# A Novel Glycoconjugation Method and Unexpected Photo-Induced Site-Selective Protein Cleavage Reaction

A thesis submitted to the Board of the Faculty of Physical Sciences
in partial fulfilment for the degree of
Doctor of Philosophy
at the
University of Oxford

Nicola Floyd
New College, University of Oxford
Trinity Term 2009

**A Novel Glycoconjugation Method and**

**Unexpected Photo-Induced Site-Selective Protein Cleavage Reaction**

A thesis submitted to the Board of the Faculty of Physical Sciences in partial fulfilment for the degree of Doctor of Philosophy at the University of Oxford

Nicola Floyd, New College, Trinity Term, 2009

Homogenous glycoforms of glycoproteins are one of the primary targets in glycobiology to allow for the determination of glycoprotein function and to create glycoprotein mimetics for use as therapeutic agents. Although *N*- and *O*-linked glycosylation is preferred in nature, *S*-linked glycoproteins are attractive synthetic targets due to their enhanced chemical stability and enzymatic resistance. This thesis reports a novel convergent method for the synthesis of *S*-linked glycoproteins through the site-specific ligation of 1-glycosyl thiols to olefin containing proteins. This strategy employs unnatural amino acid incorporation for the introduction of L-homoallylglycine (L-Hag) into proteins and free-radical addition hydrothiolation chemistry, under conditions mild enough to preserve protein activity. The unique reactivity profile of L-Hag, which has an olefinic side-chain, compared with the natural amino acids characteristically found in proteins, allows for a chemoselective chemical reaction. At the outset, methodology was optimised using Hag as a model system. The reaction was then successfully applied to proteins with different structures, for precise, site-selective glycoconjugation at single and even multiple sites.

Whilst investigating the above-mentioned photochemical modification of proteins, an unexpected photo-induced site-selective protein cleavage reaction localised to the TIM-barrel proteins from family 1 of glycosylhydrolases was discovered. Although proteins are known to be affected by UV irradiation, examples of clean, site-selective photocleavage without the use of a cleaving agent are rare. Remarkably, the β-glycosidase from *Sulfolobus solfataricus* (*Ss*βG) was found to form two daughter fragments following UV irradiation (240–308 nm). Alterations at and alanine-scanning of a sphere of residues around the cleavage site were used to establish which residues were essential for the reaction to occur and extensive analysis of the fragments allowed a reaction mechanism to be proposed. The propensity of other proteins to undergo photo-induced cleavage was also investigated. In addition, the biotechnological utility of the aforementioned photocleavage reaction has been applied to the development of a cleaved-tag affinity purification method using a GFP-*Ss*βG fusion as the model protein.

## Table of contents

**Abstract** 1

**Table of contents** 2

**List of abbreviations** 9

**Acknowledgements** 15

**Chapter 1: Introduction**

1. Biological significance of glycoproteins 17
2. Homogenous glycoforms of glycoproteins 18
3. *N*-, *O*- and *S*-linked protein glycosylation 19
4. Glycoprotein assembly 20
5. Criteria for selection of modification residue 21
6. Convergent glycoprotein synthesis 22
   - 6.1. Indiscriminate protein glycosylation 22
   - 6.2. Chemoselective and site-specific protein glycosylation 25
   - 6.3. Site-selective protein glycosylation 29
     - 6.3.1. Site-selective glycosylation of naturally occurring amino acids 29
     - 6.3.2. Site-selective glycosylation of unnatural amino acids 31
       - 6.3.2.1. Unnatural amino acid incorporation 31
       - 6.3.2.2. Site-selective glycosylation 34
   - 6.4. Thiol-ene reaction for protein modification 39
7. The use of glycoproteins as therapeutic agents 41
   - 7.1. Glycoprotein vaccines 41
     - 7.1.1. Anti-bacterial, -parasitic and -fungal vaccines 41

7.1.2. Anti-HIV vaccines 43

7.1.3. Anti-cancer vaccines 44

**Chapter 2: Thiyl-Glycosylation of Olefinic Proteins – *S*-Linked Glycoprotein Synthesis**

(Floyd, N; Vijayakrishnan, B.; Koeppe, J. R.; Davis, B. G., *Angew. Chem. Int. Ed.*, 2009, 48, 7798–7802)

1. Introduction 47

2. Homoallylglycine (Hag) synthesis 49

2.1. DL-Hag synthesis 49

2.2. L-Hag synthesis 50

2.2.1. Via *t*-butyl ester route 50

2.2.2. Via methyl ester route 52

2.3. Enantiomeric excess determination of L-Hag 55

2.3.1. Chiral HPLC 56

2.3.2. Derivatisation 56

2.3.2.1. Sanger's reagent 56

2.3.2.2. Mosher's acid chloride 57

2.3.2.3. Fmoc-L-Ala-OPfp and Fmoc-L-Asn(Trt)-OPfp 58

3. 1-Glycosyl thiol synthesis 61

3.1. 1-Thio-β-D-glucose synthesis 62

3.2. 1-Thio-α-D-glucose synthesis 63

3.3. 2-Acetamido-2-deoxy-1-thio-β-D-glucopyranose synthesis 64

3.4. 1-Thio-β-D-galactose synthesis 65

3.5. 1-Thio-β-D-mannose synthesis 66

4. pH of 1-glycosyl thiol solutions 67

5. Effect of UV light and heating on L-Hag and 1-glycosyl thiols 68

5.1. L-Hag 68

5.2. 1-Glycosyl thiols 70

6. Glyco-amino-acid synthesis 71

7. Chemoselectivity for natural amino acids 77

8. Model protein preparation 79

8.1. *Ss*βG-Hag43 80

8.1.1. M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG 80

8.1.2. M21I M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG (*Ss*βG-Met43) and M21I M43Hag M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG (*Ss*βG-Hag43) 82

8.2. Np276-Hag61 86

8.3. Qβ-$(Hag16)_{180}$ 87

9. *S*-linked glycoprotein synthesis 90

9.1. Site-selective glycosylation of *Ss*βG-Hag43 90

9.2. Site-selective glycosylation of Np276-Hag61 96

9.3. Site-selective glycosylation of Qβ-$(Hag16)_{180}$ 98

9.3.1. Ionisation efficiency 100

10. Conclusion and future work 101

**Chapter 3: An Unexpected Photo-Induced Site-Specific Protein Cleavage Reaction**

(Floyd, N., Oldham, N. J., Eyles, C. J., Taylor, S., Filatov, D. A., Brouard, M., Davis, B. G., *J. Am. Chem. Soc.* 2009, 131, 12518–19)


1. Introduction 103
2. Photo-induced cleavage reaction 108
    2.1. M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG 108
    2.2. M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG 109
    2.3. Wild-type *Ss*βG 111
3. Reactive transitions 114
4. Cleavage site identification 116
5. Probing the cleavage site 117
6. His-Trp diad 121
    6.1. S426P *Ss*βG 121
    6.2. Peptides 123
7. Alanine scanning 124
8. Characterisation of the cleavage site 134
    8.1. Characterisation of the 18 kDa fragment 134
    8.2. Characterisation of the 39 kDa fragment 138
9. Tagging the 39 kDa fragment 143
10. Oxygen related experiments 144
11. Effect of substrate, metal and EDTA 147
12. Mechanism 148

13. Generality of the photocleaved His-Trp motif ..... 149

13.1. Identification of other photocleavable proteins: 1 ..... 149

13.1.1. β-Glycosidase from Thermosphaera aggregans (TaβG) ..... 150

13.1.2. Arylsulfatase from Pseudomonas aeruginosa ..... 152

13.1.3. *Chalcone reductase from* Medicago sativa ..... 154

13.2. Identification of other photocleavable proteins: 2 ..... 155

13.2.1. β-Glucosidase from *Pyrococcus furiosus* (*Pf*βG) ..... 155

13.2.2. β-Glycosidase from *Sulfolobus solfataricus* (Y149F *Ss*βG) ..... 156

13.2.3. β-Glucosidase from *Clostridium thermocellum* (*Ct*βG) and β-Glucosidase from *Rhizobium etli* (*Re*βG) ..... 157

13.2.4. Myrosinase from *Sinapis alba* (*Sa*M) ..... 159

13.2.5. β-Glucosidase from *Streptococcus pyogenes* (*Sp*βG) ..... 159

13.3. Summary of proteins exposed to UV light ..... 160

14. Bioinformatic and phylogenetic analyses ..... 162

14.1. Multiple sequence alignments ..... 162

14.2. MEME analysis ..... 163

14.3. Phylogenetic analysis ..... 164

15. Conclusion and future work ..... 166

**Chapter 4: Cleaved-Tag Affinity Purification Using the Protein Cleavage Reaction**

1. Introduction ..... 168

2. *Ss*βG truncated proteins (T0-, T50-, T100- and T140-*Ss*βG) ..... 169

3. Development of the cleaved-tag affinity purification method ..... 173

4. Green fluorescent protein (GFP) from *Aequorea victoria* ..... 178

5. GFP-*Ss*βG fusion 180

5.1. Protein preparation and photocleavage 180

5.2. Degradation 182

5.3. Cleaved-tag affinity purification 183

5.4. GFP-*Ss*βG$_{tag}$ refolding 184

5.5. Guanidine hydrochloride concentration 187

6. Conclusion and future work 189

**Chapter 5: Summary** 191

**Chapter 6: Experimental** 195

**Chapter 7: References** 403

**Licence Agreement** 430

## List of abbreviations

| | |
|---|---|
| **Å** | angstrom |
| **Ac** | acetyl |
| **Ala / A** | alanine |
| **app** | apparent |
| **Arg / R** | arginine |
| **Asn / N** | asparagine |
| **Asp / D** | aspartic acid |
| **Boc** | *tert*-butoxy-carbonyl |
| **BSA** | bovine serum albumin |
| **c** | concentration 10 mg/mL |
| **°C** | degrees Celsius |
| **CAZy** | carbohydrate-active enzymes |
| **CAPS** | *N*-cyclohexyl-3-amino-propanesulfonic acid |
| **Calcd.** | calculated |
| **CHES** | *N*-cyclohexyl-2-aminoethanesulfonic acid |
| **conc.** | concentrated or concentration |
| **COSY** | correlation spectroscopy |
| ***Ct*βG** | β-glucosidase from *Clostridium thermocellum* |
| **CV** | column volume |
| **Cys / C** | cysteine |
| **d** | doublet or deci |
| **δ** | chemical shift |
| **Da** | Dalton |
| **DCM** | dichloromethane |
| **dd** | doublet of doublets |

| | |
|---|---|
| **ddd** | doublet of doublets of doublets |
| **DEPT** | distortion less enhancement by polarisation transfer |
| **D*i*BAL-H** | diisobutyl aluminium hydride |
| **DMAP** | 4-(*N,N*-dimethylamino)-pyridine |
| **DMF** | *N,N*-dimethyl formamide |
| **DMPU** | 1,3-dimethyl-2-oxohexa-hydropyrimidine |
| **DMSO** | dimethylsulfoxide |
| **dt** | doublet of triplets |
| **DTT** | dithiothreitol |
| **EDTA** | ethylenediamine tetraacetic acid |
| **Eq.** | equivalent(s) |
| **ESI-MS** | electrospray ionisation mass spectrometry |
| **Et** | ethyl |
| **FF** | fast flow |
| **Fmoc** | fluorenylmethyloxycarbonyl |
| **FPLC** | fast protein liquid chromatography |
| **Gal** | galactose |
| **GFP** | green fluorescent protein |
| **GH1** | family 1 glycosylhydrolase |
| **Glc** | glucose |
| **GlcNAc** | *N*-acetylglucosamine |
| **Gln / Q** | glutamine |
| **Glu / E** | glutamic acid |
| **Gly / G** | glycine |
| **Gn** | guanidine |
| **h** | hour(s) |

| | |
|---|---|
| **Hag** | homoallylglycine |
| **HEPES** | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| **His / H** | histidine |
| **HMM** | hidden Markov model |
| **HMPA** | hexamethylphosphoramide |
| **HMQC** | heteronuclear multiquantum coherence |
| **HOBt** | hydroxybenzotriazole |
| **HPLC** | high performance liquid chromatography |
| **HRMS** | high resolution mass spectrometry |
| **Hz** | hertz |
| **Ile / I** | isoleucine |
| **IPTG** | isopropyl-β-D-thio-galactopyranoside |
| **IR** | infrared |
| ***J*** | coupling constant |
| **kDa** | kilodalton |
| **L** | litre |
| **Lac** | lactose |
| **Leu / L** | leucine |
| **Lindlar's catalyst** | $Pd/Pb(OAc)_2$ on $CaCO_3$ |
| **Lit.** | literature value |
| **LRMS** | low resolution mass spectrometry |
| **Lys / K** | lysine |
| **m** | multiplet, milli or metre |
| **M** | molar or mega |
| **µ** | micro |
| **Man** | mannose |

| | |
|---|---|
| **Me** | methyl |
| **MEME** | multiple EM for motif elicitation |
| **Met / M** | methionine |
| **MHMS** | medium-high resolution mass spectrometry |
| **Mosher's acid** | 3,3,3-trifluoro-2-methoxy-2-phenyl-propanoic acid |
| **m.p** | melting point |
| **MS** | mass spectrometry |
| **MS/MS** | tandem mass spectrometry |
| **MUSCLE** | multiple sequence comparison by log-expectation |
| **MWCO** | molecular weight cut off |
| ***m/z*** | mass-to-charge ratio |
| **n** | nano |
| ***n*-Bu** | normal butyl |
| **ND-PAGE** | non-denaturing polyacrylamide gel electrophoresis |
| **NMR** | nuclear magnetic resonance |
| **OD** | optical density |
| **PBS** | phosphate-buffered saline |
| **PDB** | protein data bank |
| ***Pf*βG** | β-glucosidase from *Pyrococcus furiosus* |
| **Pfp** | pentafluorophenyl |
| **pH** | pouvoir hydrogène |
| **Ph** | phenyl |
| **Phe / F** | phenylalanine |
| **ppm** | parts per million |
| **Pro / P** | proline |

| | |
|---|---|
| **PyBOP** | benzotriazole-1-yl-oxy-tris-pyrrolidinophosphonium hexafluorophosphate |
| ***Re*βG** | β-glucosidase from *Rhizobium etli* |
| $R_f$ | retention factor |
| **Rochelle salt** | potassium sodium tartrate |
| **rpm** | revolutions per minute |
| **RT** | room temperature |
| **s** | singlet |
| ***Sa*M** | myrosinase from *Sinapis alba* |
| **Sanger's reagent** | 2,4-dinitrofluorobenzene |
| **sat.** | saturated |
| **SDS-PAGE** | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| **Ser / S** | serine |
| **SPASM** | spatial arrangements of side chains and main chains |
| ***Sp*βG** | β-glucosidase from *Streptococcus pyogenes* |
| ***Ss*βG** | β-glycosidase from *Sulfolobus solfataricus* |
| **t** | triplet |
| ***Ta*βG** | β-glycosidase from *Thermosphaera aggregans* |
| **TAE** | tris-acetate-EDTA |
| **TBS** | tris-buffered saline |
| ***t*Bu** | tertiary butyl |
| **TCEP** | tris(2-carboxyethyl)phosphine |
| **td** | triplet of doublets |
| **TEA** | triethanolamine |
| **Tebbe reagent** | bis(cyclopentadienyl)-μ-chloro-(dimethylaluminium)-μ-methylenetitanium |

| | |
|---|---|
| **temp.** | temperature |
| **TFA** | trifluoroacetic acid |
| **THF** | tetrahydrofuran |
| **Thr / T** | threonine |
| **TIM-barrel** | triosephosphateisomerase-barrel |
| **t.l.c.** | thin layer chromatography |
| **TMS** | trimethylsilyl |
| **TOF** | time of flight |
| **Tris-HCl** | tris(hydroxymethyl)aminomethane-HCl |
| **Trp / W** | tryptophan |
| **Tyr / Y** | tyrosine |
| **UV** | ultraviolet |
| **Vazo44/VA044** | 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride |
| **Val / V** | valine |
| **v/v** | volume/volume |
| **w%** | weight% |
| **WT** | wild-type |
| **X-Gal** | 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside |

**Acknowledgements**

Firstly, I would like to thank Prof. Benjamin Davis for giving me the opportunity to study for a D. Phil. in Chemistry at Oxford University. It is without question that his outstanding knowledge of science, enthusiasm, and belief that chemistry and biology can work side by side have allowed this project to get to where it is today.

I am extremely grateful to the International AIDS Vaccine Initiative (IAVI) for funding.

Plasmids were kindly supplied by Dr. Holger Kramer, University of Oxford – WT *Ss*βG and M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG; Dr. James Errey, University of Oxford – M21I Np276; Dr. Rebecca Bell, University of Oxford - E387A *Ss*βG; Prof. Joseph Noel, SALK Institute - Chalcone reductase from *Medicago sativa*; Dr. Michael Kertesz, University of Manchester – Arylsulfatase from *Pseudomonas aeruginosa*; Dr. Marco Moracci, Institute of Protein Biochemistry CNR Naples - β-Glycosidase from *Thermosphaera aggregans* (*Ta*βG); Prof. M. G. Finn, The Scripps Research Institute - K16M Qβ and Qβ-WT.

Proteins were kindly supplied by Dr. Serve Kengen, Wageningen University – β-Glucosidase from *Pyrococcus furiosus* (*Pf*βG); Miss. Samantha Cheung, University of Oxford – E206A *Ss*βG; Dr. Julia Koeppe, University of Oxford – Qβ-WT and Qβ-$(Hag16)_{180}$.

The 1-glycosyl thiols β-LacSH, $(\beta\text{-GlcS})_2$ and α/β-$Man_3SH$ were synthesised by Dr. Balakumar Vijayakrishnan, University of Oxford. Dr. Balakumar Vijayakrishnan also assisted with a number of the glycoconjugation reactions, along with Dr. Julia Koeppe.

I am indebted to Dr. Neil Oldham (University of Nottingham) for his help with mass spectrometry throughout the course of this project, and in particular for assisting with characterisation of the cleavage site of *Ss*βG.

All experiments that required the use of laser light were carried out with the help of Mr. Chris Eyles and Prof. Mark Brouard (University of Oxford), in the laboratory of Prof. Mark Brouard.

Bioinformatic and phylogenetic analyses were provided by Mr. Steve Taylor (Computational Biology Research Group, University of Oxford) and Prof. Dmitry Filatov (Dept. Plant Sciences, University of Oxford).

Fluorescence microscopy was carried out by Dr. Mark Wallace (University of Oxford).

Special thanks go to Dr. James McCullagh and Mr. Robin Procter for MS technical assistance; Dr. Tim Claridge and Dr. Barbara Odell for NMR technical assistance; Prof. Peter Ogilby (Aarhus University) and Dr. James Errey for helpful discussion; and also the CRL support staff.

Further, I would like to acknowledge the numerous other individuals I have called on for help over the past four years, not only in the Ben Davis group, but also the Chris Schofield, Hagen Bayley and Mark Wallace groups - you know who you are! A big thank you also goes to my numerous proof readers for their efforts.

Finally, my thanks go to my family and friends for their undivided support and encouragement, especially my mum, dad, sisters and love of my life, Alex.

# Chapter 1: Introduction

## 1. Biological significance of glycoproteins

Carbohydrates are versatile scaffolds able to form an unrivalled number of structures compared with other biological building blocks such as amino acids and nucleotides, due to variation in configuration (*e.g.* glucose, mannose), ring size, branching, anomeric configuration and modification (*e.g.* phosphorylation).[1, 2] In contrast, amino acids and nucleotides can form only linear structures and have limited basis sets, *i.e.* 20 amino acids for proteins and 4 nucleotides for DNA/RNA. This structural diversity is reflected in the functional diversity of carbohydrates. Glycogen and starch, for example, act as energy sources, whilst chitin and cellulose act as structural components of animals and plants, respectively. In nature an even greater variety in oligosaccharide structure is exploited by the formation of glycoproteins in which carbohydrates are covalently linked to proteins.

Over half of all proteins in nature are estimated to be glycosylated.[3] Therefore, it is unsurprising that these biomolecules play key roles in protein expression, folding, stability and activity,[4-11] and are key to various biological processes[12-18] such as cell signalling regulation,[19-21] cellular differentiation,[22] and the immune response.[23] For example, the HIV (Human Immunodeficiency Virus) envelope glycoprotein gp120, which protrudes from the surface of the HIV virion, facilitates virus entry into target host cells by binding to CD4 receptors (Figure 1).[24]

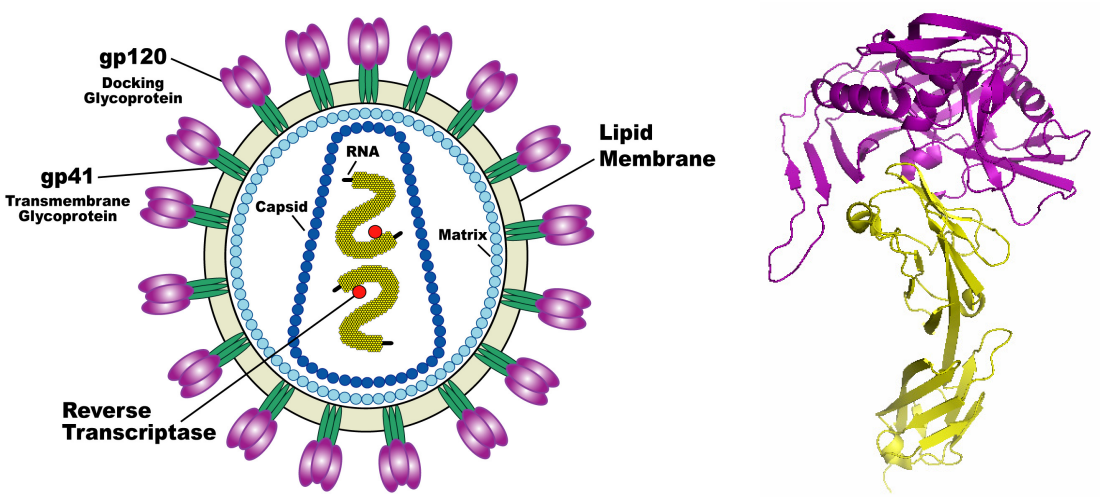



**Figure 1.** Organization of the HIV virion[25] (left) and ribbon diagram of the HIV gp120 core (purple) complexed with CD4 (yellow) (PDB ID: 1GC1) (right).

## 2. Homogenous glycoforms of glycoproteins

Glycoproteins occur naturally as mixtures of so-called glycoforms.[26] Glycoforms possess the same peptide backbone but vary in both the nature and the sites of glycosylation. The glycosylation pattern of any protein is ultimately dictated by: its amino acid sequence, local conformational properties at potential glycosylation sites, donor/substrate availability and the accessibility of carbohydrate processing enzymes such as glycosyltransferases and glycosidases. One example of a glycoprotein that occurs naturally as a mixture of glycoforms is bovine pancreatic ribonuclease B (RNase B).[27, 28] This protein consists of five different glycoforms due to variation in glycosylation at the single *N*-glycosylation site (Asn34LeuSer). At this site, five to nine mannose residues are attached to a chitobiose core ($Man_{5\text{-}9}GlcNAc_2$). Because individual constituents within microheterogeneous mixtures of glycoproteins exhibit different properties,[27, 29] the elucidation of their precise biological activity is unfeasible and their use as pharmaceutical drugs limited.[30] However, glycoprotein heterogeneity is not known to hinder the physiological role of individual glycoforms,

and indeed may even be an important contributor to glycoprotein function.[31, 32] In recent years, homogenous glycoforms[33-41] of glycoproteins have become one of the primary targets in glycobiology[17, 18] to allow for the determination of glycoprotein function and to create glycoprotein mimetics for use as therapeutic agents. One example of a therapeutic glycoprotein is erythropoietin (EPO), which can be used to treat kidney disease related anaemia.[42]

### 3. *N*-, *O*- and *S*-linked protein glycosylation

*N*- and *O*-linked protein glycosylation predominates in nature over other carbohydrate-protein linkages[43] (*e.g.* *C*-mannosyl,[44] phosphoglycosyl[45] and GPI [glycosylphosphatidyl-inositol] phosphoethanolamine bridges).[46] In *N*-linked protein glycosylation, glycans are attached to the amide nitrogen of asparagine in proteins bearing the consensus sequence of amino acids, AsnXaaSer/Thr, where Xaa is any amino acid except proline. However, this sequence does not always result in glycosylation. All *N*-linked glycans expressed by eukaryotes possess a common core structure comprising the pentasaccharide, $Man_3GlcNAc_2$ (Figure 2). Glycosylation of the side-chain of serine or threonine in proteins by an α-GalNAc residue forms the basis of *O*-linked protein glycosylation (Figure 2).[47] However, every amino acid with a hydroxyl functional group (*i.e.* tyrosine,[48] hydroxyproline[49] and hydroxylysine)[50] has been implicated in *O*-linked protein glycosylation along with the monosaccharides glucose (Glc), *N*-acetylglucosamine (GlcNAc), xylose (Xyl), galactose (Gal), fucose (Fuc), pseudaminic acid (Pse) and mannose (Man).[51] No consensus sequence is known for this type of glycosylation.

Although *N*- and *O*-linked glycosylation is preferred in nature, *S*-linked glycoproteins are attractive synthetic targets due to their enhanced chemical stability and enzymatic resistance.[52] However, naturally occuring *S*-glycosidic linkages have been identified. Lote and Weiss for example isolated peptides from human urine[53, 54] and human erythrocyte membranes[55] in which galactose and glucose, respectively, were attached to the sulphur of an N-terminal cysteine residue.

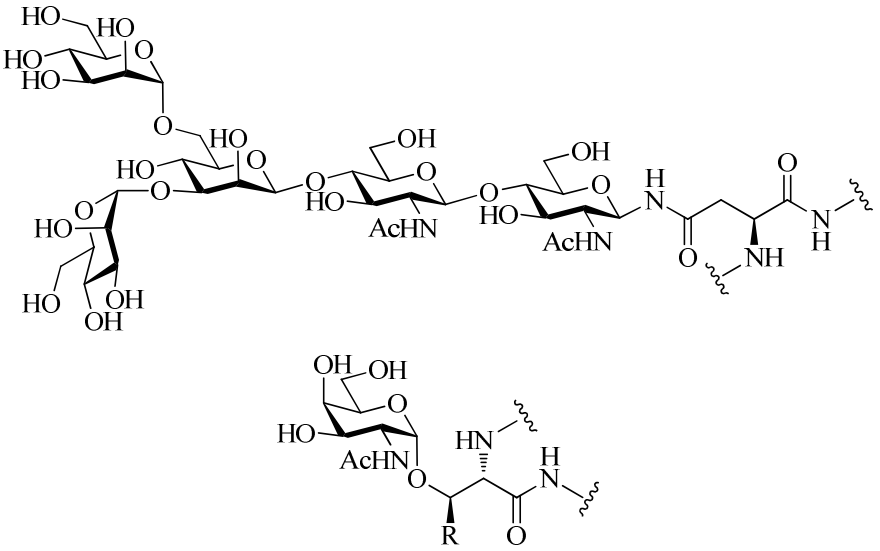



**Figure 2.** Core pentasaccharide of *N*-linked glycans (top) and representative glycopeptide linkage in *O*-linked glycoproteins (R = H or $CH_3$).

## 4. Glycoprotein assembly

As a means of investigating their activity and creating therapeutic glycoproteins there has been a concerted effort to develop new methods to synthesize homogenous glycoforms of glycoproteins in recent years. To this end, several glycoprotein assembly strategies exist, and are shown in Figure 3.[17, 18] By method A, glycopeptide fragments are ligated into proteins using enzymatic or protein ligation methods such as native chemical ligation (NCL)[56] or expressed protein ligation (EPL).[57] Alternatively, the carbohydrate can be linked directly to the protein by a convergent

method, which is depicted by method B. By method C, the glycoprotein is remodeled. Remodelling involves the combined trimming of the existing glycan structures to a common core, typically enzymatically using glycosidases, followed by elaboration to alternative ones using glycosyltransferases to build up the desired glycoprotein structure.

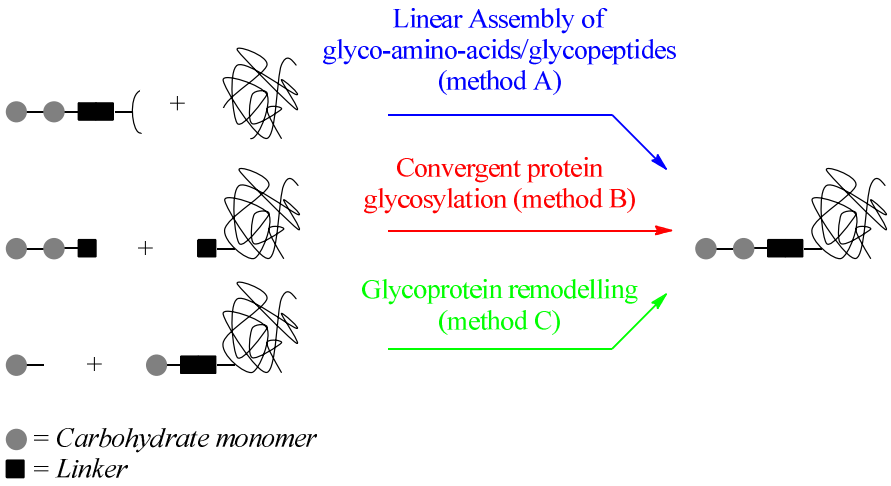



**Figure 3.** Retrosynthetic analysis of glycoprotein preparation.[18]

## 5. Criteria for selection of modification residue

When developing a method for protein conjugate synthesis it is important to select the modification residue carefully. In order to create useful protein conjugates it is vital to avoid strategies that modify residues necessary for protein function, stability or that are critical for protein architecture. For example, if cysteine is selected as the modification residue then the chemistry used must be compatible with the presence of disulfide bonds, since these are known to be critical for the maintenance of tertiary protein structures. In addition, it is important to consider the relative abundance of different amino acid residues present in natural proteins (Table 1).[58] For example, if a residue that is rare in proteins is targetted, the same chemoselective reaction that would result in indiscriminate glycosylation when used with a protein containing

multiple naturally occurring copies of that residue, can be site-specific when used with a protein containing a single naturally occuring copy.

**Table 1.** Relative abundance of amino acids residues in protein sequences.

| Residue | Abundance [%] | Residue | Abundance [%] | Residue | Abundance [%] |
|---|---|---|---|---|---|
| Trp | 1.4 | Gln | 4.5 | Lys | 6.0 |
| Cys | 1.6 | Pro | 4.5 | Val | 7.1 |
| Met | 2.3 | Arg | 5.0 | Gly | 7.6 |
| His | 2.3 | Asp | 5.8 | Ala | 8.2 |
| Tyr | 3.6 | Thr | 5.8 | Glu | 8.3 |
| Asn | 3.7 | Ile | 5.9 | Leu | 8.4 |
| Phe | 4.0 | Ser | 5.9 | | |

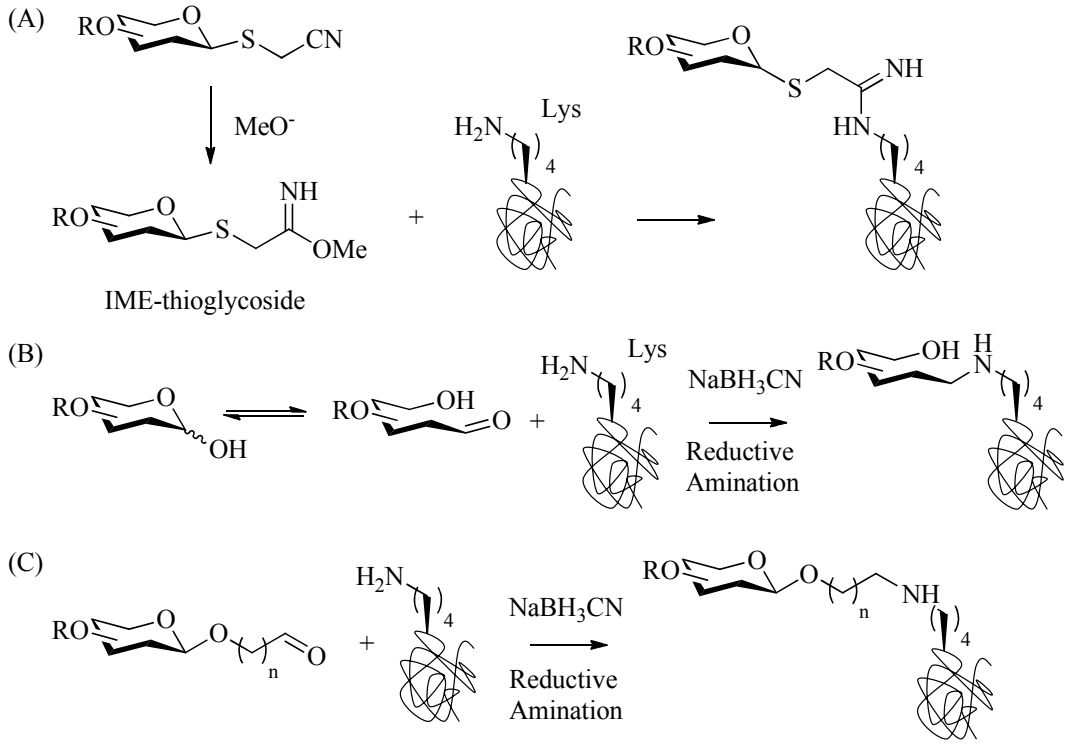
## 6. Convergent glycoprotein synthesis

### 6.1. Indiscriminate protein glycosylation

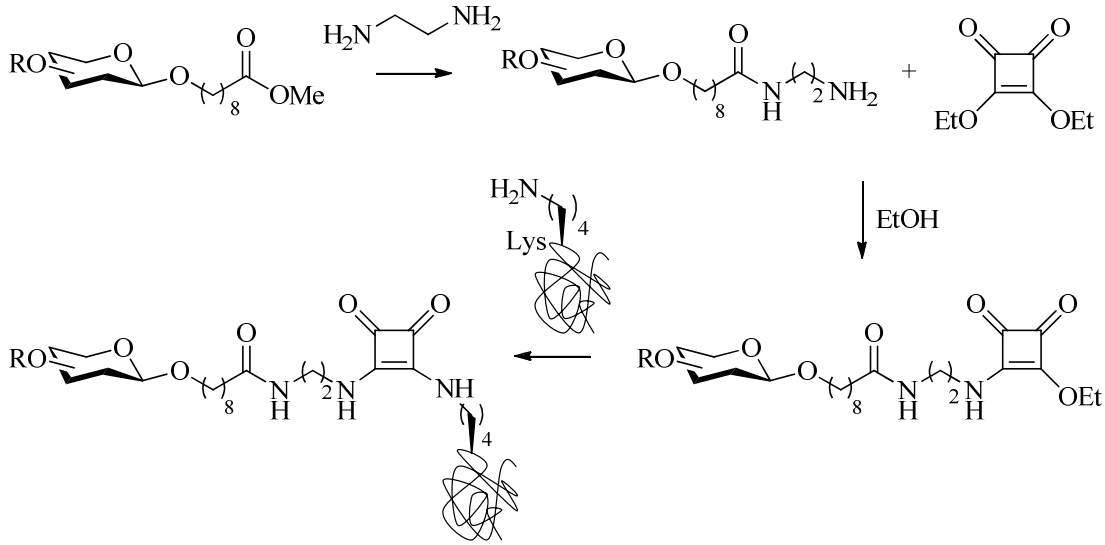
Indiscriminate protein glycosylation methods exploit functionality found naturally on the protein surface within the side-chains of proteinogenic amino acids. Although these indiscriminate glycosylation methods facilitate high levels of protein functionalisation, which can be used to increase the immune response, well-defined conjugates are often difficult to access due to a lack of residue selectivity. In addition, the intrinsic charge of the protein may be altered by these methods, which can affect the protein's properties. If the amine functional group in the side-chain of lysine, for example, is converted to an amide, lysine loses its positive charge and is converted to a neutral moiety. A selection of indiscriminate protein glycosylation methods are described below.[59]

Of the procedures employed for indiscriminate protein glycosylation, the modification of lysine side-chains using 2-imino-2-methoxyethyl 1-thioglycosides (IMEs) (Scheme 1A),[60] or by reductive amination[61, 62] are the dominating strategies (Scheme 1B). By the former method, lysine residues on the surface of adenovirus[63] and a rhamnosidase enzyme[64] have been successfully modified by Davis and co-workers. By the latter method, both Danishefsky *et al.*,[65] and Guo and *et al.*[66, 67] have generated potential carbohydrate-based anti-cancer vaccines. One feature of the reductive amination method is formation of the open-chain form of the carbohydrate from the cyclic form, which potentially can lead to loss of biological function. To avoid this problem and to address the fact that reductive amination conjugations are often accompanied by low protein loading levels due to steric hindrance caused by short spacer arms, an extra spacer arm can be added to the anomeric position of the carbohydrate (Scheme 1C).[66-69]

Squaric acid diesters can also be used for indiscriminate protein glycosylation (Scheme 2).[70-72] For example, a number of different glycans[73-75] have been conjugated to BSA using this chemistry, including a hexasaccharide fragment of the *O*-specific polysaccharide of *V. cholerae* O:1, serotype Ogawa by Saksena *et al.*.[76-78] This approach has also been used by Bundle and co-workers to construct thiooligosaccharide conjugate vaccines that evoke antibodies specific for native antigens.[79]



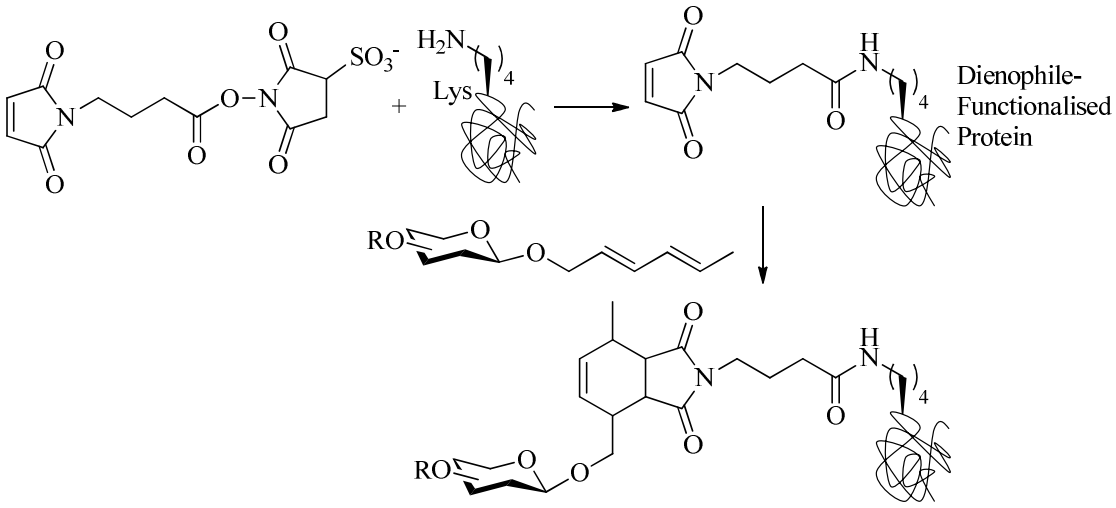

**Scheme 1.** Lysine side-chain modification using IME reagents and by reductive amination.




**Scheme 2.** Lysine side-chain modification using diethyl squarate.

In yet another approach for indiscriminate protein glycosylation, Pozsgay *et al.* demonstrated the use of the Diels-Alder cycloaddition reaction, for example to conjugate synthetic fragments of the capsular polysaccharide of *N. meningitidis* A to

human serum albumin (Scheme 3).[80, 81] Indiscriminate protein glycosylation can also be accomplished using a number of other methods which are not discussed here.[82-86]




**Scheme 3.** Lysine side-chain modification using Diels-Alder cycloaddition chemistry.

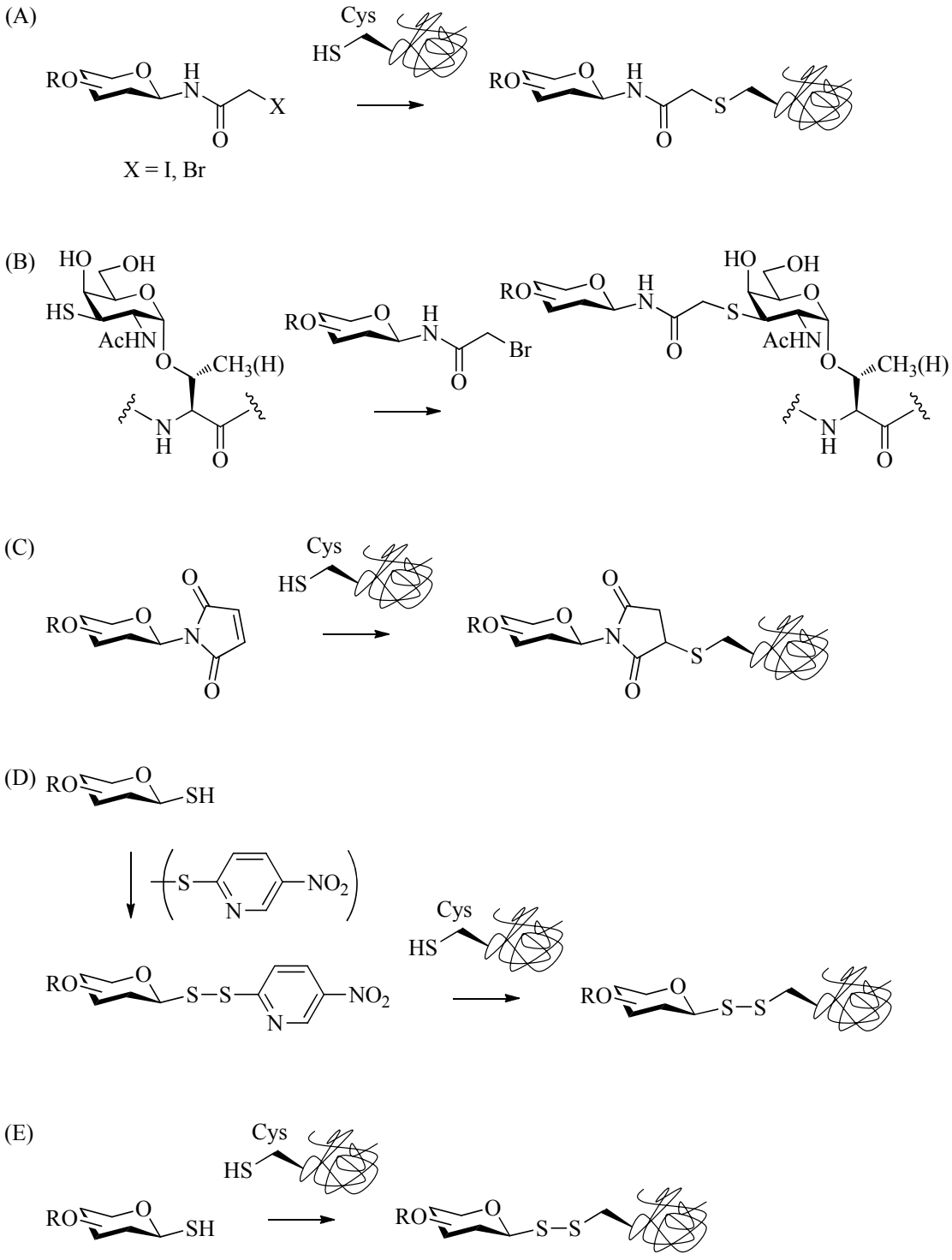
### 6.2. Chemoselective and site-specific protein glycosylation

Several approaches have been developed that exploit the chemoselectivities of a number of chemical methods, in an attempt to overcome the lack of selectivity and predictability of indiscriminate protein glycosylation methods.[59] A selection of chemoselective and site-specific protein glycosylation are described below.

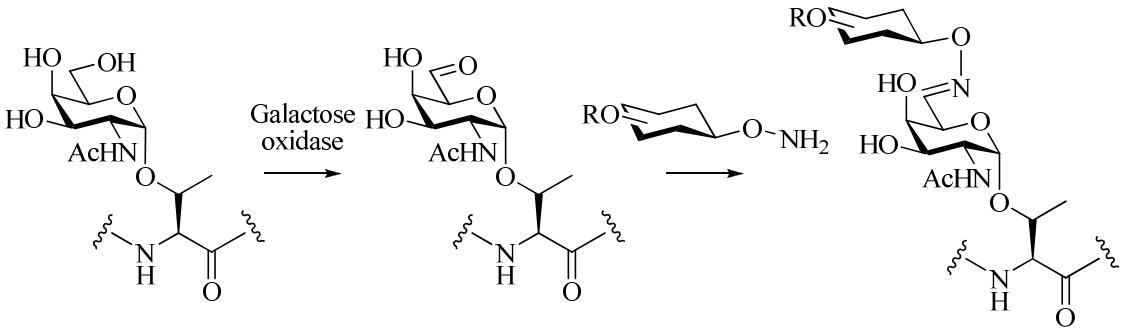
One of the most reactive and frequently modified amino acid side-chains is the thiol group of cysteine. Cysteine residues are relatively rare in proteins and because of this the same chemoselective reaction that would result in indiscriminate glycosylation when used with a protein containing multiple naturally occurring cysteine residues, can be site-specific when used with a protein containing a single naturally occurring cysteine residue. Commonly, modification of the sulfhydryl group is achieved with glycosyl haloacetamides (Scheme 4A)[87-89] or glycosyl maleimides (Scheme 4C),[90, 91]

to yield thioethers. The maleimide method has also been used with a linker between the anomeric position of the sugar and the maleimide moiety.[91] Using the iodoacetamides of *N*-acetylglucosamine[92] and chitotriose[93] for example, the single free cysteine of BSA has been successfully glycosylated. However, the haloacetamide method is known to suffer from limited site-specificity; in some cases nonspecific glycosylation of histidine and lysine residues has been oberved.[37, 94] Using the same chemistry, Bertozzi and co-workers synthesised *O*-linked glycopeptides bearing C-3 thioethers via alkylation of 3-thioGalNAc with bromoacetamido sugars (Scheme 4B).[95] Cysteine modification has also been accomplished through the use of 5-nitropyridine-2-sulfenyl activated thioglycosides (compatible only with monosaccharides) (Scheme 4D)[96] or aerial oxidation (Scheme 4E), to yield disulfide-linked glycoconjugates.[97]

Using the antibacterial 19-amino acid glycopeptide drosocin, Bertozzi and co-workers demonstrated that an aldehyde tag introduced selectivity onto the C-6 of a GalNAc residue in the peptide by use of the enzyme galactose oxidase, could be chemoselectivity coupled to a carbohydrate bearing an aminooxy group at the reducing end (Scheme 5).[98] This chemistry has also been adopted by others.[99-101]

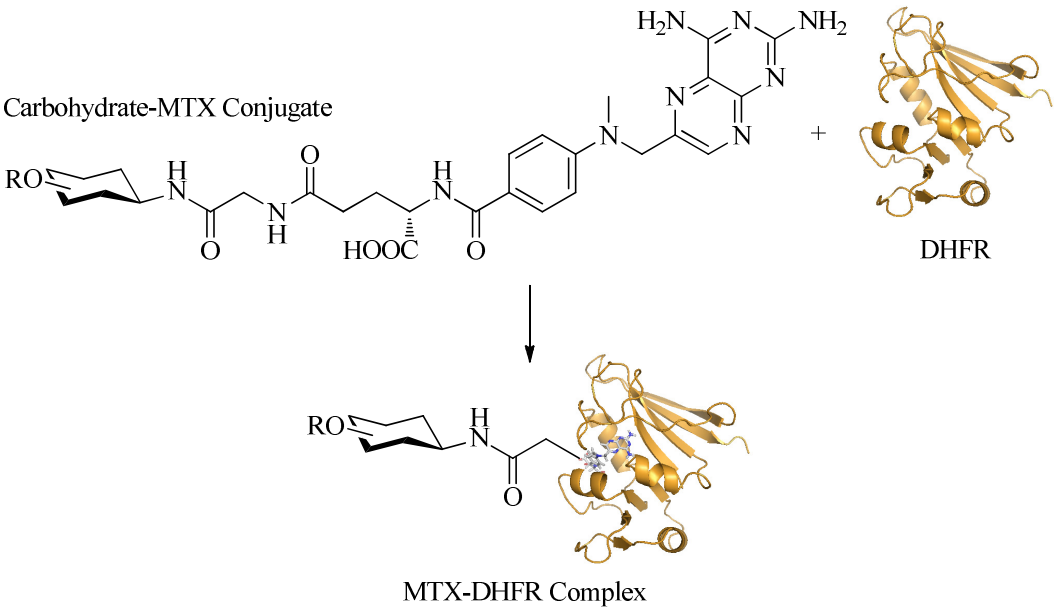



**Scheme 4.** Cysteine side-chain modification.




**Scheme 5.** Protein modification using galactose oxidase and aminooxy chemistry.

In order to create oligosaccharide-grafted proteins, a tight binding ligand approach that uses dihydrofolate reductase (DHFR) and methotrexate (MTX) as a protein-ligand pair has been developed by Totani *et al.*. This method has been used to create various carbohydrate-MTX conjugates which then tightly bind to DHFR.[102, 103]




**Scheme 6.** MTX-DHFR complex formation.

### 6.3. Site-selective protein glycosylation

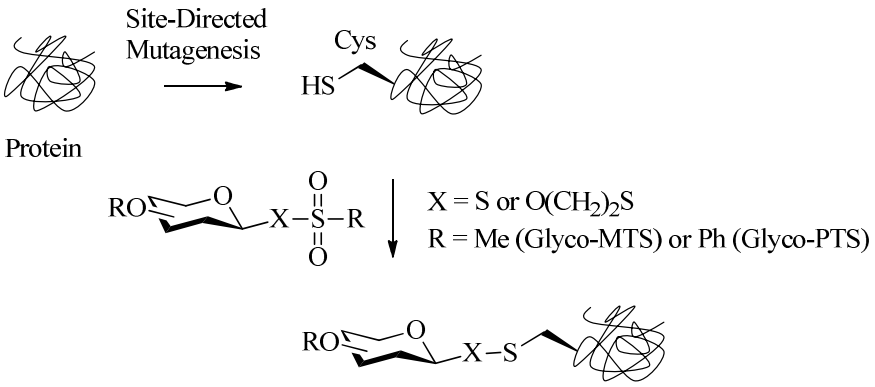
Several methods have been developed that allow complete flexibility in the choice of glycosylation site with concomitant control of the carbohydrate and nature of the sugar to protein spacer arm.[59] A selection of site-selective protein glycosylation methods are described below.

#### 6.3.1. Site-selective glycosylation of naturally occurring amino acids

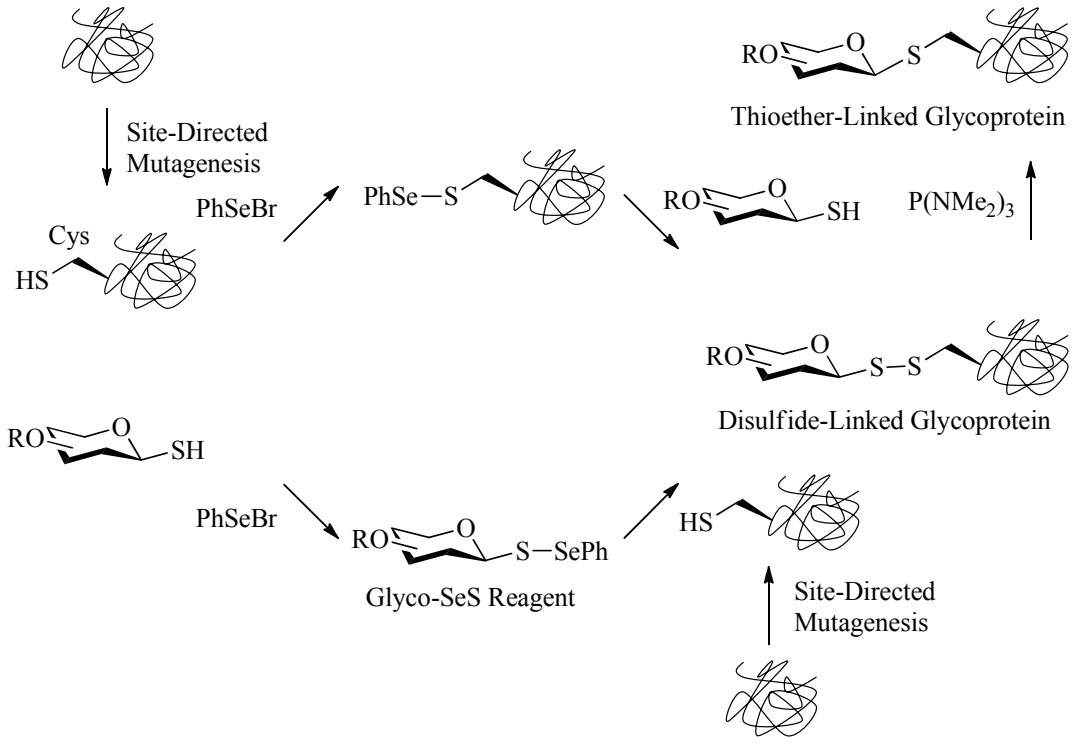
Site-directed mutagenesis in combination with chemical modification is one approach that has been used to achieve site-selective protein glycosylation. In the first example of this kind, site-directed mutagenesis was used to introduce a cysteine residue as a thiol tag at preselected positions in the serine protease subtilisin *Bacillus lentus* (SBL), a protein that does not contain any naturally occuring cysteine residues. This protein was subsequently glycosylated using thiol-selective glycomethanethiosulfonate (glyco-MTS) reagents (Scheme 7).[33-35] By employing this glyco-MTS method, the first example of a homogenous protein bearing branched multivalent glycans was demonstrated by Davis and co-workers.[104] Similarly, glycosylphenylthiosulfonate (glyco-PTS) reagents have also been used to site-selectively glycosylate proteins (Scheme 7).[105, 106]

Inspired by the natural occurrence of proteins containing selenenylsulfide (Se-S) bonds,[107] Gamblin *et al.* developed a selenenylsulfide-mediated glycosylation method (glyco-SeS) for the synthesis of disulfide-linked glycoproteins.[108] In the "glyco-SeS" method either the thiol-containing protein or sugar is preactivated using PhSeBr to provide a readily available leaving group (Scheme 8). Using the "glyco-SeS" method, the first example of multi site-selective glycoconjugation, coupling of a heptasaccharide and enzymatic elongation on a disulfide-linked glycoprotein, were

demonstrated.[108] By means of a desulfurisation reaction, disulfide-linked glycoproteins can further be converted into their thioether-linked counterparts (Scheme 8).[109]




**Scheme 7.** Site-selective protein modification using glyco-MTS and -PTS reagents.




**Scheme 8.** Site-selective protein modification using glyco-SeS reagents.

The use of a mutagenesis-modification approach ("tag-modify") has also been applied to an IgG1 Fc fragment and erythropoietin (EPO). Mutagenesis of natural Asn glycosylation sites to Cys provided a "tag", which could be modified through either oxidative coupling with glycosyl thiols (IgG1 Fc fragment) (Scheme 4E)[110] or by reaction with a glycosyl iodoacetamide (EPO) (Scheme 4A).[37] Similarly, glycosyl haloacetamides have been employed for the site-selective glycosylation of DHFR[94] and an *endo* xylanase from *Bacillus circulans*.[111]

### 6.3.2. Site-selective glycosylation of unnatural amino acids

#### 6.3.2.1. Unnatural amino acid incorporation

A number of techniques have been developed that enable the site-specific incorporation of unnatural amino acids (UAAs) into proteins, the two most common being (i) the use of a nonsense or rare codon,[112-119] and (ii) the reassignment of a sense codon.[120-127] Less often, larger (quadruplet) codons are employed for this purpose.[117, 128] Importantly, these methods enable the creation of proteins with unique handles that can be modified in an orthogonal manner to all other functional groups characteristically found in proteins (see pg. 34).

Approach (i) utilises a chemically acylated suppressor transfer RNA to insert the UAA in response to a stop codon, most frequently the amber stop codon (UAG), substituted for the codon encoding the target residue (Figure 4).[112] This method is attractive because it enables the site-selective introduction of an UAA into a protein, however, there are drawbacks. For example, protein-chain termination can occur instead of elongation during translation if suppression efficiency is low.[129] To overcome this problem, strategies to improve suppression efficiency have been developed.[130] In addition, this method requires the prior preparation of a suppressor

tRNA charged with the UAA,[131] or introduction of a suppressor tRNA/aminoacyl-tRNA synthetase pair that is orthogonal to the host system's translation machinery, into the host organism.[117, 132]

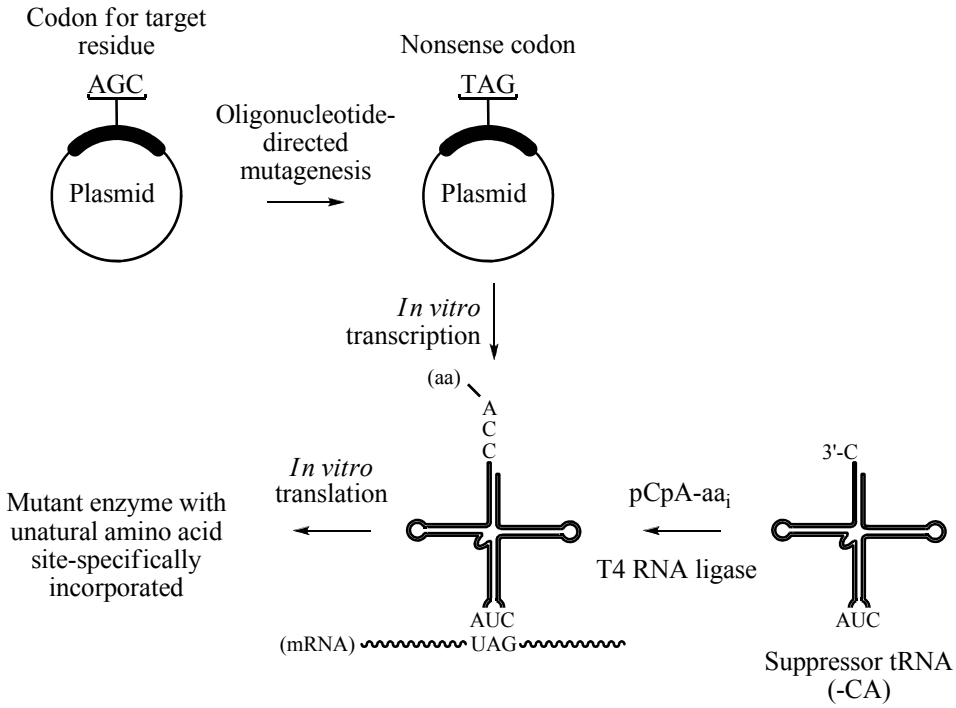



**Figure 4.** Unnatural amino acid incorporation into proteins using a nonsense codon.[112]

Approach (ii) exploits the promiscuity of wild-type aminoacyl tRNA synthetases, which can activate and charge substrates other than proteinogenic amino acids.[133] This method takes advantage of auxotrophy, the inability of an organism to synthesise a particular compound required for its growth.[134] For example, a methionine (Met) auxotrophic *E. coli* strain, unable to synthesise its own Met, has been used for the site-specific incorporation of several different Met analogues (Xaa) into proteins.[135-137] By growing the cultures in Met-containing medium prior to protein expression, and then performing a medium shift[120] into Met-deficient Xaa-containing medium, proteins incorporating the Met analogues have been obtained. Typically this method provides

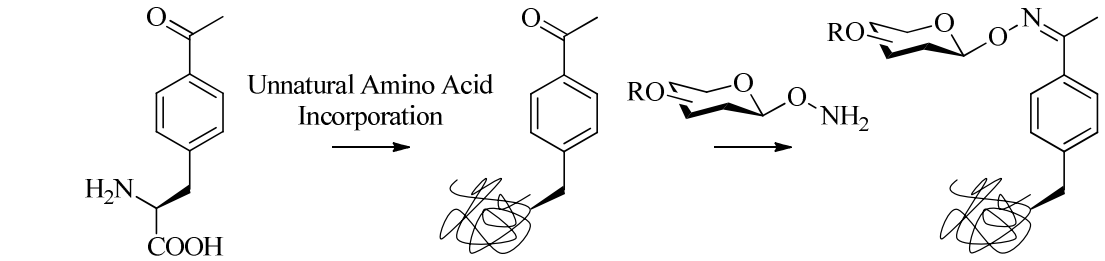
relatively large quantities of protein. Bertozzi *et al.* have reported that cultures supplemented with the Met analogue azidohomoalanine produced protein yields nearly identical to those observed from culture supplemented with methionine.[123] Global replacement of the target amino acid by its respective amino acid analogue is one disadvantage of this method. To avoid this, DNA sequences are routinely engineered by site-directed mutagenesis in order to create proteins that display the UAA at a single site. The position of the target codon can easily be controlled by the codon coding for the target amino acid, *e.g.* the "Met" triplet codon ATG in the above example. However, reassignment of the target amino acid's codon prevents the use of proteins in which this amino acid is essential for protein function. Due to the fact that only a small number of amino acid analogues exhibit translational activity *in vivo*, this method is also limited by the range of chemical functionality that can be introduced into proteins using it. In order to overcome this limitation, further analogues are being evaluated and methods that allow the incorporation of less similar analogues are being developed.[120, 138, 139]

Methionine is an attractive target for residue-specific replacement for a number of reasons. Firstly, methionine is a relatively rare amino acid, with only tryptophan and cysteine occurring less frequently in nature.[138] Methionine is also one of the most hydrophobic residues and is therefore predominantly buried inside proteins. By exploiting the limited availability of this residue, van Hest *et al.* successfully functionalised the only surface exposed methionine residue in the enzyme *Candida antarctica* lipase B.[140] Secondly, replacement of unneeded methionine residues with the near-isosteric amino acid isoleucine using site-directed mutagenesis allows for minimal impact on protein structure and function. Thirdly, incorporation of a large

number of methionine analogues into proteins has already been shown, many of which were amenable to further modification.[135, 136]

### 6.3.2.2. Site-selective glycosylation

The nonsense suppression method followed by chemical modification has been used as a strategy for site-selective protein glycosylation. In the first step of this strategy, an orthogonal *Methanococcus jannaschii* suppressor tRNA/aminoacyl-tRNA synthetase pair was used to site-specifically insert *p*-acetyl-L-phenylalanine, a ketone-containing unnatural amino acid, into *E. coli* proteins in response to the amber nonsense codon, TAG (Scheme 9).[36, 141, 142] Crucially, the suppressor tRNA/aminoacyl-tRNA synthetase pair was developed to avoid cross-reactivity with endogenous *E. coli* tRNAs and aminoacyl-tRNA synthestases. A chemoselective reaction involving an aminooxy-functionalised carbohydrate was then used to derivatise the carbonyl tagged protein (Scheme 9).[98] The resulting glycoprotein was then elaborated further using glycosyltransferases.

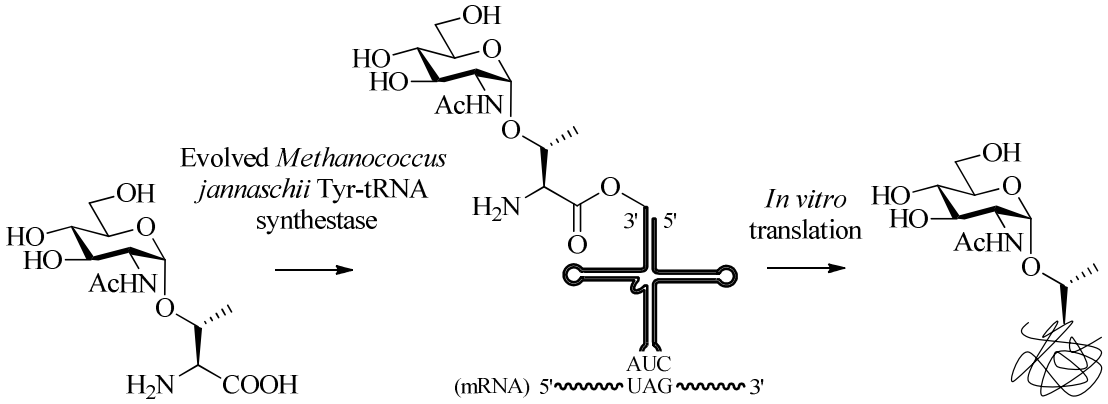



*p*-acetyl-L-phenylalanine

**Scheme 9.** Site-selective protein glycosylation.

Orthogonal *Methanococcus jannaschii* suppressor tRNA/aminoacyl-tRNA synthetase pairs have also been developed that are able to site-specifically insert the glycosyl amino acids GlcNAc-β-Ser and GalNAc-α-Thr (Scheme 10) into *E. coli* proteins in

response to the amber nonsense codon, TAG.[143, 144] (Note: These articles have been retracted) Glycosylated serine and tyrosine derivatives have also been incorporated into proteins using an *in vitro* protein biosynthesising system.[145]




**Scheme 10.** Site-selective protein glycosylation.[144]

Methionine auxotrophic *E. coli* strains have been employed for the efficient incorporation of the methionine analogues azidohomoalanine (Aha),[123] homoallylglycine (Hag) and homopropargylglycine (Hpg) into proteins (Figure 5).[120, 122] A selection of the methods that have been developed to enable selective modification of these chemical "tags" in the presence of the functional groups naturally found in proteins are described below.

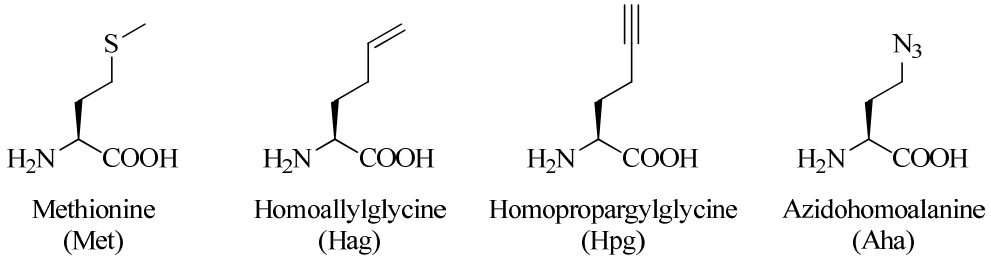
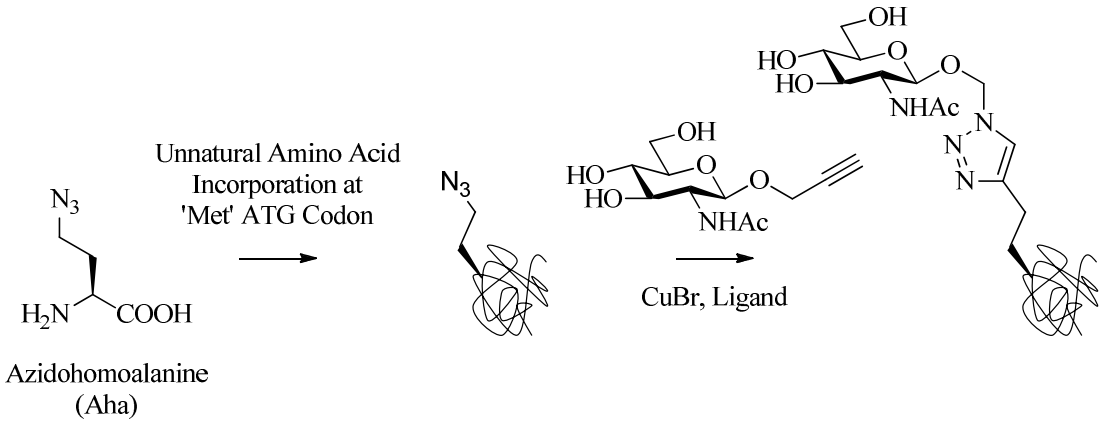


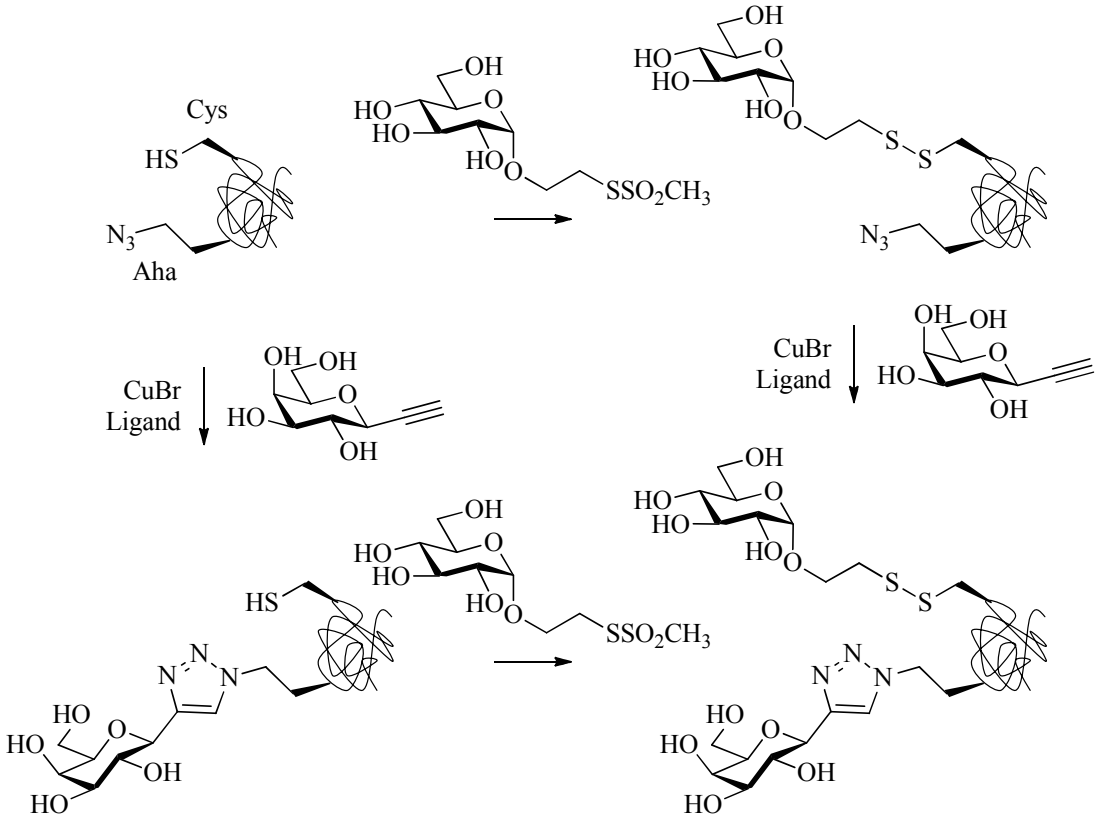

**Figure 5.** Methionine and methionine analogues.

Proteins equipped with the azide-containing unnatural amino acid, Aha, have been selectively glycosylated using the copper-catalysed Huisgen [3+2] cycloaddition (CCHC) reaction,[146, 147] in order to afford triazole linked glycoconjugates (Scheme 11).[135, 136, 148] In addition, alkyne-containing proteins have been used as substrates for "inverse" CCHC reactions with azido-derivatised sugars. [135, 136, 149]



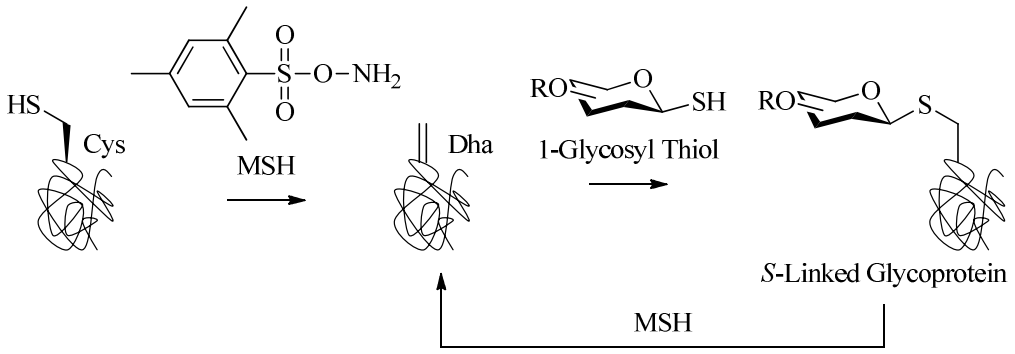

**Scheme 11.** Site-selective protein glycosylation of azido-containing proteins using the copper-catalysed Huisgen cycloaddition reaction.[135, 136]

Recently, van Kasteren *et al.* reported a dual "tag-modify" strategy, which enabled the creation of protein probes able to detect mammalian brain inflammation and disease.[135, 136] In the first step of this strategy, site-directed mutagenesis and unnatural amino acid incorporation were used to generate proteins containing (i) a natural cysteine residue 'thiol tag' and (ii) either an unnatural alkyne (Hpg) or azide (Aha) tag. The protein was then sequentially functionalised using a glyco-MTS reagent to modify the cysteine residue and then a glyco-CCHC reagent (azido- or alkyne-derivatised sugar) to modify either the Hpg or Aha residue, respectively, or *visa versa* (Scheme 12).



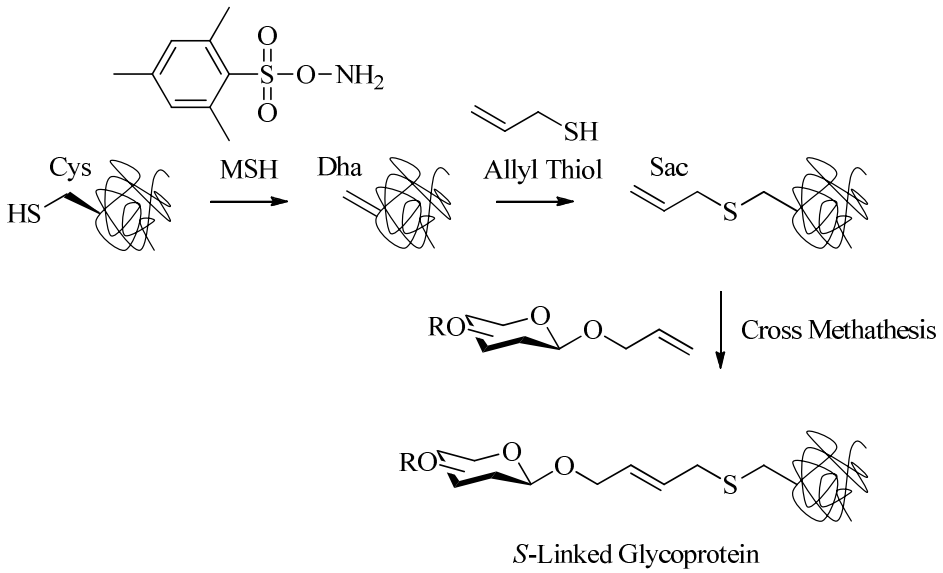

**Scheme 12.** Dual "tag-modify" strategy.[135, 136]

Site-selective protein modification has also been accomplished by employing the unnatural amino acid dehydroalanine (Dha) as a chemical tag. Dha has been incorporated into proteins by a number of methods,[150, 151] the most recent being the direct formation of Dha from Cys using *O*-mesitylenesulfonylhydroxylamine (MSH).[152] From Dha-containing proteins, *S*-linked glycoproteins can be accessed by the conjugate addition of thiol nucleophiles such as 1-glycosyl thiols to Dha (Scheme 13).[151, 152] If required, the Dha-containing protein can be regenerated from the *S*-linked glycoprotein by employing the same method used for eliminating Cys to Dha (Scheme 13).




**Scheme 13.** Site-selective protein glycosylation of Dha-containing proteins using 1-glycosyl thiols.

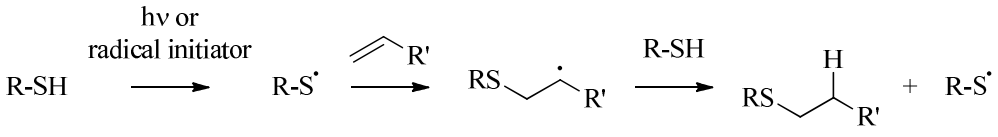
A number of methods exist for the conversion of cysteine to the unnatural amino acid *S*-allyl cysteine (Sac) on proteins. One such method involves the nucleophilic addition of allyl thiol to Dha (Scheme 14). Recently, Davis and co-workers have demonstrated that Sac-containing proteins can be selectively glycosylated using the olefin cross metathesis reaction (Scheme 14).[152, 153]




**Scheme 14.** Site-selective protein glycosylation using the olefin cross metathesis reaction.

## 6.4. Thiol-ene reaction for protein modification

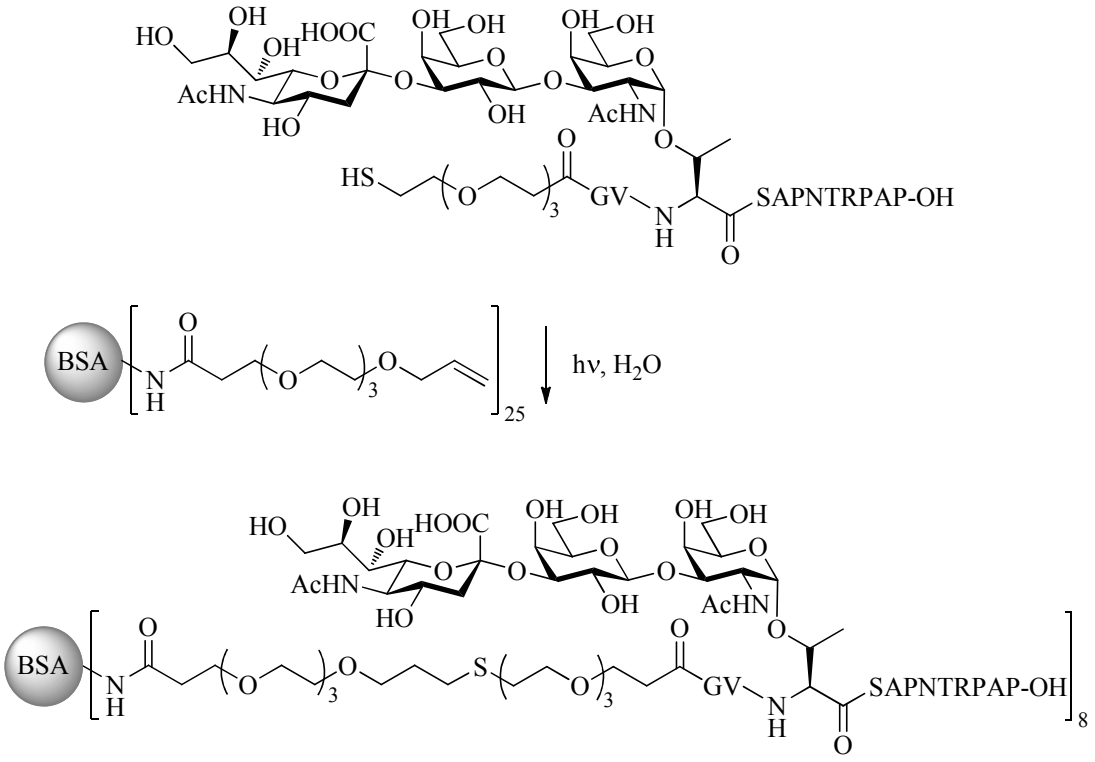
The thiol-ene reaction was discovered in 1905 by Posner,[154] and has since been used extensively in polymer and materials chemistry.[155, 156] The thiol-ene reaction involves the addition of a thiol to an alkene to form a thioether, and depending on the reaction conditions and unsaturated substrate used can proceed by either a free radical or ionic mechanism.[157] The free-radical reaction, with gives anti-Markovnikov products, is a typical chain reaction and can be initiated by the addition of a radical initiator such as a peroxide or azo compound, or by irradiation with UV light (Scheme 15). The initiation step involves the formation of a thiyl radical by hydrogen atom abstraction from the corresponding thiol. The thiyl radical then adds to the alkene to give a new carbon-centered radical, which can abstract a hydrogen atom from a thiol group to produce the final product and a new thiyl radical, thus propagating the radical chain.[158]




**Scheme 15.** The thiol-ene radical reaction.

As a reaction, the thiol-ene reaction has many positive attributes, for example it is compatible with water and oxygen, does not require the use of expensive reagents (*e.g.* transition metal catalysts), and is orthogonal to a wide range of functional groups. In addition, the reaction typically affords products in high yields and high purity.[156, 157, 159, 160] The main drawback of the addition of thiyl radicals to olefins is the strong reversibility of this step.[161]

In recent years the thiol-ene reaction has begun to attract more attention in bioorganic chemistry,[162-164] and has even been successfully applied to proteins.[161, 165] For example, in 2007, Kunz and co-workers described the application of the radical mediated thiol-ene reaction for the synthesis of vaccines. The vaccines were prepared by coupling olefin-modified BSA to thiol-terminated tumor-associated glycopeptide antigens, or *visa versa* (thiol-modified BSA to alkene-terminated tumor-associated glycopeptide antigens), in water by irradiation, one example of which is shown in Scheme 16.[166] Additionally, Niemeyer, Waldman and co-workers detailed in 2008 a photochemical method for the site-specific immobilisation and patterning of proteins that employed thiol-ene chemistry.[167]




**Scheme 16.** Conjugation of a thiol-terminated glycopeptide antigen to olefin-modified BSA.[166]

## 7. The use of glycoproteins as therapeutic agents

Increasingly, glycoproteins are being developed as therapeutic agents,[13] in particular as vaccines,[168] which in part is due to improved methods for carbohydrate synthesis and protein glycosylation. Although a number of carbohydrate-protein conjugate vaccines against several diseases already exist, great effort is being devoted to developing effective vaccines for the many diseases they are not currently available (*e.g.* HIV, malaria, tuberculosis and cancer).

The specificity of the elicited immune response to a glycoprotein vaccine is influenced by the choice of carrier protein, conjugation method, the nature of the linker, carbohydrate loading onto the protein, and immuno adjuvant.[169] As a result, attention is currently focused on developing vaccines devoid of any unnecessary immunogenic components. For example, Boons *et al.* found that a strong immunoglobulin G (IgG) antibody response was elicited against the maleimide linker region of a KLH-MI-Le(y) antigen [MI = 4-(maleimidomethyl)cyclohexane-1-carboxylate], and as a result were forced to use a smaller and more flexible 3-(bromoacetamido)propionate linker for the attachment of Le(y) to KLH in order to reduce the IgG antibody response against the linker and improve the immune response against the Le(y) antigen.[170]

### 7.1. Glycoprotein vaccines

#### 7.1.1. Anti-bacterial, -parasitic and -fungal vaccines

Bacterial polysaccharide-protein conjugate vaccines can be used to prevent infection caused by polysaccharide-encapsulated bacteria such as *Haemophilus influenzae* type b (Hib), *Streptococcus pneumoniae* (pneumococcus) and *Neisseria meningitides*

(meningo-coccus).[171] Highly effective glycoconjugate vaccines against *Haemophilus influenzae* type b for example have been available for almost 20 years, and in the countries in which Hib vaccines are in routine use, influenza type B has virtually been eliminated.[172] Existing Hib vaccines comprise of the capsular polysaccharide of Hib (Figure 6), or fragment of the capsular polysaccharide, linked to a protein carrier.[173] In 2004, Verez-Bencomo *et al.* reported a method for the large-scale synthesis of Hib vaccine comprising a fully synthetic capsular polysaccharide antigen of HiB coupled to tetanus toxoid (Figure 6).[174] Notably, the vaccine was found to be as safe and immunogenic in humans as vaccines containing the native polysaccharide.[174] This has set a precedent for the development of effective synthetic carbohydrate-based vaccines against other diseases.

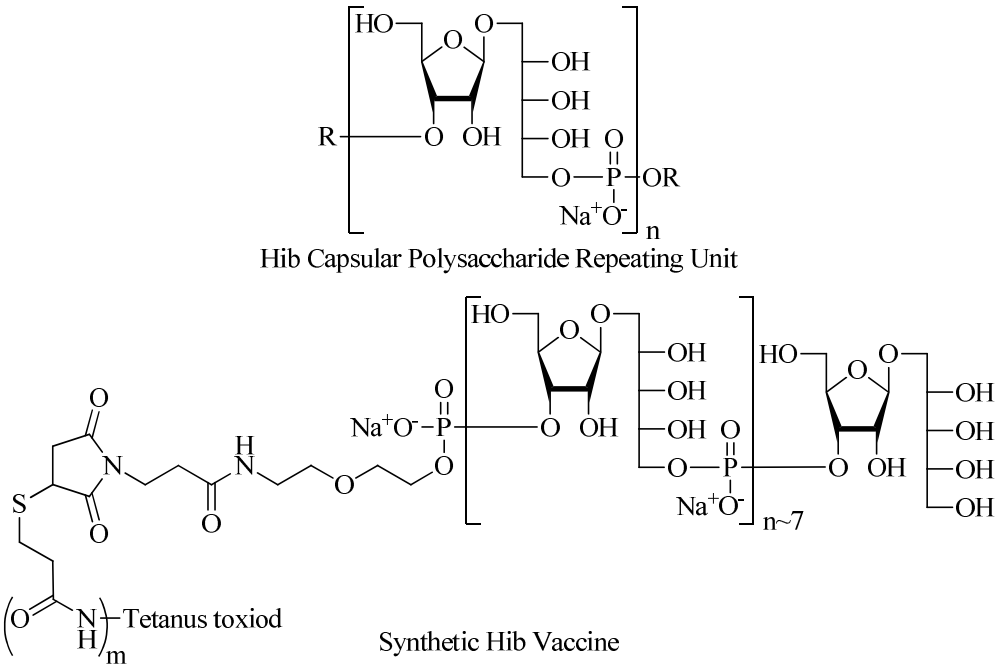



**Figure 6.** Structure of the Hib capsular polysaccharide repeating unit and synthetic Hib vaccine.[174]

Tuberculosis (TB), a bacterial infection caused by *Mycobacterium tuberculosis*, is normally treated using antibiotics, however, the emergence of drug resistant strains of

this bacteria has contributed to the urgent need for TB vaccine development.[175] Currently, protection against TB can be provided by the live attenuated bacillus Calmette-Guerin (BCG) vaccine. However, the protective efficacy of this vaccine is not consistent and ranges from 0 to 80%.[176] Approaches to construct a synthetic vaccine against TB have concentrated on using the lipoglycan components of the mycobacterial cell wall (*e.g.* lipoarabinomannan [LAM], lipomannan [LM], phosphatidylinositol mannoside [PIM]), or fragments of, as antigens.[168, 177] For example, Hamasur *et al.* described the synthesis of carbohydrate-protein conjugate vaccine candidates against TB that comprised an arabinomannan portion of LAM covalently linked to a number of carrier proteins such as tetenus toxoid.[178] The resulting conjugates showed a protective efficiency in guinea pigs and mice comparible to that obtained with the traditional BCG vaccine.[179]

The development of carbohydrate-based vaccines against other bacterial and also parasitic and fungal diseases have also been reported in the literature, including for malaria,[84, 180] shigellosis,[181] anthrax infection,[182] leishmaniasis[183] and candidiasis.

### 7.1.2. Anti-HIV vaccines

An estimated 30 million people worldwide are infected with human immunodeficiency virus (HIV),[184] the virus that causes acquired immune deficiency syndrome (AIDS). Despite decades of research there remains no preventive treatment or cure for this disease.[185] As mentioned previously (pg. 3), the HIV envelope glycoprotein gp120 which protrudes from the surface of the HIV virion, facilitates virus entry into target host cells by binding to CD4 receptors. Because of its involvement in the initial stages of HIV infection, gp120 is a key target for vaccine development. Of the small number of broadly neutralising antibodies to HIV that have

been identified,[186] the human monoclonal antibody 2G12 is known to recognise and bind to a cluster of oligomannose glycans on the surface of gp120 (Figure 7).[187] Since the discovery of 2G12, a great deal of effort has gone into developing mimics of the gp120 epitope recognised by 2G12.[188-194] However, of the glycoconjugates synthesised to date, none have successfully elicited broadly neutralising antibodies against HIV. For example, Danishefsky and co-workers recently reported the synthesis of a multivalent glycopeptide construct[193] that imitated gp120's binding to 2G12. When this peptide was conjugated to a protein carrier, the resulting glycoconjugate vaccine induced carbohydrate-specific antibodies that failed to neutralise HIV virions.[195]

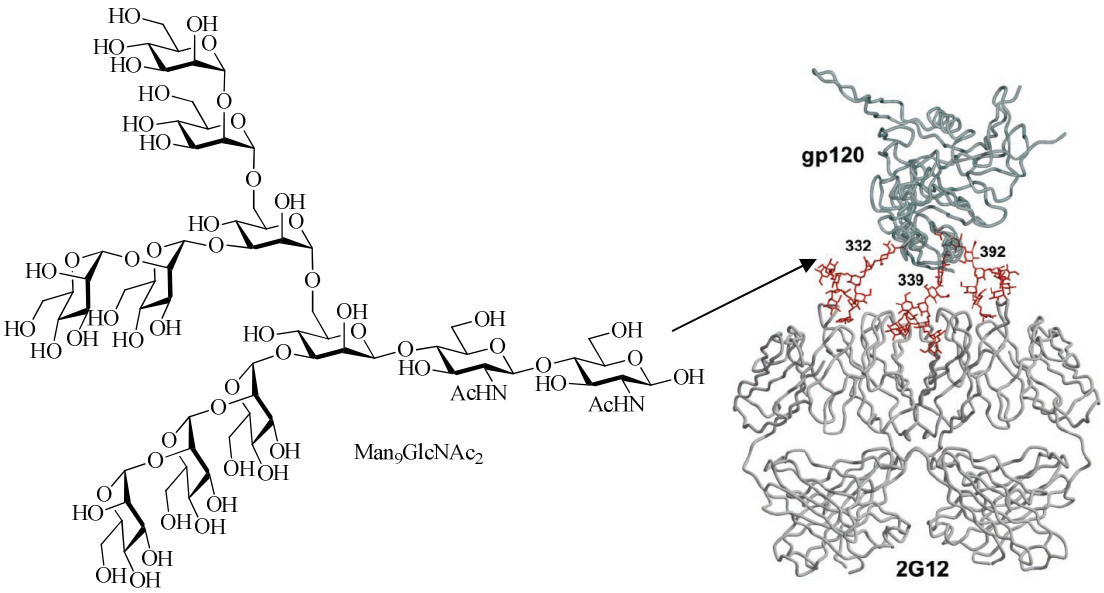



**Figure 7.** Model of 2G12 glycan recognition of gp120. From *Science*, 2003, 300, 2065-71. Reprinted with permission from AAAS.

### 7.1.3. Anti-cancer vaccines

The surface glycosylation pattern displayed by tumor cells often differs from that of normal cells.[196] For example, both the under- and over-expression of naturally

occuring glycans can occur, as well as neo-expression of glycans normally restricted to embryonic tissues. Common changes include an increase in the size and branching of *N*-linked glycans,[197] an increase in sialylation and an increase in certain terminal glycans. Globo H, sialyl Lewis x ($sLe^x$), polysialic acid (PSA) and sialyl Tn (sTn) (Figure 8) have all been identified, as well as others, as terminal glycan epitopes regularly displayed on the surface of tumor cells, and are termed tumor-associated carbohydrate antigens (TACAs).[198]

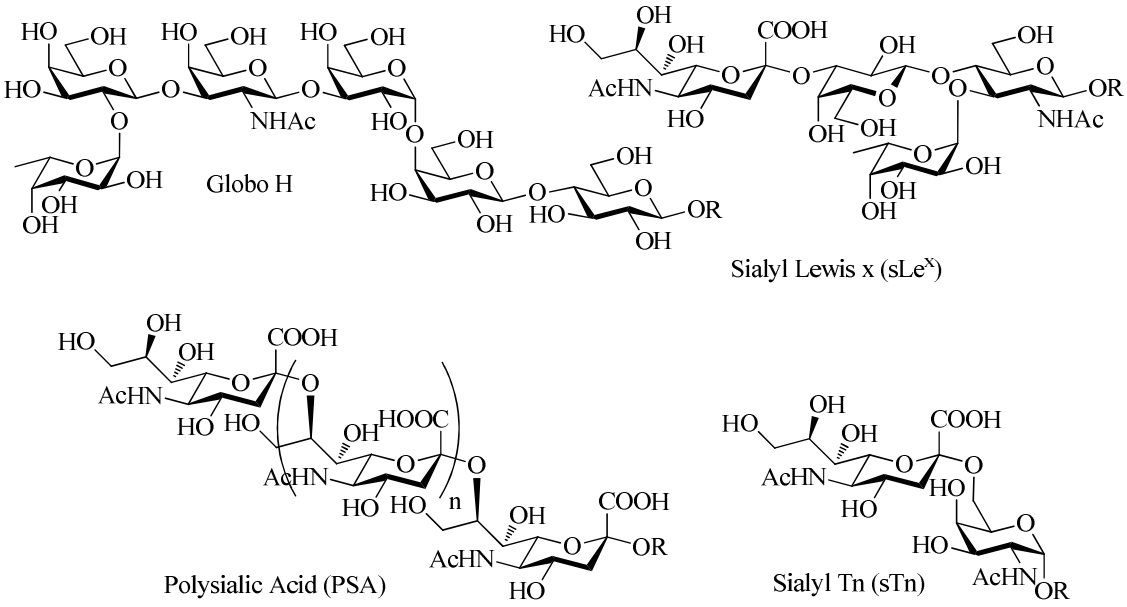



**Figure 8.** Tumor-Associated Carbohydrate Antigens.

TACAs have been employed for the development of anti-cancer vaccines.[168, 169] However, due to the fact that TACAs are "self-antigens" and therefore tolerated by the immune system, immunisation against these cancer antigens is difficult and as a result many of the vaccines generated to date have been poorly immunogenic. Self-antigens can be described as "a substance that acts as an antigen and induces antibody formation in another organism but to which the immune system of the parent organism is tolerent".

Of the numerous reports in the literature describing the preparation of carbohydrate-based anti-cancer vaccine candidates, a significant number are from Danishefsky and co-workers.[199-202] Whilst their early work focused on the synthesis of monovalent vaccine constructs comprising a single carbohydrate antigen (*e.g.* Globo H) linked to a carrier (*e.g.* KLH),[203] later work has centred on the preparation of more elaborate constructs, in which multiple repeats or clusters of a carbohydrate epitope are presented on a peptide backbone. For example, Danishefsky and co-workers recently synthesised a vaccine for ovarian cancer based on $Gb_3$, a carbohydrate known to be overexpressed on the surface of ovarian tumor cells (Figure 9).[204]

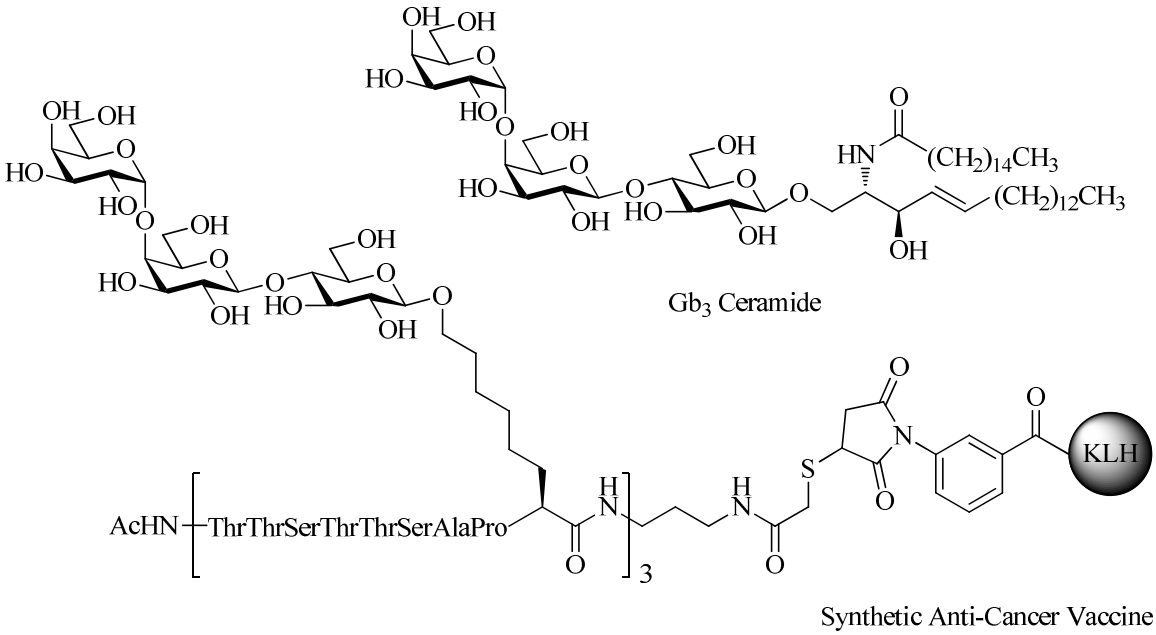
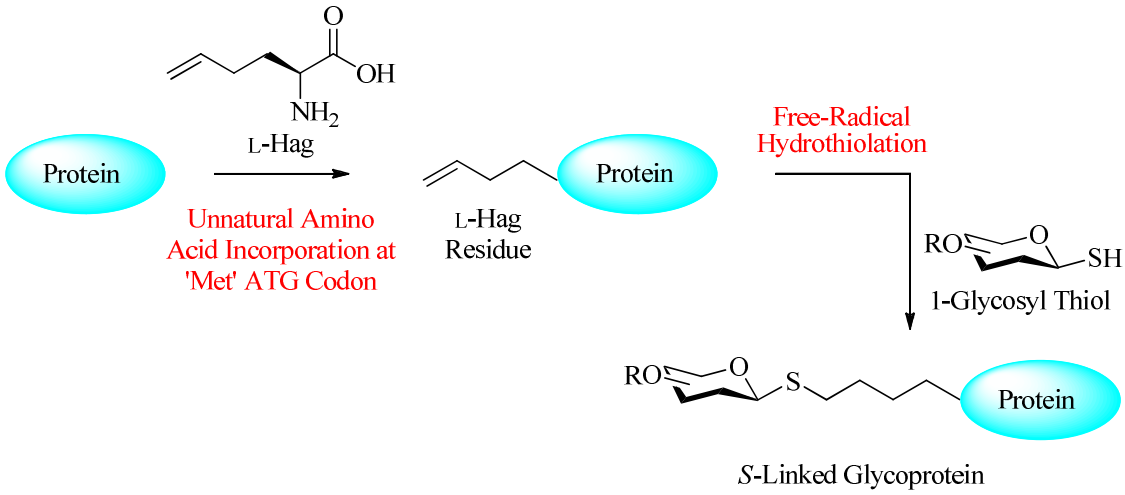



**Figure 9.** Synthetic carbohydrate-based vaccine candidate for ovarian cancer.

# Chapter 2: Thiyl-Glycosylation of Olefinic Proteins - *S*-Linked Glycoprotein Synthesis

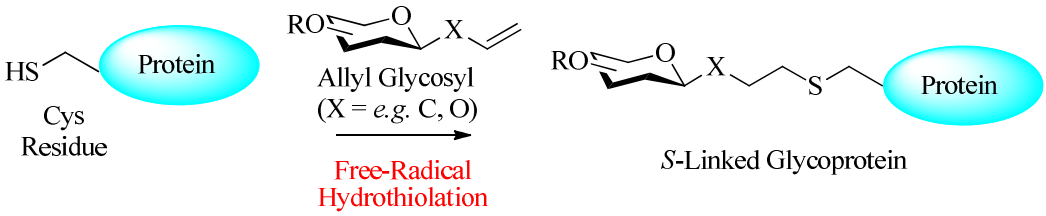
## 1. Introduction

This chapter describes the development of a convergent approach for the synthesis of a novel class of *S*-linked glycoproteins through the site-specific ligation of 1-glycosyl thiols to olefinic proteins (Scheme 17). The strategy exploits unnatural amino acid incorporation[120, 205] for the introduction of L-homoallylglycine (L-Hag) into a protein and free-radical addition hydrothiolation chemistry, under conditions mild enough to preserve protein activity. In this way, Hag functions as a new "tag" combined here with a new modification as part of a general "tag-modify" strategy for synthetic protein construction.[59] This "tag-modify" strategy has successfully been demonstrated previously using azide and alkyne tags for example (see pg. 35).[135, 136]




**Scheme 17.** Summarised strategic approach.

Prior, elegant conjugations have shown that radical addition chemistries may be successfully applied to proteins (see pg. 39).[161, 165, 166] Whilst 1-thioglycoside formation by the free-radical addition of 1-glycosyl thiols to alkenes has been reported for the synthesis of small molecules,[206-208] to date this method has not been applied to the synthesis of *S*-linked glycoproteins or bioconjugates. The unique reactivity profile of L-Hag, which has an olefinic side-chain, compared with the natural amino acids characteristically found in proteins, allows for a chemoselective chemical reaction. These selectivity advantages include the fact that Hag is inert to almost all common protein modification chemistries (see pg. 22). This allows for potential orthogonal use in combined, multi-reaction protein synthesis strategies.[135, 209] This inertness to reagents is not shared by strategies that use a protein-thiyl radical (*e.g.* derived from cysteine), although this "reverse" approach does allow for use of proteins containing only natural amino acids. In this context, during the final stages of this work,[210] it became apparent that useful complementary methods were being employed by Dondoni, Massi and co-workers for glycosylation of protein-thiyl radicals (Scheme 18).[211]



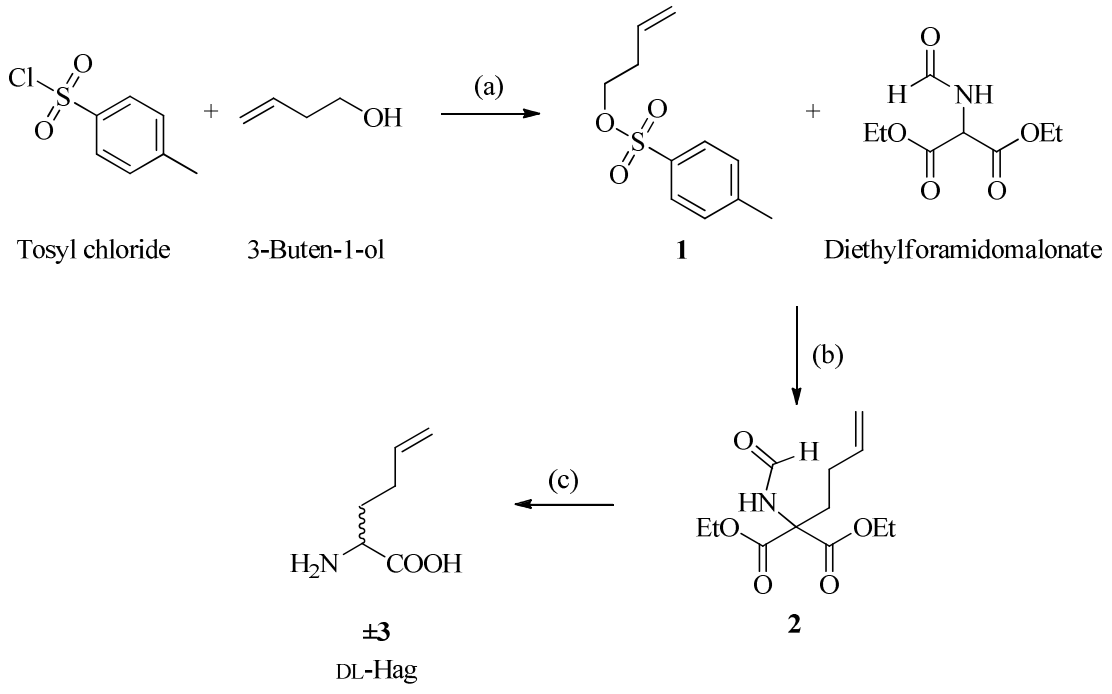

**Scheme 18.** "Reverse" approach to *S*-linked glycoprotein synthesis.

## 2. Homoallylglycine (Hag) synthesis

L- and DL-homoallylglycine were synthesised for use in model system studies and unnatural amino acid incorporation into proteins.

### 2.1. DL-Hag synthesis

DL-Hag[212-214] was synthesised in accordance with the procedure of Tirrel and co-workers.[120] Tosylate **1** was formed from a reaction involving the commercially available starting materials tosyl chloride and 3-buten-1-ol, and was used to alkylate deprotonated diethylformamidomalonate.[215] The reaction yield of 29% was similar to that previously reported in the literature.[120] Subsequent hydrolysis of **2**, followed by decarboxylation and amine deprotection afforded DL-Hag (**±3**) (Scheme 19).




**Scheme 19.** (a) Pyridine (96%); (b) (i) NaH (60% dispersion in mineral oil), DMF, (ii) **1**, 70°C (29%); (c) (i) 10% NaOH, reflux, (ii) 1.5 M HCl, reflux (100%).

Problems associated with Tirrel's acetamide strategy in the final deprotection step were circumvented by employing the inherently more labile formyl protecting group.[215]

### 2.2. L-Hag synthesis

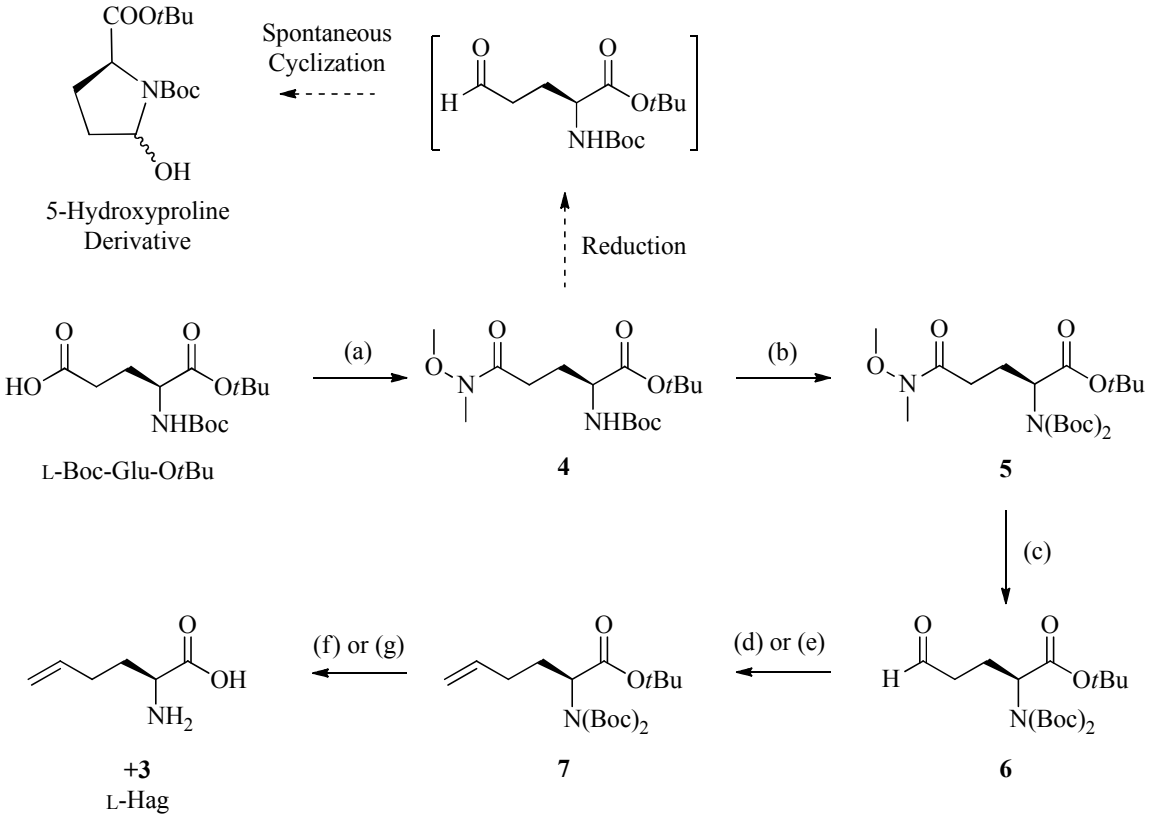
Enantiomerically pure L-Hag[216] was prepared by employing a number of routes. Protection of the Hag Cα-carboxyl group as a *t*-butyl and methyl ester was investigated (see pg. 50 and 52, respectively).

#### 2.2.1. Via *t*-butyl ester route

Commercially available L-Boc-Glu-O*t*Bu was transformed to the corresponding Weinreb amide **4** (Scheme 20).[217] Further protection of the amine group with an additional Boc group to give **5**[218] proceeded slowly (5 days) and microwave irradiation failed to enhance the reaction rate.[219] Subsequent reduction with D*i*BAL-H led to the formation of δ-aldehyde **6**.[218] This strategy successfully avoided the formation of the 5-hydroxyproline derivative, reported to result from reduction of **4**.[217] The Wittig reaction of δ-aldehyde **6** with the ylide generated from methyl triphenylphosphonium bromide and *n*-BuLi afforded alkene **7** and resulted in a yield of 6%; this was significantly lower than that reported by Adamczyk *et al*. (66%)[220] but consistent with results of Kokotos *et al*.[221] Furthermore, in accordance with Kokotos *et al*.,[221] $KN(TMS)_2$ was found to improve the reaction yield; 33% in toluene and 36% in THF. The specific optical rotation of **7** was measured to be -17.1 and -4.9 (25 $^{o}$C, c=1, MeOH) [Lit. $[\alpha]_D^{20}$ -16.2 (c=1.48, MeOH)][220] when using toluene and THF, respectively; indicating that THF may have induced epimerisation. Global deprotection of **7** was carried out using TFA/DCM with and without the presence of a

carbocation scavenger,[218, 222] to afford L-Hag (**+3**) in 15 and 77% yield, respectively. Purification of the product from the crude reaction mixture by cation exchange chromatography (DOWEX 50WX8-200, ammonium hydroxide eluent)[223] proved superior to recrystalisation (ethanol/water 2:1 v/v).[120]

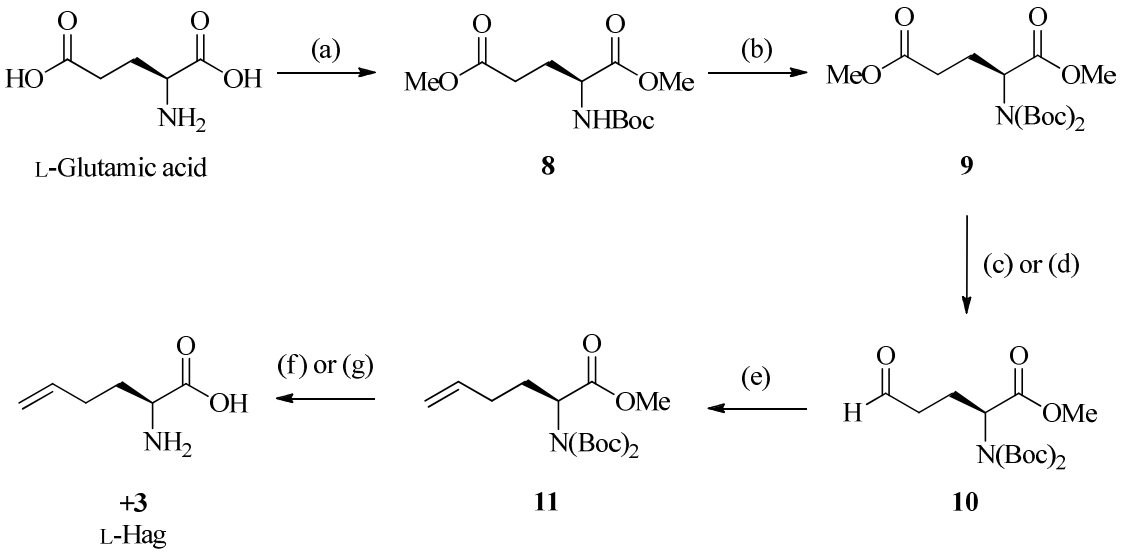
As the starting material L-Boc-Glu-O*t*Bu was expensive and the overall yield of L-Hag (**+3**) obtained from the above route was only 14% over 5 steps, an alternative more economical protocol utilising methyl ester protection of the Hag Cα-carboxyl group was explored.




**Scheme 20.** (a) (i) PyBOP, $Et_3N$, DCM (ii) HNMeOMe, $Et_3N$ (95%); (b) $(Boc)_2O$, $Et_3N$, DMAP, DCM (94%); (c) DiBAL-H (1 M in heptane), THF, -78°C (96%); (d) $CH_3PPh_3Br$, *n*-BuLi, THF, -78°C (6%); (e) $CH_3PPh_3Br$, $KN(TMS)_2$ (0.5 M in toluene), toluene, -78°C (33%); (f) TFA/DCM (1:1), 2.5% triethylsilane, 2.5% $H_2O$ (15%); (g) TFA/DCM (1:1) (77%).

### 2.2.2. Via methyl ester route

A one pot procedure to synthesise **8** from inexpensive L-glutamic acid using TMSCl and methanol in a hydrogen chloride catalysed diesterification step followed by mono-Boc protection of the amine group proved successful (Scheme 21).[221, 224] Prior to this, the use of thionyl chloride in place of TMSCl did not yield any product.[225] Boc-protection of the remaining amine resulted in the formation of **9**.[221] On a large scale (*ca.* 40 g), this reaction proceeded in a sluggish manner and under these circumstances was forced to completion by heating to 40°C and/or adding more equivalents of di-*t*-butyl dicarbonate ($Boc_2O$).




**Scheme 21.** (a) (i) TMSCl, MeOH, 0°C (ii) $(Boc)_2O$, $Et_3N$, MeOH (80%); (b) $(Boc)_2O$, DMAP, $CH_3CN$ (85%); (c) D*i*BAL-H (1 M in heptane), $Et_2O$, -78°C (52%); (d) D*i*BAL-H (1 M in hexane), $Et_2O$ -78°C (88%); (e) See Table 2; (f) (i) 0.5 M NaOH, MeOH (ii) TFA/DCM (96%); (g) (i) LiOH, $H_2O$/dioxane (ii) TFA/DCM (91%).

The regioselective D*i*BAL-H mediated δ-ester reduction was found to give markedly different outcomes depending on the choice of D*i*BAL-H stock solution solvent. Whilst D*i*BAL-H in heptane gave **10** in a 52% yield, D*i*BAL-H in hexane provided

the aldehyde in a superior yield of 88%.[221] Workup of this reaction was difficult and addition of a saturated solution of Rochelle salt (potassium sodium tartrate) was required to break up the emulsion caused by aluminum complexation.

Considerable effort was invested in optimising the Wittig reaction for the formation of **11** (Table 2). Of the conditions screened, $KN(TMS)_2$ in toluene was established to be the most successful, despite $KN(TMS)_2$ in THF yielding marginally more product (61% *vs.* 63%).[221] This was because the specific optical rotation of **11** formed using THF was measured to be 0.0 (25 $^{o}$C, c=1, $CHCl_3$) [Lit. $[\alpha]_D^{25}$ -35.3 (c=1, $CHCl_3$)],[224] indicating that the solvent may have induced epimerisation. This was also observed for the conversion of aldehyde **6** to alkene **7** (see pg. 50). Using toluene, the specific optical rotation of **11** was measured to be -31.6 (25 $^{o}$C, c=1, $CHCl_3$).

**Table 2.** Wittig reaction conditions.

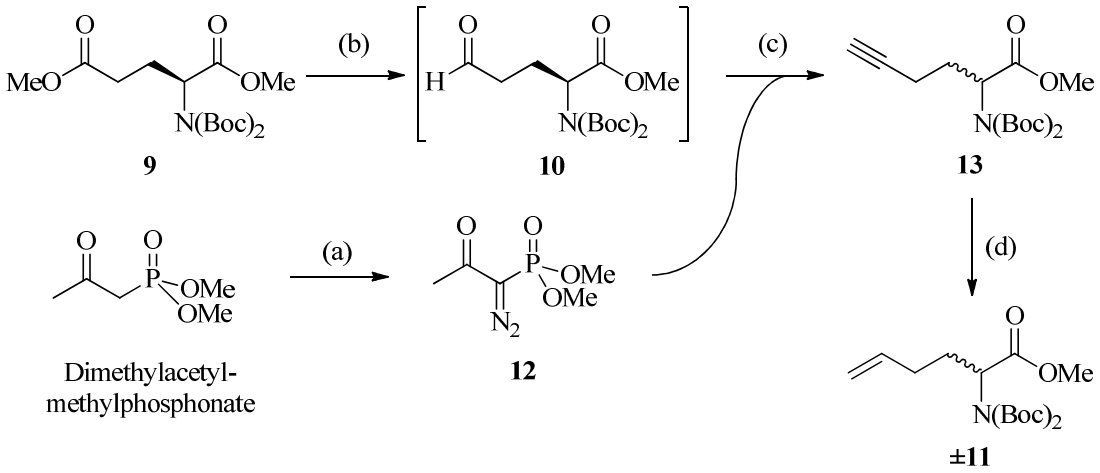
| Entry | Reagent | Solvent | Temp $^{o}$C | Isolated Yield |
|---|---|---|---|---|
| **A** | $CH_3PPh_3Br$, *n*-BuLi | THF | -78 | 0% |
| **B** | $CH_3PPh_3Br$, $KN(TMS)_2$ | THF | -78 | 63% |
| **C** | $CH_3PPh_3Br$, $KN(TMS)_2$ | Toluene | -78 | 61% |
| **D** | $CH_3PPh_3Br$, $KN(TMS)_2$ | Diethylether | -78 | 50% |
| **E** | $CH_3PPh_3Br$, $KN(TMS)_2$ | Acetonitrile | 0 | 31% |
| **F** | Tebbe reagent (0.5 M in toluene) | Toluene | RT | 13% |

A one pot procedure for the synthesis of **11** from **9** without purification of potentially unstable intermediate **10** using D*i*BAL-H (1 M in hexane) followed by reaction conditions C in Table 2, did not result in an overall improved yield for these two steps (18% *cf.* 54%). Similarly, using Tebbe reagent ($Cp_2TiCH_2ClAl(CH_3)_2$),[226, 227] similar yields were obtained irrespective of the method used; D*i*BAL-H (1 M in hexane) followed by reaction conditions F in Table 2 with purification of aldehyde **10 -** 11% *cf.*

without purification of **10** - 16%. To minimise enhancement of ester reactivity to Tebbe reagent, the commonly used Lewis basic solvent THF was avoided and toluene was chosen as the solvent.[226]

The use of HCl was precluded for ester hydrolysis of **11** in view of the fact that it is reported to add to the double bond.[120] Basic conditions may be employed, however, these needed to be sufficiently mild to avoid racemisation. The use of 0.5 M NaOH in methanol, a popular literature procedure for this purpose, resulted in the desired carboxylic acid after 7 days.[228] Subsequent Boc deprotection using TFA/DCM afforded L-Hag (**+3**) in near quantitative yield. The alternative use of LiOH in $H_2O$/dioxane resulted in a similar rate of hydrolysis, whereas no reaction was observed with hydrogen peroxide/LiOH.[229]

Simultaneous to the above route (Scheme 21), an alternative approach for the synthesis of L-Hag was pursued. Ohira's reagent (**12**) was prepared following a literature procedure (Scheme 22);[230-233] reaction of commercial dimethylacetyl-methylphosphonate with sodium hydride in THF/toluene followed by treatment with 4-acetylaminobenzene sulphonyl azide, proved superior to a one step procedure in THF (56% *vs.* 12%). Diazo compound **12** was reacted with crude aldehyde **10** (Scheme 21) to give alkyne **13**. Partial alkyne hydrogenation over Lindlar's catalyst was effective for the formation of alkene **11**.[234] Toluene was chosen as the solvent to reduce the rate of reaction,[235, 236] and typically no over reduction to the alkane was observed. The specific optical rotations of novel compound **13** and also **11** were measured to be 0.0 (25 $^{o}$C, c=1, $CHCl_3$) [-**11**: Lit. $[\alpha]_D^{25}$ -35.3 (c=1, $CHCl_3$)],[224] indicating that undesired racemisation had occurred.




**Scheme 22.** (a) (i) NaH, THF/toluene, 0°C (ii) 4-Acetylaminobenezene sulphonyl azide (56%); (b) D*i*BAL-H (1 M in hexane), $Et_2O$ -78 °C; (c) $K_2CO_3$, MeOH (57% over 2 steps); (d) Lindlar's catalyst, $H_2$, toluene (93%).

## 2.3. Enantiomeric excess determination of L-Hag

Enantiomeric excess determination is key to evaluating the enantiomeric purity of compounds such as L-Hag. A number of methods exist which allow this, including polarimetry,[237] NMR using a chiral shift or solvating agent,[238-240] gas or liquid chromatography using a chiral stationary phase and isotope dilution.[241] Whilst initially the simple method of polarimetry was chosen to assess for racemisation at Cα of L-Hag throughout its synthesis, in practice this technique is limited in terms of accuracy and reproducibility. Measurements can be distorted by small changes in solvent, concentration, wavelength and temperature. Further, it is essential that the sample is analytically pure, since the presence of even small amounts of an impurity can lead to large errors in the specific rotation. For these reasons, attention was turned to alternative techniques to determine the enantiomeric excess of L-Hag.

### 2.3.1. Chiral HPLC

The enantiomeric excess of L-Hag and similar amino acids such as allylglycine have previously been determined by chiral HPLC (*e.g.* using a Crownpak CR(+) column).[216, 242] Here, an available Chiralcel OD column, with cellulose based stationary phase, was chosen. A control separation was run in which DL-Hag (**±3**) was injected and eluted at 0.5 mL/min using an isocratic mobile phase comprising 100% ethanol. Eluent was monitored at $\lambda = 210$ nm, the wavelength of maximum absorption ($\lambda_{max}$) of DL-Hag (**±3**), but no peak(s) was observed in the UV trace. It was reasoned that DL-Hag may have gone undetected owing to its weakly absorbing nonconjugated alkene chromophore, and therefore it was decided to enhance the UV detection properties of DL-Hag using derivatising agents.

### 2.3.2. Derivatisation

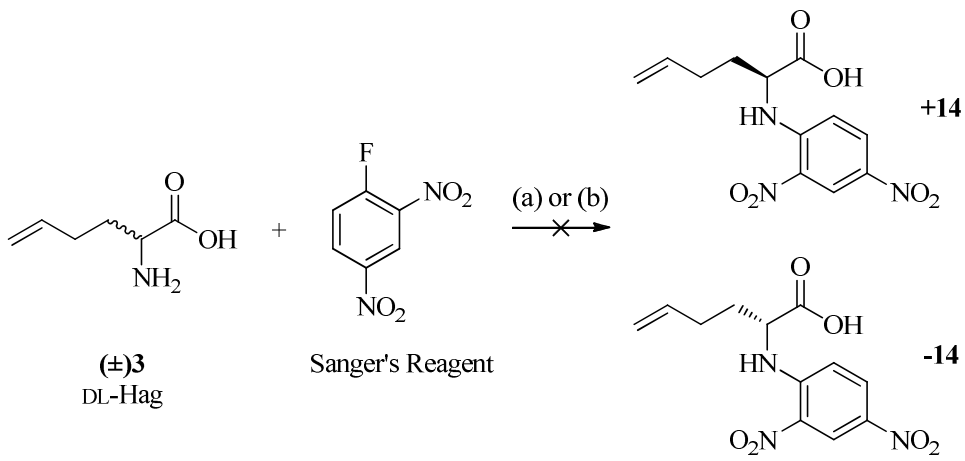
The employment of derivatising agents to determine enantiomeric excess is commonplace.[243] Here, the use of Sanger's reagent,[244] Mosher's acid chloride[245, 246] and the amino acids Fmoc-L-Ala-OPfp and Fmoc-L-Asn(Trt)-OPfp for the derivatisation of Hag are described. Reaction conditions, and chromatographic conditions where applicable, were optimised with DL-Hag (**±3**).

#### 2.3.2.1. Sanger's reagent

To enhance the UV detection properties of DL-Hag (**±3**) for chiral HPLC, it was necessary to derivatise it with a strongly absorbing chromophore. To this end, the achiral derivatising agent Sanger's reagent (1-fluoro-2,4-dinitrobenzene)[244] was chosen (Scheme 23). This compound, which is routinely used for N-terminal

sequencing of polypeptides,[247, 248] reacts readily with the amine group of amino acids to produce dinitrophenyl amino acids.

DL-Hag (**±3**) was reacted stoichiometrically with Sanger's reagent in the presence of sodium hydrogen carbonate first following a conventional procedure[244] (conditions (a) in Scheme 23) and subsequently following a modified procedure that employed the use of microwave irradiation[249] (conditions (b) in Scheme 23). Unfortunately, the reaction did not proceed cleanly or completely using either reaction conditions to yield **±14** and so an alternative derivatisation strategy was employed.



