The Linker of Endoglucanase A from *Cellulomonas fimi*: an investigation of structure, function and application

by

Linda Eileen Sandercock

B.Sc. in Cell Biotechnology, University of Alberta, 1992

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Department of Microbiology and Immunology / Biotechnology Laboratory)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
March 1999
© Linda Eileen Sandercock, 1999
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of MICROBIOLOGY & IMMUNOLOGY
The University of British Columbia
Vancouver, Canada

Date March 10/97

DE-6 (2/88)
Abstract

Endoglucanase A (CenA), a component of the multi-enzyme cellulase system of the bacterium Cellulomonas fimi (C. fimi), is typical of many cellulases, being a multi-domain protein. CenA has two domains which have been well characterized, a catalytic domain and a cellulose-binding domain (CBD) separated by a proline, threonine rich linker (PT linker). This study investigated the structure, function and possible applications of this interdomain linker.

Modification of the PT linker’s size and composition gave insights into the effect of the linker on hydrolysis of a variety of cellulosic substrates and the nature of its susceptibility to proteolysis. Modifying the linker size and composition had little effect on the hydrolysis rate. Partial deletion reduced activity on crystalline and amorphous celluloses, whereas doubling or replacing it with the fibronectin type III (Fn3) repeats from the endoglucanase B increased the rate of hydrolysis of crystalline cellulose. Altering the linker did not affect activity on 2’,4’-dinitrophenyl β-D-cellubioside (2,4-DNPC) except for the Fn3 construct which showed decreased activity. A sequence susceptible to papain and C. fimi protease was defined which could be modified to prevent proteolysis. This has potential implications for the design of linkers for stable CBD-fusion proteins.

Examining the glycosylation of CenA produced by C. fimi and Streptomyces lividans (S. lividans) gave insights into eubacterial protein glycosylation. Glycosylation was confined to the PT linker at sites conforming to an O-glycosylation consensus sequence for some mammalian glycoproteins. The oligosaccharide composition was dependent on the host, but the glycosylation mechanism is likely similar as they
discriminated between the same protein features. Oligosaccharide size and linkage information for biose units of *S. lividans* produced proteins were determined, the latter by a novel application of fluorophore-assisted carbohydrate electrophoresis. Glycosylation slowed proteolysis and slightly reduced the activity on 2,4-DNPC.

The CenA linker resembles the hinge of human immunoglobulin A1 (IgA1). A hybrid protein, CenAIgA1h - where the PT linker was replaced by the IgA1 hinge - and CenA were useful substrates for probing the specificity of a number of IgA1 proteases produced by a variety of pathogenic bacteria. The most striking finding was that only one of two IgA1 proteases which have high sequence identity and cleave the identical bond in IgA1, cleaved CenA. These substrates will be useful in future studies of IgA1 protease specificity.
Table of contents

Abstract .................................................................................................................. ii
Table of contents .................................................................................................... iv
List of tables .......................................................................................................... xi
List of figures .......................................................................................................... xiii
Abbreviations ........................................................................................................ xvii
Statement of Contribution in IgA1 Protease Collaboration ................................... xxii
Acknowledgements ................................................................................................. xxii

1. Introduction .......................................................................................................... 1

1.1. Cellulose ............................................................................................................. 1

1.2. Cellulases ............................................................................................................ 2

1.2.1. General ............................................................................................................ 2

1.2.2. Cellulase systems ............................................................................................. 2

1.2.3. Linkers ............................................................................................................ 5

1.2.4. Proteolysis ..................................................................................................... 10

1.3. Glycoproteins .................................................................................................... 11

1.3.1. General .......................................................................................................... 11

1.3.2. Protein-carbohydrate linkages ....................................................................... 11

1.3.3. Sugars present in oligosaccharide .................................................................. 12

1.3.4. Sizes of oligosaccharides ............................................................................... 13

1.3.5. Determinants of protein glycosylation ............................................................ 13

1.3.6. Prokaryotic glycoproteins .............................................................................. 15

1.3.7. Glycosylated polysaccharidases .................................................................... 16
1.3.8. Glycosylation of CenA and Cex .............................................................. 20
1.3.9. Function of glycosylation, with an emphasis on cellulases ...................... 20
1.3.10. Technological advances in carbohydrate analysis .................................. 21
1.4. IgA1 and IgA1 proteases ................................................................. 22
  1.4.1. The structure and function of IgA ................................................ 22
  1.4.2. IgA1 proteases ........................................................................... 26
1.5. Objectives ...................................................................................... 33

2. **Methods and Materials** ................................................................. 34

  2.1. Chemicals, buffers, substrates and enzymes ......................................... 34
  2.2. Bacterial strains and plasmids ......................................................... 35
  2.3. Media and growth conditions .......................................................... 37
  2.4. Recombinant DNA techniques ....................................................... 38
    2.4.1. Primer design for introducing silent mutations ............................. 39
    2.4.2. Synthetic oligonucleotides and oligonucleotide primers .............. 39
    2.4.3. DNA sequencing ..................................................................... 39
    2.4.4. PCR ....................................................................................... 41
    2.4.5. Construction of mutant PT linkers ............................................. 44
  2.5. Screening for mutants and confirmation of identity ............................. 46
    2.5.1. *E. coli* ................................................................................. 46
    2.5.2. *S. lividans* ......................................................................... 46
  2.6. Gene expression and protein purification ........................................... 47
    2.6.1. *E. coli* ................................................................................. 47
    2.6.2. *C. fimi* ................................................................................. 48
2.6.3. S. lividans ...........................................................................................................................................48

2.7. Protein detection and quantification and storage ..............................................................................50

2.8. Western blotting and N-terminal sequencing of proteins and peptides ...........................................50

2.9. Mass analysis of proteins .....................................................................................................................51

2.9.1. Predicted molecular weights ..........................................................................................................51

2.9.2. Relative mass .....................................................................................................................................51

2.9.3. Electrospray mass spectrometry .....................................................................................................51

2.9.4. Matrix-Assisted Laser-Desorption/Ionization Time-Of-Flight (MALDI-TOF) mass spectrometry ........................................................................................................................................52

2.10. Stability .............................................................................................................................................53

2.11. Enzyme assays ...................................................................................................................................53

2.11.1. PNPC hydrolysis ..........................................................................................................................53

2.11.2. BMCC hydrolysis ..........................................................................................................................53

2.11.3. PASC hydrolysis ..........................................................................................................................54

2.11.4. CMC hydrolysis ..........................................................................................................................54

2.11.5. 2,4-DNPC hydrolysis ....................................................................................................................54

2.11.6. pH profiles .....................................................................................................................................55

2.11.7. Hydrolysis products ......................................................................................................................55

2.11.8. Statistical analysis of hydrolysis data ............................................................................................55

2.12. Proteolysis .........................................................................................................................................56

2.12.1. Papain hydrolysis ..........................................................................................................................56

2.12.2. C. fimi protease hydrolysis ............................................................................................................56

2.13. IgA1 protease digestion .......................................................................................................................57
2.14. Glycosylation determination ................................................................. 57

2.14.1. Periodic acid/Schiff reagent and ConA-HRP treatment of Western blots ...... 57

2.14.2. Phenol sulfuric acid assay .................................................................. 58

2.14.3. Determination of the monosaccharide composition and linkage positions
between sugars for CenA and Cex produced by *C. fimi* .................................... 58

2.14.4. Determination of the monosaccharide composition of CenA mutants produced
by *S. lividans* and of CenA produced by *C. fimi* using FACE® ......................... 60

2.14.5. Determination of the oligosaccharide composition and linkages between
sugars for CenA2PT produced by *S. lividans* using FACE® ............................. 60

2.14.6. Location of glycans on CenA and Cex produced by *C. fimi* ..................... 61

3. Results ........................................................................................................... 63

3.1. Generation of *cenA* mutants ................................................................. 63

3.1.1. Definition of the boundaries of the PT linker of CenA ............................ 63

3.1.2. Primer design for introducing silent mutations ..................................... 64

3.1.3. Generation of pUC18-1.5cenAΔPTAS .................................................... 64

3.1.4. Construction of the *cenA* mutants ....................................................... 65

3.1.5. Subcloning *cenA* PT mutants into *E. coli* and *S. lividans* expression vectors. 66

3.2. Gene expression and protein purification .................................................. 72

3.2.1. *E. coli* ................................................................................................. 72

3.2.2. *S. lividans* ......................................................................................... 74

3.2.3. *C. fimi* ............................................................................................. 75

3.3. Protein quantification ............................................................................... 75
3.4. Molecular masses of proteins .................................................................77
3.5. Glycosylation determination .................................................................79
  3.5.1. Detection of glycosylation on proteins produced in *S. lividans* and *E. coli* ..........................79
  3.5.2. Moles of sugar per mole of protein for CenA and Cex produced in *C. fimi* and
CenAlgAlh produced in *S. lividans* ..........................................................81
  3.5.3. Location of carbohydrates on CenA and Cex produced in *C. fimi* and CenA
constructs produced in *S. lividans* ..........................................................83
  3.5.4. The monosaccharide composition and linkage positions between sugars of
CenA and Cex produced by *C. fimi* ............................................................89
  3.5.5. The monosaccharide composition of the glycans of CenA mutants produced
by *S. lividans* and CenA produced by *C. fimi* ...........................................94
  3.5.6. Oligosaccharide composition and linkage positions between sugars of
CenA2PT produced in *S. lividans* .............................................................95
3.6. Enzyme activity ......................................................................................98
  3.6.1. Protein stability ..................................................................................98
  3.6.2. Effect of method on BMCC hydrolysis by CenA ....................................98
  3.6.3. BMCC hydrolysis .............................................................................99
  3.6.4. PASC hydrolysis ............................................................................102
  3.6.5. 2,4-DNPC hydrolysis .......................................................................104
  3.6.6. pH profiles .......................................................................................104
  3.6.7. Enzyme assay products ....................................................................107
3.7. Proteolysis .............................................................................................107
  3.7.1. Papain hydrolysis of CenA constructs produced in *E. coli* .................107
3.7.2. *C. fimi* protease hydrolysis of CenA constructs produced in *E. coli* ..........111

3.7.3. Effect on papain hydrolysis of changing valine 176 to proline ...............114

3.7.4. Effect of glycosylation of CenA constructs on susceptibility to papain and *C. fimi* protease ..............................................................................................................117

3.8. IgA1 protease digestion of CenA and CenAIgA1h ........................................121

3.9. A comparison of the glycosylation of the native IgA1 hinge and that produced by *S. lividans* ......................................................................................................................125

4. Discussion ...........................................................................................................127

4.1. The effect of linker length and composition on activity ..................................127

4.1.1. The effect of linker length and composition on activity on PASC .............127

4.1.2. Difficulties in comparing cellulase activity data between studies ............127

4.1.3. The effect of linker length and composition on activity on BMCC ..........128

4.1.4. Measuring hydrolysis of an insoluble substrate ........................................130

4.1.5. The effect of linker length and composition on activity on 2,4-DNPC .......130

4.1.6. Summary of activity studies ....................................................................131

4.2. Effect of linker length and composition on sensitivity to proteases .............131

4.3. Glycosylation ..................................................................................................133

4.3.1. Sites of glycosylation ................................................................................133

4.3.2. Use of mass spectrometry for quantification of glycosylation ..........136

4.3.3. Monosaccharide composition, size of oligosaccharides and linkage positions between sugars of oligosaccharides ..............................................................137

4.3.4. Role of glycosylation on CenA and CenA derivatives .........................139

4.3.5. Summary of glycosylation studies ............................................................142
4.4. Hydrolysis of CenA and CenAIgA1h by IgA1 proteases ........................................143

4.5. Final Summary ........................................................................................................145

5. Bibliography ..............................................................................................................147
List of tables

Table 1.1: Representative linker sequences.................................................................7
Table 1.2: A selection of glycosylated polysaccharidases from bacterial and some fungal sources.........................................................................................................................17
Table 1.3: IgA1 protease producing organisms implicated in disease.........................27

Table 2.1: Bacterial strains...............................................................................................35
Table 2.2: E. coli plasmids...............................................................................................36
Table 2.3: S. lividans/E. coli shuttle plasmids.................................................................37
Table 2.4: Oligonucleotide primers used for PCR.........................................................40
Table 2.5: Oligonucleotide primers used for sequencing...............................................40
Table 2.6: Synthetic oligonucleotides for constructing mutant PT linkers.....................41
Table 2.7: HPLC program...............................................................................................62

Table 3.1: Protein concentrations of CenA constructs produced in E. coli determined by a variety of methods........................................................................................................76
Table 3.2: Molecular masses of CenA and CenA mutants produced in E. coli and S. lividans determined by three different techniques.................................................................78
Table 3.3: Summary of agitation methods studied for the hydrolysis of BMCC...........99
Table 3.4: Summary of the hydrolysis of BMCC by CenA and CenA constructs produced in E. coli and S. lividans.................................................................................................101
Table 3.5: Summary of the hydrolysis of PASC by CenA and CenA constructs produced in E. coli and S. lividans.................................................................................................103
Table 3.6: Susceptibility of CenA substrates to IgA1 proteases.
List of figures

Figure 1.1: Modular organization of cellulases from C. fimi ..............................................4
Figure 1.2: Computer generated model of the CenA PT linker between known structures
of a C. fimi cellulase catalytic domain and CBD..............................................................9
Figure 1.3: The structure of an Fn3 module from fibronectin..............................................9
Figure 1.4: Papain and C. fimi protease cleavage sites in CenA produced in E. coli .............11
Figure 1.5: Motifs proposed for O-glycosylation based on in vivo glycosylation .................14
Figure 1.6: Glycosylation of the hinge region of human myeloma serum IgA1 and human
pooled serum IgA1 .................................................................................................................25
Figure 1.7: Cleavage sites in the hinge region of myeloma serum IgA1 of some IgA1
proteases ................................................................................................................................28
Figure 1.8: Products of cleavage of IgA1 by IgA1 proteases ..................................................30
Figure 1.9: The sequence of the IgA1 hinge compared to the PT linker of CenA ..................32

Figure 2.1: Generating pUC18-1.5cenA Δ PT AS .................................................................43
Figure 2.2: Construction of cenA1.5PT and cenA2PT using synthetic oligonucleotides .......45

Figure 3.1: The PT linker of CenA as defined in various studies .......................................63
Figure 3.2: Design of silent mutation sites .........................................................................64
Figure 3.3: Construction of pUC18-1.5cenA mutants .........................................................68
Figure 3.4: Part 1 of subcloning strategy to insert cenA mutants into expression vectors
pTUG and pIJ680 ...............................................................................................................69
Figure 3.5: Part 2 of subcloning strategy to insert cenA mutants into expression vectors
pTUG and pIJ680. .......................................................................................70

Figure 3.6: DNA agarose gel showing the Nhel/HindIII fragment from each
pSL1180cenA mutant used to subclone into pTUg and pIJ680. ......................71

Figure 3.7: Congo-red stained agar plate showing halos where S. lividans colonies
expressing CenAΔpt were growing. ............................................................71

Figure 3.8: SDS-PAGE gel of CenA constructs produced in E. coli. ..................72

Figure 3.9: The domain structure and linker sequences of the CenA and CenA PT linker
constructs ..................................................................................................73

Figure 3.10: SDS-PAGE gel of CenA constructs produced in S. lividans. ...........74

Figure 3.11: SDS-PAGE gel of CenA and Cex produced in C. fimi. ..................75

Figure 3.12: SDS-PAGE gel and ConA-HRP treated blot of CenA constructs produced in
S. lividans. ..................................................................................................80

Figure 3.13: SDS-PAGE gel and ConA-HRP treated blot of CenA constructs produced in
S. lividans and E. coli. ................................................................................80

Figure 3.14: SDS-PAGE gel of CenA constructs produced in E. coli and S. lividans ......81

Figure 3.15: MALDI-TOF mass spectra of CenA and Cex produced by C. fimi...........82

Figure 3.16: Reverse phase HPLC profiles of a tryptic digest of CenA produced by C.
fimi ...........................................................................................................85

Figure 3.17: Reverse phase HPLC profiles of a tryptic digest of Cex produced by
C. fimi. .......................................................................................................86
Figure 3.18: SDS-PAGE gel and periodic acid/Schiff stained Western blot of CenA and Cex produced by *C. fimi*, and treated with papain ................................................................. 88

Figure 3.19: Glycosylation sites on the linker of CenA and Cex produced by *C. fimi* ....88

Figure 3.20: Determination of carbohydrate composition by GLC of CenA and Cex produced by *C. fimi* ........................................................................................................... 91

Figure 3.21: SDS-PAGE gel and periodic acid/Schiff stained Western blot of CenA and Cex produced by *C. fimi*, and treated with exo-glycosidases ................................... 92

Figure 3.22: Western blot treated with ConA-HRP to detect the presence of mannose on Cex and CenA produced by *C. fimi*, and treated with exo-glycosidases ............... 94

Figure 3.23: FACE® Monosaccharide composition gel of CenA produced by *C. fimi* and a subset of CenA mutants produced by *S. lividans* ........................................ 96

Figure 3.24: FACE® Oligosaccharide analysis gel of CenA2PT produced by *S. lividans*. 97

Figure 3.25: 2,4-DNPC hydrolysis at pH 6.5 by CenA linker constructs produced in *E. coli* and *S. lividans* ........................................................................................................ 105

Figure 3.26: Composite figure of 2,4-DNPC hydrolysis at pH 6.5 by CenA linker constructs produced in *E. coli* and *S. lividans* ......................................................... 106

Figure 3.27: pH profiles of 2,4-DNPC hydrolysis by CenA and CenA constructs produced in *E. coli* and *S. lividans* ................................................................. 106

Figure 3.28: Papain digestion of CenA and CenA PT constructs produced in *E. coli* ....108

Figure 3.29: Expanded gel of a papain digestion of CenA produced by *E. coli* ..........109

Figure 3.30: Papain cleavage sites in CenA and CenA PT mutants made in *E. coli* .....110

Figure 3.31: *C. fimi* protease digestion of CenA and CenA PT constructs made by *E. coli* ............................................................................................................. 112
Figure 3.32: Papain and *C. fimi* protease cleavage sites in CenA and CenA PT mutants made in *E. coli* ................................................................. 113

Figure 3.33: *C. fimi* protease digestion of IgA1................................................................. 114

Figure 3.34: Papain digestion of CenA1.5PT and CenA1.5PTmod made by *E. coli*...... 115

Figure 3.35: Papain and *C. fimi* protease cleavage sites in CenA1.5PT and CenA1.5PTmod produced in *E. coli*. ................................................................. 116

Figure 3.36: Papain digestion of CenA and CenA PT mutants made by *E. coli* and *S. lividans*. ................................................................................................. 118

Figure 3.37: *C. fimi* protease digestion of CenA and CenA PT constructs made by *E. coli* and *S. lividans*................................................................. 119

Figure 3.38: Papain and *C. fimi* protease cleavage sites in CenA and CenA PT mutants made in *E. coli* and *S. lividans*. ................................................................. 120

Figure 3.39: IgA1 protease digests of CenA IgA1h and CenA produced by *E. coli* and *S. lividans*................................................................................................. 122

Figure 3.40: Glycosylation of the hinge region of human myeloma serum IgA1 and CenA IgA1h produced in *S. lividans*. ................................................................. 126
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,4-DNPC</td>
<td>2',4'-dinitrophenyl β-D-cellobioside</td>
</tr>
<tr>
<td>x g</td>
<td>Centrifugal force relative to the gravitational force</td>
</tr>
<tr>
<td>A&lt;sub&gt;n&lt;/sub&gt;</td>
<td>Absorbance at wavelength “n”</td>
</tr>
<tr>
<td>AA</td>
<td>Amino acid</td>
</tr>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Ampicillin resistance</td>
</tr>
<tr>
<td>ANTS</td>
<td>8-aminonaphthalene-1,3,6-trisulfonic acid</td>
</tr>
<tr>
<td>Ara</td>
<td>Arabinose</td>
</tr>
<tr>
<td>BMCC</td>
<td>Bacterial micro-crystalline cellulose</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>C-</td>
<td>Carboxyl terminus</td>
</tr>
<tr>
<td>CBD</td>
<td>Cellulose-binding domain</td>
</tr>
<tr>
<td>CSD</td>
<td>Cleavage-specificity domain</td>
</tr>
<tr>
<td>CMC</td>
<td>Carboxymethylcellulose</td>
</tr>
<tr>
<td>COH</td>
<td>Carbohydrate</td>
</tr>
<tr>
<td>ConA</td>
<td>Concanavalin A</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethylsulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleic acid triphosphate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FACE</td>
<td>Fluorophore-assisted carbohydrate electrophoresis</td>
</tr>
<tr>
<td>Fn3, FNIII</td>
<td>Fibronectin type III sequence</td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast protein liquid chromatography</td>
</tr>
<tr>
<td>Fuc</td>
<td>Fucose</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>GalNAc</td>
<td>N-acetylgalactosamine</td>
</tr>
<tr>
<td>GLC</td>
<td>Gas liquid chromatography</td>
</tr>
<tr>
<td>Gle</td>
<td>Glucose</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetylgulosamine</td>
</tr>
<tr>
<td>GRASP</td>
<td>Centre for gastroenterology research on absorptive and secretory processes</td>
</tr>
<tr>
<td>HBAH</td>
<td>p-Hydroxybenzoic acid hydrazide</td>
</tr>
<tr>
<td>HPLC</td>
<td>High pressure liquid chromatography</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>IgA</td>
<td>Immunoglobulin A</td>
</tr>
<tr>
<td>IgA1</td>
<td>Immunoglobulin A1</td>
</tr>
<tr>
<td>I</td>
<td>Inositol</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactoside</td>
</tr>
<tr>
<td>Kan</td>
<td>Kanamycin</td>
</tr>
<tr>
<td>Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Kanamycin resistance</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>kDa</td>
<td>kiloDalton</td>
</tr>
<tr>
<td>$K_m$</td>
<td>Michaelis constant</td>
</tr>
<tr>
<td>LB</td>
<td>Luria broth</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix-assisted laser-desorption/ionization time-of-flight</td>
</tr>
<tr>
<td>MALDI-TOF MS</td>
<td>Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry</td>
</tr>
<tr>
<td>Man</td>
<td>Mannose</td>
</tr>
<tr>
<td>$M_r$</td>
<td>Relative molecular mass</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>N-</td>
<td>Amino terminus</td>
</tr>
<tr>
<td>NA, na</td>
<td>Not applicable</td>
</tr>
<tr>
<td>NAPS</td>
<td>UBC nucleic acid and protein synthesis unit</td>
</tr>
<tr>
<td>NeuAc</td>
<td>N-acetylneuramic acid, also called sialic acid</td>
</tr>
<tr>
<td>No., no.</td>
<td>Number</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>PASC</td>
<td>Phosphoric acid swollen cellulose</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PMSF</td>
<td>Phenylmethysulfonylfluoride</td>
</tr>
<tr>
<td>PNPC</td>
<td>Para-nitrophenyl-β-D-cellobioside</td>
</tr>
<tr>
<td>PT linker</td>
<td>Proline, threonine rich linker</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene difluoride</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>RT</td>
<td>Room temperature</td>
</tr>
<tr>
<td>Periodic acid/Schiff</td>
<td>Periodic acid/Schiff reagent treatment</td>
</tr>
<tr>
<td>stain</td>
<td>Crystalline surface-layer</td>
</tr>
<tr>
<td>S-layer</td>
<td>Polymeric immunoglobulin receptor</td>
</tr>
<tr>
<td>SC</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SELDI</td>
<td>Surface-enhanced laser desorption ionization</td>
</tr>
<tr>
<td>Sinapinic acid</td>
<td>3,5-dimethoxy-4-hydroxycinnamic acid</td>
</tr>
<tr>
<td>TPCK</td>
<td>N-tosyl-L-phenylalanylchloromethyl ketone</td>
</tr>
<tr>
<td>TSB</td>
<td>Tryptic Soy Broth</td>
</tr>
<tr>
<td>Tsr</td>
<td>Thiostrepton</td>
</tr>
<tr>
<td>Tsr&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Thiostrepton resistance</td>
</tr>
<tr>
<td>U</td>
<td>Units</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet wavelength</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
</tbody>
</table>
Statement of Contribution in IgA1 Protease Collaboration

The use of glycosylated and non-glycosylated CenA and a hybrid protein CenAlgA1h containing the IgA1 hinge region, as substrates for probing the specificity of IgA1 proteases was investigated in collaboration with the laboratory of Dr. Andrew Plaut of GRASP at Tufts-New England Medical Centre Hospital, Boston. Dr. Andrew Plaut and Jiazhou Qiu, a scientist in the laboratory of Dr. Plaut, have agreed that the results of our collaboration may be reported in my thesis and therefore I report here the nature of the contributions made by each party.

Dr. Plaut was approached by Dr. Warren about entering into this collaboration. Dr. Plaut and Jiazhou Qiu made the decisions to test the variety of IgA proteases that were tested and had the expertise to carry out the protease tests and guide the interpretation. The substrates were produced by Linda Sandercock who also analyzed the glycosylation of the substrates and the susceptibility of IgA1 to C. fimi protease. Regular consultations were held between the collaborators.
Acknowledgements

I would like to acknowledge and thank my supervisors, Drs. Doug Kilburn and Tony Warren, and their families for their good-natured and multi-faceted support, guidance and encouragement. Sincere thanks to my committee members, Drs. Julian Davies, Gerald Weeks and Steve Withers, for taking time to provide me with their unique perspectives on my project and getting me to stop! Many thanks to Dr. Andrew Plaut and Jiazhou Qiu for taking time to teach me and let me collaborate with them in their world of IgA1 proteases.

