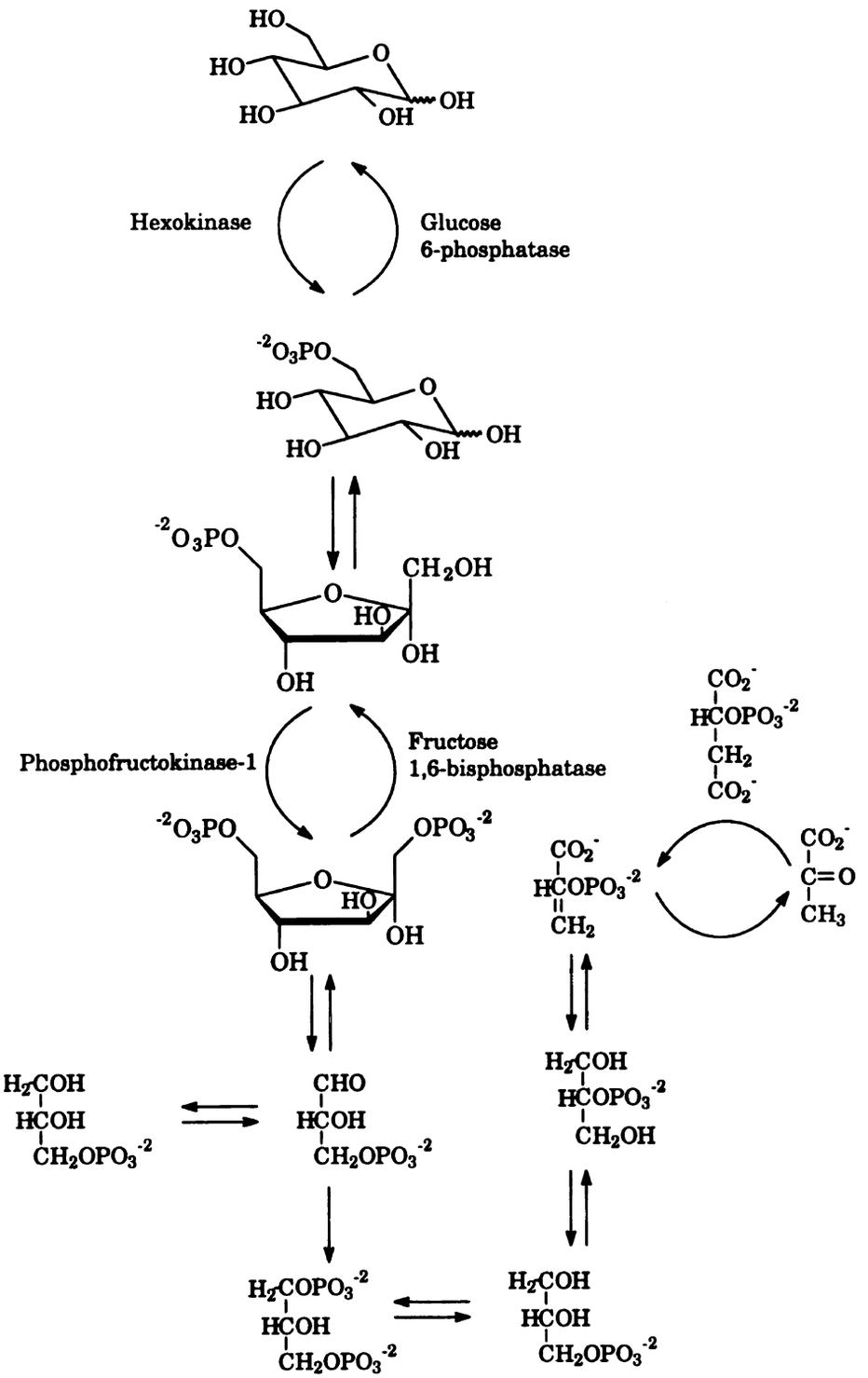
APPENDIX C (cont.)