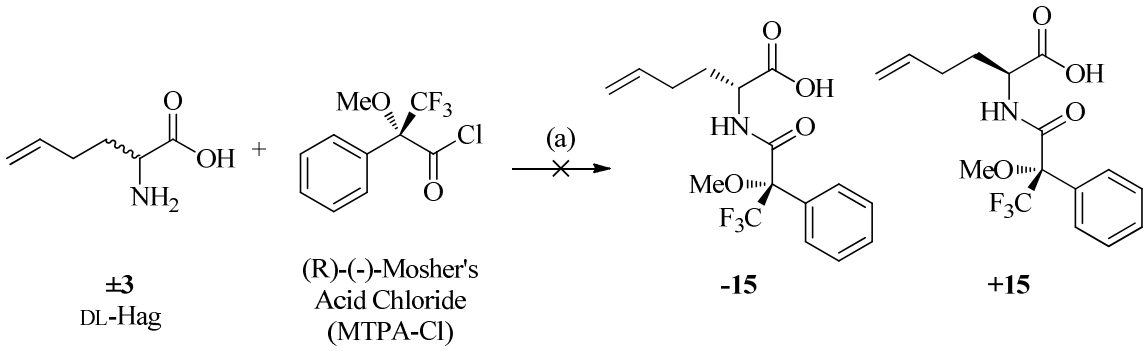

**Scheme 23.** (a) $NaHCO_3$, $EtOH/H_2O$ (1:1), RT, 16 h; (b) $NaHCO_3$, EtOH or $H_2O$, 80°C, microwave, 6 min.

### 2.3.2.2. Mosher's acid chloride

Mosher's acid (α-methoxytrifluorophenylacetic acid, MTPA)[245, 246] is one of the most widely used chiral derivatising agents and its more reactive acid chloride, MTPA-Cl, has been used to acylate vinyl glycine derivatives structurally similar to homoallylglycine.[250] The intention was to determine the enantiomeric purity of the

diastereomeric esters formed by reaction of MTPA-Cl with DL-Hag by NMR spectroscopy.

DL-Hag (**±3**) was reacted stoichiometrically with MTPA-Cl following a literature procedure.[251] Unfortunately, the reaction did not proceed cleanly to afford the desired Mosher's amides and so an alternative derivatisation strategy was employed (Scheme 24).

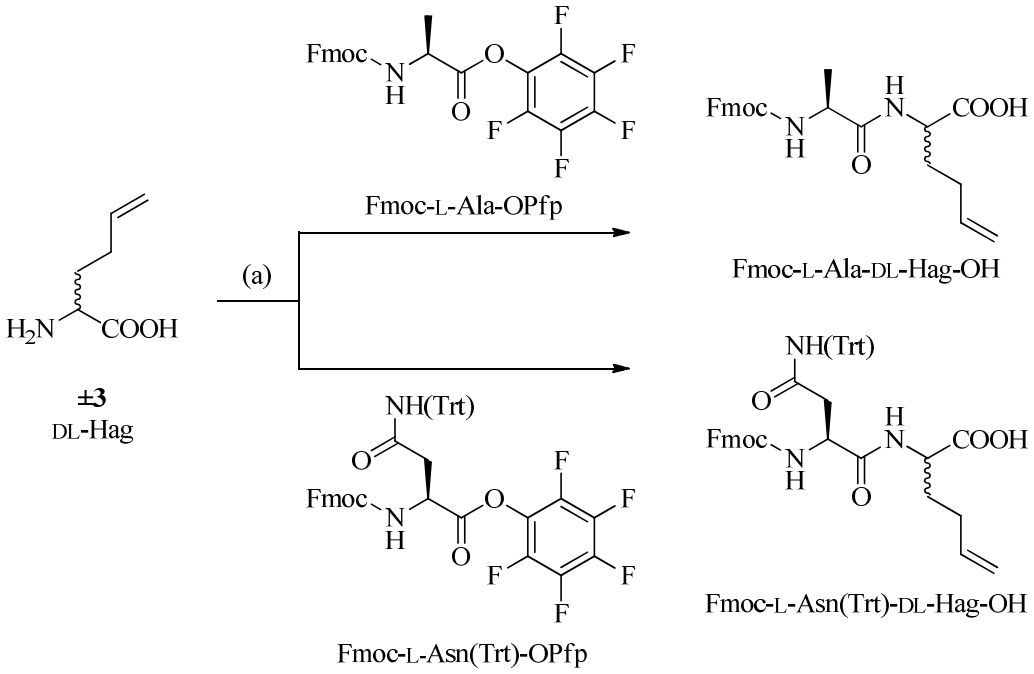



**Scheme 24.** (a) Propylene oxide (4 eq), THF, reflux, 16 h.

### 2.3.2.3. Fmoc-L-Ala-OPfp and Fmoc-L-Asn(Trt)-OPfp

Having not yet established a successful method to determine the enantiomeric excess of L-Hag (**+3**), attention was turned to the synthesis of Hag-containing dipeptides, giving diastereomers which could be resolved using achiral HPLC.

DL-Hag (**±3**) was reacted with the commercially available preactivated enantiopure Fmoc-amino acid-OPfp esters, Fmoc-L-Ala-OPfp and Fmoc-L-Asn(Trt)-OPfp, in the presence of HOBt, following the procedure of Arnold *et al.* (Scheme 25).[252] This simple protocol avoided the need for carbodiimide coupling reagents such as dicyclohexylcarbodiimide (DCC), typically used for peptide bond formation and which can cause racemisation.




**Scheme 25.** (a) HOBt, DMF.

Separation of the two diastereomers of Fmoc-L-Ala-DL-Hag-OH from the crude reaction mixture was attempted by HPLC using a Phenomenex Synergi Hydro-RP column, but was not successful. Various solvent gradient systems were employed (see pg. 269) but irrespective, both Fmoc-L-Ala-D-Hag-OH and Fmoc-L-Ala-L-Hag-OH eluted at the same retention time. This was not completely unsurprising given that the resolution of alanine containing dipeptides is known to be difficult due to the small size of the alanine side-chain.

Separation of the two diastereomers of Fmoc-L-Asn(Trt)-DL-Hag-OH from the crude reaction mixture was attempted by HPLC using a Phenomenex Synergi Hydro-RP column. Various solvent gradient systems were employed (see pg. 269), with the best separation achieved at 1 mL/min using an isocratic mobile phase comprising 55% acetonitrile in water plus 0.1% TFA. Under these conditions, Fmoc-L-Asn-D-Hag-OH and Fmoc-L-Asn-L-Hag-OH eluted at 17.6 and 19.1 minutes, respectively (Figure 10).

The ratio of Fmoc-L-Asn(Trt)-D-Hag-OH to Fmoc-L-Asn(Trt)-L-Hag-OH was determined to be practically equal; 48.9 : 51.1 at 214 nm and 48.5 : 51.5 at 254 nm.

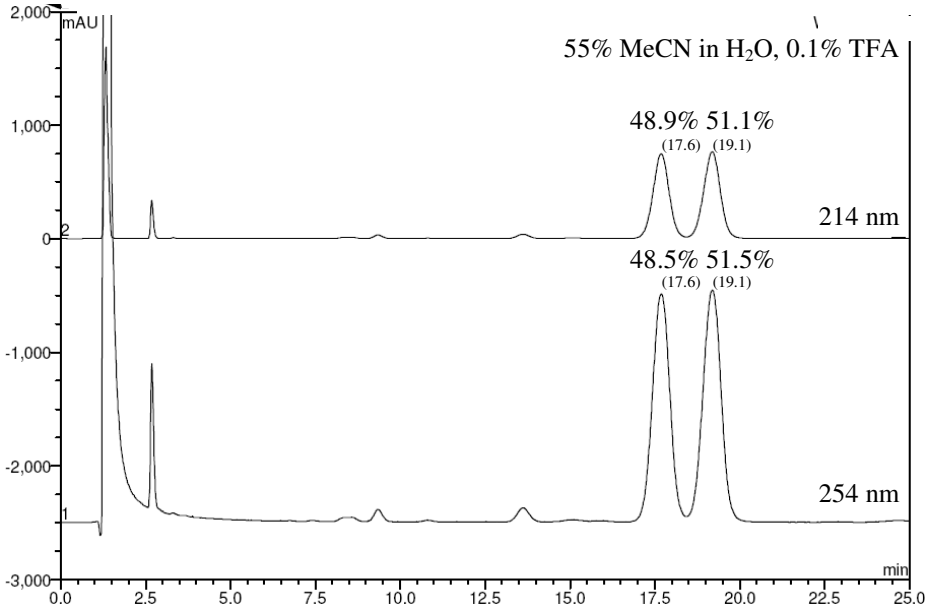



**Figure 10.** UV traces showing HPLC separation of Fmoc-L-Asn(Trt)-DL-Hag-OH.

Having established a robust method to determine the enantiomeric excess of L-Hag, Fmoc-Asn(Trt)-L-Hag-OH was synthesised and analysed by HPLC (Figure 11) in the same way as Fmoc-L-Asn(Trt)-DL-Hag-OH (Scheme 25), except that L-Hag (**+3**) was used in place of DL-Hag (**±3**).[252] The ratio of Fmoc-L-Asn(Trt)-D-Hag-OH to Fmoc-L-Asn(Trt)-L-Hag-OH was found to be 98.0 : 2.0 (214 nm) / 98.7 : 1.3 (254 nm), which enabled the enantiomeric excess of L-Hag to be determined to be > 96%.

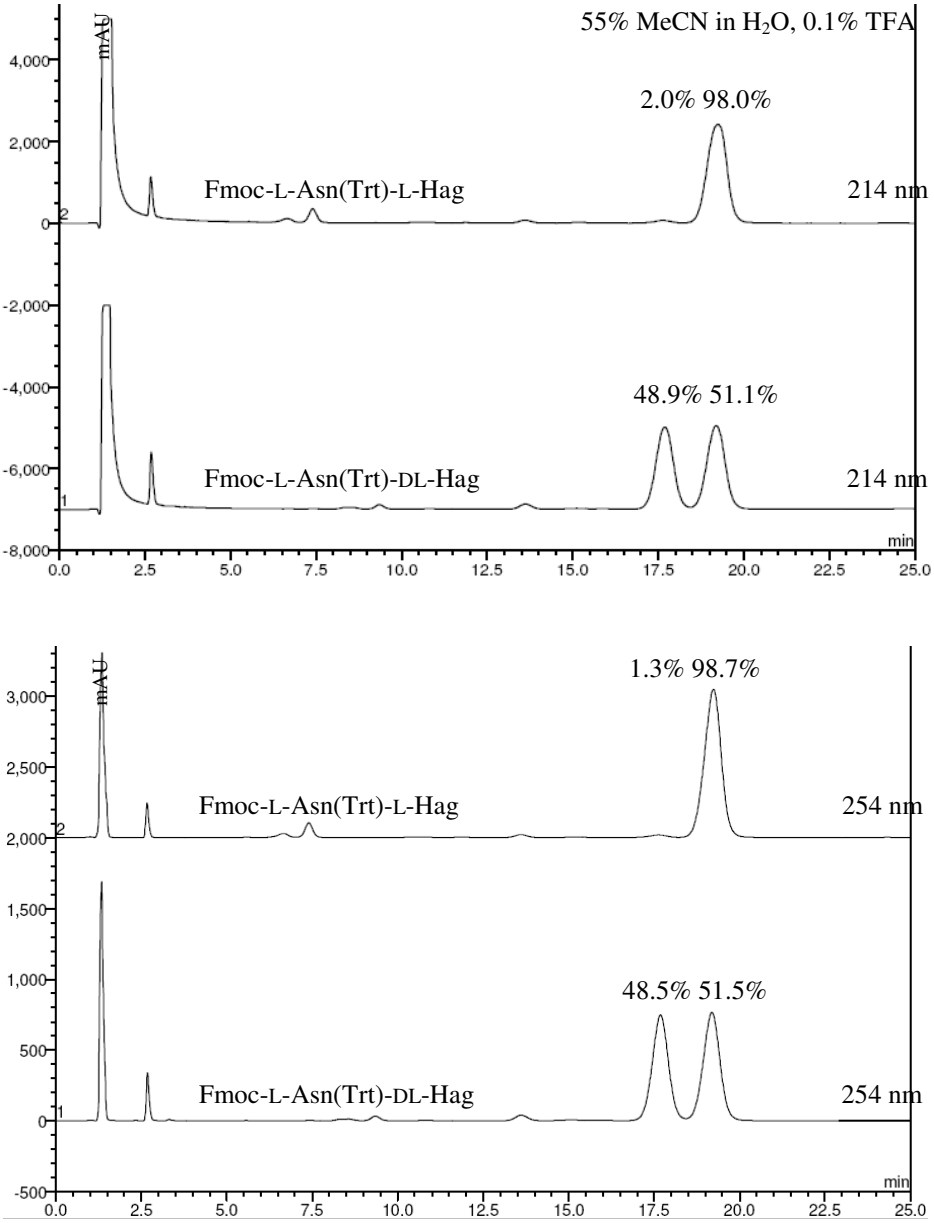



**Figure 11.** UV traces showing HPLC separation of Fmoc-L-Asn(Trt)-L-Hag-OH and Fmoc-L-Asn(Trt)-DL-Hag-OH.

## 3. 1-Glycosyl thiol synthesis

A number of structurally diverse glycosyl thiols were synthesised to use for the thiyl glycosylation of amino acid model systems containing Hag and olefinic proteins. Additionally, 4-*O*-(β-D-galactosyl)-1-thio-β-D-glucose (β-LacSH), bis-*S*-(2,3,4,6-tetra-*O*-acetyl-β-D-gluco-pyranosyl) disulfide ((β-GlcS)$_2$) and 1-thio 3,6-di-*O*-(α-D-

mannopyranosyl)-α/β-D-mannopyranoside (α/β-$Man_3SH$) (Figure 12) were kindly provided by Dr. Balakumar Vijayakrishnan (University of Oxford). The carbohydrates varied in configuration (*e.g.* glucose, mannose), branching, modification (*e.g.* protected or unprotected) and anomeric configuration.

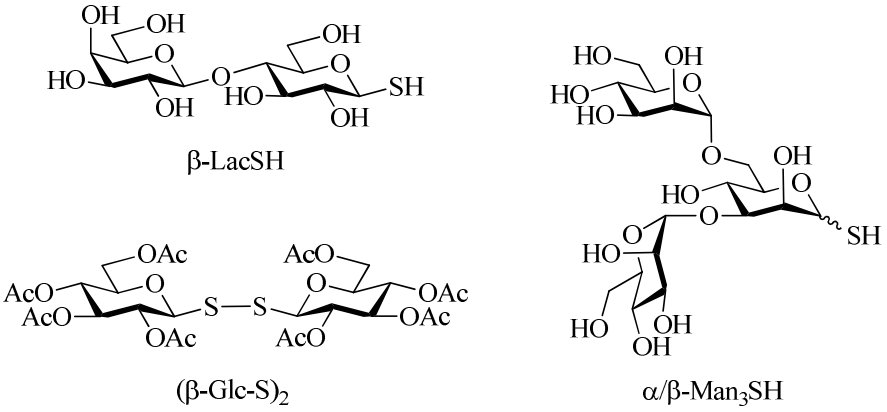


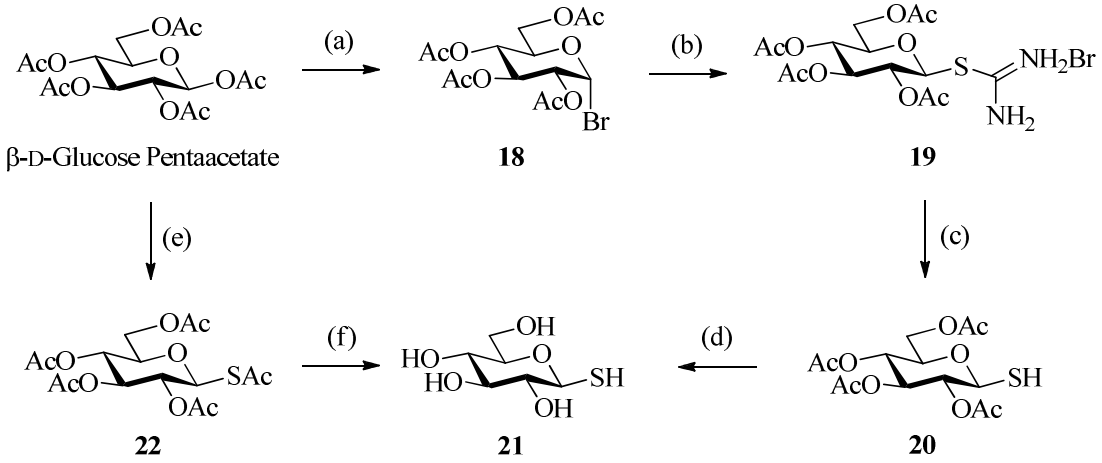

**Figure 12.** 1-Glycosyl Thiols.

### 3.1. 1-Thio-β-D-glucose synthesis

A synthetic route to access 1-thio-β-D-glucose (β-GlcSH) (**21**) was investigated. Starting from commercially available β-D-glucose pentaacetate, glycosyl bromide **18** was obtained by reaction with HBr in acetic acid.[253, 254] Subsequent treatment with thiourea to afford the 1,2-*trans* isothiouronium salt **19**,[255] due to neighboring group participation, followed by mild hydrolysis with sodium metabisulfite[256] and Zemplén deacetylation[257] afforded β-GlcSH (**21**) (Scheme 26). Each compound was purified by crystallisation, and the use of silica gel column chromatography was not required.

β-GlcSH (**21**) was also prepared in a more direct manner. Treatment of commercial β-D-glucose pentaacetate with thioacetic acid in the presence of boron trifluoride etherate afforded thioglycoside **22**. Zemplén deprotection using sodium methoxide yielded the desired product β-GlcSH (**21**) (Scheme 26). In addition, **21** was also

prepared from commercially available 1-thio-β-D-glucose sodium salt (β-GlcSNa) using cation exchange resin (DOWEX 50WX8-200).

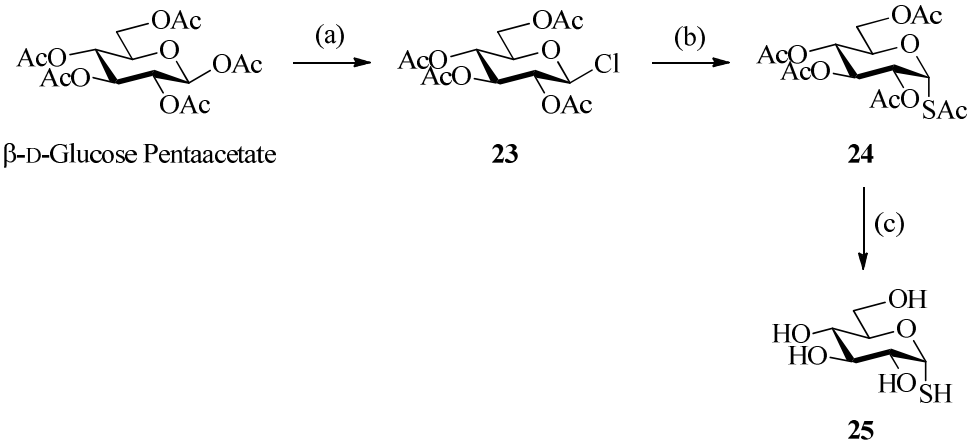



**Scheme 26.** (a) HBr (33% in acetic acid), DCM (94%); (b) Thiourea, acetone, 60 °C (64%); (c) $Na_2S_2O_5$, DCM/$H_2O$ (100%); (d) Na, MeOH (100%); (e) AcSH, $BF_3.Et_2O$, DCM (48%); (f) Na, MeOH (41%).

### 3.2. 1-Thio-α-D-glucose synthesis

For the synthesis of 1-thio-α-D-glucose (α-GlcSH) (**25**), a number of routes were investigated. Starting from commercially available β-D-glucose pentaacetate, glycosyl chloride **23** was obtained by reaction with $PCl_5$.[258] Attempts at accessing peracetylated α-GlcSH (**24**)[259-263] by conducting thioglycosylation (e) in Scheme 26 at elevated temperatures (60°C) only resulted in decomposition of material. Secondly, the preparation of the α-anomer of isothiouronium salt **19** was attempted from known β-glucosyl chloride **23**,[258] however, this resulted exclusively in formation of the undesired β-anomer **19** (80%). More successfully, **24** was prepared by KSAc substitution of chloride **23** in DMPU. This modification of a protocol by Blanc-Muesser *et al.*,[259] who carried out the reaction in highly carcinogenic HMPA, resulted in a similar yield of 40% (*cf.* 36%). Increasing the reaction to room temperature

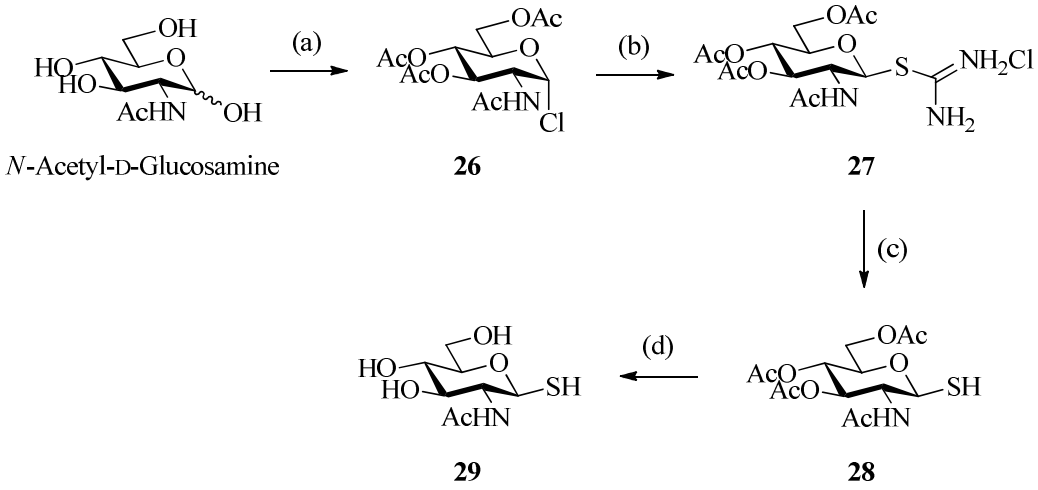
decreased reaction time from 7 days to 1 day but reduced the yield to 28%. In addition, THF and acetone were investigated as possible reaction solvents, but formation of considerable amounts of byproducts were observed, making purification difficult. Zemplén deacetylation[257] of **24** gave α-GlcSH (**25**) (Scheme 27).




**Scheme 27.** (a) $PCl_5$, $BF_3.Et_2O$, DCM (49%); (b) KSAc, DMPU, 0°C (40%); (c) NaOMe, MeOH (100%).

### 3.3. 2-Acetamido-2-deoxy-1-thio-β-D-glucopyranose synthesis

2-Acetamido-2-deoxy-1-thio-β-D-glucopyranose (β-GlcNAcSH) (**29**) was prepared in a similar fashion to β-GlcSH (**21**) (Scheme 26). Starting from commercially available *N*-acetyl-D-glucosamine, glycosyl chloride **26** was obtained by reaction with acetyl chloride.[253, 254] Subsequent treatment with thiourea[255] to afford the 1,2-*trans* isothiouronium salt **27**, due to neighboring group participation, followed by mild hydrolysis with sodium metabisulfite[256] and Zemplén deacetylation[257] afforded β-GlcNAcSH (**29**) (Scheme 28).

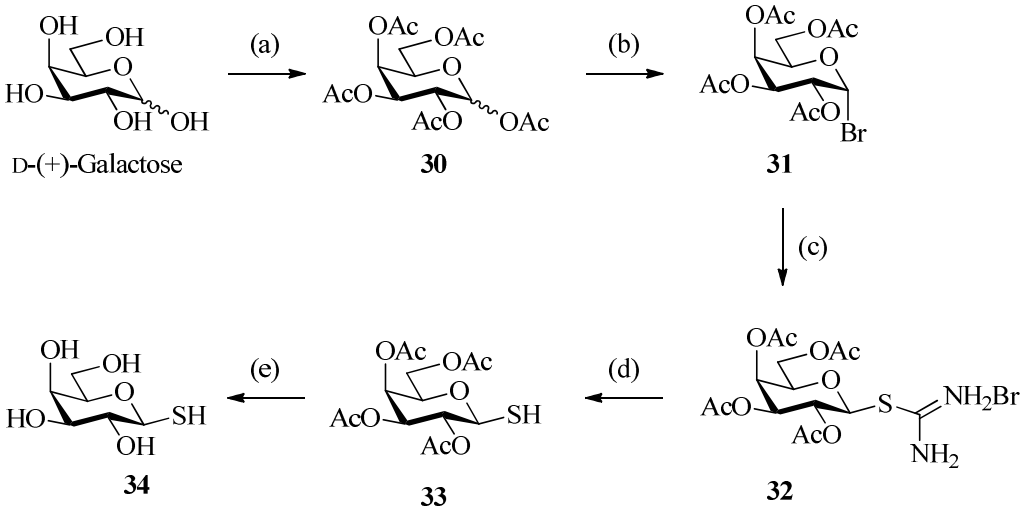



*N*-Acetyl-D-Glucosamine **26** **27** **29** **28**

**Scheme 28.** (a) Acetyl chloride (51%); (b) Thiourea, acetone, 60°C (95%); (c) $Na_2S_2O_5$, DCM/$H_2O$ (89%); (d) NaOMe, MeOH (100%).

### 3.4. 1-Thio-β-D-galactose synthesis

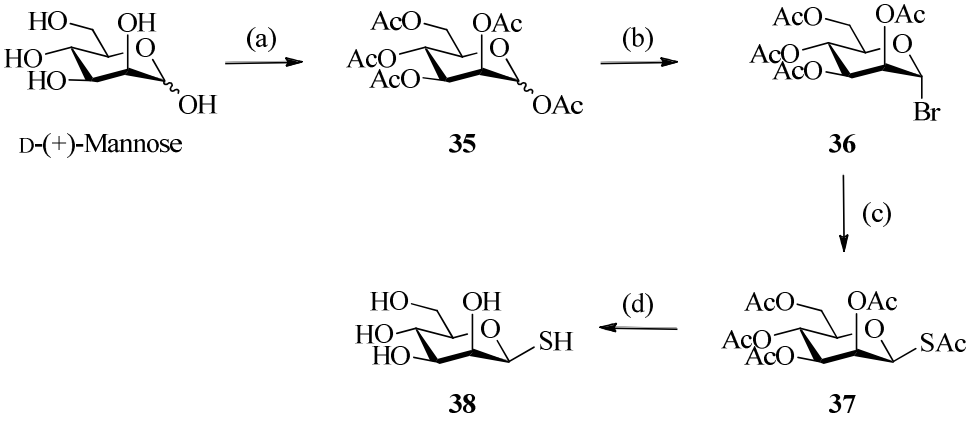
1-Thio-β-D-galactose (β-GalSH) (**34**) was prepared in a similar fashion to β-GlcSH (**21**) (Scheme 26) and β-GlcNAcSH (**29**) (Scheme 27). Starting from commercially available D-(+)-galactose, glycosyl bromide **31** was obtained in two steps following known protocols.[253, 254] Subsequent treatment with thiourea[255] to afford the 1,2-*trans* isothiouronium salt **32**, followed by mild hydrolysis with sodium metabisulfite[256] and Zemplén deacetylation[257] afforded β-GalSH (**34**) (Scheme 29).




**Scheme 29.** (a) Acetic anhydride, pyridine, 0°C (51%); (b) HBr (33% in acetic acid), DCM (92%); (c) Thiourea, acetone, 60°C (50%); (d) $Na_2S_2O_5$, DCM/$H_2O$ (92%); (e) Na, MeOH (100%).

### 3.5. 1-Thio-β-D-mannose synthesis

1-Thio-β-D-mannose (β-ManSH) (**38**) was prepared in a similar fashion to α-GlcSH (**25**) (Scheme 27). Starting from commercially available D-(+)-mannose, glycosyl bromide **36** was synthesised following a known protocol.[253, 254] Tetra-*O*-acetyl-1-*S*-acetyl-1-thio-β-D-mannopyranose (**37**) was prepared by KSAc substitution of bromide **36** in DMPU. Zemplén deacetylation[257] gave β-ManSH (**38**) (Scheme 30).




**Scheme 30.** (a) Acetic anhydride, pyridine (95%); (b) HBr (33% in acetic acid), DCM (95%); (c) KSAc, DMPU (62%); (d) NaOMe, MeOH (100%).

## 4. pH of 1-glycosyl thiol solutions

In order to investigate the thiyl glycosylation of Hag and olefin containing proteins at different pH values (Table 6 and Table 9, respectively), it was first necessary to understand the effect that the addition of 1-glycosyl thiols had on the pH of buffer solutions of various concentrations. The p$K_a$ of the thiol group of β-GlcSH (**21**) is reported to be 7.7.[264]

Commercially available 1-thio-β-D-glucose sodium salt (β-GlcSNa) was dissolved to a final concentration of 1 M and 0.1 M in increasing concentrations (0.05–1 M) of acetic acid/sodium acetate buffer (acetate buffer) (pH 4). The pH of the final solution was measured using a pH meter, and was found to be maintained only when the concentration of β-GlcSNa was 0.1 M and the buffer strength ≥ 0.25 M (Table 3, entries 6–8). This finding was also applicable at pH 7 (sodium phosphate buffer) and pH 10 (CAPS buffer) (Table 3, entries 9 and 10).

**Table 3.** pH of buffer solutions containing β-GlcSNa.

| Entry | [β-GlcSNa] | [Buffer] | Buffer | pH[a] |
|---|---|---|---|---|
| **1** | 1 M | 0.05 M | Acetate, pH 4 | 9.2 |
| **2** | 1 M | 0.25 M | Acetate, pH 4 | 7.0 |
| **3** | 1 M | 0.5 M | Acetate, pH 4 | 5.3 |
| **4** | 1 M | 1 M | Acetate, pH 4 | 4.8 |
| **5** | 0.1 M | 0.05 M | Acetate, pH 4 | 7.6 |
| **6** | 0.1 M | 0.25 M | Acetate, pH 4 | 4.0 |
| **7** | 0.1 M | 0.5 M | Acetate, pH 4 | 4.0 |
| **8** | 0.1 M | 1 M | Acetate, pH 4 | 4.0 |
| **9** | 0.1 M | 0.25 M | Sodium phosphate, pH 7 | 7.0 |
| **10** | 0.1 M | 0.25 M | CAPS, pH 10 | 10.0 |

[a] pH after addition of β-GlcSNa to buffer solution.

## 5. Effect of UV light and heating on L-Hag and 1-glycosyl thiols

As the glycoconjugation reaction was to be photo- and thermally initiated, it was necessary to confirm that L-Hag (**+3**) and 1-glycosyl thiols were individually stable under the reaction conditions, *e.g.* to racemisation and anomerisation, respectively. It should be noted that the UV light used was not quantified at any point throughout this project.

### 5.1. L-Hag

Fmoc-L-Asn(Trt)-L-Hag-OH (see pg. 58) was employed to demonstrate that UV light irradiation did not cause racemisation of L-Hag. The dipeptide dissolved in 0.25 M acetate buffer [0.25 M acetic acid/sodium acetate buffer] (pH 4)/DMF (1:1) was subjected to UV light for 8 hours. Aliquots were removed from the reaction mixture over the course of the experiment and were analysed by HPLC using a Phenomenex Synergi Hydro-RP column. The samples were eluted using an isocratic mobile phase

comprising 57.5% acetonitrile in water plus 0.1% TFA. No racemisation was observed, which confirmed that L-Hag was stable to UV irradiation (Figure 13). Fmoc-L-Asn(Trt)-DL-Hag-OH was used as a control, and under the elution conditions, Fmoc-L-Asn-D-Hag-OH and Fmoc-L-Asn-L-Hag-OH eluted at 15.3 and 16.6 minutes, respectively (Figure 13, top). Note: Irrespective of the concentration and quantity of Fmoc-L-Asn(Trt)-L-Hag-OH injected onto the Phenomenex Synergi Hydro-RP column, the peak shape was broad.

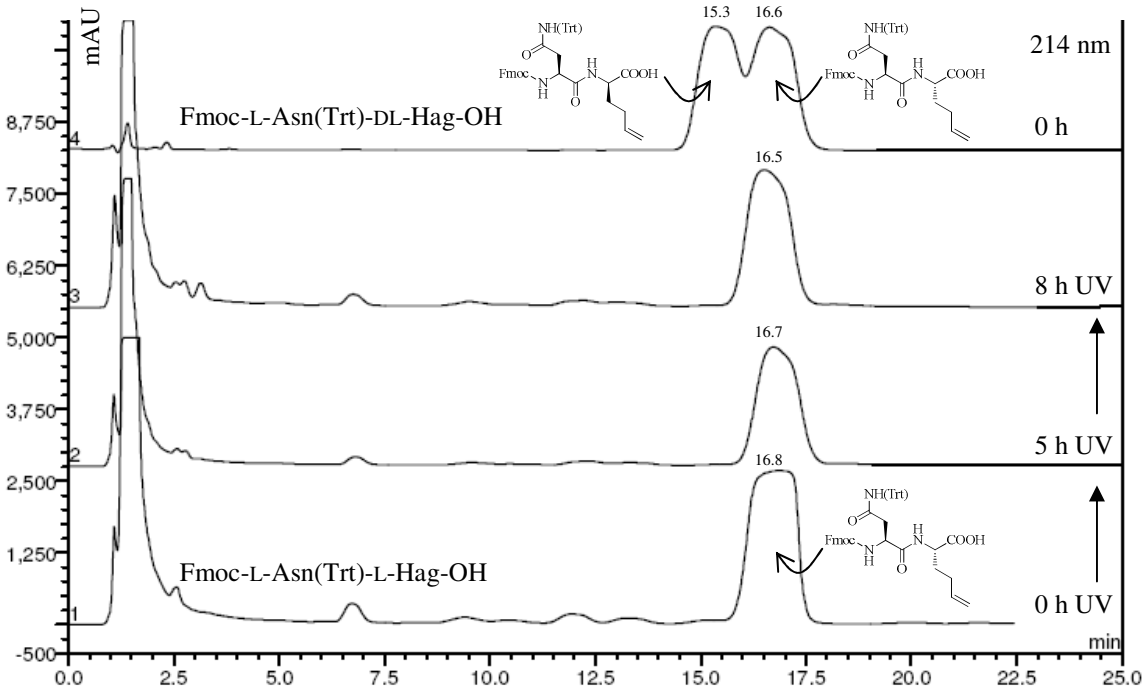



**Figure 13.** Effect of UV light on L-Hag. Note: graphs 2, 3 and 4 are offset by 2750, 5500 and 8250 mAU, respectively

## 5.2. 1-Glycosyl thiols

The effect of UV light on commercial 1-thio-β-D-glucose sodium salt (β-GlcSNa) and β-GlcNAcSH (**29**) was investigated at various pH values, to determine if irradiation caused anomerisation of 1-glycosyl thiols.

A $^{1}$H NMR of β-GlcSNa and β-GlcNAcSH (**29**) were recorded, and confirmed only the β-anomer of each to be present. β-GlcSNa and β-GlcNAcSH (**29**) were then dissolved to a final concentration of 0.1 M in 0.25 M acetate buffer (pH 4), sodium phosphate buffer (pH 7) and CAPS buffer (pH 10). The samples were incubated at room temperature (control) or irradiated with UV light for 4 hours. The samples were then analysed for anomerisation using $^{1}$H NMR by comparison of H1β/H1α integration. No anomerisation of β-GlcSNa or β-GlcNAcSH (**29**) was observed as a result of irradiation with UV light (Table 4), except in the case of β-GlcNAcSH (**29**) at pH10. This anomalous result should be eliminated or verified in the future by repeating this experiment. Anomerisation of both β-GlcSNa and β-GlcNAcSH (**29**) was observed, particularly at low pH.[262]

**Table 4.** Effect of UV light on representative 1-glycosyl thiols.

| **1-Glycosyl Thiol** | **[1-Glycosyl Thiol]** | **0.25 M Buffer** | **H1β/H1α (RT, 4 h)** | **H1β/H1α (UV, 4 h)** |
|---|---|---|---|---|
| β-GlcSNa | 0.1 M | Acetate, pH 4 | 1 : 0.1 | 1 : 0.1 |
| β-GlcSNa | 0.1 M | Sodium phosphate, pH 7 | 1 : 0.05 | 1 : 0.05 |
| β-GlcSNa | 0.1 M | CAPS, pH 10 | 1 : 0 | 1 : 0 |
| β-GlcNAcSH | 0.1 M | Acetate, pH 4 | 1 : 0.1 | 1 : 0.1 |
| β-GlcNAcSH | 0.1 M | Sodium phosphate, pH 7 | 1 : 0.05 | 1 : 0.05 |
| β-GlcNAcSH | 0.1 M | CAPS, pH 10 | 1 : 0.01 | 1 : 0.14 |

The effect of heating on 1-thio-β-D-glucose sodium salt (β-GlcSNa) and β-GlcNAcSH (**29**) was investigated at various pH values, to determine if heating caused anomerisation of 1-glycosyl thiols.

A $^1$H NMR of β-GlcSNa and β-GlcNAcSH (**29**) were recorded, and confirmed only the β-anomer of each to be present. β-GlcSNa and β-GlcNAcSH (**29**) were then dissolved to a final concentration of 0.1 M in 0.25 M acetate buffer (pH 4), sodium phosphate buffer (pH 7) and CAPS buffer (pH 10). The samples were incubated at room temperature (control) or heated to 55$^o$C for 4 hours. The samples were then analysed for anomerisation using $^1$H NMR by comparison of H1β/H1α integration. Anomerisation of β-GlcSNa and β-GlcNAcSH (**29**) was observed as a result of pH and also of heating (Table 5).

**Table 5.** Effect of heating on representative 1-glycosyl thiols.

| **1-Glycosyl Thiol** | **[1-Glycosyl Thiol]** | **0.25 M Buffer** | **H1β/H1α (RT, 4 h)** | **H1β/H1α (55$^o$C, 4 h)** |
|---|---|---|---|---|
| β-GlcSNa | 0.1 M | Acetate, pH 4 | 1 : 0.1 | 1 : 0.2 |
| β-GlcSNa | 0.1 M | Sodium phosphate, pH 7 | 1 : 0.05 | [a] |
| β-GlcSNa | 0.1 M | CAPS, pH 10 | 1 : 0 | 1 : 0.05 |
| β-GlcNAcSH | 0.1 M | Acetate, pH 4 | 1 : 0.1 | 1 : 0.5 |
| β-GlcNAcSH | 0.1 M | Sodium phosphate, pH 7 | 1 : 0.05 | 1 : 0.5 |
| β-GlcNAcSH | 0.1 M | CAPS, pH 10 | 1 : 0.01 | 1 : 0.9 |

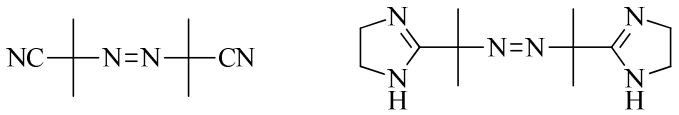
[a] Could not be determined from $^1$H NMR as H1β signal hidden.

## 6. Glyco-amino-acid synthesis

The free-radical reaction of glycosyl thiyls was investigated and optimised initially on a representative model system comprising DL-Hag (**±3**) and β-GlcSNa. As a

prerequisite for later protein stability, solubility and viability, the reaction was developed to comply with aqueous solution chemistry. Significantly, all prior protocols[206-208] for glycosyl thiyl generation have until now employed organic solvents. Together reactions probed the effect of pH (4, 7 or 10) and mode of initiation (hν or Δ [room temperature or 55 $^{o}$C], with and without water soluble initiator Vazo44[265-267]) (Table 6). UV light was generated using a borosilicate glass immersion well fitted with a medium pressure 125 W Hg lamp giving maximum emission at 365 nm.

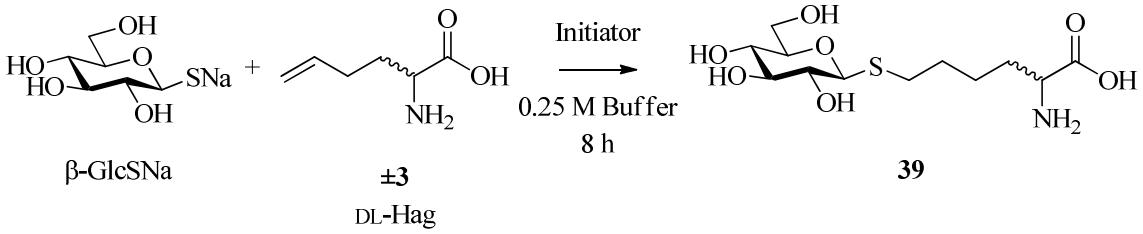
Vazo44 (VA044 or 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride) is an analogue of AIBN (azobisisobutyronitrile) (Figure 14) and is one of a number of water soluble free radical sources from DuPont. As an azo radical initiator, Vazo44 undergoes homolytic decomposition to produce two free radicals and release molecular nitrogen. The decomposition reaction is reported to proceed approximately at the rate of first order reaction with no side reactions occurring in aqueous solution. The grade number of Vazo44 is 44, which corresponds to the temperature ($^{o}$C) at which the half-life of Vazo44 in solution is ten hours. Typically reactions are run at 10 to 30$^{o}$C above this temperature. For this reason, Vazo44 was chosen in preference to other commercially available water soluble Vazo compounds that all have higher grade numbers (*e.g.* Vazo56, Vazo68), as these would require reactions to be run at higher temperatures and therefore would be less suitable for use with proteins.




AIBN Vazo44

**Figure 14.** Chemical structures of AIBN and Vazo44.

**Table 6.** Optimization of Hag Glycoconjugations.

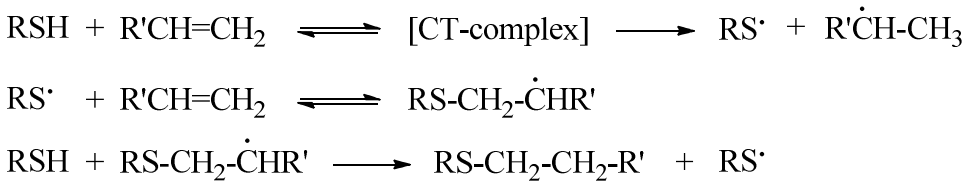



β-GlcSNa ±3 DL-Hag 39

| Initiator | 0.25 M Acetate (pH 4) | | 0.25 M Sodium Phosphate (pH 7) | 0.25 M CAPS (pH 10) |
|---|---|---|---|---|
| | Conversion Calculated From Crude NMR[c,d] | Isolated Yield[d] | | |
| hν | >95%[a] | 35%[a] | 27%[a] | Trace |
| 55°C | 55%[a] | 36%[a] | Trace | NR |
| RT | 90%[a] | 88%[a] | NR | NR |
| Vazo44 (0.2 eq), hν | >98%[b] | 88%[a] | 41%[a] | Trace |
| Vazo44 (0.2 eq), 55°C | >98%[b] | 91%[a] | 50%[a] | NR |
| Vazo44 (0.2 eq), RT | >98%[a] | 98%[a] | Trace | NR |

Trace = <5% identified from NMR of crude reaction mixture; NR = no reaction; RT = room temperature. [a] 1.2 eq 1-glycosyl thiol; [b] 10 eq 1-glycosyl thiol; [c] See pg. 265 for crude NMRs; [d] Note that the conversion calculated from crude NMR and isolated yield come from two separate reactions.

The studies indicated that both photo- and thermal-initiation allowed successful conjugation of β-GlcSNa to DL-Hag (**±3**) to yield 1-thioglycoside **39**. As thiol-ene reactions are thought to exhibit spontaneous dark reactions in the absence of initiator via a charge-transfer complex between the thiol and ene functional groups (Scheme 31),[268-271] it was not surprising that 1-thioglycoside formation at room temperature, without photo- or thermal-initiation or the presence of Vazo44, was also observed.

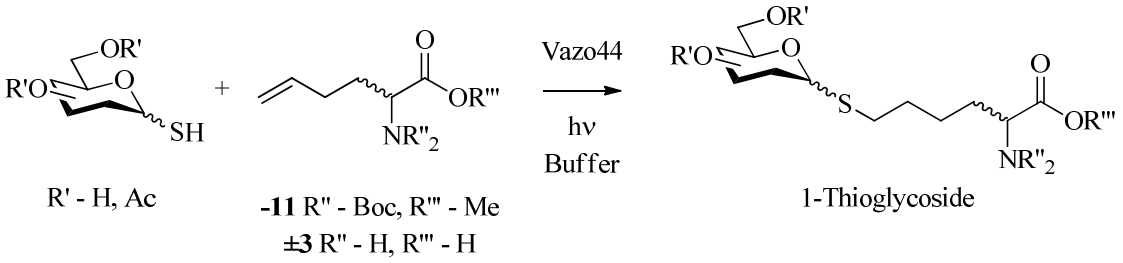


**Scheme 31.** Thiol-ene reaction via a charge-transfer (CT) complex.

Interestingly, enhancements in rate and yields were observed through photoinitiation in the presence of Vazo44. The pH of the reaction proved important, with lower pH values (4 and 7) proving superior to the use of pH 10. Importantly, ambient conditions in aqueous solutions potentially compatible with protein stability (without the need for *e.g.* degassing)[272] proved effective. For all photo-initiated reactions, good regioselectivity (> 98% ε-thiolation) and retention of anomeric configuration in the product were observed. While for the thermally (55°C) initiated reactions, only good regioselectivity (> 98% ε-thiolation) was observed and complete retention of anomeric configuration was not. This was unsurprising given that anomerisation of β-GlcSNa and β-GlcNAcSH (**29**) was previously observed as a result of pH and heating (see pg. 70). Use of excess thiol (10 eq) allowed essentially complete conversion.

Since the stability and reactivity of glycosyl thiyls can vary and to demonstrate substrate breadth and broad applicability of the method, a range of 1-glycosyl thiols were then utilised as starting materials, in both protected and unprotected forms, including β-GlcSH (**21**), α-GlcSH (**25**), β-GlcNAcSH (**29**), β-GalSH (**34**), β-ManSH (**38**) and disaccharide 4-*O*(β-D-galactosyl)-1-thio-β-D-glucose (β-LacSH). In addition, Hag was used in both protected (**-11**) and unprotected forms (**±3**). Together the reactions probed the effect of bulk, differing configuration, protecting groups and anomeric stereochemistry in the thiols, and variation around the N- and C-termini of Hag (Table 7).

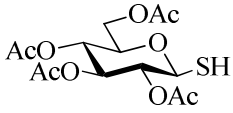
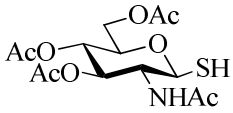
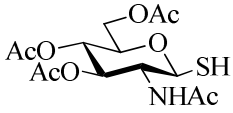
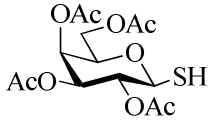
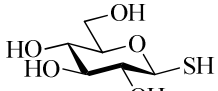
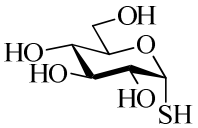
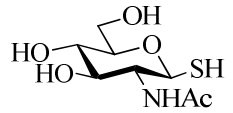
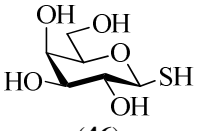
**Table 7.** Amino acid Glycoconjugation Reactions.

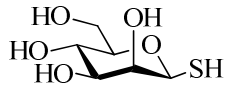
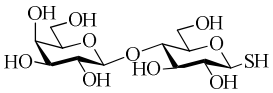



R' - H, Ac

**-11** R'' - Boc, R''' - Me
**±3** R'' - H, R''' - H

1-Thioglycoside

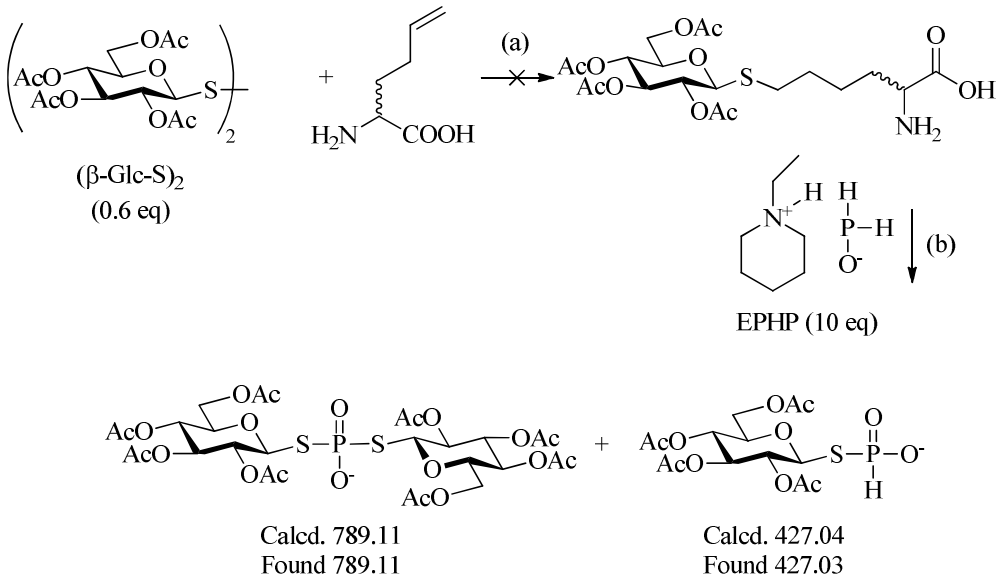
| | 1-Glycosyl Thiol (Product #) | | Olefin | Vazo-44 | Solvent | Time | Crude Yield [%][b] | Isolated Yield [%] |
|---|---|---|---|---|---|---|---|---|
| **1** |  (**40**) | 1.2 eq | -**11** | 0.4 eq | $H_2O$/MeOH (1:1) | 3 h | - | 79 |
| **2** |  (**41**) | 1.2 eq | -**11** | 0.2 eq | $H_2O$/MeOH (1:1) | 2 h | - | 71[a] |
| **3** |  (**42**) | 2.2 eq | **±3** | 0.2 eq | $H_2O$ | 17 h | - | 81 |
| **4** |  (**43**) | 2.2 eq | **±3** | 0.2 eq | $H_2O$/MeOH (1:1) | 3 h | - | 57 |
| **5** |  (**39**) | 1.2 eq | **±3** | 0.2 eq | 0.25 M Acetate pH 4 | 8 h | - | 88 |
| | | 10 eq | | | | | >98 | - |
| **6** |  (**44**) | 2.2 eq | **±3** | 0.3 eq | 0.25 M Acetate pH 4 | 12 h | - | 42 |
| | | 5 eq | | | | 8 h | ~50 | - |
| **7** |  (**45**) | 1.2 eq | **±3** | 0.2 eq | 0.25 M Acetate pH 4 | 8 h | - | 63 |
| | | | | | | | >98 | - |
| **8** |  (**46**) | 1.2 eq | **±3** | 0.2 eq | 0.25 M Acetate pH 4 | 8 h | - | 70 |
| | | 2.4 eq | | 0.3 eq | | 10 h | >80 | - |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **9** |  (**47**) | 2.2 eq | **±3** | 0.3 eq | 0.25 M Acetate pH 4 | 12 h | - | 28 |
| | | 5 eq | | | | 8 h | ~50 | - |
| **10** |  (**48**) | 1.7 eq | **±3** | 0.3 eq | 0.25 M Acetate pH 4 | 12 h | - | 56 |
| | | 3 eq | | | | 8 h | >98 | - |

[a] mono-boc; [b] conversion determined from crude reaction mixture $^{1}$H NMR (see pg. 265).

Extension of the reaction to the full range of glycosyl thiols revealed successful coupling in all cases (Table 7). Notably, rates and yields for the *cis*-1,2-hydroxythiols β-ManSH (**38**) and α-GlcSH (**25**) were significantly lower (entries 6 and 9, 42 and 28% respectively) whilst those for less hindered *trans*-hydroxythiols were good to excellent (*e.g.* entry 5 - 88% isolated yield). For these more reactive thiols, conversions were essentially complete even for larger sugars such as disaccharide β-LacSH, and the isolated yields largely reflect the difficulties in handing amphiphilic products.

Although disulfides have been suggested as putative thiyl precursors previously,[273] use of disulfide (β-GlcS)$_2$ gave no product (conditions (a) in Scheme 32). This suggests that not only do thiols provide an important hydride source in the conjugation reaction but that the lower yields observed at higher pH (10) may be a consequence of more rapid competitive oxidation to unreactive disulfide. Subsequent addition of the H$^{\bullet}$ source ethylpiperidine hypophosphite (EPHP) to the reaction mixture (conditions (b) in Scheme 32) did also not give the desired product. Instead, on complete consumption of the starting material, mass signals that might correspond to the phosphate compounds shown in Scheme 32 were observed by ESI-MS.[274, 275] Alternative hydride donor sources[276] were not investigated.



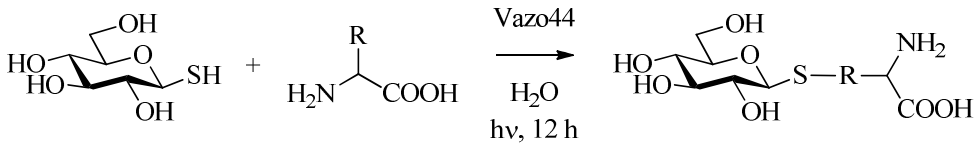

**Scheme 32.** (a) Vazo44 (0.2 eq), MeOH/$H_2O$ (2:1), hν, 10 h; (b) Vazo44 (2.6 eq), hν, 6 h.

## 7. Chemoselectivity for natural amino acids

The chemoselectivity of the glycoconjugation reaction for Hag over each of the 20 naturally occurring amino acids (excluding isoleucine as similar side chain leucine) was investigated. Homoallylglycine was used as a positive control.

When β-GlcSH (**21**) was exposed to each amino acid under optimal conditions (hν, Vazo44) disulfide bond formation between β-GlcSH and the thiol group of cysteine was observed (Table 8). As a consequence, when applying the thiyl glycoconjugation procedure to model proteins, it was necessary to use proteins devoid of free cysteine residues. No conjugation products with any amino acid other than Hag or cysteine were detected by ESI-MS.

**Table 8.** Glyco-amino-acid formation by ESI-MS.




**21** Amino Acid Glyco-Amino-Acid

| Amino Acid | MW | ESI$^+$ HRMS for Glyco-Amino Acid-H$^+$ | |
|---|---|---|---|
| | | Calcd. | Found |
| **Ala** | 89.1 | 284.1 | - |
| **Arg** | 174.2 | 369.2 | - |
| **Asn** | 132.1 | 327.1 | - |
| **Asp** | 133.1 | 328.1 | - |
| **Cys** | 121.1 | 314.1 (S-S) | 314.0 (S-S) |
| **Gln** | 147.1 | 342.1 | - |
| **Glu** | 146.1 | 341.1 | - |
| **Gly** | 75.0 | 270.0 | - |
| **His** | 155.1 | 350.1 | - |
| **Hag (positive control)** | 129.1 | 324.1 | 324.1 |
| **Leu** | 131.1 | 326.2 | - |
| **Lys** | 146.1 | 341.2 | - |
| **Met** | 149.2 | 344.2 | - |
| **Phe** | 165.1 | 360.2 | - |
| **Pro** | 115.1 | 310.1 | - |
| **Ser** | 105.0 | 300.1 | - |
| **Thr** | 119.1 | 314.1 | - |
| **Trp** | 204.2 | 399.2 | - |
| **Tyr** | 181.1 | 376.2 | - |
| **Val** | 117.1 | 312.1 | - |

## 8. Model protein preparation

The TIM-barrel protein β-glycosidase from *Sulfolobus solfataricus* (*Ss*βG) (Figure 15, left),[277] the cuboid right-handed quadrilateral β-helix protein Np276 from *Nostoc punctiforme* (Figure 15, middle)[278] and the self-assembled homomultimer virus-like bacteriophage particle Qβ (Figure 15, right)[279-281] were chosen as model proteins for the thiyl glycosylation procedure. These three proteins have great structural diversity and were selected to show the general applicability of the thiyl glycosylation reaction to several different classes of proteins.

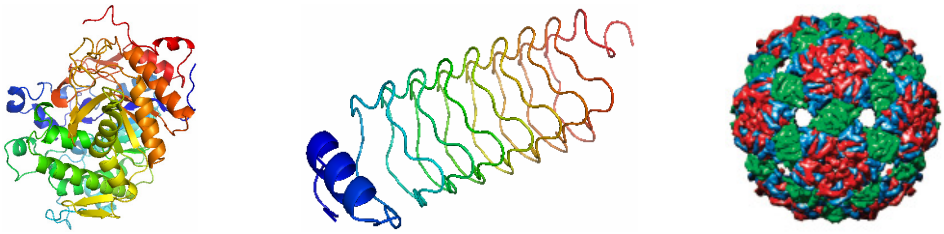


**Figure 15.** Crystal structures of β-glycosidase from *Sulfolobus solfataricus* (*Ss*βG) (left), Np276 from *Nostoc punctiforme* (middle) and bacteriophage Qβ (right). [PDB IDs: 1GOW, 2J8K and 1QBE, respectively].

The unnatural amino acid L-Hag was site-specifically introduced into these protein systems as an olefinic 'tag' using the medium shift procedure established by Tirrell *et al.*.[120] This method takes advantage of auxotrophy, the inability of an organism to synthesise a particular compound required for its growth.[134] Here, the auxotrophic *E. coli* strain B834(DE3), unable to synthesise methionine, was used. The model olefin containing proteins *Ss*βG-Hag43, Np276-Hag61 and Qβ-$(Hag16)_{180}$ were obtained by: (i) growing the culture in methionine-containing medium; (ii) performing a medium shift into methionine-deficient Hag-containing medium; (iii) inducing protein expression using IPTG. The gene sequences were designed to create proteins

displaying an olefin at a single site, the position of which could simply be controlled by the 'Met' triplet codon ATG. However, this reassignment of the Met codon unfortunately prevents the use of proteins that contain functionally essential methionines.

## 8.1. *Ss*βG-Hag43

### 8.1.1. M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG

Model protein substrate M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG, with an olefin uniquely positioned at site 43, is formally an eleven-point mutant of the wild-type β-glycosidase from *Sulfolobus solfataricus* in which all cysteine (C344S) and methionine (M21I M73I M148I M206I M236I M275I M280I M383I M439I) residues occurring in the wild-type protein, except Met43, have been mutated to serine and isoleucine, respectively.[215] This protein was chosen for a number of reasons. Firstly, the X-ray crystal structure of wild-type *Ss*βG has revealed that Met43 shows good accessibility.[277] Secondly, a cycloaddition reaction on both an incorporated azidohomoalanine and homopropargylglycine at site 43 has already been shown,[135, 136] therefore demonstrating that reactions at this position in the protein sequence are possible. Thirdly, as after unnatural amino acid incorporation the protein only has one L-Hag residue, only one alkene tag is available for modification, eliminating the possibility of formation of multiple glycoforms.

L-Hag was site-specifically introduced into M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG by unnatural amino acid incorporation to generate M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG. The N-terminal His-tagged protein was purified. The protein was judged to be

>95% pure by SDS-PAGE and no M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG was dectected by ESI-MS. The protein identity was confirmed by ESI-MS (found 57215 Da, calcd. 57215 Da) (Figure 16). In addition, the enzymatic activity of the protein was demonstrated using the X-Gal colorimetric assay.[282]

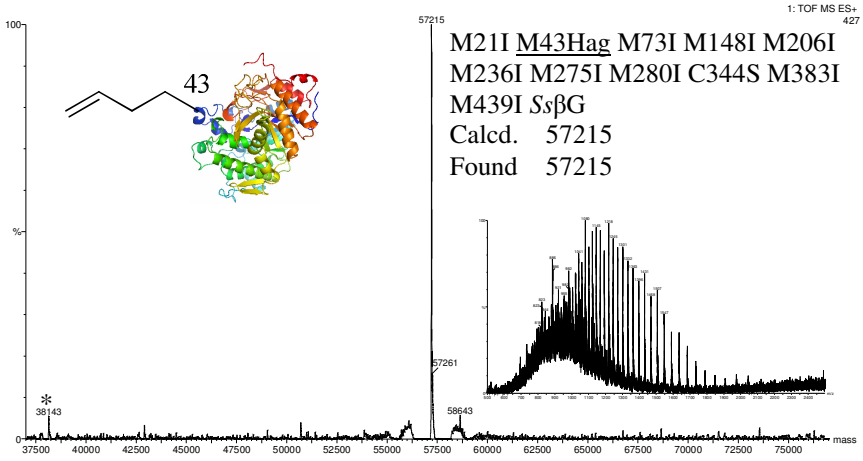



**Figure 16.** Raw (insert) and deconvoluted ESI-MS of M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

As proteins are known to be affected by UV irradiation (see pg. 103),[286-290] the stability of M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG to UV light was tested (see chapter 3, pg. 108). Unexpectedly, it was discovered that the protein cleaved to give two fragments when irradiated with UV light, therefore making it unsuitable as a model protein.

### 8.1.2. M21I M21I M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG (*Ss*βG-Met43) and M21I M43Hag M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG (*Ss*βG-Hag43)

In order to generate an *Ss*βG mutant stable to UV light, the photo-induced site-selective cleavage reaction that was first identified in M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG, was investigated in detail (see chapter 3).[210, 291] Importantly, the elucidation of key structural and sequence features associated with the unique, serendipitous photocleavage reaction in certain TIM-barrel proteins such as *Ss*βG, allowed a key stabilising Pro152Ala mutation to be introduced into WT *Ss*βG, making it stable to UV light (see pg. 124).

The same Pro152Ala mutation was then introduced into M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG by site directed mutagenesis to generate a pET24d vector carrying the gene encoding M21I M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG (*Ss*βG-Met43). The protein was expressed and purified, and the protein identity confirmed by ESI-MS (found 57209 Da, calcd. 57209 Da) (see pg. 289). In addition, the enzymatic activity of the protein was demonstrated using the X-Gal colorimetric assay.[282]

The protein was shown to be stable to UV light: protein was added to a cuvette and was irradiated with UV light for 4 hours. Aliquots were removed from the reaction mixture over the course of the experiment and were analysed by SDS-PAGE (Figure 17). M21I M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG was later used as a negative control substrate in glycoconjugation reactions to

further confirm the chemoselectivity of the reaction for the olefinic 'tag' and the lack of reaction with other amino acids in *Ss*βG.

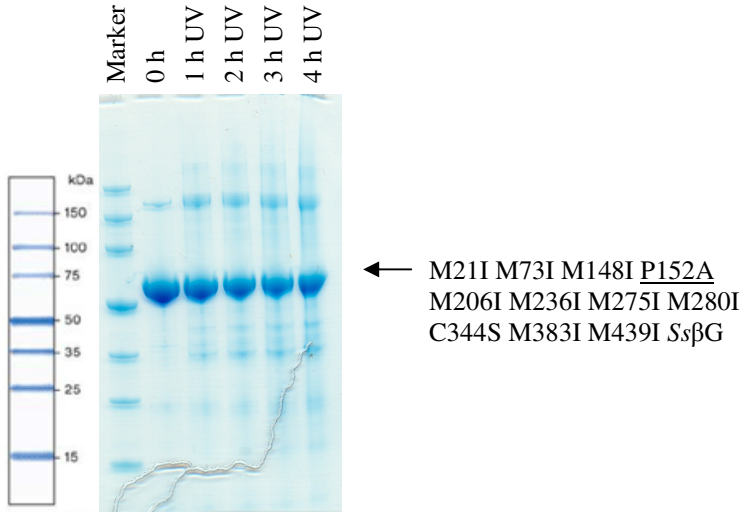



**Figure 17.** Time course analysis of M21I M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG exposed to UV light by SDS-PAGE.

In the same way as for the synthesis of M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG, the medium shift procedure established by Tirrell *et al.*[292] was used for unnatural amino acid incorporation of L-Hag at position 43 in M21I M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG to afford M21I M43Hag M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG (*Ss*βG-Hag43). The protein identity was confirmed by ESI-MS (found 57188 Da, calcd. 57189 Da) (Figure 18). In addition, the enzymatic activity of the protein was demonstrated using the X-Gal colorimetric assay.[282]

Further evidence for the incorporation of Hag was obtained by performing an in-gel trypsin digest of the protein. The precursor ion *m/z* 1052.4 $[M+H]^{3+}$ from the peptide corresponding to residues 36-64 of *Ss*βG-Hag43 was selected and fragmented in order to demonstrate the presence of Hag in place of Met at position 43. A number of b and y fragment ions of the peptide were detected by ESI-MS, and are shown in the

MS/MS spectrum in Figure 20. Additional evidence for the incorporation of Hag was not obtained by performing a gel-shift assay at any point during this project.

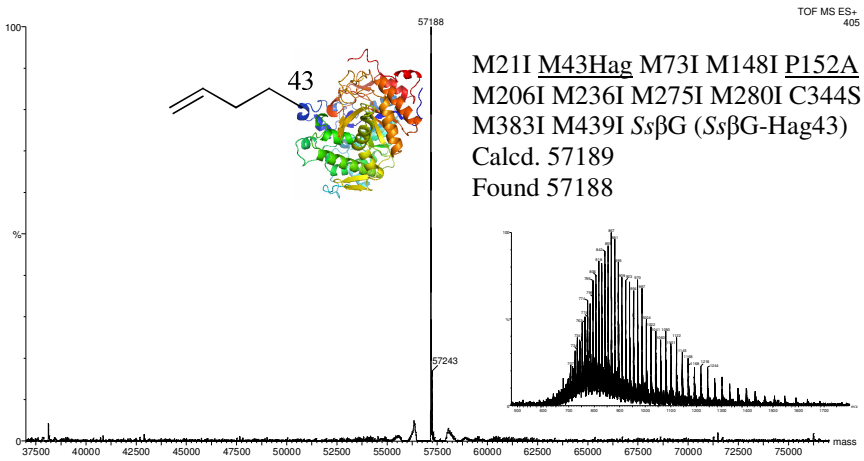



**Figure 18.** Raw (insert) and deconvoluted ESI-MS of *Ss*βG-Hag43.

It should be noted that throughout this thesis, the product ions from MS/MS experiments are named according to the nomenclature system shown in Figure 19. The different ions result from fragmentation along the peptide backbone. If a charge is retained on the N-terminal fragment, the ions are classified as a-, b- or c-, while if the charge is retained on the C-terminal fragment, the ions are classified as x-, y- or z-.[293]

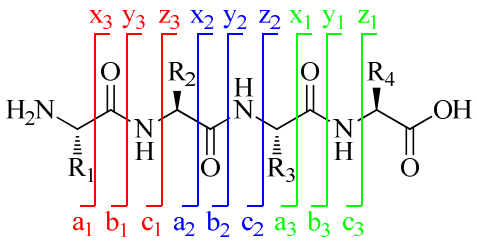



**Figure 19.** MS fragmentations of the peptide linkage.

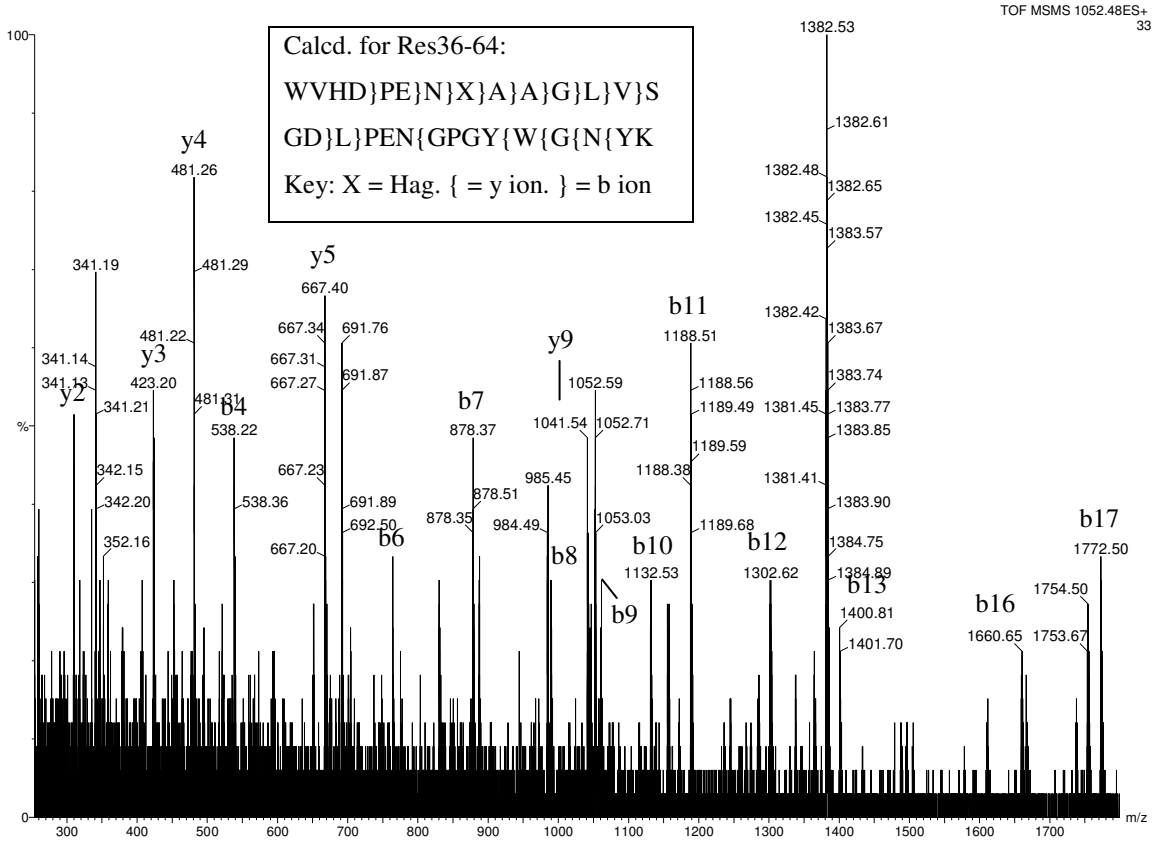



**Figure 20.** MS/MS spectrum of *m/z* 1052.4 $[M+H]^{3+}$ from the peptide corresponding to residues 36-64 of *Ss*βG-Hag43.

The stability of *Ss*βG-Hag43 to UV light was confirmed: protein was added to a cuvette, and was irradiated with UV light for 6 hours. Aliquots were removed from the reaction mixture over the course of the experiment and were analysed by SDS-PAGE (Figure 21).

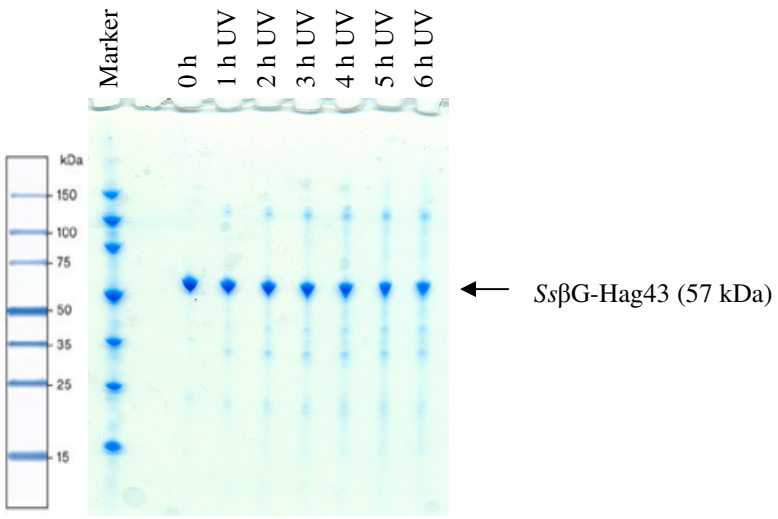



**Figure 21.** Time course analysis of *Ss*βG-Hag43 exposed to UV light by SDS-PAGE.

## 8.2. Np276-Hag61

The second model protein substrate chosen was M21I M61Hag Np276 (Np276-Hag61), which displayed a single olefin at position 61. Incorporation of Hag as the olefinic 'tag' was achieved by using a mutant of wild-type Np276 in which the only methionine residue (Met21) occurring in the wild-type protein other than Met61, was mutated to isoleucine.

L-Hag was site-specifically introduced into M21I Np276 to generate M21I M61Hag Np276 (Np276-Hag61) by unnatural amino acid incorporation. IPTG-induced protein expression at 37°C for 5 hours *vs.* 30°C for 16 hours was found to afford similar amounts of protein. The N-terminal His-tagged protein was purified. Protein was detected with an on-line detector monitoring $A_{280,}$ but due to there only being one absorbing residue (Tyr30) in Np276-Hag61 it was only possible to detect eluted protein by SDS-PAGE. Protein identity was confirmed by ESI-MS (found 20818 Da, calcd. 20818 Da) (Figure 22). However, a signal at *m/z* 20996, which might correspond to Np276-Hag61 bearing an α-N-gluconoyl-modified His-tag (+178 by

ESI-MS), was also observed. This α-N-gluconoyl modification has been reported for other proteins expressed in *E. coli* with the N-terminal sequence GSS[H]$_6$SSGLVPR, which is identical to that of Np276-Hag61.[294] Further evidence for the α-N-gluconoyl-modified His-tag was not obtained, but could be in the future by performing MS/MS analysis on a tryptic digest of the protein.

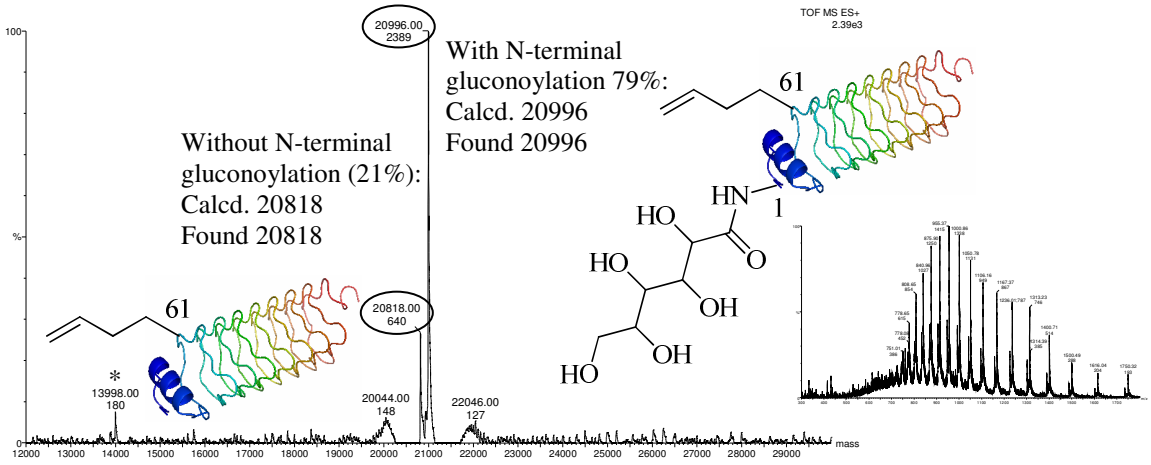



**Figure 22.** Raw (insert) and deconvoluted ESI-MS of Np276-Hag61. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

## 8.3. Qβ-(Hag16)$_{180}$

The third model protein substrate chosen was the self-assembled multimeric, virus-like particle Qβ-(Hag16)$_{180}$ displaying 180 olefins. Recently, elegant work by the Finn group[295] has used cycloaddition chemistry to partially (average levels of homopropargylglycine (Hpg) modification = 85 out of 180) glycoconjugate alkynes introduced by acylation of surface lysine side-chains or through the use of unnatural amino acid Hpg.

The plasmid for K16M Qβ was kindly donated by Prof. M. G. Finn (The Scripps Research Institute),[279] and protein Qβ-Hag16, with an olefin uniquely positioned at position 16 was generated by unnatural amino acid incorporation by Dr. Julia Koeppe (University of Oxford). Subsequent self-assembly of Qβ-Hag16 into virus-like particles and purification allowed the creation of Qβ-$(Hag16)_{180}$.

The plasmid for wild-type Qβ (Qβ-WT) was also kindly donated by Prof. M. G. Finn. The protein was expressed, assembled into virus-like particles, and purified by Dr. Julia Koeppe to afford Qβ-WT.

Due to the large size of Qβ-$(Hag16)_{180}$ (2538 kDa) and Qβ-WT, it was necessary to reduce the virus-like particles to monomers before analysis by ESI-MS. This was done using 1 M DTT and 8 M urea, and in addition in some instances the cysteine residues were capped using iodoacetamide. The identity of the protein monomer of Qβ-$(Hag16)_{180}$ was confirmed by ESI-MS (found 14105 Da, calcd. 14105 Da) (Figure 23). ESI-MS for Qβ-WT is shown on pg. 285.

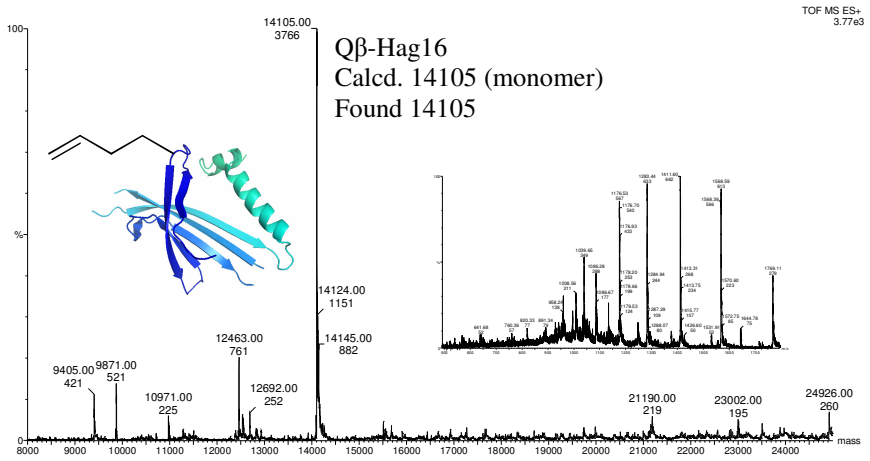



**Figure 23.** Raw (insert) and deconvoluted ESI-MS of Qβ-(Hag16).

Further evidence for the incorporation of Hag was obtained by performing an in-gel trypsin digest of the protein. The precursor ion *m/z* 612.7 $[M+H]^{2+}$ from the peptide corresponding to residues 14-24 of Qβ-Hag16 was selected and fragmented in order to demonstrate the presence of Hag in place of Met at position 16. A number of y and z fragment ions (see Figure 19) of the peptide were detected by ESI-MS, and are shown in the MS/MS spectrum in Figure 24.

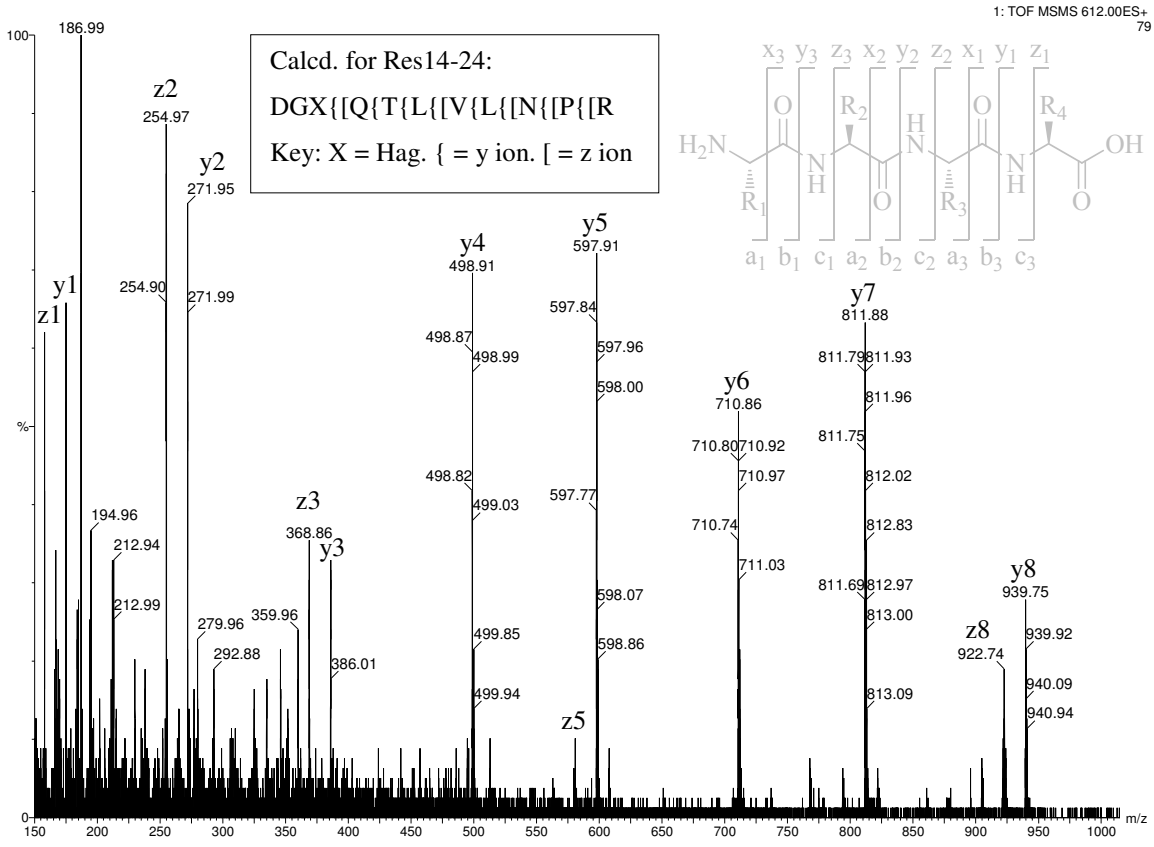



**Figure 24.** MS/MS spectrum of *m/z* 612.7 $[M+H]^{2+}$ from the peptide corresponding to residues 14-24 of Qβ-(Hag16).

The stability of Qβ-$(Hag16)_{180}$ to UV light was demonstrated: protein was added to a cuvette, and was irradiated with UV light for 6 hours. Aliquots were removed from the reaction mixture over the course of the experiment and were analysed by SDS-PAGE (Figure 25).

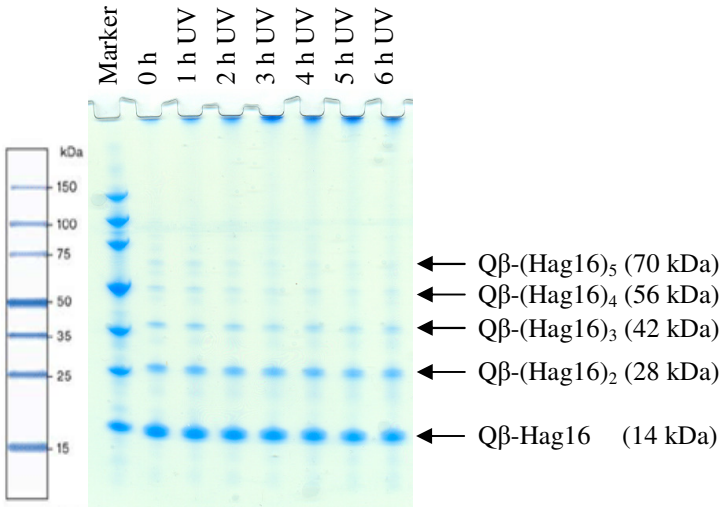


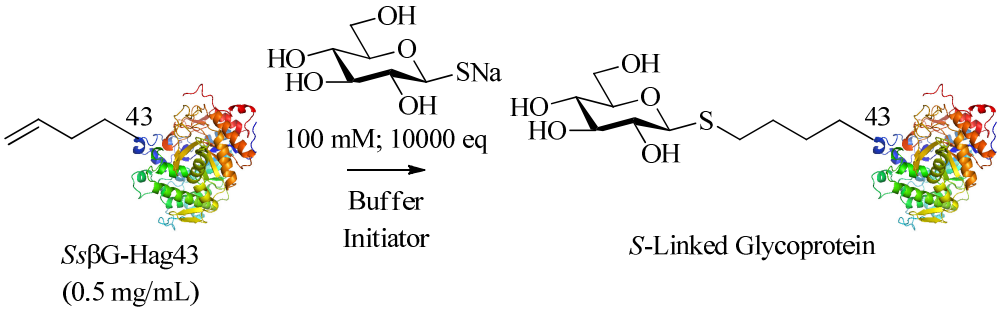

**Figure 25.** Time course analysis of Qβ-$(Hag16)_{180}$ exposed to UV light by SDS-PAGE.

## 9. *S*-linked glycoprotein synthesis

### 9.1. Site-selective glycosylation of *Ss*βG-Hag43

The thiyl glycoconjugation procedure was applied *Ss*βG-Hag43. Use of this protein as a substrate in reactions with β-GlcSH allowed optimisation of the protein chemistry and revealed conditions consistent with those determined in amino acid systems (see Table 9). Together reactions probed the effect of pH (4, 7 or 10) and mode of initiation (hν or Δ, with and without water soluble initiator Vazo44). ESI-MS was used to monitor the reactions to judge % conversion. The fastest complete conversions (> 95%) were seen at pH < 7 through the combined use of Vazo44 and photoinitiation. Importantly, all glycoconjugated *Ss*βG products were demonstrated using the X-Gal colorimetric assay[282] to be functionally (enzymatically) active, thereby confirming the benign nature of the reaction. Time courses for the *Ss*βG-Hag43 glycoconjugations shown in Table 9 are graphically represented in Figure 26.

**Table 9.** Optimisation of *Ss*βG-Hag43 Glycoconjugations.




| **Conditions** | **0.25 M acetate (pH 4)** | **0.25 M sodium phosphate (pH 7)** | **0.25 M CAPS (pH 10)** |
|---|---|---|---|
| **hν** | > 95% (5 h) | 26% (8 h) | NR (8 h) |
| **55°C** | 16% (8 h) | NR (8 h) | NR (8 h) |
| **hν, Vazo44 (0.2 eq)** | > 95% (3 h)[a] | 45% (9 h) | 25% (9 h) |
| **55°C, Vazo44 (0.2 eq)** | > 95% (3 h) | 25% (7 h) | NR (9 h) |

[a] Time course analysis of β-GlcSH conjugation to *Ss*βG-Hag43 under these conditions is shown by ESI-MS in Figure 27; NR = no reaction

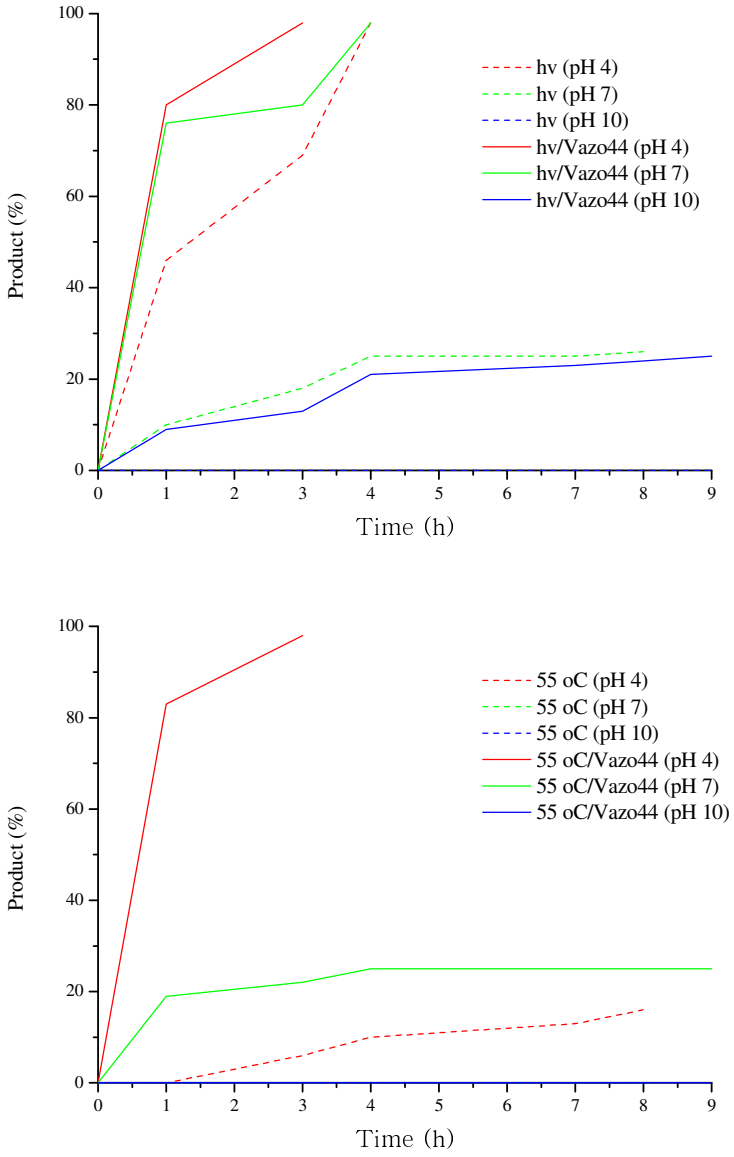



**Figure 26.** Time course of *Ss*βG-Hag43 glycoconjugations in Table 9. Product % was assessed by ESI-MS.

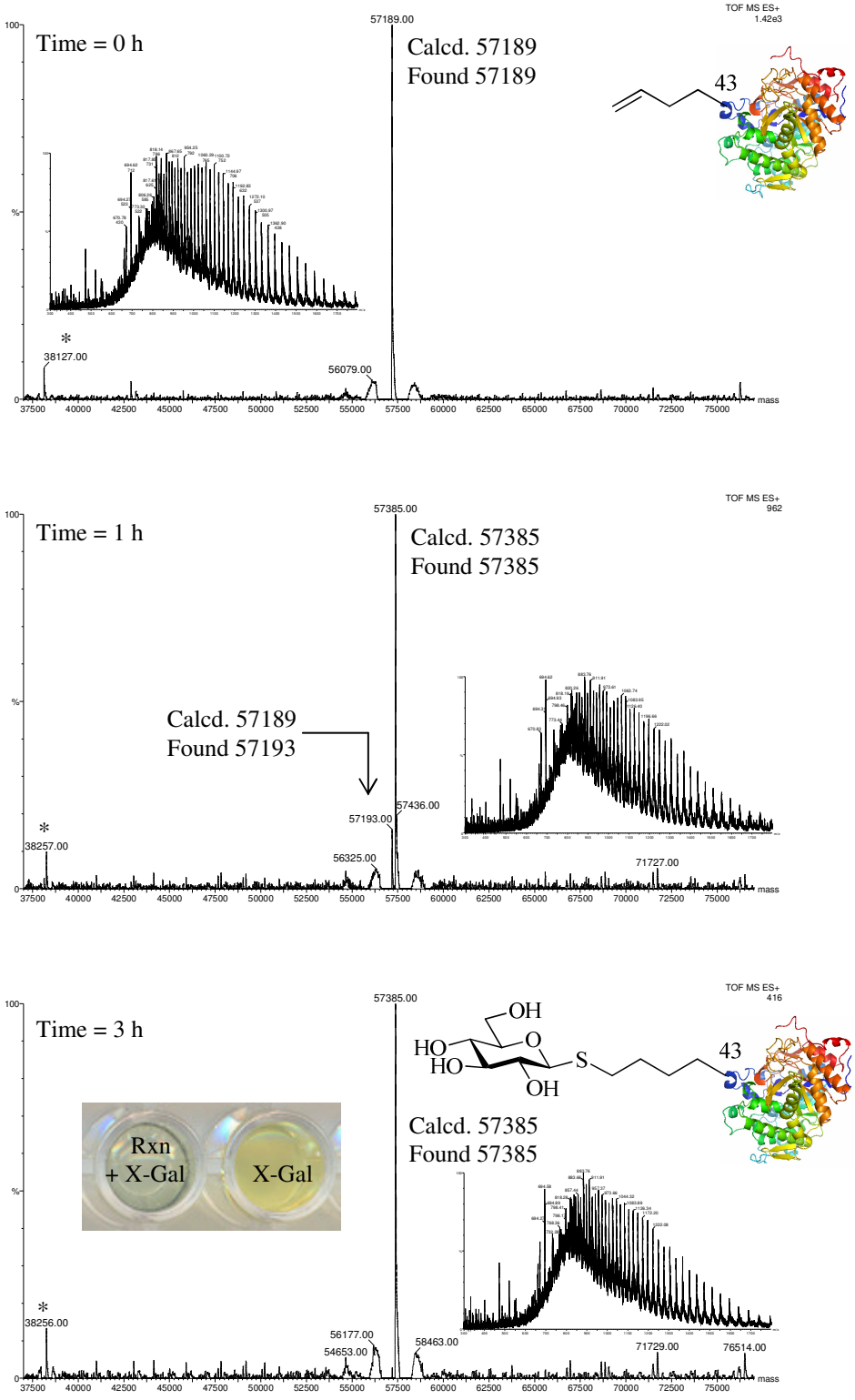
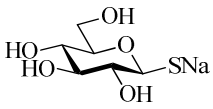
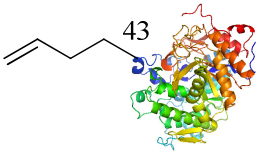
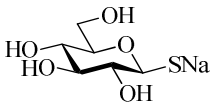
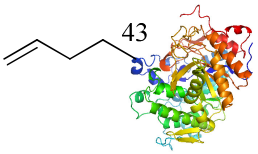
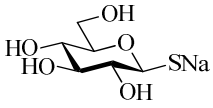
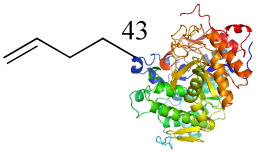
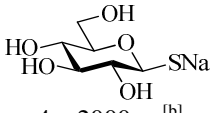
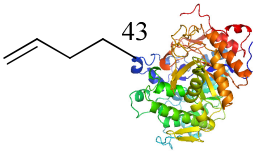
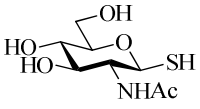
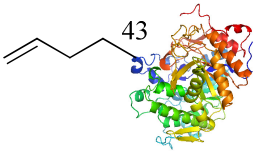
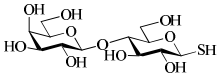
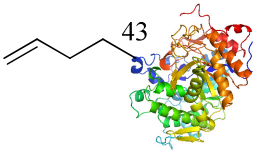
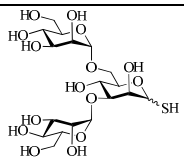
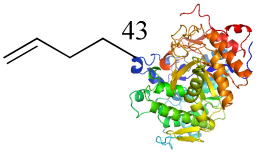
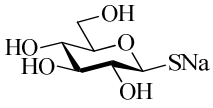
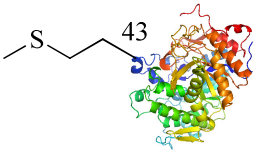


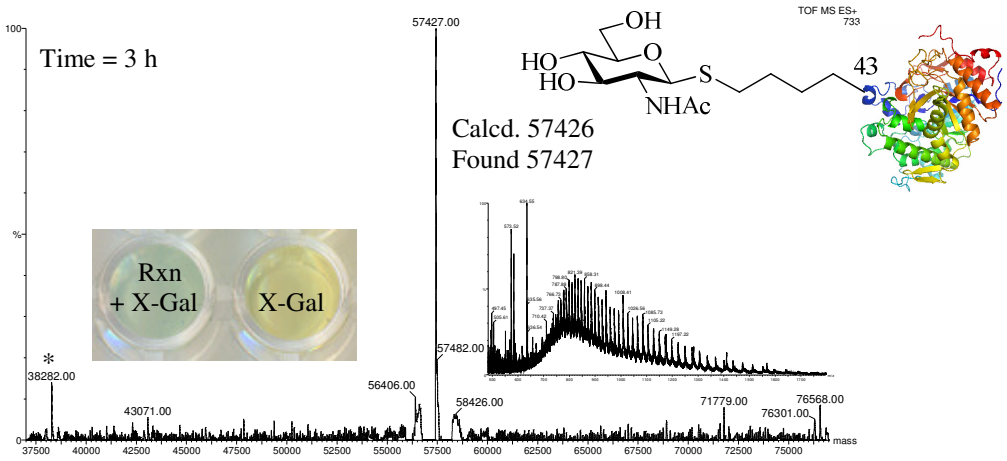

**Figure 27.** Raw (right insert) and deconvoluted ESI-MS for conjugation of β-GlcSH to *Ss*βG-Hag43 in 0.25 M sodium acetate buffer (pH 4) using hν/Vazo44. X-Gal colorimetric assay demonstrating β-galactosidase activity (left insert). * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

Having established conditions for the successful site-selective glycoconjugation to the single olefinic side-chain in *Ss*βG-Hag43, the reaction was optimised further and extension to other glycans was explored (Table 10). It was revealed that, although slightly slower, conjugation of β-GlcSNa to *Ss*βG-Hag43 at pH 6 (entry 2) was equally as successful as at pH 4 (entry 1), yielding > 95% product in 5 hours or less. Although no degradation was seen here with any of the proteins at pH 4, this more mild pH should allow glycoconjugation to even highly acid sensitive proteins. Relatively high concentrations of thiol (100 mM; 10000 eq) also proved important for rapid reaction (entry 1 *vs.* 3), and the portionwise addition of β-GlcSH was shown not to benefit the reaction (entry 4). Rates of reaction for different thiols paralleled those seen in Table 7; for example, conjugation of β-GlcSNa (entry 1 in Table 10) and β-GlcNAcSH (**29**) (entry 5 in Table 10) to *Ss*βG-Hag43 at pH 4 both yielded > 95% product in 3 hours. Notably, the reaction yield obtained for conjugation of disaccharide β-LacSH to *Ss*βG-Hag43 (entry 6 in) was significantly lower (52%) than that obtained for the monosaccharides β-GlcSNa and β-GlcNAcSH (**29**) (> 95%). Moreover, no coupling of the trisaccharide α/β-$Man_3$SH to *Ss*βG-Hag43 (entry 7 in Table 10) was observed. Importantly, use of *Ss*βG-Met43 as a negative control substrate that does not contain Hag, showed no coupling (entry 8), further confirming the chemoselectivity of the reaction for the olefinic 'tag' and the lack of reaction with other amino acids in *Ss*βG. Significantly, all glycoconjugated *Ss*βG products were functionally (enzymatically) active, thereby confirming the benign nature of the reaction.

**Table 10.** *Ss*βG-Hag43 glycoconjugation reactions.

| Entry | Glycosyl Thiol[a] | Protein (0.5 mg/mL) | Solvent | Time | Product [%] |
|---|---|---|---|---|---|
| 1 |  10000 eq |  *Ss*βG-Hag43 | 0.25 M acetate (pH 4) | 3 h | > 95 |
| 2 |  10000 eq |  *Ss*βG-Hag43 | 0.25 M sodium phosphate (pH 6) | 5 h | > 95 |
| 3 |  5000 eq |  *Ss*βG-Hag43 | 0.25 M acetate (pH 4) | 7 h | > 95 |
| 4 |  4 x 2000 eq[b] |  *Ss*βG-Hag43 | 0.25 M acetate (pH 4) | 7 h | > 95 |
| 5 |  10000 eq (**29**) |  *Ss*βG-Hag43 | 0.25 M acetate (pH 4) | 3 h | > 95[c] |
| 6 |  10000 eq (β-LacSH) |  *Ss*βG-Hag43 | 0.25 M acetate (pH 4) | 7 h | 52 |
| 7 |  5000 eq (α/β-$Man_3SH$) |  *Ss*βG-Hag43 | 0.25 M acetate (pH 4) | 8 h | 0 |
| 8 |  10000 eq |  *Ss*βG-Met43 | 0.25 M acetate (pH 4) | 9 h | 0 |

[a] Glycosyl thiol/Vazo44 (1:0.2 eq), hν, room temperature; [b] Portionwise addition; [c] ESI-MS of product is shown in Figure 28



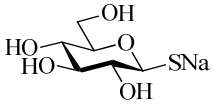
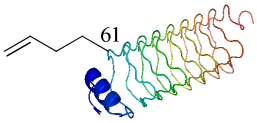
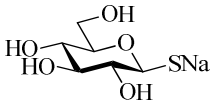
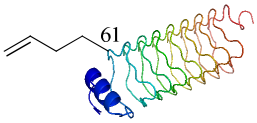
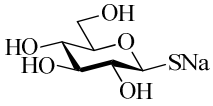
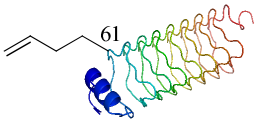

Product: 0 h (0%), 1 h (91%), 3 h (> 95%)

**Figure 28.** Raw (right insert) and deconvoluted ESI-MS for conjugation of β-GlcNAcSH to *Ss*βG-Hag43 in 0.25 M sodium acetate buffer (pH 4) using hv/Vazo44. X-Gal colorimetric assay demonstrating β-galactosidase activity (left insert). * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

## 9.2. Site-selective glycosylation of Np276-Hag61

Having established conditions for the successful site-selective conjugation to the single olefinic side-chain in *Ss*βG-Hag43, the thiyl glycoconjugation procedure was next applied to the cuboid protein Np276-Hag61. Reactions using β-GlcSNa under optimal conditions were even more rapid than for *Ss*βG-Hag43. Again, reactions at pH 4 and pH 6 gave > 95%, whilst those at pH 7 proved sluggish and gave only 40% conversion after 3 hours and did not reach completion.

**Table 11.** Np276-Hag61 glycoconjugation reactions.

| Entry | Glycosyl Thiol[a] | Protein (0.5 mg/mL) | Solvent | Time | Product [%] |
|---|---|---|---|---|---|
| 1 |  8000 eq |  Np276-Hag61 | 0.25 M acetate (pH 4) | 2 h | > 95[b] |
| 2 |  8000 eq |  Np276-Hag61 | 0.25 M sodium phosphate (pH 6) | 3 h | > 95 |
| 3 |  8000 eq |  Np276-Hag61 | 0.25 M sodium phosphate (pH 7) | 3 h | 40 |

[a] Glycosyl thiol/Vazo44 (1:0.2 eq), hν, room temperature; [b] ESI-MS of product is shown in Figure 29.

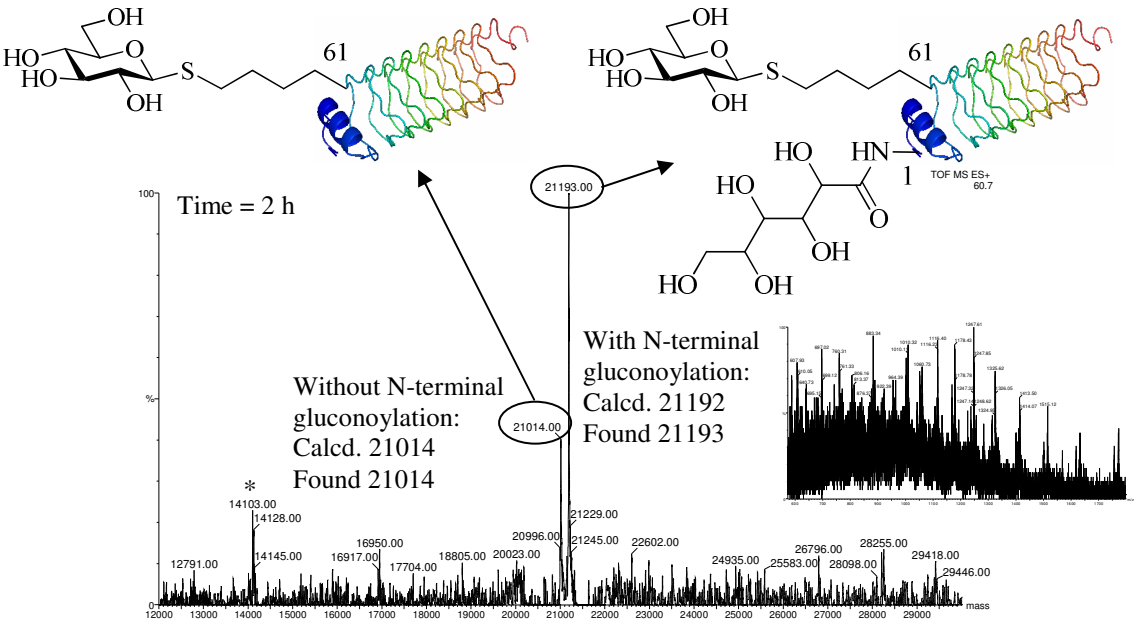
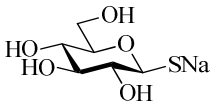
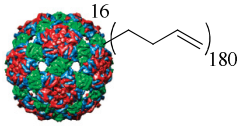
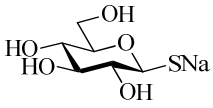
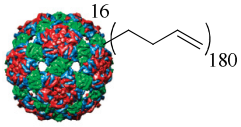
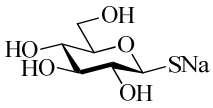
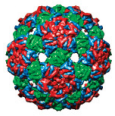



**Figure 29.** Raw (right insert) and deconvoluted ESI-MS for conjugation of β-GlcSH to NP276-Hag61 in 0.25 M sodium acetate buffer (pH 4) using hν/Vazo44. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

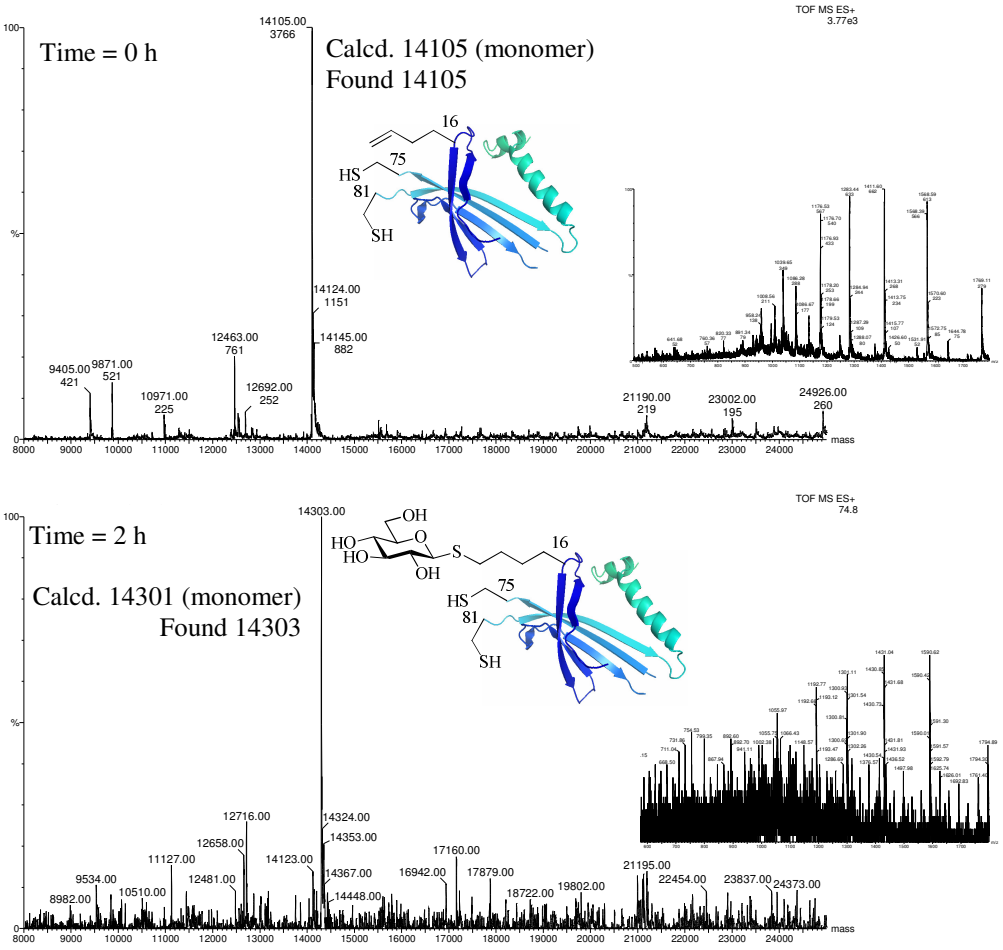
### 9.3. Site-selective glycosylation of Qβ-(Hag16)$_{180}$

The glycoconjugation reaction was extended to the self-assembled multimeric, virus-like particle Qβ-(Hag16)$_{180}$ (Table 12). Glycoconjugation under optimal conditions at either pH 4 or 6 gave rapid formation of product with complete consumption of starting protein suggesting complete glycoconjugation at all 180 sites (entries 1 and 2). Use of a negative control substrate that did not contain Hag, showed no coupling (entry 3). Furthermore, reductive disassembly of the Qβ virus-like particle after reaction and alkylative labelling confirmed that structurally key intramolecular disulfide bonds mediated by Cys75 and Cys81, which covalently link each assembled monomer, remained untouched by the hydrothiolation glycoconjugation reaction (see Figure 30). This further confirmed the excellent chemoselectivity of the reaction for the Hag olefinic site of reaction, even with cysteine containing proteins. The Qβ particle is composed of 180 copies a coat protein assembled into a icosahedral virion of average diameter 27 nm.[296] To date, this precise and complete site-selective glycoconjugation at 180 sites in a 'protein nanoparticle' is unprecedented[295] and may find utility in the creation of precise glycoconjugate vaccines in which high levels of loading may be achieved whilst precisely controlling the position of epitope display.[281, 297, 298]

**Table 12.** Qβ-(Hag16)$_{180}$ glycoconjugation reactions.

| Entry | Glycosyl Thiol[a] | Olefinic Protein (0.5 mg/mL) | Solvent | Time | Product [%] |
|---|---|---|---|---|---|
| 1 |  5000 eq |  Qβ-(Hag16)$_{180}$ | 0.25 M acetate (pH 4) | 2 h | > 95[b] |
| 2 |  5000 eq |  Qβ-(Hag16)$_{180}$ | 0.25 M sodium phosphate (pH 6) | 3 h | > 95 |
| 3 |  5000 eq |  Qβ-WT | 0.25 M acetate (pH 4) | 8 h | 0 |

[a] Glycosyl thiol/Vazo44 (1:0.2 eq), hν, room temperature; [b] ESI-MS of product is shown in Figure 30.



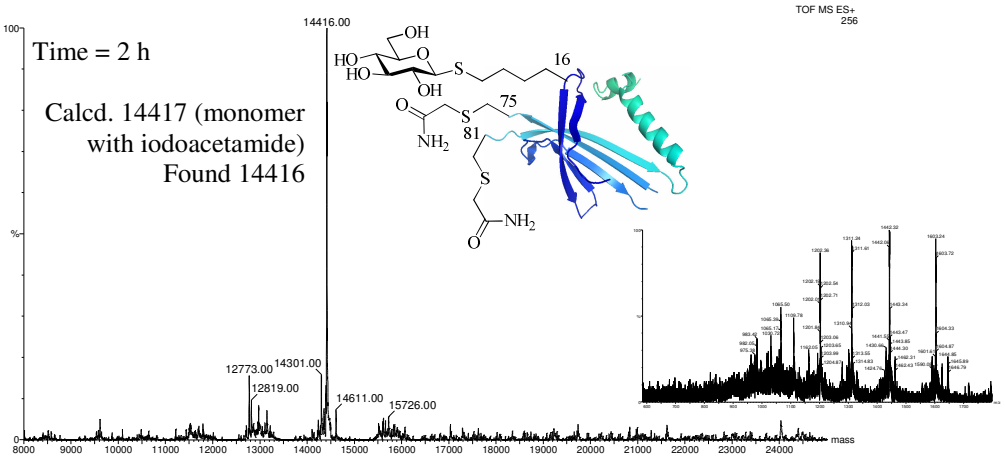



**Figure 30.** Raw (right insert) and deconvoluted ESI-MS for conjugation of β-GlcSH to Qβ-Hag16 in 0.25 M sodium acetate buffer (pH 4) using hν/Vazo44.

### 9.3.1. Ionisation efficiency

To justify conversions of > 95% for the Qβ-$(Hag16)_{180}$ glycoconjugation reactions, the ESI ionisation efficiency of non-glycosylated Qβ-Hag16 was shown to be greater than glycosylated Qβ-Hag16; Non-glycosylated and glycosylated Qβ-Hag16 at varying concentrations (determined using the BCA assay[299]) were analysed by ESI-MS in positive ion mode on a Micromass LCT mass spectrometer interfaced with a Waters 2790 Alliance HT HPLC using a Phenomenex Jupiter 5u C4 300A 250 x 4.6 mm column, as described on pg. 211 (Figure 116). The data were processed as follows: a mass spectrum of multiply charged ions was generated by combining data from the chromatogram between 14–18 minutes. The spectrum from *m/z* 800–1800 was displayed and the total ion count of the largest peak was plotted against protein concentration (Figure 31i). A deconvoluted mass spectrum was then generated using the Maximum Entropy method (MaxEnt1) with 20 iterations. The total ion count of the peak corresponding to Qβ-Hag16 protein was plotted against protein concentration (Figure 31ii).

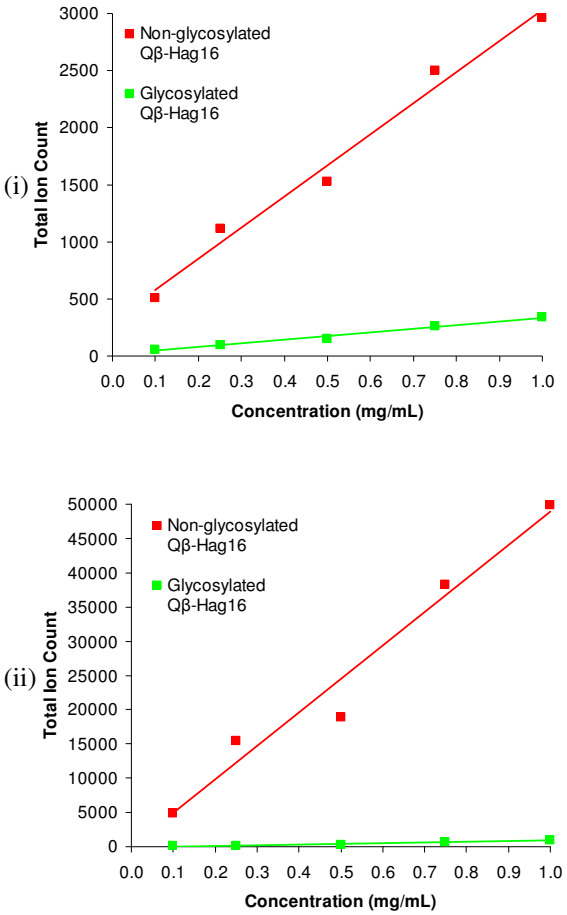



**Figure 31.** Protein concentration of Qβ-Hag16 (non-glycosylated and glycosylated) *vs.* total ion count of (i) the largest peak from mass spectrum of multiply charged ions and (ii) the peak corresponding to Qβ-Hag16 protein from deconvoluted mass spectrum.

## 10. Conclusion and future work

In summary, an effective protein glycoconjugation method based on hydrothiolation chemistry has been developed that allows rapid and complete reaction under mild conditions (ambient temperature, pH 4–6) that have been shown to be compatible with a representative selection of different proteins. Combined use of this reaction with an olefinic "tag" that may be positioned using the non-natural amino acid Hag allows site-selective glycoconjugation at pre-determined sites using chosen sugars and

proteins through a "tag–modify"[59] approach (Hag "tag"-hydrothiolation "modify"). This reaction is compatible with all other natural amino acids, which are inert under these conditions; this chemoselectivity should allow its use in orthogonal, multi-modification approaches.[135] In addition, the glycosyl thiols needed as glycoconjugation partners may be readily accessed in a number of ways, even directly from unprotected reducing sugars.[300] Here, the reaction has been applied to proteins with different structures, which retain full function, for precise, site-selective glycoconjugation at single and even high levels of multiple sites.

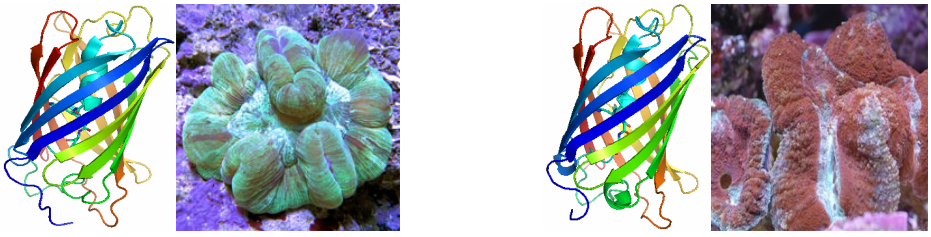
In the future, the potential of this protein glycoconjugation method as a way to create glycoconjugate immunogens as putative vaccines (see pg. 41) should be explored. However, optimisation of this reaction with larger glycans may be necessary prior to this. In addition, a glycoprotein synthesised by this method should be digested and analysed by MS/MS in order to provide further evidence for the chemoselectivity of the reaction for the olefinic 'tag' and the lack of reaction with other amino acids.

# Chapter 3: An Unexpected Photo-Induced Site-Specific Protein Cleavage Reaction

## 1. Introduction

Unexpectedly, while investigating the photochemical modification of β-glycosidase from *Sulfolobus solfataricus* (*Ss*βG) discussed in chapter 2, a photo-induced site-selective protein cleavage reaction was discovered (see pg. 80).

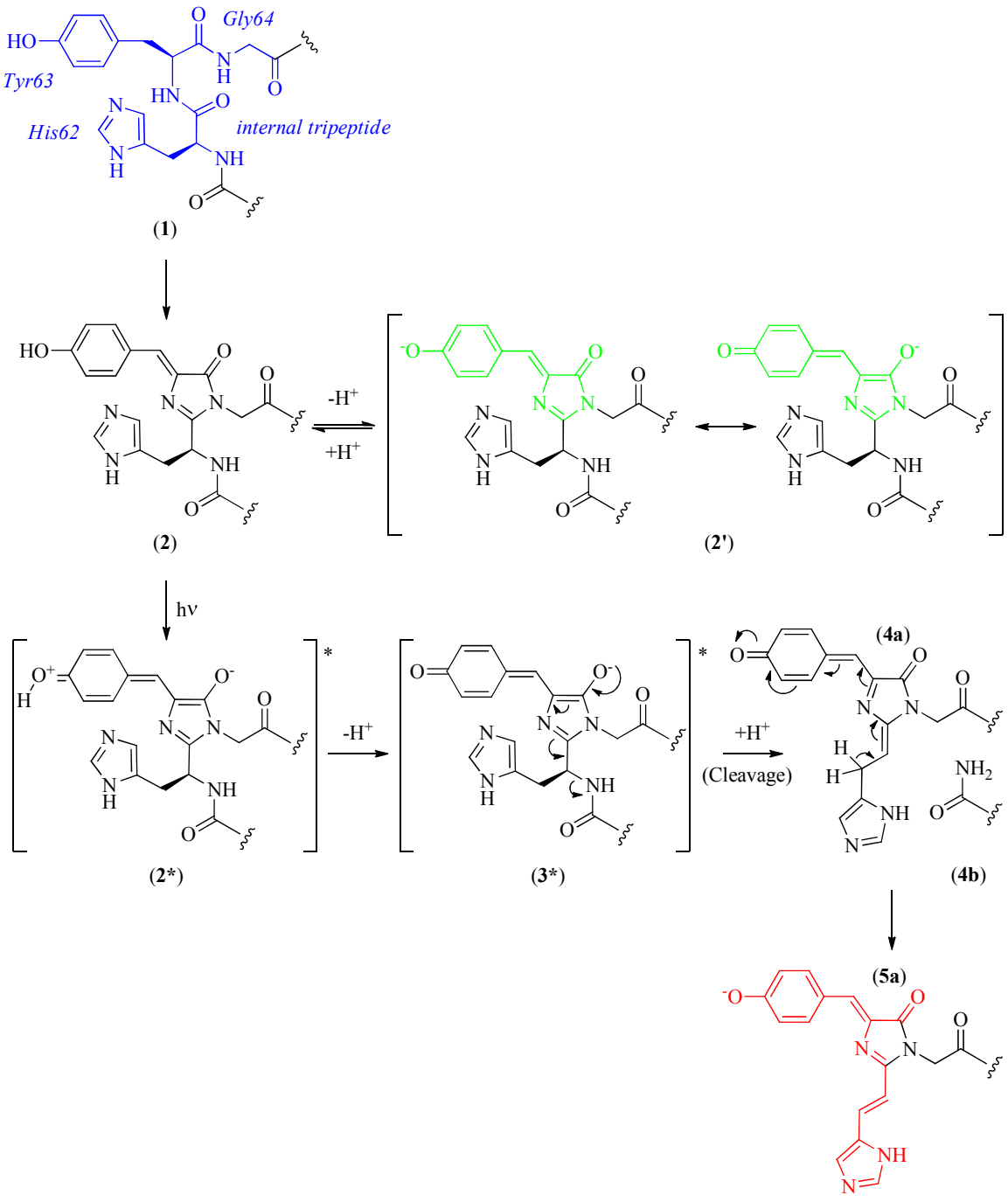
Although proteins are known to be affected by exposure to UV irradiation in processes ranging from amino acid side-chain degradation, the reduction of disulfide bridges,[286, 287] aggregation[288] and even to consequent loss of biological activity,[289, 290] natural examples of clean, site-selective photocleavage are rare. To date, the only reported examples are the coral fluorescent proteins Kaede from *Trachyphyllia geoffroyi* and EosFP from *Lobophyllia hemprichii* (Figure 32).[301-303] Irradiation of Kaede and EosFP using UV light in the range of 350–410 nm results in site-selective cleavage of their peptide backbone between Phe61-His62. This cleavage is marked by green to red photoconversion and requires prior formation of a specific chromophore.



Kaede from *Trachyphyllia geoffroyi* EosFP from *Lobophyllia hemprichii*

**Figure 32.** Crystal structures of Kaede from the coral *Trachyphyllia geoffroyi* and EosFP from the coral *Lobophyllia hemprichii* (PDB IDs: 2GW3 and 1ZUX, respectively).

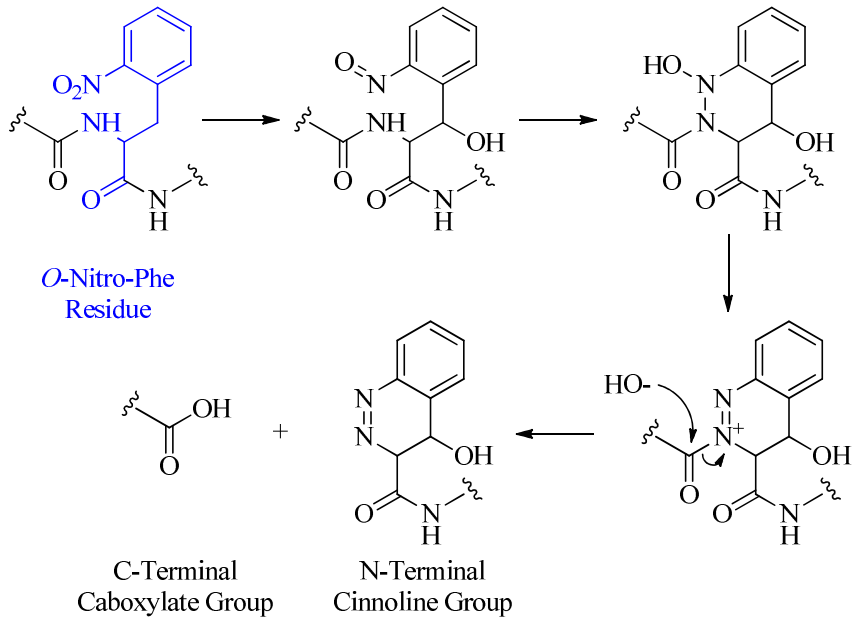
Miyawaki *et al.*[303, 304] have proposed the mechanism shown in Scheme 33 for the photo-induced cleavage of Kaede, and a similar mechanism for EosFP has also been reported.[305]




**Scheme 33.** Proposed mechanism for the photo-induced cleavage of Kaede from *Trachyphyllia geoffroyi*.

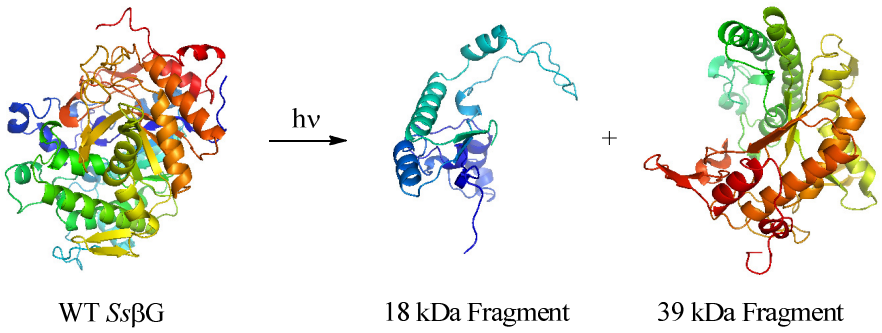
In the first step, the neutral and non-fluorescent chromophore (**2**) is formed autocatalytically from the internal tripeptide His62-Tyr63-Gly64. Subsequent deprotonation of Tyr63 results in the green fluorescent species (**2'**). After excitation of (**2**) by UV light, the excited state (**2***) releases a proton to form the excited intermediate (**3***). Cleavage then occurs at the Nα-Cα bond of His62 to eliminate the carboxamide group containing peptide (**4b**). The subsequent loss of a proton from His62-Cβ gives a *trans* double bond between His62-Cα and -Cβ, leading to the extension of the conjugated π system and formation of the red fluorescent species (**5a**).