The Cellulase Lab has provided me with a dynamic, very supportive and friendly environment to pursue more than simply science over the years. My gratitude goes out to all members past and present who have made my time in Vancouver rich indeed: especially PT, himself, for innumerable things including permission to invade his bench and ask just one more question; Greg for stimulating late night discussions and safe-walk service; Al B, my partner in glycoprotein forays; Dominik, whose high spirits helped keep my own up; Brad, my computer saviour; Helen, our organizer in the lab and on the hill; Emily and Emily who patiently taught me techniques and lent their ears many, many times; Edgar who introduced me to glycobiology and continues to inspire me; Hua and Pat whose work provided a base for my own and who fueled me even from a distance; and Henrik, Laurent and Neil for their help.

For technical assistance, I am indebted to many including Suzanne for all her protein sequencing; the Withers lab, especially Dave for the use of his bench etc. and Karen; and Dr. Eigendorf, Lina and Marshall in the mass spectrometry facility.

For keeping us graduate students organized, I commend the hard work of the microbiology graduate secretaries past and present and the department office staff.
A special thanks to the microbiology dessert women past and present for great food, wonderful friendships and chat!

Finally, I feel very fortunate and am extremely grateful for the support (from far and near!) of my husband, Mark, and our families for the duration of this project.
1. Introduction

1.1. Cellulose

Cellulose, a linear polymer of glucose residues joined β-1,4, is abundant on earth, being the major component of plant cell walls. Despite the simplicity of the linkage, cellulose may be complex and heterogeneous. Parameters such as size, shape, porosity, surface area, association with non-cellulosic components, molecular conformation and crystallinity may differ (Tomme et al., 1995). As a result there are many different forms of cellulose. Celluloses used in research include oligomers up to six glucose units which are soluble (Warren, 1996), and longer polymers made soluble by substitutions such as in carboxymethylcellulose (CMC). Insoluble celluloses include the crystalline cellulose, bacterial micro-crystalline cellulose (BMCC) made by Acetobacter, and a more amorphous, or disordered cellulose, Avicel™. Phosphoric acid treatment can render this latter cellulose even more amorphous, producing phosphoric acid swollen cellulose (PASC). PASC is considered to be the most homogeneous form of amorphous cellulose (Schülein, 1998). All the insoluble cellulosic substrates are to some degree heterogeneous however. This heterogeneity results from imperfections in packing or mechanical damage, which cause disorder in the cellulose structure. For a more detailed overview of cellulose structures see Tomme et al., 1995 (Tomme et al., 1995).
1.2. Cellulases

1.2.1. General

Cellulases are produced by a number of different organisms including bacteria, fungi (Teeri et al., 1992; Tomme et al., 1995; Warren, 1996), plants and most recently nematodes and termites (Ding et al., 1998; Smant et al., 1998; Watanabe et al., 1998). Cellulases can be broadly divided into two classes: exoglucanases or cellobiohydrolases, which cleave cellobiose units from the termini of cellulose molecules, and endoglucanases which cleave internally. The former generally have the highest apparent specific activity on crystalline cellulose (Teeri et al., 1998). Most cellulases are modular: composed of a catalytic domain associated with other domain(s) such as cellulose-binding domains (CBD), often via a linker sequence (Tomme et al., 1995). Examples of specific systems are given below.

1.2.2. Cellulase systems

To degrade such a complex and heterogeneous substrate as cellulose, organisms have evolved systems consisting of multiple cellulases, varying in substrate specificity and hydrolytic activity, which work in concert. Well studied bacterial cellulase systems include the non-complexed systems of Cellulomonas fimi (C. fimi) and Thermomonospora fusca (T. fusca), and the complexed and cell-associated systems of Clostridium thermocellum (C. thermocellum) and Clostridium cellulolyticum (C. cellulolyticum). The best characterized fungal system is that of Trichoderma reesei (T. reesei).
C. fimii is a Gram-positive, mesophilic, coryneform bacterium (Warren, 1996). This aerobe has a multicomponent cellulase system consisting of secreted endoglucanases, an exoglucanase/xylanase and celllobiohydrolases (Gilkes et al., 1991a). Four endoglucanases CenA-D, one exoglucanase/xylanase Cex, and two celllobiohydrolases have been reported (Figure 1.1) (Gilkes et al., 1991b; Shen et al., 1995; Warren, 1996).

CenA is an endo $\beta$-1,4-glucanase of about 48 kDa and Cex is a $\beta$-1,4-xylanase/exo-$\beta$-1,4-glucanase of about 46-49 kDa (Gilkes et al., 1989). The genes encoding these cellulases have been cloned and expressed in Escherichia coli (E. coli) and the recombinant proteins characterized in some detail (Damude, 1995; Din et al., 1994; Din et al., 1991; MacLeod et al., 1994; White et al., 1994). Both CenA and Cex comprise two domains: a catalytic domain and a cellulose-binding domain (CBD) separated by a proline, threonine rich linker (PT linker), but the orientation of their domains is reversed. CenB is an endo $\beta$-1,4-glucanase of about 110 kDa which has six domains: a catalytic domain, a CBD, three fibronectin type III domains (Fn3) and a second CBD, separated by linkers rich in proline and hydroxyamino acids (Meinke et al., 1991a; Meinke et al., 1991b). CenB is approximately fifty times more active than CenA on BMCC, with comparable activities on Avicel, and CMC, but CenA has four-fold higher on PASC (Tomme et al., 1996).

T. fusca is a thermophilic actinomycete which produces at least six cellulases, including four endoglucanases and two celllobiohydrolases (Warren, 1996; Wilson, 1992; Zhang et al., 1995). C. thermocellum is a Gram-positive, anaerobic, thermophilic bacterium, which degrades crystalline cellulose with high efficiency. It has a complexed, cell-associated cellulase system characterized by a multienzyme, cell-bound complex
Figure 1.1: Modular organization of cellulases from \textit{C. fimi}.

The numbers in the boxes indicate the catalytic domain family. (Cellulase catalytic domains have been divided into families based on sequence similarities.)

Acellulosomes. At least twenty-one endoglucanases, three cellobiohydrolases, and a scaffolding protein have been described to date (Béguin and Alzari, 1998; Béguin and Lemaire, 1996; Bronnenmeier and Staudenbauer, 1993). \textit{C. cellulolyticum}, like \textit{C. thermocellum}, is a Gram-positive, anaerobic bacterium which produces a cellulosome. A scaffolding protein, several endoglucanases and a potential cellobiohydrolase have been reported. The domains are not separated by defined linkers (Bronnenmeier and Staudenbauer, 1993; Gal, 1997; Warren, 1996). \textit{T. reesei}, a mesophilic fungus, has a non-complexed cellulose hydrolyzing system similar to that of \textit{C. fimi} with at least three
endoglucanases and two cellobiohydrolases (Kubicek, 1992; Schülein, 1998; Warren, 1996).

1.2.3. Linkers

Linkers rich in hydroxyamino acids and often proline separate the modules of many cellulases and hemicellulases. This type of linker is not restricted to these families of enzymes, but can also be found in a diverse range of other proteins. Some examples are shown in Table 1.1. A more extensive table was published in Gilkes et al., 1991d (Gilkes et al., 1991d). From the table it is evident that the length of linkers can vary greatly - linkers of up to three hundred amino acids have been reported (Black et al., 1996) - and that only in rare cases such as for CenA and Cex, is there high sequence identity between linkers. The linkers can be loosely divided into three classes: those with high (30% or greater) proline content; those with less proline content; and a class of interdomain linkers called Q-linkers, which are rich in arginine, glutamine, glutamate, and hydrophobic amino acids (Gilkes et al., 1991d). Proline and threonine/serine rich linker sequences can form extended helices (Argos, 1990; Williamson, 1994). A computer-derived model of the PT linker of CenA is shown in Figure 1.2. Many fungal cellulases have an O-glycosylated linker region, but the structure and flexibility of the linker region is not known, although even the less proline-rich O-glycosylated linker sequences of CBHI and CBHII from T. reesei are assumed to adopt an extended conformation (Schülein, 1998; Teeri et al., 1998). Some bacterial cellulases also have O-glycosylated linkers, such as E2 and E3 from T. fusca (Calza et al., 1985; Zhang et al., 1995). Glycosylation of proteins is dealt with later in the Introduction (Section 1.3).
Another potential spacer module is the Fn3 module. These are found in a number of different bacterial cellulases, often in tandem (Shen et al., 1995). The structure of an Fn3 module of fibronectin has been solved (Figure 1.3) as has a fibronectin segment with four Fn3 modules in tandem has also been determined (Leahy et al., 1996; Main et al., 1992). In contrast to the PT linker, the Fn3 module has a defined tertiary structure consisting of seven antiparallel β-strands forming two facing β-sheets; yet, when in tandem the Fn3 modules are reported to have non-specific interdomain interactions which provide some flexibility and elasticity (Erickson, 1994; Leahy et al., 1996).

Despite the widespread presence of linkers in cellulases and hemicellulases, a single role for all linkers has not been established (Hazelwood and Gilbert, 1998). Investigation into the role of these linkers has mainly been done by deletion. Linkerless mutants of CenA from C. fimi (Shen et al., 1991), xylanase XYLA and arabinofuranosidase XYLC from Pseudomonas fluorescens subsp. cellulosa (Black et al., 1997; Black et al., 1996), E2 from T. fusca (Wilson et al., 1995) and CBHI from T. reesei (Srisodsuk et al., 1993) all had lower activity on insoluble substrates, suggesting a role for the linker in the efficient hydrolysis of insoluble substrates. Only the CenA mutant exhibited decreased activity against a soluble substrate. It was also more difficult to desorb from Avicel™. These changes, for CenA, were attributed to a change in the relative orientation of the catalytic domain and CBD. Studies on other modular proteins suggested a role for linkers in protein stability. For example, an O-glycosylated linker separating the starch-binding domain and catalytic domain of a glucoamylase from Aspergillus awamori is important for thermostability, and resistance to degradation (Baker-Libby et al., 1994).
Table 1.1: Representative linker sequences.

A. representative linker sequences from cellulases, hemicellulases and a glucoamylase; B. two representative non-cellulase proline rich linkers; C. a representative Q linker. The bold entries indicate linkers with 30% or higher proline content. Unless otherwise noted, the sequence information is taken from Table 1 in Gilkes et al, 1991d (Gilkes et al., 1991d).
<table>
<thead>
<tr>
<th>ORGANISM</th>
<th>PROTEIN</th>
<th>LINKER SEQUENCE</th>
<th>no. AA</th>
<th>no. P</th>
<th>no. S/T</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong> Cellulomonas fimii</td>
<td>CenA</td>
<td>PT₂S(PT)₂T(PT)_₂VTPQPT</td>
<td>33</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>CenB</td>
<td>(PT)_₂T(PT)_₂T(PT)_₂S</td>
<td>21</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>CenC</td>
<td>i PTGT₁DT₂P₂T₂PGTP</td>
<td>17</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ii T₂DT₂GETEP₁T₂PGTP</td>
<td>17</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>iii T₂A₂PVTVA̅PTVPGTP</td>
<td>17</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>iv S₂PVTFT₃LPVTSTPS</td>
<td>16</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>i SLT₁SATP₂</td>
<td>10</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ii PVPTAP</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Thermononospora fusca</td>
<td>E1 *¹</td>
<td>i DAGEPG₂₂GPGD₂ETP₂S</td>
<td>20</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ii TTEKEDETTPSASC</td>
<td>15</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>E₂ *²</td>
<td>(NP)₁(TP)₂P₂GS₂GA</td>
<td>25</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>E₄ *¹</td>
<td>PE₂GE₂PG₂EGPG₂₃PGEDVTP₂SAP GS</td>
<td>31</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Streptomyces sp. strain KSM-9</td>
<td>CasA</td>
<td>i PRT₂(PT)₂P</td>
<td>9</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ii PA₁TGA(SP)₂AP₂ASPAPSADS</td>
<td>22</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Clostridium thermocellum</td>
<td>CelA</td>
<td>PLSDLSGQPTP₂SNPTPSL₂</td>
<td>21</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>CelB</td>
<td>TPSVT(PS)₂ATPSPT₁ITAP₂T</td>
<td>22</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>CelE</td>
<td>PLVS(PT)₂LMPTSPVT</td>
<td>20</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>CelH</td>
<td>i (PT)₂WTSTSP₂S</td>
<td>16</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ii PGTYPSYSKPSPTPRPTKP₂VTP</td>
<td>24</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Trichoderma reesei</td>
<td>CBHI</td>
<td>PG₁NRGT₁R₁PAT₂GS₂PGPTQS</td>
<td>26</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>CBHII</td>
<td>PGA₃S₁TRA₃ST₂SRVSPT₃S₁ATP₃</td>
<td>44</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>GST₂RV₁VG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EgII</td>
<td>P₂AS₂T₁FSTR₂S₂T₂S₁PSCTQT</td>
<td>29</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Eg1III</td>
<td>PGAT₁IT₁STRP₂SGPT₁RA(⁵TS)₂S₂₃TP₂S₃</td>
<td>34</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>Microbispora bispora</td>
<td>CelA</td>
<td>P₂TY(⁵SP)₂TPST(PS)₂QSDPGS(PS)₃</td>
<td>30</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>Dictyostelium discoideum</td>
<td>cellulase</td>
<td>270-6</td>
<td>99</td>
<td>24</td>
<td>52</td>
</tr>
<tr>
<td>Aspergillus awamori</td>
<td>GA *³</td>
<td>T₃T₄ATPTGSΓSTVTSKT₃ATAS KTSTST₃TS</td>
<td>37</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td><strong>B</strong> Human</td>
<td>IgA1</td>
<td>PVPSTP₁TPSPSTP₂T</td>
<td>16</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>OmpA</td>
<td>APV₄(AP)₄</td>
<td>12</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td><strong>C</strong> Bacillus subtilis</td>
<td>Spo0A</td>
<td>SGNAS₂VTHRAPS₂QS₂J</td>
<td>18</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

Additional references: *¹ (Jung et al., 1993), *² author's interpretation from Genbank accession number M73321, *³ (Baker-Libby et al., 1994)
**Figure 1.2:** Computer generated model of the CenA PT linker between known structures of a *C. fimi* cellulase catalytic domain and CBD. Tryptophans involved in binding are shown below the CBD. (figure courtesy of P. Tomme)

**Figure 1.3:** The structure of an Fn3 module from fibronectin. (Main *et al.*, 1992).
1.2.4. **Proteolysis**

Many cellulolytic microorganisms partially degrade their cellulases releasing active catalytic domains (Hagspiel et al., 1989; Lao and Wilson, 1996; Sandercock et al., 1996; Warren, 1996). Partial degradation may also be seen in the recombinant production of some cellulases (Fierobe et al., 1993; Fierobe et al., 1991) and may be forced using the natural proteases as well as non-specific proteases such as papain and chymotrypsin (Gilkes et al., 1989; Gilkes et al., 1988; Tomme et al., 1988). These degradation products gave insight into the domain structure of cellulases as the protease-sensitive regions lay between domains. Serine proteases have been identified for both *C. fimi* and *T. fusca* (Langsford et al., 1984; Lao and Wilson, 1996). The protease from *T. fusca* has been cloned and characterized. Partial degradation of the cellulases during hydrolysis on cellulose may be advantageous as the catalytic domains of some cellulases are more active than the full enzymes on soluble substrates (Béguin and Alzari, 1998; Fierobe et al., 1993; Fierobe et al., 1991; Ghangas and Wilson, 1988; Reverbel-Leroy et al., 1997). The separated domains may thus enhance hydrolysis of soluble cellulose preventing its accumulation during hydrolysis (Sandercock et al., 1996). The protease sensitivity, however, is a shortcoming of natural linker sequences used in the design of fusion proteins. For single-chain antibodies containing linkers derived from the *T. reesei* CBHI linker, length and amino acid composition affected linker stability in the hybrid construct. A proline placed after the proteolytic cleavage site reduced the sensitivity of these constructs to proteolysis (Alfthan et al., 1995).
Some sites in CenA cleaved by papain and \textit{C. fimi} protease have been reported previously (Gilkes \textit{et al.}, 1989; Gilkes \textit{et al.}, 1988; Langsford \textit{et al.}, 1984; Sandercock \textit{et al.}, 1996). Figure 1.4 indicates these sites.

\textbf{Figure 1.4:} Papain and \textit{C. fimi} protease cleavage sites in CenA produced in \textit{E. coli}. Arrows point to cleavage sites; P, papain; C, \textit{C. fimi} protease; the underlined sequences belong to the domains adjacent to the linker.

1.3. Glycoproteins

1.3.1. General

Glycoproteins are proteins with covalently attached carbohydrate moieties. The carbohydrates are added post-translationally or co-translationally. Glycoproteins are present in a diverse range of organisms. There are many books and reviews on glycoproteins (Dwek, 1996; Dwek, 1995; Jenkins and Curling, 1994; Lis and Sharon, 1993 and references therein). A brief overview of glycoproteins is provided here.

1.3.2. Protein-carbohydrate linkages

There are two main types of protein-carbohydrate linkages, N-linked and O-linked. N-linked refers to oligosaccharides attached to the protein through the amide of asparagine
residues. O-linked involves attachment of sugars to the hydroxyamino acids, serine and threonine, via their hydroxyl groups (Beeley, 1987). Less common linkages, for example involving the hydroxyl groups of hydroxylysine, hydroxyproline or tyrosine, occur in plants and prokaryotes (Moens and Vanderleyden, 1997; Paul and Wieland, 1986; Wilson et al., 1991 and references therein). A new linkage class, phosphoglycosylation, where the oligosaccharides are linked to serine or threonine via phosphodiesters has been proposed recently by Haynes, 1998 (Haynes, 1998).

1.3.3. Sugars present in oligosaccharide

The sugar commonly found N-linked to asparagine is N-acetylglucosamine (GlcNAc) and for O-linkages, N-acetylgalactosamine (GalNAc), but there are deviations from this pattern. For example, rhamnose and GalNAc appear N-linked to bacterial glycoproteins (Messner and Sleytr, 1988; Paul and Wieland, 1986), and mannose (Man), galactose (Gal), arabinose (Ara), glucose (Glc), GlcNAc, and fucose (Fuc) may be found in O-linkages, primarily in plant, lower eukaryotic, and prokaryotic glycoproteins (Gooley and Williams, 1994; Ong et al., 1994; Plummer et al., 1995; Wilson et al., 1991).

The sugars making up the remainder of the oligosaccharide structures also vary. Sugars commonly found in mammalian glycoproteins include Man, Gal, Fuc, GlcNAc, GalNAc, and N-acetylneuramic acid (NeuAc) (Raju et al., 1996). Sugars less commonly found include Glc, and glucuronic acid (Plummer et al., 1995). Novel sugars occur in prokaryotic, fungal, plant and mammalian glycoproteins (Gerwig et al., 1989; Lis and Sharon, 1993; Moens and Vanderleyden, 1997; van den Eijn et al., 1995). In addition to the variety of sugars, another degree of complexity arises from the number of possible
linkage positions between sugars (1-2,3,4 or 6) and the anomeric configuration of that linkage (alpha or beta) (Raju et al., 1996).

1.3.4. Sizes of oligosaccharides

The glycan components of glycoproteins can be large. The distance across a carbohydrate residue from O-1 to O-4 is 5.4 Angstroms. N-linked oligosaccharides tend to be larger branching structures compared to the less complex O-linked oligosaccharides which may be unbranched. A typical N-linked oligosaccharide consists of a pentasaccharide core with two or three outer branches of three or four sugar units. Up to five branches may be present. O-linked oligosaccharides range in size from one to twenty residues (Dwek, 1996; Lis and Sharon, 1993). The glycan groups may have considerable freedom of rotation around the glycosidic bond allowing motion that could result in large portions of a protein being shielded; even by a relatively small oligosaccharide (Dwek, 1996; Lis and Sharon, 1993).

1.3.5. Determinants of protein glycosylation

Protein glycosylation is determined by the presence of potential glycosylation sites, the structure of the protein, and the expressing cell type. The consensus sequence described for sites of N-linked glycosylation is N-X-T/S-X, where X is any amino acid except proline. This consensus sequence is common to both prokaryotes and eukaryotes (Gooley and Williams, 1994). The search for the O-linked consensus sequence has not yielded a definitive motif, although there has been some success in predicting threonine and serine glycosylation sites in eukaryotic cells within a limited sample (Figure 1.5)
(Gooley and Williams, 1994 and references therein). Less research has been done on motifs for prokaryotic O-glycosylation. D-S* and D-T*-T, where X* represents a glycosylated amino acid, have been proposed as motifs for *Chryseobacterium meningosepticum* and V-Y* for *Thermoanaerobacter kivui* (Moens and Vanderleyden, 1997).

1. $X-p-X-X$ where one $X= T$ (glycosylated)
2. $X-X-T-X$ where one $X= R/K$ and the $T$ is glycosylated
3. $S-X-X-X$ where one $X= S$ and the $S$ is glycosylated
4. $T-X-X-X$ where one $X= T$ (glycosylated) and the $T$ is glycosylated
5. $X-X-S-X-X-X$ where one $X= P$ and another $X= D/E$ and the $S$ is glycosylated
6. $C-X-S-X-P-C$ where one $X= P$ and another $X= $ any amino acid and $S$ is glycosylated
7. $C-X-X-G-G-T-C-S$ where $X= $ any amino acid and the $T$ and $S$ are glycosylated

**Figure 1.5:** Motifs proposed for O-glycosylation based on *in vivo* glycosylation.

Motifs 1-4 have been identified from some human and *Dictyostelium* proteins, motifs 5 and 6 were identified from epidermal growth factor regions of some multidomain proteins (Gooley and Williams, 1994; Pisano *et al.*, 1994).

Not all consensus sites or proposed motifs are glycosylated and some are only occasionally glycosylated. There is also variation, microheterogeneity, in sugar composition which can occur at a single site (Raju *et al.*, 1996). As a result, glycoproteins do not generally exist in a single form, but rather as a series of species called glycoforms (Dwek, 1996). This heterogeneity led to the recognition of the effect of protein conformation on glycosylation. Both the overall conformation and the environment immediately surrounding a potential
glycosylation site may have an effect (Dwek, 1996). Finally, each cell type has its own set of glycosyltransferases and glycosidases involved in the addition and processing of the oligosaccharides. These can vary with culture conditions. Glycosylation is therefore cell and cell cycle specific. Examples of this variation abound due to tissue specific studies and the use of heterologous expression systems (Dwek, 1996; Jung and Williams, 1997; Raju et al., 1996).

1.3.6. Prokaryotic glycoproteins

As alluded to above, glycoproteins are commonly produced by eukaryotic cells. In fact the majority of proteins secreted by mammalian cells are glycoproteins (Goochee et al., 1991). Glycoproteins are found in organisms across the spectrum of eukaryotic organisms and as a result, most of what is known about glycosylation has come from the study of eukaryotic systems. The presence of carbohydrates on bacterial cells was established with the description of the crystalline cell surface (S-layer) glycoprotein of Halobacterium salinarium in 1976 (Mescher and Strominger, 1976). S-layer proteins have been studied since that time and have been reviewed recently (Messner et al., 1997b). Despite publications on non-S-layer bacterial glycoproteins appearing over the last twenty years, it was only recently that they have been recognized to be widespread. This is evidenced by three minireviews appearing in the last four years (Messner, 1997a; Moens and Vanderleyden, 1997; Sandercock et al., 1994). Bacterial glycoproteins are found in members of both the archaeabacteria and eubacteria and are generally divided into S-layer glycoproteins, and non-S-layer glycoproteins. In the review by Messner (Messner, 1997a), the non-S-layer glycoproteins are further subdivided into surface-associated glycoproteins
in which flagellin and pili components are the major representatives, cellular glycoproteins, membrane glycoproteins, and secreted glycoproteins and exoenzymes. Thirty-five genera are represented in reviews, of which there are carbohydrate structures for glycoproteins from only seven: *Neisseria meningitidis, Halobacterium halobium, Chryseobacterium meningosepticum, Bacteroides cellulosolvens, Clostridium thermocellum, Mycobacterium tuberculosis* and *Thermoplasma acidophilum* (Messner, 1997a; Moens and Vanderleyden, 1997; Sandercock et al., 1994).