AMINO ACIDS AND THEIR MASSES

Amino Acid Group	Amino Acid	3 (1) Letter Code	Elemental Composition	Mono-isotopic Mass	Aver. Mass
Basic R groups	Histidine	His (H)	$C_6H_7N_2O$	137.05891	137.14
	Lysine	Lys (K)	$C_6H_{12}N_2O$	128.09496	128.17
	Arginine	Arg (r)	$C_6H_{12}N_4O$	156.10111	156.19
Acidic R groups	Aspartic Acid	Asp (D)	$C_4H_5NO_3$	115.02694	115.09
	Glutamate	Glu (E)	$C_5H_7NO_3$	129.04259	129.12

APPENDIX D

GLYCOLYSIS AND GLUCONEOGENESIS



APPENDIX E

SPECIES SPECIFIC MODIFICATIONS TO SQDG

Structures of lipooligosaccharides produced by different Rhizobium strains						
Strain	n	Q	R_1	R_2	R_3	$R_{4,5}$
R. leguminosarum bv viciae RBL5560	2,3	$C_{18:2\ 4\ 6\ 11}$ $C_{18:1\ 11}$	CH_3CO	H	H	H
R. meliloti AK41	1,2,3	$C_{16:2\ 9}$ $C_{18:2\ 4\ 9}$	H CH_3CO	SO_3H	H	H
R. meliloti 2011	2,3	$C_{16:1\ 9}$ $C_{16:2\ 9}$ $C_{18:2\ 4\ 9}$	H CH_3CO	SO_3H	H	H
R. tropici CFN299	3	$C_{18:1\ 11}$	H	SO_3H	CH_3	H
B. japonicum USDA110	3	$C_{18:1\ 9}$	H	2-O-Me-Fuc	H	H
B. japonicum USDA135	3	$C_{18:1\ 9}$ C_{16}	H CH_3CO	2-O-Me-Fuc	H	H
B. japonicum USDA61	2,3	$C_{18:1\ 9}$	H	Fuc 2-O-Me-Fuc	H	H
R. fredii USDA257	1,2,3	$C_{18:1\ 11}$	H	Fuc 2-O-Me-Fuc	H	H
R. sp. Strain NGR234	3	$C_{18:1\ 11}$ C_{18}	H	2-O-Me-Fuc 2-O-Me, 4-O-SO_3H-Fuc 2-O-Me, 3-O-CO-CH_2-Fuc		H NH_2CO
A. caulinodans ORS571	2,3	$C_{18:1\ 11}$ C_{18}	H	H	CH_3	H

BIBLIOGRAPHY

Bibliography

[1] Gene Rodenbery, StarTrek series.
[2] G. Hu and B. L. Vallee; Anal. Biochem.; **218**; 185-91 (1994).
[3] P. Jackson; Anal. Biochem.; **216**; 243-52 (1994).
[4] C. W. Sutton, J. A. O'Neill, and J. S. Cottrell; Anal. Biochem.; **218**; 34-46 (1994).
[5] D. Knapp; Analytical Derivatization Reactions; Wiley-Interscience Publication; New York; 146-224 (1979).
[6] Nimura, N.; Kinoshita, T.; Yoshida, T.; Uetake, A.; Nakai, C. Anal. Chem.; **60**, 2067-70 (1988).
[7] Z. El Rassi, J. Postlewait, Y. Mechref, G. Ostrander; Anal. Biochem.; **244**; 283-90 (1997).
[8] J. Quirke, C. Adams, and G. Van Berkel; Anal. Chem.; **66**, 1302-15 (1994).
[9] D. Knapp; Analytical Derivatization Reactions; Wiley-Interscience Publication; New York; 25-64 (1979).
[10] Bayliss, M. A. J.; Gomer, R. B.; Shepherd, J. Chromatogr.; **445;** 393-402 (1988).
[11] D. Knapp; Analytical Derivatization Reactions; Wiley-Interscience Publication; New York; 338-46 (1979).
[12] A. G. McInnes et al., J. Chromatogr. **1**, 556 (1958).
[13] C. C. Sweeley, R. Bentley, M. Makita, W. Wells; J. Am. Chem. Soc.; **85**; 2497 (1963).
[14] J. Lonngren and S. Svensson, Mass Spectrometry in Structural Analysis, 41-106
[15] H. Kwon, J. Kim; Anal. Biochem.; **215**; 243-52 (1993).
[16] D. Fu, R. O'Neill; Anal. Biochem.; **227**, 377-84 (1995).
[17] S. Honda, E. Akao, S. Suzuki, M. Okuda, K. Kakehi, J. Nakamura; Anal. Biochem.; **180**; 351-7 (1989).
[18] K. Sato, K. Sato, A. Okubo, S. Yamazaki; Anal. Biochem.; **251**; 119-21 (1997).
[19] H. Lis, N. Sharon; Eur. J. Biochem.; **218**; 1-27 (1993).
[20] M. Spellman; Anal. Chem.; **62**; 1714-22 (1990).
[21] V. Reinhold; B. Reinhold, C. Costello; Anal. Chem.; **67**; 1772-84 (1995).
[22] P. Jackson; Anal. Biochem.; **216**; 243-52 (1994).
[23] Y. Lee; Anal. Biochem.; **189**; 151-62 (1990).
[24] M. Hardy, R. Townsend; Proc. Natl. Acad. Sci. USA; **85**; 3289-93 (1988).
[25] K. Koizumi, Y. Kobata, T. Tanimoto, Y. Okada; J. Chromatogr.; **464**; 365-73 (1989).
[26] W. Wang, K. Erlansson, F. Lindh, T. Lundgren, D. Zopf; Anal. Biochem.; **190**; 182-7 (1990).
[27] A. Manzi, S. Diaz, A. Varki; Anal. Biochem.; **188**; 20-32 (1990).

