

**Carbohydrate-Interacting Proteins from Two *Nostoc*  
(Cyanobacteria) Species**

**Brian Robert Jordan**

**A Dissertation Submitted to the Faculty of the Virginia Polytechnic Institute and  
State University in Partial Fulfillment of the Requirements for the Degree of:**

**Doctor of Philosophy**

**in**

**Biochemistry**

**Malcolm Potts, Chair**

**Richard F. Helm**

**Peter J. Kennelly**

**William E. Newton**

**Kevin E. VanCott**

**Raymond E. Dessy**

**February 18, 2004**

**Blacksburg, VA**

Key words: cyanobacteria, carbohydrate, glycosyltransferase, porin

## Carbohydrate-Interacting Proteins from Two *Nostoc* (Cyanobacteria) Species

Brian Robert Jordan

### ABSTRACT

Cyanobacteria of the *Nostoc* genus are known for the thick, mucilaginous carbohydrate coatings that they produce. In this work, two examples of cyanobacterial glycobiology are considered, each of which involves a cyanobacterium of the *Nostoc* genus.

The first portion of this work details attempts to obtain amino acid sequence information from the enzymes (glycosyltransferases) that are responsible for producing the extracellular polysaccharide (EPS) of *Nostoc commune* DRH1, ultimately to allow the transfer of this capacity to another organism. Two artificial substrates were synthesized for use in a capillary electrophoresis-based enzyme assay, which was used to look for glycosyltransferase activity in *Nostoc commune* DRH1 cell extracts. Glucuronosyltransferase activity was detected in association with *Nostoc commune* membrane material. The active enzyme displayed a divalent cation metal dependence ( $Mg^{+2}$ ) that is typical of glycosyltransferase enzymes purified from other organisms. Because the enzyme responsible for this activity held the potential to be EPS-related, its purification was attempted.

The capillary electrophoresis-based enzyme assay and a  $^{32}P$ -labeled affinity tag were utilized to follow the glucuronosyltransferase enzyme through successive purification steps. The active enzyme was extracted from *Nostoc commune* membrane material using Triton X-100, and then purified by anion exchange chromatography. The active detergent extract was extremely unstable, and consequently, other purification techniques tested were unsuccessful in enriching activity. Affinity-labeling experiments indicated that the active enzyme was forming protein aggregates during these procedures, which were not amenable to in-gel protease digestion and peptide analysis by tandem mass spectrometry.

The second portion of this work describes an investigation of an *Anabaena* (*Nostoc*) PCC 7120 soluble cell extract. Upon separation by sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) and subsequent periodic acid-Schiff (PAS) staining of the resulting gel, the components of this cellular fraction produce a ladder-like pattern, which suggests that the extract may contain glycosylated protein. Analyses of several samples that were taken from within the PAS-staining region of such a gel revealed surface layer homology (SLH) domain-containing proteins, likely candidates to be covalently attached to or non-covalently interacting with carbohydrate.

Various protein sequence analyses indicated that the detected SLH domain-containing proteins belong to a family of (putative) cyanobacterial porins. Proteins in this family possess features that include a N-terminal signal sequence, a single SLH domain motif, followed by a coiled-coil region, and a C-terminal region that is homologous to the  $\beta$ -barrel-forming region of bacterial porins. All of these features were identified in the detected *Anabaena* (*Nostoc*) PCC 7120 SLH domain-containing proteins.

Smith degradation was performed on a sample that was electroeluted from the PAS-staining region of a preparative-scale SDS-PAGE gel of the soluble cell extract. Subsequent analyses of the resulting sample by SDS-PAGE and mass spectrometry indicated that at least two SLH domain-containing proteins, encoded by *all4499* and *alr4550*, were non-covalently interacting with the PAS-staining material. Following degradation, the PAS-staining material was still of sufficient size to be detected by gel electrophoresis, and it continued to migrate in the absence of an interacting protein component. Protease digestion of a similarly prepared sample, and then subsequent analysis by SDS-PAGE and mass spectrometry, revealed that the region between amino acid residues #504 and #536, in the protein encoded by the *alr4550* open reading frame, was interacting with the PAS-staining material. Monosaccharide composition analyses of this material revealed more carbohydrate constituents than are found in cyanobacterial primary (peptidoglycan) cell wall polymer alone, indicating that it contained a significant secondary cell wall polymer component as well.

## ACKNOWLEDGEMENTS

I'd like to thank my co-advisors, Dr. Richard Helm and Dr. Malcolm Potts, for their continued educational and financial support over the past several years. Additionally, I'd like to thank my remaining committee members, Dr. William Newton, Dr. Kevin VanCott, Dr. Ray Dessy, and Dr. Peter Kennelly in particular, for their advice and guidance.

Lacking the space to list everyone who has ever contributed to my general education, and the mental capacity to recall all of these people, I'd like to thank everyone that I can think of who played a direct role in assisting me to complete this project: William Bebout, Jim Biedler, Andrea DeSantis, Janet Donahue, Dr. Joe Falkingham, Dr. Karl Fisher, Tom Glass, Dr. Eugene Gregory, Kim Harich, Dr. Zebo Huang, Dave Hull, Dr. Udo Huneke, Ken Hurley, Dr. Geno Iannaccone, Jody Jervis, Karsten Klage, Dr. Brian Lower, Archana Mukhopadhyay, Ben Potters, Dr. Keith Ray, Randy Saunders, Steve Slaughter, Sue Smith, and Debbie Wright. I apologize if I have forgotten anyone.

I'd like to thank my fiancée Holly for her overall patience and her relative compliance with my demands for silence over the past several months. In retrospect, I hope that she realizes that my threats of violence were empty. The kitty was never really 'going to get it.'

Most importantly, I'd like to acknowledge the generosity of those who were so considerate with my work schedule when I was an undergraduate. Were it not for their willingness to allow me to work around both my classes, and the schedule of the opposing job, I would not have finished my first degree. Without their assistance it would have been impossible for me to achieve many of my goals in life, career-oriented or otherwise. There were many such people, but the leeway of Leonard Haddix, Alan Suskotaski, and Cassandra Bird come to mind immediately. I recall the financial generosity of Eugene Cilento as well. I am humbled by the actions of these people still.

This work was made possible by funding through the Multidisciplinary University Research Initiative (MURI) program of the Department of Defense (Biomimetic Cell and Tissue Stasis; N00014-01-1-0852), as well as the Metabolic Engineering for Cellular Stasis Program of DARPA (N00173-98-1-G005-P00004 and N00173-02-1-G016).

## TABLE OF CONTENTS

ABSTRACT .....	II
ACKNOWLEDGEMENTS .....	IV
TABLE OF CONTENTS .....	V
LIST OF FIGURES .....	X
LIST OF TABLES .....	XII
LIST OF ABBREVIATIONS .....	XIII

### CHAPTER I

#### Introduction and Background

##### *1.1 Unique Biological Properties of Cyanobacteria*

1.1.1 The Rise and Significance of Cyanobacteria .....	1
1.1.2 Survival Strategies Developed by Cyanobacteria .....	2
1.1.3 Cyanobacterial Desiccation Tolerance, A Symphony of Biological Processes .....	4
1.1.4 Sorting It All Out .....	5

##### *1.2 Nostoc commune, a Model for Prokaryotic Anhydrobiosis*

1.2.1 The Scientific Significance of Extremophiles .....	6
1.2.2 Transfer of Desiccation Survival Strategies .....	6
1.2.3 Glycosyltransferases, A Brief Overview .....	8
1.2.4 <i>Nostoc commune</i> , An Organism With Something to Offer .....	10
1.2.5 Finding and Following a Functional Enzyme .....	11
1.2.6 Specific Aims .....	13

##### *1.3 SLH Domain-Containing Proteins In Anabaena (Nostoc) PCC 7120*

1.3.1 Bacterial Protein Glycosylation .....	15
1.3.2 Characterization of Bacterial Glycoproteins .....	15
1.3.3 Functions of Bacterial Protein Glycosylation .....	16
1.3.4 Bacterial Surface Layer (S-Layer) Proteins .....	17
1.3.5 Cyanobacterial Genomics and Phylogenetics .....	20
1.3.6 An Interesting Observation .....	20
1.3.7 Specific Aims .....	21

### CHAPTER II (Results)

#### Detection of Glycosyltransferase Activity in *Nostoc commune* Using Capillary Electrophoresis with Laser-Induced Fluorescence Detection (CE-LIF)

2.1 Introduction .....	31
2.2 Synthesis of Fluorescence-Labeled Monosaccharide Acceptors .....	32
2.3 Synthesis of Fluorescence-Labeled Disaccharide Acceptors .....	32

2.4 Application of Fluorescence-Labeled Acceptors to <i>Nostoc commune</i> .....	35
2.5 Detection of an Apparent Glucuronosyltransferase Activity .....	37
2.6 Confirmation of CE-LIF Analysis Results by Preparative-Scale Thin Layer Chromatography and Matrix Assisted Laser Desorption/Ionization – Time of Flight Mass Spectrometry .....	38
2.7 Summary .....	40

### CHAPTER III (Results)

#### Partial Purification of a *Nostoc commune* Protein with Apparent Glucuronosyltransferase Activity and Attempts to Determine Its Amino Acid Sequence

3.1 Introduction .....	50
3.2 Localization of Activity .....	50
3.3 Detergent Extraction of Glucuronosyltransferase Activity .....	52
3.4 Partial Purification of Glucuronosyltransferase Activity .....	52
3.5 Stabilization of Glucuronosyltransferase Activity .....	54
3.6 Characterization of Glucuronosyltransferase Activity .....	55
3.7 Analysis of Partially Purified Fraction Containing Glucuronosyltransferase Activity .....	56
3.8 Non-radioactive Affinity Tagging of a Partially Purified Fraction Containing Glucuronosyltransferase Activity .....	56
3.9 Radioactive Affinity Tagging of a Partially Purified Fraction Containing Glucuronosyltransferase Activity .....	58
3.10 Attempts to Isolate a Radiolabeled Peptide .....	59
3.11 ‘Brute Force’ <i>de novo</i> Sequencing of Radio-labeled Spots .....	60
3.12 Summary .....	62

### CHAPTER IV (Discussion)

#### The Search for Genes That Encode *Nostoc commune* DRH1 EPS-Producing Enzymes

4.1 Introduction .....	74
4.2 Development of A Method to Detect Glycosyltransferase Activities in <i>Nostoc commune</i> DRH1 ( <i>Specific Aim #1</i> ) .....	74
4.3 Detection of EPS-Related Glycosyltransferase Activities in <i>Nostoc commune</i> DRH1 ( <i>Specific Aim #2</i> ) .....	77
4.4 Partial Purification and Identification of the Protein Responsible for Glucuronosyltransferase Activity in <i>Nostoc commune</i> DRH1 ( <i>Specific Aim #3</i> ) .....	78
4.5 <i>de novo</i> Sequencing of a Complex Sample Thought to Contain a Glucuronosyltransferase from <i>Nostoc commune</i> DRH1 ( <i>Specific Aim #4</i> ) .....	81
4.6 Conclusions .....	82
4.7 Future Directions .....	83

## CHAPTER V (Results)

### Analysis of Potential Glycoprotein(s) from *Anabaena (Nostoc)* PCC 7120

5.1 Introduction.....	86
5.2 Protein Analysis by Tandem Mass Spectrometry.....	87
5.3 Glycan Differentiation.....	87
5.4 Purification of Comigrating Protein and Carbohydrate Material.....	88
5.5 Protein Deglycosylation Attempts.....	90
5.6 Identification of a Carbohydrate Binding Domain.....	93
5.7 Protein Function Analysis of SLH Domain-Containing Proteins.....	94
5.8 The Effect of Nitrogen-Supplemented Media on SLH Domain-Containing Protein Expression.....	96
5.9 Carbohydrate Analyses.....	96
5.10 Summary.....	98

## CHAPTER VI (Discussion)

### Carbohydrate Associations of *Anabaena (Nostoc)* PCC 7120 SLH Domain-Containing Proteins

6.1 Introduction.....	115
6.2 Analysis of a Periodic Acid/Schiff-Staining Pattern ( <i>Specific Aim #1</i> ).....	115
6.3 Identification of Carbohydrate-Binding Proteins ( <i>Specific Aim #2</i> ).....	120
6.4 Carbohydrate-Binding Protein Function Analysis ( <i>Specific Aim #3</i> ).....	122
6.5 Carbohydrate Analysis of PAS-Staining Material ( <i>Specific Aim #4</i> ).....	124
6.6 Conclusions.....	127
6.7 Future Directions.....	127

## CHAPTER VII

### Instrumentation, Materials, Procedures, and Electronic Resources

#### 7.1 Instrumentation

7.1.1 Analytical Instrumentation.....	132
7.1.2 Instrumentation for Protein Isolation and Chemistry.....	132
7.1.3 General Laboratory Instrumentation.....	133

#### 7.2 Materials

7.2.1 General Materials.....	135
7.2.2 Materials for Synthesis of Tagged Acceptors.....	135
7.2.3 Materials for Protein Chemistry.....	135
7.2.4 Materials for Carbohydrate Analysis.....	136

<b>7.3 General Analytical Procedures</b>	
7.3.1 Solid Phase Extraction (SPE) of FCHASE Derivatives .....	137
7.3.2 Matrix Assisted Laser Desorption/Ionization – Time of Flight (MALDI-TOF) Mass Spectrometry .....	137
7.3.3 Glycosyltransferase Activity Reactions.....	138
7.3.4 Capillary Electrophoresis with Laser-Induced Fluorescence Detection (CE-LIF).....	138
<b>7.4 Procedures for Fluorescent-Labeled Acceptor Synthesis</b>	
7.4.1 Thin-Layer Chromatography (TLC).....	139
7.4.2 Flash (Silica Gel) Chromatography .....	139
7.4.3 Nuclear Magnetic Resonance (NMR).....	140
<b>7.5 Procedures for Protein Purification and Analysis</b>	
7.5.1 Growth of Organisms.....	140
7.5.2 Cell Lysis .....	140
7.5.3 Protein Assays.....	142
7.5.4 SDS-PAGE .....	142
7.5.5 Two-Dimensional Electrophoresis.....	143
7.5.6 Detection of Protein/Peptides in Polyacrylamide Gels.....	144
7.5.7 Protein Electroelution .....	145
7.5.8 In-Gel Trypsin Digestion .....	145
7.5.9 Liquid Chromatography – Mass Spectrometry (LC-MS <sup>n</sup> ).....	146
7.5.10 In-Solution Trypsin Digestion .....	147
7.5.11 Enzymatic Deglycosylation .....	147
7.5.12 Chemical Deglycosylation.....	147
<b>7.6 Procedures for Carbohydrate Analysis</b>	
7.6.1 Detection of Carbohydrates in Polyacrylamide Gels.....	148
7.6.2 Blotting of Protein onto PVDF Membranes .....	149
7.6.3 Glycoprotein Detection on PVDF Membranes.....	149
7.6.4 Beta-Elimination Release of Glycan / Base Hydrolysis of Protein .....	149
7.6.5 Oligosaccharide Hydrolysis.....	150
7.6.6 Enzymatic Conversion of <i>N</i> -acetylneuraminic Acid to <i>N</i> -acetylmannosamine.....	150
7.6.7 APTS (1-Aminopyrene-3,6,8-Trisulfonate) Derivatization of Carbohydrates for CE-LIF Analyses .....	150
<b>7.7 Specific Protein Purification and Labeling Procedures</b>	
7.7.1 Partial Purification of <i>Nostoc commune</i> DRH1 Glucuronosyltransferase Activity .....	151
7.7.2 <i>Nostoc commune</i> DRH1 Glucuronosyltransferase Affinity Labeling .....	152
7.7.3 Purification of <i>Anabaena (Nostoc)</i> PCC 7120 SLH Domain-Containing Proteins and SDS-PAGE Comigrating Carbohydrate Material .....	152



<b>7.8 Electronic Resources</b>	
7.8.1 Genomic Databases.....	153
7.8.2 Search Algorithms .....	153
7.8.3 Protein Function Analysis.....	154
7.8.4 Miscellaneous .....	154
<b>7.9 Synthesis of a Fluorescent-Labeled Glucose-<math>\beta</math>(1<math>\rightarrow</math>4)-Xylose Acceptor</b> .....	154
<b>7.10 Synthesis of a Fluorescent-Labeled Glucose-<math>\beta</math>(1<math>\rightarrow</math>4)-Glucose Acceptor</b> .....	161
<b>APPENDIX (Common Buffers and Solutions)</b> .....	165
<b>REFERENCES</b> .....	167
<b>CURRICULUM VITAE</b> .....	184

## LIST OF FIGURES

### *Chapter One*

Figure 1.1	Damage and Preservation of Phospholipid Bilayers During Drying.....	23
Figure 1.2	The Cyanobacterium <i>Nostoc commune</i> .....	24
Figure 1.3	Structure of <i>Nostoc commune</i> DRH1 EPS .....	25
Figure 1.4	Glycosyltransferase Reaction Mechanisms .....	26
Figure 1.5	Gene Clusters Encoding EPS-Synthesizing Proteins .....	27
Figure 1.6	Prokaryotic Cell Envelopes with S-Layers.....	28
Figure 1.7	ClustalX Alignment of Several SLH Domains .....	29
Figure 1.8	Periodic Acid-Schiff Staining Material in <i>Anabaena</i> PCC 7120 Soluble Protein Extract.....	30

### *Chapter Two*

Figure 2.1	Synthesis of Selectively Protected Xyloside Acceptor for Disaccharide Coupling Procedure.....	41
Figure 2.2	Synthesis of Disaccharide Coupling Procedure Donors.....	42
Figure 2.3	Disaccharide Coupling Attempts and Subsequent Reduction .....	43
Figure 2.4	Fluorescent-Tagging of First Disaccharide Acceptor.....	44
Figure 2.5	Fragments and Adducts of Synthesized Fluorescent-Labeled Disaccharide Acceptors .....	45
Figure 2.6	Preparation of 4-Aminophenyl Cellobiose.....	46
Figure 2.7	Fluorescent-Tagging of Second Disaccharide Acceptor .....	47
Figure 2.8	Detection of Glycosyltransferase Activity Using CE-LIF .....	48

### *Chapter Three*

Figure 3.1	Final Glucuronosyltransferase Partial Purification Protocol.....	63
Figure 3.2	Temperature Dependence of Glucuronosyltransferase Activity .....	64
Figure 3.3	Two-Dimensional Gel Analysis of a Partially Purified <i>Nostoc commune</i> DRH1 Protein Fraction Showing Glucuronosyltransferase Activity .....	65
Figure 3.4	Affinity Tagging Reaction and Subsequent Isoelectric Focusing of a Labeled <i>Nostoc commune</i> DRH1 Protein Fraction Showing Glucuronosyltransferase Activity .....	66
Figure 3.5	Liquid Scintillation Counting of Anion Exchange Fractions from a Tagged and Dialyzed <i>Nostoc commune</i> Active Fraction .....	67
Figure 3.6	Autoradiograph of a <i>Nostoc commune</i> Affinity-Tagged Fraction Separated by SDS-PAGE .....	68

Figure 3.7 Liquid Scintillation Counting of IMAC Fractions of a Tagged and Trypsin Digested <i>Nostoc commune</i> Active Fraction.....	69
Figure 3.8 <i>de novo</i> Sequenced Peptides and Peptide Fragments and Putative Matches .....	70
Figure 3.9 Putative <i>Nostoc punctiforme</i> ATCC 29133 and <i>Anabaena</i> PCC 7120 Peptide Matches .....	71
Figure 3.10 Homologous <i>Nostoc punctiforme</i> ATCC 29133 and <i>Anabaena</i> PCC 7120 ORF Groupings .....	72

### **Chapter Five**

Figure 5.1 SDS-PAGE Analysis of <i>Anabaena</i> PCC 7120 Soluble Cell Extract .....	100
Figure 5.2 Glycan Differentiation of <i>Anabaena</i> PCC 7120 Soluble Cell Extract .....	101
Figure 5.3 Lysozyme Treatment of Carbohydrate Material in <i>Anabaena</i> PCC 7120 Cell Extracts .....	102
Figure 5.4 Glycosidase Treatment of Suspected Glycoprotein(s) in <i>Anabaena</i> PCC 7120 Soluble Cell Extract .....	103
Figure 5.5 Triflic Acid Treatment of Suspected Glycoprotein(s) in an <i>Anabaena</i> PCC 7120 Cell Extract.....	104
Figure 5.6 Periodate/Acetic Acid Treatment of Carbohydrate Material in <i>Anabaena</i> PCC 7120 Soluble Cell Extract .....	105
Figure 5.7 In-Solution Trypsin Digest of Electroeluted <i>Anabaena</i> PCC 7120 Protein/Carbohydrate Material .....	106
Figure 5.8 ClustalX Alignment of <i>Anabaena</i> PCC 7120 SLH Domain-Containing Proteins and Homologs .....	107
Figure 5.9 Comparison of <i>alr4550</i> ORF Hypothetical Translation to Known Bacterial Porin Sequences.....	108
Figure 5.10 SDS-PAGE Analysis of BG11 <sub>0</sub> Media-Grown <i>Anabaena</i> PCC 7120 Soluble Cell Extract.....	109

### **Chapter Six**

Figure 6.1 SLH Domain-Containing Putative Cyanobacterial Porins .....	129
Figure 6.2 Schematic Representation of the <i>alr4550</i> ORF-Encoded Protein.....	130
Figure 6.3 Cyanobacterial SLH Domain-Containing Porins and Chloroplastic Outer Membrane Protein Homolog.....	131

## LIST OF TABLES

### *Chapter Two*

Table 2.1 Summary of Attempts to Detect Glycosyltransferase Activity in <i>Nostoc commune</i> DRH1 .....	49
---	----

### *Chapter Three*

Table 3.1 Comparison of $\gamma$ -ions Produced by TLLEGEESR and by a Peptide Found in Several <i>Nostoc commune</i> Affinity Tagged Samples .....	73
--	----

### *Chapter Five*

Table 5.1 Results of Protein Analyses by Tandem Mass Spectrometry.....	110
Table 5.1 Results of Protein Analyses by Tandem Mass Spectrometry (Continued) .....	111
Table 5.2 Protein Analyses of <i>Anabaena</i> PCC 7120 Periodate Oxidized Samples .....	112
Table 5.3 Protein Function Analyses of <i>Anabaena</i> PCC 7120 Proteins.....	113
Table 5.4 Carbohydrate Analyses of <i>Anabaena</i> PCC 7120 Samples .....	114

## LIST OF ABBREVIATIONS

### General Abbreviations

ADP-Glu	adenosine 5'-diphosphate-glucose
ADP-Rib	adenosine 5'-diphosphate-ribose
APTS	8-aminopyrene-1,3,6-trisulfonate
ATCC	American Type Culture Collection
BLAST	basic local alignment search tool
BSA	bovine serum albumin
C18	octadecyl silane
CE	capillary electrophoresis
CE-LIF	capillary electrophoresis with laser-induced fluorescence detection
CEBI	Center for Experimental Bioinformatics
CMC	critical micelle concentration
CMP-Neu5Ac	cytidine 5'-monophosphate-acetylneuraminic acid
CZE	capillary zone electrophoresis, see also FSCE
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EPS	extracellular polysaccharide
FCHASE	6-(fluorescein-5-carboxamido) hexanoic acid succinimidyl ester
FeSOD	iron-containing superoxide dismutase
FPLC	fast performance liquid chromatography
FSCE	free-solution capillary electrophoresis, see also CZE
GDP-Glu	guanosine 5'-diphosphate-glucose
GDP-Man	guanosine 5'-diphosphate-mannose
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HF	hydrofluoric acid
HPLC	high performance liquid chromatography
i.d.	internal diameter
IEF	isoelectric focusing
IMAC	immobilized metal affinity chromatography
LPS	lipopolysaccharide
MAA	mycosporine-like amino acids
MALDI-TOF	matrix assisted laser desorption ionization – time of flight (mass spectrometry)
MECC	micellar electrokinetic capillary chromatography
MES	2-( <i>N</i> -morpholino)ethanesulfonic acid
MnSOD	manganese-containing superoxide dismutase
MOPS	3-( <i>N</i> -morpholino)propanesulfonic acid
$M_r$	relative molecular mass
MS	mass spectrometry
MS/MS	tandem mass spectrometry, see also MS <sup>n</sup>

MS <sup>n</sup>	tandem mass spectrometry, see also MS/MS
<i>N</i> -linked	nitrogen-linked
NBT	4-nitro blue tetrazolium chloride
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
NosA	3- <i>O</i> -lactyl glucuronic acid
<i>O</i> -linked	oxygen-linked
ORF	open reading frame
OS-MAA	oligosaccharide-linked mycosporine-like amino acids
PAS	periodic acid-Schiff
PAS-PM	periodic acid-Schiff positive-staining material
PCC	Pasteur Culture Collection
PCR	polymerase chain reaction
PMF	peptide mass fingerprinting
PVDF	polyvinylidene fluoride
S-layer	surface-layer
SB 3-10	sulfobetaine 3-10 (detergent)
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SLH	surface-layer homology
S <sub>N</sub> 2	substitution nucleophilic bimolecular
SOD	superoxide dismutase
SPE	solid-phase extraction
SCWP	secondary cell wall polymer
TBP	tributyl phosphine
TCA	trichloroacetic acid
TFA	trifluoroacetic acid
THAM	2-amino-2-(hydroxymethyl)-1,3-propanediol, see also Tris (free base)
TLC	thin-layer chromatography
Tricine	<i>N</i> -[Tris(hydroxymethyl)methyl]glycine <i>N</i> -[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine
Tris (free base)	2-amino-2-(hydroxymethyl)-1,3-propanediol, see also THAM
Tris-HCl	Tris(hydroxymethyl)aminomethane hydrochloride
UDP	uridine 5'-disphosphate (or uridine 5'-disphospho-)
UDP-Gal	uridine 5'-diphosphate-galactose
UDP-Glu	uridine 5'-diphosphate-glucose
UDP-GluA	uridine 5'-diphosphate-glucuronic acid
UDP-Xyl	uridine 5'-diphosphate-xylose
UV	ultraviolet (radiation)
UV-A	320-400 nm ultraviolet radiation
UV-B	290-320 nm ultraviolet radiation
v/v	volume-to-volume ratio
VIGEN	Virginia Tech Institute for Genomics
w/v	weight-to-volume ratio
w/w	weight-to-weight ratio

x	times ( <i>i.e.</i> , 10 mL, 2x = with two 10 mL volumes)
X	concentration index ( <i>i.e.</i> , a 2X solution is diluted ½ for use)
X-phosphate	5-bromo-4-chloro-3-indolyl phosphate

### Greek Letter Abbreviations

$\alpha/\beta$	ratio of the $\alpha$ -anomer to the $\beta$ -anomer
$\beta/\alpha$	ratio of the $\beta$ -anomer to the $\alpha$ -anomer
$\delta_H$	$^1\text{H}$ NMR shifts
$\delta_C$	$^{13}\text{C}$ NMR shifts

### Units

A	ampere, see also amp
Å	angstrom
amp	ampere, see also A
°C	degrees Celsius or degrees Centigrade
Da	Daltons
g	gram
h	hour(s), see also hr
hr	hour(s), see also h
Hz	hertz
L	liter
m	meter
M	molar, mol/liter
min	minute
ppm	parts per million
psi	pounds per square inch
RFU	relative fluorescence units
s	second(s), see also sec
sec	second(s), see also s
V	volts
W	watts
$\Omega$	ohms

### Unit Prefixes

n	nano, $10^{-9}$
$\mu$	micro, $10^{-6}$
m	milli, $10^{-3}$
c	centi, $10^{-2}$
k	kilo, $10^3$
M	mega, $10^6$

# CHAPTER I

## Introduction and Background

### 1.1 Unique Biological Properties of Cyanobacteria

#### 1.1.1 The Rise and Significance of Cyanobacteria

Fossilized cyanobacteria are currently the oldest evidence for life on Earth (Brasier *et al.* 2002; Hedges *et al.* 2001; Schopf 1993; Schopf and Packer 1987). Fossil remnants of eleven different species of cyanobacteria have been found, dating from a time only 300 million years after the Earth's molten surface had cooled sufficiently to form continental rock and plates (Nisbet and Sleep 2001; Schopf and Packer 1987). During this time period the Earth was inhospitable to most life forms in existence today. The Earth's atmosphere was rich with gases such as ammonia, methane, and carbon dioxide (Nisbet and Sleep 2001), and its terrain was bombarded by damaging ultraviolet radiation. Cyanobacteria were one of the first organisms to evolve and flourish in this unforgiving environment. Acquiring carbon and energy through oxygenic photosynthesis, these organisms are credited with the dramatic increase in atmospheric oxygen content that occurred approximately 2.4 to 2.2 billion years ago (Catling *et al.* 2001).

Cyanobacteria have played a role in creating life as we know it today, both directly and indirectly. Cyanobacteria are believed to be the evolutionary precursor of chloroplasts, and it is thought that the photosynthetic production of oxygen by these bacteria was instrumental in the development of other organisms (Guerrero 1998). *The Endosymbiosis Hypothesis* (Margulis 1996) describes the idea that certain eukaryotic organelles (*i.e.*, chloroplasts, mitochondria, membrane-bound nuclei) arose from the invasion of one bacterial cell by another, which in some way imparted a selective advantage to the new hybrid organism. Although several variations of this hypothesis exist (Martin *et al.* 2001), most agree that a) chloroplasts are likely reduced cyanobacteria, the evolutionary remnants of a symbiotic relationship that once formed between a cyanobacterial cell and another single-celled organism and b) accumulation of



the photosynthetic by-product oxygen (O<sub>2</sub>) in the environment was the selective pressure that led to the development of aerobic organisms.

Over the first two billion years of evolution, bacterial organisms were the only inhabitants on the planet. During this time, bacterial organisms developed most of the metabolic pathways that are still in use by life today (Guerrero 1998; Nisbet and Sleep 2001). Selective pressure present during the Archean period (3.8 to 2.5 billion years ago) facilitated the development of not only aerobic respiration, but also certain phenotypically unique survival traits, some of which are still distinguishable today. Examples of these traits include resistance to ultraviolet radiation through the use of sun-screening compounds, resistance to oxidative damage through the use of radical-scavenging proteins and the creation of physical barriers, and resistance to desiccation through the synthesis of non-reducing disaccharides and extracellular polysaccharides.

#### 1.1.2 Survival Strategies Developed by Cyanobacteria

Cyanobacteria have developed an array of exceptional qualities and responses to ensure their survival in the harsh climatic conditions in which they grow. The presence of so many unique survival characteristics in some of these organisms has made it difficult for researchers to study only one trait, without taking into consideration the effects of another (Hill *et al.* 1994a; Hill *et al.* 1997; Markillie *et al.* 1999; Mattimore and Battista 1996; Scherer *et al.* 1988; Shirkey *et al.* 2002). The multiple UV-induced responses of *Nostoc commune* are an example. The biosynthesis of sun-screening pigments in this organism alone is insufficient to account for its notable resistance to damaging radiation (Ehling-Schulz *et al.* 2002; Hill *et al.* 1994a; Hill *et al.* 1994b; Scherer *et al.* 1988).

Biosynthesis of a carbohydrate shell or a protective carbohydrate coat is a common cyanobacterial tactic for self-defense against both the environment and other organisms. The peptidoglycan layers of cyanobacteria are considerably thicker and more cross-linked than those found in most other gram-negative bacteria (Hoiczkyk and Hansel 2000; Potts 1996). These layers provide mechanical protection and define the cell shape (Bryant 1994). The glycolipid and external polysaccharide layers surrounding heterocysts in organisms such as *Nostoc commune*, *Nostoc punctiforme*, *Nostoc*

*cordubensis*, and *Anabaena* are thought to serve as a barrier to O<sub>2</sub> that helps protect nitrogen-fixing enzymes from inactivation (Bryant 1994; DePhilippis and Vincenzini 1998; Xu *et al.* 1997). An extracellular polysaccharide (EPS) of *Nostoc commune* DRH1 seems to be effective in excluding other organisms from the interior of *Nostoc commune* colonies, preventing bacterial predation (Hill *et al.* 1994b). The *Nostoc commune* EPS barrier is hypothesized to also function as a ‘water buffer’ (Caiola *et al.* 1996; Shaw *et al.* 2003), and preparations of this material have been used to stabilize human cells during the drying process (Bloom *et al.* 2001; Hill *et al.* 1997).

Several strains of cyanobacteria produce ultraviolet sun-screening compounds in response to high levels of incident radiation, sometimes embedding them in their protective carbohydrate layers. The UV-A/B-absorbing, carbohydrate-linked mycosporine amino acid-like compounds found in the extracellular polysaccharide coating of *Nostoc commune* are an example (Böhm *et al.* 1995; Hill *et al.* 1994a). Various mycosporine amino acids or amino acid-like compounds have been found in fourteen other strains of cyanobacteria, including *Anabaena*, *Synechococcus*, *Gloeocapsa*, *Calothrix*, *Scytonema*, and *Chlorogloeopsis*, though sometimes only as a solute in the cytoplasm (Garcia-Pichel and Castenholtz 1993; Sinha *et al.* 1999). The extracellular sun-screening pigment scytonemin, a lipid-soluble molecule that absorbs at ultraviolet and violet (visible) wavelengths (Sinha *et al.* 1999), is often present in cyanobacteria that contain mycosporine amino acids or mycosporine amino acid-like compounds. The spectral complementation and frequent occurrence of these two types of compounds in the same organism imply that they evolved as a protective response.

Damage by ultraviolet radiation is not always a direct process. UV-A and UV-B can cause oxidative stress by generating reactive oxygen intermediates (Ehling-Schulz and Scherer 1999). Interaction of O<sub>2</sub> with excited triplet-state chlorophyll produces singlet dioxygen (<sup>1</sup>O<sub>2</sub>), which is particularly harmful to polyunsaturated fatty acids and can ultimately lead to membrane damage. Interaction of O<sub>2</sub> with the electron acceptors of photosystem I or reduced ferredoxin can result in the formation of the superoxide radical (O<sub>2</sub><sup>-•</sup>). Although this radical is not very reactive in aqueous solution, it has been implicated in lipid peroxidation, protein inactivation, and the formation of the more damaging hydroxide radical (•OH) (Regelsberger *et al.* 2002). Oxygen (O<sub>2</sub>) is physically

excluded from heterocysts (though not completely) by an outer-membrane glycolipid layer and an external polysaccharide casing, thus preventing the formation of damaging oxygen and hydroxide radicals (Bryant 1994; Murry and Wolk 1989). Carotenoids (light-harvesting pigments), well known for their ability to quench triplet-state chlorophyll, have been reported to respond to changing levels of ultraviolet radiation (Ehling-Schulz and Scherer 1999). Radical-scavenging enzymes, such as superoxide dismutase, can act as another line of defense, decreasing the level of superoxide species by converting it to O<sub>2</sub> and hydrogen peroxide, with the latter then decomposed by ascorbate peroxidase and/or catalase activities (Li *et al.* 2002; Shirkey *et al.* 2002).

### 1.1.3 Cyanobacterial Desiccation Tolerance, A Symphony of Biological Processes

The survival of some microorganisms subjected to the drying and rehydration process is due to the action of several different protective biological mechanisms, not all of which are fully understood (Crowe *et al.* 1992; Potts 1994; Potts 2001). Overproduction of an iron-containing superoxide dismutase (FeSOD) is thought to help *Nostoc commune* survive repeated cycles of desiccation and rehydration, during which the organism is more susceptible to the damaging effects of reactive oxygen species (Shirkey *et al.* 2002). Reliance upon this enzyme, and other protein or protein complexes that are stable after years in a dried state, contributes to the pronounced desiccation tolerance of *Nostoc commune* (Potts 1996; Shirkey *et al.* 2003). Synthesis of a cytoplasmic membrane rich in unsaturated fatty acids in *Nostoc* UTEX 584 maintains membrane fluidity in the dried state, circumventing damage and perhaps cell death upon rehydration (Potts 1996). Production of glass-forming sugars, such as sucrose and trehalose, and the extracellular polysaccharide coating surrounding *Nostoc commune* helps this organism survive the drying process by physically replacing water and maintaining membrane integrity (Hill *et al.* 1997; Shaw *et al.* 2003). Clearly, cyanobacteria have evolved numerous and overlapping protective strategies to survive a wide variety of environmental stresses.

#### 1.1.4 Sorting It All Out

As indicated by attempts to understand cyanobacterial resistance to dehydration or ultraviolet radiation, many of these survival strategies are interrelated. The same cyanobacterial response (*e.g.*, up-regulation of SOD) might protect the organism from two different stresses (oxidative and UV), and two different responses might be elicited by a single stress (*e.g.*, up-regulation of both EPS and sunscreens by UV) (Ehling-Schulz *et al.* 2002; Potts 1996). Indeed, several such examples have been noted. The over-produced FeSOD is thought to act as a second line of defense against UV-induced oxidative damage during repeated cycles of desiccation and rehydration in *Nostoc commune*, behind the primary shield of sun-screening compounds (Scherer *et al.* 1988; Shirkey *et al.* 2002; Shirkey *et al.* 2003). A mutant strain of *Synechococcus* (PCC 7942), lacking FeSOD activity, was shown to be more sensitive to both oxidative stress and moderate chilling (Thomas *et al.* 1999). It has been hypothesized that the resistance to ionizing radiation of *Deinococcus radiodurans*, *Acinetobacter radioresisten*, and *Chroococcidiopsis* is actually a by-product of their desiccation tolerance. This idea is supported by the observations that mutants of *Deinococcus radiodurans*, which lack the ability to repair DNA, are sensitive to both desiccation and radiation stresses, and that a *Nostoc* strain with a marked tolerance for desiccation displays a tolerance to radiation similar to that of *Deinococcus radiodurans* (Billi *et al.* 2000a; Jawad *et al.* 1998; Markillie *et al.* 1999; Mattimore and Battista 1996; Shirkey *et al.* 2003).

As increasing interest develops in transferring desirable traits among organisms (*e.g.*, to make industrially-important microbes more resistant to environmental stresses), it becomes apparent that, for some of these qualities, it is not going to be as simple as cloning a gene and expressing it in the organism of interest. Frequently, more than a single characteristic (or gene) contributes to an exceptional capacity observed in an organism, and an interrelated network of responses may need to be deciphered in the model organism to transfer that trait to another.

## **1.2 *Nostoc commune*, a Model for Prokaryotic Anhydrobiosis**

### 1.2.1 The Scientific Significance of Extremophiles

Organisms that thrive in environments that are inhospitable to most other organisms represent a potential source of wealth of scientific information because of the unique characteristics that allow them to survive. Organisms labeled *extremophiles* are known to tolerate extremes in temperature, pH, salinity, water deprivation, and radiation exposure, to name a few (Eichler 2001a).

Industrial interest in extremophiles has developed in recent years with the realization that their proteins and secondary metabolites have potential economic value. Unusually stable enzymes, sometimes called *extremozymes*, open doors for biotechnological applications in situations previously not feasible, and can provide insight to potential advances in protein design (Madigan and Marrs 1997). A landmark example is the use of the *Thermus aquaticus* (*Taq*) DNA polymerase, in the automated polymerase chain reaction (PCR) (Chien *et al.* 1976).

### 1.2.2 Transfer of Desiccation Survival Strategies

A variety of organisms can survive the stress of extreme water deprivation, often by entering a state of ‘suspended animation,’ which is sometimes referred to as *anhydrobiosis* (Hoekstra *et al.* 2001). Although a complete understanding of the survival of anhydrophiles under water deficit is lacking, desiccation tolerance is often correlated with intra- and/or extra-cellular levels of non-reducing disaccharides, such as sucrose and trehalose. Even the exogenous addition of one or both of these sugars to desiccation-sensitive cells can impart some additional capacity for survival (Crowe *et al.* 1992; Leslie *et al.* 1995; Leslie *et al.* 1994).

As water is removed from an organism’s cellular membrane, its phospholipid bilayer undergoes a phase transition from a liquid-crystalline to a gel-like state. The consequent rearrangement of phospholipid head groups increases van der Waals interactions between fatty acid chains, that are accompanied by an increase in the phase transition temperature of the membrane (Crowe *et al.* 1992).

Cellular damage or death can occur if the bilayer is rehydrated abruptly (*e.g.*, by being immersed in water). During rehydration, phase transition to the liquid-crystalline

state must occur, and in such a case (abrupt rehydration) the gel and liquid-crystalline phases coexist transiently. This coexistence leads to phase separation within the membrane, which allows cellular contents to leak, and can lead to cell death.

One of the mechanisms by which non-reducing disaccharides aid in the survival of an organism under dry conditions is described by *The Water Replacement Hypothesis* (Crowe *et al.* 1992). When a phospholipid bilayer is dried, with the exogenous addition of a non-reducing disaccharide, the disaccharide occupies space near the phospholipid head group that was previously held by water molecules. The bulky disaccharide prevents phospholipid transition to the gel-like state when water is removed, and the leakage of cellular contents upon rehydration is circumvented (Figure 1.1).

Non-reducing disaccharides also augment the desiccation tolerance of an organism by preventing membrane fusion during the drying process. A cell which has had most of its intracellular water removed will collapse into a curved shape, resembling that of a contact lens. The radius of curvature at the edge of this shape increases, resulting in a decreased packing density of the phospholipids in this area, and an increased phospholipid acyl chain exposure to water. The number of consequent thermodynamically unfavorable interactions is minimized when two of these dried cells fuse into a single phospholipid mass, but this process also destroys the organism. The cellular fusion process continues between the resulting cell aggregate and other dried (and perhaps still viable) cells, or between the resulting cell aggregate and other cell aggregates, ultimately producing single phospholipid body. This phenomenon is circumvented when samples are dehydrated in the presence of trehalose, or approximately three times as much sucrose, although the mechanism of this inhibition is not understood (Crowe *et al.* 1983).

Preparations of the *Nostoc commune* EPS, when added in small amounts, have been shown to drastically decrease the amount of non-reducing disaccharide that is required to prevent liposome fusion (Hill *et al.* 1997). Additionally, this material has stabilized human kidney cell line 293H to the drying and rehydration process, when transfection of this cell line with the sucrose-6-phosphate synthase gene was alone insufficient to do so (Bloom *et al.* 2001).

The preservation of biomedically-relevant cell lines can be of extreme importance, thus a substantial amount of effort has been invested in studying cellular preservation by freezing and drying. Both of these processes damage cells in the same manner and, in both cases, the use of non-reducing disaccharides can help a cell resist damage and maintain viability. Using an adenoviral vector to transfer *Escherichia coli* trehalose synthase genes (*otsA/otsB*) to human primary fibroblasts, it was demonstrated that techniques used by anhydrophiles to survive drying and rehydration can be applied to other cell types (Guo *et al.* 2000). The cyanobacterial sucrose phosphate synthase gene has been transferred to both *Escherichia coli* (Billi *et al.* 2000b) and human kidney cell line 293H (Bloom *et al.* 2001). In both cases the capacity of the transfected cells for survival increased, but in the latter case only with the aid of the *Nostoc commune* extracellular polysaccharide (EPS).

### 1.2.3 Glycosyltransferases, A Brief Overview

Oligosaccharides are synthesized by glycosyltransferases, which are enzymes that attach one carbohydrate constituent to another, or to some other acceptor, in a stereospecific manner. These enzymes are highly specific with respect to both the site of carbohydrate addition and the anomeric configuration of the acceptor. There is a similar, but sometimes less stringent, requirement for the carbohydrate donor. Donors are most often nucleotide diphosphate monosaccharides, though examples of nucleotide monophosphate monosaccharides, nucleotide diphosphate disaccharides, and glycosylphosphates are known (Campbell *et al.* 1997; Saxena *et al.* 1995). Typically, a single glycosyltransferase catalyzes the formation of only a single glycosidic linkage, however, there are exceptions to this rule (DeAngelis 2002).

Depending upon the stereochemistry of the final product (after carbohydrate addition), glycosyltransferases are classified as either *invertin*g or *retainin*g. A *retainin*g glycosyltransferase commonly attaches a carbohydrate donor by way of a glycosylated enzyme intermediate. *Invertin*g glycosyltransferases often form the donor-acceptor bond directly in a S<sub>N</sub>2 reaction, inverting the configuration of the donor's anomeric carbon in the process (Figure 1.4).

The genes encoding the glycosyltransferases responsible for assembling a particular extracellular bacterial polysaccharide or glycoconjugate are often associated in 10-15 kDa gene clusters. Such clusters also encode elements involved in the regulation of EPS production levels, EPS chain-length determination, EPS polymerization, EPS export, and glycosyltransferase donor modification (Levander *et al.* 2002; Stingele *et al.* 1996; Stingele *et al.* 1999a) (Figure 1.5). Such clusters exist in both chromosomal DNA and on extra-chromosomal plasmids (Gotschlich 1994; Jennings *et al.* 1995; Kolkman *et al.* 1997; Morona *et al.* 1995; Stingele *et al.* 1996; van Kranenburg *et al.* 1999a; van Kranenburg *et al.* 1999b). It is sometimes hypothesized that the gene products of these groupings exist in complexes, but this grouping has been shown in only a few cases (DeAngelis 2002).

Classically, glycosyltransferase function has been demonstrated by reacting either membrane fragments or permeabilized cells with a radiolabeled donor or acceptor. Separation of the reaction products is then carried out by either thin-layer or paper chromatography, and the products are visualized by either exposure to x-ray film or radiography (Stingele *et al.* 1999a; van Kranenburg *et al.* 1999b). More recently, genetic techniques such as gene deletion have been employed, and glycosyltransferase function was assigned by the examination of gene-knockout phenotypes (Gotschlich 1994; van Kranenburg *et al.* 1999b). A fluorescent-labeled synthetic carbohydrate acceptor/capillary electrophoresis-based assay has also been used to demonstrate glycosyltransferase activity (Wakarchuk *et al.* 1996). Glycosyltransferase functions have been assigned by structural determination of the fluorescent-labeled glycosyltransferase reaction product produced with this method (Gilbert *et al.* 2000; Wakarchuk *et al.* 1996). Sequence comparison (Stingele *et al.* 1996) has been used to infer glycosyltransferase function, but this approach is often unreliable due to the low levels of sequence similarity observed among glycosyltransferases, even when comparing genes of similar (or even the same) function within a single species (Kapitonov and Yu 1999).

Glycosyltransferase purification is usually not a trivial task, in part because these enzymes are often membrane associated (Breton *et al.* 1998). Additionally, several hundred glycosyltransferases are postulated to exist in a single living cell (Raju *et al.* 1996). In view of these difficulties, researchers often opt for a recombinant strategy



(Faik *et al.* 2000; Kumar *et al.* 1990; van Kranenburg *et al.* 1999b) for the production of active protein, when adequate genetic information is available. Alternatively, a *partial* purification approach can be utilized when working backward from the active protein to the gene (Perrin *et al.* 1999). When an attempt is made to purify a glycosyltransferase from a membrane-bound native source, a detergent extract is produced and an activity assay is utilized to follow the enzyme through successive purification steps (Leiter *et al.* 1999). When there is no activity assay, affinity-labeling techniques using radiolabeled donor analogs can be employed (Delmer and Read 1987; Frost *et al.* 1990; Zeng *et al.* 1997).

#### 1.2.4 *Nostoc commune*, An Organism With Something to Offer

The cyanobacterium *Nostoc commune* (Figure 1.2) is an anhydrophile that can remain viable for decades in a dry state as a crusty, brittle material and then regain measurable metabolic activity within ten minutes after rewetting (Scherer *et al.* 1986). This organism produces profuse amounts of an extracellular polysaccharide (EPS) that has been shown to help prevent the fusion of artificial membranes upon drying (Hill *et al.* 1997). This EPS coating has also been shown to increase the survival capacity of desiccation-sensitive human cells when added exogenously to samples (Bloom *et al.* 2001). The large amounts of EPS produced by this organism (70% of the total biomass of field material), and its demonstrated ability to stabilize desiccation-sensitive cells, warrant further study of the extracellular polysaccharide production process of *Nostoc commune*.

*Nostoc commune* can survive for long periods of time in a dried state, thus a fair amount of effort has been invested in studying this organism (and the mechanisms that it employs) as a prokaryotic model for *the anhydrobiotic cell* (Angeloni and Potts 1986; Potts 1996; Wright *et al.* 2001). The cytoplasmic membrane of *Nostoc commune* UTEX 584 was examined to determine whether this organism possesses any unusual biochemical or structural features that allow it to withstand acute water stress (Olie and Potts 1986). It was observed that the membrane possessed an unusually high concentration of 20:3 $\omega$ 3 fatty acids. This feature is suspected to contribute substantially to maintaining membrane fluidity in the dried state (Potts 1996). Just a few years after

this finding, it was reported that a group of acidic proteins were released into the cell media (without cell lysis) when *Nostoc commune* was subjected to repeated cycles of drying and rehydration (Scherer and Potts 1989). At the same time, extracellular UV-A/B absorbing pigments were detected in this organism (Scherer *et al.* 1988). Later, it was reported that the released proteins and the UV-absorbing pigments formed non-covalent, salt-dependant associations (Hill *et al.* 1994a). Electron microscopic and biochemical characterization of *Nostoc commune* DRH1 (an isolate of *Nostoc commune* CHEN) was performed (Hill *et al.* 1994b), and at least one of the UV-A/B-absorbing pigments from *Nostoc commune* Vaucher was shown to be an OS-MAA (oligosaccharide-linked mycosporine amino acid), a variant of a compound biosynthesized for use as an ultraviolet sunscreen by other organisms (Böhm *et al.* 1995). The extracellular polysaccharide (EPS) of *Nostoc commune* DRH1 helped non-reducing disaccharides inhibit the fusion of artificial liposomes (a model system) during the drying process, supporting the idea that this carbohydrate structure did indeed play a role in desiccation tolerance, though not its suspected role of suppressing membrane phase transition temperature (Hill *et al.* 1997). The average composition of this carbohydrate structure was later determined and is shown in Figure 1.3 (Helm *et al.* 2000). Recently, the responses of this organism to UV-B radiation exposure were studied. Separate shock (immediate) and acclimation (long-term) responses were noted (Ehling-Schulz *et al.* 2002).

#### 1.2.5 Finding and Following a Functional Enzyme

Identification of the gene products in *Nostoc commune* responsible for EPS production and export would facilitate the transfer of this capacity to another organism. The recipient strain might then also be endowed with an EPS coating and increased resistance to desiccation. Alternatively, transfer of these genes to a more rapidly growing host would provide the quantities of raw material necessary for the stabilization of desiccation-sensitive cell lines.

A genetic strategy could potentially be used to identify the EPS-producing enzymes of *Nostoc commune* (*e.g.*, probing a genomic library with EPS-producing enzyme homologs), but the number of these strategies that could be applied to *Nostoc*

*commune* DRH1 was limited at the time of this project's initiation. The EPS coating surrounding *Nostoc commune* defeats attempts to introduce exogenous DNA into this organism, eliminating the possibility that a transposon mutagenesis approach could be used. Also, a genomic library had not yet been successfully produced, even though attempts to do so had been underway for several years. The low sequence similarity of glycosyltransferase genes makes the design of oligonucleotide primers that could replicate a glycosyltransferase gene from genomic DNA a difficult task. This lack of sequence similarity would also hinder successful identification of a glycosyltransferase gene PCR-product by homology.

An alternative approach to this problem would be to purify one (or more) of the glycosyltransferase enzymes responsible for EPS production, and to obtain sequence information from it (or them). To purify one of these EPS-producing enzymes, a means of monitoring its presence throughout a purification scheme is required. This strategy would necessitate the development of an activity assay specific for the glycosyltransferase enzymes of *Nostoc commune*. Such a strategy would offer several advantages. Once enzymatic activity is successfully detected, there is no question of whether a functional gene-product can be obtained for *in vitro* study, as might not be the case when beginning with only a putative glycosyltransferase ORF. Information about the state and/or location of the functional protein is obtained during protein purification. Even if the protein (or proteins) is very labile and active only for a brief time *in vitro*, that time may be sufficient to label the enzyme so that it can still be purified after loss of activity. If lability is less of a problem, the active protein or protein complex (from the native source) can be studied and characterized, providing information that will aid in purification attempts. When a relatively pure protein has finally been obtained (*e.g.*, as a spot on a two-dimensional gel) and identified, characteristics of the active protein, such as cleaved signal sequences and post-translational modifications can be studied using mass spectrometry.

As with any strategy, there are also pitfalls in adopting this type of approach. The initial detection of an activity is sometimes more easily said than done, and then knowing that such activity is the correct one to pursue is essential. Purification of such an enzyme from the native source is complicated by the fact that *Nostoc commune* DRH1 does not

grow rapidly in BG11<sub>0</sub> media. Different growth conditions have been observed to alter the EPS composition of this organism, and thus the expression of the glycosyltransferases responsible for its biosynthesis. Furthermore, one must also bear in mind that an activity detected *in vitro* may or may not correspond to the same activity *in vivo*. Additionally, peptide (and thus protein) identification without a known genome can be difficult. Amino acid sequences determined from the mass spectrum of a peptide are sometimes open to interpretation, though protein identification by *de novo* peptide sequencing has been done with some degree of success for *Nostoc commune* DRH1 using the sequenced genomes of *Nostoc punctiforme* ATCC 29133 and *Anabaena (Nostoc)* PCC 7120 (Ehling-Schulz *et al.* 2002).

#### 1.2.6 Specific Aims

The preceding discussion described the ability of the *Nostoc commune* EPS to stabilize cells during the desiccation and rehydration process, and the potential benefits of knowing the genes responsible for encoding the enzymes that synthesize this material. Thus, the overall objective of this line of research is to transfer the capacity to produce the *Nostoc commune* EPS to another bacterial organism. This goal can be rationalized as attainable in view of the fact that similar transfers of EPS-producing capacity have been performed using other bacterial organisms (Germond *et al.* 2001; Stinglele *et al.* 1999b). This work is based upon the central hypothesis that the *Nostoc commune* EPS-producing genes are also clustered in a single genetic locus, or a small number of loci, which will facilitate their recombinant transfer.

A short-term goal of this project is addressed in the first portion of this work, which is devoted to obtaining amino acid sequence information that will facilitate the identification of these genes and the putative genetic locus or loci in which they reside. The sub-tasks required to achieve this goal were defined as the Specific Aims of this portion of the project:

- 1) Develop methods that can be used to detect glycosyltransferase activity in *Nostoc commune* DRH1.

- 2) Apply these methods to detect the glycosyltransferase activities responsible for EPS production in the organism.
- 3) Develop activity-based purification protocols to identify the proteins responsible for EPS biosynthesis.
- 4) Sequence the identified proteins by tandem mass spectrometry.

### **1.3 SLH Domain-Containing Proteins In *Anabaena (Nostoc) PCC 7120***

#### 1.3.1 Bacterial Protein Glycosylation

Glycobiology is generally regarded as the study of carbohydrates in biological systems. To date, considerably less progress has been made in the field of glycobiology than in the fields of protein or nucleic acid chemistry, in part because glycoconjugates are so structurally complex and technically difficult to study (Roseman 2001). For years, biological carbohydrate research was de-emphasized or overlooked; carbohydrates were assumed by many to be only structural elements of the cell. This view has since changed with the realization that minor differences in carbohydrate structures can have huge impacts on biological function (Jonietz 2001).

Though now demonstrated to be present in all domains of life, and in viral organisms, protein glycosylation was previously thought to be non-existent in prokaryotes (Upreti *et al.* 2003). The first detailed example of a prokaryotic glycoprotein was a surface-layer (S-layer) protein from *Halobacterium halobium* in 1976 (Mescher and Strominger 1976). Since that time, approximately 70 bacterial glycoproteins have been discovered, most of which are from the archaeal domain (Upreti *et al.* 2003).

#### 1.3.2 Characterization of Bacterial Glycoproteins

A bacterial protein cannot be designated as a *glycoprotein* without adequate evidence to support the hypothesis that it contains a covalent carbohydrate-protein linkage. Due to the abundance of glycans in most prokaryotic cell walls and the presence of reserve polysaccharide within the cell, extensive purification of a putative bacterial glycoprotein is normally required before its study or analysis (Moens and Vanderleyden 1997; Schaffer *et al.* 2001). SDS-PAGE separated sample bands from a bacterial lysate that respond to both protein and carbohydrate staining might only be the result of tight non-covalent protein/carbohydrate associations or LPS contamination (Upreti *et al.* 2003). Ultimately, evidence of a carbohydrate-linked amino acid or peptide is necessary to confirm that a bacterial protein is actually a true *glycoprotein*, and this evidence can be produced by chemical degradation experiments, mass spectrometry, or NMR analyses.

Classically, the existence of a carbohydrate-protein linkage has been demonstrated by protease digestion of the putative glycoprotein (*i.e.*, using trypsin or

PronaseE), purification of the resulting glycopeptides, and their subsequent chemical analyses (Mescher and Strominger 1976; Paul *et al.* 1986; Wieland *et al.* 1983). In more recent times, researchers have favored mass spectrometric techniques to identify glycosylation sites because of their lower sample requirements and increased sensitivity. Though it is difficult to identify glycosylated peptides in a complex digestion mixture and glycopeptide ionizations can be suppressed by their carbohydrate constituents (Mechref and Novotny 2002), mass spectrometry has an advantage in that the sequence context of each carbohydrate attachment is preserved (Huddleston *et al.* 1993). Carbohydrate structures can be deduced from the mass spectra of glycopeptides as well (Håkansson *et al.* 2001; Kawasaki *et al.* 2003). Techniques such as precursor/parent ion scanning (Ritchie *et al.* 2002; Wilm *et al.* 1996), isotopic labeling (Harvey 2001), and glycopeptide enrichment using lectin-affinity chromatography have been developed, specifically to overcome the disadvantages of mass spectrometry in glycopeptide analysis.

### 1.3.3 Functions of Bacterial Protein Glycosylation

Although the presence of bacterial glycoproteins as structural elements of the cell envelope is the most commonly encountered example, distinct functions of several other prokaryotic glycoproteins, present at different locations within the cell, have been noted or proposed (Schaffer *et al.* 2001). A normally unglycosylated *Bacillus*  $\beta$ -glucanase was found to be glycosylated when expressed in *Saccharomyces cerevisiae*, producing an enzyme with a higher thermotolerance. Expression of the same *Bacillus*  $\beta$ -glucanase gene in *Escherichia coli* produced an unglycosylated protein that lacked increased heat resistance. Glycanases of *Cellulomonas fimi* are still active when their unglycosylated forms are produced in *Escherichia coli*, but these enzymes show a greater susceptibility to proteolysis and a decreased affinity for cellulose when compared to the glycosylated variety (Creagh *et al.* 1996; Ong *et al.* 1994). Glycosylation has been shown to be a necessary component of a *Bacillus thuringiensis* protein that exhibits larvacidal activity, and the sugar constituents of *Pseudomonas syringae* lipoglycoprotein complexes are thought to be necessary for their ice-nucleating function (Moens and Vanderleyden 1997). Additionally, bacterial glycoproteins have been speculated to play roles in cell-

cell interactions/adhesion and surface-intracellular interactions, typically by comparison to their eukaryotic counterparts (Messner *et al.* 1990).