### 1.3.7. Glycosylated polysaccharidases

Polysaccharidases from a number of bacteria and fungi are glycoproteins. Table 1.2 summarizes the characteristics of a number of bacterial polysaccharidases which are glycoproteins and some proteins from the best characterized of the cellulolytic fungi, *T. reesei*. The structures of some of the carbohydrate moieties have been determined as indicated. Not all polysaccharidases from a given organism are necessarily glycosylated. For example, of the seven *C. fimi* cellulases described so far, only CenA and Cex are glycoproteins (Langsford, 1988; Langsford et al., 1987). Cex is also glycosylated when expressed in *Streptomyces lividans* (*S. lividans*) (MacLeod et al., 1992; Ong et al., 1994). Similarly for *T. fusca*, only two of six described cellulases are glycoproteins and they too are glycosylated when produced in *S. lividans* (Ghangas and Wilson, 1988; Zhang et al., 1995). In the case of the fungal polysaccharidases, glycosylation appears to be common. All the major cellulases of *T. reesei* are glycoproteins (Teeri et al., 1992).
Table 1.2 continued

<table>
<thead>
<tr>
<th>Organism</th>
<th>Protein</th>
<th>Linkage</th>
<th>Sugars present</th>
<th>Location of glycans</th>
<th>COH structure CARBANK Accession numbers</th>
<th>COH %</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BACTERIAL</strong>&lt;br&gt;cont...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pseudomonas fluorescens var. cellulosa</em></td>
<td>EngA</td>
<td>ND</td>
<td>Glc, Gal, Man, Fuc, amino sugars</td>
<td>ND</td>
<td>ND</td>
<td>3.3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>EngB</td>
<td>ND</td>
<td>Glc, Gal, Man, Fuc, amino sugars</td>
<td>ND</td>
<td>ND</td>
<td>11.3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>EngC</td>
<td>ND</td>
<td>Glc, Gal, Man, Fuc, amino sugars</td>
<td>ND</td>
<td>ND</td>
<td>8.5</td>
<td>9</td>
</tr>
<tr>
<td><em>Streptomyces lividans</em></td>
<td>intracellular β-1,4-glucosidase xylanase xylanase B</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>11, 17</td>
</tr>
<tr>
<td><em>Streptomyces lividans</em> recombinant</td>
<td>Cex from <em>C. fimii</em></td>
<td>T/S-Man or T/S-Gal</td>
<td>Man, Gal</td>
<td>linker</td>
<td>ND</td>
<td>5.4</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>E3 from <em>T. fusca</em></td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>2.5</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>E2 from <em>T. fusca</em></td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>&lt;1%</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Cellobiohydrolase CBHII from <em>Microbispora bispora</em></td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>~12%</td>
<td>7,12</td>
</tr>
<tr>
<td><em>Thermomonospora fusca</em></td>
<td>β-1,4-endoglucanase E2</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>2</td>
<td>1, 19</td>
</tr>
<tr>
<td></td>
<td>β-1,4-exoglucanase E3</td>
<td>ND</td>
<td>ND</td>
<td>likely on the linker but possibly on CBD</td>
<td>ND</td>
<td>2.5-5</td>
<td>19</td>
</tr>
</tbody>
</table>
Table 1.2 continued

<table>
<thead>
<tr>
<th>Organism</th>
<th>Protein</th>
<th>Linkage</th>
<th>Sugars present</th>
<th>Location of glycans</th>
<th>COH structure CARBANK Accession numbers</th>
<th>COH %</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BACTERIAL cont...</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pseudomonas fluorescens var. cellulosa</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EngA</td>
<td>ND</td>
<td>Glc, Gal, Man, Fuc, amino sugars</td>
<td>ND</td>
<td>ND</td>
<td></td>
<td>3.3</td>
<td>9</td>
</tr>
<tr>
<td>EngB</td>
<td>ND</td>
<td>Glc, Gal, Man, Fuc, amino sugars</td>
<td>ND</td>
<td>ND</td>
<td></td>
<td>11.3</td>
<td>9</td>
</tr>
<tr>
<td>EngC</td>
<td>ND</td>
<td>Glc, Gal, Man, Fuc, amino sugars</td>
<td>ND</td>
<td>ND</td>
<td></td>
<td>8.5</td>
<td>9</td>
</tr>
<tr>
<td><em>Streptomyces lividans</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intracellular β-1,4-glucosidase xylanase</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>11, 17</td>
</tr>
<tr>
<td>xylanase B</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>10, 11</td>
</tr>
<tr>
<td><em>Streptomyces lividans</em> recombinant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cex from <em>C. fimicola</em></td>
<td>T/S-Man or T/S-Gal</td>
<td>Man, Gal</td>
<td>linker</td>
<td>ND</td>
<td></td>
<td>5.4</td>
<td>12</td>
</tr>
<tr>
<td>E3 from <em>T. fusca</em></td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>2.5</td>
<td>19</td>
</tr>
<tr>
<td>E2 from <em>T. fusca</em></td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>&lt;1%</td>
<td>5</td>
</tr>
<tr>
<td>Cellulohydrolase</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>−12%</td>
<td>7, 12</td>
</tr>
<tr>
<td>CBHI from <em>Microbispora bispora</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Thermomonospora fusca</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-1,4-endoglucanase E2</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>2.5-5</td>
<td>19</td>
</tr>
<tr>
<td>β-1,4-exoglucanase E3</td>
<td>ND</td>
<td>ND</td>
<td>likely on the linker but possibly on CBD</td>
<td>ND</td>
<td>ND</td>
<td>2</td>
<td>1, 19</td>
</tr>
<tr>
<td>Organism</td>
<td>Protein</td>
<td>Linkage</td>
<td>Sugars present</td>
<td>Location of glycans</td>
<td>COH structure CARBANK Accession numbers</td>
<td>COH %</td>
<td>Reference</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td>---------</td>
<td>-------------------------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------------</td>
<td>--------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>FUNGAL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Trichoderma reesei</em></td>
<td>cellbiohydrolase</td>
<td>O-Man</td>
<td>O: Man</td>
<td>O-linked mostly in PT linker N-linkages on catalytic domain</td>
<td>ND</td>
<td>1.4-10.4</td>
<td>6, 8, 16</td>
</tr>
<tr>
<td>CBHI</td>
<td></td>
<td>N-linkages</td>
<td>N: Man, GlcNAc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>cellbiohydrolase</td>
<td>ND</td>
<td>ND</td>
<td>O-linked mostly in PT linker N-linkages on catalytic domain</td>
<td>ND</td>
<td>ND</td>
<td>18</td>
</tr>
<tr>
<td>CBHII</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>endoglucanase</td>
<td></td>
<td>ND</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGI</td>
<td>endoglucanase</td>
<td>ND</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EGIII</td>
<td>ND</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>endoglucanase</td>
<td></td>
<td>ND</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGV</td>
<td>likely N and O linked</td>
<td>ND</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>

Abbreviations: CBD, cellulose-binding domain; COH, carbohydrate; ND, not determined; Fuc, fucose; Gal, galactose; Glc, glucose; GlcNAc, N-acetylglactosamine; Man, mannose; p, pyranose; f, furanose; S, serine; T, threonine

1. (Calza et al., 1985) 11. (Moens and Vanderleyden, 1997)
2. (Gerwig et al., 1993) 12. (Ong et al., 1994)
3. (Gerwig et al., 1992) 13. (Plummer et al., 1995)
4. (Gerwig et al., 1991) 14. (Reinhold et al., 1995)
5. (Ghangas and Wilson, 1988) 15. (Saloheimo et al., 1994)
6. (Hemming, 1995) 16. (Salovuori et al., 1987)
7. (Hu et al., 1993) 17. (Sandercocck et al., 1994)
8. (Kubicek, 1992) 18. (Teeri et al., 1992)
9. (Langford, 1988) 19. (Zhang et al., 1995)
10. (Messner, 1997a)
1.3.8. *Glycosylation of CenA and Cex*

Langsford was the first to examine the glycosylation of CenA and Cex produced by *C. fimii* (Langsford, 1988). There are N-glycosylation sites but no N-glycans were believed to be present because the molecular weight of CenA or Cex were not modified by treatment with Endo-H, an enzyme used to remove mammalian N-linked glycans. Carbohydrates could be released by alkali, an indication of O-linked glycosylation. Mannose was the only sugar constituent detected. It was hypothesized that the glycans were attached to the linker region of the enzyme. Ong *et al.* later determined, for Cex recombinantly produced in *S. lividans*, that glycosylation occurred on the PT linker. Mannose and galactose were detected (Ong *et al.*, 1994).

1.3.9. *Function of glycosylation, with an emphasis on cellulases*

A comprehensive review by Dwek looks at the function of sugars of glycoproteins (Dwek, 1996). There it is stressed that no one function can be ascribed to oligosaccharides. Possible roles include use as a recognition marker or modification of properties such as susceptibility to proteases, stability, or quaternary structure.

For cellulosomes and secreted cellulases, no definitive role for glycosylation has been identified. For each example of a putative role, an exception can be found. A role for glycosylation in secretion of cellulases of *T. reesei* has been debated (Hemming, 1995), but it is unlikely that glycosylation is required for secretion of cellulases from either *C. fimii* or *T. fusca* because not all the secreted cellulases of these systems are glycoproteins. It is possible that glycans may be involved in interactions with the substrate (Sandercock *et al.*,...
Glycosylated PTCBD\textsubscript{Cex}, made in \textit{S. lividans}, had greater affinity for cellulose than non-glycosylated PTCBD\textsubscript{Cex} produced in \textit{E. coli} (Ong et al., 1994), but the ability of CipA, a scaffolding protein from \textit{C. thermocellum}, to bind its target proteins was not affected by glycosylation (Beguin \textit{et al.}, 1992). For CenA and Cex from \textit{C. fimi} and E3 from \textit{T. fusca}, the glycosylated and non-glycosylated proteins also showed no difference in their ability to bind cellulose. Their kinetic properties, thermostability, and pH stability were also unaffected (Langsford, 1988; Zhang \textit{et al.}, 1995). Glycosylation has been reported to have an impact on sensitivity to proteolysis. For example \textit{C. fimi} proteins glycosylated by \textit{C. fimi} or, in the case of Cex, also \textit{S. lividans}, are protected against the action of a \textit{C. fimi} protease when adsorbed to cellulose but to a lesser degree in solution. The same cannot be said of glycosylated E3 from \textit{T. fusca}, which retained its sensitivity to proteolysis (Langsford, 1988; Ong \textit{et al.}, 1994; Zhang \textit{et al.}, 1995). Thus, the role of glycosylation may very well be peculiar to specific proteins, and settings.

1.3.10. Technological advances in carbohydrate analysis

There is no single method for the characterization of all oligosaccharides so a combination of methods is required (Raju \textit{et al.}, 1996). Within the time frame of my research for this thesis, the technology for analyzing glycoproteins, and particularly carbohydrates, has significantly advanced. Most notably, there have been developments which make carbohydrate analysis faster and more accessible to the molecular biologist. These include the automation of carbohydrate sequencing for N-linked proteins (Dwek, 1996; Rudd \textit{et al.}, 1997), mass spectrometry (Burlingame, 1996; Costello, 1997; Harvey \textit{et al.}...
al., 1996) and fluorophore-assisted carbohydrate electrophoresis (FACE®) technology (Jackson, 1996; Raju et al., 1996).

1.4. IgA1 and IgA2 proteases

1.4.1. The structure and function of IgA

More immunoglobulin A (IgA) is produced in humans than all of the other immunoglobulins combined (Underdown and Mestecky, 1994). IgA is found in human serum and is, more importantly, the major antibody in secretions where it has an important role in mucosal immunity. The structure and function of IgA has been reviewed previously (Kerr, 1990; Lamm et al., 1995). Electron microscopy confirmed biochemical analysis which proposed that monomeric IgA has the common four-chain structure of other immunoglobulins: two light chains and two heavy chains (Svehag and Bloth, 1970). The latter are comprised of four domains: a variable domain and three constant domains. A proline-rich hinge region separates heavy constant domains 1 and 2. IgA has two isotypic forms, IgA1 and IgA2. IgA1 contains the full length hinge region of twenty amino acids which has an octapeptide direct repeat (P-S-T-P-P-T-P-S) (Figure 1.6), whereas IgA2 has two allotypes, both of which lack 13 amino acids of this hinge region (Kerr, 1990). IgA has a variety of molecular forms which are distributed in various bodily fluids in a defined pattern. Ninety percent of IgA is in the IgA1 form in serum. The percentage in the IgA1 form in other body fluids varies between thirty-five and ninety-five percent depending on location (Delacroix et al., 1982; Kerr, 1990; Kett et al., 1986). Serum IgA1 is primarily monomeric, a characteristic particular to humans (Vaerman, 1973). Secretory IgA1 is a
product of two cell types, mucosal lymphocytes and epithelial cells. Two molecules of IgA1 and the J chain glycopeptide, produced by mucosal lymphocytes, covalently associate. This complex passes through the epithelial cells lining the mucosal surfaces via a receptor-mediated secretion process. During this process a component of the polymeric immunoglobulin receptor, a glycopeptide referred to as SC, becomes covalently attached to the dimeric IgA1-J chain complex. Secretory IgA1 is therefore quite different from serum IgA1 being primarily dimeric with two covalently attached glycopeptides (Brandtzaeg, 1995; Kerr, 1990; Nezlin, 1998).

IgA1 has both N and O-linked carbohydrates. There are two potential N-linked and five potential O-linked sites. The latter are all located in the hinge region. The N-linked carbohydrates have been described previously and will not be detailed here (Tomana et al., 1972). The O-glycosylation patterns differ between myeloma and pooled serum IgA1. Myeloma IgA1, a monoclonal antibody, has minimal heterogeneity of the glycan component whereas serum IgA1 is polyclonal and has greater heterogeneity (Mattu et al., 1998) (Figure 1.6). Secreted IgA1 isolated from milk has more complex and heterogeneous oligosaccharides, which may include NeuAc, Fuc and GlcNAc (McGuire et al., 1989; Pierce-Cretel et al., 1981).

The roles of IgA are not completely understood, and likely differ between locations in the body. In secretions, the major, and most important role of IgA is to inhibit the binding of micro-organisms to mucosal surfaces. This is achieved through a number of strategies, including agglutination. Other proposed functions of secretory IgA include trapping antigens in the mucosal layer of membranes, inhibiting antigen penetration through these membranes, neutralizing bacterial toxins and enzymes, and enhancing non-
specific antibacterial factors in secretions (Kilian et al., 1996). Serum IgA is believed to be involved in the removal of antigenic material without triggering inflammation (Kilian et al., 1996). While the activation of complement by IgA was controversial (Kerr, 1990), it is now thought that IgA does not activate complement by either the classical or alternative pathways (Kilian and Russell, 1994; Nikolova et al., 1994). IgA has few effector functions. Most are dependent on the Fc portion of the antibody and thus require IgA to be intact (Kilian et al., 1996). While secreted IgA is regarded as having an important role in immunity, the importance of serum IgA has been questioned because IgA deficiency is not always associated with increased susceptibility to infection (Kerr, 1990). However, increased serum IgA concentration is associated with auto-immune diseases such as rheumatic diseases, some liver diseases and persistent infections such as AIDS (Elkon et al., 1983; Kalsi et al., 1983; Procaccia et al., 1987).
Figure 1.6: Glycosylation of the hinge region of human myeloma serum IgA1 and human pooled serum IgA1.

Hinge region of human myeloma serum IgA1 (Panel A) and human pooled serum IgA1 (Panel B). The underlined amino acids indicate sites that are partially occupied. The boxed sequences are the direct repeats. Abbreviations: GalNAc, N-acetylglactosamine; Gal, galactose; NeuAc, N-acetylneuraminic acid (sialic acid). (Lomholt, 1996; Mattu et al., 1998).
1.4.2. *IgA1 proteases*

IgA1 proteases are putative virulence factors for some pathogenic bacteria and have been comprehensively reviewed (Kilian *et al*., 1996; Lomholt, 1996). The IgA1 protease family has representatives from the metallo-, serine- and cysteine-protease families. Various bacteria implicated in diseases originating from mucosal surfaces constitutively produce IgA1 proteases (Table 1.3). Proteolysis occurs in the hinge region at post-proline peptide bonds in human IgA1. IgA2 is not a substrate for most of the proteases as it lacks 13 amino acids in the hinge. Only one other bacterial endopeptidase, that from *Flavobacterium meningosepticum*, is known to cleave post-proline bonds but it does not cleave IgA1 (Lomholt, 1996; Walter *et al*., 1980).

The specific site of proteolysis in the IgA1 hinge varies between IgA1 proteases but is well defined for each enzyme. Most of the enzymes cleave within the sequence containing the direct octapeptide repeats. It is remarkable that the proteases cleave at only one of many identical peptide bonds. Some bacteria have strains which produce proteases with different specificities (Figure 1.7). Cleavage between proline and serine (P/S) is designated type 1 and cleavage between proline and threonine (P/T) is designated type 2.

The cleavage specificity of *Neisseria* type 2 IgA1 proteases has been exploited *in vitro* by using a consensus sequence as a processing site for recombinant proteins (Pohlner *et al*., 1992). The consensus sequence for *Neisseria* type 2 IgA1 proteases has been defined as $X_1P^*X_2P$ where $X_1$ can be P or rarely P-A, P-G, or P-T; $X_2$ can be T, S, or A; and $^*$ represents the cleavage position.
Table 1.3: IgA1 protease producing organisms implicated in disease.

The * indicates that there is evidence of IgA1 protease activity in vivo. The table is based on Table 1 in Lomholt, 1996 (Lomholt, 1996).

<table>
<thead>
<tr>
<th>DISEASE</th>
<th>ORGANISM(S) IMPLICATED IN DISEASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caries, endocarditis, atopic disease</td>
<td>* Streptococcus sanguis, Streptococcus oralis, and Streptococcus mitis</td>
</tr>
<tr>
<td>Periodontis</td>
<td>* Capnocytophaga and Prevotella species</td>
</tr>
<tr>
<td>Meningitidis, respiratory disease,</td>
<td>* Neisseria meningitidis, Haemophilus influenzae, and Streptococcus pneumoniae</td>
</tr>
<tr>
<td>sinusitis, otitis media</td>
<td></td>
</tr>
<tr>
<td>Gonorrhea, cystitis</td>
<td>* Neisseria gonorrhoeae and Ureaplasma urealyticum</td>
</tr>
<tr>
<td>Conjunctivitis</td>
<td>* H. influenzae and N. gonorrhoeae</td>
</tr>
<tr>
<td>Brazilian purpuric fever</td>
<td>H. influenzae biogroup aegyptius</td>
</tr>
</tbody>
</table>

Structural features beyond the linear sequence may influence cleavage of proteins by proteases. Protein folding, length of substrate, interaction with other domains, spacing between recognition and cleavage sites, and glycosylation all have the potential to impact the efficiency of cleavage (Lomholt, 1996; Qiu et al., 1996). No single set of rules covers all IgA1 proteases. It has been proposed that mature N. gonorrhoeae and H. influenzae proteases have a cleavage-specificity domain (CSD) which differs in length between strains. The size of the CSD is proportional to the distance from the disulphide bridge
Figure 1.7: Cleavage sites in the hinge region of myeloma serum IgA1 of some IgA1 proteases.

The boxed sequences are the direct repeats. Abbreviations: GalNAc, N-acetylglactosamine; Gal, galactose. (Lomholt, 1996)
between the IgA1 alpha chains to the specific bond cleaved in the hinge. As these enzymes are thought to recognize and interact with parts of the Fcα region, the CSD may act as a spacer between the catalytic site and substrate binding site. Alternatively, the conformation of the CSD may cause interference with access to other potential cleavage sites (Lomholt, 1996; Qiu et al., 1996). Glycosylation of IgA1 affects the rate of cleavage by the streptococcal IgA1 proteases. De-glycosylated IgA1 was a poor substrate while IgA1 lacking the terminal sialic acids was a superior substrate (Reinholdt et al., 1990).

The functions of IgA1 proteases are not yet well understood. Studies have indicated that some bacterial IgA1 proteases are active in vivo during infection or colonization, but a definite role has not been established due to lack of an appropriate animal model (Lomholt, 1996) (Table 1.3). A recent development was the report of a male human challenge model for N. gonorrhoeae (Cannon, 1998). The major limitation of this model was the short time-line allowed for disease which did not allow the question of IgA1 protease function in disease maintenance to be addressed. Indirect evidence for an important role for IgA1 proteases in disease has been reviewed (Kilian et al., 1996). In brief the evidence is: IgA1 protease activity has evolved convergently from at least three independent evolutionary lines as IgA1 proteases represent a variety of enzyme families; three species of pathogenic bacteria which cause meningitis all produce IgA1 proteases, while closely related non-pathogenic species do not, although this observation is not true of all IgA protease producing bacteria; and IgA1 is important for immunity of the mucosal surfaces that are colonized by most of the IgA1 protease-producing bacteria. In vivo and in vitro hydrolysis of IgA1 results in two intact fragments, the C-terminus of the heavy chains
(Fc<sub>a</sub>) and the antigen binding fragment (Fab<sub>a</sub>) (Fig 1.8). The Fc<sub>a</sub> portion is responsible for the effector properties conferred by IgA1 thus IgA1 proteases effectively separate the antigen-binding and effector functions of IgA1. Another possible function for IgA1 proteases was put forth in 1987 (Kilian and Reinholdt, 1987) and later described (Kilian et al., 1996; Lomholt, 1996): IgA1 proteases cleave cross-reactive antibodies present from previous exposure to similar antigens. The resulting Fab<sub>a</sub> fragments of these cross-reactive antibodies might coat the surface of the pathogen, protecting it from recognition by intact antibodies.

IgA1 proteases have very narrow substrate ranges which vary with the particular protease. Three substrates are cleaved efficiently by serine IgA1 proteases from Neisseria and H. influenzae in vivo: human serum and secretory IgA1; and importantly, the precursor molecule of the IgA1 proteases themselves. The serine IgA1 proteases are the prototype for autotransporter proteins which contain within their precursor the protein component
which mediates their transport. Autotransporter proteins have been reviewed recently (Henderson et al., 1998). The IgA1 precursor contains at least five components: a leader peptide which targets the protein to the periplasmic space; the β domain which creates a pore structure through which the remaining domains pass; α and γ peptide domains which have proline-rich sequences; and the mature protease which folds and cleaves in the α and γ peptide domains. Cleavage releases the mature protein and the α and γ peptides domains from the bacterial cell outer membrane.

Alternative substrates for IgA1 proteases have been reported. In vivo cleavage has been demonstrated for IgA1 from chimpanzees, gorillas, and orangutans (Kerr, 1990; Qiu et al., 1996). There are also reports of cleavage of non-IgA1 substrates by IgA1 proteases in vitro. It needs to be stressed that the alternative substrates were not cleaved nearly as efficiently as IgA1. Reported substrates include: granulocyte/macrophage colony stimulating factor, which has functions including the up-regulation of Fcα receptors on neutrophils (Lomholt, 1996; Weisbart et al., 1988); LAMP1, a major integral membrane glycoprotein of late endosomes and lysosomes believed to help maintain the stability of these cellular compartments (Lin et al., 1997); synaptobrevin II, a protein essential for exocytosis in neurons and chromaffin cells, which contains a putative IgA1 protease cleavage site (Binscheck et al., 1995); and a hybrid protein, CenAIgA1h, where the linker of the endoglucanase CenA has been replaced by the IgA1 hinge region (Miller et al., 1992).

CenAIgA1h, produced in E. coli, was cleaved by N. gonorrhoeae type 1 and 2 proteases. However, CenA was not cleaved, despite the sequence similarity to the IgA1
hinge region (Figure 1.9) (Miller et al., 1992).

Panel A, IgA1 hinge; Panel B, CenA PT linker.

Figure 1.9: The sequence of the IgA1 hinge compared to the PT linker of CenA
1.5. Objectives

The main objective of my thesis research was to further investigate the structure, function and possible applications of the interdomain linker of CenA. More specific objectives were:

1. Modify the linker region of CenA and examine the properties, including activity and stability, of the linker mutant proteins expressed in *E. coli* and *S. lividans*.

2. Further characterize the glycosylation of CenA and Cex from *C. fimii* including definitively determining the location of glycans, re-addressing monosaccharide composition and quantity, and the role of glycosylation.

3. Characterize the glycosylation of CenA linker mutants produced in *S. lividans*.

4. Utilize glycosylated and non-glycosylated CenA and CenAIgA1h as potential substrates for IgA1 proteases to aid in determining the specificity of these enzymes.
2. Methods and Materials

2.1. Chemicals, buffers, substrates and enzymes.

Unless otherwise noted, chemicals were of analytical or high pressure liquid chromatography (HPLC) grade and purchased from Sigma (St. Louis, MO) or BDH (Toronto, ON). Solutions and buffers were prepared as described by Sambrook et al, 1989 (Sambrook et al., 1989) and were sterilized by filtration or autoclaving.

Avicel™ PHI01 (micro-crystalline cellulose) was obtained from FMC International (Cork, Ireland); bacterial micro-crystalline cellulose (BMCC) was produced by E. Kwan, Department of Microbiology and Immunology, UBC by a method previously described (Hestrin, 1963); phosphoric acid swollen cellulose (PASC) was made from Avicel™ PH101 by E. Kwan, Department of Microbiology and Immunology, UBC as described previously (Wood, 1988); para-nitrophenyl β-D-cellobioside (PNPC) and carboxymethylcellulose were from Sigma; 2’,4’-dinitrophenyl β-D-cellobioside (2,4-DNPC) was a gift from K. Rupitz and S. Withers from the Department of Chemistry, UBC

Immobilized papain was purchased from Pierce (Rockford, IL); C. fimi protease was produced from C. fimi ATCC 484 as described previously (Gilkes et al., 1988); N-tosyl-L-phenylalanylchloromethyl ketone (TPCK) trypsin was from Sigma.

Human serum IgG1 myeloma paraprotein from patient Vie (IgG1) (Qiu et al., 1996) and IgG1 proteases were gifts, as part of a collaboration, from A. Plaut and J. Qiu as described on page xx.
2.2. Bacterial strains and plasmids

The *E. coli*, *C. fimi*, and *S. lividans* strains used in this study are listed in Table 2.1.

Bacterial stocks were maintained at -70°C in media with 7% dimethylsulfoxide (DMSO) or 10% glycerol. Bacterial cells harboring constructs were stored at -70°C in media with 7% DMSO. DNA was stored in Tris-HCl/ethylenediaminetetraacetic acid (EDTA) buffer or water at -20°C. The *E. coli* plasmids and the *E. coli/S. lividans* shuttle plasmids used in this study are listed in Table 2.2. and Table 2.3 respectively.

**Table 2.1: Bacterial strains.**

<table>
<thead>
<tr>
<th>STRAIN</th>
<th>GENOTYPE</th>
<th>REFERENCE or SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH5α</td>
<td>F&lt;sup&gt;-&lt;/sup&gt; endA1, hsdR17 (rk, mk&lt;sup&gt;+&lt;/sup&gt;), supE44, thi-1, recA1, (argF-laczya)U169, φlacZ 15</td>
<td>(Hanahan, 1983)</td>
</tr>
<tr>
<td>JM101</td>
<td>supE thi Δ (lac-proAB)[F' traΔ36 proAB laq&lt;sup&gt;+&lt;/sup&gt;ΔM15]</td>
<td>(Yanisch-Perron C. <em>et al.</em>, 1985)</td>
</tr>
<tr>
<td><em>C. fimi</em></td>
<td>wild type</td>
<td>ATCC 484</td>
</tr>
<tr>
<td><em>S. lividans</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TKM31</td>
<td>a natural mutation of TK24, a basic strain of <em>S. lividans</em> 66 from D. A. Hopwood, selected for low protease activity</td>
<td>A gift from D. B. Wilson, Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York</td>
</tr>
</tbody>
</table>
### Table 2.2: *E. coli* plasmids.