[28] D. Whitfield, S. Stojkovski, H. Pang, J. Baptista, B. Sarkar; Anal. Biochem.; **194**; 259-67 (1988).
[29] R. Townsend, M. Hardy, O. Hindsgaul, Y. Lee; Anal. Biochem.; **174**; 459-70 (1988).
[30] T. Hayase, M. Sheykhanazari, V. Bhavanandan, A. Savage, Y. Lee; Anal. Biochem.; **211**; 72-80 (1993).
[31] A. Clarke; Anal. Biochem.; **212**; 344-50 (1993).
[32] Y. Lee; Anal. Biochem.; **189**; 151-62 (1990).
[33] G. Guile, S. Wong, R. Dwek; Anal. Biochem.; **222**; 231-5 (1994).
[34] M. Field, D. Papac, A. Jones; Anal. Biochem.; **239**; 92-8 (1996).
[35] J. Baenziger, M. Natowicz; Anal. Biochem.; **114**; 276-80 (1981).
[36] S. Mellis, J. Baenziger; Anal. Biochem.; **114**; 276-80 (1983).
[37] S. Mellis, J. Baenziger; Anal. Biochem; **134**; 442-9 (1983).
[38] S. Hase, S. Hara, Y. Matsushima; J. Biochem. **85**; 217-20 (1979).
[39] A. Kondo, M. Kiso, A. Hasegawa, I. Kato; Anal. Biochem.; **219**; 21-5 (1994).
[40] N. Kuraya, S. Hase; Anal. Biochem.; **233**; 205-11 (1996).
[41] Y. Makino, N. Kuraya, K. Omichi, S. Hase; Anal. Biochem.; **238**; 54-9 (1996).
[42] N. Tomiya, J. Awaya, M. Kurono, S. Endo, Y. Arata, N. Takahashi; Anal. Biochem.; **171**; 73-90 (1988).
[43] N. Takahashi, Y. Wada, J. Awaya, M. Kurono, N. Tomiya; Anal. Biochem.; **208**, 96-109 (1993).
[44] Y. Zhang, R. Cedergren, T. Nieuwenhuis, R. Hollingsworth; Anal. Biochem.; **208**; 363-71 (1993).
[45] K. Anumula; Anal. Biochem.; **230**; 24-30 (1995).
[46] J. Bigge, T. Patel, J. Bruce, P. Goulding, S. Charles, R. Parekh; Anal. Biochem.; **230**, 229-38 (1995).
[47] V. Ling, A. Guzzetta, E. Canova-Davis, J. Stults, W. Hancock; Anal. Chem.; **63**; 2909-15 (1991).
[48] P. Thibault, D. Watson, N. Young; In Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics; San Francisco, CA, 613 (1993).
[49] S. Carr, M. Huddleston, M. Bean; Protein Science; **2**; 183 (1993).
[50] K. Medzihradszky, C. Settineri, D. Maltby, A. Burlingame; In Techniques in Protein Chemistry IV; Angeletti, R. H.; Ed. Academic Press: San Diego, 117-25 (1993).
[51] M. Huddleston, M. Bean, S. Carr; Anal. Chem. **65**; 877-84 (1993).
[52] S. Carr, J. Barr, G. Roberts, K. Anumula, P. Taylor; In Methods in Enzymology, Vol. 193; McCloskey, J. A.; Ed. Academic Press: San Diego, 501-18 (1990).
[53] K. Medzihradszky, D. Maltby, S. Hall, C. Settineri, A. Burlingame; J. Am. Soc. For Mass Spectrom.; **94**; 350-8 (1994).