While a number of site-selective protein photolytic methods requiring the use of a cleaving agent have been investigated, these so far are poorly selective and low yielding.[306-317] Most recently, Schultz *et al.* have demonstrated that the peptide backbone of T4 lysozyme can be selectively photocleaved using 365 nm UV light in yields of up to ~30% (determined using optical densitometry) if the unnatural amino acid 2-nitrophenylalanine (*o*-nitro-Phe) is incorporated.[318-320] This reaction proceeds via an unusual cinnoline forming reaction to generate a C-terminal carboxylate group and N-terminal cinnoline group (Scheme 34).




**Scheme 34.** Proposed mechanism for the photo-induced cleavage of *o*-nitro-Phe containing proteins.

General, clean, photo-induced cleavage reactions have a strong potential biotechnological utility in protein preparation[321] and proteomic analysis.[322, 323] They are also of fundamental interest in the light of possible biologically relevant function, as is the case for Kaede and EosFP. For these reasons, the photo-induced cleavage reaction of *Ss*βG, a family 1 glycosylhydrolase (GH1) protein, was investigated (Scheme 35).




**Scheme 35.** Photo-induced cleavage of β-glycosidase from *Sulfolobus solfataricus* (*Ss*βG).

This chapter begins by looking at the photo-induced cleavage of M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG from chapter 2, a Hag-deficient mutant of this protein (M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG) and wild-type *Ss*βG (WT *Ss*βG) by SDS-PAGE and ESI-MS, and describes for example how the range of wavelengths of UV light responsible for the reaction were determined. In addition, this chapter describes the identification of the *Ss*βG cleavage site, the identification of which residues are essential for the reaction to occur and the extensive analyses of the daughter fragments that allowed a reaction mechanism to be proposed. Also covered is how the generality of the photocleavable motif was investigated and how a homologous scission was identified to occur in a number of other GH1 proteins. Finally, the bioinformatic and phylogenetic analyses that were carried out in an attempt to elucidate any potential biological function of this photochemical reaction are detailed.

## 2. Photo-induced cleavage reaction

### 2.1. M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG

As mentioned in chapter 2 (see pg. 80), the stability of M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG to UV light was tested. Protein was added to two cuvettes. One solution was allowed to stand at room temperature, whilst the other solution was irradiated with UV light at room temperature. Aliquots were removed from the reaction mixtures over the course of the experiment and were analysed by SDS-PAGE (Figure 34) and ESI-MS (Figure 33). Unexpectedly, it was discovered that the protein cleaved to give two fragments when irradiated with UV light. No cleavage was observed in the absence of UV light (Figure 34, right).

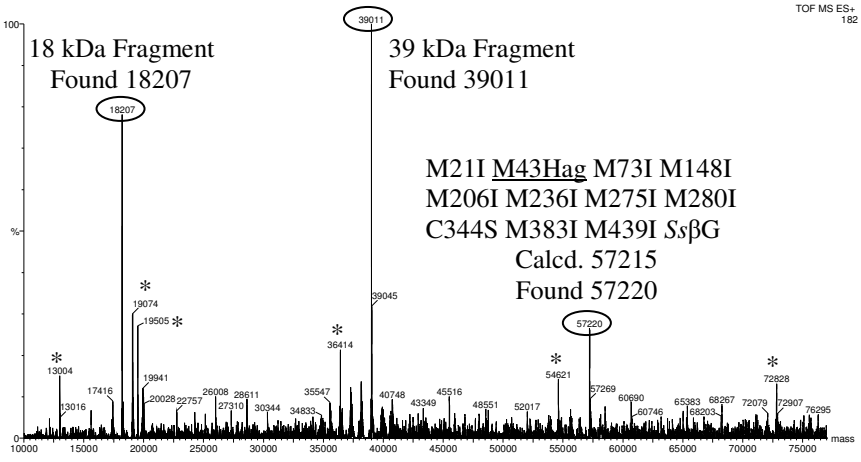



**Figure 33.** Deconvoluted ESI-MS of M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG irradiated with UV light for 1 h. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

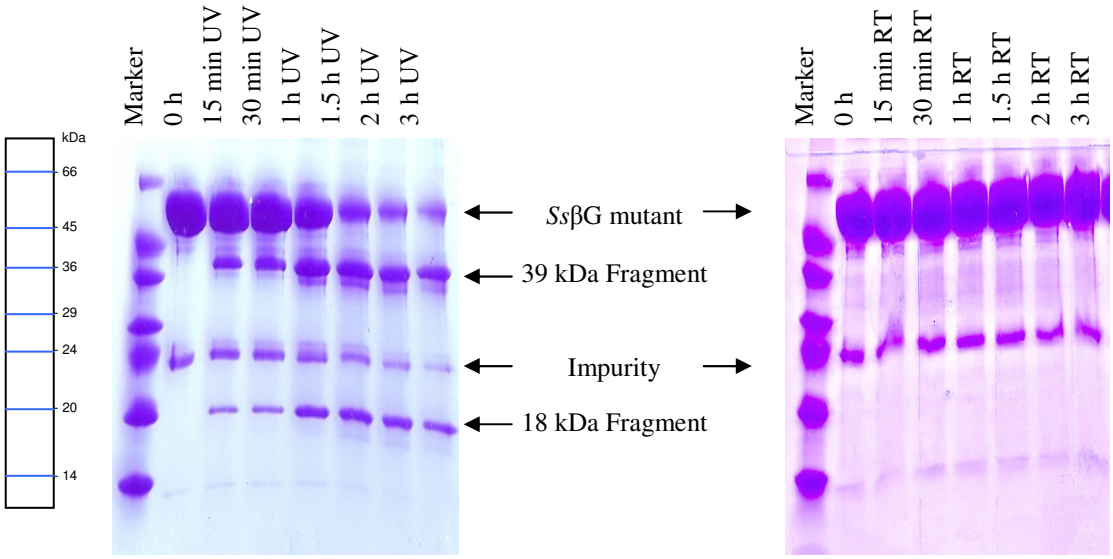



**Figure 34.** Time course analysis of M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG exposed to/not exposed to UV light by SDS-PAGE.

It should be noted that: (i) The impurity shown in Figure 34 was never identfied; and (ii) M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I SsβG was expressed again and purified so as not to contain the impurity. However, the protein still cleaved cleaved to give two fragments when irradiated with UV light.

## 2.2. M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG

M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG, a Hag-deficient mutant of M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG, was generated in order to establish if the Hag residue was responsible for the photocleavage reaction.

N-terminal His-tagged M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG was expressed and purified, and the protein identity confirmed by ESI-MS (found 57235 Da, calcd. 57235 Da) (see pg. 289). In addition, the enzymatic activity of the protein was demonstrated using the X-Gal colorimetric assay.[282]

The stability of M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG to UV light was tested in the same way as for M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG. As before, protein was added to two cuvettes. One solution was allowed to stand at room temperature, whilst the other solution was irradiated with UV light. Aliquots were removed from the reaction mixtures over the course of the experiment and were analysed by SDS-PAGE (Figure 35) and ESI-MS (Figure 36) (Note: In instances where a protein cleavage reaction has been conducted more than once, an SDS-PAGE gel for each reaction has been shown in this thesis.). A synonymous cleavage was found to occur when the protein was irradiated with UV light, which clearly demonstrated that the Hag residue was not responsible for protein cleavage. Again, no cleavage was observed in the absence of UV light (Figure 35). The bands corresponding to protein on the gel marked with a ■ were quantified by densitometry (see pg. 320).

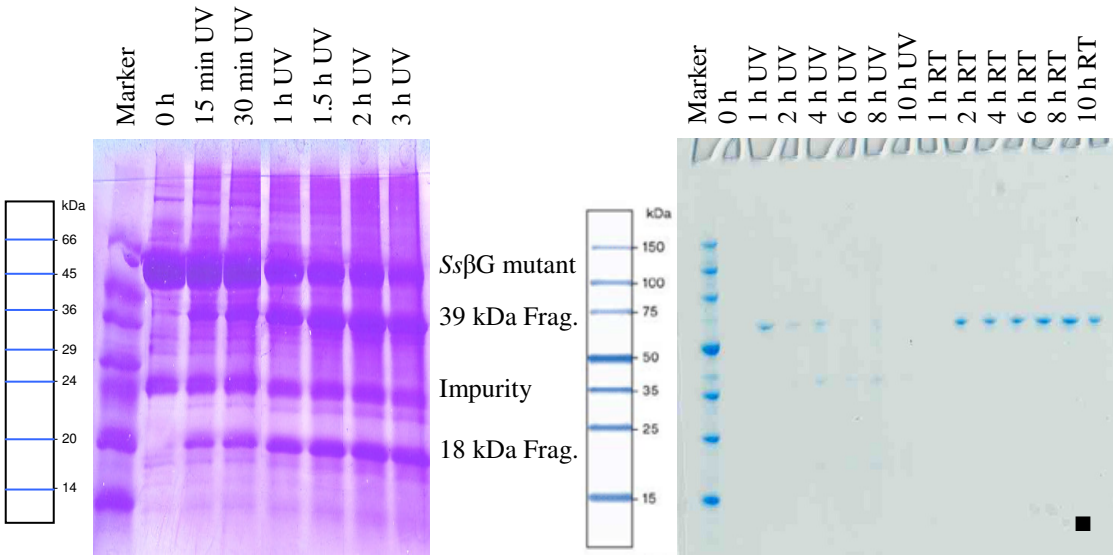



**Figure 35.** Time course analyses of M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG exposed to/not exposed to UV light by SDS-PAGE.

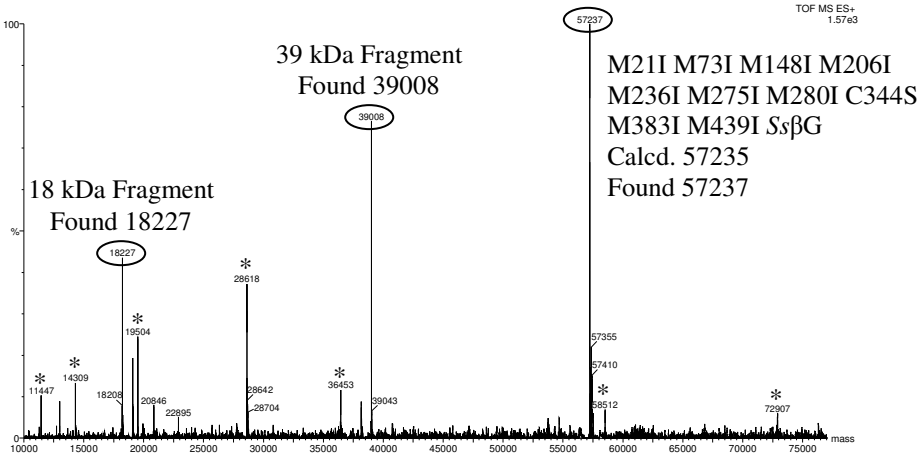



**Figure 36.** Deconvoluted ESI-MS of M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG irradiated with UV light (30 min). * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

## 2.3. Wild-type *Ss*βG

Wild-type β-glycosidase from *Sulfolobus solfataricus* (WT *Ss*βG) was employed as the model protein on which to study the photo-induced site-selective cleavage reaction.

N-terminal His-tagged WT *Ss*βG was expressed and purified, and the protein identity confirmed by ESI-MS (found 57414 Da, calcd. 57414 Da) (see pg. 289). Additionally, the enzymatic activity of WT *Ss*βG was demonstrated using the X-Gal colorimetric assay.[282]

In the same way as M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG (see pg. 80), WT *Ss*βG was found to dissociate when irradiated with UV light to form two daughter fragments that were sharp and clean by SDS-PAGE (Figure 37). This suggested that protein cleavage occured at a single site, since indiscriminate cleavage would result in diffuse PAGE smears. No cleavage was observed in the absence of UV light (Figure 37). In addition, the bands corresponding to protein on the gel marked with a ■ were quantified by densitometry (see pg. 320).

The masses of the cleanly formed fragments were determined by ESI-MS to be 18281 and 39133 Da (Figure 38), which combined equalled the mass of WT *Ss*βG (57414 Da). Note that impuritiy bands on the SDS-PAGE gels below were never identified, but may be fragments or dimers of the 18 and 39 kDa fragments.

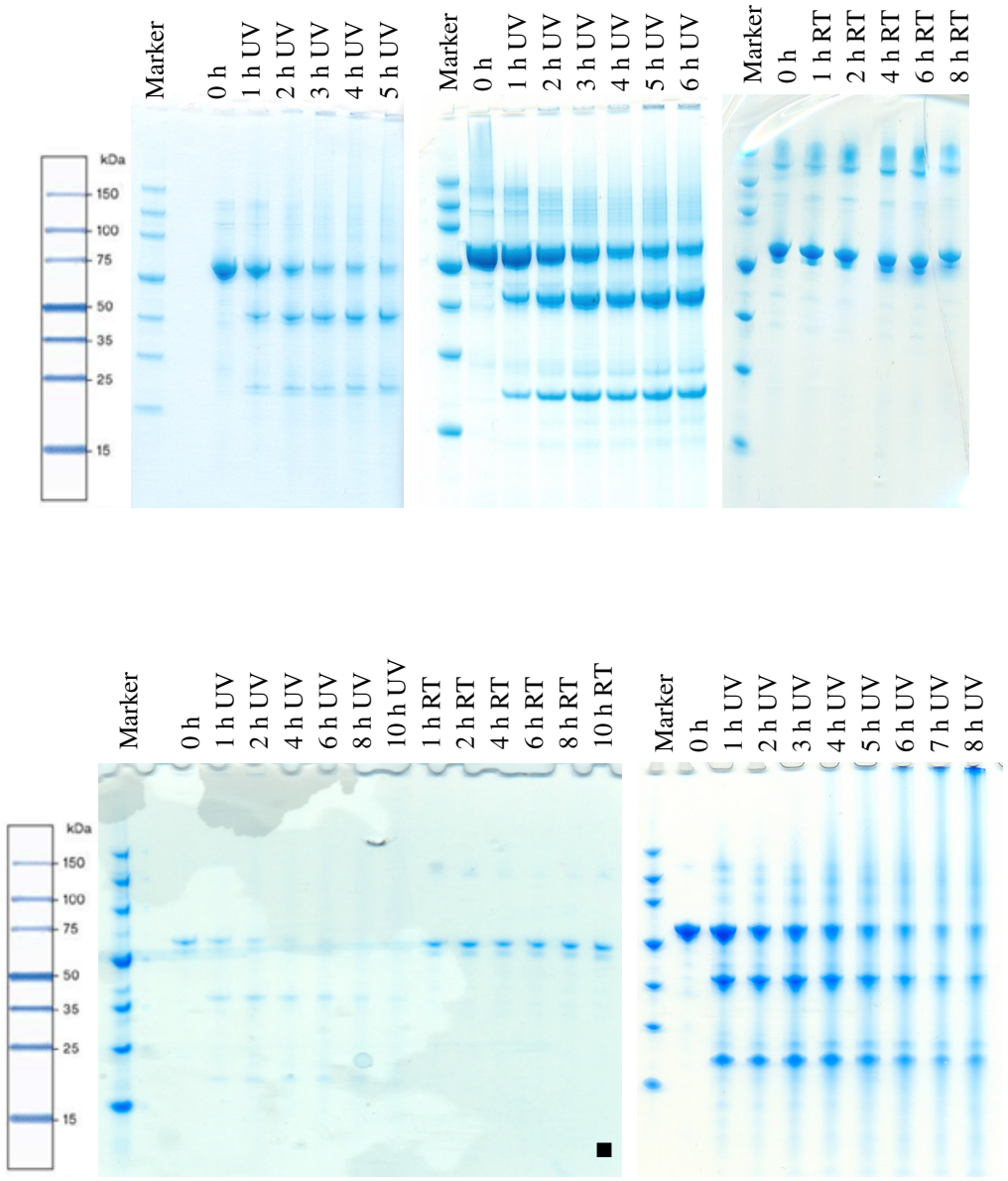



**Figure 37.** Multiple time course analyses of WT *Ss*βG exposed to/not exposed to UV light by SDS-PAGE.

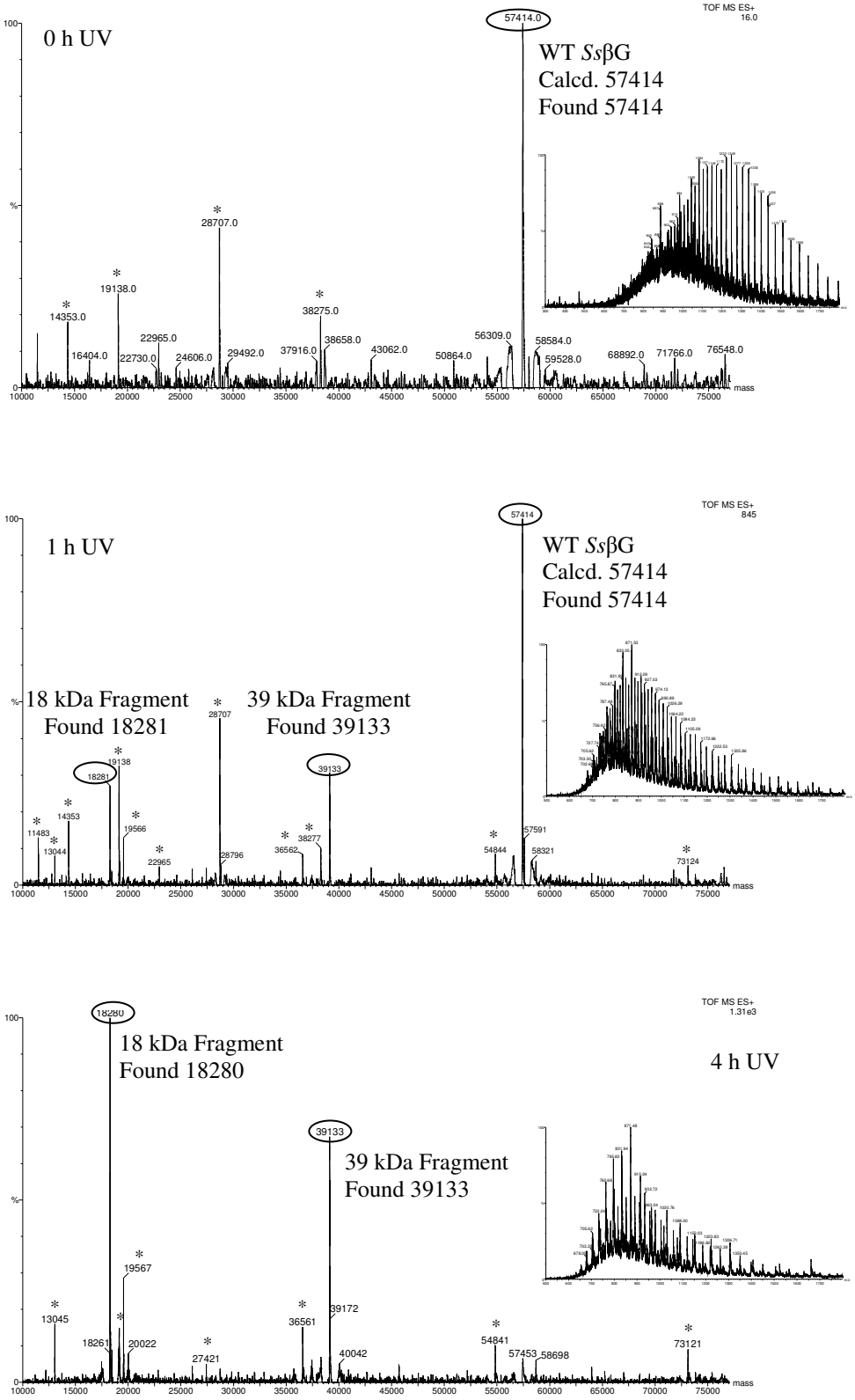



**Figure 38.** Raw (inert) and deconvoluted ESI-MS showing time course analysis of WT *Ss*βG irradiated with UV light. * = Artifact of the deconvolution algorithm.[283-285]

The intracellular photo-induced cleavage of WT *Ss*βG was attempted. This was carried out by irradiating cell pellet collected from the expression of WT *Ss*βG with UV light for 4 hours. The cells were then lysed by sonication and the lysate was purified using nickel affinity chromatography. All eluent collected from the purification was analysed by SDS-PAGE, however, only noncleaved WT *Ss*βG was observed. This suggested that the β-glycosidase protein may also be stable to UV irradiation when present inside *Sulfolobus solfataricus.*

## 3. Reactive transitions

To identify the wavelengths of light responsible for the photo-induced cleavage of WT *Ss*βG, the protein was irradiated in a quartz cuvette with transmission range 200–2500 nm and a glass cuvette with transmission range 360–2500 nm for 4 hours. Aliquots were removed from reaction mixtures during the course of the reaction and were analysed by SDS-PAGE (Figure 39). Cleavage of WT *Ss*βG was found to occur at wavelengths in the range of 200–360 nm.

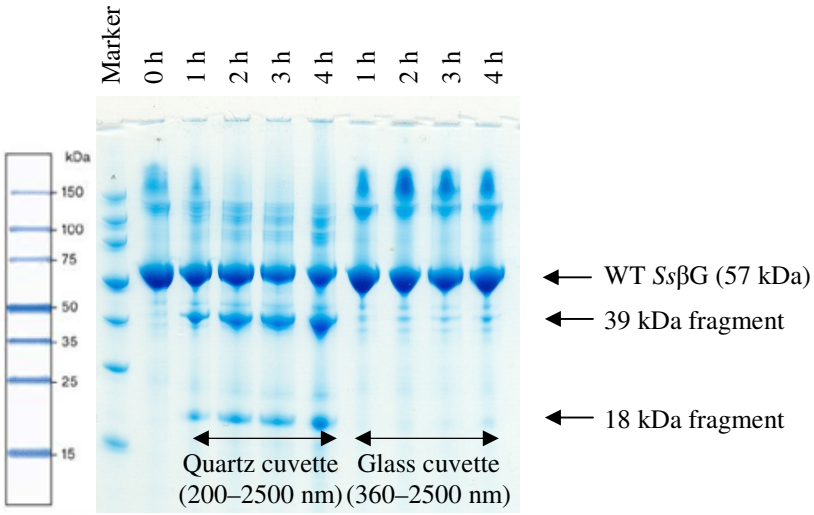



**Figure 39.** Time course analysis of the photo-induced cleavage of WT *Ss*βG using a quartz cuvette (left) and a glass cuvette (right) by SDS-PAGE.

To probe the reactive transitions further, precise frequency light (193–355 nm) was generated using a neodymium-doped yttrium aluminium garnet (Nd:YAG)-pumped tuneable dye laser, except in the case of 193 nm where an ArF-excimer laser was used and 355 nm where the frequency was tripled direct from the YAG. A range of dyes were employed to allow specific wavelengths spanning the range 222.5–355 nm to be obtained. This experiment was carried out with the help of Mr. Chris Eyles and Prof. Mark Brouard (University of Oxford), in the laboratory of Prof. Mark Brouard. Cleavage of WT *Ss*βG was found to occur at wavelengths in the range of 240–308 nm (Figure 40). The wavelengths that WT *Ss*βG was found to cleave at correspond well with the absorption spectrum of WT *Ss*βG (Figure 60), *i.e.* cleavage occurs at wavelengths that the protein absorbs strongly at.

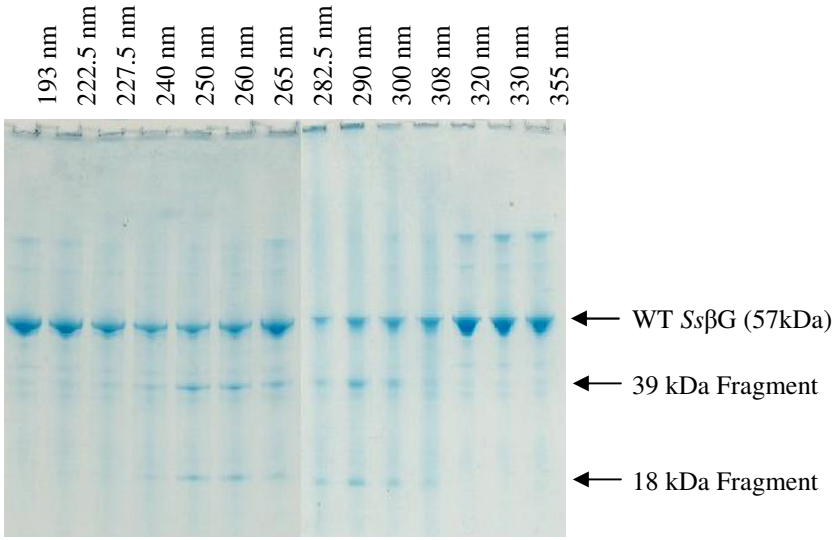



**Figure 40.** SDS-PAGE gel showing the photo-induced cleavage of WT *Ss*βG at varying wavelengths.

## 4. Cleavage site identification

The cleavage site in WT *Ss*βG was determined using N-terminal sequencing in conjunction with ESI-MS; WT *Ss*βG was irradiated with UV light for 2 hours to provoke fragmentation of the protein. The sample was then submitted to the Protein Characterisation Facility at the University of Oxford for N-terminal sequencing as a mixture containing WT *Ss*βG, the 18 kDa fragment and the 39 kDa fragment, which prior to N-terminal sequencing were separated by SDS-PAGE. N-terminal sequencing using classical Edman degradation[324] with phenylisothiocyanate revealed the 39 kDa fragment to be "blocked" by an unnatural, unreactive residue and the 18 kDa fragment to share the sequence found at the N-terminus of *Ss*βG (GHHHHH-).

Using the information gained from N-terminal sequencing in combination with the known protein sequence of WT *Ss*βG and ESI-MS data already obtained (Figure 38) [18 kDa fragment - 18281 Da, 39 kDa fragment - 39133 Da], the cleavage site was identified to lie between residues His150 and Trp151 (Figure 41).[325] The calculated masses of the N-terminal fragment (residues 1-150) and C-terminal fragment (residues 151-489) fitted well with the found masses; N-terminal fragment calcd. 18282 Da, found 18281 Da. C-terminal fragment calcd. 39150, found 39133 Da.

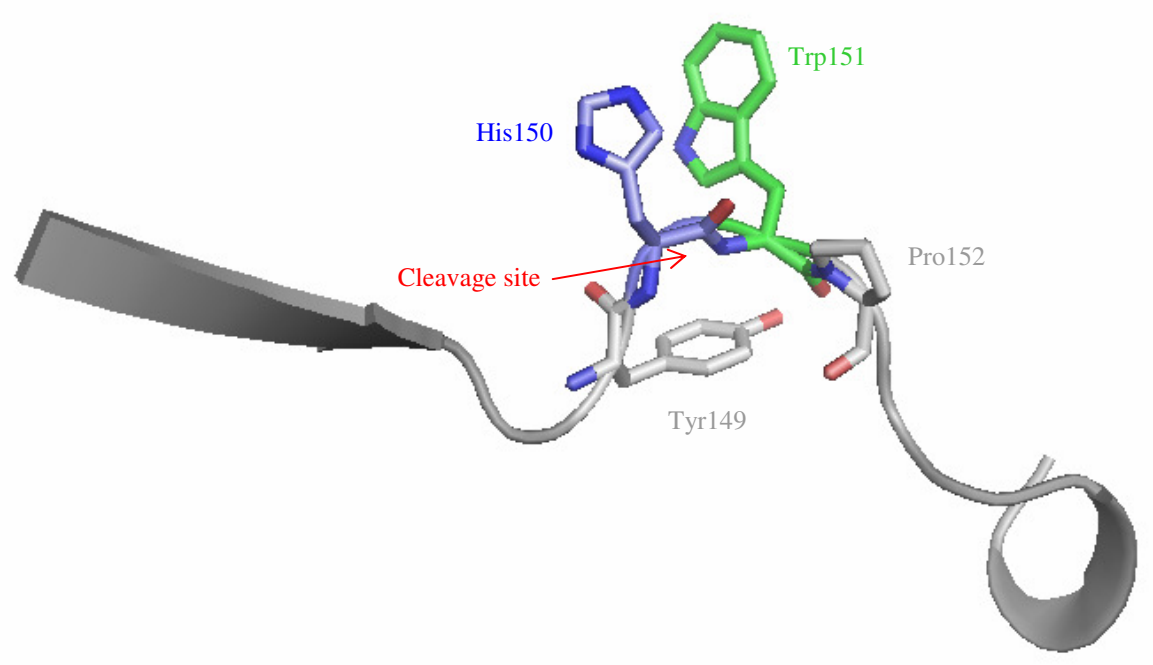


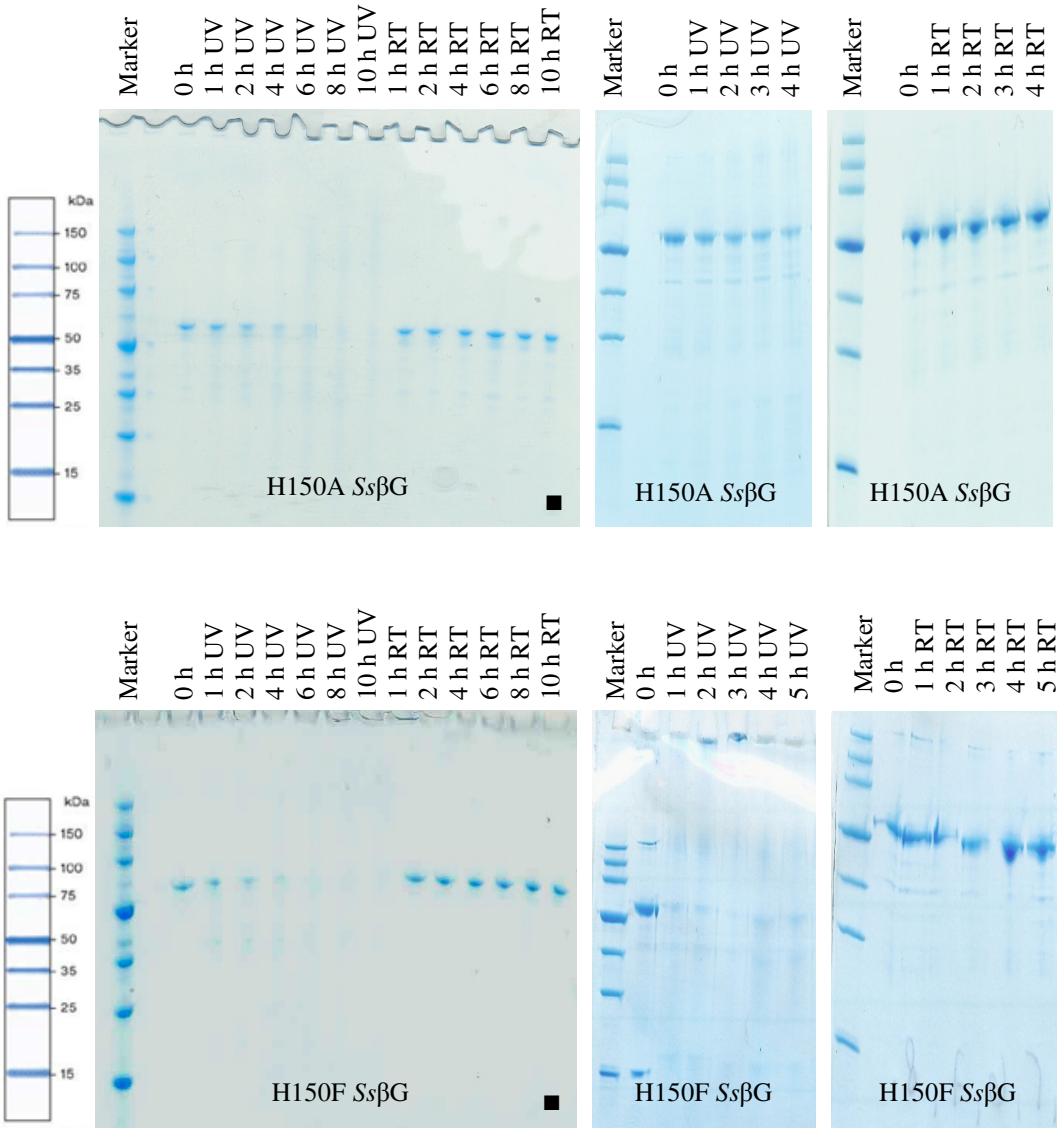

**Figure 41.** Cleavage site of *Ss*βG (PBD ID: 1GOW).

## 5. Probing the Cleavage Site

To gain additional evidence for the location of the cleavage site (His150-Trp151) (Figure 41), the *Ss*βG mutants H150A *Ss*βG, H150F *Ss*βG and W151A *Ss*βG, containing alterations at this junction were generated and tested for photocleavage activity.

Starting from a pET24d plasmid encoding WT *Ss*βG, site-directed mutagenesis was used to generate pET24d plasmids encoding H150A *Ss*βG, H150F *Ss*βG and W151A *Ss*βG. The N-terminal His-tagged proteins (H150A, H150F and W151A *Ss*βG) were then expressed and purified, and the protein identities confirmed by ESI-MS (see pg. 289); H150A *Ss*βG calcd. 57348 Da, found 57348 Da. H150F *Ss*βG calcd. 57424 Da, found 57424 Da. W151A *Ss*βG calcd. 57299 Da, found 57296 Da. Additionally, the enzymatic activities of the proteins were demonstrated using the X-Gal colorimetric assay.[282]

The proteins (H150A, H150F and W151A *Ss*βG) were tested for photocleavage activity, which was assessed for using SDS-PAGE (Figure 43). The proteins remained intact and did not fragment, giving additional evidence for the cleavage site and highlighting the need for specific residues and not simply for potential functional group equivalents (*e.g.* aromatic His150 → Phe). No cleavage was observed for any of the proteins in the absence of UV light (Figure 43). The bands corresponding to protein on the gels marked with a ■ were quantified by densitometry (see pg. 320).



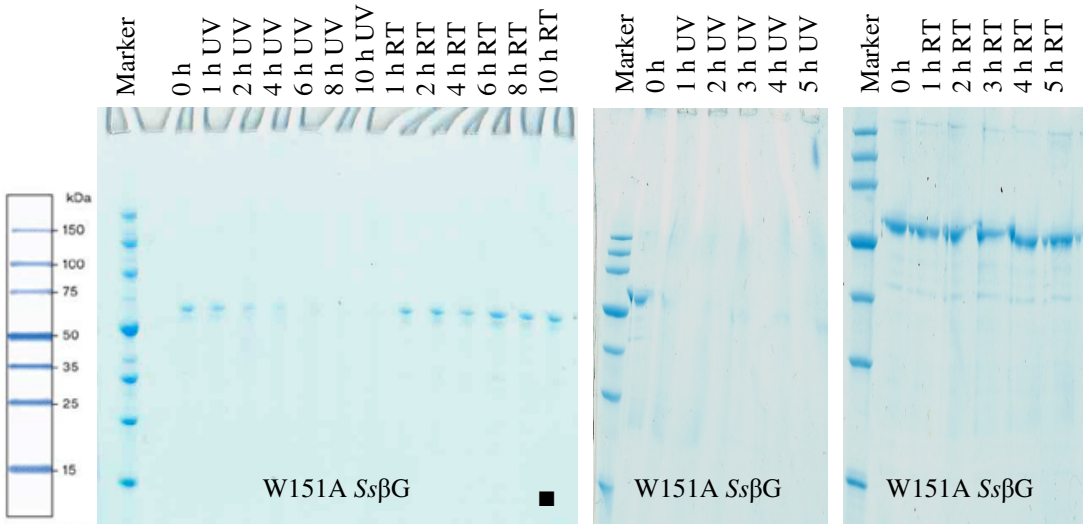



**Figure 42.** Time course analyses of H150A *Ss*βG, H150F *Ss*βG and W151A *Ss*βG exposed to/not exposed to UV light by SDS-PAGE.

Production of W151F *Ss*βG, another WT *Ss*βG mutant containing an alteration at the cleavage junction, was also desired in order to allow its photocleavage activity to be tested. In the same way as for H150A, H150F and W151A *Ss*βG, pET24d plasmid encoding W151F *Ss*βG was generated. Expression of W151F *Ss*βG was attempted, but despite employing a number of different expression conditions (LB media, LB media supplemented with 0.1% glucose, autoinduction media), the protein was not obtained. As a result, attention was turned to the production of M21I M73I M148I W151F M206I M236I M275I M280I C344S M383I M439I *Ss*βG, an *Ss*βG mutant which also contained the W151F mutation.

M21I M73I M148I W151F M206I M236I M275I M280I C344S M383I M439I *Ss*βG was generated as follows: site-directed mutagenesis was used to generate pET24d plasmid encoding M21I M73I M148I W151F M206I M236I M275I M280I C344S M383I M439I *Ss*βG from pET24d plasmid encoding M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG. DNA sequencing was used to

confirm that the desired mutation had been introduced. M21I M73I M148I W151F M206I M236I M275I M280I C344S M383I M439I *Ss*βG was expressed and purified. The protein identity was confirmed by ESI-MS (found 57196 Da, calcd. 57196 Da) (see pg. 289). The enzymatic activity of the protein was demonstrated using the X-Gal colorimetric assay.[282]

The protein was then tested for photocleavage activity, which was assessed for using SDS-PAGE (Figure 43). The protein showed a high susceptibility to indiscriminate degradation by UV light; protein was not visible by SDS-PAGE after only one hour of irradiation. No cleavage was observed in the absence of UV light (Figure 43).

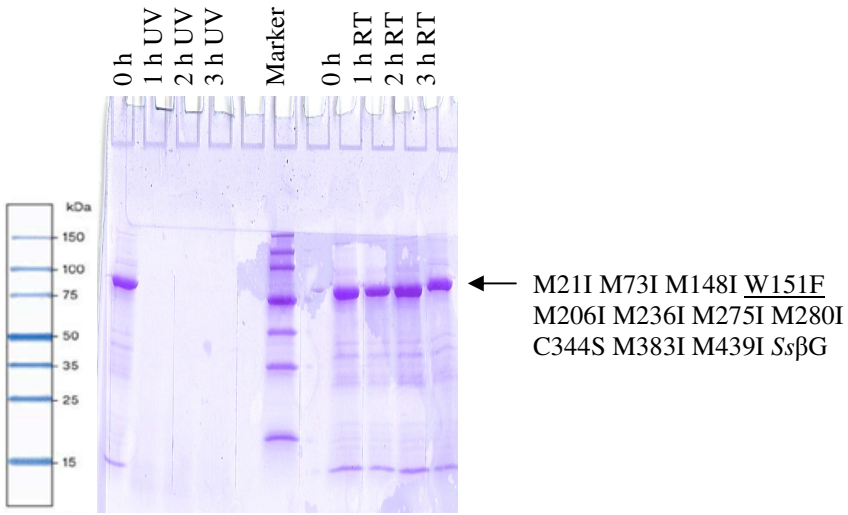



**Figure 43.** Time course analyses of M21I M73I M148I W151F M206I M236I M275I M280I C344S M383I M439I *Ss*βG exposed to/not exposed to UV light by SDS-PAGE.

## 6. His-Trp diad

Bioinformatic analyses (*e.g.* BLAST [basic local alignment search tool] search of PDB [protein data bank] proteins) revealed that the His-Trp (HW) diad is commonly found in proteins and indeed, is repeated at positions His424Trp425 in WT *Ss*βG (Figure 44). Since this site did not dissociate on exposure to UV light it was considered that an extended series of residues bordering His150 and Trp151 must contribute to the remarkable regioselectivity of this cleavage reaction. To probe the underlying molecular contributions, S426P *Ss*βG was generated (see section 6.1), as well as short peptide fragments of *Ss*βG, expected to have little inherent conformation but correct primary sequence (see section 6.2).

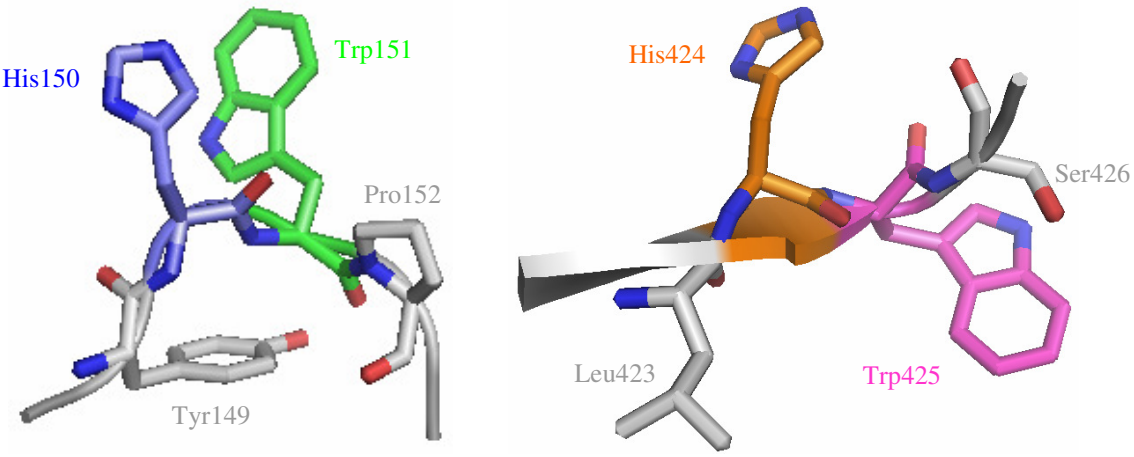



**Figure 44.** His150Trp151 (left) and His424Trp425 (right) residues in WT *Ss*βG (PDB ID: 1GOW).

### 6.1. S426P *Ss*βG

It was envisaged that by mutating Ser426 to Pro in WT *Ss*βG, to mimic the primary protein sequence at the cleavage site (His150-Trp151-Pro152), it may be possible to introduce an additional cleavage site into WT *Ss*βG.

Starting from a pET24d plasmid encoding WT *Ss*βG, site-directed mutagenesis was used to generate a pET24d plasmid encoding S426P *Ss*βG. DNA sequencing was used to confirm that the desired mutation had been introduced. The N-terminal His-tagged protein was expressed and purified, and the protein identity confirmed by ESI-MS (found 57424 Da, calcd. 57424 Da) (see pg. 289). Additionally, the enzymatic activity of S426P *Ss*βG was demonstrated using the X-Gal colorimetric assay.[282]

S426P *Ss*βG was then tested for photocleavage activity, which was assessed for using SDS-PAGE (Figure 45). The *Ss*βG mutant was found to remain intact at the His424-Trp425 junction and only fragmented at the known cleavage site His150-Trp151. No cleavage was observed in the absence of UV light (Figure 45). The bands corresponding to protein on the gel marked ■ were quantified by densitometry (see pg. 320).

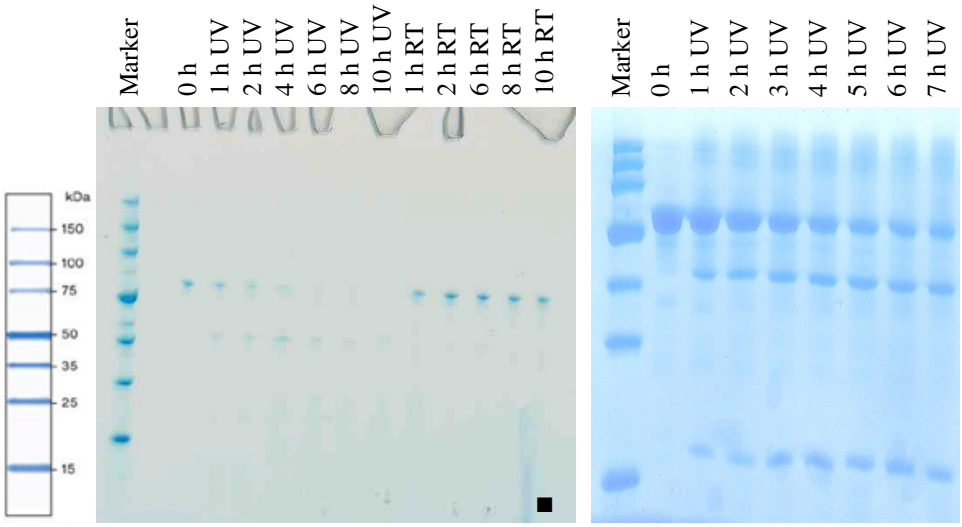



**Figure 45.** Time course analyses of S426P *Ss*βG exposed to/not exposed to UV light by SDS-PAGE.

Attempts to crystallise S426P *Ss*βG using a commercial crystal screen and the sitting drop vapor diffusion technique, in order to understand the effect that the Ser426→Pro mutation had on the spatial orientation of His424 and Trp425, were not successful (see experimental).

## 6.2. Peptides

The short peptide fragments of *Ss*βG, H-HWP-$NH_2$ and H-LNMYHWPLPL-$NH_2$ (Figure 46), representing residues 150-152 and 146-155 of *Ss*βG, respectively, were prepared on Rink Amide 4-methyl-benz-hydrylamine (MBHA) resin using standard single cycles of Fmoc solid phase peptide synthesis (SPPS). ESI-MS was used to suggest that the correct peptides had been obtained.

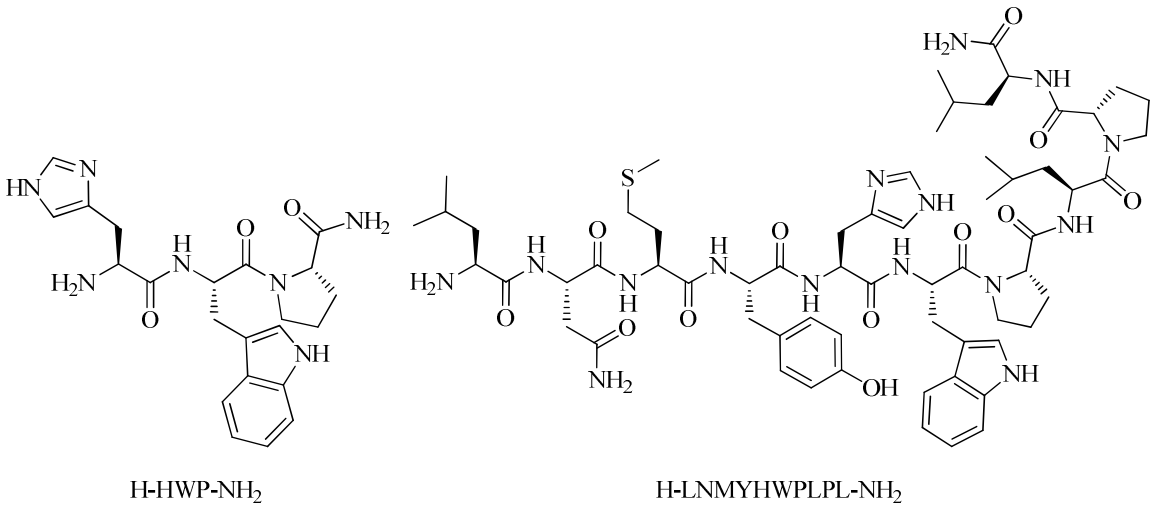



**Figure 46.** Peptide structures.

The circular dichroism spectrum of H-LNMYHWPLPL-$NH_2$ was recorded (Figure 47). By comparison to characteristic circular dichroism spectra, the peptide was found to have mostly irregular structure and negligible secondary structure.[326]

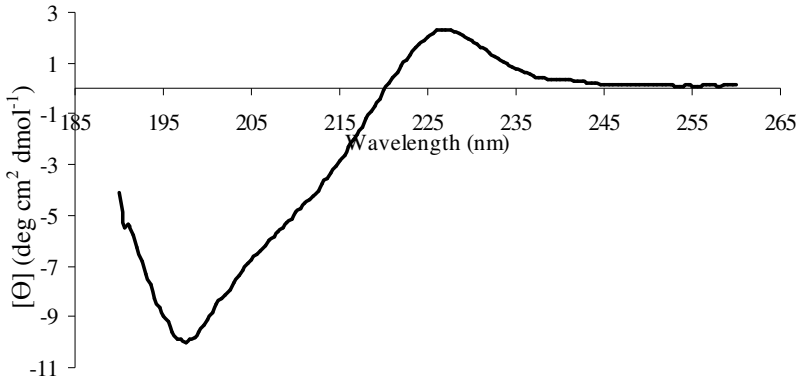



**Figure 47.** Circular dichroism spectrum of peptide H-LNMYHWPLPL-$NH_2$.

H-HWP-$NH_2$ and H-LNMYHWPLPL-$NH_2$ were then tested for photocleavage activity, which was assessed for over 7 hours using ESI-MS. Neither peptide fragmented under UV light. These results implicated conformation and secondary/tertiary structure as key determinants in fragmentation.

## 7. Alanine scanning

To investigate if the secondary/tertiary structure of *Ss*βG were key determinants in fragmentation, alanine scanning was used to probe a 3.5 Å sphere of residues around the cleavage site to establish which residues were essential for photocleavage activity (Figure 48). *Ss*βG is a TIM-barrel protein, and as shown in Figure 48 the His-Trp cleavage site is located in the centre of *Ss*βG on a loop.

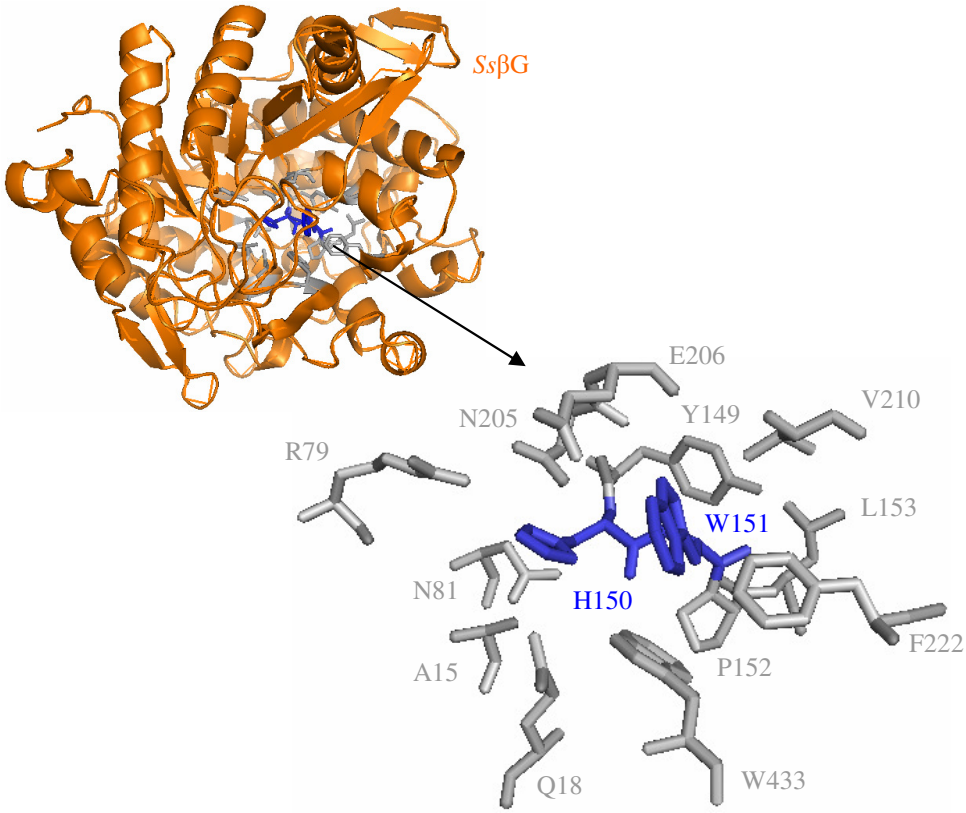
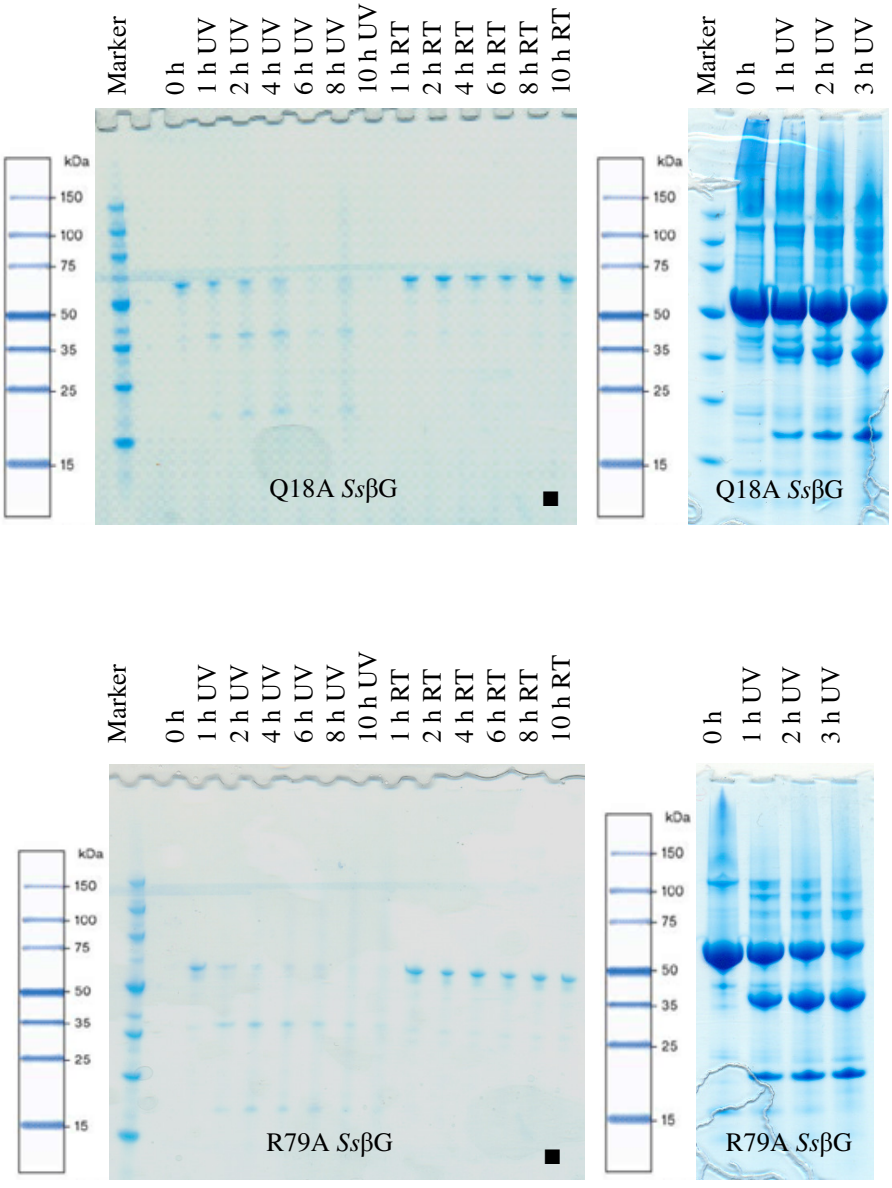
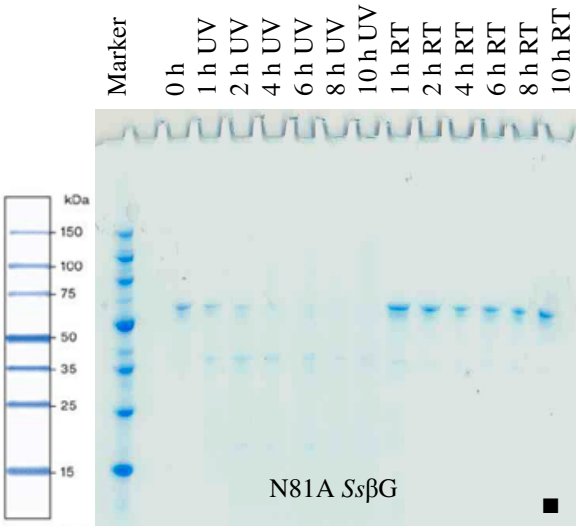


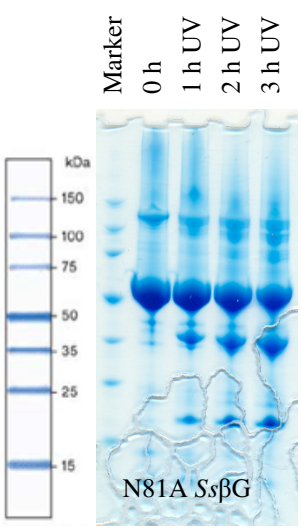

**Figure 48.** *Ss*βG (top) and *Ss*βG's photocleavage site [blue] and residues within 3.5 Å of the cleavage site [gray] (bottom). Models based on PDB ID: 1GOW.

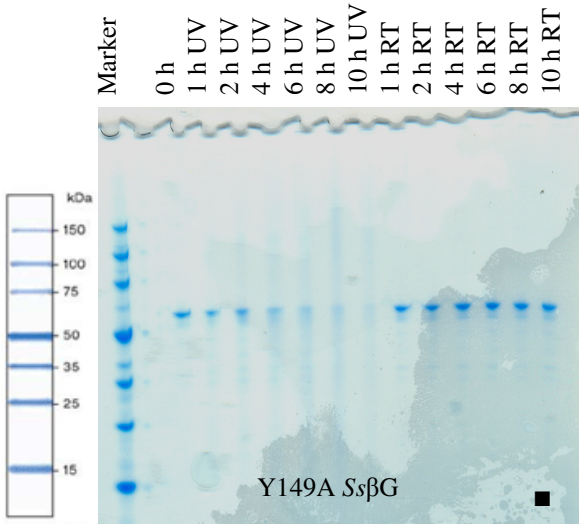
Starting from a pET24d plasmid encoding WT *Ss*βG, site-directed mutagenesis was used to generate pET24d plasmids encoding Q18A *Ss*βG, R79A *Ss*βG, N81A *Ss*βG, Y149A *Ss*βG, P152A *Ss*βG, L153A *Ss*βG, N205A *Ss*βG, V210A *Ss*βG, F222A *Ss*βG and W433A *Ss*βG. DNA sequencing was used to confirm that the desired mutations had been introduced. In addition, pET24d plasmid encoding E387A *Ss*βG (4.1 Å from cleavage site) was kindly provided by Dr. Rebecca Bell (University of Oxford). The N-terminal His-tagged proteins were expressed and purified. E206A *Ss*βG was kindly provided by Miss Samantha Cheung (Univerisity of Oxford). The protein identities were confirmed by ESI-MS (see pg. 289). Additionally, the enzymatic activities of the proteins were demonstrated using the X-Gal colorimetric assay.[282]

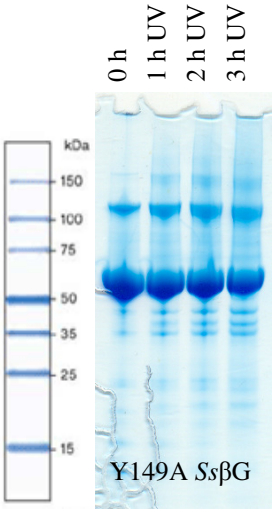
The twelve *Ss*βG mutants were then tested for photocleavage activity, which was assessed for using SDS-PAGE (Figure 49). In addition, the bands corresponding to protein on the gels marked ■ were quantified by densitometry (see pg. 320).

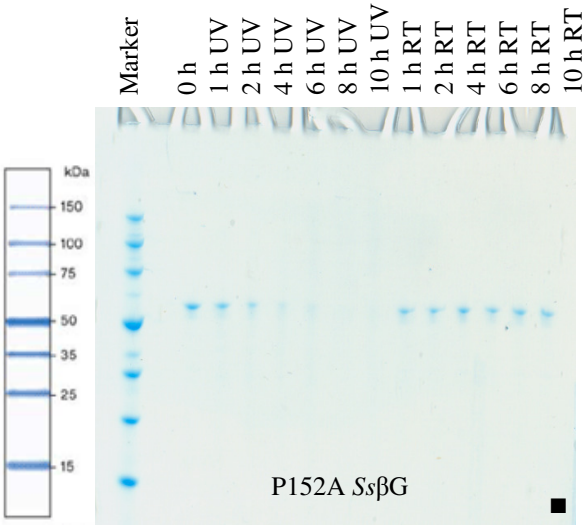


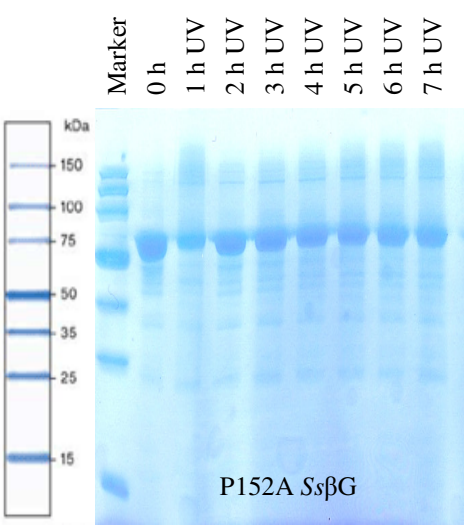

Marker
0 h
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
kDa
150
100
75
50
35
25
15
N81A SsβG

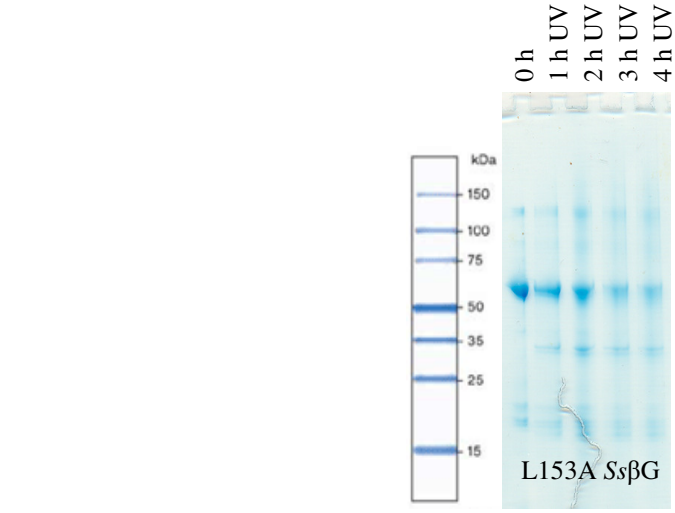

Marker
0 h
1 h UV
2 h UV
3 h UV
kDa
150
100
75
50
35
25
15
N81A SsβG

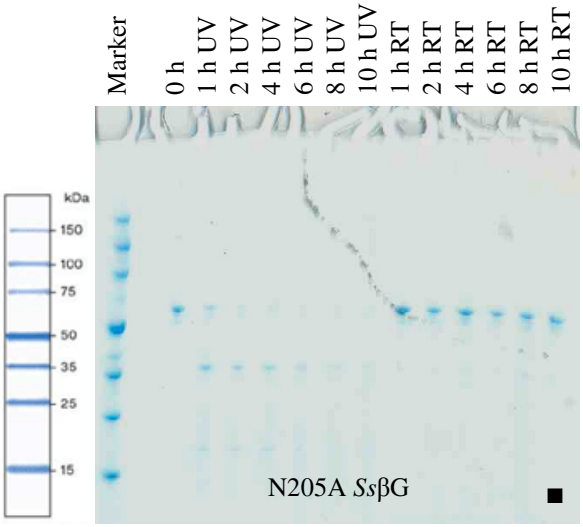

Marker
0 h
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
kDa
150
100
75
50
35
25
15
Y149A SsβG

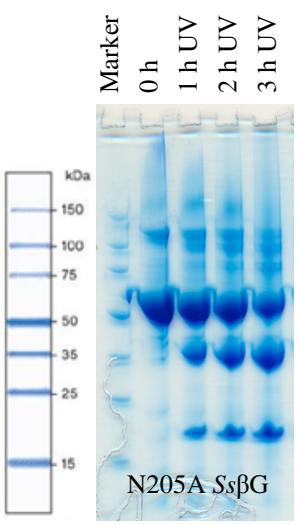

0 h
1 h UV
2 h UV
3 h UV
kDa
150
100
75
50
35
25
15
Y149A SsβG

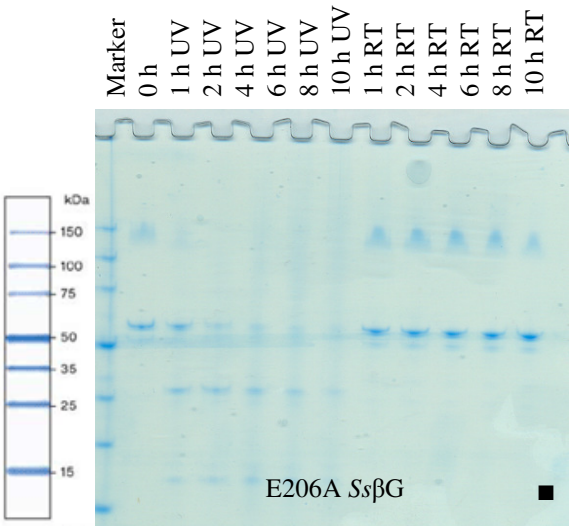

Marker
0 h
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
kDa
150
100
75
50
35
25
15
P152A SsβG

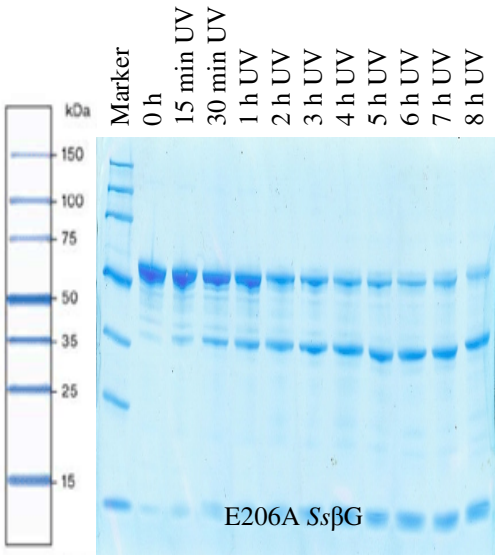

Marker
0 h
1 h UV
2 h UV
3 h UV
4 h UV
5 h UV
6 h UV
7 h UV
kDa
150
100
75
50
35
25
15
P152A SsβG

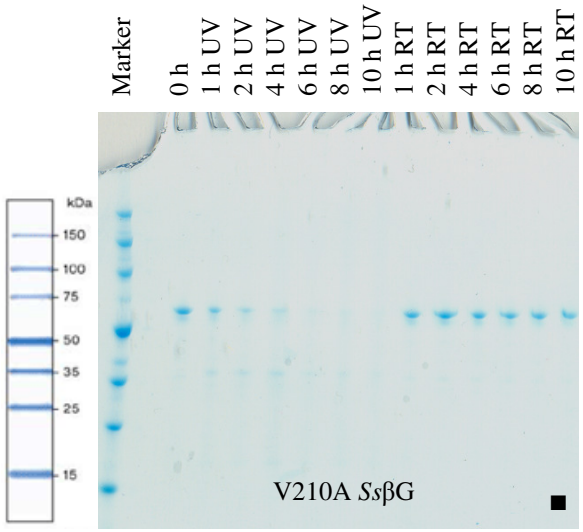

0 h
1 h UV
2 h UV
3 h UV
4 h UV
kDa
150
100
75
50
35
25
15
L153A SsβG

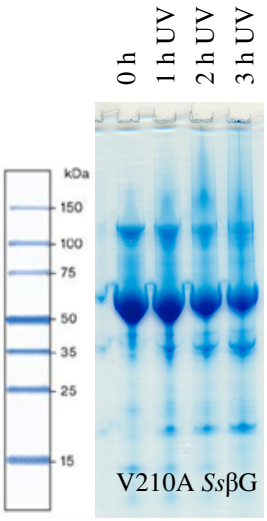

Marker
0 h
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
kDa
150
100
75
50
35
25
15
N205A SsβG

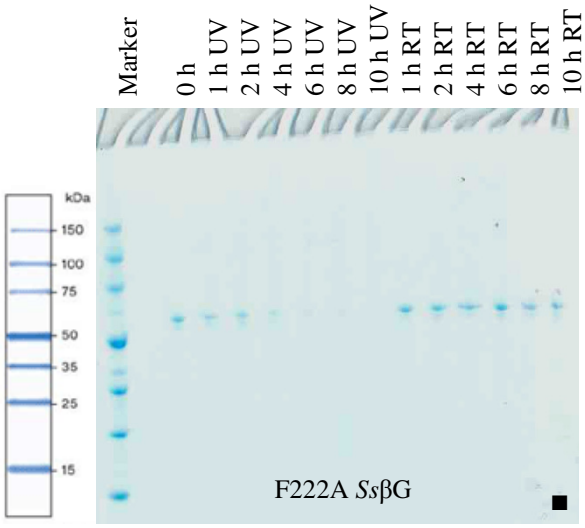

Marker
0 h
1 h UV
2 h UV
3 h UV
kDa
150
100
75
50
35
25
15
N205A SsβG

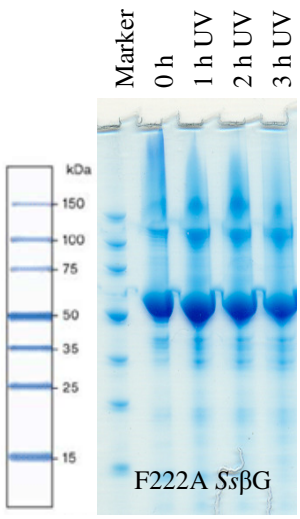

Marker
0 h
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
kDa
150
100
75
50
35
25
15
E206A SsβG

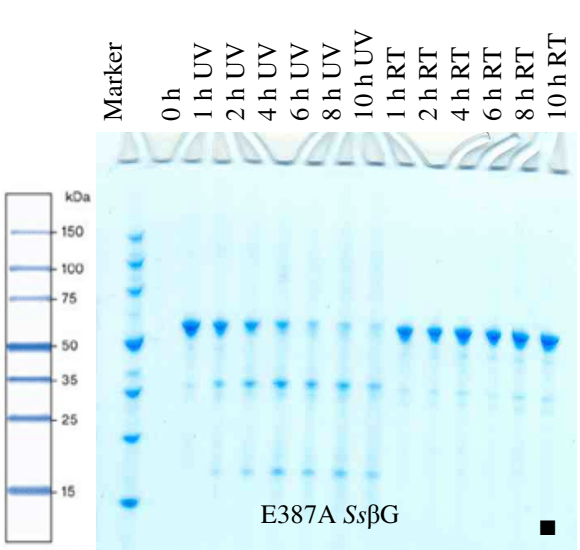

Marker
0 h
15 min UV
30 min UV
1 h UV
2 h UV
3 h UV
4 h UV
5 h UV
6 h UV
7 h UV
8 h UV
kDa
150
100
75
50
35
25
15
E206A SsβG

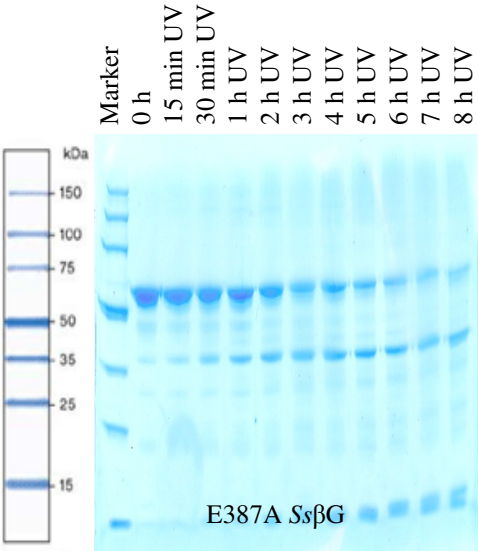

Marker
0 h
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
kDa
150
100
75
50
35
25
15
V210A SsβG


0 h
1 h UV
2 h UV
3 h UV
kDa
150
100
75
50
35
25
15
V210A SsβG


Marker
0 h
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
kDa
150
100
75
50
35
25
15
F222A SsβG


Marker
0 h
1 h UV
2 h UV
3 h UV
kDa
150
100
75
50
35
25
15
F222A SsβG


Marker
0 h
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
kDa
150
100
75
50
35
25
15
E387A SsβG


Marker
0 h
15 min UV
30 min UV
1 h UV
2 h UV
3 h UV
4 h UV
5 h UV
6 h UV
7 h UV
8 h UV
kDa
150
100
75
50
35
25
15
E387A SsβG

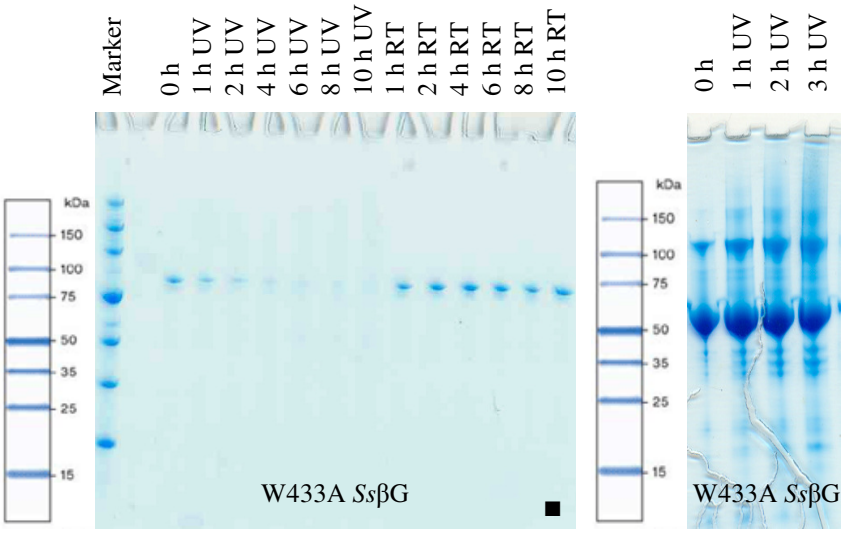



**Figure 49.** Time course analyses of proteins exposed to/not exposed to UV light by SDS-PAGE.

The *Ss*βG mutations Y149A, P152A, F222A and W433A were found to prevent fragmentation whilst alteration of the more remote residues Q18A, R79A, N81A, L153A, N205A, E206A, V210A and E387A had no effect. Fragment masses were determined by ESI-MS for all the cleaving proteins, except L153A *Ss*βG (Figure 51). Modeling of the cleavage site reveals that the essential residue groups Y149, F222, W433 and P152, cradle residue W151 in a hydrophobic, π-rich environment (Figure 50).

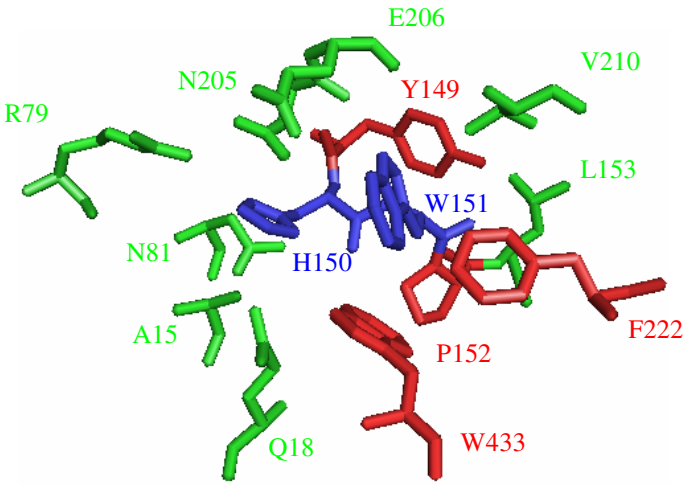
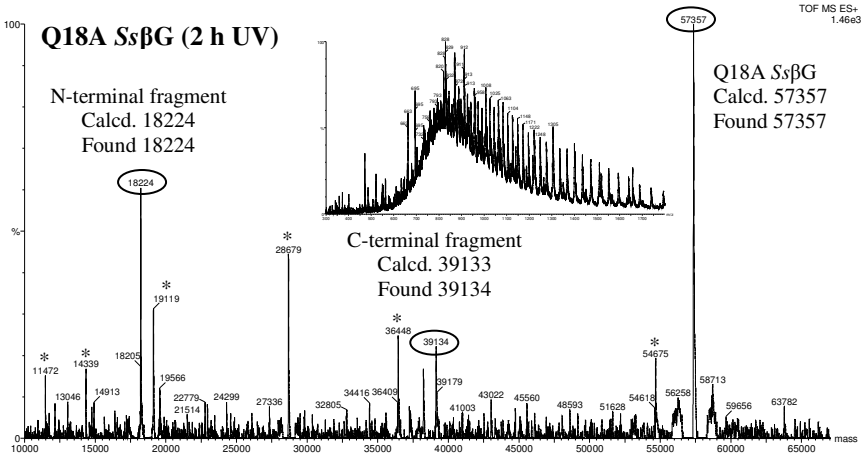
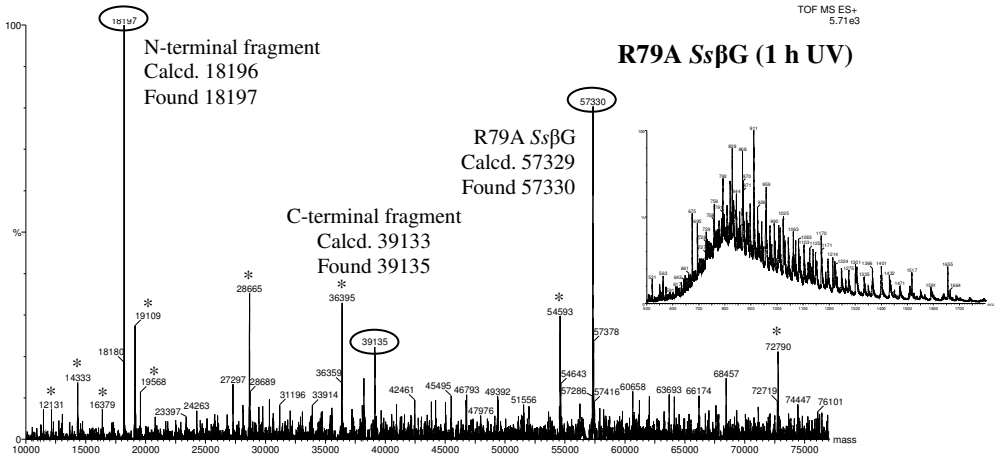
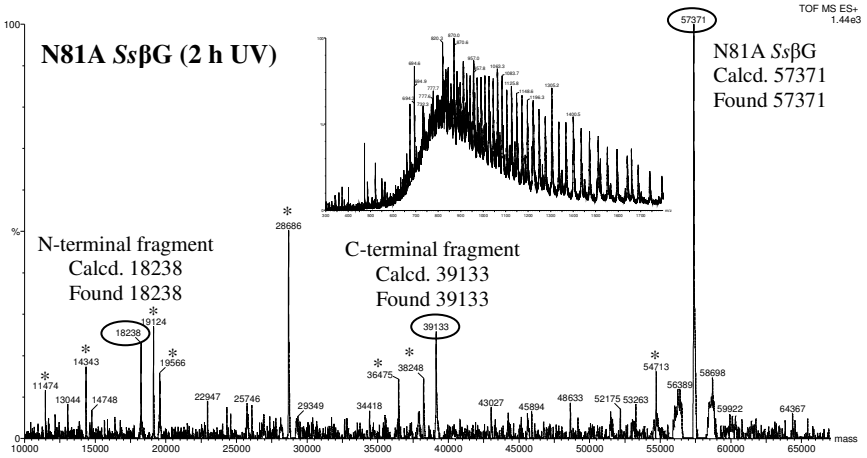
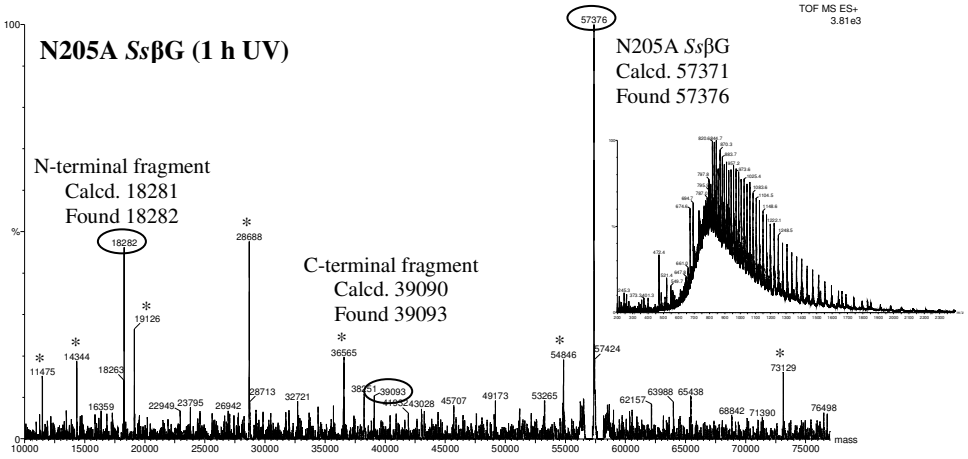


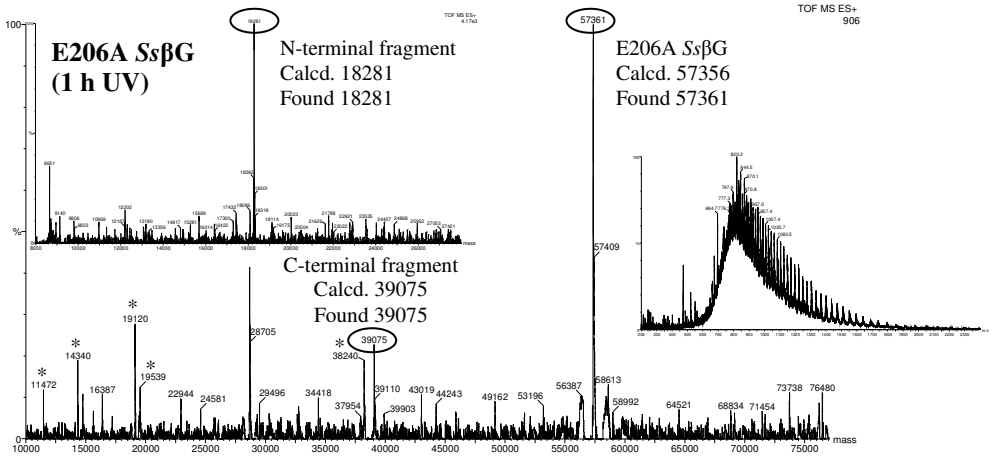

**Figure 50.** *Ss*βG's photocleavage site (blue) and associated essential (red) and non-essential (green) residues. Model based on PDB ID: 1GOW.




N81A *Ss*βG (2 h UV)
N81A *Ss*βG
Calcd. 57371
Found 57371
N-terminal fragment
Calcd. 18238
Found 18238
C-terminal fragment
Calcd. 39133
Found 39133
mass


N205A *Ss*βG (1 h UV)
N205A *Ss*βG
Calcd. 57371
Found 57376
N-terminal fragment
Calcd. 18281
Found 18282
C-terminal fragment
Calcd. 39090
Found 39093
mass


E206A *Ss*βG
(1 h UV)
N-terminal fragment
Calcd. 18281
Found 18281
E206A *Ss*βG
Calcd. 57356
Found 57361
C-terminal fragment
Calcd. 39075
Found 39075
mass

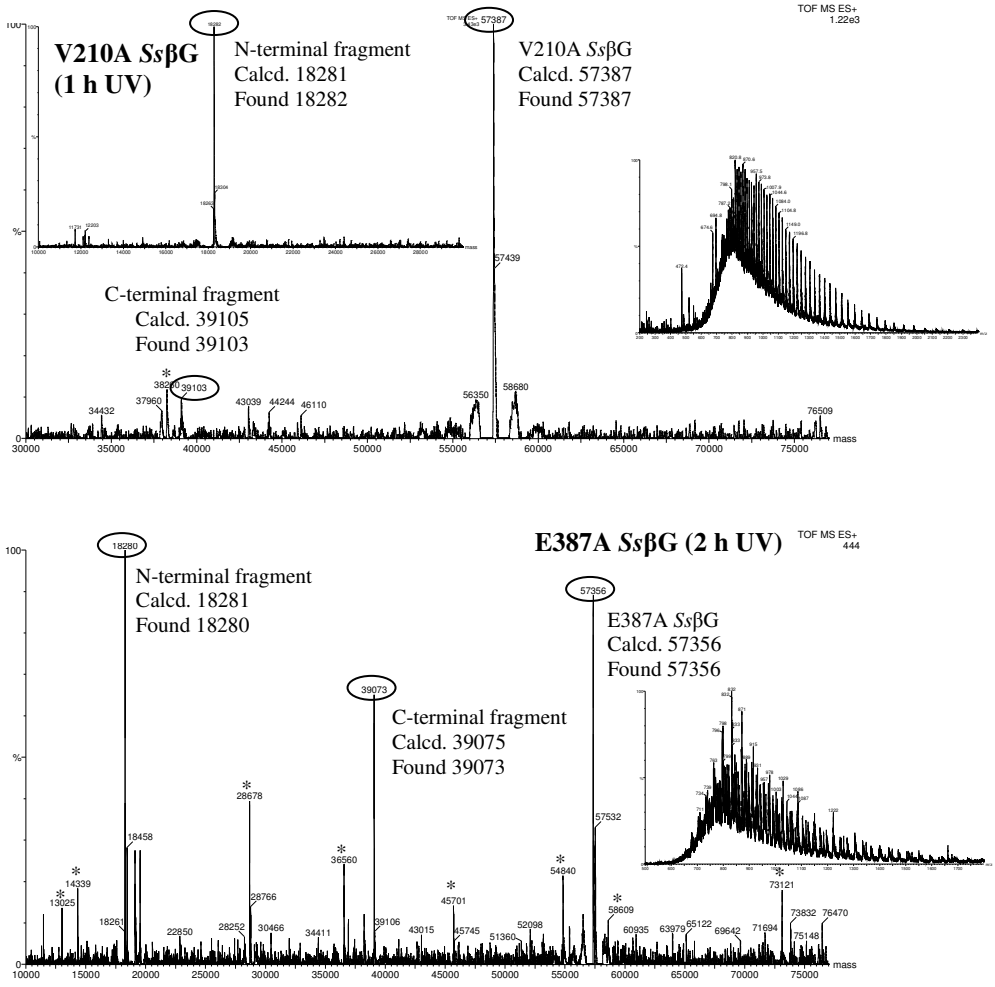
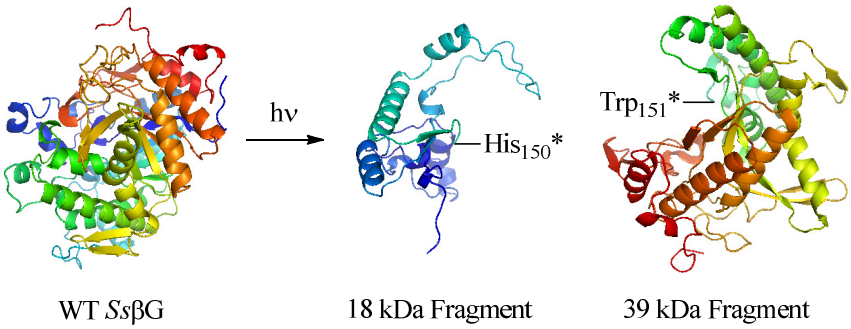



**Figure 51.** Raw (insert) and deconvoluted ESI-MS obtained for cleaving *Ss*βG proteins irradiated with UV light. * = Artifact of the deconvolution algorithm.[283-285]

Attempts to crystallise P152A *Ss*βG using a number of commercial crystal screens and the sitting drop vapor diffusion technique, in order to understand the effect that the Pro152→Ala mutation had on the spatial orientation of His150 and Trp151, were not successful (see experimental).

## 8. Characterisation of the cleavage site

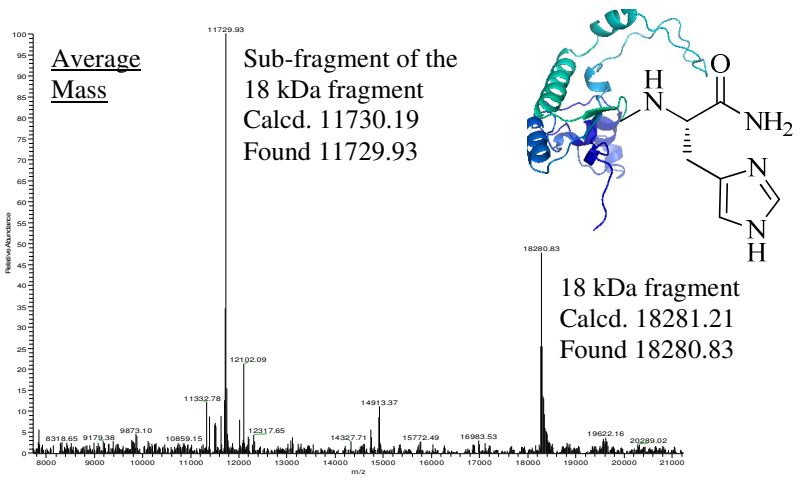
In order to postulate a mechanism for the photocleavage reaction, it was necessary to find out the identity of the C-terminal amino acid of the 18 kDa fragment ($His_{150}$* in Scheme 36) and N-terminal amino acid of the 39 kDa fragment ($Trp_{151}$* in Scheme 36). This work was carried out under the instruction of, and in the case of some experiments by, Dr. Neil Oldham currently at The University of Nottingham.




**Scheme 36.** Photocleavage of WT *Ss*βG.

### 8.1. Characterisation of the 18 kDa fragment

To begin with, the 18 kDa fragment was generated by irradiating WT *Ss*βG with UV light (Scheme 36), which was verified by SDS-PAGE. The solution containing the intact 18 kDa fragment was directly injected into an LTQ FT ULTRA hybrid Ion Trap-Fourier Transform Ion Cyclotron Resonance mass spectrometer and was analysed by ESI-MS (Figure 52). The average and monoisotopic mass of the 18 kDa fragment were found to be 18280.83 and 18269.83 Da, respectively, which corresponded well with the calculated mass of the 18 kDa fragment if it was C-terminally amidated (18281.21 and 18269.83 Da, respectively) (Figure 52).




Calcd. for: GHHHHHHHSFPNSF RFGWSQAGFQSEMGTPGSEDP NTDWYKWVHDPENMAAGLV SGDL[PENGPGYWGNYKTFHD NAQKMGLKIARLNVEWSRIFP NPLPRPQNFDESKQDVTEVEIN ENELKRLDEYANKDALNHYR EIFKDLKSRGLYFILNMYH$_{150}$*]
Key: [ ] = sub-fragment of the 18 kDa fragment induced in the mass spectrometer. H* = His-amide.

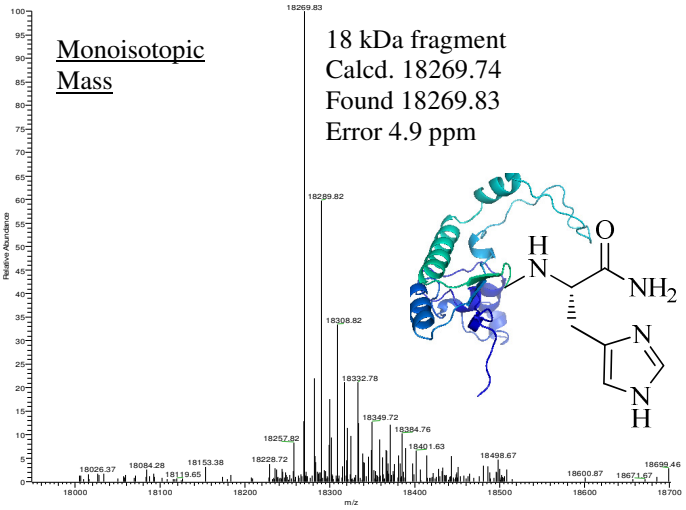
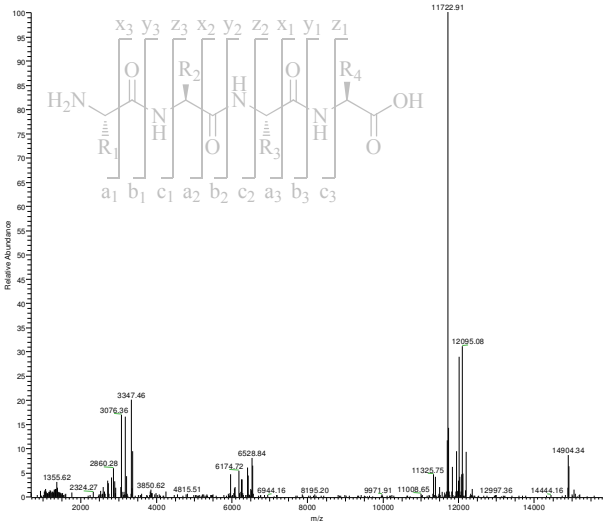



**Figure 52.** Deconvoluted ESI-MS (provided by Dr. Neil Oldham) of the 18 kDa fragment (top). Zoom-in of the 18 kDa region of the top spectrum, de-isotoped to give the monoisotopic mass (bottom). 18 kDa fragment with C-terminal amidation (insert).

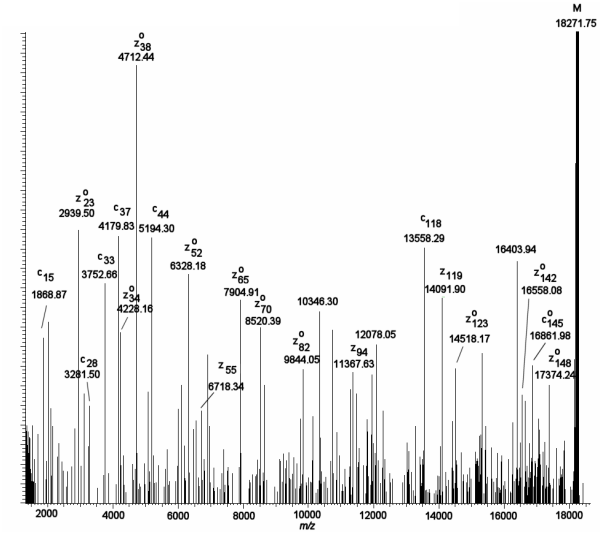
The precursor ion *m/z* 1219.7 $[M+H]^{15+}$ from the 18 kDa fragment was then selected and fragmented in the ion trap using collision-induced dissociation (CID). A number of b and y fragment ions of the 18 kDa fragment were detected by ESI-MS, and are shown in the MS/MS spectrum in Figure 53. To improve the sequence coverage and to generate alternative fragmentation information that complemented the CID-derived data,[327] the precursor ion *m/z* 962.1 $[M+H]^{19+}$ from the 18 kDa fragment was selected

and fragmented using electron capture dissociation (ECD). Numerous c and z fragment ions of the 18 kDa fragment were detected by ESI-MS, and are shown in the MS/MS spectrum in Figure 54.




Calcd. for:
GHHHHHHHSFPNSFRFGWSQA
GFQSE}M}GT}{PGSEDPNTDW
YKWVHDPENMAA{G}{L}{V}{
G{D}{L}{P{E{N{GPGYWGNY
KTFHDNAQKMGLKIARLNVE
WSRIFPNPLPRPQNFDESKQDV
TEVEINENELKRLDEYANKDA
LNHYREIFKDLKSRGLYFILNM
$YH_{150}$*
Key: { = y ion. } = b ion. H* =
His-amide.

**Figure 53.** MS/MS spectrum (provided by Dr. Neil Oldham) from in-trap CID of *m/z* 1219.7 $[M+H]^{15+}$ from the 18 kDa fragment.




Calcd. for:
GHHHHHHHSFPNSF{R}FGWS
QAGFQSEMG}TPGS{E}DPN{T
}DWYKWVH}DPENMAAGLVG
DLPENG{PGYWGNYKTFHD{N
AQKMGLKIARL{NVEWS{RIFP
NPLPRP{QNF{DESKQDVTEVE
INEN}ELK{RLDEYANKDAL{N
HYREIFKDLKSR}GLYFILNMY
$H_{150}$*
Key: { = z ion. } = c ion. H* =
His-amide.

**Figure 54.** MS/MS spectrum (provided by Dr. Neil Oldham) from ECD of *m/z* 962.1 $[M+H]^{19+}$ from the 18 kDa fragment.

The irradiated WT *Ss*βG solution containing the 18 kDa and 39 kDa fragments and noncleaved WT *Ss*βG was purified by SDS-PAGE, and the band corresponding to the 18 kDa fragment was digested in-gel using trypsin to release the C-terminal peptide GLYFILNMYH$_{151}$*. This method proved superior to the use of in-solution trypsin digest protocols. ESI-MS/MS analysis of GLYFILNMYH$_{151}$* (*m/z* 635.33 $[M+H]^{2+}$) using a ThermoFisher LTQ FT ULTRA hybrid Ion Trap-Fourier Transform Ion Cyclotron Resonance mass spectrometer allowed accurate mass and formulae determination of the C-terminal amino acid of the fragment, giving results consistent with it being a His-amide (Figure 52). The major b and y fragment ions of GLYFILNMYH$_{151}$* are shown in Figure 55.

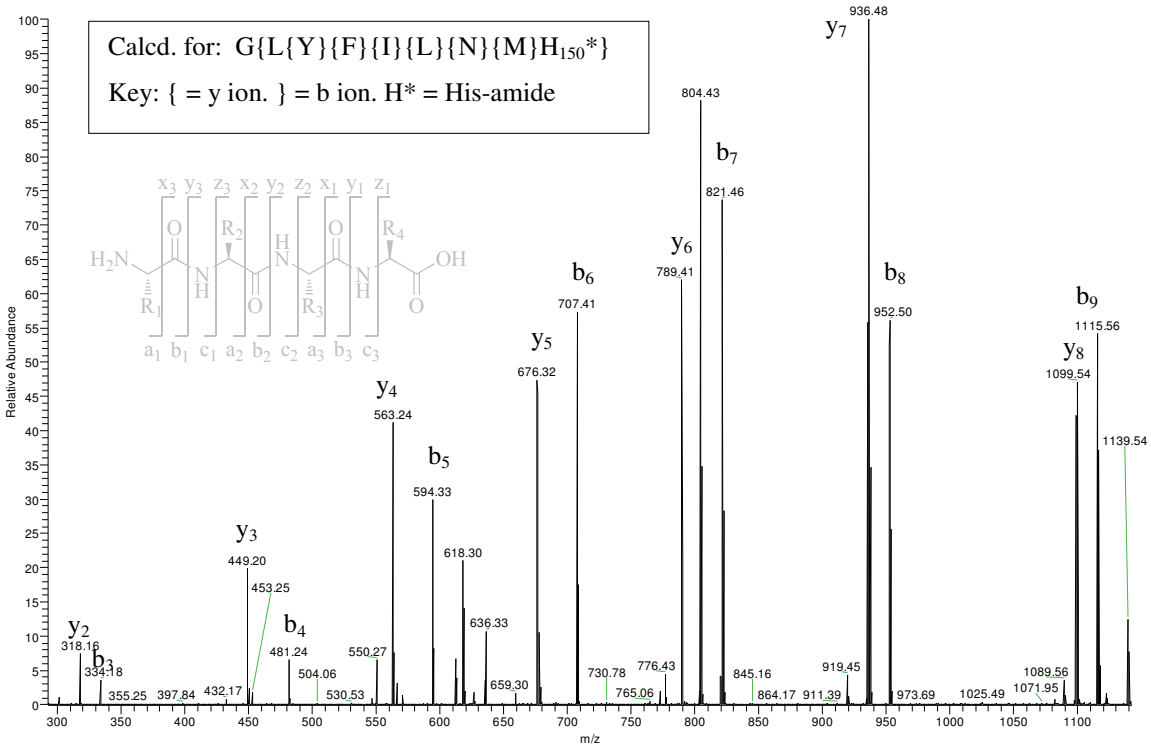



**Figure 55.** ESI-MS/MS (*m/z* 635.33 $[M+H]^{2+}$) of the C-terminal tryptic peptide GLYFILNMYH$_{150}$* from the 18 kDa fragment.

## 8.2. Characterisation of the 39 kDa fragment

Identification of the N-terminal amino acid of the 39 kDa fragment proved more challenging than for the C-terminal amino acid of the 18 kDa fragment, primarily because the mass of the fragment determined by ESI-MS deviated more from the calculated mass; N-terminal fragment found 18281 Da, calcd. 18282 Da (error 1 Da). C-terminal fragment found 39133 Da, calcd. 39150 (error 17 Da).

As for the 18 kDa fragment, the 39 kDa fragment was purified from the irradiated WT *Ss*βG solution by SDS-PAGE, and the band corresponding to the 39 kDa fragment was digested in-gel using trypsin to release the N-terminal peptide $*W_{151}$PLPLWLHDPIR. ESI-MS/MS analysis of the whole digest mixture was then carried out, from which a search of *m/z* 1146.00-1147.00 corresponding to the $y_9$ ion (*m/z* 1146.64) of $*W_{151}$PLPLWLHDPIR was performed. By combining data from the chromatogram between 28.85 and 29.25 minutes, the doubly charged ion of $*W_{151}$PLPLWLHDPIR was identified (*m/z* 763.42 $[M+H]^{2+}$), which allowed the mass of $*W_{151}$PLPLWLHDPIR to be determined as *m/z* 1524.84 (Figure 56).

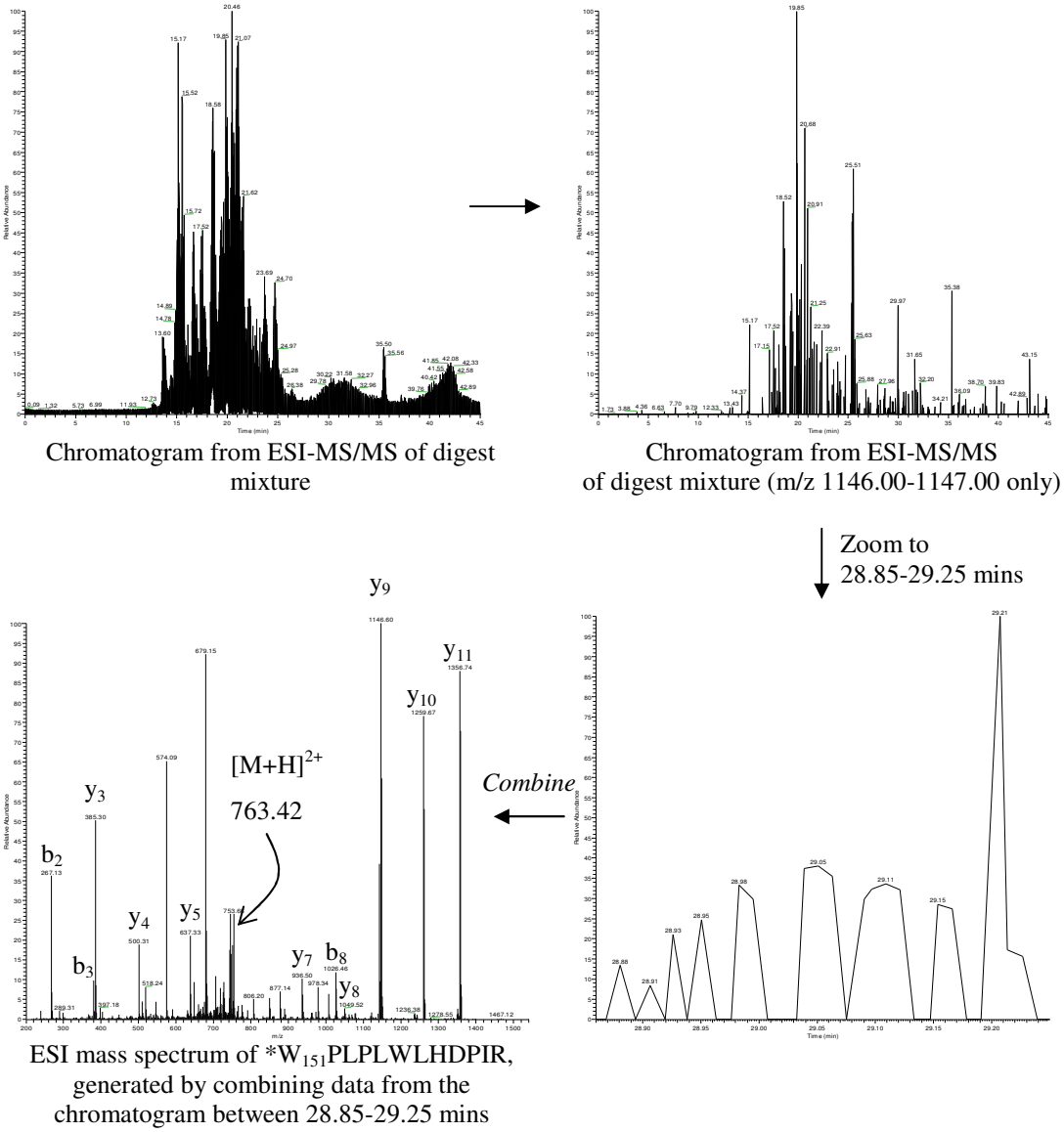



**Figure 56.** Determination of the mass of the N-terminal peptide *$W_{151}$PLPLWLHDPIR from the 39 kDa fragment.

As the b1 ion of *$W_{151}$PLPLWLHDPIR was not observed in the previous experiment, CID MS/MS/MS of *m/z* 763.42 (*$W_{151}$PLPLWLHDPIR $[M+H]^{2+}$) → *m/z* 267.11 (b2 ion of *$W_{151}$PLPLWLHDPIR) was performed (Figure 57). This allowed the accurate mass of the N-terminal amino acid (*$W_{151}$) of the 39 kDa fragment to be determined to be *m/z* 170.06004.

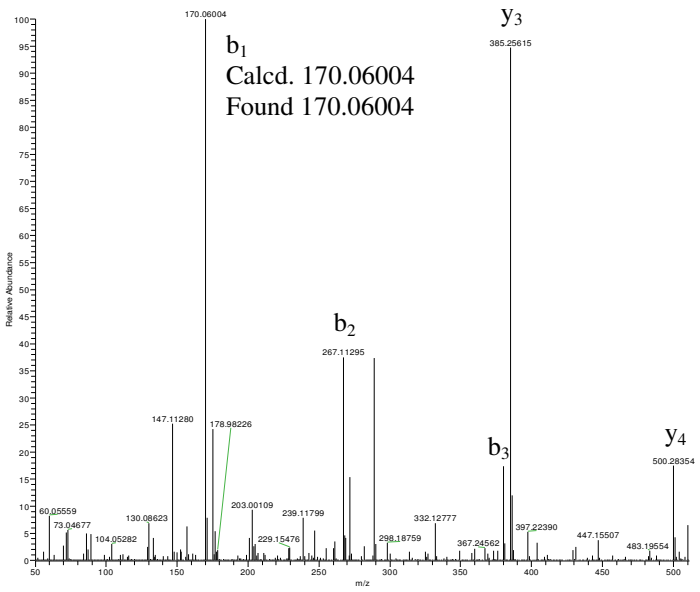



**Figure 57.** CID MS/MS/MS of *m/z* 763.42 → 267.11.

From knowledge of the accurate mass of the b1 ion, the molecular formula was determined to be $C_{11}H_8NO$, which was consistent with the enamide structure shown in Figure 58. Lack of an N-terminal amine in the 39 kDa fragment was also consistent with it's lack of reactivity under Edman degradation. Further, MS analyses of the b1 ions of *$W_{151}$PLPLWLHDPIR and 3-indole acrylic acid were identical, giving additional evidence for the structure of the N-terminal amino acid of the 39 kDa fragment; both showed characteristic loss of CO followed by loss of HCN from the indole ring (Figure 59).

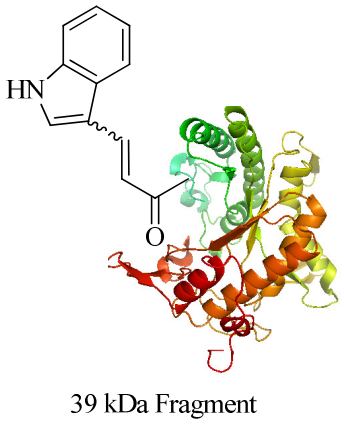



**Figure 58.** 39 kDa fragment with N-terminal enamide.

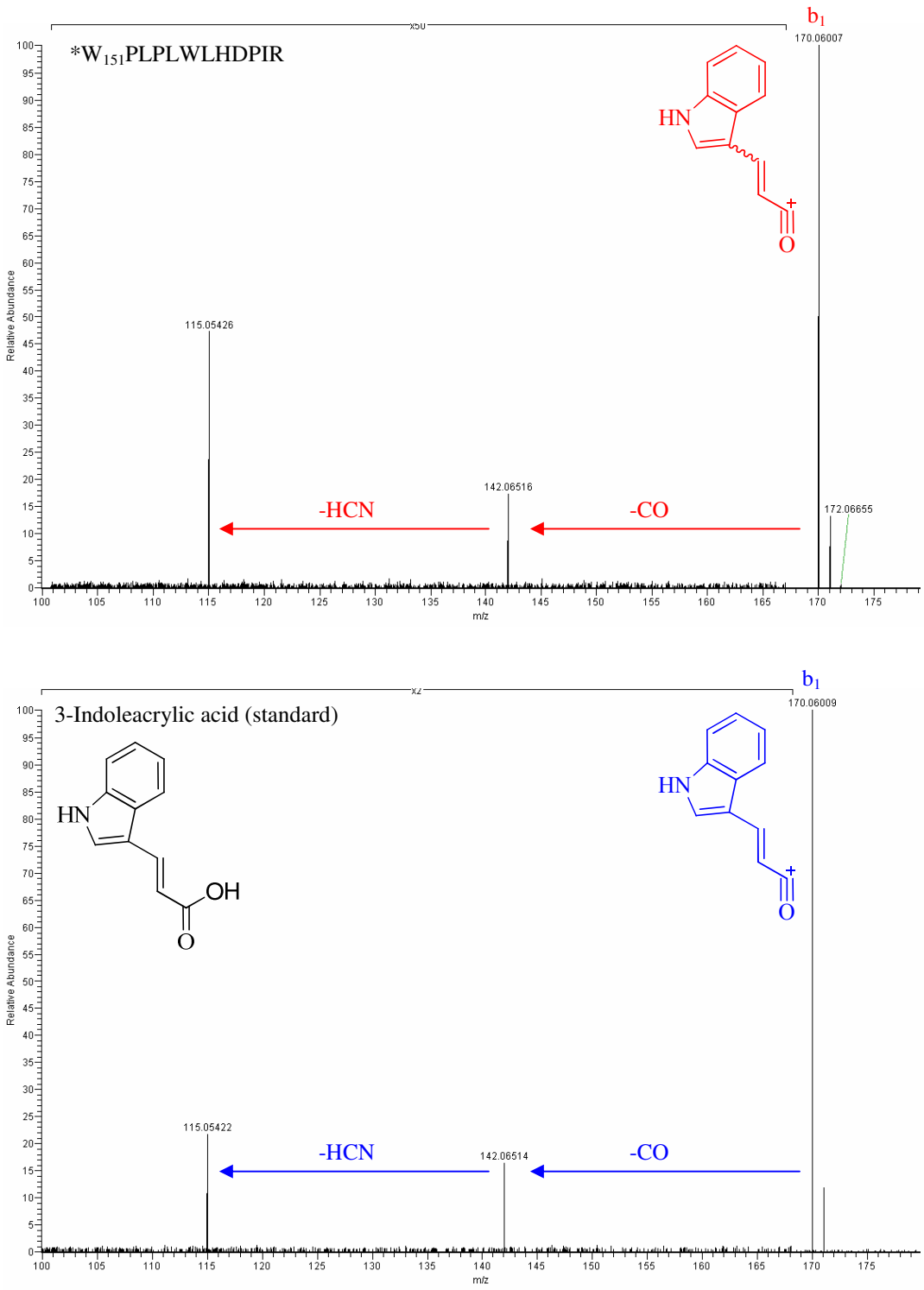



**Figure 59.** ESI-MS/MS/MS/MS (provided by Dr. Neil Oldham) of *m/z* 763.42 → 267.11 → 170.06 → products from peptide *PLPLWLHDPIR (Res# 1-12) and ESI/MS/MS/MS of *m/z* 188 → 170 → products from 3-indoleacrylic acid standard.

UV spectroscopy was employed to provide extra evidence for the presence of the indolylenamide in the 39 kDa fragment. WT *Ss*βG was irradiated with UV light to provoke protein fragmentation and the UV spectrum of the reaction mixture was recorded at time intervals during the course of the reaction (Figure 60). The increase in absorption at λ = 340 nm over time indicated the presence of dehydrotryptophan (Figure 60). This amino acid is structurally similar to the N-terminal amino acid of the 39 kDa fragment and is reported to have a characteristic absorption spectrum containing $\lambda_{max}$ at 275 and 340 nm.[328, 329]

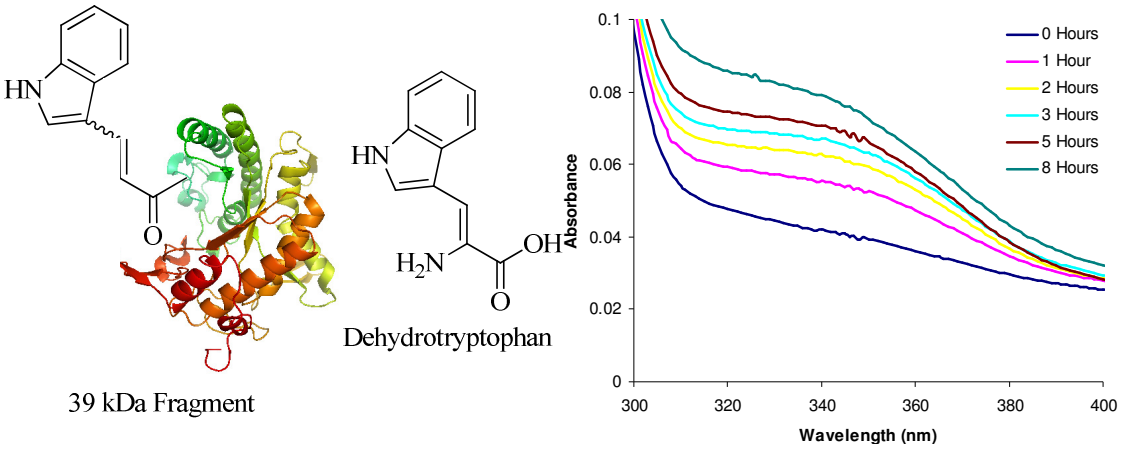



**Figure 60.** Structure of the 39 kDa fragment and dehydrotryptophan (left). UV spectra of UV irradiated WT *Ss*βG (right).

Evidence to definitively prove if the N-terminal amino acid of the 39 kDa fragment was of *cis* or *trans* configuration was not obtained. However, as previously reported,[330, 331] *trans*-3-indole acrylic acid was shown by $^{1}$H NMR to isomerise when irradiated with UV light (Figure 61), indicating that the N-terminal amino acid most likely did too.

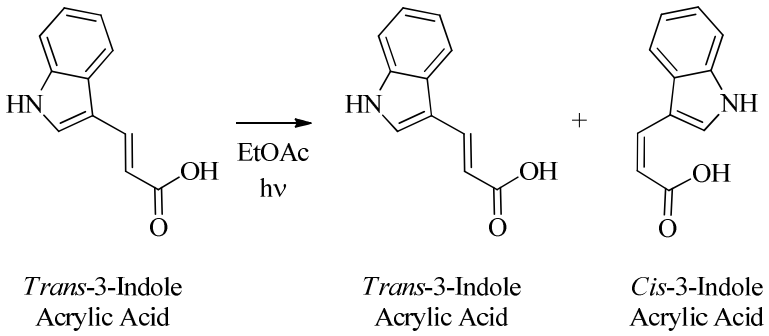



**Figure 61**. Isomerisation of *trans*-3-indole acrylic acid.

## 9. Tagging the 39 kDa fragment

Tagging of the 39 kDa fragment produced from the photocleavage of WT *Ss*βG was attempted with the fluorescent reagents 1-pyrenebutanoic acid hydrazide and fluorescein hydrazide, and tagging of the 39 kDa fragment produced from the photocleavage of M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG was attempted with thiophenol and ethanethiol (Figure 62). The later protein, in which the cysteine residue (C344) occuring in the wild-type protein has been mutated to serine, was chosen to prevent disulfide formation between the protein and thiol reagents. As electron rich indolylenamides such the N-terminal amino acid of the 39 kDa fragment are poor conjugate electrophiles,[332] it was not surprising to find that the fragments did not react with any of the reagents. It is also possible that the reaction site was not accessible because the 18 and 39 kDa fragments remained associated following protein cleavage, even after the addition of the mild denaturant Rapigest to the reaction mixture (ethanethiol only).

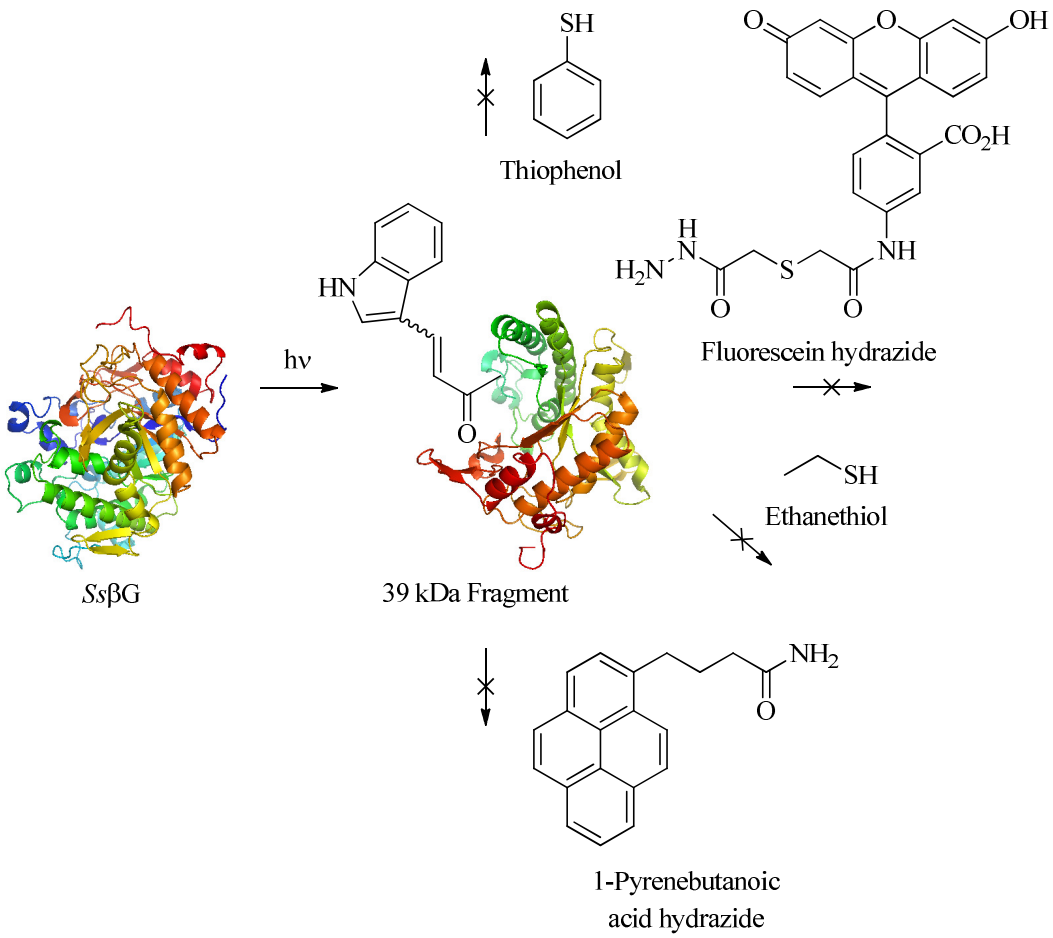



**Figure 62.** Tagging the 39 kDa fragment.

## 10. Oxygen related experiments

Prior to characterising the cleavage site, a photo-induced cleavage mechanism involving singlet oxygen was proposed (Figure 123 on pg. 338). However, evidence to suggest the involvement of singlet oxygen in the cleavage reaction was not found. This suggested a mechanism other than one involving singlet oxygen.

Exclusion of oxygen was not found to alter reaction rate. When dissolved oxygen was removed from a solution of WT *Ss*βG by degassing using vacuum - argon cycles (Figure 63, top) or when the oxygen scavenging reagent Trolox (6-hydroxy-2,5,7,8-

tetramethyl-chroman-2-carboxylic acid)[333] was added (Figure 63, bottom), protein cleavage occurred at the same rate had oxygen not been removed.

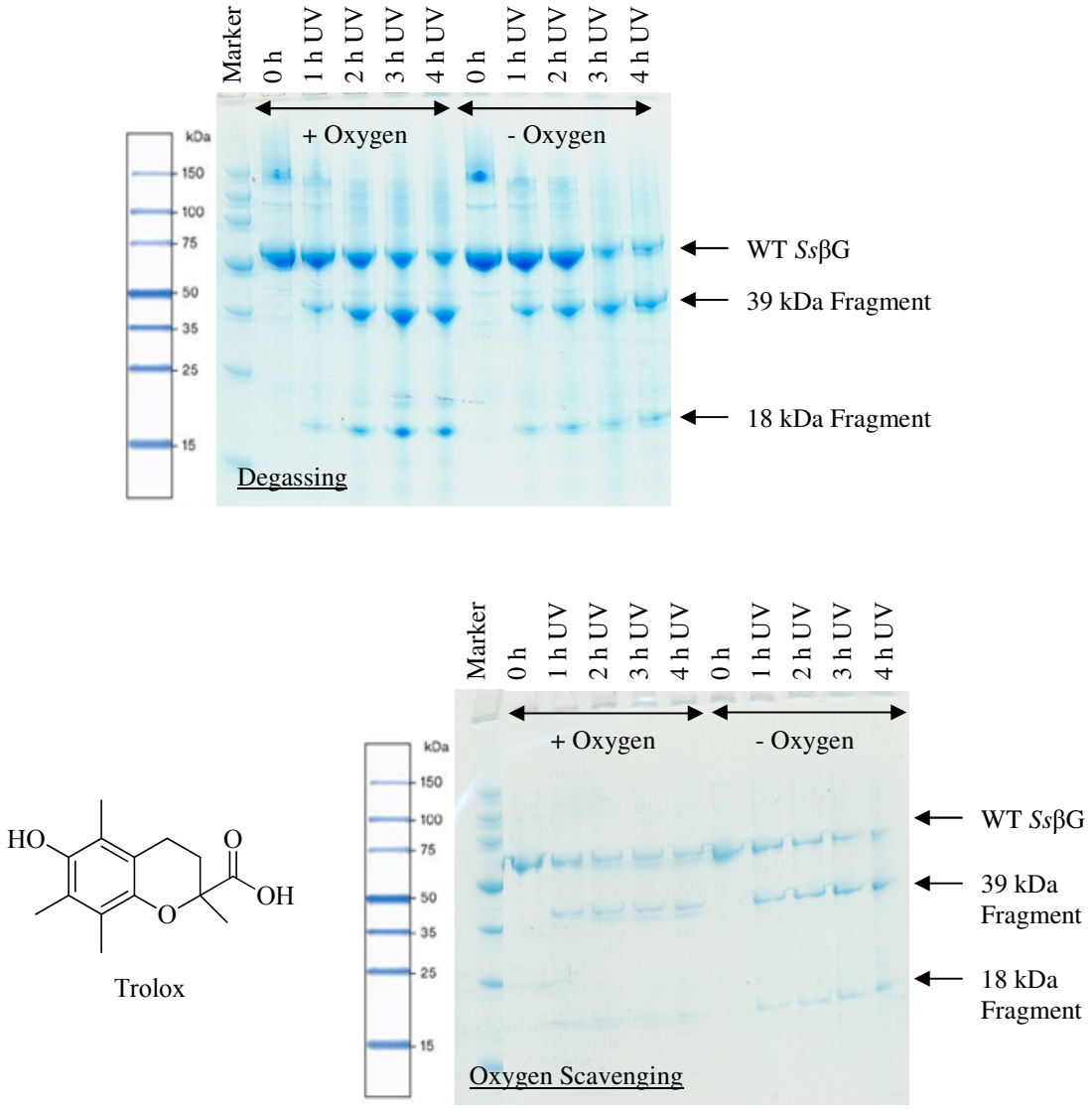



**Figure 63**. Time course analyses of oxygen exclusion experiments by SDS-PAGE.

The lifetime of singlet oxygen is approximately tenfold greater in $D_2O$ than in $H_2O$,[334, 335] however, the reaction rate was also shown to be unaffected by $D_2O$ (Figure 64).