#### 1.3.4 Bacterial Surface Layer (S-Layer) Proteins

The most studied examples of bacterial glycoproteins are the S-layer proteins (Moens and Vanderleyden 1997). Many Bacteria and Archaea possess a S-layer, which is a regularly structured protein (or glycoprotein) layer around (or attached to) the outermost portion of their cell wall. When an extracellular mucilaginous coating is also present, the S-layer is located beneath the coating (Figure 1.6) (Smarda *et al.* 2002). S-layers are composed of identical (or a small number of) protein or glycoprotein subunits that assemble into two-dimensional crystalline arrays. These arrays associate with each other, and the underlying cell envelope layer (peptidoglycan, outer membrane, or lipopolysaccharide), by non-covalent interactions (Sára *et al.* 1998a; Sára *et al.* 1998b). Such structures are thought to represent the simplest form of a biological membrane, and S-layer proteins are the most abundant of prokaryotic cellular proteins (Sleytr *et al.* 2001). S-layer proteins have been used as taxonomic criteria for differentiation between Bacteria and Archaea, although they are now recognized as being present in both (Schaffer *et al.* 2001). In the last compiled index (1996), there were 200 taxa of Bacteria and 79 taxa of Archaea possessing S-layers. S-layers have been known in cyanobacteria since 1972, and have been observed in 60 isolates of 23 different species (Smarda *et al.* 2002).

Several S-layer proteins have been found to possess a conserved sequence termed the SLH domain (S-layer homology domain) (Lupas *et al.* 1994). The SLH domain is usually comprised of three tandem repeats of a motif formed by 50-60 amino acids, or one copy of the motif followed by a coiled-coil domain (more typical of cyanobacteria), that is located towards the N-terminal end of S-layer proteins (Ilk *et al.* 1999; Smarda *et al.* 2002). SLH sequences are strongly divergent, showing an average identity of only 27%, with a single glycine residue universally conserved (Figure 1.7) (Lupas *et al.* 1994). The SLH domain is believed to anchor proteins to the cell surface, and evidence has been found to support its hypothesized role in peptidoglycan-binding (Lemaire *et al.* 1995; Sára *et al.* 1998a). Studies of several gram-positive bacteria indicate that an embedded,



HF-extractable secondary cell wall polymer (SCWP) functions as the SLH domain binding site in these organisms, rather than the primary peptidoglycan polymer (Brechtel and Bahl 1999; Chauvaux *et al.* 1999; Ilk *et al.* 1999; Sleytr *et al.* 2001). Treatment of a cell lysate with lysozyme facilitates the extraction of S-layer proteins, and bound SCWP has been shown to make these proteins water-soluble and protect them against proteolytic degradation. The binding mechanisms of SLH domains to SCWP have been compared to polysaccharide-lectin type interactions (Sára *et al.* 1998a). Proteins in which the SLH domain has been proteolytically-cleaved lose the ability to bind peptidoglycan/SCWP preparations, though not the ability to self-assemble (Lemaire *et al.* 1995). All proteins that contain SLH domains also contain signal sequences that are cleaved when these proteins pass through the cytoplasmic membrane, which has been speculated to be the site of their glycosylation as well (Eichler 2001b; Moens and Vanderleyden 1997; Smarda *et al.* 2002). Many of these proteins have been found to be cell wall-associated and, in some instances, proteins containing these domains are released into the culture medium (Lupas *et al.* 1994).

The SLH domain is not limited to S-layer proteins; it has also been found in various cell-associated exoenzymes and extracellular proteins, sometimes very close to the C-terminal end (Sára *et al.* 1998a). SLH domains occur in several extracellular carbohydrate-modifying enzymes including a xylanase from *Clostridium thermocellum*, an alkaline cellulase from *Bacillus* sp. KSM-635, an endoxylanase from *Thermoanaerobacter saccharolyticum*, an  $\alpha$ -amylase-pullulanase from *Bacillus* sp. XAL601, and a pullulanase from *Thermoanaerobacterium thermosulfurigenes* EM1 (Lemaire *et al.* 1995). This motif has also been identified in two pore-forming outer-membrane proteins of *Synechococcus* PCC 6301 (Hansel and Tadros 1998). The C-terminal SLH-domain of a *Clostridium thermocellum* protein was fused to an *Escherichia coli* maltose-binding protein (MalE) and found to confer the ability to bind peptidoglycan preparations to the chimera (Lemaire *et al.* 1995).

Glycosylation of S-layer proteins are frequently observed, though less so in Bacteria than Archaea (Schaffer *et al.* 2001). Bacterial S-layer glycan chains are fairly diverse in size, monosaccharide constituents, and types of linkages that they employ. In general, they are composed of two to six monosaccharide repeating units with an overall

degree of polymerization of up to fifty units. Archaea S-layer glycan chains are usually shorter and they lack repeats (Schaffer *et al.* 2001; Sleytr *et al.* 2001). Another distinct feature of bacterial S-layer proteins is the type of glycan attachment that they employ. Bacterial proteins are predominantly linked to glycan moieties by *O*-glycosidic linkages, whereas archaeal proteins more frequently contain *N*-linked carbohydrate chains, though both types of linkages have been found in the same S-layer protein (Schaffer *et al.* 2001). Very little is known about the biological function of S-layers, and even less is known about the function of their covalently linked carbohydrate attachments (Messner *et al.* 1990).

Even though the biological roles of S-layer proteins are still uncertain in most cases, they have been hypothesized to fulfill a variety of different functions (Beveridge *et al.* 1997). S-layers are thought to act as frameworks that help determine cell shape, and to perhaps aid in cell division (Sára *et al.* 1998a). They have been speculated to play roles in cellular adhesion and surface recognition, as well as to act as protective coatings and to provide adhesion zones for exoenzymes (Engelhardt and Peters 1998; Lupas *et al.* 1994). S-layer protein assemblies can also act as molecular sieves and molecule traps (Schaffer *et al.* 2001; Smarda *et al.* 2002).

Applied S-layer research has been directed towards the exploitation of the unique properties of these proteins/glycoproteins (Schaffer *et al.* 2001; Sleytr 1997). Genetically engineered S-layer glycoproteins hold high potential for vaccine development and manufacture; such proteins are exported from the cell, facilitating purification efforts, especially if the desired end product is a chemically cleaved carbohydrate antigen (Jahn-Schmid *et al.* 1996a; Jahn-Schmid *et al.* 1996b). S-layer proteins have been used to immobilize enzymes on solid supports (*i.e.*, in solid-phase immunoassays and biosensors) by attaching the enzyme(s) of interest to an activated carbohydrate component of an ordered S-layer glycoprotein array (Castner and Ratner 2002; Mader *et al.* 2000). S-layer proteins have been used as templates for the formation of perfectly ordered metal and semi-conductor nanoparticle arrays that are required for nanoelectronics and non-linear optics (Douglas *et al.* 1992; Sleytr *et al.* 2001), and they may soon have applications in ultra-filtration and drug delivery/targeting systems (Patel and Sprott 1999; Schuster *et al.* 2001; Upreti *et al.* 2003).

### 1.3.5 Cyanobacterial Genomics and Phylogenetics

The completion of the *Synechocystis* sp. PCC 6803 genome sequence changed the way that function and regulation of cyanobacterial genes are studied (Kotani *et al.* 1994). Because of this effort, great strides have been made in the understanding of the biological processes that take place within this organism. Another cyanobacterium, *Anabaena* (*Nostoc*) PCC 7120, has long been used to study the genetics and physiology of cell differentiation, pattern formation, and nitrogen fixation. Because cellular differentiation can be induced and techniques for genetic manipulation of this organism are available, it too was sequenced (Kaneko *et al.* 2001), creating additional assets for the delineation of poorly understood biological processes (*e.g.*, cell differentiation) in these relatively simple model organisms. Direct comparison of the two genomes permits researchers to focus on the unique characteristics of each (Kaneko *et al.* 2001).

Though there has been some debate of whether the two organisms should be classified in separate genera (Tamas *et al.* 2000), PCC 7120 has recently been reassigned from the *Anabaena* to the *Nostoc* genus, a decision supported by phylogenetic analyses using the *nifD* gene sequence (Henson *et al.* 2002). In this work, PCC 7120 will frequently be designated as *Anabaena* (*Nostoc*) PCC 7120 to avoid confusion (sequenced PCC 7120 ORFs are labeled as *all*- and *alr*-, depending on which strand they are found).

A search of the *Anabaena* (*Nostoc*) PCC 7120 genome (Chapter 7.8.1) yields fifteen proteins that contain the SLH domain, three of which appear to be S-layer associated exoenzymes. Although no studies of cyanobacterial S-layer proteins analogous to those undertaken in bacteria have been performed to date, our working hypothesis (and that of other researchers) is that bacterial results can be generalized to cyanobacteria (Smarda *et al.* 2002).

### 1.3.6 An Interesting Observation

During the course of our work with *Nostoc commune* DRH1, a comparative study with *Anabaena* (*Nostoc*) PCC 7120 was performed. In these experiments, it was observed that a SDS-PAGE gel of a soluble protein extract from PCC 7120 contained polypeptides of ~55-110 kDa that produced a positive response to a carbohydrate stain (Figure 1.8). A ladder-shaped pattern, such as the one shown in Figure 1.8, is often

indicative of a protein that is glycosylated to varying extents, which is a common product of the bacterial glycoprotein biosynthetic process. Mass spectrometric analyses of some of the major protein bands from this carbohydrate-staining region revealed the presence of several proteins of unknown function that contained the SLH domain. The migration behavior of these proteins suggested that a protein/carbohydrate association (covalent or non-covalent) was taking place. The open reading frames (ORFs) encoding these proteins contain elements identifiable as signal peptides and transmembrane domains. Together, these observations suggested the presence of a membrane-associated glycoprotein or a non-covalent carbohydrate/protein complex in the *Anabaena (Nostoc)* PCC 7120 soluble protein extract. Efforts were undertaken to determine which of these phenomena were being observed by electrophoresis and carbohydrate staining.

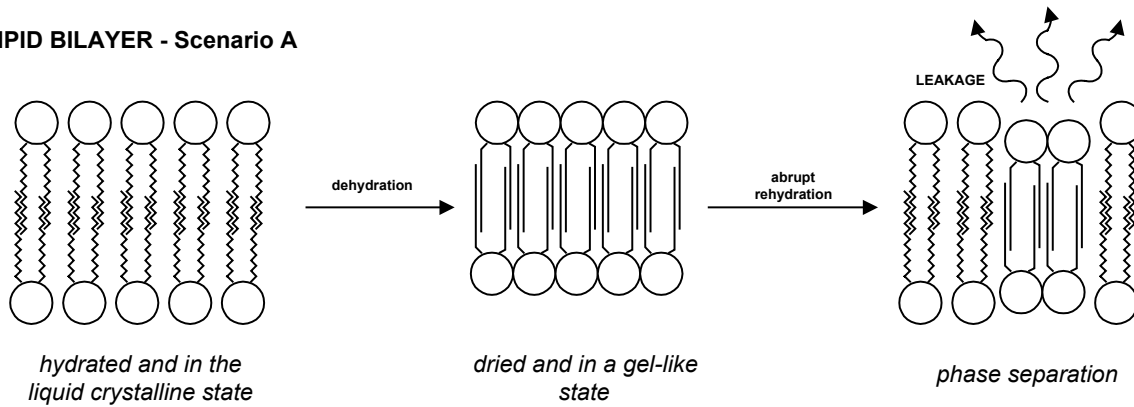
#### 1.3.7 Specific Aims

Based upon the observations described in Section 1.3.6, the overall objective of this portion of the work is to determine whether a membrane-associated glycoprotein or a non-covalent carbohydrate/protein complex is responsible for the staining pattern shown in Figure 1.8. The protein and carbohydrate analysis techniques that will be used to achieve this objective are well established. The Specific Aims of this portion of the project were chosen to achieve this overall objective and to provide additional information about the staining entity (or entities). The Specific Aims of this portion of the project are to answer the following questions:

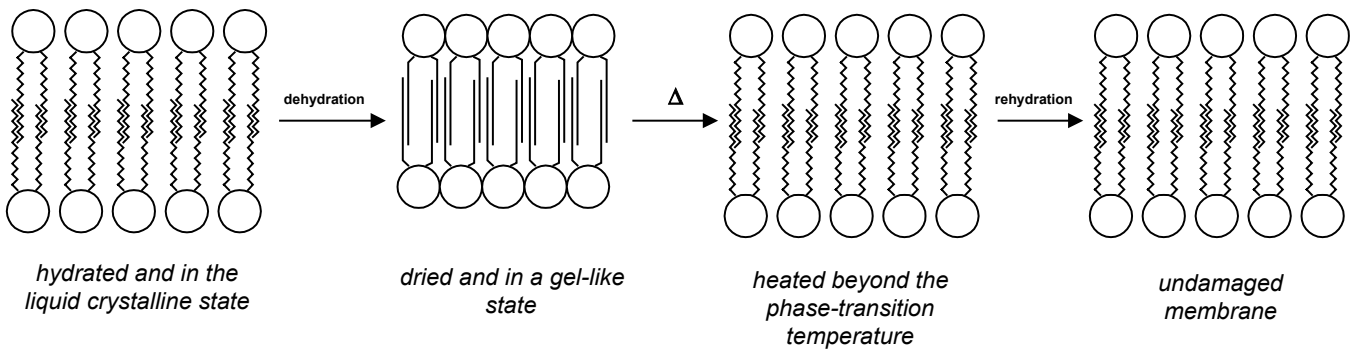
- 1) Are glycosylated S-layer (or at least SLH domain-containing) proteins present and responsible for this staining pattern?
- 2) Which of these proteins are glycosylated and/or interacting with carbohydrate material?
- 3) Can the function of these proteins be deduced by computerized sequence and structure analyses?

- 4) From where does the carbohydrate material that is hypothesized to be interacting with or attached to the SLH domain-containing proteins originate?

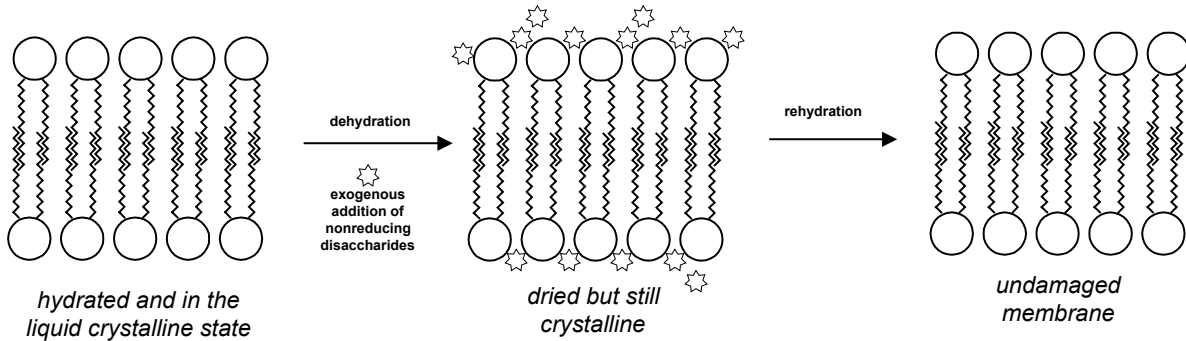
### LIPID BILAYER - Scenario A



### LIPID BILAYER - Scenario B

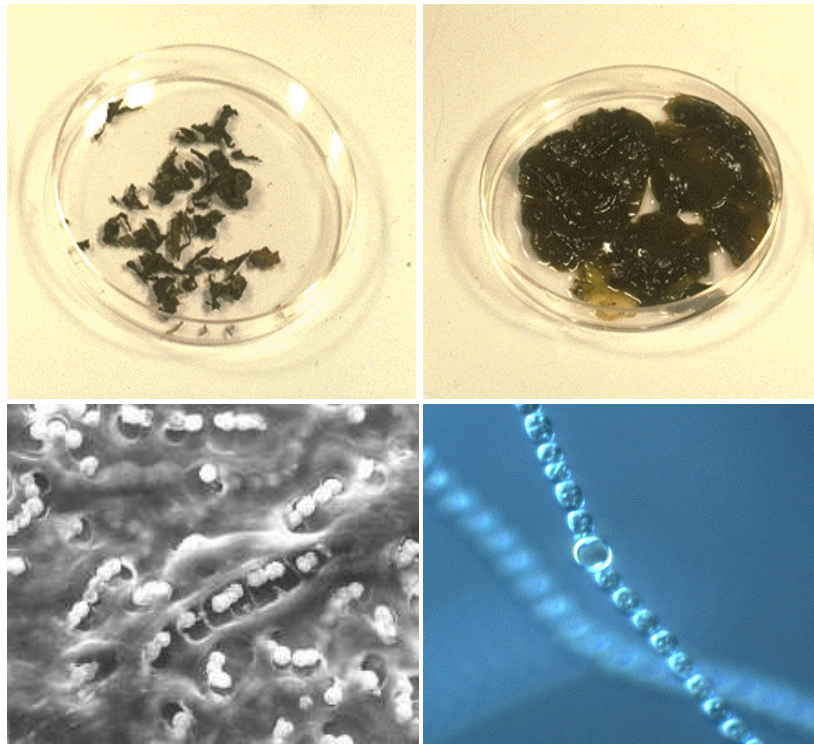


### LIPID BILAYER - Scenario C



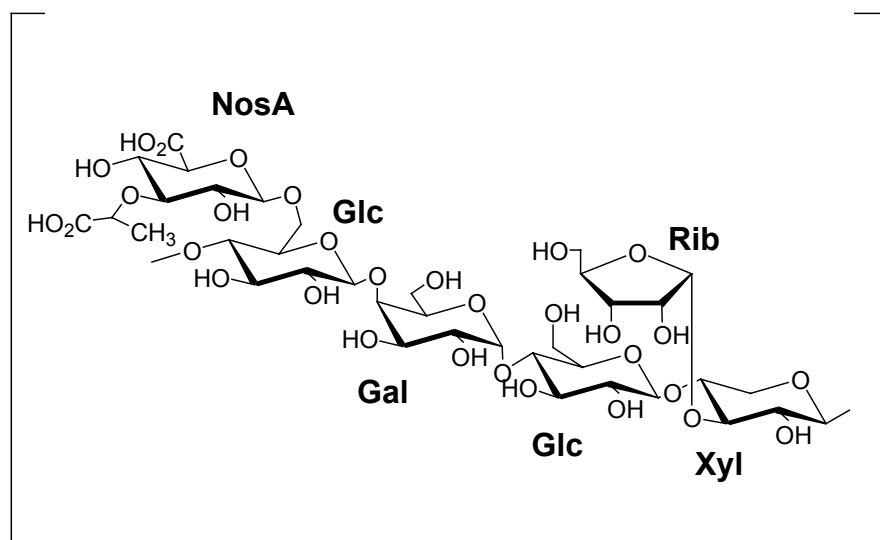
**Figure 1.1 Damage and Preservation of Phospholipid Bilayers During Drying**

Mechanisms by which phospholipid bilayers are damaged or preserved during the drying and rehydration processes. *a)* Drying and rapid rehydration of a desiccation-sensitive cell causes phase separation within the cell membrane, allowing the cell's contents to leak, oftentimes leading to cell death. *b)* Inducing the phase transition to the liquid crystalline state prior to rehydration prevents such damage, which is why it is recommended that bakers' yeast be rehydrated with *warm* water. *c)* Damage can also be averted by preventing the bilayer from ever experiencing a phase transition.



**Figure 1.2** The Cyanobacterium *Nostoc commune*

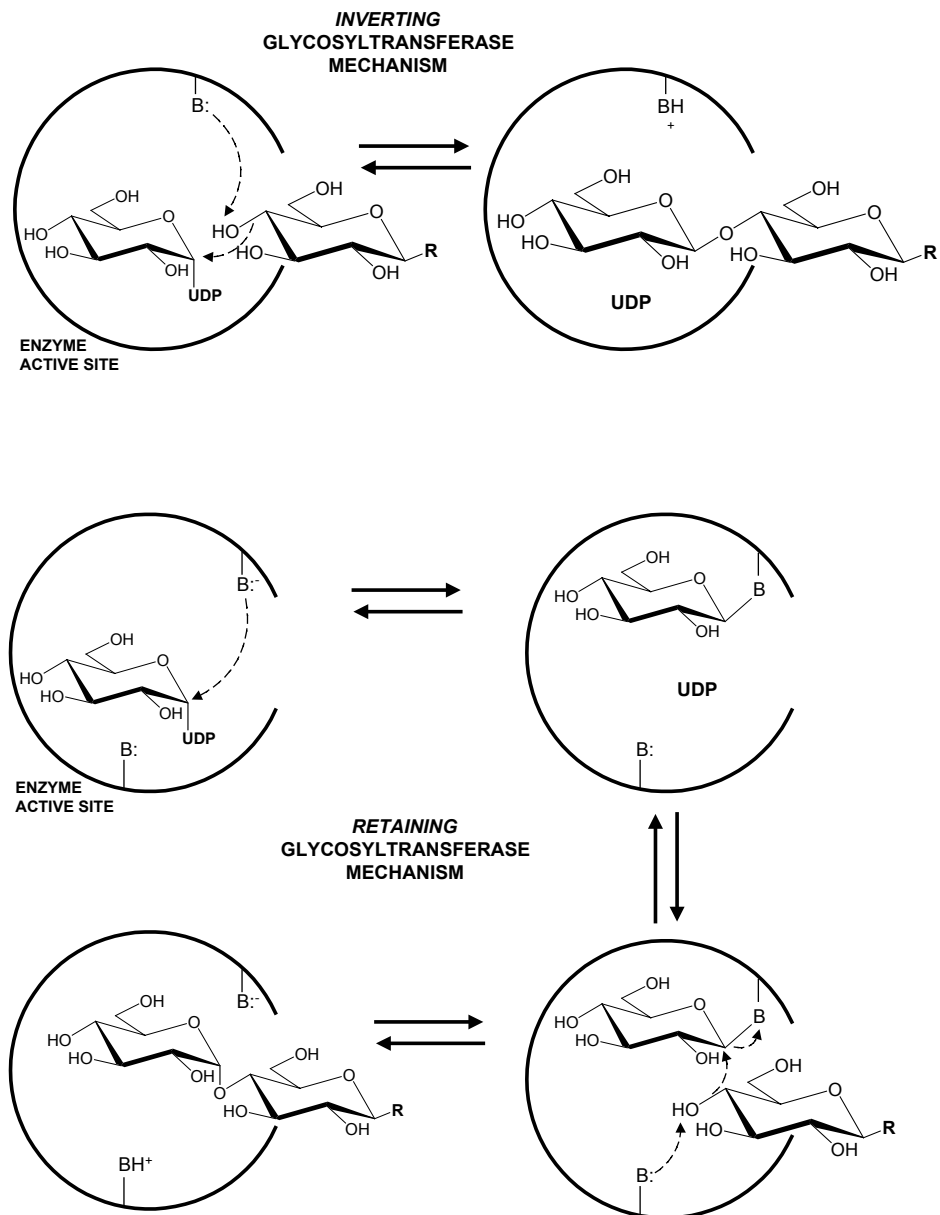
The cyanobacterium *Nostoc commune* in its dried (*upper left*) and rehydrated states (*upper right*). The organism experiences an almost instantaneous recovery of metabolic activities upon rewetting (Scherer, Chen and Böger 1986). *Lower left*: Scanning electron micrograph of desiccated field material, showing cells immobilized in an extracellular polysaccharide (EPS) matrix (Hill, Peat and Potts 1994). *Lower right*: Hydrated *N. commune* filament comprised of normal vegetative cells, capable of photosynthesis, and a heterocyst (larger, more rounded cell), capable of nitrogen fixation.



**Figure 1.3** Structure of *Nostoc commune* DRH1 EPS

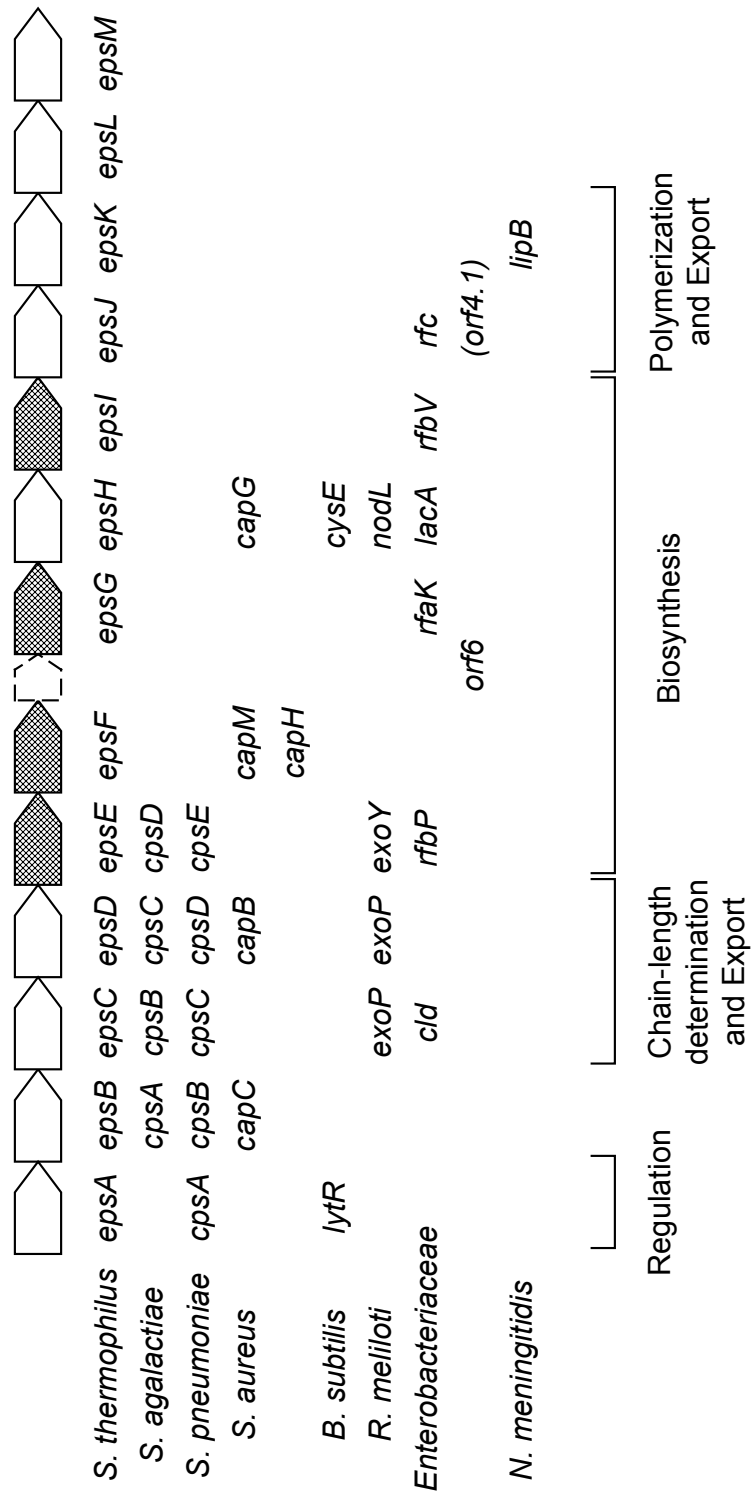
Average EPS structure of *N. commune*, grown under laboratory conditions (Helm *et al.* 2000). NosA denotes nosturonic acid, also known as 3-*O*-lactyl-D-glucuronic acid. Compositional analyses of EPS grown under different conditions (both field material and fermentor grown) show varying levels of nosturonic acid. Small amounts of mannose and glucuronic acid were also detected.





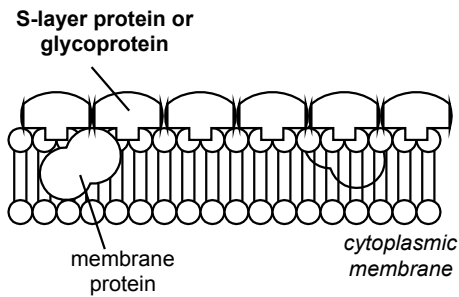
**Figure 1.4 Glycosyltransferase Reaction Mechanisms**

Examples of the *inverting* and *retaining* mechanisms of glycosyltransferases. *Retaining* glycosyltransferases proceed by way of a covalently linked enzyme-carbohydrate intermediate. Figure adapted from Persson *et al.* (2001), with permission from Nature ([www.nature.com](http://www.nature.com)).

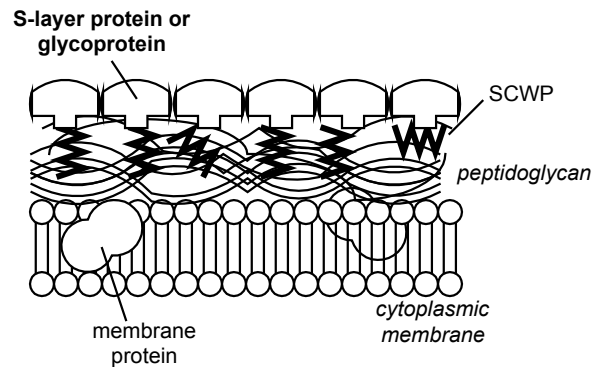


**Figure 1.5 Gene Clusters Encoding EPS-Synthesizing Proteins**

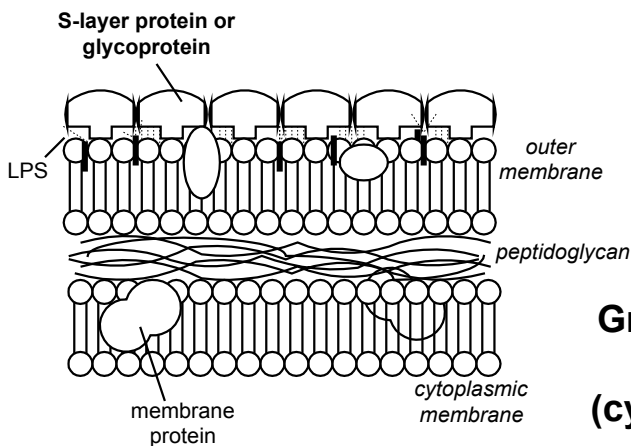
The *Streptococcus thermophilus* exopolysaccharide-producing gene cluster and homologs from several other microorganisms. Genes encoding glycosyltransferases are designated by cross-hatching. Figure adapted from Stungel, Neeser and Mollet (1996).



**Gram-negative  
Archaea**



**Gram-positive  
Archaea & Bacteria**



**Gram-negative  
Bacteria  
(cyanobacteria)**

SCWP *secondary cell wall polymer*  
LPS *lipopolysaccharide*

**Figure 1.6 Prokaryotic Cell Envelopes with S-Layers**

Representation of the three major classes of prokaryotic cell envelopes containing crystalline bacterial cell surface layers (S-layers). In Archaea the peptidoglycan layer is replaced by other cell wall polymers such as pseudomurein or methanochondroitin. In cyanobacteria, if an extracellular polysaccharide coating is present, the S-layer is located beneath of it. Figure adapted from Sleytr *et al.* (2001), with permission from Elsevier.

```

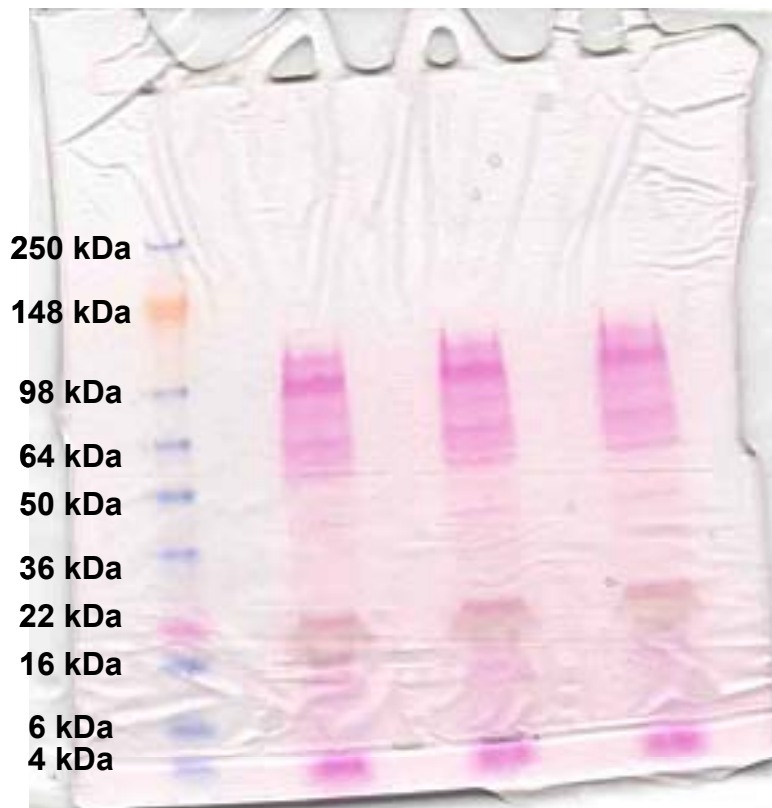
* * * : : . * . ** : :
exoglu_3 FNDDKSISDWAKNVVAN-AAKLGIIINGEPSNVFAPKGIATRAEAA 44
XynA_3 FSDDKSISDWAKNVVAN-AAKLGIVNGEPNNMFAPKDIATRAEAA 44
P100_protein FS-DVPAGHWAKEAVEALAAK-GIILGFDPGTFRGNENLTRYQAA 43
Omp_alpha FP-DVPKDHWAYEYVWKLWQR-GIFIGYPDGEFPGDRYITRYEAA 43
al14499 FS-DVQPTDWAFAQALQSLVERYGCIAGYPNGTYRGNRALTRYEFA 44
alr4550 FS-DVQPTDWAFAQALQSLVERYGCIAGYPNGTYRGNRALTRYEFA 44
exoglu_2 FS-DVKNGDWYANAIEAAYKA-GIIEG-DGKNMRPNDSITREEMT 42
XynA_2 FS-DVNSGDWYANAIEAAYKA-GIIEG-DGKNARPNDISITREEMT 42
exoglu_1 FN-DIKD-NWAKDVIEVLASR-HIVEGMTDTQYEPSKTVTRAEFT 42
XynA_1 FD-DIKN-SWAKDAIEVLASR-HIVEGMTDTQYEPNKTVTRAEFT 42
ruler 1.....10.....20.....30.....40.....

```

- \* positions which have a single, fully conserved residue
- : positions which have residues from a group of strongly similar amino acids
- .

### Figure 1.7 ClustalX Alignment of Several SLH Domains

ClustalX alignment of SLH domain sequences from *T. maritima* Omp $\alpha$  (Omp\_alpha), *T. thermophilus* 100-kDa S-layer protein (P100\_protein), *T. saccharolyticum* endoxylanase A (XynA\_1,2,3), *C. thermocellum* exo- $\beta$ -1,4-glucanase (exoglu\_1,2,3), and two *Anabaena* (*Nostoc*) PCC 7120 open reading frames of unknown function . SLH domains were identified using PFScan (Chapter 7.8.3).



**Figure 1.8 Periodic Acid-Schiff Staining Material in *Anabaena* PCC 7120 Soluble Protein Extract**

*Anabaena (Nostoc)* PCC 7120 soluble protein extract separated on a 4-15% gradient SDS-PAGE gel. The same sample was loaded in all three lanes. Each lane contains approximately 200  $\mu$ g of protein. Carbohydrate material was visualized using a periodic acid oxidation and Schiff's stain. Proteins with multiple glycosylation sites produce the same type of pattern when present as several different glycoforms.

## CHAPTER II

### Results

#### **Detection of Glycosyltransferase Activity in *Nostoc commune* Using Capillary Electrophoresis with Laser-Induced Fluorescence Detection (CE-LIF)**

##### 2.1 Introduction

Given the likelihood that the genes involved in *Nostoc commune*'s EPS production would be clustered in a single genetic locus (Stingele *et al.* 1996; Stingele *et al.* 1999), and with the recent success in applying this material to the stabilization of non-cyanobacterial cell lines (Bloom *et al.* 2001), it was decided to attempt purification of one or more of the enzymes responsible for EPS biosynthesis. Amino acid sequence information obtained from one or more active glycosyltransferases could then be used to design degenerate oligonucleotide primers for polymerase chain reactions to amplify portions of the glycosyltransferase-encoding genes. These fragments could be used to probe a genomic library, when it becomes available, to identify the (hypothetical) gene cluster that encodes the biological machinery responsible for EPS production, allowing its cloning and transfer to other organisms.

In order to purify an EPS-related glycosyltransferase, a means to detect its presence had to first be established. Use of a fluorescence-labeled, synthetic acceptor, capillary electrophoresis-based glycosyltransferase activity assay, developed by Wakarchuk (1996) has advantages over previous "classical" methods. Less sample material is required to assay for glycosyltransferase activity and activity can be detected at picomolar levels in a relatively short period of time. Because of the slow growth rate of *Nostoc commune*, it was reasoned that this method's high sensitivity would make efficient use of available biomass by enabling glycosyltransferase activity to be tracked through several successive purification steps. Additionally, the LIF detector system is extremely discriminating, and allows the selective observation of the reaction products of

a single chemical species (the fluorescent-labeled acceptor) with a *Nostoc commune* cell fraction.

The synthetic acceptor molecules of this assay consist of a carbohydrate and a fluorescein molecule, connected by a long alkyl chain. The fluorescein component facilitates the detection of this molecule, and its reaction products, in a complex reaction mixture. The alkyl linker chain is intended to project the carbohydrate component into a glycosyltransferase enzyme's binding pocket, despite the bulk of the fluorescent tag, to allow the addition of another saccharide to the acceptor. These acceptors are incubated with various donors and enzyme cofactors, and their reaction products are subsequently separated by capillary electrophoresis and/or thin-layer chromatography (Gilbert *et al.* 1996; Loughheed *et al.* 1999; Wakarchuk *et al.* 1998; Wakarchuk *et al.* 1996).

This chapter describes efforts to detect glycosyltransferase activity in *Nostoc commune* cell extracts, using this fluorescent-labeled acceptor system. The syntheses of two potential glycosyltransferase substrates for use with this system are also described.

### 2.2 Synthesis of Fluorescence-Labeled Monosaccharide Acceptors

Fluorescence-labeled monosaccharide acceptors ( $\alpha$ -D-galactose,  $\beta$ -D-glucose, and  $\beta$ -D-xylose derivatives) were synthesized by Richard F. Helm (co-advisor).

### 2.3 Synthesis of Fluorescence-Labeled Disaccharide Acceptors

In order to more effectively target a *Nostoc commune* glycosyltransferase that is responsible for production of extracellular polysaccharide (EPS), a disaccharide acceptor whose carbohydrate constituents and linkage correspond to the laboratory-grown *Nostoc commune* DRH1 EPS structure (Helm *et al.* 2000) was synthesized. The reasoning behind the use and possible necessity of a disaccharide acceptor can be explained as follows. If several promiscuous glycosyltransferase enzymes were detected using a particular monosaccharide acceptor, it would be necessary to judge which of these activities were more likely to be EPS-related, and thus should be targeted for study. Activity detected with an acceptor that more closely resembles the growing *Nostoc commune* EPS chain *in vivo* would be more likely to detect an EPS-producing enzyme. Alternatively, it has been noted that some glycosyltransferases have a stringent

requirement with regard to their carbohydrate acceptor, requiring that more than a single carbohydrate moiety or similar structural element interact with the enzyme before addition takes place (Campbell *et al.* 1997; Saxena *et al.* 1995). If this is the case with the *Nostoc commune* EPS-biosynthesizing glycosyltransferases, it will be necessary to have a disaccharide acceptor (or perhaps even a trisaccharide acceptor) to detect any activity.

A scheme for synthesizing a fluorescent-labeled disaccharide acceptor, based upon *Nostoc commune* DRH1's EPS structure (Figure 1.3), is shown in Figures 2.1, 2.2, 2.3, and 2.4. Detailed synthetic procedures are included at the end of Chapter 7 (Section 7.9 and Section 7.10). An anomeric mixture of compound **1** was prepared from D-xylose, using pyridine and acetic anhydride, and then brominated to give the  $\alpha$ -anomer of xylosyl bromide (**2**). The acetylated 4-nitrophenol derivative (**3**) was produced from the bromide (**2**) and an excess of 4-nitrophenol, and was then subjected to Zemplén deprotection to give 4-nitrophenyl  $\beta$ -D-xylopyranoside (**4**). Regioselective chloroacetylation of compound **4**, followed by benzylation, produced a fully protected monosaccharide (**6**). The weaker protecting group (chloroacetyl, ClAc) on the 4-hydroxyl of **6** was removed using thiourea to produce **7**, 4-nitrophenyl 2,3-di-*O*-benzoyl- $\beta$ -D-xylopyranoside, the monosaccharide acceptor for the coupling procedure (Figure 2.1).

The ethyl 1-thio-2,3,4,6-tetra-*O*-acetyl- $\beta$ -D-glucopyranoside donor (**9**) for the coupling procedure was synthesized from D-glucose (Figure 2.2). An anomeric mixture of 1,2,3,4,6-penta-*O*-acetyl-D-glucopyranoside (**8 $\alpha$** , **8 $\beta$** ) was prepared from D-glucose using acetic anhydride and sodium acetate catalyst. Only the  $\beta$ -anomer reacted with ethanethiol, in the presence of zirconium (IV) tetrachloride, to give compound **9** (Contour *et al.* 1989).

The coupling of compounds **7** and **9** was performed using a modified procedure of Takeo (1995), which employs silver triflate and *N*-iodosuccinimide (NIS). Several attempts to join an acetylated glucosyl bromide to the acceptor (**7**), using silver triflate/collidine and mercuric cyanide systems (Figure 2.3), were unsuccessful. The resulting, fully protected compound was subjected to the Zemplén deprotection procedure to afford compound **11**, 4-nitrophenyl *O*-( $\beta$ -D-glucopyranosyl)-(1 $\rightarrow$ 4)- $\beta$ -D-xylopyranoside.



Reduction of **11** by catalytic transfer hydrogenation (catalytic hydrogenation using a hydrogen donor instead of hydrogen gas) (Roy and Tropper 1991) produced compound **12**, the aminophenyl derivative of the disaccharide (Figure 2.3). Previous attempts to perform the reduction before deprotection on other disaccharide derivatives, in an effort to resolve solubility difficulties encountered in the reduction procedure, resulted in a compound that was difficult to either crystallize or purify by flash chromatography.

Compound **13**, a fluorescent-labeled disaccharide and a potential *Nostoc commune* glycosyltransferase substrate, was prepared by incubation of **12** with the FCHASE (6-(fluorescein-5-carboxamido) hexanoic acid succinimidyl ester) reagent from Molecular Probes (Figure 2.4). The post-reaction mixture was separated by preparative-scale thin layer chromatography (70:20:10:1 ethyl acetate-methanol-water-acetic acid). This separation revealed a single spot that both was visible under short wave ultraviolet light and produced a positive response to a TLC carbohydrate stain (naphthoresorcinol). Purification of this compound resulted in an almost unweighable quantity of material that gave an extremely weak  $^1\text{H}$  NMR signal and no detectable  $^{13}\text{C}$  NMR signal. Resonances typical of carbohydrate anomeric protons were visible in the  $^1\text{H}$  NMR spectra (doublets at 4.25 and 4.5 ppm, both with coupling constants of  $\sim 7.5$  Hz).

Mass spectrometric analysis of **13**, by direct injection into a ThermoFinnigan TSQ Quantum triple quadrupole mass spectrometer, revealed positive-ion peaks at 581.3 Da ( $[\text{M}+\text{H}]^+$ ), 714.3 Da ( $[\text{M}+\text{H}]^+$ ), 876.3 Da ( $[\text{M}+\text{H}]^+$ ), 898.3 Da ( $[\text{M}+\text{Na}]^+$ ), and 914.3 Da ( $[\text{M}+\text{K}]^+$ ), corresponding to the tag adducts and fragments shown in Figure 2.5. Tandem mass spectrometric analysis of the 876.3 Da peak revealed fragments at 581.7 Da ( $[\text{M}+\text{H}]^+$ ) and 714.1 Da ( $[\text{M}+\text{H}]^+$ ), which correspond to the phenol-terminated and monosaccharide structures, respectively. Analysis of this compound in negative-ion mode revealed a peak at 873.8 Da ( $[\text{M}-\text{H}]^-$ ), which likely the result of this compound containing an ionized carboxyl.

Taking into consideration that a glycosyltransferase might demonstrate a slightly different activity *in vitro* than *in vivo*, a potential *Nostoc commune* glycosyltransferase acceptor was also prepared from a commercially available disaccharide (cellobiose). Although preparation of a non-EPS-related structure seems counterintuitive to the

philosophy behind the first disaccharide acceptor synthesis, such an acceptor is much more quickly and easily prepared. At the time of this second potential acceptor's synthesis, only one of the other fluorescent-labeled acceptors had been used successfully, and the activity detected with that acceptor did not directly correspond to the EPS structure of *Nostoc commune* DRH1.

The fully protected  $\alpha$ -D-cellobiosyl bromide (**14**) was prepared from D-cellobiose using the procedure of Kartha and Jennings (1990). The strategy for synthesis of this molecule is shown in Figure 2.6 and Figure 2.7. In contrast to the xylosyl bromide (**2**), this compound would not form a 4-nitrophenol derivative using potassium carbonate in acetone, but instead had to be joined to the phenol moiety using a two-phase reaction system and phase-transfer catalysis (Roy *et al.* 1992). Subsequent reduction and deprotection of **15** afforded the unprotected aminophenyl derivative (**17**), which was also tagged using the FCHASE reagent to produce a fluorescent-labeled disaccharide (**18**). Mass spectrometric analysis of **18**, by direct injection into a ThermoFinnigan TSQ Quantum triple quadrupole mass spectrometer, revealed a positive-ion peak at 905.3 Da ( $[M+H]^+$ ), which produced additional peaks at 581.0 Da ( $[M+H]^+$ ) and 742.8 Da ( $[M+H]^+$ ) upon further fragmentation (Figure 2.5).

#### 2.4 Application of Fluorescence-Labeled Acceptors to *Nostoc commune*

Although it was anticipated that any EPS-related glycosyltransferases in *Nostoc commune* would be membrane-associated, both water-soluble and insoluble cell fractions were tested for activity. Following lysis of *Nostoc commune* cells, using a French pressure cell, residual whole cells were removed by a low speed centrifugation (5000 x g, 20 min, 22°C) and intact cell envelope material was removed by a medium speed centrifugation (30,000 x g, 20 min, 22°C). A high-speed centrifugation (150,000 x g, 60 min, 22°C) was then used to segregate water-soluble and insoluble cell material, referred to as the soluble fraction and membrane fraction, respectively.

Soluble protein was released from the cell into a lysis buffer of 50 mM Tris, 10 mM MgCl<sub>2</sub>, and 20 mM KCl, at a pH of 7.8 (extraction buffer). Samples were then concentrated using a 15 mL Centriprep YM-10 centrifugal concentrator, at the maximum speed recommended by the manufacturer, until they had reached the point of saturation

(4°C). The membrane fraction was resuspended by vortexing, and the resulting suspension was homogenized with a spatula, in one or more of the following buffers: 50 mM HEPES (pH 6.7), 50 mM HEPES (pH 7.0), 50 mM MES (pH 7.5), 50 mM Tris (pH 7.8), 50 mM Tris (pH 8.0), or 100 mM Tris (pH 8.0). Glycosyltransferase activity assays of both soluble and membrane fractions were performed using freshly prepared cell extracts, which were subsequently stored at 4°C. Assays were performed in the presence of 10 mM DTT, typically with the addition of MgCl<sub>2</sub> to a final concentration of 50 mM, if Mg<sup>+2</sup> was not already present in the resuspension buffer.

Glycosyltransferase activity assays were performed as described in Chapter 7.3.3 using the above-mentioned fluorescent-labeled monosaccharides and disaccharides as potential acceptors. The following commercially available nucleotide donors and glycosyl phosphates were used to test for activity: UDP-galactose (UDP-Gal), UDP-glucose (UDP-Glu), UDP-xylose (UDP-Xyl), UDP-glucuronic acid (UDP-GluA), ADP-ribose (ADP-Rib), ADP-glucose (ADP-Glu), GDP-glucose (GDP-Glu), GDP-mannose (GDP-Man), CMP-acetylneuraminic acid (CMP-Neu5Ac),  $\alpha$ -D-galactose-1-phosphate ( $\alpha$ -D-Gal-1-p),  $\alpha$ -D-glucose-1-phosphate ( $\alpha$ -D-Glu-1-p),  $\beta$ -D-glucose-1-phosphate ( $\beta$ -D-Glu-1-p),  $\alpha$ -D-xylose-1-phosphate ( $\alpha$ -D-Xyl-1-p), and  $\alpha$ -D-glucuronic acid-1-phosphate ( $\alpha$ -D-GluA-1-p). All nucleotide donors tested were in the  $\alpha$ -anomeric conformation, as they are found to occur in nature (Breton and Imberty 1999; Kapitonov and Yu 1999).

Analysis of activity assay reaction products by capillary electrophoresis was performed in three different buffer systems. This was done to increase the probability that an adequate separation of precursors from product molecules would be achieved, so that any potential glycosyltransferase reaction products would not go undetected. Sodium dodecyl sulfate (SDS) is a common detergent additive used for the separation of neutral molecules in capillary electrophoresis, and cyclodextrins are frequently employed for the separation of chiral molecules (Deyl *et al.* 1998). Borate buffers that contained one of these additives, as well as an additive-free buffer, were utilized.

A summary of attempts to detect glycosyltransferase activity is presented in Table 2.1. All nine nucleotide donors were tested with all five fluorescent-labeled saccharides, using a 50 mM HEPES-suspended membrane fraction, and the subsequent analyses were performed in three distinct capillary electrophoresis buffer systems. Phosphate donors

were tested with the three fluorescent-labeled monosaccharides, using an insoluble fraction resuspended in 50 mM HEPES (pH 6.7) or 50 mM Tris (pH 7.8). The products from these reactions were analyzed in a borate buffer system with no additives. In some phosphate donor experiments, nucleotide monophosphates were also included, in an attempt to mimic an activity rescue experiment. It should be noted that such experiments were unsuccessful, even when tried with the below-mentioned glycosyltransferase activity (Section 2.5).

The soluble fraction and a concentrate of the cell growth media (BG11<sub>o</sub>) were tested using the fluorescent-labeled monosaccharides and uridine-diphosphate (UDP) donors only. Reactions were analyzed in a borate buffer system without additives. Attempts to detect glycosyltransferase activity in several different detergent extracts, including one of whole *Nostoc commune* DRH1 cells, will be discussed in the following chapter.

During this series of experiments, only one potential glycosyltransferase activity was detected. Incubation of fluorescent-labeled galactose with UDP-glucuronic acid and a membrane fraction suspension produced small quantities of a compound that could be separated from the substrate using a borate buffer system, with or without SDS. This compound was not present in a control reaction in which water had been substituted for the donor. This compound appeared regardless of which of the previously mentioned buffer systems had been used to resuspend the membrane fraction (Figure 2.8). Personal communications with Dr. W. Wakarchuk indicated that the reaction product-to-precursor ratio (estimated as ~1/100 by peak heights) observed for this activity is typical of other successful applications of this assay.

### 2.5 Detection of an Apparent Glucuronosyltransferase Activity

Incubation of the fluorescent-labeled galactose acceptor with UDP-glucuronic acid (UDP-GluA) and a freshly prepared *Nostoc commune* membrane homogenate produced a chemical species, which separated in a borate buffer system (without additives), and which was not present in a control reaction containing water in place of the donor (Figure 2.8). Production of this species appeared metal ion independent when this fraction was assayed.

No peaks were detectable when (control) reaction mixtures, without the fluorescent-labeled acceptor, were analyzed. This indicated that all peaks in the electropherogram were the result of interaction or reaction of the potential acceptor with the biological extract or the separation buffer. Control reactions confirmed that the production of the activity reaction product required both the activated glucuronic acid donor and the membrane material. Although designation of this UDP-GluA-dependent phenomenon as glucuronosyltransferase activity assumes that the uronic acid moiety is being transferred to the acceptor, the UDP-GluA-dependent reaction product's later elution reflects a greater mobility, consistent with the idea that an additional charge has been attached to the substrate. The faster migration of this peak, even in the absence of detergent additive, also supports the idea that this molecule has gained additional charge as a result of inclusion of the nucleotide donor in the reaction mixture.

Detection of a potential glycosyltransferase in a membrane fraction (where an EPS-related glycosyltransferase would be expected to be located) is a significant find, regardless of the fact that it does not immediately show a direct relationship to the EPS structure of *Nostoc commune*. The rationale behind investing research effort into purification of the enzyme (or enzymes) responsible for this activity is discussed in Chapter 3 and Chapter 4.

#### 2.6 Confirmation of CE-LIF Analysis Results by Preparative-Scale Thin Layer Chromatography and Matrix Assisted Laser Desorption/Ionization – Time of Flight Mass Spectrometry

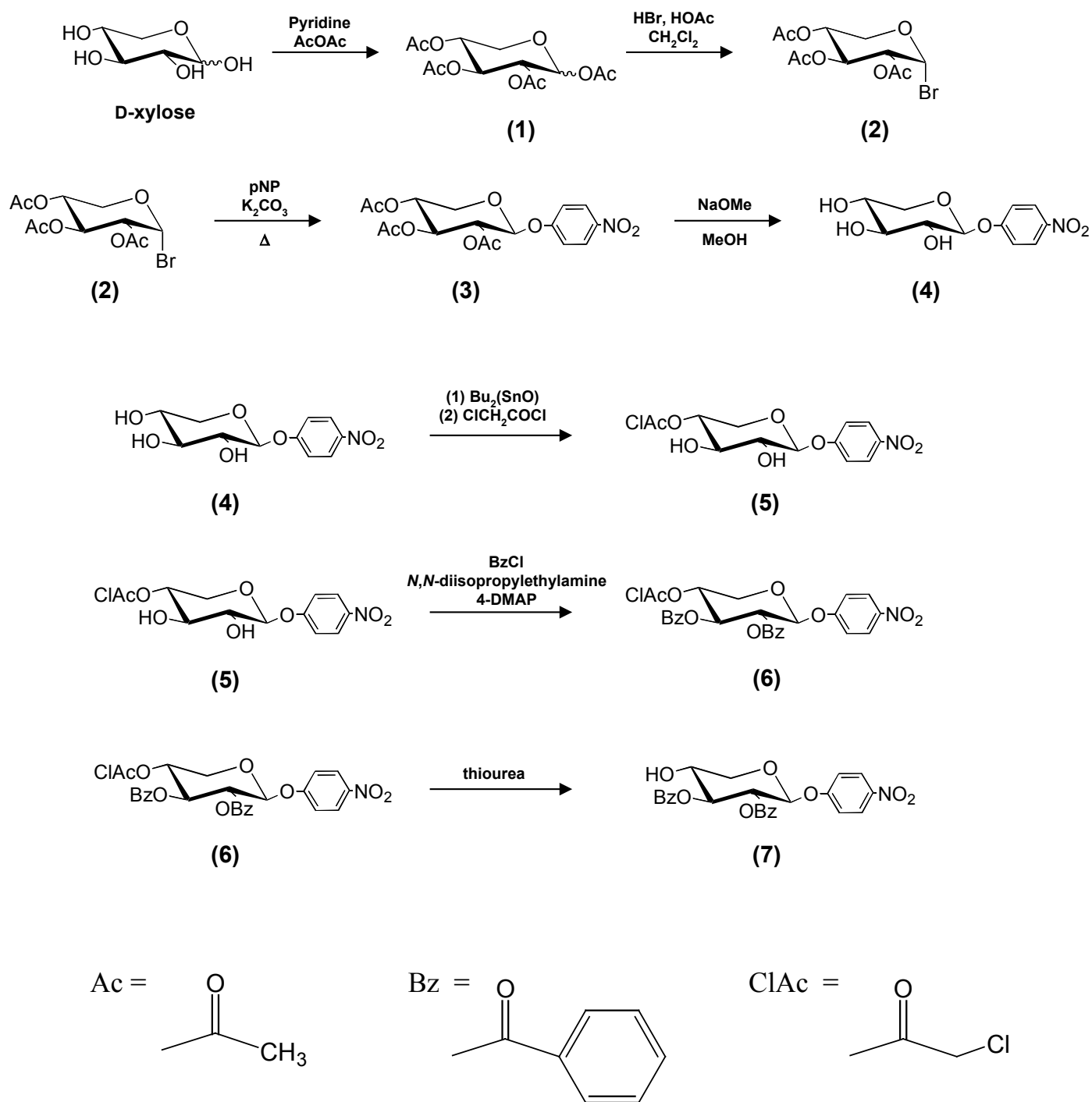
A glycosyltransferase activity assay reaction mixture was fractionated by preparative-scale thin-layer chromatography (TLC) and analyzed by matrix assisted laser desorption/ionization – time of flight (MALDI-TOF) mass spectrometry to determine whether products that had not been detected by capillary electrophoresis were present. Additionally, structural confirmation of the reaction product discussed in Section 2.5 was desired. A glycosyltransferase activity assay reaction mixture, containing the fluorescent-labeled galactose acceptor and UDP-glucuronic acid donor, was assembled on a 900  $\mu$ L scale (45 times larger than 'normal' assay reaction mixtures) and incubated overnight. Analysis of this reaction mixture by capillary electrophoresis revealed the

presence of the UDP-glucuronic acid-dependent species. This product was purified using a Bond Elut C18 solid-phase extraction cartridge, a silica gel TLC plate (70:20:10:1 ethyl acetate-methanol-water-acetic acid), and a C18 TLC plate (1:2 acetonitrile-water). The mass spectrum of a sample that contained the UDP-GluA-dependent species (detectable by CE-LIF), ionized with either 2,5-dihydroxybenzoic acid or  $\alpha$ -cyano-4-hydroxycinnamic acid, did not reveal any peaks that were not also present in a matrix control.

Prior to the experiment described above, it had been assumed that there would be no difficulty in detecting the fluorescent-labeled carbohydrate compounds or their reaction products by MALDI-TOF mass spectrometry. To determine the validity of this assumption, a saturated solution of the fluorescent-labeled galactose acceptor (in water) and a saturated solution of 2,5-dihydroxybenzoic acid matrix (in 9:1 acetonitrile-water) were combined in varying volumetric ratios and analyzed by MALDI-TOF mass spectrometry. A 1:2 volumetric mixture of saturated matrix and sample solutions resulted in the most intense (sample) ion signals in these experiments. Three positive-ion peaks (740.9 Da, 762.6 Da, and 784.4 Da at 30%, 40%, and 15% of the most intense (matrix) ion in the spectrum, respectively) that could not be reconciled to the formula weights of various ionic adducts of the galactose acceptor ( $[M+H]^+=743.7$  Da,  $[M+Na]^+=765.7$  Da,  $[M+K]^+=781.8$  Da) were detected in these samples, but not in the control (matrix only). No peaks were detected in negative-ion mode. Post-source decay analysis of the 740.9 Da peak revealed a degradation product at 581.3 Da, which could correspond to a fragment of the fluorescent-labeled acceptor ( $[M+H]^+=581.6$  Da). Considering that the height of the UDP-glucuronic acid-dependent species peak is approximately  $1/100^{\text{th}}$  of the height of its fluorescent-labeled precursor's peak in the electropherogram, it is not likely that this reaction product would ever be detectable using MALDI-TOF mass spectrometry, and so additional attempts to use this technique to find a glycosyltransferase reaction product were not made.