[54] T. Creighton; Proteins, Structures and Molecular Principles; W. H. Freeman and Co.; New York, 1984.
[55] P. Edman, G. Begg; Eur. J. Biochem; **1**; 80-91 (1967).
[56] W. Pearson; In Methods in Enzymology, Vol. 183; Doolittle, R. F., Ed.; Academic: San Diego, 1990; pp. 63-98.
[57] K. Biemann; Ann. Rev. Biochem.; **61**; 977-1010 (1992).
[58] T. Covey, R. Bonner, B. Shushan, J. Henion; Rapid Commun. Mass Spectrom.; **2**; 249 (1988).
[59] J. Loo, H. Udseth, R. Smith; Anal. Biochem.; **176**; 404 (1989).
[60] R. Smith, J. Loo, C. Edmonds, C. Barinaga, H. Udseth; Anal. Chem.; **62**; 882-99 (1990).
[61] A. Rockwood, M. Busman, R. Smith; Int. J. Mass Spectrom. Ion Processes; **111**; 103-29 (1991).
[62] M. Busman, A. Rockwood, R. Smith; J. Phys. Chem.; **96**; 2397-400 (1992).
[63] J. Loo, C. Edmonds, R. Smith; Anal. Chem.; **63**; 2488-99 (1989).
[64] B. Domon, C. Costello; Glycoconjugate J.; **5**; 397-409 (1988).
[65] A. Varki; Glycobiology; **3**; 97 (1993).
[66] V. Reinhold, B. Reinhold, C. Costello; Anal. Chem.; **67**; 1772-84 (1995).
[67] M. R. Hardy, R. R. Townsend; Y. C. Lee; Methods Enzym. **179**, 65-76 (1989).
[68] L. J. Basa; M. W. Spellman; J. Chromatogr.; **499**; 205-11 (1990).
[69] K. Yamashida; T. Mizuochi; A. Kobata; Methods Enzymol.; **83**; 105-26 (1982).
[70] M. Spellman; Anal. Chem.; **62**; 1714-22 (1990).
[71] S. Hoffstetteer-Kuhn; A. Paulus; E. Gassmann, H. M. Widmer; Anal. Chem.; **63;** 1541-7 (1991).
[72] R. A. Evangelista, M. Liu, F. Chen; Anal. Chem.; **67**; 2239-45 (1995).
[73] R. Townsend, M. Hardy, D. Cuming, J. carver, B. Bendiak; Anal. Biochem.; **182**; 1-8 (1989).
[74] L. Snyder; J. Kirkland; (1979) Introduction to Modern Liquid Chromatography, 2nd ed., pp. 281-9, Wiley, New York.
[75] Y. Lee; Anal. Biochem.; **189;** 151-62 (1990).
[76] C. Yulp; V. Manam; R. Hepler; J. Hennessey; Anal. Biochem.; **201**; 343-9 (1992).
[77] Y. Lee in Methods in Enzymology (Ginsburg, V., Ed.) Vol. 28, pp. 63-73 (1972) Academic Press, New York.
[78] G. Neuburger, D. Johnson; Anal. Chem.; **59**; 150-4 (1987).
[79] W. Lu, R. Cassidy; Anal. Chem.; **65**; 2878-81 (1993).
[80] L. Colon, R. Dadoo, R. Zare; Anal. Chem.; **65**; 476-81 (1993).
[81] T. O'Shea, S. Lunte; Anal. Chem.; **65**; 948-51 (1993).
[82] R. Rocklin, A. Clarke, M. Weitzhandler; Anal. Chem.; **70**; 1496-1501 (1998).
[83] H. Kwon; J. Kim; Anal. Biochem.; **215**; 243-52 (1993).
[84] P. Jackson; Anal. Biochem.; **216**; 243-52 (1994).