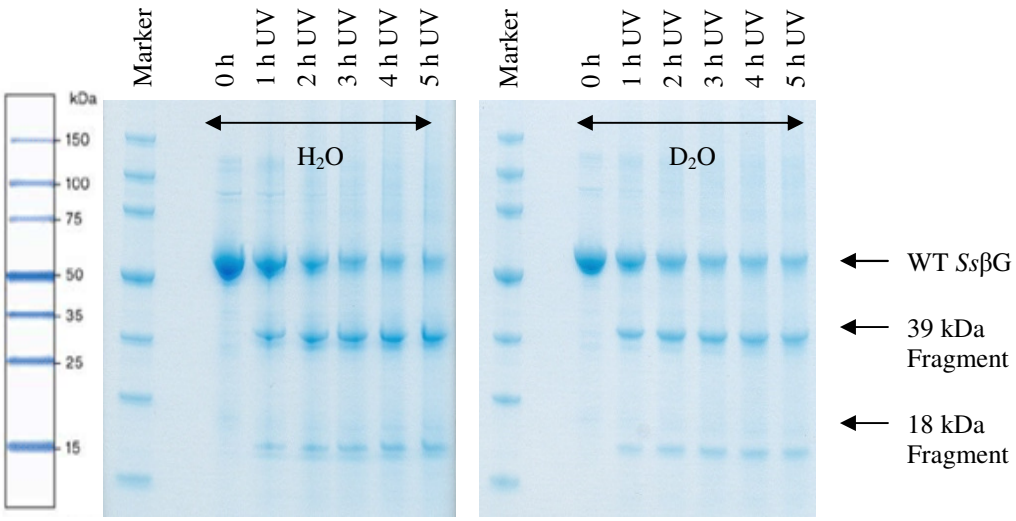



**Figure 64.** Time course analysis of the photo-induced cleavage of WT *Ss*βG in phosphate buffered $H_2O$ and $D_2O$ by SDS-PAGE.

Furthermore, one-photon photosensitised production of singlet oxygen using TMPyP (5,10,15,20-tetrakis(*N*-methyl-4-pyridyl)-21*H*,23*H*-porphine)[336] at 390 nm, a wavelength known not to cause protein fragmentation, did not result in the cleavage of WT *Ss*βG (Figure 65). TMPyP is a water soluble porphyrin that acts as a sensitiser for singlet oxygen upon irradiation at 390 nm.[337, 338] UV light at 390 nm was generated using a neodymium-doped yttrium aluminium garnet (Nd:YAG)-pumped tuneable dye laser. This experiment was carried out with the help of Mr. Chris Eyles and Prof. Mark Brouard (University of Oxford), in the laboratory of Prof. Mark Brouard.

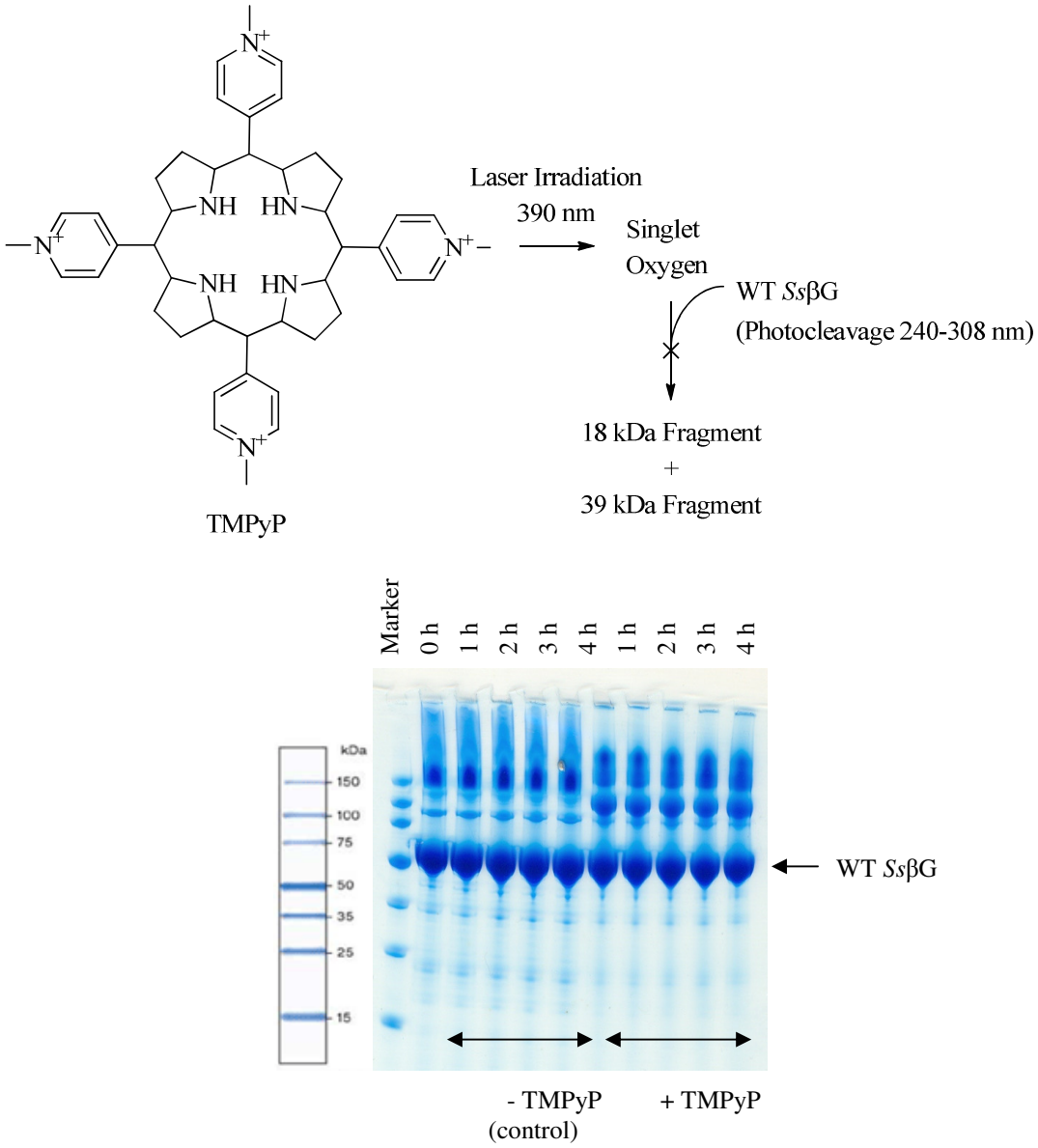



**Figure 65.** Reaction scheme (top). Time course analysis of singlet oxygen generation by TMPyP experiment by SDS-PAGE (bottom).

## 11. Effect of substrate, metal and EDTA

The effect of substrate, metal and EDTA (ethylenediamine tetraacetic acid) on the photocleavage reaction was probed. WT *Ss*βG and lactose, manganese(II) chloride, iron(II) chloride tetrahydrate, iron(III) chloride hexahydrate or EDTA were irradiated with UV light and the reactions were followed by SDS-PAGE (Figure 66). The

reaction was found to be unaffected by the addition of any of the reagents. No reductants were tested.

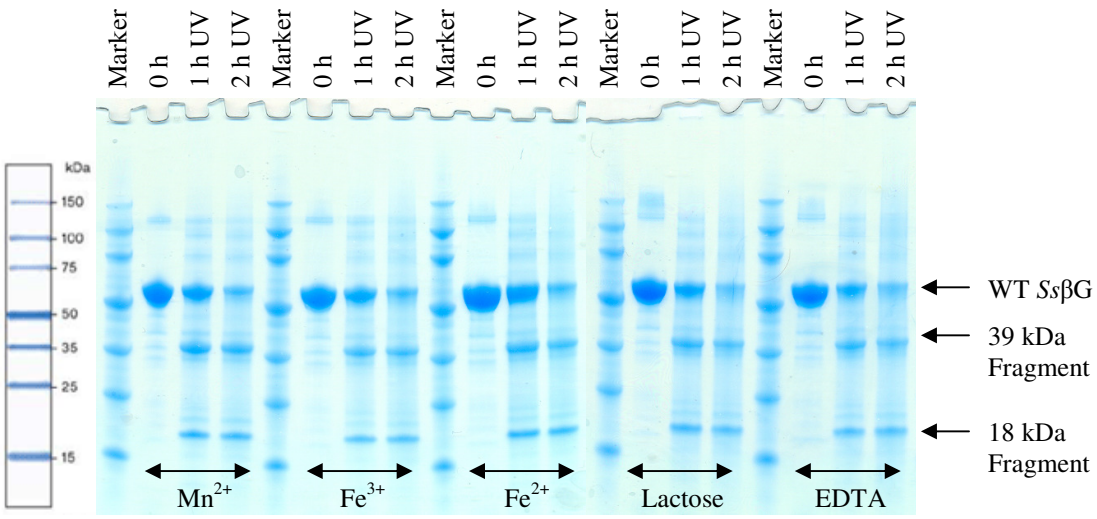
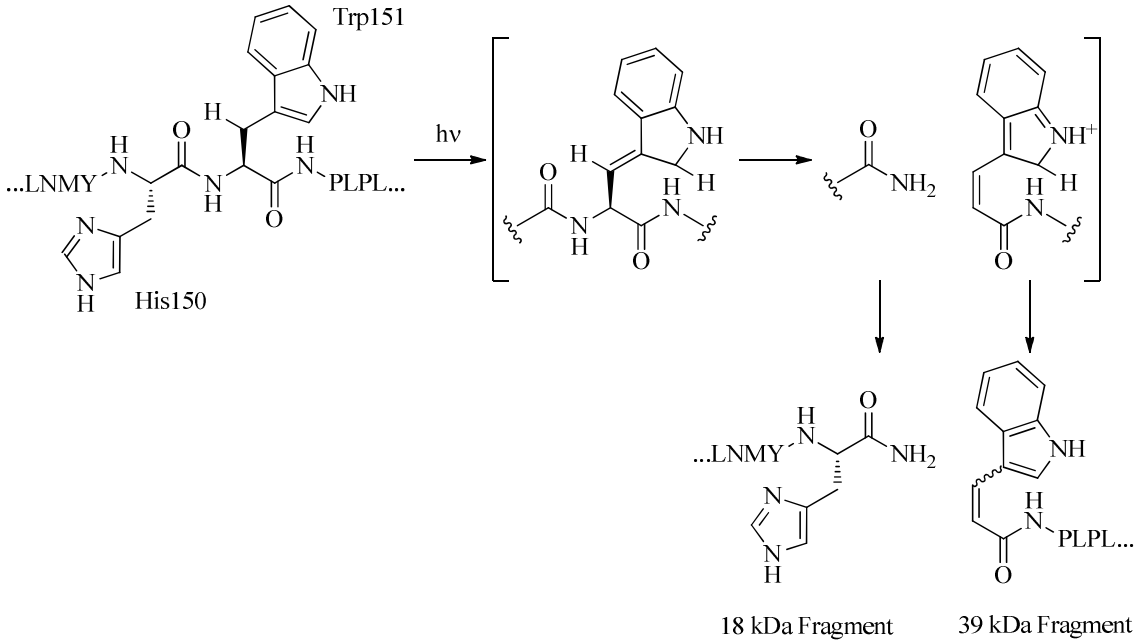



**Figure 66.** SDS-PAGE gel showing the effect of substrate, metal and EDTA on the photo-induced cleavage of WT *Ss*βG.

## 12. Mechanism

A plausible sequence consistent with 240-308 nm light-induced cleavage is shown in Scheme 37: suprafacial 1,3-β-H sigmatropic shift,[339] α-elimination of the 18 kDa fragment and rearomatisation. Nearby critical residues might serve to lower the transition state of energy of the shift and/or provide general base assistance in the elimination. Other mechanisms, *e.g.* photoactivated electron transfer–elimination,[340] cannot be excluded.[291] It should be noted that the secondary structure of *Ss*βG plays an essential role in enabling the reaction mechanism to result in protein cleavage, possibly due to steric or electronic reasons.




**Scheme 37.** Putative photo-induced cleavage mechanism of *Ss*βG.

## 13. Generality of the photocleaved His-Trp motif

The generality of the photocleaved His-Trp motif was investigated through the various methods described in this section and also in section 14 on pg. 162 (Bioinformatic and phylogeneitc analyses). This work was carried out in order to identify other cleaving proteins and also in an attempt to elucidate any potential biological function of this photochemical reaction.

### 13.1. Identification of other photocleavable proteins: 1

Using the atomic coordinates of the motif, all proteins with known 3D structure containing a similarly spatially arranged HWP-triad to that of WT *Ss*βG were identified using SPASM (SPatial Arrangements of Side chains and Main chains).[341] Hits generated with a root mean square deviation (RMSD) < 1.1 are shown in Table 13. The RMSD is a measure of the difference between values predicted by the model and the actual values of the HWP sequence being modelled.

**Table 13.** Hits generated from SPASM search (motif H150-W151-P152 [PBD ID: 1GOW]).

| PDB ID | Protein from *Organism* | Motif | RMSD (Å) |
|---|---|---|---|
| 1GOW | β-Glycosidase from *Sulfolobus solfataricus* | -YHWP- | 0.0 |
| 1QVB | β-Glycosidase from *Thermosphaera aggregans* | -YHWP- | 0.16 |
| 1UWI | Mutated β-Glycosidase from *Sulfolobus solfataricus* | -YHWP- | 0.17 |
| 1UWS | β-Glycosidase from *Sulfolobus solfataricus* in complex with 2-deoxy-2-fluoro-glucose | -YHWP- | 0.18 |
| 1HDH | Arylsulfatase from *Pseudomonas aeruginosa* | -PHWP- | 1.08 |

The β-glycosidase from *Thermosphaera aggregans* (*Ta*βG)[342] and the arylsulfatase from *Pseudomonas aeruginosa*[343] were identified as the proteins possessing an HW fold most similar to that of WT *Ss*βG, and were chosen to be tested for photocleavage activity, as well as the chalcone reductase from *Medicago sativa* (motif -IHWP-),[344] which was identified using the protein data bank (PDB).

### 13.1.1. β-Glycosidase from *Thermosphaera aggregans* (*Ta*βG)

The β-glycosidase from *Thermosphaera aggregans* (*Ta*βG),[345, 346] like *Ss*βG, is an archaic family 1 glycosylhydrolase (GH1) protein. The two proteins have 54% sequence similarity and the tetrapeptide region that corresponds to Y149-H150-W151-P152 in *Ss*βG is identical in *Ta*βG (see pg. 160).

DNA sequencing of the plasmid carrying the *Ta*βG gene, kindly donated by Marco Moracci (Institute of Protein Biochemistry CNR Naples), revealed that in addition to the mutations described by Moracci *et al.*[277] the mutation F51L was also present. *Ta*βG was expressed and purified. The protein identity was confirmed by ESI-MS (found 55355 Da, calcd. 55356 Da) (see pg. 310). Additionally the enzymatic activity of the protein was demonstrated using the X-Gal colorimetric assay.[282]

The photocleavage activity of *Ta*βG was tested, along with the stability of the protein at room temperature. Like WT *Ss*βG, *Ta*βG was found to dissociate between His and Trp when irradiated with UV light to form two daughter fragments that were sharp and clean by SDS-PAGE (Figure 67) and ESI-MS (Figure 68). No cleavage was observed in the absence of UV light (Figure 67). The bands corresponding to protein on the gel marked with a ■ were quantified by densitometry (see pg. 320).

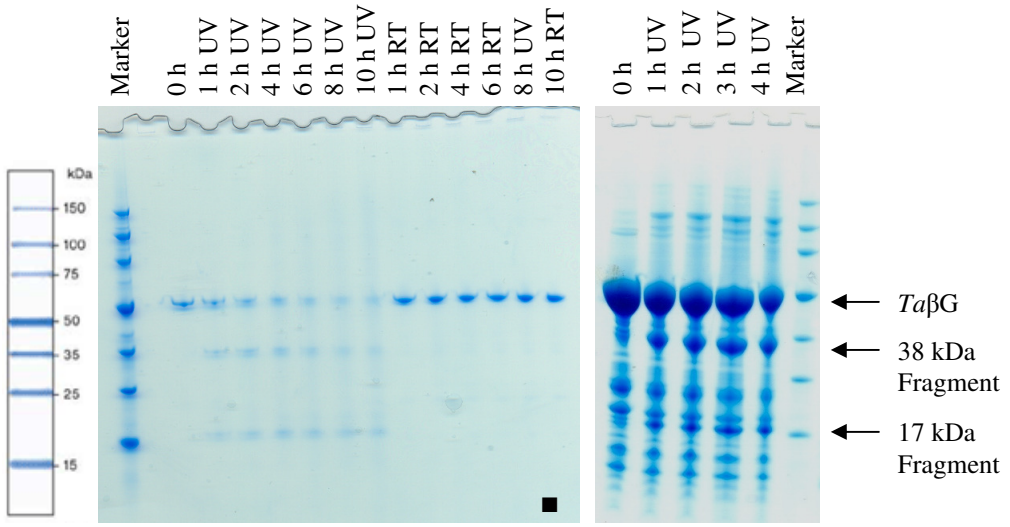



**Figure 67.** Time course analysis of *Ta*βG exposed to UV light by SDS-PAGE. Note that the SDS-PAGE gel on the right is heavily overloaded and therefore bands of impurities are

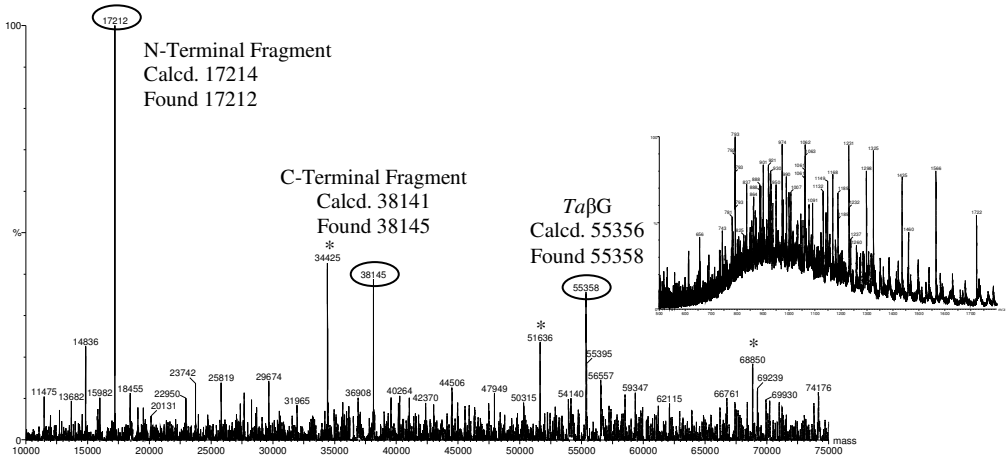



**Figure 68.** Time course analysis of *Ta*βG exposed to UV light by ESI-MS [Raw (insert) and deconvoluted ESI-MS at 1 h]. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

### 13.1.2. Arylsulfatase from *Pseudomonas aeruginosa*

The arylsulfatase from *Pseudomonas aeruginosa*[347-349] and WT *Ss*βG have negligible sequence similarity and the tetrapeptide region that corresponds to Y149-H150-W151-P152 in *Ss*βG is one residue different in the arylsulfatase (PHWP) (see pg. 160).

*E. coli* DH5α harbouring pME4322 plasmid carrying the arylsulfatase gene was kindly donated by Dr Michael Kertesz (University of Manchester). The protein was expressed and purified. The protein identity was confirmed by ESI-MS (found 59768 Da, calcd. 59768 Da) (see pg. 272). In addition, the enzymatic activity of the protein was demonstrated by 4-nitrocatechol release from 4-nitrocatechol sulfate, as reported by Beil *et al*.[343]

The photocleavage activity of the arylsulfatase was tested, along with the stability of the protein at room temperature. Unlike WT *Ss*βG and *Ta*βG, the protein was not found to dissociate between His and Trp when irradiated with UV light (Figure 69) and instead the protein was indiscriminately degraded. Whilst this result does not lend any direct insight into the cleavage mechanism, it does emphasize the fact that both the primary and/ or secondary protein structure must be essential for *Ss*βG protein cleavage to occur. The bands corresponding to protein on the gel marked with a ■ were quantified by densitometry (see pg. 320).

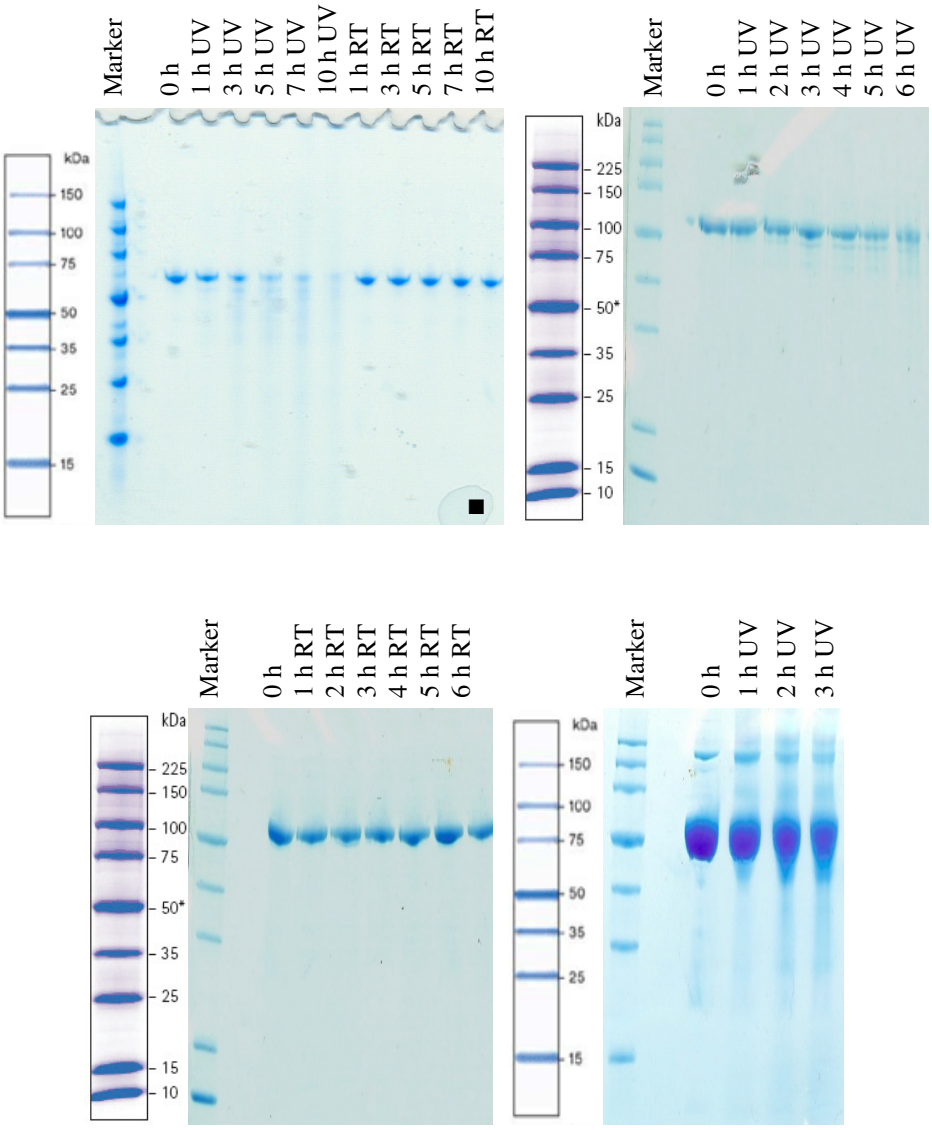



**Figure 69.** Time course analyses of arylsulfatase exposed to/not exposed to UV light by SDS-PAGE.

### 13.1.3. Chalcone reductase from *Medicago sativa*

The chalcone reductase from *Medicago sativa*[344] and WT *Ss*βG have negligible sequence similarity and the tetrapeptide region that corresponds to Y149-H150-W151-P152 in *Ss*βG is one residue different in the chalcone reductase (IHWP) (see pg. 160).

*E. coli* harbouring plasmid carrying the chalcone reductase gene was kindly donated by Joseph Noel (SALK Institute). The N-terminal His-tagged protein was expressed and purified, and the protein identity confirmed by ESI-MS (found 37107 Da, calcd. 37108 Da) (see pg. 274).

The photocleavage activity of the chalcone reductase was tested, along with the stability of the protein at room temperature. Unlike WT *Ss*βG and *Ta*βG, the protein was not found to dissociate between His and Trp when irradiated with UV light (Figure 70) and instead the protein was indiscriminately degraded. The bands corresponding to protein on the gel marked with a ■ were quantified by densitometry (see pg. 320).

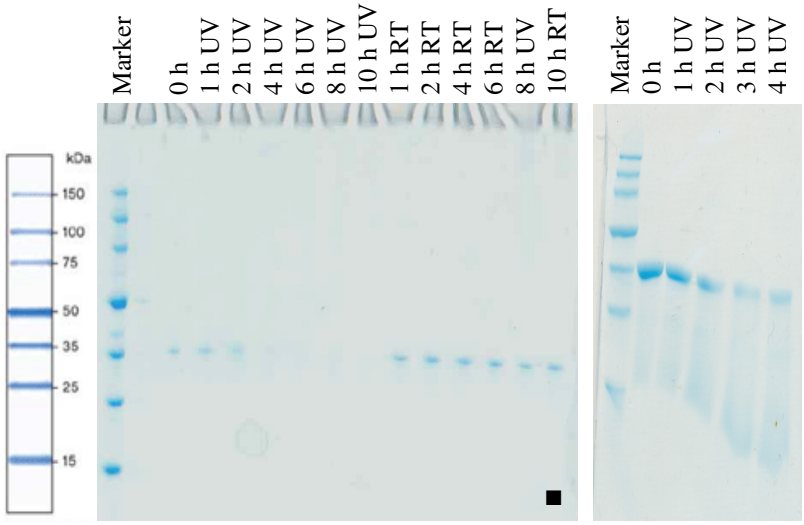



**Figure 70.** Time course analyses of chalcone reductase exposed to/not exposed to UV light by SDS-PAGE.

## 13.2. Identification of other photocleavable proteins: 2

Motif-pattern analysis (using multiple EM for motif elicitation [MEME][350] with hidden Markov model [HMM]) and sequence alignments (using multiple sequence comparison by log-expectation [MUSCLE][351] and carbohydrate-active enzymes [CAZy][352]) revealed ~1800 proteins with the highest similarity across a variety of species (see pg. 162 for all bioinformatic analysis). Phylogenetic analysis of this entire grouping (using Quick Neighbor Joining algorithm) based on clustering of the full protein sequence revealed branches focused according to the tetrapeptide region that correspond to Tyr149-His150-Trp151-Pro152 in WT *Ss*βG. Representative proteins from these branches were evaluated for photocleavage activity. Together these analyses correctly predicted other cleaving proteins (Y/FHWP, *e.g.*, *Pf*βG, Y149F *Ss*βG) and revealed branches and motifs with weak (YHWD, *e.g.*, *Re*βG, *Ct*βG) and no cleavage activity (HHFD/FHWD, *e.g.*, *Sp*βG, *Sa*M). Each protein is discussed in turn below and a summary of all proteins exposed to UV light is shown on pg. 160.

### 13.2.1. β-Glucosidase from *Pyrococcus furiosus* (*Pf*βG)

The archaic GH1 protein β-glycosidase from *Pyrococcus furiosus* (*Pf*βG) was kindly donated by Dr. Servé Kengen (Wageningen University) and was used without further purification. The tetrapeptide region that corresponds to Y149-H150-W151-P152 in *Ss*βG is identical in *Pf*βG (see pg. 160). The protein identity was confirmed by ESI-MS (found 54610 Da, calcd. 54605 Da) (see pg. 284). The photocleavage activity of *Pf*βG was tested. Like *Ss*βG, *Pf*βG was found to dissociate between His and Trp when irradiated with UV light to form two daughter fragments that were sharp and clean by SDS-PAGE (Figure 71) and ESI-MS (Figure 71).

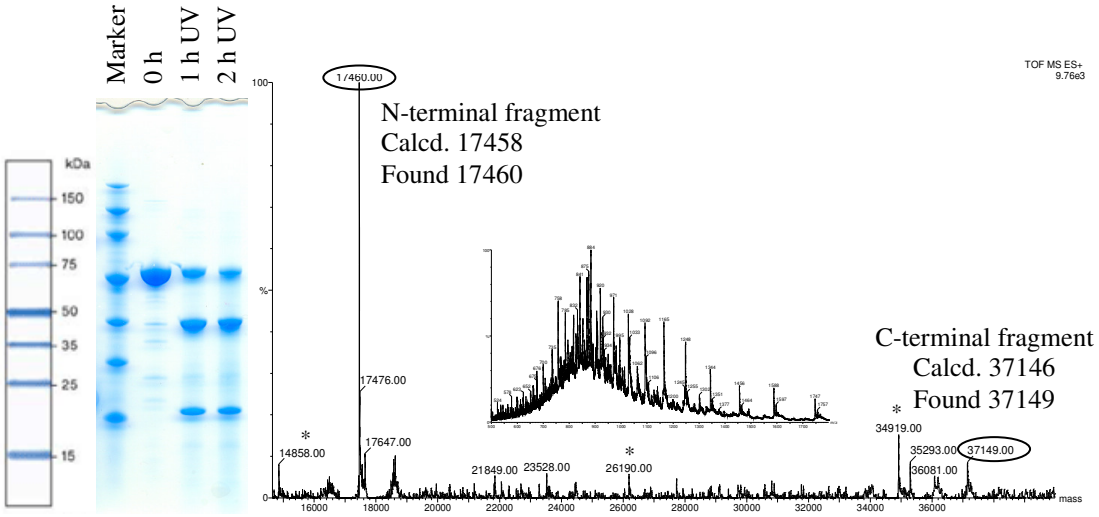



**Figure 71.** Time course analysis of *Pf*βG exposed to UV light by SDS-PAGE (left) and ESI-MS [Raw (insert) and deconvoluted ESI-MS at 1 h] (right). * = Artifact of the MaxtEnt1 deconvolution algorithm.[283-285]

### 13.2.2. β-Glycosidase from *Sulfolobus solfataricus* (Y149F *Ss*βG)

Starting from a pET24d plasmid encoding WT *Ss*βG, site-directed mutagenesis was used to generate a pET24d plasmid encoding Y149F *Ss*βG. DNA sequencing was used to confirm that the desired mutation had been introduced. N-terminal His-tagged Y149F *Ss*βG was expressed and purified, and the protein identity confirmed by ESI-MS (found 57398 Da, calcd. 57398 Da) (see pg. 289). In addition, the enzymatic activity of the protein was demonstrated using the X-Gal colorimetric assay.[282] The photocleavage activity of Y149F *Ss*βG was tested, along with the stability of the protein at room temperature. Like WT *Ss*βG, Y149F *Ss*βG was found to dissociate between His and Trp when irradiated with UV light to form two daughter fragments that were visible by SDS-PAGE (Figure 72). No cleavage was observed in the absence of UV light. The bands corresponding to protein on the gel marked with a ■ were quantified by densitometry (see pg. 320).

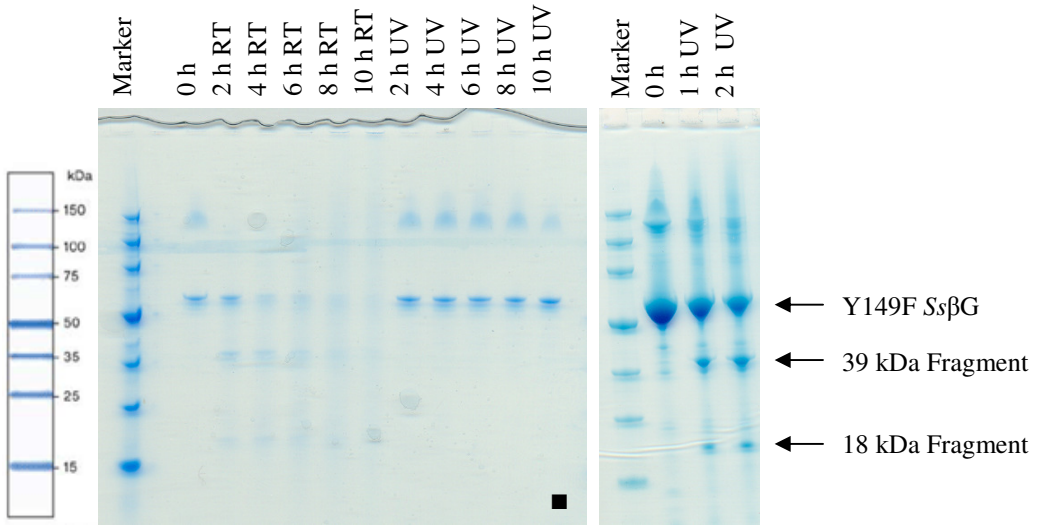



**Figure 72.** Time course analysis of Y149F *Ss*βG exposed to UV light by SDS-PAGE.

### 13.2.3. β-Glucosidase from *Clostridium thermocellum* (*Ct*βG) and β-Glucosidase from *Rhizobium etli* (*Re*βG)

The bacterial GH1 proteins β-glycosidases from *Clostridium thermocellum* (*Ct*βG) and *Rhizobium etli* (*Re*βG) were obtained from a commercial source and were used without further purification. The tetrapeptide region that corresponds to Y149-H150-W151-P152 in *Ss*βG is one residue different in both *Ct*βG and *Re*βG (YHWD) (see pg. 160). The protein identities were confirmed by ESI-MS (*Ct*βG: found 52568 Da, calcd. 52574 Da. *Re*βG: found 53158 Da, calcd. 53153 Da) (see pgs. 276 and 286, respectively). The photocleavage activity of *Ct*βG and *Re*βG were tested. Like *Ss*βG, *Ct*βG was found to dissociate between His and Trp when irradiated with UV light to form two daughter fragments that were visible by SDS-PAGE (Figure 73) and ESI-MS (Figure 73).

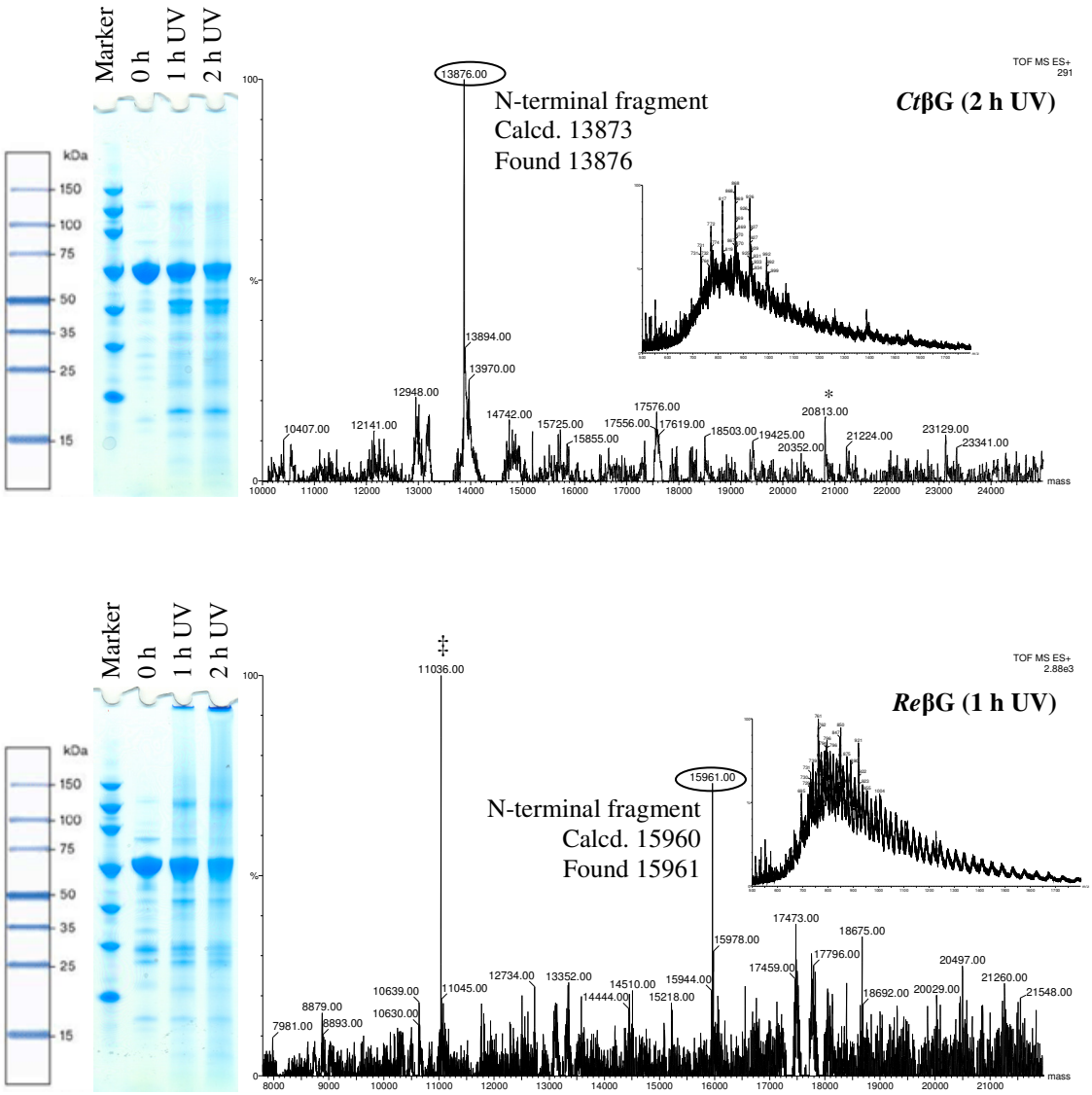



**Figure 73.** Time course analysis of *Re*βG and *Ct*βG exposed to UV light by SDS-PAGE (left) and ESI-MS [Raw (insert) and deconvoluted ESI-MS] (right). * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]. ‡ = Sub-fragment of the N-terminal fragment induced in the mass spectrometer.

### 13.2.4. Myrosinase from *Sinapis alba* (*Sa*M)

The eukaryotic GH1 protein myrosinase from *Sinapis alba* (*Sa*M) was obtained from a commercial source and was used without further purification. The tetrapeptide region that corresponds to Y149-H150-W151-P152 in *Ss*βG is two residues different in *Sa*M (FHWD) (see pg. 160). The enzymatic activity of the protein was demonstrated using the X-Gal colorimetric assay.[282] The photocleavage activity of *Sa*M was tested and unlike *Ss*βG, *Sa*M was found not to dissociate between His and Trp when irradiated with UV light (Figure 74).

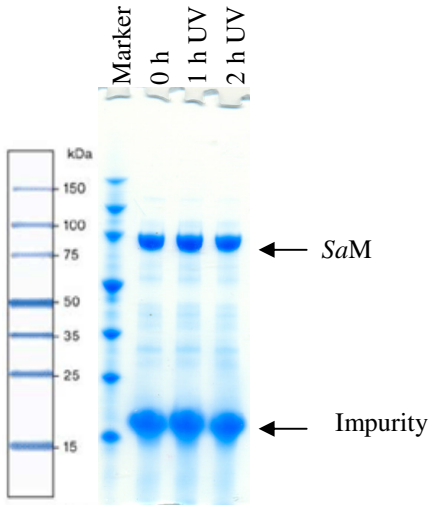



**Figure 74.** Time course analysis of *Sa*M exposed to UV light by SDS-PAGE.

### 13.2.5. β-Glucosidase from *Streptococcus pyogenes* (*Sp*βG)

The bacterial GH1 protein β-glycosidase from *Streptococcus pyogenes* (*Sp*βG) was obtained from a commercial source and was used without further purification. The tetrapeptide region that corresponds to Y149-H150-W151-P152 in *Ss*βG is three residues different in *Sp*βG (HHFD) (see pg. 160). The protein identity was confirmed by ESI-MS (found 57183 Da, calcd. 57183 Da) (see pg. 288). The photocleavage activity of *Sp*βG was tested. Unlike *Ss*βG, *Sp*βG was not found to dissociate between His and Trp when irradiated with UV light (Figure 75).

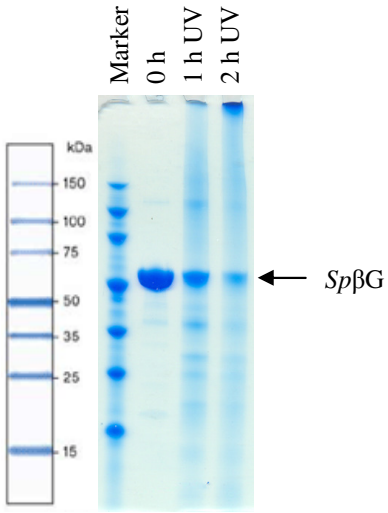
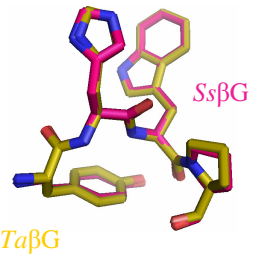
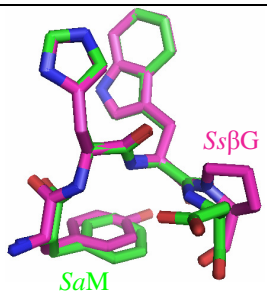
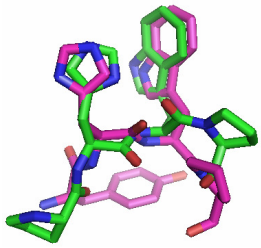
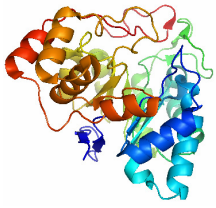
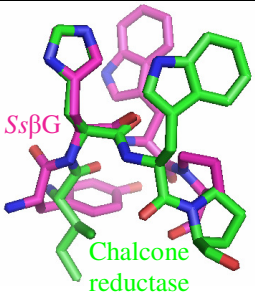



**Figure 75.** Time course analysis of *Sp*βG exposed to UV light by SDS-PAGE.

## Summary of proteins exposed to UV light

| **Protein from *Organism* [GenBank#]** | **Crystal Structure + PDB Code** | ***Ss*βG/Protein Overlay** | **Motif** | **Cleavage** |
|---|---|---|---|---|
| β-Glycosidase from *Sulfolobus solfataricus* (*Ss*βG) [AAA72843] |  PDB ID: 1GOW | --- | -YHWP- | ☑ |
| β-Glycosidase from *Thermosphaera aggregans* (*Ta*βG) [AAD43138] |  PDB ID: 1QVB |   | -YHWP- | ☑ |
| β-Glucosidase from *Pyrococcus furiosus* (*Pf*βG) [AAC25555] | --- | --- | -YHWP- | ☑ |

| | | | | |
|---|---|---|---|---|
| β-Glucosidase from *Clostridium thermocellum* (*Ct*βG) [CAA42814] | --- | --- | -YHWD- | ☑ (slow) |
| β-Glucosidase from *Rhizobium etli* (*Re*βG) [ABC92395] | --- | --- | -YHWD- | ☑ (slow) |
| β-Glucosidase from *Streptococcus pyogenes* (*Sp*βG) [AAM79900] | --- | --- | -HHFD- | ☒ |
| Myrosinase from *Sinapis alba* (*Sa*M) [CAA42534] |  PDB ID: 1DWA |   | -FHWD | ☒ |
| Arylsulfatase from *Pseudomonas aeruginosa* [CAA88421] |  PDB ID: 1HDH |   | -PHWP- | ☒ |
| Chalcone reductase from *Medicago sativa* [CAA57782] |  PDB ID: 1ZGD |   | -IHWP- | ☒ |

## 14. Bioinformatic and phylogenetic analyses

Bioinformatic and phylogenetic analyses were provided by Mr. Steve Taylor (Computational Biology Research Group, University of Oxford) and Prof. Dmitry Filatov (Dept. Plant Sciences, University of Oxford).

### 14.1. Multiple sequence alignments

Multiple sequence alignments of (i) 1654 archaic, bacterial and eukaryotic GH1 proteins, (ii) 49 archaic GH1 proteins, (iii) 1204 bacterial GH1 proteins and (iv) 395 eukaryotic GH1 proteins, obtained from the CAZy database (http://www.cazy.org/fam/GH1.html),[352] were made using MUSCLE.[351] These are available at http://www.cbrg.ox.ac.uk/data.shtml. It is recommended that the alignments are viewed in Jalview[353] and the appropriate jar data file is included. The core motif is marked in red. After inspection of alignment (i), partial peptides and proteins that showed poor conservation were removed and the sequences were realigned. This left 1582 sequences in alignment (v). In alignment (v) there was a strong conserved domain containing the conserved H residue[277] shown in Figure 76 across all sequences.

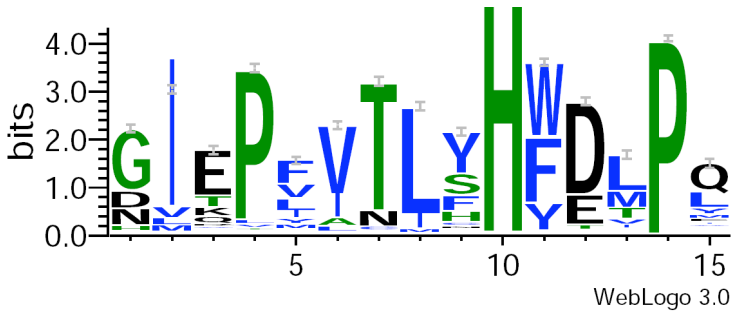



**Figure 76.** Sequence logo[354, 355] generated from 1582 aligned GH1 sequences (alignment (v)) that were trimmed to 15 amino acids.

This 15 amino acid peptide region was then extracted for all GH1 sequences and a HMM using HMMER (http://hmmer.janelia.org/)[356] was constructed. Next the Genpept database, excluding the GH1 protein set, was searched to look for novel putative photocleavable proteins. The list is available at http://www.cbrg.ox.ac.uk/data.shtml. Lastly, 206 proteins were selected from this new list, all with score $\geq$ 30.0 / E $\leq$ 0.0068, and were realigned with the GH1 proteins to generate alignment (vi).

### 14.2. MEME analysis

MEME analysis[350] of 1654 archaic, bacterial and eukaryotic GH1 protein sequences is available at: http://www.cbrg.ox.ac.uk/data/photocleave/meme_allGH1/meme.html. The conserved H residue[277] was found in motif 2 (1533 sequences).

### 14.3. Phylogenetic analysis

The evolutionary relationships of 49 archaic, 1129 bacterial, 295 eukaryotic, and 1654 archaic, bacterial and eukaryotic GH1 proteins obtained from the CAZy database (http://www.cazy.org/fam/GH1.html),[352] are shown in Figure 124, Figure 125, Figure 126 and Figure 127, respectively. In addition, the evolutionary relationships of the 1860 proteins (archaic, bacterial and eukaryotic GH1 proteins and additional proteins) from alignment (vi) are shown in Figure 77, and is also available to view and download at http://www.cbrg.ox.ac.uk/data.shtml. It is recommended that the file is viewed in Dendroscope.[357] The phylogenetic relationship was inferred using the Quick Neighbour-Joining algorithm as part of the Topali 2 package (http://www.topali.org/).[358] Phylogenetic visualisation was conducted in Dendroscope.[357] An enlarged view of the arm containing the cleaving proteins (*Ss*βG, *Ta*βG, *Pf*βG) is shown in Figure 78.

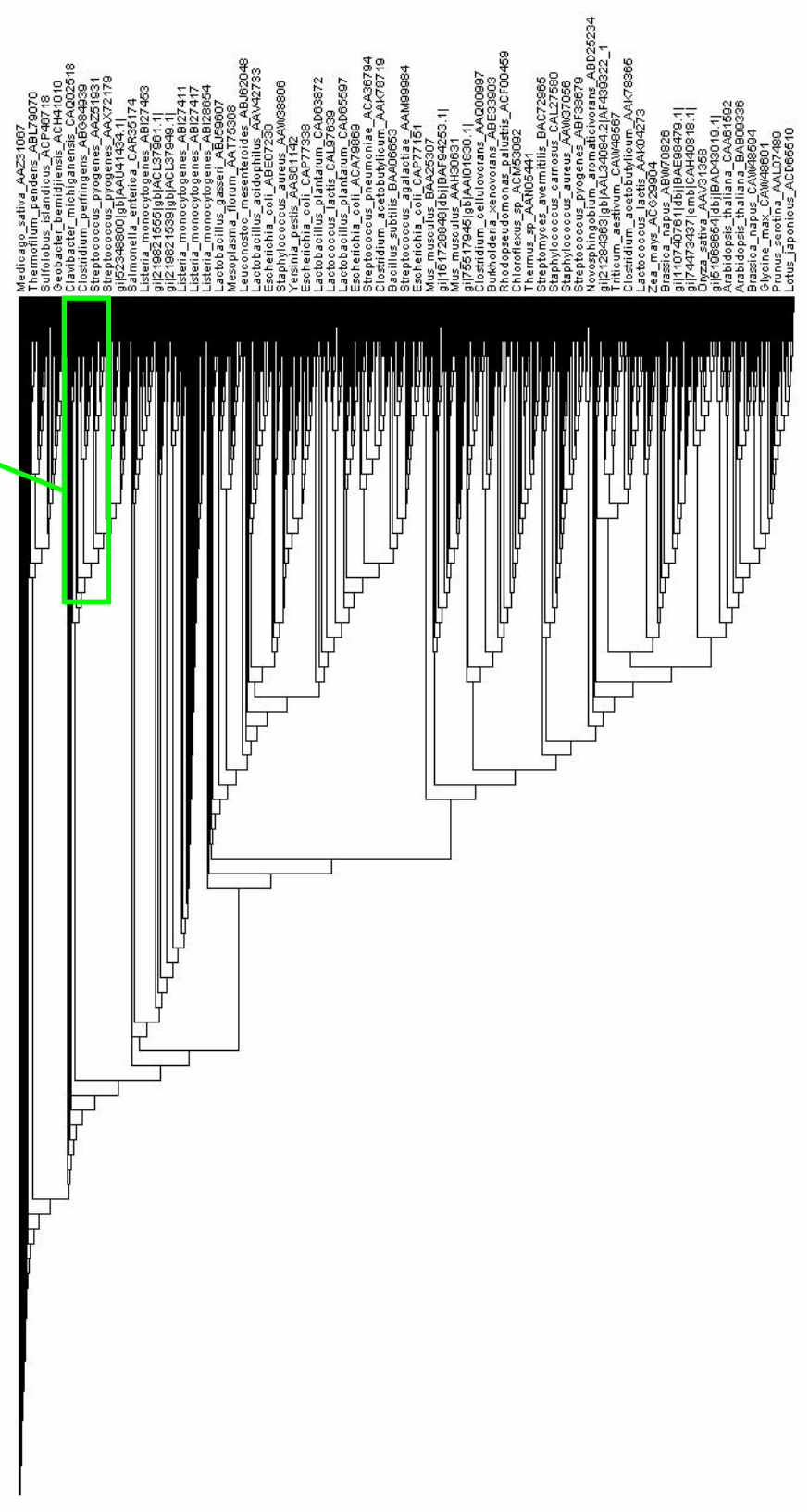



**Figure 77.** Phylogenetic analysis of 1860 proteins (archaic, bacterial and eukaryotic GH1 proteins, and additional proteins).

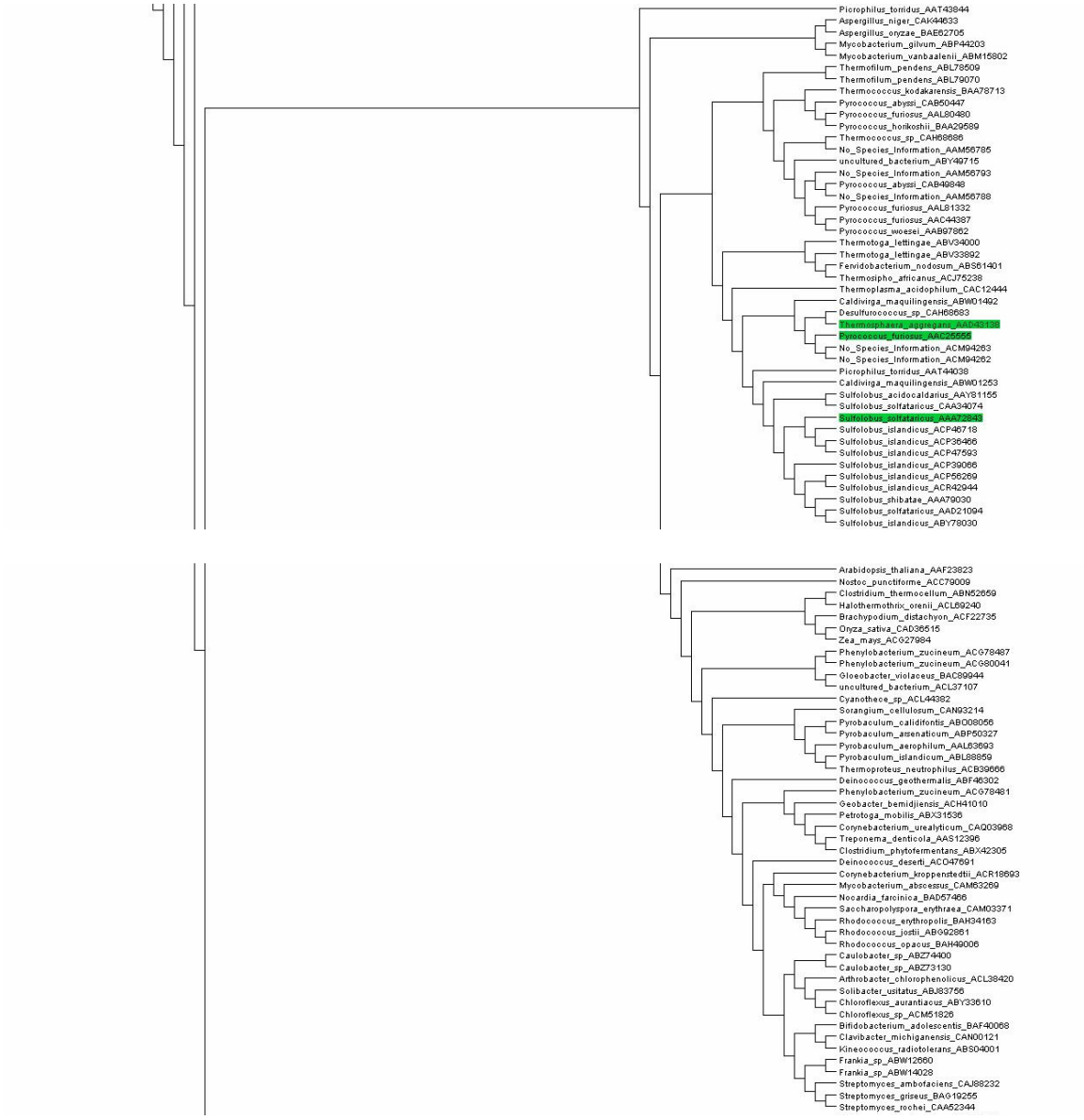



**Figure 78.** An enlarged view from Figure 77 of the arm containing the cleaving proteins (*Ss*βG, *Ta*βG and *Pf*βG).

## 15. Conclusion and future work

In conclusion, while investigating the photochemical modification of proteins described in chapter 2, a novel and efficient photo-induced site-selective protein cleavage reaction present in a natural motif of some GH1 proteins was unexpectedly discovered. Remarkably, the β-glycosidase from *Sulfolobus solfataricus* (*Ss*βG) was found to form two daughter fragments following UV irradiation (240-308 nm).

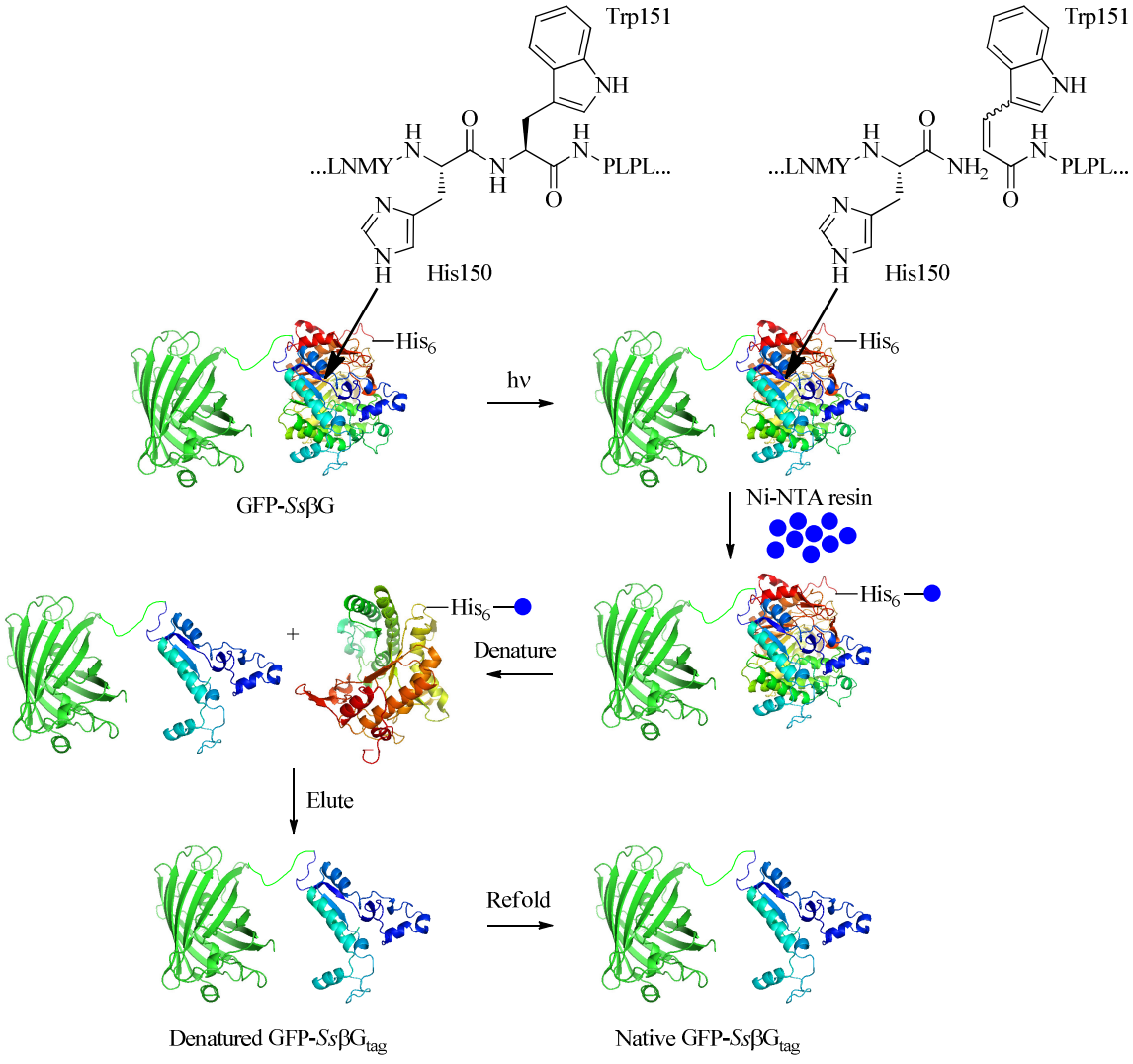
Alterations at and alanine-scanning of a sphere of residues around the cleavage site were used to establish which residues were essential for the reaction to occur and extensive analysis of the fragments allowed a reaction mechanism to be proposed. The proposed reaction mechanism comprises a suprafacial 1,3-β-H sigmatropic shift, α-elimination of the 18 kDa fragment and rearomatisation. However, other mechanisms such as photoactivated electron transfer–elimination were not excluded. The propensity of other proteins that contained the same or similar motif to undergo photo-induced cleavage was also investigated and identified more cleaving proteins (*e.g. Ta*βG, *Pf*βG, *Re*βG and *Ct*βG), in addition to several non-cleaving proteins (*e.g.* arylsulfatase from *Pseudomonas aeruginosa,* chalcone reductase from *Medicago sativa, Sp*βG and *Sa*M). Interestingly, the cleaving proteins are typically found in environments naturally devoid of UV light, which might imply a yet to be discovered biological function or the maintenance of this photosensitive motif in the absence of UV stability as a selection pressure. Perhaps, as is the case for the inducible-chromophore proteins Kaede and EosFP, a clear biological role does exist.

In the future, further investigations should be carried out to determine the biological function, if any, of this protein photocleavage reaction as it is not yet clear. Other future work could include obtaining additional evidence to support the putative photo-induced cleavage mechanism of *Ss*βG shown on pg. 148 and optimising the reaction to improve the yield.

# Chapter 4: Cleaved-Tag Affinity Purification Using the Protein Cleavage Reaction

## 1. Introduction

This chapter describes the application of the aforementioned photocleavage reaction to the development of a cleaved-tag affinity purification method using a GFP-*Ss*βG fusion as the model protein. A schematic overview is shown in Scheme 38.




**Scheme 38.** Schematic overview of cleaved-tag affinity purification of GFP-*Ss*βG.

## 2. *Ss*βG truncated proteins (T0-, T50-, T100- and T140-*Ss*βG)

A primary objective of this purification method was to minimise the size of the *Ss*βG-($His_6$)-tag fused to GFP subsequent to photo-release. To this end, a mutant of WT *Ss*βG in which the N-terminal His-tag was moved to the C-terminus was generated, named T0-*Ss*βG. This strategy dramatically reduced the size of the *Ss*βG-($His_6$)-tag that would be fused to GFP subsequent to photo-release from 39 kDa to 18 kDa.

Starting from a pET24d plasmid encoding WT *Ss*βG, a T0-*Ss*βG plasmid DNA insert was generated and then ligated into a pET24a(+) vector. The T0-*Ss*βG plasmid was successfully sequenced and the C-terminal His-tagged protein was expressed and purified. The protein identity was confirmed by ESI-MS (found 58888 Da, calcd. 58886) (see pg. 308). In addition, the enzymatic activity of T0-*Ss*βG was demonstrated using the X-Gal colorimetric assay.[282]

The photocleavage activity of T0-*Ss*βG was tested, along with the stability of the protein at room temperature. Like WT *Ss*βG, T0-*Ss*βG was found to dissociate between His and Trp when irradiated with UV light to form two daughter fragments that were sharp and clean by SDS-PAGE and ESI-MS (Figure 79). No cleavage was observed in the absence of UV light (Figure 79). The bands corresponding to protein on the gel marked with a ■ were quantified by densitometry (see pg. 320).

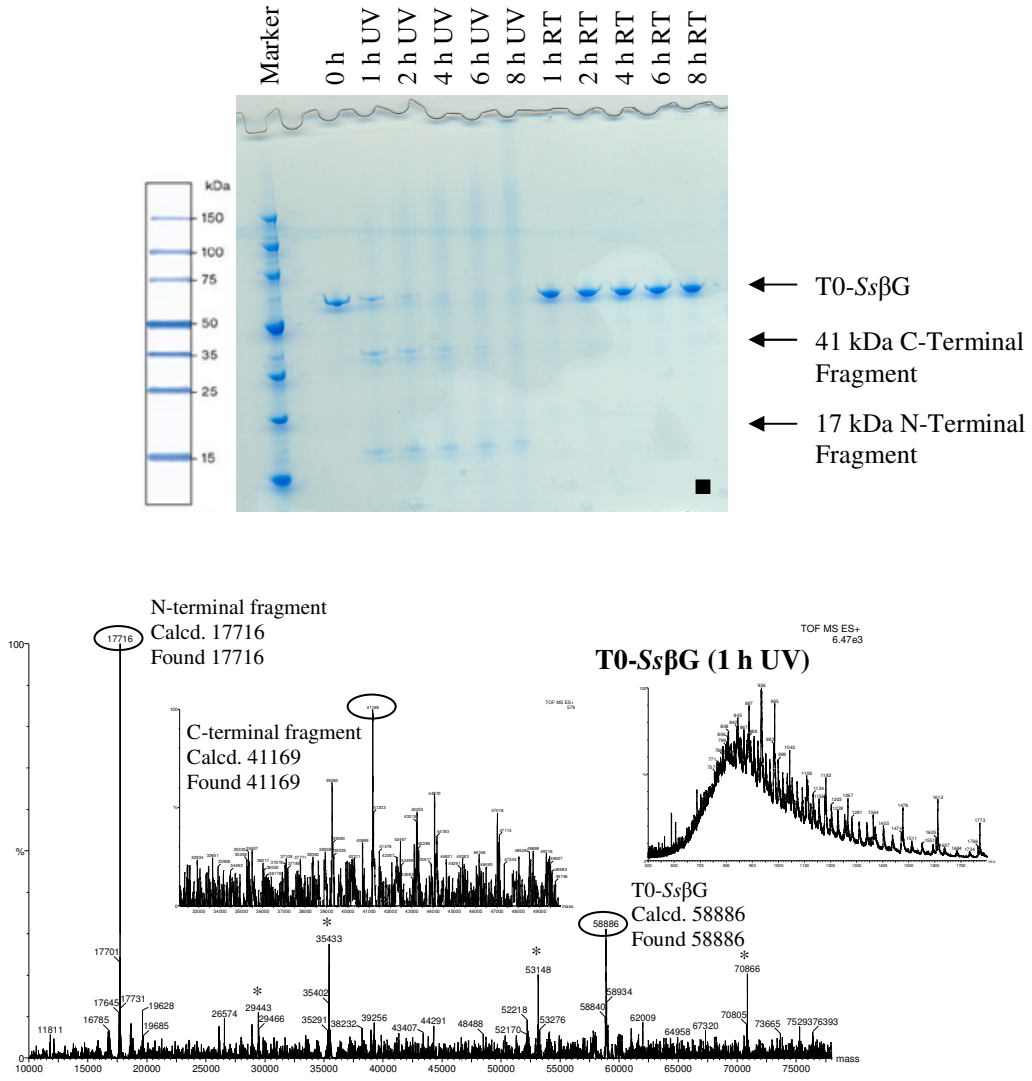
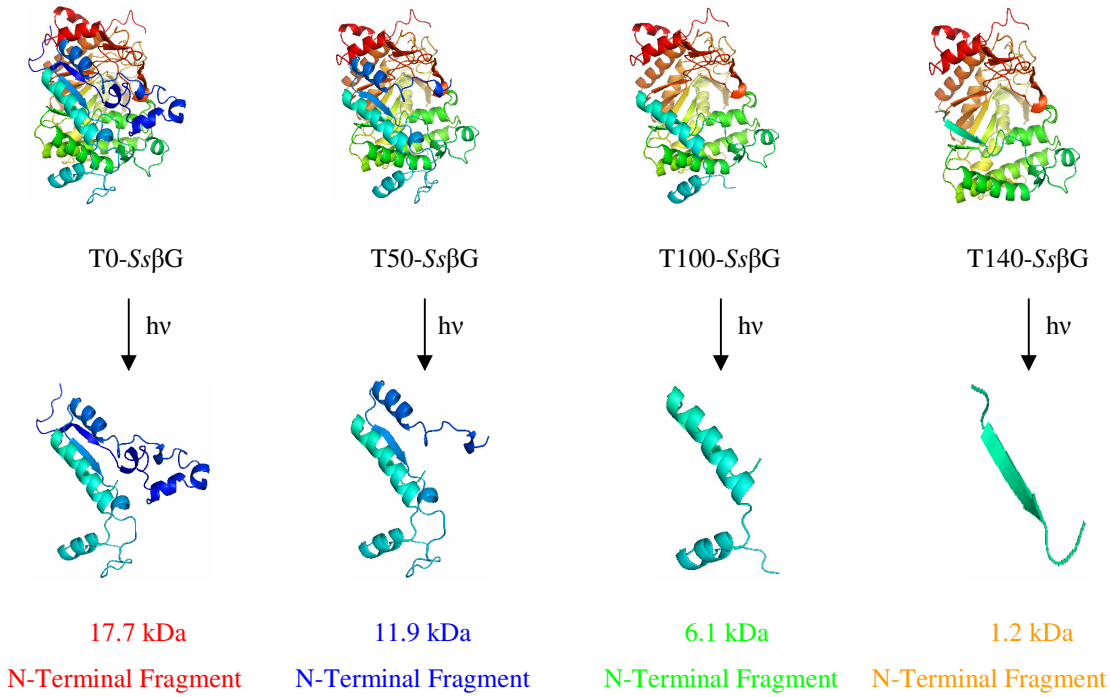



**Figure 79.** Time course analysis of T0-*Ss*βG exposed to/not exposed to UV light by SDS-PAGE (top) and ESI-MS [Raw (insert) and deconvoluted ESI-MS] (bottom). * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

The generation of three truncated T0-*Ss*βG proteins, named T50-, T100- and T140-*Ss*βG, was next attempted in order to reduce the size of the *Ss*βG-($His_6$)-tag further (Figure 80). The numbers 50, 100 and 140 refer to the number of residues removed from the N-terminus of T0-*Ss*βG.



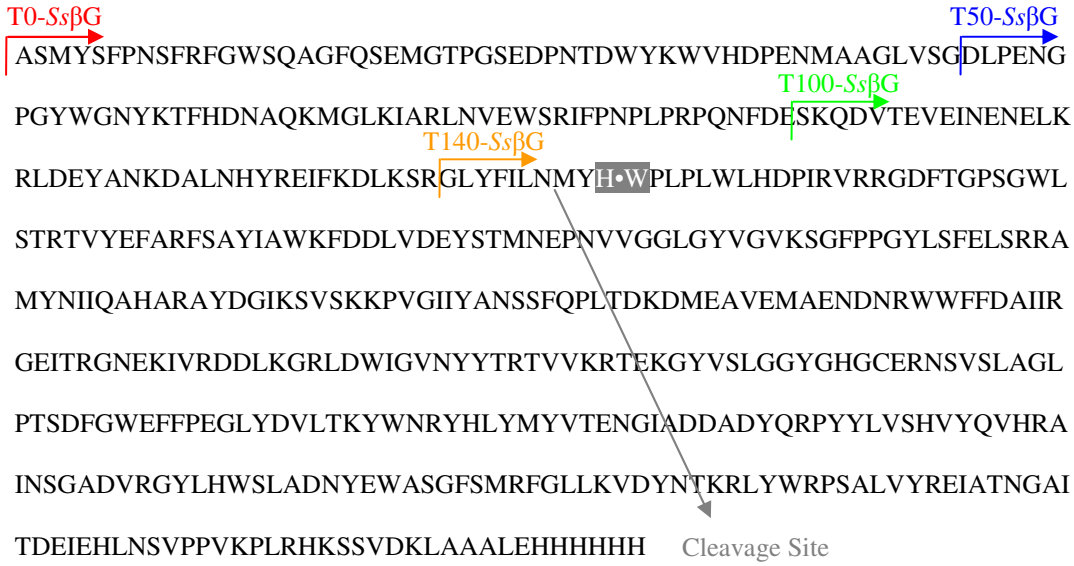




**Figure 80.** Photo-induced cleavage of T0-, T50, T100- and T140-*Ss*βG (top). Primary protein sequence of T0-*Ss*βG (bottom).

Starting from a pET24d plasmid encoding WT *Ss*βG, T50-, T100- and T140-*Ss*βG plasmid DNA inserts were generated and then each ligated into a pET24a(+) vector. The three plasmids were successfully sequenced. Expression and purification of these C-terminal His-tagged proteins was attempted, but despite employing a number of

different expression conditions and purification methods, pure T50-, T100- and T140-*Ss*βG were never obtained. However, the first elution fractions from the manual purification of T50-, T100- and T140-*Ss*βG were shown by anti-His western blot to contain the proteins (Figure 81). The identities of T50-, T100- and T140-*Ss*βG were not able to be confirmed by ESI-MS, probably because the proteins were not pure.

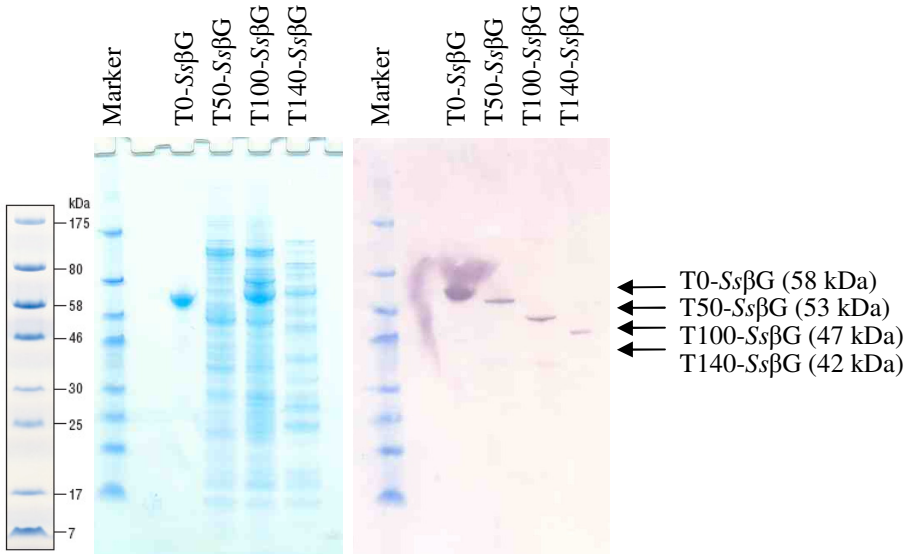



**Figure 81.** SDS-PAGE (left) and corresponding anti-His western blot (right) of the first elution fraction from the manual purification of T50-, T100- and T140-*Ss*βG, as well as T0-*Ss*βG.

The photocleavage activities of T50-, T100- and T140-*Ss*βG were tested, which was assessed for by SDS-PAGE and anti-His western blot. Unlike T0-*Ss*βG, the proteins were not found to dissociate between His and Trp when irradiated with UV light (Figure 82), and as a result T0-*Ss*βG was chosen as the protein to be the *Ss*βG-($His_6$)-tag.

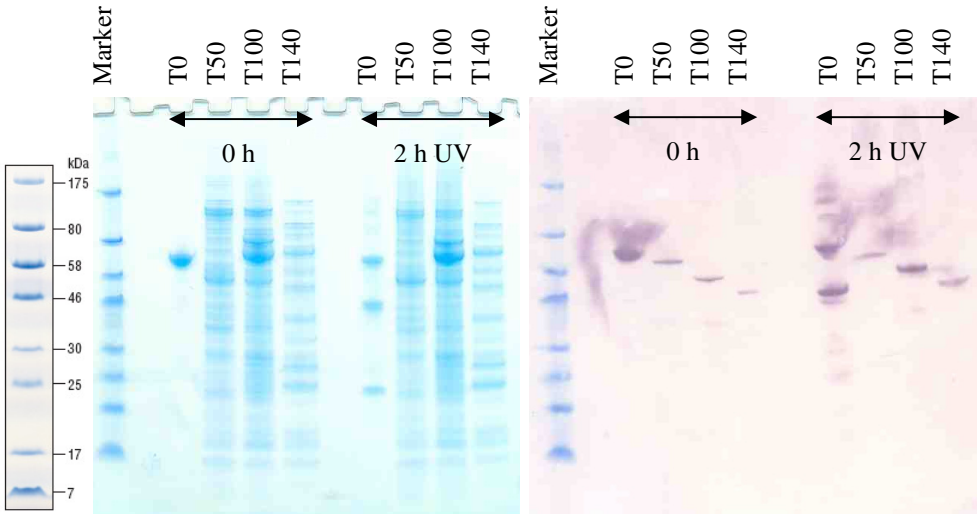



**Figure 82.** Time course analysis of T0-, T50-, T100- and T140-*Ss*βG exposed to UV light by SDS-PAGE (left) and corresponding anti-His western blot (right).

## 3. Development of the cleaved-tag affinity purification method

In order to identify an efficient method to separate the 17 kDa N-terminal non-His-tagged fragment from the photocleavage of T0-*Ss*βG that in due course would be fused to GFP, from the 41 kDa C-terminal His-tagged fragment and His-tagged uncleaved T0-*Ss*βG. For this purpose, the His-tag was exploited and nickel affinity chromatography used.

T0-*Ss*βG was irradiated with UV light to induce protein cleavage and the reaction mixture was purified by nickel affinity chromatography using a HisTrap FF column. Analysis by SDS-PAGE of the eluent collected from the purification revealed that separation of the 17 kDa N-terminal non-His-tagged fragment from the 41 kDa C-terminal His-tagged fragment and His-tagged uncleaved T0-*Ss*βG was not achieved (Figure 83). It was reasoned that this was because the two fragments resulting from the photocleavage of T0-*Ss*βG remained intact subsequent to fragmentation of the protein.

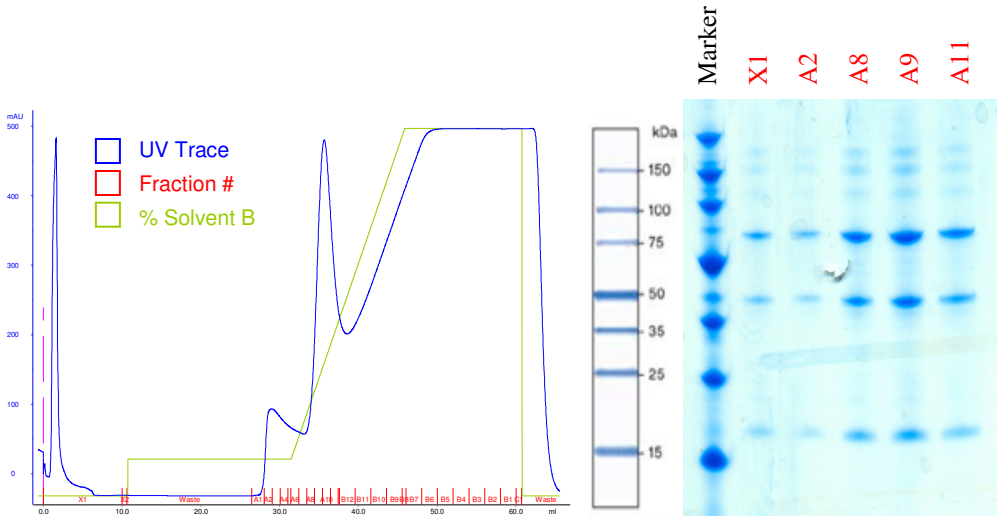



**Figure 83.** FPLC trace (left) and corresponding SDS-PAGE (right) for the cleaved-tag affinity purification of T0-*Ss*βG.

In order to separate the two fragments resulting from the photocleavage of T0-*Ss*βG, protein denaturation methods were employed. Herein, WT *Ss*βG was used in place of T0-*Ss*βG as the model protein due to it being readily available.

Initially, protein denaturation by heating was attempted: WT *Ss*βG was irradiated with UV light to induce protein cleavage and the reaction mixture containing the 17 kDa N-terminal non-His-tagged fragment, 41 kDa C-terminal His-tagged fragment and His-tagged uncleaved T0-*Ss*βG was then heated to 100°C. Unfortunately, the proteins precipitated, making this denaturation technique unsuitable to use for the cleaved-tag affinity purification method.

Subsequently, sodium dodecyl sulphate (SDS) in combination with dithiothreitol (DTT), and urea, were employed: WT *Ss*βG was irradiated with UV light to induce protein cleavage and then to the reaction mixture was added either 2% w/v SDS and 200 mM dithiothreitol, or 8 M urea. The former reaction mixture was also heated to 100°C. The reaction mixtures were purified by nickel affinity chromatography using a

HisTrap FF column, but analysis by SDS-PAGE of the eluents collected from the purifications revealed that separation of the 39 kDa C-terminal non-His-tagged fragment from the 18 kDa N-terminal His-tagged fragment and His-tagged uncleaved WT *Ss*βG was not achieved. In the case of urea, this was despite supplementing the binding and elution buffers with 8 M urea. Separation of the products from the photocleavage of WT *Ss*βG was also not accomplished using hydrophobic interaction chromatography, in which the denaturant ammonium sulphate was present in the elution buffer. As *Ss*βG is reported to be highly stable to denaturing conditions,[359-362] it was not unsurprising to find that photocleaved *Ss*βG proved difficult to denature.

Next, the strong denaturant guanidine hydrochloride (GnHCl) was used: WT *Ss*βG was irradiated with UV light to induce protein cleavage and then 7 M GnHCl was added. The reaction mixture was incubated overnight before being purified by nickel affinity chromatography using a HisTrap FF column. Protein binding and elution was performed using buffers supplemented with 7 M GnHCl. Pleasingly, analysis by SDS-PAGE of the eluent collected from the purification revealed that separation of the 39 kDa C-terminal non-His-tagged fragment (fractions A1-A2 in Figure 84) from the 18 kDa N-terminal His-tagged fragment (fractions A6-B12 in Figure 84) was achieved. As GnHCl forms a precipitate with SDS, it proved necessary to purify any GnHCl-containing samples prior to analysis by SDS-PAGE, as otherwise protein bands were not observed. This was done using the Pierce SDS-PAGE Sample Prep Kit.

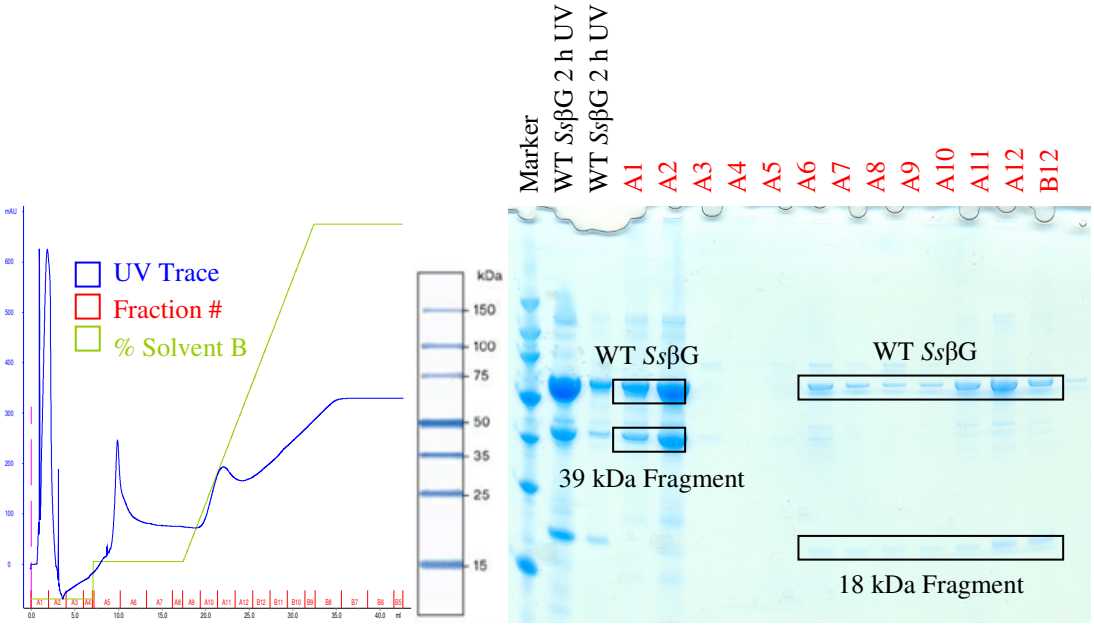



**Figure 84.** FPLC trace (left) and corresponding SDS-PAGE (right) for the cleaved-tag affinity purification of WT *Ss*βG when using 7 M GnHCl as the denaturant.

As seen in Figure 84, the fractions containing the 39 kDa C-terminal non-His-tagged fragment also contained His-tagged uncleaved WT *Ss*βG. It was identified that this was due to the presence of GnHCl, which is known to reduce the binding capacity of Ni-NTA agarose resin and was not because of overloading of the column: non-irradiated WT *Ss*βG that was (i) not denatured and (ii) denatured using 7 M GnHCl, was purified by nickel affinity chromatography using a HisTrap FF column. WT *Ss*βG was only recovered in the flow-through fraction for the sample (ii) (Figure 85). Additionally, a new protein band was observed below the WT *Ss*βG band in the flow-through fraction (Figure 85). Attempts to determine its identity, *e.g.* by ESI-MS, were unsuccessful.

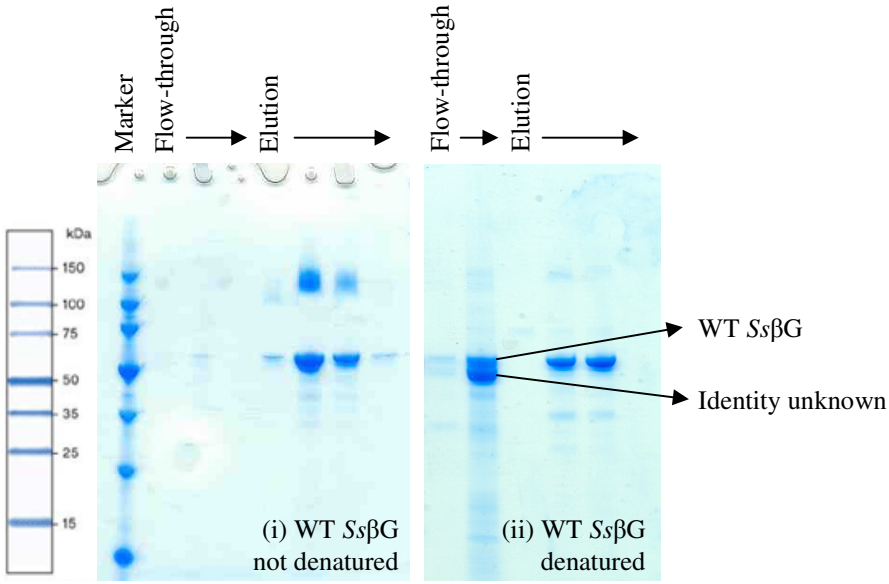



**Figure 85.** SDS-PAGEs for the purification of WT *Ss*βG not denatured/denatured using 7 M GnHCl.

The purification method for the photocleaved WT *Ss*βG reaction mixture using 7 M GnHCl as the denaturant was optimised so as to reduce the quantity of His-tagged uncleaved WT *Ss*βG present in the flow-through fraction alongside the 39 kDa C-terminal non-His-tagged fragment. The strategies used here to do this were targeted towards improving the binding efficiency of His-tagged uncleaved WT *Ss*βG to the Ni-NTA agarose resin.

Significantly, incubation of the photocleaved WT *Ss*βG reaction mixture with Ni-NTA agarose resin and 7 M GnHCl overnight prior to purification proved superior to the previous protocol used, in which the reaction mixture was incubated overnight with only 7 M GnHCl and was subsequently purified using a HisTrap FF column. Clearly this was because the His-tagged uncleaved WT *Ss*βG was in contact with the resin for a longer period of time. An identical outcome was accomplished when (i) Ni-NTA agarose resin and 7 M GnHCl were added simultaneously to photocleaved WT *Ss*βG and the reaction mixture was incubated overnight or (ii) Ni-NTA agarose resin

was added to photocleaved WT *Ss*βG 3 hours prior to 7 M GnHCl and the reaction mixture was incubated overnight (Figure 86). Moreover, when the ratio of Ni-NTA agarose resin to protein quantity increased, better purification of the 39 kDa C-terminal non-His-tagged fragment from the 18 kDa N-terminal His-tagged fragment and His-tagged uncleaved T0-*Ss*βG was accomplished.

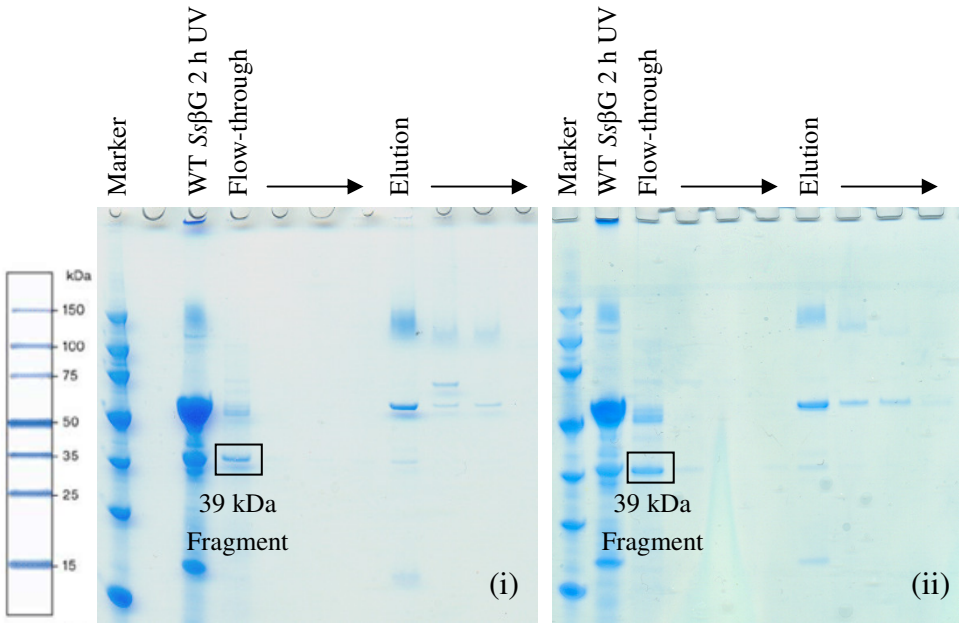



**Figure 86.** Cleaved-tag affinity purification of WT *Ss*βG by SDS-PAGE.

## 4. Green fluorescent protein (GFP) from *Aequorea victoria*

The green fluorescent protein (GFP) from *Aequorea victoria*[363] was selected as the model protein to fuse to the *Ss*βG-($His_6$)-tag in order to develop the cleaved-tag affinity purification method. This protein was chosen for a number of reasons. Firstly, because of its fluorescent properties. Secondly, GFP fusion proteins have been constructed previously.[364-366] Thirdly, as the protein has been reported to be highly stable; Ward and colleagues determined that GFP can withstand exposure to high temperatures, organic solvents, denaturants such as 8 M urea, preservatives and several proteases.[367-371]

To confirm the stability of GFP to UV light, it was first necessary to generate purified GFP protein. GFP was expressed and was purified. IPTG-induced expression at 30°C for 4 hours (A in Figure 87) proved superior to 16 hours at 23°C (B in Figure 87). The protein identity was confirmed by ESI-MS (found 26763 Da, calcd. 26765 Da) (see pg. 279).

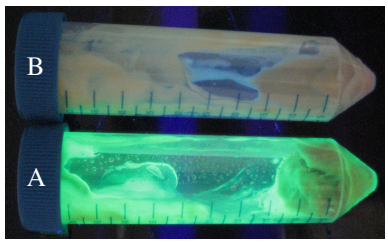



**Figure 87.** GFP cell pellets viewed using a UV transilluminator; A = 30°C for 4 h. B = 23°C for 16 h.

The photocleavage activity of GFP was tested, along with the stability of the protein at room temperature. In a similar way to other non-cleaving proteins, *e.g.* the chalcone reductase from *Medicago sativa*, GFP was not found to dissociate at a specific site when irradiated with UV light and instead was indiscriminately degraded (Figure 88).

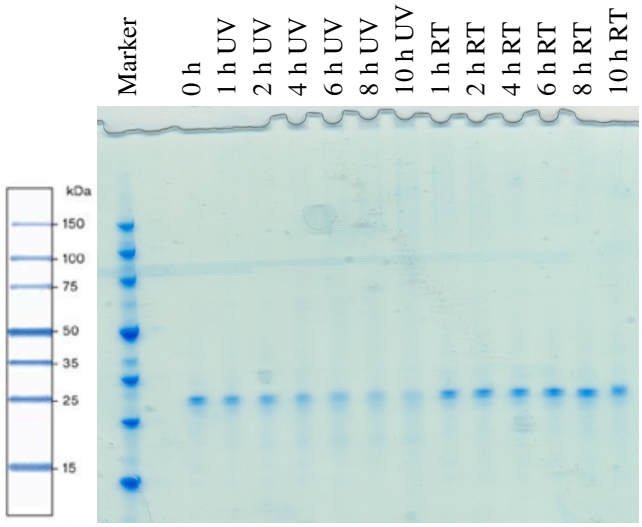



**Figure 88.** Time course analysis of GFP exposed to/not exposed to UV light by SDS-PAGE.

## 5. GFP-*Ss*βG fusion

To demonstrate the cleaved-tag affinity purification method (see pg. 168), a GFP-*Ss*βG fusion, with 10 glycine residue linker, was selected to be the model protein.

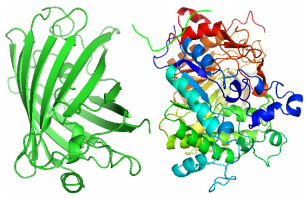


**Figure 89.** Hypothetical crystal structure of GFP-*Ss*βG (PDB IDs: 1EMA and 1GOW, respectively).

### 5.1. Protein preparation and photocleavage

After numerous unsuccessful attempts to generate a plasmid encoding GFP-*Ss*βG, the C-terminal His-tagged protein was expressed from a pET24a plasmid encoding GFP-*Ss*βG that was synthesised by GenScript, and was purified. IPTG-induced expression at 23°C for 16 hours proved superior to 37°C for 6 hours. The protein identity was confirmed by ESI-MS (found 86337 Da, calcd. 86335 Da) (see pg. 281). In addition, the enzymatic activity of the protein was demonstrated using the X-Gal colorimetric assay.[282] GFP-*Ss*βG was tested for photocleavage activity. As expected, the protein was found to dissociate when irradiated with UV light to form two daughter fragments that were clear to see by SDS-PAGE and ESI-MS (Figure 90).

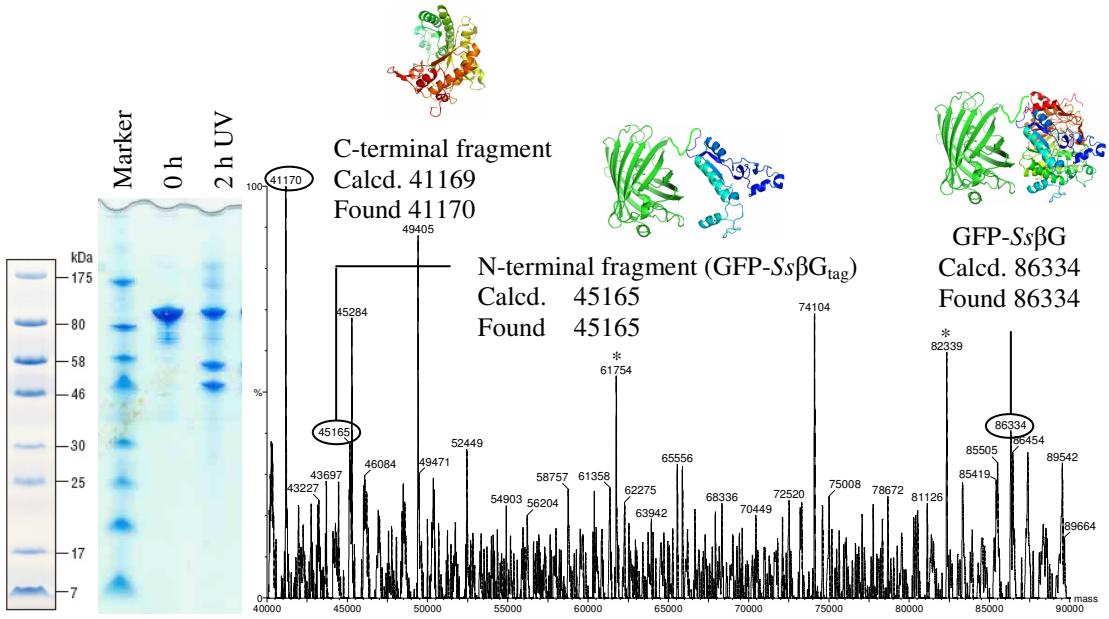



**Figure 90.** Time course analysis of GFP-*Ss*βG exposed to UV light (2 h) by SDS-PAGE (left) and ESI-MS [Raw (insert) and deconvoluted ESI-MS] (right). * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

In addition, photocleaved GFP-*Ss*βG was analysed by non-denaturing PAGE (ND-PAGE) (Figure 91), however, a protein corresponding to the fluorescent GFP-containing 45 kDa fragment was not observed. It was presumed this was because it remained associated with the 41 kDa fragment under the non-denaturing conditions. Note that *Ss*βG is a tetramer under the non-detauring conditions.

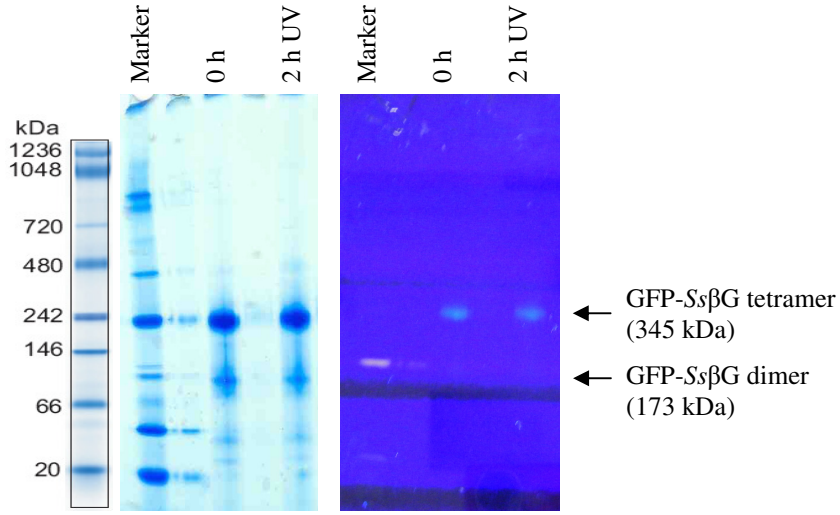



**Figure 91.** GFP-*Ss*βG exposed to UV light by ND-PAGE [visualised by Coomassie staining (right) and UV light (right)].

## 5.2. Degradation

GFP-*Ss*βG which had been stored at 4°C for one month in 50 mM sodium phosphate buffer (pH 6.5) was found by SDS-PAGE to have degraded to two fragments (Figure 92), the masses of which were determined by ESI-MS to be 26694 and 59801 Da. These corresponded roughly with the masses of GFP (26765 Da) and T0-*Ss*βG (58886 Da), and did not imply fragmentation at the His-Trp cleavage site in T0-SsβG that would result in 45 and 41 kDa fragments. The 26 kDa fragment was fluorescent by ND-PAGE, which confirmed that it was GFP-containing (Figure 92). This was further verified by purifying degraded GFP-*Ss*βG using nickel affinity chromatography: the non-His-tagged GFP-containing 26 kDa fragment was eluted in the flow-through, whereas the His-tagged T0-*Ss*βG-containing fragment was recovered only by elution with imidazole.

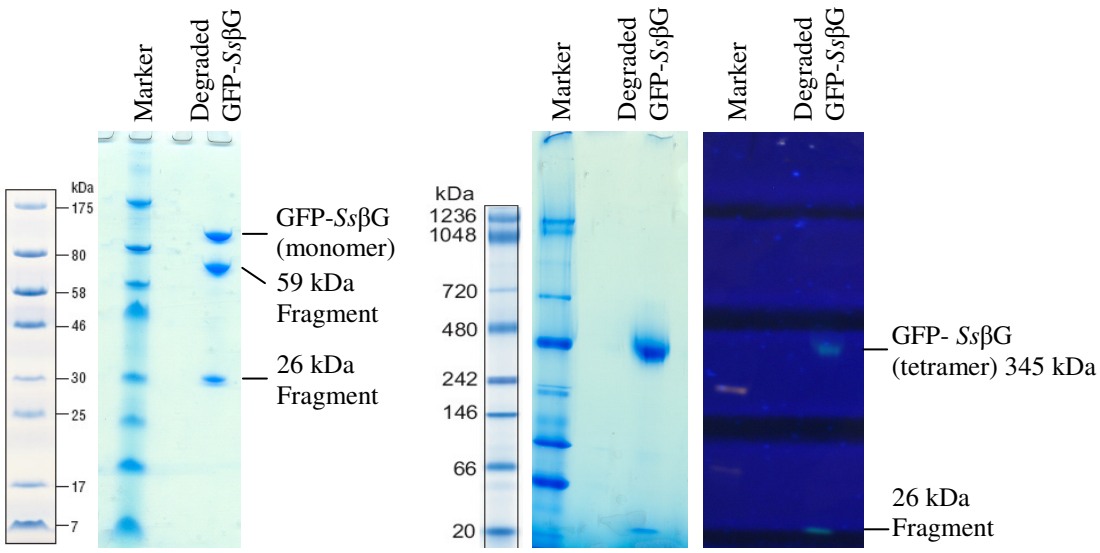



**Figure 92.** Degraded GFP-*Ss*βG by SDS-PAGE (left) and ND-PAGE (right) [visualised by Coomassie staining (right) and UV light (right)].

## 5.3. Cleaved-tag affinity purification

The optimised conditions developed for purification of the photocleaved T0-*Ss*βG reaction mixture were next applied to the purification of the reaction mixture resulting from the photocleavage of GFP-*Ss*βG. The protein was irradiated with UV light to induce protein cleavage and Ni-NTA agarose resin was added. The reaction mixture was then incubated for 4 hours before 7 M GnHCl was added. After overnight incubation, the reaction mixture was purified manually using nickel affinity chromatography. Analysis by SDS-PAGE of the eluent collected from the purification revealed that separation of the target 45 kDa N-terminal non-His-tagged fragment (GFP-*Ss*βG$_{tag}$) from the 41 kDa C-terminal His-tagged fragment and His-tagged uncleaved GFP-*Ss*βG was achieved (Figure 93). As for T0-*Ss*βG, if a higher ratio of Ni-NTA agarose resin to protein quantity was used, less uncleaved GFP-*Ss*βG was recovered in the flow-through.

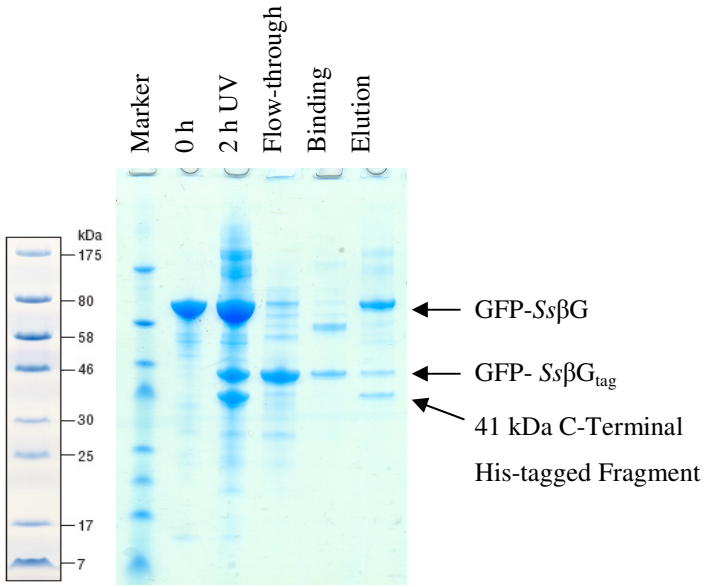


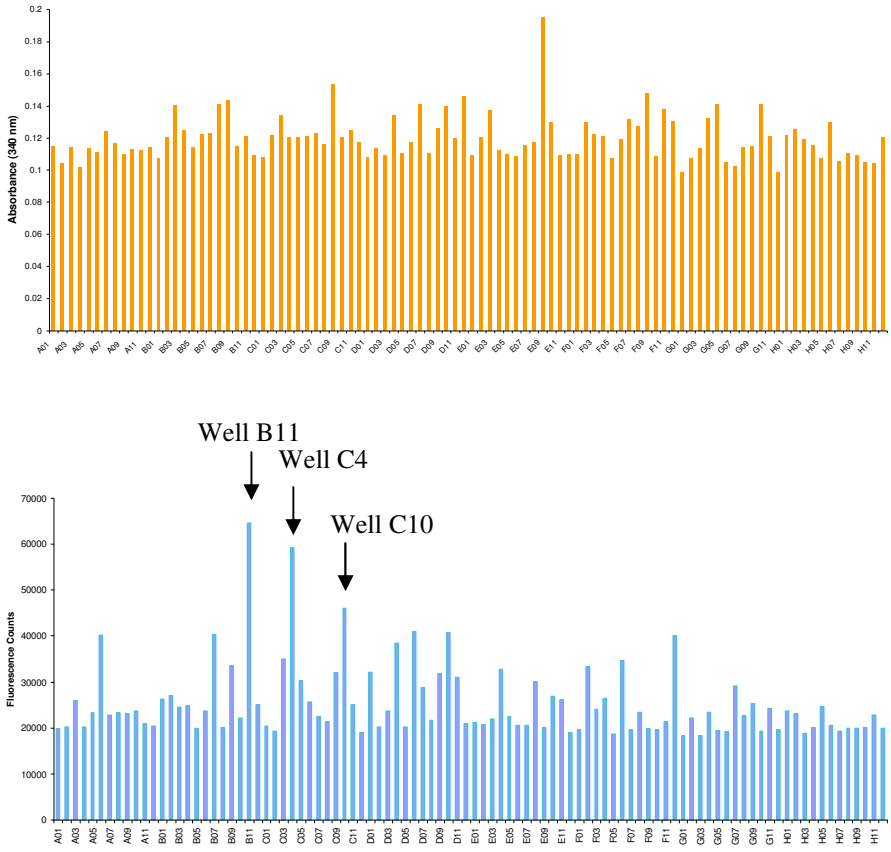

**Figure 93.** Cleaved-tag affinity purification of GFP-*Ss*βG by SDS-PAGE.

Purification of the photocleaved GFP-*Ss*βG reaction mixture was also tried under non-denaturing conditions, but separation of GFP-*Ss*βG$_{tag}$ from the 41 kDa C-terminal His-tagged fragment and His-tagged uncleaved GFP-*Ss*βG was not achieved.

## 5.4. GFP-*Ss*βG$_{tag}$ refolding

Because the cleaved-tag affinity purification method required the use of denaturing conditions, GFP-*Ss*βG$_{tag}$ was recovered unfolded. Therefore, in order to recover protein activity it was necessary to refold GFP-*Ss*βG$_{tag}$ to its native state. A number of well established methods exist to do this including direct dilution and membrane controlled denaturant removal, *e.g.* by dialysis, as well as less common ones such as chromatographic techniques including size exclusion and hydrophobic interaction chromatography, and affinity based processes.[372-374] Successful refolding of GFP-*Ss*βG$_{tag}$ could easily be monitored by fluorescence, as only correctly folded GFP is fluorescent.[375]

Initially, GFP-*Ss*βG$_{tag}$ refolding by direct dilution was tried using an iFOLD Protein Refolding System. Protein refolding was initiated by rapid dilution of denatured GFP-*Ss*βG$_{tag}$ into an iFOLD System 2 matrix, which comprised of a 96 well plate containing 95 unique refolding solutions and a water control. Each refolding reaction was screened for protein activity and protein precipitation by measuring fluorescence and absorbance at 340 nm (smaller $A_{340}$ values indicate less protein precipitation),[376] respectively (Figure 94). It was necessary to use both methods as soluble protein does not always correspond to correctly folded and active protein.




B11: 50 mM CHES, 1 M NDSB 256, 1 mM EDTA, 9 mM GSH, 1 mM GSSG, 240 mM NaCl, 10 mM KCl (pH 9.0)
C4: 50 mM HEPES, 1 M NDSB 256, 1 mM EDTA, 9 mM GSH, 1 mM GSSG, 240 mM NaCl, 10 mM KCl (pH 7.5)
C10: 50 mM TAPS, 1 M NDSB 256, 1 mM EDTA, 9 mM GSH, 1 mM GSSG, 240 mM NaCl, 10 mM KCl (pH 8.5)

**Figure 94.** Analysis of protein refolding by absorbance at 340 nm (top) and fluorescence (excitation 390 nm, emission 510 nm) (bottom).

The three wells shown to possess the most fluorescence (B11, C4 and C10), in which the refolding solution interestingly varied only in buffer and pH, were analysed by ND-PAGE. The gel was visualised by Coomassie staining and UV light but no protein material was observed by either technique, which it was thought may have been due to insufficient protein.

A larger scale refolding of denatured GFP-*Ss*βG$_{tag}$ was performed by dialysis using the optimum conditions identified from the iFOLD refolding screen. As before, the refolding reaction was screened for protein activity and precipitation by measuring fluorescence and absorbance at 340 nm (Figure 95). The increase in fluorescence observed after dialysis suggested successful protein refolding, however, it was also possible that it was due to contaminating uncleaved GFP-*Ss*βG. Refolded GFP-*Ss*βG$_{tag}$ was also analysed by ND-PAGE, but again no protein band corresponding to GFP-*Ss*βG$_{tag}$ was observed when the gel was visualised by Coomassie staining or UV light. Refolded GFP-*Ss*βG$_{tag}$ was also not observed by ND-PAGE when buffer exchanged into storage buffer (50 mM sodium phosphate, pH 6.5) and analysed.

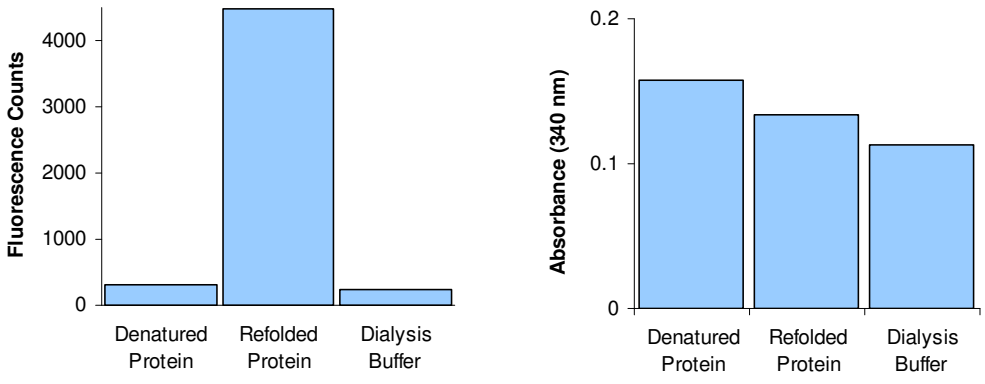



**Figure 95.** Analysis of protein refolding by fluorescence (excitation 390 nm, emission 510 nm) (left) and absorbance (340 nm) (right).

Refolding of GFP-SsβG$_{tag}$ by dialysis or direct dilution into buffer solutions reported to refold GFP,[368, 377] or by hydrophobic interaction chromatography using a HiTrap Phenyl HP 1 mL column did not improve on the result shown above. It is worth noting that GFP refolding is reported in the literature to be challenging.[368, 378-380]

### 5.5. Guanidine hydrochloride concentration

The cleaved-tag affinity purification of GFP-*Ss*βG using different guanidine hydrochloride (GnHCl) concentrations was investigated.

GFP-*Ss*βG was irradiated with UV light to invoke protein cleavage and was then bound to Ni-NTA agarose resin. The supernatant containing any unbound protein was removed and to the resin was added 0, 0.5, 1, 3, 5 or 7 M GnHCl. The reaction mixtures were incubated overnight and then the beads (45-165 μm) were imaged by Dr. Mark Wallace (University of Oxford) under white light before being excited with an Argon ion laser (488 nm) to view GFP fluorescence (Figure 96).

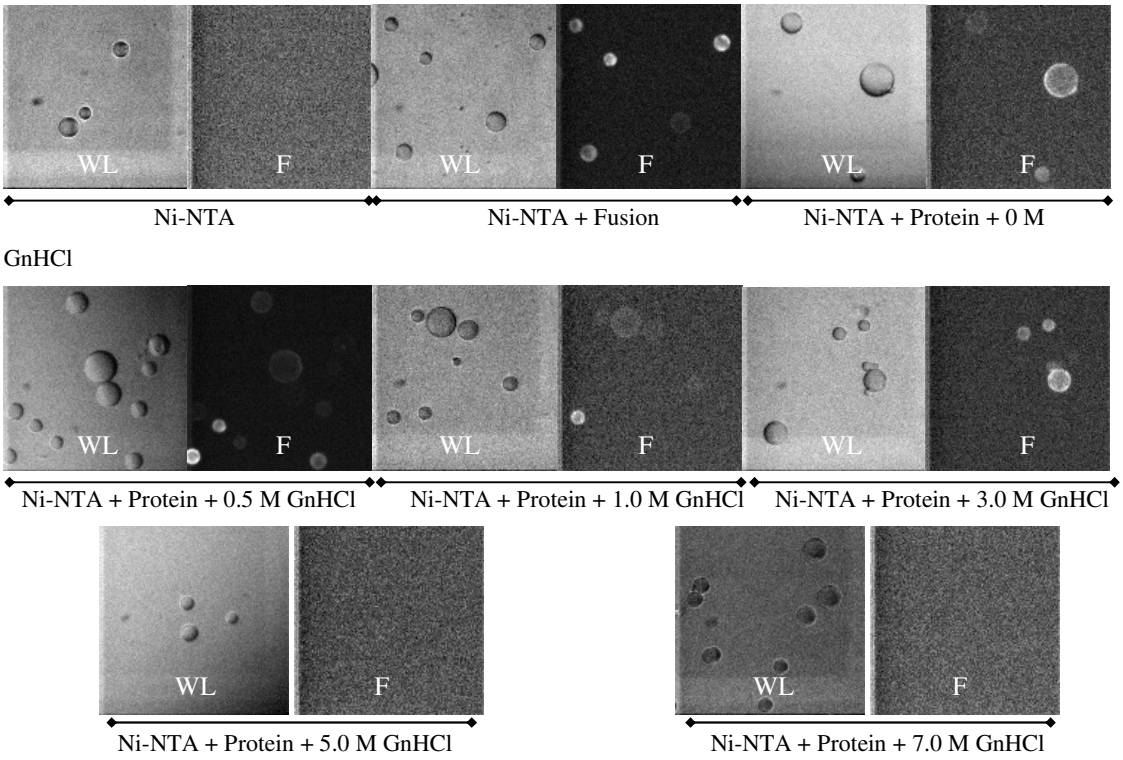



**Figure 96.** Analysis by fluorescence microscopy of irradiated GFP-*Ss*βG bound to Ni-NTA agarose resin denatured with GnHCl (0 -7 M). WL = white light. F = fluorescence.

The reaction mixtures were additionally viewed using UV light (Figure 97). Fluorescence was only observed when 0-3 M GnHCl was used, indicating the need for GnHCl concentrations ≥ 5 M in order to denature GFP.

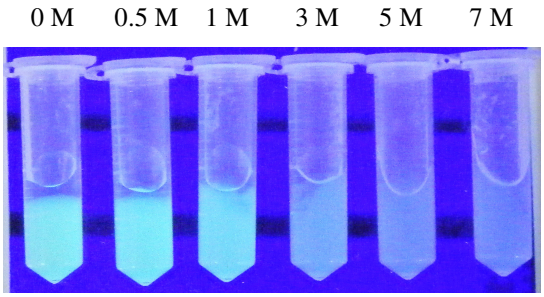



**Figure 97.** Irradiated GFP-*Ss*βG bound to Ni-NTA agarose resin denatured with GnHCl (0 - 7 M), viewed using UV light.

SDS-PAGE analysis of the eluent collected from purification of the reaction mixtures (Figure 98) revealed that GFP-*Ss*βG$_{tag}$ was purified from the 41kDa C-terminal His-

tagged fragment and His-tagged uncleaved GFP-*Ss*βG only when 7 M GnHCl was used. This clearly demonstrated the requirement of 7 M GnHCl for the cleaved-tag affinity method.

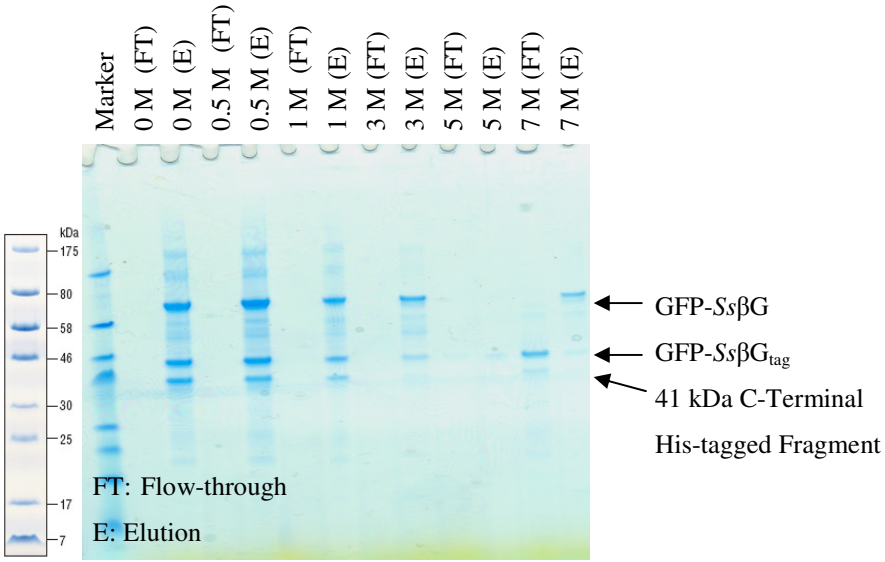



**Figure 98.** Cleaved-tag affinity purification of GFP-*Ss*βG using 0 - 7 M GnHCl as the denaturant, by SDS-PAGE.

## 6. Conclusion and future work

In conclusion, the biotechnological utility of the photocleavage reaction of *Ss*βG described in chapter 3 has been applied here to the development of a cleaved-tag affinity purification method using a GFP-*Ss*βG fusion as the model protein. As shown in Scheme 38, this method employs the use of Ni-NTA resin to "capture" photocleaved GFP-*Ss*βG and then the protein denaturant guanidine hydrochloride to "release" GFP-*Ss*βG$_{tag}$. Because GFP-*Ss*βG$_{tag}$ is recovered unfolded, it is subsequently necessary to refold GFP-*Ss*βG$_{tag}$ to its native state in order to recover GFP activity.

The requirement for denaturing conditions is currently one limitation of this method. However, this limitation could be overcome by engineering *Ss*βG so that the daughter fragments produced from its photocleavage disassemble readily without the need for

denaturants. Another drawback of the cleaved-tag affinity purification method is the size (17.7 kDa) of the tag (*Ss*βG$_{tag}$) that remains fused to GFP following purification. Although attempts to reduce the size of this tag have to date been unsuccessful, further attempts should be made in the future.

## Chapter 5: Summary

In summary, this D. Phil. project has made a contribution to human knowledge by adding a new method for the synthesis of *S*-linked glycoproteins to the ever-growing repetoire of protein glycoconjugation methods, by resulting in the discovery of a photo-induced site-selective protein cleavage reaction and by developing of a cleaved-tag affinity purification method.

The effective novel convergent protein glycoconjugation method for the synthesis of *S*-linked glycoproteins that has been developed here is based on hydrothiolation chemistry, and allows rapid and complete reaction under mild conditions (ambient temperature, pH 4–6) that have been shown to be compatible with a representative selection of different proteins, namely the TIM-barrel protein β-glycosidase from *Sulfolobus solfataricus* (*Ss*βG), the cuboid right-handed quadrilateral β-helix protein Np276 from *Nostoc punctiforme* and the self-assembled homomultimer virus-like bacteriophage particle Qβ. *S*-linked glycoproteins are attractive synthetic targets due to their enhanced chemical stability and enzymatic resistance compared with *N*- and *O*-linked glycoproteins. Combined use of this reaction with an olefinic "tag" that can be positioned using the non-natural amino acid Hag allowed site-selective glycoconjugation at pre-determined sites using chosen sugars and proteins through a "tag–modify" approach (Hag "tag"-hydrothiolation "modify"). Importantly, this reaction is compatible with all other natural amino acids, which proved to be inert under these conditions; this chemoselectivity should allow the use of this reaction in orthogonal, multi-modification approaches. In addition, the glycosyl thiols needed as glycoconjugation partners for this reaction may be readily accessed in a number of ways, even directly from unprotected reducing sugars. Although only mono- and di-

saccharides were successfully coupled to proteins here, this reaction has the potential to also couple larger glycans to proteins. Here, the reaction has been applied to proteins with different structures, which retain full function, for precise, site-selective glycoconjugation at single and even high levels of multiple sites in the case of the self-assembled homomultimer virus-like bacteriophage particle Qβ. New protein glycoconjugation methods such as this one, are particulary important if the function of glycoproteins are to be determined and if glycoprotein mimetics are to be created for use as, *e.g.* therapeutic agents. To this end, the potential of this method to create glycoconjugate immunogens as putative vaccine is currently being explored.

Whilst investigating the above-mentioned photochemical modification of proteins, a novel and efficient photo-induced site-selective protein cleavage reaction present in a natural motif of some TIM-barrel proteins from family 1 of glycosylhydrolases was unexpectedly discovered, and although proteins are known to be affected by UV irradiation, examples of clean, site-selective photocleavage without the use of a cleaving agent are rare. Remarkably, the β-glycosidase from *Sulfolobus solfataricus* (*Ss*βG) was found to form two daughter fragments following UV irradiation (240–308 nm). As protein fragmentation was detected using both SDS-PAGE and ESI-MS, it is believed not to be a result of either technique. However, a crystal structure of cleaved *Ss*βG should be obtained in the future to confirm that this is in fact the case. Alterations at and alanine-scanning of a sphere of residues around the cleavage site were used to establish which residues were essential for the reaction to occur and extensive analysis of the fragments allowed a reaction mechanism to be proposed. The proposed reaction mechanism comprises a suprafacial 1,3-β-H sigmatropic shift, α-elimination of the 18 kDa fragment and rearomatisation. However, other mechanisms such as photoactivated electron transfer–elimination were not excluded and further

work is needed in order to verify the proposed reaction mechanism. The propensity of other proteins that contained the same or similar motif to undergo photo-induced cleavage was also investigated and identified more cleaving proteins (*e.g. Ta*βG, *Pf*βG, *Re*βG and *Ct*βG), in addition to several non-cleaving proteins (*e.g.* arylsulfatase from *Pseudomonas aeruginosa,* chalcone reductase from *Medicago sativa, Sp*βG and *Sa*M). The biological function, if any, of this photochemical post-translational processing if not clear. However, early phylogenic analyses revealed a cluster of photocleavable proteins that are products of isolated genes or genes in operons associated with peptide transport.[291] Interestingly, the cleaving proteins are typically found in environments naturally devoid of UV light, which might imply a yet to be discovered biological function or the maintenance of this photosensitive motif in the absence of UV stability as a selection pressure. Perhaps, as is the case for the inducible-chromophore proteins Kaede and EosFP, a clear biological role does exist. In the future, investigations to determine the biological function, if any, of this protein photocleavage reaction should be prioritised. It should also be noted that photo-induced cleavage of WT *Ss*βG was not observed *in vivo* in E. coli, possibly suggesting a requirement for the protein to be *ex vivo* if cleavage is to occur.

In addition, the biotechnological utility of the aforementioned photocleavage reaction of *Ss*βG was applied to the development of a cleaved-tag affinity purification method using a GFP-*Ss*βG fusion as the model protein during this project. This method employed the use of Ni-NTA resin to "capture" photocleaved GFP-*Ss*βG and then the protein denaturant guanidine hydrochloride to "release" GFP-*Ss*$\beta G_{tag}$. Because GFP-*Ss*$\beta G_{tag}$ is recovered unfolded, it was subsequently necessary to refold GFP-*Ss*$\beta G_{tag}$ to its native state in order to recover GFP activity. The requirement for denaturing conditions is currently one limitation of this method. However, this limitation could

be overcome by engineering *Ss*βG so that the daughter fragments produced from its photocleavage disassemble readily without the need for denaturants. Given the stability of the *Ss*βG complex (*i.e.* that the 39 kDa and 18 kDa fragments do not dissociate) following photo-cleavage and the location of the cleavage site (*i.e.* His151 and Trp152 are not part of the active site), it is highly likely that the *in vivo* function of *Ss*βG does not alter as a result of the UV irradiation [assuming that *Ss*βG cleavage does in fact occur *in vivo*!]. Another drawback of the cleaved-tag affinity purification method is the size (17.7 kDa) of the tag (*Ss*βG$_{tag}$) that remains fused to GFP following purification. Although attempts to reduce the size of this tag during this project were unsuccessful, further attempts should be made in the future.

# Chapter 6: Experimental

## 1. General procedures – chemical synthesis

Proton ($\delta_H$) and carbon ($\delta_C$) nuclear magnetic resonances were recorded on a Bruker DPX 200 spectrometer (200 MHz proton), Bruker DPX 400 spectrometer (400 MHz proton), Bruker AV400 spectrometer (400 MHz proton, 101 MHz carbon) or Bruker AVII500 spectrometer equipped with $^{13}C$ cryoprobe (500 MHz proton, 125 MHz carbon). Spectra were assigned using COSY, HMQC and DEPT 135. All chemical shifts are quoted in parts per million (ppm), referenced to residual solvent frequencies ($^1H$ NMR: $CDCl_3$ = 7.26, $D_2O$ = 4.79, methanol-$d_4$ = 4.87, DMSO-$d_6$ = 2.50, acetone-$d_6$ = 2.05 and $^{13}C$ NMR: $CDCl_3$ = 77.0; methanol-$d_4$ = 49.0; DMSO-$d_6$ = 39.5). The following splitting abbreviations were used: s = singlet, d = doublet, t = triplet, q = quartet, a = apparent.

Low resolution electrospray ionisation (ESI) mass spectra (LRMS) were recorded on a Micromass Platform spectrometer, medium-high resolution ESI mass spectra (MHMS) were recorded on a Waters Micromass LCT premier TOF spectrometer and high resolution ESI mass spectra (HRMS) were recorded on a Bruker MicroTOF spectrometer. Methanol was used as the carrier solvent. Nominal and exact *m/z* values are reported in Daltons.