## 2.7 Summary

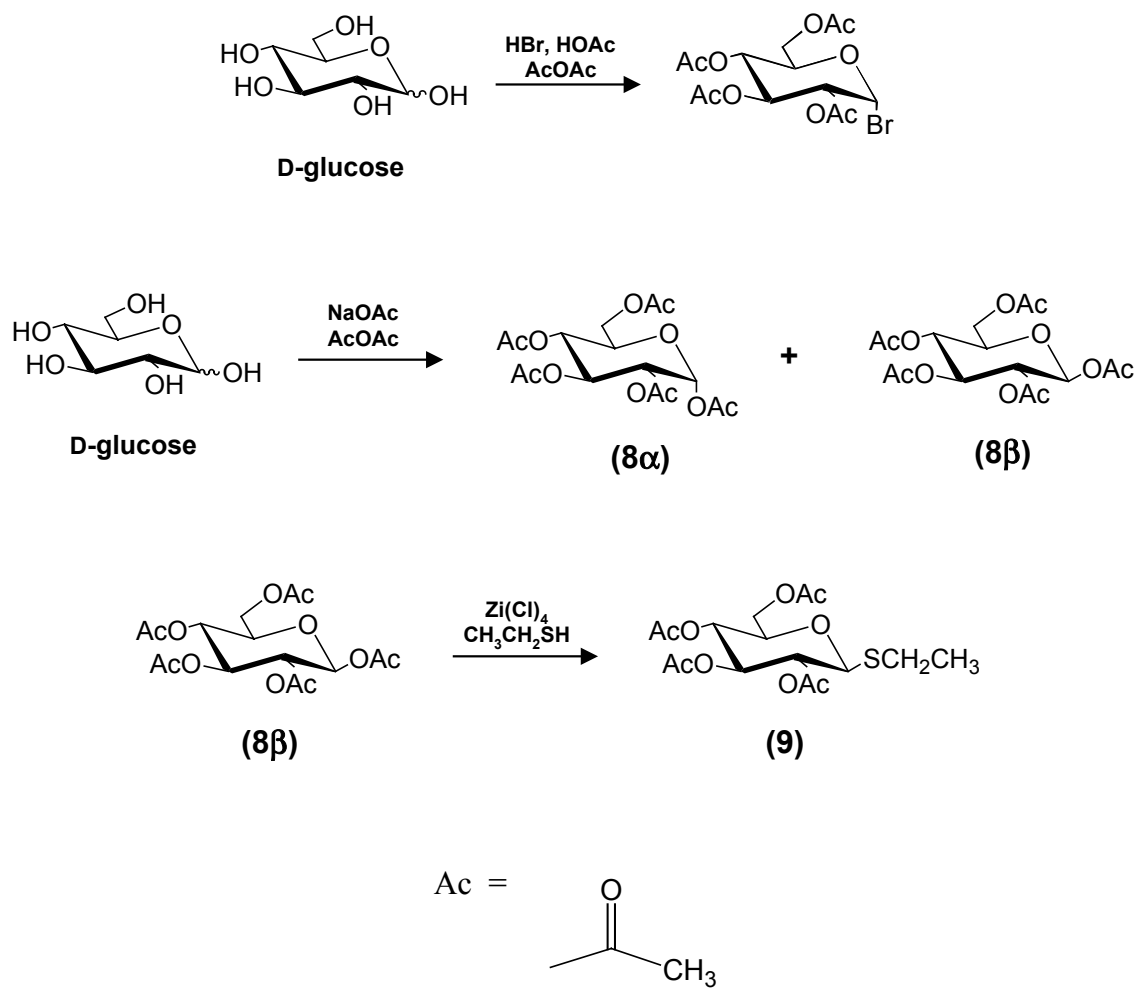
In this portion of the work, Specific Aims #1 and #2, presented in Chapter 1.2.6, were addressed. Five fluorescent-labeled carbohydrate compounds that could potentially serve as acceptors for *Nostoc commune* DRH1 glycosyltransferases were prepared. These compounds were used to assay cell fractions for the activities of the EPS-producing enzymes of *Nostoc commune* DRH1. Trials of various assay and separation conditions resulted in the detection of a putative glucuronosyltransferase activity. The UDP-GluA-dependent reaction product produced in this activity assay could not be detected by mass spectrometry, but its electrophoretic migration behavior supports the idea that transfer of glucuronic acid to the synthetic galactose acceptor has taken place.



**Figure 2.1 Synthesis of Selectively Protected Xyloside Acceptor for Disaccharide Coupling Procedure**

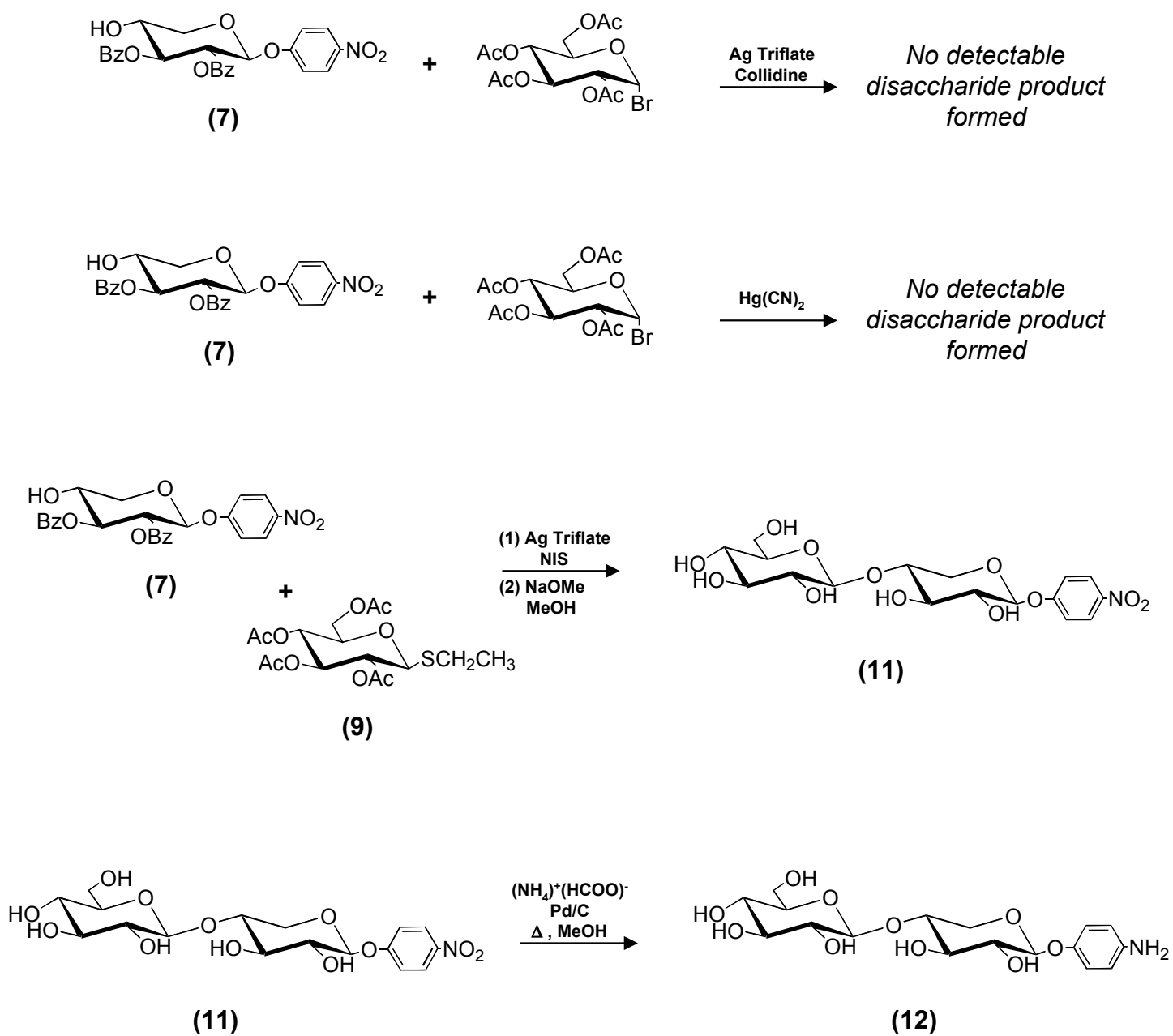
Synthesis of the 4-nitrophenyl xyloside acceptor. The xylose 4-hydroxyl group is selectively deprotected for participation in a coupling reaction with a glucose derivative. Attachment of the 4-nitrophenol moiety early in the synthetic scheme facilitates detection of the compound with ultraviolet light when reaction mixtures are separated by thin-layer chromatography.





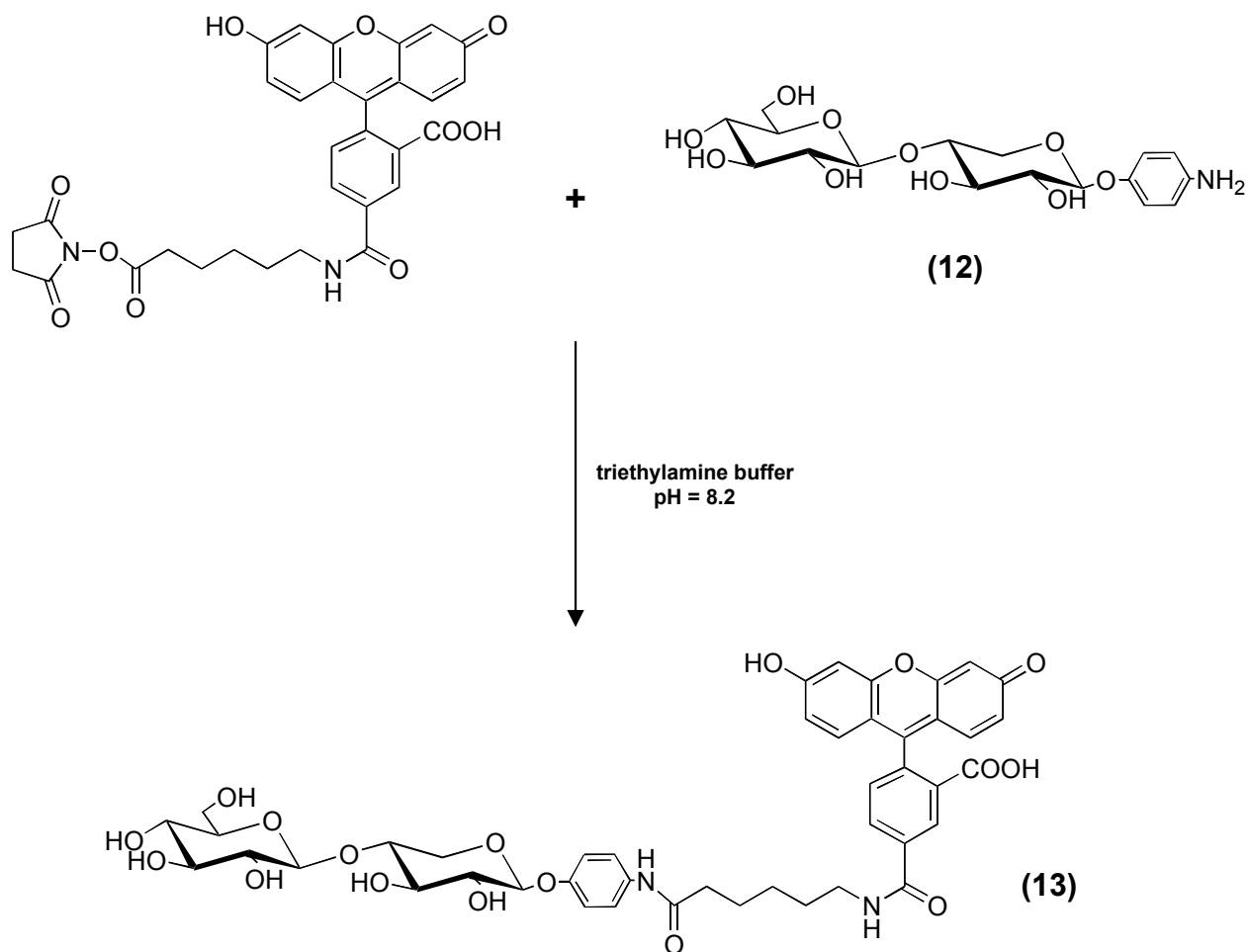
**Figure 2.2 Synthesis of Disaccharide Coupling Procedure Donors**

Synthesis of glucose derivatives for creation of a fluorescent-labeled disaccharide acceptor based on the EPS structure of *Nostoc commune*. Acetobromoglucose (*above*) was prepared using the one-pot method of Kartha and Jennings (1990); this compound was not successfully coupled to the acceptor (7).



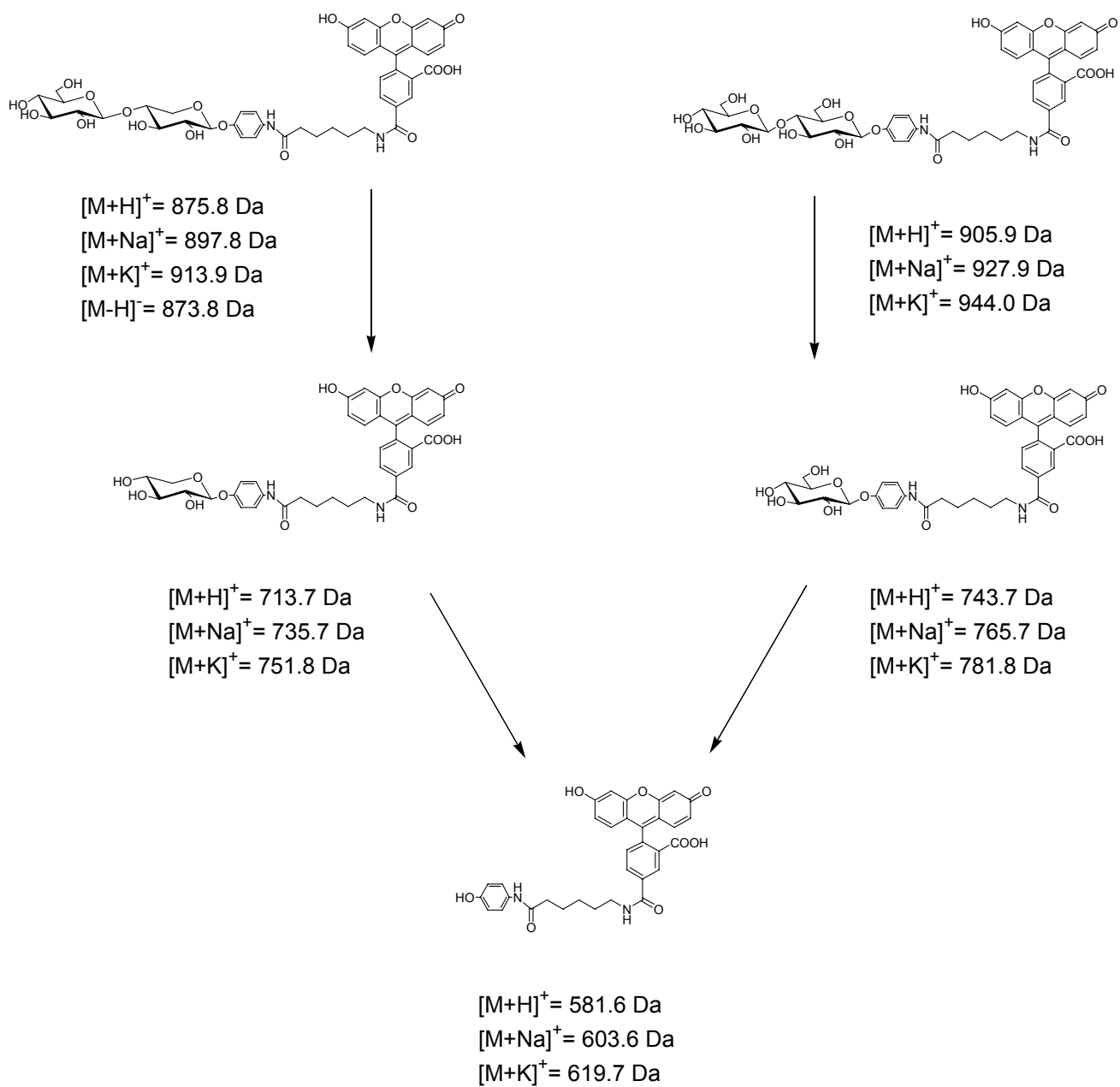
**Figure 2.3 Disaccharide Coupling Attempts and Subsequent Reduction**

Attempts to form the nitrophenyl derivative of *O*-( $\beta$ -D-glucopyranosyl)-(1 $\rightarrow$ 4)- $\beta$ -D-xylopyranoside and subsequent reduction to the aminophenyl derivative. Formation of the disaccharide was successful when performed with silver triflate/NIS, after several unsuccessful attempts using two other coupling systems.



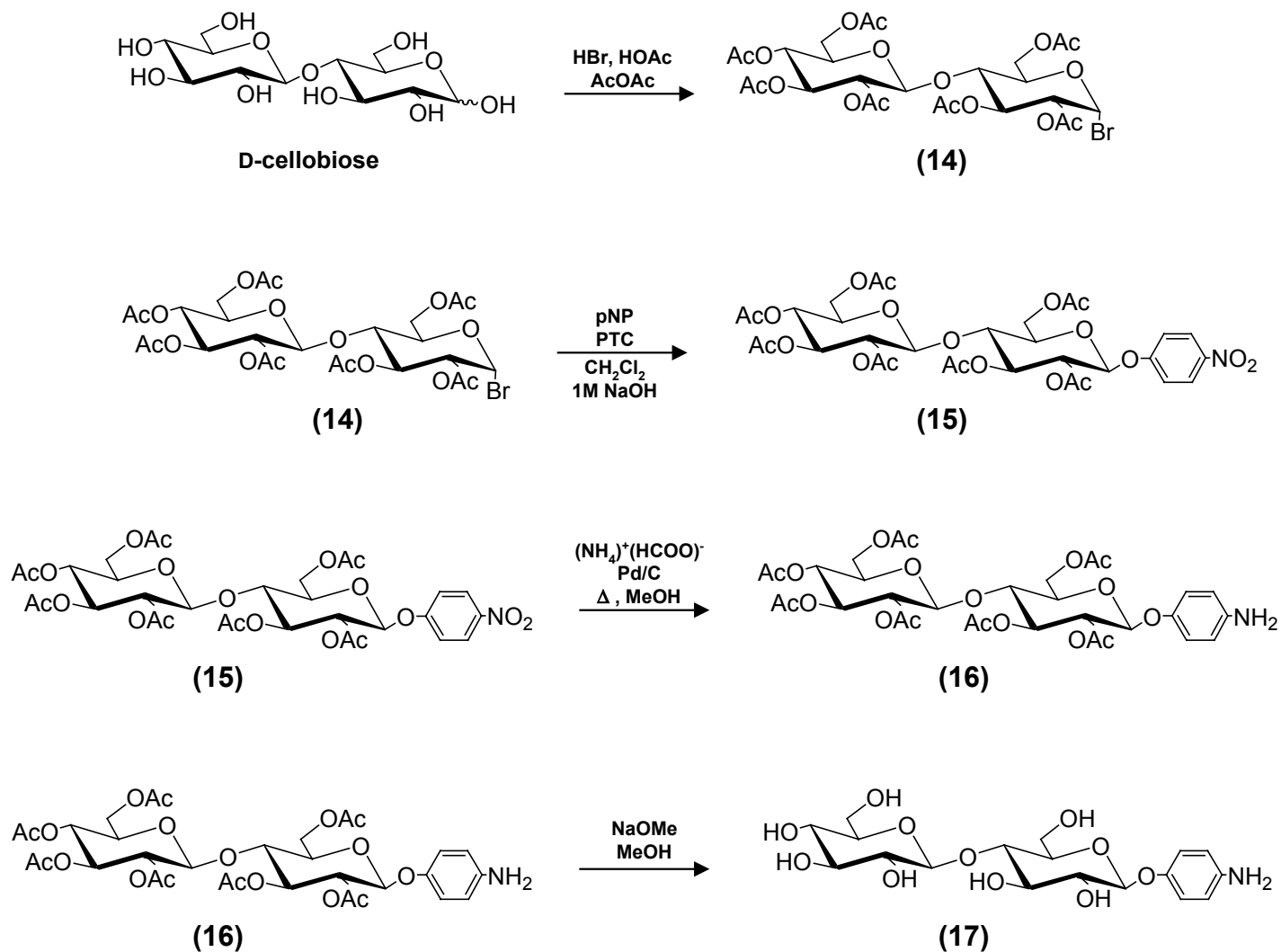
**Figure 2.4 Fluorescent-Tagging of First Disaccharide Acceptor**

Preparation of the first synthetic, fluorescent-labeled disaccharide acceptor. The FCHASE (6-(Fluorescein-5-Carboxamido) Hexanoic Acid Succinimidyl Ester) tagging reagent is available from Molecular Probes (Eugene, OR).



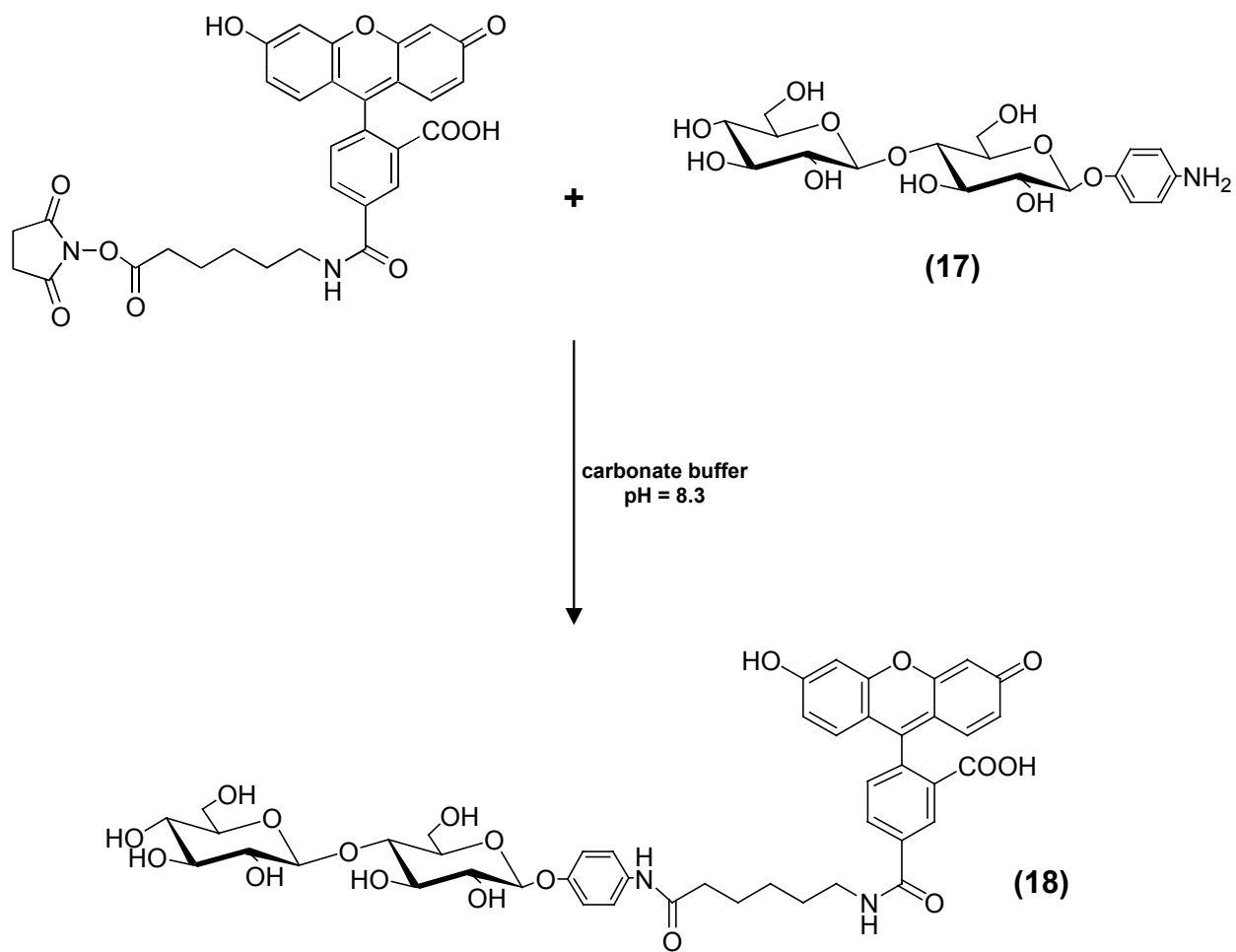
**Figure 2.5 Fragments and Adducts of Synthesized Fluorescent-Labeled Disaccharide Acceptors**

Anticipated tag fragments, resulting from mass spectrometric analyses of synthesized disaccharide acceptors. Formula weights of detectable ionic species are listed. Analyses were performed by direct injection into a ThermoFinnigan TSQ Quantum triple quadrupole mass spectrometer.



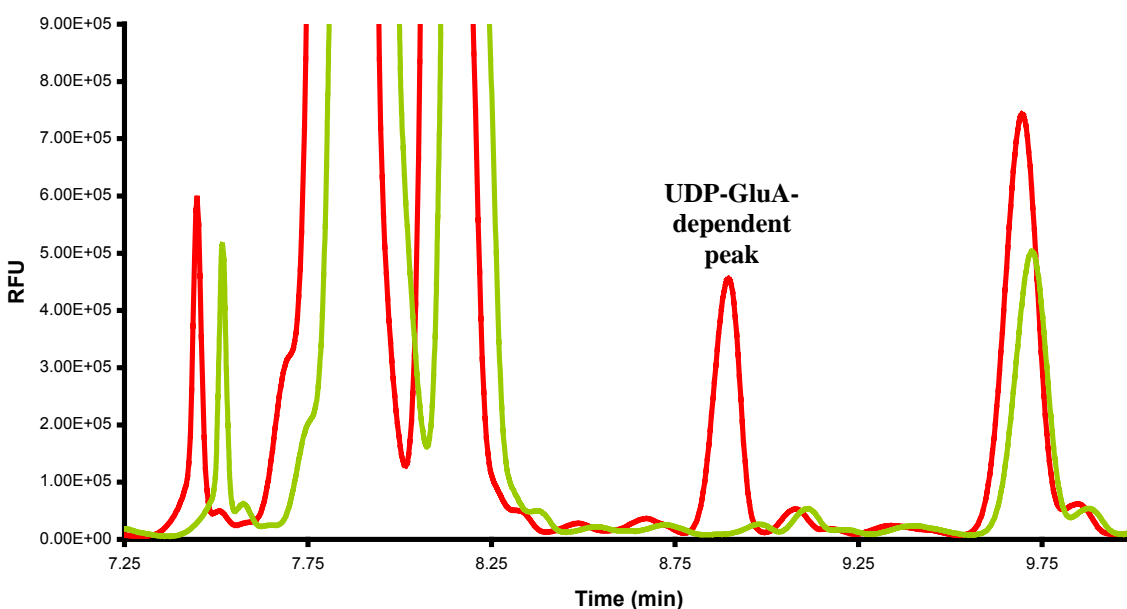
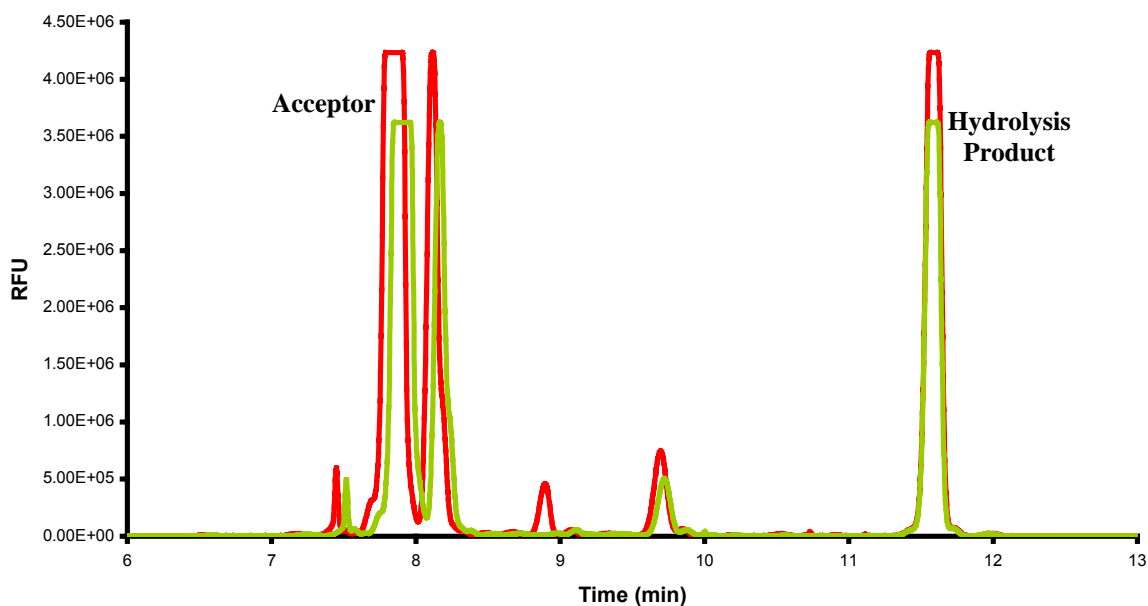
**Figure 2.6 Preparation of 4-Aminophenyl Cellobiose**

Preparation of the 4-aminophenyl derivative of cellobiose for labeling with the FCHASE reagent. Synthesis of a disaccharide acceptor requires fewer steps when starting from a commercially available disaccharide, but also requires a unique reaction system for 4-nitrophenyl derivatization.



**Figure 2.7 Fluorescent-Tagging of Second Disaccharide Acceptor**

Preparation of the second synthetic, fluorescent-labeled disaccharide acceptor. No significant difference in the use carbonate and triethylamine buffers for this procedure was observed.



**Figure 2.8 Detection of Glycosyltransferase Activity Using CE-LIF**

Glycosyltransferase activity is assayed using capillary electrophoresis. In the control reaction (no nucleotide donor, green), a glycosyltransferase reaction product is not detected; peaks for the synthetic acceptor and species resulting from its interactions with membrane fragments and buffer are present. When the donor is included, an additional peak (red) appears in the electropherogram (activity is detected). *Above:* Separation of the glucuronosyltransferase reaction product, produced from the fluorescent-labeled galactose acceptor and UDP-GluA. Peaks for the tagged acceptor and its alkaline hydrolysis products are the most prominent. *Below:* Close up of the UDP-glucuronic acid-dependent peak, the size of which is substantially smaller than that of the peak for the unreacted galactose acceptor.

<i>Acceptors (right) and Donors (below)</i>	Galactose	Glucose	Xylose	Glucose- $\beta(1\rightarrow4)$ - Xylose	Glucose- $\beta(1\rightarrow4)$ - Glucose
<i>UDP-Gal</i>	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$B^\Phi$	$B^\Phi$
<i>UDP-Glu</i>	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$B^\Phi$	$B^\Phi$
<i>UDP-Xyl</i>	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$B^\Phi$	$B^\Phi$
<i>UDP-GluA</i>	$A^X, B^\Phi, C^X, D^X, E^X, F^X, G^X, H^X$	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$A^X, B^\Phi, C^X, E^X, G^X, H^X$	$B^\Phi$	$B^\Phi$
<i>ADP-Rib</i>	$A^X, B^\Phi, D^X, F^X$	$A^X, B^\Phi, D^X, F^X$	$A^X, B^\Phi, D^X, F^X$	$B^\Phi$	$B^\Phi$
<i>ADP-Glu</i>	$A^X, B^\Phi, D^X, F^X$	$A^X, B^\Phi, D^X, F^X$	$A^X, B^\Phi, D^X, F^X$	$B^\Phi$	$B^\Phi$
<i>GDP-Glu</i>	$A^X, B^\Phi, D^X, F^X$	$A^X, B^\Phi, D^X, F^X$	$A^X, B^\Phi, D^X, F^X$	$B^\Phi$	$B^\Phi$
<i>GDP-Man</i>	$B^\Phi, D^X$	$B^\Phi, D^X$	$B^\Phi, D^X$	$B^\Phi$	$B^\Phi$
<i>CMP-Neu5Ac</i>	$B^\Phi, D^X$	$B^\Phi, D^X$	$B^\Phi, D^X$	$B^\Phi$	$B^\Phi$
<i><math>\alpha</math>-D-Gal-1-p</i>	$A^X, D^X$	$A^X, D^X$	$A^X, D^X$		
<i><math>\alpha</math>-D-Glu-1-p</i>	$A^X, D^X$	$A^X, D^X$	$A^X, D^X$		
<i><math>\beta</math>-D-Glu-1-p</i>	$A^X, D^X$	$A^X, D^X$	$A^X, D^X$		
<i><math>\alpha</math>-D-Xyl-1-p</i>	$A^X, D^X$	$A^X, D^X$	$A^X, D^X$		
<i><math>\alpha</math>-D-GluA-1-p</i>	$A^X, D^X$	$A^X, D^X$	$A^X, D^X$		

**Buffer System/Protein Fraction:**

- A 50 mM HEPES (pH 6.7) + 20 mM MgCl<sub>2</sub>, *membrane fraction*
- B 50 mM HEPES (pH 7.0), *membrane fraction*
- C 50 mM MES (pH 7.5) + 10 mM MnCl<sub>2</sub>, *membrane fraction*
- D 50 mM TRIS (pH 7.8) + 10 mM MgCl<sub>2</sub> + 20 mM KCl, *membrane fraction*
- E 50 mM TRIS (pH 8.0) + 10 mM MgCl<sub>2</sub>, *membrane fraction*
- F 100 mM TRIS (pH 8.0) + 10 mM MgCl<sub>2</sub>, *membrane fraction*
- G 50 mM TRIS (pH 7.8) + 10 mM MgCl<sub>2</sub> + 20 mM KCl, *soluble fraction*
- H concentrated BG11<sub>o</sub> cell media

**CE Buffer System (superscripts):**

- X 25 mM sodium borate buffer (pH 9.4)
- Y 25 mM sodium borate buffer (pH 9.4) + 25 mM SDS
- Z 25 mM sodium borate buffer (pH 9.4) + 20 mM  $\beta$ -cyclodextrin
- $\Phi$  analysis conducted using all three of the above systems (X, Y, and Z)

**Table 2.1 Summary of Attempts to Detect Glycosyltransferase Activity in *Nostoc commune* DRH1**

Summary of reaction conditions used to detect glycosyltransferase activity in *Nostoc commune* DRH1, using a synthetic, fluorescent-tagged acceptor assay. Phosphate donors were not tested with synthetic disaccharide acceptors. Only the combination of the galactose monosaccharide acceptor and UDP-GluA donor resulted in detectable activity (bold).



## CHAPTER III

### Results

#### **Partial Purification of a *Nostoc commune* Protein with Apparent Glucuronosyltransferase Activity and Attempts to Determine Its Amino Acid Sequence**

##### 3.1 Introduction

In the previous chapter, attempts to detect EPS-related glycosyltransferase activities were described. An extensive search, using a capillary electrophoresis-based assay, uncovered only a single potential glycosyltransferase activity. This activity did not result in the formation of a carbohydrate linkage that is present in the laboratory-grown EPS of *Nostoc commune* DRH1 (Helm *et al.* 2000), however, the possibility that this activity was EPS-related could not be ruled out. The enzyme responsible for the detected activity catalyzed the transfer of glucuronic acid, a biosynthetic precursor of the NosA EPS constituent, which suggested that this enzyme might play a role in the biosynthesis of this structure. The possibility that this enzyme might not be exhibiting its natural activity *in vitro* was considered, as well as the possibility that the use of a synthetic substrate might be affecting the enzyme's specificity. For these reasons, purification of the enzyme responsible for this activity was attempted.

This chapter provides the protocols explored for purification of the putative glucuronosyltransferase and describes efforts to identify this protein on a polyacrylamide gel using affinity-labeling techniques. Attempts to gain sequence information from this enzyme by tandem mass spectrometry are also presented.

##### 3.2 Localization of Activity

The detection of a glycosyltransferase activity in the membrane fraction, but not in the soluble fraction, was an encouraging result. Bacterial EPS-synthesizing enzymes are commonly membrane-associated (Breton *et al.* 1998; Stingele *et al.* 1996). In order

to further explore the localization of the enzyme responsible for this activity and the manner in which this enzyme might be interacting with the membrane, additional cell fractions were tested for activity.

Attempts to detect extracellular glycosyltransferase activity in both the cell envelope and the growth media were made (details of the activity assays are presented in Chapter 7.3.3). Taking into consideration two examples in which sialyltransferase activity was extracted from whole cells using a non-ionic detergent (Gilbert *et al.* 1996), whole *Nostoc commune* DRH1 cells were extracted with 50 mM HEPES (pH 7.0) containing 0.4% (v/v) Triton X-100. The detergent extract was concentrated and assayed for transferase activity with all three potential monosaccharide acceptors and all four UDP-donors. The intact cell envelope material from the medium speed centrifugation step (following cell lysis) was tested also, but only for glucuronosyltransferase activity. Both fractions failed to reveal any activity. Following a *Nostoc commune* cell harvest, a concentrate of BG11<sub>o</sub> culture media was combined with an equal volume of 50 mM HEPES (pH 6.7), containing 20 mM MgCl<sub>2</sub> and assayed for transferase activity. One carbohydrate-modifying activity has already been noted to be released into the media by this organism (Hill *et al.* 1994), but this attempt was also unsuccessful in detecting glycosyltransferase activity.

In an attempt to better understand how this activity was associating with the membrane, a catalytically active membrane fragment suspension (demonstrating glucuronosyltransferase activity) was ultracentrifuged (150,000 x g, 1 hr), and the suspension buffer was removed from the pellet. Membrane fragments were washed with 20 mM sodium acetate buffer (pH 5.0), containing 0.5 M NaCl, and resuspended in 50 mM HEPES buffer (pH 6.7). No glucuronosyltransferase activity was detected after this wash, in either the resuspended membrane fragments or in the wash buffer. In a similar experiment in which membrane fragments were washed with 50 mM HEPES buffer (pH 6.7), containing 1.0 M NaCl, activity was retained by the membrane fragments. These results suggest that the glucuronosyltransferase activity is demonstrated by a protein or protein complex that either strongly adheres to or perhaps is embedded in the membrane.

### 3.3 Detergent Extraction of Glucuronosyltransferase Activity

A catalytically-active membrane fraction from *Nostoc commune* DRH1 was extracted with each of the following non-ionic detergents in 50 mM HEPES buffer (pH 6.7), overnight, at 4°C: 4.2 mg/mL Triton X-100 (33 x CMC), 0.69 mg/mL Tween 20 (11.5 x CMC), and 1.9 mg/mL Brij 58 (22 x CMC). Detergent extracts were separated from insoluble membrane material by centrifugation at 150,000 x g for 1 hour, at room temperature, and the soluble extract was assayed for activity. Only the Tween 20 detergent extract was unsuccessful in solubilizing the active protein (or proteins). The glucuronosyltransferase reaction product peak produced with the Triton X-100-extract was much more prominent than that produced with the Brij 58-extract. Based upon these results, the glucuronosyltransferase was purified as a Triton X-100 extract. No other transferase activity was detected in these extracts.

### 3.4 Partial Purification of Glucuronosyltransferase Activity

Fifty millimolar HEPES buffer (pH 6.7), containing 0.4% (v/v) Triton X-100, was used to prepare a detergent extract of a catalytically active membrane fraction. The 30 mL detergent extract was concentrated using two 15 mL Centriplus YM-50 centrifugal concentrators (50 kDa cutoff), and then the retentate (4 mL) and filtrate fractions were assayed for glucuronosyltransferase activity. Activity was detected only in the retentate fraction of the YM-50 concentrator, but was detected in both the filtrate fraction and the retentate when a concentrator with a 100 kDa cut-off membrane was used.

Material that demonstrated glucuronosyltransferase activity adhered to a Q-Sepharose strong anion exchange column that was equilibrated in a 50 mM HEPES buffer (pH 7.0), containing 0.4% (v/v) Triton X-100. After the column was washed with 0.1 M NaCl, active material was recovered by isocratic elution with 0.25 M NaCl in the same buffer/detergent system. Active fractions were pink in color, which suggested the presence of phycoerythrin. Although protein samples could be loaded onto the Q-Sepharose column at flow rates higher than 1 mL/min, attempts to continue the experiment at these higher flow rates resulted in the presence of active material in both the 0.1 M and 0.25 M NaCl eluates. Surprisingly, no activity could be found in either the

flow-through fraction or a 2 M NaCl eluant when active detergent extracts were fractionated by a strong anion exchange column (MonoQ) in the same buffer system.

Attempts to find another resin to which active material would bind, and then could be successfully recovered, were largely unsuccessful. Active material would not bind to a weak anion exchange resin (Sephadex DEAE A-25), though it did successfully adhere to and elute from Sephadex QAE A-25 resin and the Bond Elut (Varian, Inc.) Solid Phase Extraction (SPE) system's strong anion exchange resin (trimethylaminopropyl-derivatized silica). Active material could not be found in the flow-through fraction or a 2 M NaCl wash from a strong cation exchange column (MonoS), on two separate occasions, and attempts to bind active material to several other Bond Elut silica-based solid phase extraction resins (aminopropyl, diol, silica gel, C18, carboxylic acid, and benzenesulfonic acid) were unsuccessful. Active material did successfully bind to a MonoP chromatofocusing column, in HEPES buffer at pH 7.0, and elute with a 2 M NaCl wash. However, because of the previous problems encountered with the loss of enzyme activity in buffers with acidic pH values, the use of this column was not explored any further.

Affinity chromatography, a technique that takes advantage of an enzyme's specificity for a particular substrate, has previously been applied to glycosyltransferase purification with excellent success (Faik *et al.* 2000; Perrin *et al.* 1999). Several affinity-interaction resins were ordered or prepared. Pre-prepared single-atom spacer affinity resins, available from Sigma, were obtained including UDP (Sigma #U9751), 4-aminophenyl- $\alpha$ -D-galactose (Sigma #A2655, now discontinued), and 4-aminophenyl- $\alpha$ -D-glucose (Sigma #A2905, now discontinued). A UDP-affinity resin with a much longer spacer between the UDP portion and the agarose bead was created by attaching UDP-hexanolamine (Sigma #U2627) to 6-aminohexanoic acid *N*-hydroxysuccinimide ester-activated-Sepharose 4B (Sigma #A9019). Several other 4-aminophenol derivatized resins were created using the 6-aminohexanoic acid *N*-hydroxysuccinimide ester-activated-Sepharose 4B resin also; 4-aminophenyl- $\alpha$ -D-galactose, 4-aminophenyl- $\beta$ -D-galactose, 4-aminophenyl- $\beta$ -D-glucose, and 4-aminophenyl- $\beta$ -D-xylose were individually attached.

Two affinity resin preparation kits were utilized. Purified *Nostoc commune* EPS fragments were immobilized using the BioRad Affi-gel 102 kit. The Pierce Ultralink Immobilization kit was used to attach 4-aminophenyl- $\alpha$ -D-galactose,  $\alpha$ -D-glucuronic acid-1-phosphate, D-galactose, UDP-glucuronic acid,  $\alpha$ -D-glucose-1-phosphate, and  $\beta$ -D-glucose-1-phosphate. All of these affinity resins were equilibrated in a 50 mM HEPES buffer (pH 7.0) containing 0.4% (v/v) Triton X-100 and used to attempt to bind active material from a detergent extract, however, none were successful.

The final protocol for partial purification of the glucuronosyltransferase enzyme (or enzymatic complex) is shown in Figure 3.1, and described in Chapter 7.7.1. Most efforts to purify the protein (or proteins) responsible for this activity were unsuccessful because a) the detected activity was not stable in a detergent extract, and b) a chromatographic resin to which the protein (or proteins) responsible for this activity would bind could not be found.

### 3.5 Stabilization of Glucuronosyltransferase Activity

One benefit to using an ion-exchange step in the glucuronosyltransferase purification protocol was the apparent stabilizing effect that this partial purification and the final 250 mM NaCl concentration had upon glucuronosyltransferase activity. Previously, crude Triton X-100-detergent extracts could be stored for only a week before activity had declined to the point that it was no longer detectable, even in extracts to which a protease-inhibitor cocktail had been added. The active fraction from the ion-exchange column could be stored for as long as three weeks, at 4°C, and still show detectable activity. When this fraction was stored as a 50% (v/v) glycerol solution, detectable activity (though only barely detectable) could be retained for as long as two or three months.

It was recognized in the development of the purification protocol that freezing of *Nostoc commune* membrane material, or a detergent extract produced from this material, had a detrimental effect on activity. Whole cell mass that had been frozen at -20°C generated detectable activity in only rare cases. Cells that had been in long-term storage at -70°C did not produce any activity, and were extremely difficult to lyse using either the French pressure cell or by grinding with alumina. For these reasons, lyophilized

material was not used and all purification work was performed immediately after *Nostoc commune* DRH1 cell cultures had been harvested.

### 3.6 Characterization of Glucuronosyltransferase Activity

The dependence of activity on divalent metal cations was again addressed using a partially purified fraction that displayed glucuronosyltransferase activity. Several glycosyltransferases have been noted to show a dependence on divalent metal cations for activity, thus the frequent inclusion of  $Mg^{+2}$ ,  $Mn^{+2}$ , and sometimes  $Ca^{+2}$  in membrane resuspension buffers when assaying for a previously undetected glycosyltransferase activity (Gilbert *et al.* 2000; Jolly *et al.* 2002; Peugnet *et al.* 2001). Depending upon the strength of the interaction between a metal cation and the detected glucuronosyltransferase, the cation could potentially be removed during the ion-exchange step in the purification protocol. A set of activity assays were performed, in the presence of 20 mM DTT, with the following monovalent and divalent metal salts at a 40 mM concentration:  $MgCl_2$ ,  $MnSO_4$ ,  $CuSO_4$ ,  $LiBr$ ,  $KI$ ,  $NaCl$ ,  $ZnCl_2$ , and  $CaCl_2$ . Activity was detected in only the reactions that contained  $MgCl_2$  or  $MnSO_4$ . The reaction product peak from the  $MgCl_2$  -supplemented reaction was of greater intensity, suggesting that  $Mg^{+2}$  is the naturally preferred cofactor of this glycosyltransferase.

In a separate experiment, activity assays were performed on an active fraction that was previously purified by anion exchange (250 mM  $NaCl$  eluant), in the presence of 50 mM  $MgCl_2$  and 10 mM DTT, both separately and together. Only activity assay reactions that contained  $MgCl_2$  resulted in the detection of glucuronosyltransferase activity. The inclusion of DTT in activity assay reaction mixtures resulted in an increase in the size of the UDP-GluA-dependant reaction product peak in the electropherogram, but the presence of the reducing agent was not required for activity to be detected.

The dependence of this activity on temperature was also studied. The results of these experiments are shown in Figure 3.2. Activity assay reaction mixtures were prepared as described in Chapter 7.3.3, with the addition of  $MgCl_2$  and DTT, but these reactions were allowed to proceed for five days before analysis. Percentage activity was calculated by comparing the area of the UDP-GluA-dependant reaction product peak in the electropherogram to one from a room temperature (control) reaction run in parallel.

As previously noted, activity was irreversibly lost when whole cells, membrane fragments, and detergent extracts were frozen ( $< 0^{\circ}\text{C}$ ). Activity assay reactions that were incubated in the refrigerator ( $4^{\circ}\text{C}$ ) for five days produced no detectable reaction product, until they were allowed to incubate at room temperature. The amount of reaction product produced at  $37^{\circ}\text{C}$  was comparable to the amount produced at room temperature ( $22^{\circ}\text{C}$ ) at the end of this time period. This result is surprising since cultures of *Nostoc commune* DRH1 die when incubated at this temperature ( $37^{\circ}\text{C}$ ).

### 3.7 Analysis of Partially Purified Fraction Containing Glucuronosyltransferase Activity

The analysis of membrane proteins by two-dimensional electrophoresis is known to be difficult and problematic, however, methods have been developed to perform such analyses (Herbert 1999; Herbert *et al.* 1998; Molloy 2000; Rabilloud *et al.* 1997; Santoni *et al.* 2000). One never knows for sure whether a membrane protein will successfully separate by two-dimensional electrophoresis until after it has been attempted. An active ion-exchange fraction from *Nostoc commune* was precipitated with TCA, redissolved in rehydration buffer, and analyzed by two-dimensional gel electrophoresis to judge its purity. Details of the electrophoresis procedure are presented in Chapter 7.5.5. The resulting Coomassie-stained gel is shown in Figure 3.3. Approximately 30 to 50 spots are visible with this stain (Chapter 7.5.6), but there is no indication of which, if any of these are the protein responsible for the detected glucuronosyltransferase activity. Several spots from this gel were excised, subjected to in-gel digestion (Chapter 7.5.8), and analyzed by MALDI-TOF mass spectrometry (Chapter 7.3.2). An attempt to identify the proteins in these spots by Peptide Mass Fingerprinting (PMF) was made, using the *Anabaena (Nostoc)* PCC 7120 and *Nostoc punctiforme* ATCC 29133 genomes, but only a large smear of ATPase  $\beta$ -chain protein was successfully identified.

### 3.8 Non-radioactive Affinity Tagging of a Partially Purified Fraction Containing Glucuronosyltransferase Activity

The experiments that were performed up to this point revealed that the putative glucuronosyltransferase enzyme was relatively unstable. Furthermore, partial enrichment of activity followed by two-dimensional electrophoresis and mass spectrometric protein

analyses did not uncover any proteins with glycosyltransferase homology. Efforts thus turned to finding another way of identifying the protein that was responsible for activity, when the activity assay could no longer be used.

Following several published examples of glycosyltransferase identification by affinity-labeling (Delmer and Read 1987; Frost *et al.* 1990; Zeng *et al.* 1997), non-radiolabeled 5'-azido-UDP-glucose was purchased from Affinity Labeling Technologies and oxidized to 5'-azido-UDP-glucuronic acid with UDP-glucose dehydrogenase (EC 1.1.1.22), in 25 mM glycine-NaOH buffer (pH 8.7, 0.5  $\mu$ mol tag, 0.5 units of enzyme, 3 mM MgCl<sub>2</sub>, 17 mM NAD<sup>+</sup>). Reaction progress was monitored by thin-layer chromatography, on silica gel plates, in a methanol solvent system, in the dark. Compounds were visualized by charring with naphthoresorcinol spray. Within two hours, 5'-azido-UDP-glucose had oxidized completely to 5'-azido-UDP-glucuronic acid. The resulting reaction mixture was concentrated to dryness under vacuum, redissolved in 60  $\mu$ L water, and used in the tagging experiment described below.

Affinity tagging was performed with the enzymatically-synthesized substrate analog on 180  $\mu$ L of a partially purified *Nostoc commune* DRH1 active fraction. Samples were precipitated after the labeling procedure by the addition of one volume of 20% (w/v) TCA. The precipitate was redissolved in 100  $\mu$ L of 1X SDS-PAGE loading buffer and a 20  $\mu$ L portion was applied to a 12.5% SDS-PAGE gel. Lanes of the gel containing labeled protein, an unlabeled control sample, and ovalbumin were stained by a periodic acid-Schiff (PAS) staining procedure and a periodic acid-based silver stain for carbohydrates (Chapter 7.6.1) (Dubray and Bezard 1982). The only material visible by PAS staining was in the extremely low molecular weight region of the gel, and thus was likely to be derived from cell wall or storage polysaccharide that was carried along through the purification procedure. The ovalbumin standard was not visible by PAS staining. It could, however, be visualized using the periodic acid-based silver stain for carbohydrates. This technique manifested several species from both the tagged and untagged protein samples, but none of them appeared to be unique to the addition of the tag. Tandem mass spectrometric protein analyses of several samples from the silver-stained gel were unsuccessful in identifying any proteins, which was not surprising since



silver stains are known for interfering with the ionization of peptides in a mass spectrometer (Patton 2002).

### 3.9 Radioactive Affinity Tagging of a Partially Purified Fraction Containing Glucuronosyltransferase Activity

<sup>32</sup>P-labeled 5'-azido-UDP-glucuronic acid was purchased directly from Affinity Labeling Technologies and used to label a partially purified active fraction, according to the tag manufacturer's instructions (Chapter 7.7.2). Like its non-radiolabeled analog, this compound should interact with and attach to any UDP-GluA-binding proteins in a cell fraction. The <sup>32</sup>P isotope in this compound allows detection of these labeled proteins at extremely high sensitivity. In contrast to the previous experiment, excess tag was removed when labeled protein was precipitated with acetone, instead of TCA, to prevent acid hydrolysis of protein or the attached affinity tag.

The details of the labeling procedure are described in Chapter 7.7.2. Initially, precipitated protein was redissolved in rehydration buffer and analyzed by two-dimensional electrophoresis (Chapter 7.5.5). Autoradiographs of two-dimensional gels thought to contain the tagged protein showed little radioactivity. Radiography of a tagged sample, after being subjected to isoelectric focusing, revealed a focused spot, as well as a much larger amount of radioactivity at one of the electrodes that presumably represents excess and/or hydrolyzed tag (Figure 3.4). Excision of a portion of the IEF strip that contained the focused spot, incubation of this segment in SDS-equilibration buffer, and radioimaging revealed that the radiolabel had smeared across its length. These results indicated that although isoelectric focusing of the tagged protein or proteins had been successful, labeled protein did not actually enter the IEF strip. This is a common problem encountered in trying to analyze membrane proteins by two-dimensional electrophoresis (Molloy 2000; Santoni *et al.* 2000). Attempts to actively rehydrate the IEF strip at a low voltage, in an effort to assist the entry of protein into the IEF strip (Garfin 2003), were unsuccessful in producing a focused radiolabeled spot that did not wash away during the SDS-equilibration step. Later attempts to focus samples and then extract the tagged material from several IEF strips, and then separate this

material by one-dimensional electrophoresis (SDS-PAGE), were unsuccessful in producing a species that was detectable by radioimaging.

In another experiment, a partially purified active fraction was tagged according to the manufacturer's instructions, but excess tag was removed by dialysis instead of precipitation. The dialyzed mixture was subjected to anion-exchange (Q-Sepharose) chromatography in the same buffer system as the crude detergent-extract, as described in Section 3.4. Affinity-labeled material was eluted isocratically with 0.2 M NaCl (in HEPES/Triton), after a 0.1 M NaCl wash, and radioactivity was measured in each of the 1 mL fractions (10 total) collected, by liquid scintillation counting. The results of the liquid scintillation counting are shown in Figure 3.5. Protein in fractions 2 through 5 was precipitated by the addition of an equal volume of 5.3 M  $(\text{NH}_4)_2\text{SO}_4$ . Protein pellets were redissolved in 50 mM HEPES buffer (pH 7.0) containing 0.4% (v/v) Triton X-100 and combined with an equal volume of 2X SDS-PAGE loading buffer. The resulting mixture was heated and applied to a 7.5% SDS-PAGE gel. This was the most successful of several experimental attempts to coax radiolabeled protein into a SDS-PAGE gel, including one in which 5M guanidine-HCl was first used to disperse protein aggregates. An autoradiograph of the gel is shown in Figure 3.6. This gel was mounted on cardboard (in plastic) with several radioactive marker spots, radioimaged, and the resulting image was printed on cellulose acetate film. Overlaying the markers on the cardboard with those on the transparency allowed the radioactive portion of the unstained gel to be identified and excised. Two spots that were later submitted for protein analysis are indicated. Attempts to find a glycosyltransferase-related peptide in these samples are described in Section 3.11.

### 3.10 Attempts to Isolate a Radiolabeled Peptide

Although some components of the *Nostoc commune* active fraction were being labeled by the affinity tag, they were refractory to isolation by SDS-PAGE. Thus, an attempt was made to digest the tagged material with trypsin (in-solution) and isolate the affinity-labeled peptides using enrichment procedures for phosphorylated peptides (Ficarro *et al.* 2002; Raska *et al.* 2002; Salvucci *et al.* 1992). A  $^{32}\text{P}$ -labeled 5'-azido-UDP-glucuronic acid-tagged fraction (Chapter 7.7.2) was subjected to an in-solution

trypsin digestion (Chapter 7.5.10). The resulting peptide mixture was applied to an immobilized metal affinity column (IMAC) that had been charged with  $\text{Fe}^{+3}$  as described by Ficarro (2002). The column was equilibrated in 25 mM ammonium bicarbonate buffer at pH 8.0, and washed with the same buffer after sample application. Fractions were eluted with 0.5 M NaCl, 4 M urea, and 2 M sodium phosphate buffer at pH 8.0, and collected for analysis. Radioactivity in the flow-through fraction and eluates was measured by liquid scintillation counting. The majority of the radioactive material passed through the column and was detectable in the flow-through/column wash fraction (Figure 3.7). Attempts to concentrate radioactive material from this fraction using a C18 cartridge were unsuccessful. The phosphate eluate contained a small amount of radioactivity (approximately three times more than the background control sample). The phosphate eluate was dialyzed and submitted for analysis by mass spectrometry, however, nothing resembling a peptide was detected.

### 3.11 'Brute Force' de novo Sequencing of Radio-labeled Spots

Previous experiments indicated that one or more components of the active fraction from *Nostoc commune* could be labeled with a glucuronosyltransferase substrate analog. The labeled product, however, was not amenable to electrophoresis. Attempts to purify the protein responsible for the detected glucuronosyltransferase activity were hindered by the instability of this activity in a detergent extract. Because of these difficulties, a more direct, and perhaps more time consuming, approach was adopted. Five radiolabeled spots were chosen for protein analyses (Figure 3.6). The amino acid sequence of every peptide that was produced from these samples, for which a tandem mass spectrum had been generated, was determined. Peptide amino acid sequences were determined *de novo* from the b-ions (fragments that contain the N-terminal end of the peptide) and y-ions (fragments that contain the C-terminal end of the peptide) detected in the tandem mass spectrum of each peptide. These sequences were surveyed against the *Anabaena (Nostoc)* PCC 7120 and *Nostoc punctiforme* ATCC 29133 genomes, using the FASTA and FASTS algorithms (Chapter 7.8.2) at the Virginia Institute for Genomics (VIGEN) website ([vigen.biochem.vt.edu](http://vigen.biochem.vt.edu)).