[85] R. Pertel; Int. J. Mass Spectrom. Ion Processes; **16**; 39 (1975).
[86] R. Voykner, T. Pack; Rapid Commun. Mass Spectrom.; **63**; 580 (1991).
[87] V. Katta, S. Chowdhury, B. Chait; Anal. Chem.; **63**; 174 (1991).
[88] Fisons Platform Guide (1993).
[89] C. Cresenzi, A. Di Corcia, S. Marchese, R. Samperi; Anal. Chem.; **67**, 1968 (1995).
[90] F. McLafferty; Registry of Mass Spectral Data.
[91] V. Reinhold; B. Reinhold, C. Costello; Anal. Chem.; 1995; **67**; 1772-84.
[92] C. Barinaga, C. Edmonds, H. Udseth, R. Smith; Rapid Commun. Mass Spectrom.; **3**; 160-4 (1989).
[93] S. Carr, M. Hemling, M. Bean, G. Roberts; Anal. Chem.; **63**; 2802-24 (1991).
[94] R. Smith, J. Loo; R. Ogorzalek Loo; M. Busman, H. Udseth; Mass Spectrom. Rev.; **10**; 249-51 (1991).
[95] V. Datta, S. Chowdhury, B. Chait; Anal. Chem.; **63**; 174-8 (1991).
[96] C. Meng, C. McEwen, B. Larsen; Rapid Commun. Mass Spectrom.; **4**; 151-5 (1990).
[97] M. Barber, D. Bell, M. Morris, L. Tetler, M. Woods, J. Monaghan, W. Morden; Org. Mass Spectrom.; **24**; 504-10 (1989).
[98] R. Orlando, C. Fenselau; Org. Mass Spectrom.; **26**; 664-6 (1991).
[99] R. Orlando, R. Boyd; Org. Mass Spectrom.; **27**; 151-5 (1992).
[100] R. Pertel; Int. J. Mass Spectrom. Ion Processes; **16**; 39 (1975).
[101] C. Cresenzi, A. Di Corcia, S.. Marchese, R. Samperi; Anal. Chem.; **67**; 1968 (1995).
[102] R. Voykner, T. Pack; Rapid Commun. Mass Spectrom.; **63**; 580 (1991).
[103] V. Katta, S. Chowdhury, B. Chait; Anal. Chem.; **63**; 174 (1991).
[104] G. Neubauer, R. Anderegg; Anal. Chem.; **66**; 1056-61 (1994).
[105] P. Schnier, D. Gross, E. Williams; J. Am. Soc. Mass Spectrom.; **6**; 1086-97 (1995).
[106] S. Hunt, M. Sheil, M. Belov, P. Derrick; Anal. Chem.; **70**; 1812-22 (1998).
[107] D. Noble; Anal. Chem.; **66**; 1239A-42A (1994).
[108] R. Atlas; Microbiology; Macmillan Publishing Co.; N.Y.; 757-71 (1984).
[109] I. Vijn, L. das Neves, A. van Kammen, H. Franssen, T. Bisseling; Science; **260**; 1764-5 (1993).
[110] J. Denarie, J. Cullimore; Cell; **74**; 951-4 (1993).
[111] R. Dixon, C. Wheeler; *Nitrogen Fixation in Plants*; Chapman and Hall; NY (1986).
[112] I. Tikhonovich, N. Provorov, V. Romanov, W. Newton; *Nitrogen Fixation: Fundamentals and Applications*; Dluwer Academic Publishers; Boston (1995).
[113] Kloepper (1992).
[114] Gunnera
[115] Nap and Bisseling, 1990.