Optical rotations were measured on a Perkin Elmer 241 polarimeter at 589 nm (Na D-line) with a path length of 1 dm and are reported with implied units of $10^{-1}$ deg $cm^2$ $g^{-1}$. Concentrations (c) are given in g/100 mL.

Melting points (m.p.) were determined using a Leica Galen III hot stage microscope equipped with a Testo 720 thermocouple probe and are uncorrected.

Infrared (IR) spectra were recorded on a Bruker Tensor 27 Fourier Transform spectrometer (liquids and oils were examined as thin films on NaCl plates and solids as KBr disks) or a Bio-Rad FTS-6000 Fourier Transform spectrometer fitted with a DuraSamplIR II diamond ATR head (oils and solids were examined neat). Absorption maxima ($\nu_{max}$) are recorded in wavenumbers ($cm^{-1}$).

Thin layer chromatography (t.l.c.) was performed using Merck aluminum backed plates coated with silica gel 60 $F_{254}$. When flash chromatography was performed using Biotage KP-Sil SNAP cartridges, method development by t.l.c. used glass backed plates coated with the same. Plates were visualised using a UV lamp ($\lambda_{max}$ = 254 nm) and/or stained using one of: ammonium molybdate (5% in 2 M $H_2SO_4$), ninhydrin (0.2% in ethanol), potassium permanganate (5% in 1 M NaOH) or ammonium molybdate (20 g) and cerium(IV) sulphate (0.4 g) in 10% sulphuric acid (400 mL).

Flash column chromatography was conducted using BDH silica gel 40-63 μm (VWR) or was performed on a Biotage SP4 Flash Purification System using Biotage KP-Sil SNAP cartridges.

Unless stated, reagents and starting materials were obtained from standard commercial sources and used without further purification. Alcohol free DCM and THF were dried over activated alumina. Other dry solvents were purchased in sure-seal bottles over molecular sieves. All other solvents were used as supplied. "Petrol" refers to the fraction of petroleum ether boiling in the range of 40-60°C. All reactions using anhydrous conditions were performed using flame dried apparatus under an atmosphere of argon. Distilled $H_2O$ was used for chemical reactions.

DOWEX 50WX8-200 was conditioned as follows: the commercial resin (100 g) was placed in a sintered filter funnel and was washed successively with methanol (500

mL), $H_2O$ (500 mL), 1 M HCl (500 mL) and then $H_2O$ until the pH of the final filtrate was ~ 7.

## 2. General procedures – molecular biology

All molecular biology techniques are based upon procedures described in Molecular Cloning: A Laboratory Manual.[282]

Standard sterile techniques were used for experimental work when appropriate. All growth media and applicable equipment was autoclaved at 121°C for 20 minutes. For procedures requiring a sterile atmosphere a Heraeus HERAsafe sterile cabinet was used. $H_2O$ was purified using a Milli-Q purification system.

### 2.1. Growth media, antibiotics and IPTG

Kanamycin (50 mg/mL), carbenicillin (50 mg/mL), ampicillin (100 mg/mL) and IPTG (0.1 M) stock solutions were sterile filtered (0.2 μm) and stored at 4°C.

**Luria broth (LB) -** Miller's LB broth base (25 g) was dissolved in $H_2O$ (1 L) and was autoclaved.

**Luria broth agar (LB agar) -** Miller's LB broth base (10 g tryptone, 10 g, NaCl, 5 g yeast extract) (25 g) and agar (15 g) were dissolved in $H_2O$ (1 L). The media was autoclaved.

**LB agar plates -** LB agar (1 L) was melted and antibiotic (1 mL) was added. The solution was poured into petri dishes and allowed to set. The plates were stored at 4°C.

**SelenoMet** - SelenoMet Medium Base (Molecular Dimensions) (21.6 g) was dissolved in $H_2O$ (1 L) and was autoclaved. Prior to use, sterile filtered (0.2 μm)

SelenoMet Nutrient Mix (Molecular Dimensions) (5.1 g) dissolved in $H_2O$ (50 mL) was added.

**NZY+ -** NZ amine (casein hydrolysate) (10 g), yeast extract (5 g) and NaCl (5 g) were dissolved in $H_2O$ (1 L). The solution was adjusted to pH 7.5 using NaOH and was autoclaved. Prior to use, the following sterile filtered (0.2 μm) supplements were added: 1 M $MgCl_2$ (12.5 mL), 1 M $MgSO_4$ (12.5 mL) and 20% w/v glucose (20 mL) or 2 M glucose (10 mL).

**Terrific broth (TB)** - Tryptone (12 g), yeast extract (24 g) and glycerol (4 mL) were dissolved in $H_2O$ (900 mL). The solution was shaken until all the solutes had dissolved and was autoclaved. A potassium phosphate solution comprising 0.17 M $KH_2PO_4$ and 0.72 M $K_2HPO_4$ was prepared; $KH_2PO_4$ (2.31 g) and $K_2HPO_4$ (12.5 g) were dissolved in $H_2O$ (90 mL), the volume was adjusted to 100 mL with $H_2O$ and the solution was autoclaved. The potassium phosphate solution (100 mL) was added to the tryptone containing solution.

**SB -** Solution A (Tryptone (12 g), yeast extract (24 g) and glycerol (5 mL)) (900 mL) and solution B ($KH_2PO_4$ (1.7 g) and $K_2HPO_4$ (15.3 g)) (100 mL) were autoclaved separately and then mixed together.

**Autoinduction media** – $Na_2HPO_4$ (6 g), $KH_2PO_4$ (3 g), tryptone (20 g), yeast extract (5 g) and NaCl (5 g) were dissolved in $H_2O$ (1 L). The solution was adjusted to pH 7.2 using NaOH and was autoclaved. Immediately prior to use, sterile filtered (0.2 μm) and autoclaved 60% glycerol (10 mL), 10% glucose (5 mL) and 8% lactose (25 mL) were added to each flask.

### 2.2. Glycerol stock preparation

Bacterial culture (900 μL) was preserved by the addition of 80% autoclaved glycerol (100 μL). After gentle mixing, the stocks were stored at -80°C.

### 2.3. Purification of plasmid DNA

Plasmid DNA was purified using a QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's protocol, and was eluted using $H_2O$ (30 μL) prewarmed to 40°C.

### 2.4. DNA sequencing

DNA sequencing was carried out by the Department of Biochemistry, University of Oxford (before January 2007) or Geneservice Ltd. (after January 2007) on an ABI Prism 3730xl DNA analyser. T7T and T7P primers, and in some instances the internal *Ss*βG primers Int F (5' CGTAGGCATATGTATAACATC 3') and Int R (5' GGAATGAGCTATTAGC 3' were used.

### 2.5. DNA electrophoresis

DNA was analysed using agarose gel electrophoresis. 0.8% w/v agarose gels were prepared as follows: 50x TAE buffer [Tris-HCl (242 g), glacial acetic acid (57 mL), 0.5 M EDTA (100 mL), $H_2O$ (1 L), pH 8] (16 mL) was diluted by the addition of $H_2O$ (784 mL). Agarose (1.2 g) was dissolved in TAE buffer (150 mL) by microwave heating. After the addition of ethidium bromide (16 μL) the gel was poured, allowed to set and then submerged in TAE buffer (650 mL). The DNA sample (20 μL) was mixed with loading dye (5x GelPilot Loading Dye, Qiagen) and along with a marker was applied to the gel, which was then run at 100 V for 75 minutes before being visualised using a Mini Benchtop UV transilluminator.

## 2.6. DNA gel purification

DNA was purified from a 0.8% agarose gel using a QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's protocol, and was eluted using $H_2O$ (30 μL) prewarmed to 50°C.

## 2.7. Protein purification

**AKTA FPLC** - Protein purifications using an AKTA FPLC were performed using a system comprising an AKTA FPLC, a UPC-900 UV detector (eluent monitored at λ = 280 nm), a P-920 pump and Frac-950 fraction collector.

**Nickel affinity chromatography** - Protein purification by nickel affinity chromatography was carried out using an AKTA FPLC system and HisTrap FF 1 or 5 mL column (GE Healthcare). The column was charged with 0.1 M $NiSO_4$ before use. The protein was injected and eluted at 1.0 or 1.5 mL/min, respectively, using the solvent gradient system shown in Figure 99. All buffers were filtered before use. In some instances, Ni-NTA Agarose resin (Qiagen) was used. This was loaded into an empty stoppered (cotton wool) syringe. The resin was washed with $H_2O$ (5 CV), charged with 0.1 M $NiSO_4$ (5 CV) and washed with binding buffer (5 CV). The protein was loaded and eluted using a consecutive sequence of binding buffer (10 CV), wash buffer (10 CV) and elution buffer (10 CV). Alternatively, the protein sample was incubated with the resin for a period of time prior to elution.

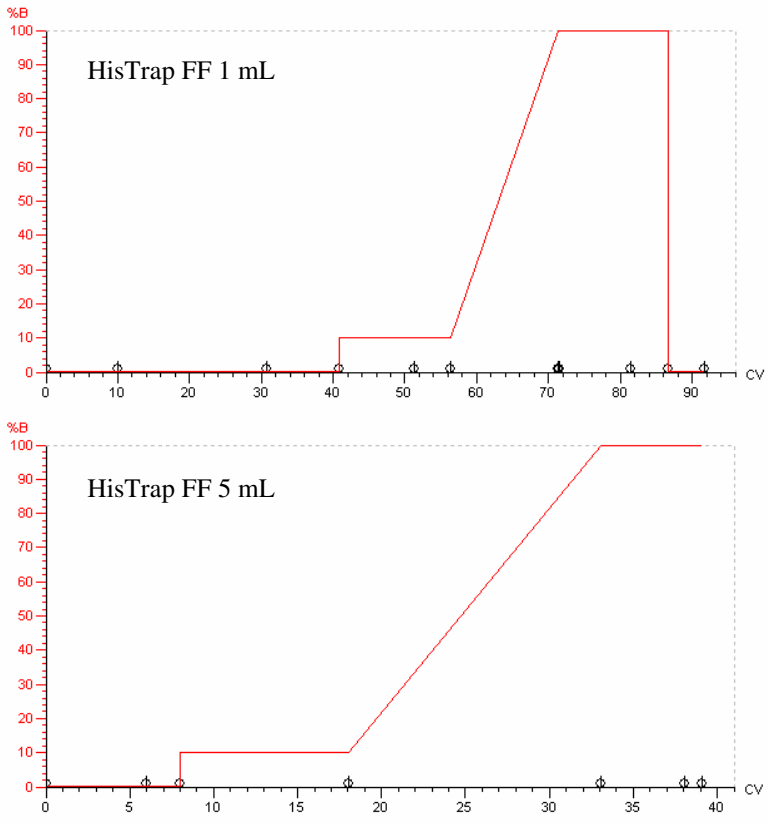



**Figure 99.** Gradient of elution buffer.

**Size exclusion chromatography** - Protein purification by size exclusion chromatography was carried out using an AKTA FPLC system and HiLoad 16/60 Superdex 75 or HiLoad 26/60 Superdex 200 column (Amersham Biosciences). The column was washed with elution buffer before use. The protein was injected and eluted isocratically for 1.5 CV at 1.0 or 2.5 mL/min, respectively. All buffers were filtered before use.

**Anion exchange chromatography -** Protein purification by anion exchange chromatography was carried out using an AKTA FPLC system and FastQ (Q-Sepharose) 50 mL column (Amersham Biosciences). The column was washed with 1 M NaOH (2 CV) and buffer B [50 mM TEA, 2 M NaCl, pH 7.5] (2 CV), and was equilibrated with buffer A [50 mM TEA, pH 7.5] (2 CV). The protein was injected

and eluted at 2.0 mL/min using the solvent gradient system shown in Figure 100. All buffers were filtered before use.

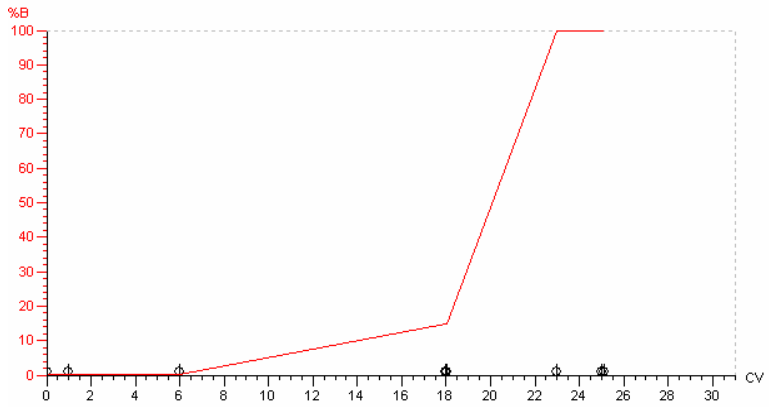



**Figure 100.** Gradient of elution buffer.

**Hydrophobic interaction chromatography -** Protein purification by hydrophobic interaction chromatography was carried out using an AKTA FPLC system and HiTrap Phenyl HP 1 mL column (GE Healthcare). The column was washed with $H_2O$ (2 CV). Ammonium sulphate was added to the protein to a final concentration of 2 M, and was injected and eluted at 0.75 mL/min using the solvent gradient system shown in (Figure 101). All buffers were filtered before use.

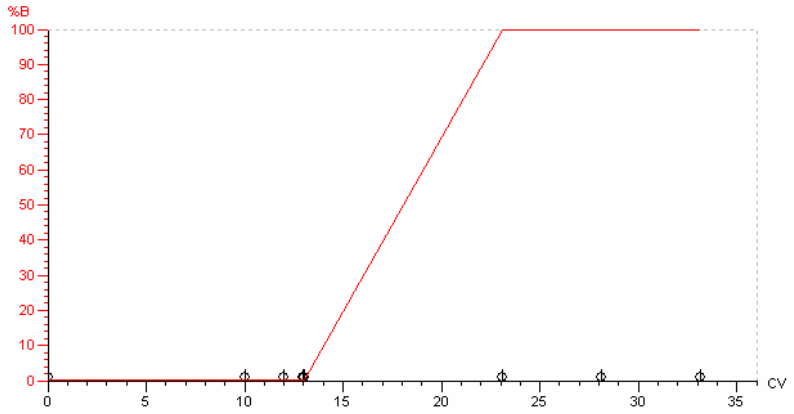



**Figure 101.** Gradient of elution buffer.

Alternatively an AKTA FPLC system and a Phenyl Sepharose FF 75 mL column was used. The column was washed with $H_2O$ (2 CV). The protein was diluted (4x) with buffer A and was injected and eluted at 1.0 mL/min using a gradient from 75% buffer A to 100% buffer B (Figure 102). All buffers were filtered before use.

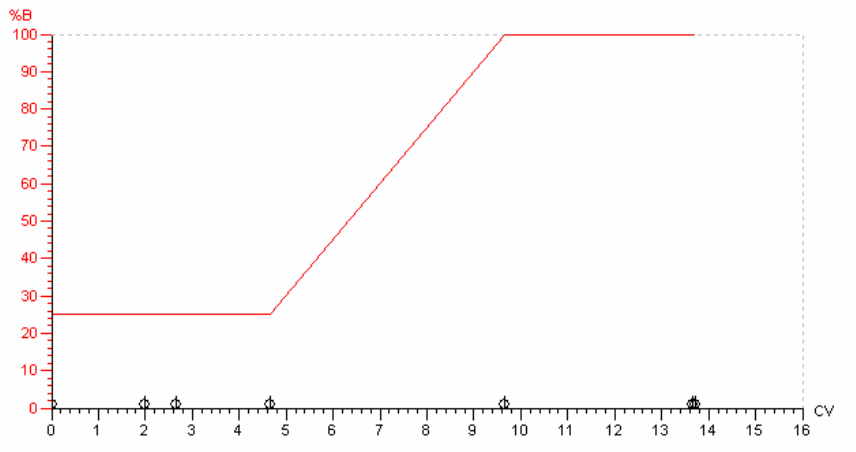



**Figure 102.** Gradient of elution buffer.

**Ammonium sulphate precipitation -** The quantity of solid ammonium sulphate required to make the desired percentage concentration was calculated using the ammonium sulphate calculator at http://www.encorbio.com/protocols/AM-SO4.htm and was added to the protein sample. The solution was stirred at 4°C overnight and was centrifuged for 30 minutes (30 minutes, 10000 rpm, 4°C) to pellet out the protein. The protein was resuspended in the desired buffer.

## 2.8. Pierce SDS-PAGE sample prep kit

The PAGE-prep protein binding resin was vortexed to evenly disperse the resin to a slurry. The slurry (20 µL) was transferred into the center of a spin cup and sample (300 µL) was added. The Pierce SDS-PAGE Sample Prep Kit was then used

following the manufacturer's protocol. The purified samples were prepared for SDS-PAGE analysis as normal.

### 2.9. SDS-PAGE electrophoresis

To protein sample (20 μL) was added sample loading buffer [Tris-HCl (1.5 g), glycerol (20 mL), SDS (2 g), bromophenol blue (2 mg), $H_2O$ (adjust to 100 mL), pH 6.8 (4x)]. The samples were vortexed, centrifuged (1 minute, 13000 rpm, 4$^oC$), heated (100$^oC$, 10 minutes) and centrifuged (1 minute, 13000 rpm, 4$^oC$). Protein marker (Perfect Protein Markers 15-150 or 10-225 kDa (Novagen), Dalton Mark VII-L (Sigma-Aldrich) or Prestained Protein Marker Broad Range (New England Biolabs) (for western blot)) and the samples were loaded on to a gel. Bio-Rad Tris-HCl ready gels were eluted for ~ 45 minutes (120 V, 60 mA) with Bio-Rad Tris/Glycine/SDS running buffer (600 mL). Invitrogen NuPAGE Novex Bis-Tris gels were eluted for ~ 45 minutes (200 V) with NuPAGE MES running buffer (800 mL).

### 2.10. ND-PAGE electrophoresis

ND-PAGE electrophoresis was carried out using the Invitrogen NativePAGE Novex Bis-Tris Gel System; To protein (20 μL) was added NativePAGE sample buffer (4x) (6.7 μL). The samples were vortexed and centrifuged (1 minute, 13000 rpm, 4$^oC$). Protein marker (NativeMark Unstained Protein Standard) (10 μL) and the samples were loaded (10 μL) on to an Invitrogen NativePAGE Novex 3–12% Bis-Tris gel which was eluted for 1 hour at 150 V followed by 30-90 minutes at 250 V with anode buffer [NativePAGE running buffer] (600 mL) in the lower chamber and light blue cathode buffer [20x NativePAGE running buffer (10 mL), 20x cathode buffer additive (1 mL), $H_2O$ (189 mL)] (200 mL) in the upper chamber, at 4$^oC$.

## 2.11. Gel staining and visualisation

Gels were visualised by Coomassie staining or UV light.

**Coomassie staining -** The gel was incubated in gel stain [Coomassie Brilliant Blue R-250 (2.5 g), methanol (500 mL), glacial acetic acid (400 mL), $H_2O$ (100 mL)] with gentle shaking. After 30 minutes, gel stain was replaced with gel destain [methanol (400 mL), glacial acetic acid (70 mL), $H_2O$ (530 mL)] and the gel was incubated overnight with gentle shaking. Alternatively SimplyBlue SafeStain (Invitrogen) or InstantBlue (Expedeon) were used; The gel was washed 3 times for 5 minutes with $H_2O$ (100 ml) (SimplyBlue SafeStain only). The gel was stained with stain (~ 20 ml) for 1 hour with gentle shaking.

**UV Light -** Fluorescence was visualised using a Mini Benchtop UV transilluminator or Bio-Rad Gel Doc XR system.

## 2.12. Dialysis

Dialysis tubing was prepared by boiling for 20 minutes in $H_2O$ containing 0.5 M EDTA (24 mL) and $NaHCO_3$ (5 g). Tubing was washed with $H_2O$ and stored in 20% ethanol. Protein samples were dialysed into buffer (4 L) overnight, the buffer was changed (4 L) and the protein sample dialysed for a further 4 hours.

## 2.13. Protein desalting and buffer exchange

Protein desalting and buffer exchange were carried using a Vivaspin 6, 15 or 500 concentrator (typically 10 kDa MWCO) (Viva Science) or PD-10 desalting column (GE Healthcare). The PD-10 column was washed with buffer (25 mL), the protein sample (2.5 mL) was loaded and eluted with buffer (3.5 mL).

## 2.14. Protein concentration

Protein concentration was carried out using a Vivaspin 6, 15 or 500 concentrator (typically 10 kDa MWCO) (Viva Science) or Amicon stirred ultrafiltration cell.

## 2.15. Determination of protein concentration

**Bradford assay**[381] - Bradford reagent (1 mL) was diluted in $H_2O$ (4 mL), filtered through a Whatman No. 1 filter paper and stored in the dark (for up to 2 weeks at room temperature). Eight bovine serum albumin (BSA) standards (0 – 1.0 mg/mL) were prepared by diluting a stock BSA solution (1.0 mg/mL). The protein was diluted X fold. To the wells of a microplate were added the BSA standards (10 μL) and protein (3 x 10 μL). Bradford reagent (200 μL) was added to each well and the plate was shaken. After 5 minutes incubation at room temperature, absorbances were read at 595 nm. Protein concentration was determined by comparison to a calibration curve produced from the BSA standards.

**Absorbance assay (280 nm) -** The UV lamp of a spectrophotomer (SpectraMax Plus, Molecular Devices) was warmed up (~ 15 minutes) and the wavelength adjusted to 280 nm. The absorbance reading at 280 nm was calibrated to zero using a buffer only solution (1 mL). The absorbance of the protein solution (10 μL protein + 990 μL buffer) was measured. Protein concentration was calculated using http://www.mrc-lmb.cam.ac.uk/ms/methods/proteincalculator.html.

**BCA (bicinchoninic acid) assay**[299] **-** BCA Protein Assay Reagent A (Pierce) (10 mL) was added to BCA Protein Assay Reagent B (200 μL) (50:1). Eight bovine serum albumin (BSA) standards (0 – 1.0 mg/mL) were prepared by diluting a stock BSA solution (1.0 mg/mL). The protein was diluted X fold. To the wells of a microplate

were added the BSA standards (10 μL) and protein (3 x 10 μL). BCA Protein Assay Reagent A + B (200 μL) was added to each well and the plate shaken. After 30 minutes incubation at 37°C, absorbances were read at 562 nm. Protein concentration was determined by comparison to a calibration curve produced from the BSA standards.

## 2.16. Buffers

**Acetate buffer -** Glacial acetic acid (12.7 g) and sodium acetate trihydrate (5.13 g) were made up to 1 L with $H_2O$ to give a 0.25 M solution at pH 4.

**Sodium phosphate** - According to the Henderson-Hasselbach equation, monobasic dihydrogen phosphate ($NaH_2PO_4$) and dibasic dihydrogen phosphate ($Na_2HPO_4$) were mixed in the following ratios, were made up to 1 L with $H_2O$ to give 50 mM solutions, and were pH adjusted if necessary (Table 14).

**Table 14.** Preparation of 50 mM sodium phosphate buffer at 25°C.

| pH | Volume of 0.5 M $NaH_2PO_4$ (mL) | Volume of 0.5 M 0.5 M $Na_2HPO_4$ (mL) |
|---|---|---|
| **6.5** | 70 | 35 |
| **7.5** | 19 | 80 |
| **7.8** | 10.4 | 89.6 |
| **8.5** | (0.15 g) | (13.1 g) |

## 2.17. Determination of pH

The pH values quoted throughout refer to that of aqueous buffer before the addition of any organic solvent. The pH values quoted refer to that of the solutions recorded at room temperature, except where otherwise stated.

### 2.18. β-Galactosidase activity screening

X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) (40 mg) dissolved in DMSO (1 mL) was added to PBS buffer (40 mL) containing $K_2FeCN_6$ (5 mM) and $K_3FeCN_6$ (5 mM). β-Galactosidase activity was measured qualitatively. X-Gal solution (100 μL) was added to protein sample (50 μL) and the mixture was incubated at 37°C. A solution colour change from yellow to blue was considered a positive result.[282]

### 2.19. Western blot

10x TBS buffer was prepared: Trizma (24.2 g) and NaCl (80.0 g) were made up to 1 L with $H_2O$, and was adjusted to pH 7.6. An SDS-PAGE gel was blotted using the semi-dry method (XCell II Blot Module, Invitrogen); The blotting pads and filter papers were immersed in transfer buffer [39 mM glycine, 48 mM Tris-HCl, 0.037% SDS, 20% methanol, $H_2O$ (adjust to 1 L), pH 8.3] (100 mL). Hybond-P membrane (Amersham) was washed with 100% methanol (10 seconds), $H_2O$ (1 minute) and transfer buffer (1 minute). The module was assembled following the manufacturer's protocol and the transfer conducted at 25 V for 2 hours. The membrane was incubated at room temperature for at least 1 hour in blocking buffer [BSA fraction V (2.5 g), TBST buffer (50 mL), sterile filtered (0.2 μm)] (50 mL), washed 3 times for 5 minutes with TBST buffer [10x TBS buffer (100 mL), Tween 20 (1 mL), $H_2O$ (adjust to 1L), store at 4°C for up to 1 week] (50 mL) and incubated overnight at 4°C with primary antibody solution [blocking buffer (50 mL), monoclonal anti-polyhistidine alkaline phosphatase conjugate clone His-1 (1:2500 dilution, 20 μL) (Sigma-Aldrich)] (50 mL). The membrane was washed 3 times for 5 minutes with TBST buffer (50 mL)

followed by 5 minutes with $H_2O$ (50 mL) and was developed using BCIP/NBT reagent (Sigma-Aldrich) (10 mL).

### 2.20. In-gel trypsin digestion

**Excision and initial wash -** The gel was washed with $H_2O$ extensively. A clean razor blade was used to excise the desired band from the gel. The gel slice was placed into a microcentrifuge tube and was cut into ~ 1 $mm^3$ pieces. The gel slices were washed with the following with shaking on a whirlimixer – a) $H_2O$ (100 μL, 15 minutes), b) $H_2O$/acetonitrile (1:1) (100 μL, 30 minutes), c) 0.1 M ambic (ammonium bicarbonate)/acetonitrile (1:1) (100 μL, 30 minutes), d) $H_2O$ (100 μL, 15 minutes), e) 0.1 M ambic/acetonitrile (1:1) (100 μL, 30 minutes), f) $H_2O$ (100 μL, 15 minutes), g) 0.1 M ambic/acetonitrile (1:1) (100 μL, 30 minutes), h) $H_2O$ (100 μL, 5 minutes). The liquid was removed and the gel slices were dehydrated using 100% acetonitrile (50 μL) with shaking. When they became white and stuck together (~ 5 minutes) the acetonitrile was removed and the gel slices were rehydrated using 0.1 M ambic (50 μL). After 5 minutes with shaking, acetonitrile (50 μL) was added. The gel slices were left for 15 minutes without shaking before being dried in a Speed Vac for 1 hour at room temperature.

**Reduction and alkylation -** This step was omitted if no cysteine was present in the protein. 10 mM DTT (in 0.1 M ambic) (50 μL) was added and the gel slices were incubated for 45 mins at 56$^{o}$C. The liquid was removed. 55 mM iodoacetamide (in 0.1 M ambic) (50 μL) was added and the gel slices were incubated for 30 minutes at room temperature in the dark. The liquid was removed. 0.1 M ambic (50 μL) was added and the gel slices were incubated for 5 mins. The liquid was removed. Acetonitrile (50 μL) was added to the gel slices. When they became white and stuck together (~ 5 minutes),

the acetonitrile was removed and the gel slices were rehydrated using 0.1 M ambic (50 μL). After 5 minutes, acetonitrile (50 μL) was added. The gel slices were left for 15 minutes without shaking before being dried in a Speed Vac for 30 minutes at room temperature.

**Digestion -** Trypsin (Promega) (20 μg) was suspended in 50 mM ambic (1 mL). The gel slices were rehydrated in trypsin solution (50 μL) and were incubated at 37$^{o}$C overnight.

**Peptide extraction** - The liquid was transferred to a microcentrifuge tube and 25 mM ambic (100 μL) was then added to the gel slices. After 5 minutes of shaking, acetonitrile (100 μL) was added and the gel slices were shaken for 1 hour. The liquid was transferred to the microcentrifuge tube and 0.1% formic acid (100 μL) was added. After 5 minutes of shaking, acetonitrile (100 μL) was added and the gel slices were shaken for 1 hour. The liquid was transferred to the microcentrifuge tube and the combined liquids were freeze dried and resuspended in 0.1% formic acid (10 μL).

### 2.21. Fluorescence

Fluorescence measurements were recorded using a Varian Cary Eclipse fluorescence spectrophotometer or POLARstar OPTIMA.

### 2.22. HPLC

HPLC was performed using a system comprising a Dionex UltiMate 3000 instrument coupled to an UltiMate 3000 variable wavelength detector. Chromeleon software (version 6.80.) was used.

## 3. General procedures - mass spectrometry

### 3.1. Intact proteins

Intact proteins were analysed by liquid chromatography electrospray ionisation mass spectrometry (LC-ESI-MS) using either a Micromass LCT or Micromass LCT Premier XE mass spectrometer.

**Micromass LCT mass spectrometer -** Proteins were analysed by LC-ESI-MS in positive ion mode on a Micromass LCT mass spectrometer interfaced with a Waters 2790 Alliance HT HPLC using a Phenomenex Jupiter 5u C4 300A 250 x 4.6 mm column. The protein (typically 1 mg/mL, 20 μL) was injected and eluted at 1.0 mL/min using a gradient from 95% Solvent A (99.9% $H_2O$, 0.1% formic acid) to 95% Solvent B (99.9% acetonitrile, 0.1% formic acid) (Table 15).

**Table 15.** Solvent gradient system.

| Time (Mins) | Solvent A (%) | Solvent B (%) |
|---|---|---|
| 0 | 95 | 5 |
| 5 | 95 | 5 |
| 20 | 5 | 95 |
| 25 | 5 | 95 |
| 30 | 95 | 5 |
| 35 | 95 | 5 |

All solvents were degassed by sonication for 15 minutes prior to use. The eluent was split mass spectrometer/waste (2:8). The following MS parameters were used: capillary voltage, 3000 V; sample cone voltage, 25 V or 35 V; desolvation temperature, 250 $^{o}$C; extraction cone, 4 V; RF lens, 250; source temperature 80 $^{o}$C;

desolvation flow ($N_2$), 540 L/h; no cone flow; *m/z* scan range 300 to 1800; scan time, 1 s. The data were processed using the Maximum Entropy method (MaxEnt1). Myoglobin (equine heart) (1 mg/mL) was used as the calibrant; Calcd. 16951. Protein Sequence: [M]GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEK FDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKGHHEAELKPLAQSH ATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKY KELGFQG.

**Micromass LCT premier XE (UPLC) mass spectrometer** - Proteins were analysed by LC-ESI-MS in positive ion mode on a Micromass LCT Premier XE mass spectrometer interfaced with a Waters Acquity Ultra Performance LC (UPLC) using a Waters Acquity UPLC BEH 300 C18 1.7µm 2.1 x 100 mm column at 50$^{o}$C. The protein (1 mg/mL, 1 µL) was injected and eluted at 0.3 mL/min using a gradient from 95% Solvent A (99.9% $H_2O$, 0.1% formic acid) to 95% Solvent B (99.9% acetonitrile, 0.1% formic acid) (Table 16).

**Table 16.** Solvent gradient system.

| Time (Mins) | Solvent A (%) | Solvent B (%) |
|---|---|---|
| 0 | 95 | 5 |
| 8 | 5 | 95 |
| 8.01 | 95 | 5 |
| 10 | 95 | 5 |

The eluent was injected directly into the mass spectrometer. The following MS parameters were used: capillary voltage, 3000 V; sample cone voltage, 35 V; desolvation temperature, 150 $^{o}$C; source temperature 80 $^{o}$C; desolvation flow ($N_2$), 200 L/h; cone flow 50 L/h; *m/z* scan range 300 to 3500; scan time, 1 s. The data were

processed using the Maximum Entropy method (MaxEnt1). Sodium formate was used as the calibrant.

### 3.2. Digest peptides

Digest samples were analysed using either a ThermoFisher LTQ FT ULTRA hybrid ion trap-fourier transform ion cyclotron resonance, Micromass Q-Tof *micro* or Waters Synapt High Definition mass spectrometer.

**ThermoFisher LTQ FT ULTRA hybrid ion trap-fourier transform ion cyclotron resonance mass spectrometer** - Digest samples were analysed using electrospray ionisation (ESI) on a ThermoFisher LTQ FT ULTRA hybrid ion trap-fourier transform ion cyclotron resonance mass spectrometer by in-line LC/MS using a Dionex 3000 nano-LC system equipped with a Dionex PepMap100 C18 column (15 cm, 3 μm, 100 Å). The peptides were injected and eluted at 350 nL/min using a gradient from 98% Solvent A (99.9% $H_2O$, 0.1% formic acid) to 90% Solvent B (99.9% acetonitrile, 0.1% formic acid) (Table 17).

**Table 17.** Solvent gradient system.

| Time (Mins) | Solvent A (%) | Solvent B (%) |
|---|---|---|
| 0 | 98 | 2 |
| 30 | 30 | 70 |
| 30.1 | 10 | 90 |
| 35 | 10 | 90 |
| 35.1 | 100 | 0 |
| 40.0 | 100 | 0 |

**Micromass Q-Tof micro mass spectrometer** - Digest samples were analysed by electrospray ionisation time-of-flight MS (ESI-Tof-MS) in both MS and MS/MS mode. Experiments were performed in positive ion mode on a Micromass Q-Tof *micro* mass spectrometer interfaced with an Agilent 1100 series HPLC using a Phenomenex Jupiter 5u C18 300Å 150 x 0.5 mm column. The peptides were injected and eluted at 15 µL/min using a gradient from 95% solvent A (99.9% $H_2O$, 0.1% formic acid) to 95% solvent B (99.9% acetonitrile, 0.1% formic acid) (Table 18).

**Table 18.** Solvent gradient system.

| Time (Mins) | Solvent A (%) | Solvent B (%) |
|---|---|---|
| 0 | 95 | 5 |
| 5 | 95 | 5 |
| 30 | 60 | 40 |
| 60 | 5 | 95 |
| 70 | 95 | 5 |
| 90 | 95 | 5 |

The eluent was injected directly into the mass spectrometer. The following MS parameters were used: capillary voltage, 3000 V; cone voltage, 35 V; source temperature, 80 °C; collision energy, 35 V; desolvation temperature, 200 °C; *m/z* scan range 200 to 2800.

**Waters synapt high definition mass spectrometer -** Digest samples were analysed by LC-ESI-MS in both MS and MS/MS mode. Experiments were performed in positive ion mode on a Waters Synapt High Definition mass spectrometer interfaced with a Waters Acquity Ultra Performance LC (UPLC) using a Waters Acquity UPLC BEH 300 C18 1.7µm 2.1 x 100 mm column at 50$^{o}$C. The peptides were injected and

eluted at 0.2 mL/min using a gradient from 95% solvent A (99.9% $H_2O$, 0.1% formic acid) to 95% solvent B (99.9% acetonitrile, 0.1% formic acid) (Table 19).

**Table 19.** Solvent gradient system.

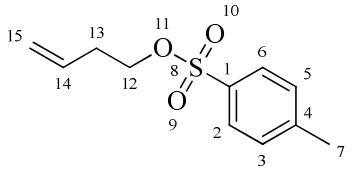
| Time (Mins) | Solvent A (%) | Solvent B (%) |
|---|---|---|
| 0 | 95 | 5 |
| 60.00 | 40 | 60 |
| 60.01 | 95 | 5 |
| 80.00 | 95 | 5 |

The eluent was injected directly into the mass spectrometer. The following MS parameters were used: capillary voltage, 3000 V; cone voltage, 35 V; source temperature, 120 °C; collision energy ramp, 35 to 110 V; desolvation temperature, 250 °C; *m/z* scan range 200 to 2800.

## 4. Chemical synthesis

### 4.1. Homoallylglycine (Hag)

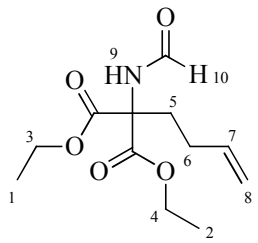
**Toluene-4-sulfonic acid but-3-enyl ester (1)**




A solution of 3-buten-1-ol (5.0 g, 69 mmol, 1 eq) in pyridine (10 mL) was cooled in an ice bath. Tosyl chloride (13.2 g, 69 mmol, 1 eq) was added. After 3 hours, t.l.c. (ethyl acetate/petrol (3:7)) indicated the formation of product ($R_f$ 0.4) with complete consumption of the starting material ($R_f$ 0.2). The reaction mixture was poured into an ice/conc. HCl 4/1 v/v solution (50 mL), extracted with diethyl ether (100 mL), dried

over $MgSO_4$, filtered and evaporated *in vacuo* to yield toluene-4-sulfonic acid but-3-enyl ester (**1**) (15.1 g, 96%) as a yellow oil.[120] IR (neat) $\nu_{max}$ 3082 (C-H), 2988 (C-H), 1599 (C=C), 1356 ($SO_2$), 1175 ($SO_2$) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) 2.36-2.43 (m, 2H, *J* 6.5 Hz, $CH_2$, H-13), 2.45 (s, 3H, $CH_3Ar$, H-7), 4.06 (app t, 2H, *J* 6.5 Hz, $CH_2OSO_2$, H-12), 5.04-5.10 (m, 2H, $CH_2$, H-15), 5.61-5.72 (m, 1H, $J_{14,15trans}$ 17.0 Hz, $J_{14,15cis}$ 10.4 Hz, $J_{13,14}$ 6.6 Hz, $CH_2\underline{C}H$, H-14), 7.35 (d, 2H, *J* 8.2 Hz, ArH, H-3/H-5), 7.79 (d, 2H, *J* 8.0 Hz, ArH, H-2/H-6); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ ppm 21.6 ($CH_3$, C-7), 33.1 ($CH_2$, C-13), 69.4 ($OCH_2$, C-12), 118.2 ($\underline{C}H_2CH$, C-15), 127.9 (ArH, C-2/C-6), 129.8 (ArH, C-3/C-5), 132.4 ($CH_2\underline{C}H$, C-14), 133.1 (ArS, C-1), 144.8 ($ArCH_3$, C-4); $ESI^+$ LRMS: calcd. for $C_{12}H_{16}O_2S$ ($M+Na^+$) 249.05, found 248.87.

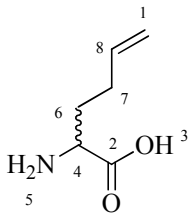
**But-3-enyl-formylamino-malonic acid diethyl ester (2)**




Diethylformylaminomalonate (7.1 g, 35.2 mmol, 0.95 eq) was dissolved in DMF (100 mL) at room temperature under Ar. NaH (60% dispersion in mineral oil, 1.4 g, 35.2 mmol, 0.95 eq) was added under vigorous stirring. After 30 minutes, toluene-4-sulfonic acid but-3-enyl ester (**1**) (8.4 g, 37.1 mmol, 1 eq) was added. The reaction was heated to 70°C. After 18 hours, t.l.c. (ethyl acetate/petrol (3:7)) indicated the formation of product ($R_f$ 0.4) with incomplete consumption of the starting material ($R_f$ 0.4). The reaction mixture was concentrated *in vacuo* and was partitioned between $H_2O$ (100 mL) and ethyl acetate (200 mL). The aqueous phase was re-extracted with ethyl acetate (3 x 200 mL). The combined organics were washed with brine (200 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting yellow oil was

purified by silica gel chromatography (diethyl ether/petrol (3:2)). The off white solid was recrystallised (diethyl ether/petrol) to afford but-3-enyl-formylamino-malonic acid diethyl ester (**2**) (2.74 g, 29%) as a white crystalline solid. m.p 90-92°C [Lit. m.p 94-95.5 °C];[382, 383] IR (neat) $\nu_{max}$ 3293 (N-H), 2984 (C-H), 1742 (C=O), 1674 (C=C), 1651 (C=O) $cm^{-1}$; $^1$H NMR (400 MHz, $CDCl_3$) δ ppm 1.27 (t, 6H, *J* 7.1 Hz, $CH_3$, H-1/H-2), 1.90-1.98 (m, 2H, $CH_2$, H-6), 2.44-2.50 (m, 2H, $CH_2$, H-5), 4.27 (q, 4H, *J* 7.0 Hz, $OCH_2$, H-3/H-4), 4.98 (dd, 1H, *J* 10.1 Hz, *J* 1.3 Hz, C$\underline{H_2}$CH, H-8 cis), 5.03 (dd, 1H, *J* 17.1 Hz, *J* 1.4 Hz, CHC$\underline{H_2}$, H-8 trans), 5.69-5.81 (m, 1H, C$\underline{H}$CH$_2$, H-7), 6.93 (br s, 1H, NH, H-9), 8.18 (s, 1H, CHO, H-10); $^{13}$C NMR (101 MHz, $CDCl_3$) δ ppm 13.9 ($CH_3$, C-1/C-2), 27.8 ($CH_2$, C-6), 31.2 ($CH_2$, C-5), 62.7 ($OCH_2$, C-3/C-4), 65.5 ($CH_2$, C-5), 115.6 (CH$\underline{C}$H$_2$, C-8), 136.6 ($\underline{C}$HCH$_2$, C-7), 159.6 (CHO, C-11), 167.6 (C=O, C-12/C-13); ESI$^+$ LRMS: calcd. for $C_{12}H_{19}NO_5$ (M+$Na^+$) 280.11, found 280.07.

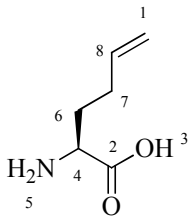
**DL-Homoallylglycine (DL-Hag) (±3)**




Scheme 19, method c: But-3-enyl-formylamino-malonic acid diethyl ester (**2**) (3.37 g, 6.2 mmol, 1 eq) was refluxed in 10 wt% NaOH (20 mL). After 3 hours, t.l.c. (diethyl ether/petrol (3:2)) indicated the formation of product ($R_f$ 0.0) with complete consumption of the starting material ($R_f$ 0.3). The solution was neutralised with 6 M HCl and then evaporated *in vacuo*. To the residue was added methanol (50 mL). The resulting white precipitate (NaCl) was filtered off under vacuum and washed with methanol (3 x 10 mL). The filtrate was evaporated *in vacuo* and then refluxed with 1.5 M HCl (20 mL). After 3 hours, t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) indicated

the formation of product with complete consumption of the dicarboxylate intermediate. The solvent was evaporated *in vacuo* and the residue was taken up in methanol (50 mL). The resulting white precipitate (NaCl) was filtered under vacuum and washed with methanol (3 x 10 mL). The filtrate was evaporated *in vacuo*. The crude was purified by cation exchange chromatography (DOWEX 50WX8-200, ammonium hydroxide eluent). Lyophilisation afforded DL-Hag (**±3**) (788 mg, 100%) as a white solid. $R_f$ 0.5 (propan-2-ol/acetic acid/ $H_2O$ (3:1:1)); IR (KBr) $\nu_{max}$ 3400 (N-H), 2986 (C-H), 2943 (C-H), 1583 (N-H) $cm^{-1}$; $^1H$ NMR (400 MHz, $D_2O$) δ ppm 1.81-1.99 (m, 2H, $CH_2$, H-6), 2.08-2.15 (m, 2H, $CH_2$, H-7), 3.70 (m, 1H, H-4), 5.02 (dd, *J* 10.2 Hz, 1H, C<u>H</u>$_2$CH, trans H-1), 5.08 (dd, *J* 17.3 Hz, *J* 1.6 Hz, 1H, C<u>H</u>$_2$CH, cis H-1), 5.82 (ddt, *J* 17.1 Hz, *J* 10.4 Hz, *J* 6.5 Hz, 1H, $CH_2$C<u>H</u>, H-8); $^{13}C$ NMR (101 MHz, $D_2O$) δ ppm 28.6 ($CH_2$, C-7), 29.6 ($CH_2$, C-6), 54.2 (CH, C-4), 115.9 (<u>C</u>$H_2$CH, C-1), 136.9 (<u>C</u>HC$H_2$, C-8), 174.6 (COOH, C-2); $ESI^+$ LRMS: calcd. for $C_6H_{11}NO_2$ ($M+Na^+$) 152.07, found 152.06.

**L-Homoallylglycine (L-Hag) (+3)**




Scheme 20, method f: To (*S*)-(-)-1-*t*-butyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (**7**) (100 mg, 0.3 mmol, 1 eq) dissolved in DCM (5 mL) was added TFA (5 mL), triethylsilane (125 µL) and $H_2O$ (125 µL). The reaction mixture was stirred at room temperature. After 2 hours, t.l.c. (propan-2-ol/acetic acid/$H_2O$ (3:1:1)) indicated the formation of product ($R_f$ 0.5) with complete consumption of starting material ($R_f$ 1.0). The reaction mixture was concentrated *in vacuo*. The resulting residue was

purified by silica gel chromatography (methanol/ethyl acetate (15:85) to 100%) to afford L-Hag (**+3**) (5 mg, 15%) as an off white solid. $R_f$ 0.5 (propan-2-ol/acetic acid/$H_2O$ (3:1:1)).

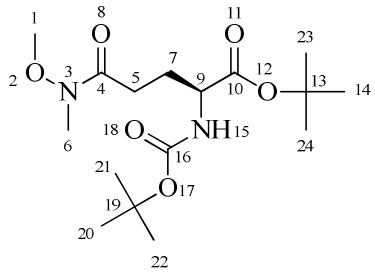
Scheme 20, method g: To (*S*)-(-)-1-*t*-butyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (**7**) (730 mg, 0.19 mmol, 1 eq) dissolved in DCM (5 mL) was added TFA (5 mL). The reaction mixture was stirred at room temperature under an atmosphere of Ar. After 1 hour, t.l.c. (propan-2-ol/acetic acid/$H_2O$ (3:1:1)) indicated the formation of product ($R_f$ 0.5) with complete consumption of starting material ($R_f$ 1.0). The reaction mixture was concentrated *in vacuo* and the resulting residue was dissolved in ethanol (10 mL). To this was added 1 M NaOH (10 mL). The reaction mixture was stirred at 40$^o$C for 1 hour, cooled to room temperature and concentrated *in vacuo*. The crude was purified by cation exchange chromatography (DOWEX 50WX8-200, ammonium hydroxide eluent)[223] to afford L-Hag (**+3**) (240 mg, 77%) as a white solid. m.p. 250-253$^o$C [Lit. m.p 250–255 $^o$C];[212] $[\alpha]_D^{25}$ +8.0 (c=1, $H_2O$) [Lit.$[\alpha]_D^{25}$ +13.1 (c=1, $H_2O$)].[384] IR (neat) $\nu_{max}$ 3474 (N-H), 2986 (C-H), 2945 (C-H), 1580 (N-H) cm$^{-1}$; $^1$H NMR (400 MHz, $D_2O$) $\delta$ ppm 1.79-1.98 (m, 2H, $CH_2$, H-6), 2.06-2.14 (m, 2H, $CH_2$, H-7), 3.61 (app t, *J* 6.6 Hz, 1H, H-4), 5.01 (dd, *J* 10.3 Hz, 1H, C$\underline{H_2}$CH, H-1 trans), 5.07 (dd, *J* 17.2 Hz, 1H, C$\underline{H_2}$CH, H-1 cis), 5.75-5.87 (ddt, *J* 17.1 Hz, *J* 10.4 Hz, *J* 6.5 Hz, 1H, $CH_2C\underline{H}$, H-8); $^{13}$C NMR (101 MHz, $D_2O$) $\delta$ ppm 28.9 ($CH_2$, C-7), 30.1 ($CH_2$, C-6), 54.6 (CH, C-4), 116.2 ($\underline{C}H_2CH$, C-1), 137.3 ($\underline{C}HCH_2$, C-8), 175.2 (COOH). ESI$^+$ LRMS: calcd. for $C_6H_{11}NO_2$ (M+Na$^+$) 152.07, found 152.06.

Scheme 21, method f: To methyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (**11**) (50 mg, 0.2 mmol, 1 eq) dissolved in methanol (5 mL) at room temperature under an Ar atmosphere was added 0.5 M NaOH (10 mL). After 18 hours, the reaction mixture

was acidified using DOWEX 50WX8-200, filtered and evaporated *in vacuo*. Then, DCM (5 mL) and TFA (5 mL) were added. After 30 minutes, t.l.c. (propan-2-ol/acetic acid/$H_2O$ (3:1:1)) indicated the formation of product ($R_f$ 0.5) with complete consumption of the starting material ($R_f$ 1.0). The reaction mixture was evaporated *in vacuo*. The crude was purified by cation exchange chromatography (DOWEX 50WX8-200, 1 M ammonium hydroxide eluent). Freeze drying afforded L-Hag (**+3**) (19 mg, 96%) as a white solid. $[\alpha]_D^{25}$ +8.9 (c=1, $H_2O$) [Lit.$[\alpha]_D^{25}$ +13.1 (c=1, $H_2O$)].[384]

Scheme 21, method g: To methyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (**11**) (250 mg, 0.7 mmol. 1 eq) dissolved in $H_2O$ (1 mL) and dioxane (1 mL) at 0$^oC$ was added LiOH monohydrate (183 mg, 0.7 mmol, 1.05 eq). After 30 minutes, the reaction mixture was warmed to room temperature and dioxane (5 mL) was added. After 18, 36 and 54 hours, LiOH monohydrate (183 mg, 0.7 mmol, 1.05 eq) was added to the reaction mixture. Then, the reaction mixture was concentrated *in vacuo* and re-dissolved in DCM (2 mL). TFA (2 mL) was added to the reaction mixture. After 5 minutes, t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) indicated the formation of product with complete consumption of the intermediate. The reaction mixture was concentrated *in vacuo*. The crude was purified by cation exchange chromatography (DOWEX 50WX8-200, ammonium hydroxide eluent) to afford L-Hag (**+3**) (85 mg, 91%) as a white solid. $[\alpha]_D^{19}$ +5.8 (c=1, $H_2O$) [Lit. $[\alpha]_D^{25}$ +13.1 (c=1, $H_2O$)].[384]

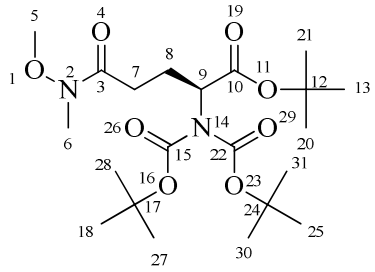
**(*S*)-(-)-2-(*t*-butoxycarbonylamino)-4-(*N*-methoxy-*N*-methylcarbamoyl) butanoic acid *t*-butyl ester (4)**



To a solution of L-Boc-Glu-O*t*Bu (Novabiochem) (5.0 g, 16.4 mmol, 1 eq) in DCM (50 mL) was added triethylamine (2.55 mL, 18.3 mmol, 1.1 eq) and PyBOP (9.45 g, 18.1 mmol, 1.1 eq). The reaction mixture was stirred for 10 minutes at room temperature under argon. Then, *N*,*O*-dimethylhydroxylamine hydrochloride (1.85 g, 19.0 mmol, 1.1 eq) and triethylamine (2.55 mL, 18.3 mmol, 1.1 eq) were added. After 21 hours, t.l.c. (ethyl acetate/pentane (3:7)) indicated the formation of product with complete consumption of the starting material. The reaction mixture was partitioned between 10% aqueous citric acid (75 mL) and DCM (75 mL) and the aqueous phase was re-extracted with DCM (2 x 75 mL). The combined organics were washed with sodium bicarbonate (100 mL), brine (100 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol; gradient (4:6) to (1:1)) to afford (S)-(-)-2-(*t*-butoxycarbonylamino)-4-(*N*-methoxy-*N*-methylcarbamoyl) butanoic acid *t*-butyl ester (**4**) (5.71 g, 95%) as a white solid. $R_f$ 0.4 (ethyl acetate/pentane (1:1)) [Lit. $R_f$ 0.4 (ethyl acetate/pentane (1:1)];[217] $[\alpha]_D^{25}$ -18.1 (c=1, EtOH) [Lit. $[\alpha]_D^{25}$ -18.9 (c=1, EtOH)];[217] IR (neat) $\nu_{max}$ 3350 (N-H), 2980 (C-H), 2940 (C-H), 1715 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) $\delta$ ppm 1.45 (s, 9H, $C(CH_3)_3$), 1.47 (rotamer d, *J* 1.0 Hz, 9H, $C(CH_3)_3$), 1.88-1.98 (m, 1H, $CH_2$, H-7), 2.10-2.21 (m, 1H, $CH_2$, H-7), 2.43 - 2.62 (m, 2H, $CH_2$, H-5), 3.18 (s, 3H, $NCH_3$, H-6), 3.68 (rotamer d, *J* 1.0 Hz, 3H, $OCH_3$, H-1),

4.15 - 4.25 (m, 1H, CH, H-9), 5.19 (d, *J* 8.1 Hz, 1H, NH, H-15); $^{13}$C NMR (101 MHz, $CDCl_3$) δ ppm 27.6 ($CH_2$, C-7), 27.9 ($C(CH_3)_3$), 28.0 ($CH_2$, C-5), 28.3 ($C(CH_3)_3$), 32.2 ($NCH_3$, C-6), 53.7 (NCH, C-9), 61.2 ($OCH_3$, C-1), 79.5/81.9 ($O\underline{C}(CH_3)_3$, C-13/C-19), 155.4 (NHCOO, C-16), 171.5, 173.6 (C=O, C-4/C-10); $ESI^+$ LRMS: calcd. for $C_{16}H_{30}N_2O_6$ ($M+Na^+$) 369.20, found 369.17.

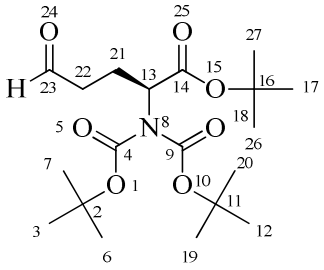
**(*S*)-(-)-2-[bis(*t*-butoxycarbonylamino)]-4-(*N*-methoxy-*N*-methylcarbamoyl) butanoic acid *t*-butyl ester (5)**



To (*S*)-(-)-2-(*t*-butoxycarbonylamino)-4-(*N*-methoxy-*N*-methylcarbamoyl) butanoic acid *t*-butyl ester (**4**) (5.0 g, 14.4 mmol, 1 eq) dissolved in DCM (50 mL) was added $(Boc)_2O$ (15.7 g, 72.2 mmol, 5 eq), triethylamine (43.2 mL, 0.3 mol, 20 eq) and DMAP (176 mg, 1.4 mmol, 0.1 eq). The reaction mixture was stirred at room temperature under argon. After 4 days, t.l.c. showed incomplete consumption of the starting materials. The reaction mixture was heated to 40°C and $(Boc)_2O$ (4.50 g, 20.6 mmol, 1.4 eq), triethylamine (21.6 mL, 0.1 mol, 10 eq) and DMAP (704 mg, 5.6 mmol, 0.4 eq) were added. After a further 1 day, t.l.c. (ethyl acetate/pentane (1:1)) indicated the formation of product ($R_f$ 0.5) with complete consumption of the starting material ($R_f$ 0.4). The reaction mixture was partitioned between 0.35 M $KHSO_4$ (100 mL) and DCM (50 mL) and the aqueous phase was re-extracted with DCM (2 x 100 mL). The combined organics were washed with $H_2O$ (100 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol gradient (1:9) to (1:1)) to afford (S)-(-)-2-[bis(*t*-

butoxycarbonylamino)]-4-(*N*-methoxy-*N*-methylcarbamoyl) butanoic acid *t*-butyl ester (**5**) (6.04 g, 94%) as a yellow oil.[218] $[\alpha]_D^{25}$ -19.7 (c=1, $CHCl_3$); IR (neat) $\nu_{max}$ 2982 (C-H), 2941 (C-H), 1740 (C=O), 1703 (C=O), 1668 (C=O) $cm^{-1}$; $^1H$ NMR (200 MHz, $CDCl_3$) δ ppm 1.45 (s, 9H, $C(CH_3)_3$, H-13/H-20/H-21), 1.50 (s, 18H, $C(CH_3)_3$, H-18/H-25/H-27/H-28/H-30/H-31), 2.06-2.25 (m, 1H, $CH_2$, H-8), 2.33-2.64 (m, 3H, $CH_2$, H-8/H-7), 3.17 (s, 3H, $NCH_3$, H-6), 3.67 (s, 3H, $OCH_3$, H-5), 4.76-4.89 (m, 1H, CH, H-9); $^{13}C$ (101 MHz, $CDCl_3$) δ ppm 24.3 ($CH_2$, C-8), 27.9 ($C(CH_3)_3$, C-13/C-20/C-21), 27.9 ($C(CH_3)_3$, C-18/C-25/C-27/C-28/C-30/C-31), 28.5 ($CH_2$, C-7), 32.1 ($NCH_3$, C-6), 58.3 (NCH, C-9), 61.0 ($OCH_3$, C-5), 81.1 ($O\underline{C}(CH_3)_3$, C-17/C-24 or C-12), 82.7 ($O\underline{C}(CH_3)_3$, C-17/C-24 or C-12), 152.2 (NHCOO, C-15/C-22), 169.3, 171.0 (C=O, C-3/C-10); $ESI^+$ LRMS: calcd. for $C_{21}H_{38}N_2O_8$ ($M+Na^+$) 915.52, found 915.48.

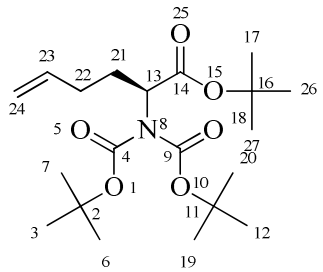
**(*S*)-(-)-2-[bis(*t*-butoxycarbonylamino)]-4-(formyl) butanoic acid *t*-butyl ester (6)**




To (*S*)-(-)-2-[bis(*t*-butoxycarbonylamino)]-4-(*N*-methoxy-*N*-methylcarbamoyl) butanoic acid *t*-butyl ester (**5**) (5.83 g, 13.1 mmol, 1 eq) dissolved in THF (75 mL) at -78°C under argon was added dropwise diisobutylaluminium hydride (D*i*BAL-H 1.0 M solution in heptane, 19.9 mL, 19.9 mmol, 1.5 eq). The reaction mixture was warmed to room temperature. After 1 hour, t.l.c. (ethyl acetate/pentane (1:1)) indicated the formation of product ($R_f$ 0.6) with complete consumption of the starting material ($R_f$ 0.5). The reaction mixture was partitioned between 0.35 M $NaHSO_4$ (200 mL) and diethyl ether (100 mL) and the aqueous phase was re-extracted with diethyl ether (2 x 100 mL). The combined organics were washed with 1 M HCl (2 x 100 mL),

sodium bicarbonate (2 x 100 mL), brine (2 x 100 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo* to afford (S)-(-)-2-[bis(*t*-butoxycarbonylamino)]-4-(formyl) butanoic acid *t*-butyl ester (**6**) (4.90 g, 96%) as a yellow oil. $R_f$ 0.8 (ethyl acetate/hexane (2:8));[220] $[\alpha]_D^{25}$ –24.5 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{20}$ -23.1 (c=0.8, $CHCl_3$)];[220] IR (neat) $\nu_{max}$ 2982 (C-H), 2940 (C-H), 1738 (C=O), 1701 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 1.45 (s, 9H, $C(CH_3)_3$, H-17/H-26/H-27) 1.50 (s, 18H, $C(CH_3)_3$, H-3/H-6/H-7/H-12/H-19/H-20), 2.08-2.20 (m, 1H, $CH_2$, H-21), 2.38-2.64 (m, 3H, $CH_2$, H-21/H-22), 4.72- 4.76 (dd, *J* 9.5 Hz, *J* 5.2 Hz, 1H, CH, H-13), 9.76 (m, 1H, CHO, H-23); $^{13}C$ (101 MHz, $CDCl_3$) δ ppm 21.9 ($CH_2$, C-21), 27.9 ($C(CH_3)_3$, C-17/C-26/C-27), 27.9 ($C(CH_3)_3$, C-3/C-6/C-7/C-12/C-19/C-20), 40.6 ($CH_2$, C-22), 58.1 (CHN, C-13), 81.5, 83.0 ($OC(CH_3)_3$, C-2/C-11/C-16), 152.3 (OCON, C-4/C-19), 169.1 (OC=O, C-14), 201.2 (HCO, C-23); $ESI^+$ LRMS: calcd. for $C_{21}H_{38}N_2O_8$ ($M+NH_4+CH_3CN^+$) 446.29, found 446.25.

**(*S*)-(-)-1-*t*-butyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (7)**



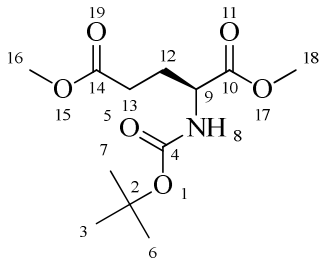

Scheme 20, method d: To a stirring solution of methyltriphenylphosphonium bromide (8.28 g, 23.2 mmol, 2 eq) in THF (50 mL) at room temperature under argon was added dropwise *n*-BuLi (1.6 M in heptane, 14.5 mL, 23.2 mmol, 2 eq). After 30 minutes, the resulting orange ylide solution was cooled to -78°C and (*S*)-(-)-2-[bis(*t*-butoxycarbonyl) amino]-4-(formyl) butanoic acid *t*-butyl ester (**6**) (4.5 g, 11.6 mmol, 1 eq) dissolved in THF (10 mL) was added. The reaction mixture was warmed to 0°C.

After 2 hours, t.l.c. (ethyl acetate/pentane (1:1)) indicated the formation of product ($R_f$ 0.8) with complete consumption of the starting material ($R_f$ 0.6). The reaction mixture was quenched with a saturated aqueous $NH_4Cl$ solution (50 mL) and was partitioned between $H_2O$ (100 mL) and ethyl acetate (100 mL). The aqueous phase was re-extracted with ethyl acetate (3 x 100 mL). The combined organics were dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol gradient (1:19) to (1:9)) to afford (*S*)-(-)-1-*t*-butyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (**7**) (260 mg, 6%) as a yellow oil.

Scheme 20, method e: To a stirring solution of methyltriphenylphosphonium bromide (1.1 g, 3.0 mmol, 1.2 eq) in toluene (25 mL) at 0°C under argon was added dropwise $KN(TMS)_2$ (0.5 M in toluene, 5.67 mL, 2.8 mmol, 1.1 eq). After 15 minutes, the resulting bright yellow ylide solution was cooled to -78°C and (*S*)-(-)-2-[bis(*t*-butoxycarbonyl) amino]-4-(formyl) butanoic acid *t*-butyl ester (**6**) (1.0 g, 2.6 mmol, 1 eq) dissolved in toluene (2 mL) was added dropwise. After 3 hours, t.l.c. (ethyl acetate/pentane (1:1)) indicated the formation of product ($R_f$ 0.8) with complete consumption of the starting material ($R_f$ 0.6) The reaction mixture was quenched with a saturated aqueous $NH_4Cl$ solution (100 mL) and was partitioned between $H_2O$ (100 mL) and diethyl ether (100 mL). The aqueous phase was re-extracted with diethyl ether (2 x 100 mL). The combined organics were washed with brine (100 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (1:9)) to afford (S)-(-)-1-*t*-butyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (**7**) (330 mg, 33%) as a colourless oil. $R_f$ 0.8 (ethyl acetate/pentane (1:1)) $[\alpha]_D^{25}$ -17.1 (c=1, MeOH) [Lit.$[\alpha]_D^{20}$ -16.2 (c=1.48, MeOH)];[220] IR (neat) $\nu_{max}$ 2982 (C-H), 2940 (C-H), 1740 (C=O), 1703 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) $\delta$ ppm 1.44 (s, 9H, $C(CH_3)_3$, H-17/H-26 /H-27), 1.49 (s,

18H, $C(CH_3)_3$, H-3/H-6/H-7/H-12/H-19/H-20), 1.89-1.98 (m, 1H, $CH_2$, H-21), 2.07-2.20 (m, 3H, $CH_2$, H-21/H-22), 4.72 (dd, *J* 9.5 Hz, *J* 4.8 Hz,1H, CH, H-13), 4.96-5.01 (m, 1H, $\underline{H_2}CCH$, H-24 trans), 5.01-5.05 (m, 1H, $\underline{H_2}CCH$, H-24 cis), 5.74-5.85 (m, 1H, $C\underline{H_2}CH$, C-23); $^{13}C$ (101 MHz, $CDCl_3$) δ ppm 27.9 ($C(CH_3)_3$, C-17/C-26/C-27), 28.0 ($C(CH_3)_3$, C-3/C-6/C-7/C-12/C-19/C-20), 28.6 ($CH_2$, C-21), 30.5 ($CH_2$, C-22), 58.2 (NCH, C-13), 81.1, 82.6 ($O\underline{C}(CH_3)_3$, C-2/C-11/C-16), 115.2 ($\underline{C}H_2CH$, C-24), 137.6 ($\underline{C}HCH_2$, C-23), 152.3 (C=O, C-4/C-9), 169.8 (C=O, C-14); $ESI^+$ LRMS: calcd. for $C_{20}H_{35}NO_6$ ($M+Na^+$) 408.24, found 408.38.

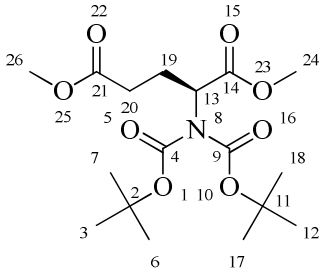
**(*S*)-(+)-1,5-dimethyl-2-(*t*-butyloxycarbonyl)-glutamate (8)**




To L-glutamic acid (1.0 g, 5.4 mmol, 1 eq) in methanol (25 mL) at 0°C under Ar was added chlorotrimethylsilane (3.03 mL, 23.7 mmol, 4.4 eq). The reaction mixture was warmed to room temperature. After 24 hours, t.l.c. (methanol/ethyl acetate (1:1)) indicated the formation of intermediate ($R_f$ 0.6) with complete consumption of L-glutamic acid ($R_f$ 0.0). Triethylamine (4.89 mL, 35.1 mmol, 6.6 eq) and $(Boc)_2O$ (1.3 g, 5.9 mmol, 1.1 eq) were added to the reaction mixture. After the evolution of gas was complete, the reaction mixture was concentrated *in vacuo*. The resulting white precipitate was filtered off and washed with diethyl ether (2 x 50 mL). The filtrate was concentrated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (3:7)) to afford (S)-(+)-1,5-dimethyl-2-(*t*-butyloxycarbonyl)-glutamate (**8**) (1.18 g, 80%) as a colourless oil. $R_f$ 0.7 (ethyl acetate/petrol (1:1)); $[\alpha]_D^{25}$ +14.7 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{25}$ +12.2 (c=1, $CHCl_3$)];[385]

IR (neat) $\nu_{max}$ 3381 (N-H), 2982 (C-H), 2957 (C-H), 1740 (C=O), 1711 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 1.43 (s, 9H, $CH_3$, H-3/H-6/H-7), 1.90-2.0 (m, 1H, $CH_2$, H-12), 2.12-2.24 (m, 1H, $CH_2$, H-12), 2.33-2.48 (m, 2H, $CH_2$, H-13), 3.68 (s, 3H, $OCH_3$, H-16 or H-18), 3.74 (s, 3H, $OCH_3$, H-16 or H-18), 4.29-4.39 (m, 1H, CH, H-9), 5.13 (d, *J* 7.5 Hz, 1H, NH, H-8); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ ppm 27.7 ($CH_2$, C-12), 28.2 ($C(CH_3)_3$, C-3/C-6/C-7), 30.0 ($CH_2$,C-13), 51.7, 52.4 ($OCH_3$, C-16/C-18), 52.8 (NCH, C-9), 80.0 ($O\underline{C}(CH_3)_3$, C-2), 155.3 (NHCOO, C-4), 172.6, 173.1 (C=O, C-10/C-14); $ESI^+$ LRMS: calcd. for $C_{12}H_{29}NO_6$ ($M+Na^+$) 298.13, found 298.10.

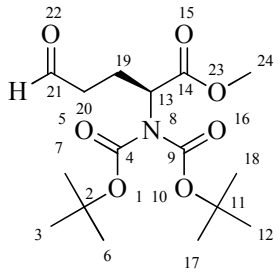
**(*S*)-(-)-1,5-dimethyl-2-[bis(*t*-butoxycarbonyl)amino] pentanedioate (9)**



To (S)-(+)-1,5-dimethyl-2-(*t*-butyloxycarbonyl)-glutamate (**8**) (1.0 g, 3.6 mmol, 1 eq) and DMAP (88 mg, 0.7 mmol, 0.2 eq) dissolved in acetonitrile (25 mL) was added $(Boc)_2O$ (1.2 g, 5.4 mmol, 1.5 eq). The reaction mixture was stirred at room temperature under Ar. After 24 hours, t.l.c. (ethyl acetate/petrol (4:6)) indicated the formation of product ($R_f$ 0.9) with complete consumption of the starting material ($R_f$ 0.8). The reaction mixture was concentrated *in vacuo* and the resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (1:9)) to afford (S)-(-)-1,5-dimethyl-2-[bis(*t*-butoxycarbonyl)amino] pentanedioate (**9**) (1.15 g, 85%) as a white solid. $R_f$ 0.8 (pentane/ethyl acetate (1:1)); m.p 48-59 $^oC$; $[\alpha]_D^{25}$ -41.0 (c=1, $CHCl_3$) [Lit. -37.2 (c=1, $CHCl_3$)];[224] IR (neat) $\nu_{max}$ 2988 (C-H), 2957 (C-H), 1780 (C=O), 1757 (C=O), 1734 (C=O) $cm^{-1}$; $^1H$ (400 MHz, $CDCl_3$) δ ppm 1.50 (s, 18H, $C(CH_3)_3$,

H-3/H-6/H-7/H-12/H-17/H-18), 2.13-2.26 (m, 1H, $CH_2$, H-19), 2.33-2.55 (m, 3H, $CH_2$, H-19/H-20), 3.68 (s, 3H, $OCH_3$, H-24 or H-26), 3.72 (s, 3H, $OCH_3$, H-24 or H-26), 4.89-4.99 (m, 1H, NCH, H-13); $^{13}C$ (101 MHz, $CDCl_3$) δ ppm 25.1 ($CH_2$, C-19), 27.9 ($C(CH_3)_3$, C-3/C-6/C-7/C-12/C-17/C-18), 30.6 ($CH_2$, C-20), 51.6, 52.2 ($OCH_3$, C-24/C-26), 57.3 (NCH, C-13), 83.3 ($O\underline{C}(CH_3)_3$, C-2/C-11), 151.8 (C=O, C-4/C-9), 170.8, 173.0 (C=O, C-14/C-21); $ESI^+$ LRMS: calcd. for $C_{17}H_{29}NO_8$ ($M_2+Na^+$) 773.37, found 773.36.

**(*S*)-(-)-1-methyl-2-[bis-(*t*-butoxycarbonylamino)]-5-oxopentanoate (10)**

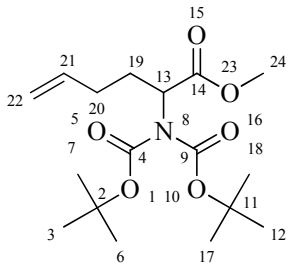


Scheme 21, method c: To (S)-(-)-1,5-dimethyl-2-[bis(*t*-butoxycarbonyl)amino] pentanedioate (**9**) (1.0 g, 2.7 mmol, 1 eq) dissolved in diethyl ether (25 mL) at -78°C under Ar was added dropwise diisobutylaluminium hydride (D*i*BAL-H 1.0 M solution in heptane, 3.0 mL, 3.0 mmol, 1.1 eq). After 2 hours, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.6) with complete consumption of the starting material ($R_f$ 0.7). The reaction mixture was quenched with $H_2O$ (1 mL), warmed to room temperature, dried over $MgSO_4$, filtered through a pad of Celite and concentrated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (15:85)) to afford (*S*)-(-)-1-Methyl-2-[bis-(*t*-butoxycarbonylamino)]-5-oxopentanoate (**10**) (474 mg, 52%) as a colourless oil.

Scheme 21, method d: To (S)-(-)-1,5-dimethyl-2-[bis(*t*-butoxycarbonyl)amino] pentanedioate (**9**) (5.0 g, 13.3 mmol, 1 eq) dissolved in diethyl ether (150 mL) at -

78°C under Ar was added by cannula diisobutylaluminium hydride (D*i*BAL-H 1.0 M solution in hexane, 16.0 mL, 16.0 mmol, 1.2 eq) pre-cooled to -78°C. After 6 hours, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.6) with complete consumption of the starting material ($R_f$ 0.7). The reaction mixture was stirred with Rochelle salt (sat., 500 mL) for 1 hour, then diluted with DCM (200 mL). The aqueous phase was re-extracted with DCM (2 x 100 mL). The combined organics were washed with brine (250 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (1:9)) to afford (S)-(-)-1-methyl-2-[bis-(*t*-butoxycarbonylamino)]-5-oxopentanoate (**10**) (4.04 g, 88%) as a pale yellow oil. $[\alpha]_D^{25}$ -39.5 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{25}$ -35.3 (c=1, $CHCl_3$)];[224] IR (neat) $\nu_{max}$ 2984 (C-H), 2959 (C-H), 1748 (C=O), 1703 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 1.50 (s, 18H, $C(CH_3)_3$, H-3/H-6/H-7/H-12/H-17/H-18), 2.11-2.23 (m, 1H, $CH_2$, H-19), 2.38-2.68 (m, 3H, $CH_2$, H-19/H-20), 3.73 (s, 3H, $OCH_3$, H-24), 4.90 (dd, *J* 9.5 Hz, *J* 5.0 Hz, 1H, NCH, H-13), 9.78 (s, 1H, CHO, H-21); $^{13}C$ (101 MHz, $CDCl_3$) δ ppm 22.5 ($CH_2$, C-19), 27.9 ($C(CH_3)_3$, C-3/C-6/C-7/C-12/C-17/C-18), 40.4 ($CH_2$, C-20), 52.2 ($OCH_3$, C-24), 57.2 (NCH, C-13), 83.4 ($OC(CH_3)_3$, C-2/C-11), 151.9 (C=O, C-4/C-9), 170.7 (C=O, C-14), 200.9 (CHO, C-21); $ESI^+$ LRMS: calcd. for $C_{16}H_{27}NO_7$ ($M_2$+$Na^+$) 713.35, found 712.96.

**1-methyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (11)**



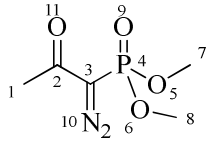
Scheme 22: To a solution of methyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexynoate (**13**) (100 mg, 0.3 mmol, 1 eq) in toluene was added Lindlar's catalyst (30 mg, 0.03 mmol, 0.1 eq). The reaction mixture was stirred in an atmosphere of $H_2$. After 2 hours, t.l.c. (ethyl acetate/petrol (3:7)) indicated the formation of product with complete consumption of the starting material. The reaction mixture was filtered through a pad of Celite, which was washed with ethyl acetate (3 x 25 mL). The combined organics were evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (2.5:97.5)) to afford (*RS*)-1-methyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (**±11**) (93 mg, 93%) as a colourless oil. $[\alpha]_D^{25}$ 0.0 (c=1, $CHCl_3$) [Lit. (**-11**) $[\alpha]_D^{25}$ -35.3 (c=1, $CHCl_3$)][224].

Scheme 21, method e, entries A-E in Table 2: To methyltriphenylphosphonium bromide (0.95 eq) dissolved in solvent (*A/B* = THF, *C* = toluene, *D* = diethylether, *E* - acetonitrile) at 0°C under an Ar atmosphere was added base (*A* - *n*-BuLi, *B-E* - $KN(TMS)_2$ (0.5 M in toluene), 0.9 eq) dropwise. After 15 minutes, the reaction mixture was cooled (*A-D* to -78 °C; *E* to 0°C). In a separate flask, (*S*)-(-)-1-Methyl-2-[bis-(*t*-butoxycarbonylamino)]-5-oxopentanoate (**10**) (1 eq) dissolved in solvent was cooled (*A-D* to -78 °C; *E* to 0°C). Then, the aldehyde was added dropwise to the ylide solution via cannula. After 18 hours, t.l.c. (ethyl acetate/petrol (2:8)) indicated the formation of product ($R_f$ 0.6) with complete consumption of the starting material ($R_f$ 0.2). The yellow reaction mixture was partitioned between a saturated aqueous $NH_4Cl$ solution (200 mL) and diethyl ether (150 mL). The aqueous phase was re-extracted with diethyl ether (2 x 150 mL). The combined organics were washed with brine (250 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (1:9)) to afford 1-methyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (**11**) (*A* - 2% {not pure}, *B* - 63%, *C* - 61%,

*D* - 50%, *E* - 31%) as a pale yellow oil; *B* - $[\alpha]_D^{25}$ 0.0 (c=1, $CHCl_3$), *C* - $[\alpha]_D^{25}$ -31.6 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{25}$ -35.3 (c=1, $CHCl_3$)].[224]

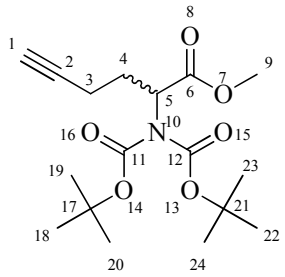
Scheme 21, method e, entry F in Table 2: To (*S*)-(-)-1-methyl-2-[bis-(*t*-butoxy-carbonylamino)]-5-oxopentanoate (**10**) (1.5 g, 4.3 mmol, 1 eq) dissolved in toluene (10 mL) at 0°C was added Tebbe reagent (0.5 M in toluene) (9.0 mL, 4.5 mmol, 1.05 eq). The reaction was warmed to room temperature. After 30 minutes, t.l.c. (ethyl acetate/petrol (2:8)) indicated the formation of product ($R_f$ 0.6) with complete consumption of the starting material ($R_f$ 0.2). The reaction mixture was diluted with diethyl ether (60 mL). Then, NaOH (0.1 M, 1 mL) was added dropwise. The reaction mixture was dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (1:9)) to afford (**-11**) (194 mg, 13%) as a pale yellow oil. $[\alpha]_D^{25}$ -39.5 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{25}$ -35.3 (c=1, $CHCl_3$)];[224] IR (neat) $\nu_{max}$ 2984 (C-H), 2959 (C-H), 1748 (C=O), 1703 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 1.50 (s, 18H, $C(CH_3)_3$, H-3/H-6/H-7/H-12/H-17/H-18), 2.11-2.23 (m, 1H, $CH_2$, H-19), 2.38-2.68 (m, 3H, $CH_2$, H-19/H-20), 3.73 (s, 3H, $OCH_3$, H-24), 4.90 (dd, *J* 9.5 Hz, *J* 5.0 Hz, 1H, NCH, H-13), 9.78 (s, 1H, CHO, H-21); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ ppm 22.5 ($CH_2$, C-19), 27.9 ($C(CH_3)_3$, C-3/C-6/C-7/C-12/C-17/C-18), 40.4 ($CH_2$, C-20), 52.2 ($OCH_3$, C-24), 57.2 (NCH, C-13), 83.4 ($O\underline{C}(CH_3)_3$, C-2/C-11), 151.9 (C=O, C-4/C-9), 170.7 (C=O, C-14), 200.9 (CHO, C-21); $ESI^+$ LRMS: calcd. for $C_{16}H_{27}NO_7$ ($M+Na^+$) 366.19, found 366.18.

**Dimethyl-(1-diazo-2-oxopropylphosphonate) (Ohira's Reagent) (12)**



NaH (60% in mineral oil, 2.6 g, 66.2 mmol, 1.1 eq) was dissolved in THF (100 mL) and toluene (100 mL). The reaction mixture was cooled to 0°C. Then, dimethylacetylmethylphosphonate (10.0 g, 60.2 mmol, 1 eq) in THF (20 mL) was added via cannula. The reaction mixture was stirred for 1 hour. Then, the sulphonyl azide (15.9 g, 66.2 mmol, 1.1 eq) in THF (20 mL) was added via cannula. The reaction mixture was warmed to room temperature. After 6 hours, t.l.c. (100% ethyl acetate) indicated the formation of product ($R_f$ 0.3) with complete consumption of the starting material ($R_f$ 0.1). The reaction mixture was filtered through a pad of Celite, which was washed with ethyl acetate (3 x 150 mL). The combined organics were evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (2:8)) to afford Ohira's reagent (**12**) (6.52 g, 56%) as a bright yellow oil. $R_f$ 0.3 (100% ethyl acetate) [Lit. $R_f$ 0.3 (100% ethyl acetate)];[230] IR (neat) $\nu_{max}$ 2960 (C-H), 2121 (C=$N_2$), 1656 (C=O) $cm^{-1}$; $^1$H NMR (400 MHz, $CDCl_3$) δ ppm 2.26 (s, 3H, $CH_3$, H-1), 3.82, 3.85 (2 x s, 2 x 3H, $OCH_3$, H-7/H-8); $^{13}$C NMR (101 MHz, $CDCl_3$) δ ppm 27.0 ($CH_3$, C-1), 53.5, 53.5 ($OCH_3$, C-7/C-8), 189.77, 189.90 (d, 1C, *J* 12.6 Hz); $ESI^+$ LRMS: calcd. for $C_5H_9N_2O_4P$ ($M+Na^+$) 215.01, found 214.87.

**Methyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexynoate (13)**




To a solution of (S)-(-)-1,5-dimethyl-2-[bis(*t*-butoxycarbonyl)amino] pentanedioate (**9**) (450 mg, 1.2 mmol, 1 eq) in diethyl ether (50 mL) at -78°C under an argon atmosphere was added dropwise diisobutylaluminium hydride (D*i*BAL-H 1.0 M

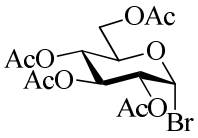
solution in hexane, 1.3 mL, 1.3 mmol, 1.1 eq). After 3 hours, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of aldehyde intermediate (**10**) ($R_f$ 0.6) with complete consumption of the starting material ($R_f$ 0.7). The reaction mixture was partitioned between Rochelle salt (sat., 100 mL) and diethyl ether (100 mL). The aqueous phase was re-extracted with diethyl ether (2 x 50 mL). The combined organics were washed with brine (250 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo* to afford the aldehyde intermediate. Then, to a solution of the aldehyde intermediate in methanol (50 mL) at room temperature under an argon atmosphere was added $K_2CO_3$ (663 mg, 4.8 mmol, 4 eq) and Ohira's reagent (**12**) (268 mg, 1.4 mmol, 1.2 eq). After 18 hours, t.l.c. (pentane/ethyl acetate (1:1)) indicated the formation of product ($R_f$ 0.8) with complete consumption of the starting material ($R_f$ 0.5). The reaction mixture was partitioned between $H_2O$ (100 mL) and ethyl acetate (100 mL). The aqueous phase was re-extracted with ethyl acetate (3 x 100 mL). The combined organics were washed with brine (250 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (2.5:97.5)) to afford methyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexynoate (**13**) (232 mg, 57% over 2 steps) as a colourless oil. $[\alpha]_D^{25}$ ±0.0 (c=1, $CHCl_3$); IR (neat) $\nu_{max}$ 3280 (C≡C-H), 2981 (C-H), 2943 (C-H), 1745 (C=C), 1720 (C=C), 1367 (C=C) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 1.50 (s, 18H, $CCH_3$, H-18/H-19/H-20/H-22/H-23/H-24), 1.97 (t, $J_{1,3}$ 2.5 Hz, 1H, HC≡, H-1), 2.06-2.15 (m, 1H, $CH_2$, H-4), 2.23-2.44 (m, 3H, $CH_2$, H-3/H-4), 3.72 (s, 3H, $OCH_3$, H-9), 5.01 (dd, *J* 8.8 Hz, *J* 4.7 Hz, 1H, CH, H-5); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ ppm 15.6 ($CH_2$, C-3), 27.9 ($C(CH_3)_3$, C-18/C-19/C-20/C-22/C-23/C-24), 29.2 ($CH_2$, C-4 ), 52.2 ($OCH_3$, C-9), 57.1 (NCH, C-5), 69.1 (HC≡, C-1), 83.1, 83.2 (≡C- and $OC(CH_3)_3$, C-2/C-17/C-21), 151.8 (C=O, C-

11/C-12), 170.9 (C=O, C-6); $ESI^+$ LRMS: calcd. for $C_{17}H_{27}NO_6$ ($M+Na^+$) 364.17, found 364.20.

### 4.2. 1-Glycosyl thiols

Carbohydrates and their derivatives have been named in accordance with IUPAC recommendations and numbered according to the carbohydrate convention.[386]

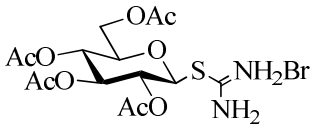
**2,3,4,6-Tetra-*O*-acetyl-α-D-glucopyranosyl bromide (18)**




Pentaacetyl-β-D-glucopyranose (20.0 g, 51.2 mmol, 1 eq) was dissolved in DCM (150 mL). To this, hydrogen bromide (33% in acetic acid, 120 mL) was added. The reaction mixture was stirred at room temperature under argon. After 2 hours, t.l.c (ethyl acetate/petrol (1:1)) indicated the formation of a product ($R_f$ 0.6) with complete consumption of the starting material ($R_f$ 0.5). The reaction mixture was partitioned between DCM (200 mL) and $H_2O$ (200 mL), and the aqueous layer re-extracted with DCM (3 x 200 mL). The combined organic layers were washed with sodium hydrogen carbonate (1 L of a saturated aqueous solution) until pH 8 was obtained, washed with brine (500 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. Crystallisation from ethyl acetate/petrol afforded 2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl bromide (**18**) (19.7 g, 94%) as a white crystalline solid. $R_f$ 0.6 (ethyl acetate/petrol (1:1)) [Lit. $R_f$ 0.6 (ethyl acetate/petrol (1:1)];[387] m.p. 81-83°C [Lit. m.p 89-91 °C];[388] $[\alpha]_D^{25}$ +190.9 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{25}$ +197.0 (c=1, $CHCl_3$)];[389] IR (neat) $\nu_{max}$ 2967 (C-H), 1742 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 2.04, 2.06, 2.10, 2.11 (4 x s, 12H, $COCH_3$), 4.09-4.17 (m, 1H, H-6b), 4.15-4.36 (m, 2H, H-

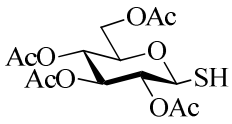
5/H-6a), 4.83 (dd, $J_{2,3}$ 9.9 Hz, $J_{2,1}$ 4.1 Hz, 1H, H-2), 5.15 (app t, $J_{2,1}$ 9.8 Hz, 1H, H-4), 5.53-5.60 (m, 1H, H-3), 6.60 (d, $J_{1,2}$ 3.9 Hz, 1H, H-1); $^{13}$C NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.5, 20.6, 20.6 (4 x CO*C*$H_3$), 60.9 (C-6), 67.1 (C-4), 70.1 (C-3), 70.5 (C-2), 72.0 (C-5), 86.5 (C-1), 169.4, 169.7, 169.8, 170.4 (4 x *C*OCH$_3$).

**2,3,4,6-Tetra-*O*-acetyl-β-D-glucopyranosyl-1-isothiouronium bromide (19)**



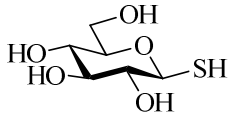

Thiourea (5.5 g, 72.0 mmol, 1.5 eq) and 2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl bromide (**18**) (19.7 g, 48.0 mmol, 1 eq) were dissolved in acetone (100 mL) under argon. The reaction mixture was heated to 60°C. After 2 hours, a white solid precipitated. This was removed by filtration and the filtrate returned to reflux. This process was repeated until the solid ceased to precipitate. The precipitates were combined and re-crystallised from acetone/petrol to afford 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl-1-isothiouronium bromide (**19**) (15.0 g, 64%) as a white crystalline solid. $R_f$ 0.7 (propan-2-ol/acetic acid/$H_2O$ (3:1:1)); m.p. 203-204°C [Lit. 191 °C];[390] $[\alpha]_D^{25}$ -7.5 (c=1, $H_2O$) [Lit. $[\alpha]_D^{25}$ -7.6 (c=1.4, $H_2O$)];[390] IR (neat) $\nu_{max}$ 3267 (N-H), 3026 (C-H), 1751 (C=O), 1655 (C=N) cm$^{-1}$; $^1$H NMR (400 MHz, DMSO-$d_6$) δ ppm 1.98, 2.00, 2.02, 2.06 (4s, 4 x COC*H*$_3$), 4.04-4.12 (m, 1H, H-6b), 4.15-4.24 (m, 2H, H-5/H-6a), 5.11 (m, 2H, H-2/H-4), 5.32 (app t, *J* 8.9 Hz, 1H, H-3), 5.70 (d, $J_{1,2}$ 9.8 Hz, 1H, H-1), 9.08 (br s, 2H, $NH_2$), 9.26 (br s, 2H, $NH_2$); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ ppm 21.1, 21.1, 21.2, 21.4 (4 x CO*C*$H_3$), 62.0 (C-6), 68.2, 69.5 (C-2/C-4) 73.2 (C-3), 76.1 (C-5), 80.5 (C-1), 167.1, 170.1, 170.2, 170.3, 170.8 (4 x *C*OCH$_3$, 1 x C=N); ESI$^-$ LRMS: calcd. for $C_{15}H_{23}BrN_2O_9S$ (M-Br$^+$) 407.11, found 407.14.

**2,3,4,6-Tetra-*O*-acetyl-1-thio-β-D-glucopyranoside (20)**




$Na_2S_2O_5$ (6.0 g, 31.7 mmol, 1.4 eq) and 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl-1-isothiouronium bromide (**19**) (11.0 g, 22.6 mmol, 1 eq) were added to a stirring mixture of DCM (100 mL) and $H_2O$ (50 mL). The reaction mixture was heated to reflux under argon. After 3 hours, t.l.c. (petrol/ethyl acetate (1:1)) indicated the formation of product ($R_f$ 0.5) with complete consumption of the starting material ($R_f$ 0.0). The reaction mixture was cooled to room temperature and the phases separated. The aqueous layer was re-extracted with DCM (2 x 100 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo* to afford 2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-glucopyranoside (**20**) (8.38 g, 100%) as a white crystalline solid. m.p. 69-70°C [Lit. 113-115 °C];[391] $[\alpha]_D^{25}$ +5.5 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{25}$ +6.5 (c=1, $CHCl_3$)];[392] IR (neat) $\nu_{max}$ 2941 (C-H), 2567 (S-H), 1738 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 2.01, 2.02, 2.08, 2.09 (4s, 4 x COC$H_3$), 2.31 (d, $J_{1,SH}$ 10.0 Hz, 1H, SH), 3.73 (ddd, $J_{5,4}$ 9.9 Hz $J_{5,6b}$ 4.8 Hz, $J_{5,6a}$ 2.1 Hz, 1H, H-5), 4.12 (dd, $J_{6a,6b}$ 12.5 Hz, $J_{5,6a}$ 2.3 Hz, 1H, H-6a), 4.25 (dd, $J_{6b,6a}$ 12.5 Hz $J_{5,6b}$ 4.8 Hz, 1H, H-6b), 4.55 (d, *J* 10.0 Hz, 1H, H-1), 4.97 (d, *J* 9.7 Hz, 1H, H-2), 5.10 (app t, *J* 9.7 Hz, 1H, H-4), 5.19 (app t, *J* 9.3 Hz, 1H, H-3); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.5, 20.7, 20.7 (4 x CO*C*$H_3$), 61.0 (C-6), 68.0 (C-4) 73.5, 73.5 (C-2/C-3), 76.3 (C-5), 78.7 (C-1), 169.3, 169.6, 170.1, 170.6 (4 x *C*OCH$_3$); $ESI^+$ LRMS: calcd. for $C_{146}H_{20}O_9S$ ($M+Na^+$) 387.07, found 387.09.

**1-Thio-β-D-glucose (β-GlcSH) (21)**

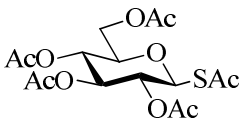



Scheme 26, method e: To 2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-glucopyranoside (**20**) (1.0 g, 2.7 mmol, 1 eq) was added sodium methoxide [sodium (95 mg, 4.1 mmol, 1.5 eq) dissolved in methanol (10 mL)]. The reaction mixture was stirred at room temperature. After 10 minutes, t.l.c (100% ethyl acetate) indicated the formation of product ($R_f$ 0.0) with complete consumption of the starting material ($R_f$ 0.9). The reaction mixture was acidified with DOWEX 50WX8-200, filtered and concentrated *in vacuo* to afford 1-thio-β-D-glucose (β-GlcSH) (**21**) (570 mg, 100%) as a pale yellow oil. $R_f$ 0.3 ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)); $[\alpha]_D^{19}$ +35.7 (c=1, $H_2O$); IR (neat) $\nu_{max}$ 3337 (O-H), 2933 (C-H), 2499 (S-H) $cm^{-1}$; $^1H$ NMR (400 MHz, methanol-$d_4$) δ ppm 3.28-3.33 (m, 4H, H-2/H-3/H-4/H-5), 3.64 (dd, $J_{6b,6a}$ 12.0 Hz, $J_{6b,5}$ 5.1 Hz, 1H, H-6b), 3.84 (dd, $J_{6a,6b}$ 12.1 Hz, $J_{6a,5}$ 1.8 Hz, 1H, H-6a), 4.41 (d, $J_{1,2}$ 9.3 Hz, 1H, H-1); $^{13}C$ NMR (101 MHz, methanol-$d_4$)[393, 394] δ ppm 62.9 (C-6), 71.5, 78.3, 79.3, 82.0 (C-2/C-3/C-4/C-5), 82.5 (C-1); $ESI^-$ LRMS: calcd. for $C_6H_{12}O_9S$ ($M-H^+$) 195.03, found 194.95.

Scheme 26, method f: To 1-*S*-acetyl-2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-glucopyranoside (**22**) (50 mg, 0.1 mmol, 1 eq) was added sodium methoxide [sodium (4 mg, 0.2 mmol, 1.5 eq) dissolved in methanol (10 mL)]. The reaction mixture was stirred at room temperature. After 10 minutes, t.l.c. (100% ethyl acetate) indicated the formation of product ($R_f$ 0.0) with complete consumption of the starting material ($R_f$ 0.9). The reaction mixture was acidified with DOWEX 50WX8-200, filtered and concentrated *in vacuo* to afford 1-thio-β-D-glucose (β-GlcSH) (**21**) (9.6 mg, 41%) as a pale yellow oil.

1-Thio-β-D-glucose sodium salt (Sigma-Aldrich) was dissolved in $H_2O$ (50 mL) and was acidified with DOWEX 50WX8-200, filtered, concentrated *in vacuo*, redissolved in $H_2O$ (10 mL) and then lyophilised to afford 1-thio-β-D-glucose (β-GlcSH) (**21**) (100%) as a white solid.

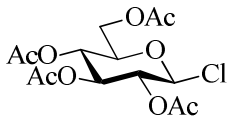
**1-*S*-Acetyl-2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-glucopyranoside (22)**




To a solution of β-D-glucose pentaacetate (5.0 g, 12.8 mmol, 1 eq) dissolved in DCM (200 mL) and under an atmosphere of argon, was added thioacetic acid (1.3 mL, 25.6 mmol, 2 eq), followed by $BF_3.Et_2O$ (3.7 mL, 32 mmol, 2.5 eq). After 18 hours, thioacetic acid (325 μL, 6.4 mmol, 0.5 eq) and $BF_3.Et_2O$ (740 μL, 6.4 mmol, 0.5 eq) were added to the reaction mixture. After a further 18 hours, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.4) with complete consumption of the starting material ($R_f$ 0.5). Triethylamine was added dropwise until effervescence ceased. The reaction mixture was partitioned between sodium hydrogen carbonate (250 mL) and DCM (200 mL). The aqueous phase was re-extracted with DCM (2 x 100 mL). The combined organics were washed with brine (250 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by flash column chromatography (ethyl acetate/petrol (4:6)). Crystallisation from methanol/ethyl acetate afforded 1-*S*-acetyl-2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-glucopyranoside (**22**) (2.48 g, 48%) as a white crystalline solid. $R_f$ 0.5 (toluene/diethyl ether (1:2)) [Lit. $R_f$ 0.5 (toluene/diethyl ether (1:2))];[395] m.p 116-117°C [Lit. m.p. 121 °C];[396] $[\alpha]_D^{19}$ +14.5 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{25}$ +12.2 (c=0.6, $CHCl_3$];[397] IR (KBr) $\nu_{max}$ 2967 (C-H), 1744 (C=O), 1706 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 2.01, 2.03, 2.04, 2.09 (4s, 4 x $COCH_3$), 2.40 (s, $CSCH_3$), 3.85 (ddd, $J_{5,4}$ 10.1 Hz, $J_{5,6b}$ 4.3

Hz, $J_{5,6a}$ 2.1 Hz , 1H, H-5), 4.11 (dd, $J_{6a,6b}$ 12.6 Hz, $J_{6a,5}$ 2.0 Hz, 1H, H-6a), 4.27 (dd, $J_{6b,6a}$ 12.5 Hz, $J_{6b,5}$ 4.4 Hz, 1H, H-6b), 5.09-5.17 (m, 2H, H-3/H-4), 5.25-5.31 (m, 2H, H-1/H-2); $^{13}$C NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.6, 20.7 (4 x CO*C*$H_3$), 30.8 (CS*C*$H_3$), 61.6 (C-6), 67.8, 68.9 (C-3/C-4), 73.9 (C-2), 76.3 (C-5), 80.1 (C-1), 169.3, 169.3, 170.0, 170.6 (4 x *C*O$CH_3$), 192.0 (*C*S$CH_3$); ESI$^+$ LRMS: calcd. for $C_{16}H_{22}O_{10}S$ (M+$Na^+$) 429.08, found 429.12.

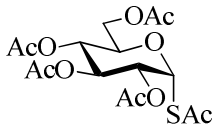
**2,3,4,6-Tetra-*O*-acetyl-β-D-glucopyranosyl chloride (23)**




To pentaacetyl-β-D-glucopyranose (5.0 g, 12.8 mmol, 1 eq) and phosphorus pentachloride (2.9 g, 14.1 mmol, 1.1 eq) in DCM (50 mL) under an atmosphere of argon was added $BF_3.Et_2O$ (20 μL). After 5 minutes, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of a product ($R_f$ 0.6) with complete consumption of the starting material ($R_f$ 0.5). The reaction mixture was partitioned between DCM (30 mL) and $H_2O$ (100 mL) and the aqueous phase was re-extracted with DCM (2 x 50 mL). The combined organics were washed with sodium hydrogen carbonate (100 mL), brine (100 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The resulting solid was co-evaporated with toluene, triturated under cyclohexane and crystallised from diethyl ether/petrol to afford 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl chloride (**23**) (2.31 g, 49%) as a white crystalline solid. m.p 78-79$^o$C [Lit. m.p. 99-100 $^o$C];[398] IR (KBr) $\nu_{max}$ 2965 (C-H), 1756 (C=O) $cm^{-1}$ $cm^{-1}$; $^1$H NMR (400 MHz, $CDCl_3$) δ ppm 2.03, 2.04, 2.10, 2.12 (4s, 4 x CO*C*$H_3$), 3.79-3.84 (m, 1H, H-5), 4.18 (dd, $J_{6a,6b}$ 12.5 Hz $J_{5,6b}$ 2.3 Hz, 1H, H-6b), 4.28 (dd, $J_{6a,6b}$ 12.5 Hz, $J_{5,6a}$ 4.7 Hz, 1H, H-6a), 5.16-5.22 (m, 3H, H-2/H-3/H-4), 5.31 (dd, $J_{1,2}$ 8.1Hz, 1H, H-1); $^{13}$C NMR (101 MHz, $CDCl_3$) δ

ppm 20.5, 20.5, 20.7, 20.7 (4 x CO*C*$H_3$), 61.5 (C-6), 67.5, 72.7, 73.3 (C-2/C-3/C-4), 75.5 (H-5), 87.5 (H-1), 169.0, 169.2, 170.0, 170.6 (4 x *C*O$CH_3$).

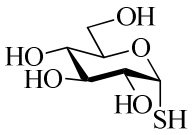
**1-*S*-acetyl-2,3,4,6-tetra-*O*-acetyl-1-thio-α-D-glucopyranoside (24)**




To a solution of 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl chloride (**23**) (200 mg, 2.7 mmol, 1 eq) in DMPU (1 mL) at 0°C under an atmosphere of argon was added potassium thioacetate (68 mg, 0.5 mmol, 1.1 eq). After 3 days, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.4) with almost complete consumption of starting material ($R_f$ 0.6). The reaction mixture was partitioned between ethyl acetate (50 mL) and $H_2O$ (50 mL) and the organic phase was washed with $H_2O$ (2 x 50 mL). The combined aqueous layers were re-extracted with ethyl acetate (3 x 50 mL) and the combined organics washed with brine (100 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (3:7)) to afford 1-*S*-acetyl-2,3,4,6-tetra-*O*-acetyl-1-thio-α-D-glucopyranoside (**24**) (80 mg, 40%) as a orange coloured solid. m.p 121-122°C [Lit. m.p. 128-129°C];[263] $[\alpha]_D^{25}$ +122.2 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{20}$ +135 (c=1, $CHCl_3$];[399] IR (KBr) $\nu_{max}$ 2950 (C-H), 1753 (C=O) $cm^{-1}$; $^1$H NMR (400 MHz, $CDCl_3$) δ ppm 2.02, 2.03, 2.04, 2.09 (4s, 4 x $COCH_3$), 2.44 (s, 3H, SCO*C*$H_3$), 3.96 (ddd, $J_{5,4}$ 9.9 Hz, $J_{5,6a}$ 4.0 Hz, $J_{5,6b}$ 2.4 Hz, 1H, H-5), 4.06 (dd, $J_{6b,6a}$ 12.5 Hz, $J_{6b,5}$ 2.2 Hz, 1H, H-6b), 4.29 (dd, $J_{6a,6b}$ 12.5 Hz, $J_{6a,5}$ 4.1 Hz, 1H, H-6a), 5.11 (t, $J_{4,5}$ 9.8 Hz, 1H, H-4), 5.19 (t, $J_{3,2}$ 10.1 Hz, 1H, H-3), 5.25 (dd, $J_{2,3}$ 10.1 Hz, $J_{2,1}$ 5.1 Hz, 1H, H-2), 6.22 (d, $J_{1,2}$ 5.1 Hz, 1H, H-1); $^{13}$C NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.6, 20.7 (4 x CO*C*$H_3$), 31.4 (SCO*C*$H_3$), 61.5 (C-6), 67.9 (C-4), 69.0 (C-2), 71.1 (C-3), 71.5 C-5),

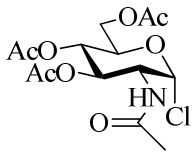
80.3 (C-1), 169.3, 169.4, 170.0, 170.6 (4 x *C*OCH$_3$), 191.3 (*C*SCH$_3$); ESI$^+$ LRMS: calcd. for $C_{16}H_{22}O_{10}S$ (M+Na$^+$) 429.08, found 429.18.

**1-Thio-α-D-glucopyranose (α-GlcSH) (25)**



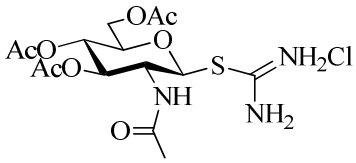

To 1-*S*-acetyl-2,3,4,6-tetra-*O*-acetyl-1-thio-α-D-glucopyranoside (**24**) (0.22 g, 0.5 mmol, 1 eq) was added sodium methoxide (25% wt. solution, 0.19 ml) dissolved in methanol (5 mL). The reaction mixture was stirred at room temperature. After 15 minutes, t.l.c (ethyl acetate; 100%) indicated the formation of product with complete consumption of the starting material ($R_f$ 0.9). The reaction mixture was acidified with DOWEX 50WX8-200, filtered and concentrated *in vacuo* to afford 1-thio-α-D-glucopyranose (α-GlcSH) (**25**) (100 mg, 100%) as a pale yellow solid. $R_f$ 0.4 ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)); m.p 119 $^oC$; $[\alpha]_D^{19}$ +106.0 (c=1, $H_2O$); IR (neat) $\nu_{max}$ 3315 (O-H), 2922 (C-H), 2479 (S-H) cm$^{-1}$; $^1$H NMR (400 MHz, methanol-$d_4$) δ ppm 3.29-3.31 (m, 1H, H-4), 3.56 (app t, *J* 9.6 Hz, 1H, H-3), 3.64 (dd, $J_{2,3}$ 9.6 Hz, $J_{1,2}$ 5.1 Hz, 1H, H-2), 3.68 (dd, $J_{6b,6a}$ 11.8 Hz, $J_{6b,5}$ 5.4 Hz, 1H, H-6b), 3.79 (dd, $J_{6a,6b}$ 11.7 Hz, $J_{6a,5}$ 2.2 Hz, 1H, H-6a), 3.95 (ddd, $J_{5,4}$ 9.9 Hz, $J_{5,6b}$ 5.3 Hz, $J_{5,6a}$ 2.2 Hz, 1H, H-5), 5.63 (d, $J_{1,2}$ 5.1 Hz, H-1);[400] $^{13}$C NMR (101 MHz, methanol-$d_4$) δ ppm 62.6 (C-6), 71.7 (C-4), 72.9 (C-2), 74.6 (C-5), 75.0 (C-3), 82.1 (C-1);[394] ESI$^-$ LRMS: calcd. for $C_6H_{12}O_9S$ (M-H$^+$) 195.03, found 195.00.

**2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl chloride (26)**



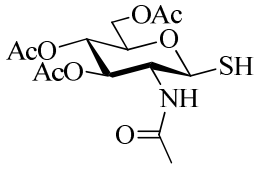

*N*-Acetyl-D-glucosamine (5.0 g, 22.6 mmol, 1 eq) was suspended in acetyl chloride (30 mL, 140.1 mmol, 6.2 eq). After 12 hours, t.l.c. (100% ethyl acetate) indicated the formation of product ($R_f$ 0.6) with complete consumption of starting material ($R_f$ 0.0). The solution was diluted with DCM (100 mL) and poured into ice water (100 mL). The phases were separated and the aqueous layer re-extracted with DCM (3 x 50 mL). The combined organics were washed with sodium hydrogen carbonate (100 mL of a saturated aqueous solution), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (2:8)) to afford 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl chloride (**26**) (4.04 g, 51%) as a pale yellow solid. $R_f$ 0.2 (ethyl acetate/petrol (1:1)); m.p 128-129°C [Lit. m.p. 133-134°C];[401] IR (KBr) $\nu_{max}$ 3246 (N-H), 2979 (C-H), 1742 (C=O), 1644 (C=O), 1542 (C=O) $cm^{-1}$; $^1$H NMR (400 MHz, $CDCl_3$) δ ppm 1.99 (s, NHCOC$H_3$), 2.06, 2.06, 2.11 (3s, 3 x COC$H_3$), 4.14 (dd, *J* 12.3 Hz, *J* 1.9 Hz, 1H, H-6b), 4.25-4.32 (m, 2H, H-5/H6a), 4.54 (ddd, $J_{2,3}$ 12.5 Hz, $J_{2,NH}$ 8.8 Hz, $J_{2,1}$ 3.7 Hz, 1H, H-2), 5.22 (t, *J* 9.6 Hz, 1H, H-4), 5.33 (app t, 1H, H-3), 5.83 (d, $J_{NH,1}$ 8.7 Hz, 1H, NH), 6.19 (d, $J_{2,1}$ 3.6 Hz, 1H, H-2); $^{13}$C NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.6 (2s, 3 x CO*C*$H_3$), 23.0 (NHCO*C*$H_3$), 53.4 (C-2), 61.1 (C-6), 66.8 (C-4), 70.1 (C-3), 70.8 (C-5), 93.5 (C-1), 169.1, 170.1, 170.5, 171.4 (3 x *C*OCH$_3$ and NH*C*OCH$_3$).

**2,3,4,6-Tetra-*O*-acetyl-β-D-glucopyranosyl-1-isothiouronium chloride (27)**



2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl chloride (**26**) (4.0 g, 11.3 mmol, 1 eq) and thiourea (1.5 g, 20.1 mmol, 1.7 eq) were dissolved in acetone (60 mL) under argon. The reaction mixture was heated to 60°C. After 2 hours, a white solid precipitated. The precipitate was removed by filtration, the filtrate was returned to reflux and this process was repeated until the solid ceased to precipitate. The solids were combined and re-crystallised from acetone/petrol to afford 2-acetamido-3,4,6-tri-*O*-acetyl-β-D-glucopyranosyl-1-isothiouronium chloride (**27**) (4.59 g, 95%) as a white crystalline solid. $R_f$ 0.0 (100% ethyl acetate); m.p 177-178°C [Lit. m.p. 135°C];[402] $[\alpha]_D^{25}$ -22.8 (c=1, $H_2O$) [Lit. $[\alpha]_D^{25}$ -29.3 (c=1, $H_2O$];[395] IR (KBr) $\nu_{max}$ 3230 (N-H), 3053 (C-H), 1751 (C=O), 1654 (C=N) $cm^{-1}$; $^1H$ NMR (400 MHz, DMSO-$d_6$) δ ppm 1.81 (s, 3H, NHCOC$H_3$), 1.94, 1.98, 2.02 (3s, 3 x COC$H_3$), 3.97-4.09 (m, $J_{6b,5}$ 1.7 Hz, $J_{6b,6a}$ 12.0 Hz, 2H, H-2/H-6b), 4.15-4.26 (m, $J_{6a,6b}$ 12.2 Hz, $J_{6a,5}$ 4.8 Hz, 2H, H-5/H-6a); 4.94 (app t, *J* 9.7 Hz, 1H, H-4), 5.13 (app t, *J* 9.8 Hz, 1H, H-3), 5.68 (d, $J_{1,2}$ 9.7 Hz, 1H, H-1), 8.43 (d, $J_{NH,2}$ 9.0 Hz, 1H, NH), 9.21, 9.41 (2s, 2 x $NH_2$); $^{13}C$ NMR (101 MHz, DMSO-$d_6$) δ ppm 20.3, 20.4, 20.6 (3 x CO*C*$H_3$), 22.5 (NHCO*C*$H_3$), 52.0 (C-2), 61.5 (C-6), 67.9 (C-4), 72.7 (C-3), 74.7 (C-5), 80.6 (C-1), 169.3, 169.6, 169.9, 170.1 (3 x *C*O$CH_3$ and NH*C*O$CH_3$); ESI$^-$ LRMS: calcd. for $C_{15}H_{24}ClN_3O_8S$ (M-$Cl^+$) 406.13, found 406.16.

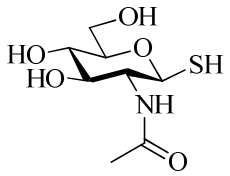
**2-Acetamino-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio-β-D-glucopyranose (28)**




2-acetamido-3,4,6-tri-*O*-acetyl-β-D-glucopyranosyl-1-isothiouronium chloride (**27**) (4.5 g, 10.6 mmol, 1 eq) and sodium metabisulfite (2.5 g, 12.7 mmol, 1.2 eq) were added to a stirring mixture of DCM (30 mL) and $H_2O$ (20 mL). The mixture was heated to reflux under argon. After 2 hours, t.l.c (100% ethyl acetate) indicated the formation of product with complete consumption of the starting material. The reaction mixture was cooled to room temperature and the phases separated. The aqueous layer was re-extracted with DCM (2 x 50 mL). The combined organic layers were washed with $H_2O$ (50 mL), brine (50 mL), dried over $MgSO_4$, filtered and the solvent removed *in vacuo*. The resulting white solid was crystallised from ethyl acetate/petrol to afford 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio-β-D-glucopyranose (**28**) (3.30 g, 89%) as a white solid. $R_f$ 0.7 ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)); m.p. 145-146°C [Lit. m.p. 167-168°C];[403] $[\alpha]_D^{25}$ -15.5 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{25}$ -14.5 (c=0.9, $CHCl_3$];[403] IR (KBr) $\nu_{max}$ 3338 (N-H), 2946 (C-H), 2557 (S-H), 1746 (C=O), 1663 (C=O), 1533 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 1.99 (s, 3H, NHCOC$H_3$), 2.03, 2.05, 2.10 (3s, 3 x COC$H_3$), 2.57 (d, $J_{SH,1}$ 9.3 Hz, 1H, SH), 3.69 (ddd, $J_{5,4}$ 9.6 Hz, $J_{5,6a}$ 4.6 Hz, $J_{5,6b}$ 2.1 Hz, 1H, H-5), 4.08-4.17 (m, $J_{6b,6a}$ 12.5 Hz, $J_{6b,5}$ 2.3 Hz, 2H, H-2/H-6b), 4.24 (dd, $J_{6a,6b}$ 12.5 Hz, $J_{6a,5}$ 4.8 Hz, 1H, H-6a), 4.59 (app t, *J* 9.8 Hz, 1H, H-1), 5.06-5.15 (m, 2H, H-3/H-4), 5.72 (d, $J_{NH,2}$ 9.4 Hz, 1H, NH); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.6, 20.7 (CO*C*$H_3$), 23.2 (s, 1C, NHCO*C*$H_3$), 56.7 (C-2), 62.1 (C-6), 68.0, 73.4 (C-3/C-4), 76.2 (C-5), 80.0 (C-1), 169.2, 170.4,

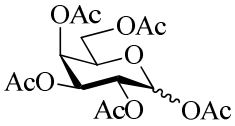
170.7, 171.2 (3 x *C*OCH$_3$ and NH*C*OCH$_3$); ESI$^+$ LRMS: calcd. for $C_{14}H_{21}NO_8S$ (M+Na$^+$) 386.09, found 386.15.

**_N_-Acetyl-1-thio-β-D-glucosamine (β-GlcNAcSH) (29)**



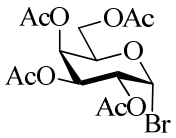

To 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio-β-D-glucopyranose (**28**) (0.22 g, 0.5 mmol, 1 eq) was added sodium methoxide (25 wt% solution, 0.18 mL) and methanol (5 mL). The reaction mixture was stirred at room temperature. After 15 minutes, t.l.c ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) indicated the formation of product ($R_f$ 0.5) with complete consumption of the starting material. The reaction mixture was acidified with DOWEX 50WX8-200, filtered and concentrated *in vacuo* to afford 2-acetamido-2-deoxy-1-thio-β-D-glucopyranose (β-GlcNAcSH) (**29**) (0.12 g, 100%) as a white solid. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)); m.p 167-168 $^oC$; IR (KBr) $\nu_{max}$ 3291 (O-H), 2882 (C-H), 1635 (C=O), 1547 (C=O), cm$^{-1}$; $^1$H NMR (400 MHz, $D_2O$) δ ppm 2.00 (s, 3H, NHCOC*H*$_3$), 3.40-3.50 (m, 3H, H-3/H-4/H-5), 3.64-3.73 (m, 2H, H-2/H-6b), 3.84 (dd, $J_{6a,6b}$ 12.3 Hz, $J_{6a,5}$ 1.1 Hz, 1H, H-6a), 4.63 (d, $J_{1,2}$ 10.2 Hz, 1H, H-1); $^{13}$C NMR (101 MHz, $D_2O$) δ ppm 22.5 (NHCO*C*H$_3$), 58.2 (C-2), 61.0 (C-6), 69.9 (C-4), 75.2 (s, C-3), 79.3 (C-1), 80.5 (C-5), 174.9 (NH*C*OCH$_3$); ESI$^-$ MHMS: calcd. for $C_6H_{20}O_9S$ (M-H$^+$) 236.06, found 236.08.

**1,2,3,4,6-Pentaacetyl galactose (30)**



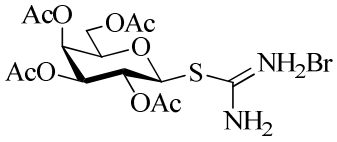

To a solution of D-(+)-galactose (50.0 g, 462.2 mmol, 1 eq) in pyridine (200 mL) at 0°C and under an argon atmosphere, was slowly added acetic anhydride (400 mL). After 24 hours, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.5) with complete consumption of the starting material ($R_f$ 0.0). The reaction mixture was concentrated *in vacuo* and re-dissolved in toluene. This was repeated a further two times. Then, the reaction mixture was concentrated *in vacuo* and re-dissolved in ethanol before being concentrated *in vacuo* again to give a pale yellow solid. Crystallisation from ethanol afforded 1,2,3,4,6-pentaacetyl galactose (**30**) (92.2 g, 51%) as a white solid. IR (KBr) $\nu_{max}$ 2961 (C-H), 1763 (C=O) $cm^{-1}$; $^1$H NMR (400 MHz, acetone-$d_6$) δ ppm 1.95, 1.97, 1.99, 2.15, 2.16 (5s x COC*H*$_3$), 4.05 (dd, $J_{6a,6b}$ 11.2 Hz, $J_{5,6b}$ 6.6 Hz, 1H, H-6b), 4.12 (dd, $J_{6a,6b}$ 11.2 Hz, $J_{5,6a}$ 6.4 Hz, 1H, H-6a), 4.52 (m, 1H, H-5), 5.22 (dd, $J_{2,3}$ 10.9 Hz, $J_{1,2}$ 3.6 Hz, H-2), 5.35 (dd, $J_{2,3}$ 10.9 Hz, $J_{3,4}$ 3.2 Hz, H-3), 5.50 (dd, $J_{3,4}$ 3.2 Hz, H-4), 6.31 (d, $J_{1,2}$ 3.6 Hz, H-1); $^{13}$C NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.6, 20.6, 20.6, 20.8 (CO*C*H$_3$), 61.2 (C-6), 66.4 (C-4), 67.3, 67.3 (C-2/C-3), 68.7 (C-5), 89.7 (C-1), 170.1 (5 x *C*OCH$_3$); ESI$^+$ LRMS: calcd. for $C_{14}H_{21}NO_8S$ (M+$Na^+$) 413.11, found 413.18.

**2,3,4,6-tetra-*O*-Acetyl-α-D-galactopyranosyl bromide (31)**



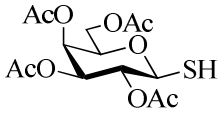

1,2,3,4,6-Pentaacetyl galactose (**30**) (20.0 g, 51.2 mmol, 1 eq) was dissolved in DCM (120 mL). To this, hydrogen bromide (33% in acetic acid, 150 mL) was added. The reaction mixture was stirred under argon at room temperature. After 2 hours, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.6) with complete consumption of the starting material ($R_f$ 0.5). The reaction mixture was partitioned between DCM (100 mL) and $H_2O$ (100 mL) and the aqueous phase was re-extracted with DCM (2 x 50 mL). The combined organics were washed with sodium hydrogen carbonate until pH 8 was obtained, brine (100 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo* to afford 2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl bromide (**31**) (19.3 g, 92%) as a yellow oil. IR (neat) $\nu_{max}$ 2950 (C-H), 1741 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 2.00, 2.05, 2.11 2.14 (4s, 4 x COC$H_3$), 4.10 (dd, $J_{6a,6b}$ 11.3 Hz, $J_{5,6b}$ 6.6 Hz, 1H, H-6b), 4.18 (dd, $J_{6a,6b}$ 11.3 Hz, $J_{5,6a}$ 6.3 Hz, 1H, H-6a), 4.48 (dd, $J_{5,6}$ 6.4 Hz, 1H, H-5), 5.04 (dd, $J_{2,3}$ 10.7 Hz, $J_{1,2}$ 4.0 Hz, 1H, H-2), 5.39 (dd, $J_{2,3}$ 10.6 Hz, $J_{3,4}$ 3.2 Hz, 1H, H-3), 5.51 (dd, $J_{3,4}$ 3.2 Hz, $J_{4,5}$ 1.0 Hz, 1H, H-4), 6.7 (d, $J_{1,2}$ 4.0 Hz, 1H, H-1); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.5, 20.5, 20.6 (CO$CH_3$), 60.7 (C-6), 66.9 (C-4), 67.7 (C-2), 67.9 (C-3), 70.9 (C-5), 88.0 (C-1), 169.7, 169.8, 170.0, 170.2 ($C$OCH$_3$).

**2,3,4,6-Tetra-*O*-acetyl-β-D-galactopyranosyl-1-isothiouronium bromide (32)**



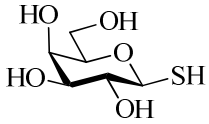

2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl bromide (**31**) (19.0 g, 46.2 mmol, 1 eq) and thiourea (5.3 g, 69.3 mmol, 1.5 eq) were dissolved in acetone (50 mL) and heated to 60°C under argon. After 2 hours, t.l.c. (100% ethyl acetate) indicated the formation of product ($R_f$ 0.0) with complete consumption of the starting material ($R_f$ 0.9). The reaction mixture was concentrated *in vacuo*. Crystallisation from acetone/petrol afforded 2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl-1-isothiouronium bromide (**32**) (11.1 g, 50%) as a white solid. m.p 167-168°C [Lit. m.p. 169.5°C][255]; $[\alpha]_D^{25}$ +13.7 (c=1, EtOH) [Lit. $[\alpha]_D^{25}$ +16.0 (c=1.56, EtOH)];[255] IR (KBr) 3459 (N-H), 2950 (C-H), 1756 (C=O), 1650 (C=N) $cm^{-1}$; $^1H$ NMR (400 MHz, DMSO-$d_6$) δ ppm 1.95, 2.01, 2.08, 2.14 (4s, 4 x COC$H_3$), 4.08 (m, 2H, H-6a/H-6b), 4.44 (app t, *J* 6.3 Hz, 1H, H-5), 5.11 (dd, $J_{2,1}$ 9.9 Hz, 1H, H-2), 5.23 (dd, $J_{3,2}$ 9.9 Hz, $J_{3,4}$ 3.4 Hz, 1H, H-3), 5.38 (dd, $J_{4,3}$ 3.4 Hz, 1H, H-4), 5.71 (d, $J_{1,2}$ 9.9 Hz, 1H, H-1), 9.12, 9.35 (2s, 4H, 2 x $NH_2$); $^{13}C$ NMR (101 MHz, DMSO-$d_6$) δ ppm 20.3, 20.4, 20.5 (3s, 4 x COC$H_3$), 61.2 (C-6), 66.3 (C-2), 67.1 (C-4), 70.4 (C-3), 74.4 (C-5), 80.1 (C-1), 166.3 (SC=N), 169.3, 169.7, 169.8, 169.9 (*C*OCH$_3$); ESI$^-$ LRMS: calcd. for $C_{15}H_{23}BrN_2O_9S$ ($M-Br^+$) 407.11, found 407.24.

**2,3,4,6-Tetra-*O*-acetyl-1-thio-β-D-galactopyranose (33)**



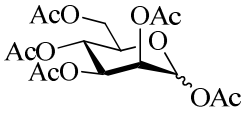

2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl-1-isothiouronium bromide (**32**) (10.0 g, 20.6 mmol, 1 eq) and sodium metabisulfite (5.5 g, 28.8 mmol, 1.4 eq) were added to a stirring solution of DCM (50 mL) and $H_2O$ (50 mL). The mixture was heated to reflux under argon. After 3 hours, t.l.c (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.5) with complete consumption of starting material ($R_f$ 0.0). The reaction mixture was cooled to room temperature and the phases separated. The aqueous layer was re-extracted with DCM (2 x 100 mL). The combined organic layers were washed with $H_2O$ (50 mL), brine (50 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo* to afford 2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-galactopyranose (**33**) (7.5 g, 92%) as a pale yellow oil. [Lit. m.p. 86.5-88 $^oC$];[404] $[\alpha]_D^{19}$ +26.4 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{23}$ +32.0 (c=1, $CHCl_3$];[405] IR (neat) $\nu_{max}$ 2950 (C-H), 2550 (S-H), 1749 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 1.98, 2.04, 2.08, 2.16 (4s, 4 x COC$H_3$), 2.36 (d, $J_{SH,1}$ 9.8 Hz, 1H, SH), 3.94 (t, *J* 6.5Hz, 1H, H-5), 4.12 (d, *J* 6.4Hz, 2H, H-6a/H-6b), 4.53 (t, $J_{1,SH}$ 9.8 Hz, 1H, H-1), 5.01 (dd, $J_{3,2}$ 10.1 Hz, $J_{3,4}$ 3.4 Hz, 1H, H-3), 5.17 (t, *J* 9.8 Hz, 1H, H-2), 5.42 (d, $J_{4,3}$ 3.2 Hz, 1H, H-4); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.6, 20.6, 20.8 ($CH_3$), 61.4 (C-6), 67.2 (C-4), 70.7 (C-2), 71.5 (C-3), 74.9 (C-5), 79.1 (C-1), 169.8, 169.9, 170.1, 170.3 ($COCH_3$); $ESI^+$ LRMS: calcd. for $C_{14}H_{20}O_9S$ ($M+Na^+$) 387.07, found 387.15.