An attempt was made to sequence approximately three hundred different peptides from the five samples. Of these, only sixteen decipherable sequences (excluding keratin peptides) were found, possibly due to low levels of protein in the samples, or possibly because of inadequate protease digestion. Of these sixteen, potential sequence matches for only three were identified in the *Anabaena (Nostoc)* PCC 7120 or *Nostoc punctiforme* ATCC 29133 genomic databases (Figure 3.8). Each of these three peptides appeared to be derived from a different protein, and none of them exhibited any recognizable sequence similarity to known glycosyltransferases. One of these peptides was noted to be in four of the five tagged samples analyzed, suggesting that it originated from the affinity-labeled protein or one that associates with it. The gene product of the ORF to which this peptide matched was therefore examined as a potential source of glucuronosyltransferase activity.

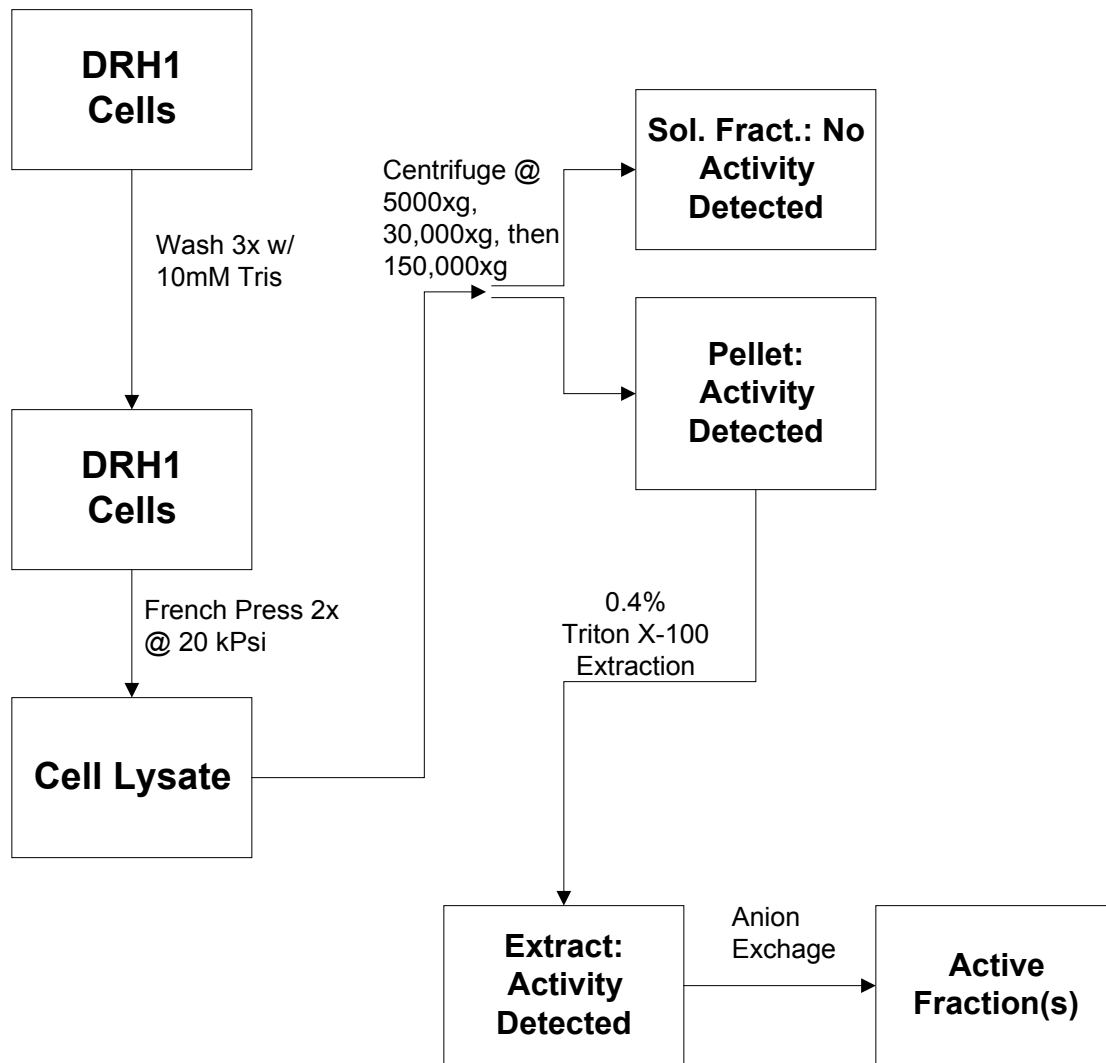
One possible *de novo* sequence interpretation of the peptide that appears in four of the five tagged and sequenced samples is LTLESVESR (Leucine-Threonine-Leucine-Glutamate-Serine-Valine-Glutamate-Serine-Arginine). This peptide is an excellent match for potential tryptic-cleaved peptides from *Anabaena (Nostoc)* PCC 7120 and *Nostoc punctiforme* ATCC 29133 (Figure 3.9), differing by two and one conservative amino acid substitutions, respectively. The GenBank database returned no other matches for this sequence. This peptide matches to homologous ORFs in both organisms, which lie in homologous gene clusters that may possibly constitute operons (Figure 3.10). Unfortunately, no information on the function of these ORFs or the proteins that they potentially encode is presently available.

Based upon this information and the inability to produce any more sequence information from the tagged samples, no additional effort was invested in this project. Several months later, in an online publication of a poster prepared for an American Society for Mass Spectrometry meeting by CEBI ([www.pil.sdu.dk](http://www.pil.sdu.dk)), it was noted that the ions from the possible LTLESVESR peptide corresponded to what the authors describe as a series of b-ions from a common keratin fragment, with the sequence TLLEGEESR (Table 3.1). The idea that this is only a keratin peptide is supported by similar information from the Proteome Sciences web site ([www.proteinworks.com](http://www.proteinworks.com)), thus closing the door on the idea that, in the worst-case scenario, glycosyltransferase sequence

information could be found by *de novo* sequencing every possible peptide in a crude sample and looking for glycosyltransferase homology.

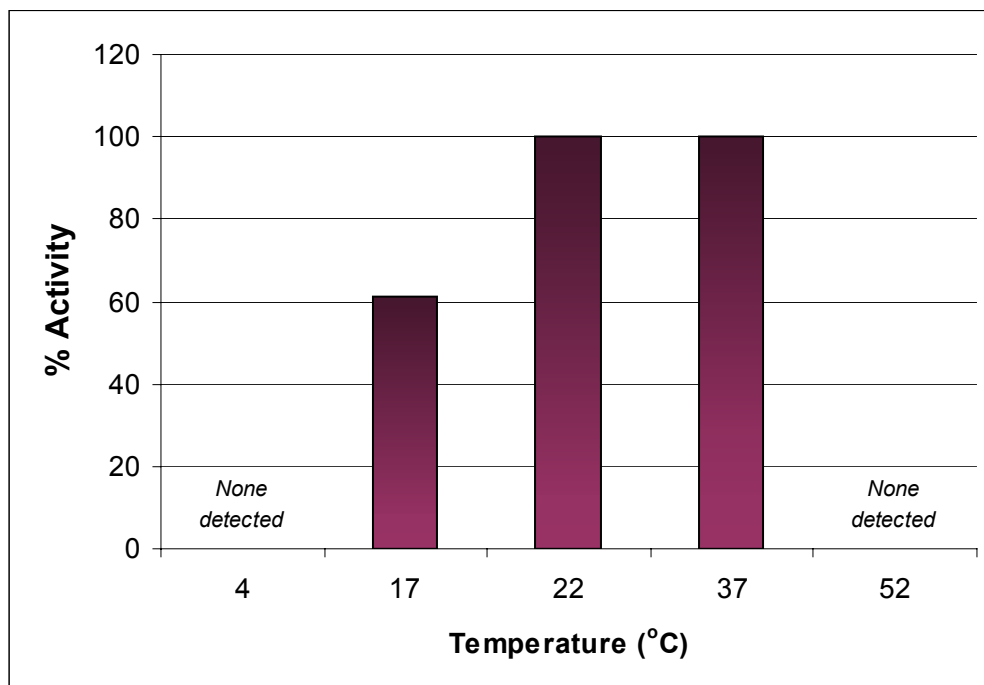
### 3.12 Summary

In this portion of the work, Specific Aims #3 and #4, presented in Chapter 1.2.6, were addressed. Attempts to purify a *Nostoc commune* DRH1 enzyme responsible for a detected glucuronosyltransferase activity met with limited success. An active detergent extract was produced from a membrane fraction, and partially purified, but not to a degree that the enzyme responsible for activity could be identified. Affinity labeling experiments indicated that the active protein forms aggregates that will not enter a polyacrylamide gel. Efforts to find the amino acid sequence of the enzyme responsible for glucuronosyltransferase activity by tandem mass spectrometric protein analyses of several of these aggregates were unsuccessful.



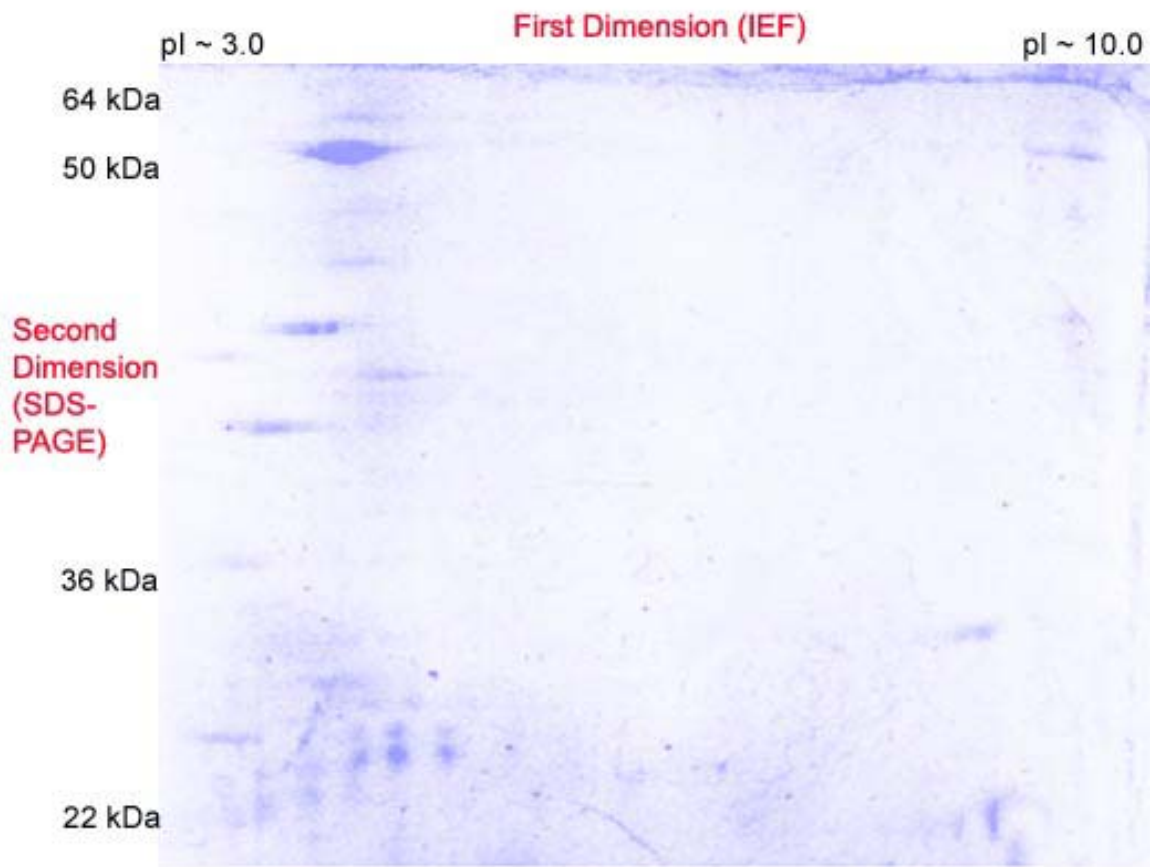
**Figure 3.1 Final Glucuronosyltransferase Partial Purification Protocol**

Purification protocol developed for the glucuronosyltransferase enzyme. The instability of the enzyme in a detergent extract made it difficult to find a purification technique that could be used.



**Figure 3.2 Temperature Dependence of Glucuronosyltransferase Activity**

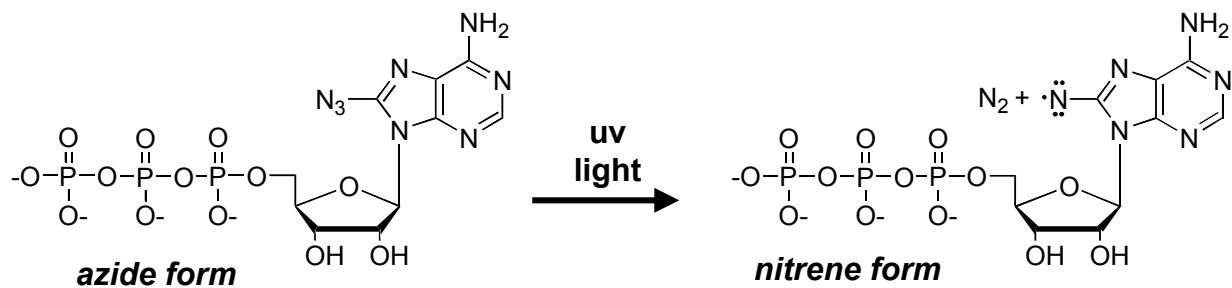
Glucuronosyltransferase activity measured as a function of temperature, relative to a room temperature control (22 °C). Detergent-extracted activity was stable for more than a day only if the extract was refrigerated. Freezing irreversibly destroyed activity. Surprisingly, activity measured at 37 °C is comparable to activity measured at room temperature, though *Nostoc commune* DRH1 will not grow at this temperature. Activity assays were prepared as described in Chapter 7.3.3 and allowed to proceed for five days.



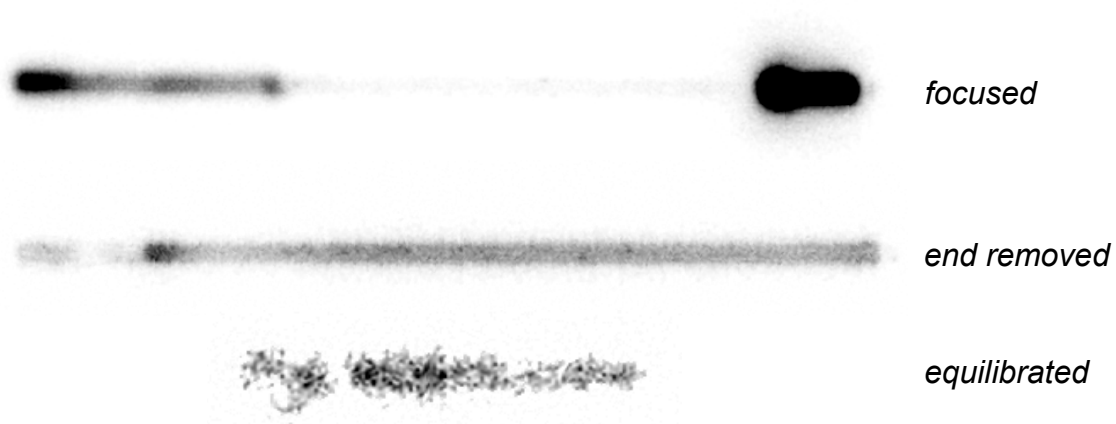
**Figure 3.3 Two-Dimensional Gel Analysis of a Partially Purified *Nostoc commune* DRH1 Protein Fraction Showing Glucuronosyltransferase Activity**

Coomassie stained two-dimensional gel of a *Nostoc commune* active fraction. There is no indication of which, if any of these proteins are responsible for the detected glucuronosyltransferase activity. Note the largest protein smear, in the upper left-hand corner; peptide mass fingerprinting along its length revealed the presence of large amounts of ATPase  $\beta$ -chain protein.





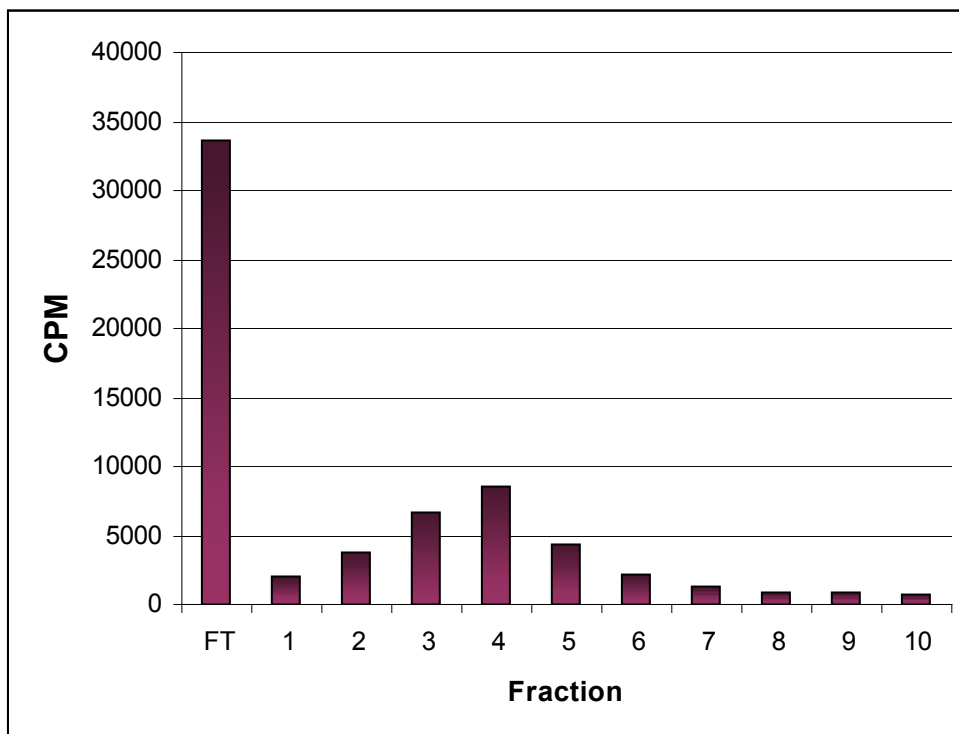
**AFFINITY TAGGING  
REACTION**



**Immobilized pH gradient  
IEF strips**

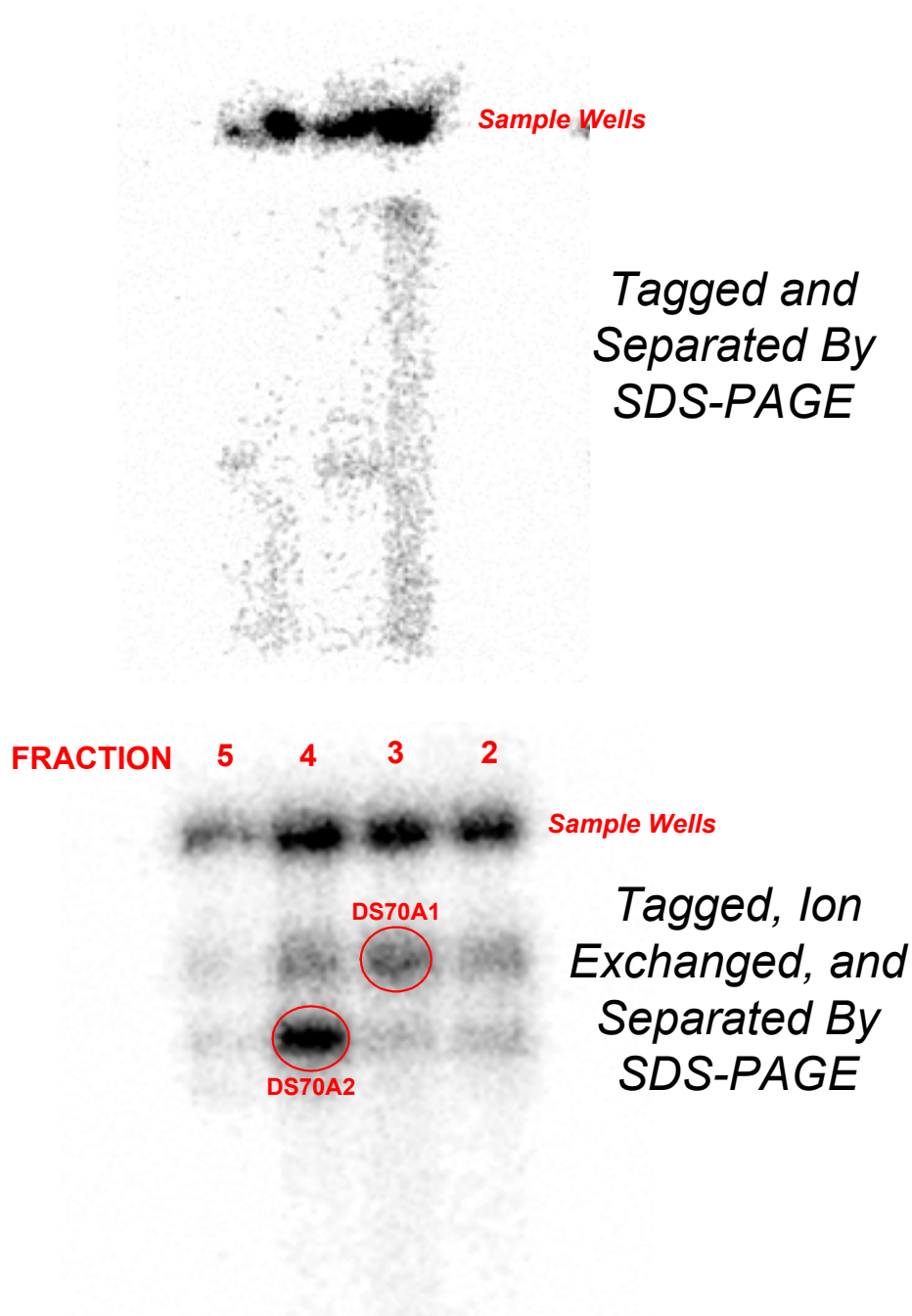
**Figure 3.4 Affinity Tagging Reaction and Subsequent Isoelectric Focusing of a Labeled *Nostoc commune* DRH1 Protein Fraction Showing Glucuronosyltransferase Activity**

*Above:* Mechanism by which  $^{32}\text{P}$ -labeled 5'-azido-UDP-glucuronic acid affinity tag is activated by ultraviolet light. Irradiation at 254 nm causes the azido group to lose dinitrogen; the resulting nitrene radical attaches to the enzyme's binding site. *Below:* Results of isoelectric focusing of a tagged fraction showing glucuronosyltransferase activity. Protein spots are more easily visualized when the end at which excess tag focuses is removed. Equilibration in SDS buffer disrupts focused spots, indicating that although labeled protein focused, it never actually entered the IEF gel.



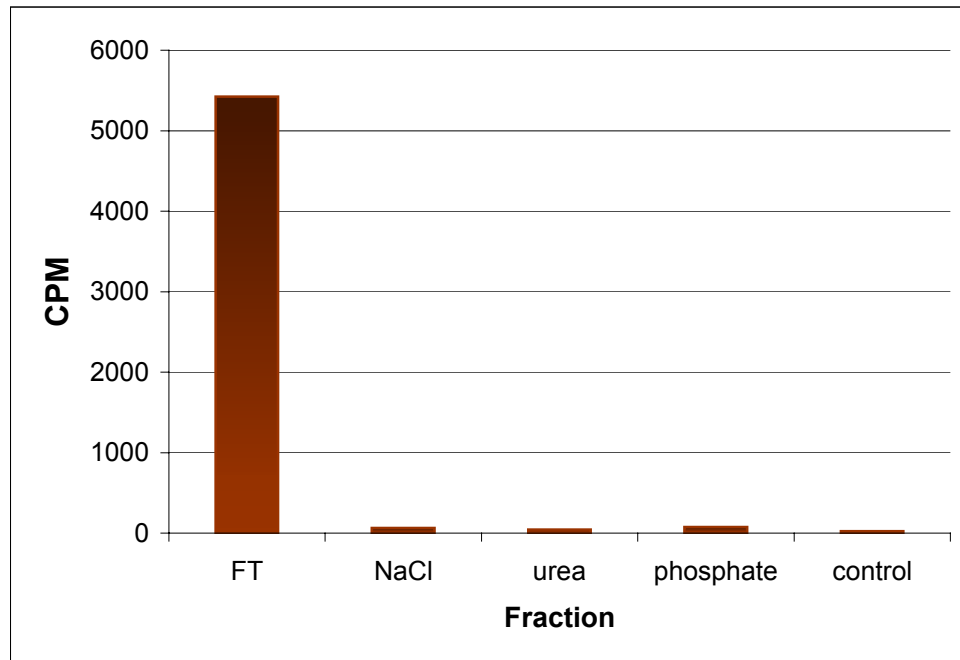
**Figure 3.5 Liquid Scintillation Counting of Anion Exchange Fractions from a Tagged and Dialyzed *Nostoc commune* Active Fraction**

A tagged *Nostoc commune* active fraction was purified by anion exchange (Q-Sepharose) to confirm that the correct protein had been labeled, and to enrich the sample in unaggregated material before SDS-PAGE. Fractions #2-5 were pooled and precipitated for electrophoresis. Fraction #3 contained the greatest amount of pink color, indicating that a pink protein was not responsible for glycosyltransferase activity. Twenty microliters of each fraction was combined with one milliliter of scintillation fluid for counting.



**Figure 3.6** Autoradiograph of a *Nostoc commune* Affinity-Tagged Fraction Separated by SDS-PAGE

The majority (*below*), if not all (*above*), of the affinity-labeled material from a *Nostoc commune* protein fraction showing glucuronosyltransferase activity stays at the top of the gel when subject to SDS-PAGE. Similar difficulties in getting the sample to enter the gel were experienced in isoelectric focusing experiments. A second anion exchange purification after tagging enriches for labeled material that will enter the gel (*below*). In addition to the two circled bands, three spots from the above gel (or a similar preparation) were sampled for protein analyses.



**Figure 3.7 Liquid Scintillation Counting of IMAC Fractions of a Tagged and Trypsin Digested *Nostoc commune* Active Fraction**

A tagged *Nostoc commune* active fraction was subjected to an in-solution trypsin digestion and an attempt was made to isolate the resulting affinity-labeled peptides by immobilized metal affinity chromatography (IMAC). Although the majority of the radiolabeled material was present in the flow-through and column wash (FT fraction), the phosphate eluate showed three times as much radioactivity as the background. Mass spectrometric analysis later indicated this radioactivity was only due to degraded tag material that adhered to the column.

<u>Sample</u>	<u>Peptide</u>	<u>Mass</u>	<u>Putative Match</u>
DS138A2	QQDQLK	758.5	
DS138A2	LTLESVESR	1034.5	gi23129182 (hypothetical protein), <i>N. punct.</i> , contig 497, gene 84
DS138A2	TVVNLAEK	875.5	
DS138A2	AEPEE	1265.7	
DS138A2	GELVDA	1713.8	
DS138A2	LNEQVDAVT	1477.8	
DS70A2	LTLESVESR	1034.1	gi23129182 (hypothetical protein), <i>N. punct.</i> , contig 497, gene 84
DS70A2	DVLE	1567.2	
DS138A3	LTLESVESR	1034.6	gi23129182 (hypothetical protein), <i>N. punct.</i> , contig 497, gene 84
DS138A3	LELSELAAK	973.5	
DS138A3	WTVLQEQER	1203.7	
DS138A3	VWDNVAR	1157.9	
DS70A1	LTLESVESR	1034.1	gi23129182 (hypothetical protein), <i>N. punct.</i> , contig 497, gene 84
DS70A1	VDPELQQSK	1042.2	alr0504 (unknown protein)
DS70A1	VTLGDTR	1038.2	
DS70A1	LPLGHAYLENK	1252.3	
DS70A1	WEAWFGFK	1109.2	
DS70A1	MESSGNR	1076.2	
DS70A1	[AQ]YEDLVVK	1065.3	gi4099045 (S-layer associated multidomain endoglucanase) alr2090 (unknown protein)

### Figure 3.8 *de novo* Sequenced Peptides and Peptide Fragments and Putative Matches

Peptides and peptide fragments determined by *de novo* sequencing of affinity tagged *Nostoc commune* protein samples. Sequences that compare to keratin contaminants have been omitted. Note that one peptide is present in all four tagged samples, suggesting that it originated from the affinity-labeled protein. A fifth sample produced only keratin peptides.

**LTLESVESR**

>>*hypothetical\_1629, Contig497, 84*

*(1385 aa)*

QUERY

**LTLESVESR**

hypoth QVVGTEPPVTVQVVVRGAPQ**KL**LTLET**VESR**IPADGRSTATIRGQLIDASGNRS

>>*all3289 ID:4143 {3966530 <--- 3970399} unknown protein*

*(1289 aa)*

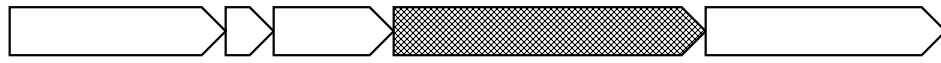
QUERY

**LTLESVESR**

all328 QIVGGTEAPVTLRVVVRGTAK**KL**TVET**VESR**VPADGRSTATIRGQLLDENGNR

**Figure 3.9 Putative *Nostoc punctiforme* ATCC 29133 and *Anabaena* PCC 7120 Peptide Matches**

Putative matches for a peptide that appeared in four *Nostoc commune* affinity-tagged samples that were *de novo* sequenced. Corresponding tryptic peptides from *Nostoc punctiforme* (*above*) and *Anabaena* (*below*) are shown. Sequence comparisons were made using the FASTS algorithm.



<b><i>N. punctiforme</i></b>	Npun5481	Npun5482	Npun5483	<b>Npun5484</b>	Npun5485
<b><i>Anabaena</i></b>	all3292	all3291	all3290	<b>all3289</b>	all3288
<b>ClustalW</b>	55	71	76	<b>74</b>	26

*all3284: homologous to cytosine deaminase*  
*all3286: hypothetical protein*  
*all3288: homologous to argininosuccinate lyase*  
*all3289: homologous to argininosuccinate lyase*  
*all3290: homologous to chromosome partitioning ATPase*  
*all3291: homologous to inositol monophosphatase family protein*  
*all3292: hypothetical protein*  
*alr3285: homologous to inosine-adenosine-guanosine-nucleoside hydrolase*  
*alr3287: homologous to nitrate transport protein NrtB*  
*alr3293: homologous to threonine synthase*  
*alr3296: homologous to low specificity L-threonine aldolase*

*Npun5481: homology to membrane glycoprotein*  
*Npun5482: hypothetical protein*  
*Npun5483: hypothetical protein*  
*Npun5484: hypothetical protein*  
*Npun5485: homology to RTX protein*

**Figure 3.10 Homologous *Nostoc punctiforme* ATCC 29133 and *Anabaena* PCC 7120 ORF Groupings**

The ORF clusters surrounding the ORFs to which a peptide from affinity-labeled samples matched (*Npun5484/all3289*). Corresponding ORFs in each organism and their ClustalW comparison scores are shown.

<b>TLLEGEESR</b>	<b>[LT]LESVESR (?)</b>
175.12	<i>corresponding b-ion</i>
262.15	262.1
391.19	391.2
520.24	520.3
577.26	577.3
706.30	706.3
819.38	819.3
932.47	932.3

**Table 3.1 Comparison of y-ions Produced by TLLEGEESR and by a Peptide Found in Several *Nostoc commune* Affinity Tagged Samples**

A comparison of the y-ions produced by a common keratin peptide (TLLEGEESR) and those produced by the peptide found in four affinity-labeled *Nostoc commune* samples. Although the  $y_1$  ion in the *N. commune* peptide was not observed, its complementary b-ion was seen.



## CHAPTER IV

### Discussion

#### **The Search for Genes That Encode *Nostoc commune* DRH1 EPS-Producing Enzymes**

##### 4.1 Introduction

The previous two chapters describe the results obtained from efforts to detect, purify, and determine the amino acid sequence of EPS-synthesizing glycosyltransferase enzymes. The purpose of this chapter is to review these results and to put them into perspective in terms of the overall goal of this project, which was defined as four Specific Aims in Chapter 1.2.6. Conclusions drawn from this work and suggestions for future experiments are also discussed.

##### 4.2 Development of A Method to Detect Glycosyltransferase Activities in *Nostoc commune* DRH1 (Specific Aim #1)

The synthetic acceptor, capillary electrophoresis-based glycosyltransferase activity assay of Wakarchuk has previously been used to detect glycosyltransferase activity in both *Campylobacter jejuni* and *Neisseria meningitidis* cell extracts (Gilbert *et al.* 1996; Wakarchuk *et al.* 1998; Wakarchuk *et al.* 1996). Additionally, it has been employed to test the use of glycosyl fluoride donors with glycosyltransferase enzymes (Lougheed *et al.* 1999). Glycosyltransferase activity is detected by monitoring the presence of a reaction product formed from the synthetic acceptor by this enzyme. Detection of activity with this assay requires not only that an active glycosyltransferase be present in the assay reaction mixture (likely as part of a cell extract), but also that electrophoretic conditions that are adequate for separation of the acceptor and any glycosyltransferase reaction product molecules be identified. Finding a combination of reaction and separation conditions that will successfully detect activity requires trial and

error, though some insight can be obtained by considering similar work performed with other organisms.

In prior works in which this capillary electrophoresis-based assay was utilized, reaction components were combined to produce a 20  $\mu$ L reaction volume, which was incubated 37  $^{\circ}$ C. Cell membrane material that had been separated from soluble protein by centrifugation at 100,000 x g for 1 hour was added to these mixtures as a buffer suspension (Wakarchuk *et al.* 1998; Wakarchuk *et al.* 1996). Assay reaction components were present at the following final concentrations: 25-50 mM buffer, 10 mM metal chloride, 5 mM DTT, 0.1-5.0 mM labeled acceptor (depending on its solubility in water), and 1 mM nucleotide donor, with various concentrations of enzyme. Activity assays were normally allowed to proceed between 1 and 15 hours, after which time they were either heated with an equal volume of 2% (w/v) SDS at 75  $^{\circ}$ C or diluted with 10 mM NaOH to terminate the reaction. Samples were then diluted with water prior to analysis by capillary electrophoresis. It should be noted that the use of this assay, or any other type of glycosyltransferase activity assay performed on ruptured cell material, relies upon the presence of a functional enzyme in the cell lysate; this is frequently not the case (Wakarchuk *et al.* 1998).

The final activity assay developed to detect glycosyltransferase activities in *Nostoc commune* was essentially identical to that described above, with minor modifications. Assay reaction components were combined to the following final concentrations, as described in Chapter 7.3.3: 25 mM HEPES buffer (pH 7.0), 50 mM  $MgCl_2$ , 10 mM DTT, 0.6-1.0 mM fluorescent labeled acceptor (depending on its solubility in water), and 2.5 mM nucleotide-diphosphate donor, also with various concentrations of enzyme. The reaction was performed at room temperature (22  $^{\circ}$ C), in a 22  $\mu$ L reaction volume, using 10  $\mu$ L of buffer-suspended cell membrane material that had been sedimented from a cell lysate by centrifugation at 150,000 x g for 1 hour. A control reaction, identical to the activity assay, except that water had been substituted for the nucleotide donor, was run at the same time and analyzed immediately following the activity assay reaction so that their electropherograms could be overlaid to identify the UDP-GluA-dependent reaction product peak.

The glycosyltransferase reaction reported herein proceeded more slowly than those reported for *Neisseria meningitidis* or *Campylobacter jejuni* extracts (Gilbert *et al.* 1996; Wakarchuk *et al.* 1998; Wakarchuk *et al.* 1996). Activity assays were thus normally allowed to proceed at least 16 hours, after which time they were diluted by a factor of 40 with borate buffer and analyzed by capillary electrophoresis. Because the rate of reaction was not being quantified, it was unnecessary to add any reagents to terminate the reaction before analysis.

A 1/40 dilution of reaction mixtures with CE separation buffer typically resulted in a UDP-GluA-dependent reaction product peak with a height of 1.0 RFUs in the electropherogram, when freshly prepared membrane material was assayed. This dilution level was sufficient to ensure that there was an adequate separation of the UDP-GluA-dependent reaction product peak and neighboring peaks in the electropherogram. The peak derived from unreacted acceptor was frequently off scale (capped at approximately 50 RFUs). Using larger sample injections (to look for other activities and additional transfers), the system was sometimes unable to separate the UDP-GluA-dependent peak from others with a similar migration time.

Alkaline borate buffer (pH 9.4) was chosen as the medium for capillary electrophoresis because it is used in the separation of other fluorescent-labeled molecules (APTS derivatives) that differ by a single carbohydrate (Chen and Evangelista 1995; Evangelista *et al.* 1996). SDS was initially added to the separation buffer when scouting for glycosyltransferase activity, but this detergent was later omitted after the observation that it was not necessary for the separation of the UDP-GluA-dependent reaction product was made. Faster migration of the reaction product in this buffer system, without SDS, supports the idea that a negatively charged moiety, presumably glucuronic acid, has been added to the acceptor.

At the onset of this project, three (potential) monosaccharide acceptors and four UDP-donors were used to assay for activity. The failure of the original three monosaccharide-derivatives to be used in detecting more than one glycosyltransferase activity reaction product suggested that most of the *Nostoc commune* glycosyltransferase enzymes present in the cell lysate could have fairly stringent structural requirements with regard to their acceptors. Thus, the synthesis and use of fluorescent-labeled acceptors

that more accurately mimicked a growing *Nostoc commune* EPS chain *in vivo* would be appropriate. Although some consideration was given to the idea of synthesizing a range of donors, this task would have been longer and more involved than the syntheses of alternative acceptors. Also, no information that indicated which of these donors would be successful in detecting activity could be derived from the EPS structure, whereas potentially successful acceptor structures are suggested. Further complicating the notion of donor synthesis was the possibility that nucleotide-diphosphate oligosaccharide donors could be utilized by this organism, though their presence has been observed only in eukaryotic systems to date (Upreti *et al.* 2003).

To summarize, although an extensive search for glycosyltransferase activity in *Nostoc commune* DRH1 was performed, only a single reaction product (implying that only a single activity) was detected. This glucuronosyltransferase reaction product could be detected regardless of the choice of assay reaction buffer, and it was formed at several pH values between 6.7 and 8.0. Standardized conditions were chosen to assay for this activity, which were later used in protein purification attempts. The reaction constituents used and separation conditions employed to assay for this activity are described above. A synopsis of various other reaction and separation conditions tested (without success) was presented in Table 2.1.

#### 4.3 Detection of EPS-Related Glycosyltransferase Activities in *Nostoc commune* DRH1

(Specific Aim #2)

Considering the relative instability of the glucuronosyltransferase activity, it came as no surprise that additional activities were not detected. The speed at which this activity deteriorated to non-detectable levels lent credibility to the idea that other enzymes, responsible for similar activities, could be even less stable in a cell lysate. Alternatively, it could be hypothesized that several such enzymes exist as an *in vivo* complex that is required for activity, and that disruption of this complex occurs during cell lysis. Because only a single glycosyltransferase reaction product was detected, there was no difficulty in deciding on which activity to focus purification efforts.

The product of the detected activity, glucuronic acid-linked galactose, does not correspond to the EPS structure of *Nostoc commune* DRH1 (Helm *et al.* 2000). Aside

from the enzyme's physical location and the presence of a glucuronic acid derivative in the EPS, there is little to suggest that this enzyme plays a role in EPS production. It would, however, be reasonable to hypothesize that this enzyme is EPS-related and that it exhibits an altered activity *in vitro*, or that the synthetic acceptor is interacting with this enzyme in a manner that mimics the glucose acceptor end of a growing EPS chain. However, the primary reason that this activity was pursued was that there were no other choices to consider.

#### 4.4 Partial Purification and Identification of the Protein Responsible for Glucuronosyltransferase Activity in *Nostoc commune* DRH1 (Specific Aim #3)

The detection of a membrane-associated glycosyltransferase, where EPS-producing enzymes would be expected, held the potential to lead this project towards its desired goal. Though only a single activity was detected, it was investigated further. Once the gene for the protein responsible for this activity had been found, adjacent EPS-producing machinery in the putative gene cluster might be identifiable by nucleic acid sequence similarity.

Typically, when a glycosyltransferase is purified from a membrane fraction, a detergent extract is produced, which is then fractionated using various forms of chromatography (Perrin *et al.* 1999). Triton X-100 is a non-ionic detergent commonly chosen to produce such an extract. Triton X-100 is advantageous in that it is uncharged and so it does not have to be removed prior to isoelectric focusing. Initial attempts to extract activity were more successful with Triton X-100 than with other non-ionic detergents. The active protein (or proteins) in a Triton X-100 detergent extract was retained by a 50 kDa molecular weight cut-off membrane during centrifugal concentration, but not by a 100 kDa molecular weight cut-off membrane. Although these membranes do not create a rigid molecular weight segregation, these results can be used to infer some information about the approximate size of the extracted enzyme, which is comparable to previously described glycosyltransferases (van de Rijn and Drake 1992; Wu *et al.* 1991; Yamashiki *et al.* 2002).

The adhesion of active material to an anion exchange column indicated that the glycosyltransferase might have an acidic isoelectric point. Enzymes are frequently

inactivated or precipitated at their isoelectric points and experimental observations indicating that this enzyme denatured in an acidic environment (pH 5.0 buffer wash) support this idea. This presumption, coupled with the above-mentioned indication that the glucuronosyltransferase enzyme has a molecular weight of greater than 50 kDa, led peptide mass fingerprinting attempts to focus on polypeptides located in the upper left-hand region of the gel shown in Figure 3.3. Assuming that the <sup>32</sup>P-labeled affinity tag bound to the protein of interest, this enzyme's isoelectric point can be estimated to a value between 6.6 and 6.8 (Figure 3.4), which explains why a strong anion-exchange column was successful in binding active material. It never became clear why the detergent-extracted enzyme would not bind to other chromatographic matrices.

Insight into the mechanism of this enzyme's inactivation can be gained by comparing radiolabeled affinity tagging experiments to activity-based protein purification work. During the course of these experiments, the activity assays, or any other procedure performed at room temperature, the activity of the detergent extract steadily decreased, which is ironic because this temperature was required for activity to be observed at all. Instantaneous deactivation occurred during protein precipitation techniques; no attempts to solubilize active material were ever successful after it had been precipitated from solution, regardless of whether acetone, TCA, or ammonium sulfate was used. Any activity remaining in a partially inactivated extract, perhaps one that had been in storage for a little too long could, however, subsequently be enriched by anion exchange chromatography.

The protein fraction subjected to affinity labeling was enriched in glucuronosyltransferase activity, making it likely that any protein (or proteins) in this fraction that was affinity-labeled was also responsible for such activity. Interestingly, the protein (or proteins) that bound the affinity tag displayed behavior that is characteristic of a protein that is denaturing and aggregating, and thus losing activity. Aggregates of radiolabeled material that would not migrate into a SDS-PAGE gel or enter an IEF strip were formed at some step between the affinity labeling procedure and electrophoresis. Any protein that recognizes the affinity tag must be in its native (active) conformation to bind this molecule, thus these aggregates were not formed until after the labeling had taken place. Whether these aggregates were forming because of conditions created

during the excess tag-removal procedures (dialysis or protein precipitation) or because of sample exposure to laboratory temperature was not determined. Perhaps both played a role. After the labeling procedure, the entire sample did not adhere to an anion exchange column again. SDS-PAGE analysis revealed that material that would not adhere to the column during this procedure produced radiolabeled aggregates at the top of a gel, while NaCl eluates were enriched in species that would at least partially enter. This phenomenon is reminiscent of the enrichment of activity from a partially inactivated extract by ion-exchange chromatography described above.

Assuming that the enzyme responsible for the detected glucuronosyltransferase activity was labeled, these results indicate that enzymatic activity was being lost through protein unfolding and aggregation. These observations are consistent with what is presently known about protein folding and the ability of detergent additives to influence the process. Temperature induced denaturation of a protein is often the result of protein unfolding, followed by an irreversible process such as aggregation (Van den Burg *et al.* 1998). During the course of *in vitro* folding and unfolding reactions, transient states that expose hydrophobic surfaces are formed. When these states are present in moderate concentrations, intermolecular aggregation can occur by way of these exposed hydrophobic patches (Agard 1993).

Protein aggregates, the end result of hydrophobic interactions between incompletely folded polypeptide chains, should form less often at lower temperatures, as the folding/unfolding/aggregation processes slow (Goldberg *et al.* 1995). Detergents are frequently used to encourage protein refolding to recover activity, and are added to shift the balance in the folding/unfolding/aggregation process. The actual effect of a detergent on protein structure depends upon its interaction with folding intermediates, which must be considered on a case-by-case basis. Most non-ionic detergents (*e.g.*, Triton X-100) are non-denaturing, suggesting that the use of such would help to preserve an enzyme in its active state, but these detergents suffer a limitation in that they are less effective at disrupting protein aggregation. No attempts at protein refolding were made when it was discovered that neither 7 M urea nor 5 M guanidine-HCl could disrupt the affinity-labeled aggregates.

Aggregate formation may have been unavoidable. Although labeling experiments such as these are typically performed with the protein sample on ice, it was assumed that tagging of the glucuronosyltransferase enzyme must be undertaken at room temperature since this enzyme shows no detectable activity at 4°C. During a room temperature labeling procedure, protein unfolding and aggregation processes would be expected to accelerate. Although the formation of aggregates could potentially be reduced by lowering the sample temperature immediately after labeling, the precipitation and dialysis procedures used to remove excess radiolabeled affinity tag were also conducive to aggregate formation.

Without a means to narrow down the number of glucuronosyltransferase candidates, and without a way to separate and identify this protein on a gel, this project came to a standstill. Previous plans simply to try and identify every spot on a gel, if necessary, to find the transferase, were abandoned when experimental evidence indicated that this protein likely was not even in the gel. With few, if any, other options for pursuing the transferase, several affinity-labeled spots (mostly well-immobilized protein aggregates) were chosen and digested with trypsin. Every peptide detected in these samples was *de novo* sequenced, in order to determine if a protein with glycosyltransferase homology could be found in the mixture.

#### 4.5 *de novo* Sequencing of a Complex Sample Thought to Contain a Glucuronosyltransferase from *Nostoc commune* DRH1 (Specific Aim #4)

Five affinity-labeled protein spots from three different gels were sampled. Two of these spots were observed to enter the gel, at least partially. The other three spots, which appeared to be protein aggregates, were immobilized at the entry point of the gel. There was some question of whether the protease used to produce peptides prior to mass spectrometry could digest enough aggregated protein material for analysis, but no way to disrupt these aggregates had been found, so the normal digestion procedure was attempted.

A surprisingly low number of decipherable sequences was found for the number of peptides examined (300), the majority of which were keratin contaminants. The signals generated in most mass spectra were very faint, perhaps because of an insufficient



amount of digested protein. Of the few (non-keratin) peptide sequences that were successfully determined, no two matched to the same open reading frame in either the *Nostoc punctiforme* or the *Anabaena (Nostoc)* genomes. None of these peptide or partial peptide sequences matched to an open reading frame that was identifiable as a glycosyltransferase.

#### 4.6 Conclusions

In terms of the overall goal of this work, the primary problem encountered was the inability to detect glycosyltransferase activity. The question of which glycosyltransferase activities were more likely EPS-related was not addressed because only a single reaction product was detected. Although the fluorescent-labeled synthetic acceptor, capillary electrophoresis-based glycosyltransferase activity assay offered advantages over alternative methods of glycosyltransferase activity detection (*i.e.*, TLC-based assays), it may not have been suitable for use with this organism, at least not with only the present set of synthetic acceptors. Whether the instability of the *Nostoc commune* DRH1 glycosyltransferase enzymes in a cell lysate or the inability of the present set of synthetic acceptors to produce the appropriate enzyme-substrate interactions for addition to occur was the true stumbling block remains unclear. It could be speculated that this detection system would have more success with different fluorescent-labeled acceptors, but unlike other works (Gilbert *et al.* 1996; Wakarchuk *et al.* 1998), this project was impeded by the commercial unavailability of disaccharide compounds corresponding to the *Nostoc commune* DRH1 EPS structure.

The pursuit of a glycosyltransferase enzyme in this work was based upon the assumption that a glucuronosyltransferase activity was responsible for the reaction product detected by capillary electrophoresis. Although the UDP-GluA-dependent reaction product was never successfully isolated and characterized, both its electrophoretic migration behavior and affinity-tagging experiment results support the idea that a glucuronosyltransferase reaction took place. The possibility that the fluorescent-labeled galactose acceptor was uridinylated still exists, however this possibility seems unlikely when it is considered that the same reaction product was not observed when other UDP-activated donors were substituted for UDP-GluA. This

synthetic acceptor-based method has been used not only to detect glycosyltransferase enzymes in other organisms, but also to demonstrate glycosyltransferase reaction product structure. In one instance this was done by NMR (Gilbert *et al.* 2000), and in another by successive transferase and glycosidase reactions with a synthetic acceptor (Wakarchuk *et al.* 1996). Low activity levels prevented the formation of sufficient reaction product for structural characterization by NMR, and the inability to find a commercially available glucuronosidase blocked efforts to enzymatically demonstrate that a particular carbohydrate linkage was formed. These fluorescent-labeled compounds did not ionize well enough for structural confirmation by MALDI-TOF mass spectrometry either.

Loss of glucuronosyltransferase activity seemed to be a consequence of protein aggregation. Although an affinity tag appeared to be successful in radiolabeling the enzyme responsible for this activity, the labeled enzyme soon after formed protein aggregates that were not amenable to electrophoresis.

The formation of protease-resistant protein aggregates are a common phenomenon (Callahan *et al.* 2001; Dickson 2002). These structures tend to form during mass spectrometry-based protein analysis procedures, upon the removal of SDS from gel-trapped membrane proteins. The formation of these structures limits the accessibility of the trypsin protease and results in low peptide ion intensities (van Montfort 2001). Additionally, high sequence coverage of a membrane protein is normally difficult to obtain because of the lack, or limited accessibility, of proteolytic cleavage sites in membrane-spanning segments (van Montfort *et al.* 2002). Thus, the inability of the trypsin protease to cleave a sufficient number of peptides from these affinity-labeled membrane protein aggregates for protein identification is understandable.

#### 4.7 Future Directions

If purification of the glucuronosyltransferase were to be continued, some means of preventing or minimizing enzyme aggregation must be developed. It is possible that the use of a different detergent would help maintain the protein in its active state, or at least prevent it from forming aggregates. Experimental evidence indicates that these aggregates limit the accessibility of the trypsin protease during the in-gel digestion procedure, however the same may not necessarily be true if the affinity-labeled protein

fraction were subjected to an in-solution digestion procedure, after which the resulting radiolabeled peptides could be separated on a Tris-Tricine gel. Alternatively, CNBr digestion is more suited to the large hydrophobic stretches that are characteristic of integral membrane proteins (van Montfort 2001; van Montfort *et al.* 2002), and its use in an in-gel digestion procedure would be hindered by less, if any of the same problems of accessibility posed by the protein aggregate. The use of gel filtration, perhaps in the presence of a high concentration of chaotropic agent, could present an alternative to size separation by electrophoretic techniques.

Even more problematic and significant to the ultimate goal of this project was the lack of successfully detected glycosyltransferase activities. When *Nostoc commune* DRH1 cell cultures are harvested, the organism is still actively producing and releasing EPS, as indicated by a steady increase in the viscosity of the culture media. If the enzymes responsible for EPS biosynthesis are present, and yet they cannot be detected, either those enzymes are being inactivated before the activity assay is performed, or the activity assay is inappropriate.

The *Nostoc commune* DRH1 cell is extremely sturdy. Attempts to lyse this organism's cells by means other than French pressure cell (*i.e.*, freeze-thaw cycles, grinding with alumina) have met with limited success. No gentler way of cell lysis for this organism has yet been identified. Assuming that it is possible to obtain an active glycosyltransferase enzyme in the cell lysate, attempts to find other ways to lyse *Nostoc commune* cells might circumvent enzyme inactivation, and thus result in the detection of additional activities. Alternatively, it is possible that something as simple as the use of additional protease inhibitors during cell lysis would do the same.

The EPS-related glycosyltransferase enzymes are not necessarily denatured upon lysis. It is very likely that the lack of detected activity stems from the inability of the synthetic acceptors to produce the appropriate enzyme-substrate interactions to allow carbohydrate transfer to take place. Purified glycan fragments from the *Nostoc commune* EPS (following derivatization) could be labeled with the FCHASE reagent to create larger glycosyltransferase substrates, so producing acceptors that are more representative of those used *in vivo*. These acceptors might participate in enzyme-substrate interactions that the monosaccharide acceptors were unable to, allowing carbohydrate transfer to

occur and activity to be detected. It is also possible that some other reaction condition, such as the correct donor or metal cofactor, has not been met. If this were the case, another method of identifying enzymes that interact with nucleotide donors, such as the affinity labeling procedure used with the glucuronosyltransferase, could provide a more successful approach. Its apparent success in identifying the glucuronosyltransferase enzyme might have led this project to a better outcome if the problem of protein aggregation had been effectively addressed. A TLC-based assay using radiolabeled acceptors (not fluorescent-labeled) or donors might successfully detect glycosyltransferase activity. However, a series of acceptors that correspond to *Nostoc commune* DRH1's EPS structure would first have to be synthesized, and there is always the possibility that the significant amount of carbohydrate material in this organism's membrane fraction would cause interference.

Abandoning the 'protein-to-gene' approach, the genetic locus (or loci) for EPS production might be identified by probing a genomic library of *Nostoc commune* DRH1 with homologous EPS-producing genes, once such a library is completed. This approach has already been successfully applied to *Neisseria meningitidis* MC58 and several *Lactococcus lactis* strains (Jennings *et al.* 1995; van Kranenburg *et al.* 1999). The low level of homology between the *Haemophilus influenzae lic2A* gene fragment used as a probe and the identified *Neisseria meningitidis* locus emphasizes the sensitivity and potential for success of this technique. As an alternative approach, gene sequences expected to be homologous to those involved in *Nostoc commune* DRH1 EPS production could be aligned and used to design oligonucleotide primers for PCR. Cloning and sequencing of any gene fragments so obtained might provide information that can be used to locate the genetic stretches encoding EPS production in this organism. Given the low levels of sequence similarity observed among other glycosyltransferases (Kapitonov and Yu 1999), this approach would be somewhat more risky than the previously suggested one. To date, no means to transfect *Nostoc commune* DRH1 with exogenous DNA has been found, but if this difficulty were overcome, perhaps by the use of a gene gun, a transposon mutagenesis strategy could be attempted. Considering the copious amounts of EPS produced by this organism, it is likely that a mutant deficient in EPS production could be identified by visual inspection alone.

## CHAPTER V

### Results

#### **Analysis of Potential Glycoprotein(s) from *Anabaena (Nostoc)* PCC 7120**

##### 5.1 Introduction

During the search for glycosyltransferase activity in *Nostoc commune* DRH1, several parallel experiments were performed on a closely related organism, *Anabaena (Nostoc)* PCC 7120. Although *Anabaena (Nostoc)* PCC 7120 does not produce the great quantities of extracellular polysaccharide for which *Nostoc commune* DRH1 is known, it does produce an extracellular polysaccharide coating around its heterocystous cells. The *Anabaena (Nostoc)* PCC 7120 genome sequence has been determined (Bryant 1994; Kaneko *et al.* 2001; Xu *et al.* 1997), which allows for protein identification by mass spectrometry, but without reliance upon time-consuming *de novo* sequencing procedures. It was reasoned that an examination of the proteins expressed in *Anabaena (Nostoc)* might reveal homologs of the proteins sought in *Nostoc commune*, thereby providing insight into EPS biosynthesis in the latter, phylogenetically similar organism.

This experimental work revealed that a soluble protein extract from *Anabaena (Nostoc)* PCC 7120 contained components that responded to both carbohydrate (PAS, periodic acid-Schiff) and protein stains, following separation by SDS-PAGE (Figure 1.8 and Figure 5.1). The apparent masses of these components ranged from 55-120 kDa. Ladder-shaped patterns of carbohydrate staining, such as those shown in Figure 1.8 and Figure 5.1, are sometimes indicative of proteins that are glycosylated to differing extents (Schaffer *et al.* 2002; Thornton *et al.* 2002). These observations suggested that one or more abundant cyanobacterial glycoproteins were present in this cell fraction.

In this chapter, experiments designed to address the Specific Aims presented in Chapter 1.3.7 are described. Putative glycoproteins, found in an *Anabaena (Nostoc)* PCC 7120 soluble cell extract, were investigated. Insight into the structure, function, and

association of the sample components (both protein and carbohydrate) that caused the ladder-like staining pattern was sought.

### 5.2 Protein Analysis by Tandem Mass Spectrometry

Protein-containing spots were picked from the SDS-PAGE gel shown in Figure 5.1, subjected to the in-gel trypsin digestion procedure (Chapter 7.5.8), and analyzed by tandem mass spectrometry (Chapter 7.5.9). The results of these analyses are listed in Table 5.1. Four proteins that contained putative SLH domains were identified, which were encoded by open reading frames *all4499*, *all7614*, *alr0834*, and *alr4550*. SLH domains are characteristic of (but not exclusive to) S-layer proteins, which are often extensively glycosylated and/or strongly associated with carbohydrate material (Engelhardt and Peters 1998; Moens and Vanderleyden 1997; Sleytr *et al.* 2001). Two of the identified proteins, the one encoded by *alr4550* in particular, spanned a relatively broad region of the gel. The PAS stain is not very sensitive and often requires high sample loadings to produce unambiguous results. Although a wide banding pattern such as this could potentially be the result of the considerable amount of material that had to be loaded onto the gel to produce a clear PAS-staining pattern, the presence of peptides corresponding to the *Anabaena (Nostoc) alr4550* ORF in samples 1, 4, 5, 6, and 10, but not in samples 2, 3, or 9 suggests that the migration of the *alr4550*-encoded protein varies because of its interaction with or attachment to the PAS-staining material.