[116] A. van Brussel, K. Recourt, E. Pees, H. Spaink, T. Tak, C. Wijffelman, J. Kijne, B. Lugtenberg; J. Bacteriol.; **172**; 5394-401 (1990).
[117] A. van Brussel, S. Zaat, H. Canter-Cremers, C. Wijfelman, E. Pees, B. Lugtenberg; J. Bacteriol.; **165**; 517-22 (1986).
[118] M. Bec-Ferte, H. Krishnan, D. Prome, A. Savagnac, S. Pueppke, J. Prome; Biochemistry; **33**; 11782-88 (1994).
[119] R. Carlson, S. Juan, U. Bhat, J. Glushka, H. Spaink, A. Wijfjes, A. van Brussel, T. Stokkermans, N. Peters, G. Stacey; J. Biol. Chem.; **268**; 18372-81 (1993).
[120] P. Lerouge, P. Roche, C. Faucher, F. Maillet, G. Truchet, J. Prome, J. Denarie; Nature; **344**; 781-4 (1990).
[121] P. Mergaert, M. Van Montagu, J. Prome, M. Holsters; Proc. Matl. Acad. Sci.; **90**; 1551-5 (1993).
[122] R. Poupot, E. Martinez-Romero, J. Prome; Biochemistry; **32**; 10430-5 (1993).
[123] N. Price, B. Relic, F. Talmont, A. Lewin, D. Prome, S. Pueppke, F. Maillet, J. Denarie, J. Prome, W. Broughton; Mol. Microbiol.; **6**; 3575-84 (1992).
[124] J. Sanjuan, W. Carlson, H. Spaink, U. Bhat, W. Barbour, J. Glushka, G. Stacey; Proc. Natl. Acad. Sci.; **89**; 8789-93 (1992).
[125] M. Schultze, B. Quiclet-Sir, E. Kondorosi, H. Virelizier, J. Glushka, G. Endre, S. Gero, A. Kondorosi; Proc. Natl. Acad. Sci. **89**; 192-6 (1992).
[126] H. Spaink, D. Sheeley, A. van Brussel, J. Glushka, W. York, T. Tak, O. Geiger, E. Kennedy, V. Reinhold, B. Lugtenberg; Nature; **354**; 125-30 (1991).
[127] J. Denarie, J. Cullimore; Cell; **74**; 951-4 (1993).
[128] B. Bohool, E. Schmidt; Science; **185**; 269-71 (1974).
[129] J. Hamblin, S. Kent; Nature; **254**; 28-30 (1973).
[130] S. Puepppke, Plant Physiol.; **75**; 924-8 (1984).
[131] S. Vesper, W. Bauer; Appl. Environ. Microbiol.; **52**; 134-41 (1986).
[132] C. Benning; Annu. Rev. Plant Physiol. Plant Mol. Biol.; **49**; 53-75 (1998).
[133] P. van Rhijn, J. Vanderleyden; Microbiological Reviews; **59**; 124-42 (1995).
[134] M. Schultze, A. Kondorosi; Annu Rev Genet; **32**; 33-57 (1998).
[135] S Jabbouri, B. Relic, M. Hanin, P. Kamalaprija, U. Burger, D. Prome, J. Prome, W. Broughton; The J. of Biol. Chem.; **273**; 12047-55 (1998).
[136] R. Poupot, E. Martinez-Romero, F. Maillet, J. Prome; FEBS Letters; **368**; 536-40 (1995).
[137] P. Lerouge, P. Roche, C. Faucher, F. Maillet, G. Truchet, J. Prome, J. Denarie; Nature; **344**; 781-4 (1990).
[138] J. Denarie, F. Debelle, J. Prome; Annu. Rev. Biochem.; **65**; 503-36 (1996).
[139] P. Mergaert, M. van Montague, M. Holsters; Mol. Microbiol.; **25**; 811-7 (1997).
[140] J. Joyard, E. Blee, R. Douce; Biochemica et Biophysica Acta; **879**; 78-87 (1986).
[141] C. Pugh, A. Roh, T. Hawkes, J. Harwood; Biochem. J.; **309**; 513-9 (1995).