<table>
<thead>
<tr>
<th>PLASMID</th>
<th>RELEVANT CHARACTERISTICS</th>
<th>REFERENCE OR SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC18-1.5cenAΔPT</td>
<td>Amp&lt;sup&gt;+&lt;/sup&gt;, the PT linker of <em>cenA</em> has been deleted</td>
<td>A gift from H. Shen, Department of Microbiology and Immunology, UBC (Shen, 1990)</td>
</tr>
<tr>
<td>pUC18-1.5cenAigAlh</td>
<td>Amp&lt;sup&gt;+&lt;/sup&gt;, the <em>iga1</em> hinge replaced the PT linker in <em>cenA</em></td>
<td>A gift from P. Miller, Department of Microbiology and Immunology, UBC (Miller <em>et al.</em>, 1992)</td>
</tr>
<tr>
<td>pUC18-1.5cenAΔPTAS</td>
<td>pUC18-1.5cenAΔPT with <em>A</em>geI and SpeI sites flanking the deletion to allow the insertion of mutant linkers</td>
<td>This study</td>
</tr>
<tr>
<td>pTAL3</td>
<td>Amp&lt;sup&gt;+&lt;/sup&gt;, pTZ18R/PTIS with <em>cenB</em></td>
<td>A gift from A. Meinke, Department of Microbiology and Immunology, UBC (Meinke <em>et al.</em>, 1991a)</td>
</tr>
<tr>
<td>pUC18-1.5cenA1.5PT, pUC18-1.5cenA2PT, pUC18-1.5cenAFn3</td>
<td>first intermediates in construction of pTUg and pIJ680 variants; derived from pUC18-1.5cenAΔPTAS</td>
<td>This study</td>
</tr>
<tr>
<td>pSL1180</td>
<td>plasmid with superpolylinker used in this study for subcloning</td>
<td>Pharmacia, (Brosius, 1989)</td>
</tr>
<tr>
<td>pSL1180-cenAΔPT, pSL1180-cenAigAlh, pSL1180-cenA1.5PT, pSL1180-cenA2PT, pSL1180-cenAFn3</td>
<td>intermediates in construction of pTUg and pIJ680 variants</td>
<td>This study</td>
</tr>
<tr>
<td>pTUgKRG-1.5cenA.N</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt;, <em>lacI</em>&lt;sub&gt;q&lt;/sub&gt;, <em>tac</em> promoter, <em>cel</em> leader peptide sequence followed by a sequence encoding a six histidine tail with and IEGR site and wild type <em>cenA</em>; high expression vector</td>
<td>A gift from H. Damude, Department of Microbiology and Immunology, UBC (Damude, 1995)</td>
</tr>
<tr>
<td>pTUgKRG-cenAΔPT, pTUgKRG-cenAigAlh, pTUgKRG-cenA1.5PT, pTUgKRG-cenA2PT, pTUgKRG-cenAFn3 pTUgKRG-&lt;i&gt;cenA&lt;/i&gt;1.5PTmod</td>
<td>pTUgKRG-1.5cenA.N where <em>cenA</em> was replaced with <em>cenA</em> mutants</td>
<td>This study</td>
</tr>
</tbody>
</table>
Table 2.3: *S. lividans/E. coli* shuttle plasmids.

<table>
<thead>
<tr>
<th>PLASMID</th>
<th>RELEVANT CHARACTERISTICS</th>
<th>REFERENCE OR SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>pIJ680-CBD&lt;sub&gt;cex&lt;/sub&gt;</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, Tsr&lt;sup&gt;R&lt;/sup&gt;, an <em>S. lividans</em> /<em>E. coli</em> shuttle vector with <em>aph</em> promoter sequence, <em>E. coli</em> origin of replication, <em>cex</em> leader peptide sequence and the <em>cexCBD</em> gene; a cloning and expression vector for use in <em>Streptomyces</em></td>
<td>A gift from E. Ong, Department of Microbiology and Immunology, UBC (Ong et al., 1994)</td>
</tr>
<tr>
<td>pIJ680-cenAΔPT, pIJ680-cenAlgA1h, pIJ680-cenA</td>
<td>pIJ680-CBD&lt;sub&gt;cex&lt;/sub&gt; where CBD&lt;sub&gt;cex&lt;/sub&gt; was replaced with cenA mutants</td>
<td>This study</td>
</tr>
<tr>
<td>pIJ680 -cenA2PT, pIJ680-cenAFn3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.3. Media and growth conditions

*E. coli* strains were grown in Luria broth (LB) (Sigma) at 37° C for DNA manipulations and Terrific broth® (Sigma) at 30° C for protein production. Media were supplemented with either 50 µg kanamycin or 100 µg ampicillin mL<sup>-1</sup> and, for recombinant protein induction, isopropyl-β-D-thiogalactoside (IPTG) to a final concentration of 0.1 mM. 1.5% agar was added to make solid media.

*C. fimi* was grown in modified Leatherwood medium (1 g K<sub>2</sub>HPO<sub>4</sub>, 0.5 g KCl, 0.5 g MgSO<sub>4</sub>, 1 g NaNO<sub>3</sub>, 5 g peptone and 5 g yeast extract per Litre; pH 7.2) (Stewart and Leatherwood, 1976) supplemented with 0.1 % (w/v) carbon source at 30° C and shaken at 250 revolutions minute<sup>-1</sup> (rpm). Solid medium used for *C. fimi* was low salt LB (0.5 g NaCl L<sup>-1</sup>) with 1.5% agar.
*S. lividans* was grown in tryptic soy broth (TSB) (BDH), pH 7.2, supplemented with 5 μg thiostrepton (Tsr) mL⁻¹ at 30°C in baffled flasks at 250 rpm to increase aeration. To prevent foaming during protein production, 1 drop (about 30 μL) of antifoam 289 from Sigma was added to 500 mL. Solid medium used for *S. lividans* was TSB with 50 μg Tsr mL⁻¹ and 1.5 % agar.

### 2.4. Recombinant DNA techniques

Recombinant deoxyribonucleic acid (DNA) techniques for *E. coli* were as described in Sambrook *et al.*, 1989 (Sambrook *et al.*, 1989) and for *S. lividans* as described in Hopwood *et al.*, 1985 (Hopwood *et al.*, 1985) unless otherwise specified.

Restriction endonucleases, ligases, polymerases and nucleotides were from New England BioLabs (Beverly, MA), Boehringer Mannheim Canada, or Gibco BRL (Gaithersburg, MD), and used as recommended by the suppliers.

DNA was extracted from agarose gels using either the Geneclean® II kit (BIO101, Inc., Vista, CA) or Qiaex® II kit (Qiagen, Chatsworth, CA) (Qiagen Inc., 1995). *E. coli* plasmid DNA was prepared using the Qiagen system (Qiagen) (Crowe, 1992). *S. lividans* plasmid DNA was prepared by a modification of the *E. coli* alkaline lysis method (Birnboim and Doly, 1979), developed by the Jensen Laboratory (University of Alberta, AB); where 1.5 mL of culture is spun to recover the mycelia which are then resuspended in 100 μL of 50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA and 2 mg lysozyme mL⁻¹; dispersed with a vortex mixer and incubated at 37°C for 30 minutes prior to the alkaline lysis method. Ligations were performed with a DNA insert to vector ratio of approximately 10:1 in a total volume of 12 μL and incubated at 16°C overnight. Ligated
DNA was desalted by butanol precipitation (Thomas, 1994). Electroporation, with a GenePulser electroporator (Bio-Rad, Mississauga, ON), was used to transform *E. coli*. *E. coli* DH5α was used for DNA manipulations and JM101 for gene expression. *S. lividans* protoplasts were transformed by polyethylene glycol-mediated DNA uptake.

2.4.1. Primer design for introducing silent mutations

The program Reverse Translator® written by D. Trimbur, Department of Microbiology and Immunology, UBC was used to determine the silent mutations that could be introduced flanking the PT linker of CenA.

2.4.2. Synthetic oligonucleotides and oligonucleotide primers

Synthetic oligodeoxyribonucleotides and oligodeoxyribonucleotide primers were synthesized by the UBC Nucleic Acid and Protein Synthesis Unit (NAPS) with an Applied Biosystem DNA synthesizer and purified by precipitation with n-butanol (Sawadogo and van Dyke, 1991). The oligonucleotides used in this study are described in Tables 2.4, 2.5 and 2.6.

2.4.3. DNA sequencing

The AmpliTaq polymerase dye termination protocol with the addition of 5% DMSO was used to sequence DNA because of the high GC content of *C. fimi* DNA. The sequencing service was provided by the UBC NAPS unit using an Applied Biosystem DNA sequencer Model 377 (Perkin-Elmer, Norwalk, CT).
Table 2.4: Oligonucleotide primers used for PCR.

<table>
<thead>
<tr>
<th>NAME</th>
<th>OLIGONUCLEOTIDE SEQUENCE (5’-3’)</th>
<th>SOURCE AND IMPORTANT CHARACTERISTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSMUTA</td>
<td>GGG TCG ACG TAG AAG CCA CTA GTC GGC TGC GGC GTG ACG GTA ACG CTG GTC GTC GGC ACG GTA CCG GTG CAG GTG GTG CCG TT</td>
<td>This study; introduces silent mutations forming AgeI and SpeI sites flanking the PT linker</td>
</tr>
<tr>
<td>LSMUTC</td>
<td>AAC GGC ACC ACC TGC ACC GGT ACC GTG CCG ACG ACC AGC GTT AAC GTC ACG CCG CAG CCG ACT AGT GCC TTC TAC GTC GAC CC</td>
<td>This study; introduces silent mutations forming AgeI and SpeI sites flanking the PT linker</td>
</tr>
<tr>
<td>SAENH</td>
<td>AGG TCT ACT AGT CCC GGC TGC CGC GTC GAC</td>
<td>A gift from P. Tomme, Department of Microbiology and Immunology, UBC</td>
</tr>
<tr>
<td>CenAR</td>
<td>TCA CCA CCT GGC GTT GCG CGC CAT</td>
<td>A gift from H. Damude, Department of Microbiology and Immunology, UBC</td>
</tr>
<tr>
<td>FNIIIDN</td>
<td>GAA GCC ACT AGT CGG GTG ACG TGT TAC CGA CGG GTT GCT CGT CAC CGG CAG CGG ACT</td>
<td>This study: introduces a SpeI site just downstream of the CenBFn3 repeats</td>
</tr>
<tr>
<td>FNIIIP</td>
<td>CCT GCA CCG GTA CTG TAC CGA CCG GGA CGA CGA CGG AC</td>
<td>This study: introduces a AgeI site just upstream of the CenBFn3 repeats</td>
</tr>
<tr>
<td>PTCCP</td>
<td>GAA GCC ACT AGT CGG CTG AGG TGT TGG CGT AG</td>
<td>This study: introduces a P (bold and underlined) at the C terminus of the PT linker</td>
</tr>
</tbody>
</table>

Table 2.5: Oligonucleotide primers used for sequencing.

<table>
<thead>
<tr>
<th>NAME</th>
<th>OLIGONUCLEOTIDE SEQUENCE (5’-3’)</th>
<th>SOURCE AND IMPORTANT CHARACTERISTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CenAP3</td>
<td>GTC ACG ATC ACC AAC CT</td>
<td>A gift from A. Meinke, Department of Microbiology and Immunology, UBC</td>
</tr>
<tr>
<td>FNIIIDN</td>
<td>as in Table 2.4</td>
<td></td>
</tr>
<tr>
<td>FNIIIP</td>
<td>as in Table 2.4</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.6: Synthetic oligonucleotides for constructing mutant PT linkers.

<table>
<thead>
<tr>
<th>NAME</th>
<th>OLIGONUCLEOTIDE SEQUENCE (5'-3')</th>
<th>SOURCE AND IMPORTANT CHARACTERISTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2PTN</td>
<td>CCT GCA CCG GTA CTG TA</td>
<td>This study</td>
</tr>
<tr>
<td>2PTNC</td>
<td>GTC GGT ACA GTA CCG GTG CAG G</td>
<td>This study: 5’phosphorylated</td>
</tr>
<tr>
<td>2PT1</td>
<td>CCG ACT ACC TCA CCT ACA CCA ACT</td>
<td>This study: 5’phosphorylated</td>
</tr>
<tr>
<td>2PT1C</td>
<td>GTC GGA GTT GTC GGC GTA GGA GTT GGT GTA GGT GAG GTA</td>
<td>This study: 5’phosphorylated</td>
</tr>
<tr>
<td>2PT2</td>
<td>CCG AGC CCA ACT CCA ACC CCG ACA CCA ACA CCT ACC CCT ACG</td>
<td>This study: 5’phosphorylated</td>
</tr>
<tr>
<td>2PT2C</td>
<td>GGG TAG GTG TTG GTG TCG GGG TTG GAG TTG GC</td>
<td>This study: 5’phosphorylated</td>
</tr>
<tr>
<td>2PT3</td>
<td>CCT ACC ACT TCG CCG ACT CCT ACT CCG ACG CCG ACC ACG</td>
<td>This study: 5’phosphorylated</td>
</tr>
<tr>
<td>2PT3C</td>
<td>GTC GGC GTG GTC GGC GGA GTA GGA GTC GGC GAA GTG GTA GGC GTA G</td>
<td>This study: 5’phosphorylated</td>
</tr>
<tr>
<td>2PTC</td>
<td>GTA ACA CCT CAG CCG ACT AGT GGC TTC</td>
<td>This study: 5’phosphorylated</td>
</tr>
<tr>
<td>2PTCC</td>
<td>GAA GCC ACT AGT CGG CTG AGG TGT TAC CGT AG</td>
<td>This study</td>
</tr>
</tbody>
</table>

2.4.4. PCR

The polymerase chain reaction (PCR) was performed in a Twinblock™ System Easycycler™ thermocycler (Ericomp, San Diego, CA). 1-10 ng of template DNA, 20-50 pmol of each primer and 5 μL of DMSO were combined in a 0.5 mL Eppendorf tube in a total volume of 10 μL. The mixture was heated for 1 minute at 96° C and then, while still
at 96° C, 40 µL of a mixture containing 0.5 µL Vent™ polymerase (1 Unit), 5 µL of 10x Vent™ polymerase reaction buffer, and 4 mM of each deoxyribonucleic acid triphosphate (dNTP) (dATP, dCTP, dGTP and dTTP) were added. The mixture was covered with one drop of mineral oil and cycled twenty times through program 83 (1 minute at 96° C, 45 seconds at 55° C, 90 seconds at 72° C) and once through program 84 (7 minutes at 72° C). The reactions were then kept at 21° C until removed.

pUC18-1.5cenAΔprAS was constructed by using the four primer PCR method described in Figure 2.1. Two PCR reactions using the template pUC18-1.5cenAΔpr and primer pairs LSMUTA, SAENH and LUSMUTC, CenARev, respectively were carried out as described above. 1 µL of each of the products of these reactions were then combined and mixed with 5 µL DMSO in a 0.5 mL Eppendorf tube in a total volume of 10 µL. The mixture was heated for 1 minute at 96° C. While still hot, 38 µL of a mixture containing 0.5 µL Vent™ polymerase (1 Unit), 5 µL of 10x Vent™ polymerase reaction buffer, and 4 mM of each dNTP (dATP, dCTP, dGTP and dTTP) were added. The mixture was then cycled once through program 83 (1 minute at 96° C, 45 seconds at 55° C, 90 seconds at 72° C) and heated to 1 96° C for 1 minute. While still at 96° C, 2 µL of a mixture containing 50 pmol each of primers SAENH and CenARev were introduced and the mixture was covered with one drop of mineral oil and cycled twenty times through program 83 (1 minute at 96° C, 45 seconds at 55° C, 90 seconds at 72° C) and once through program 84 (7 minutes at 72° C). The reaction was kept at 21° C until removed.

Enzyme and unincorporated nucleotides from the PCR were separated from the PCR products by phenol:chloroform extraction and ethanol precipitation. The final PCR
product and vector pUC18-1.5cenAΔpr were digested with MluI and PflmI, and the PCR fragment was ligated into the vector forming pUC18-1.5cenAΔprAS.

**Figure 2.1:** Generating pUC18-1.5cenAΔprAS.

The four primer PCR method was used to engineer in silent AgeI (A) and SpeI (S) sites flanking the deleted PT linker site (Δ) to generate pUC18-1.5cenAΔprAS. // indicates DNA sequence that has been omitted for clarity.
2.4.5. Construction of mutant PT linkers

cenAAΔpt and cenAigalh were obtained from pUC18-1.5cenAAΔpt and pUC18-1.5cenAigalh, respectively.

cenAFn3 and cenA1.5PTmod were constructed by PCR using the FNIIIDW and FNIIUP primer set with the pTAL3 template, and the PTCCP and SAENH primer set with the pTUgKRG-cenA1.5PT template, respectively. Both products encoded silent AgeI and SpeI sites which allowed insertion into pUC18-1.5cenAAΔptAS and pTUgKRG-cenAFn3, respectively.

cenA1.5PT and cenA2PT were constructed by hybridizing and ligating synthetic oligonucleotides (Table 2.6) together as shown (Figure 2.2). Each pair of oligonucleotides (2PTN and 2PTNC = N, 2PT1 and 2PT1C = 1, 2PT2 and 2PT2C = 2, 2PT3 and 2PT3C = 3, 2PTC and 2PTCC = C) were first hybridized together to form a double stranded piece of DNA. The short double stranded DNA pieces were precipitated with ethanol in the presence of 0.01 M MgCl₂ and recovered by ultracentrifugation at 27000 rpm for 1-2 hours at 4° C in a Beckman TL-100 Ultracentrifuge with TLA 45 rotor (Sambrook et al., 1989). Selected double stranded pieces were then ligated overnight using the ratios indicated in the brackets (N and 1 (5:1), 2 and 3 (5:1)). The products were separated and the correct sized pieces isolated from a 2% agarose gel. To form 1.5PT, pieces N, 2-3-2, and C were ligated; and to form 2PT, pieces N-1, 2-3-2, and C were ligated and concentrated by ethanol precipitation as above. The N- and C-terminus oligonucleotide pairs encoded silent AgeI and SpeI sites which allowed insertion into pUC18-1.5cenAAΔptAS.
Figure 2.2: Construction of cenA1.5PT and cenA2PT using synthetic oligonucleotides
For gene expression in *E. coli* and *S. lividans*, the *cenA* constructs were first subcloned, using *MluI* and *PflmI* sites, from pUC18 into pSL1180*cenA* (created by subcloning an *NheI/HindIII* fragment from pTUgKRG-1.5*cenA*.*N* into pSL1180). Then an *NheI/HindIII* fragment from each pSL1180*cenA* mutant was subcloned into either pTUgKRG-1.5*cenA*.*N* to create pTUgKRG-*cenA* mutants for expression in *E. coli* or pIJ680-CBD-*cen* producing pIJ680-*cenA* mutants for expression in *S. lividans*.

**2.5. Screening for mutants and confirmation of identity**

**2.5.1. *E. coli***

Antibiotic-resistant colonies were screened by restriction endonuclease digestions of plasmid DNA preparations. Positive clones were confirmed by DNA sequencing.

**2.5.2. *S. lividans***

To obtain stable clones, antibiotic-resistant colonies were placed in 1 mL TSB supplemented with 5 µg Tsr mL⁻¹ at 30°C, 250 rpm. Those that grew were plated onto TSB agarose plates with high viscosity CMC (1%) and grown for 3 days. The clones were tested for CMCase activity using the Congo red staining procedure for carbohydrates (Teather and Wood, 1982). Positive clones were confirmed by either production of protein or sequencing of plasmids. Plasmids isolated from *S. lividans* were difficult to sequence so were transformed into *E. coli* DH5α and re-isolated for sequencing.
2.6. Gene expression and protein purification

2.6.1. *E. coli*

CenA was a gift from E. Kwan, Department of Microbiology and Immunology, UBC. This protein had been purified by affinity chromatography on CF1™ cellulose (fibrous, long cellulose), (Sigma) with elution with 6M guanidinium hydrochloride (ICN Biomedicals Inc, Aurora, OH) and subsequent buffer exchange into 50 mM potassium phosphate buffer, pH 7.0. A 10 mg mL\(^{-1}\) stock solution was made by diluting CenA into 50 mM sodium citrate buffer, pH 7.0. This stock was used to make dilutions for the enzymatic assays and for use in proteolysis and other experiments.

p30 was derived from CenA by papain hydrolysis as described in Gilkes *et al.*, 1989 (Gilkes *et al.*, 1989) followed by gel filtration chromatography on a Superose 12 column (Pharmacia, Uppsala, Sweden) in 150 mM NaCl, 100 mM Tris-HCl pH 8.0 in HPLC grade H\(_2\)O at a flow rate of 0.3 mL min\(^{-1}\).

For expression, plasmids were electroporated into *E. coli* JM101. *E. coli* JM101, carrying *cenA* mutant plasmids, were grown for 24 - 48 hours at 30°C and 250 rpm to an optical density at 600 nm (OD\(_{600}\)) of 5-6 in Terrific broth® medium, supplemented with 50 mg kanamycin mL\(^{-1}\) and 0.1 mM IPTG at the time of inoculation. Cells were separated from the supernatant by centrifugation. Phenylmethylsulfonylfluoride (PMSF), sodium azide and EDTA were added to the supernatant to respective concentrations of 0.1 mM, 0.02 %, and 3 mM. The supernatant was concentrated in an Ultrasette™ tangential flow concentrator (3 kDa cut off, Filtron, Northborough, MA) and then buffer was added to a final concentration of 50 mM potassium phosphate, pH 7.0. The concentrate was clarified.
by a 45 minute spin at 23700 x g and then loaded onto a CF1™ cellulose column and eluted
and exchanged as for CenA. CenA mutants were further purified by nickel affinity
chromatography using the HisBind® resin and protocol from Novagen (Milwaukee, MI)
(Novagen, 1994). The proteins were exchanged into 50 mM sodium citrate buffer, pH 7
using an Amicon stirred cell concentrator (5 kDa cutoff), Amicon, (Beverly, MA). A final
centrifugation step at 40 000 rpm for 45 minutes in a Beckman TL-100 Ultracentrifuge
with TLA 45 rotor was done to remove any protein which had precipitated during the
process.

2.6.2. C. fimii

Native CenA and Cex were purified from the supernatant of C. fimii grown on
modified Leatherwood medium supplemented with 0.1 % (w/v) aspen holocellulose (a gift
from K. Wong and J. Saddler of the Department of Wood Science, Faculty of Forestry,
University of British Columbia). Cellulose-binding proteins were recovered by batch
adsorption to Avicel™ PH101 and elution with 8M guanidinium hydrochloride.
Glycoproteins were recovered by binding to ConA Sepharose (Pharmacia, Uppsala,
Sweden). CenA and Cex were then resolved by anion exchange chromatography
(Langsford, 1988).

2.6.3. S. lividans

Methods were basically as provided by D. B. Wilson, Section of Biochemistry,
Molecular and Cell Biology, Cornell University, Ithaca, New York, which are
modifications of published methods (Hopwood et al., 1985).
Production of the CenA mutants: 500 µL of an *S. lividans* TKM31 DMSO stock from -70°C carrying one of the pIJ680cenA constructs was diluted into 20 mL of TSB broth (BDH) supplemented with 5 µg Tsr mL⁻¹ and grown at 30°C in baffled flasks to increase aeration until the cultures were dense but not pink (approximately 24 hours). 8 mL of these pre-cultures were used to inoculate each of four 500 mL preparations of TSB, 5 µg Tsr mL⁻¹ and 1 drop (about 30 µL) of antifoam in 2 L baffled flasks. These were grown at 30°C and 250 rpm for 48 to 72 hours. At this time the cells were harvested by centrifugation and the supernatant filtered through glass wool and then through a Type A/E glass fibre filter (Gelman Sciences, Ann Arbor, MI) to remove any remaining mycelium. 150 mL (packed volume) of CF1™; potassium phosphate buffer, pH 7.0 (final concentration 50mM); PMSF (final concentration 0.1 mM); and sodium azide (final concentration 0.02%) were added to the supernatant. This was put at 4°C overnight with stirring to allow binding to cellulose. After 24 hours the CF1™ was recovered by settling and packed into an fast liquid chromatography (FPLC) column. The proteins were washed, eluted, and buffer exchanged as described for the *E. coli* proteins except the proteins were exchanged into sodium citrate buffer. No further purification procedures were done. A final centrifugation step at 40 000 rpm for 45 minutes in a Beckman TL-100 Ultracentrifuge with TLA 45 rotor removed any protein which had precipitated during the process.
2.7. Protein detection and quantification and storage

Protein purity was estimated from proteins resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (12%) and visualized with Coomassie-Blue (Laemmli, 1970). Approximately 5 µg of protein were analyzed in this manner.

The spectrophotometric methods of Mach, Middaugh and Lewis (A280/A320/A350) (Mach et al., 1992) and Scopes (A280/A205) (Scopes, 1974); the Bradford dye binding assay (Bio-Rad) (Bio-Rad, 1993; Bradford, 1976); and amino acid analysis by the UBC NAPS unit were used according to the referenced publications or the supplier’s directions to determine the most appropriate routine method to be used in this study. Spectrophotometry was done using a Hitachi U2000 spectrophotometer (Toyko, Japan). Where appropriate, the theoretical masses were used in the equations. Purified proteins were stored at 4° C.

2.8. Western blotting and N-terminal sequencing of proteins and peptides

Western blotting was done using a Bio-Rad mini-protean trans-blot® apparatus (0.5 A, 20 minutes). Proteins were transferred from SDS-PAGE gels to polyvinylidene difluoride (PVDF) membranes (Immobilon™, Millipore Corporation). Prestained standards from Bio-Rad were: phosphorylase B (106 kDa), bovine serum albumin (BSA) (80 kDa), ovalbumin (50 kDa), carbonic anhydrase (33 kDa), soybean trypsin inhibitor (28 kDa), and lysozyme (19 kDa).