### 5.3 Glycan Differentiation

Before purification of the potential glycoproteins, the *Anabaena (Nostoc)* PCC 7120 soluble fraction and a Triton X-100 extract of membrane material were probed using the DIG Glycan Differentiation Kit, as described in Chapter 7.6.2 and Chapter 7.6.3. This kit provides reagents that can be used to detect several different types of protein glycosylation, in a procedure similar to the Western blotting protocol, but different in that it employs a digoxigenin-conjugated lectin in place of the primary antibody. Digoxigenin-labeled lectins provided in the kit include *Galanthus nivalis* agglutinin (recognizes terminal mannose, linked  $\alpha(1\rightarrow3)$ ,  $\alpha(1\rightarrow6)$ , or  $\alpha(1\rightarrow2)$  to mannose), *Sambucus nigra* agglutinin (recognizes sialic acid, linked  $\alpha(2\rightarrow6)$  to

galactose), *Maackia amurensis* agglutinin (recognizes sialic acid, linked  $\alpha(2\rightarrow3)$  to galactose), peanut agglutinin (recognizes Gal- $\beta(1\rightarrow3)$ -GalNAc), and *Datura stramonium* agglutinin (recognizes Gal- $\beta(1\rightarrow4)$ -GluNAc). Stained PVDF strips, containing individually probed samples and corresponding lectin control protein lanes, resulting from this procedure, are shown in Figure 5.2. Only very faint bands, if any, appeared in the sample lanes upon incubation with the alkaline phosphatase-conjugated digoxigenin antibody and the NBT/X-phosphate solution. These bands became almost undetectable when the PVDF membrane sections had dried, indicating the extremely low abundance of the proteins contained within them, even at high sample protein concentrations; therefore, the use of lectin-affinity chromatography to purify potential glycoproteins from the sp. PCC 7120 soluble fraction was not explored. The bands detected in this procedure were of lower apparent molecular weight than those detected by the PAS-staining procedure.

#### 5.4 Purification of Comigrating Protein and Carbohydrate Material

The cell fraction found to contain PAS-staining material and SLH domain-containing proteins was purified, as described in Chapter 7.5.1 and Chapter 7.7.3, before further analyses. *Anabaena (Nostoc)* PCC 7120 cells, grown in BG11 media, were harvested and processed, using the same general procedures as those employed for *Nostoc commune* DRH1. Harvested cell material was ruptured preferably by grinding with alumina, because of the larger capacity of this technique in comparison to French press. Unlike those from *Nostoc commune* DRH1, the insoluble protein/membrane fragment pellets produced by the 100,000 x g centrifugation of this lysate were fragile and easily disrupted. Material that responded to the PAS staining procedure was identified in all four of the soluble, Triton-extracted, SDS-extracted, and SDS-suspended cell fractions. No immediate determination could be made as to whether the compositions of the PAS-staining materials in these four fractions were the same. In comparison to PAS-staining material that was obtained from cells that were lysed by freeze-thawing or by using the French pressure cell, the soluble and Triton X-100-extracted PAS-staining material was smaller in size when it originated from alumina-ground cells.

Upon separation from insoluble material, the soluble cell fraction possessed a protein concentration of roughly 20 mg/mL. A portion of this extract was fractionated with a Q-Sepharose column that was equilibrated in 50 mM Tris (pH 7.0). SDS-PAGE gel analysis of samples of the flow-through fraction and a concentrated 50 mM Tris (pH 7.0)/2 M NaCl eluate revealed that the carbohydrate material (and thus the putative glycoproteins) did not adhere to the column. It is therefore likely that, at pH 7.0, the PAS-staining material is either uncharged or positively charged. Only 18% (by wt.) of the protein loaded onto the column remained in the flow-through fraction, along with the PAS-staining material. The rich blue color of the soluble fraction, reflecting the presence of large quantities of phycocyanin, was retained by the exchange resin. A similar purification, utilizing a strong cation exchange resin (MonoS), was attempted. Carbohydrate material did not bind to this resin, and the amount of protein retained on the column was almost insignificant, thus purification procedures involving a MonoS ion exchange resin were not explored any further.

The combined Q-Sepharose flow-through and column wash fractions, which both contained the PAS-staining material, were collectively added to an equal volume of saturated (4 M) ammonium sulfate solution and incubated overnight at 4 °C. Precipitated material was collected by centrifugation and washed once with saturated ammonium sulfate. Experiments in which the retentate and flow-through fractions of centrifugal concentrators were analyzed by SDS-PAGE and PAS-staining indicated that the carbohydrate material would not pass through a 100 kDa cut-off membrane. Attempts to wash the majority of non-carbohydrate associated or linked proteins out of the sample, using a Centriplus concentrator, were unsuccessful as these samples tended to clog the membrane. An attempt to concentrate the PAS-staining material and associating or linked proteins using acetone, as a precipitant, was also unsuccessful.

Analyses of specific protein and carbohydrate-staining bands, or entire regions of a gel that contained these bands, were performed on samples purified by preparative-scale SDS-PAGE and subsequent electroelution. Following ion exchange chromatography and an ammonium sulfate precipitation, samples were dissolved in 1X SDS-PAGE loading buffer and separated on BioRad Criterion SDS-PAGE gels with 800  $\mu$ L sample wells. PAS-staining regions that formed individual bands, or entire PAS-



staining regions between particular molecular weight markers, were excised. Carbohydrate and protein material was recovered from these gel slices, using ProteoPLUS electroelution tubes, and concentrated using Centriprep and/or Centriplus centrifugal concentrators.

Electroeluted samples were lyophilized, in preparation for carbohydrate composition analyses. These dried samples were subjected to a base hydrolysis (of protein material) and an acid hydrolysis (of carbohydrate material), preceding APTS derivatization of the resulting monosaccharides. Samples that were not purified by ion exchange chromatography or preparative-scale electrophoresis were subjected to the same protocols. A protein/carbohydrate-containing sample from cells ruptured by freeze-thawing was also analyzed.

### 5.5 Protein Deglycosylation Attempts

A control experiment was performed to ensure that the PAS-staining material observed was not incompletely degraded peptidoglycan. A soluble fraction and a Triton X-100 extract of membrane material from *Anabaena (Nostoc)* PCC 7120 were each combined with equal volumes of a 10.1 mg/mL solution of egg white lysozyme, dissolved in 1 M Tris (pH 8.5). Enzymatic reactions, and control samples containing buffer in place of the lysozyme solution, were incubated overnight at 37 °C in a heating block. Following incubation, samples were heated to 100 °C with an equal volume of 2X SDS-PAGE loading buffer and electrophoresed on a BioRad Criterion 4-15% gradient SDS-PAGE gel. The resulting PAS and Coomassie-stained gels are shown in Figure 5.3. The presence and size of the carbohydrate material was unaffected by the additional lysozyme treatment, indicating that the primary peptidoglycan polymer was not the material responding to the PAS stain.

In order to examine the possibility that the ‘ladder-like’ staining pattern, observed with the PAS stain, reflected a series of glycoforms of one or more proteins (perhaps S-layer), several glycoprotein deglycosylation procedures were tried. Two distinct sample bands, located between the 50 kDa and 64 kDa molecular weight markers on a SDS-PAGE gel, and which stained with both the PAS and Electro-Blue stains, were individually electroeluted and concentrated. The two resulting samples were labeled

‘upper’ and ‘lower,’ based upon their position between the two molecular weight markers, and contained 58  $\mu\text{g}$  and 8  $\mu\text{g}$  of protein, respectively. These samples were subjected to the denaturing protocols of Prozyme’s GlycoPro Enzymatic Deglycosylation Kit and prO-Link Extender Enzymatic Deglycosylation Kit, which collectively contained the following glycosidases for removal of *N*-linked and Core 1 and Core 2 type *O*-linked oligosaccharides: PNGaseF (removes *N*-linked oligosaccharides, unless fucosylated at the core), Endo-*O*-Glycosidase (removes the Core 1 disaccharide, sometimes the result of enzymatic degradation of a Core 2 structure), Sialidase A (removes non-reducing terminal, branched and unbranched, sialic acids),  $\beta(1-4)$ -Galactosidase (releases non-reducing terminal,  $\beta(1-4)$ -linked galactose, common to some Core 2 oligosaccharides), and Glucosaminidase (releases non-reducing terminal,  $\beta$ -linked GluNAc residues, a feature of Core 2 oligosaccharides). Following incubation, treated samples and controls were combined with 2X SDS-PAGE loading buffer, as described in Chapter 7.5.11, and separated on a BioRad Criterion 4-15% gradient SDS-PAGE gel. The resulting Coomassie-stained gel is shown in Figure 5.4. No increase in migration for either of the electroeluted samples was observed. The protein quantities of these reactions were too low to attempt PAS-staining of a similarly run gel.

The failure of the above-mentioned, commercially available glycosidases to produce faster migrating protein bands did not necessarily indicate that Core 1, Core 2, and unsubstituted *N*-linkages were absent in the putative glycoprotein (or glycoproteins), but it did indicate that these types of linkages did not attach the majority of the detected carbohydrate, if indeed there were a point or several points of attachment. Thus, a more general (chemical) deglycosylation procedure that employs anhydrous triflic acid (trifluoromethanesulfonic), described in Chapter 7.5.12, was attempted on protein/carbohydrate material electroeluted from the 98-120 kDa range of a preparative-scale SDS-PAGE gel. The results of triflic acid treatment of this material and a bovine fetuin control are shown in Figure 5.5. Chemical deglycosylation of the control protein occurred within 30 minutes, producing a protein band that migrated to a lower molecular weight position on an SDS-PAGE gel, however, no similar protein was ultimately apparent in the *Anabaena (Nostoc)* PCC 7120 samples. In the sample subjected to a 60 min triflic acid treatment, no protein was detectable at all.

In both the samples and in the controls, increasing protein degradation with treatment time can be seen, presumably due to acid hydrolysis of the peptide chain. Although some degree of protein degradation is incurred when using this reagent, carbohydrate hydrolysis occurs much faster, as indicated by the successful deglycosylation of the control protein. The disappearance of carbohydrate-staining material upon triflic acid treatment of the *Anabaena (Nostoc)* samples was expected. The disappearance of protein-staining material, as well, indicated that the amount of protein in these samples (electroeluted from the 98-120 kDa region) was very low. In regard to the presence of a covalent protein-carbohydrate linkage in these samples, the results of this experiment were inconclusive.

Evidence to support the hypothesis that a non-covalent interaction was taking place between the carbohydrate material and two SLH domain-containing proteins (encoded by the *all4499* and *alr4550* ORFs) was obtained from a procedure that more gently degraded the carbohydrate component of a protein/carbohydrate sample, using sodium periodate. Following electroelution from a SDS-PAGE gel, material with an apparent mass of 98-120 kDa was treated with sodium periodate, using the procedure detailed in Chapter 7.5.12. When samples subjected to this treatment were electrophoresed, few protein bands were detected, the number of which depended upon the concentration of oxidizing agent employed (Figure 5.6). A PAS-stained gel of the same samples showed that the carbohydrate material migrated to a lower molecular weight region, following treatment (50 mM and 100 mM samples), which is consistent with the idea that the carbohydrate material had been degraded and the interactions necessary for carbohydrate-protein association were disrupted.

The four samples designated in Figure 5.6 were subjected to the in-gel trypsin digestion procedure (Chapter 7.5.8) and analyzed by tandem mass spectrometry (Chapter 7.5.9). The results of these analyses are listed in Table 5.2. Treatment of the sample oxidized with 50 mM periodate resulted in three major protein-staining sample bands, two of which displayed an increased electrophoretic mobility. The band that ran fastest produced peptides from the SLH domain-containing proteins encoded by open reading frames *all4499* and *alr4550*. No other peptides were detected in this sample.

### 5.6 Identification of a Carbohydrate Binding Domain

The ability of carbohydrate material to affect the behavior and chemical properties of SLH domain-containing proteins has been noted (Sára *et al.* 1998a). Additionally, the utilization of protein glycosylation, as a means of protection against proteolysis, by bacterial organisms has been observed (Moens and Vanderleyden 1997; Ong *et al.* 1994; Tomme *et al.* 1995). If the carbohydrate in an electroeluted sample were indeed interacting with one or more sample protein components, certain parts of these proteins might be protected from proteolysis by this association (Sára *et al.* 1998a). Evidence to support the role of the SLH domain in binding to and interaction with peptidoglycan and/or peptidoglycan-embedded secondary cell wall polymer exists (Sára *et al.* 1998b), and it would not be surprising to find that these regions, in particular, are shielded from enzymatic degradation (Sára *et al.* 1998a). To test the hypothesis that SLH domains are involved in the apparent carbohydrate-protein interaction observed, an in-solution trypsin digest was performed, on material electroeluted from the 55-64 kDa region of a SDS-PAGE gel, using the procedure described in Chapter 7.5.10. Following overnight incubation, the digested sample and a protease-free control were separated on a BioRad Criterion 16.5% Tris-Tricine SDS-PAGE gel. If there were a particular region of a protein (*e.g.*, the SLH domain) that was interacting with carbohydrate, this segment would be expected to demonstrate the same slowed electrophoretic migration that was shown by the *all4499* and *alr4550* ORF-encoded proteins, in the presence of this material, when cleaved from the remainder of the protein.

The resulting Tris-Tricine gel, stained with both Coomassie G-250 and the PAS stain, is shown in Figure 5.7. A series of peptides produced by the digestion can be seen in the 4-10 kDa region of the gel, however, only a single Coomassie-staining band remains in the region containing carbohydrate. Analysis of this band by in-gel trypsin digestion (Chapter 7.5.8) and tandem mass spectrometry (Chapter 7.5.9) revealed two peptides, derived from the product of a single open reading frame. The sequences of these two peptides were GNVLGIFAGAQPYPAR (amino acid residue #504-518, 1533.75 Da average molecular weight) and GVQAGANVPYQVEGFYK (amino acid residue #519-536, 1956.14 Da average molecular weight), both of which were encoded by ORF *alr4550*. It is not clear whether this Coomassie-staining band is a single peptide that had

been protected from proteolysis by carbohydrate during the initial (in-solution) trypsin digestion, or whether it is comprised of two comigrating peptides. The apparent molecular weight (~30 kDa) of the peptide (or peptides) in this band is far greater than the calculated molecular weights of the above-mentioned species, whether compared individually or combined. This drastic apparent molecular weight difference is likely due to non-covalent peptide interaction with the PAS-staining material.

Contrary to expectations, these two carbohydrate-interacting peptides were not derived from the SLH domain (amino acid residue #78-121 or #91-134, see Table 5.3) of the *alr4550*-encoded protein. These two peptides were, however, derived from a region of the protein which is similar to the OprB carbohydrate-selective porin of *Pseudomonas aeruginosa* (Wylie *et al.* 1993; Wylie and Worobec 1995). This observation will be discussed further in Section 5.7.

### 5.7 Protein Function Analysis of SLH Domain-Containing Proteins

Using sequence and structure similarity, an attempt to predict the function of the SLH domain-containing proteins (encoded by *all4499*, *all7614*, *alr0834*, and *alr4550*) that were mentioned in Section 5.2 was made. The databases and programs utilized are noted in Chapter 7.8.3. The results of these analyses are summarized in Table 5.3.

As previously discussed, these four proteins contain a putative SLH domain, identified by sequence similarity. Judging by the hypothetical amino acid sequences of their open reading frames, all four of these proteins possess N-terminal signal peptides upon translation, which suggests that their final locations are outside of the cell cytoplasm. This prediction is not surprising because SLH domain-containing proteins are usually embedded in, attached to, or assembled around the outside of the outer membranes of gram-negative bacteria (Engelhardt and Peters 1998; Hansel and Tadros 1998; Sleytr *et al.* 2001; Smarda *et al.* 2002). All four of these proteins possess a C-terminal region that is similar to the OprB carbohydrate-selective porin of *Pseudomonas aeruginosa* (Wylie *et al.* 1993; Wylie and Worobec 1995). This observation is particularly interesting because a portion of the *alr4550*-encoded protein was shown to interact with carbohydrate in Section 5.6, and an OprB carbohydrate-binding site has been suggested (Wylie *et al.* 1993).

The hypothetical amino acid sequences of the proteins encoded by these four open reading frames are typical of known cyanobacterial porins although, in some of these examples, the assignment of *in vivo* porin functionality is only speculative (Hoiczky and Hansel 2000). A search of the SYSTERS (SYSTEMatic Re-Searching) database shows that all four of these proteins can be grouped into a single family that includes the SomA and SomB porins of *Synechococcus* PCC 6301 and PCC 7942, members of which all possess similar features and structural organization (signal peptide, SLH domain, typical gram-negative bacterial porin  $\beta$ -barrel region). Proteins in this family are thought to play a role in attaching the cyanobacterial outer membrane to the peptidoglycan layer (Hansel *et al.* 1998; Hansel *et al.* 1994; Hansel and Tadros 1998; Sauer *et al.* 2001). A ClustalX alignment of these four proteins and their *Synechococcus* homologs is shown in Figure 5.8.

Conventional hydropathy methods that are often used for predicting membrane topology are not applicable to gram-negative bacterial porins, though other methods of predicting the 14 to 16 membrane-spanning, anti-parallel  $\beta$ -sheet structures (10 to 18 such structures for autotransporter proteins) that form their barrel-shaped pores do exist (Henderson *et al.* 1998; Jeanteur *et al.* 1991; Schleiff *et al.* 2003). An indication of the number and sequence location of these  $\beta$ -sheet segments was sought by aligning a truncated *alr4550* hypothetical amino acid sequence, lacking the signal peptide and SLH domain, with known *Neisseria meningitidis* porins. This alignment is shown in Figure 5.9. Some, but not all, of the boxed regions ( $\beta$ -sheet segments) show sufficient sequence similarity to allow speculation that the same structural element is formed by both amino acid stretches. The *alr4550*-encoded protein does not possess all of the amino acids that are typically conserved in bacterial porins (Figure 5.9, red amino acids), or the bacterial porin PEFXG consensus sequence.

N-terminal signal peptides and C-terminal  $\beta$ -barrel regions are also characteristic of gram-negative bacterial autotransporter proteins. An additional attempt to delineate the individual structural elements of the putative *alr4550*-encoded porin was made, using the consensus sequences of autotransporter proteins instead, but was unsuccessful (Henderson *et al.* 1998).

### 5.8 The Effect of Nitrogen-Supplemented Media on SLH Domain-Containing Protein Expression

Changes in physiological state (e.g., cell surface hydrophobicity or cell wall composition and thickness) have been noted to occur in certain organisms in response to the presence or absence of various nutrients in the medium in which they are grown (Aguilar-Uscanga and Francois 2003; Huang *et al.* 1998). The possibility that sp. PCC 7120's four previously identified SLH domain-containing proteins might not be expressed in nitrogen-free media, or that additional SLH domain-containing proteins would be, was examined. A 15 L culture of *Anabaena (Nostoc)* PCC 7120 cells was grown in BG11<sub>o</sub> media and processed using the same protocols employed for the preceding (BG11-grown) harvest and purification. A PAS and Coomassie-stained SDS-PAGE gel, used to separate a sample of the soluble cell extract, in preparation for protein analysis, is shown in Figure 5.10. A composite sample of gel pieces that were excised from areas indicated by the arrows in this Figure was subjected to in-gel trypsin digestion (Chapter 7.5.8) and peptide analysis by tandem mass spectrometry (Chapter 7.5.9). Peptides corresponding to open reading frame *alr4550* were identified. These results indicate that the protein encoded by the *alr4550* open reading frame is the most abundant in the regions of the gel sampled, when sp. PCC 7120 is grown under nitrogen-fixing conditions. However, no determination of whether or not this protein was actually up-regulated, like the closely-related *Synechococcus* PCC 7942 SomA and SomB proteins, could be made with only this information (Sauer *et al.* 2001).

### 5.9 Carbohydrate Analyses

In order to gain some insight into the type of carbohydrate material with which the *alr4550*-encoded and other SLH domain-containing proteins were interacting, carbohydrate analyses were performed on samples electroeluted from PAS-staining regions of a SDS-PAGE gel. Several less purified samples were analyzed, in order to judge the degree to which this material contributed to the overall carbohydrate composition of the soluble cell fraction. Samples were prepared to examine how the compositions of these materials changed, when the organism was subjected to different growth conditions, as well. Analyses were performed using the procedures detailed in

Chapter 7.3.4 and Chapter 7.6.4 through Chapter 7.6.7. Prior to carbohydrate analysis, an ammonia-based  $\beta$ -elimination procedure, described in Chapter 7.6.4, was used to selectively hydrolyze the protein component of protein/carbohydrate-containing samples. Carbohydrate-derived peaks, observed in analysis electropherograms, were identified by comigration with APTS derivatives of known compounds. Derivatized sugars that migrated as a single peak in the borate system (*i.e.*, glucose and ribose) were separated in a MOPS system, in order to measure the relative amount of each compound present. Carbohydrate response factors were measured by derivatization and analysis of known amounts of various monosaccharides, and then used to adjust sample peak areas when calculating composition.

The carbohydrate compositions of five samples that were analyzed are shown in Table 5.4. An additional (soluble fraction) sample, which was not subjected to ion exchange chromatography or preparative-scale SDS-PAGE, produced an indecipherable electropherogram upon analysis, presumably due to the large number of additional (non-carbohydrate) compounds in this impure fraction that were able to react with the APTS labeling reagent. No indication (faster migrating, but later eluting peaks that were not present in the control reaction) of any negatively charged carbohydrate residues, such as uronic acids, was observed in the electropherograms of these samples. In most of these samples, *N*-acetyl mannosamine was the most abundant constituent. In several samples this monosaccharide was present with rhamnose, in equal molar quantities, which suggested their participation in a repeating unit of a larger structure. This pattern was apparent in a sample (BJ38F) that was not subjected to preparative-scale SDS-PAGE, as well as ones that were electroeluted from both narrow (BJ6F) and wide (BJ23F) regions of a gel. The detection of these saccharides in equal molar amounts, regardless of from where the sample was taken, would be expected in the analysis of material from a large cellular structure, wall, or coating (*e.g.*, *O*-antigen, secondary cell wall polymer), but not of material originating from smaller localized saccharides, like those attached to glycoproteins. Although the size of the PAS-staining material, observed on a SDS-PAGE gel, was somewhat smaller when cells were ruptured by grinding with alumina, rather than by French pressure cell or by freeze-thawing, its composition was largely



unaffected, which is consistent with the idea that this material is a size distribution of fragments emanating from a larger structure.

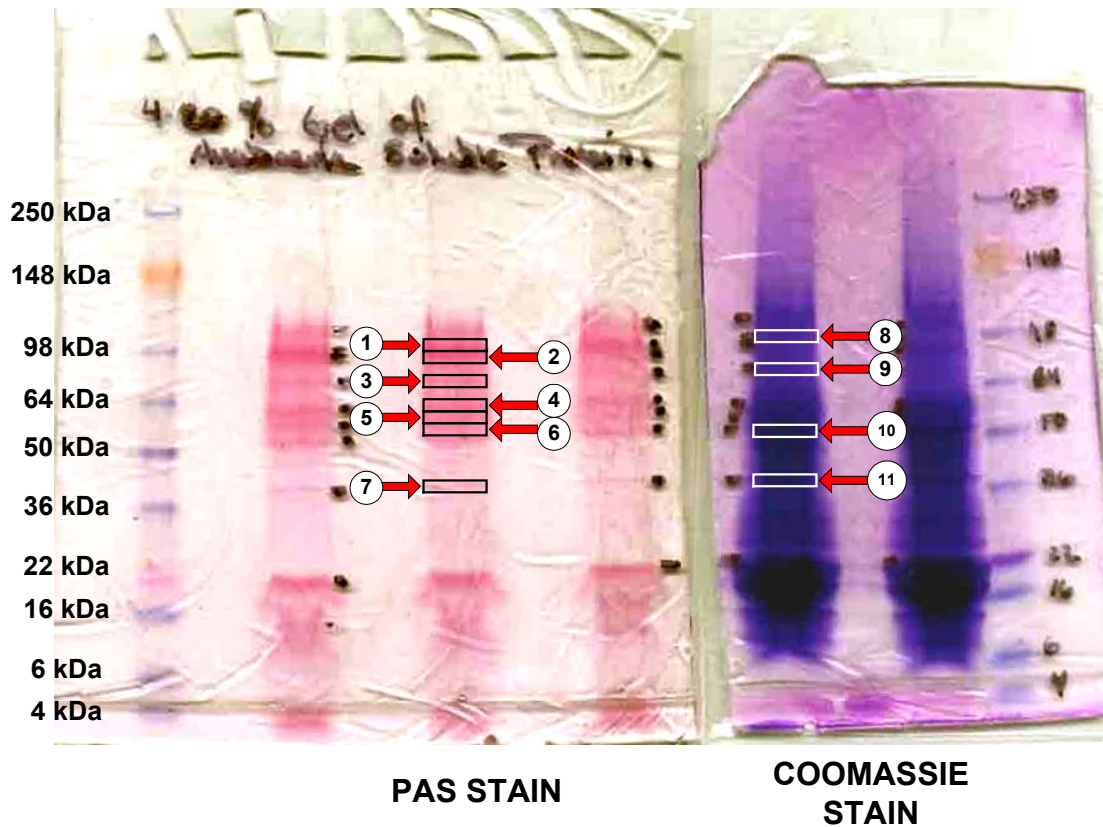
Somewhat perplexing was the unexpected high ribose content of BJ37F, especially when compared to an almost identically prepared sample (BJ38F). These samples were generated from the same cell harvest, and differ only in the method of cell rupturing used and in minor purification steps. BJ37F is the only sample in Table 5.4 that was not subjected to an ammonium sulfate precipitation. During such a procedure, it would be anticipated that any degraded nucleic acid material present in the sample would be removed by its tendency to remain in the high salt supernatant, rather than precipitating with the PAS-staining material and interacting protein. It could be hypothesized that the high ribose content discovered in BJ37F is the result of hydrolyzed nucleic acids or ATP that are normally removed from samples by the ammonium sulfate precipitation procedure. However, this hypothesis does not explain the high ribose content of BJ31F, or offer any explanation as to why *N*-acetyl mannosamine and rhamnose are not present in equal amounts in that sample.

Measurement of sialic acid content, by enzymatic conversion to *N*-acetyl mannosamine, is described in Chapter 7.6.6. All samples, with the exception of BJ37F, elicited a positive response for *N*-acetylneuraminic acid (a sialic acid), at levels approximately 30-fold less than that of the other carbohydrate constituents assayed. Whether this sugar originates from lipopolysaccharide, cell wall, or glycoprotein material is uncertain. The consistent detection of this saccharide, at levels 30-fold less than other detected saccharides, and in different high molecular weight regions of a preparative-scale SDS-PAGE gel, however, suggests that it is linked to the PAS-staining material.

### 5.10 Summary

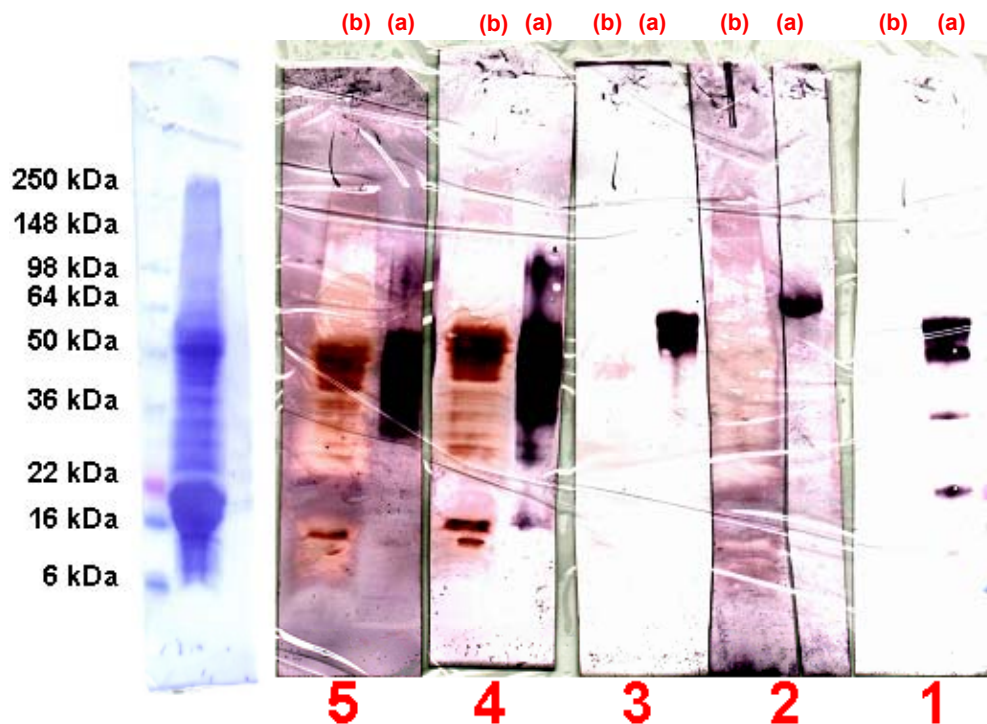
In this portion of the work, the Specific Aims presented in Chapter 1.3.7 were addressed. Evidence to support the idea that proteins encoded by the *Anabaena* (*Nostoc*) PCC 7120 *alr4550* and *all4499* open reading frames interact with carbohydrate material was found. Further evidence indicated that the bulk of, if not all of, this carbohydrate material is bound in a non-covalent manner. For the *alr4550*-encoded protein, this interaction occurs with a region formed by amino acid residues #504-536.

Monosaccharide composition analyses of this carbohydrate material were performed. Protein function analyses revealed features, in addition to the SLH domain, that indicate these proteins are located outside of the cytoplasmic membrane *in vivo*, and noted homologies with the *Pseudomonas aeruginosa* carbohydrate-selective outer membrane porin, OprB. The *alr4550*-encoded protein was expressed in both nitrogen-free and ammonium-containing media.



**Figure 5.1 SDS-PAGE Analysis of *Anabaena* PCC 7120 Soluble Cell Extract**

*Anabaena (Nostoc)* soluble cell extract, separated on 4-20% gradient SDS-PAGE gels, and stained with either Coomassie R-250 (*right*) or a periodic acid-Schiff's stain (*left*). All five lanes that are shown above were loaded with an equal amount of protein (~200  $\mu$ g). Spots chosen for protein analysis by protease digestion and tandem mass spectrometry are indicated.



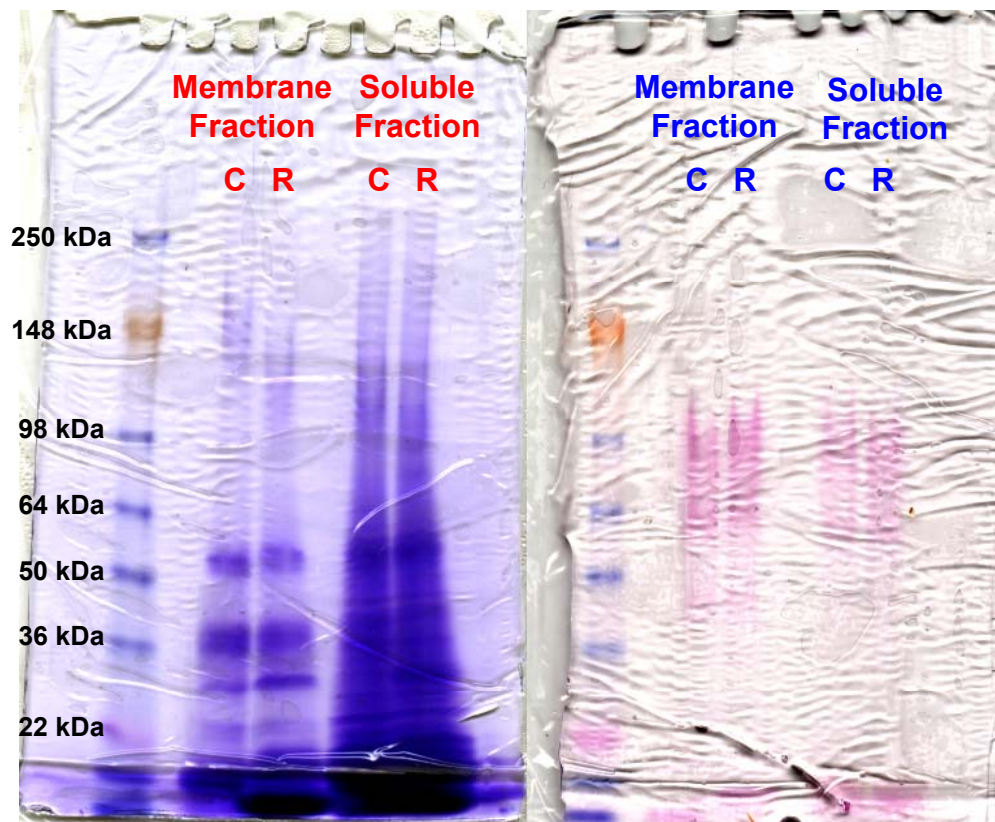
<u>Strip</u>	<u>Lectin Probe</u>
1	<i>Galanthus nivalis</i> agglutinin (GNA)
2	<i>Sambucus nigra</i> agglutinin (SNA)
3	<i>Maackia amurensis</i> agglutinin (MAA)
4	Peanut agglutinin (PNA)
5	<i>Datura stramonium</i> agglutinin (DSA)

<u>Lane</u>	<u>Control</u>
1a	glycoprotein carboxypeptidase Y
2a	glycoprotein transferrin
3a	glycoprotein fetuin
4a	glycoprotein asialofetuin
5a	glycoprotein asialofetuin

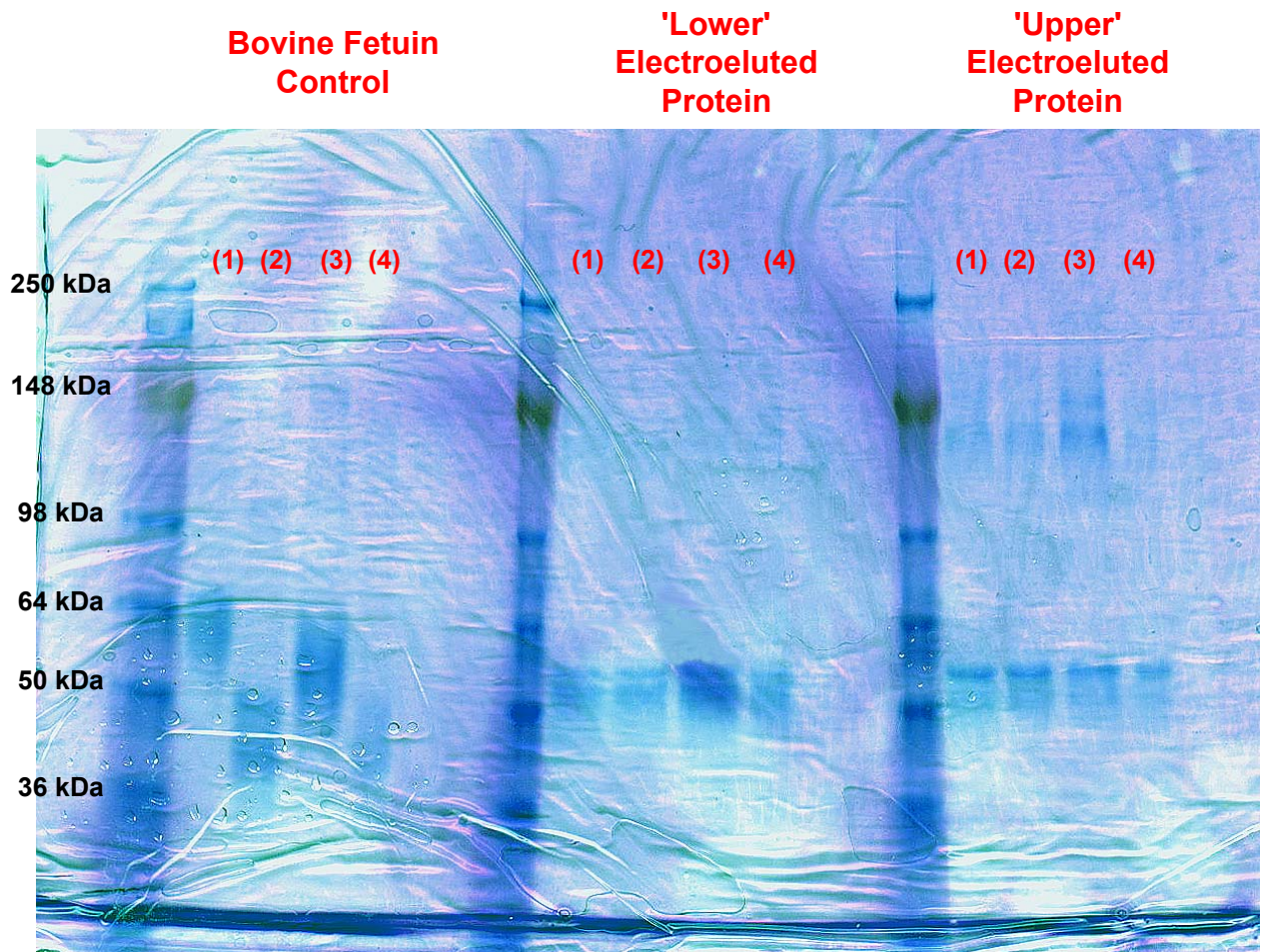
**Figure 5.2 Glycan Differentiation of *Anabaena* PCC 7120 Soluble Cell Extract**

Application of the DIG Glycan Differentiation Kit to the *Anabaena* (*Nostoc*) PCC 7120 soluble fraction. A faint positive response is seen in lanes 4 and 5, when the extract is probed with digoxigenin-conjugated peanut and *Datura stramonium* agglutinins. These lectin-detected bands are in an apparent molecular weight region of the gel below the area where the PAS-staining material was found. No positive responses were produced when a Triton X-100 extract of membrane material was tested.



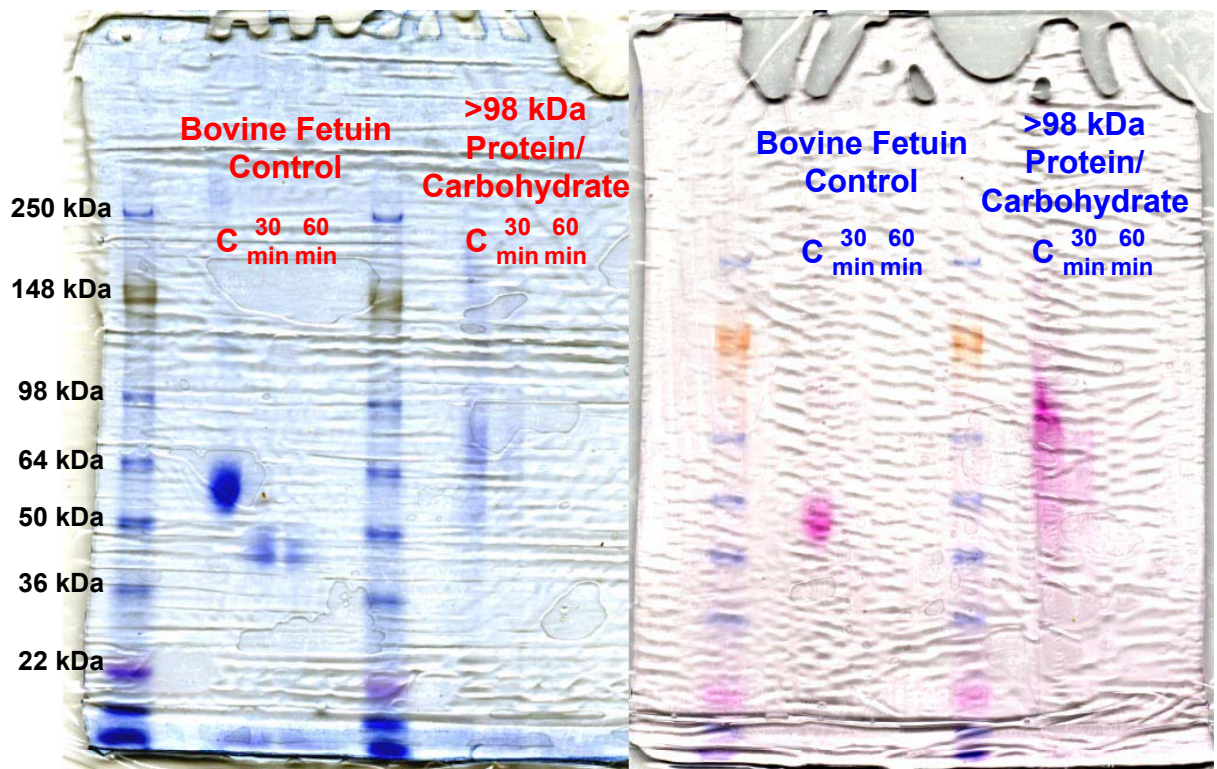
**Figure 5.3 Lysozyme Treatment of Carbohydrate Material in *Anabaena* PCC 7120 Cell Extracts**

Attempted lysozyme digestion of carbohydrate material detected in *Anabaena* (*Nostoc*) PCC 7120 cell fractions. Coomassie (*left*) and PAS-stained (*right*) SDS-PAGE gels of the enzymatic reaction mixtures (containing 5 mg/mL lysozyme) and their corresponding controls (containing only buffer) are shown. Digestion of the carbohydrate material was not observed.



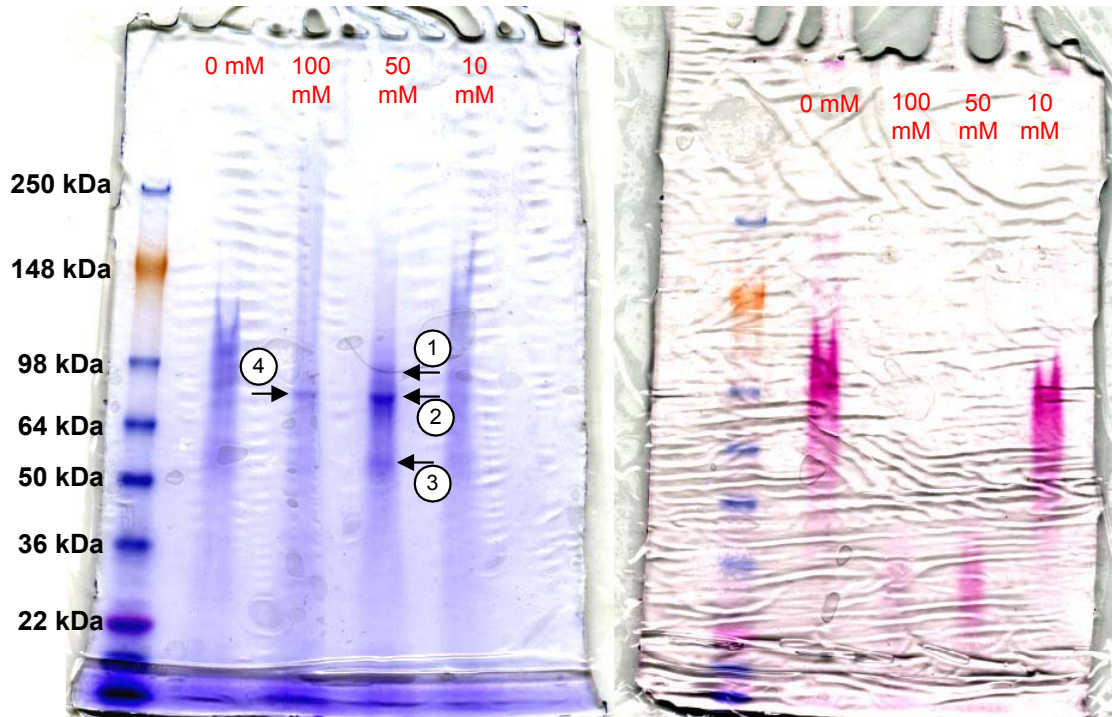
**Figure 5.4 Glycosidase Treatment of Suspected Glycoprotein(s) in *Anabaena* PCC 7120 Soluble Cell Extract**

Attempted deglycosylation of suspected glycoprotein(s), purified from an *Anabaena* (*Nostoc*) PCC 7120 soluble extract, using PNGaseF and glycosidases that are specific for core I (lane 2) and core II (lanes 3 & 4) type *O*-linked saccharides. No change in migration was observed for either sp. PCC 7120 sample following enzyme treatment. **(Lane 1: no glycosidases; Lane 2: PNGaseF + Endo-*O*-Glycosidase + Sialidase A; Lane 3:  $\beta$ (1-4)-Galactosidase + Glucosaminidase; Lane 4: PNGaseF + Endo-*O*-Glycosidase + Sialidase A +  $\beta$  (1-4)-Galactosidase + Glucosaminidase)**



**Figure 5.5 Triflic Acid Treatment of Suspected Glycoprotein(s) in an *Anabaena* PCC 7120 Cell Extract**

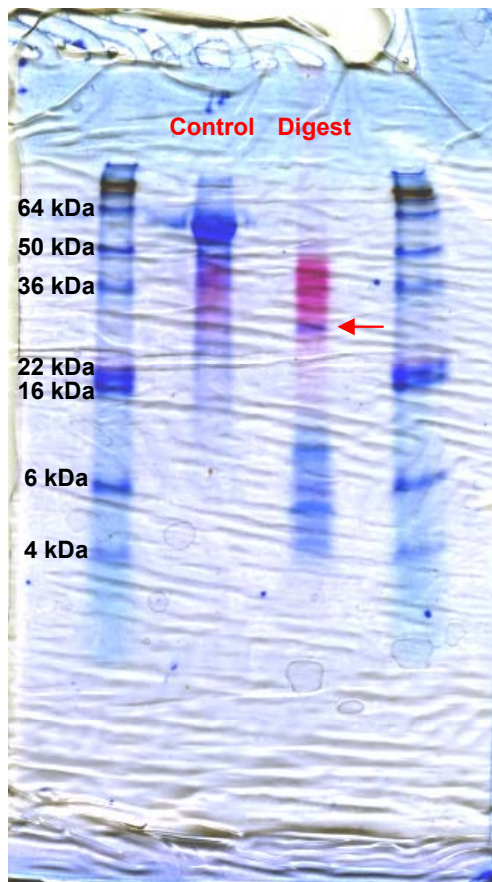
Attempted deglycosylation of suspected glycoprotein(s) purified from an *Anabaena* (*Nostoc*) PCC 7120 soluble cell extract, using anhydrous triflic acid. Longer treatment times resulted in protein degradation also, as observed in the bovine fetuin control samples. Treatment of electroeluted *Anabaena* (*Nostoc*) proteins do not result in a faster moving (deglycosylated) protein band, suggesting that there is no covalent linkage between the sample protein(s) and the PAS-staining carbohydrate material.



**Figure 5.6 Periodate/Acetic Acid Treatment of Carbohydrate Material in *Anabaena* PCC 7120 Soluble Cell Extract**

Attempted chemical degradation of carbohydrate material in an *Anabaena* (*Nostoc*) PCC 7120 soluble fraction, using periodate/acetic acid. Coomassie (*left*) and PAS-stained (*right*) SDS-PAGE gels of treated samples are shown. Higher concentrations of periodate caused slight protein degradation as well. Protein bands, analyzed by protease digestion and mass spectrometry, are indicated. The results of these analyses are listed in Table 5.2.



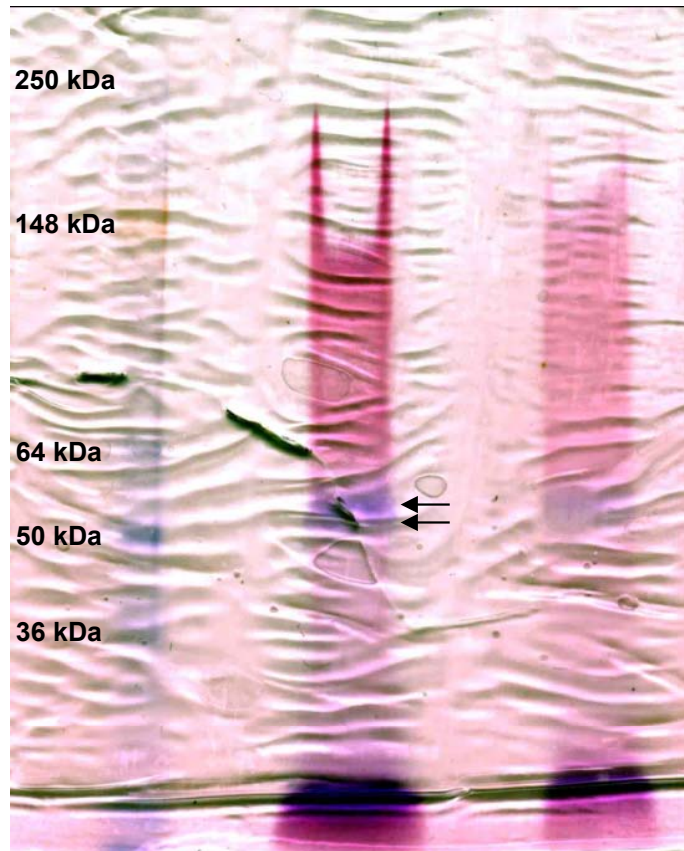


**Figure 5.7 In-Solution Trypsin Digest of Electroeluted *Anabaena* PCC 7120 Protein/Carbohydrate Material**

An in-solution trypsin digest was performed on an electroeluted protein/carbohydrate sample, and the resulting peptides were separated on a 16.5% Tris-tricine SDS-PAGE gel. Coomassie and PAS-staining of the gel indicated one peptide band (red arrow) that appeared to be slowed by the carbohydrate material. Analysis of this band revealed two peptides, both of which are encoded by the same open reading frame (*alr4550*). Neither peptide is from the protein's SLH domain.







**Figure 5.10 SDS-PAGE Analysis of BG11<sub>0</sub> Media-Grown *Anabaena* PCC 7120 Soluble Cell Extract**

SDS-PAGE separated soluble cell extract, from cells grown in nitrogen-free media. A composite sample, taken from protein bands indicated by the arrows, was analyzed by protease digestion and tandem mass spectrometry.

<u>Sample</u>	<u>ORF</u>	<u>MW<sub>H</sub> (kDa)</u>	<u>~M<sub>r</sub> (kDa)</u>	<u>Δ MW</u>	<u>Homology</u>
<b>[1]</b> PAS	alr0020	126.9	90	-36.9	phycobilisome core-membrane linker protein ApcE
	alr2269	89.64	90	0.36	chloroplastic outer envelope membrane protein homolog
	all4414	78.29	90	11.71	unknown protein
	all1237	76.74	90	13.26	hypothetical protein
	alr1742	67.91	90	22.09	DnaK-type molecular chaperone
	<b>alr4550</b>	60.58	90	29.42	<b>*hypothetical protein</b>
	all4563	38.62	90	51.38	fructose-1,6-bisphosphate aldolase
<b>[8]</b> Coomassie	alr0020	126.9	90	-36.9	phycobilisome core-membrane linker protein ApcE
	alr2269	89.64	90	0.36	chloroplastic outer envelope membrane protein homolog
	all1237	76.74	90	13.26	hypothetical protein
	alr3602	76.73	90	13.27	Putative long-chain-fatty-acid--CoA ligase
	alr1742	67.91	90	22.09	DnaK-type molecular chaperone
<b>[2]</b> PAS	alr2269	89.64	81	-8.64	chloroplastic outer envelope membrane protein homolog
	alr3344	72.07	81	8.93	transketolase
	alr1742	67.91	81	13.09	DnaK-type molecular chaperone
<b>[3]</b> PAS	alr2999	93.14	72.5	-20.64	endopeptidase Clp ATP-binding chain
	alr2887	80.84	72.5	-8.34	hypothetical protein
	alr3344	72.07	72.5	0.43	transketolase
	all4071	63.68	72.5	8.82	lysyl-tRNA synthetase
	alr1896	58.97	72.5	13.53	chaperonin GroEL
	alr3662	57.95	72.5	14.55	chaperonin GroEL
<b>[9]</b> Coomassie	alr2887	80.84	72.5	-8.34	hypothetical protein
	alr3344	72.07	72.5	0.43	transketolase
	alr1896	58.97	72.5	13.53	chaperonin GroEL
	alr3662	57.95	72.5	14.55	chaperonin GroEL
<b>[4]</b> Coomassie	<b>alr4550</b>	60.58	60.5	-0.08	<b>*hypothetical protein</b>
	alr1896	58.97	60.5	1.53	chaperonin GroEL
	alr3608	58.27	60.5	2.23	similar to endoglucanase
	alr3662	57.95	60.5	2.55	chaperonin GroEL
	alr1890	55.84	60.5	4.66	phosphoglycerate dehydrogenase
	alr0237	51.92	60.5	8.58	leucine aminopeptidase
	all3964	51.82	60.5	8.68	phosphoglucomutase/phosphomannomutase

**Table 5.1 Results of Protein Analyses by Tandem Mass Spectrometry**

*Anabaena (Nostoc)* PCC 7120 open reading frames (ORFs) matching peptides identified from the samples designated in Figure 5.1. Hypothetical molecular weights (MW<sub>H</sub>) and estimated relative molecular weights (~M<sub>r</sub>) are listed. Proteins identified as containing SLH domains are shown in bold and designated by an asterisk (\*). (Continued on next page)

<u>Sample</u>	<u>ORF</u>	<u>MWH (kDa)</u>	<u>~Mr (kDa)</u>	<u>Δ MW</u>	<u>Homology</u>
[5]	alr1381	71.53	57	-14.53	trypsin; PrcA
PAS	<b>alr4550</b>	60.58	57	-3.58	<b>*hypothetical protein</b>
	alr1896	58.97	57	-1.97	chaperonin GroEL
	<b>all4499</b>	58.57	57	-1.57	<b>*hypothetical protein</b>
	alr3662	57.95	57	-0.95	chaperonin GroEL
	all3093	54.04	57	2.96	bifunctional purine biosynthesis protein PurH
	alr1524	53.04	57	3.96	ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit
	alr2328	53	57	4	glutamate--ammonia ligase
	all4968	49.48	57	7.52	glutathione reductase
[6]	<b>alr4550</b>	60.58	53.5	-7.08	<b>*hypothetical protein</b>
PAS	alr1896	58.97	53.5	-5.47	chaperonin GroEL
	<b>all4499</b>	58.57	53.5	-5.07	<b>*hypothetical protein</b>
	<b>all7614</b>	58.35	53.5	-4.85	<b>*hypothetical protein</b>
	alr3662	57.95	53.5	-4.45	chaperonin GroEL
	<b>alr0834</b>	54.44	53.5	-0.94	<b>*porin; major outer membrane protein</b>
	alr2328	53	53.5	0.5	glutamate--ammonia ligase
	all2303	48.03	53.5	5.47	dihydroorotase
	alr0672	46.98	53.5	6.52	similar to vanadium chloroperoxidase
	all4337	44.81	53.5	8.69	translation elongation factor EF-Tu
[10]	<b>alr4550</b>	60.58	53.5	-7.08	<b>*hypothetical protein</b>
Coomassie	<b>all4499</b>	58.57	53.5	-5.07	<b>*hypothetical protein</b>
	alr5275	52.18	53.5	1.32	6-phosphogluconate dehydrogenase
	alr0672	46.98	53.5	6.52	similar to vanadium chloroperoxidase
	all4337	44.81	53.5	8.69	translation elongation factor EF-Tu
[7]	all0122	36.35	40	3.65	pyruvate dehydrogenase E1 beta subunit
PAS	alr0530	32.21	40	7.79	phycocyanin-associated rod linker protein CpcC
	alr3301	30.43	40	9.57	unknown protein
[11]	all0122	36.35	40	3.65	pyruvate dehydrogenase E1 beta subunit
Coomassie	alr0543	34.48	40	5.52	ABC transporter; substrate-binding protein
	alr0530	32.21	40	7.79	phycocyanin-associated rod linker protein CpcC
	alr0525	31.3	40	8.7	phycoerythrocyanin-associated rod linker protein PecC

**Table 5.1 Results of Protein Analyses by Tandem Mass Spectrometry (Continued)**

A continuation of Table 5.1, begun on the previous page.

<u>Sample</u>	<u>ORF</u>	<u>MW<sub>H</sub> (kDa)</u>	<u>~M<sub>r</sub> (kDa)</u>	<u>Δ MW</u>	<u>Homology</u>
[1]	alr2153	89.56	90	0.44	outer membrane heme receptor
	alr0397	94.31	90	-4.31	similar to ferric aerobactin receptor
[2]	alr2269	89.64	85	-4.64	chloroplastic outer envelope membrane protein homolog
[3]	<b>all4499</b>	58.57	57	-1.57	<b>hypothetical protein</b>
	<b>alr4550</b>	60.58	57	-3.58	<b>hypothetical protein</b>
[4]	alr2269	89.64	85	-4.64	chloroplastic outer envelope membrane protein homolog

**Table 5.2 Protein Analyses of *Anabaena* PCC 7120 Periodate Oxidized Samples**

*Anabaena (Nostoc)* PCC 7120 open reading frames (ORFs) matching peptides identified from the protein bands designated in Figure 5.6. Periodate/acetic acid degradation of the carbohydrate component of a sample, electroeluted from the 98-120 kDa region of a SDS-PAGE gel, produced 2-3 proteins that continued to migrate to the same general region of the gel and two whose migration increased substantially (~40 kDa shift in M<sub>r</sub>). SLH domain-containing proteins are shown in bold. Note that the *all4499* and *alr4550*-encoded proteins, previously electroeluted from a much higher molecular weight region of a gel, now show relative molecular weights similar to their hypothetical values.





Sample	BJ6F	BJ23F	BJ38F	BJ31F	BJ37F
<b>IX'd</b>	yes	yes	yes	no	yes
<b>SDS-PAGE</b>	yes	yes	no	yes	no
<b>Region</b>	60-64 kDa	50-120 kDa	N/A	50-120 kDa	N/A
<b>Media</b>	BG11	BG11	BG11 <sub>0</sub>	BG11 <sub>0</sub>	BG11 <sub>0</sub>
<b>Lysis</b>	alumina	alumina	freeze-thaw	alumina	alumina
<b>ManNAc</b>	27.9 %	33.8 %	16.1 %	32.2 %	-
<b>GluNAc</b>	13.8 %	10.4 %	-	9 %	-
<b>Rha</b>	28.2 %	30.4 %	15.1 %	11.9 %	-
<b>Man</b>	12.8 %	17.2 %	5.3 %	11.4 %	9.5 %
<b>Rib</b>	4.1 %	-	6 %	24.8 %	48.8 %
<b>Xyl</b>	-	0.3 %	-	6.6 %	-
<b>Gal</b>	3.2 %	5 %	49 %	-	19.3 %
<b>Glu</b>	10 %	2.8 %	8.5 %	4.1 %	22.4 %

**Table 5.4 Carbohydrate Analyses of *Anabaena* PCC 7120 Samples**

Carbohydrate analyses of several *Anabaena* (*Nostoc*) PCC 7120 samples were performed to determine the composition of the material thought to be interacting with SLH domain-containing proteins, and to see how this material contributed to the overall carbohydrate composition of the soluble fraction. Samples were purified by anion exchange chromatography, preparative-scale SDS-PAGE, or both prior to analysis. When preparative-scale SDS-PAGE was performed, the molecular weight region of the gel from which samples were electroeluted is indicated. Attempts to perform analysis on the soluble protein fraction directly, without any additional purification, were unsuccessful. Samples were subjected to acid hydrolysis and a reacetylation procedure, before APTS derivatization and separation by CE-LIF.

## CHAPTER VI

### Discussion

#### **Carbohydrate Associations of *Anabaena (Nostoc)* PCC 7120 SLH Domain-Containing Proteins**

##### 6.1 Introduction

In the previous chapter, a series of experiments was performed to investigate a glycoprotein-like PAS staining pattern, produced by an *Anabaena (Nostoc)* PCC 7120 soluble cell extract. The results of these experiments, as they pertain to the Specific Aims of this portion of the work (stated in Chapter 1.3.7), are discussed.