**1-Thio-β-D-galactose (β-GalSH) (34)**



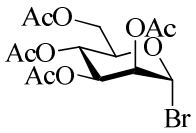

To 2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-galactopyranose (**33**) (1.0 g, 2.7 mmol, 1 eq) was added sodium methoxide [sodium (95 mg, 4.1 mmol, 1.5 eq) dissolved in methanol (10 mL)]. The reaction mixture was stirred at room temperature. After 30 minutes, t.l.c (ethyl acetate; 100%) indicated the formation of product ($R_f$ 0.0) with complete consumption of the starting material ($R_f$ 0.9). The reaction mixture was acidified with DOWEX 50WX8-200, filtered and concentrated *in vacuo* to afford 1-thio-β-D-galactose (β-GalSH) (**34**) (540 g, 100%) as a white solid. $^1$H NMR (400 MHz, $D_2O$) δ ppm 3.45 (t, *J* 9.4 Hz, 1H, H-2), 3.57 (dd, $J_{3,2}$ 9.6 Hz, $J_{3,4}$ 3.4 Hz, 1H, H-3), 3.61-3.72 (m, 3H, H-5/H-6), 3.92 (d, 1H, $J_{4,3}$ 3.8 Hz, H-4), 4.43 (d, $J_{1,2}$ 9.4 Hz, 1H, H-1); $^{13}$C NMR (101 MHz, $D_2O$) δ ppm 61.4 (C-6), 69.4 (C-4), 73.7 (C-2), 73.9 (C-3), 79.2 (C-5), 80.8 (C-1); ESI$^-$ LRMS: calcd. for $C_6H_{12}O_5S$ ($M-H^+$) 195.03, found 195.19.

**1,2,3,4,6-Penta-*O*-acetyl-mannopyranoside (35)**




D-(+)-Mannose (10.0 g, 55.5 mmol, 1 eq) was dissolved in pyridine (120 mL) and cooled to 0°C under an atmosphere of argon. Then, acetic anhydride (100 mL) was added slowly. The reaction mixture was stirred at room temperature. After 1 hour, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.5) with complete consumption of the starting material ($R_f$ 0.0). The reaction mixture was partitioned between $H_2O$ (500 mL) and ethyl acetate (500 mL). The aqueous phase was re-extracted with ethyl acetate (2 x 100 mL). The combined organics were washed with 1 M HCl (500 mL) until acidic pH was reached, saturated aqueous sodium hydrogen carbonate (250 mL) and brine (250 mL). The combined organics were dried over $MgSO_4$, filtered and evaporated *in vacuo* to afford 1,2,3,4,6-penta-*O*-acetyl-mannopyranoside (**35**) (20.5 g, 95%) as a white solid. IR (neat) $\nu_{max}$ 3027 (C-H), 1754 (C=O) $cm^{-1}$; $^1$H NMR (400 MHz, $CDCl_3$) δ ppm 1.99, 2.04, 2.08, 2.16, 2.16 (5s x $COCH_3$), 4.01-4.06 (m, 1H, H-5), 4.07 (dd, $J_{6b,6a}$ 12.5 Hz, $J_{6b,5}$ 2.5 Hz, 1H, H-6b), 4.27 (dd, $J_{6a,6b}$ 12.4 Hz, $J_{6a,5}$ 4.9 Hz, 1H, H-6a), 5.24 (m, 1H, H-2), 5.33 (m, 2H, H-3/H-4), 6.07 (d, $J_{1,2}$ 1.9 Hz, 1H, H-1); $^{13}$C NMR (101 MHz, $CDCl_3$) δ ppm 20.6, 20.6, 20.7, 20.7, 20.8 ($COCH_3$), 62.0 (C-6), 65.5, 68.7 (C-3/C-4), 68.3 (C-2), 70.5 (C-5), 90.5 (C-1), 168.0, 169.5, 169.7, 169.9, 170.6 ($COCH_3$); $ESI^+$ LRMS: calcd. for $C_{16}H_{22}O_{11}$ ($2M+Na^+$) 803.22, found 803.17.

**2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl bromide (36)**



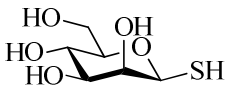

1,2,3,4,6-Penta-*O*-acetyl-mannopyranoside (**35**) (5.0 g, 12.8 mmol, 1 eq) was dissolved in DCM (60 mL). To this, hydrogen bromide (33% in acetic acid, 75 mL) was added and the reaction mixture was stirred under argon at room temperature. After 2 hours, t.l.c (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.6) with complete consumption of the starting material ($R_f$ 0.5). The reaction mixture was partitioned between DCM (100 mL) and ice water (100 mL) and the aqueous layer was re-extracted with DCM (3 x 75 mL). The combined organic layers were washed with sodium hydrogen carbonate (750 mL of a saturated aqueous solution) until pH 8 was obtained, washed with brine (250 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo* to afford 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl bromide (**36**) (5.00 g, 95%) as a pale yellow oil. [Lit. m.p. 48-50 $^oC$];[406] $[\alpha]_D^{19}$ +111.8 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{19}$ +122.0 (c=1, $CHCl_3$];[406] IR (neat) $\nu_{max}$ 2950 (C-H), 1751 (C=O $cm^{-1}$; $^1$H NMR (400 MHz, $CDCl_3$) δ ppm 2.00, 2.06, 2.09, 2.16 (4s, 4 x COC$H_3$), 4.13 (dd, $J_{6b,5}$ 2.1 Hz, $J_{6b,6a}$ 12.4 Hz, 1H, H-6b), 4.21 (ddd, $J_{5,4}$ 10.1 Hz, $J_{5,6a}$ 4.9 Hz, $J_{5,6b}$ 1.7 Hz, 1H, H-5), 4.32 (dd, $J_{6a,6b}$ 12.5 Hz, $J_{6a,5}$ 4.8 Hz, 1H, H-6a), 5.36 (app t, $J$ 10.1 Hz, 1H, H-4), 5.43 (dd, $J_{2,3}$ 3.5 Hz, $J_{2,1}$ 1.5 Hz, 1H, H-2), 5.70 (dd, $J_{3,4}$ 10.2 Hz, $J_{3,2}$ 3.4 Hz, 1H, H-3), 6.29 (d, $J_{1,2}$ 1.0 Hz, 1H, H-1) ; $^{13}$C NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.5, 20.6, 20.7 (CO*C*$H_3$), 61.3 (C-6), 65.2 (C-4), 67.8 (C-3), 72.0 (C-5), 72.7 (C-2), 83.0 (C-1), 169.4, 169.6, 170.4 (3s, 4 x *C*O$CH_3$).

**2,3,4,6-Tetra-*O*-acetyl-1-*S*-acetyl-1-thio-β-D-mannopyranose (37)**




To a solution of 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl bromide (**36**) (1.07 g, 2.6 mmol, 1 eq) in DMPU (6 mL) at room temperature and under an atmosphere of argon was added potassium thioacetate (328 mg, 2.9 mmol, 1.1 eq). After 15 hours, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.4) with almost complete consumption of starting material ($R_f$ 0.6). The reaction mixture was partitioned between ethyl acetate (50 mL) and $H_2O$ (50 mL) and the organic phase was re-extracted with $H_2O$ (2 x 50 mL). The combined aqueous layers were washed with ethyl acetate (3 x 50 mL). The combined organic layers were washed with brine (100 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (3:7)) to afford 2,3,4,6-tetra-*O*-acetyl-1-*S*-acetyl-1-thio-β-D-mannopyranose (**37**) (650 mg, 62%) as a white solid. m.p 122°C [Lit. m.p. 130-131°C];[407] $[\alpha]_D^{19}$ -17.2 (c=1, $CHCl_3$) [Lit. $[\alpha]_D^{19}$ -29.1 (c=1, $CHCl_3$)];[407] IR (neat) $\nu_{max}$ 1739 (C=O), 1638 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) δ ppm 1.99, 2.06, 2.10, 2.20 (4s, 4 x $COCH_3$), 2.38 (s, 3H, $SOCH_3$), 3.82 (ddd, $J_{5,4}$ 9.9 Hz, $J_{5,6a}$ 5.2 Hz, $J_{5,6b}$ 2.2 Hz, 1H, H-5), 4.13 (dd, $J_{6b,6a}$ 12.4 Hz, $J_{6b,5}$ 2.3 Hz, 1H, H-6b), 4.27 (dd, $J_{6a,6b}$ 12.4 Hz, $J_{6b,5}$ 5.2 Hz, 1H, H-6a), 5.15 (dd, $J_{3,4}$ 10.0 Hz, $J_{3,2}$ 3.2 Hz, 1H, H-3), 5.27 (app t, $J$ 10.0 Hz, 1H, H-4), 5.49 (m, 2H, H-1/H-2); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.6, 20.6, 20.7 ($COCH_3$), 30.6 ($SHCOCH_3$), 62.3 (C-6), 65.1 (C-4), 70.5 (C-2), 71.7 (C-3), 76.9 (C-5), 79.2 (H-1), 169.5, 169.8, 169.9, 170.6 ($COCH_3$), 191.5 ($SCOCH_3$); $ESI^+$ LRMS: calcd. for $C_{16}H_{22}O_{10}S$ ($M+Na^+$) 429.08, found 429.06.

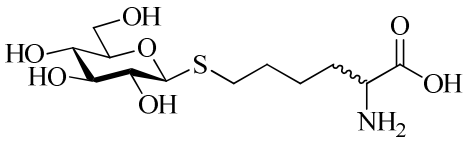
**1-Thio-β-D-mannose (β-ManSH) (38)**



To a stirring solution of 2,3,4,6-tetra-*O*-acetyl-1-*S*-acetyl-1-thio-β-D-mannopyranose (**37**) (304 mg, 0.748 mmol, 1.0 eq) in MeOH (7 ml) was added NaOMe (25% wt. solution, 0.26 ml) in room temperature. After 30 minutes, t.l.c. (100% ethyl acetate) indicated the complete consumption of starting material. The reaction mixture was diluted with MeOH/$H_2O$ (20 ml, 2:1), neutralised with DOWEX 50WX8-200, filtered and concentrated *in vacuo*. Then the residue was redissolved in $H_2O$ followed by lyophilisation to afford 1-thio-β-D-mannose (β-ManSH) (**38**) (140 mg, 100%) as a pale yellow solid. $^1$H NMR (400 MHz, $CDCl_3$) δ ppm 3.29 (ddd, $J_{5,6a}$ 2.0 Hz, $J_{5,6b}$ 6.4 Hz, $J_{5,4}$ 10.0 Hz, 1H, H-5), 3.44 (app t, *J* 10.0 Hz, 1H, H-4), 3.50 – 3.59 (m, 4H, H-3, H-6a,b, SH), 3.83 (d, $J_{2,1}$ 2.4 Hz, 1H, H-2), 4.82 (app s, 1H, H-1); $^{13}$C NMR (100 MHz, $D_2O$) δ ppm 60.9 (C-6), 66.8 (C-4), 70.6 (C-5), 72.4 (C-2), 73.8 (C-3), 80.5 (C-1); ESI$^-$ MHMS: calcd for $C_6H_{11}NO_5$ ($M-H^+$) 195.03, found: 195.03.

## 4.3. 1-Thioglycosides

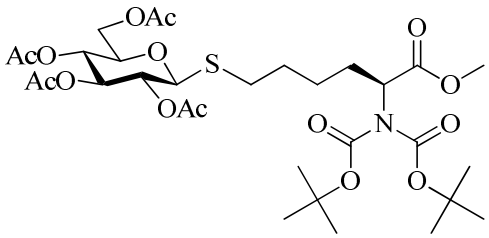
**5-Amino-6-hexanoic acid-(1-thio-β-glucopyranose) (39)**



To a quartz flask was added DL-Hag (**±3**) (20 mg, 0.15 mmol, 1 eq), 1-thio-β-D-glucose sodium salt (Sigma-Aldrich) (40 mg, 0.18 mmol, 1.2 eq), in some examples Vazo44 (10 mg, 0.03 mmol, 0.2 eq), and 0.25 M sodium acetate buffer (pH 4) or sodium phosphate buffer (pH 7) or CAPS buffer (pH 10) (2 mL). The reaction mixture was subjected to hν or room temperature or Δ (55ºC) and was monitored by

t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) (product $R_f$ 0.2). Conditions for individual reactions are presented in Table 6. The reaction mixture was lyophilised and the resulting residue purified by silica gel chromatography ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) and cation exchange chromatography (DOWEX 50WX8-200, ammonium hydroxide eluent) to afford 5-amino-6-hexanoic acid-(1-thio-β-glucopyranose) (**39**). Individual reaction yields are presented in Table 6. $^1$H NMR (400 MHz, $D_2O$) δ ppm 1.35-1.52 (m, 2H, $CHCH_2CH_2$), 1.60-1.69 (m, 2H, $SCH_2CH_2$), 1.77-1.87 (m, 2H, $CHCH_2$), 2.55-2.83 (m, 2H, $SCH_2$); 3.25 (t, *J* 9.7 Hz, 1H, H-2), 3.31-3.45 (m, 3H, H-3/H-4/H-5), 3.61-3.70 (m, 2H, $CH$-$NH_2$/H-6b), 3.84 (dd, $J_{6a,6b}$ 12.3 Hz, $J_{6a,5}$ 2.0 Hz, 1H, H-6a), 4.45 (d, $J_{1,2}$ 9.8 Hz, $J_{C1,H1}$ 156.3 Hz, 1H, H-1); $^{13}$C NMR (101 MHz, $D_2O$) δ ppm 23.7 ($CHCH_2CH_2$), 29.2 ($SCH_2CH_2$), 29.8 ($CHCH_2$), 30.2 ($SCH_2$), 55.0 ($CH$-$NH_2$), 61.2 (C-6), 69.8 (C-2), 72.6 (C-4), 77.5 (C-3), 80.1 (C-5) 85.6 (C-1), 175.1 (*C*OOH); ESI$^+$ MHMS: calcd for $C_{12}H_{23}NO_7S$ (M+$Na^+$) 348.11, found: 384.19.

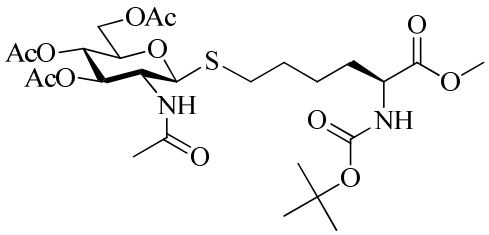
**1-[5-(bis(*t*-butoxycarbonyl)amino)-6-methoxy-6-oxohexylthio]-β-D-glucose tetraacetate (40)**




To a quartz flask was added (S)-1-methyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (**-11**) (50 mg, 0.1 mmol, 1 eq), 2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-glucopyranoside (**20**) (62 mg, 0.12 mmol, 1.2 eq), Vazo44 (9 mg, 0.02 mmol, 0.2 eq), methanol (3 mL) and $H_2O$ (3 mL). The reaction mixture was irradiated with UV light. After 1 hour, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$

0.6) with almost complete consumption of the starting materials. Vazo44 (9 mg, 0.02 mmol, 0.2 eq) and methanol (5 mL) were added to the reaction mixture. After a further 2 hours, t.l.c. (ethyl acetate/petrol (1:1)) indicated the formation of product ($R_f$ 0.6) with complete consumption of the starting materials. The reaction mixture was partitioned between $H_2O$ (50 mL) and ethyl acetate (50 mL). The aqueous phase was re-extracted with ethyl acetate (2 x 50 mL). The combined organics were washed with brine (75 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (3:7)) to afford 1-[5-(bis(*t*-butoxycarbonyl)amino)-6-methoxy-6-oxohexylthio]-β-D-glucose tetraacetate (**40**) (78 mg, 79%) as a gummy white solid. $[\alpha]_D^{25}$ -26.7 (c=1, $CHCl_3$); IR (neat) $\nu_{max}$ 2984 (C-H), 1746 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $CDCl_3$) 1.35-1.49 (m, 2H, $CHCH_2CH_2$), 1.46 (s, 18H, 2 x $C(CH_3)_3$), 1.53-1.69 (m, 2H, $SCH_2CH_2$), 1.78-1.91 (m, 1H, $CHCH_2$), 1.96-2.22 (m, 1H, $CHCH_2$), 1.97, 1.99, 2.02, 2.05 (4 s, 4 x $COCH_3$), 2.64 (m, 2H, $SCH_2$), 3.67 (s, 3H, $OCH_3$), 3.67-3.72 (m, 1H, H-5), 4.05-4.12 (m, 1H, H-6b), 4.22 (dd, $J_{6a,6b}$ 12.3 Hz, $J_{6a,5}$ 4.8 Hz, 1H, H-6a), 4.46 (d, $J_{1,2}$ 9.8 Hz, 1H, H-1), 4.81 (app t, *J* 9.3 Hz, *J* 5.2 Hz, 1H, CH-N), 4.98 (app t, $J_{2,1}$ 9.7 Hz, 1H, H-2), 5.04 (app t, *J* 9.5 Hz, 1H, H-4), 5.18 (app t, *J* 9.3 Hz, 1H, H-3); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ ppm 20.5, 20.5, 20.6, 20.7 ($COCH_3$), 25.4 ($CHCH_2CH_2$), 29.2 ($SCH_2CH_2$), 29.3 ($CHCH_2$), 29.7 ($SCH_2$), 27.9 (2 x $C(CH_3)_3$), 52.1 ($OCH_3$), 57.8 (CH-N), 62.0 (C-6), 68.2 (C-4), 69.8 (C-2), 73.8 (C-3), 75.7 (C-5), 83.0 ($OC(CH_3)_3$), 83.5 (C-1), 152.0 ($COOCH_3$), 169.3, 169.3, 170.1, 170.6, 171.0, 171.2 (4 x $COCH_3$, 2 x NCOO); $ESI^+$ HRMS: calcd for $C_{31}H_{49}NO_{15}S$ ($M+Na^+$) 730.2721, found 730.2714.

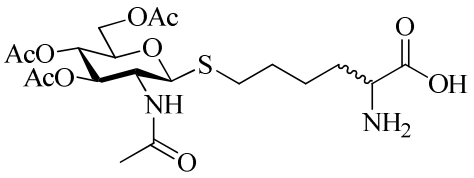
**1-[5-(bis(*t*-butoxycarbonyl)amino)-6-methoxy-6-oxohexylthio]-2-acetamido-3,4,6-tri-*O*-acetyl-β-D-glucose (41)**



To a quartz flask was added (S)-1-methyl-2-[bis(*t*-butoxycarbonyl)amino]-5-hexenoate (**-11**) (50 mg, 0.1 mmol, 1 eq), 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio-β-D-glucopyranose (**28**) (63 mg, 0.12 mmol, 1.2 eq), Vazo44 (9 mg, 0.02 mmol, 0.2 eq), methanol (2 mL) and $H_2O$ (2 mL). The reaction mixture was irradiated with UV light. After 2 hours, t.l.c. (ethyl acetate/petrol; 1:1) indicated the formation of product ($R_f$ 0.4) with complete consumption of the starting materials. The reaction mixture was partitioned between $H_2O$ (50 mL) and ethyl acetate (50 mL). The aqueous phase was re-extracted with ethyl acetate (2 x 50 mL). The combined organics were washed with brine (75 mL), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The resulting residue was purified by silica gel chromatography (ethyl acetate/petrol (4:6)) to afford 1-[5-(bis(*t*-butoxycarbonyl)amino)-6-methoxy-6-oxohexylthio]-2-acetamido-3,4,6-tri-*O*-acetyl-β-D-glucose (**41**) (60 mg, 71%) as a white solid. m.p 154-157 $^{o}$C; $[\alpha]_D^{25}$ -25.6 (c=1, $CHCl_3$); IR (neat) $\nu_{max}$ 3362 (N-H), 3331 (N-H), 2961 (C-H), 1746 (C=O), 1676 (C=O), 1653 (C=O) cm$^{-1}$; $^1$H NMR (400 MHz, DMSO-$d_6$) 1.40 (br. s, 11H, $CHCH_2CH_2/C(CH_3)_3$), 1.52-1.59 (m, 2H, $SCH_2CH_2$), 1.57-1.73 (m, 2H, $CHCH_2$), 1.78 (s, 3H, $COOCH_3$), 1.92, 1.98, 2.01 (3s, 3 x $COCH_3$), 2.54-2.69 (m, 2H, $SCH_2$), 3.64 (s, 3H, OMe), 3.77-3.88 (m, 2H, H-2/H-5), 3.93-4.00 (m, 1H, C*H*-NH), 4.08 (dd, $J_{6b,6a}$ 12.1 Hz, $J_{6b,5}$ 5.0 Hz, 1H, H-6b), 4.14 (dd, $J_{6a,6b}$ 11.8 Hz, $J_{6a,5}$ 2.2 Hz, 1H, H-6a), 4.72 (d, $J_{1,2}$ 10.2 Hz, 1H, H-1), 4.84 (app t, $J$ 9.6 Hz, 1H, H-4), 5.12 (app t, $J$ 9.7 Hz, 1H, H-3), 6.68 (br s, 1H, NH), 7.63 (d, $J$ 8.6 Hz,

1H, NH); $^{13}$C NMR (101 MHz, DMSO-$d_6$) 19.5, 19.6, 19.6 (3 x C($CH_3$)$_3$), 21.9 (NHCO*C*H$_3$), 24.0 (CHCH$_2$*C*H$_2$), 28.3 (SCH$_2$*C*H$_2$), 28.4 (CH*C*H$_2$), 29.9 (S*C*H$_2$), 50.8 (O*C*H$_3$), 52.2 (C-2), 53.2 (*C*H-NH), 61.7 (C-6), 68.7 (C-4), 73.4 (C-3), 74.3 (C-5), 77.7 (O*C*(CH$_3$)$_3$), 82.9 (C-1), 152.2 (*C*OOCH$_3$), 168.4, 168.4, 168.8, 169.1, 172.2 (3 x CO*C*H$_3$, 2 x NH*C*OO); ESI$^+$ HRMS: calcd for $C_{26}H_{42}N_2O_{12}S$ (M+Na$^+$) 629.2356, found 629.2356.

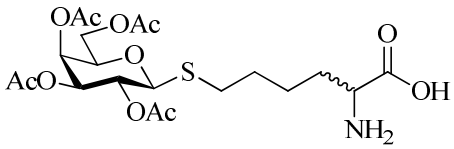
**5-Amino-6-hexanoic acid-(2-acetamido-3,4,6-tri-*O*-acetyl-1-thio-β-D-glucopyranose) (42)**




To a quartz flask was added DL-Hag (**±3**) (50 mg, 0.4 mmol, 1 eq), 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio-β-D-glucopyranose (**28**) (155 g, 0.44 mmol, 1.1 eq), Vazo44 (25 mg, 0.08 mmol, 0.2 eq) and $H_2O$ (5 mL). The reaction mixture was irradiated with UV light. After 3 hours, t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) showed the formation of product ($R_f$ 0.5) with almost complete consumption of the starting materials. 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio-β-D-glucopyranose (**28**) (155 g, 0.44 mmol, 1.1 eq) was added to the reaction mixture. After a further 14 hours, t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) showed the formation of product ($R_f$ 0.5) with almost complete consumption of the starting materials. The reaction mixture was concentrated *in vacuo*, re-dissolved in $H_2O$ and freeze dried. The resulting residue was purified by silica gel chromatography (methanol/DCM (1:9) to (2:8)) to afford 5-amino-6-hexanoic acid-(2-acetamido-3,4,6-tri-*O*-acetyl-1-thio-β-D-glucopyranose) (**42**) (187 mg, 81%) as a pale brown solid. IR (neat) $\nu_{max}$ 3296 (N-H), 1744 (C=O), 1661 (C=O), 1636 (C=O) cm$^{-1}$; $^1$H NMR (400 MHz, $D_2O$) δ ppm 1.34-

1.51 (m, 2H, SCHCH$_2$C$H_2$), 1.58-1.68 (m, 2H, SCH$_2$C$H_2$), 1.76-1.88 (m, 2H, CHC$H_2$), 1.93 (s, 3H, NHCOC$H_3$), 2.01, 2.04, 2.08 (3s, 3 x COC$H_3$), 2.61-2.79 (m, 2H, SC$H_2$), 3.65-3.71 (m, 1H, C*H*-NH$_2$), 3.94-4.01 (m, 2H, H-2/H-5), 4.15 (dd, $J_{6b,6a}$ 12.6 Hz, $J_{6b,5}$ 1.9 Hz, 1H, H-6b), 4.33 (dd, $J_{6a,6b}$ 12.7 Hz, $J_{6a,5}$ 4.0 Hz, 1H, H-6a), 4.79 (d, $J_{1,2}$ 10.6 Hz, 1H, H-1), 5.03 (app t, *J* 9.8 Hz, 1H, H-4), 5.20 (app t, *J* 9.7 Hz, 1H, H-3); $^{13}$C NMR (101 MHz, D$_2$O) δ ppm 20.3, 20.4, 20.5 (CO*C*H$_3$), 22.3 (NHCO*C*H$_3$), 23.9 (CHCH$_2$*C*H$_2$), 29.1, 29.2, (SCH$_2$*C*H$_2$), 30.1, 30.0 (CH*C*H$_2$), 30.2, 30.3 (S*C*H$_2$), 53.1 (C-2), 55.0 (*C*H-NH$_2$), 62.4 (C-6), 68.8 (C-4), 74.4 (C-3), 75.3 (C-5), 84.2 (H-1), 173.1, 173.5, 174.1, 174.6, 175.08 (4 x *C*OCH$_3$, NH*C*OCH$_3$); ESI$^+$ HRMS: calcd for C$_{20}$H$_{32}$N$_2$O$_{10}$S (M+Na$^+$) 493.1856, found 493.1850.

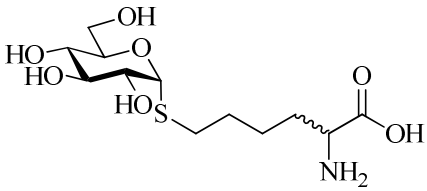
**5-Amino-6-hexanoic acid-(2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-galactopyranose) (43)**




To a quartz flask was added DL-Hag (**±3**) (50 mg, 0.38 mmol, 1 eq), 2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-galactopyranose (**33**) (155 mg, 0.42 mmol, 1.1 eq), Vazo44 (25 mg, 0.07 mmol, 0.2 eq), methanol (5 mL) and H$_2$O (5 mL). The reaction mixture was irradiated with UV light. After 1 hour, t.l.c (H$_2$O/propan-2-ol/ethyl acetate (1:2:2)) showed the formation of product ($R_f$ 0.6) with almost complete consumption of the starting materials. 2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-galactopyranose (**33**) (155 mg, 0.42 mmol, 1.1 eq) was added to the reaction mixture. After a further 2 hours, t.l.c (H$_2$O/propan-2-ol/ethyl acetate (1:2:2)) indicated the formation of product ($R_f$ 0.6) with complete consumption of starting materials. The reaction mixture was concentrated *in vacuo*. The resulting residue was purified by silica gel chromatography (methanol/DCM (1:9)) to afford 5-amino-6-hexanoic acid-(2,3,4,6-

tetra-*O*-acetyl-1-thio-β-D-galactopyranose) (**43**) (108 mg, 57%) as a white solid. IR (neat) $\nu_{max}$ 2948 (C-H), 1743 (C=O) $cm^{-1}$; $^1H$ NMR (400 MHz, $D_2O$) δ ppm 1.29-1.47 (m, 2H, $CHCH_2CH_2$), 1.55-1.65 (m, 2H, $SCH_2CH_2$), 1.71-1.84 (m, 2H, $CHCH_2$), 1.92, 1.99, 2.03, 2.12 (4s, 4 x $COCH_3$), 2.59-2.74 (m, 2H, $SCH_2$), 3.59-3.65 (m, 1H, *C*H-$NH_2$), 4.07-4.20 (m, 3H, H-5/H-6a/H-6b), 4.86 (d, $J_{1,2}$ 9.9 Hz, 1H, H-1), 5.09 (app t, *J* 9.9 Hz, 1H, H-2), 5.20 (dd, $J_{3,2}$ 9.8 Hz, $J_{3,4}$ 3.4 Hz, 1H, H-3), 5.46 (app d, *J* 3.2 Hz, 1H, H-4); $^{13}C$ NMR (101 MHz, $D_2O$) δ ppm 20.3, 20.3, 20.5, 20.6 ($COCH_3$), 23.9 ($CHCH_2CH_2$), 29.4 ($SCH_2CH_2$), 30.3 ($CHCH_2$), 30.3 ($SCH_2$), 55.0 (*C*H-$NH_2$), 62.4 (C-6), 68.4 (C-2), 68.6 (C-4), 72.4 (C-3), 74.4 (C-5), 83.8 (C-1), 173.0, 173.0, 173.4, 173.6, 173.9 (4 x $COCH_3$, COOH); $ESI^+$ HRMS: Calcd. for $C_{20}H_{31}NO_{11}S$ ($M+H^+$) 494.1696. Found 494.1690.

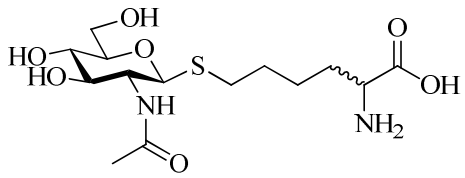
**5-Amino-6-hexanoic acid-(1-thio-α-D-glucopyranose) (44)**




To a quartz flask was added DL-Hag (**±3**) (20 mg, 0.15mmol, 1 eq), α-GlcSH (**25**) (35 mg, 0.18 mmol, 1.2 eq), Vazo44 (10 mg, 0.03 mmol, 0.2 eq) and 0.25 M acetate buffer (pH 4) (2 mL). The reaction mixture was irradiated with UV light. After 8 hours, t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) indicated the formation of product ($R_f$ 0.2) with almost complete consumption of the starting materials. An additional portion of α-GlcSH (**25**) (1 eq) and Vazo44 (0.1 eq) were added, and the reaction mixture was again irradiated. After 4 hours, t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) indicated the formation of product ($R_f$ 0.2). The reaction mixture was lyophilised. The resulting residue was purified by cation exchange (DOWEX 50WX8-200, ammonium hydroxide eluent) and silica gel chromatography (Biotage)

($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) to afford 5-amino-6-hexanoic acid-(1-thio-α-D-glucopyranose) (**44**) (21 mg, 42%) as a colourless solid. $^1$H NMR (400 MHz, $D_2O$) δ ppm 1.28-1.46 (m, 2H, $CHCH_2CH_2$), 1.52-1.65 (m, 2H, $SCH_2CH_2$), 1.69-1.83 (m, 2H, $CHCH_2$), 2.48-2.75 (m, 2H, $SCH_2$), 3.25 (m, 1H, H-4), 3.53 (app t, *J* 9.3 Hz, 1H, H-3) 3.55-3.82 (m, 5H, H-2/H-5/H-6ab/$CH$-$NH_2$), 5.36 (d, $J_{1,2}$ 5.5 Hz, $J_{C1,H1}$ 167.0 Hz, 1H, H-1); $^{13}$C NMR (101 MHz, $D_2O$) 23.8 ($CHCH_2CH_2$), 28.8, 29.2, 30.2 ($SCH_2CH_2$, $CHCH_2$, $SCH_2$), 55.0 ($CH$-$NH_2$), 60.9 (C-6), 70.1 (C-2), 71.3 (C-4), 74.0 (C-3), 80.1 (C-5), 85.8 (C-1); ESI$^-$ MHMS: calcd for $C_{12}H_{23}NO_7S$ ($M$-$H^+$) 324.11, found 324.23.

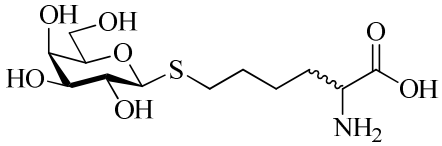
**5-Amino-6-hexanoic acid-(2-acetamido-1-thio-β-D-glucopyranose) (45)**




To a quartz flask was added DL-Hag (**±3**) (20 mg, 0.15 mmol, 1 eq), β-GlcNAcSH (**29**) (43 mg, 0.18 mmol, 1.2 eq), Vazo44 (10 mg, 0.03 mmol, 0.2 eq) and 0.25 M acetate buffer (pH 4) (2 mL). The reaction mixture was irradiated with UV light. After 8 hours, t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) indicated the formation of product ($R_f$ 0.2) with almost complete consumption of the starting materials. The reaction mixture was lyophilised, and the resulting residue was purified by cation exchange (DOWEX 50WX8-200, ammonium hydroxide eluent) and silica gel chromatography (Biotage) ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) to afford 5-amino-6-hexanoic acid-(2-acetamido-1-thio-β-D-glucopyranose) (**45**) (36 mg, 63%) as a colourless solid. IR (KBr) $\nu_{max}$ 3423 (N-H), 2926 (C-H), 1735 (C=O), 1646 (C=O), cm$^{-1}$; $^1$H NMR (400 MHz, $D_2O$) δ ppm 1.33-1.52 (m, 2H, $CHCH_2CH_2$), 1.53-1.68 (m, 2H, $SCH_2CH_2$), 1.78-1.98 (m, 2H, $CHCH_2$), 1.95 (s, 3H, $NHCOCH_3$), 2.50-2.77 (m, 2H, $SCH_2$), 3.43-3.41 (m, 2H, H-4/H-5), 3.49 (t, *J* 8.4 Hz, 1H, H-3), 3.60-3.70 (m, 2H,

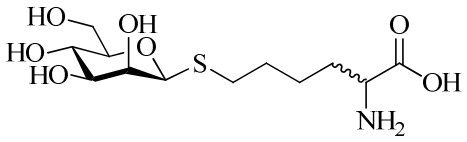
H-2/H-6b), 3.83 (d, *J* 12.3 Hz, 1H, H-6a), 4.00 (t, *J* 6.1 Hz, 1H, C*H*-$NH_2$), 4.53 (d, $J_{1,2}$ 10.4 Hz, 1H, H-1); $^{13}$C NMR (101 MHz, $D_2O$) δ ppm 22.5 (NHCO*C*$H_3$), 23.5, 28.8, 29.5, 29.9 ($CH_2$, $CHCH_2CH_2$, $SCH_2CH_2$, $CHCH_2$, $SCH_2$), 53.1 (*C*H-$NH_2$), 55.0 (C-2), 61.2 (C-6), 70.1 (C-4), 75.4 (C-3), 80.2 (H-5), 84.6 (H-1); ESI$^+$ HRMS calcd. for $C_{14}H_{26}N_2O_7S$ (M+$Na^+$) 389.1358. found 389.1366.

**5-Amino-6-hexanoic acid-(1-thio-β-D-galactopyranose) (46)**



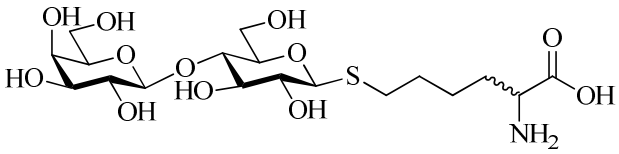

To a quartz flask was added DL-Hag (**±3**) (20 mg, 0.15 mmol, 1 eq), β-GalSH (**34**) (35 mg, 0.18 mmol, 1.2 eq), Vazo44 (10 mg, 0.03 mmol, 0.2 eq) and 0.25 M acetate buffer (pH 4) (2 mL). The reaction mixture was irradiated with UV light. After 8 hours, t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) indicated the formation of product ($R_f$ 0.2) with almost complete consumption of the starting materials. The reaction mixture was lyophilised, and the resulting residue was purified by cation exchange (DOWEX 50WX8-200, ammonium hydroxide eluent) and silica gel chromatography (Biotage) ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) to afford 5-amino-6-hexanoic acid-(1-thio-β-D-galactopyranose) (**46**) (30 mg, 70%) as a colourless solid. $^1$H NMR (400 MHz, $D_2O$) δ ppm 1.36-1.54 (m, 2H, $CHCH_2CH_2$), 1.61-1.71 (m, 2H, $SCH_2CH_2$), 1.76-1.88 (m, 2H, $CHCH_2$), 2.64-2.82 (m, 2H, $SCH_2$), 3.50 (t, *J* 9.6 Hz, 1H, H-2), 3.60 (dd, $J_{3,2}$ 9.6 Hz, $J_{3,4}$ 3.4 Hz, 1H, H-3), 3.63-3.73 (m, 4H, H-5/H-6a/H-6b/C*H*-$NH_2$), 3.92 (d, $J_{1,2}$ 9.7 Hz, 1H, H-1) ; $^{13}$C NMR (101 MHz, $D_2O$) δ ppm 23.8 ($CHCH_2CH_2$), 29.2 ($SCH_2CH_2$), 29.8 ($CHCH_2$), 30.2 ($SCH_2$), 55.0 (*C*H-$NH_2$), 61.4 (C-6), 69.1 (C-4), 69.9 (C-2), 74.2 (C-3), 79.2 (C-5), 86.2 (C-1), *C*OOH not observed; ESI$^-$ MHMS calcd. for $C_{12}H_{23}NO_7S$ (M-$H^+$) 324.11, found 324.26.

**5-Amino-6-hexanoic acid-(1-thio-β-D-mannopyranose) (47)**



To a quartz flask was added DL-Hag (**±3**) (20 mg, 0.15 mmol, 1.0 eq), β-ManSH (**38**) (35 mg, 0.18 mmol, 1.2 eq), Vazo44 (1 mg, 0.03 mmol, 0.2 eq) and 0.25 M acetate buffer (2 mL). The reaction mixture was irradiated using UV light. After 8 hours, t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) indicated the formation of product ($R_f$ 0.2) with almost complete consumption of the starting materials. An additional portion of β-ManSH (**38**) (1 eq) and Vazo44 (0.1 eq) were added, and the reaction mixture was again irradiated. After 4 hours, the reaction mixture was lyophilised and the resulting residue was purified by cation exchange (DOWEX 50WX8-200, ammonium hydroxide eluent) and silica gel chromatography (Biotage) ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) afforded 5-amino-6-hexanoic acid-(1-thio-β-D-mannopyranose) (**47**) (10 mg, 28%) as a colorless solid. $R_f$ 0.2 ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)); $^1$H NMR (400 MHz, $D_2O$) $\delta$ ppm 1.36 – 1.44 (m, 2H, C$H_2$CH$_2$CH), 1.58 – 1.64 (m, 2H, SC$H_2$CH$_2$), 1.75 – 1.79 (m, 2H, CHC$H_2$), 2.62 – 2.72 (m, 2H, SC$H_2$), 3.30 – 3.35 (m, 1H, H-5), 3.48 – 3.53 (m, 1H, H-4), 3.57 (dd, $J_{2,3}$ 2.4 Hz, $J_{3,4}$ 8.0 Hz, 1H, H-3), 3.60 – 3.67 (m, 3H, C*H*-NH$_2$, H-6a,b), 3.94 (app d, $J_{2,3}$ 2.4 Hz, 1H, H-2), 4.76 (app s, 1H, H-1); $^{13}$C NMR (100 MHz, $D_2O$) δ ppm 23.4 (*C*H$_2$CH$_2$CH), 28.4 (SCH$_2$*C*H$_2$), 29.9 (*C*H$_2$CH), 30.4 (S*C*H$_2$), 54.7 (*C*H-NH$_2$), 60.8 (C-6), 66.7 (C-4), 70.7 (C-5), 72.2 (C-2), 73.8 (C-3), 80.2 (C-1); ESI$^+$ HRMS calcd for: $C_{12}H_{22}NO_{10}S$ (M+Na$^+$) 381.0826; found: 381.0826.

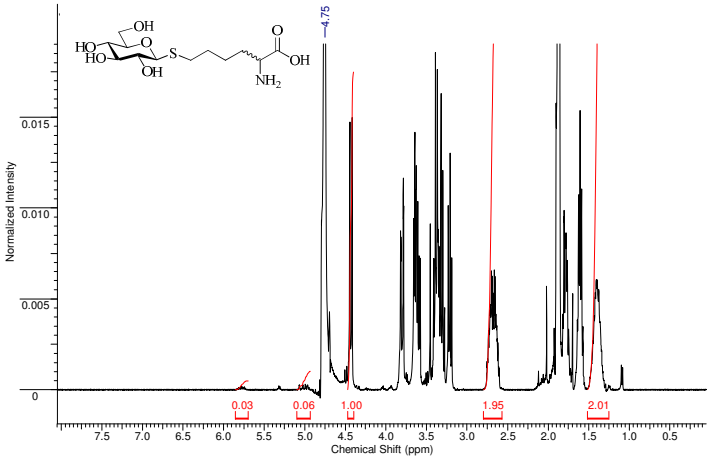
**5-Amino-6-hexanoic acid galactopyranosyl-β-D-(1→4)-1-thio-β-D-glucopyranoside (48)**



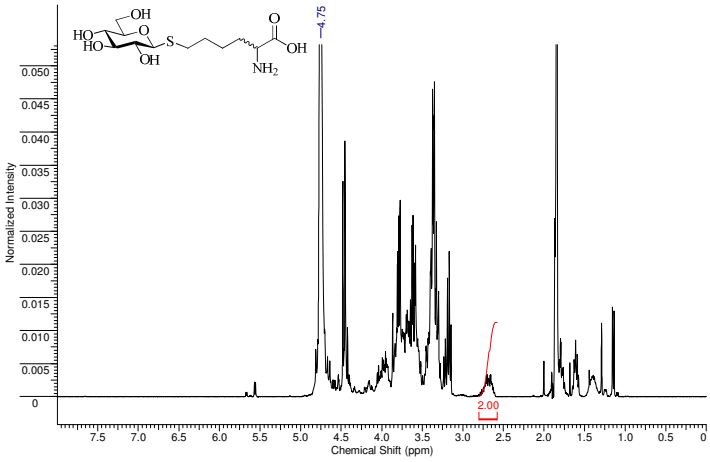
To a quartz flask was added DL-Hag (**±3**) (20 mg, 0.15 mmol, 1.0 eq), LacSH (65 mg, 0.18 mmol, 1.2 eq), Vazo44 (10 mg, 0.03 mmol, 0.2 eq) and 0.25 M acetate buffer (2 mL). The reaction mixture was irradiated with UV light. After 8 hours, an additional portion of LacSH (0.5 eq) and Vazo44 (0.1 eq) were added, and the reaction mixture was again irradiated. After 4 hours, t.l.c. ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) indicated the formation of product ($R_f$ 0.1). The resulting residue was purified by cation exchange (DOWEX 50WX8-200, ammonium hydroxide eluent) and silica gel chromatography (Biotage) ($H_2O$/propan-2-ol/ethyl acetate (1:2:2)) afforded 5-amino-6-hexanoic acid galactopyranosyl-β-D-(1→4)-1-thio-β-D-glucopyranoside (**48**) (40 mg, 56%) as a colorless solid. $^1$H NMR (400 MHz, $D_2O$) $\delta$ ppm 1.32 – 1.41 (m, 2H, C$H_2$CH$_2$CH), 1.53 – 1.60 (m, 2H, SC$H_2$CH$_2$), 1.71 – 1.77 (m, 2H, CHC$H_2$), 2.57 – 2.71 (m, 2H, SC$H_2$), 3.23 (app t, *J* 8.9 Hz, 1H, H-2), 3.41 (app t, *J* 8.8 Hz, 1H, H-2'), 3.48 – 3.54 (m, 4H, H-3, H-5, H-3',H-5'), 3.57 – 3.60 (m, 2H, H-4, C*H*-NH$_2$), 3.63 (app d, $J_{6'a,6'b}$ 4.2 Hz, 2H, H-6'a,b), 3.67 (app d, $J_{6a,6b}$ 4.9 Hz, 1H, H-6b), 3.79 (d, $J_{4'3'}$ 2.8 Hz, 1H, H-4'), 3.83 (dd, $J_{6a,5}$ 12.4 Hz, 1H, H-6a), 4.32 (d, $J_{1'2'}$ 7.7 Hz, 1H, H-1'), 4.42 (d, $J_{1,2}$ 9.9 Hz, 1H, H-1); $^{13}$C NMR (100 MHz, $D_2O$) δ ppm 23.8 (*C*H$_2$CH$_2$CH), 29.2 (SCH$_2$*C*H$_2$), 29.8 (*C*H$_2$CH), 30.4 (S*C*H$_2$), 55.0 (*C*H-NH$_2$), 60.5 (C-6), 61.4 (C-6'), 68.9 (C-4'), 71.3 (C-2'), 72.4 (C-2), 72.8 (C-3'), 75.7 (C-4), 76.1 (C-5'), 78.5 (C-3), 78.9 (C-5), 85.5 (C-1), 103.2 (C-1'), 175.4 (CO); ESI$^+$ HRMS calcd for: $C_{18}H_{33}NO_{12}S$ (M+Na$^+$) 510.1616; found: 510.1629.

### 4.3.1. Crude NMRs

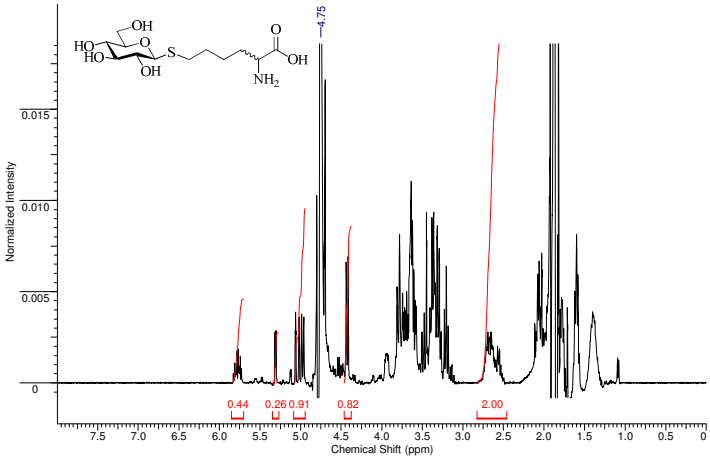
See Table 6. Crude refers to the reaction mixture prior to purification.



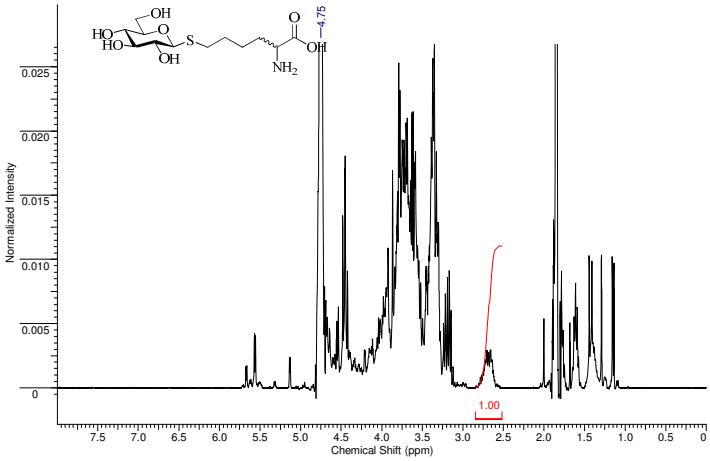

Reaction conditions: hν, DL-Hag (1 eq), β-GlcSH (1.2 eq)



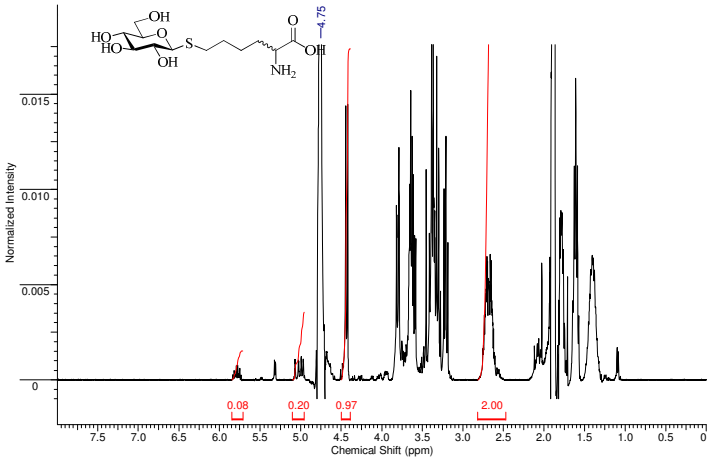

Reaction conditions: hν, DL-Hag (1 eq), β-GlcSH (10 eq), Vazo44 (0.2 eq)



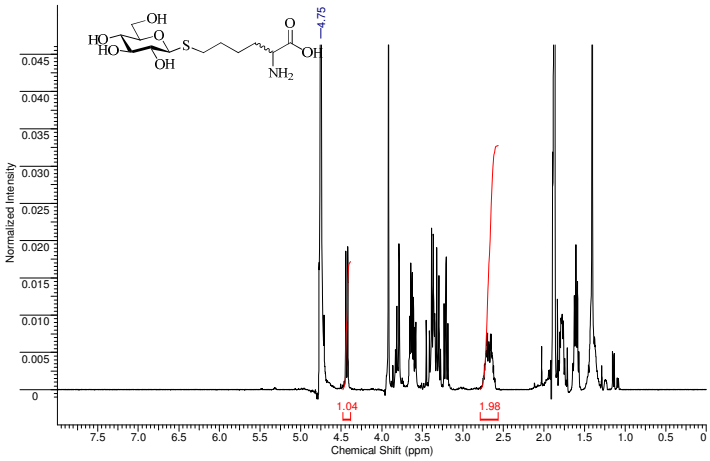

Reaction conditions: 55°C, DL-Hag (1 eq), β-GlcSH (1.2 eq)



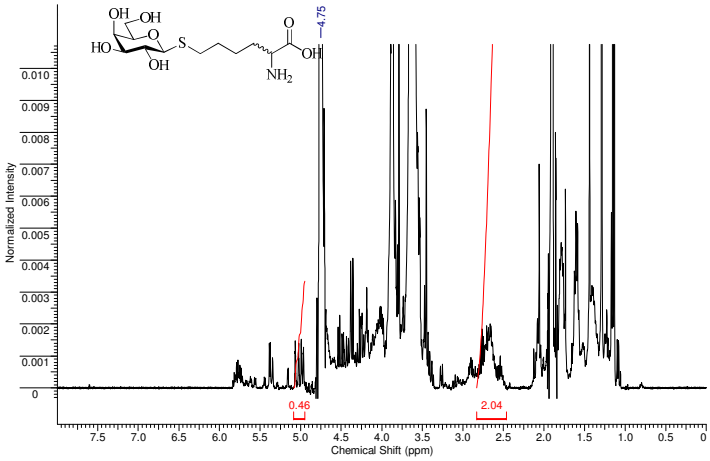

Reaction conditions: 55°C, DL-Hag (1 eq), β-GlcSH (10 eq), Vazo (0.2 eq)



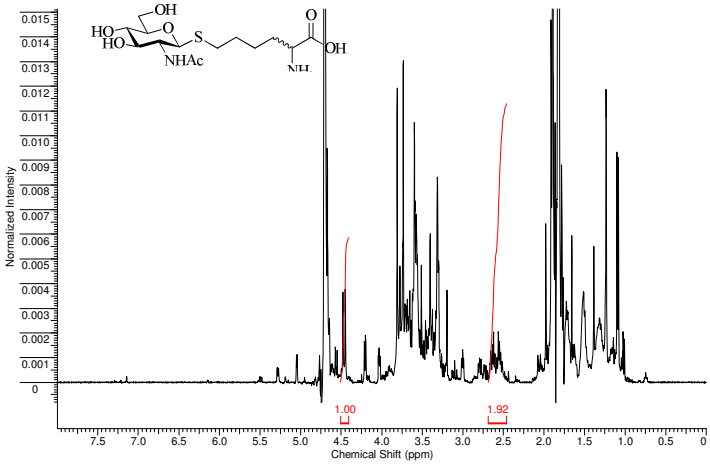

Reaction conditions: RT, DL-Hag (1 eq), β-GlcSH (1.2 eq)



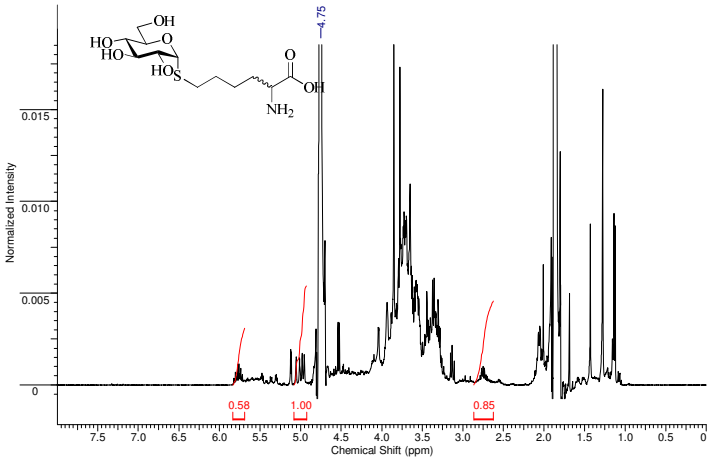

Reaction conditions: RT, DL-Hag (1 eq), β-GlcSH (1.2 eq), Vazo (0.2 eq)



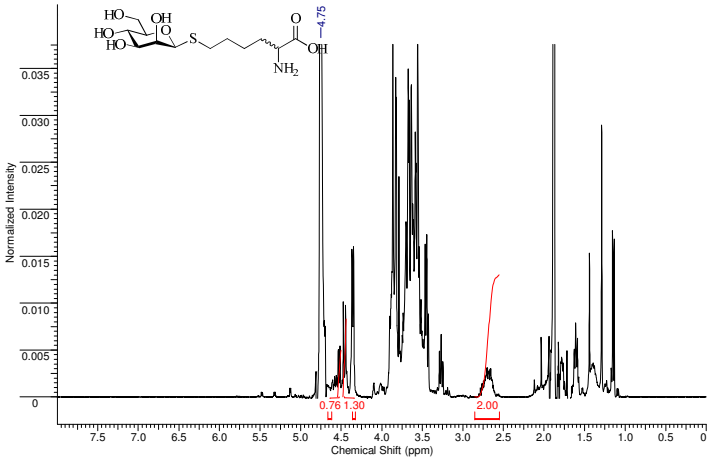

Reaction conditions: hν, DL-Hag (1 eq), β-GalSH (2.2 eq), Vazo44 (0.3 eq)



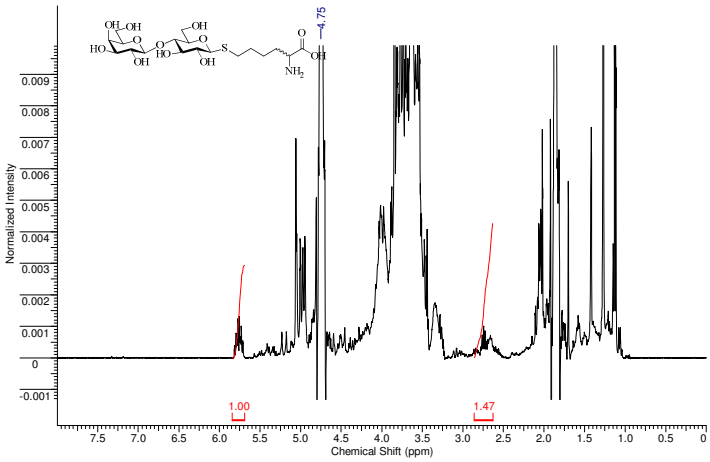

Reaction conditions: hν, DL-Hag (1 eq), β-GlcNAcSH (1.2 eq), Vazo44 (0.2 eq)




Reaction conditions: hν, DL-Hag (1 eq), α-GlcSH (5 eq), Vazo44 (0.3 eq)



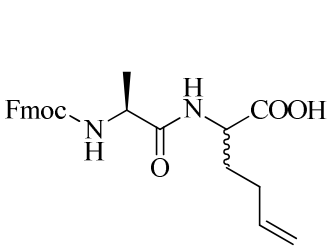

Reaction conditions: hv, DL-Hag (1 eq), α-ManSH (5 eq), Vazo44 (0.3 eq)



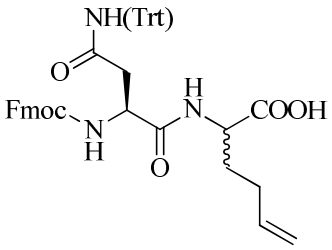

Reaction conditions: hv, DL-Hag (1 eq), LacSH (3 eq), Vazo44 (0.3 eq)

## 4.4. Peptides

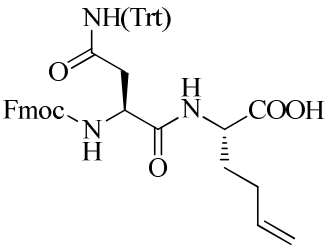
### Fmoc-L-Ala-DL-Hag-OH, Fmoc-L-Asn(Trt)-DL-Hag-OH and Fmoc-L-Asn(Trt)-L-Hag-OH




Fmoc-L-Ala-DL-Hag-OH




Fmoc-L-Asn(Trt)-DL-Hag-OH




Fmoc-L-Asn(Trt)-L-Hag-OH

*Synthesis*: DL- or L-Hag (25 mg, 0.19 mmol, 1 eq), Fmoc-L-Ala-OPfp or Fmoc-L-Asn(Trt)-OPfp (0.19 mmol, 1 eq) and HOBt (0.19 mmol, 1 eq) were dissolved in DMF (10 mL). The reaction mixture was stirred at room temperature. After 16 hours, t.l.c. (methanol/DCM (2:8)) indicated the formation of product with complete consumption of the starting materials.

*Fmoc-L-Ala-DL-Hag-OH HPLC Analysis*: Crude reaction mixture (5 μL) was injected onto an analytical Phenomenex Synergi 4u Hydro-RP 80Å 100 x 4.6 mm column. Elution was performed at 1 mL/min using an solvent gradient system from 100% solvent A (99.9% water, 0.1% TFA) to 100% solvent B (99.9% acetonitrile, 0.1% TFA), or using an isocratic mobile phase (60, 50, 45 or 40% acetonitrile in $H_2O$, 0.1% TFA). Eluent was monitored at $\lambda$ = 214 and 254 nm. Collected peak fractions were analysed by ESI-MS. $ESI^+$ LRMS: calcd. for $C_{24}H_{26}N_2O_5$ ($M+Na^+$) 445.17, found 445.20.

*Fmoc-L-Asn(Trt)-DL-Hag-OH and Fmoc-L-Asn(Trt)-L-Hag-OH HPLC Analysis*: Crude reaction mixture (5 μL) was injected onto an analytical Phenomenex Synergi 4u Hydro-RP 80Å 100 x 4.6 mm column. For Fmoc-L-Asn(Trt)-DL-Hag-OH elution

was performed at 1 mL/min using an solvent gradient system from 100% solvent A (99.9% water, 0.1% TFA) to 100% solvent B (99.9% acetonitrile, 0.1% TFA), or using an isocratic mobile phase (75, 70, 65, 60, 57.5 and 55% acetonitrile in $H_2O$, 0.1% TFA). For Fmoc-L-Asn(Trt)-L-Hag-OH elution was performed at 1 mL/min using an isocratic mobile phase (55% acetonitrile in $H_2O$, 0.1% TFA). Eluent was monitored at $\lambda$ = 214 and 254 nm. Collected peak fractions were analysed by ESI-MS. Fmoc-L-Asn(Trt)-DL-Hag-OH and Fmoc-L-Asn(Trt)-L-Hag-OH: $ESI^-$ LRMS: calcd. for $C_{44}H_{41}N_3O_6$ ($M-H^+$) 706.29, found 706.29.

**H-HWP-$NH_2$ and H-LNMYHWPLPL-$NH_2$**

Peptides H-HWP-$NH_2$ and H-LNMYHWPLPL-$NH_2$ (Figure 46) were prepared using a CEM Liberty Peptide Synthesiser on Rink Amide MBHA resin on a 1 mmol and 0.25 mmol scale, respectively, using standard single cycles of Fmoc microwave assisted automated SPPS, and were cleaved from the resin. The following groups were used to protect amino acid side-chains: *t*-butyl for Tyr, trityl for Asn and His, and BOC for Trp. Procedure for 1 mmol preparation:

(i) Removal of Fmoc protection: The resin was washed to the bottom of the reaction vessel, drained and treated with 20% piperidine in DMF. Nitrogen was bubbled through the reaction mixture and the vessel was irradiated with microwave energy at 25 W for 30 seconds with the temperature cut off set at 75°C. The resin was drained then treated with a further portion of 20% piperidine in DMF and irradiated at 25 W for 180 seconds with a 75°C cut off. The resin was then drained and washed several times with DMF.

(ii) Chain extension with Fmoc-protected amino acids: The resin was washed to the bottom of the reaction vessel, drained and treated with a 0.2 M solution of the Fmoc-protected amino acid in DMF (16 mL) (or 10% DMSO in DMF for Fmoc-protected histidine (16 mL)), 0.45 M *O*-(benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HBTU) in DMF (19 mL) and a 2 M solution of *N*,*N*-diisopropylethylamine (DIEA) (3.5 mL) in *N*-methylpyrrolidinone (NMP) (6.5 mL). Nitrogen was bubbled through the reaction mixture and the vessel was irradiated with microwave energy at 25 W for 300 seconds with the temperature cut off set at 70°C. The resin was then drained and washed several times with DMF.

(iii) Cleavage from TFA labile supports as the deprotected peptide: The resin was washed several times with DCM to remove DMF, drained, treated with 95% TFA / 2.5% triisopropylsilane (TIS) / 2.5% $H_2O$, and irradiated at 15 W with a temperature cut off of 30°C for 1100 seconds. The cleavage cocktail was delivered to the product tube along with two portions of DCM with which the resin was washed.

(iv) Peptide precipitation: The peptides were precipitated on ice from cold diethyl ether and were pelleted by centrifugation. The ether was decanted and the solids washed by two further suspensions in ether, followed by centrifugation. The pellets were dissolved in $H_2O$ and lyophilised to give H-HWP-$NH_2$ as a purple solid and H-LNMYHWPLPL-$NH_2$ as a white solid; $ESI^+$ HRMS: calcd. for H-HWP-$NH_2$ ($M+H^+$) 438.23, found 438.23. $ESI^+$ HRMS: calcd. for H-LNMYHWPLPL-$NH_2$ ($M+H^+$) 1282.67, found. 1282.67.

Circular Dichroism: H-LNMYHWPLPL-$NH_2$ (8.7 mg) was dissolved in buffer (10 mM sodium phosphate, 10 mM NaCl, pH 6.5) (8.7 mL). The solution was diluted 10 fold. Peptide solution (350 μL, 0.1 mg/mL) was added to a quartz cuvette and the

circular dichroism was recorded on an Applied Photophysics Chirascan Circular Dichroism (CD) Spectrometer (Figure 47).

## 5. Protein preparation

### 5.1. Arylsulfatase from *Pseudomonas aeruginosa*

LB media (kanamycin, final conc. 50 μg/mL) (20 mL) was inoculated from a glycerol stock of *E. coli* DH5α containing pME4322 plasmid carrying the arylsulfatase gene. The culture was incubated for 16 hours at 37°C with shaking (220 rpm). Plasmid DNA was purified (Qiagen Spin Miniprep Kit) and transformed into BL21(DE3) competent cells (Novagen) following the manufacturer's protocol. A single colony was used to inoculate LB media (kanamycin, final conc. 50 μg/mL) (100 mL). The culture was incubated for 16 hours at 37°C with shaking (220 rpm). LB media (kanamycin, final conc. 50 μg/mL) (750 mL x 2), prewarmed to 30°C, was inoculated with overnight culture (10 mL). The cultures were grown at 30°C with shaking (220 rpm) to an optical density ($OD_{600}$) of 0.6 - 0.8. Expression was induced by addition of IPTG (final conc. 1 mM) and was continued for 3 hours at 30°C. The cells were harvested by centrifugation (10 minutes, 8000 rpm, 4°C) to give a pellet (~ 10 g), which was collected and stored at -20°C. The cell pellet was thawed; 50 mM TEA (pH 7.5) (40 mL), DNase (3 μL), lysozyme (10 mg) and a Complete Protease Inhibitor Cocktail Tablet (Roche) were added and the cell suspension was incubated for 1 hour on ice before being sonicated (15% power in bursts of 20 seconds with a rest time of 40 seconds between each burst). The cell debris was removed from the lysis mixture by centrifugation (20 minutes, 18000 rpm). The protein was purified from the cleared cell lysate using an AKTA FPLC system and FastQ (Q-Sepharose)

50 mL column. Protein was injected and eluted using a gradient from 100% buffer A (50 mM TEA, pH 7.5) to 100% buffer B (50 mM TEA, 2 M NaCl, pH 7.5) (Figure 100). The eluent was analysed by SDS-PAGE and the fractions shown to contain pure protein were combined and concentrated to 25 mL. A solution comprising 50 mM TEA and 2 M ammonium sulphate (pH 7.5) (25 mL) was slowly added and the protein solution was centrifuged to remove precipitate. Additional protein purification used an AKTA FPLC system and Phenyl Sepharose 75 mL column. Protein was injected and eluted using a gradient from 75% buffer A (50 mM TEA, 2 M ammonium sulphate, pH 7.5) to 100% buffer B (50 mM TEA, pH 7.5) (Figure 102). The eluent was analysed by SDS-PAGE and the fractions shown to contain pure protein were combined and dialysed against 50 mM sodium phosphate buffer (pH 7.5). Enzymatic activity of the protein was demonstrated by 4-nitrocatechol release from 4-nitrocatechol sulfate:[343] A solution of 4-nitrocatechol sulfate containing 70 mM Tris-HCl (211 mg), 70 mM sodium acetate (143 mg) and 10 mM 4-nitrocatechol sulfate dipotassium salt (78 mg) was made up to 25 mL with $H_2O$ and was adjusted to pH 8.8. 4-nitrocatechol sulfate solution (500 μL) was added to a UV cuvette and the reaction was started by addition of arylsulfatase (50 μL). The reaction was incubated at 37 $^{o}$C and the liberated p-nitrocatechol was monitored by measurement of $A_{515}$, after correction for a blank (4-nitrocatechol sulfate solution). ESI-MS of the protein was obtained (Figure 103).

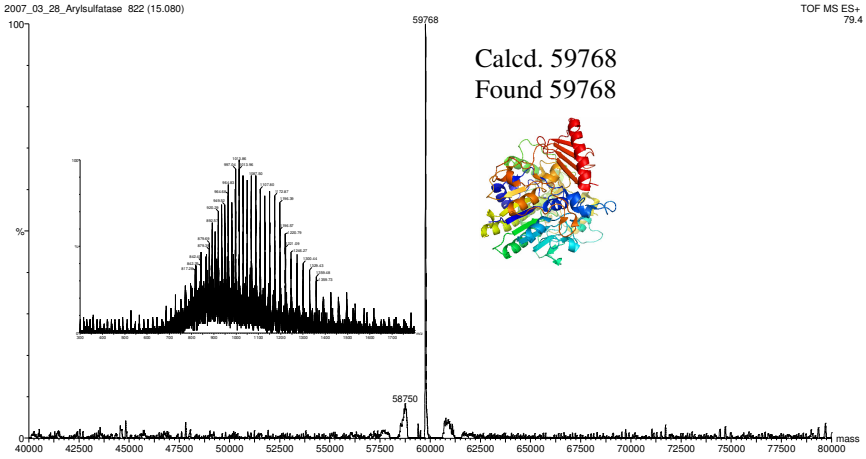



**Figure 103.** Raw (insert) and deconvoluted ESI-MS of Arylsulfatase from *Pseudomonas aeruginosa.*

Protein sequence: [M]SKRPNFLVIVADDLGFSDIGAFGGEIATPNLDALAIAGL RLTDFHTASTGSPTRSMLLTGTDHHIAGIGTMAEALTPELEGKPGYEGHLNER VVALPELLREAGYQTLMAGKWHLGLKPEQTPHARGFERSFSLLPGAANHYG FEPPYDESTPRILKGTPALYVEDERYLDTLPEGFYSSDAFGDKLLQYLKERDQ SRPFFAYLPFSAPHWPLQAPREIVEKYRGRYDAGPEALRQERLARLKELGLVE ADVEAHPVLALTREWEALEDEERAKSARAMEVYAAMVERMDWNIGRVVD YLRRQGELDNTFVLFMSDNGAEGALLEAFPKFGPDLLGFLDRHYDNSLENIG RANSYVWYGPRWAQAATAPSRLYKAFTTQGGIRVPALVRYPRLSRQGAISH AFATVMDVTPTLLDLAGVRHPGKRWRGREIAEPRGRSWLGWLSGETEAAHD ENTVTGWELFGMRAIRQGDWKAVYLPAPVGPATWQLYDLARDPGEIHDLA DSQPGKLAELIEHWKRYVSETGVVEGASPFLVR

[ ] – cleaved during protein expression

## 5.2. Chalcone reductase from *Medicago sativa*

BL21(DE3) competent cells (Novagen) were transformed following the manufacturer's protocol with plasmid carrying the chalcone reductase gene. A single colony was used to inoculate Terrific Broth (TB) media (kanamycin, final conc. 50 µg/mL) (100 mL). The culture was grown for 16 hours at 37°C with shaking (220 rpm). TB media (kanamycin, final conc. 50 µg/mL) (750 mL x 4), prewarmed to 37°C

was inoculated with overnight culture (10 mL). The cultures were grown at 37°C with shaking (220 rpm) to an optical density ($OD_{600}$) of 1.0. Expression was induced by addition of IPTG (final conc. 1 mM) and was continued for 6 hours at 22°C. The cells were harvested by centrifugation (8 minutes, 8000 rpm, 4°C) to give pellets, which were collected and stored at -20°C. The cell pellet was thawed; lysis buffer [500 mM NaCl, 50 mM Tris-HCl, 20 mM imidazole, 10% v/v glycerol, 1% v/v Tween 20, 10 mM β-mercaptoethanol, pH 8.0] (40 mL), DNase (10 μL), lysozyme (10 mg) and a Complete Protease Inhibitor Cocktail Tablet (Roche) were added and the cell suspension was incubated for 1 hour on ice before being sonicated (15% power in bursts of 20 seconds with a rest time of 40 seconds between each burst). The cell debris was removed from the lysis mixture by centrifugation (20 minutes, 18000 rpm). The protein was purified from the cleared cell lysate using an AKTA FPLC system and HisTrap FF 5 mL column (GE Healthcare) equilibrated in lysis buffer. The protein was injected and eluted using a gradient from 100% buffer A (50 mM Tris-HCl, 20 mM imidazole, 500 mM NaCl, 10 mM β-mercaptoethanol, pH 8.0) to 100% buffer B (50 mM Tris-HCl, 250 mM imidazole, 500 mM NaCl, 10 mM β-mercaptoethanol, pH 8.0) (Figure 99). The eluent was analysed by SDS-PAGE and the fractions shown to contain pure protein were combined and dialysed against a solution comprising 50 mM Tris-HCl, 500 mM NaCl and 10 mM β-mercaptoethanol (pH 8.0). ESI-MS of the protein was obtained (Figure 104).

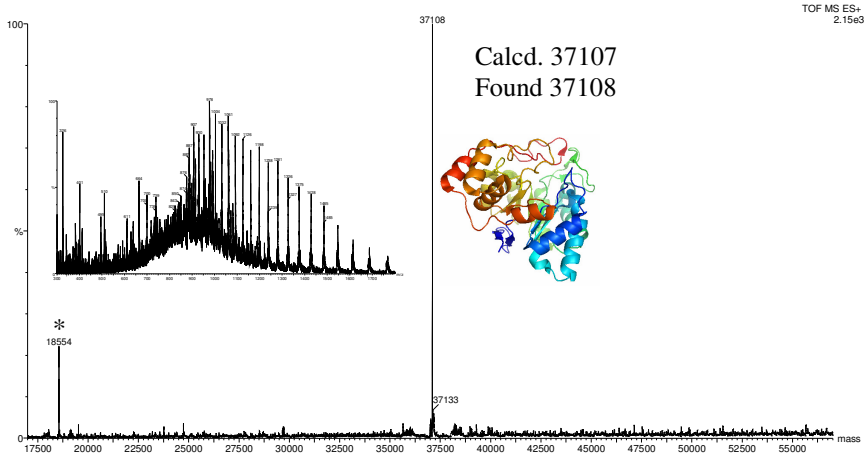



**Figure 104.** Raw (insert) and deconvoluted ESI-MS of Chalcone reductase from *Medicago sativa.* * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

Protein sequence: [M]ASVEIPTKVLTNTSSQLKMPVVGMGSAPDFTCKKDTKD AIIEAIKQGYRHFDTAAAYGSEQALGEALKEAIELGLVTREELFVTSKLWVTE NHPHLVIPALQKSLKTLQLDYLDLYLIHWPLSSQPGKFSFPIDVADLLPFDVKG VWESMEESLKLGLTKAIGVSNFSVKKLENLLSVATVLPAVNQVEMNLAWQQ KKLREFCNANGIVLTAFSPLRKGASRGPNEVMENDMLKEIADAHGKSVAQIS LRWLYEQGVTFVPKSYDKERMNQNLCIFDWSLTKEDHEKIDQIKQNRLIPGPT KPGLNDLYDDDPNSSSVDKLAAALEHHHHHH

[ ] - cleaved during protein expression

## 5.3. β-Glucosidase from *Clostridium thermocellum* (*Ct*βG)

β-Glucosidase from *Clostridium thermocellum* (*Ct*βG) was purchased from Prozomix. The protein was buffer exchanged (PD-10 column) into 50 mM sodium phosphate buffer (pH 6.5). ESI-MS of the protein was obtained (Figure 105).

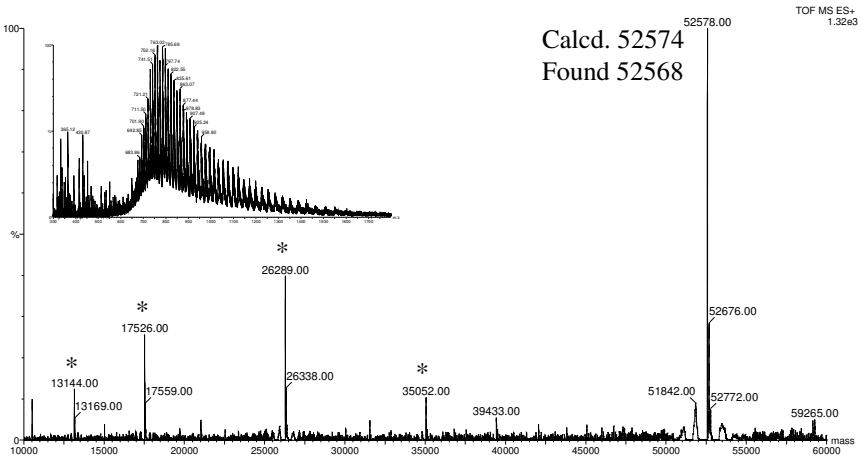



**Figure 105.** Raw (insert) and deconvoluted ESI-MS of *Ct*βG. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

Protein sequence: [M]ASSKITFPKDFIWGSATAAYQIEGAYNEDGKGESIWD RFSHTPGNIADGHTGDVACDHYHRYEEDIKIMKEIGIKSYRFSISWPRIFPEGT GKLNQKGLDFYKRLTNLLLENGIMPAITLYHWDLPQKLQDKGGWKNRDTTD YFTEYSEVIFKNLGDIVPIWFTHNEPGVVSLLGHFLGIHAPGIKDLRTSLEVSH NLLLSHGKAVKLFREMNIDAQIGIALNLSYHYPASEKAEDIEAAELSFSLAGR WYLDPVLKGRYPENALKLYKKKGIELSFPEDDLKLISQPIDFIAFNNYSSEFIK YDPSSESGFSPANSILEKFEKTDMGWIIYPEGLYDLLMLLDRDYGKPNIVISEN GAAFKDEIGSNGKIEDTKRIQYLKDYLTQAHRAIQDGVNLKAYYLWSLLDNF EWAYGYNKRFGIVHVNFDTLERKIKDSGYWYKEVIKNNGFLEHHHHHH
[ ] - cleaved during protein expression

## 5.4. Green fluorescent protein (GFP) from *Aequorea victoria*

NovaBlue GigaSingles competent cells (Novagen) were transformed following the manufacturer's protocol with pGFP plasmid (Clontech) carrying the GFP gene. LB media (ampicillin, final conc. 100 µg/mL) (30 mL) was inoculated with a single colony and the culture was grown for 16 hours at 37°C with shaking (220 rpm). The plasmid DNA was purified (Qiagen QIAprep Spin Miniprep Kit) and transformed into BL21(DE3) competent cells (Novagen) following the manufacturer's protocol. The

colonies were suspended in LB media (carbenicillin, final conc. 50 μg/mL) (4 mL), and were used to inoculate LB media (carbenicillin, final conc. 50 μg/mL) (1 L x 4) prewarmed to 37°C. The cultures were grown at 37°C with shaking (220 rpm) to an optical density ($OD_{600}$) of 0.7. Expression was induced by addition of IPTG (final conc. 1 mM) and was continued for 4 hours at 30°C. The cells were harvested by centrifugation (8 mins, 8000 rpm, 4°C) to give pellets, which were collected and stored at -20°C in 10 mM Tris-EDTA (TE) buffer (pH 7.8) (30 mL). The cell pellet was thawed; DNase (10 μL), lysozyme (10 mg) and a Complete Protease Inhibitor Cocktail Tablet (Roche) were added and the cell suspension was stirred on ice for 1 hour before being lysed by sonication (15% power in bursts of 20 seconds with a rest time of 40 seconds between each burst). The cell debris was removed from the lysis mixture by centrifugation (20 minutes, 18000 rpm). The protein was subjected to a 75% ammonium sulphate cut at 0°C. The precipitant was dissolved in a minimal amount of buffer (10 mM TE, 2 M ammonium sulphate, pH 7.8). Additional protein purification used an AKTA FPLC system and Phenyl Sepharose 75 mL column. Protein was injected and eluted using a gradient from 75% buffer A (10 mM TEA (triethanolamine), 2 M ammonium sulphate, pH 7.8) to 100% buffer B (10 mM TEA, pH 7.8) (Figure 102). The eluent was analysed by SDS-PAGE and the fractions shown to contain pure protein were combined and concentrated to 6 mL. Additional protein purification used an AKTA FPLC system and HiLoad 26/60 Superdex 200 column. The protein was injected and eluted using an isocratic mobile phase (50 mM TEA, 150 mM NaCl, pH 7.5). The eluent was analysed by SDS-PAGE and the fractions shown to contain pure protein were combined and dialysed against 50 mM Tris-HCl (pH 8.0). ESI-MS of the protein was obtained (Figure 103). Protein activity was confirmed qualitatively by demonstrating its ability to fluoresce under UV light.

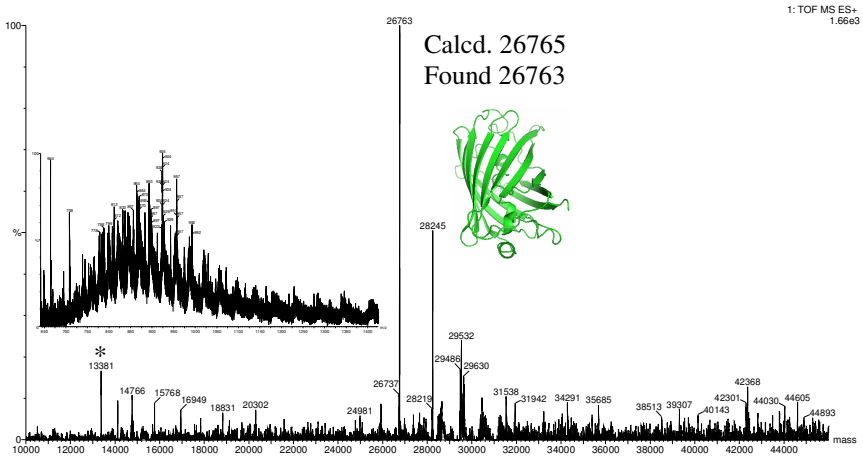



**Figure 106.** Raw (insert) and deconvoluted ESI-MS of GFP. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

Protein sequence: [M]SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGD ATYGKLTLKFICTTGKLPVPWPTLVTTF{SYG}VQCFSRYPDHMKRHDFFKSA MPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHK LEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGP VLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

[ ] - cleaved during protein expression; { } - fluorophore (formed post-translationally)

## 5.5. GFP-*Ss*βG fusion

LB media (kanamycin, final conc. 50 μg/mL) (30 mL) was inoculated from a backstab of TOP10 *E. coli* containing pET24a plasmid carrying the GFP-*Ss*βG fusion gene (synthesised by GenScript). The culture was incubated for 16 hours at 37°C with shaking (220 rpm). Plasmid DNA was purified (Qiagen QIAprep Spin Miniprep Kit) and the concentration determined (NanoDrop ND-1000). DNA sequencing was carried out. BL21(DE3) competent cells (Novagen) were transformed with plasmid DNA following the manufacturer's protocol. The colonies were dissolved in LB media (kanamycin, final conc. 50 μg/mL) (6 mL). LB media (kanamycin, final conc.

50 μg/mL) (1 L x 6), prewarmed to 37°C, was inoculated with the bacterial colonies and the cells were grown at 37°C with shaking (220 rpm) until an optical density ($OD_{600}$) of 0.6 - 0.8 was obtained. Expression was induced by addition of IPTG (final conc. 1 mM) and was continued for 16 hours at 23°C. The cells were harvested by centrifugation (8 mins, 8000 rpm, 4°C) to give pellets, which were collected and stored (-20°C) in binding buffer (20 mM Tris-HCl, 5 mM imidazole, 0.5 M NaCl, pH 7.8) (30 mL). The cell pellet was thawed; a Complete Protease Inhibitor Cocktail Tablet (Roche) was added and the cell suspension was incubated for 1 hour on ice before being sonicated (15% power in 10 bursts of 20 seconds with a rest time of 40 seconds between each burst). The cell debris was removed from the lysis mixture by centrifugation (20 minutes, 18000 rpm, 4°C). The lysate was sterile filtered (0.4 μm). The protein was purified from the cleared cell lysate using an AKTA FPLC system and HisTrap FF 1 mL column. Protein was injected and eluted using a gradient from 100% binding buffer to 100% elution buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, pH 7.8) (Figure 99). All eluent collected from the purification was analysed by SDS-PAGE and that shown to contain pure protein was combined. The protein was additionally purified using an AKTA FPLC system and HiLoad 26/60 Superdex 200 column. The protein was injected and eluted isocratically using an isocratic mobile phase (50 mM TEA, 150 mM NaCl, pH 7.5). All eluent collected from the purification was analysed by SDS-PAGE and that shown to contain pure protein was combined. The protein was further purified using an AKTA FPLC system and HiLoad 16/60 Superdex 75 column. The protein was injected and eluted using an isocratic mobile phase (50 mM TEA, 150 mM NaCl, pH 7.5). All eluent collected from the purification was analysed by SDS-PAGE and that shown to contain pure

protein was combined and desalted (PD-10 column) against 50 mM sodium phosphate buffer (pH 6.5). ESI-MS of the protein was obtained (Figure 107).

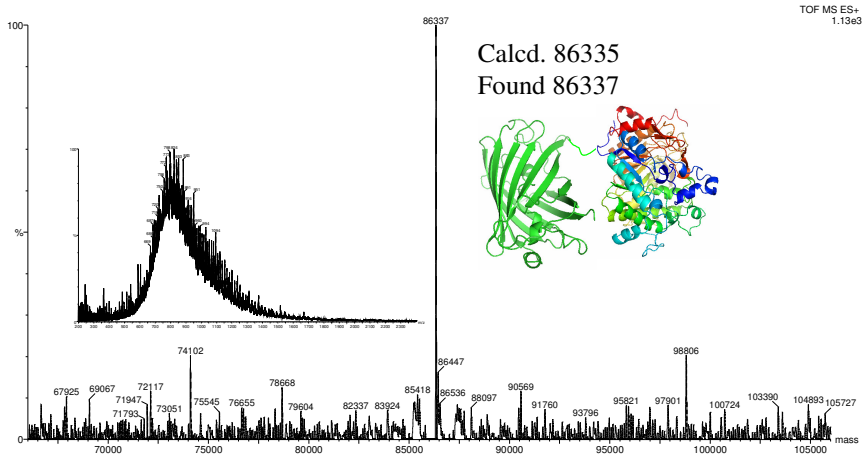



**Figure 107.** Raw (insert) and deconvoluted ESI-MS of GFP-*Ss*βG fusion.

Nucleotide sequence: GCTAGCATGAGTAAAGGAGAAGAACTTTTCACTGGAG
TTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTT
CTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTT
AAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTC
ACTACTTTCTCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATG
AAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGA
AAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAG
TCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATT
GATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTA
TAACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCA
AAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTA
GCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTA
CCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAAC
GAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGAT
TACACATGGCATGGATGAACTATACAAAGGTGGTGGTGGTGGCGGTGGTG
GTGGTGGTATGTACTCATTTCCAAATAGCTTTAGGTTTGGTTGGTCCCAGG
CCGGATTTCAATCAGAAATGGGAACACCAGGGTCAGAAGATCCAAATACT
GACTGGTATAAATGGGTTCATGATCCAGAAAACATGGCAGCGGGATTAGT
AAGTGGAGATCTACCAGAAAATGGGCCAGGCTACTGGGGAAACTATAAG
ACATTTCACGATAATGCACAAAAAATGGGATTAAAAATAGCTAGACTAAA

TGTGGAATGGTCTAGGATATTTCCTAATCCATTACCAAGGCCACAAAACTT
TGATGAATCAAAACAAGATGTGACAGAGGTTGAGATAAACGAAAACGAG
TTAAAGAGACTTGACGAGTACGCTAATAAAGACGCATTAAACCATTACAG
GGAAATATTCAAGGATCTTAAAAGTAGAGGACTTTACTTTATACTAAACA
TGTATCATTGGCCATTACCTCTATGGTTACACGACCCAATAAGAGTAAGA
AGAGGAGATTTTACTGGACCAAGTGGTTGGCTAAGTACTAGAACAGTTTA
CGAATTCGCTAGATTCTCAGCTTATATAGCTTGGAAATTCGATGATCTAGT
GGATGAGTACTCAACAATGAATGAACCTAACGTTGTTGGAGGTTTAGGAT
ACGTTGGTGTTAAGTCCGGTTTTCCCCCAGGATACCTAAGCTTTGAACTTT
CCCGTAGGGCAATGTATAACATCATTCAAGCTCACGCAAGAGCGTATGAT
GGGATAAAGAGTGTTTCTAAAAAACCAGTTGGAATTATTTACGCTAATAG
CTCATTCCAGCCGTTAACGGATAAAGATATGGAAGCGGTAGAGATGGCTG
AAAATGATAATAGATGGTGGTTCTTTGATGCTATAATAAGAGGTGAGATC
ACCAGAGGAAACGAGAAGATTGTAAGAGATGACCTAAAGGGTAGATTGG
ATTGGATTGGAGTTAATTATTACACTAGGACTGTTGTGAAGAGGACTGAA
AAGGGATACGTTAGCTTAGGAGGTTACGGTCACGGATGTGAGAGGAATTC
TGTAAGTTTAGCGGGATTACCAACCAGCGACTTCGGCTGGGAGTTCTTCCC
AGAAGGTTTATATGACGTTTTGACGAAATACTGGAATAGATATCATCTCTA
TATGTACGTTACTGAAAATGGTATTGCGGATGATGCCGATTATCAAAGGC
CCTATTATTTAGTATCTCACGTTTATCAAGTTCATAGAGCAATAAATAGTG
GTGCAGATGTTAGAGGGTATTTACATTGGTCTCTAGCTGATAATTACGAAT
GGGCTTCAGGATTCTCTATGAGGTTTGGTCTGTTAAAGGTCGATTACAACA
CTAAGAGACTATACTGGAGACCCTCAGCACTAGTATATAGGGAAATCGCC
ACAAATGGCGCAATAACTGATGAAATAGAGCACTTAAATAGCGTACCTCC
AGTAAAGCCATTAAGGCACAAGAGCTCCGTCGACAAGCTTGCGGCCGCAC
TCGAGCACCACCACCACCACCACTGAGAGCTC

Protein sequence: [M]ASMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGD
ATYGKLTLKFICTTGKLPVPWPTLVTTF{SYG}VQCFSRYPDHMKRHDFFKSA
MPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHK
LEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGP
VLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGGGGG
GGGGGMYSFPNSFRFGWSQAGFQSEMGTPGSEDPNTDWYKWVHDPENMAA
GLVSGDLPENGPGYWGNYKTFHDNAQKMGLKIARLNVEWSRIFPNPLPRPQ

NFDESKQDVTEVEINENELKRLDEYANKDALNHYREIFKDLKSRGLYFILNM
YHWPLPLWLHDPIRVRRGDFTGPSGWLSTRTVYEFARFSAYIAWKFDDLVDE
YSTMNEPNVVGGLGYVGVKSGFPPGYLSFELSRRAMYNIIQAHARAYDGIKS
VSKKPVGIIYANSSFQPLTDKDMEAVEMAENDNRWWFFDAIIRGEITRGNEKI
VRDDLKGRLDWIGVNYYTRTVVKRTEKGYVSLGGYGHGCERNSVSLAGLPT
SDFGWEFFPEGLYDVLTKYWNRYHLYMYVTENGIADDADYQRPYYLVSHV
YQVHRAINSGADVRGYLHWSLADNYEWASGFSMRFGLLKVDYNTKRLYWR
PSALVYREIATNGAITDEIEHLNSVPPVKPLRHKSSVDKLAAALEHHHHHH
[ ] - cleaved during protein expression; { } - fluorophore (formed post-translationally)

### 5.6. Np276-Hag61

B834(DE3) competent cells (Novagen) were transformed following the manufacturer's protocol with pET28a plasmid carrying the M21I Np276 gene (supplied by Dr. James Errey, University of Oxford)). A glycerol stock of *E. coli* B834(DE3) containing pET28a plasmid carrying the M21I Np276 gene, was used to inoculate SelenoMet media (kanamycin, final conc. 50 μg/mL) (200 mL) supplemented with L-methionine (final conc. 40 μg/mL). An overnight culture was grown for 16 hours at 37°C with shaking (220 rpm). Aliquots (25 mL) of the overnight culture were used to inoculate SelenoMet media (kanamycin, final conc. 50 μg/mL) (1 L x 6) supplemented with L-methionine (final conc. 40 μg/mL). The cells were incubated at 37°C with shaking (220 rpm) and the optical density ($OD_{600}$) was monitored until a value of 0.6 was obtained. The cells were then pelleted by centrifugation (8 minutes, 8000 rpm, 4°C). The pellets were washed twice with SelenoMet media (kanamycin, final conc. 50 μg/mL) (2 x 400 mL) and then transferred to prewarmed SelenoMet media (kanamycin, final conc. 50 μg/mL) (1 L x 6) supplemented with DL-Hag (final conc. 80 μg/mL). The culture was incubated for 30 minutes at 37°C, followed by 30 minutes at 30°C. Expression was induced by

addition of IPTG (final conc. 1 mM) and was continued for 16 hours at 30°C with shaking (220 rpm). The cells were harvested by centrifugation (8 minutes, 8000 rpm, 4°C) to give pellets, which were collected and stored at -20°C in binding buffer (20 mM Tris-HCl, 15 mM imidazole, 0.3 M NaCl, pH 7.8) (30 mL). The cell pellet was thawed; DNase (10 μL) and lysozyme (10 mg) were added and the cell suspension was incubated for 1 hour on ice before being sonicated (20% power in bursts of 30 seconds with a rest time of 1 minute between each burst). The cell debris was removed from the lysis mixture by centrifugation (30 minutes, 20,000 rpm). The protein was purified by incubation of the cleared cell lysate with Ni-NTA agarose resin (10 mL) overnight. The resin was washed with binding buffer (20 mM Tris-HCl, 15 mM imidazole, 0.3 M NaCl, pH 7.8) (150 mL) and the protein was eluted from the resin using elution buffer (20 mM Tris-HCl, 500 mM imidazole, 0.3 M NaCl, pH 7.8) (150 mL). All eluent was analysed by SDS-PAGE and that shown to contain pure protein was combined. ESI-MS of the protein was obtained (Figure 22): calcd. 20818 (with N-terminal gluconoylation 20996),[294] found 20818 (20996).

Protein sequence: [M]GSSHHHHHHSSGLVPRGSHIDVGKLRQLYAAGERDFSI
VDLRGAVLENINLSGAILHGA[Hag]LDEANLQQANLSRADLSGATLNGADLR
GANLSKADLSDAILDNAILEGAILDEAVLNQANLKAANLEQAILSHANIREAD
LSEANLEAADLSGADLAIADLHQANLHQAALERANLTGANLEDANLEGTILE
GGNNNLAT

[ ] - cleaved during protein expression

### 5.7. β-Glucosidase from *Pyrococcus furiosus* (*Pf*βG)

β-Glucosidase from *Pyrococcus furiosus* (*Pf*βG) was kindly donated by Dr. Servé Kengen (Wageningen University). ESI-MS of the protein was obtained Figure 108.

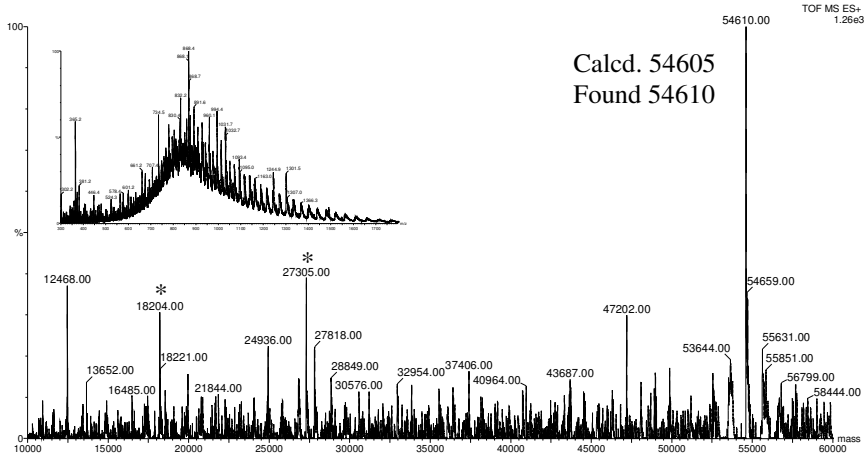



**Figure 108.** Raw (insert) and deconvoluted ESI-MS of *Pf*βG. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

Protein sequence: [M]AKFPKNFMFGYSWSGFQFEMGLPGSEVESDWWVWV HDKENIASGLVSGDLPENGPAYWHLYKQDHDIAEKLGMDCIRGGIEWARIFP KPTFDVKVDVEKDEEGNIISVDVPESTIKELEKIANMEALEHYRKIYSDWKER GKTFILNLYHWPLPLWIHDPIAVRKLGPDRAPAGWLDEKTVVEFVKFAAFVA YHLDDLVDMWSTMNEPNVVYNQGYINLRSGFPPGYLSFEAAEKAKFNLIQA HIGAYDAIKEYSEKSVGVIYAFAWHDPLAEEYKDEVEEIRKKDYEFVTILHSK GKLDWIGVNYYSRLVYGAKDGHLVPLPGYGFMSERGGFAKSGRPASDFGWE MYPEGLENLLKYLNNAYELPMIITENGMADAADRYRPHYLVSHLKAVYNAM KEGADVRGYLHWSLTDNYEWAQGFRMRFGLVYVDFETKKRYLRPSALVFR EIATQKEIPEELAHLADLKFVTRK

[ ] - cleaved during protein expression

## 5.8. Qβ virus-like particles

B834(DE3) competent cells (Novagen) were transformed following the manufacturer's protocol with p75M plasmid carrying the K16M Qβ gene. Then unnatural amino acid incorporation of DL-Hag into K16M Qβ was carried out by Dr. Julia Koeppe to generate Qβ-Hag16. Qβ-WT was kindly provided by Dr. Julia Koeppe.

Preparation of Qβ samples for MS analysis: Virus-like particles (VLP) samples (50 µL, 0.5 mg/mL) were mixed with 8 M urea (75 µL) and 1 M DTT (7.5 µL) and incubated at 37ºC for 1 hour prior to analysis by ESI-MS. In some instances the samples were further incubated with 1 M iodoacetamide (7.5 µL) at 37ºC for up to 16 hours to cap any free cysteines prior to analysis by ESI-MS. ESI-MS of the Qβ proteins were obtained: Qβ-$(Hag16)_{180}$ calcd. 14105 (monomer), found 14105 (Figure 23). See Figure 109 for Qβ-WT.

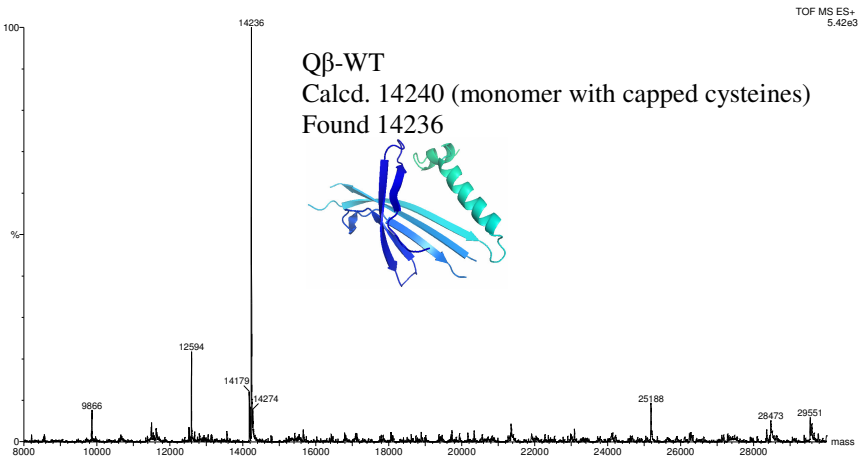



**Figure 109.** Raw (insert) and deconvoluted ESI-MS of Qβ-WT.

Protein sequence for Qβ-WT: [M]AKLETVTLGNIGKDGKQTLVLNPRGVNPTN
GVASLSQAGAVPALEKRVTVSVSQPSRNRKNYKVQVKIQNPTACTANGSCDP
SVTRQAYADVTFSFTQYSTDEERAFVRTELAALLASPLLIDAIDQLNPAY
[ ] - cleaved during protein expression

## 5.9. β-Glucosidase from *Rhizobium etli* (*Re*βG)

β-Glucosidase from *Rhizobium etli* (*Re*βG) was purchased from Prozomix. The protein was buffer exchanged (PD-10 column) into 50 mM sodium phosphate buffer

(pH 6.5). Enzymatic activity of the protein was demonstrated by reaction with X-Gal. ESI-MS of the protein was obtained (Figure 110).

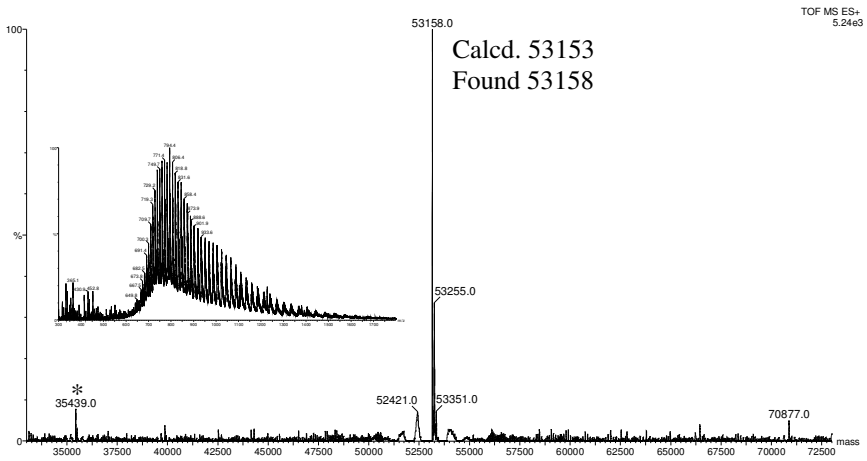



**Figure 110.** Raw (insert) and deconvoluted ESI-MS of *Re*βG. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

Protein sequence: [M]GSSHHHHHHSSGLVPRGSHMIDANTLAGRFPGDFTFGV ATAAFQIEGASKADGRKPSIWDAFCNMPGRVHNRDNGDVACDHYNRLEQDL DLVKEMGVEAYRFSIAWPRIIPDGTGPVNEAGLDFYDRLVDGCKARGIKTFA TLYHWDLPLLLAGDGGWTARSTAYAFQRYAKTVMSRLGDRLDAVATFNEP WCIVWLSHLYGIHAPGERNMQAALHAMHYMNLAHGLGVEAIRSEAPNVPV GLVLNAASIIASSTSPADLAAAERAHQFHNGAFFDPVFKGEYPKAFVQALGDR MPVIEDGDLKVISQKLDWWGLNYYTPERVTDDADRKGDFPWTVKAPPASEV KTDIGWEIYAPGLKLLVEDLYRRYELPDCYITENGACDNTGVADGEVDDTMR LDYLGDHLDVLADLIKDGYPLRGYFAWSLMDNFEWAEGYRMRFGLVHVDY ETQRRTVKKSGKWYRDLAGQFPKGNHKPG

[ ] - cleaved during protein expression

### 5.10. Myrosinase from *Sinapis alba* (*Sa*M)

Myrosinase from *Sinapis alba* (*Sa*M) was purchased from Sigma-Aldrich. The protein (54 mg) was dissolved in 50 mM sodium phosphate buffer (150 µL). Enzymatic activity of the protein was demonstrated by reaction with X-Gal.

### 5.11. β-Glucosidase from *Streptococcus pyogenes* (*Sp*βG)

β-Glucosidase from *Streptococcus pyogenes* (*Sp*βG) was purchased from Prozomix. ESI-MS of the protein was obtained (Figure 111).

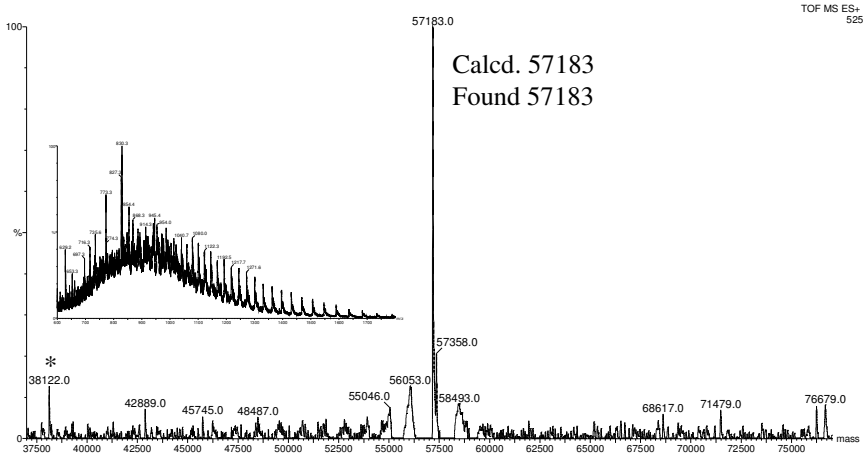


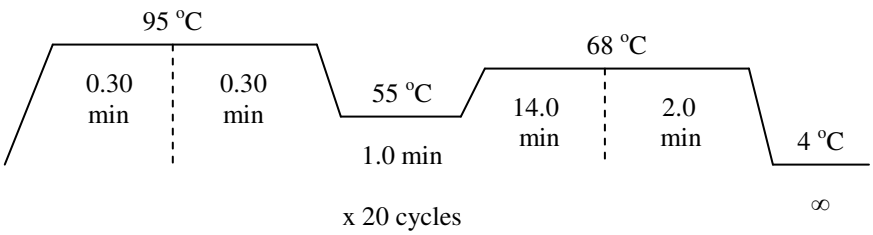

**Figure 111.** Raw (insert) and deconvoluted ESI-MS of *Sp*βG. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

Protein sequence: [M]GSSHHHHHHSSGLVPRGSHMLAFPKEFWWGGATSGPQ SEGRFAKQHRNLFDYWYEEEPDLFYDYVGPDTASDAYHQIESDLALLASLGH NSYRTSIQWTRLIDDFEQATINPDGLAYYNRVINACLANGIRPVINLHHFDLPI ALYQAYGGWESKHVVDIFVAFSKVCFEQFGDRVKDWFVHNEPMVVVEGSY LMQFHYPAIVDGKKAVQVAYNLALATAKVIQAYRRGPAELSNGRIGTILNLT PAYPASQSEADMAAAHFTELWNNDLFMEAAVHGKFPEELVAVLKKDGVLW QSTPEELALIAENRVDYLGLNFYHPKRVKAPDVIPVISPSWSPEWYYDPYLMP GRRMNVDKGWEIYPEAVYDIAIKMRDHYDNIPWFLSENGVGISGEERYRDET

GQIQDDYRIQFLKEHLTYLHKGIEAGSNCFGYHVWTPIDGWSWLNAYKNRY
GLVENNIHTQVRRPKASAYWFKKVATHNRLISLEVMEEFGGSASHLSD
[ ] - cleaved during protein expression

### 5.12. β-Glycosidase from *Sulfolobus solfataricus* (*Ss*βG)

**Site-Directed Mutagenesis -** LB media (kanamycin, final conc. 50 μg/mL) (15 mL) was inoculated from a glycerol stock of *E. coli* BL21(DE3) containing pET24d plasmid carrying the WT *Ss*βG or M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG gene (both supplied by Dr. Holger Kramer, University of Oxford). The culture was incubated for 16 hours at 37°C with shaking (220 rpm). Plasmid DNA was purified (Qiagen Spin Miniprep Kit) and the concentration determined using a NanoDrop ND-1000 spectrophotometer. Mutated plasmid DNA was formed using a Stratagene QuikChange II Site-Directed Mutagenesis Kit following the manufacturer's protocol: To a polymerase chain reaction (PCR) tube were added 10x reaction buffer (100 mM KCl, 100 mM $(NH_4)_2SO_4$, 200 mM Tris-HCl (pH 8), 20 mM $MgCl_2$, 1% Triton X-100, 100 μg/ml nuclease-free BSA) (5 μL), dsDNA template (150 ng), forward primer (125 ng), reverse primer (125 ng) and dNTP mix (1 μL). The reaction mixture was made up to 50 μL with $H_2O$, and PfuTurbo DNA polymerase (1 μL) or PfuUltra HF DNA polymerase (1 μL) was added. The reaction mixture was placed in an Applied Biosystems Geneamp PCR system 9700 thermocycler on the following program:



Then, Dpn I restriction enzyme (1 μL) was added. After gentle and thorough mixing, the reaction was centrifuged (1 minute, 13000 rpm, 4°C) and then incubated for 1 hour at 37°C.

Primers for site-directed mutagenesis were generated using the website http://bioinformatics.org/primerx/cgi-bin/protein_1.cgi and were ordered from Sigma Genosys. The primers used to introduce the desired mutations are listed below (F = forward primer, R = reverse primer):

Q18A *Ss*βG

F 5' GTCCCAGGCCGGATTTGCATCAGAAATGGGAACAC 3'

R 5' GTGTTCCCATTTCTGATGCAAATCCGGCCTGGGAC 3'

R79A *Ss*βG

F 5' GGGATTAAAAATAGCTGCGCTAAATGTGGAATGGTC 3'

R 5' GACCATTCCACATTTAGCGCAGCTATTTTTAATCCC 3'

N81A *Ss*βG

F 5' GGGATTAAAAATAGCTAGACTAGCCGTGGAATGGTCTAGGATATT TC 3'

R 5' GAAATATCCTAGACCATTCCACGGCTAGTCTAGCTATTTTTAATCC C 3'

Y149A *Ss*βG

F 5' CTTTACTTTATACTAAACATGGCGCATTGGCCATTACCTCTATGG 3'

R 5' CCATAGAGGTAATGGCCAATGCGCCATGTTTAGTATAAAGTAAAG 3'

Y149F *Ss*βG

F 5' CTTTATACTAAACATGTTTCATTGGCCATTACCTC 3'

R 5' GAGGTAATGGCCAATGAAACATGTTTAGTATAAAG 3'

H150A *Ss*βG

F 5' CTTTATACTAAACATGTATGCGTGGCCATTACCTCTATGG 3'

R 5' CCATAGAGGTAATGGCCACGCATACATGTTTAGTATAAAG 3'

H150F *Ss*βG

F 5' CTTTATACTAAACATGTATTTTTGGCCATTACCTCTATGG 3'

R 5' CCATAGAGGTAATGGCCAAAAATACATGTTTAGTATAA 3'

W151A *Ss*βG

F 5' CTAAACATGTATCATGCGCCATTACCTCTATGG 3'

R 5' CCATAGAGGTAATGGCGCATGATACATGTTTAG 3'

W151F *Ss*βG

F 5' CTTTATACTAAACATGTATCATTTTCCATTACCTCTATGGTTACCGA C 3'

R 5' GTCGTGTAACCATAGAGGTAATGGAAAATGATACATGTTAGATAG 3'

M21I M73I M148I W151F M206I M236I M275I M280I C344S M383I M439I *Ss*βG

F 5' CTTTATACTAAACATCTATCATTTTCCATTACCTCTATGGTTACAC GAC 3'

R 5' GTCGTGTAACCATAGAGGTAATGGAAAATGATAGATGTTTAGTATA AAG 3'

P152A *Ss*βG

F 5' CTAAACATGTATCATTGGGCGTTACCTCTATGGTTACAC 3'

R 5' GTGTAACCATAGAGGTAACGCCCAATGATACATGTTTAG 3'

M21I M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG

F 5' CTAAACATCTATCATTGGGCGTTACCTCTATGGTTACACG 3'

R 5' CGTGTAACCATAGAGGTAACGCCCAATGATAGATGTTTAG 3'

L153A *Ss*βG

F 5' CTAAACATGTATCATTGGCCAGCGCCTCTATGGTTACACGACCC 3'

R 5' GGGTCGTGTAACCATAGAGGCGCTGGCCAATGATACATGTTTAG 3'

N205A *Ss*βG

F 5' GATGAGTACTCAACAATGGCGGAACCTAACGTTGTTGGAG 3'

R 5' CTCCAACAACGTTAGGTTCCGCCATTGTTGAGTACTCATC 3'

V210A *Ss*βG

F 5' GAATGAACCTAACGTTGCGGGAGGTTTAGGATACG 3'

R 5' CGTATCCTAAACCTCCCGCAACGTTAGGTTCATTC 3'

F222A *Ss*βG

F 5' GTTGGTGTTAAGTCCGGTGCGCCCCCAGGATACCTAAG 3'

R 5' CTTAGGTATCCTGGGGGCGCACCGGACTTAACACCAAC 3'

S426P *Ss*βG

F 5' GTTAGAGGGTATTTACATTGGCCGCTAGCTGATAATTACGAATG 3'

R 5' CATTCGTAATTATCAGCTAGCGGCCAATGTAAATACCCTCTAAC 3'

W433A *Ss*βG

F 5' CTAGCTGATAATTACGAAGCGGCTTCAGGATTCTCTATG 3'

R 5' CATAGAGAATCCTGAAGCCGCTTCGTAATTATCAGCTAG 3'

XL10-Gold ultracompetent cells (Novagen) or NovaBlue GigaSingles competent cells (Novagen) were transformed with plasmid DNA following the manufacturer's protocol. A single colony was used to inoculate LB media (kanamycin, final conc. 50 μg/mL) (15 mL). The culture was grown for 16 hours at 37$^{o}$C with shaking (220 rpm)**.** Plasmid DNA was purified (Qiagen QIAprep Spin Miniprep Kit) and the concentration determined (NanoDrop ND-1000). The plasmid DNA was sequenced, and then transformed into BL21(DE3) competent cells (Novagen) following the manufacturer's protocol. A single colony was used to inoculate LB media (kanamycin, final conc. 50 μg/mL) (15 mL). The culture was grown for 16 hours at 37$^{o}$C with shaking (220 rpm), after which time a glycerol stock was prepared**.**

**Protein Expression** - An LB agar plate (kanamycin, final conc. 50 μg/mL) was streaked with a glycerol stock of *E. coli* BL21(DE3) containing pET24d plasmid carrying the desired *Ss*βG gene. The bacterial plate culture was incubated for 16 hours at 37$^{o}$C without shaking. A single colony was used to inoculate LB media (kanamycin,

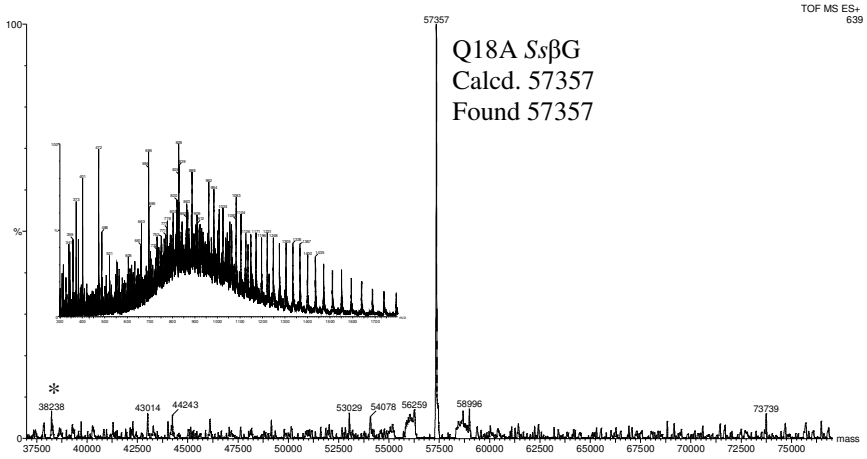
final conc. 50 μg/mL) (250 mL). The culture was grown for 16 hours at 37°C with shaking (220 rpm). LB media (kanamycin, final conc. 50 μg/mL) (1 to 12 L), prewarmed to 37°C, was inoculated with overnight culture (20 mL). The cultures were grown at 37°C with shaking (220 rpm) to an optical density ($OD_{600}$) of 0.6-0.8. Expression was induced by addition of IPTG (final conc. 1 mM) and was continued for 6 hours at 37°C. The cells were harvested by centrifugation (8 minutes, 8000 rpm, 4°C) to give pellets, which were collected and stored at -20°C in binding buffer (20 mM Tris-HCl, 5 mM imidazole, 0.5 M NaCl, pH 7.8) (35 mL).

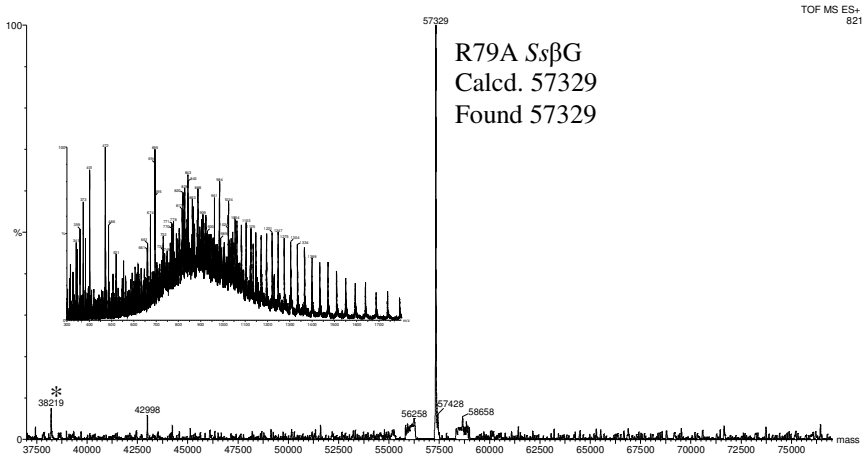
**Unnatural amino acid incorporation** - B834(DE3) competent cells (Novagen) were transformed following the manufacturer's protocol with pET28d plasmid carrying the desired *Ss*βG gene (M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I or M21I M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG). A single colony was used to inoculate LB media (kanamycin, final conc. 50 μg/mL) (15 mL) supplemented with L-methionine (final conc. 40 μg/mL). The culture was grown for 16 hours at 37°C with shaking (220 rpm), after which time a glycerol stock was prepared. The glycerol stock of *E. coli* B834(DE3) containing pET28d plasmid carrying the desired *Ss*βG gene was used to inoculate SelenoMet media (kanamycin, final conc. 50 μg/mL) (200 mL) supplemented with L-methionine (final conc. 40 μg/mL). The culture was grown for 16 hours at 37°C with shaking (220 rpm). SelenoMet media (kanamycin, final conc. 50 μg/mL) (1 L x 6) supplemented with L-methionine (final conc. 40 μg/mL), prewarmed to 37°C, was inoculated with overnight culture (25 mL). The culture was grown at 37°C with shaking (220 rpm) to an optical density ($OD_{600}$) of 1.2. The medium shift was performed as follows: The cells were pelleted by centrifugation (8 minutes, 8000 rpm, 4°C), washed twice with SelenoMet media (kanamycin, final conc. 50 μg/mL) (2 x 400 mL) and then

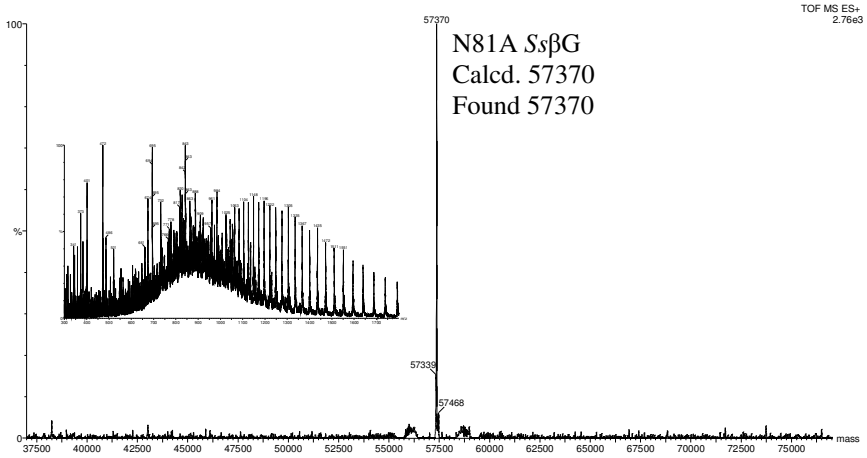
transferred to prewarmed SelenoMet media (kanamycin, final conc. 50 μg/mL) (1 L x 6) supplemented with DL- or L-Hag (final conc. 80 or 40 μg/mL, respectively). The culture was incubated for 30 minutes at 37°C, followed by 30 minutes at 30°C. Expression was induced by addition of IPTG (final conc. 1 mM) and was continued for 16 hours at 30°C. The cells were harvested by centrifugation (8 minutes, 8000 rpm, 4°C) to give pellets, which were collected and stored at -20°C in binding buffer (20 mM Tris-HCl, 5 mM imidazole, 0.5 M NaCl, pH 7.8) (35 mL).

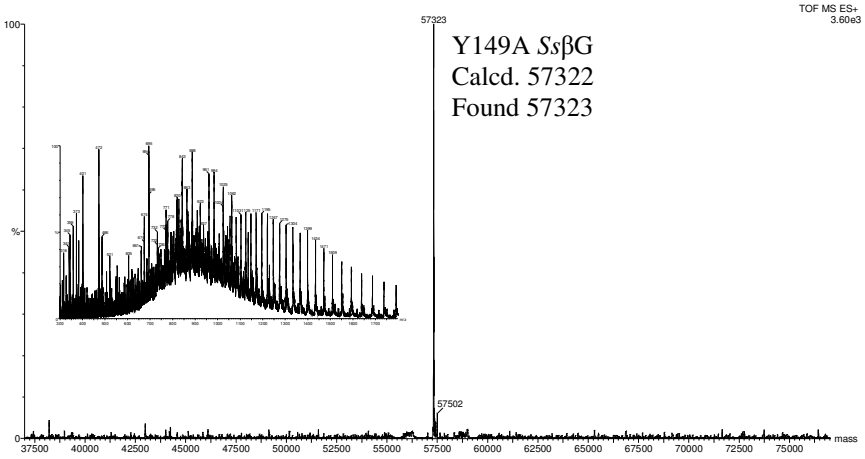
**Protein Purification** -The cell pellet was thawed; DNase (10 μL) and lysozyme (10 mg) were added and the cell suspension was incubated for 1 hour on ice before being sonicated (15% power in bursts of 20 seconds with a rest time of 40 seconds between each burst). The cell debris was removed from the lysis mixture by centrifugation (20 minutes, 18000 rpm). The protein was purified from the cleared cell lysate using an AKTA FPLC system and a HisTrap FF 1 or 5 mL column, or Ni-NTA Agarose resin. Protein was injected and eluted using a gradient from 100% binding buffer to 100% elution buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, pH 7.8) (Figure 99). In addition, P152A *Ss*βG and in some instances WT *Ss*βG for *e.g.* protein crystalisation, were also purified using an AKTA FPLC system and Phenyl Sepharose FF 75 mL column. The protein was diluted (4x) with buffer A (50 mM TEA, 2 M ammonium sulfate, pH 7.5) and was injected and eluted using a gradient system from 75% buffer A to 100% buffer B (50 mM TEA, pH 7.5) (Figure 102). All eluent from purification was analysed by SDS-PAGE and that shown to contain pure protein was combined and dialysed or desalted (PD-10 column) against 50 mM sodium phosphate buffer (pH 6.5). The enzymatic activity of each protein was demonstrated by reaction with X-Gal. ESI-MS of the *Ss*βG proteins were obtained (Figure 16 for M21I M43Hag M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG,

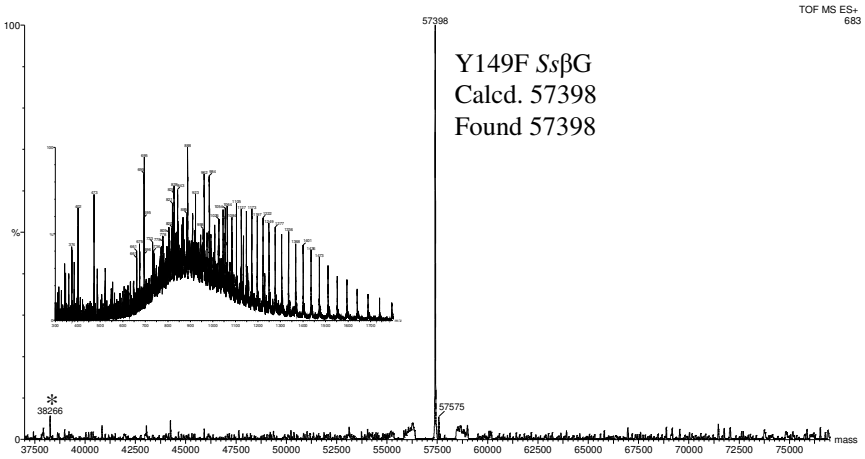
Figure 18 for M21I M43Hag M73I M148I P152A M206I M236I M275I M280I C344S M383I M439I *Ss*βG, Figure 38 for WT *Ss*βG, and Figure 112 for all other *Ss*βG proteins).