[142] M. Olsthoorn, I. Lopez-Lara, B. Peterson, K. Bock, J. Haverkamp, H. Spaink, J. Thomas-Oates; Biochem.; **37**; 9024-32 (1998).
[143] A. Roy, J. Harwood; Biochem. J.; **344**; 185-7 (1999).
[144] C. Benning, C. Somerville; J. Bacteriology; **174**; 6479-87 (1992).
[145] R. Carlson, J. Sanjuan, U. Bhatt, J. Glushka, H. Spaink, A. Wijfjes, A. van Brussel, T. Stokkermans, N. Peters, G. Stacey; The J. of Biol. Chem., **268**; 18372-81 (1993).
[146] P. Mergaert, M. Van Montagu, J. Prome, M. Holsters; Proc. Natl. Aca. Sci. USA; **90**; 1551-5 (1993).
[147] R. Lee, D. Monsey, A. Weston, K. Duncan, C. Rithner, M. McNeil; Anal. Biochem.; **242**: 1-7 (1996).
[148] S. Hull, R. Montgomery; Anal. Biochem.; **222**; 49-54 (1994).
[149] H. Smits, A. Cohen, T. Buttler, J. Nielsen, L. Olsson; Anal. Biochem.; **261**; 36-42 (1998).
[150] J. Sanjuan, R. Carlson, H. Spaink, U. Bhat, W. Barbour, J. Glushka, G. Stacey; Proc. Natl. Acad. Sci. USA; **89**; 8789-93 (1992).
[151] M. Rossak, C. Tietji, E. Heinz, C. Benning; The J. of Biol. Chem.; **270**; 25792-7 (1995).
[152] N. Karamanos, A. Syrokou, P. Vanky, M. Nurminen, A. Hjerpe; Anal. Biochem.; **221**; 189-99 (1994).
[153] M. Stefansson, M. Novotny; Anal. Chem.; **66**; 1134-40 (1994).
[154] M. Stefansson, M. Novotny; Anal. Chem.; **66**; 3466-71 (1994).
[155] J. Sudor, M. Novotny; Anal. Chem.; **67**; 4205-9 (1995).
[156] M. Schultze, B. Quiclet-Sire, E. Kondorosi, H. Virelizier, J. Glushka, G. Endre, S. Gero, A. Kondorosi; Proc. Natl. Acad. Sci. USA; **89**; 192-6 (1992).
[157] R. Poupot, E. Martinez-Romero, J. Prome; Biochemistry; **32**; 10430-5 (1993).
[158] R. Poupot, E. Martinez-Romero, F. Maillet, J. Promet; FEBS letters; **386**; 536-40 (1995).
[159] P. Mergaert, M Van Montagu, J Prome, M. Holsters; Proc. Natl. Acad. Sci. USA; **90**; 1551-5 (1993).
[160] R. Carlson, J. Sanjuan, U. Bhat, J. Glushka, H. Spaink, A. Wijfjes, A. van Brussel, T. Stokkermans, N. Peters, G. Stacey; The J. of Biol. Chem.; **268**; 18372-81 (1993).
[161] A. Lehninger; Principles of Biochemistry; Worth Publishers, Inc., NY, NY; p. 348 (1982).
[162] J. McMurry; *Organic Chemistry*; Brooks/Cole Publishing Company; CA; 178 (1984).
[163] A. Roy, M. Hewlins; Carbohydr. Res.; **302**; 113-118 (1997).
[164] F. Wang, L Sayre; J. Amer. Chem. Soc.; **114**; 248-55 (1992).
[165] J. March; Advanced Organic Chemistry; pp. 363-4; John Wiley & Sons; NY (1985).

[166] Gilbert; "Sulfonation and Related Reactions;" pp 136-48; Interscience, NY (1965).
[167] R. Reed, H. Tartar; J. Am. Chem. Soc.; **57**; 570 (1935).
[168] J. McMurry; Organic Chemistry; p 300; Brooks/Cole; CA (1984).
[169] J. Slagle, T. Huang, B. Franzus; J. Org. Chem.; **46**; 3526-30 (1981).
[170] R. Appel; Angw. Chem. Internat. Edit.; **14**; 801-11 (1975).
[171] Y. Lee; Anal. Biochem.; **189**; 151-62 (1990).
[172]
[173] R. Cedergren, R. Hollingsworth;