##### 6.2 Analysis of a Periodic Acid/Schiff-Staining Pattern (Specific Aim #1)

The PAS-staining pattern, originally presented in Figure 1.8, indicates the presence of a significant amount of carbohydrate material in the *Anabaena (Nostoc)* PCC 7120 soluble cell extract. Although there is a large “smear” of this PAS stain across the 55-120 kDa molecular weight region of the gel, several distinct bands can be seen. Such a “ladder-like pattern,” or “ladder of bands,” is sometimes indicative of glycosylated protein, which is present as several different glycoforms, each of which whose migration by SDS-PAGE is inversely proportional to its degree of glycosylation (Arranz *et al.* 1997; Schaffer *et al.* 2002; Thornton *et al.* 2002). This pattern is intentionally generated at times, by premature termination of a glycosidase reaction, in order to estimate the number of glycosylation sites present on a glycoprotein.

Detection of this type of pattern does not guarantee that a glycoprotein is present in a bacterial cell extract. Bacterial cells lack the high degree of compartmentalization that eukaryotic cells possess, and thus may present more opportunities for their protein constituents to interact non-covalently with cellular saccharides. Such interactions can be strong enough to withstand the denaturing conditions of SDS-PAGE, thereby creating the false impression that a bacterial protein is glycosylated. Alternatively, the PAS-staining

material may not be interacting with protein at all; it may be comigrating with a particular protein or proteins. Larger and more branched bacterial carbohydrate structures (*i.e.*, cell wall material, storage polysaccharide) would be expected to fragment into a size distribution, upon cell lysis, resembling that observed herein. However, carbohydrate polymers possess little capacity for binding SDS, so successful entry of any such material into a SDS-PAGE gel would require that its structure contain negatively charged functionalities.

Assuming that any carbohydrate-linked or carbohydrate-associated proteins would be detectable in multiple bands in the gel, tandem mass spectrometric protein analyses of several different areas within the staining region, designated in Figure 5.1, were performed. Common features shared by several of the proteins detected in the sampled areas (indicated in Table 5.1) were N-terminal SLH domains and sequence similarity to known cyanobacterial porins (Hansel and Tadros 1998; Hoiczky and Hansel 2000; Umeda *et al.* 1996). The significance of the SLH domain in both covalent and non-covalent protein/carbohydrate interactions has been discussed in previous chapters (Chapter 1.3.4 and Chapter 5.6). The frequent occurrences of SLH domain-containing proteins, as either glycosylated or carbohydrate-associated entities (sometimes as both), point to these proteins as the most likely candidates to be linked to or associating with the PAS-staining material, if an interaction is indeed occurring (Brechtel and Bahl 1999; Chauvaux *et al.* 1999; Hoiczky and Hansel 2000; Lemaire *et al.* 1995; Schaffer *et al.* 2002; Smarda *et al.* 2002). The distribution of these proteins over the PAS-stained region of the gel, and at the extreme ends, supports this idea.

Although SLH domain-containing proteins were not found in *every* sample band along the length of the stain, there were two proteins (encoded by *all4499* & *alr4550*) that appeared in adjacent bands at the lower molecular weight end of the PAS stain. Because of the low sensitivity of the PAS stain and the extremely high sensitivity of the ion-trap mass spectrometer, it is likely that a region of a gel in which a glycoprotein is present in sufficient quantities to result in a positive PAS stain response would possess sufficient protein for identification. However, the possibility that sufficient peptides would not be generated, due to the sequestering of protease cleavage sites upon glycosylation, does exist. Additionally, glycosylated peptides are known to be much

more difficult to ionize in a mass spectrometer than their unmodified counterparts (Mechref and Novotny 2002). A highly glycosylated protein in a complex sample could easily be missed in the presence of non-glycosylated protein, and this possibility is even more likely in the higher molecular weight regions of the gel, where more extensively glycosylated forms of the protein would be found.

The presence of the same protein in several adjacent bands (*i.e.*, the presence of the *all4499*-encoded protein in both sample #5 and #6 in Figure 5.1) could be due to a sample tailing phenomenon, rather than the retarding effect of the carbohydrate on SDS-binding and protein migration. This phenomenon might occur during the electrophoresis, as a consequence of the high protein loading used. However, the protein encoded by the *alr4550* open reading frame was detected not only in three adjacent sample bands (samples #4, #5, & #6, in the 55-65 kDa  $M_r$  region of the gel), but also in a sample from the ~100 kDa  $M_r$  region of the gel (sample #1). Such a drastic difference in apparent molecular weight indicated that something was reducing the mobility of this protein. The high  $\beta$ -sheet contents (~40%) and hydrophobicity, typical of bacterial porins, are known to cause retention of secondary structure during SDS-PAGE, resulting in a reduced mobility, but not on the scale of that being observed for the *alr4550*-encoded protein (Engelhardt and Peters 1998; Hansel *et al.* 1994; Jeanteur *et al.* 1991; Troll *et al.* 1992; Wylie *et al.* 1993).

The simplest and most obvious explanation for the observed differences in protein migration was that the carbohydrate material was influencing the electrophoretic behavior of at least some of the SLH domain-containing proteins. To test this hypothesis, and to determine whether this carbohydrate was either covalently linked to one or more of these proteins or merely interacting with them in a non-covalent manner, a series of deglycosylation and carbohydrate degradation procedures was attempted on a soluble cell fraction that was enriched in the PAS-staining material (and thus any associating proteins). The results of these procedures are described in Chapter 5.5. The inability of the enzymatic (Figure 5.4) deglycosylation procedure to produce faster moving protein bands, which did not stain with the PAS stain, was the first indication that the observed protein/carbohydrate association might be a non-covalent interaction.

An experiment in which the material from which the PAS-stain elicits a positive response (presumably carbohydrate; hereinafter referred to as the PAS-PM, or PAS positive-staining material) was oxidized by sodium periodate produced results that also supported the above-mentioned idea of a non-covalent interaction (Figure 5.6). In samples oxidized by 50 mM and 100 mM concentrations of periodate, the PAS-PM appeared to have been degraded to a degree that it could no longer participate in the interactions necessary for protein association. In these two samples, the protein and carbohydrate sample components migrated to non-overlapping regions of the gel after treatment. The untreated protein “smear,” which can be seen in the control sample lane, has collapsed to individual protein bands. A lower concentration of oxidizing agent (10 mM) proved insufficient to produce the same effect. In the 50 mM periodate sample, less carbohydrate and incidental protein degradation has taken place, and one resulting sample band in particular shows an increase in migration. Tandem mass spectrometric protein analysis of this band revealed peptides matching to both the *alr4550* and *all4499* open reading frames of *Anabaena (Nostoc) PCC 7120*. Sodium periodate oxidizes the vicinal diol groups of compounds, and it is unable to cleave a glycoprotein carbohydrate-protein linkage, without damaging the remainder of the carbohydrate structure, which could potentially produce results similar to those shown in Figure 5.6. So, although these results did not rule out the possibility that a glycoprotein was present, it did indicate that the bulk of the carbohydrate material in these samples was not covalently attached to protein.

An additional clue that a non-covalent carbohydrate/protein association was occurring was found when electroeluted samples were subjected to a second electrophoresis (Figure 5.5 and Figure 5.6). A tendency of these samples to spread out over a region greater than that from which they were electroeluted was observed. This phenomenon is likely an effect of the additional competition between the negatively-charged SDS-bound protein and the SDS-unassociated carbohydrate material that sample material experienced when subjected to further electrophoresis. Although the carbohydrate-staining and the protein-staining materials in these samples did not separate completely during this procedure, the densest regions of protein and the densest regions

of carbohydrate (judged by stain intensity) no longer overlapped. Such an observation would not be made if the PAS-PM were covalently attached to protein.

It is proposed here that the PAS-PM is a large polymeric structure, rather than many smaller moieties that are bound to the protein (or proteins) of interest. Support of this hypothesis can be seen in the PAS-stained gel, shown in Figure 5.6. In two of the oxidized samples, the PAS-PM is significantly, but not completely, decomposed. It apparently can no longer interact with the protein components of the sample, but it is not so completely degraded that it cannot be observed by gel electrophoresis. If the PAS-PM were a multiplicity of several smaller saccharides, the PAS-PM oxidation product would not be visible by SDS-PAGE, except in bands or regions that also contain protein.

It is curious that this PAS-staining material migrates by SDS-PAGE, even after separation from proteins with which it allegedly interacts. This behavior suggests that it is negatively-charged, or that it has a hydrophobic component with an affinity for SDS. Such behavior would be expected of peptidoglycan material. Anion exchange experiments indicate that this material does not carry a substantial negative charge, and this idea is also supported by the results of carbohydrate analyses.

The studies described herein have shown that the large quantity of PAS-PM in an *Anabaena (Nostoc)* soluble cell extract is not covalently attached to protein, and that this material is present as a polydisperse, high molecular weight structure. The results of these experiments indicate that two or more SLH domain-containing proteins are involved in non-covalent interactions with the PAS-PM, and that these associations are not disrupted by the denaturing conditions of SDS-PAGE. The proteins interacting with the PAS-PM are discussed further in Section 6.3.

As previously stated, these results cannot rule out the possibility that a glycoprotein was present in the described *Anabaena (Nostoc)* soluble cell fractions, and that it migrated to somewhere within the PAS-staining regions of the SDS-PAGE gels, shown in Figure 1.8 and Figure 5.1. Additionally, the possibility that SLH domain-containing proteins detected in this extract, which are thought to be interacting with the PAS-PM, were also glycosylated cannot be dismissed with the information gathered in these experiments alone.

### 6.3 Identification of Carbohydrate-Binding Proteins (Specific Aim #2)

The results of the experiment in which samples were treated with sodium periodate indicated that there was at least one protein, and perhaps as many as three proteins, interacting with the PAS-staining material. Analysis of band #3 in Figure 5.6 revealed two proteins (encoded by *alr4550* and *all4499*) whose migration had increased substantially after this procedure, presumably because of the degradation of the PAS-staining material. A tendency of homologs of these two proteins to dimerize has been noted (Hansel *et al.* 1998), but the various regions of the gel in which these two proteins were detected (and the relative molecular weights to which they correspond) indicated that dimerization was not occurring. If such an association were taking place, it would be possible, albeit unlikely, that it would abruptly cease upon periodate treatment.

Surprisingly, although the protein encoded by *alr4550* had been detected in the region of the gel from which material used in the periodate experiment was electroeluted, the protein encoded by *all4499* had not (Table 5.1). The cell lysate used in the original protein identification experiment (Figure 5.1) and the periodate oxidation experiment (Figure 5.6) was of the same preparation. Hence, detection of the *all4499*-encoded protein in the latter experiment indicated that: 1) it was present in the 98-120 kDa  $M_r$  region all along, but undetectable because of the other proteins also present (material used in the original protein identification experiment was not purified by ion exchange chromatography); 2) it was previously present in the 98-120 kDa  $M_r$  region, but not in the specific portion of the gel sampled for protein analysis; or 3) the in-solution interactions of the SLH domain-containing proteins with the PAS-PM are transient, and at some point between the first and second analyses, a quantity of *all4499*-encoded protein, sufficient to be detected by tandem mass spectrometry, began associating with larger PAS-PM fragments. If the first alternative cited above is the case, it indicates that the *alr4550*-encoded protein was expressed at higher levels than the *all4499*-encoded one, an idea supported by the results of the experiment described in Chapter 5.8. If the third alternative is true, further studies could potentially reveal additional proteins that interact with the PAS-PM.

An in-solution trypsin digest of material that was electroeluted from a lower molecular weight PAS-staining region (~55-64 kDa) of the gel was performed, in an

attempt to identify the carbohydrate-binding domains of the SLH domain-containing proteins. This experiment produced some unexpected results. Following Tris-Tricine SDS-PAGE, and then both Coomassie G-250 and PAS staining of the same gel, a single peptide band (containing a potentially missed/protected cleavage site) was found to be comigrating with the PAS-PM. Analysis of this band by an additional (in-gel) trypsin digestion and mass spectrometry uncovered two peptides, both encoded by *alr4550*, which did not correspond to the SLH domain, as was anticipated. A search of the NCBI NonRedundant and SwissProt protein databases, using the sequence of these peptides with several different BLAST algorithms, returns only other cyanobacterial hypothetical proteins and putative porins as matches.

A ClustalX alignment (Figure 5.9) of a portion of *alr4550*'s hypothetical protein sequence with several characterized bacterial porins (Jeanteur *et al.* 1991) was generated. The *alr4550*-encoded GNVLGIFAGAQPYAR (amino acid residue #504-518) and GVQAGANEVPYQVEGFYK (amino acid residue #519-536) peptides, discovered in the in-solution digestion experiment, described in the previous paragraph, span a sequence region between, and overlapping part of, two probable  $\beta$ -sheet stretches. This region of the protein has weak homology to the OprB carbohydrate-selective porin of *Pseudomonas aeruginosa*, indicated by the PFAM (E=0.43, amino acid residue #240-575) and InterPro (E=2.2e-7, amino acid residue #529-575) databases. Unfortunately, no determination of whether this portion of the *alr4550*-encoded protein contains an OprB-like carbohydrate binding site can be made with the presently published information alone (Hancock and Brinkman 2002; Wylie *et al.* 1993; Wylie and Worobec 1995).

The discovery of only *alr4550*-encoded peptides comigrating with the PAS-PM does not preclude the possibility that the *all4499*-encoded protein, or even several of the other SLH domain-containing proteins, also interact with this material. Experimental evidence has indicated that the *alr4550*-encoded protein was more abundant in cell extracts than the other SLH domain-containing proteins, suggesting that a digest of a greater quantity of material could have revealed additional peptides from other proteins whose electrophoretic migration was also influenced by the PAS-PM.

Although the increase in migration of bands #2 and #4 (containing only the *alr2269*-encoded protein) in Figure 5.6 was not of the same magnitude as that observed



for band #3, there did appear to be a  $M_r$  decrease of 10-15 kDa for this protein upon periodate-treatment. Whether this change was due to a disruption of an interaction between this protein and the PAS-PM, or perhaps another protein, is unclear. Few chloroplast proteins have been characterized (Rolland *et al.* 2003), and examples of known chloroplast glycoproteins are limited (Burkey and Mathis 1998; Maione and Jagendorf 1984). Thus, it is not known whether the *alr2269*-encoded protein should be expected to participate in covalent or non-covalent protein/carbohydrate interactions, based upon studies of its homologs. The presence of the *alr2269*-encoded protein in samples #1 and #2 in Figure 5.1 suggests that this protein may have only originally been present in the area of the gel from which it was electroeluted as a consequence of sample overloading. However, the possibility that the *alr2269*-encoded protein is also interacting with the PAS-PM exists.

To summarize, the interaction of the *alr4550*-encoded protein with carbohydrate material from an *Anabaena (Nostoc)* soluble cell extract has been demonstrated. The specific portion of the protein that interacts with this material (the PAS-PM) is formed by amino acid residues #504 through #536. Experimental evidence has indicated that an *all4499*-encoded protein was also interacting with this material, but a protease digestion experiment was unable to confirm this result. The possibility that other proteins are participating in similar, non-covalent interactions with the PAS-PM, including a protein that does not contain a SLH domain (encoded by *alr2269*), was also suggested.

#### 6.4 Carbohydrate-Binding Protein Function Analysis (Specific Aim #3)

The protein profiles of the outer membranes of several different cyanobacteria are dominated by two major polypeptides, each having an approximate molecular weight of 50 kDa, as judged by SDS-PAGE. In the case of *Anacystis nidulans* R-2 (*Synechococcus* PCC 7942), it has been suggested that one of these proteins is glycoprotein (Jürgens *et al.* 1989; Jürgens and Weckesser 1985b; Resch and Gibson 1983; Umeda *et al.* 1996; Woitzik *et al.* 1988). These proteins are often presumed to be porins, but only rarely have attempts been made to experimentally determine their specific functions (or molecular selectivity) (R Benz and Bohme 1985; Sauer *et al.* 2001). These putative cyanobacterial porins are greater in size than other gram-negative bacterial porins, which

is explained by the presence of an additional ~120 amino acid, N-terminal sequence region, identifiable as a SLH domain (or SLH domain and associated coiled-coil region). These proteins are thought to also function in attaching the outer membrane to the peptidoglycan layer (Hansel *et al.* 1998; Hoiczky and Hansel 2000).

BLAST analyses and ClustalX alignments of the *alr4550* hypothetical protein sequence show a high level of sequence similarity to other putative cyanobacterial porins (Figure 6.1). The sequences of these proteins all seem to follow the same (signal peptide)-(SLH domain)-(β-barrel region) arrangement. Consideration was given to the idea that the sp. PCC 7120 proteins might serve as autotransporters (proteins that are known to export portions of their sequence from the cell, often to act as virulence factors). Some autotransporter proteins are known to be carbohydrate-associated, however, autotransporters are generally greater in size than the *alr4550*-encoded protein and its homologs (I Benz and Schmidt 2001; Henderson *et al.* 1998; Laarmann and Schmidt 2003; Loveless and Saier 1997). If the SLH domain-containing, sp. PCC 7120 proteins are indeed autotransporters, they export little more than an SLH domain outside of the cell (Figure 6.2).

One of the putative cyanobacterial porins to which the sp. PCC 7120 proteins show homology, a *Synechocystis* PCC 6803 protein (*sll0772*), is encoded by an open reading frame that is located next to the gene for a known non-phosphorylating carbohydrate transporter (*sll0771*, *glcP*) (Zhang *et al.* 1989). These two ORFs are located on the same strand, and they are near another membrane transport protein-encoding ORF, which is oriented in the opposite direction (*slr0774*). Considering the sequence similarity of the *alr4550*-encoded protein to the OprB carbohydrate-selective porin, it is tempting to speculate that the *sll0772*-encoded protein also plays a role in carbohydrate transport, and that the *alr4550*-encoded protein and its homologs comprise a family of carbohydrate-selective porins (Wylie *et al.* 1993; Wylie and Worobec 1995). However, further experimentation would be necessary to test this hypothesis. Several experiments, performed with a strain of *Synechocystis* PCC 6803 in which the *sll0772* gene was deleted, were unsuccessful in assigning function to the protein that it encodes (Zhang *et al.* 1989).

The possibility that the chloroplastic outer membrane protein homolog, encoded by *alr2269*, is also associated with PAS-PM was suggested by experimental data (Figure 5.6). It stands to reason that a hydrophobic membrane protein might require an associating hydrophilic moiety (*e.g.*, carbohydrate structure) to partition into a soluble cell fraction, but strong evidence of this was not found. Very few regions of sequence similarity exist among the *alr2269*-encoded protein and the SLH domain-containing porins (Figure 6.3). If there is an association that is taking place between the *alr2269*-encoded protein and the PAS-staining material, it is probably dissimilar to the interaction in which the SLH domain-containing proteins are involved. Functional assignment of the protein encoded by the *alr2269* ORF was made on the basis of its high degree of sequence similarity to the *Arabidopsis thaliana* chloroplast import-associated channel protein (direct submission to NCBI by The Institute for Genomic Research, Rockville MD) and the *Pisum sativum* (garden pea) chloroplastic outer envelope membrane protein OEP75 precursor (Tranel *et al.* 1995). The *alr2269*-encoded protein's function is easier to speculate upon than that of the putative cyanobacterial porins, however, the *alr2269*-encoded protein's interaction with the PAS-PM is still in question.

#### 6.5 Carbohydrate Analysis of PAS-Staining Material (Specific Aim #4)

Although the PAS-PM could potentially be the carbohydrate component of protein-bound lipopolysaccharide, which still associates with the carbohydrate-binding domains of exocellular proteins upon cell lysis, this possibility seems unlikely. The strong sequence similarity between the SLH domain-containing proteins and other bacterial porins indicates that a large portion of their structure are embedded in the outer membrane. That sufficient LPS could bind to a protein to cause a ~50 kDa increase in  $M_r$  is also unlikely. In view of this, it was anticipated that analyses of the PAS-PM carbohydrate composition would indicate that this material had originated in the peptidoglycan layer (Hansel *et al.* 1998).

Known cyanobacterial peptidoglycan layers are of the A1 $\gamma$  type (Jürgens *et al.* 1983; Jürgens *et al.* 1989; Jürgens and Weckesser 1985b; Schleifer and Kandler 1972; Woitzik *et al.* 1988). These cell wall structures possess some of the characteristics of gram-positive peptidoglycan layers, including a greater thickness (~10 nm average) and a

higher degree of cross-linking in comparison to other gram-negative bacterial peptidoglycans. These cyanobacterial cell wall structures frequently contain a phosphodiester-linked, covalently bound polysaccharide (or secondary cell wall polymer, SCWP) (Jürgens *et al.* 1989; Jürgens and Weckesser 1986; Woitzik *et al.* 1988). In some bacteria, this hydrofluoric acid-extractable SCWP is the component that interacts with SLH domain-containing proteins, instead of the *N*-acetyl glucosamine/*N*-acetyl muramic acid primary carbohydrate polymer (Brechtel and Bahl 1999; Chauvaux *et al.* 1999; Sára *et al.* 1998a; Sára *et al.* 1998b).

Cyanobacterial peptidoglycans consist primarily of an *N*-acetyl glucosamine/*N*-acetyl muramic acid polymer, which has a 40-60% degree of cross-linking (Jürgens *et al.* 1983; Woitzik *et al.* 1988). Lesser quantities of muramic acid-6-phosphate are sometimes found in these structures, serving as the site of attachment for any covalently bound secondary polysaccharides present (Jürgens *et al.* 1989; Jürgens and Weckesser 1986; Woitzik *et al.* 1988). Although carbohydrate standards for *N*-acetyl muramic acid were unavailable at the time of these analyses, no indications of any negatively charged sugars, such as *N*-acetyl muramic acid or muramic acid-6-phosphate, were observed. It is possible that either (or both) of these carbohydrates would be indistinguishable from the monosaccharides listed in Table 5.4, with the mode of analysis employed, but there are clearly more than two or three components to all of the five samples. Thus, even if peptidoglycan material is present in these samples, there is an additional carbohydrate component as well.

An assumption that is made in many carbohydrate analysis methods is that all amino sugars were present in their acetylated forms prior to chemical treatment (*i.e.*, strong acid hydrolysis, hydrazinolysis). They are frequently re-acetylated during the analysis, partially because many separation techniques are incompatible with a molecule that contains an unsubstituted amino group. Unfortunately, this assumption does not always hold true. But, because an additional assay to quantify the extent of acetylation is not always performed, a detected sugar is sometimes referred to only as glucosamine, thus giving no indication of whether *N*-acetyl glucosamine or glucosamine was actually present in the sample (Araki *et al.* 1972; Jürgens and Weckesser 1985a). Bearing this caveat in mind, of the monosaccharides detected here, rhamnose, mannose, glucose,

galactose, glucosamine, and mannosamine have all been previously detected in LPS preparations from cyanobacteria (Jürgens *et al.* 1989; Jürgens and Weckesser 1985b; Weckesser *et al.* 1974; Woitzik *et al.* 1988). Mannose, glucose, galactose, glucosamine, galactosamine, and mannosamine have all been previously detected in preparations of cyanobacterial secondary cell wall polymers, examples of which have been estimated to possess average molecular weights between 25 kDa and 33 kDa (Jürgens and Burger-Wiersma 1989; Jürgens *et al.* 1983; Jürgens *et al.* 1989; Jürgens and Weckesser 1986; Woitzik *et al.* 1988).

Size considerations make it unlikely that the PAS-staining material is membrane LPS, although the detection of rhamnose favors this idea. The CyanoBase web site lists several putative rhamnose-related ORFs in *Anabaena (Nostoc)*, under the category of “surface polysaccharides, lipopolysaccharides, and antigens.” However, no ORFs that are associated with SCWP biosynthesis are identified, leaving open the possibility that the SCWP also contains rhamnose. The relatively low amounts of glucose in the two samples that were purified most extensively (BJ6F and BJ23F) indicate that storage polysaccharide is not the most abundant carbohydrate compound in these samples, if it is indeed in them. *Anabaena (Nostoc)* PCC 7120 produces an extracellular polysaccharide layer around its heterocystous cells, which comprise only a small portion of those analyzed, particularly for cultures grown in BG11 media (Bryant 1994). It is, thus, also unlikely that the PAS-staining material is heterocystous EPS.

The origin of the PAS-PM cannot be determined using the information collected in this work alone, but some inferences about this material can be made. The probable location of the SLH domain-containing proteins *in vivo* suggests that the PAS-PM originates in the cell wall. Carbohydrate composition analyses of samples from both small and large regions of a SDS-PAGE gel yield similar results, indicating that the PAS-staining material is a size distribution of fragments from a larger structure. Considering the number of carbohydrate constituents uncovered, this material is clearly neither the *N*-acetyl glucosamine/*N*-acetyl mannosamine peptidoglycan biopolymer alone, nor is it solely storage polysaccharide. The monosaccharides detected in these analyses are typical of those found in cyanobacterial peptidoglycan-associated secondary cell wall polymers (SCWPs) and LPS, but the apparent size of the PAS-PM makes it unlikely that

it is LPS. The propensity of this material to migrate by SDS-PAGE, even when freed from interacting proteins (Figure 5.6), implies that the PAS-PM possesses a SDS-binding component itself. These results collectively suggest that the PAS-PM is composed primarily of secondary cell wall polymer (peptidoglycan material), perhaps with a lesser degree of the primary polymer and/or peptide material still attached.

## 6.6 Conclusions

Although the possibility that several different proteins are interacting with the PAS-PM has been suggested by the experimental evidence generated in this study, only the interactions of the proteins encoded by the *alr4550* and *all4499* open reading frames are supported by direct evidence. These two proteins are members of a family of putative cyanobacterial porins whose specificity has been in question for years (Hansel *et al.* 1998; Hansel *et al.* 1994; Hansel and Tadros 1998). The carbohydrate/protein interactions of non-SLH domain-containing bacterial porins have been noted, but not studied in detail (Lugtenberg and Van Alphen 1983). Evidence for carbohydrate interaction with the  $\beta$ -barrel region of a putative cyanobacterial porin has been presented here. It is possible that these proteins are involved in carbohydrate transport, and that their observed association with carbohydrate (the PAS-PM) is the result of an interaction between an OprB-like binding domain and random carbohydrate fragments produced upon cell lysis. However, the results of carbohydrate analyses (discussed in Section 6.5) suggest that a non-specific interaction is not likely the case. Further work is needed to gain insight into the specific function of these proteins *in vivo*.

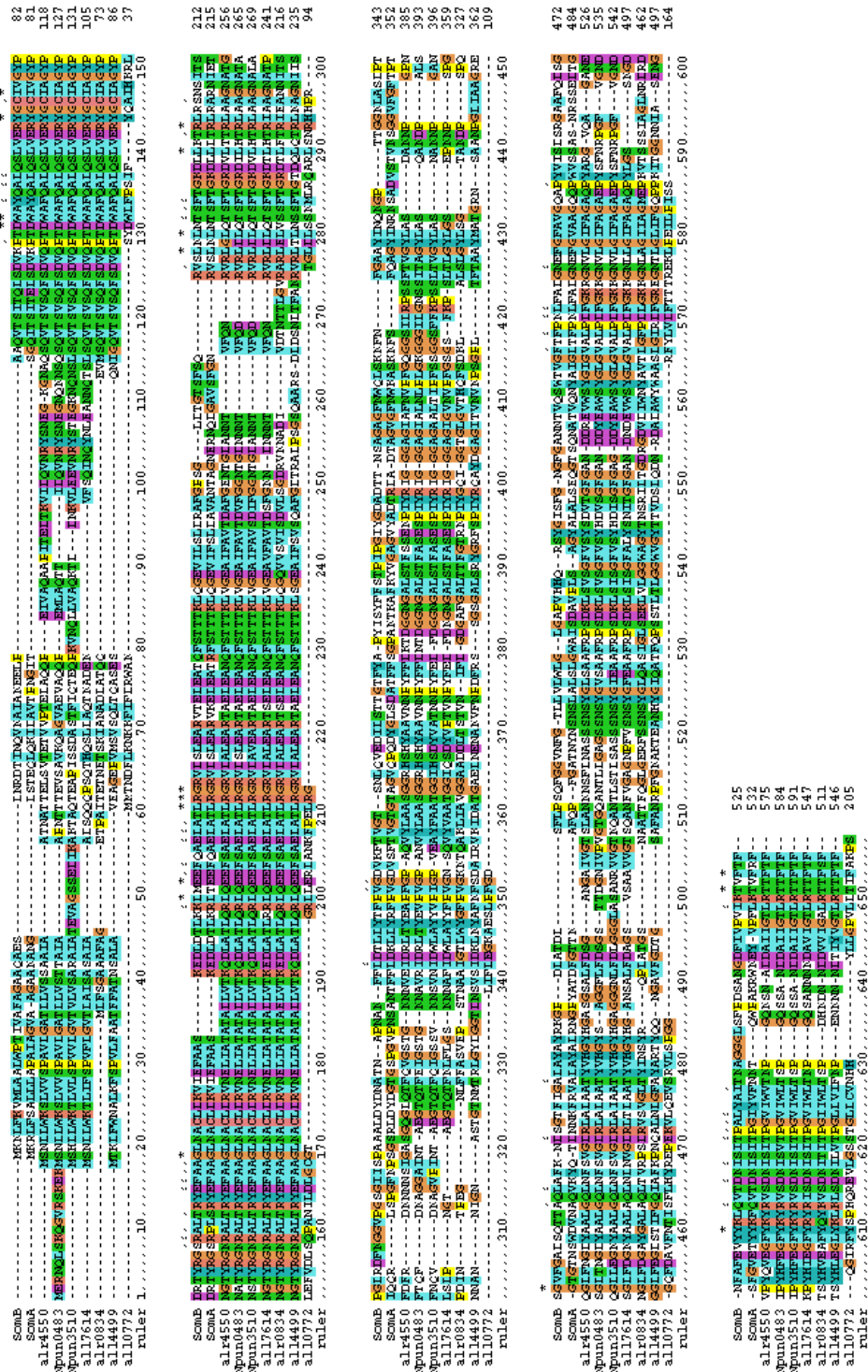
## 6.7 Future Directions

Additional experiments, using either cloned or purified cyanobacterial SLH domain-containing proteins and various cell wall (or LPS) preparations, could be performed to further demonstrate the protein/carbohydrate interactions of these proteins, and to examine their specificity for the carbohydrates with which they interact. However, of greater concern to researchers studying these putative porins is the question of pore selectivity. This issue has not been an easy one to address (Sauer *et al.* 2001; Zhang *et al.* 1989). Some porins facilitate the diffusion of many different types of molecules into

the cell, discriminating only by molecular weight (Lugtenberg and Van Alphen 1983). Molecular modeling studies could reveal a particular size limitation (or charge preference) imposed by the structure of these proteins, and perhaps suggest a particular class of compounds or molecules selected by them.

Analysis of the promoter sequences for the genes encoding the SLH domain-containing proteins might reveal certain specific conditions under which these putative porins are expressed or up-regulated, and so indicate a common functional theme among them. If these proteins respond to a particular stimulus or serve a specific function, it might be demonstrated by genetic knockout and complementation experiments, as was attempted with the *sl10771* glucose transporter gene (Zhang *et al.* 1989).

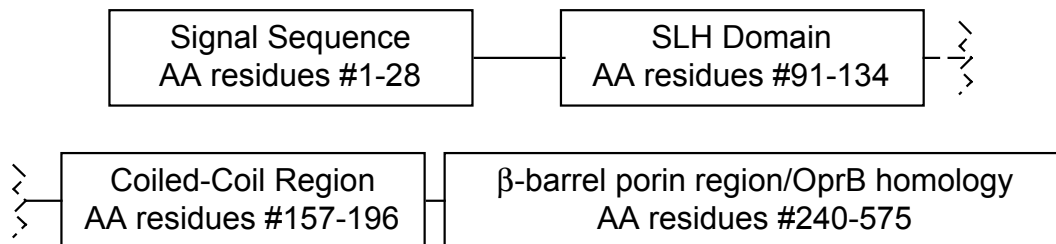
The interaction between the PAS-PM and the *alr4550*-encoded protein could potentially provide insight into this particular protein's function, if investigated further. Although carbohydrate/protein interactions involving SLH domains have been demonstrated for several different organisms, very little such work has been performed with the putative cyanobacterial porins, and no reports of carbohydrate/protein interactions with a specific portion of the presumed  $\beta$ -barrel region of these proteins exist (Brechtel and Bahl 1999; Chauvaux *et al.* 1999; Hansel *et al.* 1994; Lemaire *et al.* 1995; Sára *et al.* 1998b). Additional studies of this interaction and the sequence region involved could also provide additional evidence of its physical location and/or orientation *in vivo*.



**Figure 6.1 SLH Domain-Containing Putative Cyanobacterial Porins**

Hypothetical amino acid sequences of cyanobacterial ORFs that possess sequence similarity to bacterial porins. These proteins contain a SLH domain (Surface Layer protein Homology domain, and associated coiled-coil region), which accounts for their additional size in comparison to other gram-negative porin proteins. It is thought that the SLH domain (amino acid residues #91-196 in *atr4550*) is responsible for connecting the peptidoglycan layer (or its embedded secondary cell wall polymer) to the outer membrane.





(AA residues #29-90)

ATNATTELSVTETVVPTELAQQPEIVAQAAPITEDTKVIDQVNRYSNE  
GKGNAQSQVTSVSQ

(AA residues #197-239)

ATNATTELSVTETVVPTELAQQPEIVAQAAPITEDTKVIDQVNRYSNE  
GKGNAQSQVTSVSQ

**Figure 6.2 Schematic Representation of the *alr4550*-Encoded Protein**

Representation of the *alr4550*-encoded protein and the amino acid sequences of its functionally unassigned regions. If this protein were a gram-negative autotransporter protein, the region between amino acids #29 and #239 (or #197) would form an extracellular protein, perhaps a virulence factor. BLAST searches could reveal no specific function for this protein, based upon the two amino acid sequences shown below the illustration.



## CHAPTER VII

### Instrumentation, Materials, Procedures, and Electronic Resources

#### 7.1 Instrumentation

##### 7.1.1 Analytical Instrumentation

NMR was performed using a Varian Inova 400 MHz spectrometer (Palo Alto, CA) or a JEOL Eclipse-plus 500 MHz spectrometer (Peabody, MA). Capillary electrophoresis was performed using a Beckman-Coulter P/ACE MDQ capillary electrophoresis unit with a laser-induced fluorescence (LIF) detection system that utilized a Beckman P/ACE 488 nm argon-ion laser module (Fullerton, CA). Protein identification by liquid chromatography-tandem mass spectrometry was performed using a ThermoFinnigan LCQ DecaXP mass spectrometer (San Jose, CA) and a LC Packings Ultimate HPLC pump (Sunnyvale, CA). MALDI-TOF was performed using a Kompact SEQ MALDI-TOF MS instrument equipped with a nitrogen UV laser (337 nm) from Kratos Analytical (Chestnut Ridge, NY). Analyses of synthesized disaccharide acceptors were performed by direct injection into a ThermoFinnigan TSQ Quantum triple quadrupole mass spectrometer (San Jose, CA).

##### 7.1.2 Instrumentation for Protein Isolation and Chemistry

Fifteen liter cell cultures were grown in a Bellco Biotechnology bioreactor, which was stirred with a 1 amp Heavy Duty Overhead Drive motor (Vineland, NJ). Cell mass was collected from *Nostoc commune* DRH1 and *Anabaena (Nostoc)* PCC 7120 15 L fermentor-grown cultures using a Beckman J6HC High Capacity Refrigerated Floor Centrifuge, with 1 L bottles (Fullerton, CA). Cells were ruptured using a 35 mL capacity Thermo Spectronic French pressure cell (Madison, WI), with a Power Laboratory Press from American Instruments (Silver Spring, MD). Protein samples were subject to low (5000 x g) and medium (30,000 x g) speed centrifugation in a Sorvall RC-5B Refrigerated Superspeed Centrifuge, with a SS-34 or GSA rotor (Newtown, CT).

Ultracentrifugation was performed with a Beckman Class H Ultracentrifuge and a 50-Ti rotor (Fullerton, CA).

A Fisher Scientific Marathon 12kBr centrifuge (Pittsburgh, PA) and a Jouan model CR412 centrifuge (Winchester, VA) were used with 15 mL centrifugal concentrators. FPLC was performed using an Amersham-Pharmacia P-500 pump and an LCC-501 Plus controller (Uppsala, Sweden). Protein electroelution was carried out in a Gibco BRL Horizontal Gel Electrophoresis Apparatus (model 58) from Life Technologies (Gaithersburg, MD). Protein assays were performed using a Milton Roy Spectronic 601 (Ivyland, PA) or a Hitachi UV-2000 spectrophotometer (Tokyo, Japan). Samples contained in 0.6 mL, 1.5 mL, and 2.0 mL Eppendorf tubes were centrifuged using an Eppendorf model 5415C centrifuge (Westbury, NY).

A Fisher Scientific Water Bath was used to incubate enzyme reactions at 37 °C, and a Fisher Isotemp Drybath Model 145 (Pittsburgh, PA), set to 105 °C, was used to boil protein samples with loading buffer before SDS-PAGE. An E-C Apparatus Corporation Model EC4000P Series 90 programmable power supply (Holbrook, NY) was used for SDS-PAGE. A BioRad PowerPac 200 power supply (Hercules, CA) and an Electrophoretic Blotting Unit Model EBU-1000 from American BioNuclear (Emeryville, CA) were used for blotting of protein onto PVDF membranes, before analyses with the DIG-Glycan Differentiation Kit. Isoelectric Focusing was performed using an Amersham-Pharmacia IPGPhor IEF unit (Uppsala, Sweden). Gels and blots were agitated during staining and developing procedures using a Warner-Chilcott shaker (Rockaway, NJ). Gels and membranes were scanned using a flatbed UMAX Astra 1220S Scanner (Fremont, CA).

### 7.1.3 General Laboratory Instrumentation

A Ludlum Measurements (Sweetwater, TX) Survey Meter (Geiger counter) was used for detection of the <sup>32</sup>P radioisotope in laboratory work areas. Electronic autoradiography was performed using a Packard Instant Imager (Meriden, CT). Liquid scintillation counting was performed using a Beckman Model LS 6500 Multipurpose liquid scintillation counter (Fullerton, CA).

An Accumet Basic AB15 pH Meter from Fisher Scientific (Pittsburgh, PA) was used for measuring the pH of buffers and solutions. A Barnstead E-Pure system (Dubuque, IA) was used to generate 18 M $\Omega$  water for use in buffers, solutions, and enzyme assays. Mettler AE50 and PE3600 electronic balances (Anaheim, CA) were used to weigh chemicals and laboratory materials. A Labconco Lyph-Lock 12 Freeze Dry System (Cambridge, MA) was used to lyophilize samples.

A Thermolyne Type 17600 Dri-Bath (Dubuque, IA) was used to heat beta-elimination and oligosaccharide hydrolysis reactions. A Jouan RC 10.22 Speedvac Centrifuge with Titan Vapor Trap (Winchester, VA) and a LH Leybold Trivac pump (Export, PA) was used to concentrate samples to dryness following both of these procedures. APTS derivatizations of mono- and oligosaccharides were carried out in a PE Applied Biosystems GeneAmp PCR System 9700 with dual 384-well (20  $\mu$ L) sample block module (Foster City, CA).

An I<sup>2</sup>R (Instruments for Research and Industry) Thermo-Watch L6-1000SS controller (Cheltenham, PA) was used to regulate reaction temperature during synthesis of tagged acceptors. A Gilson FC 203B fraction collector (Middleton, WI) was used to collect samples during flash chromatography. A system consisting of a Brinkmann Instruments rotary evaporator (Westbury, NY), with Büchi model 461 heated water bath (Flawil, Switzerland), Neslab RTE 221 recirculating chiller (Waltham, MA), and KNF Neuberger model UN7263 FTP vacuum pump (Trenton, NJ) was used to concentrate samples during the syntheses of tagged acceptors and the purification of glycosyltransferase reaction products. Synthetic samples that were to be dried under high vacuum were placed on a Kontes vacuum manifold with Cold Finger Condenser (Vineland, NJ), which was evacuated by a Fisher Scientific Maxima high-vacuum pump (Pittsburgh, PA). Aluminum-backed TLC plates were dried and charred using a MHT 1200 W Industrial Model 500 heat gun (New Berlin, WI). TLC plates with spots/bands that were visible under ultraviolet light were examined in a Fisher Biotech Ultraviolet Viewing Cabinet, with 254 nm and 365 nm 15 W lamps (Pittsburgh, PA); this 254 nm lamp was also used to activate 5'-azido affinity tags during procedures to radiolabel the glucuronosyltransferase enzyme.

## 7.2 Materials

### 7.2.1 General Materials

Unless otherwise stated, all chemicals were obtained from Sigma-Aldrich (St. Louis, MO). ScintiSafe Plus 50% scintillation mixture was obtained from Fisher Scientific (Pittsburgh, PA). Quartz cuvettes with 10 mm light paths were purchased from VWR International (West Chester, PA). Immobilon P Transfer Membranes were obtained from Millipore (Billerica, MA).

### 7.2.2 Materials for Synthesis of Tagged Acceptors

Chromatographic silica gel (200-425 mesh) was obtained from Fisher Scientific (Pittsburgh, PA). Whatman preparative-scale, 20 cm x 20 cm C18 (200  $\mu\text{m}$  thickness, 60  $\text{\AA}$  pore size) and silica gel (1000  $\mu\text{m}$  thickness, 60  $\text{\AA}$  pore size) TLC plates, with fluorescent indicator backing, were obtained from Fisher Scientific (Pittsburgh, PA). Whatman aluminum-backed, 20 cm x 20 cm silica gel (250  $\mu\text{m}$  thickness, 60  $\text{\AA}$  particle size) TLC plates, with fluorescent indicator backing, were obtained from Fisher Scientific (Pittsburgh, PA). Merck aluminum-backed RP-18F<sub>254S</sub> C18 TLC plates (5 cm x 7.5 cm) were obtained from Fisher Scientific (Pittsburgh, PA). The 6-(Fluorescein-5-Carboxamido) Hexanoic Acid Succinimidyl Ester (FCHASE) fluorescent labeling compound was obtained from Molecular Probes (Eugene, OR). All deuterated NMR solvents were purchased from Cambridge Isotope Laboratories (Andover, MA).

### 7.2.3 Materials for Protein Chemistry

RQ1 RNase-free DNase and porcine trypsin were obtained from Promega (Madison, WI). RNaseA from bovine pancreas was obtained from Sigma-Aldrich (St. Louis, MO). Triton X-100 and egg white lysozyme were obtained from Fisher Scientific (Pittsburgh, PA). Protein assays of samples that contained detergents, chaotropes, or high concentrations of salt were performed using a membrane protein micro-Bradford method (Zuo and Lundahl 2000), or using the 2-D Quant Kit from Amersham Biosciences (Piscataway, NJ). Polystyrene disposable cuvettes with 10 mm light paths, for use with the Bradford reagent, were purchased from Sigma-Aldrich (St. Louis, MO).

Q-Sepharose strong anion exchange resin was obtained from Amersham Biosciences (Piscataway, NJ). 5'-azido-UDP-glucose and <sup>32</sup>P-labeled 5'-azido-UDP-glucuronic acid were purchased from Affinity Labeling Technologies (Lexington, KY). Slide-A-Lyzer Dialysis Cassettes were obtained from Pierce (Rockford, IL). Centriprep and Centriplus centrifugal concentrators were obtained from Millipore (Billerica, MA). ProteoPLUS electroelution tubes and non-fixing Electro-Blue staining solution were obtained from Qbiogene (Carlsbad, CA). Eppendorf Safe-Lock Tubes, for use with radioisotopes, carbohydrate hydrolysis and beta-elimination reactions, and sample heating prior to SDS-PAGE, were obtained from VWR International (West Chester, PA). Sherlock microcentrifuge tube caps were obtained from USA Scientific (Ocala, FL).

Immobiline DryStrip Gels (isoelectric focusing strips) were obtained from Amersham Biosciences (Piscataway, NJ). Criterion IEF, SDS-PAGE (Tris-glycine), and SDS-PAGE (Tris-Tricine) pre-cast gels, and the Criterion gel running box, were obtained from BioRad Laboratories (Hercules, CA). ProtoGel, a stabilized solution of 30% acrylamide/bisacrylamide (37.5:1 ratio), was obtained from National Diagnostics (Atlanta, GA). SeeBlue Plus2 pre-stained molecular weight standards, for SDS-PAGE, were obtained from Invitrogen (Carlsbad, CA). Coomassie Brilliant Blue G-250 (for peptides) and R-250 (for proteins) were purchased from BioRad Laboratories (Hercules, CA). The Silver Stain Plus kit was purchased from BioRad Laboratories (Hercules, CA).

#### 7.2.4 Materials for Carbohydrate Analysis

Accustain Schiff's reagent was obtained from Sigma-Aldrich (St. Louis, MO). The GlycoPro Enzymatic Deglycosylation Kit (*N*-linked and core I *O*-linked glycosidases) and pr*O*-Link Extender Enzymatic Deglycosylation kit (core II *O*-linked glycosidases) were obtained from Prozyme (San Leandro, CA). The DIG-Glycan Differentiation Kit was obtained from Roche (Mannheim, Germany).

APTS (8-aminopyrene-1,3,6-trisulfonate) was obtained from Beckman-Coulter (Fullerton, CA). *N*-acetylneuraminic acid from *Escherichia coli*, for use as a control and a standard in the sialic acid analysis assay, was obtained from Fluka (Buchs, Switzerland). Fused-silica eCAP polyacrylamide-coated capillaries with pre-burned windows were obtained from Beckman-Coulter (Fullerton, CA).

### 7.3 General Analytical Procedures

#### 7.3.1 Solid Phase Extraction (SPE) of FCHASE Derivatives

Following purification by preparative-scale thin-layer chromatography, bands containing the synthesized FCHASE-derivatives were scraped and extracted with water (10 mL, 5x). This solution was applied to a C18 solid-phase extraction cartridge, previously rinsed with 2 mL methanol and equilibrated with 5 mL water. The bound material was rinsed with water (10 mL) to remove any salts, and the product compound was subsequently eluted with 1:1 acetonitrile-water, until no more yellow color rinsed from the column (~6 mL).

Preceding purification by preparative-scale TLC and analysis by MALDI-TOF, the 900  $\mu$ L glucuronosyltransferase reaction mixture was centrifuged and the resulting supernatant was loaded onto a C18 solid-phase extraction cartridge, previously rinsed with 2 mL methanol and equilibrated with 5 mL water. The bound material was rinsed with water (20 mL) to remove any salts, and the product compound was eluted with 1:1 acetonitrile-water, until no more yellow color rinsed from the column (~6 mL).

#### 7.3.2 Matrix Assisted Laser Desorption/Ionization – Time of Flight (MALDI-TOF) Mass Spectrometry

The reaction product from the 900  $\mu$ L glucuronosyltransferase reaction, described in Chapter 2.6, was freeze-dried and redissolved in 50  $\mu$ L of a 1:1 mixture of acetonitrile-water. Various volumes of either this solution, or a saturated solution of the fluorescent-labeled galactose acceptor in water, were combined with a saturated solution of either 2,5-dihydroxybenzoic acid in 9:1 acetonitrile-water, or  $\alpha$ -cyano-4-hydroxycinnamic acid in 1:1 acetonitrile-water + 0.1% trifluoroacetic acid. One-microliter spots were applied to the MALDI sample plate, and then dried under a stream of nitrogen. The laser was scanned across the sample to find the regions of greatest ion intensity, in both positive and negative ion mode. A total of 50 shots were averaged to give the final sample spectrum.

Peptide solutions from the in-gel trypsin digestion procedure (Section 7.5.8) were combined with an equal volume of  $\alpha$ -cyano-4-hydroxycinnamic acid in 1:1 acetonitrile-water + 0.1% trifluoroacetic acid. One-microliter spots of this solution were applied to



the MALDI sample plate, and then dried under a stream of nitrogen. The laser was scanned across the sample to find the regions of greatest ion intensity, in positive ion mode. A total of 50 shots were averaged to give the final sample spectrum. Spectra were internally calibrated using the 842.5 Da and 2211.1 Da autolysis products of the trypsin protease.

### 7.3.3 Glycosyltransferase Activity Reactions

Glycosyltransferase activity assay reactions were typically prepared by combining 5  $\mu$ L of a saturated solution of the fluorescent-labeled carbohydrate acceptor, 5  $\mu$ L of either an activated donor solution or water (control), 10  $\mu$ L of either a suspension of membrane fragments in buffer or a detergent extract of such fragments, and 1  $\mu$ L of 200 mM DTT (21  $\mu$ L final volume). Reactions to assay for glucuronosyltransferase activity, specifically, were prepared with a 10 mM UDP-glucuronic acid (donor) solution, and included an additional 1  $\mu$ L of 1 M  $MgCl_2$  (22  $\mu$ L final volume). These reactions were carried out in 0.6 mL Eppendorf tubes and allowed to proceed overnight, at room temperature.

### 7.3.4 Capillary Electrophoresis with Laser-Induced Fluorescence Detection (CE-LIF)

Capillary electrophoresis of glycosyltransferase activity reactions were performed on a Beckman-Coulter P/ACE MDQ unit, equipped with a laser-induced fluorescence detector using a 3 mW argon-ion laser as the excitation source. Software was provided by the manufacturer. Separations were performed in a 75  $\mu$ m x 57 cm fused silica capillary, using a 5 s, 0.5 psi pressure injection. Electrophoresis was performed at 17.5 kV, for 20-30 min, using a 25 mM sodium borate running buffer adjusted to pH 9.4, sometimes with the inclusion of either 20 mM SDS, or 20 mM  $\beta$ -cyclodextrin. The capillary was sequentially washed with 0.1 M sodium hydroxide, water, and borate running buffer, each for two minutes at 20 psi, preceding each run, and then with water for two minutes at 20 psi, after the separation was complete.

Separations of APTS-derivatized carbohydrates were performed on the same instrument, in a 19  $\mu$ m x 37 cm fused silica capillary, using a 20 s, 0.5 psi pressure injection. Electrophoresis was performed at 25 kV, for 20 min, using either a 240 mM

sodium borate running buffer (pH 9.0), or a 240 mM MOPS running buffer (pH 7.0) for the separation of glucose and ribose derivatives. The capillary was washed with running buffer at 20 psi for 5 min prior to, and for 2 min following, separation. Derivatization procedures are described in Section 7.6.7.

## **7.4 Procedures for Fluorescent-Labeled Acceptor Synthesis**

### 7.4.1 Thin-Layer Chromatography (TLC)

Small (approximately 2 cm x 6 cm) rectangular sections from a 20 cm x 20 cm aluminum-backed TLC plate were cut and used for analytical TLC procedures. Sample spots were applied to a pencil-marked origin line, drawn horizontal to and approximately 1 cm from the 2 cm wide base of the section. Plate sections were developed in a 60 mL glass screw-top jar and dried, and spots were then visualized under ultraviolet light or by charring with a solution of naphthoresorcinol (1,3-dihydroxynaphthalene) in sulfuric acid/ethanol.

Purification of a compound by preparative-scale thin-layer chromatography was conducted by application of the sample across the width of a pencil-drawn origin line, drawn approximately 3 cm from the base of the 20 cm x 20 cm plate. Plates were developed in glass developing chambers and partially dried. Separated bands were visualized under ultraviolet light, in a viewing hood, and marked with a pencil. The silica gel, containing the compound of interest, was scraped from the glass TLC plate, using an Exacto knife, subsequent to extraction.

### 7.4.2 Flash (Silica Gel) Chromatography

Approximately 37 g of silica gel was suspended in as little mobile phase as possible, and rapidly poured into a 15 mm i.d., 450 mm (79 mL) Ace glass column with attached slurry packer/solvent reservoir. Columns were packed at 8-10 psi, using a nitrogen tank that fed a four-way valve at the top of the assembly, and then operated at 2-4 psi. Samples were loaded by manual addition to the top of the packing. Fractions were collected, at a flow rate of approximately 2 mL/min, by a Gilson fraction collector into 10 mL glass test tubes. These (general) operating conditions were adjusted when necessary to facilitate separation of the compound of interest.

### 7.4.3 Nuclear Magnetic Resonance (NMR)

NMR spectra were recorded on either a Varian Inova 400 MHz spectrometer or a JEOL Eclipse-plus 500 MHz spectrometer, at 25 °C, using the software supplied by the manufacturer. Samples that contained numerous hydroxyl groups were subjected to a H/D atom exchange by dissolution in D<sub>2</sub>O (99.9%) and freeze-drying, prior to their analyses. Samples analyzed in D<sub>2</sub>O (99.996%) were referenced to acetone (<sup>1</sup>H shift, 2.225 ppm; <sup>13</sup>C shift, 31.07 ppm) added at the time of, or in a subsequent, analysis. Samples that were analyzed in CD<sub>3</sub>OD were referenced to the central solvent peak (<sup>1</sup>H shift, 3.30 ppm; <sup>13</sup>C shift, 49.0 ppm). Samples that were analyzed in CD<sub>3</sub>COCD<sub>3</sub> were also referenced to the central solvent peak (<sup>1</sup>H shift, 2.04 ppm; <sup>13</sup>C shift, 29.8 ppm).

## **7.5 Procedures for Protein Purification and Analysis**

### 7.5.1 Growth of Organisms

*Nostoc commune* DRH1, a clonal axenic isolate of material collected in the Hunan province of China in 1981 (*Nostoc commune* CHEN), was originally described by Hill (Hill *et al.* 1994). Fifteen-liter *Nostoc commune* DRH1 cultures were grown with aeration for five to six weeks, at 25 °C, in BG11<sub>o</sub> media (Rippka *et al.* 1979). Cultures were grown under an incident photon flux density of approximately 300 μmol/(m<sup>2</sup>s), after which time cells were harvested by centrifugation.

*Anabaena* (*Nostoc*) PCC 7120 was obtained from Dr. Jeff Elhai, Department of Biology, VA Commonwealth University. Fifteen-liter cultures of *Anabaena* (*Nostoc*) PCC 7120 were grown with aeration for either one week in BG11 media, or for four weeks in BG11<sub>o</sub> media, at 25 °C (Rippka *et al.* 1979). Cultures were grown under an incident photon flux density of approximately 300 μmol/(m<sup>2</sup>s), after which time cells were harvested by centrifugation.

### 7.5.2 Cell Lysis

*Nostoc commune* DRH1 cells, from the 15 L fermentor, harvested by centrifugation at 5000 x g for 10 min, were washed three times with 10 mM Tris (pH 7.8), and resuspended in as little extraction buffer (50mM Tris, 10 mM MgCl<sub>2</sub>, 20 mM KCl, 1 mM PMSF, 50 μg/mL DNase, and 50 μg/mL RNaseA at a pH of 7.8) as was

necessary to produce a free-flowing solution (approximately 1:1 volumetric ratio with wet cell mass). Cells were ruptured by two passes through a pre-cooled French pressure cell, at 20,000 psi. The cell lysate was stored on ice between passes. The cell lysate was successively centrifuged at 5000 x g for 20 min, and then at 30,000 x g for 20 min, to remove whole cells and intact cell envelope material, respectively. The resulting supernatant was subsequently ultracentrifuged at 150,000 x g for 1 hour, at room temperature, to sediment insoluble proteins and membrane fragments. The supernatant from the ultracentrifugation procedure (soluble cell fraction) was withdrawn and concentrated, before immediately being assayed for glycosyltransferase activity. Insoluble material was washed with extraction buffer, containing 1 M NaCl, and collected again by ultracentrifugation. Salt-washed insoluble material was homogenized in various buffers, using a spatula, for glycosyltransferase activity assays. Similarly prepared insoluble material could be extracted with 50 mM HEPES (pH 7.0), containing 0.4% Triton X-100, overnight at 4 °C, and then ultracentrifuged to produce a membrane detergent extract for subsequent glucuronosyltransferase purification procedures.

Alternatively, harvested and washed *Nostoc commune* DRH1 cells were resuspended in extraction buffer and ground with gradually increasing amounts of sterile alumina. The extent of cell lysis could be judged by centrifuging a small portion of this mixture. A blue supernatant was produced when cells had ruptured. The mixture was centrifuged at 5000 x g for 20 min, to remove whole cells and alumina, and then treated as the French pressure lysate described above. A 200X protease inhibitor cocktail concentrate (Appendix) was added to both alumina-ground and French pressure cell-lysed cell fractions, before subsequent storage at 4 °C.

*Anabaena (Nostoc)* PCC 7120 cells were subjected to the procedures described above, to achieve cell lysis, with the following exceptions. Cells harvested by centrifugation were not washed before lysis, and it was unnecessary to concentrate the soluble extract. The protein concentration of this fraction, upon cell lysis, was already approximately 20 mg/mL. Furthermore, the insoluble cell fraction, produced by the 150,000 x g centrifugation, was not washed with salt before extraction with 50 mM HEPES (pH 7.0), containing 0.4% Triton X-100. Material that was not solubilized by the Triton X-100 extraction was boiled with 1X SDS-PAGE loading buffer to produce an

additional detergent extract. The remaining insoluble material was suspended in 1X SDS-PAGE loading buffer for analysis.

### 7.5.3 Protein Assays

Soluble protein concentration was determined by the method of Bradford (Bradford 1976), using a Bradford Reagent available from Sigma-Aldrich. Sample protein concentration was estimated by comparison to a standard curve generated with 0-140  $\mu\text{g}$  of bovine serum albumin. Protein and standard solutions were combined with water (if necessary), to a final volume of 50  $\mu\text{L}$ , with which 1.5 mL of the Bradford Reagent was mixed, and the assay mixture was allowed to incubate for 5 minutes. The absorbances of these solutions at 595 nm were measured relative to a control sample, containing 50  $\mu\text{L}$  of water. The volumes used in this procedure were doubled, when necessary, to compensate for lower protein concentrations (*i.e.*, when 50  $\mu\text{L}$  of sample contained less than 10  $\mu\text{g}$  of protein).

Protein concentrations of samples that contained high concentrations of salt or detergents (*i.e.*, a membrane detergent extract) were measured using the micro-Bradford membrane protein assay of Zuo and Lundahl (Zuo and Lundahl 2000), or with the 2-D Quant Kit from Amersham Biosciences. The 2-D Quant Kit was used according to the instructions of the manufacturer. The micro-Bradford membrane protein assay was executed as suggested by the authors, except that 400  $\mu\text{L}$  of Bradford reagent was used to dissolve the protein-calcium phosphate precipitate and 600  $\mu\text{L}$  of water was added before sample absorbance was measured. In both methods, sample protein concentration was estimated by comparison to a bovine serum albumin standard.