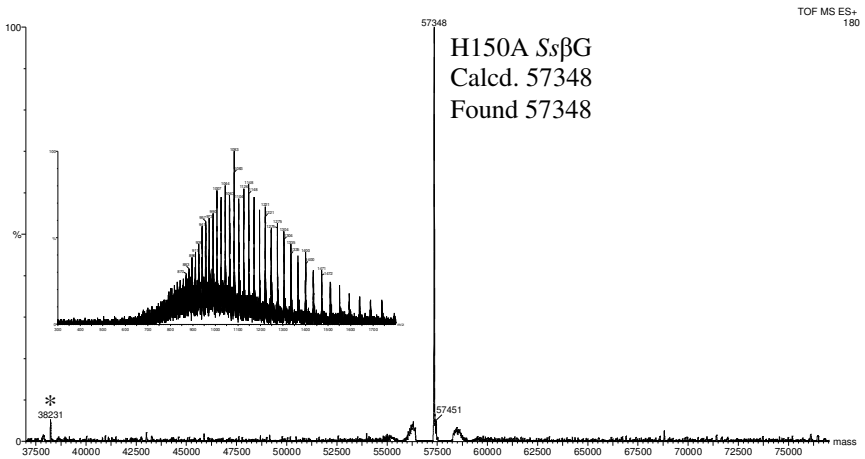

TOF MS ES+
639
57357
Q18A *Ss*βG
Calcd. 57357
Found 57357
100
%
0
38238
43014
44243
53029
54078
56259
58996
73739
mass
37500 40000 42500 45000 47500 50000 52500 55000 57500 60000 62500 65000 67500 70000 72500 75000

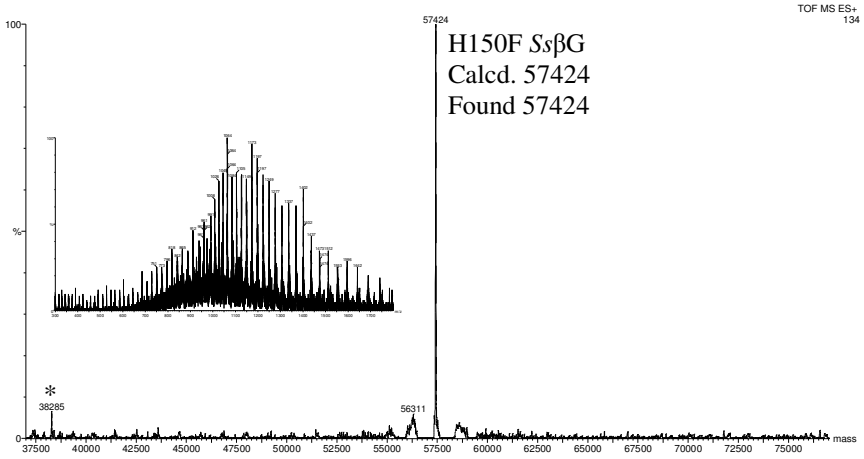

TOF MS ES+
821
57329
R79A *Ss*βG
Calcd. 57329
Found 57329
100
%
0
38219
42998
56258
57428
58658
mass
37500 40000 42500 45000 47500 50000 52500 55000 57500 60000 62500 65000 67500 70000 72500 75000

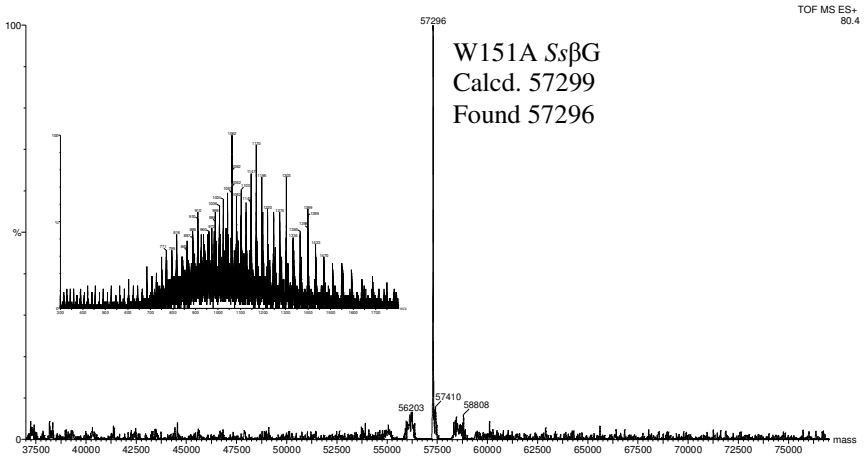

TOF MS ES+
2.76e3
57370
N81A *Ss*βG
Calcd. 57370
Found 57370
100
%
0
57339
57468
mass
37500 40000 42500 45000 47500 50000 52500 55000 57500 60000 62500 65000 67500 70000 72500 75000

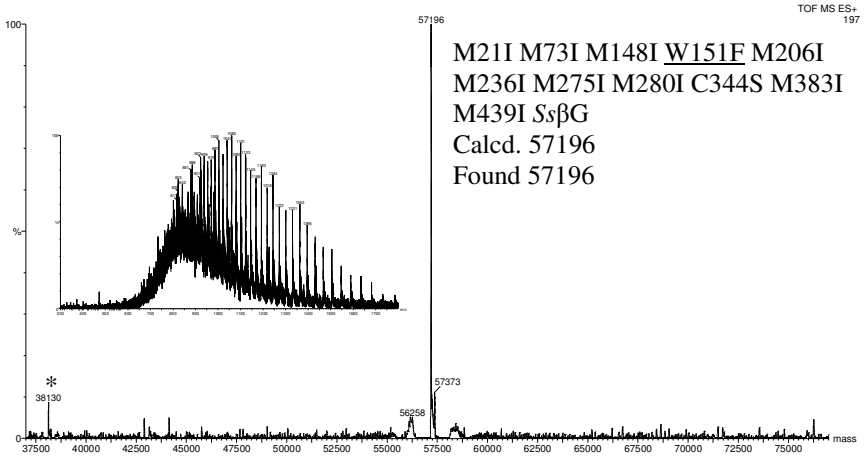

TOF MS ES+
3.60e3
57323
Y149A *Ss*βG
Calcd. 57322
Found 57323
57502
mass

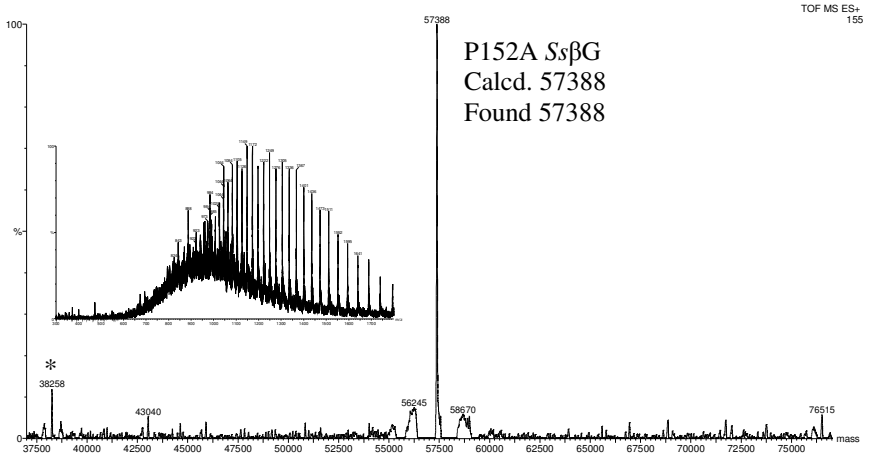

TOF MS ES+
683
57398
Y149F *Ss*βG
Calcd. 57398
Found 57398
38266
57575
mass

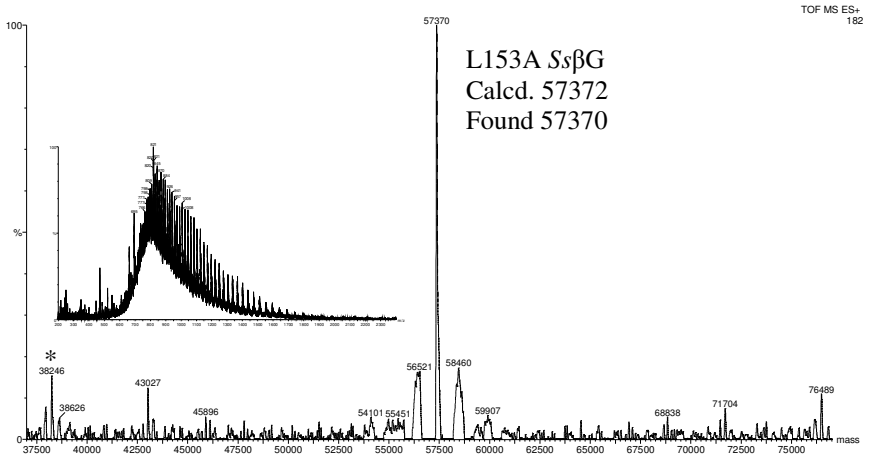

TOF MS ES+
180
57348
H150A *Ss*βG
Calcd. 57348
Found 57348
38231
57451
mass

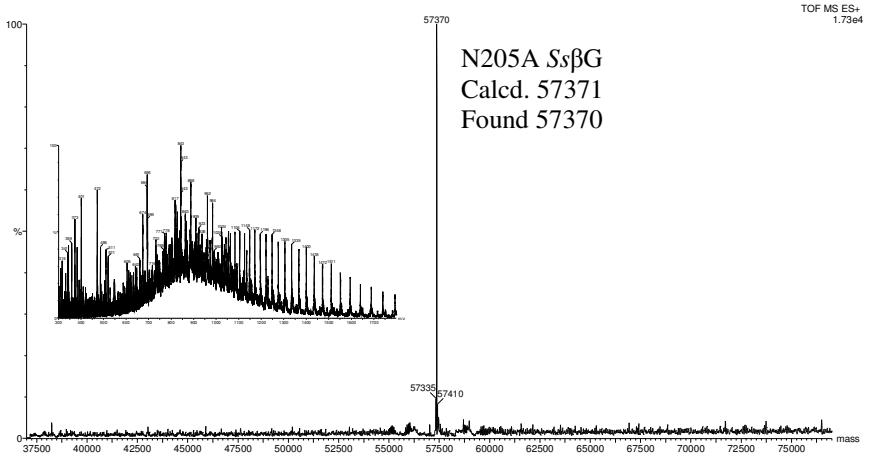

TOF MS ES+
134
57424
H150F *Ss*βG
Calcd. 57424
Found 57424
38285
56311
mass

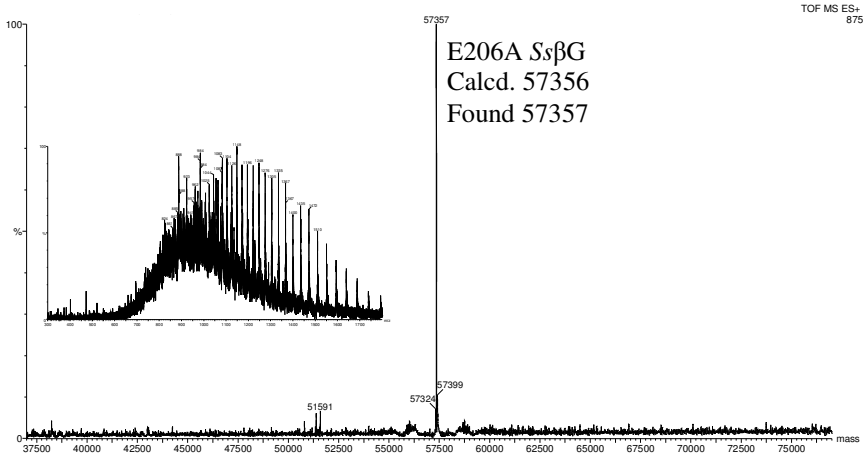

TOF MS ES+
80.4
57296
W151A *Ss*βG
Calcd. 57299
Found 57296
56203
57410
58808
mass

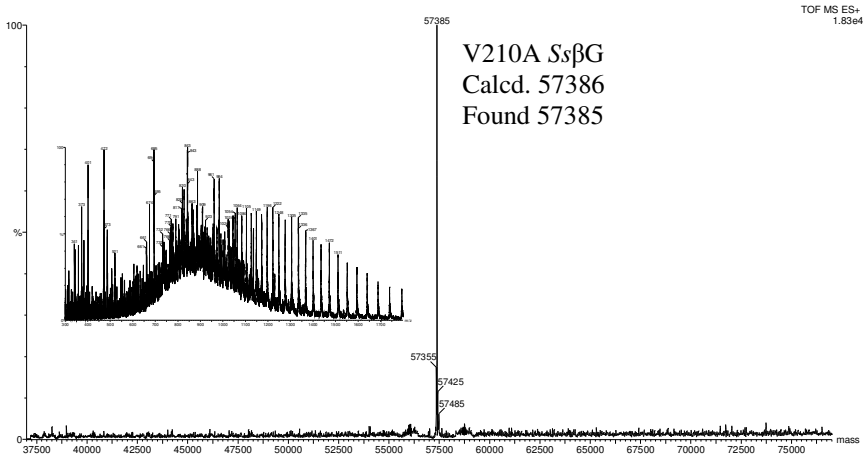

TOF MS ES+
197
57196
M21I M73I M148I W151F M206I
M236I M275I M280I C344S M383I
M439I *Ss*βG
Calcd. 57196
Found 57196
38130
57373
mass

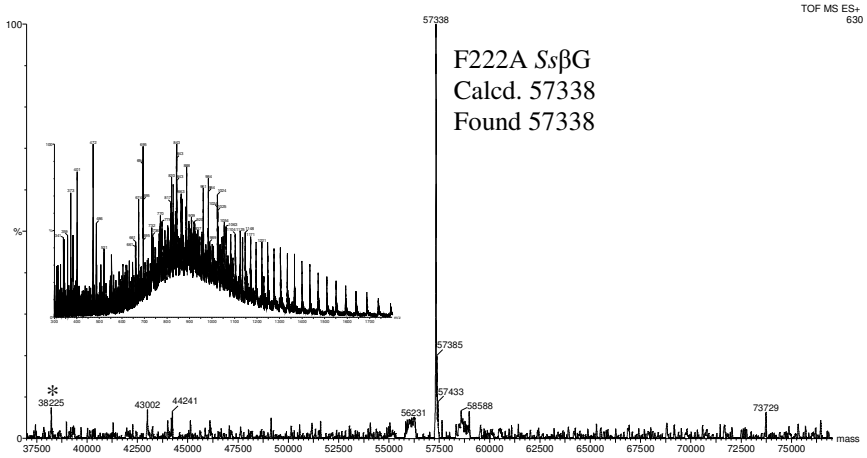

TOF MS ES+
155
100
%
0
57388
P152A *Ss*βG
Calcd. 57388
Found 57388
*
38258
43040
56245
58670
76515
37500 40000 42500 45000 47500 50000 52500 55000 57500 60000 62500 65000 67500 70000 72500 75000
mass

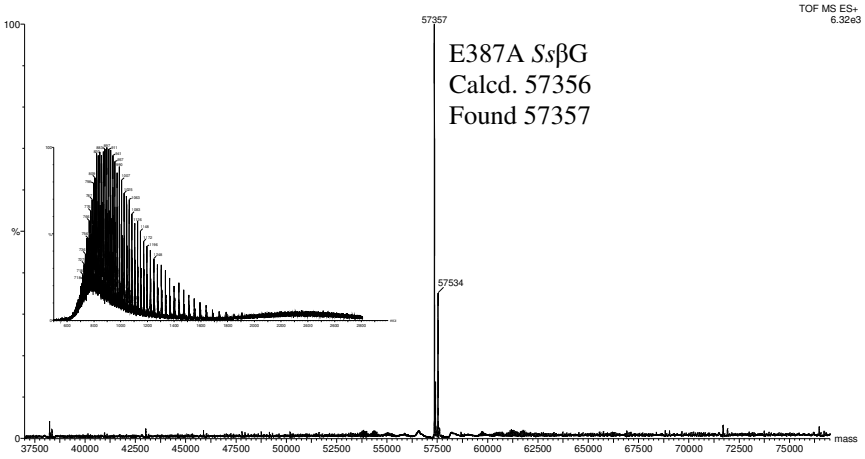

TOF MS ES+
182
100
%
0
57370
L153A *Ss*βG
Calcd. 57372
Found 57370
*
38246
38626
43027
45896
54101
55451
56521
58460
59907
68838
71704
76489
37500 40000 42500 45000 47500 50000 52500 55000 57500 60000 62500 65000 67500 70000 72500 75000
mass

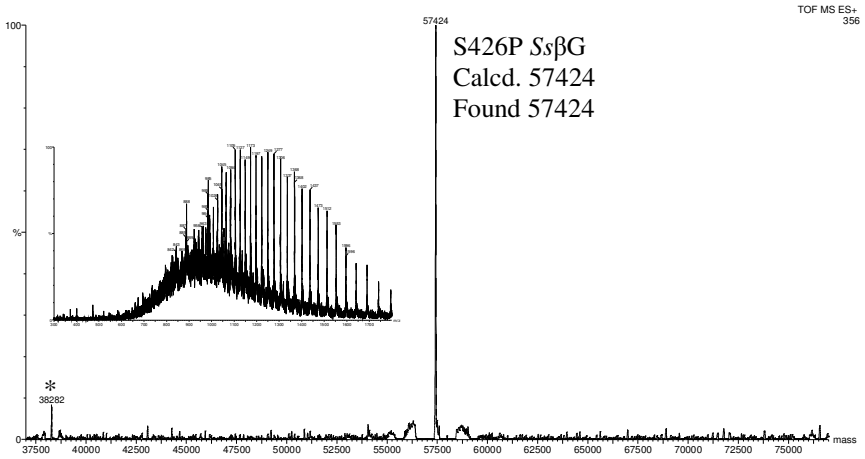

TOF MS ES+
1.73e4
100
%
0
57370
N205A *Ss*βG
Calcd. 57371
Found 57370
57335
57410
37500 40000 42500 45000 47500 50000 52500 55000 57500 60000 62500 65000 67500 70000 72500 75000
mass

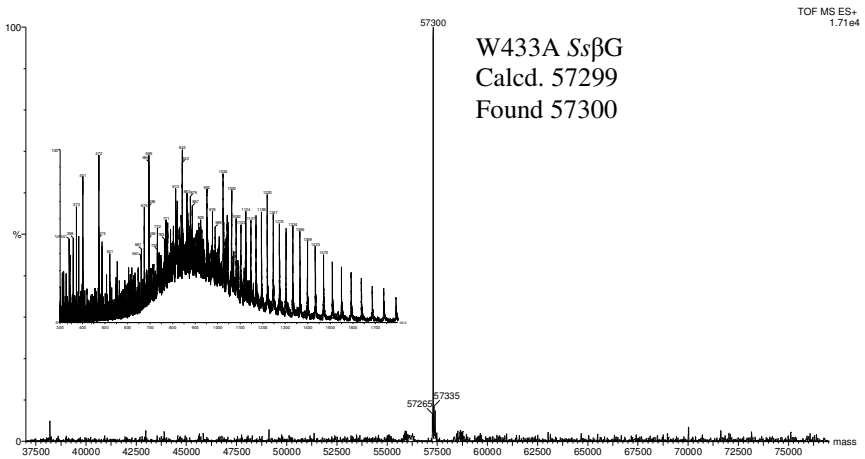

TOF MS ES+
875
E206A *Ss*βG
Calcd. 57356
Found 57357
57357
57399
57324
51591
mass


TOF MS ES+
1.83e4
V210A *Ss*βG
Calcd. 57386
Found 57385
57385
57355
57425
57485
mass


TOF MS ES+
630
F222A *Ss*βG
Calcd. 57338
Found 57338
57338
57385
57433
38225
43002
44241
56231
58588
73729
mass


TOF MS ES+
6.32e3
57357
E387A *Ss*βG
Calcd. 57356
Found 57357
57534
100
%
0
37500 40000 42500 45000 47500 50000 52500 55000 57500 60000 62500 65000 67500 70000 72500 75000
mass


TOF MS ES+
356
57424
S426P *Ss*βG
Calcd. 57424
Found 57424
*
38282
100
%
0
37500 40000 42500 45000 47500 50000 52500 55000 57500 60000 62500 65000 67500 70000 72500 75000
mass


TOF MS ES+
1.71e4
57300
W433A *Ss*βG
Calcd. 57299
Found 57300
57265
57335
100
%
0
37500 40000 42500 45000 47500 50000 52500 55000 57500 60000 62500 65000 67500 70000 72500 75000
mass

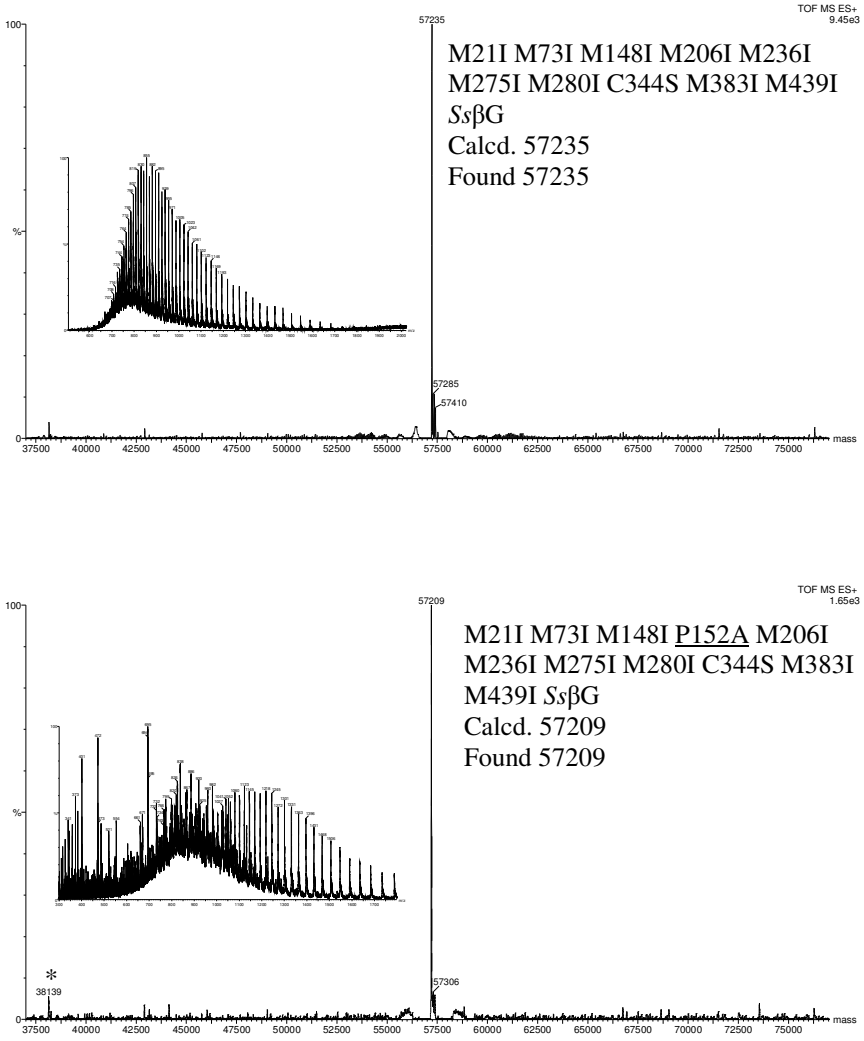



**Figure 112.** Raw (insert) and deconvoluted ESI-MS of *Ss*βG Proteins. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

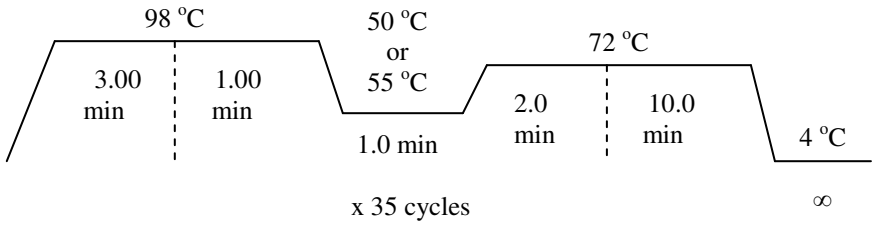
Nucleotide sequence (WT *Ss*βG without N-terminal his-tag): ATGTACTCATTTCCA
AATAGCTTTAGGTTTGGTTGGTCCCAGGCCGGATTTCAATCAGAAATGGG
AACACCAGGGTCAGAAGATCCAAATACTGACTGGTATAAATGGGTTCATG
ATCCAGAAAACATGGCAGCGGGATTAGTAAGTGGAGATCTACCAGAAAA
TGGGCCAGGCTACTGGGGAAACTATAAGACATTTCACGATAATGCACAAA
AAATGGGATTAAAAATAGCTAGACTAAATGTGGAATGGTCTAGGATATTT
CCTAATCCATTACCAAGGCCACAAAACTTTGATGAATCAAAACAAGATGT
GACAGAGGTTGAGATAAACGAAAACGAGTTAAAGAGACTTGACGAGTAC
GCTAATAAAGACGCATTAAACCATTACAGGGAAATATTCAAGGATCTTAA
AAGTAGAGGACTTTACTTTATACTAAACATGTATCATTGGCCATTACCTCT
ATGGTTACACGACCCAATAAGAGTAAGAAGAGGAGATTTTACTGGACCAA

GTGGTTGGCTAAGTACTAGAACAGTTTACGAATTCGCTAGATTCTCAGCTT
ATATAGCTTGGAAATTCGATGATCTAGTGGATGAGTACTCAACAATGAAT
GAACCTAACGTTGTTGGAGGTTTAGGATACGTTGGTGTTAAGTCCGGTTTT
CCCCCAGGATACCTAAGCTTTGAACTTTCCCGTAGGCATATGTATAACATC
ATTCAAGCTCACGCAAGAGCGTATGATGGGATAAAGAGTGTTTCTAAAAA
ACCAGTTGGAATTATTTACGCTAATAGCTCATTCCAGCCGTTAACGGATAA
AGATATGGAAGCGGTAGAGATGGCTGAAAATGATAATAGATGGTGGTTCT
TTGATGCTATAATAAGAGGTGAGATCACCAGAGGAAACGAGAAGATTGTA
AGAGATGACCTAAAGGGTAGATTGGATTGGATTGGAGTTAATTATTACAC
TAGGACTGTTGTGAAGAGGACTGAAAAGGGATACGTTAGCTTAGGAGGTT
ACGGTCACGGATGTGAGAGGAATTCTGTAAGTTTAGCGGGATTACCAACC
AGCGACTTCGGCTGGGAGTTCTTCCCAGAAGGTTTATATGACGTTTTGACG
AAATACTGGAATAGATATCATCTCTATATGTACGTTACTGAAAATGGTATT
GCGGATGATGCCGATTATCAAAGGCCCTATTATTTAGTATCTCACGTTTAT
CAAGTTCATAGAGCAATAAATAGTGGTGCAGATGTTAGAGGGTATTTACA
TTGGTCTCTAGCTGATAATTACGAATGGGCTTCAGGATTCTCTATGAGGTT
TGGTCTGTTAAAGGTCGATTACAACACTAAGAGACTATACTGGAGACCCT
CAGCACTAGTATATAGGGAAATCGCCACAAATGGCGCAATAACTGATGAA
ATAGAGCACTTAAATAGCGTACCTCCAGTAAAGCCATTAAGGCAC

Protein sequence (WT SsβG): [M]GHHHHHHHSFPNSFRFGWSQAGFQSEMGTPG
SEDPNTDWYKWVHDPENMAAGLVSGDLPENGPGYWGNYKTFHDNAQKMG
LKIARLNVEWSRIFPNPLPRPQNFDESKQDVTEVEINENELKRLDEYANKDAL
NHYREIFKDLKSRGLYFILNMYHWPLPLWLHDPIRVRRGDFTGPSGWLSTRT
VYEFARFSAYIAWKFDDLVDEYSTMNEPNVVGGLGYVGVKSGFPPGYLSFEL
SRRAMYNIIQAHARAYDGIKSVSKKPVGIIYANSSFQPLTDKDMEAVEMAEN
DNRWWFFDAIIRGEITRGNEKIVRDDLKGRLDWIGVNYYTRTVVKRTEKGYV
SLGGYGHGCERNSVSLAGLPTSDFGWEFFPEGLYDVLTKYWNRYHLYMYVT
ENGIADDADYQRPYYLVSHVYQVHRAINSGADVRGYLHWSLADNYEWASG
FSMRFGLLKVDYNTKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRH
[ ] - cleaved during protein expression

### 5.13. T0-, T50-, T100- and T140-*Ss*βG

LB media (kanamycin, final conc. 50 μg/mL) (30 mL) was inoculated from a glycerol stock of *E. coli* BL21(DE3) containing pET24d plasmid carrying the WT *Ss*βG gene. The culture was incubated for 16 hours at 37°C with shaking (220 rpm). Plasmid DNA was purified (Qiagen Spin Miniprep Kit) and the concentration determined using a NanoDrop ND-1000 spectrophotometer. T0-*Ss*βG, T50-*Ss*βG, T100-*Ss*βG and T140-*Ss*βG plasmid DNA inserts were generated: To a PCR tube were added Pfu buffer (10x) (5 μL), dsDNA template (1 μL), forward primer (10 pm/μL, 5 μL), reverse primer (10 pm/μL, 5 μL), dNTP mix (2 mM, 5 μL) and DMSO (1 μL). The reaction mixture was made up to 50 μL with $H_2O$ and *PfuTurbo* DNA polymerase (1 μL) was added. The reaction mixture was placed in an Applied Biosystems Geneamp PCR system 9700 thermocycler on the following program (Annealing temperatures: T0-*Ss*βG / T50-*Ss*βG = 50°C, T100-*Ss*βG / T140-*Ss*βG = 55°C):




Primers were generated using the website http://bioinformatics.org/primerx/cgi-bin/protein_1.cgi and were ordered from Sigma Genosys. The primers used to introduce the desired mutations are listed below (F = forward primer, R = reverse primer):

T0-*Ss*βG

F 5' TTTTTAGCTAGCATGTACTCATTTCCAAATAGC 3'

R 5' TTTTTTGAGCTCTTGTGCCTTAATGGCTTTACT G 3'

T50-*Ss*βG

F 5' TTTTTAGCTAGCATGGATCTACCAGAAAATGGGCCA 3'

R 5' TTTTTTGAGCTCTTGTGCCTTAATGGCTTTACTG 3'

T100-*Ss*βG

F 5' TTTTTAGCTAGCATGTCAAAACAAGATGTGACAGAG 3'

R 5' TTTTTTGAGCTCTTGTGCCTTAATGGCTTTACTG 3'

T140-*Ss*βG

F 5' TTTTTAGCTAGCATGGGACTTTACTTTATACTAAAC 3'

R 5' TTTTTTGAGCTCTTGTGCCTTAATGGCTTTACTG 3'

The DNA was purified from a 0.8% agarose gel (Qiagen QIAquick Gel Extraction Kit) (Figure 113) and was cut using the restriction enzymes SacI and Nhe1: To a PCR tube were added insert (30 µL), 10x NEB buffer (4 µL), BSA (10 mg/mL, 0.4 µL), SacI (1 µL), NheI (1 µL) and $H_2O$ (3.6 µL). The digest was carried out for 4 hours at 37$^{o}$C.

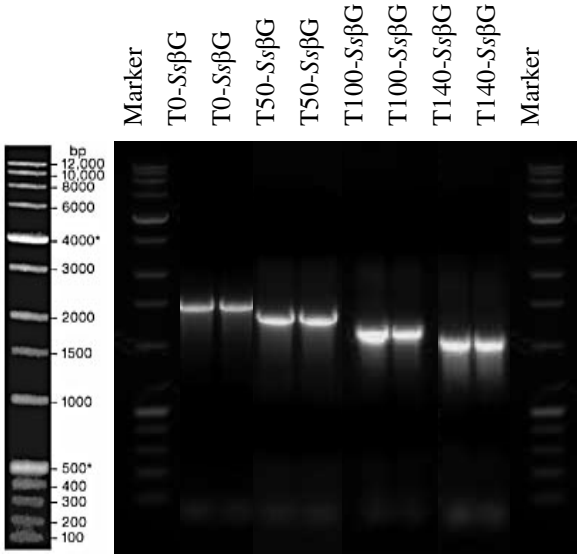



**Figure 113.** DNA gel of T0-*Ss*βG, T50-*Ss*βG, T100-*Ss*βG and T140-*Ss*βG plasmid DNA inserts.

pET-24a(+) plasmid DNA (Novagen) was cut using the restriction enzymes SacI and Nhe1: To a PCR tube were added vector (1 μg/μL, 1 μL), NEB buffer 1 (10x) (4 μL), BSA (10 mg/mL, 0.4 μL), SacI (1 μL), NheI (1 μL) and $H_2O$ (32.6 μL). The digest was carried out for 4 hours at 37°C and the resulting digested DNA was purified from a 0.8% agarose gel (Qiagen QIAquick Gel Extraction Kit). Ligation of T0-*Ss*βG, T50-*Ss*βG, T100-*Ss*βG and T140-*Ss*βG plasmid DNA inserts into pET-24a(+) was carried out at 16°C overnight: To a PCR tube were added cut pET-24a(+) (1 μL), cut insert (4 μL), 10x T4 DNA ligase buffer (Promega) (1 μL), high concentration (HC) T4 DNA Ligase (Promega) (1 μL) and $H_2O$ (3 μL). The ligation mixtures were transformed into NovaBlue GigaSingles competent cells (Novagen) following the manufacturer's protocol. A single colony from each plate was used to separately inoculate LB media (kanamycin, final conc. 50 μg/mL) (4 x 30 mL). The cultures were grown overnight at 37°C with shaking (220 rpm). Plasmid DNA was purified (Qiagen QIAprep Spin Miniprep Kit) and subjected to an analytical restriction digest to assess if the DNA insert was present: To a PCR tube were added plasmid DNA (6.5 μL), 10x NEB buffer 1 (1.5 μL), SacI (1 μL), NheI (1 μL) and $H_2O$ (5 μL). The digest was carried out for 4 hours at 37°C. The plasmid DNA was also sequenced. BL21(DE3) competent cells (Novagen) were transformed with plasmid DNA following the manufacturer's protocol. A single colony was used to inoculate LB media (kanamycin, final conc. 50 μg/mL) (100 mL). An overnight culture was grown at 37°C with shaking (220 rpm). LB media (kanamycin, final conc. 50 μg/mL) (1 L x 2), prewarmed to 37°C, was inoculated with overnight culture (25 mL). The cells were grown at 37°C with shaking (220 rpm) to an optical density ($OD_{600}$) of 0.6 - 0.8. Expression was induced by addition of IPTG (final conc. 1 mM) and was continued for 5 hours at 37°C [additionally, in a separate grow up, IPTG induction of T50-*Ss*βG,

T100-*Ss*βG and T140-*Ss*βG was continued for 16 hours at 25 $^{o}$C]. The cells were harvested by centrifugation (8 minutes, 8000 rpm, 4$^{o}$C) to give pellets, which were collected and stored at -20$^{o}$C in binding buffer (20 mM Tris-HCl, 5 mM imidazole, 0.5 M NaCl, pH 7.8) (30 mL). The cell pellets were thawed; DNase (10 μL) and lysozyme (10 mg) were added and the cell suspensions were incubated for 1 hour on ice before being sonicated (15% power in bursts of 20 seconds with a rest time of 40 seconds between each burst). The cell debris was removed from the lysis mixtures by centrifugation (20 minutes, 18000 rpm).

**Purification of T0-*Ss*βG** - T0-*Ss*βG was purified from the cleared cell lysate using an AKTA FPLC system and HisTrap FF 5 mL column. Protein was injected and eluted using a gradient from 100% binding buffer to 100% elution buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, pH 7.8) (Figure 99). All eluent collected from the purification was analysed by SDS-PAGE and that shown to contain pure protein was combined and dialysed against 50 mM sodium phosphate buffer (pH 6.5). Enzymatic activity of the protein was demonstrated by reaction with X-Gal. ESI-MS of the protein was obtained (Figure 114).

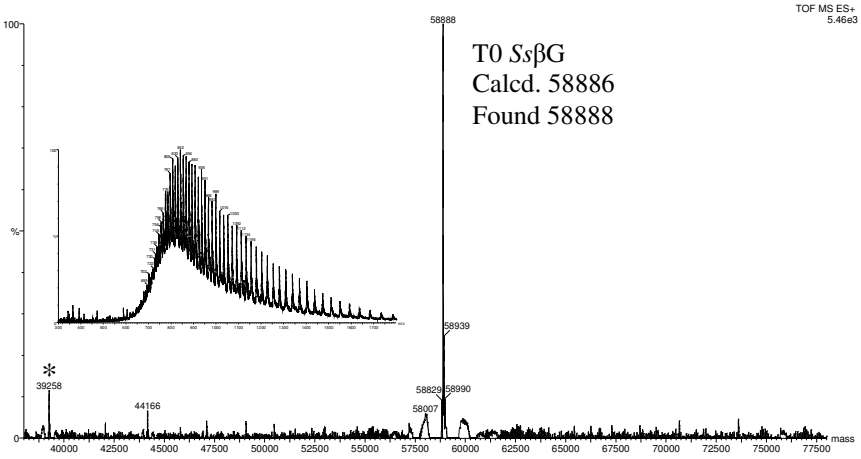



**Figure 114.** Raw (insert) and deconvoluted ESI-MS of T0-*Ss*βG. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

**Purification of T50-, T100- and T140-*Ss*βG** - Protein purification used Ni-NTA Agarose resin (5 mL), as purification of T50-, T100- and T140-*Ss*βG using an AKTA FPLC system and HisTrap FF 5 mL column were unsuccessful. The resin was added to the crude lysate and the slurry mixed gently for > 3 hours at 4°C. The resin was washed with binding buffer (20 mM Tris-HCl, 5 mM imidazole, 0.5 M NaCl, pH 7.8) (25 mL) and the protein eluted using elution buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, pH 7.8) (40 mL). All eluent collected from the purification was analysed by SDS-PAGE and the first elution fraction additionally by anti-his western blot (Figure 81). ESI-MS of the proteins were unable to be obtained: T50-*Ss*βG calcd. 53127; T100-*Ss*βG calcd. 47273; T140-*Ss*βG calcd. 42439.

Protein sequence (T0-*Ss*βG): [M]ASMYSFPNSFRFGWSQAGFQSEMGTPGSEDPN
TDWYKWVHDPENMAAGLVSGDLPENGPGYWGNYKTFHDNAQKMGLKIAR
LNVEWSRIFPNPLPRPQNFDESKQDVTEVEINENELKRLDEYANKDALNHYRE
IFKDLKSRGLYFILNMYHWPLPLWLHDPIRVRRGDFTGPSGWLSTRTVYEFAR
FSAYIAWKFDDLVDEYSTMNEPNVVGGLGYVGVKSGFPPGYLSFELSRRAM
YNIIQAHARAYDGIKSVSKKPVGIIYANSSFQPLTDKDMEAVEMAENDNRWW
FFDAIIRGEITRGNEKIVRDDLKGRLDWIGVNYYTRTVVKRTEKGYVSLGGYG
HGCERNSVSLAGLPTSDFGWEFFPEGLYDVLTKYWNRYHLYMYVTENGIAD

DADYQRPYYLVSHVYQVHRAINSGADVRGYLHWSLADNYEWASGFSMRFG
LLKVDYNTKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRHKSSVDK
LAAALEHHHHHH

[ ] - cleaved during protein expression

Protein sequence (T50-*Ss*βG): MDLPENGPGYWGNYKTFHDNAQKMGLKIARLN
VEWSRIFPNPLPRPQNFDESKQDVTEVEINENELKRLDEYANKDALNHYREIF
KDLKSRGLYFILNMYHWPLPLWLHDPIRVRRGDFTGPSGWLSTRTVYEFARF
SAYIAWKFDDLVDEYSTMNEPNVVGGLGYVGVKSGFPPGYLSFELSRRAMY
NIIQAHARAYDGIKSVSKKPVGIIYANSSFQPLTDKDMEAVEMAENDNRWWF
FDAIIRGEITRGNEKIVRDDLKGRLDWIGVNYYTRTVVKRTEKGYVSLGGYG
HGCERNSVSLAGLPTSDFGWEFFPEGLYDVLTKYWNRYHLYMYVTENGIAD
DADYQRPYYLVSHVYQVHRAINSGADVRGYLHWSLADNYEWASGFSMRFG
LLKVDYNTKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRHKSSVDK
LAAALEHHHHHH

Protein sequence (T100-*Ss*βG): MSKQDVTEVEINENELKRLDEYANKDALNHYR
EIFKDLKSRGLYFILNMYHWPLPLWLHDPIRVRRGDFTGPSGWLSTRTVYEFA
RFSAYIAWKFDDLVDEYSTMNEPNVVGGLGYVGVKSGFPPGYLSFELSRRA
MYNIIQAHARAYDGIKSVSKKPVGIIYANSSFQPLTDKDMEAVEMAENDNRW
WFFDAIIRGEITRGNEKIVRDDLKGRLDWIGVNYYTRTVVKRTEKGYVSLGG
YGHGCERNSVSLAGLPTSDFGWEFFPEGLYDVLTKYWNRYHLYMYVTENGI
ADDADYQRPYYLVSHVYQVHRAINSGADVRGYLHWSLADNYEWASGFSMR
FGLLKVDYNTKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRHKSSV
DKLAAALEHHHHHH

Protein sequence (T140-*Ss*βG): MGLYFILNMYHWPLPLWLHDPIRVRRGDFTGP
SGWLSTRTVYEFARFSAYIAWKFDDLVDEYSTMNEPNVVGGLGYVGVKSGF
PPGYLSFELSRRAMYNIIQAHARAYDGIKSVSKKPVGIIYANSSFQPLTDKDME
AVEMAENDNRWWFFDAIIRGEITRGNEKIVRDDLKGRLDWIGVNYYTRTVV
KRTEKGYVSLGGYGHGCERNSVSLAGLPTSDFGWEFFPEGLYDVLTKYWNR
YHLYMYVTENGIADDADYQRPYYLVSHVYQVHRAINSGADVRGYLHWSLA
DNYEWASGFSMRFGLLKVDYNTKRLYWRPSALVYREIATNGAITDEIEHLNS
VPPVKPLRHKSSVDKLAAALEHHHHHH

### 5.14. β-Glycosidase from *Thermosphaera aggregans* (*Ta*βG)

Plasmid carrying the *Ta*βG gene was transformed into TOP10 OneShot competent cells (Invitrogen) following the manufacturer's protocol. A single colony was used to inoculate LB media (kanamycin, final conc. 50 μg/mL) (15 mL). The culture was grown for 16 hours at 37°C with shaking (220 rpm). Plasmid DNA was purified (Qiagen QIAprep Spin Miniprep Kit) and sequenced. Rosetta 2 (DE3)pLysS competent cells (Novagen) were transformed with plasmid DNA following the manufacturer's protocol. A single colony was used to inoculate SB media [Solution A (tryptone (12 g), yeast extract (24 g), glycerol (5 mL), $H_2O$ (900 mL)) and solution B ($KH_2PO_4$ (1.7 g), $K_2HPO_4$ (15.3 g) and $H_2O$ (900 mL)) were autoclaved separately and then mixed together] (kanamycin, final conc. 50 μg/mL) (100 mL). A culture was grown for 16 hours at 37°C with shaking (220 rpm). SB media (kanamycin, final conc. 50 μg/mL) (750 mL x 6), prewarmed to 37°C with shaking (220 rpm), was inoculated with overnight culture (15 mL). The cells were grown at 37°C with shaking (220 rpm) to an optical density ($OD_{600}$) of 1.0. Expression was induced by addition of IPTG (final conc. 1 mM) and was continued for 16 hours at 37°C. The cells were harvested by centrifugation (8 minutes, 8000 rpm, 4°C) to give pellets, which were collected and stored at -20°C. The cell pellet was thawed; 50 mM PBS buffer (pH 7.5) (40 mL), DNase (10 μL), lysozyme (10 mg) and a Complete Protease Inhibitor Cocktail Tablet (Roche) were added, and the cell suspension was incubated for 1 hour on ice before being sonicated (15% power in bursts of 20 seconds with a rest time of 40 seconds between each burst). The cell debris was removed from the lysis mixture by centrifugation (20 minutes, 18000 rpm). The cleared cell lysate was incubated for 30 minutes at 75°C and then centrifuged (10 minutes, 3750 rpm, 4°C). The pellet containing the denatured proteins was discarded and the supernatant was incubated for

10 minutes at 85°C and then centrifuged (10 minutes, 3750 rpm, 4°C). The protein pellet was discarded. 1 M ammonium sulfate (25 mL) was added to the supernatant (25 mL) and the protein was purified using an AKTA FPLC system and Phenyl Sepharose FF 75 mL column. The protein was injected and eluted using a gradient system from 75% buffer A (50 mM sodium phosphate, 1 M ammonium sulfate, pH 6.8) to 100% buffer B ($H_2O$, pH 6.8) (Figure 102). The eluent was analysed by SDS-PAGE and the fractions shown to contain pure protein were combined. The protein was additionally purified using an AKTA FPLC system and HiLoad 26/60 Superdex 200 column. The protein was injected and eluted using an isocratic mobile phase (50 mM TEA, 150 mM NaCl, pH 7.5). The eluent was analysed by SDS-PAGE and that shown to contain pure protein was combined and dialysed against 50 mM sodium phosphate buffer (pH 6.5). Enzymatic activity of the protein was demonstrated by reaction with X-Gal. ESI-MS of the protein was obtained (Figure 115).

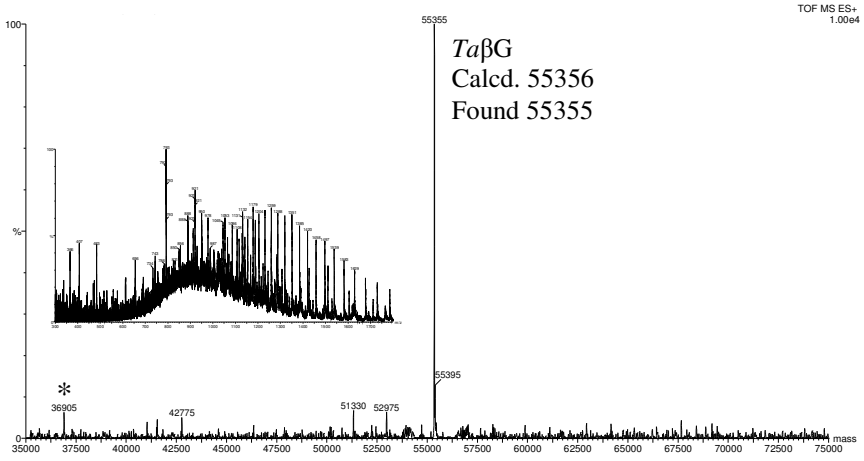



**Figure 115.** Raw (insert) and deconvoluted ESI-MS of *Ta*βG. * = Artifact of the MaxEnt1 deconvolution algorithm.[283-285]

Protein sequence: [M]KFPKDFMIGYSSSPFQFEAGIPGSEDPNSDWWVWVHDPE NTAAGLVSGDLPENGPGYWNLYKNDHDLAEKLGVNTIRVGVEWSRIFPKPTF NVKVPVERDENGSIVHVDVDDKAVERLDELANKEAVNHYVEMYKDWVERG

RKLILNLYHWPLPLWLHNPIMVRRMGPDRAPSGWLNEESVVEFAKYAAYIA
WKMGELPVMWSTMNEPNVVYEQGYMFVKGGFPPGYLSFEAADKARRNMIQ
AHARAYDNIKRFSKKPVGLIYAFQWFELLEGPAEVFDKFKSSKLYYFTDIVSK
GSSIINAEYRRDLANRLDWLGVNYYSRLVYKIVDDKPIILHGYGFLCTPGGISP
AENPCSDFGWEVYPEGLYLLLKELYNRYGVDLIVTENGVSDSRDALRPAYLV
SHVYSVWKAVNEGIPVKGYLHWSLTDNYEWAQGFRQKFGLVMVDFKTKKR
YLRPSALVFREIATHNGIPDELQHLTLIQ

[ ] - cleaved during protein expression

## 6. Effect of UV light and heating on L-Hag and 1-glycosyl thiols

**L-Hag** - Fmoc-L-Asn(Trt)-L-Hag-OH was synthesised as described on pg. 269. The crude reaction mixture was concentrated *in vacuo*, redissolved in 0.25 M acetate buffer (pH 4)/DMF (1:1) (1 mL) and irradiated with UV light. Aliquots (5 μL) were removed from the reaction mixture during the course of the reaction and were analysed by HPLC using an analytical Phenomenex Synergi 4u Hydro-RP 80Å 100 x 4.6 mm column. Elution was performed at 1 mL/min using an isocratic mobile phase (55% acetonitrile in $H_2O$, 0.1% TFA). Eluent was monitored at $\lambda$ = 214 nm.

**1-Glycosyl Thiols -** 1-Thio-β-D-glucose sodium salt (β-GlcSNa) or β-GlcNAcSH (**29**) was dissolved to the desired final concentration (0.1 or 1 M) in the desired concentration (0.05 or 0.25 M) of acetate buffer (pH 4), sodium phosphate buffer (pH 7) or CAPS buffer (pH 10). The solution was irradiated with UV light, heated to 55°C, or incubated at room temperature. After 4 hours, the reaction mixture was lyophilised and analysed by $^1$H NMR (MeOD or $D_2O$).

## 7. Chemoselectivity for natural amino acids

A single amino acid selected from the 20 naturally occurring amino acids (excluding isoleucine) and homoallylglycine (*ca.* 5 mg), Vazo44 (*ca.* 1 mg), 1-thio-β-D-glucose (*ca.* 5 mg) and $H_2O$ (1 mL) were added to a cuvette. The reaction mixture was irradiated with UV light for 12 hours. Aliquots (3 μL) were removed from the reaction vessel at 4 and 12 hours, and were analysed by ESI-MS using a MicroTOF spectrometer.

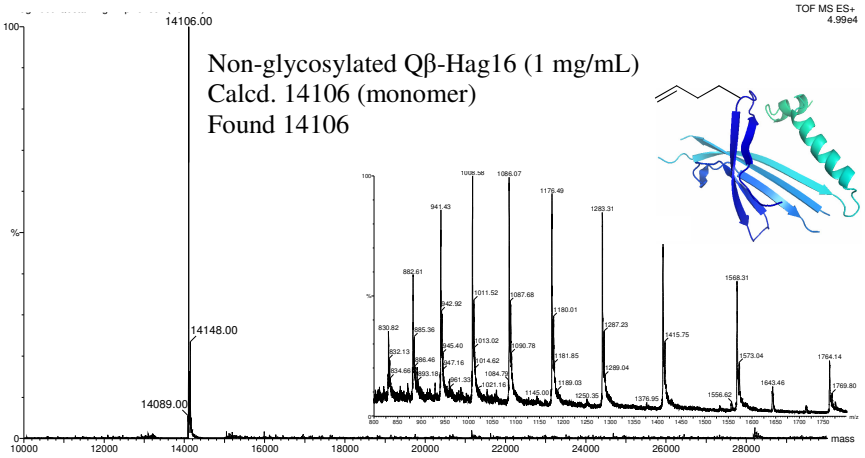
## 8. *S*-linked glycoprotein synthesis

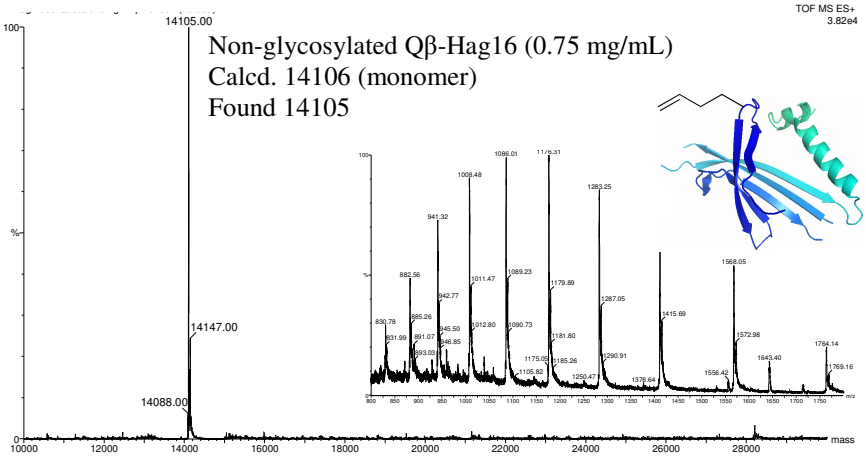
General procedure: To a solution of protein (typically 500 μL) in 0.25 M acetate buffer (pH 4), sodium phosphate buffer (pH 6 or 7) or CAPS buffer (pH 10) were added the 1-glycosyl thiol and in some instances Vazo44 (1 eq 1-glycosyl thiol : 0.2 eq Vazo44). The reaction mixture was placed in a cuvette and was irradiated with UV light, heated to 55 $^{o}$C, or incubated at room temperature. Aliquots (50 μL) were removed from the reaction mixture over the course of the experiment and were analysed by ESI-MS. In some instances, enzymatic activity of the protein was demonstrated by reaction with X-Gal at the end of the reaction.

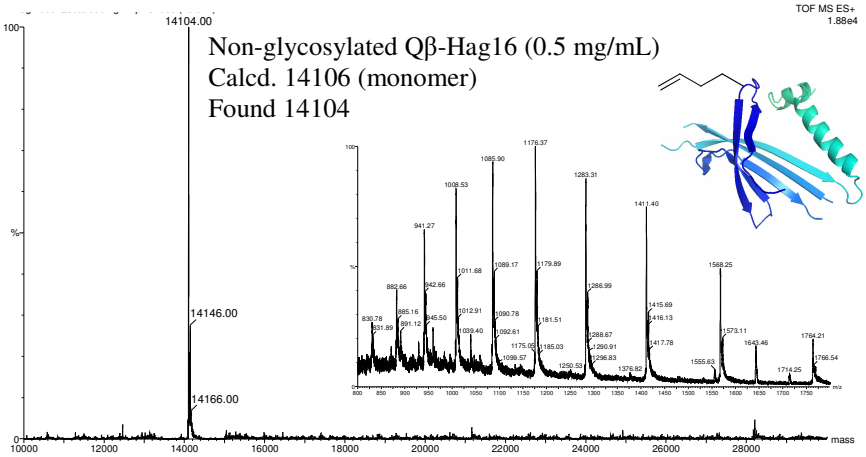
## 9. Ionisation efficiency

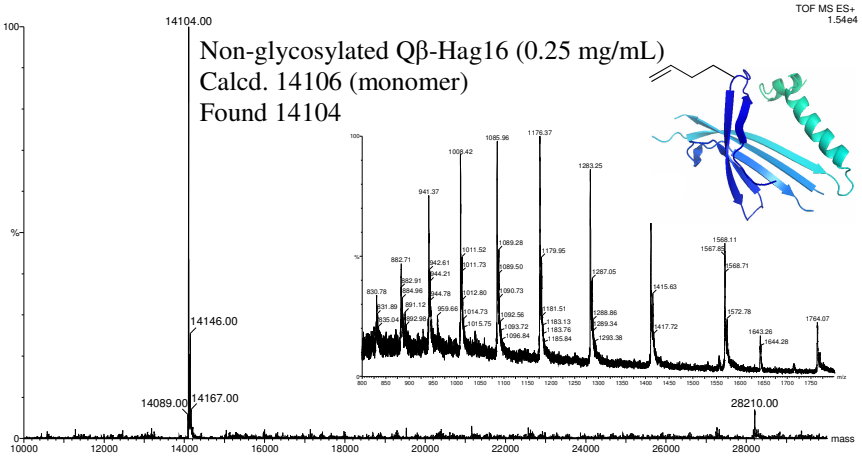
Non-glycosylated and glycosylated Qβ-Hag16 at 1, 0.75, 0.5, 0.25 and 0.1 mg/mL were analysed by ESI-MS in positive ion mode on a Micromass LCT mass spectrometer interfaced with a Waters 2790 Alliance HT HPLC using a Phenomenex Jupiter 5u C4 300A 250 x 4.6 mm column, as described on pg. 211 (Figure 116). The data were processed as follows: A mass spectrum of multiply charged ions was

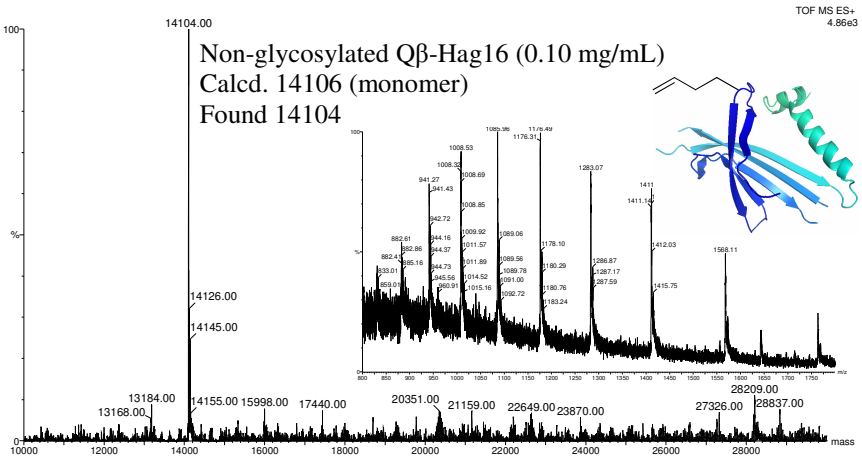
generated by combining data from the chromatogram between 14–18 minutes. The spectrum from *m/z* 800–1800 was displayed and the total ion count of the largest peak was plotted against protein concentration (Figure 31). A deconvoluted mass spectrum was then generated using the Maximum Entropy method (MaxEnt1) with 20 iterations. The total ion count of the peak corresponding to Qβ-Hag16 protein was plotted against protein concentration (Figure 31). Myoglobin (equine heart) (1 mg/mL) was used as the calibrant.

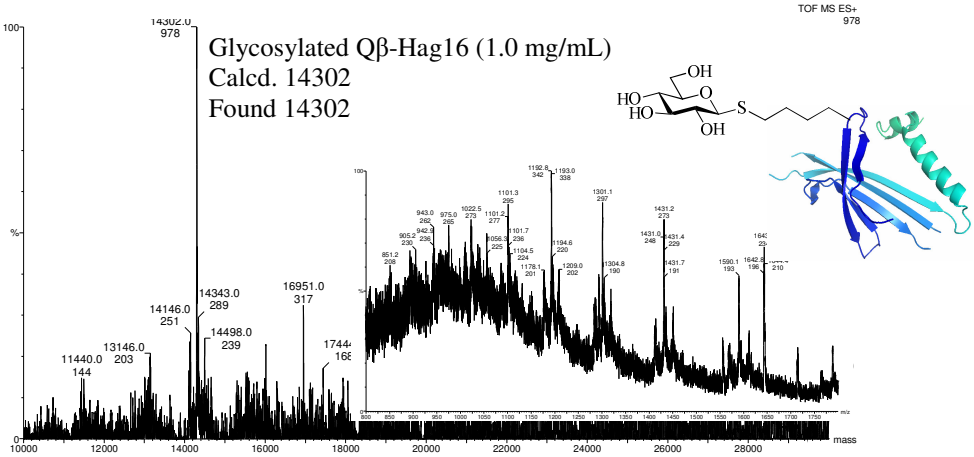

TOF MS ES+
4.99e4
14106.00
Non-glycosylated Qβ-Hag16 (1 mg/mL)
Calcd. 14106 (monomer)
Found 14106
14148.00
14089.00
mass

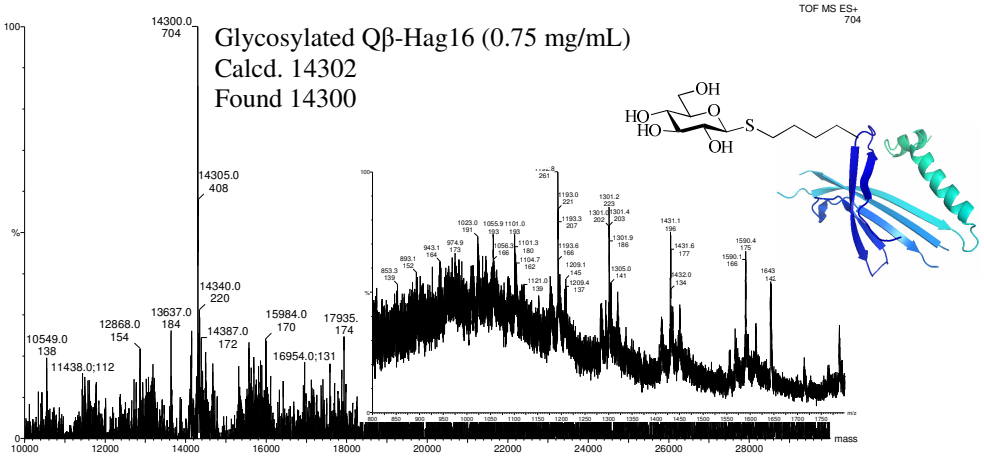

TOF MS ES+
3.82e4
14105.00
Non-glycosylated Qβ-Hag16 (0.75 mg/mL)
Calcd. 14106 (monomer)
Found 14105
14147.00
14088.00
mass

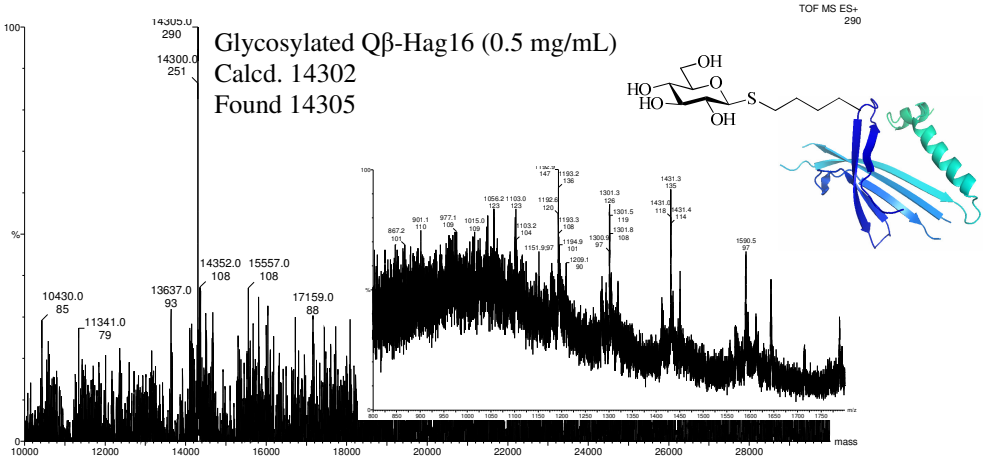

TOF MS ES+
1.88e4
14104.00
Non-glycosylated Qβ-Hag16 (0.5 mg/mL)
Calcd. 14106 (monomer)
Found 14104
14146.00
14166.00
mass

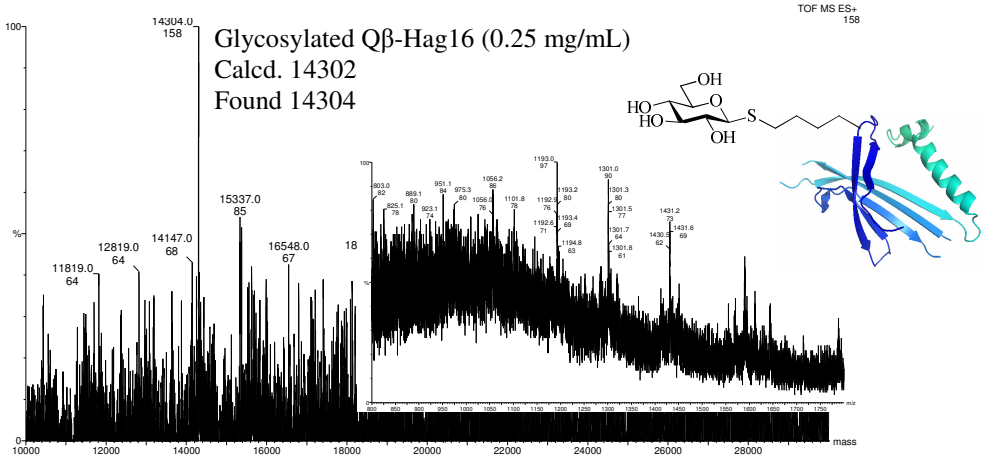

TOF MS ES+
1.54e4
Non-glycosylated Qβ-Hag16 (0.25 mg/mL)
Calcd. 14106 (monomer)
Found 14104
14104.00
14146.00
14089.00
14167.00
28210.00
mass


TOF MS ES+
4.86e3
Non-glycosylated Qβ-Hag16 (0.10 mg/mL)
Calcd. 14106 (monomer)
Found 14104
14104.00
14126.00
14145.00
13184.00
13168.00
14155.00
15998.00
17440.00
20351.00
21159.00
22649.00
23870.00
28209.00
27326.00
28837.00
mass


TOF MS ES+
978
Glycosylated Qβ-Hag16 (1.0 mg/mL)
Calcd. 14302
Found 14302
OH
HO
O
HO
S
OH
14343.0
289
16951.0
317
14146.0
251
14498.0
239
13146.0
203
11440.0
144
mass


Glycosylated Qβ-Hag16 (0.75 mg/mL)
Calcd. 14302
Found 14300
TOF MS ES+
704
14300.0
704
14305.0
408
14340.0
220
13637.0
184
12868.0
154
10549.0
138
11438.0;112
14387.0
172
15984.0
170
16954.0;131
17935.
174
OH
HO
HO
O
S
OH
mass


Glycosylated Qβ-Hag16 (0.5 mg/mL)
Calcd. 14302
Found 14305
TOF MS ES+
290
14305.0
290
14300.0
251
14352.0
108
15557.0
108
13637.0
93
10430.0
85
11341.0
79
17159.0
88
OH
HO
HO
O
S
OH
mass


Glycosylated Qβ-Hag16 (0.25 mg/mL)
Calcd. 14302
Found 14304
TOF MS ES+
158
14304.0
158
15337.0
85
14147.0
68
12819.0
64
11819.0
64
16548.0
67
OH
HO
HO
O
S
OH
mass

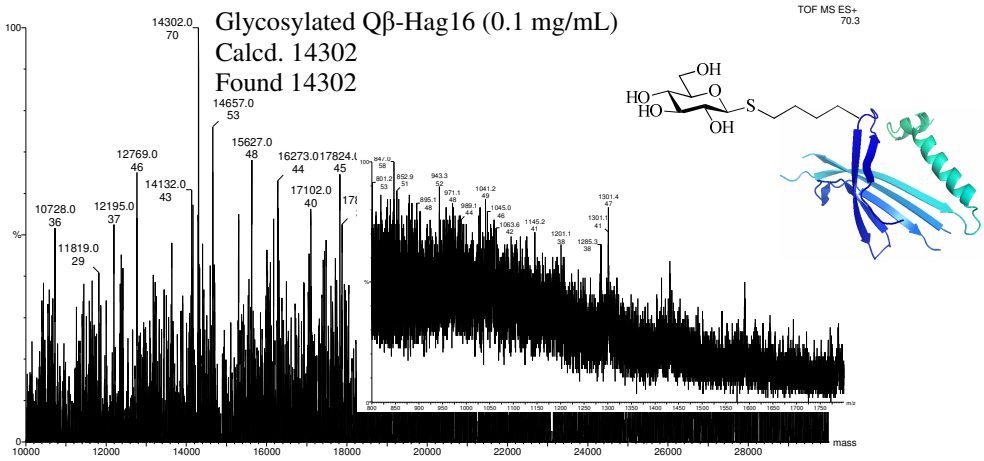



**Figure 116.** Raw (insert) and deconvoluted ESI-MS for non-glycosylated and glycosylated Qβ-Hag16.

## 10. Photo-induced cleavage reactions

UV light was generated using a borosilicate glass immersion well fitted with a medium pressure Hg lamp (125 W) (Hanovia or Photochemical Reactors Ltd.).

### 10.1. H-HWP-$NH_2$ and H-LNMYHWPLPL-$NH_2$

H-HWP-$NH_2$ or H-LNMYHWPLPL-$NH_2$ (10 mg) was dissolved in 50 mM sodium phosphate buffer (pH 6.5) (1 mL) and was added to a cuvette. The peptide solution was irradiated with UV light for 7 hours at room temperature. Aliquots (20 µL) were removed from the reaction mixture during the course of the reaction and were analysed by ESI-MS in both positive- and negative-ion mode. No photo-induced cleavage of either peptide was observed.

### 10.2. Proteins (excluding T50-, T100- and T140-*Ss*βG)

General procedure: protein (typically 500 μL, 0.5–2 mg/mL) in 50 mM sodium phosphate buffer (pH 6–7.5) was added to two BRAND UV cuvettes (Sigma-Aldrich). One solution was allowed to stand at room temperature, while the other solution was irradiated with UV light at room temperature. Aliquots (20-50 μL) were removed from the reaction mixtures during the course of the reaction and were analysed by SDS-PAGE (20 μL) (see *e.g.* Figure 49 for *Ss*βG proteins) and/or ESI-MS (30 μL) (see *e.g.* Figure 51 for *Ss*βG proteins). In some instances, the bands on the gel corresponding to protein were quantified by densitometry (pg. 320).

### 10.3. T0-, T50-, T100- and T140-*Ss*βG

T0-*Ss*βG (100 μL, 50 mM sodium phosphate (pH 6.5)), T50-*Ss*βG (100 μL, first elution fraction), T100-*Ss*βG (100 μL, first elution fraction) and T140-*Ss*βG (100 μL, first elution fraction) were each added to a cuvette. The protein solutions were irradiated with UV light at room temperature. Aliquots (20 μL) were removed from the reaction mixtures at 0 and 2 hours, and were analysed by SDS-PAGE (Figure 82) and anti-his western blot (Figure 82). For T0-*Ss*βG photo-induced cleavage also see pg. 319.

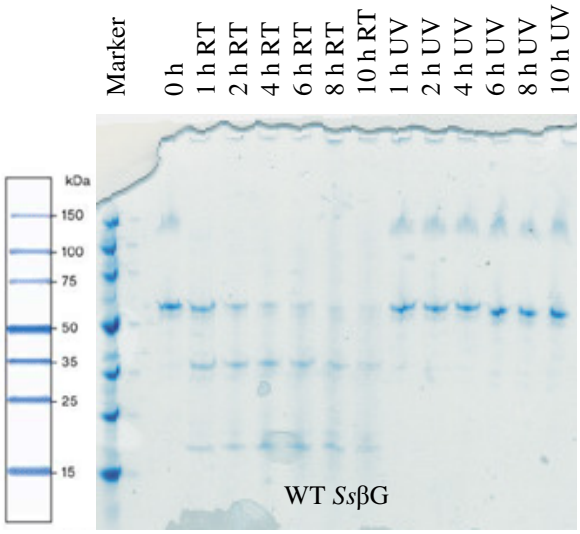
### 10.4. Intracellular photo-induced cleavage of WT *Ss*βG

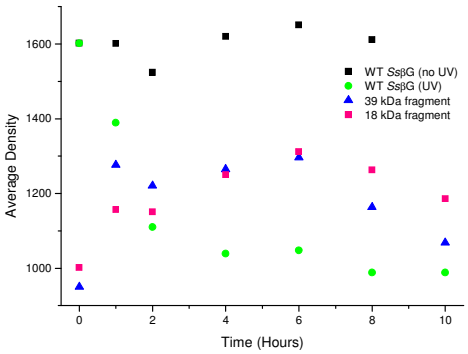
Cell pellet collected from the expression of WT *Ss*βG (2 mL) was added to a cuvette and was irradiated with UV light for 4 hours at room temperature. The cells were stored at -20$^{o}$C in binding buffer (20 mM Tris-HCl, 5 mM imidazole, 0.5 M NaCl, pH 7.8) (5 mL) overnight. The cell pellet was incubated for 1 hour on ice before being sonicated (15% power in 5 bursts of 20 seconds with a rest time of 40 seconds

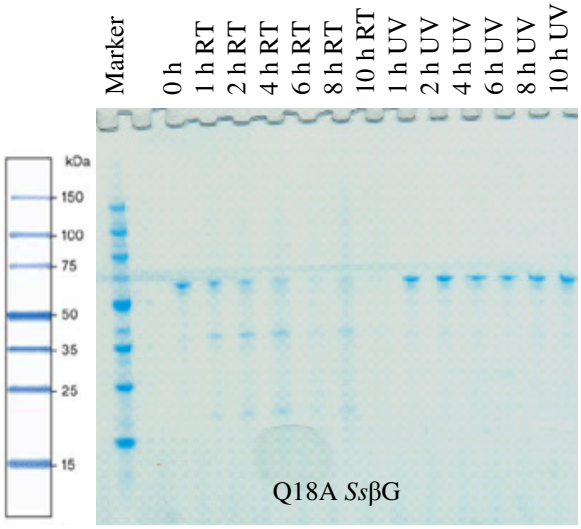
between each burst). The cell debris was removed from the lysis mixture by centrifugation (20 minutes, 18000 rpm, 4°C) and the lysate was sterile filtered (0.4 μm). The protein was purified from the cleared cell lysate using an AKTA FPLC system and HisTrap FF 1 mL column. Protein was injected and eluted using a gradient from 100% binding buffer to 100% elution buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, pH 7.8) (Figure 99). All eluent collected from the purification was analysed by SDS-PAGE.

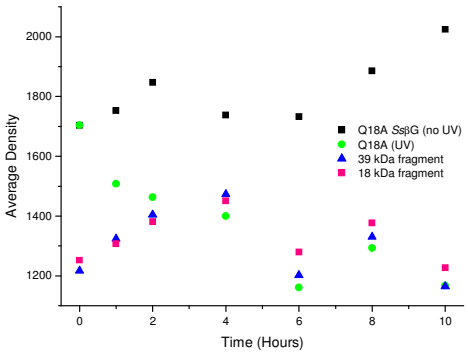
**11. Densitometry**

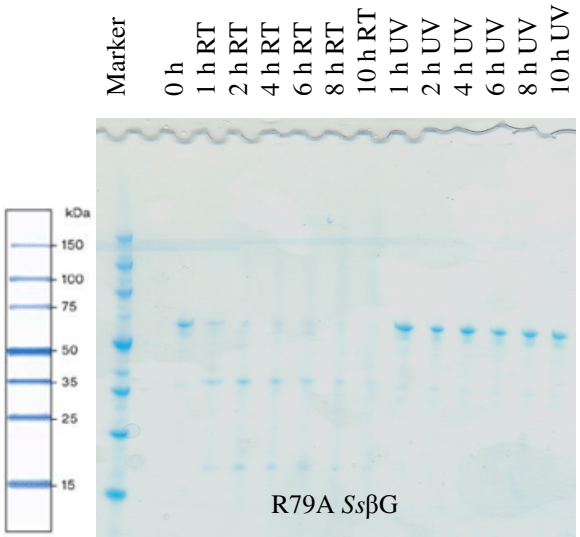
The bands corresponding to protein on certain gels (see below, Figure 117) were quantified by densitometry, using a Bio-Rad Gel Doc XR system with Quantity One software. The average peak density for each data point minus the background was recorded and plotted.

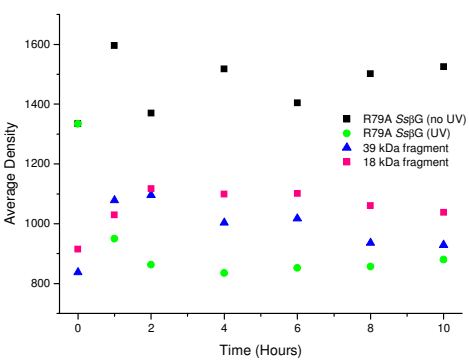

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
WT SsβG

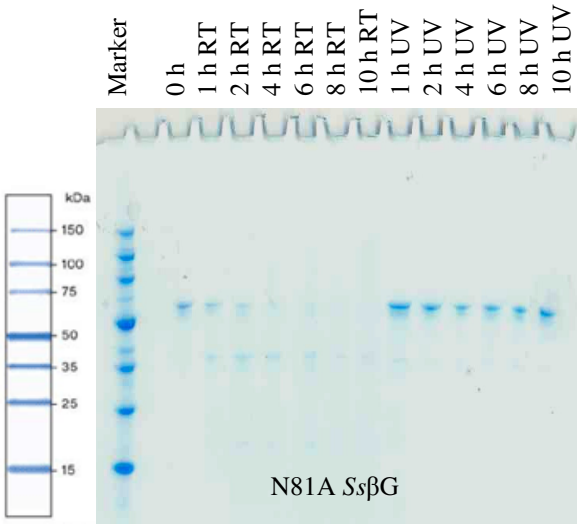

Average Density
Time (Hours)
WT SsβG (no UV)
WT SsβG (UV)
39 kDa fragment
18 kDa fragment

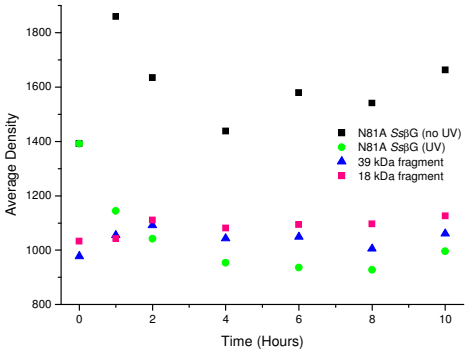

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
Q18A SsβG

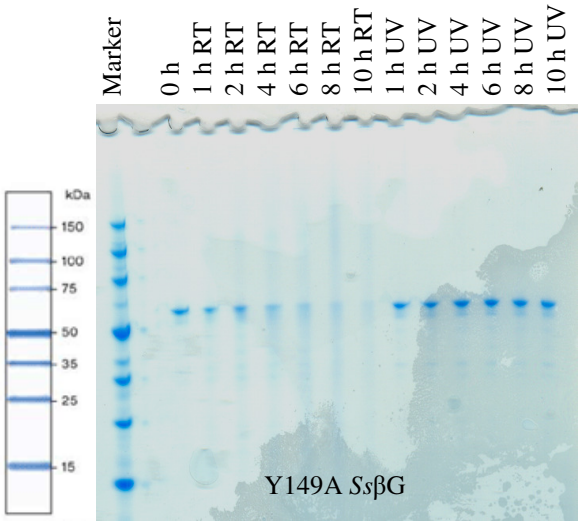

Average Density
Time (Hours)
Q18A SsβG (no UV)
Q18A (UV)
39 kDa fragment
18 kDa fragment

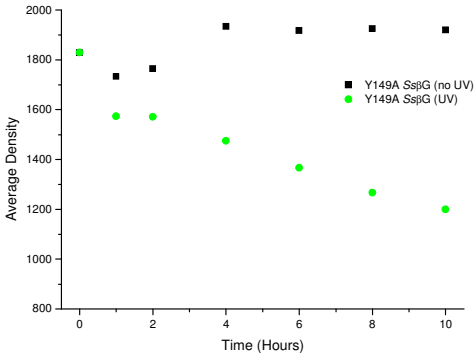

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
R79A SsβG

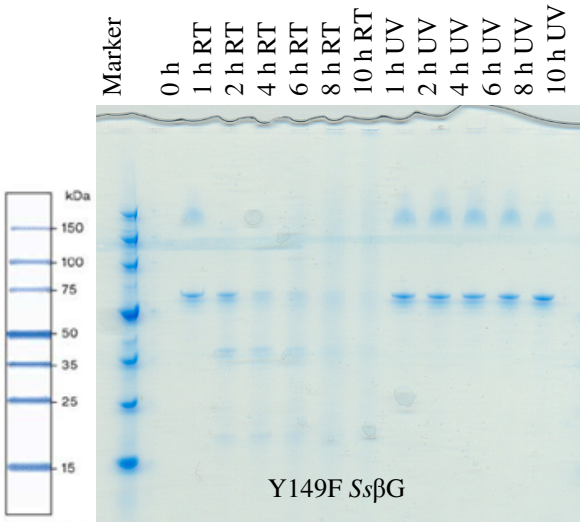

Average Density
Time (Hours)
R79A SsβG (no UV)
R79A SsβG (UV)
39 kDa fragment
18 kDa fragment

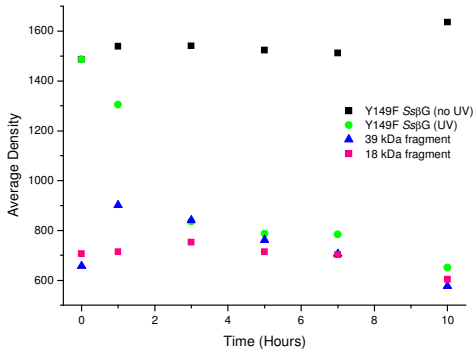

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
N81A SsβG

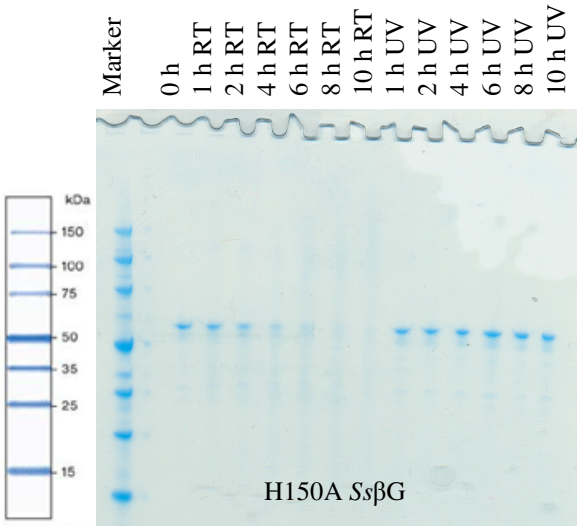

Average Density
Time (Hours)
N81A SsβG (no UV)
N81A SsβG (UV)
39 kDa fragment
18 kDa fragment

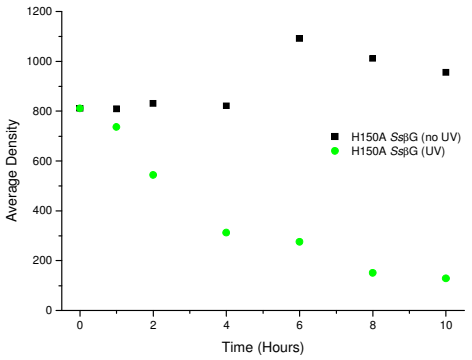

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
Y149A SsβG

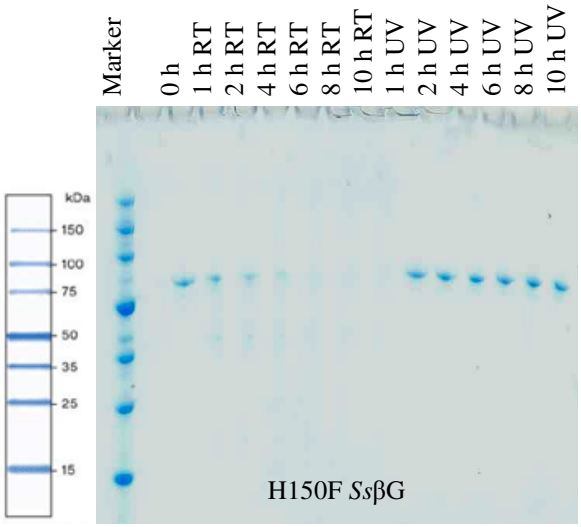

Average Density
Time (Hours)
Y149A SsβG (no UV)
Y149A SsβG (UV)

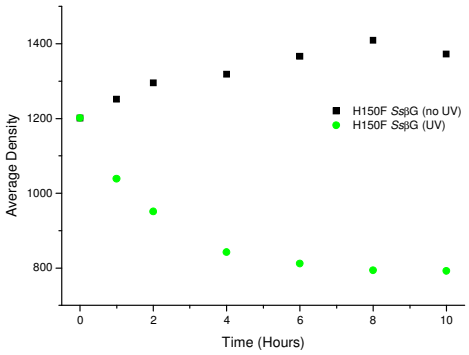

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
Y149F SsβG

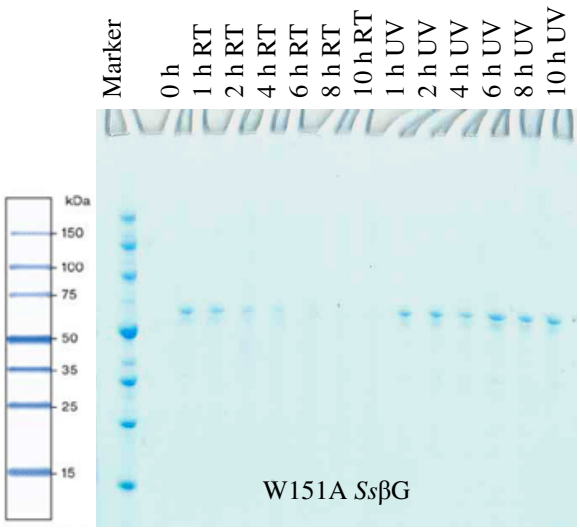

Average Density
Time (Hours)
Y149F SsβG (no UV)
Y149F SsβG (UV)
39 kDa fragment
18 kDa fragment

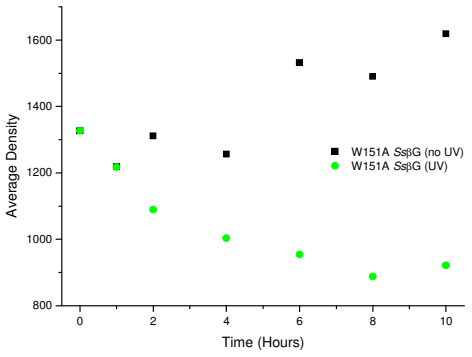

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
H150A SsβG

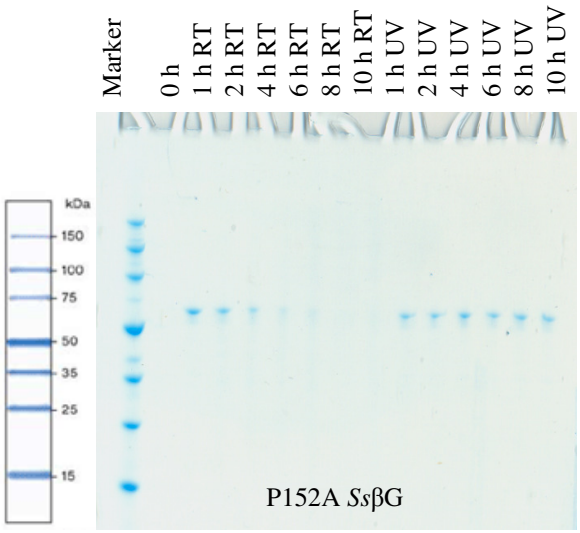

Average Density
Time (Hours)
H150A SsβG (no UV)
H150A SsβG (UV)

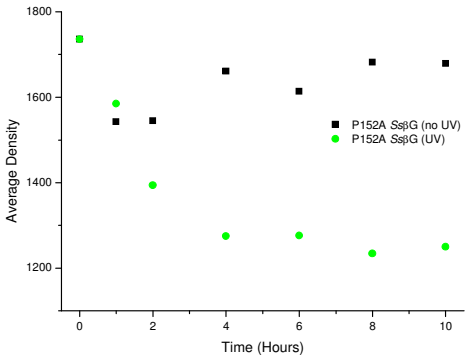

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
H150F SsβG

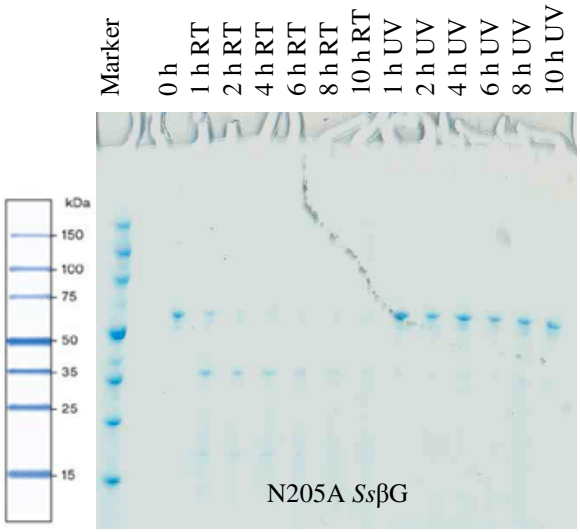

Average Density
Time (Hours)
H150F SsβG (no UV)
H150F SsβG (UV)

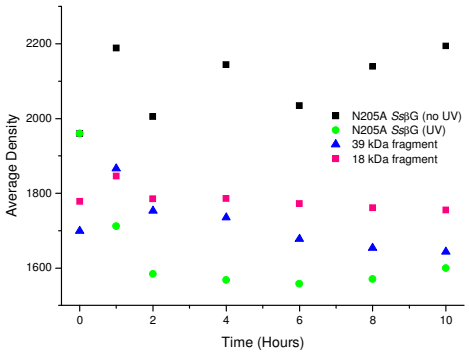

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
W151A SsβG

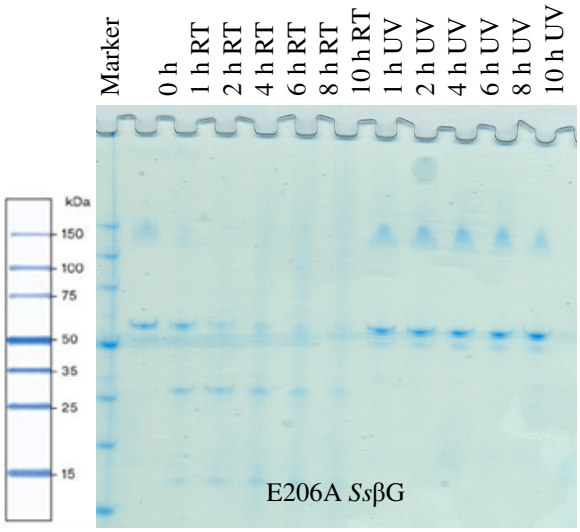

Average Density
Time (Hours)
W151A SsβG (no UV)
W151A SsβG (UV)

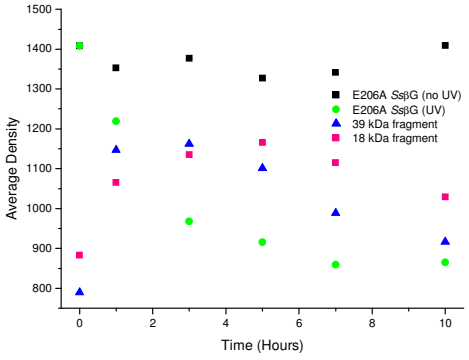

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
P152A SsβG

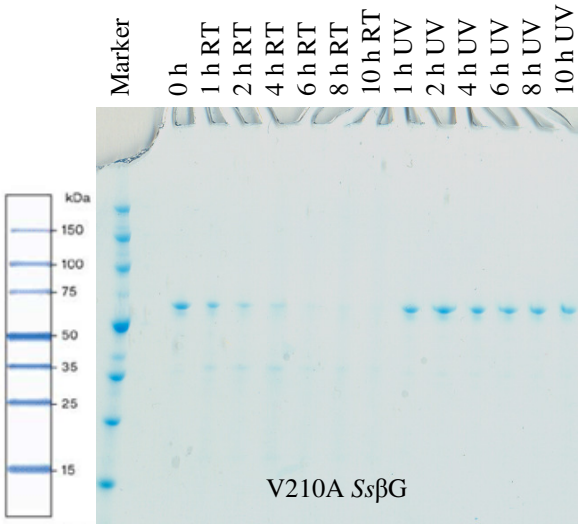

Average Density
Time (Hours)
P152A SsβG (no UV)
P152A SsβG (UV)

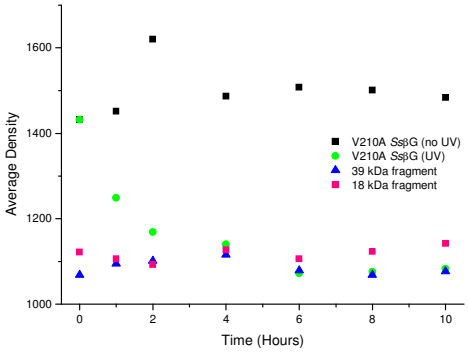

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
N205A SsβG

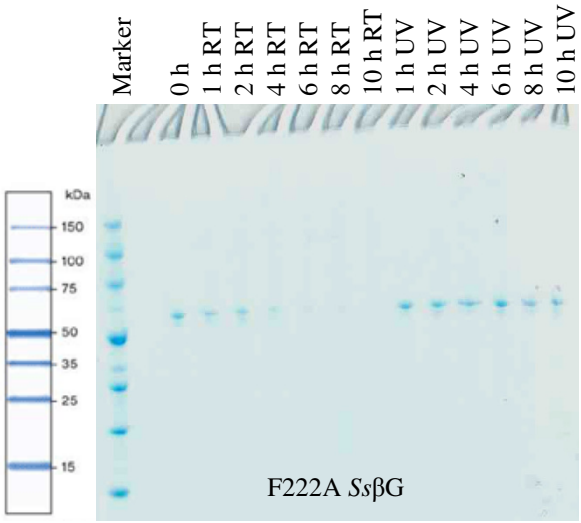

Average Density
Time (Hours)
N205A SsβG (no UV)
N205A SsβG (UV)
39 kDa fragment
18 kDa fragment

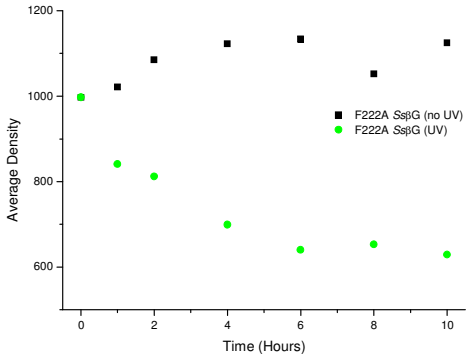

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
E206A SsβG

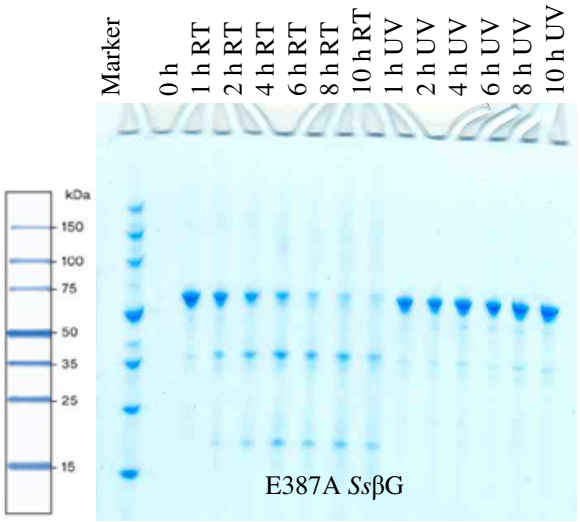

Average Density
Time (Hours)
E206A SsβG (no UV)
E206A SsβG (UV)
39 kDa fragment
18 kDa fragment

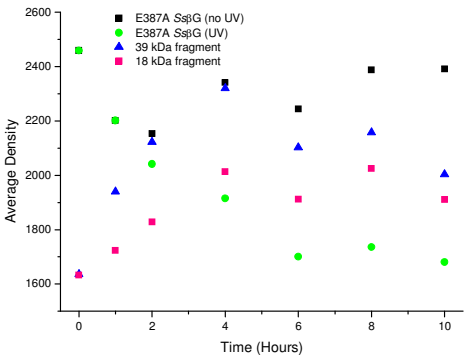

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
V210A SsβG

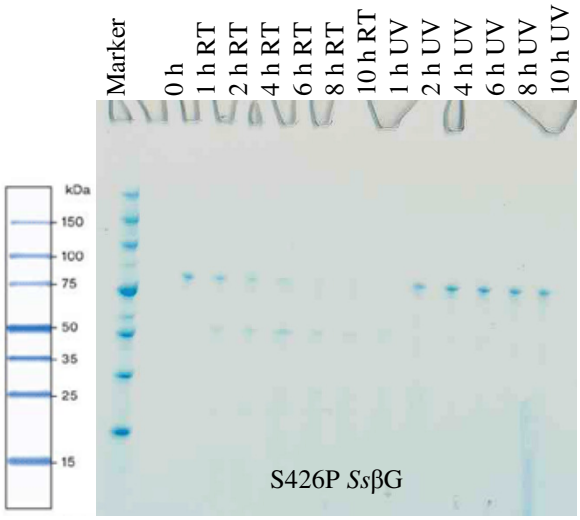

Average Density
Time (Hours)
V210A SsβG (no UV)
V210A SsβG (UV)
39 kDa fragment
18 kDa fragment

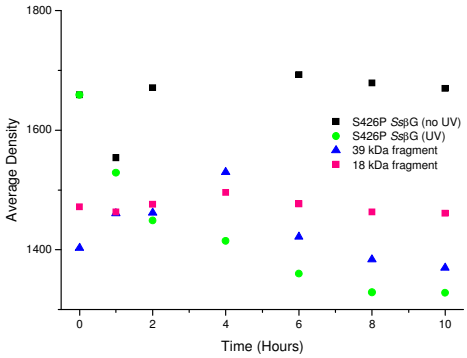

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
F222A SsβG

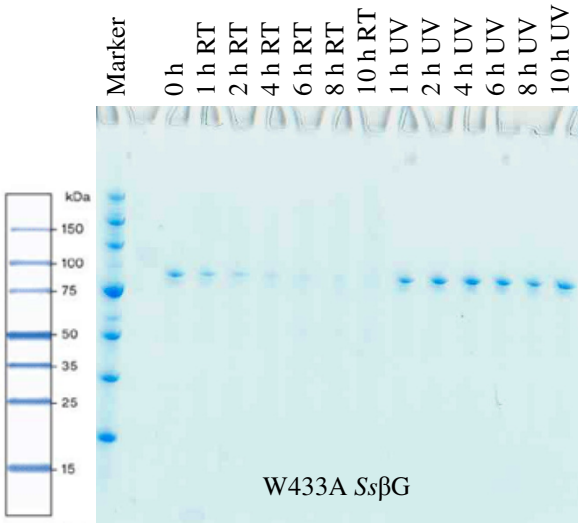

Average Density
Time (Hours)
F222A SsβG (no UV)
F222A SsβG (UV)

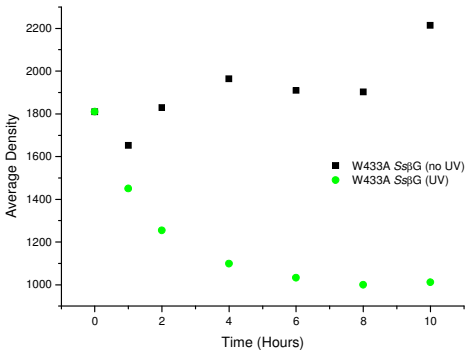

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
E387A SsβG

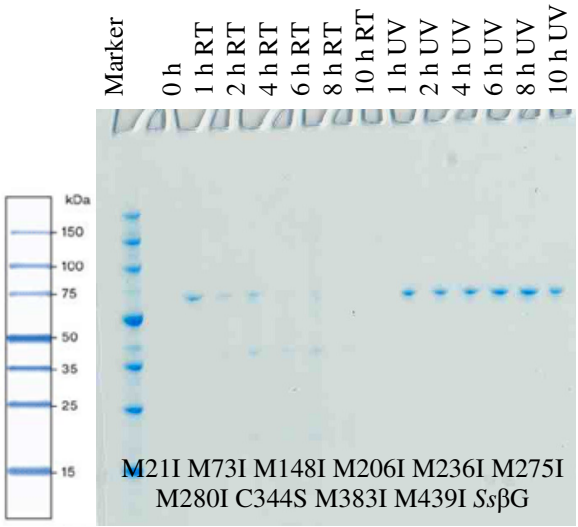

Average Density
Time (Hours)
E387A SsβG (no UV)
E387A SsβG (UV)
39 kDa fragment
18 kDa fragment

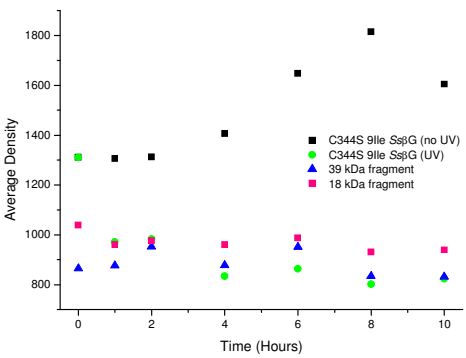

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
S426P SsβG

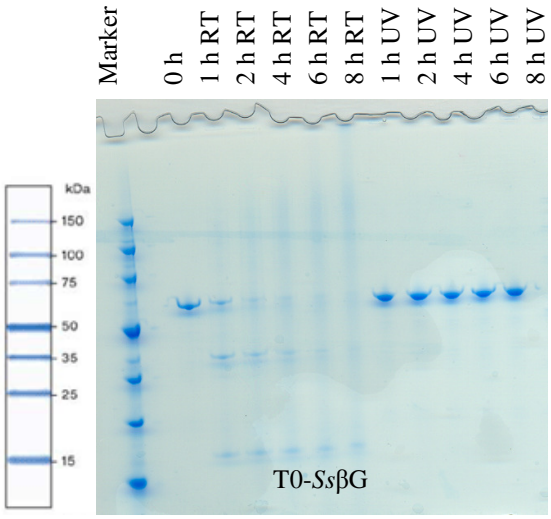

Average Density
Time (Hours)
S426P SsβG (no UV)
S426P SsβG (UV)
39 kDa fragment
18 kDa fragment

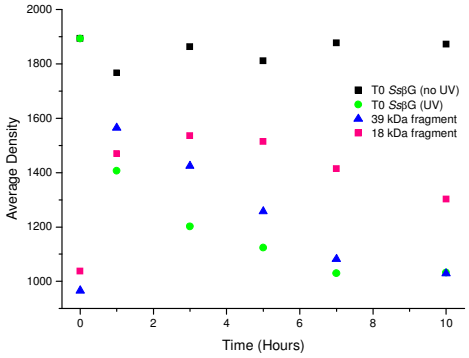

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
W433A SsβG

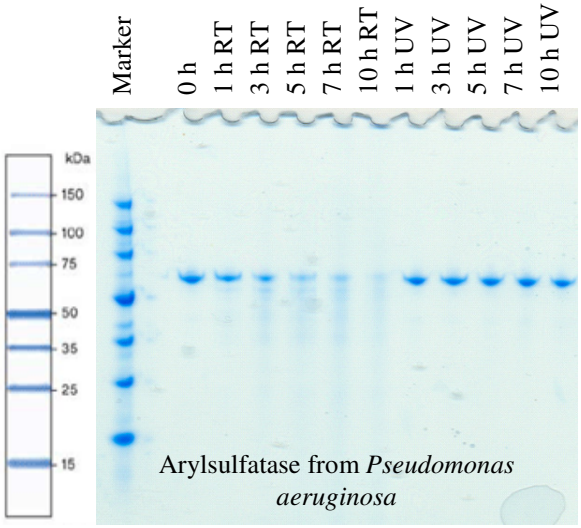

Average Density
Time (Hours)
W433A SsβG (no UV)
W433A SsβG (UV)

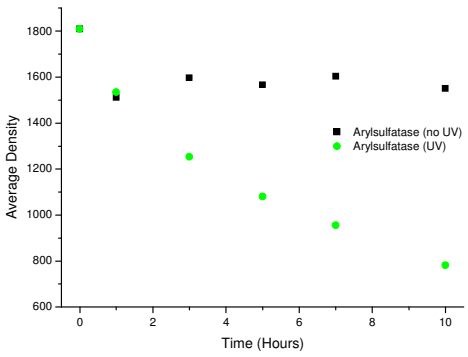

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
M21I M73I M148I M206I M236I M275I
M280I C344S M383I M439I SsβG

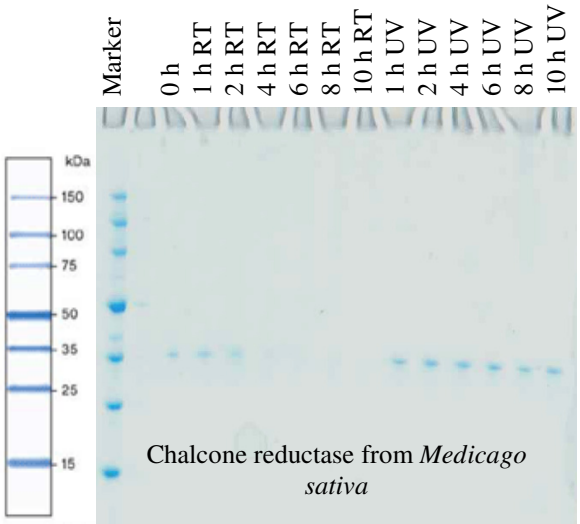

Average Density
Time (Hours)
C344S 9Ile SsβG (no UV)
C344S 9Ile SsβG (UV)
39 kDa fragment
18 kDa fragment

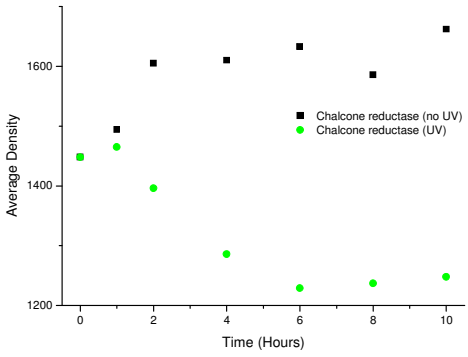

Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
kDa
150
100
75
50
35
25
15
T0-SsβG


Average Density
Time (Hours)
T0 SsβG (no UV)
T0 SsβG (UV)
39 kDa fragment
18 kDa fragment


Marker
0 h
1 h RT
3 h RT
5 h RT
7 h RT
10 h RT
1 h UV
3 h UV
5 h UV
7 h UV
10 h UV
kDa
150
100
75
50
35
25
15
Arylsulfatase from Pseudomonas aeruginosa


Average Density
Time (Hours)
Arylsulfatase (no UV)
Arylsulfatase (UV)


Marker
0 h
1 h RT
2 h RT
4 h RT
6 h RT
8 h RT
10 h RT
1 h UV
2 h UV
4 h UV
6 h UV
8 h UV
10 h UV
kDa
150
100
75
50
35
25
15
Chalcone reductase from Medicago sativa


Average Density
Time (Hours)
Chalcone reductase (no UV)
Chalcone reductase (UV)

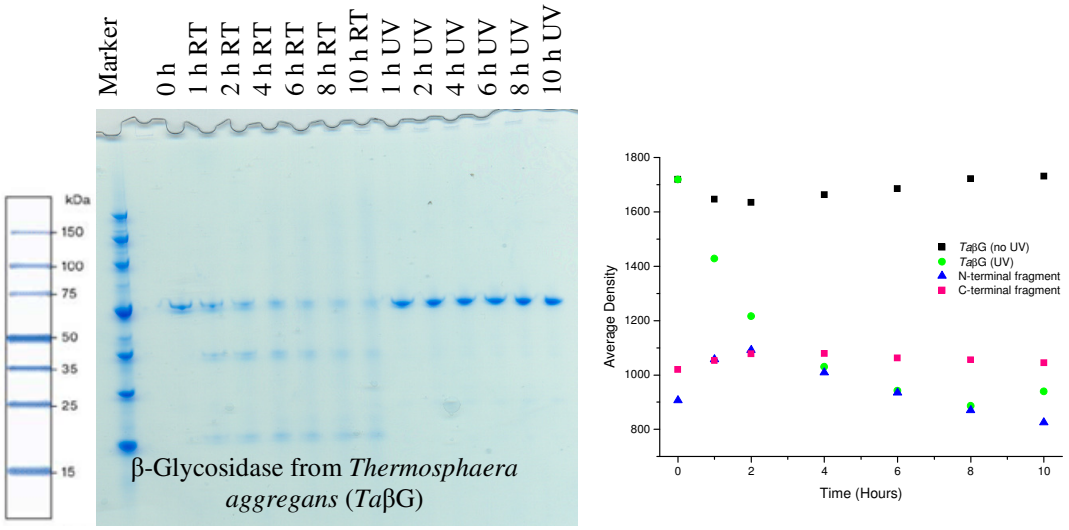



**Figure 117.** Time course analysis of proteins exposed to/not exposed to UV light by SDS-PAGE (left). Densitometry (right). Note: lines

Whilst densitometry allowed the reaction course of the photo-induced cleavage of a number of proteins to be followed, this technique is limited in terms of accuracy.[408, 409] For example, in the case of Q18A *Ss*βG the background reading taken at two different points on the gel varied significantly (see Figure 118).

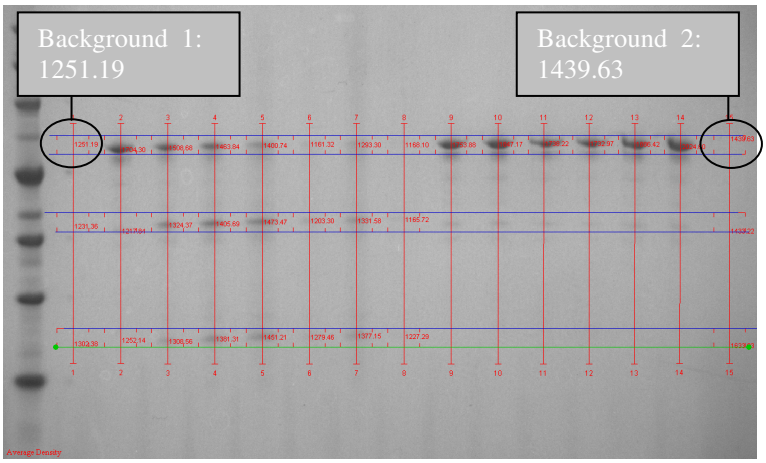



**Figure 118.** Densitometry of Q18A *Ss*βG SDS-PAGE from Figure 117.

## 12. Reactive transitions

WT *Ss*βG (200 μL, 5 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a quartz cuvette (transmission range 200 – 2500 nm) and glass cuvette (transmission range 360 – 2500 nm). The two protein solutions were irradiated with UV light for 4 hours at room temperature. Aliquots (20 μL) were removed from the reaction mixtures during the course of the reaction and were analysed by SDS-PAGE (Figure 39).

WT *Ss*βG (500 μL, 1 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to 14 quartz cuvettes. Laser light (193 nm [direct from an ArF laser], 222.5 nm [Coumarin 2], 227.5 nm [Coumarin 2], 240 nm [Coumarin 307], 250 nm [Coumarin 307], 260 nm [Coumarin 307], 265 nm [Coumarin 307], 282.5 nm [Rhodamine 6G], 290 nm [Rhodamine 6G], 300 nm [Sulforhodamin B], 308 nm [Sulforhodamin B], 320 nm [Rhodamine 101], 330 nm [Rhodamine 101], 355 nm [direct from YAG]) was generated by a neodymium-doped yttrium aluminium garnet (Nd:YAG)-pumped tunable dye laser, except in the case of 193nm, where an $Ar/F_2$ excimer laser was used. A constant energy of 0.5 mJ / pulse was employed at 10 pulses / s (giving a mean power of 5 mW, pulse times: 5ns for YAG, 20 ns for ArF). The samples were exposed to the radiation for 15 minutes each, after which time an aliquot (20 μL) was removed from each reaction vessel for analysis by SDS-PAGE (Figure 40).

## 13. N-terminal sequencing

WT *Ss*βG (175 μL, 1.2 mg/mL, 50 mM sodium phosphate buffer, pH 7.5) was added to a cuvette, and was irradiated with UV light for 2 hours at room temperature. An aliquot (10 μL) of the reaction mixture was submitted to the Protein Characterisation

Facility, University of Oxford for N-terminal sequencing. The 18 kDa and 39 kDa fragments were separated by SDS-PAGE and classical Edman degradation with phenylisothiocyanate was used. Results: 18 kDa fragment sequence – GHHHHH, 39 kDa fragment sequence – no sequence.

**14. Crystallisation screens**

**Hampton Research crystal screen kit HR2-110** - A Cryschem plate was sprayed with canned air to remove dust. Crystallisation reagent 1 (0.5 mL) was pipetted into reservoir A1 of the crystallisation plate. P152A or S426P *Ss*βG (2.5 mg/mL, 50 mM sodium phosphate buffer, pH 6.5, 1μL) was pipetted into the post of reservoir A1. Crystallisation reagent (1μL) was pipetted from reservoir A1 on to the drop in post A1. These steps were repeated for the crystallisation reagents 2-48, excluding 33. The plate was sealed with 2 strips of clear sealing tape. Wells were checked after 1 hour, 1 day, 1 week and 1 month for crystals, using a microscope connected to a digital camera.

**JCSG+, PACT premier, Stura, Natrix, Hampton Index, HR2-110, HR Gridscreens and Wizard 1+2 crystal screen kits** - P152A *Ss*βG was dialysed into 10 mM Tris-HCl (pH 7.5), concentrated to 10 mg/mL, sterile filtered (0.2 μm) and centrifuged (13000 rpm, 10 minutes). The protein was screened against the commercial crystal screens JCSG+, PACT premier and Stura (Molecular Dimensions), Natrix, Hampton Index, HR2-110 and HR Gridscreens (Hampton Research) and Wizard 1 + 2 (Emerald BioSystems). The 96 well plates were set up using a Tecan Genesis 150 crystallisation robot in the Laboratory of Molecular Biophysics, University of Oxford. The sitting drop vapor diffusion technique was used;

crystallisation reagent (100 μL) was added to the reservoir, and protein (0.1 μL) and crystallisation reagent (0.1 μL) to the platform. The best conditions (well C5, JCSG+ screen [0.1 M sodium HEPES, 0.8 M sodium dihydrogen phosphate 0.8 M potassium dihydrogen phosphate, pH 7.5], which gave long needle like crystals being formed after only 24 hours, was optimised. The following stock solutions were prepared: 0.1 M sodium HEPES (pH 6.8, pH 7.0, pH 7.2, pH 7.4, pH 7.5, pH 7.6, pH 7.8, pH 8.0 and pH 8.2), 2 M sodium dihydrogen phosphate and 2 M potassium dihydrogen phosphate. The plate was designed using www.biop.ox.ac.uk/www/noble/traymaker.html.

## 15. UV spectroscopy of WT *Ss*βG

WT *Ss*βG (1 mL, 0.6 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a cuvette, and was irradiated with UV light for 8 hours at room temperature. The UV spectrum of the protein solution was recorded using a Varian Cary 4000 UV-Vis spectrophotometer at time intervals during the course of the reaction (Figure 60).

## 16. Isomerisation of *trans*-3-indole acrylic acid

To a quartz flask were added *trans*-3-indole acrylic acid (200 mg) and ethyl acetate (30 mL). The reaction mixture was irradiated with UV light under argon. After 24 hours, t.l.c. (ethyl acetate/petrol (3:7)) indicated the formation of *cis*-3-indole acrylic acid (Figure 61). The reaction mixture was concentrated *in vacuo*. The observed $^{1}$H NMR chemical shifts and coupling constants for the vinylic protons of *trans*- and *cis*-3-indole acrylic acid were similar to those reported in the literature - $^{1}$H NMR (DMSO-$d_6$) *trans*-3-indole acrylic acid 6.30 ppm ($J_{1,2}$ 15.9 Hz) [$^{1}$H NMR ($CDCl_3$) Lit.

7.81 ppm ($J_{1,2}$ 15.9 Hz)][330]; $^1$H NMR (DMSO-$d_6$) *cis*-3-indole acrylic acid 5.68 ppm ($J_{1,2}$ 12.2 Hz) [$^1$H NMR ($CDCl_3$) Lit. 7.30 ppm ($J_{1,2}$ 12.6 Hz)][330]. *Trans*-/*cis*-3-indole acrylic acid (1:0.13).

## 17. Tagging the 39 kDa fragment of WT *Ss*βG

### 17.1. 1-Pyrenebutanoic acid hydrazide

WT *Ss*βG (1 mL, 1 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a cuvette, and was irradiated with UV light for 3 hours at room temperature to provoke protein fragmentation. The protein solution was transferred to a microcentrifuge tube and 1-pyrenebutanoic acid hydrazide (5 mg, 0.016 mmol, 1000 eq) was added. The reaction mixture was incubated for 3 hours at room temperature, then buffer exchanged (PD-10 column) into 50 mM sodium phosphate buffer (pH 6.5) to remove any unreacted 1-pyrenebutanoic acid hydrazide. The eluent was concentrated using a spin concentrator (10 kDa MWCO) and washed with $H_2O$ (3 x 1 mL). The reaction was followed using SDS-PAGE (Figure 119). The gel was visualised under UV light but no band corresponding to the tagged 39 kDa fragment was observed.

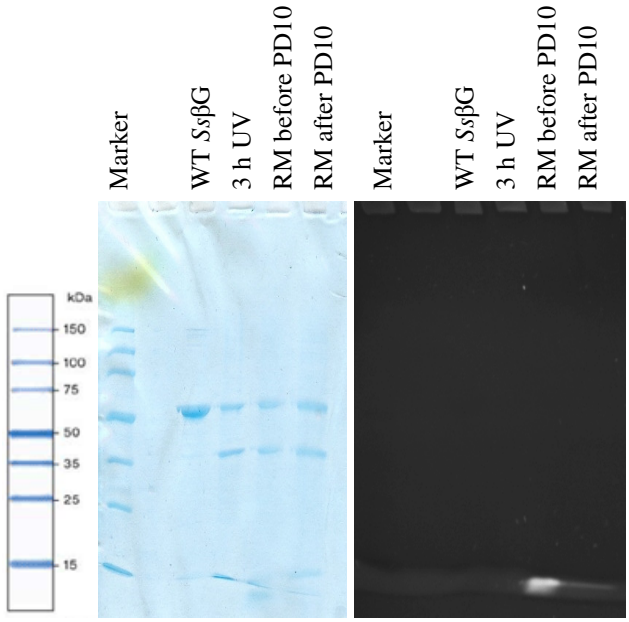



**Figure 119.** Time course analysis of tagging experiment (1-pyrenebutanoic acid hydrazide) by SDS-PAGE.

In order to improve the solubility of 1-pyrenebutanoic acid hydrazide the reaction was repeated: WT *Ss*βG (500 µL, 5.9 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a cuvette, and was irradiated with UV light for 2 hours at room temperature to provoke protein fragmentation. The following reactions were set up in microcentrifuge tubes:

| Rxn | Irradiated Protein Solution | 1-Pyrenebutanoic Acid Hydrazide | Solvent |
|---|---|---|---|
| 1 | 60 µL | 4 mg, 0.013 mmol, 2000 eq | methanol (40 µL) |
| 2 | 80 µL | 4 mg, 0.013 mmol, 1600 eq | acetonitrile (20 µL) |
| 3 | 80 µL | 4 mg, 0.013 mmol, 1600 eq | DMF (20 µL) |
| 4 | 60 µL | 4 mg, 0.013 mmol, 2000 eq | acetonitrile (40 µL) |
| 5 | 120 µL | 4 mg, 0.013 mmol, 4000 eq | DMF (80 µL) |

The reaction mixtures were agitated on a rocking table for 1 hour at room temperature, then buffer exchanged (PD-10 column) into 50 mM sodium phosphate buffer (pH 6.5). The eluents were concentrated using spin concentrators (10 kDa MWCO) and washed with $H_2O$ (3 x 1 mL). The reaction was followed using SDS-PAGE (Figure 64). The gel was visualised under UV light but no band corresponding to the tagged 39 kDa fragment was observed.

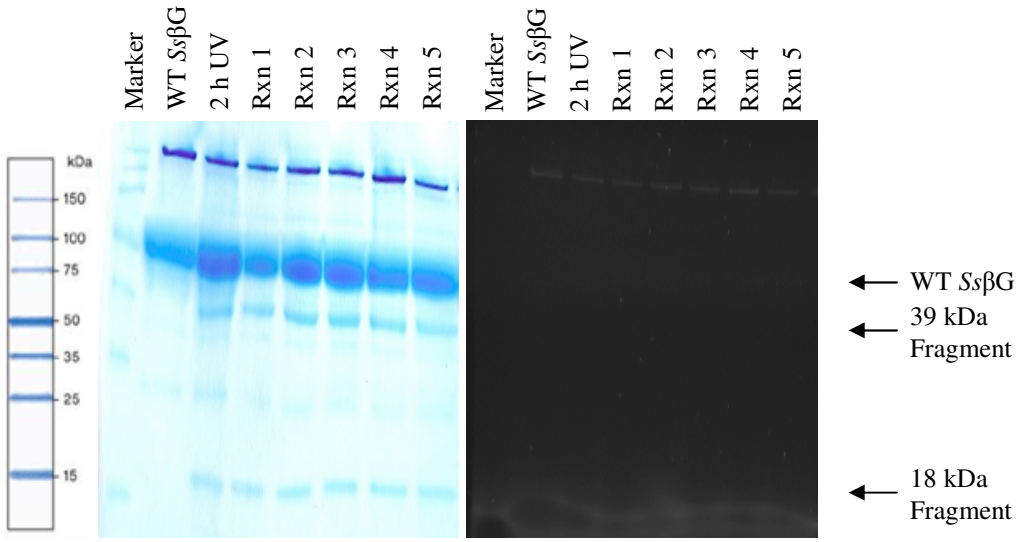



**Figure 120.** Time course analysis of tagging experiment (1-pyrenebutanoic acid hydrazide) by SDS-PAGE.

## 17.2. Fluorescein hydrazide

WT *Ss*βG (200 µL, 5.9 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a cuvette, and was irradiated with UV light for 2 hours at room temperature to provoke protein fragmentation. Irradiated protein solution (126 µL) was transferred to a microcentrifuge tube and fluorescein hydrazide (5 mg, 0.01 mmol, 750 eq) dissolved in DMSO (84 µL) was added (Rxn 1). To another microcentrifuge tube was added WT *Ss*βG (126 µL, 5.9 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) and fluorescein hydrazide (5 mg, 0.01 mmol) dissolved in DMSO (84 µL) (Rxn 2). The reaction vessels were covered in foil and incubated for 16 hours at room temperature.

The reaction mixtures were buffer exchanged (PD-10 column) into 50 mM sodium phosphate buffer (pH 6.5) to remove any unreacted fluorescein hydrazide. The eluents were concentrated using a spin concentrator (10 kDa MWCO) and washed with $H_2O$ (3 x 1 mL). The reaction was followed using SDS-PAGE (Figure 121). The gel was visualised under UV light (Bio-Rad Gel Doc XR system) but no band corresponding to the tagged 39 kDa fragment was observed.

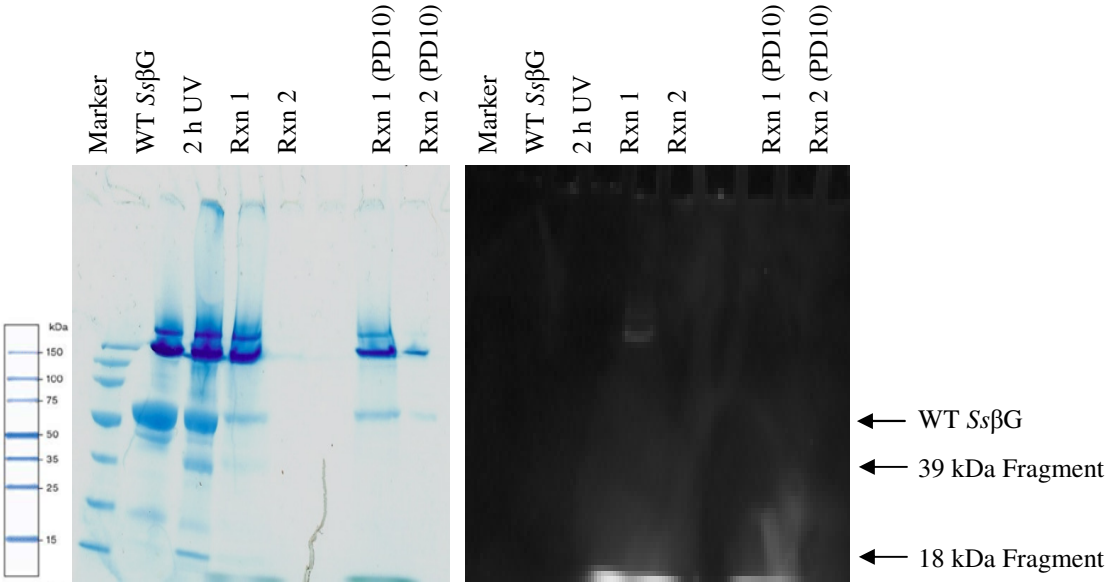



**Figure 121.** Time course analysis of tagging experiment (fluorescein hydrazide) by SDS-PAGE.

WT *Ss*βG (200 µL, 5.9 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a cuvette, and was irradiated with UV light for 5 hours at room temperature to provoke protein fragmentation. Irradiated protein solution (120 µL) was transferred to a microcentrifuge tube and fluorescein hydrazide (1 mg, 2.03 µmol, 150 eq) dissolved in DMSO (80 µL) was added. The reaction vessel was covered in foil and incubated for 16 hours at room temperature. The reaction mixture was buffer exchanged (PD-10 column) into 50 mM sodium phosphate buffer (pH 6.5) to remove any unreacted fluorescein hydrazide. The eluent was concentrated using a spin concentrator (10 kDa

MWCO) and washed with $H_2O$ (3 x 1 mL). The reaction was followed using SDS-PAGE (Figure 122). The gel was visualised under UV light (Mini Benchtop UV transilluminator) but all the bands appeared fluorescent, indicating non-specific interaction between fluorescein hydrazide and the proteins (WT *Ss*βG, 39 kDa fragment and 18 kDa fragment).

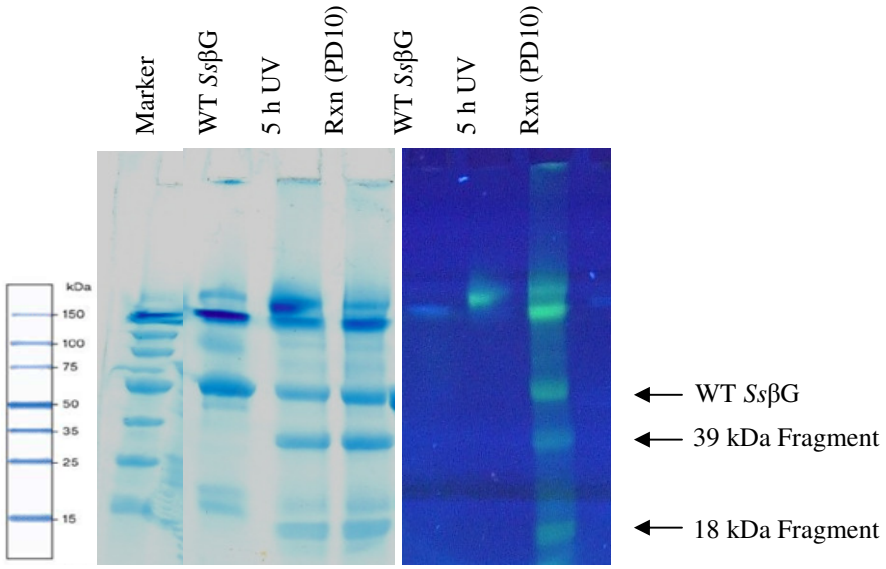



**Figure 122.** Time course analysis of tagging experiment (fluorescein hydrazide) by SDS-PAGE.

### 17.3. Thiophenol

M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG (100 µL, 1.7 mg/mL, 50 mM phosphate buffer, pH 6.5) was added to a cuvette, and irradiated with UV light for 4 hours at room temperature to provoke protein fragmentation. Irradiated protein solution (50 µL) was transferred to a microcentrifuge tube, and thiophenol sodium salt (3 mg, 7500 eq) dissolved in 50 mM sodium phosphate buffer (pH 6.5) (30 µL) was added. The reaction mixture was incubated for 4 hours at room temperature. The reaction was followed using ESI-MS, but no peak corresponding to the tagged 39 kDa fragment was observed.