N-terminal amino acid sequencing was done by S. Perry of the UBC NAPS unit using automated Edman degradation for 6 or more cycles on an Perkin Elmer Applied Biosystems 476A gas-phase sequencer (Matsudaira, 1990) or by S. Kielland of the
Department of Biochemistry and Microbiology, University of Victoria using automated Edman degradation for 6 or more cycles on a Perkin Elmer Applied Biosystems 470A gas-phase sequencer.

2.9. Mass analysis of proteins

2.9.1. Predicted molecular weights

The predicted molecular weights of the proteins were calculated from DNA-derived amino acid sequences with the computer program DNA Strider (version 2) (CEA, France).

2.9.2. Relative mass

The relative molecular masses ($M_r$) of the proteins were estimated from Coomassie-Blue stained SDS-PAGE separated samples using the following molecular mass standards from Bio-Rad: myosin (200 kDa), β-galactosidase (116 kDa), phosphorylase B (97 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (31 kDa), soybean trypsin inhibitor (22 kDa), lysozyme (14 kDa) and aprotinin (7 kDa).

2.9.3. Electrospray mass spectrometry

Masses of non-glycosylated proteins and protease products were determined by S. He, Department of Chemistry, UBC, using electrospray mass spectrometry (MS). Mass spectra were recorded on a PE-Sciex API 300 triple quadrupole mass spectrometer from Sciex (Thorhill, Ontario).
2.9.4. Matrix-Assisted Laser-Desorption/Ionization Time-Of-Flight (MALDI-TOF) mass spectrometry

Masses of glycosylated proteins were determined using two different mass spectrometers.

The masses of CenA and Cex produced by *C. fimii* were determined using a TofSpec® (Fisons Instruments, MA) matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) mass spectrometer by the demonstration staff of Fisons. Proteins were analyzed in a matrix consisting of the protein in HPLC water and 50 μM sinapinic acid (3,5-dimethoxy-4-hydroxycinnamic acid) in water-acetonitrile-trifluoroacetic acid (50:50:0.1). Equal volumes of the sample (10 pmol μL\(^{-1}\)) and the matrix were combined and 0.5 μL were applied to a stainless steel target. The sample was allowed to dry at room temperature before insertion into the mass spectrometer. The average number of shots at the target was 50, and the laser intensity was set at 3.95. The signals from both the singly and doubly charged ions of lysozyme were used for internal mass calibration.

The mass of CenA1gA1h produced by *S. lividans* was determined using a surface-enhanced laser desorption ionization (SELDI) mass analyzer from Ciphergen Biosystems (Palo Alto, CA). 1 μL (10 pmol) of the protein in 70% acetonitrile, 0.1% TFA was placed onto a SELDI ProteinChip™ array cationic binding target. The sample was air-dried before adding the matrix (sinapinic acid) in 70% acetonitrile, 0.1% TFA. This was air-dried before the chip was put into the mass analyzer. The signals from both the singly and doubly charged ions of horse radish peroxidase (43240 D) and bovine serum albumin (66410 D) were used for calibration. The mass spectra were obtained from the signal
averaging of approximately 50 laser shots. Mass calibration and conversion to mass/charge distributions were performed using software supplied by Ciphergen (Ciphergen Biosystems, 1997).

2.10. Stability

Protein stability was assessed under assay conditions by comparing the CMCase activity of the enzymes sampled at 0 and 24 hours from a mock BMCC hydrolysis experiment where the assay was set up without the BMCC and the samples were not boiled.

2.11. Enzyme assays

All enzyme assays were done in 50 mM citrate buffer, pH 7 in a total volume of 1.5 mL at 37°C unless noted. Glucose was used to generate standard curves.

2.11.1. PNPC hydrolysis

*C. fimi* Cex activity was followed by a dot blot assay using PNPC. 10 μL of 12.5 mM PNPC (in 50 mM potassium phosphate buffer, pH 7 and 0.02 % NaN₃) were pipetted onto a piece of parafilm. 10 μL of the enzyme containing solution were added and mixed. Development of a yellow colour within a few minutes was indicative of PNPCase activity.

2.11.2. BMCC hydrolysis

Five trials were performed with samples set up in triplicate. 1 mg of substrate and 0.15 nmole enzyme were combined, mixed for 5 seconds on a vortex mixer and incubated at 37°C without further agitation. After 24 hours, the tubes were placed in boiling water
for 3 minutes and then spun twice at 16600 x g, saving the supernatant after each spin. The supernatants were then placed at -20°C until the time the samples were analyzed. For analysis, the samples were allowed to thaw on ice and then 200 to 500 µL aliquots were analyzed in duplicate for reducing sugar content using the p-hydroxybenzoic acid hydrazide (HBAH) reducing sugar assay (Lever, 1973).

2.11.3. PASC hydrolysis

This was done as for BMCC hydrolysis, except three trials were performed with samples set up in triplicate, 10 mg of substrate were used and 10 to 20 µL samples were analyzed in duplicate.

2.11.4. CMC hydrolysis

Reactions were set up in duplicate. An enzyme sample was mixed with pre-warmed 0.5 % CMC in 50 mM citrate buffer in a total volume of 0.5 mL and incubated at 30°C without further agitation for 30 minutes. 1 mL of HBAH reagent was added to stop the reaction and then the HBAH reducing sugar assay was carried out as for BMCC hydrolysis (Shen, 1990).

2.11.5. 2,4-DNPC hydrolysis

2,4-DNPC hydrolysis was carried out as described previously (Damude, 1995). A concentration of approximately 10 x Michaelis constant (K_m) (1.5 mM) 2,4-DNPC was used in the assays. The assays were carried out at pH 7.0.
2.11.6. pH profiles

pH profiles of the enzymes were compiled from 2,4-DNPC hydrolysis reactions over a pH range of 4.5 to 8.0 at 0.5 pH unit intervals. The method was described previously (Damude, 1995). A concentration of 5 x $K_m$ (840 μM) 2,4-DNPC was used in the assays. Absorbance (using the method of Mach et al., Section 2.7) was measured for diluted enzyme samples after each set was completed to get the most accurate protein concentration reading for calculations.

2.11.7. Hydrolysis products

The products of BMCC hydrolysis by CenA and CenAFn3 were analyzed as described previously (Stålbrand et al., 1998). The samples analyzed were taken at 10 and 96 hours post enzyme addition. HPLC analysis was done by S. Mansfield, Forest Product Biotechnology, Department of Wood Science, UBC.

2.11.8. Statistical analysis of hydrolysis data

The hydrolysis data were expressed as a ratio of linker variant specific activity to wild type (WT) specific activity. A statistical method which can be used to determine whether ratios differ significantly from each other is the paired t-test of the logs of the specific activities, because the log of a ratio is equal to the difference between the logs of the individual components (Equation 2.1). Using the logarithms produces a distribution more likely to be Gaussian than simply utilizing the raw ratios (Motulsky, 1995).

Equation 2.1

$$\log\left(\frac{SACenA\text{ mutant}}{SACenA}\right) = \log(SACenA\text{ mutant}) - \log(SACenA)$$
2.12. Proteolysis

2.12.1. Papain hydrolysis

The papain reactions contained 108 μg protein, 6.36 x 10^6 units (U) papain mol⁻¹ protein, 5 mM L-cysteine (made fresh), 2 mM EDTA, 100 mM Tris-HCl pH 7.0, and 0.02 % NaN₃ in a total volume of 330 μL. The reactions were incubated at 37° C with agitation. 30 μL (approximately 10 μg protein) samples were taken at 0 minutes, 10 minute intervals for an hour, then at 90, 120 minutes and 24 hours. The samples were immediately combined with 10 μL of 4x SDS loading buffer and stored at -20° C. 20 μL (approximately 5 μg) of the samples were analysed by SDS-PAGE (16% acrylamide). To determine the proteolysis sites, selected samples were prepared for N-terminal sequencing by Western blotting as described in section 2.8. The membranes were stained with 0.1% Coomassie-Blue in 50% HPLC grade MeOH and destained with 50% HPLC grade MeOH. Selected bands were excised and sequenced.

2.12.2. C. fimi protease hydrolysis

C. fimi protease was prepared from Cellulomonas fimi ATCC 484 grown on glycerol medium, and assayed with hide powder azure as described previously (Gilkes et al., 1988). The C. fimi protease reactions contained 70 μg protein, 2.2 x 10⁸ U C. fimi protease mol⁻¹ protein, 50 mM citrate buffer pH 7.0 in a total volume of 225 μL. The reactions were incubated at 37° C with agitation. 30 μL (approximately 10 μg protein) samples were taken at 0, 1, 5 or 6, 24, 48, 72 and sometimes 96 hours. The samples were
immediately combined with 10 μL of 4x SDS loading buffer and stored at -20°C. 20 μL (approximately 5 μg) of the samples were analysed by SDS-PAGE (12% acrylamide). To determine the proteolysis sites, selected samples were prepared for N-terminal sequencing by Western blotting as described for papain hydrolysis (Section 2.12.1).

2.13. IgA1 protease digestion

These reactions were carried out by J. Qiu, GRASP at Tufts University, Boston, MA. The IgA1 proteases used in this study were recombinant *H. influenzae* type 1 from *H. influenzae* Rd, and impure preparations of *N. meningitidis* type 1 from clinical strain *N. meningitidis* S-8273, *N. meningitidis* type 2 from clinical strain *N. meningitidis* S-7941, *N. gonorrhoeae* type 2 from clinical strain *N. gonorrhoeae* MS11 and *S. pneumoniae* from *S. pneumoniae* ATCC 6314. The reaction mixtures contained 0.75 mg substrate (CenAIgA1h or CenA from *E. coli* or *S. lividans.*) mL⁻¹; 15 U IgA1 protease mL⁻¹; 50 mM Tris-HCl, pH 7.5; and 0.01 % sodium azide (an IgA1 protease unit equals the amount of enzyme which can cleave 1 μg of IgA1 sec⁻¹ at 37° C, pH 7.5). 5 mM CaCl₂ and MgCl₂ were added for the *S. pneumoniae* digests. The reaction mixtures were placed at 37° C for up to 7 days. 15 μg of the digested substrates were analyzed by 10% SDS-PAGE.

2.14. Glycosylation determination

2.14.1. Periodic acid/Schiff reagent and ConA-HRP treatment of Western blots

Periodic acid/Schiff reagent treatment (periodic acid/Schiff stain) of Western blots of SDS-PAGE gels was done according to the procedure described previously (Stromqvist...
and Gruffman, 1974) using periodic acid, acetic acid (Fisher), premixed Schiff reagent (Sigma) and sodium metabisulfite (BDH).

Western blots were treated with Concanavalin A horseradish peroxidase (ConA-HRP) (Seikagaku, Tokyo, Japan) and bands visualized using the HRP development reagent from Bio-Rad (Hercules, CA) or the enhanced chemiluminescence (ECL) detection system from Amersham.

2.14.2. Phenol sulfuric acid assay

The phenol sulfuric acid assay for quantification of sugars was performed as described previously (Chaplin, 1986). Glucose was used as a standard.

2.14.3. Determination of the monosaccharide composition and linkage positions between sugars for CenA and Cex produced by C. fimi

GLC-MS: Monosaccharides were released from CenA and Cex produced by C. fimi by methanolysis using the methanolic HCl kit (Alltech, Deerfield, IL), silver carbonate (Aldrich) and acetic anhydride, and P₂O₅ (Aldrich). The dried monosaccharide mixture was derivatized by trimethylsilylation using Tri-Sil® (Pierce) and pyridine (BDH) at 60° C. Samples were analyzed by gas liquid chromatography (GLC) with a DB1701 30 metre microbore (0.252 mm diameter) column (J and W Scientific) on a Carlo Erba Series 4160 GLC (Kratos, New York). Conditions: injector temperature 250° C and detector temperature 280° C; program, 140° C start for 25 minutes then an increase of 3° C min⁻¹ to 230° C followed by 40 minutes at 230° C; carrier gas was Helium (1 kg cm⁻² s⁻¹). Standards for GLC were D-mannose (Aldrich - a 99% mixture of anomers), α-D-glucose and inositol
(Aldrich), D-(+)-galactose (Sigma - Sigma grade), N-acetyl-D-glucosamine and N-acetyl-
D-galactosamine (Sigma) and L-(-)-fucose (Sigma - 99%). Inositol was used as an internal
standard. The GLC-mass spectrometry of CenA and Cex samples was done by L. Madilao,
Department of Chemistry, University of British Columbia. The GLC conditions were as
above, the interphase temperature was 280° C, and the mass was determined using a MS80
RFA mass spectrometer with a DS-55 data acquisition system (Kratos, New York).

Exo-glycosidase digestions of CenA and Cex produced by C. fimi: Exo-
glycosidases used were α-mannosidase from Jack bean meal (specificity α1-2,3,6), and α-
galactosidase from green coffee beans (specificity α1-3,4,6) both from Oxford
GlycoSystems (Abingdon, UK). The enzymes were used according to the manufacturer’s
specifications. The reaction mixtures contained 0.25 mg of CenA or Cex mL⁻¹, 5 U exo-
glycosidase mL⁻¹ in buffers recommended by the suppliers, in a final volume of 0.4 mL.
The mixtures were incubated at 37° C for 24 hours. The α-mannosidase was removed from
the reaction mixtures by gel filtration (Superose 12, Pharmacia) to prevent a subunit with
similar apparent Mₜ as CenA and Cex from interfering with the interpretation of data from
SDS-PAGE. A fraction of the mannosidase-free CenA or Cex was subsequently digested
with α-galactosidase. The proteins in the sample were concentrated by binding to Avicel™
followed by release from Avicel™ by boiling in SDS-PAGE loading buffer. The
supernatant was retained following centrifugation and analyzed by SDS-PAGE and
periodic acid/Schiff staining as described in Section 2.14.1.

The fluorophore-assisted carbohydrate electrophoresis (FACE®) monosaccharide composition kit supplied by Glyko (Novato, CA) was used according to the manufacturer's specifications. Monosaccharides were released by acid hydrolysis and then labeled using 8-aminonaphthalene-1,3,6-trisulfonic acid (ANTS) by reductive amination chemistry. The labeled carbohydrates were separated on high-percent pre-cast polyacrylamide gels. Monosaccharides separate on FACE® gels as a result of the formation of charged complexes of borate with vicinal hydroxyl groups during electrophoresis and can be visualized by long wave UV (Glyko, 1997; Glyko, 1995a).

2.14.5. *Determination of the oligosaccharide composition and linkages between sugars for CenA2PT produced by S. lividans using FACE®*

The FACE® O-linked oligosaccharide profiling kit from Glyko (Novato, CA) was used according to the manufacturer’s specifications with the addition of the use of the following standards: α1-2 mannobiose, α1-3 mannobiose, α1-4 mannobiose, α1-6 mannobiose, and α1-3,1-6 mannotriose, all from Dextra Laboratories Ltd. (UK), and α-lactose and D-(+)-cellobiose from Sigma (MO). Oligosaccharides were released by hydrazinolysis and then labeled using ANTS as described for monosaccharide composition analysis in Section 2.14.4. The labeled carbohydrates were separated on high-percent (up to 40 %), pre-cast polyacrylamide gels. Separation of oligosaccharides was affected by the charge/mass ratio of the saccharide as well as its hydrodynamic volume. The separation of
neutral oligosaccharides was based on size. The bands were visualized by long wave UV (Glyko, 1997; Glyko, 1995b).

2.14.6. Location of glycans on Cen A and Cex produced by C. fimi

200 µg of Cen A or Cex produced by C. fimi were combined with urea and ammonium carbonate to a final concentration of 6.4 M and 0.32 M, respectively. The mixture was heated at 50° C for 10 minutes to denature the protein. Distilled water and 80 µg TPCK trypsin in 10 mM Tris-HCl, pH 8.0 were then added to a final volume of 1032 µL. The mixture was incubated at 37° C with agitation for 17 hours. Digestion was stopped by the addition of PMSF to a final concentration of 0.1 mM. Reduction of the trypsinized protein was performed by the addition of 50 µL of a 50 mM dithiothreitol (DTT) solution to 500 µL of the digest and incubation at 50° C for 10 minutes. In order to separate out the glycosylated peptides, MnCl₂·4 H₂O and CaCl₂·2 H₂O were added to 1 mM final concentration and this mixture was added to a 100 µL ConA sepharose affinity column packed into an approximately 500 µL spin column. The sample was loaded by gravity, washed 2 times with 200 µL Buffer A (150 mM NaCl, 1 mM CaCl₂ 2 H₂O, 1 mM MnCl₂ 4 H₂O, 0.01 % NaN₃ and 50 mM Tris-HCl, pH 7.3) and eluted with 200 µL Buffer B (0.5 M methyl-α-D-mannopyranoside in Buffer A). Buffer B was allowed to soak the column for at least an hour to facilitate elution. HPLC was performed with the assistance of P. Tomme, Department of Microbiology and Immunology, UBC using a Vydac 218TP54 4.6 x 25 cm C18 reverse phase column (300 Å, 5 µ support) with a resolve C18 guard pak column on a Shimadzu HPLC system with a SIL6B autoinjector, an RID6A refractive index detector and a CRS01 chromatopac integrator-plotter. 150 µL
(approximately 1.5 nmoles protein) were injected. The flow rate was 0.8 mL min⁻¹. The solvents were all HPLC grade. Solvent A contained 0.08 % TFA in water and Solvent B comprised 0.072 % TFA and 80 % acetonitrile in water. The elution program used is shown in Table 2.7. The computer program PC gene was used to predict the elution times for the peptides. The peptides were detected by $A_{230}$ and collected by hand. Attenuation was 5 and the fullscale was 0.08 absorbance units for the integrator-plotter. Glycosylated peptides were attached to PVDF membranes in spin columns (Pro Spin™ sample preparation cartridge, Applied Biosystems, Foster City, CA) and N-terminal sequenced. HPLC was also performed on samples prior to the ConA affinity chromatography step, and on tryptic digests of CenA and Cex produced by $E. coli$ as controls.

Table 2.7: HPLC program.

<table>
<thead>
<tr>
<th>TIME (minutes)</th>
<th>%B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>65</td>
<td>75</td>
</tr>
<tr>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>90</td>
<td>5</td>
</tr>
<tr>
<td>91</td>
<td>stop</td>
</tr>
</tbody>
</table>
3. Results


3.1.1. Definition of the boundaries of the PT linker of CenA

For this study the wild type (WT) PT linker was considered to be 27 amino acids (P_{108} - P_{134}). In previous studies the PT linker was defined as 23 amino acids (P_{112} - T_{134}) (Shen, 1990), and 33 amino acids (P_{108} - S_{141}) (Gilkes et al., 1991d). The linkers are shown in Figure 3.1. As other cellulase PT rich linkers contain serine (Gilkes et al., 1991d), it was decided that the PTTS sequence should be included in the definition of the linker. Valine 135 through threonine 140 was not included because the N-terminal amino acid of a C. fimii protease product, p30, is valine 135 (Gilkes et al., 1989). As p30 is stable against further proteolysis by the C. fimii protease, valine 135 is likely the beginning of the catalytic domain.

![Figure 3.1](image)

Figure 3.1: The PT linker of CenA as defined in various studies.
This study (Panel A); previous studies (Shen, 1990) (Panel B), and (Gilkes et al., 1991d) (Panel C).
3.1.2. Primer design for introducing silent mutations

The program Reverse Translator®, as noted in Section 2.4.1, was used to determine appropriate silent mutations to flank the PT linker of CenA. The exact DNA bases changed to introduce the silent *A*gel and *Spe*I sites are shown in Figure 3.2.

```
104
TGTVPTTSPTPTPTPTP
ACC GGC ACC GTG CCG ACG ACC AGC CCC ACC CCG ACG CCG ACC GGT
Agel

141
TVPQPTS
ACG GTC ACG CCG ACG AGC CCC ACC CCC ACC CCC ACC ACC CCG CCG
Act AGT
SpeI
```

**Figure 3.2:** Design of silent mutation sites.

The PT linker amino acid sequence is in bold. The base pair changes needed to introduce the silent mutations are also in bold and the restriction endonuclease recognition sites are underlined with the restriction endonuclease cited below.

3.1.3. Generation of pUC18-1.5cenAAΔpAAS

In order to introduce silent mutations using only four primers as detailed in Section 2.4.4, pUC18-1.5cenAAΔpA (Shen et al., 1991) was used as a template for PCR. The deletion in pUC18-1.5cenAAΔpA from amino acids 108-134 made it possible to design a single set of complementary mutagenic primers, shown in Table 2.4, which spanned the region where the PT linker variants would be inserted. Fewer PCR reactions were required
using the four primer method. This reduced the probability of introducing unwanted mutations. The product from the four primer PCR method was digested with Mlul and PflmI and subcloned into pUC18-1.5cenAΔrT to form pUC18-1.5cenAΔrT,AS.

3.1.4. Construction of the cenA mutants

Construction of pUC18-1.5cenA1.5PT and pUC18-1.5cenA2PT:

Oligodeoxyribonucleotides and oligodeoxyribonucleotide primers were synthesized as described in Section 2.4.2. The synthetic oligomers were optimized for two properties. First, the GC content was reduced because gene expression was to be done in E. coli, which has a lower GC content than C. fimi (48 - 52 % vs. 71 - 76 %) (Ørskov, 1984; Stackebrandt and Keddie, 1984). The changes reduced the GC content of the linker region from 77% to 58%. Second, the sequences were manipulated to minimize erroneous oligomer hybridization using the overlap detection function of DNA Strider™. The synthetic oligonucleotides were hybridized and ligated together as outlined in Section 2.4.5. 1.5PT and 2PT oligonucleotides were digested with AgeI and SpeI and inserted into pUC18-1.5cenAΔrT,AS to form pUC18-1.5cenA1.5PT and pUC18-1.5cenA2PT (Figure 3.3).

Construction of pUC18-1.5cenAFn3: PCR was used to engineer AgeI and SpeI sites into the N- and C- terminus, respectively, of the linkers adjacent to the Fn3 repeats of CenB, as described in Section 2.4.5. This allowed replacement of the PT linker of CenA with the Fn3 repeats. Plasmid pTAL3, encoding CenB, was used as the Fn3 template in the PCR. The PCR product of 945 bp was digested with AgeI and SpeI and ligated into pUC18-1.5cenAΔPTAS to form pUC18-1.5cenAFn3 (Figure 3.3).
Subcloning cenA PT mutants into E. coli and S. lividans expression vectors

pTUgKRGcenA mutants: pUC18-1.5cenAAprAS did not have the unique Nhel and HindIII sites needed for subcloning the cenA linker constructs into the expression vectors pTUG for expression in E. coli (Graham et al., 1995), and pIJ680CBDex for expression in S. lividans (Ong et al., 1994). Therefore, a subcloning strategy was required. The strategy that was used is outlined in Figures 3.4 and 3.5. A Pfil/MluI DNA fragment from each pUC18-1.5cenA mutant (sizes in base pairs: WT: 516; A pr : 453; 1.5pr: 555; igalh: 510; 2PT: 597; Fn3: 1350) was first subcloned into pSL1180cenA, created by subcloning a 1.5 kb Nhel/HindIII fragment from pTUgKRG-1.5cenAN into pSL1180. Then an Nhel/HindIII fragment (Figure 3.6) from each pSL1180cenA mutant was subcloned into pTUgKRG-1.5cenAN to create pTUgKRGcenA mutants. pTUgKRGcenA mutants were then electroporated into E. coli DH5α. Positive clones were selected and confirmed as detailed in Section 2.5.1. For expression, DNA constructs were electroporated into E. coli JM101.

pIJ680cenA mutants: Nhel/HindIII fragments from all the pSL1180cenA mutants except pSL1180-cenA1.5PT, were subcloned into pIJ680CBDex to create pIJ680cenA mutants (Figure 3.5). pIJ680cenA mutants were then transformed into S. lividans TKM31 and stable clones selected as described in Sections 2.4 and 2.5.2. The CMCase activity of S. lividans colonies expressing CenAΔPT is shown in Figure 3.7. S. lividans colonies
which did not express a recombinant cellulase exhibited no CMCase activity (data not shown).

Construction of pTUgKRG-\textit{cen}A1.5PTmod: Site-directed mutagenesis as described in Section 2.4.5 was used to change valine 176 to proline. Positive clones were selected and confirmed as detailed in Section 2.5.1. For expression, the plasmid was electroporated into \textit{E. coli} JM101.
Figure 3.3: Construction of pUC18-1.5cenA mutants.
Silent mutations were introduced in pUC18-1.5cenAΔPT to introduce Age I and Spe I sites flanking the PT linker (1); PCR was used to generate cenB Fn3 repeats from pTAL3 (2); the synthetic oligonucleotides for 1.5 and 2 PT linkers and the PCR product encoding the Fn3 repeats from cenB were digested with Age I and Spe I and inserted into pUC18-1.5cenAΔPTAS to form pUC18-1.5cenA mutants (3).
Figure 3.4: Part 1 of subcloning strategy to insert *cenA* mutants into expression vectors pTUG and pIJ680.

*cenA* was subcloned from pTuGKR-1.5cenA.N into pSL1180 using *Nhe* I and *Hind* III to form pSL1180cenA (1); DNA fragments of *cenA* containing the PT mutants were subcloned from pUC18-1.5cenA into pSL1180cenA using *Mlu* I and *Pfl* MI to form pSL1180cenA mutants (2).
Figure 3.5: Part 2 of subcloning strategy to insert cenA mutants into expression vectors pTUG and pIJ680. 
cenA mutants were subcloned into pTUG-KRG-1.5cenA.N and pIJ60CBD_ex using Nhe I and Hind III from pSL1180cenA mutants (1) to form pTUG-KRG-1.5cenA mutants and pIJ680cenA mutants, (2) and (3), respectively.
Figure 3.6: Agarose gel showing the Nhel/HindIII DNA fragment from each pSL1180cenA mutant used to subclone into pTUg and pIJ680.

cenAΔPT (Lane 1, 1470 bp), cenA (Lane 2, 1533 bp), cenA1.5PT (Lane 3, 1572 bp),
cenA2PT (Lane 4, 1614 bp), cenAFn3 (Lane 5, 2367 bp), cenAigalh (Lane 6, 1527 bp).