### 7.5.4 SDS-PAGE

Tris-glycine polyacrylamide gels for SDS-PAGE separation of proteins were prepared as described by Laemmli (Laemmli 1970) using the ProtoGel acrylamide/bisacrylamide stock solution or were purchased from BioRad. Protein samples were prepared for electrophoresis by mixing with an equal volume of 2X SDS-PAGE loading buffer and heating at 100  $^{\circ}\text{C}$  for at least 5 minutes. Samples were loaded into the sample wells of the polymerized gel, and were separated, alongside pre-stained

molecular weight markers, at either a constant current of 40 mA/gel, or a constant voltage of 200 V (BioRad Criterion gels only). Gels were typically stained immediately after the run, or covered with Saran wrap for radioimaging. A printout of the radioimage on cellulose acetate film, overlaid upon a gel, allowed radiolabeled bands to be identified and excised, without staining. Detection of proteins by Coomassie staining or with the Electro-Blue stain is described below (Section 7.5.6), after which gels were typically covered with Saran wrap, scanned, and stored at room temperature.

Tris-Tricine polyacrylamide gels for SDS-PAGE separation of peptides were either prepared as described by Schägger and von Jagow (Schägger and von Jagow 1987), or purchased from BioRad Laboratories. The 16.5% (w/v) acrylamide/6 M urea gel was run at 30V for 1 hour, and then at a constant current of 40 mA for 1 hour, to separate trypsin-digested peptides. The 16.5% and 10-20% gradient Tris-Tricine SDS-PAGE gels obtained from BioRad were run at a constant voltage of 100 V, for approximately 100 minutes. In both cases, normal SDS-PAGE loading buffer was used to dilute samples, with the addition of a small amount of Coomassie G-250 as a tracking dye. Both Coomassie dyes were removed from the gel in a 50% methanol-10% glacial acetic acid fixing step, performed after electrophoresis and before Coomassie G-250, PAS, and/or silver staining procedures.

#### 7.5.5 Two-Dimensional Electrophoresis

Proteins were typically precipitated with either cold acetone, or cold TCA, preceding electrophoresis, to remove any salts and/or charged detergents, and to concentrate the sample. Protein pellets were then solubilized in two-dimensional electrophoresis lysis buffer (Appendix), and diluted to a concentration appropriate for the procedure with two-dimensional electrophoresis rehydration buffer (Appendix). Both of these solutions are based on formulations, described by Rabilloud, to take advantage of the chaotropic power of thiourea (Rabilloud *et al.* 1997). Additionally, tributyl phosphine (TBP) was substituted for DTT in these procedures because it is uncharged (Herbert *et al.* 1998).

Sample volumes employed for isoelectric focusing were dependent on IPG (immobilized pH gradient) strip length; the volumes utilized were those recommended by

the manufacturer. Carrier ampholytes (0.5% v/v IPG buffer) and reducing agent (2 mM tributyl phosphine) were added to the diluted protein sample, which was pipetted into the center of a ceramic gel strip holder. The IPG strip was positioned, gel side down, in the sample, and both were covered with IPG cover fluid (silicone oil). The covered ceramic strip holder containing strip, sample, and cover fluid, was placed into the IPGphor unit, to rehydrate at room temperature for 10-16 hours, after which time the focusing program recommended by the manufacturer was executed. Strips were stored at -20 °C if the second dimension was not immediately run after isoelectric focusing.

Focused strips were equilibrated twice for 15 minutes in SDS-equilibration buffer, initially with the addition of 8 mM TBP, and then with the addition of 135 mM iodoacetamide, before the second dimension was run. Reduced and alkylated strips were then positioned horizontally, on top of a SDS-PAGE stacking and resolving gel, and sealed in place with 0.5% (w/v) agarose. A pre-stained molecular weight standard-soaked piece of filter paper was sometimes also sealed in the agarose. Alternatively, BioRad Criterion gels have pre-formed wells for molecular weight markers, which were used when this type of gel was employed. The remainder of the second dimension was carried out as described in Section 7.5.4.

#### 7.5.6 Detection of Protein/Peptides in Polyacrylamide Gels

Following SDS-PAGE or two-dimensional electrophoresis, gels and/or IEF strips containing whole proteins were stained, overnight, in a filtered solution of 0.1% (w/v) Coomassie R-250 in 50% methanol and 10% glacial acetic acid. After being briefly rinsed with water, gels were destained in a mixture of 25% methanol and 5% glacial acetic acid, until no background staining was observed. Gels were shaken in plastic containers containing Kimwipe paper towels to adsorb released stain. Gels were typically wrapped in Saran wrap, scanned, and stored at room temperature, after destaining was complete. PVDF membrane sections, stained for whole proteins, were subjected to the same procedure except that destaining was performed in a mixture of 50% methanol and 10% glacial acetic acid (40% water).

Following electrophoresis, Tris-Tricine SDS-PAGE gels, containing peptides, were stained in a filtered solution of 0.005% (w/v) Coomassie G-250 in 10% acetic acid,

for 1-2 hours, and then destained in a solution of 10% acetic acid. Gels were shaken in plastic containers that contained Kimwipe paper towels to adsorb released staining reagent. Gels were not immersed in stain/destain any longer than necessary to prevent diffusion of lower molecular weight peptides from the gel.

Prior to electroelution, proteins in preparative-scale SDS-PAGE gels were sometimes stained with the Electro-Blue non-fixing staining solution (Qbiogene). Silver staining of proteins in SDS-PAGE and two-dimensional gels was performed using the Silver Stain Plus kit (BioRad). In both cases, staining was performed according to the manufacturer's instructions.

#### 7.5.7 Protein Electroelution

Qbiogene ProteoPLUS electroelution tubes were filled with approximately 800  $\mu$ L 1X SDS-PAGE running buffer and several 0.5 cm x 1 cm gel slices, cut with a clean razor blade. Four tubes were placed in the tube holder, which was positioned in a 1X SDS-PAGE running buffer-filled Gibco horizontal gel electrophoresis unit, and oriented so that dialysis membranes were in-line with electrodes. Electroelution was carried out at 50 V for 2-3 hours, after which time the tubes were vortexed to remove protein stuck to the membranes. Samples were then pipetted directly into a Centriprep unit for concentration. A 600  $\mu$ L water wash of each tube and the spent gel slices contained within was also added to the concentrator before the gel slices were removed and the tubes were reused. The running buffer in the electrophoresis unit was changed whenever buffer discoloration was observed.

#### 7.5.8 In-Gel Trypsin Digestion

Protein spots, to be analyzed by either peptide mass fingerprinting or by liquid chromatography-tandem mass spectrometry, were excised and destained using several wash cycles of 25 mM ammonium bicarbonate (pH 8.0) in 50% (v/v) acetonitrile. Gel-immobilized proteins were then reduced with DTT (10 mM in 25 mM ammonium bicarbonate) and alkylated using iodoacetamide (55 mM in 25 mM ammonium bicarbonate), at room temperature. Gel pieces were rinsed several times with 100  $\mu$ L 25 mM ammonium bicarbonate, and then dehydrated with 100  $\mu$ L acetonitrile. A solution of



Promega MS grade Trypsin Gold (20 ng/ $\mu$ l) was added to the dehydrated gel pieces and allowed to absorb into the gel, for 15 minutes on ice. Excess trypsin solution was withdrawn and replaced with 25  $\mu$ L 25 mM ammonium bicarbonate. Gel pieces were subsequently incubated for 4-12 hours at 37 °C in a heating block. Proteolysis was terminated by the addition of 5  $\mu$ L of 5% TFA (or acetic acid). Samples were vortexed for 15 minutes, and then stored at -20 °C until analysis.

#### 7.5.9 Liquid Chromatography – Mass Spectrometry (LC-MS<sup>n</sup>)

Capillary columns were packed (in-house) using 75  $\mu$ m i.d. fused silica tubing and Jupiter C18 packing material (Phenomenex, 10  $\mu$ m, 300 Å), to a total length of approximately 10 cm, using a pressure cell (500 psi He). The column tip was gravity-pulled using a micro-torch so that no frit was required. A flow rate of 50  $\mu$ L/min from an LC Packings Ultimate HPLC pump was split to 250 nL/min using a low dead volume, stainless steel tee. The gradient program for peptide separation ran from 5% acetonitrile to 80% acetonitrile (diluted with 0.5% (v/v) acetic acid), in 15 minutes, after a 5 minute isocratic wash, following sample loading. Samples were loaded onto the column using a pressure bomb, and typically had a volume of 1  $\mu$ L (3-5% of the total digest).

Mass spectra were acquired on a ThermoFinnigan LCQ DecaXP quadrupole ion trap mass spectrometer, using a retrofitted nanospray source (ThermoFinnigan). An electrospray voltage of 2.3 kV was applied at the waste line of the stainless steel tee. Capillary operating temperature was 150 °C, and capillary voltage was 10 V. Three microscans were acquired for each spectrum recorded (scan time ~1.5 seconds). The threshold for tandem mass spectrum acquisition was set at 5E6, with previously fragmented ions dynamically excluded for 2 minutes. Peptides were identified using the turboSEQUENT algorithm, searching the protein database originally acquired from the Oak Ridge National Laboratory. All identified peptides were manually confirmed by comparison of the potential peptide fragments (obtained *in silico*, using an in-house copy of the MS-Product function of Protein Prospector) with the experimentally obtained MS<sup>2</sup> spectra. The FASTA and FASTS programs (Mackey, Haystead and Pearson 2002; Pearson and Lipman 1988) were also used to insure that identified peptides were not

present in other hypothetical proteins. All programs used can be found at the VA Tech Institute for Genomics web site (proteomics.biotech.vt.edu).

#### 7.5.10 In-Solution Trypsin Digestion

Electroeluted samples, containing approximately 200 µg of protein, were dissolved in a solution of 100 mM ammonium carbonate (pH 7.8), containing 20 µg of Promega modified porcine trypsin. The mixture was incubated in a water bath at 37 °C for 24 hours, and then combined with an equal volume of 2X SDS-PAGE loading buffer for electrophoresis.

#### 7.5.11 Enzymatic Deglycosylation

Attempts to enzymatically deglycosylate protein were made using the GlycoPro Enzymatic Deglycosylation Kit (*N*-linked and core I *O*-linked glycosidases) and prO-Link Extender Enzymatic Deglycosylation kit (core II glycosidases) from Prozyme. Enzymatic reactions were carried out, according to the denaturing protocol of the manufacturer's instructions, and allowed to proceed overnight. Upon completion, reactions were combined with an equal volume of 2X SDS-PAGE loading buffer for electrophoresis. Bovine fetuin was used as a control glycoprotein.

#### 7.5.12 Chemical Deglycosylation

Attempts to chemically deglycosylate protein, using anhydrous triflic (trifluoromethanesulfonic) acid, were made with a modified version of a procedure described by Sojar and Bahl (Sojar and Bahl 1987). Lyophilized protein samples were dissolved in pre-cooled triflic acid (30 µL) and incubated for 30-60 minutes, under nitrogen and on ice. Upon completion, the reaction mixture was cooled to -20 °C in an ethanol-dry ice bath, and a pre-cooled solution of 60% pyridine in water was gradually added to terminate the reaction. Reaction mixtures were diluted with 300 µL 1 M Tris (pH 6.8) and centrifuged to dryness on a 3 mL YM-10 Centriplus centrifugal concentrator. Samples were removed from the concentrator membrane using 20 µL 1X SDS-PAGE loading buffer.

Periodate-oxidation (and subsequent reduction of the resulting free aldehydes) was also used to degrade carbohydrate material. A range of electroeluted proteins (98-120 kDa in size, 20 µg total protein per reaction) that produced a positive PAS stain was incubated with 0 mM, 10 mM, 50 mM, and 100 mM sodium meta-periodate, at 4 °C for 23 hours. Upon reaction completion, excess periodate was neutralized by the addition of glycerol, and the mixture was subsequently incubated for 15 minutes at room temperature. Free aldehydes were reduced by the addition of sodium borohydride, to a final concentration of 0.1 M. Samples were incubated with sodium borohydride, for 2 hours at 4 °C, before being diluted with 10% (v/v) glacial acetic acid (to quench excess borohydride), and then centrifuged to dryness on a 3 mL YM-10 Centriplus centrifugal concentrator. Concentrator membranes were vortexed with 100 µL 1 M Tris (pH 6.8) and again centrifuged to dryness. Samples were removed from the concentrator membrane using 40 µL 1X SDS-PAGE loading buffer prior to electrophoresis.

## **7.6 Procedures for Carbohydrate Analysis**

### ***7.6.1 Detection of Carbohydrates in Polyacrylamide Gels***

The periodic acid-Schiff reagent (PAS) staining procedure, used for the detection of the 1,2-diol groups of carbohydrates, was a modified version of a procedure presented by Dubray and Bezard (Dubray and Bezard 1982). Following electrophoresis, gels to be PAS stained were shaken in a 50% methanol-10% acetic acid solution at room temperature, overnight, and then in 7.5% glacial acetic acid at room temperature, for 30 minutes. Following these washes, the gel was placed in a 1% (w/v) periodic acid solution, at 4 °C for 1 hour, with a small stir bar in the staining tray for liquid agitation. Next, the gel was placed into a separate staining tray, containing Schiff's reagent at 4 °C, for a 1-hour incubation, also with gentle agitation. Destaining was performed at room temperature, using several changes of 7.5% acetic acid. Staining trays were covered with foil to exclude light during the oxidation and staining steps.

The carbohydrate silver staining procedure utilized was also that of Dubray and Bezard (Dubray and Bezard 1982). Staining was performed as described by the authors, with the gel being immersed in ammoniacal silver solution for 15 minutes.

### 7.6.2 Blotting of Protein onto PVDF Membranes

Following SDS-PAGE, proteins were blotted onto an Immobilon P PVDF membrane, for detection of potential glycoproteins using the DIG-Glycan Differentiation Kit. A wet transfer electroblotting system was utilized. In preparation for transfer, the membrane was immersed in methanol for 30 s, in water for 2 min, and in blotting buffer for at least 5 min. The polyacrylamide gel was equilibrated for 15-20 min in blotting buffer. The gel and the membrane were arranged in contact, sandwiched between 8 pieces of filter paper (4 on each side) that was prewetted in blotting buffer; air bubbles in the stack were rolled out using a glass stirring rod. The gel-membrane-filter paper stack was inserted in the wet transfer apparatus, in the appropriate orientation to allow migration of the negatively charged SDS-bound proteins out of the gel and onto the PVDF membrane. The unit was operated at 30 V (0.3 A current limit), overnight, at 4 °C. After transfer, gels were stained with either Coomassie, or PAS stain, to judge the efficiency of the process. The PVDF membrane was rinsed briefly in methanol and allowed to dry, overnight, for use with the DIG-Glycan Differentiation Kit.

### 7.6.3 Glycoprotein Detection on PVDF Membranes

The DIG Glycan Differentiation Kit (Roche) was used, according to the manufacturer's instructions. Upon separation of samples and glycosylated controls (included with the kit) by SDS-PAGE, protein was blotted onto a 9 cm x 12 cm Immobilon P PVDF membrane. Each control and sample lane pair was cut into a separate strip for probing with one of five different digoxigenin-labeled lectin probes, performed in 50 mL Falcon tubes. Following detection, strips were wrapped in Saran wrap and scanned while still wet.

### 7.6.4 Beta-Elimination Release of Glycan / Base Hydrolysis of Protein

Release of glycan and/or selective hydrolysis of protein, in carbohydrate/protein containing samples, was accomplished using an ammonia-based  $\beta$ -elimination procedure (Huang, Mechref and Novotny 2001). The procedure was performed as instructed by the authors, in a heating block set to 60 °C and in Eppendorf Safe-Lock tubes, held closed by Sherlock microcentrifuge tube caps. Solvents were evaporated during the procedure

using a Jouan Speedvac Centrifuge. The resulting carbohydrate solution was divided into smaller portions, dried, and hydrolyzed for monosaccharide composition analyses.

#### 7.6.5 Oligosaccharide Hydrolysis

Dried oligosaccharide samples, from the ammonia-based  $\beta$ -elimination procedure, were hydrolyzed with 200  $\mu$ L 2 M TFA (trifluoroacetic acid), for 5 hours at 100 °C, for monosaccharide composition analysis. Separate aliquots were hydrolyzed with 200  $\mu$ L 0.1 M TFA for 1 hour at 80 °C for sialic acid analysis. Samples were dried after hydrolysis reactions using a Speedvac Centrifuge. Samples subjected to strong acid hydrolysis (2 M TFA) were subsequently reacylated in 20  $\mu$ L 25 mM sodium bicarbonate buffer (pH 9.5), with the addition of 8  $\mu$ L of acetic anhydride, for 1 hour at room temperature, and then dried prior to APTS derivatization.

#### 7.6.6 Enzymatic Conversion of *N*-acetylneuraminic Acid to *N*-acetylmannosamine

Sialic acid content was analyzed by enzymatic conversion to *N*-acetylmannosamine, using the method of Chen, Dobashi, and Evangelista (Chen, Dobashi and Evangelista 1998). Twenty-five units *Escherichia coli* *N*-acetylneuraminic acid aldolase was dissolved in 50  $\mu$ L 50 mM phosphate buffer (pH 7.7). Five microliters of this solution was added to each of the 0.1 M TFA-hydrolyzed samples that had been redissolved in 20  $\mu$ L 50 mM phosphate buffer (pH 7.7), and also to a *N*-acetylneuraminic acid standard. Control reactions received buffer in place of the aldolase solution. Reactions were incubated at room temperature for 1 hour before being dried for APTS derivatization procedures.

#### 7.6.7 APTS (1-Aminopyrene-3,6,8-Trisulfonate) Derivatization of Carbohydrates for CE-LIF Analyses

APTS (1-aminopyrene-3,6,8-trisulfonate) derivatization and analysis of monosaccharides was performed using a modified version of several procedures described by Beckman Instruments (Chen, Dobashi and Evangelista 1998; Chen and Evangelista 1995; Evangelista, Guttman and Chen 1996). Carbohydrate samples and standards, to be derivatized and analyzed by capillary electrophoresis, were separately

evaporated to dryness in 1.5 mL Eppendorf tubes, using a Speedvac Centrifuge. Samples were redissolved in 6  $\mu$ L 1.8 M citric acid, to which 6  $\mu$ L 0.1 M APTS and 6  $\mu$ L 1 M sodium cyanoborohydride in tetrahydrofuran were added. The 18  $\mu$ L reactions were loaded into a sample tray and held at 55 °C for two hours in a PE Applied Biosystems GeneAmp PCR System, to prevent the reactions from drying, before being diluted with 580  $\mu$ L of water and stored at -20 °C until analysis. CE running conditions are described in Section 7.3.4. Sample peaks were identified by comigration with APTS derivatives of known carbohydrate compounds.

## **7.7 Specific Protein Purification and Labeling Procedures**

### 7.7.1 Partial Purification of *Nostoc commune* DRH1 Glucuronosyltransferase Activity

A 15 L culture of *Nostoc commune* DRH1 was inoculated and grown under the conditions described in Section 7.5.1. Harvested cyanobacterial cell mass was lysed using a French pressure cell, and processed as described in Section 7.5.2, to produce a Triton X-100 detergent extract of the membrane material. Activity assays (Section 7.3.3) were performed on both the solid membrane fragments and the concentrated detergent extract to ensure that glucuronosyltransferase activity was present before proceeding.

Following concentration with a 15 mL YM-50 Centriprep centrifugal concentrator, the active detergent extract was loaded onto a 2 mL Q-Sepharose column, equilibrated in 50 mM HEPES (pH 7.0), containing 0.4% Triton X-100, using a FPLC instrument, and at a flow rate of 1 mL/min. The majority of the protein material that was loaded did not adhere to the column, as judged by the relative UV absorbances of the flow-through fraction and the salt eluates in the chromatogram. The column was washed with 50 mM HEPES (pH 7.0), containing 0.4% Triton X-100 and 0.1 M NaCl, until UV absorbance returned to the baseline (10-20 mL, depending on sample loading). Glycosyltransferase activity was eluted using 50 mM HEPES (pH 7.0), containing 0.4% Triton X-100 and 0.25 M NaCl. The purified active fraction (0.25 M NaCl eluate), frequently pink in color, was concentrated and tested for activity. This sample was typically combined with an equal volume of glycerol and stored at 4 °C.

### 7.7.2 *Nostoc commune* DRH1 Glucuronosyltransferase Affinity Labeling

The purified *Nostoc commune* DRH1 membrane detergent extract, which showed glucuronosyltransferase activity, was labeled using a <sup>32</sup>P-labeled 5'-azido-UDP-glucuronic acid affinity tag from Affinity Labeling Technologies, according to the manufacturer's instructions. An aliquot of the affinity tag stock solution in methanol, sufficient to produce a final concentration of 10-20 μM, was evaporated to dryness under a stream of nitrogen, in a glass test tube, and then redissolved in water. A 0.5-1 mL sample of *Nostoc commune* DRH1 membrane detergent extract, containing 50 mM MgCl<sub>2</sub>, was added to the tag solution and the mixture was vortexed for 30 s. The sample was irradiated for 60 s with 254 nm UV light, before excess tagging reagent was inactivated by the addition of DTT to a final concentration of 20 mM or greater. Excess tag was normally separated by an acetone precipitation of labeled protein, after which protein material was redissolved in either 1X SDS-PAGE loading buffer, or two-dimensional lysis buffer, for electrophoresis (Appendix).

### 7.7.3 Purification of *Anabaena* (*Nostoc*) PCC 7120 SLH Domain-Containing Proteins and SDS-PAGE Comigrating Carbohydrate Material

A 15 L culture of *Anabaena* (*Nostoc*) PCC 7120 was grown under the conditions described in Section 7.5.1. Procedures for fermentor harvesting, cell lysis, and soluble/insoluble protein fractionation were the same as those employed for *Nostoc commune* DRH1, described in Section 7.5.2. Following ultracentrifugation, the blue soluble fraction was passed through a strong ion exchange column (Q-Sepharose), equilibrated with 50 mM Tris at pH 7.0. The flow-through fraction, now a brown color, was combined with an equal volume of saturated ammonium sulfate, and the resulting mixture was allowed to precipitate at 4 °C, overnight. Precipitated material was collected by centrifugation, washed once with saturated ammonium sulfate, and redissolved in 50 mM Tris at pH 7.0. This solution was then combined with an equal volume of 2X SDS-PAGE loading buffer and separated by SDS-PAGE, using BioRad Criterion 4-15% gradient gels with preparative scale (800 μL) sample wells. Protein/carbohydrate-containing samples were prepared from these gels, for further analyses, by electroelution of the material in particular gel regions, defined by the molecular weight markers.

Alternatively, gels were sometimes stained with the non-fixing Electro-Blue protein stain, so that individual bands could be identified and excised, for electroelution and further analyses. A small section of each of these gels was PAS-stained to confirm the presence of carbohydrate.

## **7.8 Electronic Resources**

### 7.8.1 Genomic Databases

The *Anabaena (Nostoc)* PCC 7120 genome sequence is available at the CyanoBase website ([www.kazusa.or.jp/cyanobase](http://www.kazusa.or.jp/cyanobase)). The *Nostoc punctiforme* ATCC 29133 genome sequence is available at the National Center for Biotechnology Information (NCBI) website ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Both of these genomes were searched using the BLAST, FASTA, and FASTS algorithms at the VA Tech Institute for Genomics website ([vigen.biochem.vt.edu](http://vigen.biochem.vt.edu)). The genome of *Nostoc commune* DRH1 is presently unsequenced.

### 7.8.2 Search Algorithms

Protein sequences, determined by tandem mass spectrometry, were compared to cyanobacterial genomes using the BLAST, FASTA, and FASTS algorithms (Altschul *et al.* 1997; Mackey, Haystead and Pearson 2002; Pearson and Lipman 1988). The BLAST (Basic Local Alignment Search Tool) algorithm compares small segments of a query sequence to pre-indexed databases, and looks for non-overlapping stretches of these small segments. A score for each database entry is generated, based upon the sequence similarity of the database entry to the query sequence, and database entries with statistically significant scores are reported. Scores generated for each database entry by the BLAST algorithm are based upon amino acid similarity, as well as identity, unlike the FASTA algorithm. The FASTA algorithm also compares query sequences to a pre-indexed database, but unlike the BLAST algorithm, the FASTA algorithm compares all possible segments of the query sequence (of user specified length). The FASTS algorithm is a variant of the FASTA algorithm, designed for comparison of short peptide fragments to a database, such as those determined by mass spectrometric analysis of a protein. Its use is described by Mackey *et al.* (Mackey, Haystead and Pearson 2002).



### 7.8.3 Protein Function Analysis

Protein function analyses were performed using the Simple Modular Architecture Research Tool ([smart.embl-heidelberg.de](http://smart.embl-heidelberg.de)), the ProfileScan Server ([hits.isb-sib.ch/cgi-bin/PFSCAN](http://hits.isb-sib.ch/cgi-bin/PFSCAN)), the SYSTEMatic Re-Searching database ([systers.molgen.mpg.de](http://systers.molgen.mpg.de)), and the InterPro database ([www.ebi.ac.uk/interpro](http://www.ebi.ac.uk/interpro)). Protein signal sequences were also identified using the SignalP World Wide Web Prediction Server ([www.cbs.dtu.dk/services/SignalP](http://www.cbs.dtu.dk/services/SignalP)). Sequences were aligned using the ClustalX interface for the ClustalW mutiple sequence alignment program ([www-igbmc.u-strasbg.fr/BioInfo](http://www-igbmc.u-strasbg.fr/BioInfo)). Potential membrane-spanning regions of proteins were identified using the TMPred program ([www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)).

### 7.8.4 Miscellaneous

Manual (*de novo*) sequencing of peptide data was facilitated by the use of the KVL Aminosequencer program, available at [www.dina.dk/~sestoft/aminoseq](http://www.dina.dk/~sestoft/aminoseq).

## **7.9 Synthesis of a Fluorescent-Labeled Glucose- $\beta$ (1 $\rightarrow$ 4)-Xylose Acceptor**

*1,2,3,4-tetra-O-acetyl-D-xylopyranoside* (**1**). – A mixture of D-xylose (10.03g, 66.8 mmol), acetic anhydride (58 mL, 603 mmol), and freshly distilled pyridine (55 mL) were stirred, under nitrogen, at room temperature, until all solid material had dissolved. After co-evaporation with toluene (125 mL, 2x), under reduced pressure, at 55 °C, the residue was dissolved in methylene chloride (200 mL), washed with cold, aqueous 3% HCl (100 mL, 2x) and then 1 M ammonium chloride (100 mL). The organic layer was dried with sodium sulfate, filtered, and concentrated to dryness under reduced pressure. Crystallization from ether yielded a sticky, white, low-melting solid from which the mother liquor was withdrawn. The solid material was rinsed with hexane, dried under high vacuum, and taken on to the next step. Integration of  $^1\text{H}$  NMR anomeric proton resonances revealed this material to be a 3.85-to-1  $\alpha/\beta$  mixture of the title compound. **1 $\alpha$** :  $\delta_{\text{H}}$ (acetone- $d_6$ ) 2.02(s, COCH<sub>3</sub>); 2.03(s, COCH<sub>3</sub>); 2.04(s, COCH<sub>3</sub>); 2.21(s, COCH<sub>3</sub>); 3.77(1H, t, J<sub>4,5b</sub>11.08 Hz, J<sub>5a,5b</sub>11.33 Hz, H-5b); 3.94(1H, dd, J<sub>4,5a</sub>5.95 Hz, H-5a); 5.03(1H, ddd, J<sub>3,4</sub>9.89 Hz, H-4); 5.04(1H, dd, J<sub>1,2</sub>3.66 Hz, J<sub>2,3</sub>9.89 Hz, H-2); 5.48(1H, t,

H-3); 6.24(1H, d, H-1);  $\delta_{\text{C}}$ (acetone- $d_6$ ) 20.35-20.64(COCH<sub>3</sub>), 61.13, 69.24, 69.84, 70.27, 89.75(C-1), and 169.3-170.3(COCH<sub>3</sub>). **1 $\beta$** :  $\delta_{\text{H}}$ (acetone- $d_6$ ) 2.04(s, COCH<sub>3</sub>); 2.05(s, COCH<sub>3</sub>); 2.10(s, COCH<sub>3</sub>); 2.24(s, COCH<sub>3</sub>); 3.68(1H, dd,  $J_{4,5b}$ 8.70 Hz,  $J_{5a,5b}$ 11.91 Hz, H-5b); 4.15(1H, dd,  $J_{4,5a}$ 4.95 Hz, H-5a); 4.98(1H, dt,  $J_{3,4}$ 8.43 Hz, H-4); 5.07(1H, dd,  $J_{1,2}$ 6.96 Hz,  $J_{2,3}$ 8.43 Hz, H-2); 5.30(1H, t, H-3); 5.82(1H, d, H-1);  $\delta_{\text{C}}$ (acetone- $d_6$ ) 20.35-20.64(COCH<sub>3</sub>), 63.15, 69.11, 70.31, 71.54, 92.68(C-1), and 169.3-170.3(COCH<sub>3</sub>).

*2,3,4-tri-O-acetyl- $\alpha$ -D-xylopyranosyl bromide (2)*. – Hydrogen bromide (30% w/w in acetic acid, 90 mL) was added to a solution of compound **1** in methylene chloride (30 mL). After four hours, the reaction mixture was poured into a stirred beaker of ice water (150 mL) and methylene chloride (150 mL). The organic layer was separated, washed several times with cold, aqueous, saturated sodium bicarbonate, dried with sodium sulfate, filtered, and concentrated to dryness under reduced pressure. The resulting brown oil was dissolved in warm ether and allowed to stand at -12 °C, overnight, to produce a white crystalline solid. This material was collected by filtration and washed with hexane to give the title compound (11.7g, 52% from D-xylose).  $\delta_{\text{H}}$ (acetone- $d_6$ ) 2.00(s, COCH<sub>3</sub>); 2.01(s, COCH<sub>3</sub>); 2.04(s, COCH<sub>3</sub>); 3.81(1H, t,  $J_{4,5b}$ 10.90 Hz,  $J_{5a,5b}$ 11.36 Hz, H-5b); 4.11(1H, dd,  $J_{4,5a}$ 6.05 Hz, H-5a); 4.90(1H, dd,  $J_{1,2}$ 3.85 Hz,  $J_{2,3}$ 9.93 Hz, H-2); 5.10(1H, ddd,  $J_{3,4}$ 9.71 Hz, H-4); 5.51(1H, t, H-3); 6.72(1H, d, H-1);  $\delta_{\text{C}}$ (acetone- $d_6$ ) 20.47-20.51(COCH<sub>3</sub>), 63.47, 68.41, 70.17, 71.30, 90.39(C-1), and 170.11-170.13(COCH<sub>3</sub>).

*4-nitrophenyl 2,3,4-tri-O-acetyl- $\beta$ -D-xylopyranoside (3)*. – A mixture of 4-nitrophenol (8.17g, 58.7 mmol), potassium carbonate (16.6g, 120.3 mmol), and acetone (200 mL) were brought to reflux using a Liebig condenser. A solution of **2** (11.64g, 34.3 mmol) in acetone (68 mL) was then added drop-wise down the condenser over the course of 75 minutes. The mixture was refluxed for 18 hours, cooled, filtered, and concentrated to dryness under reduced pressure. The residue was dissolved in acetone (100 mL), combined with an equal volume of cold water, and allowed to precipitate at 4 °C for several hours. The resulting white solid was washed with cold water until no visible trace of 4-nitrophenol remained, dissolved in acetone, dried with sodium sulfate, filtered, concentrated, and crystallized from warm absolute ethanol to give the title compound

(7.38g, 54%).  $\delta_{\text{H}}$ (acetone- $d_6$ ) 2.02(s, COCH<sub>3</sub>); 2.03(s, COCH<sub>3</sub>); 2.08(s, COCH<sub>3</sub>); 3.80(1H, dd,  $J_{4,5b}$ 8.43 Hz,  $J_{5a,5b}$ 11.91 Hz, H-5b); 4.20(1H, dd,  $J_{4,5a}$ 4.95 Hz, H-5a); 5.02(1H, dt,  $J_{3,4}$ 8.33 Hz, H-4); 5.18(1H, dd,  $J_{1,2}$ 6.59 Hz,  $J_{2,3}$ 8.43 Hz, H-2); 5.31(1H, t, H-3); 5.63(1H, d, H-1); 7.26(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>); 8.24(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>);  $\delta_{\text{C}}$ (acetone- $d_6$ ) 20.52-20.58(COCH<sub>3</sub>), 62.78, 69.13, 70.83, 71.45, 98.54(C-1), 117.58(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 126.50(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 143.83(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 162.35(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 169.72(COCH<sub>3</sub>), 170.14(COCH<sub>3</sub>), and 170.22(COCH<sub>3</sub>).

*4-nitrophenyl β-D-xylopyranoside* (**4**). – Sodium methoxide (93.7 mg) was added to a suspension of **3** (6.61g, 16.6 mmol) in dry methanol (100 mL), and the mixture was stirred overnight. The solution was neutralized with cation exchange resin (Dowex HCRW2, H<sup>+</sup>), filtered, and concentrated to dryness under reduced pressure. Crystallization from acetone gave the title compound (4.14g, 92%).  $\delta_{\text{H}}$ (acetone- $d_6$ ) 3.48-3.65 (4H, m); 3.94(1H, dd,  $J_{4,5a}$ 5.04 Hz,  $J_{5a,5b}$ 11.26 Hz, H-5a); 4.42(2H, s, OH); 5.16(1H, d,  $J_{1,2}$ 6.96 Hz, H-1); 7.23(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>); 8.21(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>);  $\delta_{\text{C}}$ (acetone- $d_6$ ) 66.57, 70.45, 74.07, 77.28, 101.76(C-1), 117.36(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 126.36(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 143.26(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), and 163.39(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>).

*4-nitrophenyl 4-O-chloroacetyl-β-D-xylopyranoside* (**5**). – This compound was prepared according to a procedure of Takeo *et al.* (Takeo *et al.* 1995), with minor modifications. A suspension of **4** (1.01g, 3.72 mmol) and dibutyltin oxide (917.8 mg, 3.69 mmol) in methanol (20 mL) was refluxed for two hours, during which time it was concentrated to approximately half of its original volume using a Dean-Stark condenser. Upon cooling, the mixture was concentrated to dryness under reduced pressure and then dissolved in methylene chloride (20 mL). A solution of chloroacetyl chloride (310 μL, 3.89 mmol) in methylene chloride (4 mL) was added, drop-wise at 0 °C, with stirring, over the course of 40 minutes. The mixture was allowed to warm over the next 3.5 hours, at which point it was again concentrated to dryness. The resulting residue was dissolved in acetonitrile (50 mL), washed with petroleum ether (50 mL, 3x) and concentrated to dryness. Flash chromatography (1:3 toluene-ethyl acetate) gave the title compound as an oil, which crystallized upon standing (507.4 mg, 39.2%).  $\delta_{\text{H}}$ (acetone- $d_6$ ) 3.64-3.70(2H, m,  $J_{1,2}$ 7.33

Hz,  $J_{2,3}$ 8.98 Hz,  $J_{4,5b}$ 9.43 Hz,  $J_{5a,5b}$ 11.63 Hz, H-2, H-5b); 3.81(1H, dt,  $J_{3,4}$ 8.98 Hz, H-3); 4.09(1H, dd,  $J_{4,5a}$ 5.40 Hz, H-5a); 4.30(1H, d,  $J_{CH_2}$ 15.20 Hz, COCH<sub>2</sub>Cl); 4.33(1H, d, COCH<sub>2</sub>Cl); 4.74(1H, d, OH); 4.91(1H, ddd, H-4); 4.98(1H, d, OH); 5.26(1H, d, H-1); 7.25(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>); 8.22(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>);  $\delta_C$ (acetone-d<sub>6</sub>) 30.11(COCH<sub>2</sub>Cl), 63.19, 73.77, 74.19, 74.21, 101.59(C-1), 117.41(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 126.40(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 143.44(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 163.21(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), and 167.63(COCH<sub>2</sub>Cl).

*4-nitrophenyl 2,3-di-O-benzoyl-4-O-chloroacetyl- $\beta$ -D-xylopyranoside (6)*. – Benzoyl chloride (340  $\mu$ L, 2.93 mmol), diisopropylethylamine (725  $\mu$ L), and 4-dimethylaminopyridine (231.2 mg, 1.89 mmol) were successively added to a suspension of **5** (488.3 mg, 1.4 mmol) in methylene chloride (5 mL) at 0 °C. After addition the mixture was removed from the ice bath and allowed to warm for 30 minutes, at which point it was returned, and methanol added drop-wise to quench the reaction. The mixture was diluted with methylene chloride, washed with cold, aqueous 3% HCl (2x) and 1 M ammonium sulfate (2x). The organic layer was dried with sodium sulfate, filtered, concentrated, and crystallized from warm absolute ethanol to give the title compound (493.8 mg, 63%).  $\delta_H$ (acetone-d<sub>6</sub>) 4.04(1H, dd,  $J_{4,5b}$ 7.00 Hz,  $J_{5a,5b}$ 12.36 Hz, H-5b); 4.28(1H, d,  $J_{CH_2}$ 15.20 Hz, COCH<sub>2</sub>Cl); 4.33(1H, d, COCH<sub>2</sub>Cl); 4.46(1H, dd,  $J_{4,5a}$ 4.40 Hz, H-5a); 5.42(1H, dt,  $J_{3,4}$ 7.21 Hz, H-4); 5.66(1H, dd,  $J_{1,2}$ 5.50 Hz,  $J_{2,3}$ 7.37 Hz, H-2); 5.79(1H, t, H-3); 6.01(1H, d, H-1); 7.30(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>); 7.48(m, COC<sub>6</sub>H<sub>5</sub>); 7.62(m, COC<sub>6</sub>H<sub>5</sub>); 8.02(m, COC<sub>6</sub>H<sub>5</sub>); 8.23(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>);  $\delta_C$ (acetone-d<sub>6</sub>) 41.46(COCH<sub>2</sub>Cl), 62.17, 70.28, 70.68, 71.10, 98.16(C-1), 117.59(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 126.55(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 129.46(COC<sub>6</sub>H<sub>5</sub>), 130.10(COC<sub>6</sub>H<sub>5</sub>), 130.19(COC<sub>6</sub>H<sub>5</sub>), 130.46(COC<sub>6</sub>H<sub>5</sub>), 130.49(COC<sub>6</sub>H<sub>5</sub>), 134.42(COC<sub>6</sub>H<sub>5</sub>), 134.47(COC<sub>6</sub>H<sub>5</sub>), 143.83(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 162.23(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 165.53(COC<sub>6</sub>H<sub>5</sub>), 165.73(COC<sub>6</sub>H<sub>5</sub>), and 167.37(COCH<sub>2</sub>Cl).

*4-nitrophenyl 2,3-di-O-benzoyl- $\beta$ -D-xylopyranoside (7)*. – Compound **6** (401.3mg, 0.72 mmol) and thiourea (284.6mg, 3.74 mmol) were refluxed in methanol (10 mL) for 7.5 hours then concentrated to dryness under reduced pressure. The residue was partitioned between methylene chloride and water to remove excess thiourea, and then the organic layer was separated and concentrated to dryness. Flash chromatography (6:1 toluene-

ethyl acetate) and crystallization from warm absolute ethanol at -12 °C yielded the title compound (259.2mg, 75%).  $\delta_{\text{H}}$ (acetone- $d_6$ ) 3.87(1H, dd, H-5b); 4.20-4.26(2H, m, H-4, H-5a); 4.99(1H, d, OH); 5.55(1H, dd,  $J_{1,2}$ 7.14 Hz,  $J_{2,3}$ 9.25 Hz, H-2); 5.64(1H, t,  $J_{3,4}$ 8.24 Hz, H-3); 5.84(1H, d, H-1); 7.26(2H, m,  $C_6H_4NO_2$ ); 7.44(m,  $COC_6H_5$ ); 7.57(m,  $COC_6H_5$ ); 7.94(m,  $COC_6H_5$ ); 8.20(2H, m,  $C_6H_4NO_2$ );  $\delta_{\text{C}}$ (acetone- $d_6$ ) 66.61, 68.58, 72.39, 76.17, 99.25(C-1), 117.55( $C_6H_4NO_2$ ), 126.50( $C_6H_4NO_2$ ), 129.31( $COC_6H_5$ ), 129.39( $COC_6H_5$ ), 130.30( $COC_6H_5$ ), 130.33( $COC_6H_5$ ), 130.36( $COC_6H_5$ ), 130.79( $COC_6H_5$ ), 134.06( $COC_6H_5$ ), 134.27( $COC_6H_5$ ), 143.78( $C_6H_4NO_2$ ), 162.56( $C_6H_4NO_2$ ), 165.71( $COC_6H_5$ ), and 166.22( $COC_6H_5$ ).

*1,2,3,4,6-penta-O-acetyl-D-glucopyranoside (8)*. – Anhydrous sodium acetate (997.6mg, 12.2 mmol) was added to acetic anhydride (24 mL, 250 mmol) at 100 °C, with stirring. D-glucose (2.03g, 11.3 mmol) was slowly added to this mixture, producing a homogenous solution within 5 minutes. The solution was stirred, with heating, for 2 hours, at which point water (32 mL) was slowly added, before being allowed to cool overnight. The aqueous phase was extracted with chloroform (30 mL, 2x), and the extracts pooled and concentrated to dryness under reduced pressure. Crystallization from warm absolute ethanol at -12 °C gave the title compound, collected by filtration and washed with hexane. Integration of  $^1\text{H}$  NMR anomeric proton resonances revealed this material to be a 3.48-to-1  $\beta/\alpha$  anomeric mixture (4.00g, 91%).  $\delta_{\text{H}}$ (acetone- $d_6$ ) 1.94-2.18( $COCH_3'$ ); 4.01-4.14(m, H-5' $_{\beta}$ , H-6b' $_{\alpha}$ , H-6b' $_{\beta}$ ); 4.18-4.29(m, H-6a' $_{\alpha}$ , H-6a' $_{\beta}$ ); 5.03(dd,  $J_{1',2'(\alpha)}$  3.85 Hz,  $J_{2',3'(\alpha)}$  10.26 Hz, H-2' $_{\alpha}$ ); 5.04(dd,  $J_{1',2'(\beta)}$  8.24 Hz,  $J_{2',3'(\beta)}$  9.71 Hz, H-2' $_{\beta}$ ); 5.09(t,  $J_{3',4'(\beta)}$  9.61 Hz, H-4' $_{\beta}$ ); 5.15(t,  $J_{3',4'(\alpha)}$  9.89 Hz, H-4' $_{\alpha}$ ); 5.38(t, H-3' $_{\beta}$ ); 5.47(t, H-3' $_{\alpha}$ ); 5.88(d, H-1' $_{\beta}$ ); 6.27(d, H-1' $_{\alpha}$ ); **1 $\alpha$** :  $\delta_{\text{C}}$ (acetone- $d_6$ ) 20.34-20.65( $COCH_3'$ ), 62.28, 68.73, 70.14, 70.36, 70.62, 89.49(C-1'), 169.49( $COCH_3'$ ), 169.87( $COCH_3'$ ), 170.26( $COCH_3'$ ), 170.27( $COCH_3'$ ), and 72.15( $COCH_3'$ ). **1 $\beta$** :  $\delta_{\text{C}}$ (acetone- $d_6$ ) 20.34-20.65( $COCH_3'$ ), 62.39, 68.82, 71.13, 73.17, 92.28(C-1'), 169.33( $COCH_3'$ ), 169.68( $COCH_3'$ ), 169.94( $COCH_3'$ ), 170.19( $COCH_3'$ ), and 170.62( $COCH_3'$ ).

*Ethyl 1-thio-2,3,4,6-tetra-O-acetyl- $\beta$ -D-glucopyranoside (9)*. – This compound was prepared according to a procedure of Contour *et al.* (Contour *et al.* 1989), with minor

modifications. Two 0.5 mL aliquots of ethanethiol (13.5 mmol total) were added to an anomeric mixture of **8** (2.13g, 1.50g 1 $\beta$ , 3.85 mmol 1 $\beta$ ) and zirconium tetrachloride (1.43g, 6.14 mmol) in methylene chloride (85 mL), at 0 °C, one hour apart. The mixture was maintained at 0 °C for an additional 21 hours, at which point it was partitioned between cold methylene chloride and water. The organic layer was separated, washed successively with cold water, saturated sodium bicarbonate, and water, then dried with sodium sulfate, filtered, and concentrated to dryness under reduced pressure. Flash chromatography (1:1 ether-hexane) and crystallization from absolute ethanol at -20 °C gave the title compound in 24% yield from the  $\beta$ -anomer (377.6mg). Product co-eluting with later compounds was not used in calculating percent yield.  $\delta_{\text{H}}$ (acetone- $d_6$ ) 1.24(1H, t,  $J_{\text{CH}_2',\text{CH}_3'}$ 7.51 Hz,  $\text{CH}_2\text{CH}_3'$ ); 1.94-2.01(COCH $_3'$ ); 2.64-2.78(2H, m,  $\text{CH}_2\text{CH}_3'$ ); 3.96(1H, ddd,  $J_{4',5'}$ 9.89 Hz,  $J_{5',6a'}$ 5.40 Hz,  $J_{5',6b'}$ 2.38 Hz, H-5'); 4.11(1H, dd,  $J_{6a',6b'}$ 12.27 Hz, H-6b'); 4.22(1H, dd, H-6a'); 4.81(1H, d,  $J_{1',2'}$ 10.07 Hz, H-1'); 4.94(1H, dd,  $J_{2',3'}$ 9.43 Hz, H-2'); 5.02(1H, t,  $J_{3',4'}$ 9.62 Hz, H-4'); 5.27(1H, t, H-3');  $\delta_{\text{C}}$ (acetone- $d_6$ ) 15.39( $\text{CH}_2\text{CH}_3'$ ), 20.48-20.59(COCH $_3'$ ), 24.52( $\text{CH}_2\text{CH}_3'$ ), 62.96, 69.37, 70.79, 74.40, 76.22, 83.65(C-1'), 169.69(COCH $_3'$ ), 169.95(COCH $_3'$ ), 170.19(COCH $_3'$ ), and 170.62(COCH $_3'$ ).

*4-nitrophenyl O-(2,3,4,6-tetra-O-acetyl- $\beta$ -D-glucopyranosyl)-(1 $\rightarrow$ 4)-2,3-di-O-benzoyl- $\beta$ -D-xylopyranoside (10).* – This compound was prepared according to a procedure of Takeo *et al.* (Takeo *et al.* 1995), with minor modifications. A solution of silver triflate (22.6 mg, 88  $\mu\text{mol}$ ) in toluene (600  $\mu\text{L}$ ) was added drop-wise, at -5 °C to a mixture of **9** (225.7 mg, 575  $\mu\text{mol}$ ), **7** (212.0 mg, 440  $\mu\text{mol}$ ), 4 Å molecular sieves (309.1mg), and *N*-iodosuccinimide (126.4 mg, 560  $\mu\text{mol}$ ) in methylene chloride (4 mL). After stirring under nitrogen for approximately ten minutes, the mixture was diluted with additional methylene chloride and filtered through Celite into ice water. The organic layer was then separated and washed successively with 10% sodium thiosulfate, saturated sodium bicarbonate, and water. The methylene chloride layer was dried with sodium sulfate and concentrated to dryness under vacuum. The resulting oil was separated by flash chromatography (1:1 chloroform-ethyl acetate, 500 mL) and eluted from the column with methanol (100 mL). The concentrated methanol eluant was again separated by flash

chromatography (19:1 chloroform-ethyl acetate, 500 mL) and eluted from the column with acetone (100 mL). The concentrated acetone eluant was purified by preparative-scale thin-layer chromatography (1:1 chloroform-ethyl acetate) to give the title compound (31.4mg, 8%).  $\delta_{\text{H}}$ (acetone- $d_6$ ) 1.88-2.41(COCH<sub>3</sub>' ); 3.76(1H, dd,  $J_{5',6b}$ 2.40 Hz,  $J_{6a',6b}$ 12.10 Hz, H-6b'); 3.87(1H, ddd,  $J_{4',5'}$ 9.80 Hz,  $J_{5',6a}$ 4.90 Hz, H-5'); 3.90-3.97(2H, m, H-5b, H-6a'); 4.31-4.36(2H, m,  $J_{3,4}$ 7.33 Hz, H-4, H-5a); 4.86(1H, t,  $J_{3',4'}$ 9.62 Hz, H-4'); 4.89(1H, dd,  $J_{1',2'}$ 8.06 Hz,  $J_{2',3'}$ 9.71 Hz, H-2'); 5.03(1H, d, H-1'); 5.23(1H, t, H-3'); 5.25(1H, dd,  $J_{1,2}$ 5.86 Hz,  $J_{2,3}$ 7.51 Hz, H-2); 5.76(1H, t, H-3); 5.90(1H, d, H-1); 7.28(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>); 7.48(4H, m, COC<sub>6</sub>H<sub>5</sub>); 7.61(2H, m, COC<sub>6</sub>H<sub>5</sub>); 8.02(4H, m, COC<sub>6</sub>H<sub>5</sub>); 8.22(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>);  $\delta_{\text{C}}$ (acetone- $d_6$ ) 20.48-20.55(COCH<sub>3</sub>' ), 62.53(C6'), 63.32(C5), 69.09(C4'), 71.10, 71.92, 72.07, 72.46(C5'), 73.31(C3'), 75.92(C4), 98.28(C1), 101.5(C1'), 117.51(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 126.51(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 129.33(COC<sub>6</sub>H<sub>5</sub>), 129.37(COC<sub>6</sub>H<sub>5</sub>), 130.26(COC<sub>6</sub>H<sub>5</sub>), 130.44(COC<sub>6</sub>H<sub>5</sub>), 130.55(COC<sub>6</sub>H<sub>5</sub>), 130.67(COC<sub>6</sub>H<sub>5</sub>), 134.17(COC<sub>6</sub>H<sub>5</sub>), 134.32(COC<sub>6</sub>H<sub>5</sub>), 143.73(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 162.34(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), 165.58(COC<sub>6</sub>H<sub>5</sub>), 165.69(COC<sub>6</sub>H<sub>5</sub>), 169.64(COCH<sub>3</sub>' ), 169.84(COCH<sub>3</sub>' ), 170.24(COCH<sub>3</sub>' ), and 170.55(COCH<sub>3</sub>' ).

*4-nitrophenyl O-(β-D-glucopyranosyl)-(1→4)-β-D-xylopyranoside (11)*. – A solution of **10** and sodium methoxide (10 mg) in dry methanol (10 mL) was stirred under nitrogen, overnight, after which time thin layer chromatography indicated that the reaction was complete (1:1 chloroform-ethyl acetate). The solution was neutralized and concentrated as in **4**. The crude material was dried under high vacuum and taken onto the next step (19.0 mg). An amber contaminant obtained from the cation exchange resin interfered with the calculation of this and successive yields.  $\delta_{\text{H}}$ (1:1 acetone- $d_6$ -methanol- $d_4$ ) 3.13-3.39(m); 3.61-3.66(4H, m); 3.77(1H, m); 3.84(1H, dt); 4.06(1H, dd); 4.39-4.42(2H, m); 5.08(1H, d); 7.19(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>); 8.16(2H, m, C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>);  $\delta_{\text{C}}$ (1:1 acetone- $d_6$ -methanol- $d_4$ ) 62.18, 64.57, 71.18, 74.08, 74.19, 75.47, 77.51, 77.87(wide), 101.68, 103.31, 117.51(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>), and 126.43(C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>).

*4-aminophenyl O-(β-D-glucopyranosyl)-(1→4)-β-D-xylopyranoside (12)*. – This compound was prepared according to a procedure of Roy and Tropper (Roy and Tropper

1991), with minor modifications. Ammonium formate (42.4 mg, 672  $\mu\text{mol}$ ), 5% palladium on carbon (15.6 mg), and crude **11** (19.0 mg) were refluxed in methanol (5 mL) for approximately ten minutes, at which point thin layer chromatography indicated that the reaction was complete (85:15 acetonitrile-water). The solution was filtered, concentrated, and dried under high vacuum. The resulting material was dissolved in deuterium oxide (1 mL) and lyophilized twice before NMR analysis.  $\delta_{\text{H}}$ (deuterium oxide) 3.27-3.31(1H, dd); 3.38-3.57(5H, m); 3.68(1H, t); 3.74(1H, dd); 3.90-3.94(2H, m); 4.13(1H, d); 4.55(1H, d); 4.94(1H, d); 6.82(2H, m,  $\text{C}_6\text{H}_4\text{NH}_2$ ); 6.98(2H, m,  $\text{C}_6\text{H}_4\text{NH}_2$ ); 8.45(s,  $\text{C}_6\text{H}_4\text{NH}_2$ );  $\delta_{\text{C}}$ (deuterium oxide) 61.24, 63.51, 70.11, 73.33, 73.41, 74.25, 76.09, 76.52, 76.87, 101.71, 102.48, 118.05( $\text{C}_6\text{H}_4\text{NH}_2$ ), 118.94( $\text{C}_6\text{H}_4\text{NH}_2$ ), 142.50( $\text{C}_6\text{H}_4\text{NH}_2$ ), 150.28( $\text{C}_6\text{H}_4\text{NH}_2$ ), and 163.37.

*N*-(4-phenyl O-( $\beta$ -D-glucopyranosyl)-(1 $\rightarrow$ 4)- $\beta$ -D-xylopyranoside) 6-(fluorescein-5-carboxamido)hexanamide (**13**). – This compound was prepared according to a procedure of Wakarchuk *et al.* (Wakarchuk *et al.* 1996), with minor modifications. 6-(Fluorescein-5-Carboxamido)Hexanoic Acid Succinimidyl Ester (5 mg, 8.5  $\mu\text{mol}$ , Molecular Probes F-6106) was dissolved in methanol (250  $\mu\text{L}$ ) and added to a solution of **12** in 0.2 M triethylamine acetate buffer (500  $\mu\text{l}$ , pH 8.2). This mixture was shaken in the dark for three days, frozen, and lyophilized under high vacuum. The resulting yellow material was purified by preparative-scale thin-layer chromatography (70:20:10:1 ethyl acetate-methanol-water-acetic acid). The product band was extracted with water (10 mL, 5x), concentrated, de-salted with a Bond Elut C18 solid-phase extraction cartridge, concentrated again, and dried under high vacuum to give the FCHASE derivative of **12** (1.6 mg).

#### 7.10 Synthesis of a Fluorescent-Labeled Glucose- $\beta$ (1 $\rightarrow$ 4)-Glucose Acceptor

2,3,6,2',3',4',6'-hepta-O-acetyl- $\alpha$ -D-cellobiosyl bromide (**14**). – This compound was prepared according to a procedure of Kartha and Jennings (Kartha and Jennings 1990), with modifications. Hydrogen bromide (30% w/w in acetic acid, 11 mL) was added to a mixture of  $\beta$ -D-cellobiose (5.07g, 14.8 mmol), water (4 drops), and acetic anhydride (25



mL, 262.4 mmol), which was then stirred overnight. Additional hydrogen bromide (30% w/w in acetic acid, 92.5 mL) and acetic anhydride (29 mL, 304 mmol) were added over the course of the next three days, until the solution became homogenous. After stirring for an additional 3.5 hours, the reaction mixture was poured into a well-stirred mixture of ice water (150 mL) and cold methylene chloride (150 mL). The organic layer was separated, washed several times with cold, aqueous, saturated sodium bicarbonate and water, dried with sodium sulfate, filtered, and concentrated to dryness under reduced pressure. The title compound was crystallized from the resulting oil in methylene chloride-ether, collected by filtration, washed with ether, and dried under high vacuum (8.61g, 83%).  $\delta_{\text{H}}$ (acetone- $d_6$ ) 1.93(s, COCH<sub>3</sub>); 1.98(s, COCH<sub>3</sub>); 2.03(s, COCH<sub>3</sub>); 2.05(s, COCH<sub>3</sub>); 2.06(s, COCH<sub>3</sub>); 2.07(s, COCH<sub>3</sub>); 2.09(s, COCH<sub>3</sub>); 4.02(1H, ddd); 4.07(1H, dd); 4.12(1H, t); 4.23(1H, m); 4.27(1H, dd); 4.39(1H, dd); 4.55(1H, dd); 4.83-4.91(3H, m,  $J_{1,2}$ 4.03 Hz,  $J_{1',2'}$ 8.06 Hz, H-2, H-1', H-2'); 5.05(1H, t); 5.23(1H, t); 5.50(1H, t); 6.71(1H, d, H-1);  $\delta_{\text{C}}$ (acetone- $d_6$ ) 20.46-20.65(COCH<sub>3</sub>), 62.04, 62.44, 68.85, 70.32, 71.29, 72.33, 72.43, 73.60, 74.52, 76.00, 89.29(C-1), 101.34(C-1'), 169.63(COCH<sub>3</sub>), 169.85(COCH<sub>3</sub>), 169.96(COCH<sub>3</sub>), 170.15(COCH<sub>3</sub>), 170.29(COCH<sub>3</sub>), 170.68(COCH<sub>3</sub>), and 170.70(COCH<sub>3</sub>).

*4-nitrophenyl 2,3,6,2',3',4',6'-hepta-O-acetyl- $\beta$ -D-cellobiose (15)*. – This compound was prepared according to a procedure of Roy, Tropper, and Romanowska (Roy, Tropper and Romanowska 1992), with minor modifications. 4-nitrophenol (402.8 mg, 2.9 mmol), **14** (993.2 mg, 1.4 mmol), and tetrabutylammonium hydrogen sulfate (501.1 mg, 1.5 mmol) was added to a two-phase mixture of 1 M sodium hydroxide (10 mL) and methylene chloride (10 mL). The mixture was gently warmed to dissolve the phase-transfer catalyst and allowed to stir for two hours at room temperature. The organic layer was separated, washed with 1 M sodium hydroxide and water, then dried with sodium sulfate, filtered, and concentrated to dryness under vacuum. The resulting material was purified by flash chromatography (1:1 ethyl acetate-hexane + 0.5% isopropyl alcohol) to produce the title compound as an oil, which crystallized upon standing (208.1 mg, 19%).  $\delta_{\text{H}}$ (acetone- $d_6$ ) 1.89(s, COCH<sub>3</sub>); 1.94(s, COCH<sub>3</sub>); 1.98(s, COCH<sub>3</sub>); 2.00(s, COCH<sub>3</sub>); 2.02(s, COCH<sub>3</sub>); 2.06(s, COCH<sub>3</sub>); 3.98(1H, ddd); 4.02(1H, t); 4.05(1H, dd); 4.13-4.20(2H, m); 4.36(1H,

dd); 4.56(1H, dd); 4.83(1H, dd,  $J_{1',2'}8.06$  Hz,  $J_{2',3'}9.34$  Hz, H-2'); 4.88(1H, d, H-1'); 5.02(1H, t); 5.14(1H, dd,  $J_{1,2}7.88$  Hz,  $J_{2,3}9.71$  Hz, H-2); 5.21(1H, t,  $J_{3',4'}9.53$  Hz, H-3'); 5.30(1H, t,  $J_{3,4}9.16$  Hz, H-3); 5.60(1H, d, H-1); 7.24(2H, m,  $C_6H_4NO_2$ ); 8.22(2H, m,  $C_6H_4NO_2$ );  $\delta_C$ (acetone- $d_6$ ) 20.45-20.70(COCH<sub>3</sub>), 62.42, 62.79, 68.88, 71.87, 72.39, 72.94, 73.57, 73.89, 77.14, 98.27(C-1), 101.31(C-1'), 117.54( $C_6H_4NO_2$ ), 126.47( $C_6H_4NO_2$ ), 143.91( $C_6H_4NO_2$ ), 162.46( $C_6H_4NO_2$ ), 169.62(COCH<sub>3</sub>), 169.78(COCH<sub>3</sub>), 169.87(COCH<sub>3</sub>), 170.15(COCH<sub>3</sub>), 170.27(COCH<sub>3</sub>), 170.69(COCH<sub>3</sub>), and 170.78(COCH<sub>3</sub>).