Figure 3.7: Congo-red stained agar plate showing halos where S. lividans colonies expressing CenAΔPT were growing.

The black marks inside the halos indicate positions of colonies that were scraped off.
3.2. Gene expression and protein purification.

3.2.1. *E. coli*

*E. coli* harbouring pTUgKRG-*cenA* constructs were grown and proteins purified as outlined in Section 2.6.1. The yield was up to 5 mg purified protein L\(^{-1}\). Some slight yellow/brown colour was retained through the purification. A method of protein measurement was chosen which accounted for the colour (Section 3.3). CenA had been purified as described in Section 2.6.1. p30 was obtained by digesting CenA with papain as described in Section 2.6.1., followed by size exclusion chromatography as described previously (Gilkes *et al.*, 1991c).

All proteins were purified to greater than 90% homogeneity as judged from a Coomassie-Blue stained SDS-PAGE gel (Figure 3.8). The modular structures of the mature proteins and the final linker sequences are detailed in Figure 3.9.

![SDS-PAGE gel of CenA constructs produced in *E. coli*.](image)

**Figure 3.8:** SDS-PAGE gel of CenA constructs produced in *E. coli*.

CenA (Lane 1), CenAΔPT (Lane 2), CenA1gA1h (Lane 3), CenA1.5PT (Lane 4), CenA2PT (Lane 5), CenAFn3 (Lane 6) and p30 (Lane 7). The SDS-PAGE gel was stained with Coomassie-Blue.
**Figure 3.9:** The domain structure and linker sequences of the CenA and CenA PT linker constructs.

The underlined sequences belong to the domains adjacent to the linker.
3.2.2. *S. lividans*

Proteins were produced in and purified from *S. lividans* as described in Section 2.6.3. The yield was up to 10 mg purified protein L$^{-1}$. Some slight yellow/brown colour was retained through the purification, as for proteins produced in *E. coli*.

All proteins were purified to greater than 90% homogeneity as judged from a Coomassie-Blue-stained SDS-PAGE gel (Figure 3.10). To eliminate the possibility that the cellulose-binding proteins were of *S. lividans* origin, they were tested with an antibody against CenA and were all reactive (data not shown).

<table>
<thead>
<tr>
<th>kDa</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 3. 10:* SDS-PAGE gel of CenA constructs produced in *S. lividans.*

CenA$\Delta$PT (Lane1), CenA$\Delta$gA1h (Lane2), CenA (Lane3), CenA$\Delta$PT (Lane 4) and CenA$\Delta$Fn3 (Lane 5). The SDS-PAGE gel is stained with Coomassie-Blue.
3.2.3. *C. fimi*

CenA and Cex were purified from the supernatant of *C. fimi* as described in Section 2.6.2. 0.8 mg CenA L⁻¹ and 0.3 mg Cex L⁻¹ were purified. The proteins were purified to greater than 90% homogeneity as judged from an SDS-PAGE gel stained with Coomassie-Blue (Figure 3.11).

![SDS-PAGE gel of CenA and Cex produced in *C. fimi*. CenA (Lane 1) and Cex (Lane 2). The SDS-PAGE gel is stained with Coomassie-Blue.](image)

3.3. Protein quantification

A number of protein quantification methods was explored using a subset of the proteins produced in *E. coli*. Protein concentrations were determined by the spectrophotometric methods of Mach, Middaugh and Lewis ($A_{280\text{nm}}, A_{320\text{nm}}, A_{350\text{nm}}$) and Scopes ($A_{280\text{nm}}, A_{205\text{nm}}$), by the chemical method of Bradford, and by amino acid analysis, as described in Section 2.7. The results are shown in Table 3.1. It was assumed that the amino acid analysis was the most accurate assay but, as it was not practical for routine analysis, it was used as a benchmark against which the other methods were compared. The
Bradford dye-binding assay values were furthest from the amino acid analysis values for all but one sample. The method of Mach, Middaugh and Lewis was chosen for routine measurements over the Scopes method because it accounted for interference by absorbance due to colour retained during protein purification.

Table 3.1: Protein concentrations of CenA constructs produced in *E. coli* determined by a variety of methods.

<table>
<thead>
<tr>
<th>PROTEIN DETECTION METHOD</th>
<th><em>E. coli</em> protein (concentration in mg mL⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CenA</td>
</tr>
<tr>
<td>Bradford Dye-Binding Assay*</td>
<td>6.4</td>
</tr>
<tr>
<td>Scopes (A280,A205)</td>
<td>8.9</td>
</tr>
<tr>
<td>Mach (A280,A320,A350)</td>
<td>7.2</td>
</tr>
<tr>
<td>Amino Acid Analysis</td>
<td>8.2</td>
</tr>
</tbody>
</table>

* used BSA as the protein standard
3.4. Molecular masses of proteins

The sizes of the proteins produced in *E. coli* and *S. lividans* were determined by three methods as described in Section 2.9: predicted molecular weight, $M_r$, and electrospray MS. The results are summarized in Table 3.2. The larger than expected $M_r$ results are due to the PT linker or the IgA1 hinge. It has been shown previously that proteins containing PT rich linkers or the IgA1 hinge result in aberrantly slow migration during SDS-PAGE (Miller *et al.*, 1992). The difference between the predicted molecular weight and the electrospray MS values for CenA1.5PT construct likely reflects misprocessing by one amino acid at the N-terminus. The N-terminus was not sequenced to confirm this.
Table 3.2: Molecular masses of CenA and CenA mutants produced in *E. coli* and *S. lividans* determined by three different techniques.

All masses are in Da. The electrospray MS values are accurate to 0.1%.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Organism used for protein production</th>
<th>Predicted (DNA Strider)</th>
<th>Relative Mobility</th>
<th>Electrospray MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CenAΔPT</td>
<td><em>E. coli</em></td>
<td>43276</td>
<td>50400</td>
<td>43287</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td></td>
<td>50400</td>
<td></td>
</tr>
<tr>
<td>CenAΔgA1h</td>
<td><em>E. coli</em></td>
<td>45133</td>
<td>55800</td>
<td>45149</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td></td>
<td>56900</td>
<td></td>
</tr>
<tr>
<td>CenA</td>
<td><em>E. coli</em></td>
<td>43820</td>
<td>55800</td>
<td>43826</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td></td>
<td>59000</td>
<td></td>
</tr>
<tr>
<td>CenA1.5PT</td>
<td><em>E. coli</em></td>
<td>46622</td>
<td>57200</td>
<td>46745</td>
</tr>
<tr>
<td>CenA1.5PTmod</td>
<td><em>E. coli</em></td>
<td>46620</td>
<td>57200</td>
<td></td>
</tr>
<tr>
<td>CenA2PT</td>
<td><em>E. coli</em></td>
<td>48009</td>
<td>63600</td>
<td>48026</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td></td>
<td>67900</td>
<td></td>
</tr>
<tr>
<td>CenAFl3</td>
<td><em>E. coli</em></td>
<td>72466</td>
<td>75000</td>
<td>72498</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td></td>
<td>75000</td>
<td></td>
</tr>
<tr>
<td>p30</td>
<td><em>E. coli</em></td>
<td>30077</td>
<td>29500*</td>
<td>30078</td>
</tr>
</tbody>
</table>

* this was determined from a separate gel
3.5. Glycosylation determination

3.5.1. Detection of glycosylation on proteins produced in \textit{S. lividans} and \textit{E. coli}

The proteins produced in \textit{E. coli} and \textit{S. lividans} were tested for the presence of covalently attached sugars by the periodic acid/Schiff stain and ConA-HRP as described in section 2.14.1. The proteins produced in \textit{E. coli} were not glycosylated as they did not react with the periodic acid/Schiff stain (data not shown). All but CenAFn3 were glycosylated by \textit{S. lividans} (Figure 3.12). The lower bands in Lanes 3 and 4 (Figure 3.12) indicate a low level of degradation has occurred. These degradation products are apparent due to the high sensitivity of ConA-HRP. To confirm that CenAFn3 was not glycosylated by \textit{S. lividans}, approximately 10 \textmu g of protein made by \textit{E. coli} and \textit{S. lividans} were blotted and tested with ConA-HRP. Neither protein reacted (Figure 3.13). Figure 3.14 shows the molecular mass difference between the proteins produced by \textit{E. coli} and \textit{S. lividans}. The more diffuse bands seen for proteins produced by \textit{S. lividans} are typical of glycoproteins, which exist as a set of glycoforms varying by as little as one sugar unit. The lower molecular weight bands in Lane 1 (Figure 3.13) are degradation products of an old preparation of CenAFn3 and the lower molecular weight bands in Lanes 1 and 2 (Figure 3.14 B) are degradation products evident from overloading these samples of CenAFn3.
Figure 3.12: SDS-PAGE gel and ConA-HRP treated blot of CenA constructs produced in *S. lividans*.

SDS-PAGE gel stained with Coomassie-Blue (Panel A) and ConA-HRP treated blot (Panel B), CenAΔPT (Lane 1), CenA IgA1h (Lane 2), CenA (Lane 3), CenA2PT (Lane 4) and CenAFn3 (Lane 5).

Figure 3.13: SDS-PAGE gel and ConA-HRP treated blot of CenA constructs produced in *S. lividans* and *E. coli*.

SDS-PAGE gel stained with Coomassie-Blue (Panel A) and ConA-HRP treated blot (Panel B), CenAFn3 produced by *E. coli* (Lane 1), CenAFn3 (Lane 2), CenAΔPT (Lane 3) and CenA (Lane 4) produced by *S. lividans*. The arrow indicates the size of CenAFn3 on the blot.
3.5.2. Moles of sugar per mole of protein for CenA and Cex produced in C. fimi and CenAIgA1h produced in S. lividans

The masses of CenA and Cex produced by C. fimi were determined by MALDI-TOF MS as detailed in Section 2.9.4. Masses of 48519 and 47874 Da were determined in two runs for CenA and a mass of 49623 Da was obtained for Cex (Figure 3.15). This information was used to determine the number of moles of sugars on the proteins by

Figure 3.14: SDS-PAGE gels of CenA constructs produced in E. coli and S. lividans. E. coli (odd lanes) S. lividans (even lanes), Panel A: CenAΔPT (Lanes 1 and 2), CenAIgA1h (Lanes 3 and 4), CenA (Lanes 5 and 6), and CenA2PT (Lanes 7 and 8); Panel B: CenAFn3 (Lanes 1 and 2). The SDS-PAGE gels were stained with Coomassie-Blue.
comparing the masses of the glycosylated proteins with their non-glycosylated recombinant counterparts. CenA and Cex had 15-19 and 16 moles of carbohydrate per mole of protein, respectively.

The phenol sulfuric acid assay, outlined in Section 2.14.2, was also used to quantify sugars. The moles of carbohydrate measured per mole of protein were 29 and 30 for CenA and Cex produced in *C. fimii*, respectively. The MALDI-TOF MS data are considered the most accurate.

The masses of CenAIgA1h produced by *S. lividans* and *E. coli* were determined by SELDI MS, as described in Section 2.9.4, to be 46637 and 45024 Da, respectively (data not shown). This indicated the presence of approximately 7 moles of sugar per mole of protein.

![MALDI-TOF mass spectra of CenA and Cex produced by *C. fimii*.](image)

**Figure 3.15:** MALDI-TOF mass spectra of CenA and Cex produced by *C. fimii*. CenA (Panel A: a and b) and Cex (Panel B). Masses are shown in Da.
3.5.3. Location of carbohydrates on CenA and Cex produced in C. fimi and CenA constructs produced in S. lividans.

To determine the location of the glycans on CenA and Cex produced in C. fimi, HPLC profiles of tryptic digests of native and recombinant cellulases were compared as detailed in Section 2.14.6. Potential trypsin cleavage sites indicated that the PT linker should be contained within a single peptide for both CenA and Cex. The tryptic digests of the native and recombinant cellulases gave very similar HPLC profiles (data not shown). For CenA and Cex produced in C. fimi, one main peptide bound to ConA (labeled 2 in Figure 3.16 and 1 in Figure 3.17). The peak labeled 2 in Figure 3.16 is undigested CenA. The retention times of these peptides matched those predicted by the computer program PC gene for the peptides containing the PT linker. This was confirmed by N-terminal sequencing of the major retained peaks. Minor peaks in both profiles were not sequenced. For CenA, the peptide containing the PT linker is 108 amino acids long and contains 19 amino acids of the catalytic domain and 62 amino acids of the CBD. The 19 amino acids of the catalytic domain include 4 threonines and 1 serine residue as potential O-glycosylation sites and no potential N-linked sites; the 62 amino acids of the CBD include 10 threonines and 9 serines and 5 potential N-linked sites. The remaining CBD sequences are found in peptides of 5, 10 and 30 amino acids in size. The smaller ones may have been hard to detect, but the larger one should have been apparent if glycosylated. For Cex, the peptide containing the PT linker is 60 amino acids long, consisting of 15 amino acids of the catalytic domain with no potential O- or N-linked glycosylation sites, and 26 amino acids of the CBD with 3 threonines residues as potential O-linked sites and 1 potential N-
linked site. The remainder of the Cex CBD is found on 2 peptides of 40 and 43 amino acids, respectively. If these peptides were glycosylated, they would have been concentrated on the ConA column to a similar level as the one containing the PT linker, but no other peaks concentrated to that level.
Figure 3.16: Reverse phase HPLC profiles of a tryptic digest of CenA produced by C. fimii.

Acetonitrile was used as Solvent B. The peptide profile prior to ConA affinity chromatography (Panel A) and the peptides which were retained by ConA (Panel B). N-terminal sequences were determined for the numbered peaks.
Figure 3.17: Reverse phase HPLC profiles of a tryptic digest of Cex produced by C. fimii. Acetonitrile was used as Solvent B. The peptide profile prior to ConA affinity chromatography (Panel A) and the peptides which were retained by ConA (Panel B). The N-terminal sequence was determined for the numbered peak.
Papain digests of CenA and Cex produced by \textit{C. fimi}, as described in Section 2.12.1, followed by periodic acid/Schiff staining confirmed that no carbohydrates were located on the CBD of either glycoprotein (Figure 3.18). The proteins produced in \textit{C. fimi} were not cleaved efficiently by papain. However, the CBDs were just distinguishable in the gel stained with Coomassie-Blue (Figure 3.18 Panel A), but did not react with the more sensitive periodic acid/Schiff stain (Figure 3.18 Panel B). N-terminal amino acid sequencing of papain hydrolysis products and degradation products of CenA and Cex produced in \textit{C. fimi}, respectively, gave an indication of some glycosylated sites in the PT linker (Figure 3.19).

For the proteins produced in \textit{S. lividans}, it became apparent that \textit{S. lividans} does not glycosylate the catalytic domain or CBD as CenAFn3 was not glycosylated (Figure 3.13).
**Figure 3.18:** SDS-PAGE gel and periodic acid/Schiff stained Western blot of CenA and Cex produced by *C. fimi*, and treated with papain.

SDS-PAGE gel stained with Coomassie-Blue (Panel A) and periodic acid/Schiff stained Western blot (Panel B). CenA untreated (Lane 1), CenA treated with papain (Lane 2), CBD<sub>cex</sub> produced in *E. coli* (Lane 3), Cex treated with papain (Lane 4) and Cex untreated (Lane 5).

**Figure 3.19:** Glycosylation sites on the linker of CenA and Cex produced by *C. fimi*.

The underlined sequences belong to the domains adjacent to the linker and the bold sequence was N-terminal sequenced.
3.5.4. The monosaccharide composition and linkage positions between sugars of CenA and Cex produced by C. fimi

The types of monosaccharides of the glycan component of CenA and Cex produced by C. fimi were determined by GLC-MS as described in Section 2.14.3. Mannose was the major sugar detected. A small amount of glucose was also detected but no galactose (Figure 3.20). Glucose is regarded as a common contaminant in sugar analysis reactions. Therefore, mannose is the sole sugar on these glycoproteins. This was confirmed later by FACE® (Section 3.5.5).

To gain information about the anomeric configuration (α, β) and the linkage positions between sugars (1-2,3,4, or 6) of native CenA and Cex, treatment with exoglycosidases was employed as described in Section 2.14.3: α-mannosidase (Jack bean meal) (specificity α-man-1-2,3,6), and α-galactosidase (green coffee beans) (specificity α-gal-1-3,4,6) (Figure 3.21). Treatment with α-mannosidase reduced the Mr of CenA to that of recombinant CenA produced in E. coli and abolished its reactivity with the periodic acid/Schiff stain. This confirmed the presence of α-D-mannose. A glycosylated CenA species of partially reduced Mr, remained that reacted with ConA-HRP (Figure 3.22). The Mr of Cex was partially reduced by α-mannosidase. This product reacted with both the periodic acid/Schiff stain and ConA-HRP (Figures 3.21 and 3.22). α-galactosidase had no effect on the Mr of CenA or Cex, even following pretreatment with α-mannosidase.

Deglycosylation of papain hydrolysis products of CenA and Cex, which lack the CBD, was tried because the PT linker of these peptides should be more accessible to the exoglycosidases. Total deglycosylation was still not attained, however (data not shown).
Jack bean $\alpha$-mannosidase cleaves mannose linked $\alpha$-1-2,3, or 6. From the digests and the information regarding the specificity of the mannosidase it is evident that CenA consists of a heterogeneous population of glycoproteins some of which contain only mannose linked $\alpha$-1-2,3,6. All the CenA and Cex glycoforms contain terminal $\alpha$-mannose. The protein-carbohydrate linkages in Cex and in some glycoforms of CenA remain to be elucidated.
Figure 3.20: Determination of carbohydrate composition by GLC of CenA and Cex produced by *C. fimii*. Standards (Panel A); fucose (F), mannose (M), galactose (G), glucose (Glc), Inositol (I), N-Acetyl galactosamine (GalNAc), N-Acetylglucosamine (GlcNAc); Cex (Panel B); CenA (Panel C).
Figure 3.21: SDS-PAGE gel and periodic acid/Schiff stained Western blot of CenA and Cex produced by *C. fimii*, and treated with exo-glycosidases.

Coomassie-Blue stained SDS-PAGE gel (top) and periodic acid/Schiff stained Western blot (bottom) of CenA (Panel A) and Cex (Panel B) produced by *C. fimii* and treated with exo-glycosidases: α-mannosidase (Jack bean meal) (Lane 4), α-mannosidase and then α-galactosidase (green coffee beans) (Lane 5), α-galactosidase (Lane 6). Controls: Molecular weight markers in kDa (Lane 1), α-mannosidase (Lane 2), non-glycosylated cellulase produced in *E. coli* (Lane 3), glycosylated CenA (Panel A) or Cex (Panel B) produced by *C. fimii* (Lane 7), α-galactosidase (Lane 8).
Figure 3.22: Western blot treated with ConA-HRP to detect the presence of mannose on Cex and CenA produced by *C. fimi*, and treated with exo-glycosidases.

Cex (Lanes 2-5) and CenA (Lanes 6-9); exo-glycosidases: α-mannosidase from Jack bean meal (Lane 3 and 8), and α-mannosidase and then α-galactosidase from green coffee beans (Lane 4 and 7). Controls: Molecular weight markers in kDa (Lane 1), untreated Cex and CenA produced by *C. fimi* (Lanes 2 and 9), untreated Cex and CenA produced in *E. coli* (Lanes 5 and 6).

3.5.5. The monosaccharide composition of the glycans of CenA mutants produced by *S. lividans* and CenA produced by *C. fimi*

The analysis was done using the FACE® Monosaccharide composition kit from Glyko as described in Section 2.14.3. Figure 3.23 shows the FACE® Monosaccharide composition gel. Glucose was detected in the distilled water sample and in the sample of CenA produced in *E. coli* so was regarded as a contaminant in all cases. CenA produced by *C. fimi* contained only mannose. The CenA mutants produced by *S. lividans* (CenAIgA1h, CenA2PT, and CenAΔPT) all had mannose like CenA from *C. fimi* but also contained galactose. From the intensity of the bands there appeared to be more mannose than galactose. The intensity of the signal increased with an increase of potential glycosylation sites, with CenAΔPT < CenAIgA1h < CenA2PT. Quantification is
theoretically possible using the kit but, due to the number of steps involved, it was decided it was likely to be unreliable.

3.5.6. **Oligosaccharide composition and linkage positions between sugars of CenA2PT produced in *S. lividans***

The FACE® O-linked oligosaccharide profiling kit was utilized to determine the size and linkages between the sugars of the oligosaccharides of CenA2PT produced in *S. lividans* as detailed in Section 2.14.5. Figure 3.24 shows the FACE® oligosaccharide analysis gels. The mannobiose standards, which only differed in the linkage between carbohydrate moieties, migrated to different positions in the gel (Figure 3.24 Panel A, Lanes 1, 3-6). However, substituting one neutral sugar, galactose, for another, glucose, did not affect migration (Figure 3.24 Panel A, Lanes 7-8). The anomeric configuration also appears to affect migration in the case of cellobiose (β1-4 glucose) which did not migrate at the same level as the starch (α1-4 glucose) biose standard on the hydrolysis ladder (Figure 3.24 Panel A, Lanes 8-9). The oligosaccharide profile of CenA2PT produced in *S. lividans* is shown in Figure 3.24 Panel B. Using the relative mobility of the mannobiose and triose standards to the starch hydrolysis ladder; a standard curve generated from the biose standards; and the relative mobility of cellobiose to the starch hydrolysis ladder, the size of the oligosaccharides and the type of linkage in the disaccharide were determined. CenA2PT had single neutral sugars; biose units linked mostly α1-2, and some α1-3; two triose units; and a tetrose unit of undetermined linkage. *S. lividans* attached both mannose and galactose. Therefore, which carbohydrates were involved in the linkages could not be ascertained.
Figure 3.23: FACE® Monosaccharide composition gel of CenA produced by *C. fimi* and a subset of CenA mutants produced by *S. lividans.*

Panel A: CenA produced by *E. coli* (Lane 1), CenA produced by *C. fimi* (Lane 2), and water (Lane 3); Panel B: CenAIgAlh (Lane 1), CenA2PT (Lane 2), and CenAΔPT (Lane 3) all produced in *S. lividans.* Monosaccharide standards are indicated to the left of the gel.
Figure 3.24: FACE® Oligosaccharide analysis gel of CenA2PT produced by *S. lividans*.

Panel A: Standards: α-1-3, 1-6 mannotriose (Lane1), α-1-6 mannobiose (Lane3), α-1-4 mannobiose (Lane4), α-1-3 mannobiose (Lane5), α-1-2 mannobiose (Lane6), lactose (galactose-β-1-4 glucose) (Lane7), cellobiose (glucose-β-1-4 glucose) (Lane8), Panel B: Standards: α-1-2 mannobiose (Lane3), α-1-6 mannobiose (Lane4), SgCenA2PT (Lane 2). Starch hydrolysis glucose standard ladder (Panel A: Lanes 2 and 9; Panel B: Lane 1). The number of α-1-4 linked glucose units is indicated to the left of the panels.
3.6. Enzyme activity

3.6.1. Protein stability

Protein stability was determined under assay conditions by comparing the CMCase activity, as described in Section 2.11.4, of the enzymes sampled at 0 and 24 hours from a mock BMCC hydrolysis experiment, where the assay was set up without the BMCC and the samples were not boiled. No loss of activity was detected (data not shown).

3.6.2. Effect of method on BMCC hydrolysis by CenA

BMCC and PASC are both insoluble substrates. The effect of agitation on activity was analyzed for BMCC hydrolysis. Methods used included: no agitation, mixing on a tube roller, shaking on platform shakers at 250 and 300 rpm, mixing using an end over end labquake™ apparatus and agitation with a Pierce magnetic stirring apparatus. Earlier studies (data not shown) indicated that the presence of BSA had no effect on the activity of CenA on CMC. Therefore, in this study reactions set up in parallel with and without BSA were compared as a means of comparing the different agitation methods. From Table 3.3, it is apparent that no agitation, mixing on the tube roller and shaking at 300 rpm all gave comparable readings between samples with and without BSA. The other three methods had larger variation between samples. The method of agitation also affected the absolute activity measured, with end over end mixing giving the lowest activity reading, and the Pierce magnetic stirring apparatus the greatest. No stirring was chosen instead of agitation for the activity studies because the method was simple and the substrate stayed well dispersed in suspension over the time period of the assay. All the other agitation
methods resulted in the substrate aggregating into clumps. When evaluating samples set up in triplicate later, this method proved reliable (data not shown).

Table 3.3: Summary of agitation methods studied for the hydrolysis of BMCC. All values are presented as μmole glucose μmole enzyme$^{-1}$ min$^{-1}$.

<table>
<thead>
<tr>
<th>AGITATION METHOD</th>
<th>CenA -BSA</th>
<th>CenA + BSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>no agitation</td>
<td>.248</td>
<td>.214</td>
</tr>
<tr>
<td>tube roller</td>
<td>.089</td>
<td>.095</td>
</tr>
<tr>
<td>shaker (250 rpm)</td>
<td>.117</td>
<td>.161</td>
</tr>
<tr>
<td>shaker (300 rpm)</td>
<td>.299</td>
<td>.306</td>
</tr>
<tr>
<td>end over end labquake™ apparatus</td>
<td>.084</td>
<td>.050</td>
</tr>
<tr>
<td>PIERCE magnetic stirring apparatus</td>
<td>.496</td>
<td>.913</td>
</tr>
</tbody>
</table>

3.6.3. **BMCC hydrolysis**

Trials were performed with samples set up in triplicate. Substrate (1 mg) and enzyme (0.15 nmole) were combined, mixed by vortexing and placed at 37°C without further agitation for 24 hours. Sampling and analysis were as described in Section 2.11.2. CenA produced in *E. coli* was always included as a reference. Results are presented in Table 3.4. The activities are presented as relative activities - ratios of the activity of CenA mutant “x” to the activity of CenA produced in *E. coli* - because the ratios were reproducible but the absolute values varied between trials. Paired t-tests were performed on the ratios as described in Section 2.11.8. A P value of less than 0.05 is generally taken
to be significant (Motulsky, 1995). The P value for this test reflects the probability that the means are the same. Thus a P value of 0.004 for a ratio of 1.22, as for CenAFn3 produced in *E. coli*, indicates a 0.4% chance that CenA and CenAFn3 have the same activity.

Modifying the length of the linker of non-glycosylated CenA had a modest effect on its activity on BMCC. Removing the linker and CBD, or shortening the linker decreased the activity: p30 had 48% of the activity of wild type CenA and CenAΔPT had 74%. Increasing the length of the linker by 150%, as for CenA1.5PT; or modifying the composition minimally while retaining a similar length linker as CenA, exhibited by CenAlgA1h, had no effect on activity. Doubling the linker length or replacing the linker with the Fn3 repeats increased activity: CenA2PT had a 9% increase, while CenAFn3 had a 20% increase.

The effect of glycosylation of CenA and CenA linker mutants on their activity on BMCC was inconsistent. The glycosylated variants of CenA and CenA2PT had increased activities of 26% and 20%, respectively. In contrast, the glycosylated variants of CenAΔPT and CenAlgA1h had decreased activities: 89% and 91%, respectively, of their non-glycosylated counterparts. Both CenA2PT and CenAlgA1h had P values above 0.05.
Table 3.4: Summary of the hydrolysis of BMCC by CenA and CenA constructs produced in *E. coli* and *S. lividans*.

Activity is presented as a relative activity: the mean ratio (calculated for the number of trials indicated in the table) of the activity of CenA mutant “x” to the activity of CenA produced by *E. coli*. Each trial consisted of samples set up in triplicate. (SD = standard deviation; na = not applicable)

<table>
<thead>
<tr>
<th>Protein</th>
<th>Organism protein produced in</th>
<th>Relative Activity (mean ratio)</th>
<th>SD of mean ratio</th>
<th>Paired t-test two tailed P-value</th>
<th>Number of trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>CenA</td>
<td><em>E. coli</em></td>
<td>1.00</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td>1.26</td>
<td>0.006</td>
<td>0.01</td>
<td>2</td>
</tr>
<tr>
<td>CenAAPT</td>
<td><em>E. coli</em></td>
<td>0.74</td>
<td>0.01</td>
<td>2 x 10^{-7}</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td>0.55</td>
<td>0.07</td>
<td>0.01</td>
<td>3</td>
</tr>
<tr>
<td>CenAIgA1h</td>
<td><em>E. coli</em></td>
<td>0.95</td>
<td>0.17</td>
<td>0.50</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td>0.84</td>
<td>0.10</td>
<td>0.12</td>
<td>3</td>
</tr>
<tr>
<td>CenA1.5PT</td>
<td><em>E. coli</em></td>
<td>1.03</td>
<td>0.09</td>
<td>0.53</td>
<td>4</td>
</tr>
<tr>
<td>CenA2PT</td>
<td><em>E. coli</em></td>
<td>1.09</td>
<td>0.06</td>
<td>0.03</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td>1.29</td>
<td>0.22</td>
<td>0.12</td>
<td>3</td>
</tr>
<tr>
<td>CenAFn3</td>
<td><em>E. coli</em></td>
<td>1.22</td>
<td>0.06</td>
<td>0.004</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td>1.20</td>
<td>0.06</td>
<td>0.03</td>
<td>3</td>
</tr>
<tr>
<td>p30</td>
<td><em>E. coli</em></td>
<td>0.48</td>
<td>na</td>
<td>na</td>
<td>1</td>
</tr>
</tbody>
</table>
3.6.4. *PASC hydrolysis*

Trials were performed with samples set up in triplicate. 10 mg of substrate and 0.15 nmole enzyme were combined, mixed by vortexing and placed at 37°C without further agitation for 24 hours. Sampling and analysis were as described in Section 2.11.3. CenA produced in *E. coli* was always included as a reference. Results are presented in Table 3.5. The activities are presented as ratios to CenA produced by *E. coli* as for BMCC hydrolysis (Section 3.6.3). Statistical analysis was as described in Sections 2.11.8 and 3.6.3.

CenAΔPT was the only mutant to show a significant difference from CenA with a decrease of 13%. p30 had 33% more activity in the single trial performed. The other constructs had increases up to 7% but all had high P values. Glycosylation had little or no effect on the activity of the proteins, with the greatest difference being for CenAIgA1h which showed a decrease of 13% upon glycosylation, but with a high P value.

Shortening the linker or removing the CBD were the only modifications to have notable effects on the hydrolysis of PASC.
Table 3.5: Summary of the hydrolysis of PASC by CenA and CenA constructs produced in *E. coli* and *S. lividans*.

Activity is presented as a relative activity: the mean ratio (calculated for the number of trials indicated in the table) of the activity of CenA mutant “x” to the activity of CenA produced by *E. coli*. Each trial consisted of samples set up in triplicate. (SD = standard deviation; na = not applicable)

<table>
<thead>
<tr>
<th>Protein</th>
<th>Organism protein produced in</th>
<th>Relative Activity (mean ratio)</th>
<th>SD of mean ratio</th>
<th>Paired t-test two tailed P value</th>
<th>Number of trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>CenA</td>
<td><em>E. coli</em></td>
<td>1.00</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td>1.00</td>
<td>0.03</td>
<td>0.78</td>
<td>2</td>
</tr>
<tr>
<td>CenAAPT</td>
<td><em>E. coli</em></td>
<td>0.87</td>
<td>0.05</td>
<td>0.02</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td>0.92</td>
<td>0.03</td>
<td>0.05</td>
<td>3</td>
</tr>
<tr>
<td>CenAIgA1h</td>
<td><em>E. coli</em></td>
<td>1.03</td>
<td>0.18</td>
<td>0.84</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td>0.90</td>
<td>0.06</td>
<td>0.12</td>
<td>3</td>
</tr>
<tr>
<td>CenA1.5PT</td>
<td><em>E. coli</em></td>
<td>1.07</td>
<td>0.14</td>
<td>0.37</td>
<td>4</td>
</tr>
<tr>
<td>CenA2PT</td>
<td><em>E. coli</em></td>
<td>1.03</td>
<td>0.09</td>
<td>0.61</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td>1.06</td>
<td>0.11</td>
<td>0.44</td>
<td>3</td>
</tr>
<tr>
<td>CenAFn3</td>
<td><em>E. coli</em></td>
<td>1.00</td>
<td>0.07</td>
<td>0.99</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td><em>S. lividans</em></td>
<td>1.02</td>
<td>0.02</td>
<td>0.17</td>
<td>3</td>
</tr>
<tr>
<td>p30</td>
<td><em>E. coli</em></td>
<td>1.33</td>
<td>na</td>
<td>na</td>
<td>1</td>
</tr>
</tbody>
</table>
3.6.5. 2,4-DNPC hydrolysis

2,4-DNPC hydrolysis was carried out as described in Section 2.11.5. A concentration of approximately $10 \times K_m$ for 2,4-DNPC was used in the assays and the assays were done at pH 6.5. A single trial was carried out with triplicate samples. Protein concentration determination was performed on the diluted enzyme samples to get the most accurate protein concentration readings for calculations. The results are depicted graphically in Figures 3.25 and 3.26.

CenAFn3 activity was 77% that of CenA, while p30 was 117% that of CenA. All the other enzymes had activities similar to that of CenA. Glycosylated proteins all had decreased activity when compared to their non-glycosylated counterparts, although the level of the decrease varied, being largest for CenAIgAlh (45%). There was a negative correlation between the size of the protein and the activity on 2,4-DNPC.

3.6.6. pH profiles

pH profiles of the enzymes were compiled from 2-4 DNPC hydrolysis reactions over a pH range of 4.5 - 8.0 at 0.5 pH unit intervals as described in Section 2.11.6. A concentration of 2,4-DNPC of $5 \times K_m$ was used. All the proteins showed the same profile with maximum activity at pH 6.5(Figure 3.27).
Figure 3.25: 2,4-DNPC hydrolysis at pH 6.5 by CenA linker constructs produced in *E. coli* and *S. lividans*.

*E. coli* (Panel A) and *S. lividans* (denoted by the prefix sg) (Panel B)
**Figure 3.26:** Composite figure of 2,4-DNPC hydrolysis at pH 6.5 by CenA linker constructs produced in *E. coli* and *S. lividans*.

Proteins produced in *S. lividans* denoted by the prefix sg

**Figure 3.27:** pH profiles of 2,4-DNPC hydrolysis by CenA and CenA constructs produced in *E. coli* and *S. lividans*.

CenA (1); sgCenA (2); CenAΔPT (3); CenAΔPT (4); CenA1gAlh (5); sgCenA1gAlh (6); CenA1.5PT (7); CenA2PT (8); sgCenA2PT (9); CenAFn3 (10); sgCenAFn3 (11); p30 (12); proteins produced in *S. lividans* are denoted with the prefix sg.
3.6.7. Enzyme assay products

The soluble products of BMCC hydrolysis by CenA and CenAFn3 produced in *E. coli* were analyzed as outlined in Section 2.11.7. These constructs were chosen because they had the greatest difference in activity. There was no difference in the soluble products formed (data not shown).

3.7. Proteolysis

3.7.1. Papain hydrolysis of CenA constructs produced in *E. coli*

Papain was chosen for this study because it was known to cleave in the linker sequence of CenA and Cex produced by *E. coli* (Gilkes *et al.*, 1991c). Papain hydrolysis was carried out and analyzed as described in Section 2.12.1. Figures 3.28 and 3.29 show the SDS-PAGE gels of papain digests of CenA constructs made by *E. coli*. Sites of hydrolysis were determined by N-terminal sequencing of selected products. These sites are detailed in Figure 3.30. CenAFn3 was most resistant to papain over a two hour time period, and the hinge region of IgAl in the construct CenAlgAlh was as susceptible as the PT linker. CenA and the modified CenA-like PT linkers yielded various products during the digestion. CenA itself yielded three products. Two transient products resulted from cleavage near the N-terminus of the linker at the site PT*TSP*, where * indicates the cleavage site, and from cleavage further into the linker, at the site PT*TP* (confirmed by electrospray MS). A final stable product resulted from cleavage near the C-terminus of the linker at the site PTV*TP*. All the modified CenA-linker constructs were cleaved in these same sequences, even when they appeared multiple times as in CenA2PT or when they...
appeared in the middle of the linker as in CenA1.5PT and CenA2PT. When the cleavage sites in CenAlgA1h and CenAFn3 were examined, a pattern of susceptible sites was discernible. The sequences cleaved in CenAlgA1h were PTTSVS*TP, PS*TP and PLQ*SN*VTP; and in CenAFn3 were PTG*TTTDTTP, VTF*TTDTTGETEP, PVTS*TP and PS*VTP. The first two cleavage sites in CenAFn3 were determined from N-terminal sequencing of the *S. lividans* produced protein. *S. lividans* does not glycosylate CenAFn3 so the sites were assumed to be the same for CenAFn3 produced in *E. coli*. In most of the proteins, the cleavage sites were located in the consensus sequence P(T/S/V)$_i$*(T/S/V)$_j$P. The location of the sequence did not have to be near the termini of what is defined in this study as the PT linker.

**Figure 3.28:** Papain digestion of CenA and CenA PT constructs produced in *E. coli*. The numbered arrows indicate those products which were N-terminal sequenced and are shown in Figure 3.30. The SDS-PAGE gels were stained with Coomassie-Blue.
**Figure 3.29:** Expanded gel of a papain digestion of CenA produced by *E. coli*.

The arrows indicate products which were N-terminal sequenced and are shown in Figure 3.30. The SDS-PAGE gel was stained with Coomassie-Blue.
Figure 3.30: Papain cleavage sites in CenA and CenA PT mutants made in *E. coli*.

Arrows indicate the site(s) of cleavage and the numbers indicate the band sequenced from the gels in Figures 3.28 and 3.29. The source of the data is indicated by the various arrow types. The underlined sequences belong to the domains adjacent to the linker.
3.7.2. *C. fimi* protease hydrolysis of CenA constructs produced in *E. coli*

*C. fimi* protease was used in this study because it cleaves in the linker sequence of CenA and Cex produced by *E. coli* and *C. fimi* (Gilkes *et al.*, 1989). *C. fimi* protease hydrolysis was carried out and analyzed as described in Section 2.12.2. Figure 3.31 shows the SDS-PAGE gels of the CenA constructs made by *E. coli* digested with *C. fimi* protease. Sites of hydrolysis were determined by N-terminal sequencing of selected products. These sites are detailed in Figure 3.32. The papain sites are noted for comparison. The sites of cleavage by *C. fimi* protease were very similar to those of papain. CenAFn3 was most resistant to papain over a two hour time period but was as susceptible as the CenA-like linkers to *C. fimi* protease. The hinge region of IgA1 in CenAIgA1h was as susceptible as the PT linker to *C. fimi* protease. However, the native IgA1 hinge was not susceptible to *C. fimi* protease. Proteolysis or degradation did occur at a location other than the hinge by 5 hours (Figure 3.33).
Figure 3.31: *C. fimii* protease digestion of CenA and CenA PT constructs made by *E. coli*.

The arrows point to sequenced products which are shown in Figure 3.32. The SDS-PAGE gels were stained with Coomassie-Blue.
Figure 3.32: Papain and *C. fimi* protease cleavage sites in CenA and CenA PT mutants made in *E. coli*.

Arrows indicate the site(s) of cleavage. The enzyme used to cleave is indicated by the arrow type. The underlined sequences belong to the domains adjacent to the linker.
3.7.3. Effect on papain hydrolysis of changing valine 176 to proline

The mutant CenA1.5PTmod was made to explore the possibility of eliminating the C-terminal susceptible hydrolysis site in CenA1.5PT by changing valine 135 to proline. The internal sites were maintained to test the activity of the protease. Papain hydrolysis was carried out and analyzed as described in Section 2.12.1. Figure 3.34 shows the SDS-PAGE gels of the CenA constructs made by *E. coli* and digested with papain. Hydrolysis sites, detailed in Figure 3.35, were determined by N-terminal sequencing. Two hydrolysis products were seen for CenA1.5PTmod and three for CenA1.5PT. The first two products were identical for both constructs. The second cleavage site at the sequence PT*TP was confirmed by electrospray MS. CenA1.5PTmod yielded no further products whereas CenA1.5PT was cleaved once more near the C-terminus of the PT linker. Changing the sequence of a site susceptible to cleavage by papain can abolish sensitivity to papain. The final *C. fimi* protease cleavage site was also abolished in CenA1.5PTmod (data not shown).
Figure 3.34: C. fimi protease digestion (Panel A) and papain digestion of CenA1.5PT and CenA1.5PTmod made by E. coli. The sequenced products are numbered and are shown in figure 3.23.
Figure 3.35: Papain and *C. fim* protease cleavage sites in CenA1.5PT and CenA1.5PTmod produced in *E. coli*.

Arrows indicate the sites of cleavage. The enzyme used to cleave is indicated by the arrow types. The boxed amino acid is the site mutated in the CenAPTmod construct. The underlined sequences belong to the adjacent domains.
3.7.4. **Effect of glycosylation of CenA constructs on susceptibility to papain and C. fimi protease**

Papain and *C. fimi* protease digests were carried out and analyzed as described in Section 2.12.1 and 2.12.2, respectively. Figure 3.36 and 3.37 show the SDS-PAGE gels of the CenA constructs made by *E. coli* and *S. lividans* digested with papain and *C. fimi* protease, respectively. Hydrolysis sites of selected products were determined by N-terminal sequencing and are detailed in Figure 3.38. Only a subset of the glycosylated products was analyzed because the patterns appeared to be the same as that seen for their non-glycosylated counterparts. Of the products sequenced, it became apparent that the cleavage of the glycosylated and non-glycosylated products occurred in the same sequences; although not always between the same amino acids. Digestion also took longer. Thus glycosylation serves to slow down the proteolysis by these two proteases.
Figure 3.36: Papain digestion of CenA and CenA PT mutants made by \textit{E. coli} and \textit{S. lividans}.

Arrows point to the sequenced products which are shown in figure 3.38. The SDS-PAGE gels are stained with Coomassie-Blue.
Figure 3.37: *C. fimi* protease digestion of CenA and CenA PT constructs made by *E. coli* and *S. lividans*.

Arrows point to the sequenced products which are shown in Figure 3.38. The SDS-PAGE gels were stained with Coomassie-Blue.
Figure 3.38: Papain and \textit{C. fimi} protease cleavage sites in CenA and CenA PT mutants made in \textit{E. coli} and \textit{S. lividans}.

Arrows indicate the site(s) of cleavage. The enzyme used to cleave and the nature of the substrate (\textit{E. coli}, \textit{S. lividans}) is indicated by the arrow types. The underlined sequences belong to the domains adjacent to the linker.
3.8. IgA1 protease digestion of CenA and CenAIgA1h

In a previous study, CenA and CenAIgA1h were tested as substrates for type 1 and type 2 IgA1 proteases from *N. gonorrhoeae* (Miller *et al.*, 1992). Despite the sequence similarity between the IgA1 hinge and the CenA PT linker, only the CenAIgA1h construct was cleaved. It was of interest to expand on this experiment by utilizing both non-glycosylated and glycosylated CenA and CenAIgA1h as substrates for a broader selection of IgA1 proteases, as the native IgA1 hinge is glycosylated. IgA1 protease digestions were carried out and analyzed as detailed in Section 2.13. Figure 3.39 shows the SDS-PAGE gels of the digests of CenA and CenAIgA1h, produced by *E. coli* and *S. lividans*, by recombinant *H. influenzae* type 1, and impure preparations of *N. meningitidis* type 1, *N. meningitidis* type 2, *N. gonorrhoeae* type 2 and *S. pneumoniae* IgA1 proteases. The results are summarized in Table 3.6. Glycosylated substrates were cleaved less efficiently than their non-glycosylated counterparts and all were cleaved much less efficiently than native IgA1. The *H. influenzae* type 1 IgA1 protease had the broadest activity, slowly cleaving all substrates. *N. meningitidis* IgA1 proteases cleaved both non-glycosylated substrates. The *N. gonorrhoeae* type 2 and *S. pneumoniae* IgA1 proteases only cleaved non-glycosylated CenAIgA1h. The *N. meningitidis* type 1 enzyme cleaved the CenA PT linker though it doesn’t have the sequence that is cleaved in IgA1. The CenA linker has no P-S bonds. Of the two type 2 enzymes, only the *N. meningitidis* enzyme cleaved CenA while both cleaved CenAIgA1h.
Figure 3.39: IgA1 protease digests of CenAlgA1h and CenA produced by *E. coli* and *S. lividans*.

Digests of CenAlgA1h (A Panels) and CenA (B Panels) by the IgA1 proteases noted above the gels. Lanes 1, 2, 5 and 6 show glycosylated proteins made in *S. lividans* and Lanes 3, 4, 7 and 8 show the non-glycosylated proteins produced in *S. lividans*. Odd lanes are time = 0 and even lanes are time = 4 days.
Table 3.6: Susceptibility of CenA substrates to IgA1 proteases.

+++ = complete cleavage in 4 days; +/- = minor cleavage by 7 days; - = no cleavage.

<table>
<thead>
<tr>
<th>ENZYMES</th>
<th>CenA</th>
<th>CenA</th>
<th>CenA1gA1h</th>
<th>CenA1gA1h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>glyc</td>
<td>ng</td>
<td>glyc</td>
<td>ng</td>
</tr>
<tr>
<td>H. influenzae</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>type 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. meningitidis</td>
<td>-</td>
<td>++</td>
<td>+/-</td>
<td>+++</td>
</tr>
<tr>
<td>type 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. meningitidis</td>
<td>-</td>
<td>++</td>
<td>+/-</td>
<td>+++</td>
</tr>
<tr>
<td>type 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. gonorrhoeae</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>++</td>
</tr>
<tr>
<td>type 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. pneumoniae</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
3.9. A comparison of the glycosylation of the native IgA1 hinge and that produced by
S. lividans

The monosaccharide composition of the hinge region differed between CenAIgA1h
produced in S. lividans and the native hinge. CenAIgA1h had mannose and galactose
(Figure 3.23) while the human myeloma serum IgA1 hinge was reported to have GalNAc
and Gal (Baenziger and Kornfeld, 1974). N-terminal sequencing of a papain digestion
product allowed the determination of some glycosylation sites of CenAIgA1h produced in
S. lividans (Figure 3.40). Modified sites are indicated by the absence of a signal in the
amino acid sequence. As the protein is known to be glycosylated, the unrecovered amino
acids were assumed to be glycosylated amino acids. All the serines in the hinge region of
human myeloma serum IgA1 are fully occupied glycosylation sites (Baenziger and
Kornfeld, 1974). In contrast, the hinge region of CenAIgA1h, made in S. lividans, had
fully occupied sites on all the serines and the last threonine and a partially occupied site on
the second threonine. CenAIgA1h produced in S. lividans has approximately 7 moles of
sugar per mole of protein (Section 3.5.2) which is similar to that of the human myeloma
IgA1 hinge (Baenziger and Kornfeld, 1974).
**Figure 3.40:** Glycosylation of the hinge region of human myeloma serum IgA1 and CenAlgA1h made in *S. lividans*.

Hinge region of myeloma serum IgA1 (Panel A), and CenAlgA1h made in *S. lividans* (Panel B). The underlined amino acids indicate sites that are partially occupied. Abbreviations: GalNAc, N-acetylgalactosamine; Gal, galactose; Man, mannose
4. Discussion

4.1. The effect of linker length and composition on activity

4.1.1. The effect of linker length and composition on activity on PASC

Only the activities of the linker deletion mutant, CenAAPT, and the catalytic domain, p30, differed from wild type CenA (Section 3.6.4). This study confirmed the findings of Shen et al., 1991 that deleting most of the linker of CenA caused a slight decrease in activity on PASC whereas the catalytic domain alone (p30) had increased activity. However, the extent of the decrease reported varied between the two studies. The activities from Shen et al., 1991 were converted to ratios - activity of linker mutant to activity of wild type CenA - in order to compare the data. For CenAAPT, ratios of 0.87 and 0.52 were calculated for this study and Shen et al., respectively, and the ratios for p30 were 1.33 and 2.0. These differences are likely a reflection of the heterogeneity of the substrate, different substrate preparations and possibly different controls. If wild type CenA was not included as a control for each activity measurement made by Shen et al., then the ratios calculated from their data are imprecise. This is a clear illustration of the difficulties encountered when trying to compare data from separate studies in the cellulase field.

4.1.2. Difficulties in comparing cellulase activity data between studies

Variance in the absolute activity values for hydrolysis of insoluble cellulosic substrates makes it difficult to compare results from different studies. This variance is
primarily due to significant differences in structure and composition of cellulosic substrates, even from similar sources such as BMCC, which are expected due to different preparation treatments by various suppliers and/or researchers. Reproducibility in dispensing the substrate for each new set of experiments adds to the difficulties. Activities in this study were reported as relative activities - ratios of the activity of the mutant to wild type - because it was the most consistent measurement and allowed meaningful comparison of activities of the different mutants. Ratios are not commonly reported in the cellulase literature and therefore only general trends can be compared. This drawback has been noted previously (Gal, 1997). More wide-spread use of a relative measure (ratio) in reporting activities would have the advantage of minimizing the variance due to substrate heterogeneity and preparation allowing for more consistent comparison of the activities of different cellulase enzymes within and between research groups.

4.1.3. The effect of linker length and composition on activity on BMCC

Various modifications to the linker region of CenA had a modest impact on the hydrolysis of the more crystalline substrate, BMCC (Section 3.6.3). Both CenAΔPT and p30 had decreased activity relative to CenA which confirms the importance of the CBD itself and its orientation in the hydrolysis of crystalline cellulose (Gilkes et al., 1988; Shen et al., 1991). In this thesis, it was hypothesized that the activity of CenA against crystalline cellulose could be increased by increasing the length of the linker or replacing the linker with the Fn3 repeats from CenB. A longer linker sequence might increase the number of glycosidic bonds available to the catalytic domain while tethered to the substrate by the CBD (Black et al., 1997) and the Fn3 module might confer properties which
enhance activity on crystalline cellulose (H. Stålbrand, manuscript in preparation). CenB has fifty times the activity of CenA on BMCC (Tomme et al., 1996). This difference in activity is probably mainly due to differences in their catalytic domains (they are from different families), but there is evidence that the Fn3 modules of CenB may be important for hydrolysis of the insoluble substrates PASC and BMCC. When the Fn3 repeats of CenB were deleted, the activity compared to WT-CenB on PASC and BMCC decreased six-fold and eighteen-fold, respectively (H. Stålbrand, manuscript in preparation). It has been suggested that the combination of the rigidity of an individual Fn3 domain and the flexibility imparted by repeats may assist cooperativity between the CBD and catalytic domain by limiting unproductive binding (H. Stålbrand, manuscript in preparation).

Doubling the length of the linker and replacing the linker with the Fn3 repeats from CenB resulted in only a modest gain in activity against BMCC: 10% and 22%, respectively. This difference seen for BMCC hydrolysis is statistically significant, supported by the paired t-test (Section 3.6.3), but probably not biologically significant since the activity of the enhanced CenA constructs on BMCC is still low in comparison to other cellulases, such as CenB. The small increase in activity observed for CenAFn3 may be due to the Fn3 domains assisting cooperativity between the CBD and catalytic domain as proposed above for CenB. If the linker length increased the range of the tethered catalytic domain, one might expect there to be a change in the profile of the products released. The HPLC analysis of the soluble products showed no such change (Section 3.6.7). Analysis of the insoluble products was not performed. Only one other study on cellulases looked at the effect of increasing the linker length. Wilson et al. found that doubling the length of the
linker of endoglucanase E2 caused a slight increase in activity. The assay was done on filter paper and no values were reported (Wilson et al., 1995).