*4-aminophenyl 2,3,6,2',3',4',6'-hepta-O-acetyl- $\beta$ -D-cellobiose (16)*. – This compound was prepared according to a procedure of Roy and Tropper (Roy and Tropper 1991), with minor modifications. A mixture of **15** (208.1 mg, 275  $\mu$ mol), ammonium formate (89.4 mg, 1.4 mmol), and 5% palladium on carbon (48.4 mg) in methanol (21 mL) was refluxed for 15 minutes, and then filtered through Celite and concentrated to dryness under reduced pressure. The resulting tan solid was dissolved in ethyl acetate and washed with aqueous, saturated sodium bicarbonate (2x) and water (2x). The organic layer was dried with sodium sulfate, filtered, and evaporated to dryness under reduced pressure. Preparative-scale thin-layer chromatography (ethyl acetate) and crystallization from warm ethanol yielded the title compound (106.7 mg, 53%).  $\delta_H$ (acetone- $d_6$ ) 1.91(3H, s, COCH<sub>3</sub>); 1.96(3H, s, COCH<sub>3</sub>); 2.00(6H, s, COCH<sub>3</sub>); 2.01(3H, s, COCH<sub>3</sub>); 2.03(3H, s, COCH<sub>3</sub>); 2.07(3H, s, COCH<sub>3</sub>); 3.95(1H, m); 4.00(1H, ddd); 4.06(1H, dd); 4.20(1H, ddd); 4.37(2H, m); 4.52(1H, dd); 4.83(1H, dd,  $J_{1',2'}8.06$  Hz,  $J_{2',3'}9.34$  Hz, H-2'); 4.87(1H, d, H-1'); 5.02(1H, dd,  $J_{1,2}8.06$  Hz,  $J_{2,3}9.53$  Hz, H-2); 5.03(1H, t); 5.07(1H, d, H-1); 5.22(1H, t); 5.24(1H, m); 6.58(2H, m,  $C_6H_4NH_2$ ); 6.78(2H, m,  $C_6H_4NH_2$ );  $\delta_C$ (acetone- $d_6$ ) 20.46(COCH<sub>3</sub>), 20.53(COCH<sub>3</sub>), 20.61(COCH<sub>3</sub>), 20.62(COCH<sub>3</sub>), 20.70(COCH<sub>3</sub>), 20.73(COCH<sub>3</sub>), 62.44, 63.00, 68.90, 72.36, 72.40, 73.21, 73.42, 73.62, 77.52, 100.84, 101.25, 115.73( $C_6H_4NH_2$ ), 119.22( $C_6H_4NH_2$ ), 145.26( $C_6H_4NH_2$ ), 149.91( $C_6H_4NH_2$ ), 169.62(COCH<sub>3</sub>), 169.81(COCH<sub>3</sub>), 169.88(COCH<sub>3</sub>), 170.11(COCH<sub>3</sub>), 170.27(COCH<sub>3</sub>), 170.68(COCH<sub>3</sub>), and 170.77(COCH<sub>3</sub>).

*4-aminophenyl β-D-cellobiose (17)*. – Sodium methoxide (18.6 mg) was added to a solution of **16** (106.7 mg, 147 μmol) in dry methanol (10 mL) and acetone (2 mL), which was then stirred overnight. The solution was neutralized with cation exchange resin (Dowex HCRW2, H<sup>+</sup>), filtered, and concentrated to dryness under reduced pressure, and then under high vacuum. The resulting tan oil was dissolved in deuterium oxide (1 mL) and lyophilized twice under high vacuum before NMR analysis (96.4 mg). An amber contaminant obtained from the cation exchange resin interfered with the calculation of this and successive yields. δ<sub>H</sub>(deuterium oxide) 3.32(1H, dd); 3.42(1H, t); 3.48-3.57(3H, m); 3.58-3.76(4H, m); 3.83(1H, dd); 3.95(2H, ddd); 4.53(1H, d, H-1'); 4.98(1H, d, H-1); 6.82(2H, m, C<sub>6</sub>H<sub>4</sub>NH<sub>2</sub>); 6.99(2H, m, C<sub>6</sub>H<sub>4</sub>NH<sub>2</sub>); 8.44(s, C<sub>6</sub>H<sub>4</sub>NH<sub>2</sub>); δ<sub>C</sub>(deuterium oxide) 60.53, 61.23, 70.12, 73.42, 73.83, 74.82, 75.54, 76.16, 76.66, 79.05, 101.87, 103.24, 118.18(C<sub>6</sub>H<sub>4</sub>NH<sub>2</sub>), 118.86(C<sub>6</sub>H<sub>4</sub>NH<sub>2</sub>), 139.63(C<sub>6</sub>H<sub>4</sub>NH<sub>2</sub>), and 150.71(C<sub>6</sub>H<sub>4</sub>NH<sub>2</sub>).

*N-(4-phenyl β-D-cellobiose) 6-(fluorescein-5-carboxamido)hexanamide (18)*. – 6-(Fluorescein-5-Carboxamido)Hexanoic Acid Succinimidyl Ester (5 mg, 8.5 μmol, Molecular Probes F-6106) was dissolved in methanol (250 μL) and added to a solution of **17** (9.1 mg, 21 μmol) in 0.2 M sodium bicarbonate buffer (250 μL, pH 8.3). This mixture was stirred in the dark for two days, frozen, and lyophilized under high vacuum. The resulting yellow material was purified by preparative-scale thin-layer chromatography (1:1 acetonitrile-water). The product band was extracted with water (10 mL, 5x), concentrated, de-salted with a Bond Elut C18 solid-phase extraction cartridge, concentrated again, and dried under high vacuum to give the FCHASE derivative of **17** (1.7 mg). The resulting yellow solid was dissolved in deuterium oxide (1 mL) and lyophilized twice under high vacuum before NMR analysis. δ<sub>H</sub>(deuterium oxide) 1.44-1.50(2H, m); 1.68-1.80(4H, m); 2.42(2H, t); 3.24(1H, dd); 3.28-3.57(11H, m); 3.73-3.80(2H, m); 3.97(1H, dd); 4.38(1H, d, J<sub>1',2'</sub>7.88 Hz, H-1'); 4.56(1H, d, J<sub>1,2</sub>7.88 Hz, H-1); 6.68(1H, dd, fluorescein); 6.74-6.77(3H, m, fluorescein); 6.84(2H, m, C<sub>6</sub>H<sub>4</sub>NH); 7.00-7.03(1H, m, fluorescein); 7.09(1H, d, fluorescein); 7.14(1H, d, fluorescein); 7.27(2H, m, C<sub>6</sub>H<sub>4</sub>NH); 7.47(1H, dd, fluorescein); 7.85(m); 8.11(1H, d, fluorescein); 8.44(s).

## APPENDIX

### Common Buffers and Solutions

#### *Extraction Buffer (for cell lysis)*

50 mM Tris (free base, pH 7.8)  
10 mM MgCl<sub>2</sub>  
20 mM KCl

#### *Protease Inhibitor Cocktail (1X, prepared as 200X)*

1 μM leupeptin  
0.07 μM benzamidine hydrochloride (BZA)  
0.1 μM DTT  
1 mM tetrasodium ethylenediamine tetraacetate (EDTA)  
125 μM Tris-HCl (pH 7.2)  
25 μM diisopropyl fluorophosphate (DFP)  
50 μM phenylmethylsulfonyl fluoride (PMSF)

#### *SDS-PAGE Sample Loading Buffer (2X)*

50 mM Tris (free base, pH 6.8)  
2% (w/v) SDS  
10% (v/v) glycerol  
0.025% (w/v) bromophenol blue

#### *SDS Electrophoresis Buffer (1X, prepared as 5X)*

25 mM Tris (free base)  
192 mM glycine  
0.1% (w/v) SDS

#### *Lysis Buffer - Surfactants\**

5 M urea  
2 M thiourea  
2% (w/v) Triton X-100  
2% (w/v) SB 3-10  
40 mM Tris (free base)

*Lysis Buffer - Chaotropes*\*

7 M urea  
2 M thiourea  
4% (w/v) Triton X-100  
40 mM Tris (free base)

*Rehydration Buffer - Surfactants*\*

5 M urea  
2 M thiourea  
2% (w/v) Triton X-100  
2% (w/v) SB 3-10  
40 mM Tris (free base)  
a trace of bromophenol blue

*Rehydration Buffer - Chaotropes*\*

7 M urea  
2 M thiourea  
4% (w/v) Triton X-100  
40 mM Tris (free base)  
a trace of bromophenol blue

*SDS Equilibration Buffer*

50 mM Tris-HCl  
6 M urea  
30% (v/v) glycerol  
2% (w/v) SDS  
a trace of bromophenol blue

*Agarose Sealing Solution*

0.5% agarose  
a trace of bromophenol blue  
(in 1X SDS Electrophoresis Buffer, instead of water)

\*2 mM TBP and 0.5% (v/v) IPG Buffer (carrier ampholytes) added before IEF procedures

## REFERENCES

- Agard, D. (1993) "To fold or not to fold," *Science* **260**: 1903-1904.
- Aguilar-Uscanga, B., and J. Francois. (2003) "A study of the yeast cell wall composition and structure in response to growth conditions and mode of cultivation," *Lett. Appl. Microbiol.* **37**: 268-274.
- Altschul, S., T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res.* **25**: 3389-3402.
- Angeloni, S., and M. Potts. (1986) "Polysome turnover in immobilized cells of *Nostoc commune* (cyanobacteria) exposed to water stress," *J. Bacteriol.* **168**: 1036-1039.
- Araki, Y., T. Nakatani, K. Nakayama, and E. Ito. (1972) "Occurrence of *N*-nonsubstituted glucosamine residues in peptidoglycan of lysozyme-resistant cell walls from *Bacillus cereus*," *J. Biol. Chem.* **247**: 6312-6322.
- Arranz, S., I. Albertali, and M. Cabada. (1997) "*Bufo arenarum* egg jelly coat: purification and characterization of two highly glycosylated proteins," *Biochem. J.* **323**: 307-312.
- Benz, I., and M. Schmidt. (2001) "Glycosylation with heptose residues mediated by the *aah* gene product is essential for adherence of the AIDA-I adhesin," *Mol. Microbiol.* **40**: 1403-1413.
- Benz, R., and H. Bohme. (1985) "Pore formation by an outer membrane protein of the cyanobacterium *Anabaena variabilis*," *Biochim. Biophys. Acta* **812**: 286-292.
- Beveridge, T., P. Pouwels, M. Sára, A. Kotiranta, K. Lounatmaa, K. Kari, E. Kerosuo, M. Haapasalo, E. Egelseer, I. Schocher, U. Sleytr, L. Morelli, M. Callegari, J. Nomellini, W. Bingle, J. Smit, E. Leibovitz, M. Lemaire, I. Miras, S. Salamiou, P. Beguin, H. Ohayon, P. Gounon, M. Matuschek, and S. Koval. (1997) "Functions of S-layers," *FEMS Microbiol. Rev.* **20**: 99-149.
- Billi, D., E. Friedmann, K. Hofer, M. Caiola, and R. Ocampo-Friedmann. (2000a) "Ionizing-radiation resistance in the desiccation-tolerant cyanobacterium *Chroococcidiopsis*," *Appl. Environ. Microbiol.* **66**: 1489-1492.
- Billi, D., D. Wright, R. Helm, T. Prickett, M. Potts, and J. Crowe. (2000b) "Engineering desiccation tolerance in *Escherichia coli*," *Appl. Environ. Microbiol.* **66**: 1680-1684.

- Bloom, F., P. Price, G. Lao, J. Xia, J. Crowe, J. Battista, R. Helm, S. Slaughter, and M. Potts. (2001) "Engineering mammalian cells for solid-state sensor applications," *Biosens. Bioelectron.* **16**: 603-608.
- Böhm, G., W. Pfeleiderer, P. Böger, and S. Scherer. (1995) "Structure of a novel oligosaccharide-mycosporine-amino acid ultraviolet A/B sunscreen pigment from the terrestrial cyanobacterium *Nostoc commune*," *J. Biol. Chem.* **270**: 8536-8539.
- Bradford, M. (1976) "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding," *Anal. Biochem.* **72**: 248-254.
- Brasier, M., O. Green, A. Jephcoat, A. Kleppe, M.V. Kranendonk, J. Lindsay, A. Steele, and N. Grassineau. (2002) "Questioning the evidence for Earth's oldest fossils," *Science* **416**: 76-81.
- Brechtel, E., and H. Bahl. (1999) "In *Thermoanaerobacterium thermosulfurigenes* EM1 S-layer homology domains do not attach to peptidoglycan," *J. Bacteriol.* **181**: 5017-5023.
- Breton, C., E. Bettler, D. Joziasse, R. Geremia, and A. Imberty. (1998) "Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases," *J. Biochem. (Tokyo)* **123**: 1000-1009.
- Breton, C., and A. Imberty. (1999) "Structure/function studies of glycosyltransferases," *Curr. Opin. Struct. Biol.* **9**: 563-71.
- Bryant, D., ed. The molecular biology of cyanobacteria. Dordrecht: Kluwer Academic Publishers, 1994.
- Burkey, K., and J. Mathis. (1998) "Identification of a novel isoform of the chloroplast-coupling factor  $\alpha$ -subunit," *Plant Physiol.* **116**: 703-708.
- Caiola, M., D. Billi, and E. Friedmann. (1996) "Effect of desiccation on envelopes of the cyanobacterium *Chroococcidiopsis* sp. (*Chroococcales*)," *Eur. J. Phycol.* **31**: 97-105.
- Callahan, M., L. Xiong, and B. Caughey. (2001) "Reversibility of scrapie-associated prion protein aggregation," *J. Biol. Chem.* **276**: 28022-28028.
- Campbell, J., G. Davies, V. Bulone, and B. Henrissat. (1997) "A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities," *Biochem. J.* **326**: 929-939.
- Castner, D., and B. Ratner. (2002) "Biomedical surface science: foundations to frontiers," *Surf. Sci.* **500**: 28-60.

- Catling, D., K. Zahnle, and C. McKay. (2001) "Biogenic methane, hydrogen escape, and the irreversible oxidation of early Earth," *Science* **293**: 839-843.
- Chauvaux, S., M. Matuschek, and P. Beguin. (1999) "Distinct affinity of binding sites for S-layer homologous domains in *Clostridium thermocellum* and *Bacillus anthracis* cell envelopes," *J. Bacteriol.* **181**: 2455-2458.
- Chen, F., T. Dobashi, and R. Evangelista. (1998) "Quantitative analysis of sugar constituents of glycoproteins by capillary electrophoresis," *Glycobiology* **8**: 1045-1052.
- Chen, F., and R. Evangelista. (1995) "Analysis of mono- and oligosaccharide isomers derivatized with 9-aminopyrene-1,4,6-trisulfonate by capillary electrophoresis with laser-induced fluorescence," *Anal. Biochem.* **230**: 273-280.
- Chien, A., D. Edgar, and J. Trela. (1976) "Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*," *J. Bacteriol.* **127**: 1550-1557.
- Contour, M.-O., J. Defaye, M. Little, and E. Wong. (1989) "Zirconium(IV) chloride-catalyzed synthesis of 1,2-*trans*-1-thioglycopyranosides," *Carbohydr. Res.* **193**: 283-287.
- Creagh, A., E. Ong, E. Jarvis, D. Kilburn, and C. Haynes. (1996) "Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven," *Proc. Natl. Acad. Sci. U.S.A.* **93**: 12229-12234.
- Crowe, J., L. Crowe, and S. Jackson. (1983) "Preservation of structural and functional activity in lyophilized sarcoplasmic reticulum," *Arch. Biochem. Biophys.* **220**: 477-484.
- Crowe, J., F. Hoekstra, and L. Crowe. (1992) "Anhydrobiosis," *Annu. Rev. Physiol.* **54**: 579-599.
- DeAngelis, P. (2002) "Microbial glycosaminoglycan glycosyltransferases," *Glycobiology* **12**: 9R-16R.
- Delmer, D., and S. Read. (1987) "Inhibition of mung bean UDP-glucose:(1-3)- $\beta$ -glucan synthase by UDP-pyridoxal," *Plant Physiol.* **85**: 1008-1015.
- DePhilippis, R., and M. Vincenzini. (1998) "Exocellular polysaccharides from cyanobacteria and their possible application," *FEMS Microbiol. Rev.* **22**: 151-175.
- Deyl, Z., I. Mikik, and F. Tagliaro. (1998) "Advances in capillary electrophoresis," *Forensic Sci. Int.* **92**: 89-124.



- Dickson, D. (2002) "Misfolded, protease-resistant proteins in animal models and human neurodegenerative disease," *J. Clin. Invest.* **110**: 1403-1405.
- Douglas, K., G. Devaud, and N. Clark. (1992) "Transfer of biologically derived nanometer-scale patterns to smooth substrates," *Science* **257**: 642-644.
- Dubray, G., and G. Bezard. (1982) "A highly sensitive periodic acid-silver stain for 1,2-diol groups of glycoproteins and polysaccharides in polyacrylamide gels," *Anal. Biochem.* **119**: 325-329.
- Ehling-Schulz, M., and S. Scherer. (1999) "UV protection in cyanobacteria," *Eur. J. Phycol.* **34**: 329-338.
- Ehling-Schulz, M., S. Schulz, R. Wait, A. Görg, and S. Scherer. (2002) "The UV-B stimulon of the terrestrial cyanobacterium *Nostoc commune* comprises early shock proteins and late acclimation proteins," *Mol. Microbiol.* **46**: 827-843.
- Eichler, J. (2001a) "Biotechnological uses of archaeal extremozymes," *Biotechnol. Adv.* **19**: 261-278.
- Eichler, J. (2001b) "Post-translational modification of the S-layer glycoprotein occurs following translocation across the plasma membrane of the haloarchaeon *Haloferax volcanii*," *Eur. J. Biochem.* **268**: 4366-4373.
- Engelhardt, H., and J. Peters. (1998) "Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer-cell wall interactions," *J. Struct. Biol.* **124**: 276-302.
- Evangelista, R., A. Guttman, and F. Chen. (1996) "Acid-catalyzed reductive amination of aldoses with 8-aminopyrene-1,3,6-trisulfonate," *Electrophoresis* **17**: 347-351.
- Faik, A., M. Bar-Peled, A. DeRocher, W. Zeng, R. Perrin, C. Wilkerson, N. Raikhel, and K. Keegstra. (2000) "Biochemical characterization and molecular cloning of an  $\alpha$ -1,2-fucosyltransferase that catalyzes the last step of cell wall xyloglucan biosynthesis in pea," *J. Biol. Chem.* **275**: 15082-15089.
- Ficarro, S., M. McClelland, P. Stukenberg, D. Burke, M. Ross, J. Shabanowitz, D. Hunt, and F. White. (2002) "Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*," *Nat. Biotechnol.* **20**: 301-305.
- Frost, D., S. Read, R. Drake, B. Haley, and B. Wasserman. (1990) "Identification of the UDP-glucose-binding polypeptide of callose synthase from *Beta vulgaris* L. by photoaffinity labeling with 5-azido-UDP-glucose," *J. Biol. Chem.* **265**: 2162-2167.
- Garcia-Pichel, F., and R. Castenholtz. (1993) "Occurrence of UV-absorbing, mycosporine-like compounds among cyanobacterial isolates and an estimate of their screening capacity," *Appl. Environ. Microbiol.* **59**: 163-169.

- Garfin, D. (2003) "Two-dimensional gel electrophoresis: an overview," *TrAC Trend. Anal. Chem.* **22**: 263-272.
- Germond, J., M. Delley, N. D'Amico, and S. Vincent. (2001) "Heterologous expression and characterization of the exopolysaccharide from *Streptococcus thermophilus* Sfi39," *Eur. J. Biochem.* **268**: 5149-5156.
- Gilbert, M., J. Brisson, M. Karwaski, J. Michniewicz, A. Cunningham, Y. Wu, N. Young, and W. Wakarchuk. (2000) "Biosynthesis of ganglioside mimics in *Campylobacter jejuni* OH4384. Identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-MHz  $^1\text{H}$  and  $^{13}\text{C}$  NMR analysis," *J. Biol. Chem.* **275**: 3896-3906.
- Gilbert, M., D. Watson, A. Cunningham, M. Jennings, N. Young, and W. Wakarchuk. (1996) "Cloning of the lipooligosaccharide  $\alpha$ -2,3-sialyltransferase from the bacterial pathogens *Neisseria meningitidis* and *Neisseria gonorrhoeae*," *J. Biol. Chem.* **271**: 28271-28276.
- Goldberg, M., N. Expert-Bezancon, L. Vuillard, and T. Rabilloud. (1995) "Non-detergent sulphobetaines: a new class of molecules that facilitate in vitro protein renaturation," *Fold. Des.* **1**: 21-27.
- Gotschlich, E. (1994) "Genetic locus for the biosynthesis of the variable portion of *Neisseria gonorrhoeae* lipooligosaccharide," *J. Exp. Med.* **180**: 2181-2190.
- Guerrero, R. (1998) "Crucial crises in biology: life in the deep biosphere," *Inter. Microbiol.* **1**: 285-294.
- Guo, N., I. Puhlev, D. Brown, J. Mansbridge, and F. Levine. (2000) "Trehalose expression confers desiccation tolerance on human cells," *Nat. Biotechnol.* **18**: 168-171.
- Håkansson, K., H. Cooper, M. Emmett, C. Costello, A. Marshall, and C. Nilsson. (2001) "Electron capture dissociation and infrared multiphoton dissociation MS/MS of an *N*-glycosylated tryptic peptic to yield complementary sequence information," *Anal. Chem.* **73**: 4530-4536.
- Hancock, R., and F. Brinkman. (2002) "Function of *Pseudomonas* porins in uptake and efflux," *Annu. Rev. Microbiol.* **56**: 17-38.
- Hansel, A., F. Pattus, U. Jürgens, and M. Tadros. (1998) "Cloning and characterization of the genes coding for two porins in the unicellular cyanobacterium *Synechococcus* PCC 6301," *Biochim. Biophys. Acta* **1399**: 31-39.
- Hansel, A., A. Schmid, M. Tadros, and U. Jürgens. (1994) "Isolation and characterisation of porin from the outer membrane of *Synechococcus* PCC 6301," *Arch. Microbiol.* **161**: 163-167.

- Hansel, A., and M. Tadros. (1998) "Characterization of two pore-forming proteins isolated from the outer membrane of *Synechococcus* PCC 6301," *Curr. Microbiol.* **36**: 321-326.
- Harvey, D. (2001) "Identification of protein-bound carbohydrates by mass spectrometry," *Proteomics* **1**: 311-328.
- Hedges, S., H. Chen, S. Kumar, D. Yang, A. Thompson, and H. Watanabe. (2001) "A genomic timescale for the origin of eukaryotes," *BMC Evol. Biol.* **1**: 1-10.
- Helm, R., Z. Huang, D. Edwards, H. Leeson, W. Peery, and M. Potts. (2000) "Structural characterization of the released polysaccharide of desiccation-tolerant *Nostoc commune* DRH-1," *J. Bacteriol.* **182**: 974-982.
- Henderson, I., F. Navarro-Garcia, and J. Nataro. (1998) "The great escape: structure and function of the autotransporter proteins," *Trends Microbiol.* **6**: 370-378.
- Henson, B., L. Watson, and S. Barnum. (2002) "Molecular differentiation of the heterocystous cyanobacteria, *Nostoc* and *Anabaena*, based on complete *nifD* sequences," *Curr. Microbiol.* **45**: 161-164.
- Herbert, B. (1999) "Advances in protein solubilisation for two-dimensional electrophoresis," *Electrophoresis* **20**: 660-3.
- Herbert, B., M. Molloy, A. Gooley, B. Walsh, W. Bryson, and K. Williams. (1998) "Improved protein solubility in two-dimensional electrophoresis using tributyl phosphine as reducing agent," *Electrophoresis* **19**: 845-851.
- Hill, D., S. Hladun, S. Scherer, and M. Potts. (1994a) "Water stress proteins of *Nostoc commune* (cyanobacteria) are secreted with UV-A/B-absorbing pigments and associate with 1,4- $\beta$ -D-xylanxylanohydrolase activity," *J. Biol. Chem.* **269**: 7726-7734.
- Hill, D., T. Keenan, R. Helm, M. Potts, L. Crowe, and J. Crowe. (1997) "Extracellular polysaccharide of *Nostoc commune* (cyanobacteria) inhibits fusion of membrane vesicles during desiccation," *J. Appl. Phycol.* **9**: 237-248.
- Hill, D., A. Peat, and M. Potts. (1994b) "Biochemistry and structure of the glycan secreted by desiccation-tolerant *Nostoc commune* (cyanobacteria)," *Protoplasma* **182**: 126-148.
- Hoekstra, F., E. Golovina, and J. Buitink. (2001) "Mechanisms of plant desiccation tolerance," *Trends Plant Sci.* **6**: 431-438.
- Hoiczuk, E., and A. Hansel. (2000) "Cyanobacterial cell walls: news from an unusual prokaryotic envelope," *J. Bacteriol.* **182**: 1191-1199.

- Huang, Y., Y. Mechref, and M. Novotny. (2001) "Microscale nonreductive release of O-linked glycans for subsequent analysis through MALDI mass spectrometry and capillary electrophoresis," *Anal. Chem.* **73**: 6063-6069.
- Huang, Z., Y. Liu, B. Paulsen, and D. Klaveness. (1998) "Studies on polysaccharides from three edible species of *Nostoc* (cyanobacteria) with different colony morphologies: comparison of monosaccharide compositions and viscosities of polysaccharides from field colonies and suspension cultures," *J. Phycol.* **34**: 962-968.
- Huddleston, M., M. Bean, and S. Carr. (1993) "Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests," *Anal. Chem.* **65**: 877-884.
- Ilk, N., P. Kosma, M. Puchberger, E. Egelseer, H. Mayer, U. Sleytr, and M. Sára. (1999) "Structural and functional analyses of the secondary cell wall polymer of *Bacillus sphaericus* CCM 2177 that serves as an S-layer-specific anchor," *J. Bacteriol.* **181**: 7643-7646.
- Jahn-Schmid, B., M. Graninger, M. Glozik, S. Kupcu, C. Ebner, F. Unger, U. Sleytr, and P. Messner. (1996a) "Immunoreactivity of allergen (Bet v 1) conjugated to crystalline bacterial cell surface layers (S-layers)," *Immunotechnology* **2**: 103-113.
- Jahn-Schmid, B., P. Messner, F. Unger, U. Sleytr, O. Scheiner, and D. Kraft. (1996b) "Toward selective elicitation of T<sub>H</sub>1-controlled vaccination responses: vaccine applications of bacterial surface layer proteins," *J. Biotechnol.* **44**: 225-231.
- Jawad, A., A. Snelling, J. Heritage, and P. Hawkey. (1998) "Exceptional desiccation tolerance of *Acinetobacter radioresistens*," *J. Hosp. Infect.* **39**: 235-240.
- Jeanteur, D., J. Lakey, and F. Pattus. (1991) "The bacterial porin superfamily: sequence alignment and structure prediction," *Mol. Microbiol.* **5**: 2153-2164.
- Jennings, M., D. Hood, I. Peak, M. Virji, and E. Moxon. (1995) "Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*," *Mol. Microbiol.* **18**: 729-740.
- Jolly, L., J. Newell, I. Porcelli, S. Vincent, and F. Stingle. (2002) "*Lactobacillus helveticus* glycosyltransferases: from genes to carbohydrate synthesis," *Glycobiology* **12**: 319-327.
- Jonietz, E. "Glycomics: sugars could be biology's next sweet spot." *Technol. Rev.* October 2001: 33.

- Jürgens, U., and T. Burger-Wiersma. (1989) "Peptidoglycan-polysaccharide complex in the cell wall of the filamentous prochlorophyte *Prochlorothrix hollandica*," J. Bacteriol. **171**: 498-502.
- Jürgens, U., G. Drews, and J. Weckesser. (1983) "Primary structure of the peptidoglycan from the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6714," J. Bacteriol. **154**: 471-478.
- Jürgens, U., C. Martin, and J. Weckesser. (1989) "Cell wall constituents of *Microcystis* sp. PCC 7806," FEMS Microbiol. Lett. **53**: 47-51.
- Jürgens, U., and J. Weckesser. (1985a) "Carotenoid-containing outer membrane of *Synechocystis* sp. strain PCC 6714," J. Bacteriol. **164**: 384-389.
- Jürgens, U., and J. Weckesser. (1985b) "The fine structure and chemical composition of the cell wall and sheath layers of cyanobacteria," Ann. Inst. Pasteur/Microbiol. **136**: 41-44.
- Jürgens, U., and J. Weckesser. (1986) "Polysaccharide covalently linked to the peptidoglycan of the cyanobacterium *Synechocystis* sp. strain PCC 6714," J. Bacteriol. **168**: 568-573.
- Kaneko, T., Y. Nakamura, C.P. Wolk, T. Kuritz, S. Sasamoto, A. Watanabe, M. Iriguchi, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, M. Kohara, M. Matsumoto, A. Matsuno, A. Muraki, N. Nakazaki, S. Shimpo, M. Sugimoto, M. Takazawa, M. Yamada, M. Yasuda, and S. Tabata. (2001) "Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120," DNA Res. **8**: 205-213.
- Kapitonov, D., and R. Yu. (1999) "Conserved domains of glycosyltransferases," Glycobiology **9**: 961-978.
- Kartha, K., and H. Jennings. (1990) "A simplified, one-pot preparation of acetobromosugars from reducing sugars," J. Carbohydr. Chem. **9**: 777-781.
- Kawasaki, N., S. Itoh, M. Ohta, and T. Hayakawa. (2003) "Microanalysis of *N*-linked oligosaccharides in a glycoprotein by capillary liquid chromatography/mass spectrometry and liquid chromatography/tandem mass spectrometry," Anal. Biochem. **316**: 15-22.
- Kolkman, M., B. van der Zeijst, and P. Nuijten. (1997) "Functional analysis of glycosyltransferases encoded by the capsular polysaccharide biosynthesis locus of *Streptococcus pneumoniae* serotype 14," J. Biol. Chem. **272**: 19502-19508.
- Kotani, H., T. Kaneko, T. Matsubayashi, S. Sato, M. Sugiura, and S. Tabata. (1994) "A physical map of the genome of a unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803," DNA Res. **1**: 303-307.

- Kumar, R., J. Yang, R. Larsen, and P. Stanley. (1990) "Cloning and expression of *N*-acetylglucosaminyltransferase I, the medial golgi transferase that initiates complex *N*-linked carbohydrate formation," *Proc. Natl. Acad. Sci. U.S.A.* **87**: 9948-9952.
- Laarmann, S., and M. Schmidt. (2003) "The *Escherichia coli* AIDA autotransporter adhesin recognizes an integral membrane glycoprotein as receptor," *Microbiology* **149**: 1871-1882.
- Laemmli, U. (1970) "Cleavage of structural proteins during the assembly of the head of bacteriophage T4," *Nature* **227**: 680-685.
- Leiter, H., J. Mucha, E. Staudacher, R. Grimm, J. Glossl, and F. Altmann. (1999) "Purification, cDNA cloning, and expression of GDP-L-Fuc:Asn-linked GlcNAc  $\alpha$ -1,3-fucosyltransferase from mung beans," *J. Biol. Chem.* **274**: 21830-21839.
- Lemaire, M., H. Ohayon, P. Gounon, T. Fujino, and P. Beguin. (1995) "OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domains to components of the cell envelope," *J. Bacteriol.* **177**: 2451-2459.
- Leslie, S., E. Israeli, B. Lighthart, J. Crowe, and L. Crowe. (1995) "Trehalose and sucrose protect both membranes and proteins in intact bacteria during drying," *Appl. Environ. Microbiol.* **61**: 3592-3597.
- Leslie, S., S. Teter, L. Crowe, and J. Crowe. (1994) "Trehalose lowers membrane phase transitions in dry yeast cells," *Biochim. Biophys. Acta* **1192**: 7-13.
- Levander, F., M. Svensson, and P. Radstrom. (2002) "Enhanced Exopolysaccharide Production by Metabolic Engineering of *Streptococcus thermophilus*," *Appl. Environ. Microbiol.* **68**: 784-790.
- Li, T., X. Huang, R. Zhou, Y. Liu, B. Li, C. Nomura, and J. Zhao. (2002) "Differential expression and localization of Mn and Fe superoxide dismutases in the heterocystous cyanobacterium *Anabaena* sp. strain PCC 7120," *J. Bacteriol.* **184**: 5096-5103.
- Lougheed, B., H. Ly, W. Wakarchuk, and S. Withers. (1999) "Glycosyl fluorides can function as substrates for nucleotide phosphosugar-dependent glycosyltransferases," *J. Biol. Chem.* **274**: 37717-37722.
- Loveless, B., and M. Saier, Jr. (1997) "A novel family of channel-forming, autotransporting, bacterial virulence factors," *Mol. Membr. Biol.* **14**: 113-123.
- Lugtenberg, B., and L. Van Alphen. (1983) "Molecular architecture and functioning of the outer membrane of *Escherichia coli* and other gram-negative bacteria," *Biochim. Biophys. Acta* **737**: 51-115.

- Lupas, A., H. Engelhardt, J. Peters, U. Santarius, S. Volker, and W. Baumeister. (1994) "Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis," *J. Bacteriol.* **176**: 1224-1233.
- Mackey, A., T. Haystead, and W. Pearson. (2002) "Getting more from less: algorithms for rapid protein identification with multiple short peptide sequences," *Mol. Cell. Prot.* **1**: 139-147.
- Mader, C., S. Kupcu, U. Sleytr, and M. Sára. (2000) "S-layer-coated liposomes as a versatile system for entrapping and binding target molecules," *Biochim. Biophys. Acta* **1463**: 142-150.
- Madigan, M., and B. Mairs. (1997) "Extremophiles," *Sci. Am.* **276**: 82-87.
- Maione, T., and A. Jagendorf. (1984) "Partial deglycosylation of chloroplast coupling factor 1 (CF1) prevents the reconstitution of photophosphorylation," *Proc. Natl. Acad. Sci. U.S.A.* **81**: 3733-3736.
- Margulis, L. (1996) "Archaeal-Eubacterial mergers in the origin of Eukarya: phylogenetic classification of life," *Proc. Natl. Acad. Sci. U.S.A.* **93**: 1071-1076.
- Markillie, L., S. Varnum, P. Hradecky, and K.-K. Wong. (1999) "Targeted mutagenesis by duplication insertion in the radioresistant bacterium *Deinococcus radiodurans*: radiation sensitivities of catalase (*kataA*) and superoxide dismutase (*sodA*) mutants," *J. Bacteriol.* **181**: 666-669.
- Martin, W., M. Hoffmeister, C. Rotte, and K. Henze. (2001) "An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle," *Biol. Chem.* **382**: 1521-1539.
- Mattimore, V., and J. Battista. (1996) "Radioresistance of *Deinococcus radiodurans*: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation," *J. Bacteriol.* **178**: 633-637.
- Mechref, Y., and M. Novotny. (2002) "Structural investigations of glycoconjugates at high sensitivity," *Chem. Rev.* **102**: 321-369.
- Mescher, M., and J. Strominger. (1976) "Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarum*," *J. Biol. Chem.* **251**: 2005-2014.
- Messner, P., K. Bock, R. Christian, G. Schulz, and U. Sleytr. (1990) "Characterization of the surface layer glycoprotein of *Clostridium symbiosum* HB25," *J. Bacteriol.* **172**: 2576-2583.
- Moens, S., and J. Vanderleyden. (1997) "Glycoproteins in prokaryotes," *Arch. Microbiol.* **168**: 169-175.

- Molloy, M. (2000) "Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients," *Anal. Biochem.* **280**: 1-10.
- Morona, R., L. van den Bosch, and P. Manning. (1995) "Molecular, genetic, and topological characterization of *O*-antigen chain length regulation in *Shigella flexneri*," *J. Bacteriol.* **177**: 1059-1068.
- Murry, M., and C. Wolk. (1989) "Evidence that the barrier to the penetration of oxygen into heterocysts depends upon two layers of the cell envelope," *Arch. Microbiol.* **151**: 469-474.
- Nisbet, E., and N. Sleep. (2001) "The habitat and nature of early life," *Nature* **409**: 1083-1091.
- Olie, J., and M. Potts. (1986) "Purification and biochemical analysis of the cytoplasmic membrane from the desiccation-tolerant cyanobacterium *Nostoc commune* UTEX 584," *Appl. Environ. Microbiol.* **52**: 706-710.
- Ong, E., D. Kilburn, R. Miller, Jr, and R. Warren. (1994) "*Streptomyces lividans* glycosylates the linker region of a  $\beta$ -1,4-glycanase from *Cellulomonas fimi*," *J. Bacteriol.* **176**: 999-1008.
- Patel, G., and G. Sprott. (1999) "Archaeobacterial ether lipid liposomes (archaeosomes) as novel vaccine and drug delivery systems," *Crit. Rev. Biotechnol.* **19**: 317-357.
- Patton, W. (2002) "Detection technologies in proteome analysis," *J Chromatogr B* **771**: 3-31.
- Paul, G., F. Lottspeich, and F. Wieland. (1986) "Asparaginyln-*N*-acetylgalactosamine. Linkage unit of halobacterial glycosaminoglycan," *J. Biol. Chem.* **261**: 1020-1024.
- Pearson, W., and D. Lipman. (1988) "Improved tools for biological sequence comparison," *Proc. Natl. Acad. Sci. U.S.A.* **85**: 2444-2448.
- Perrin, R., A. DeRocher, M. Bar-Peled, W. Zeng, L. Norambuena, A. Orellana, N. Raikhel, and K. Keegstra. (1999) "Xyloglucan fucosyltransferase, an enzyme involved in plant cell wall biosynthesis," *Science* **284**: 1976-1979.
- Persson, K., H. Ly, M. Dieckelmann, W. Wakarchuk, S. Withers, and N. Strynadka. (2001) "Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs," *Nat. Struct. Biol.* **8**: 166-175.
- Peugnet, I., F. Goubet, M. Bruyant-Vannier, B. Thoiron, C. Morvan, H. Schols, and A. Voragen. (2001) "Solubilization of rhamnogalacturonan I galactosyltransferases from membranes of a flax cell suspension," *Planta* **213**: 435-445.



- Potts, M. (1996) "The anhydrobiotic cyanobacterial cell," *Physiol. Plant.* **97**: 788-794.
- Potts, M. (1994) "Desiccation tolerance of prokaryotes," *Microbiol. Rev.* **58**: 755-805.
- Potts, M. (2001) "Desiccation tolerance: a simple process?," *Trends Microbiol.* **9**: 553-559.
- Rabilloud, T., C. Adessi, A. Giraudel, and J. Lunardi. (1997) "Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients," *Electrophoresis* **18**: 307-316.
- Raju, T., L. Lerner, and J. O'Connor. (1996) "Glycopinoin: biological significance and methods for the analysis of complex carbohydrates of recombinant glycoproteins," *Biotechnol. Appl. Biochem.* **24**: 191-194.
- Raska, C., C. Parker, Z. Dominski, W. Marzluff, G. Glish, R. Pope, and C. Borchers. (2002) "Direct MALDI-MS/MS of phosphopeptides affinity-bound to immobilized metal ion affinity chromatography beads," *Anal. Chem.* **74**: 3429-3433.
- Regelsberger, G., W. Atzenhofer, F. Rüker, G. Peschek, C. Jakopitsch, M. Paumann, P. Furtmüller, and C. Oblinger. (2002) "Biochemical characterization of a membrane-bound manganese-containing superoxide dismutase from the cyanobacterium *Anabaena* PCC 7120," *J. Biol. Chem.* **277**: 43615-43622.
- Resch, C., and J. Gibson. (1983) "Isolation of the carotenoid-containing cell wall of three unicellular cyanobacteria," *J. Bacteriol.* **155**: 345-350.
- Rippka, R., J. Deruelles, J. Waterbury, M. Herdman, and R. Stanier. (1979) "Generic assignments, strain histories and properties of pure cultures of cyanobacteria," *J. Gen. Microbiol.* **111**: 1-61.
- Ritchie, M., A. Gill, M. Deery, and K. Lilley. (2002) "Precursor ion scanning for detection and structural characterization of heterogeneous glycopeptide mixtures," *J. Am. Soc. Mass Spectrom.* **13**: 1065-1077.
- Rolland, N., M. Ferro, D. Seigneurin-Berny, J. Garin, R. Douce, and J. Joyard. (2003) "Proteomics of chloroplast envelope membranes," *Photosynth. Res.* **78**: 205-230.
- Roseman, S. (2001) "Reflections on glycobiology," *J. Biol. Chem.* **276**: 41527-41542.
- Roy, R., and F. Tropper. (1991) "Carbohydrate protein interactions. Synthesis of agglutination inhibitors of wheat germ agglutinin by phase transfer catalysis.," *Can. J. Chem.* **69**: 817-821.
- Roy, R., F. Tropper, and A. Romanowska. (1992) "New strategy in glycopolymer syntheses. Preparation of antigenic water-soluble poly(acrylamide-*co-p*-acrylamidophenyl  $\beta$ -lactoside)," *Bioconjug. Chem.* **3**: 256-261.

- Salvucci, M., A. Chavan, and B. Haley. (1992) "Identification of peptides from the adenine binding domains of ATP and AMP in adenylate kinase: isolation of photoaffinity-labeled peptides by metal chelate chromatography," *Biochemistry* **31**: 4479-4487.
- Santoni, V., M. Molloy, and T. Rabilloud. (2000) "Membrane proteins and proteomics: *un amour impossible?*," *Electrophoresis* **21**: 1054-1070.
- Sára, M., C. Dekitsch, H. Mayer, E. Egelseer, and U. Sleytr. (1998a) "Influence of the secondary cell wall polymer on the reassembly, recrystallization, and stability properties of the S-layer protein from *Bacillus stearothermophilus* PV72/p2," *J. Bacteriol.* **180**: 4146-4153.
- Sára, M., E. Egelseer, C. Dekitsch, and U. Sleytr. (1998b) "Identification of two binding domains, one for peptidoglycan and another for a secondary cell wall polymer, on the N-terminal part of the S-layer protein SbsB from *Bacillus stearothermophilus* PV72/p2," *J. Bacteriol.* **180**: 6780-6783.
- Sauer, J., U. Schreiber, R. Schmid, U. Volker, and K. Forchhammer. (2001) "Nitrogen starvation-induced chlorosis in *Synechococcus* PCC 7942. Low-level photosynthesis as a mechanism of long-term survival," *Plant Physiol.* **126**: 233-243.
- Saxena, I., R. Brown, Jr., M. Fevre, R. Geremia, and B. Henrissat. (1995) "Multidomain architecture of  $\beta$ -glycosyl transferases: implications for mechanism of action," *J. Bacteriol.* **177**: 1419-1424.
- Schaffer, C., M. Graninger, and P. Messner. (2001) "Prokaryotic glycosylation," *Proteomics* **1**: 248-261.
- Schaffer, C., T. Wugeditsch, H. Kahlig, A. Scheberl, S. Zayni, and P. Messner. (2002) "The surface layer (S-layer) glycoprotein of *Geobacillus stearothermophilus* NRS 2004/3a. Analysis of its glycosylation," *J. Biol. Chem.* **277**: 6230-6239.
- Schägger, H., and G. von Jagow. (1987) "Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa," *Anal. Biochem.* **166**: 368-379.
- Scherer, S., T. Chen, and P. Böger. (1988) "A new UV-A/B protecting pigment in the terrestrial cyanobacterium *Nostoc commune*," *Plant Physiol.* **88**: 1055-1057.
- Scherer, S., T. Chen, and P. Böger. (1986) "Recovery of adenine-nucleotide pools in terrestrial blue-green algae after prolonged drought periods," *Oecologia* **68**: 585-588.
- Scherer, S., and M. Potts. (1989) "Novel water stress protein from a desiccation-tolerant cyanobacterium: purification and partial characterization," *J. Biol. Chem.* **264**: 12546-12553.

- Schleifer, K., and O. Kandler. (1972) "Peptidoglycan types of bacterial cell walls and their taxonomic implications," *Bacteriol. Rev.* **36**: 407-477.
- Schleiff, E., L. Eichacker, K. Eckart, T. Becker, O. Mirus, T. Stahl, and J. Soll. (2003) "Prediction of the plant beta-barrel proteome: a case study of the chloroplast outer envelope," *Protein Sci.* **12**: 748-759.
- Schopf, J. (1993) "Microfossils of the early Archean apex chert - New evidence of the antiquity of life," *Science* **260**: 640-646.
- Schopf, J., and B. Packer. (1987) "Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona group, Australia," *Science* **237**: 70-73.
- Schuster, B., D. Pum, M. Sára, O. Braha, H. Bayley, and U. Sleytr. (2001) "S-layer ultrafiltration membranes: a new support for stabilizing functionalized lipid membranes," *Langmuir* **17**: 499-503.
- Shaw, E., D. Hill, N. Brittain, D. Wright, U. Tauber, H. Marand, R. Helm, and M. Potts. (2003) "Unusual water flux in the extracellular polysaccharide of the cyanobacterium *Nostoc commune*," *Appl. Environ. Microbiol.* **69**: 5679-5684.
- Shirkey, B., D. Kovarcik, D. Wright, G. Wilmoth, T. Prickett, R. Helm, E. Gregory, and M. Potts. (2002) "An active Fe-containing superoxide dismutase and abundant *sodF* mRNA in *Nostoc commune* (cyanobacteria) after years of desiccation," *J. Bacteriol.* **182**: 189-197.
- Shirkey, B., N. McMaster, S. Smith, D. Wright, H. Rodriguez, P. Jaruga, M. Birincioglu, R. Helm, and M. Potts. (2003) "Genomic DNA of *Nostoc commune* (cyanobacteria) becomes covalently modified during long-term (decades) desiccation but is protected from oxidative damage and degradation," *Nucleic Acids Res.* **31**: 2995-3005.
- Sinha, R., M. Klisch, and D. Häder. (1999) "Induction of a mycosporine-like amino acid (MAA) in the rice-field cyanobacterium *Anabaena* sp. by UV irradiation," *J. Photoch. Photobio. B* **52**: 59-64.
- Sleytr, U. (1997) "Basic and applied S-layer research: an overview," *FEMS Microbiol. Rev.* **20**: 5-12.
- Sleytr, U., M. Sára, D. Pum, and B. Schuster. (2001) "Characterization and use of crystalline bacterial cell surface layers," *Prog. Surf. Sci.* **68**: 231-278.
- Smarda, J., D. Smajs, J. Komrska, and V. Krzyzanek. (2002) "S-layers on cell walls of cyanobacteria," *Micron* **33**: 257-277.
- Sojar, H., and O. Bahl. (1987) "Chemical deglycosylation of glycoproteins," *Methods Enzymol.* **138**: 341-350.

- Stingele, F., J. Neeser, and B. Mollet. (1996) "Identification and characterization of the eps (exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6," J. Bacteriol. **178**: 1680-1690.
- Stingele, F., J. Newell, and J. Neeser. (1999a) "Unraveling the function of glycosyltransferases in *Streptococcus thermophilus* Sfi6," J. Bacteriol. **181**: 6354-6360.
- Stingele, F., S. Vincent, E. Faber, J. Newell, J. Kamerling, and J. Neeser. (1999b) "Introduction of the exopolysaccharide gene cluster from *Streptococcus thermophilus* Sfi6 into *Lactococcus lactis* MG1363: production and characterization of an altered polysaccharide," Mol. Microbiol. **32**: 1287-1295.
- Takeo, K., Y. Ohguchi, R. Hasegawa, and S. Kitamura. (1995) "Synthesis of 2- and 4-nitrophenyl glycosides of  $\beta$ -(1-4)-D-xylo-oligosaccharides of dp 2-4," Carbohydr. Res. **277**: 231-244.
- Tamas, I., Z. Svircev, and S. Andersson. (2000) "Determinative value of a portion of the *nifH* sequence for the genera *Nostoc* and *Anabaena* (cyanobacteria)," Curr. Microbiol. **41**: 197-200.
- Thomas, D., J. Thomas, S. Prier, N. Nasso, and S. Herbert. (1999) "Iron superoxide dismutase protects against chilling damage in the cyanobacterium *Synechococcus* sp. PCC 7942," Plant Physiol. **120**: 275-282.
- Thornton, C., D. Pitt, G. Wakley, and N. Talbot. (2002) "Production of a monoclonal antibody specific to the genus *Trichoderma* and closely related fungi, and its use to detect *Trichoderma* spp. in naturally infested composts," Microbiology **148**: 1263-1279.
- Tomme, P., R. Warren, and N. Gilkes. (1995) "Cellulose hydrolysis by bacteria and fungi," Adv. Microb. Physiol. **37**: 1-81.
- Tranel, P., J. Froehlich, A. Goyal, and K. Keegstra. (1995) "A component of the chloroplastic protein import apparatus is targeted to the outer envelope membrane via a novel pathway," EMBO J. **14**: 2436-2446.
- Troll, H., D. Malchow, A. Muller-Taubenberger, B. Humbel, F. Lottspeich, M. Ecke, G. Gerisch, A. Schmid, and R. Benz. (1992) "Purification, functional characterization, and cDNA sequencing of mitochondrial porin from *Dictyostelium discoideum*," J. Biol. Chem. **267**: 21072-21079.
- Umeda, H., H. Aiba, and T. Mizuno. (1996) "*somA*, a novel gene that encodes a major outer-membrane protein of *Synechococcus* sp. PCC 7942," Microbiology **142**: 2121-2128.
- Upreti, R., M. Kumar, and V. Shankar. (2003) "Bacterial glycoproteins: Functions, biosynthesis and applications," Proteomics **3**: 363-379.

- van de Rijn, I., and R. Drake. (1992) "Analysis of the streptococcal hyaluronic acid synthase complex using the photoaffinity probe 5-azido-UDP-glucuronic acid," *J. Biol. Chem.* **267**: 24302-24306.
- Van den Burg, B., G. Vriend, O. Veltman, G. Venema, and V. Eijsink. (1998) "Engineering an enzyme to resist boiling," *Proc. Natl. Acad. Sci. U.S.A.* **95**: 2056-2060.
- van Kranenburg, R., I. van Swam, J. Marugg, M. Kleerebezem, and W. de Vos. (1999a) "Exopolysaccharide biosynthesis in *Lactococcus lactis* NIZO B40: functional analysis of the glycosyltransferase genes involved in synthesis of the polysaccharide backbone," *J. Bacteriol.* **181**: 338-340.
- van Kranenburg, R., H. Vos, I. van Swam, M. Kleerebezem, and W. de Vos. (1999b) "Functional analysis of glycosyltransferase genes from *Lactococcus lactis* and other gram-positive cocci: complementation, expression, and diversity," *J. Bacteriol.* **181**: 6347-6353.
- van Montfort, B. "Identification of domain interfaces in the *Escherichia coli* mannitol permease." University of Groningen, 2001.
- van Montfort, B., M. Doeven, B. Canas, L. Veenhoff, B. Poolman, and G. Robillard. (2002) "Combined in-gel tryptic digestion and CNBr cleavage for the generation of peptide maps of an integral membrane protein with MALDI-TOF mass spectrometry," *Biochim. Biophys. Acta* **1555**: 111-115.
- Wakarchuk, W., A. Cunningham, D. Watson, and N. Young. (1998) "Role of paired basic residues in the expression of active recombinant galactosyltransferases from the bacterial pathogen *Neisseria meningitidis*," *Protein Eng.* **11**: 295-302.
- Wakarchuk, W., A. Martin, M. Jennings, E. Moxon, and J. Richards. (1996) "Functional relationships of the genetic locus encoding the glycosyltransferase enzymes involved in expression of the lacto-*N*-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*," *J. Biol. Chem.* **271**: 19166-19173.
- Weckesser, J., A. Katz, G. Drews, H. Mayer, and I. Fromme. (1974) "Lipopolysaccharide containing L-acofriose in the filamentous blue-green alga *Anabaena variabilis*," *J. Bacteriol.* **120**: 672-678.
- Wieland, F., R. Heitzer, and W. Schaefer. (1983) "Asparaginyglucose: novel type of carbohydrate linkage," *Proc. Natl. Acad. Sci. U.S.A.* **80**: 5470-5474.
- Wilm, M., G. Neubauer, and M. Mann. (1996) "Parent ion scans of unseparated peptide mixtures," *Anal. Chem.* **68**: 527-533.
- Woitzik, D., J. Weckesser, and U. Jürgens. (1988) "Isolation and characterization of cell wall components of the unicellular cyanobacterium *Synechococcus* sp. PCC 6307," *J. Gen. Microbiol.* **134**: 619-627.

- Wright, D., T. Prickett, R. Helm, and M. Potts. (2001) "Form species *Nostoc commune* (cyanobacteria)," *Int. J. Syst. Evol. Microbiol.* **51**: 1839-1852.
- Wu, A., R. Harriman, D. Frost, S. Read, and B. Wasserman. (1991) "Rapid enrichment of CHAPS-solubilized UDP-glucose:(1,3)- $\beta$ -glucan (callose) synthase from *Beta vulgaris* L. by product entrapment," *Plant Physiol.* **97**: 684-692.
- Wylie, J., C. Bernegger-Egli, J. O'Neil, and E. Worobec. (1993) "Biophysical characterization of OprB, a glucose-inducible porin of *Pseudomonas aeruginosa*," *J. Bioenerg. Biomembr.* **25**: 547-556.
- Wylie, J., and E. Worobec. (1995) "The OprB porin plays a central role in carbohydrate uptake in *Pseudomonas aeruginosa*," *J. Bacteriol.* **177**: 3021-3026.
- Xu, X., I. Khudyakov, and P. Wolk. (1997) "Lipopolysaccharide dependence of cyanophage sensitivity and aerobic nitrogen fixation in *Anabaena* sp. strain PCC 7120," *J. Bacteriol.* **179**: 2884-2891.
- Yamashiki, N., H. Yokota, M. Sakamoto, and A. Yuasa. (2002) "Presence of phenol UDP-glucuronosyltransferase in bovine alveolar macrophages and bronchial epithelial cells," *Toxicology* **176**: 221-227.
- Zeng, Y., G. Bannon, V. Thomas, K. Rice, R. Drake, and A. Elbein. (1997) "Purification and specificity of  $\beta$ -1,2-xylosyltransferase, an enzyme that contributes to the allergenicity of some plant proteins," *J. Biol. Chem.* **272**: 31340-31347.
- Zhang, C., M. Durand, R. Jeanjean, and F. Joset. (1989) "Molecular and genetical analysis of the fructose-glucose transport system in the cyanobacterium *Synechocystis* PCC 6803," *Mol. Microbiol.* **3**: 1221-1229.
- Zuo, S.-S., and P. Lundahl. (2000) "A micro-Bradford membrane protein assay," *Anal. Biochem.* **284**: 162-164.

## BRIAN R. JORDAN

Department of Biochemistry  
Virginia Tech  
Blacksburg, VA 24061  
(540) 231-3062

143 Oak Creek Lane  
Pilot, VA 24138  
(540) 381-9098  
Email: brjordan@vt.edu

### EDUCATION

---

**Virginia Polytechnic Institute and State University** **Blacksburg, VA**  
*Ph.D. Candidate, Biochemistry (present, anticipated graduation in May 2004)*

Dissertation work in the labs of Drs. Richard Helm and Malcolm Potts, on extracellular polysaccharide production of *Nostoc commune* and protein glycosylation in *Anabaena* PCC 7120: Carbohydrate-Interacting Proteins from Two *Nostoc* (Cyanobacteria) Species.

**West Virginia University** **Morgantown, WV**  
*M.S. Chemical Engineering (December 1998)*  
*B.S. Chemical Engineering (May 1997)*

Thesis work in the lab of Dr. Hisashi Kono, on rheological properties of aerated fine powders: Rheological Properties of Candle Filter Deposits at Elevated Temperatures and Pressures.

### PROFESSIONAL EXPERIENCE

---

**Mylan Pharmaceuticals** **Morgantown, WV**  
(February 1994-August 1997) Quality Control Chemist. Responsible for performing identification, impurity, and moisture assays on manufactured products. Analytical methods employed include thin-layer chromatography, infrared spectroscopy, and ultraviolet spectroscopy.

### TEACHING EXPERIENCE

---

Teaching Assistant  
(Spring 2001) BCHM 2144: Organic Biochemistry  
Department of Biochemistry, Virginia Polytechnic Institute and State University

Tutor  
(Fall 1995) General Science, Mathematics, and Engineering  
West Virginia University Learning Center, West Virginia University

## LAB SKILLS

---

Experience in various aspects of protein chemistry/purification including two-dimensional electrophoresis, MALDI-TOF analysis of peptides/proteins, MS/MS *de novo* peptide sequencing, and use of radioisotopes. Experience in various aspects of analytical chemistry including high performance liquid chromatography (HPLC), nuclear magnetic resonance spectroscopy (NMR), and capillary electrophoresis (CE). Experience in basic aspects of synthetic carbohydrate chemistry.

## PUBLICATIONS

---

Kono, H.O., B. Jordan, T. Otake, and D.H. Smith, "Formation and Measurement of the Porosities, Tensile Strengths, and Deformation Coefficients of Gasification Filter Cakes at Operating Temperatures and Pressures", *Aerosol Science and Technology*, vol.29, pp. 236-245 (1998).

Kono, H.O., L. Richman, B. Jordan, J. Su, E. Koresawa, and T. Otake, "Rheological parameters of homogeneously aerated fine powders (Its Physical Meaning and Fracture Model)", *Fluidization XI*, pp. 283-292, published by Engineering Foundation, NY (1998).

## HONORS & AWARDS

---

Chemical Engineering Academy Scholarship

December 1995

Brian Robert Jordan grew up in Morgantown, West Virginia, where he attended University High School until 1991. Upon high-school graduation, he attended West Virginia University, from which he received Bachelor of Science (May 1997) and Master of Science (December 1998) degrees in Chemical Engineering. In January of 1999, he moved to Blacksburg, Virginia, in order to pursue a doctoral degree in the field of Biochemistry. Brian is presently studying protein phosphorylation in *Synechocystis* PCC 6803, in the laboratory of Dr. Peter Kennelly, in the Department of Biochemistry at Virginia Tech.