4.1.4. Measuring hydrolysis of an insoluble substrate

The approach I chose to evaluate the hydrolysis of BMCC, that of quantifying the soluble products released by either a reducing sugar assay or HPLC, has limitations. First, not all products of hydrolysis of BMCC by an endoglucanase are soluble; only oligosaccharides of up to six sugar units. Hydrolysis products of greater size are not measured by this method as they remain with the insoluble fraction. Second, an enzyme tethered to an insoluble substrate can potentially hydrolyze soluble products and thus alter their profile (Black et al., 1996). Therefore, the absence of change in the HPLC profile of soluble products may not actually reflect that there has been no change. Analysis of the insoluble products was not performed. One method for looking at the insoluble fraction is high performance size-exclusion chromatography where the insoluble fraction is derivatized to make it soluble. This method was not used because it is very labour intensive and requires a high percentage of substrate to be hydrolyzed to get meaningful results (Stålbrand et al., 1998; Kleman-Leyer et al., 1994). A direct, sensitive method for analyzing hydrolysis at the cellulose surface is needed to gain further insights.

4.1.5. The effect of linker length and composition on activity on 2,4-DNPC

The activity on 2,4-DNPC decreased only for the CenAFn3 construct. This is discussed further in the glycosylation discussion (Section 4.3). The pH profiles on soluble substrates did not change with the modifications to the linker length and composition.
4.1.6. Summary of activity studies

In summary, deleting most of the linker decreases the activity on BMCC and PASC whereas doubling the length or substituting Fn3 repeats for the linker result in a modest increase in activity against BMCC. Advances in analyzing hydrolysis at the cellulose surface are needed to achieve a more complete understanding of enzymatic cellulose degradation.

4.2. Effect of linker length and composition on sensitivity to proteases

All the nonglycosylated CenA linker mutants were susceptible to proteolysis and allowed insights into the nature of this susceptibility (Section 3.7). CenAFn3 was the construct most resistant to papain during the first two hours of incubation but it was eventually cleaved to p30. However, it was no more resistant to the C. fimi protease than the other constructs. This suggests that these proteases interact with CenAFn3 differently. It was known previously that that they attacked similar sequences in CenA (Gilkes et al., 1989) but these data urge caution in extrapolating data from papain hydrolysis to C. fimi protease hydrolysis. The other proteins were equally susceptible but were cleaved at different sites. In all cases, cleavage was in the linker sequences and not in the catalytic and cellulose-binding domains. Sequence analysis of the proteolysis products gave insight into the nature of this susceptibility. Sequences near the boundaries of the domains were previously shown to be susceptible to proteolysis (Gilkes et al., 1989; Langsford et al., 1987). Sequencing of the N-terminal region of selected proteolytic fragments showed that it is not the structure formed at the junction that makes this region susceptible but rather the primary amino acid sequences which form a susceptible structure. When these
sequences were moved from the N-terminus to a more central location in the linker or when the C-terminal sequence was modified (changing a valine to a proline), there was no hydrolysis at the junction, only at the susceptible sequences. This suggests that the boundaries of the linker are truly defined by these susceptible sequences. The placement of these sequences may not be a coincidence. Partial degradation of the cellulases during hydrolysis of cellulose may be advantageous since the catalytic domains of some cellulases are more active than the intact enzymes on soluble substrates (Béguin and Alzari, 1998; Fierobe et al., 1993; Fierobe et al., 1991; Ghangas and Wilson, 1988; Reverbel-Leroy et al., 1997). The separated domains may enhance hydrolysis of soluble cellulose which accumulates during hydrolysis (Sandercock et al., 1996).

The results of this study have implications in the design of stable linkers based on natural cellulase linkers, for CBD-fusion proteins. Although each fusion protein is unique, this study suggests that using specific sequences may allow control over protease susceptibility. As seen for the single-chain antibodies containing linkers derived from the *T. reesei* CBHI linker, replacing part of a susceptible site with proline or adding proline adjacent to a susceptible site may reduce or eliminate proteolysis (Alfthan et al., 1995). The effects of higher proline content on the activity and independence of the domains would have to be determined. There are not many bacterial proteases which cleave adjacent to proline; IgA1 proteases and a protease from *Flavobacterium meningosepticum* are the only examples reported, which may account for the stability of sequences with alternating prolines (Lomholt, 1996; Walter et al., 1980).

Analysis of the products of proteolysis of the various CenA linker constructs by *C. fimi* protease and papain gave insights into the specificity of papain and *C. fimi* protease for
these substrates, the boundaries of the natural CenA linker, and strategies to prevent proteolysis of hybrid proteins with proline rich linkers.

4.3. Glycosylation

4.3.1. Sites of glycosylation

The sites of glycosylation on CenA and Cex from C. fimi were examined (Section 3.5.3). Langsford et al. previously described the C. fimi proteins CenA and Cex as glycoproteins and suggested, based on susceptibility to alkali and the resistance to Endo-H, that the glycosylation was O-linked and likely located on the PT linker (Langsford, 1988). In this study, analysis of the sites of glycosylation was carried further using carbohydrate detection methods on a variety of proteolysis products to confirm that the glycosylation occurs solely on the PT linker for Cex. For CenA, the possibility of O-glycosylation on a small part of the catalytic domain could not be ruled out. There are no potential N-linkage sites in the glycosylated regions therefore glycosylation of CenA and Cex produced by C. fimi proteins is restricted to O-linkages. This agrees with the results for Cex, recombinantly produced in S. lividans (Ong et al., 1994), but is in contrast to the fungal cellulases which show both N- and O-linked glycosylation (Salovuori et al., 1987).

By taking advantage of a glycosylated degradation product of Cex and a proteolysis product of CenA, limited N-terminal sequencing revealed that C. fimi glycosylates both serines and threonines in the PT linker. However, not all the serines and threonines of the linker get glycosylated (Section 3.5.3). The sites of glycosylation fit one of a number of predicted O-glycosylation sequences, XPXX (where one X is a glycosylated T and the
others are any amino acid except C, F, H, M, Y, or W), put forth by Gooley and Williams, 1994 and Pisano et al., 1994 for eukaryotic protein glycosylation. Some potential sites were not occupied. The 45 kDa glycopeptide from Mycobacterium tuberculosis has glycans at sites which fit the above motif, and as for the C. fimii proteins, some potential sites are unoccupied. The glycosylated sequence is DPEPAPPVP'TTA, where the bold letter indicates a glycosylated amino acid and the underlined T reflects a potential site which is reported not to be glycosylated (Dobos et al., 1995). The linkers of CBHI and CBHII from T. reesei are rich in hydroxyamino acids but do not have as many prolines as the linkers of CenA and Cex (Gilkes et al., 1991d). Therefore, glycosylation may not be restricted to the predictive sequence discussed for the C. fimii proteins.

There are limits to the technique that was used to determine the specific sites. Glycosylation makes the amino acid more polar and as a result, in the standard gas phase Edman chemistry used for sequencing, the modified amino acids are not soluble in the solvents used to transfer the amino acid to the reaction chamber. Therefore, no amino acid is detected for that cycle if it is glycosylated. Using the knowledge of the sequence, can determine what amino acid should have appeared and can infer that it was glycosylated. Only sites which are modified greater than 50% of the time show up as blanks. The less frequently glycosylated sites appear to be non-glycosylated. Pisano and colleagues (Gooley and Williams, 1994; Pisano et al., 1994) developed solid phase Edman chemistry to allow the use of more polar solvents to enable recovery of the glycosylated amino acids. Comparing recovery of the non-glycosylated and glycosylated variants can give an idea of site occupancy. Our data therefore reflect sites which are 50-100% occupied. Both
secondary and tertiary structure immediately surrounding the glycosylation site may affect its occupancy; possibly by affecting accessibility of glycosyltransferases.

For the proteins produced in *S. lividans*; the CenA linker, CenA linker analogs and the IgA1 hinge sequences were the only sites of glycosylation. The absence of glycosylation on the CenAFn3 construct ruled out the possibility of glycosylation on the CBD and catalytic domains assuming that the Fn3 repeats did not change the conformation of the protein preventing the catalytic domain and/or the CBD from being glycosylated (Section 3.5.1). This conclusion is consistent with the data obtained for CenA produced by *C. fimi* and Cex produced by *C. fimi* and *S. lividans* (Ong *et al.*, 1994). Specific amino acid occupancy was determined for the CenAIgA1h construct and part of the CenA2PT construct (data not shown). Again, the sites of glycosylation agree with the predictive sequence discussed above. Interestingly, there are potential glycosylation sites of this type in the linkers separating the Fn3 repeats, yet neither CenB produced by *C. fimi* nor CenAFn3 produced in *S. lividans* are glycosylated. The other *C. fimi* enzymes with Fn3 repeats, CenD, CbhA and CbhB, are not glycoproteins either (Sandercock *et al.*, 1996). The linker sequences of these regions are more heterogeneous, contain less prolines, and do not follow the pattern of alternating prolines and threonines seen for CenA and Cex. These differences, coupled with the location of the linkers next to domains different than those which are found in CenA and Cex, suggest that it is the protein structure surrounding these sites which prevents glycosylation of this region. The linkers may be more flexible and/or the Fn3 domains may interact with each other when in tandem (Erickson, 1994; Leahy *et al.*, 1996). This freedom of motion might be enough to allow the surrounding domains to obscure potential glycosylation sites from the glycosylation machinery. Even
the shortest linker, found in the CenAPT construct, is glycosylated. This further strengthens the argument that the surrounding domains and flexibility of the linker sequence have an impact on glycosylation. The glycosylation pattern of glycoproteins produced by *C. fimi* and *S. lividans* are similar and the two organisms discriminate between the same protein features for glycosylation. Therefore, it is likely that their glycosylation systems are similar.

### 4.3.2. Use of mass spectrometry for quantification of glycosylation

Despite the advances in mass spectrometry as a tool for analyzing glycosylation, I was unable to fully take advantage of this technique (Sections 1.3.10, 3.5.2). Even with these advances, large glycoproteins are still quite difficult to work with and thus only a subset of the proteins were analyzed in this manner. The masses of CenA and Cex from *C. fimi* revealed the presence of on average 15-19 (9%w/w) and 16 (5%w/w) moles of sugar per mole of protein, respectively. The range seen for CenA reflects the results of two different trials. The MALDI-TOF spectra are actually composed of a set of peaks from all the glycoforms of CenA which have been smoothed by computer software. The average mass reported is the highest point on that smoothed peak. The 1.3% difference in mass falls within the standard error and within the high range of the smoothed peaks. The difference is likely the result of calibration or an artifact of the software smoothing of the peak. The percent carbohydrate determined by mass spectrometry is similar to that determined by GLC by Langsford, 10% and 8% for CenA and Cex, respectively, and to that determined by mass spectrometry of Cex recombinantly produced in *S. lividans*, 5.4% (Langsford, 1988; Ong *et al.*, 1994). The values obtained by mass spectrometry were
about half of those determined by the phenol sulfuric acid sugar assay (Section 3.5.2). This discrepancy is due to two factors: the phenol sulfuric acid sugar assay detects non-covalently attached sugars including any contaminating glucose from the environment or that was carried through from the cellulose column purification step, and glucose was used as a standard which reacts differently than mannose in the phenol sulfuric acid sugar assay (A. Boraston, personal communication). The mass of CenAIgA1h produced in S. lividans was also determined. An average of 7 moles of sugar were detected (Section 3.5.2). Mass spectrometry has the potential to be a powerful tool for more detailed analysis of the carbohydrates of these bacterial glycoproteins.

4.3.3. Monosaccharide composition, size of oligosaccharides and linkage positions between sugars of oligosaccharides

The monosaccharide composition of CenA and Cex produced in C. fimi was re-addressed after Ong et al. found low amounts of galactose present on Cex produced in S. lividans (Ong et al., 1994). Overloading of the GLC sample showed definitively that no galactose was present on CenA and Cex produced by C. fimi. This was confirmed for CenA using the FACE® technology. CenA and the CenA glycosylated analogs, including the IgA1h hinge, all contained galactose and mannose when produced in S. lividans. This agrees with the observations of Ong et al. for Cex (Sections 3.5.4 and 3.5.5) (Ong et al., 1994). The difference in monosaccharide composition of glycans from glycoproteins produced in C. fimi and S. lividans likely reflects the absence of a particular galactosyltransferase in C. fimi. This is not surprising because glycosylation in bacteria appears to be more diverse than in eukaryotes (Moens and Vanderleyden, 1997).
Oligosaccharide size and some sugar linkage information for CenA2PT produced in *S. lividans* was obtained using FACE® technology (Section 3.5.6). The oligosaccharides ranged in size from 1 to 4 sugar moieties. The linkage positions for the disaccharides were determined to be mostly α1-2 with some α1-3. Ong *et al.*, using exoglycosidases to probe the sugar linkage positions, hypothesized the existence of mainly α1-6 linkages on PTCBD<sub>rex</sub> produced by *Streptomyces* (Ong *et al.*, 1994). This interpretation was from the observation that the α1-2 specific exo-mannosidase from *Aspergillus saitoi* (*A. satoi*) did not affect the mobility of the protein as judged from SDS-PAGE. The discrepancy between the two studies may be the result of the specificity of the α1-2 mannosidase or may simply reflect the location of the α1-2 linkages in these oligosaccharides. The *A. saitoi* enzyme is highly specific for non-reducing terminal mannose α1-2 linkages (Oxford Glycosystems catalogue). Information is not available as to whether the enzyme will cleave this linkage when it is in close proximity to the protein such as is found for a disaccharide. Therefore, it is possible that these terminal sugar linkage positions would not be cleaved. Enough sugars need to be removed to result in a sufficient decrease in MW to be able to detect a mobility shift (by SDS-PAGE) for samples treated with exoglycosidases when compared to untreated controls. An undetectable shift would result if the α1-2 linkages are not at the terminus or are not in high enough quantity to change the MW, and hence the mobility significantly.

This is the first time, to the best of my knowledge, that the configuration of the linkage of disaccharides has been deduced directly from FACE® gels. This was possible due to the observation that linkage and anomeric configurations, but not the type of neutral
sugar in the linkage affected the mobility in the FACE® gels of disaccharides containing combinations of mannose, galactose and glucose. Further trials with other neutral sugar combinations need to be done to determine whether this observation reflects a more general phenomenon. Previously, the anomeric and linkage configuration were determined using exo-glycosidases. These enzymes are limited, however, by their inability to distinguish between certain linkage positions between sugars. An example is Jack Bean α-mannosidase which cleaves α1-2,3 and 6 linked mannose. The FACE® system proved to be a fast and simple method. With the correct battery of standards, linkage data could be generated quickly for simple O-linked sugars from bacterial, yeast or fungal proteins.

Structural analysis of the carbohydrate moieties of N-linked proteins of mammalian origin has been accomplished using a series of exo-glycosidase treated proteins followed by FACE® analysis. Standard mobility shifts reflecting the loss of specific sugars allowed the determination of the sequence of the oligosaccharide but not the linkage positions between sugars. The linkage information was inferred from known N-linked structures previously characterized by other means (Kumar et al., 1996). With FACE® technology it is not possible to garner information for linkage positions between sugars or structures not previously described (Raju et al., 1996).

4.3.4. **Role of glycosylation on CenA and CenA derivatives**

Determination of the role of glycosylation on CenA and Cex is challenging as glycosylation could potentially impact many aspects of cellulose degradation. Previous results from Langsford for CenA and Cex from *C. fimii* (glycosylated) and *E. coli* (non-glycosylated) showed no difference in their kinetic properties, thermostability, pH stability
or ability to bind cellulose (Langsford, 1988). It is also unlikely that glycosylation is required for secretion because not all the secreted cellulases of the *C. fimi* system are glycoproteins. However, when PTCBD<sub>ex</sub> was produced in *S. lividans* it had enhanced binding properties over the non-glycosylated variant produced in *E. coli*, and when Cex was produced in *Saccharomyces cerevisiae*, which hyperglycosylates the enzyme, the thermostability was enhanced (Curry et al., 1988; Ong et al., 1994). Thus the effects reported for glycosylation of cellulases have been variable.

The glycosylated *C. fimi* proteins are protected against the action of a *C. fimi* protease when adsorbed to cellulose (Langsford, 1988). When the enzymes are in solution, the rate of proteolysis is merely slowed down. Proteolysis studies of the glycosylated and non-glycosylated variants of CenA and CenA derivatives using papain and *C. fimi* protease confirmed the results of Langsford, but also showed that, in solution, the amount of protection provided by glycosylation varied with both the quantity of glycosylation and the linker composition. The partial deletion construct, CenAΔPT, which had the least glycosylation; and the CenA<sub>IgAlh</sub> construct which had glycosylation comparable to CenA, but with a more variable amino acid composition, were degraded most rapidly (Section 3.7.4). The limited protection of these two constructs by glycosylation is likely to be the result of too few carbohydrates to effectively mask the protease sensitive sites in the case of CenAΔPT, and greater flexibility of the linker due to less prolines in the case of CenA<sub>IgAlh</sub>. The increased movement of the amino acids relative to the carbohydrates might allow the protease greater opportunity to access the cleavage sites.
When the effect of glycosylation on cellulolytic activity was addressed, a difference was found for one substrate, 2,4-DNPC, a small, soluble, synthetic substrate (Sections 3.6.3 - 3.6.6). This was surprising at first as no activity difference on a soluble substrate had been previously ascribed to glycosylation of CenA. The catalytic domain of CenA alone (p30) was more active than the whole enzyme and the least active enzyme variant was CenAFn3, the largest enzyme. Therefore, it may not be an effect of glycosylation per se that influences activity, but of the mass and volume that glycosylation adds to proteins. The observation that the catalytic domain is more active than the whole enzyme on soluble substrates has also been observed for CMCase activity of *T. fusca* E2, made in *S. lividans*, and an endoglucanase from *Sporotrichum pulverulentum* (Ghangas and Wilson, 1988), and *C. cellulolyticum* CelA, CelC and CelF (Fierobe *et al.*, 1993; Fierobe *et al.*, 1991; Reverbel-Leroy *et al.*, 1997). The catalytic domains of *T. reesei* CBHI and CBHII had no change in activity on soluble substrates (Tomme *et al.*, 1988; Zhang *et al.*, 1995). In addition, a hybrid enzyme consisting of the *T. fusca* E2 CBD and the *Prevotella ruminicola* CMCase produced an enzyme, an endoglucanase, which had higher specific activity on CMC than the CMCase alone (Maglione *et al.*, 1992). Like the dockerin and CBD domains, the sugar moieties add mass and volume to the protein. The extended linker region and the carbohydrate moieties are thought to be dynamic and have some flexibility, and so could conceivably block portions of the catalytic domain at times, reducing the activity. Flexibility could be tested using fluorescence quenching of labels on the protein's extremities (i.e. on the catalytic domain and CBD). The possibility that there is a conformational change resulting in decreased catalytic activity associated with having the CBD and linker attached cannot be discounted as there are no structures known for intact
cellulases, just for their domains alone. Future determination of the $K_m$ and $V_{max}$ values for pairs of glycosylated and non-glycosylated cellulases may give further insight into these results. Glycosylation of the CenA constructs did not have a clear effect on the hydrolysis of the insoluble substrates, BMCC and PASC. This is likely due to the CBD binding the enzyme to the substrate which conceivably would limit the freedom of motion of the enzyme compared to when it is in solution, as for 2,4-DNPC hydrolysis. A role of glycosylation, if any, beyond proteolysis protection for CenA remains elusive.

4.3.5. Summary of glycosylation studies

This study confirms that *C. fimi* and *S. lividans* differ in the type of sugars added to glycoproteins but suggests that they are likely to have a similar mechanism for glycosylation. To date, no evidence of N-glycosylation by these organisms has been found. Amino acid sequencing showed that predictive sequences for O-glycosylation in eukaryotes may be applicable in some bacterial systems but that the structure of the protein strongly influences the occupancy of potential glycosylation sites. The role of glycosylation on CenA and CenA derivatives produced in *S. lividans* was examined in this study, but results indicated no expanded role beyond the protective function described previously. Surprisingly, examination of glycosylation of CenA and CenA derivatives gave some insights into the observation that the catalytic domain by itself has increased activity on soluble substrates. Finally, this study introduced the FACE® technology as a fast, general means for determining the sugar linkage positions for disaccharides of bacterial glycoproteins.
4.4. Hydrolysis of CenA and CenAIgAIh by IgAI proteases

This study built on a previous study in which *N. gonorrhoeae* type 1 and 2 proteases were shown to cleave CenAIgAIh but not CenA (Miller *et al.*, 1992). Cleavage of CenAIgAIh was much slower than for IgAI. This reduction in proteolysis rate was hypothesized to be due to the conformation within and on either side of the target sequence. It could also have been due to the lack of glycosylation on CenAIgAIh produced in *E. coli* as human IgAI has a glycosylated hinge region. This study tested whether glycosylation of the hinge of CenAIgAIh would increase the efficiency of cleavage. However, the substrates glycosylated by *S. lividans* were cleaved less efficiently than their non-glycosylated counterparts, except by the *H. influenzae* protease (Section 3.8). Glycosylation is organism- and cell-specific, so it is reasonable that the glycosylation pattern is different for human IgAI and the hybrid protein produced in *S. lividans*. The percent carbohydrate on the linkers did not differ greatly between CenAIgAIh and myeloma IgAI, and so it is likely the other differences in glycosylation, including sites and composition, may have hampered the proteolysis (Section 3.9). The glycosylation sites of pooled serum IgAI more closely resemble those of CenAIgAIh produced in *S. lividans*, but this form of IgAI was not tested in this study. This gives further evidence that the glycosylation pattern may be important for proteolysis (Reinholdt *et al.*, 1990).

Results from this study support previous reports that features beyond amino acid sequence may be important determinants of cleavage efficiency for IgAI proteases (Lomholt, 1996; Miller *et al.*, 1992). Regardless of glycosylation, the CenA and CenAIgAIh substrates were all cleaved much less efficiently than human IgAI. Other
substrates, such as LAMP1 and some ape IgAl molecules, were also poor substrates (Lin et al., 1997; Qiu et al., 1996).

Despite the poor cleavage rate, interesting insights were obtained using CenA and CenAIgAlh as substrates. *H. influenzae* protease cleaved at two sites in CenA and CenAIgAlh, but only at one site in IgAl. This is likely to be due to conformational differences of the domains surrounding the hinge or linker. The *H. influenzae* protease interacts with the Fcα domain of IgAl and the site of cleavage correlates with the length of the cleavage specificity domain. Without the Fcα domain, control over the cleavage specificity may be lost. Interestingly, the *H. influenzae* protease does not require a P/S bond for cleavage; demonstrated by cleavage of the CenA linker which lacks a P/S bond. Cleavage of bonds by specific IgAl proteases different from that cleaved in human IgAl has been demonstrated previously. This was the case for the *C. ramosum* protease cleavage of gorilla IgAl, and the *N. meningitidis* type 1 and 2 proteases cleavage of IgAl from orangutans (Qiu et al., 1996). The *N. meningitidis* and *N. gonorrhoeae* type 2 IgAl proteases cleave the identical peptide bond in IgAl. Both cleaved CenAIgAlh. However, the susceptibility of CenA to these enzymes reveals a difference. Only the *N. meningitidis* type 2 protease could cleave CenA. This is the first example of such differentiation and is particularly striking and worth following up as the two enzymes have high sequence identity.

The IgAl proteases had different substrate profiles and therefore likely have unique substrate requirements. Further studies need to be done to elucidate the intricacies of IgAl protease cleavage, especially as the proteases are considered to be good vaccine candidates.
There are few alternative substrates to IgA1 for studying IgA1 proteases. CenA and CenAIgA1h are good candidates for alternative substrates to help define properties of substrates and proteases responsible for the specificity of some IgA1 proteases because the ability to cleave the substrates differed between the proteases tested, most importantly between some proteases with high sequence identity, and they are readily produced. Future experiments using CenA and CenAIgA1h as substrates for naturally occurring and designed IgA1 protease mutants will likely yield further insights into their intricacies.

In summary, this study has defined CenA as an IgA1 protease substrate, demonstrated that an IgA1 hinge homologue, CenAIgA1h, was cleaved by a variety of IgA1 proteases, and that some oligosaccharides may interfere with cleavage. CenA and CenAIgA1h will be useful in defining the properties beyond the cleaved region that contribute to specificity of IgA1 proteases.

4.5. Final Summary

This study has shown that modifying the size and composition of the PT linker of CenA had a limited effect on hydrolysis rates on cellulosic substrates. Changes in composition, however, did have an effect on the susceptibility to proteolysis by papain and C. fimi protease for both glycosylated and non-glycosylated proteins. This knowledge could be applied in the design of linkers for CBD-fusion proteins to prevent cleavage, or to add cleavage sites. The analysis of the glycan components of the linkers added to the knowledge base of the fledgling field of prokaryotic glycosylation and introduced the FACE® technology as a fast, general means for determining sugar linkage positions for
disaccharides of bacterial glycoproteins. Having established some of the linkage positions of disaccharides on glycoproteins made by *S. lividans*, it would be valuable to identify the glycosyltransferases involved since glycosyltransferases have potential commercial value. Finally, this study has defined CenA and CenAlgA1h as useful substrates for future investigations into properties influencing the specificity of IgA1 proteases.
5. Bibliography


Novagen. TB#5. 4. 1994.


Qiagen Inc. Qiaex II handbook. 1995. Chatsworth, CA, Qiagen Inc.


domains of endoglucanase A (CenA) of Cellulomonas fimi alters its conformation and catalytic activity. Journal of Biological Chemistry 266, 11335-11340.