### 17.4. Ethanethiol

M21I M73I M148I M206I M236I M275I M280I C344S M383I M439I *Ss*βG (200 μL, 1 mg/mL, 50 mM sodium phosphate buffer, pH 8.5) was added to a cuvette, and was irradiated with UV light at room temperature to provoke protein fragmentation. The protein solution was transferred to two microcentrifuge tubes (Rxn 1 and Rxn 2, respectively) (2 x 80 μL). To Rxn 1 was added ethanethiol (0.25 μL, 1000 eq). The mild denaturant Rapigest SF (20 μL, 1 mg/mL in $H_2O$) was added to Rxn 2, which was then incubated at 37°C for 30 minutes before the addition of ethanethiol (0.25 μL, 1000 eq). Both reaction mixtures were incubated with gentle shaking at room temperature. After 16 hours, an additional portion of ethanethiol (1 μL, 4000 eq) was added, and both reaction mixtures were incubated at 37°C for a further 3 hours. The reaction was followed using ESI-MS, but no peak corresponding to the tagged 39 kDa fragment was observed.

## 18. Oxygen related experiments

### 18.1. Photo-induced cleavage mechanism of *Ss*βG involving oxygen

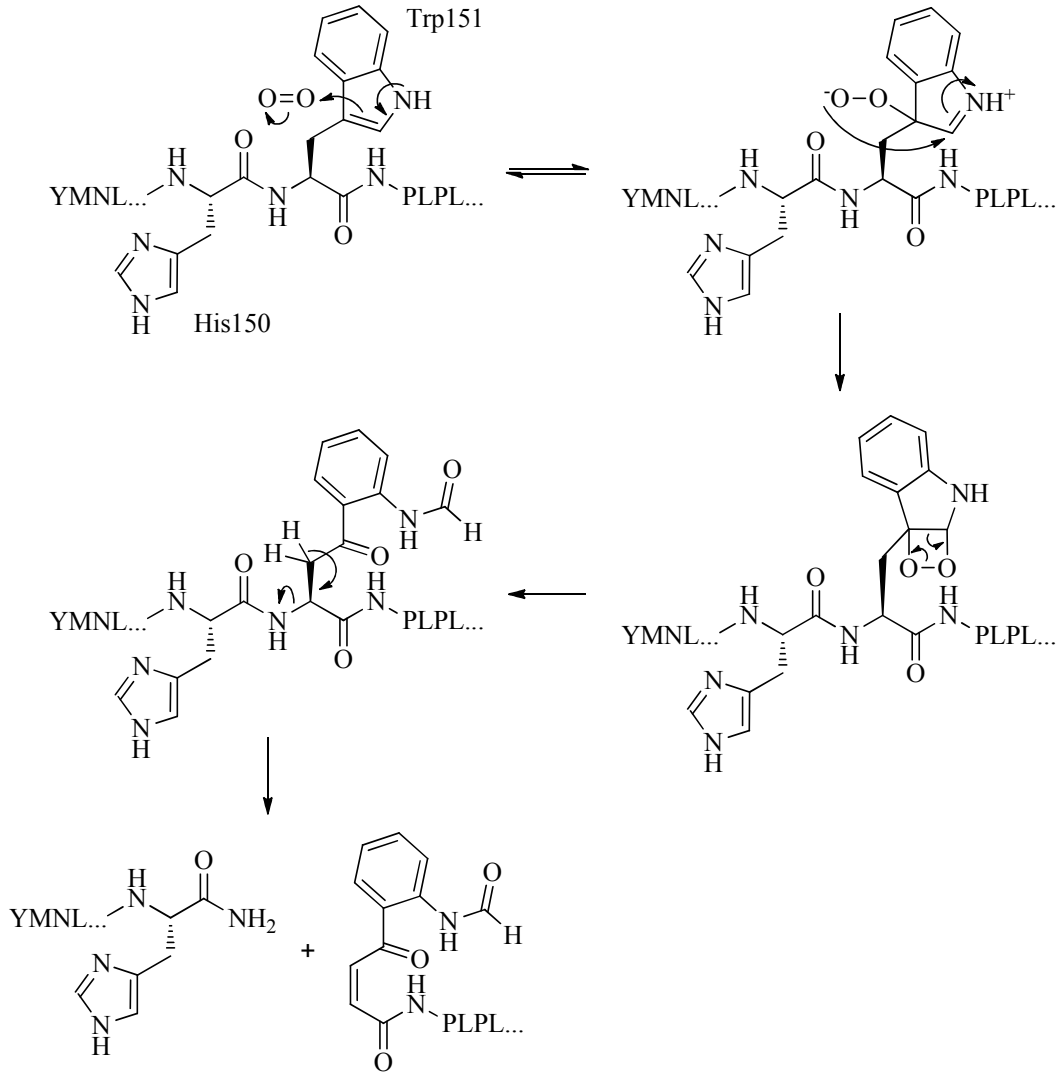



**Figure 123.** Putative photo-induced cleavage mechanism of *Ss*βG involving singlet oxygen that was disproved.

### 18.2. Bubbling with argon

WT *Ss*βG (500 μL, 5 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to two cuvettes. One protein solution was left open to the atmosphere for the duration of the experiment, while the other protein solution was degassed using vacuum / argon cycles for 5 minutes, and was then kept under an inert atmosphere for the remainder of the experiment by using a balloon filled with argon. The two protein solutions were irradiated with UV light for 4 hours at room temperature. Aliquots (20 μL) were removed from the reaction mixtures at hourly intervals during the course of the reaction and were analysed by SDS-PAGE (Figure 63).

### 18.3. Trolox

Trolox (0.5 mg) was dissolved in degassed $H_2O$ (1 mL) to give a 2 mM solution. WT *Ss*βG (500 μL, 2.5 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) and trolox solution (500 μL) were added to two cuvettes. One reaction mixture was left open to the atmosphere, while the other protein solution was degassed using argon for 30 minutes and was kept under argon for the remainder of the experiment. The two protein solutions were irradiated with UV light for 4 hours at room temperature. Aliquots (20 μL) were removed from the reaction mixtures at hourly intervals during the course of the reaction and were analysed by SDS-PAGE (Figure 63).

### 18.4. Deuterium oxide

WT *Ss*βG was buffer exchanged into 50 mM sodium phosphate in $D_2O$ (pH 6.5); The protein (400 μL, 10 mg/mL, 50 mM sodium phosphate in $H_2O$, pH 6.5) was added to a spin concentrator (10 kDa MWCO) and was reduced to ~ 50 μL. $D_2O$ (1 mL) was added and the solution was reduced to ~ 50 μL. This was repeated a further two times.

The solution was transferred to a microcentrifuge tube, was made up to 300 μL with 50 mM sodium phosphate in $D_2O$ (pH 6.5) and was incubated overnight at 4°C. Then WT *Ss*βG (100 μL, 14 mg/mL, 50 mM sodium phosphate in $D_2O$, pH 6.5) and 50 mM sodium phosphate in $D_2O$ (pH 6.5) (233 μL) were added to a cuvette. To another cuvette was added WT *Ss*βG (60 μL, 10 mg/mL, 50 mM sodium phosphate in $H_2O$, pH 6.5) and 50 mM sodium phosphate in $H_2O$ (135 μL). The two solutions were irradiated with UV light and aliquots (20 μL) were removed from the reaction mixtures at hourly intervals during the course of the reaction for analysis by SDS-PAGE (Figure 64).

### 18.5. Singlet oxygen generation by TMPyP

WT *Ss*βG (400 μL, 1 mg/mL, 50 mM sodium phosphate in $H_2O$, pH 6.5) with and without the addition of TMPyP (100 μL, 10 mg/mL in $H_2O$) was irradiated with laser light (390 nm). Aliquots (20 μL) were removed from the reaction mixtures at hourly intervals during the course of the reaction and were analysed by SDS-PAGE (Figure 65).

## 19. Effect of substrate, metal and EDTA

WT *Ss*βG (200 μL, 1.0 mg/mL, 50 mM sodium phosphate, pH 6.5) and lactose (10 eq), manganese(II) chloride (10 eq), iron(II) chloride tetrahydrate (10 eq), iron(III) chloride hexahydrate (10 eq) or EDTA (10 eq) were added to 6 cuvettes. The solutions were irradiated with UV light for 2 hours at room temperature. Aliquots (20 μL) were removed from the reaction mixtures at hourly intervals during the course of the reaction and were analysed by SDS-PAGE (Figure 66).

## 20. SPASM

A search for similar motifs using an online SPASM server (http://portray.bmc.uu.se/ci-bin/spasm/scripts/spasm.pl)[341] in default mode and the full library was performed on residues H150-W151-P152 from WT *Ss*βG (PDB ID: 1GOW). Hits generated with a RMSD <1.1 are shown in Table 13. The root mean square deviation (RMSD) is a measure of the difference between values predicted by the model and the actual values of the HWP sequence being modelled.

## 21. Phylogenetic analysis

### 21.1. Archaic GH1 proteins

The evolutionary relationships of 49 archaic GH1 proteins obtained from the CAZy database (http://www.cazy.org/fam/GH1.html),[352] are shown in Figure 124, and is also available to view and download at http://www.cbrg.ox.ac.uk/data.shtml. It is recommended that the file is viewed in Dendroscope.[357] The phylogenetic relationship was inferred using the Neighbour-Joining method.[410] The evolutionary distances (shown next to the branches) were computed using the Dayhoff matrix based method[411] and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 228 positions in the final dataset. Phylogenetic visualisation was conducted in Molecular Evolutionary Genetics Analysis 4 (MEGA4).[412]

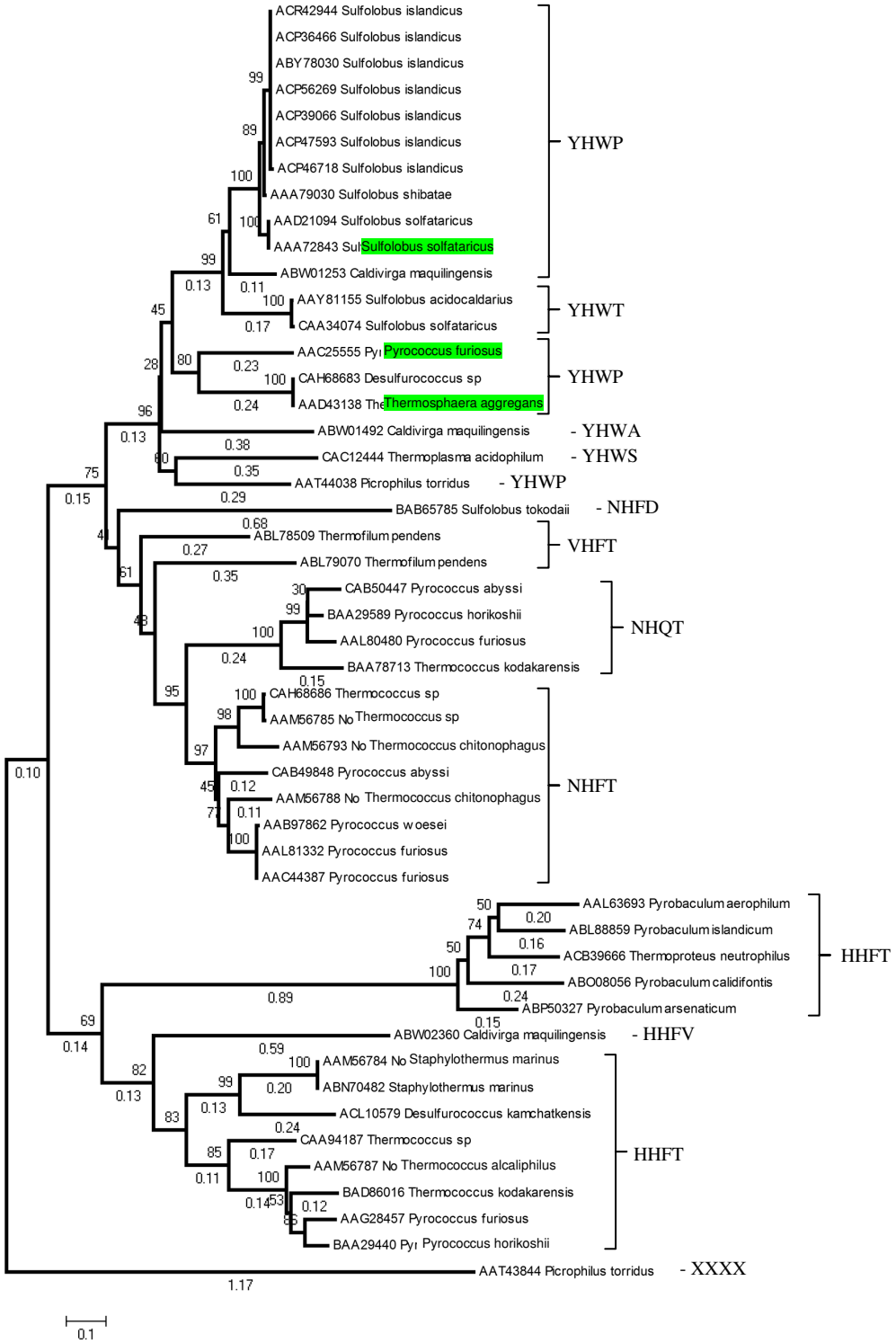
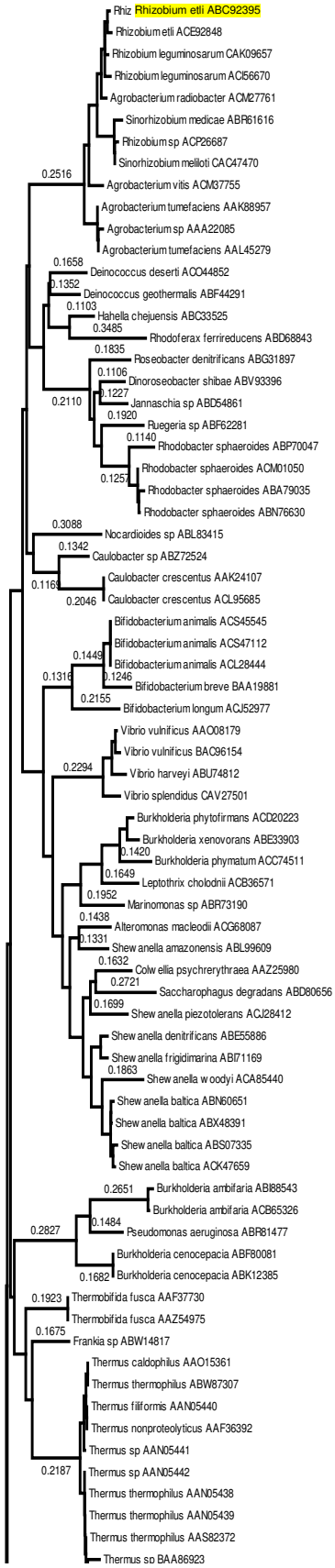
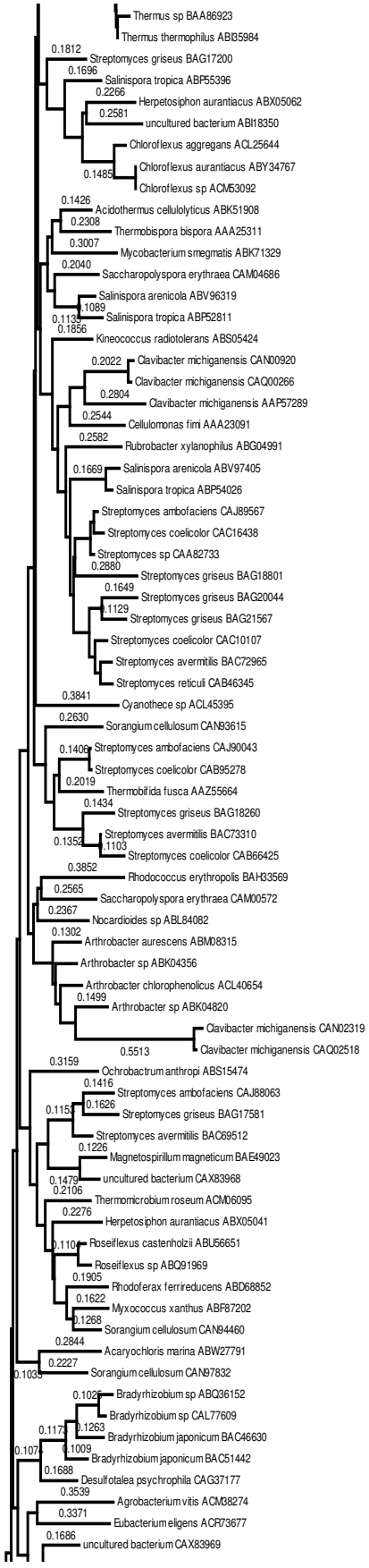


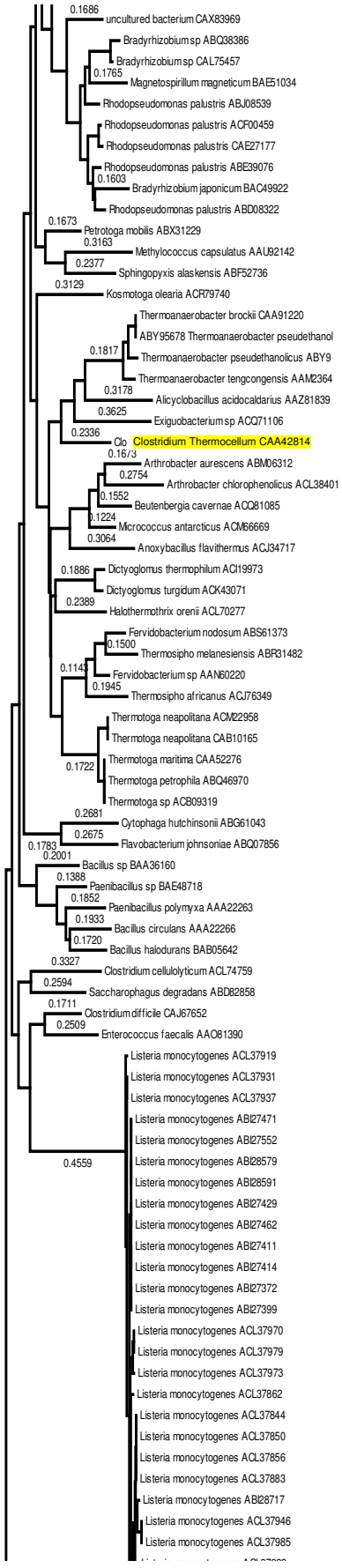

**Figure 124.** Phylogenetic analysis of 49 archaic GH1 proteins ( = cleaving proteins).

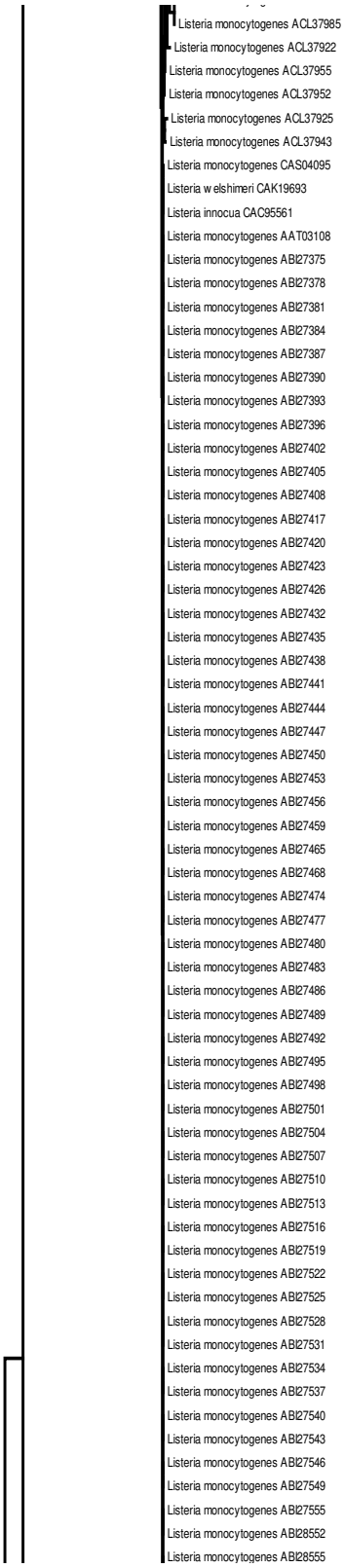
## 21.2. Bacterial GH1 proteins

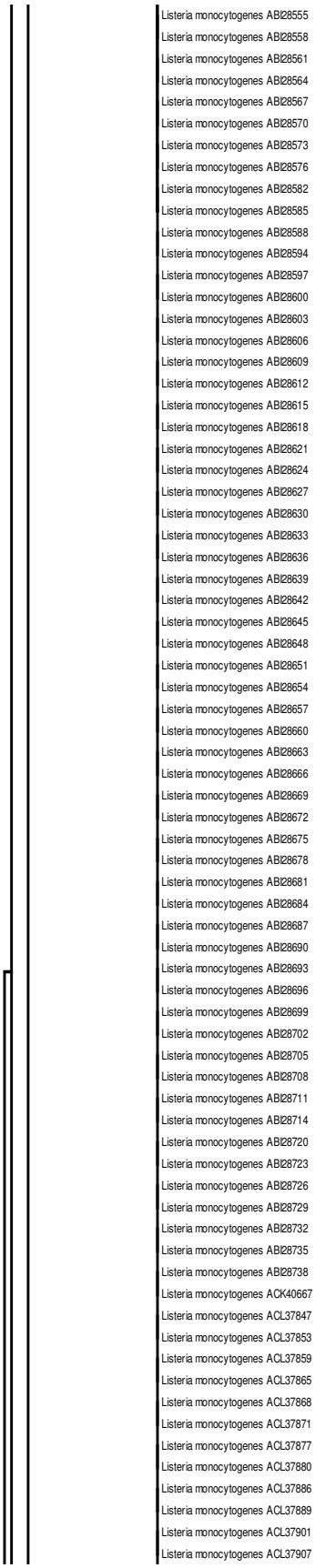
The evolutionary relationships of 1129 bacterial GH1 proteins obtained from the CAZy database (http://www.cazy.org/fam/GH1.html),[352] are shown in Figure 125, and is also available to view and download at http://www.cbrg.ox.ac.uk/data.shtml. It is recommended that the file is viewed in Dendroscope.[357] The phylogenetic relationship was inferred using the Neighbour-Joining method.[410] The evolutionary distances between the proteins were computed using the Dayhoff matrix based method[411] and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were excluded from the analysis, which left 157 gapless positions for the analysis. Phylogenetic visualisation was conducted in MEGA4.[412]

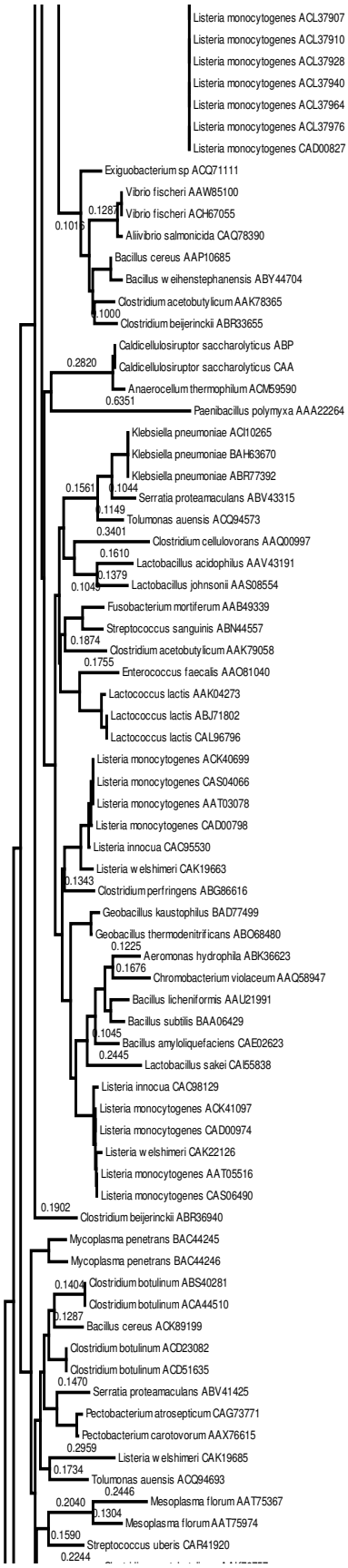

Rhiz Rhizobium etli ABC92395
Rhizobium etli ACE92848
Rhizobium leguminosarum CAK09657
Rhizobium leguminosarum ACI56670
Agrobacterium radiobacter ACM27761
Sinorhizobium medicae ABR61616
Rhizobium sp ACP26687
Sinorhizobium meliloti CAC47470
0.2516
Agrobacterium vitis ACM37755
Agrobacterium tumefaciens AAK88957
Agrobacterium sp AAA22085
Agrobacterium tumefaciens AAL45279
0.1658
Deinococcus deserti ACO44852
0.1352
Deinococcus geothermalis ABF44291
0.1103
Hahella chejuensis ABC33525
0.3485
Rhodoferax ferrireducens ABD68843
0.1835
Roseobacter denitrificans ABG31897
0.1106
Dinoroseobacter shibae ABV93396
0.1227
Jannaschia sp ABD54861
0.2110
0.1920
Ruegeria sp ABF62281
0.1140
Rhodobacter sphaeroides ABP70047
Rhodobacter sphaeroides ACM01050
0.1257
Rhodobacter sphaeroides ABA79035
Rhodobacter sphaeroides ABN76630
0.3088
Nocardioides sp ABL83415
0.1342
Caulobacter sp ABZ72524
Caulobacter crescentus AAK24107
0.1168
0.2046
Caulobacter crescentus ACL95685
Bifidobacterium animalis ACS45545
Bifidobacterium animalis ACS47112
0.1449
Bifidobacterium animalis ACL28444
0.1316
0.1246
Bifidobacterium breve BAA19881
0.2155
Bifidobacterium longum ACJ52977
Vibrio vulnificus AAO08179
Vibrio vulnificus BAC96154
0.2294
Vibrio harveyi ABU74812
Vibrio splendidus CAV27501
Burkholderia phytofirmans ACD20223
Burkholderia xenovorans ABE33903
0.1420
Burkholderia phymatum ACC74511
0.1649
Leptothrix cholodnii ACB36571
0.1952
Marinomonas sp ABR73190
0.1438
Alteromonas macleodii ACG68087
0.1331
Shewanella amazonensis ABL99609
0.1632
Colwellia psychrerythraea AAZ25980
0.2721
Saccharophagus degradans ABD80656
0.1699
Shewanella piezotolerans ACJ28412
Shewanella denitrificans ABE55886
Shewanella frigidimarina ABI71169
0.1863
Shewanella woodyi ACA85440
Shewanella baltica ABN60651
Shewanella baltica ABX48391
Shewanella baltica ABS07335
Shewanella baltica ACK47659
0.2651
Burkholderia ambifaria ABI88543
Burkholderia ambifaria ACB65326
0.1484
0.2827
Pseudomonas aeruginosa ABR81477
Burkholderia cenocepacia ABF80081
0.1682
Burkholderia cenocepacia ABK12385
0.1923
Thermobifida fusca AAF37730
Thermobifida fusca AAZ54975
0.1675
Frankia sp ABW14817
Thermus caldophilus AAO15361
Thermus thermophilus ABW87307
Thermus filiformis AAN05440
Thermus nonproteolyticus AAF36392
Thermus sp AAN05441
0.2187
Thermus sp AAN05442
Thermus thermophilus AAN05438
Thermus thermophilus AAN05439
Thermus thermophilus AAS82372
Thermus sp BAA86923

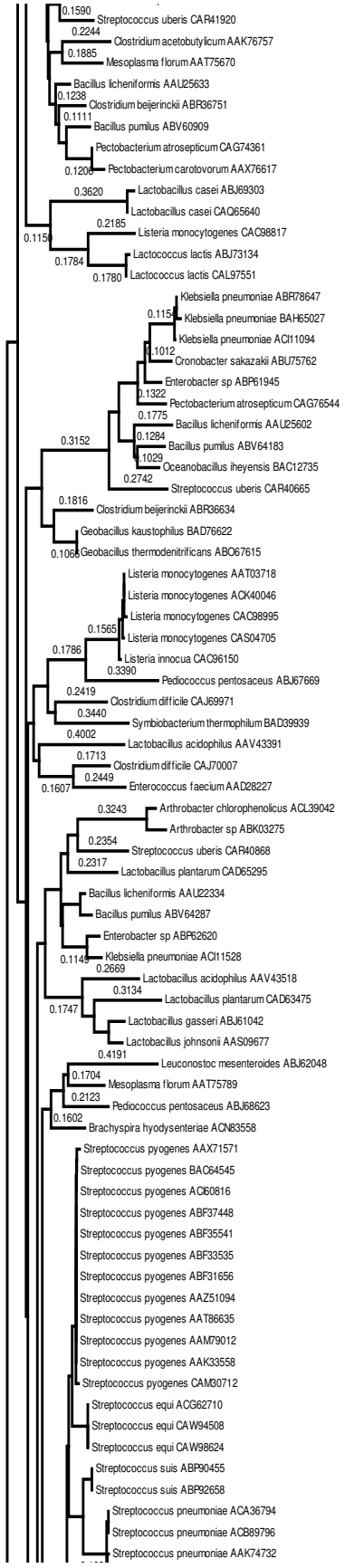

Thermus sp BAA86923
Thermus thermophilus ABI35984
0.1812
Streptomyces griseus BAG17200
0.1696
Salinispora tropica ABP55396
0.2266
Herpetosiphon aurantiacus ABX05062
0.2581
uncultured bacterium ABI18350
Chloroflexus aggregans ACL25644
0.1485
Chloroflexus aurantiacus ABY34767
Chloroflexus sp ACM53092
0.1426
Acidothermus cellulolyticus ABK51908
0.2308
Thermobispora bispora AAA25311
0.3007
Mycobacterium smegmatis ABK71329
0.2040
Saccharopolyspora erythraea CAM04686
Salinispora arenicola ABV96319
.1089
0.113
Salinispora tropica ABP52811
0.1856
Kineococcus radiotolerans ABS05424
0.2022
Clavibacter michiganensis CAN00920
Clavibacter michiganensis CAQ00266
0.2804
Clavibacter michiganensis AAP57289
0.2544
Cellulomonas fimi AAA23091
0.2582
Rubrobacter xylanophilus ABG04991
0.1669
Salinispora arenicola ABV97405
Salinispora tropica ABP54026
Streptomyces ambofaciens CAJ89567
Streptomyces coelicolor CAC16438
Streptomyces sp CAA82733
0.2880
Streptomyces griseus BAG18801
0.1649
Streptomyces griseus BAG20044
.1129
Streptomyces griseus BAG21567
Streptomyces coelicolor CAC10107
Streptomyces avermitilis BAC72965
Streptomyces reticuli CAB46345
0.3841
Cyanothece sp ACL45395
0.2630
Sorangium cellulosum CAN93615
0.1406
Streptomyces ambofaciens CAJ90043
Streptomyces coelicolor CAB95278
0.2019
Thermobifida fusca AAZ55664
0.1434
Streptomyces griseus BAG18260
Streptomyces avermitilis BAC73310
0.1352
.1103
Streptomyces coelicolor CAB66425
0.3852
Rhodococcus erythropolis BAH33569
0.2565
Saccharopolyspora erythraea CAM00572
0.2367
Nocardioides sp ABL84082
0.1302
Arthrobacter aurescens ABM08315
Arthrobacter sp ABK04356
Arthrobacter chlorophenolicus ACL40654
0.1499
Arthrobacter sp ABK04820
Clavibacter michiganensis CAN02319
0.5513
Clavibacter michiganensis CAQ02518
0.3159
Ochrobactrum anthropi ABS15474
0.1416
Streptomyces ambofaciens CAJ88063
0.1626
Streptomyces griseus BAG17581
0.115
Streptomyces avermitilis BAC69512
0.1226
Magnetospirillum magneticum BAE49023
0.1479
uncultured bacterium CAX83968
0.2106
Thermomicrobium roseum ACM06095
0.2276
Herpetosiphon aurantiacus ABX05041
0.110
Roseiflexus castenholzii ABU56651
Roseiflexus sp ABQ91969
0.1905
Rhodoferax ferrireducens ABD68852
0.1622
Myxococcus xanthus ABF87202
0.1268
Sorangium cellulosum CAN94460
0.2844
Acaryochloris marina ABW27791
0.2227
0.1035
Sorangium cellulosum CAN97832
0.1025
Bradyrhizobium sp ABQ36152
Bradyrhizobium sp CAL77609
0.1263
0.117
Bradyrhizobium japonicum BAC46630
0.107
0.1009
Bradyrhizobium japonicum BAC51442
0.1688
Desulfotalea psychrophila CAG37177
0.3539
Agrobacterium vitis ACM38274
0.3371
Eubacterium eligens ACR73677
0.1686
uncultured bacterium CAX83969

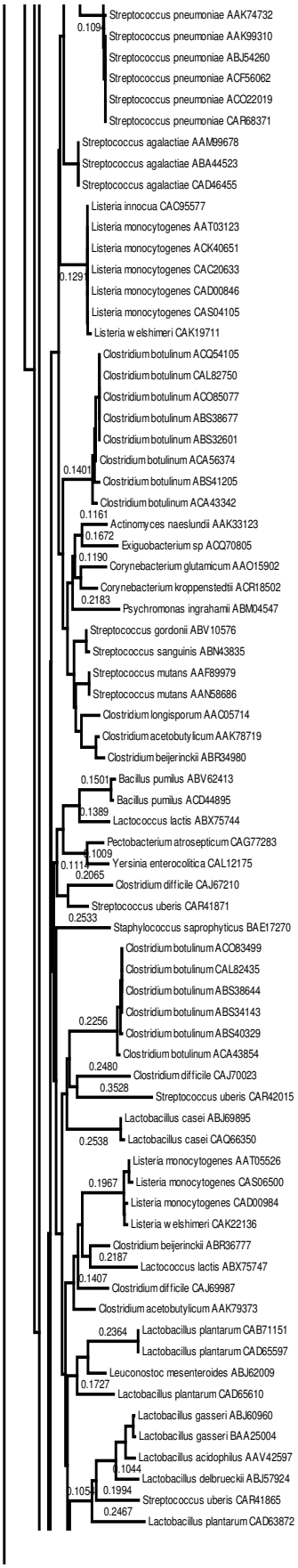

0.1686
uncultured bacterium CAX83969
Bradyrhizobium sp ABQ38386
Bradyrhizobium sp CAL75457
0.1765
Magnetospirillum magneticum BAE51034
Rhodopseudomonas palustris ABJ08539
Rhodopseudomonas palustris ACF00459
Rhodopseudomonas palustris CAE27177
Rhodopseudomonas palustris ABE39076
0.1603
Bradyrhizobium japonicum BAC49922
Rhodopseudomonas palustris ABD08322
0.1673
Petrotoga mobilis ABX31229
0.3163
Methylococcus capsulatus AAU92142
0.2377
Sphingopyxis alaskensis ABF52736
0.3129
Kosmotoga olearia ACR79740
Thermoanaerobacter brockii CAA91220
ABY95678 Thermoanaerobacter pseudethanol
0.1817
Thermoanaerobacter pseudethanolicus ABY9
Thermoanaerobacter tengcongensis AAM2364
0.3178
Alicyclobacillus acidocaldarius AAZ81839
0.3625
Exiguobacterium sp ACQ71106
0.2336
Clo Clostridium Thermocellum CAA42814
0.1673
Arthrobacter aurescens ABM06312
0.2754
Arthrobacter chlorophenolicus ACL38401
0.1552
Beutenbergia cavernae ACQ81085
0.1224
Micrococcus antarcticus ACM66669
0.3064
Anoxybacillus flavithermus ACJ34717
0.1886
Dictyoglomus thermophilum ACI19973
Dictyoglomus turgidum ACK43071
0.2389
Halothermothrix orenii ACL70277
Fervidobacterium nodosum ABS61373
0.1500
Thermosipho melanesiensis ABR31482
0.1143
Fervidobacterium sp AAN60220
0.1945
Thermosipho africanus ACJ76349
Thermotoga neapolitana ACM22958
Thermotoga neapolitana CAB10165
0.1722
Thermotoga maritima CAA52276
Thermotoga petrophila ABQ46970
Thermotoga sp ACB09319
0.2681
Cytophaga hutchinsonii ABG61043
0.2675
Flavobacterium johnsoniae ABQ07856
0.1783
0.2001
Bacillus sp BAA36160
0.1388
Paenibacillus sp BAE48718
0.1852
Paenibacillus polymyxa AAA22263
0.1933
Bacillus circulans AAA22266
0.1720
Bacillus halodurans BAB05642
0.3327
Clostridium cellulolyticum ACL74759
0.2594
Saccharophagus degradans ABD82858
0.1711
Clostridium difficile CAJ67652
0.2509
Enterococcus faecalis AAO81390
Listeria monocytogenes ACL37919
Listeria monocytogenes ACL37931
Listeria monocytogenes ACL37937
Listeria monocytogenes ABI27471
Listeria monocytogenes ABI27552
0.4559
Listeria monocytogenes ABI28579
Listeria monocytogenes ABI28591
Listeria monocytogenes ABI27429
Listeria monocytogenes ABI27462
Listeria monocytogenes ABI27411
Listeria monocytogenes ABI27414
Listeria monocytogenes ABI27372
Listeria monocytogenes ABI27399
Listeria monocytogenes ACL37970
Listeria monocytogenes ACL37979
Listeria monocytogenes ACL37973
Listeria monocytogenes ACL37862
Listeria monocytogenes ACL37844
Listeria monocytogenes ACL37850
Listeria monocytogenes ACL37856
Listeria monocytogenes ACL37883
Listeria monocytogenes ABI28717
Listeria monocytogenes ACL37946
Listeria monocytogenes ACL37985

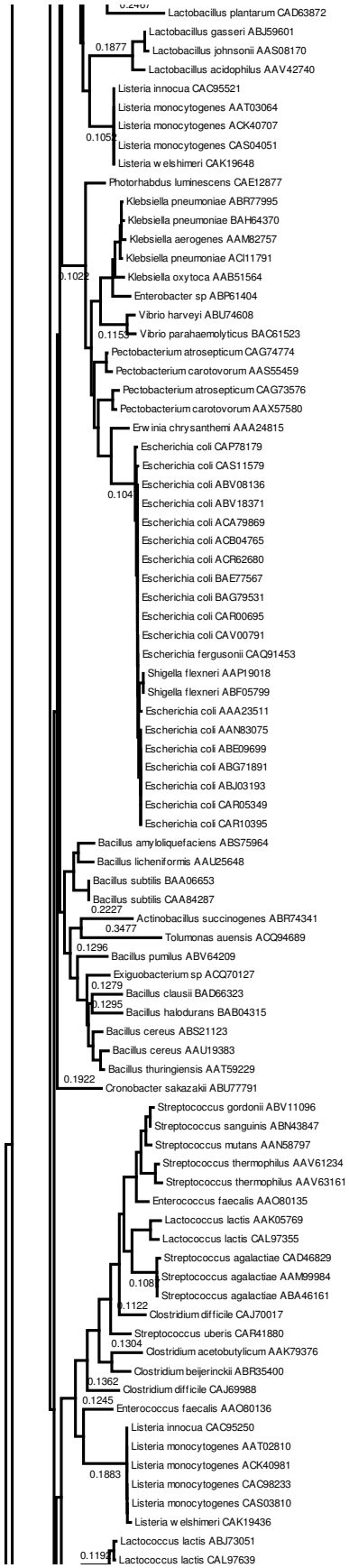

Listeria monocytogenes ACL37985
Listeria monocytogenes ACL37922
Listeria monocytogenes ACL37955
Listeria monocytogenes ACL37952
Listeria monocytogenes ACL37925
Listeria monocytogenes ACL37943
Listeria monocytogenes CAS04095
Listeria w elshimeri CAK19693
Listeria innocua CAC95561
Listeria monocytogenes AAT03108
Listeria monocytogenes ABI27375
Listeria monocytogenes ABI27378
Listeria monocytogenes ABI27381
Listeria monocytogenes ABI27384
Listeria monocytogenes ABI27387
Listeria monocytogenes ABI27390
Listeria monocytogenes ABI27393
Listeria monocytogenes ABI27396
Listeria monocytogenes ABI27402
Listeria monocytogenes ABI27405
Listeria monocytogenes ABI27408
Listeria monocytogenes ABI27417
Listeria monocytogenes ABI27420
Listeria monocytogenes ABI27423
Listeria monocytogenes ABI27426
Listeria monocytogenes ABI27432
Listeria monocytogenes ABI27435
Listeria monocytogenes ABI27438
Listeria monocytogenes ABI27441
Listeria monocytogenes ABI27444
Listeria monocytogenes ABI27447
Listeria monocytogenes ABI27450
Listeria monocytogenes ABI27453
Listeria monocytogenes ABI27456
Listeria monocytogenes ABI27459
Listeria monocytogenes ABI27465
Listeria monocytogenes ABI27468
Listeria monocytogenes ABI27474
Listeria monocytogenes ABI27477
Listeria monocytogenes ABI27480
Listeria monocytogenes ABI27483
Listeria monocytogenes ABI27486
Listeria monocytogenes ABI27489
Listeria monocytogenes ABI27492
Listeria monocytogenes ABI27495
Listeria monocytogenes ABI27498
Listeria monocytogenes ABI27501
Listeria monocytogenes ABI27504
Listeria monocytogenes ABI27507
Listeria monocytogenes ABI27510
Listeria monocytogenes ABI27513
Listeria monocytogenes ABI27516
Listeria monocytogenes ABI27519
Listeria monocytogenes ABI27522
Listeria monocytogenes ABI27525
Listeria monocytogenes ABI27528
Listeria monocytogenes ABI27531
Listeria monocytogenes ABI27534
Listeria monocytogenes ABI27537
Listeria monocytogenes ABI27540
Listeria monocytogenes ABI27543
Listeria monocytogenes ABI27546
Listeria monocytogenes ABI27549
Listeria monocytogenes ABI27555
Listeria monocytogenes ABI28552
Listeria monocytogenes ABI28555

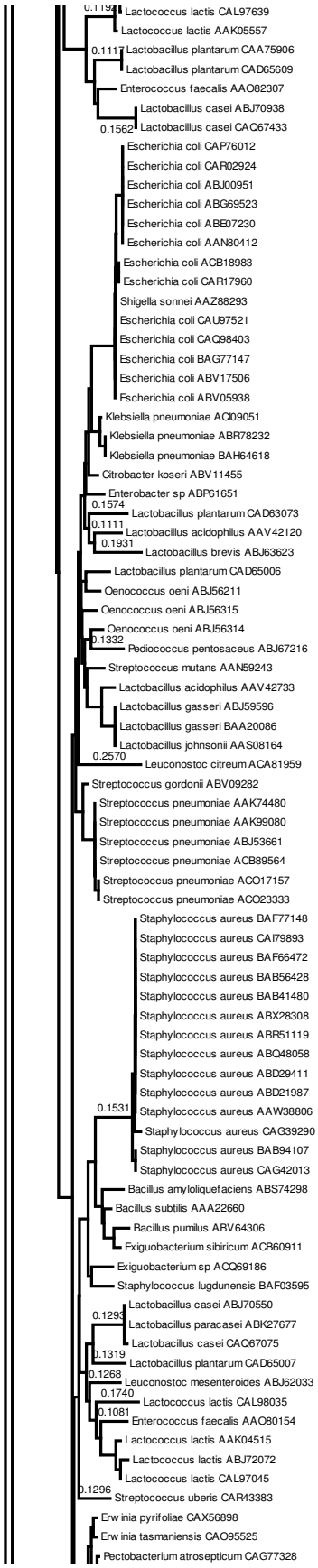

Listeria monocytogenes ABI28555
Listeria monocytogenes ABI28558
Listeria monocytogenes ABI28561
Listeria monocytogenes ABI28564
Listeria monocytogenes ABI28567
Listeria monocytogenes ABI28570
Listeria monocytogenes ABI28573
Listeria monocytogenes ABI28576
Listeria monocytogenes ABI28582
Listeria monocytogenes ABI28585
Listeria monocytogenes ABI28588
Listeria monocytogenes ABI28594
Listeria monocytogenes ABI28597
Listeria monocytogenes ABI28600
Listeria monocytogenes ABI28603
Listeria monocytogenes ABI28606
Listeria monocytogenes ABI28609
Listeria monocytogenes ABI28612
Listeria monocytogenes ABI28615
Listeria monocytogenes ABI28618
Listeria monocytogenes ABI28621
Listeria monocytogenes ABI28624
Listeria monocytogenes ABI28627
Listeria monocytogenes ABI28630
Listeria monocytogenes ABI28633
Listeria monocytogenes ABI28636
Listeria monocytogenes ABI28639
Listeria monocytogenes ABI28642
Listeria monocytogenes ABI28645
Listeria monocytogenes ABI28648
Listeria monocytogenes ABI28651
Listeria monocytogenes ABI28654
Listeria monocytogenes ABI28657
Listeria monocytogenes ABI28660
Listeria monocytogenes ABI28663
Listeria monocytogenes ABI28666
Listeria monocytogenes ABI28669
Listeria monocytogenes ABI28672
Listeria monocytogenes ABI28675
Listeria monocytogenes ABI28678
Listeria monocytogenes ABI28681
Listeria monocytogenes ABI28684
Listeria monocytogenes ABI28687
Listeria monocytogenes ABI28690
Listeria monocytogenes ABI28693
Listeria monocytogenes ABI28696
Listeria monocytogenes ABI28699
Listeria monocytogenes ABI28702
Listeria monocytogenes ABI28705
Listeria monocytogenes ABI28708
Listeria monocytogenes ABI28711
Listeria monocytogenes ABI28714
Listeria monocytogenes ABI28720
Listeria monocytogenes ABI28723
Listeria monocytogenes ABI28726
Listeria monocytogenes ABI28729
Listeria monocytogenes ABI28732
Listeria monocytogenes ABI28735
Listeria monocytogenes ABI28738
Listeria monocytogenes ACK40667
Listeria monocytogenes ACL37847
Listeria monocytogenes ACL37853
Listeria monocytogenes ACL37859
Listeria monocytogenes ACL37865
Listeria monocytogenes ACL37868
Listeria monocytogenes ACL37871
Listeria monocytogenes ACL37877
Listeria monocytogenes ACL37880
Listeria monocytogenes ACL37886
Listeria monocytogenes ACL37889
Listeria monocytogenes ACL37901
Listeria monocytogenes ACL37907

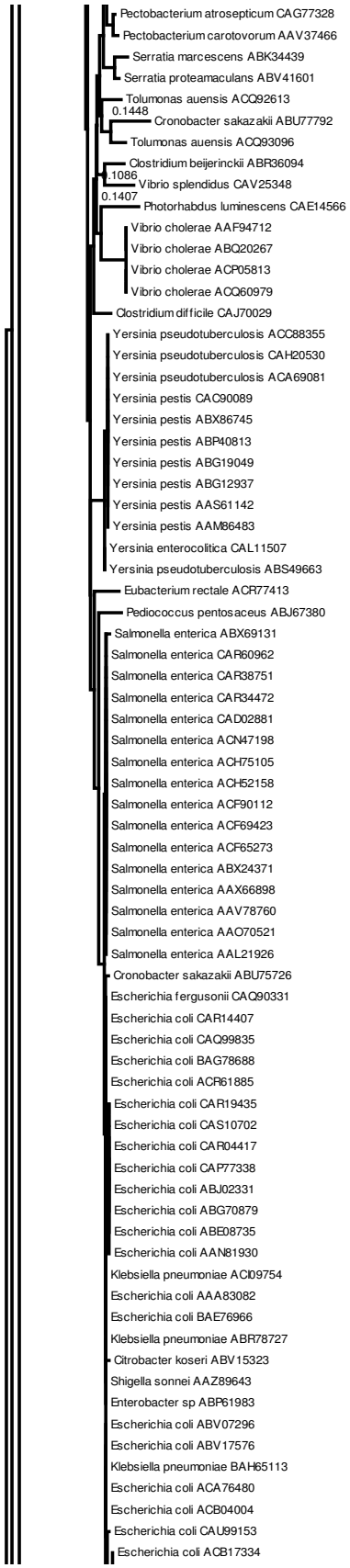

Listeria monocytogenes ACL37907
Listeria monocytogenes ACL37910
Listeria monocytogenes ACL37928
Listeria monocytogenes ACL37940
Listeria monocytogenes ACL37964
Listeria monocytogenes ACL37976
Listeria monocytogenes CAD00827
Exiguobacterium sp ACQ71111
Vibrio fischeri AAW85100
0.1287
Vibrio fischeri ACH67055
0.1016
Aliivibrio salmonicida CAQ78390
Bacillus cereus AAP10685
Bacillus weihenstephanensis ABY44704
Clostridium acetobutylicum AAK78365
0.1000
Clostridium beijerinckii ABR33655
Caldicellulosiruptor saccharolyticus ABP
0.2820
Caldicellulosiruptor saccharolyticus CAA
Anaerocellum thermophilum ACM59590
0.6351
Paenibacillus polymyxa AAA22264
Klebsiella pneumoniae ACI10265
Klebsiella pneumoniae BAH63670
Klebsiella pneumoniae ABR77392
0.1561
0.1044
Serratia proteamaculans ABV43315
0.1149
Tolumonas auensis ACQ94573
0.3401
Clostridium cellulovorans AAQ00997
0.1610
Lactobacillus acidophilus AAV43191
0.1379
0.1045
Lactobacillus johnsonii AAS08554
Fusobacterium mortiferum AAB49339
Streptococcus sanguinis ABN44557
0.1874
Clostridium acetobutylicum AAK79058
0.1755
Enterococcus faecalis AAO81040
Lactococcus lactis AAK04273
Lactococcus lactis ABJ71802
Lactococcus lactis CAL96796
Listeria monocytogenes ACK40699
Listeria monocytogenes CAS04066
Listeria monocytogenes AAT03078
Listeria monocytogenes CAD00798
Listeria innocua CAC95530
Listeria welshimeri CAK19663
0.1343
Clostridium perfringens ABG86616
Geobacillus kaustophilus BAD77499
Geobacillus thermodenitrificans ABO68480
0.1225
Aeromonas hydrophila ABK36623
0.1676
Chromobacterium violaceum AAQ58947
Bacillus licheniformis AAU21991
Bacillus subtilis BAA06429
0.1045
Bacillus amyloliquefaciens CAE02623
0.2445
Lactobacillus sakei CAI55838
Listeria innocua CAC98129
Listeria monocytogenes ACK41097
Listeria monocytogenes CAD00974
Listeria welshimeri CAK22126
Listeria monocytogenes AAT05516
Listeria monocytogenes CAS06490
0.1902
Clostridium beijerinckii ABR36940
Mycoplasma penetrans BAC44245
Mycoplasma penetrans BAC44246
0.1404
Clostridium botulinum ABS40281
Clostridium botulinum ACA44510
0.1287
Bacillus cereus ACK89199
Clostridium botulinum ACD23082
Clostridium botulinum ACD51635
0.1470
Serratia proteamaculans ABV41425
Pectobacterium atrosepticum CAG73771
Pectobacterium carotovorum AAX76615
0.2959
Listeria welshimeri CAK19685
0.1734
Tolumonas auensis ACQ94693
0.2446
0.2040
Mesoplasma florum AAT75367
0.1304
Mesoplasma florum AAT75974
0.1590
Streptococcus uberis CAR41920
0.2244

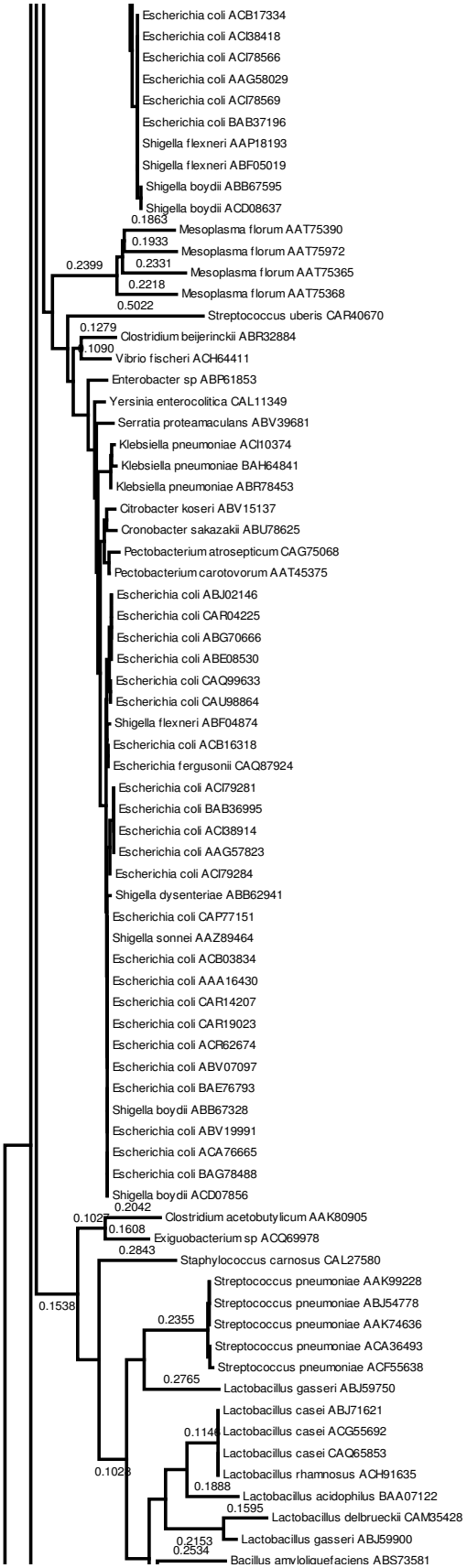

0.1590
Streptococcus uberis CAR41920
0.2244
Clostridium acetobutylicum AAK76757
0.1885
Mesoplasma florum AAT75670
Bacillus licheniformis AAU25633
0.1238
Clostridium beijerinckii ABR36751
0.1111
Bacillus pumilus ABV60909
Pectobacterium atrosepticum CAG74361
0.1206
Pectobacterium carotovorum AAX76617
0.3620
Lactobacillus casei ABJ69303
Lactobacillus casei CAQ65640
0.2185
Listeria monocytogenes CAC98817
0.1150
Lactococcus lactis ABJ73134
0.1784
0.1780
Lactococcus lactis CAL97551
Klebsiella pneumoniae ABR78647
0.1154
Klebsiella pneumoniae BAH65027
Klebsiella pneumoniae ACI11094
0.1012
Cronobacter sakazakii ABU75762
Enterobacter sp ABP61945
0.1322
Pectobacterium atrosepticum CAG76544
0.1775
Bacillus licheniformis AAU25602
0.1284
Bacillus pumilus ABV64183
0.3152
0.1029
Oceanobacillus iheyensis BAC12735
0.2742
Streptococcus uberis CAR40665
0.1816
Clostridium beijerinckii ABR36634
Geobacillus kaustophilus BAD76622
0.1066
Geobacillus thermodenitrificans ABO67615
Listeria monocytogenes AAT03718
Listeria monocytogenes ACK40046
Listeria monocytogenes CAC98995
0.1565
Listeria monocytogenes CAS04705
0.1786
Listeria innocua CAC96150
0.3390
Pediococcus pentosaceus ABJ67669
0.2419
Clostridium difficile CAJ69971
0.3440
Symbiobacterium thermophilum BAD39939
0.4002
Lactobacillus acidophilus AAV43391
0.1713
Clostridium difficile CAJ70007
0.1607
0.2449
Enterococcus faecium AAD28227
0.3243
Arthrobacter chlorophenolicus ACL39042
Arthrobacter sp ABK03275
0.2354
Streptococcus uberis CAR40868
0.2317
Lactobacillus plantarum CAD65295
Bacillus licheniformis AAU22334
Bacillus pumilus ABV64287
Enterobacter sp ABP62620
0.1149
Klebsiella pneumoniae ACI11528
0.2669
Lactobacillus acidophilus AAV43518
0.3134
Lactobacillus plantarum CAD63475
0.1747
Lactobacillus gasseri ABJ61042
Lactobacillus johnsonii AAS09677
0.4191
Leuconostoc mesenteroides ABJ62048
0.1704
Mesoplasma florum AAT75789
0.2123
Pediococcus pentosaceus ABJ68623
0.1602
Brachyspira hyodysenteriae ACN83558
Streptococcus pyogenes AAX71571
Streptococcus pyogenes BAC64545
Streptococcus pyogenes ACI60816
Streptococcus pyogenes ABF37448
Streptococcus pyogenes ABF35541
Streptococcus pyogenes ABF33535
Streptococcus pyogenes ABF31656
Streptococcus pyogenes AAZ51094
Streptococcus pyogenes AAT86635
Streptococcus pyogenes AAM79012
Streptococcus pyogenes AAK33558
Streptococcus pyogenes CAM30712
Streptococcus equi ACG62710
Streptococcus equi CAW94508
Streptococcus equi CAW98624
Streptococcus suis ABP90455
Streptococcus suis ABP92658
Streptococcus pneumoniae ACA36794
Streptococcus pneumoniae ACB89796
Streptococcus pneumoniae AAK74732

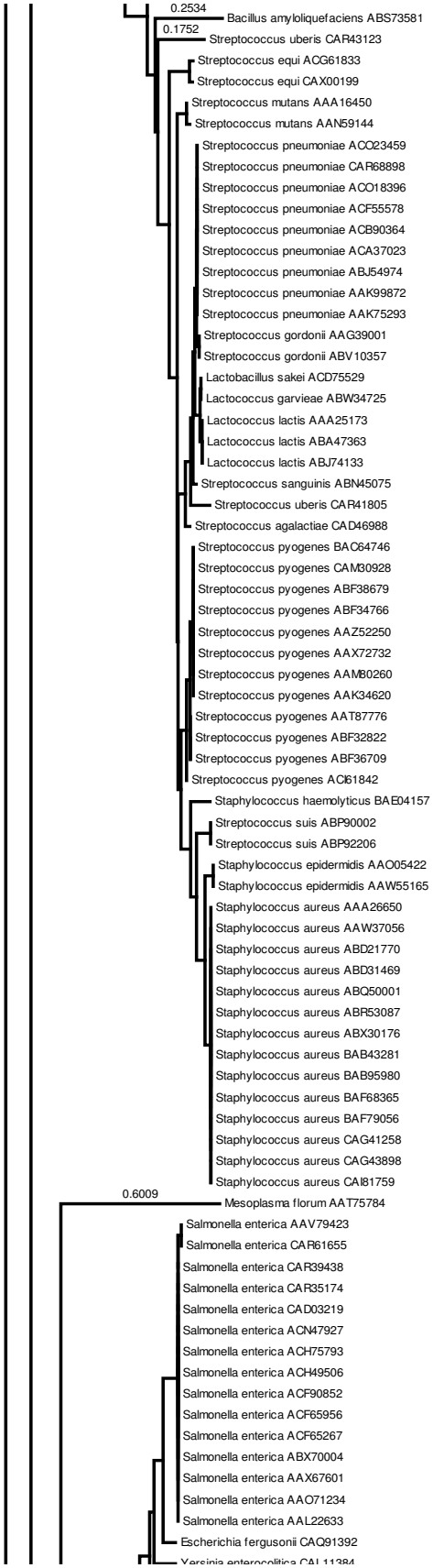

Streptococcus pneumoniae AAK74732
0.1094
Streptococcus pneumoniae AAK99310
Streptococcus pneumoniae ABJ54260
Streptococcus pneumoniae ACF56062
Streptococcus pneumoniae ACO22019
Streptococcus pneumoniae CAR68371
Streptococcus agalactiae AAM99678
Streptococcus agalactiae ABA44523
Streptococcus agalactiae CAD46455
Listeria innocua CAC95577
Listeria monocytogenes AAT03123
Listeria monocytogenes ACK40651
Listeria monocytogenes CAC20633
0.1291
Listeria monocytogenes CAD00846
Listeria monocytogenes CAS04105
Listeria welshimeri CAK19711
Clostridium botulinum ACQ54105
Clostridium botulinum CAL82750
Clostridium botulinum ACO85077
Clostridium botulinum ABS38677
Clostridium botulinum ABS32601
Clostridium botulinum ACA56374
0.1401
Clostridium botulinum ABS41205
Clostridium botulinum ACA43342
0.1161
Actinomyces naeslundii AAK33123
0.1672
Exiguobacterium sp ACQ70805
0.1190
Corynebacterium glutamicum AAO15902
Corynebacterium kroppenstedtii ACR18502
0.2183
Psychromonas ingrahamii ABM04547
Streptococcus gordonii ABV10576
Streptococcus sanguinis ABN43835
Streptococcus mutans AAF89979
Streptococcus mutans AAN58686
Clostridium longisporum AAC05714
Clostridium acetobutylicum AAK78719
Clostridium beijerinckii ABR34980
0.1501
Bacillus pumilus ABV62413
Bacillus pumilus ACD44895
0.1389
Lactococcus lactis ABX75744
Pectobacterium atrosepticum CAG77283
0.1009
0.1114
Yersinia enterocolitica CAL12175
0.2065
Clostridium difficile CAJ67210
Streptococcus uberis CAR41871
0.2533
Staphylococcus saprophyticus BAE17270
Clostridium botulinum ACO83499
Clostridium botulinum CAL82435
Clostridium botulinum ABS38644
Clostridium botulinum ABS34143
0.2256
Clostridium botulinum ABS40329
Clostridium botulinum ACA43854
0.2480
Clostridium difficile CAJ70023
0.3528
Streptococcus uberis CAR42015
Lactobacillus casei ABJ69895
Lactobacillus casei CAQ66350
0.2538
Listeria monocytogenes AAT05526
Listeria monocytogenes CAS06500
0.1967
Listeria monocytogenes CAD00984
Listeria welshimeri CAK22136
Clostridium beijerinckii ABR36777
0.2187
Lactococcus lactis ABX75747
0.1407
Clostridium difficile CAJ69987
Clostridium acetobutylicum AAK79373
0.2364
Lactobacillus plantarum CAB71151
Lactobacillus plantarum CAD65597
Leuconostoc mesenteroides ABJ62009
0.1727
Lactobacillus plantarum CAD65610
Lactobacillus gasseri ABJ60960
Lactobacillus gasseri BAA25004
Lactobacillus acidophilus AAV42597
0.1044
Lactobacillus delbrueckii ABJ57924
0.105
0.1994
Streptococcus uberis CAR41865
0.2467
Lactobacillus plantarum CAD63872

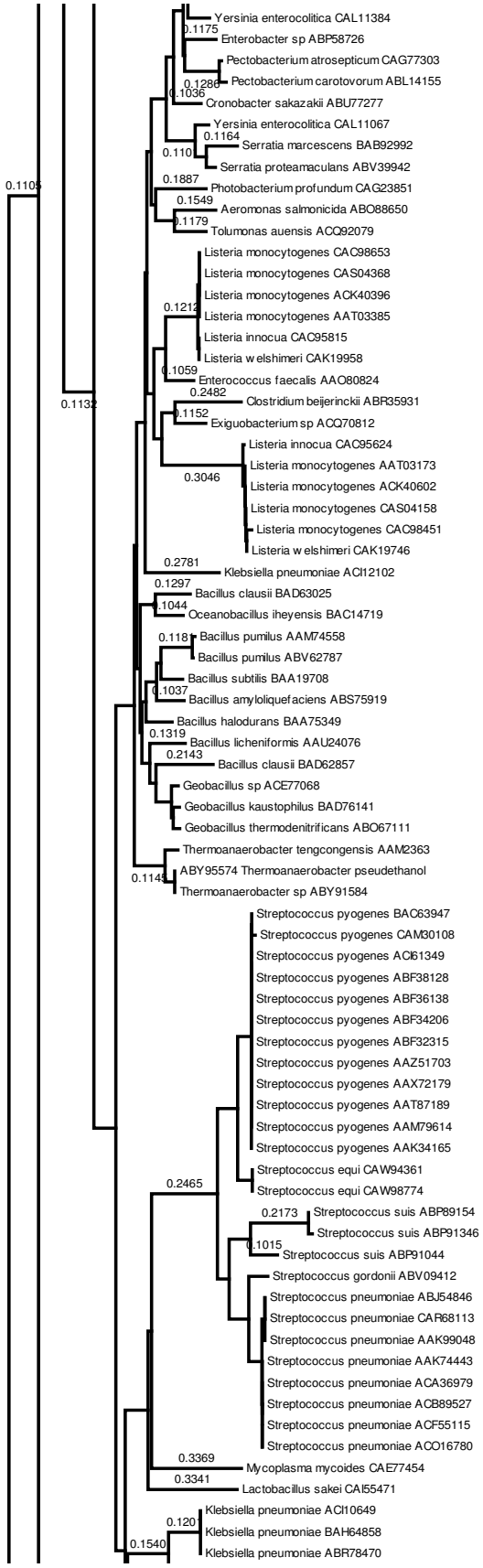

Lactobacillus plantarum CAD63872
Lactobacillus gasseri ABJ59601
0.1877
Lactobacillus johnsonii AAS08170
Lactobacillus acidophilus AAV42740
Listeria innocua CAC95521
Listeria monocytogenes AAT03064
Listeria monocytogenes ACK40707
0.1052
Listeria monocytogenes CAS04051
Listeria welshimeri CAK19648
Photorhabdus luminescens CAE12877
Klebsiella pneumoniae ABR77995
Klebsiella pneumoniae BAH64370
Klebsiella aerogenes AAM82757
Klebsiella pneumoniae ACI11791
0.1022
Klebsiella oxytoca AAB51564
Enterobacter sp ABP61404
Vibrio harveyi ABU74608
0.1153
Vibrio parahaemolyticus BAC61523
Pectobacterium atrosepticum CAG74774
Pectobacterium carotovorum AAS55459
Pectobacterium atrosepticum CAG73576
Pectobacterium carotovorum AAX57580
Erwinia chrysanthemi AAA24815
Escherichia coli CAP78179
Escherichia coli CAS11579
Escherichia coli ABV08136
0.104
Escherichia coli ABV18371
Escherichia coli ACA79869
Escherichia coli ACB04765
Escherichia coli ACR62680
Escherichia coli BAE77567
Escherichia coli BAG79531
Escherichia coli CAR00695
Escherichia coli CAV00791
Escherichia fergusonii CAQ91453
Shigella flexneri AAP19018
Shigella flexneri ABF05799
Escherichia coli AAA23511
Escherichia coli AAN83075
Escherichia coli ABE09699
Escherichia coli ABG71891
Escherichia coli ABJ03193
Escherichia coli CAR05349
Escherichia coli CAR10395
Bacillus amyloliquefaciens ABS75964
Bacillus licheniformis AAU25648
Bacillus subtilis BAA06653
Bacillus subtilis CAA84287
0.2227
Actinobacillus succinogenes ABR74341
0.3477
Tolumonas auensis ACQ94689
0.1296
Bacillus pumilus ABV64209
Exiguobacterium sp ACQ70127
0.1279
Bacillus clausii BAD66323
0.1295
Bacillus halodurans BAB04315
Bacillus cereus ABS21123
Bacillus cereus AAU19383
Bacillus thuringiensis AAT59229
0.1922
Cronobacter sakazakii ABU77791
Streptococcus gordonii ABV11096
Streptococcus sanguinis ABN43847
Streptococcus mutans AAN58797
Streptococcus thermophilus AAV61234
Streptococcus thermophilus AAV63161
Enterococcus faecalis AAO80135
Lactococcus lactis AAK05769
Lactococcus lactis CAL97355
Streptococcus agalactiae CAD46829
0.1081
Streptococcus agalactiae AAM99984
Streptococcus agalactiae ABA46161
0.1122
Clostridium difficile CAJ70017
Streptococcus uberis CAR41880
0.1304
Clostridium acetobutylicum AAK79376
Clostridium beijerinckii ABR35400
0.1362
Clostridium difficile CAJ69988
0.1245
Enterococcus faecalis AAO80136
Listeria innocua CAC95250
Listeria monocytogenes AAT02810
Listeria monocytogenes ACK40981
0.1883
Listeria monocytogenes CAC98233
Listeria monocytogenes CAS03810
Listeria welshimeri CAK19436
Lactococcus lactis ABJ73051
0.1194
Lactococcus lactis CAL97639

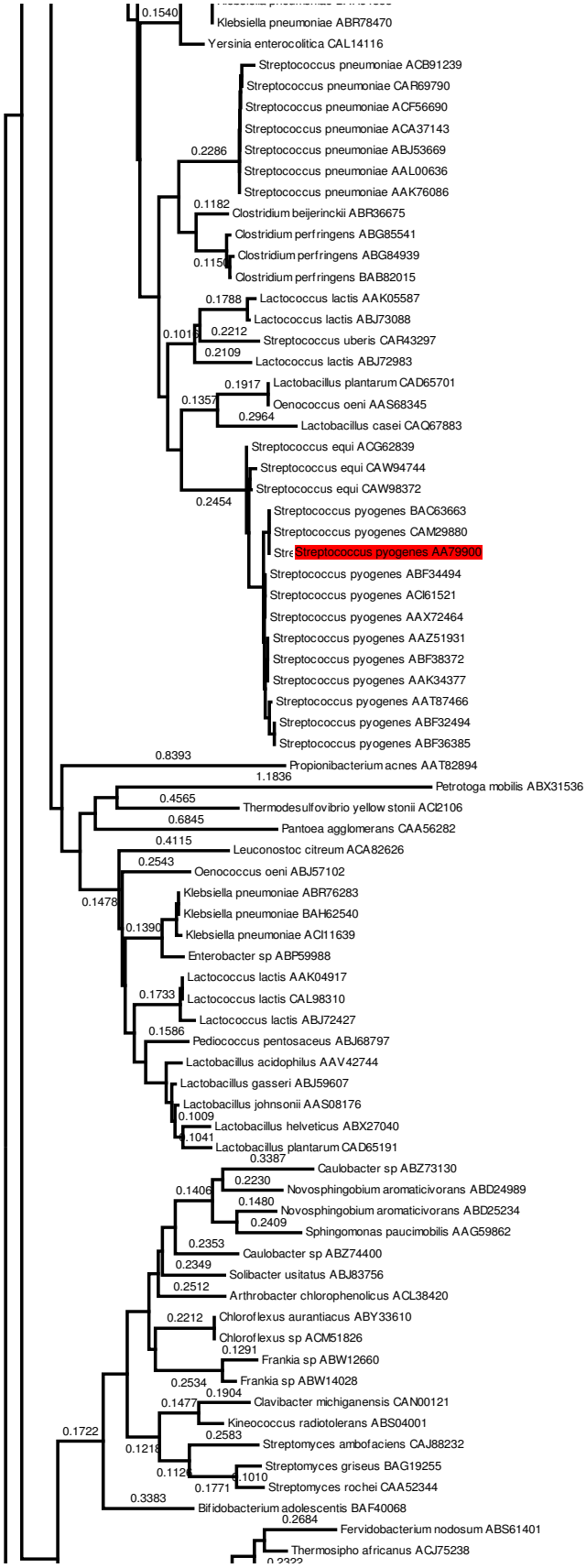

0.1198
Lactococcus lactis CAL97639
Lactococcus lactis AAK05557
0.1117
Lactobacillus plantarum CAA75906
Lactobacillus plantarum CAD65609
Enterococcus faecalis AAO82307
Lactobacillus casei ABJ70938
0.1562
Lactobacillus casei CAQ67433
Escherichia coli CAP76012
Escherichia coli CAR02924
Escherichia coli ABJ00951
Escherichia coli ABG69523
Escherichia coli ABE07230
Escherichia coli AAN80412
Escherichia coli ACB18983
Escherichia coli CAR17960
Shigella sonnei AAZ88293
Escherichia coli CAU97521
Escherichia coli CAQ98403
Escherichia coli BAG77147
Escherichia coli ABV17506
Escherichia coli ABV05938
Klebsiella pneumoniae ACI09051
Klebsiella pneumoniae ABR78232
Klebsiella pneumoniae BAH64618
Citrobacter koseri ABV11455
Enterobacter sp ABP61651
0.1574
Lactobacillus plantarum CAD63073
0.1111
Lactobacillus acidophilus AAV42120
0.1931
Lactobacillus brevis ABJ63623
Lactobacillus plantarum CAD65006
Oenococcus oeni ABJ56211
Oenococcus oeni ABJ56315
Oenococcus oeni ABJ56314
0.1332
Pediococcus pentosaceus ABJ67216
Streptococcus mutans AAN59243
Lactobacillus acidophilus AAV42733
Lactobacillus gasseri ABJ59596
Lactobacillus gasseri BAA20086
Lactobacillus johnsonii AAS08164
0.2570
Leuconostoc citreum ACA81959
Streptococcus gordonii ABV09282
Streptococcus pneumoniae AAK74480
Streptococcus pneumoniae AAK99080
Streptococcus pneumoniae ABJ53661
Streptococcus pneumoniae ACB89564
Streptococcus pneumoniae ACO17157
Streptococcus pneumoniae ACO23333
Staphylococcus aureus BAF77148
Staphylococcus aureus CAI7893
Staphylococcus aureus BAF66472
Staphylococcus aureus BAB56428
Staphylococcus aureus BAB41480
Staphylococcus aureus ABX28308
Staphylococcus aureus ABR51119
Staphylococcus aureus ABQ48058
Staphylococcus aureus ABD29411
Staphylococcus aureus ABD21987
0.1531
Staphylococcus aureus AAW38806
Staphylococcus aureus CAG39290
Staphylococcus aureus BAB94107
Staphylococcus aureus CAG42013
Bacillus amyloliquefaciens ABS74298
Bacillus subtilis AAA22660
Bacillus pumilus ABV64306
Exiguobacterium sibiricum ACB60911
Exiguobacterium sp ACQ69186
Staphylococcus lugdunensis BAF03595
Lactobacillus casei ABJ70550
0.1293
Lactobacillus paracasei ABK27677
Lactobacillus casei CAQ67075
0.1319
Lactobacillus plantarum CAD65007
0.1268
Leuconostoc mesenteroides ABJ62033
0.1740
Lactococcus lactis CAL98035
0.1081
Enterococcus faecalis AAO80154
Lactococcus lactis AAK04515
Lactococcus lactis ABJ72072
Lactococcus lactis CAL97045
0.1296
Streptococcus uberis CAR43383
Erw inia pyrifoliae CAX56898
Erw inia tasmaniensis CAO95525
Pectobacterium atrosepticum CAG77328


Pectobacterium atrosepticum CAG77328
Pectobacterium carotovorum AAV37466
Serratia marcescens ABK34439
Serratia proteamaculans ABV41601
Tolumonas auensis ACQ92613
0.1448
Cronobacter sakazakii ABU77792
Tolumonas auensis ACQ93096
Clostridium beijerinckii ABR36094
.1086
Vibrio splendidus CAV25348
0.1407
Photorhabdus luminescens CAE14566
Vibrio cholerae AAF94712
Vibrio cholerae ABQ20267
Vibrio cholerae ACP05813
Vibrio cholerae ACQ60979
Clostridium difficile CAJ70029
Yersinia pseudotuberculosis ACC88355
Yersinia pseudotuberculosis CAH20530
Yersinia pseudotuberculosis ACA69081
Yersinia pestis CAC90089
Yersinia pestis ABX86745
Yersinia pestis ABP40813
Yersinia pestis ABG19049
Yersinia pestis ABG12937
Yersinia pestis AAS61142
Yersinia pestis AAM86483
Yersinia enterocolitica CAL11507
Yersinia pseudotuberculosis ABS49663
Eubacterium rectale ACR77413
Pediococcus pentosaceus ABJ67380
Salmonella enterica ABX69131
Salmonella enterica CAR60962
Salmonella enterica CAR38751
Salmonella enterica CAR34472
Salmonella enterica CAD02881
Salmonella enterica ACN47198
Salmonella enterica ACH75105
Salmonella enterica ACH52158
Salmonella enterica ACF90112
Salmonella enterica ACF69423
Salmonella enterica ACF65273
Salmonella enterica ABX24371
Salmonella enterica AAX66898
Salmonella enterica AAV78760
Salmonella enterica AAO70521
Salmonella enterica AAL21926
Cronobacter sakazakii ABU75726
Escherichia fergusonii CAQ90331
Escherichia coli CAR14407
Escherichia coli CAQ99835
Escherichia coli BAG78688
Escherichia coli ACR61885
Escherichia coli CAR19435
Escherichia coli CAS10702
Escherichia coli CAR04417
Escherichia coli CAP77338
Escherichia coli ABJ02331
Escherichia coli ABG70879
Escherichia coli ABE08735
Escherichia coli AAN81930
Klebsiella pneumoniae ACI09754
Escherichia coli AAA83082
Escherichia coli BAE76966
Klebsiella pneumoniae ABR78727
Citrobacter koseri ABV15323
Shigella sonnei AAZ89643
Enterobacter sp ABP61983
Escherichia coli ABV07296
Escherichia coli ABV17576
Klebsiella pneumoniae BAH65113
Escherichia coli ACA76480
Escherichia coli ACB04004
Escherichia coli CAU99153
Escherichia coli ACB17334


Escherichia coli ACB17334
Escherichia coli ACI38418
Escherichia coli ACI78566
Escherichia coli AAG58029
Escherichia coli ACI78569
Escherichia coli BAB37196
Shigella flexneri AAP18193
Shigella flexneri ABF05019
Shigella boydii ABB67595
Shigella boydii ACD08637
0.1863
Mesoplasma florum AAT75390
0.1933
Mesoplasma florum AAT75972
0.2399
0.2331
Mesoplasma florum AAT75365
0.2218
Mesoplasma florum AAT75368
0.5022
Streptococcus uberis CAR40670
0.1279
Clostridium beijerinckii ABR32884
.1090
Vibrio fischeri ACH64411
Enterobacter sp ABP61853
Yersinia enterocolitica CAL11349
Serratia proteamaculans ABV39681
Klebsiella pneumoniae ACI10374
Klebsiella pneumoniae BAH64841
Klebsiella pneumoniae ABR78453
Citrobacter koseri ABV15137
Cronobacter sakazakii ABU78625
Pectobacterium atrosepticum CAG75068
Pectobacterium carotovorum AAT45375
Escherichia coli ABJ02146
Escherichia coli CAR04225
Escherichia coli ABG70666
Escherichia coli ABE08530
Escherichia coli CAQ99633
Escherichia coli CAU98864
Shigella flexneri ABF04874
Escherichia coli ACB16318
Escherichia fergusonii CAQ87924
Escherichia coli ACI79281
Escherichia coli BAB36995
Escherichia coli ACI38914
Escherichia coli AAG57823
Escherichia coli ACI79284
Shigella dysenteriae ABB62941
Escherichia coli CAP77151
Shigella sonnei AAZ89464
Escherichia coli ACB03834
Escherichia coli AAA16430
Escherichia coli CAR14207
Escherichia coli CAR19023
Escherichia coli ACR62674
Escherichia coli ABV07097
Escherichia coli BAE76793
Shigella boydii ABB67328
Escherichia coli ABV19991
Escherichia coli ACA76665
Escherichia coli BAG78488
Shigella boydii ACD07856
0.2042
0.102
Clostridium acetobutylicum AAK80905
0.1608
Exiguobacterium sp ACQ69978
0.2843
Staphylococcus carnosus CAL27580
0.1538
Streptococcus pneumoniae AAK99228
Streptococcus pneumoniae ABJ54778
0.2355
Streptococcus pneumoniae AAK74636
Streptococcus pneumoniae ACA36493
Streptococcus pneumoniae ACF55638
0.2765
Lactobacillus gasseri ABJ59750
Lactobacillus casei ABJ71621
0.114
Lactobacillus casei ACG55692
Lactobacillus casei CAQ65853
0.102
Lactobacillus rhamnosus ACH91635
0.1888
Lactobacillus acidophilus BAA07122
0.1595
Lactobacillus delbrueckii CAM35428
0.2153
Lactobacillus gasseri ABJ59900
0.2534
Bacillus amyloliquefaciens ABS73581


0.2534
Bacillus amyloliquefaciens ABS73581
0.1752
Streptococcus uberis CAR43123
Streptococcus equi ACG61833
Streptococcus equi CAX00199
Streptococcus mutans AAA16450
Streptococcus mutans AAN59144
Streptococcus pneumoniae ACO23459
Streptococcus pneumoniae CAR68898
Streptococcus pneumoniae ACO18396
Streptococcus pneumoniae ACF55578
Streptococcus pneumoniae ACB90364
Streptococcus pneumoniae ACA37023
Streptococcus pneumoniae ABJ54974
Streptococcus pneumoniae AAK99872
Streptococcus pneumoniae AAK75293
Streptococcus gordonii AAG39001
Streptococcus gordonii ABV10357
Lactobacillus sakei ACD75529
Lactococcus garvieae ABW34725
Lactococcus lactis AAA25173
Lactococcus lactis ABA47363
Lactococcus lactis ABJ74133
Streptococcus sanguinis ABN45075
Streptococcus uberis CAR41805
Streptococcus agalactiae CAD46988
Streptococcus pyogenes BAC64746
Streptococcus pyogenes CAM30928
Streptococcus pyogenes ABF38679
Streptococcus pyogenes ABF34766
Streptococcus pyogenes AAZ52250
Streptococcus pyogenes AAX72732
Streptococcus pyogenes AAM80260
Streptococcus pyogenes AAK34620
Streptococcus pyogenes AAT87776
Streptococcus pyogenes ABF32822
Streptococcus pyogenes ABF36709
Streptococcus pyogenes ACI61842
Staphylococcus haemolyticus BAE04157
Streptococcus suis ABP90002
Streptococcus suis ABP92206
Staphylococcus epidermidis AAO05422
Staphylococcus epidermidis AAW55165
Staphylococcus aureus AAA26650
Staphylococcus aureus AAW37056
Staphylococcus aureus ABD21770
Staphylococcus aureus ABD31469
Staphylococcus aureus ABQ50001
Staphylococcus aureus ABR53087
Staphylococcus aureus ABX30176
Staphylococcus aureus BAB43281
Staphylococcus aureus BAB95980
Staphylococcus aureus BAF68365
Staphylococcus aureus BAF79056
Staphylococcus aureus CAG41258
Staphylococcus aureus CAG43898
Staphylococcus aureus CAI81759
0.6009
Mesoplasma florum AAT75784
Salmonella enterica AAV79423
Salmonella enterica CAR61655
Salmonella enterica CAR39438
Salmonella enterica CAR35174
Salmonella enterica CAD03219
Salmonella enterica ACN47927
Salmonella enterica ACH75793
Salmonella enterica ACH49506
Salmonella enterica ACF90852
Salmonella enterica ACF65956
Salmonella enterica ACF65267
Salmonella enterica ABX70004
Salmonella enterica AAX67601
Salmonella enterica AAO71234
Salmonella enterica AAL22633
Escherichia fergusonii CAQ91392


Yersinia enterocolitica CAL11384
0.1175
Enterobacter sp ABP58726
Pectobacterium atrosepticum CAG77303
0.1280
Pectobacterium carotovorum ABL14155
0.1036
Cronobacter sakazakii ABU77277
Yersinia enterocolitica CAL11067
0.1164
Serratia marcescens BAB92992
0.110
Serratia proteamaculans ABV39942
0.1105
0.1887
Photobacterium profundum CAG23851
0.1549
Aeromonas salmonicida ABO88650
0.1179
Tolumonas auensis ACQ92079
Listeria monocytogenes CAC98653
Listeria monocytogenes CAS04368
Listeria monocytogenes ACK40396
0.1212
Listeria monocytogenes AAT03385
Listeria innocua CAC95815
Listeria welshimeri CAK19958
0.1059
Enterococcus faecalis AAO80824
0.1132
0.2482
Clostridium beijerinckii ABR35931
0.1152
Exiguobacterium sp ACQ70812
Listeria innocua CAC95624
Listeria monocytogenes AAT03173
0.3046
Listeria monocytogenes ACK40602
Listeria monocytogenes CAS04158
Listeria monocytogenes CAC98451
Listeria welshimeri CAK19746
0.2781
Klebsiella pneumoniae ACI12102
0.1297
Bacillus clausii BAD63025
0.1044
Oceanobacillus iheyensis BAC14719
0.1181
Bacillus pumilus AAM74558
Bacillus pumilus ABV62787
Bacillus subtilis BAA19708
0.1037
Bacillus amyloliquefaciens ABS75919
Bacillus halodurans BAA75349
0.1319
Bacillus licheniformis AAU24076
0.2143
Bacillus clausii BAD62857
Geobacillus sp ACE77068
Geobacillus kaustophilus BAD76141
Geobacillus thermodenitrificans ABO67111
Thermoanaerobacter tengcongensis AAM2363
0.1145
ABY95574 Thermoanaerobacter pseudethanol
Thermoanaerobacter sp ABY91584
Streptococcus pyogenes BAC63947
Streptococcus pyogenes CAM30108
Streptococcus pyogenes ACI61349
Streptococcus pyogenes ABF38128
Streptococcus pyogenes ABF36138
Streptococcus pyogenes ABF34206
Streptococcus pyogenes ABF32315
Streptococcus pyogenes AAZ51703
Streptococcus pyogenes AAX72179
Streptococcus pyogenes AAT87189
Streptococcus pyogenes AAM79614
Streptococcus pyogenes AAK34165
Streptococcus equi CAW94361
Streptococcus equi CAW98774
0.2465
0.2173
Streptococcus suis ABP89154
Streptococcus suis ABP91346
0.1015
Streptococcus suis ABP91044
Streptococcus gordonii ABV09412
Streptococcus pneumoniae ABJ54846
Streptococcus pneumoniae CAR68113
Streptococcus pneumoniae AAK99048
Streptococcus pneumoniae AAK74443
Streptococcus pneumoniae ACA36979
Streptococcus pneumoniae ACB89527
Streptococcus pneumoniae ACF55115
Streptococcus pneumoniae ACO16780
0.3369
Mycoplasma mycoides CAE77454
0.3341
Lactobacillus sakei CAI55471
Klebsiella pneumoniae ACI10649
0.1201
Klebsiella pneumoniae BAH64858
0.1540
Klebsiella pneumoniae ABR78470


0.1540
Klebsiella pneumoniae ABR78470
Yersinia enterocolitica CAL14116
Streptococcus pneumoniae ACB91239
Streptococcus pneumoniae CAR69790
Streptococcus pneumoniae ACF56690
Streptococcus pneumoniae ACA37143
0.2286
Streptococcus pneumoniae ABJ53669
Streptococcus pneumoniae AAL00636
Streptococcus pneumoniae AAK76086
0.1182
Clostridium beijerinckii ABR36675
Clostridium perfringens ABG85541
Clostridium perfringens ABG84939
Clostridium perfringens BAB82015
0.1788
Lactococcus lactis AAK05587
Lactococcus lactis ABJ73088
0.2212
Streptococcus uberis CAR43297
0.2109
Lactococcus lactis ABJ72983
0.1917
Lactobacillus plantarum CAD65701
Oenococcus oeni AAS68345
0.2964
Lactobacillus casei CAQ67883
Streptococcus equi ACG62839
Streptococcus equi CAW94744
Streptococcus equi CAW98372
0.2454
Streptococcus pyogenes BAC63663
Streptococcus pyogenes CAM29880
Streptococcus pyogenes AA79900
Streptococcus pyogenes ABF34494
Streptococcus pyogenes ACI61521
Streptococcus pyogenes AAX72464
Streptococcus pyogenes AAZ51931
Streptococcus pyogenes ABF38372
Streptococcus pyogenes AAK34377
Streptococcus pyogenes AAT87466
Streptococcus pyogenes ABF32494
Streptococcus pyogenes ABF36385
0.8393
Propionibacterium acnes AAT82894
1.1836
Petrotoga mobilis ABX31536
0.4565
Thermodesulfovibrio yellowstonii ACI2106
0.6845
Pantoea agglomerans CAA56282
0.4115
Leuconostoc citreum ACA82626
0.2543
Oenococcus oeni ABJ57102
Klebsiella pneumoniae ABR76283
Klebsiella pneumoniae BAH62540
Klebsiella pneumoniae ACI11639
Enterobacter sp ABP59988
Lactococcus lactis AAK04917
0.1733
Lactococcus lactis CAL98310
Lactococcus lactis ABJ72427
0.1586
Pediococcus pentosaceus ABJ68797
Lactobacillus acidophilus AAV42744
Lactobacillus gasseri ABJ59607
Lactobacillus johnsonii AAS08176
0.1009
Lactobacillus helveticus ABX27040
0.1041
Lactobacillus plantarum CAD65191
0.3387
Caulobacter sp ABZ73130
0.2230
Novosphingobium aromaticivorans ABD24989
0.1480
Novosphingobium aromaticivorans ABD25234
0.2409
Sphingomonas paucimobilis AAG59862
0.2353
Caulobacter sp ABZ74400
0.2349
Solibacter usitatus ABJ83756
0.2512
Arthrobacter chlorophenolicus ACL38420
0.2212
Chloroflexus aurantiacus ABY33610
Chloroflexus sp ACM51826
0.1291
Frankia sp ABW12660
0.2534
Frankia sp ABW14028
0.1904
0.1477
Clavibacter michiganensis CAN00121
Kineococcus radiotolerans ABS04001
0.1722
0.2583
Streptomyces ambofaciens CAJ88232
Streptomyces griseus BAG19255
0.1010
0.1771
Streptomyces rochei CAA52344
0.3383
Bifidobacterium adolescentis BAF40068
0.2684
Fervidobacterium nodosum ABS61401
Thermosipho africanus ACJ75238

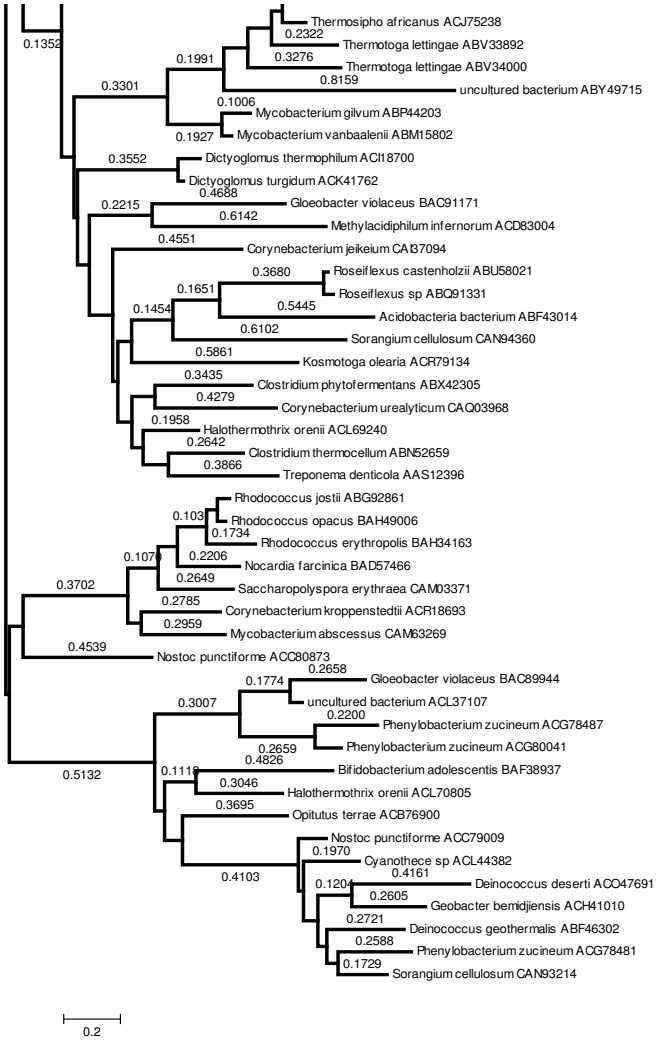
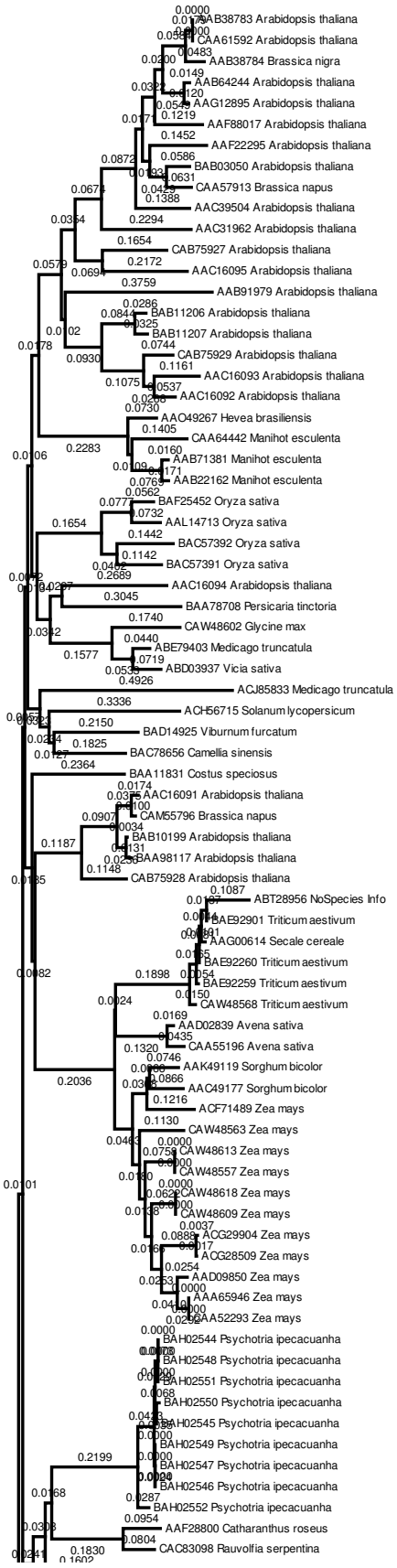
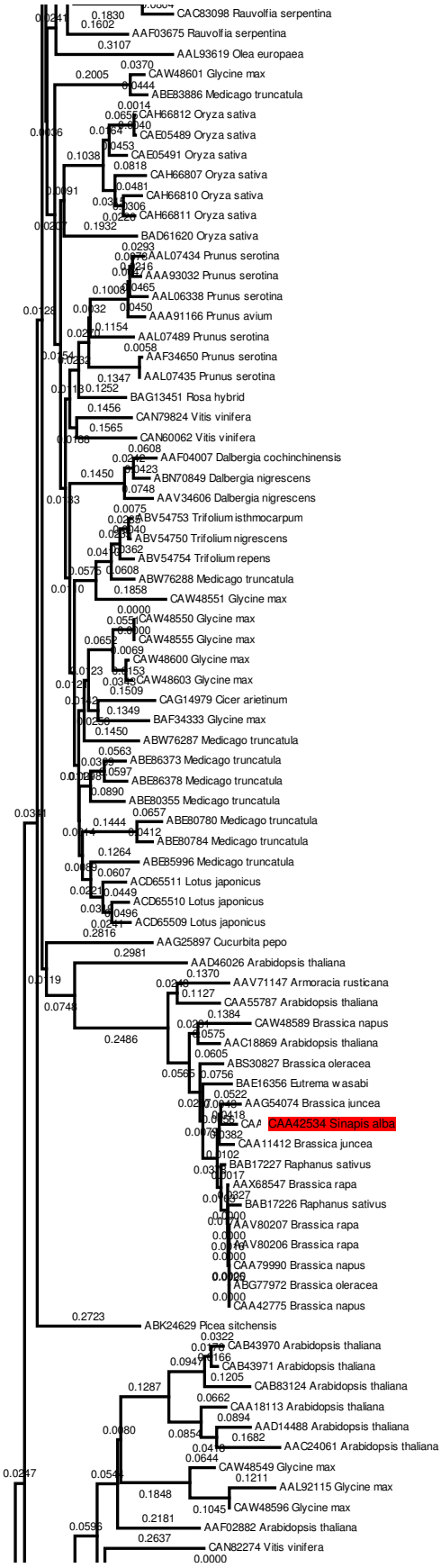


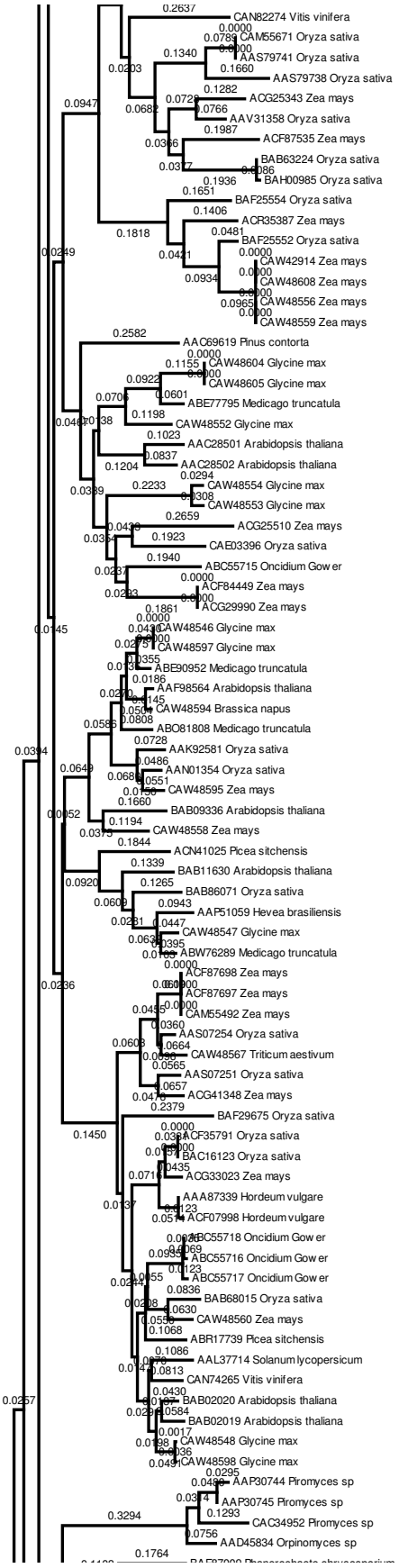

**Figure 125.** Phylogenetic analysis of 1129 bacterial GH1 proteins (■ = slow-cleaving proteins; ■ = non-cleaving proteins).

### 21.3. Eukaryotic GH1 proteins

The evolutionary relationships of 295 eukaryotic GH1 proteins obtained from the CAZy database (http://www.cazy.org/fam/GH1.html),[352] are shown in Figure 126, and is also available to view and download at http://www.cbrg.ox.ac.uk/data.shtml. It is recommended that the file is viewed in Dendroscope.[357] The phylogenetic relationship was inferred using the Neighbour-Joining method.[410] The evolutionary distances (shown next to the branches) were computed using the Dayhoff matrix based method[411] and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 186 positions in the final dataset. Phylogenetic visualisation was conducted in MEGA4.[412]


AAB38783 Arabidopsis thaliana
CAA61592 Arabidopsis thaliana
AAB38784 Brassica nigra
AAB64244 Arabidopsis thaliana
AAG12895 Arabidopsis thaliana
AAF88017 Arabidopsis thaliana
AAF22295 Arabidopsis thaliana
BAB03050 Arabidopsis thaliana
CAA57913 Brassica napus
AAC39504 Arabidopsis thaliana
AAC31962 Arabidopsis thaliana
CAB75927 Arabidopsis thaliana
AAC16095 Arabidopsis thaliana
AAB91979 Arabidopsis thaliana
BAB11206 Arabidopsis thaliana
BAB11207 Arabidopsis thaliana
CAB75929 Arabidopsis thaliana
AAC16093 Arabidopsis thaliana
AAC16092 Arabidopsis thaliana
AAO49267 Hevea brasiliensis
CAA64442 Manihot esculenta
AAB71381 Manihot esculenta
AAB22162 Manihot esculenta
BAF25452 Oryza sativa
AAL14713 Oryza sativa
BAC57392 Oryza sativa
BAC57391 Oryza sativa
AAC16094 Arabidopsis thaliana
BAA78708 Persicaria tinctoria
CAW48602 Glycine max
ABE79403 Medicago truncatula
ABD03937 Vicia sativa
ACJ85833 Medicago truncatula
ACH56715 Solanum lycopersicum
BAD14925 Viburnum furcatum
BAC78656 Camellia sinensis
BAA11831 Costus speciosus
AAC16091 Arabidopsis thaliana
CAM55796 Brassica napus
BAB10199 Arabidopsis thaliana
BAA98117 Arabidopsis thaliana
CAB75928 Arabidopsis thaliana
ABT28956 NoSpecies Info
BAE92901 Triticum aestivum
AAG00614 Secale cereale
BAE92260 Triticum aestivum
BAE92259 Triticum aestivum
CAW48568 Triticum aestivum
AAD02839 Avena sativa
CAA55196 Avena sativa
AAK49119 Sorghum bicolor
AAC49177 Sorghum bicolor
ACF71489 Zea mays
CAW48563 Zea mays
CAW48613 Zea mays
CAW48557 Zea mays
CAW48618 Zea mays
CAW48609 Zea mays
ACG29904 Zea mays
ACG28509 Zea mays
AAD09850 Zea mays
AAA65946 Zea mays
CAA52293 Zea mays
BAH02544 Psychotria ipecacuanha
BAH02548 Psychotria ipecacuanha
BAH02551 Psychotria ipecacuanha
BAH02550 Psychotria ipecacuanha
BAH02545 Psychotria ipecacuanha
BAH02549 Psychotria ipecacuanha
BAH02547 Psychotria ipecacuanha
BAH02546 Psychotria ipecacuanha
BAH02552 Psychotria ipecacuanha
AAF28800 Catharanthus roseus
CAC83098 Rauvolfia serpentina


CAC83098 Rauvolfia serpentina
AAF03675 Rauvolfia serpentina
AAL93619 Olea europaea
CAW48601 Glycine max
ABE83886 Medicago truncatula
CAH66812 Oryza sativa
CAE05489 Oryza sativa
CAE05491 Oryza sativa
CAH66807 Oryza sativa
CAH66810 Oryza sativa
CAH66811 Oryza sativa
BAD61620 Oryza sativa
AAL07434 Prunus serotina
AAA93032 Prunus serotina
AAL06338 Prunus serotina
AAA91166 Prunus avium
AAL07489 Prunus serotina
AAF34650 Prunus serotina
AAL07435 Prunus serotina
BAG13451 Rosa hybrid
CAN79824 Vitis vinifera
CAN60062 Vitis vinifera
AAF04007 Dalbergia cochinchinensis
ABN70849 Dalbergia nigrescens
AAV34606 Dalbergia nigrescens
ABV54753 Trifolium isthmocarpum
ABV54750 Trifolium nigrescens
ABV54754 Trifolium repens
ABW76288 Medicago truncatula
CAW48551 Glycine max
CAW48550 Glycine max
CAW48555 Glycine max
CAW48600 Glycine max
CAW48603 Glycine max
CAG14979 Cicer arietinum
BAF34333 Glycine max
ABW76287 Medicago truncatula
ABE86373 Medicago truncatula
ABE86378 Medicago truncatula
ABE80355 Medicago truncatula
ABE80780 Medicago truncatula
ABE80784 Medicago truncatula
ABE85996 Medicago truncatula
ACD65511 Lotus japonicus
ACD65510 Lotus japonicus
ACD65509 Lotus japonicus
AAG25897 Cucurbita pepo
AAD46026 Arabidopsis thaliana
AAV71147 Armoracia rusticana
CAA55787 Arabidopsis thaliana
CAW48589 Brassica napus
AAC18869 Arabidopsis thaliana
ABS30827 Brassica oleracea
BAE16356 Eutrema wasabi
AAG54074 Brassica juncea
CAA42534 Sinapis alba
CAA11412 Brassica juncea
BAB17227 Raphanus sativus
AAX68547 Brassica rapa
BAB17226 Raphanus sativus
AAV80207 Brassica rapa
AAV80206 Brassica rapa
CAA79990 Brassica napus
ABG77972 Brassica oleracea
CAA42775 Brassica napus
ABK24629 Picea sitchensis
CAB43970 Arabidopsis thaliana
CAB43971 Arabidopsis thaliana
CAB83124 Arabidopsis thaliana
CAA18113 Arabidopsis thaliana
AAD14488 Arabidopsis thaliana
AAC24061 Arabidopsis thaliana
CAW48549 Glycine max
AAL92115 Glycine max
CAW48596 Glycine max
AAF02882 Arabidopsis thaliana
CAN82274 Vitis vinifera


CAN82274 Vitis vinifera
CAM55671 Oryza sativa
AAS79741 Oryza sativa
AAS79738 Oryza sativa
ACG25343 Zea mays
AAV31358 Oryza sativa
ACF87535 Zea mays
BAB63224 Oryza sativa
BAH00985 Oryza sativa
BAF25554 Oryza sativa
ACR35387 Zea mays
BAF25552 Oryza sativa
CAW42914 Zea mays
CAW48608 Zea mays
CAW48556 Zea mays
CAW48559 Zea mays
AAC69619 Pinus contorta
CAW48604 Glycine max
CAW48605 Glycine max
ABE77795 Medicago truncatula
CAW48552 Glycine max
AAC28501 Arabidopsis thaliana
AAC28502 Arabidopsis thaliana
CAW48554 Glycine max
CAW48553 Glycine max
ACG25510 Zea mays
CAE03396 Oryza sativa
ABC55715 Oncidium Gower
ACF84449 Zea mays
ACG29990 Zea mays
CAW48546 Glycine max
CAW48597 Glycine max
ABE90952 Medicago truncatula
AAF98564 Arabidopsis thaliana
CAW48594 Brassica napus
ABO81808 Medicago truncatula
AAK92581 Oryza sativa
AAN01354 Oryza sativa
CAW48595 Zea mays
BAB09336 Arabidopsis thaliana
CAW48558 Zea mays
ACN41025 Picea sitchensis
BAB11630 Arabidopsis thaliana
BAB86071 Oryza sativa
AAP51059 Hevea brasiliensis
CAW48547 Glycine max
ABW76289 Medicago truncatula
ACF87698 Zea mays
ACF87697 Zea mays
CAM55492 Zea mays
AAS07254 Oryza sativa
CAW48567 Triticum aestivum
AAS07251 Oryza sativa
ACG41348 Zea mays
BAF29675 Oryza sativa
ACF35791 Oryza sativa
BAC16123 Oryza sativa
ACG33023 Zea mays
AAA87339 Hordeum vulgare
ACF07998 Hordeum vulgare
ABC55718 Oncidium Gower
ABC55716 Oncidium Gower
ABC55717 Oncidium Gower
BAB68015 Oryza sativa
CAW48560 Zea mays
ABR17739 Picea sitchensis
AAL37714 Solanum lycopersicum
CAN74265 Vitis vinifera
BAB02020 Arabidopsis thaliana
BAB02019 Arabidopsis thaliana
CAW48548 Glycine max
CAW48598 Glycine max
AAP30744 Piromyces sp
AAP30745 Piromyces sp
CAC34952 Piromyces sp
AAD45834 Orpinomyces sp

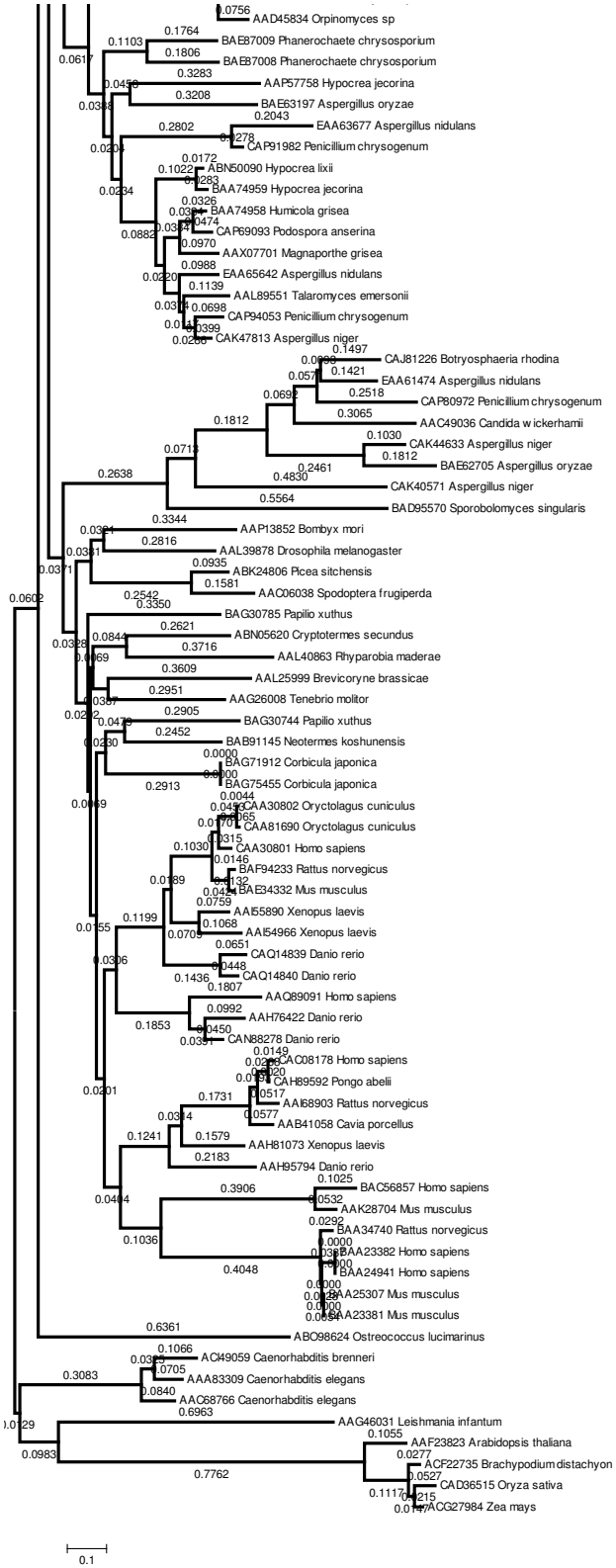
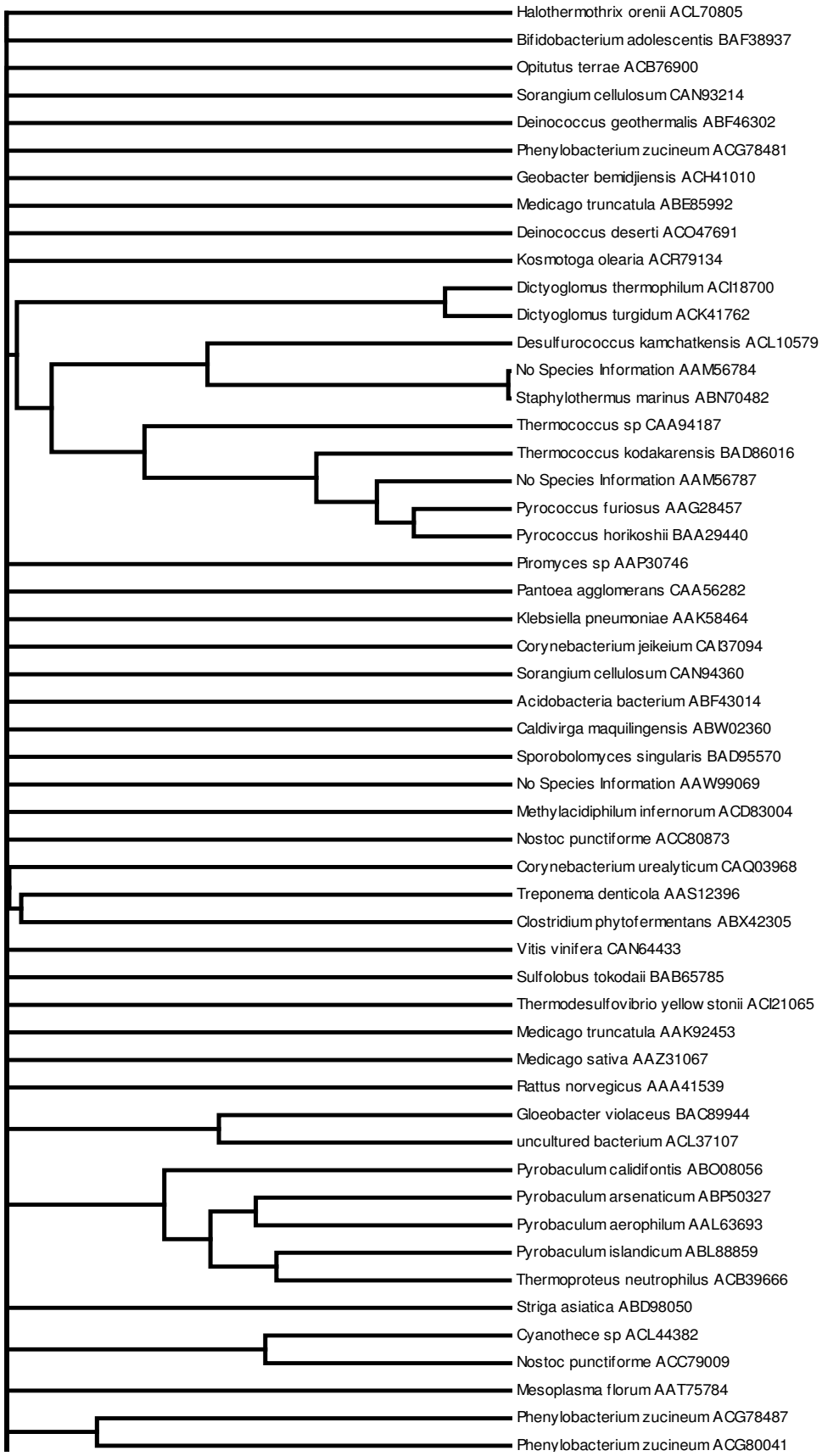
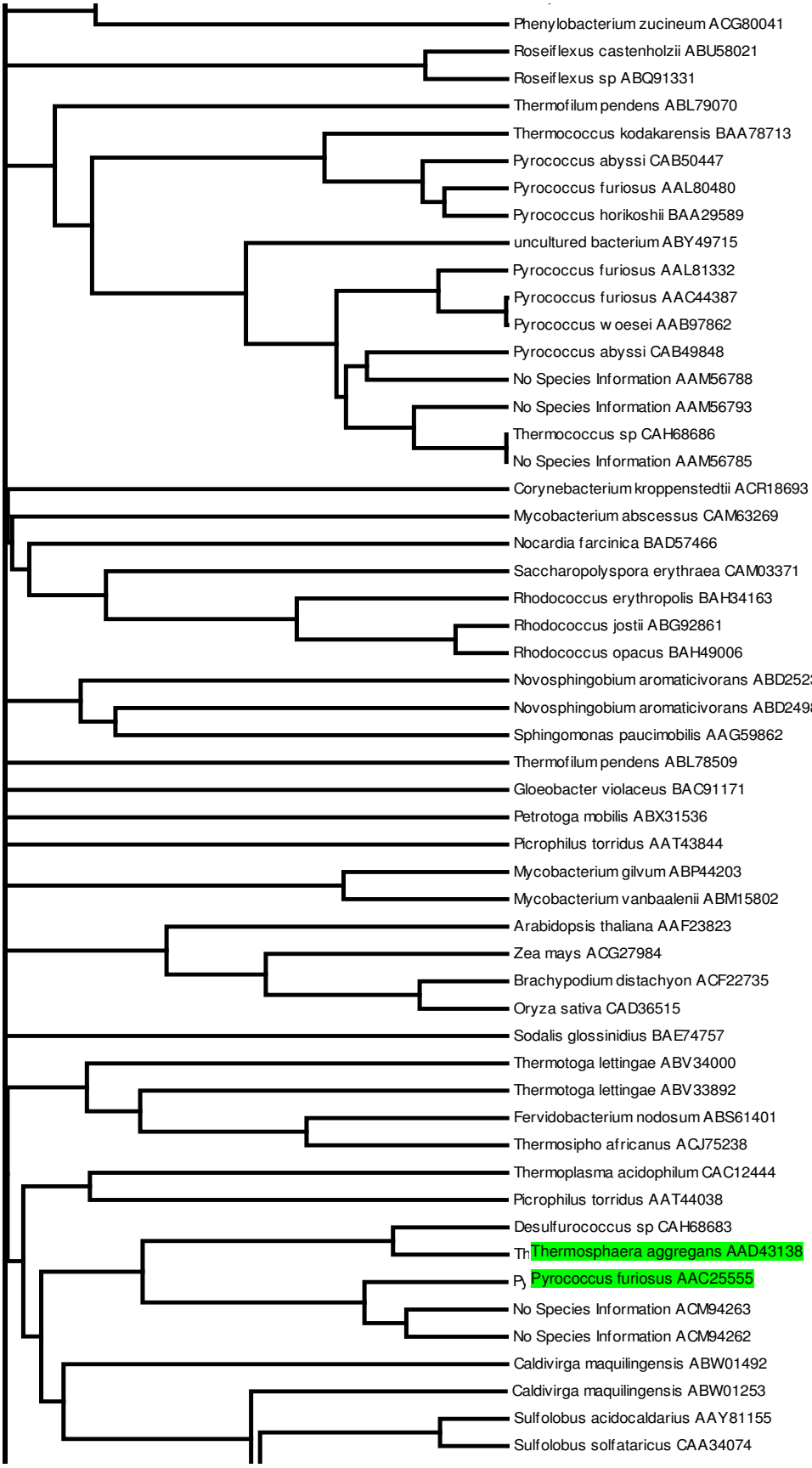


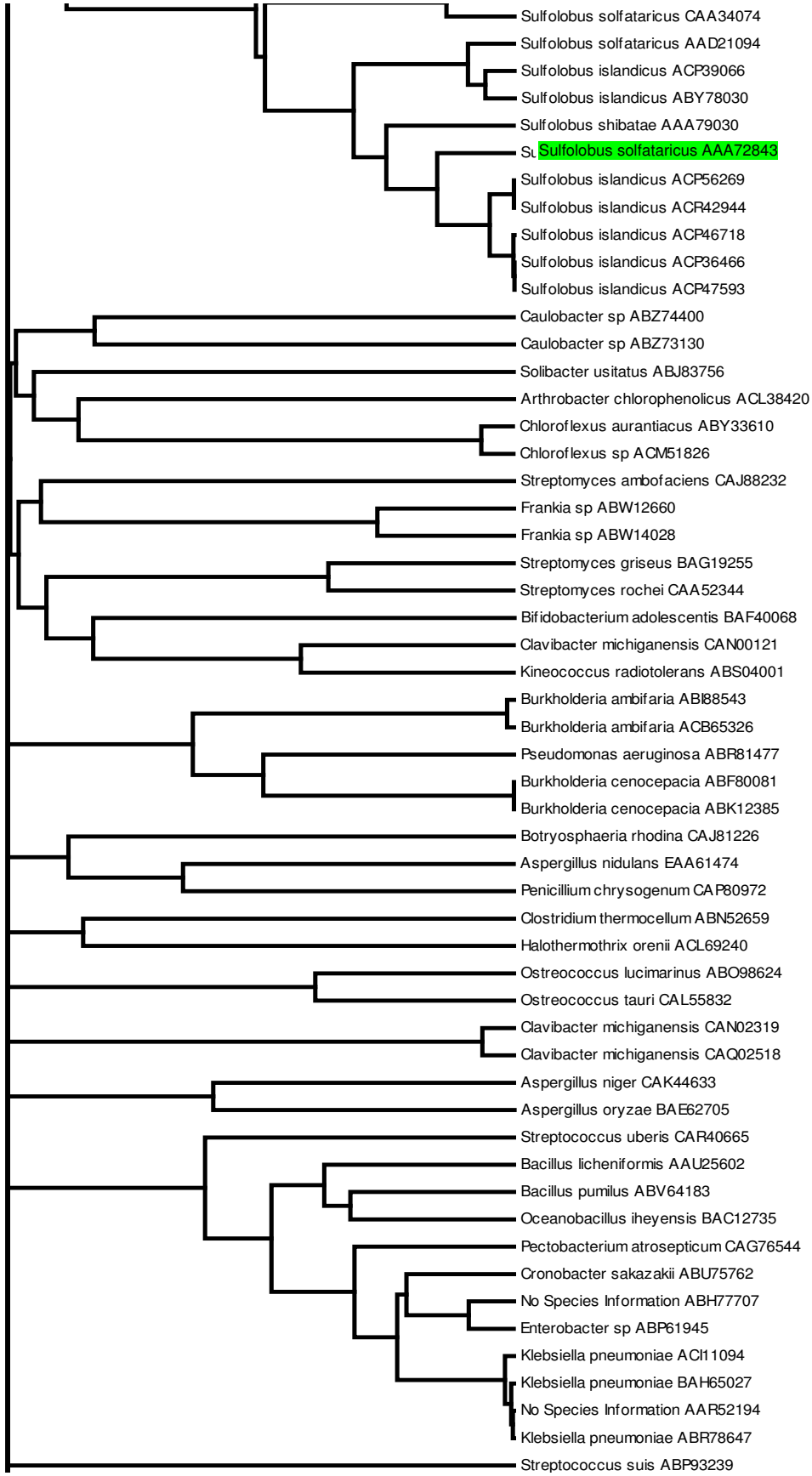

**Figure 126.** Phylogenetic analysis of 295 eukaryotic GH1 proteins ( = non-cleaving proteins).

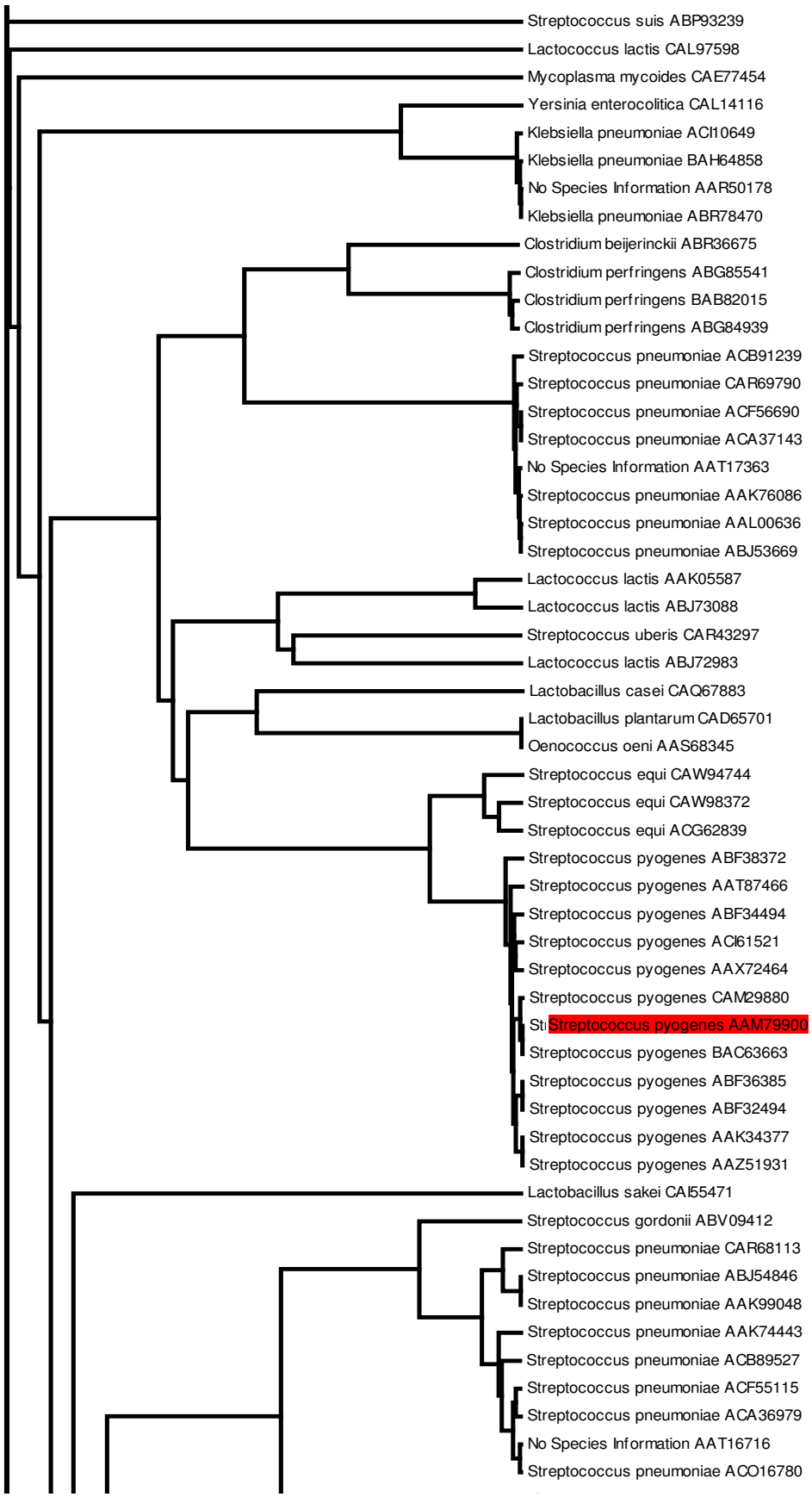
### 21.4. Archaic, bacterial and eukaryotic GH1 proteins

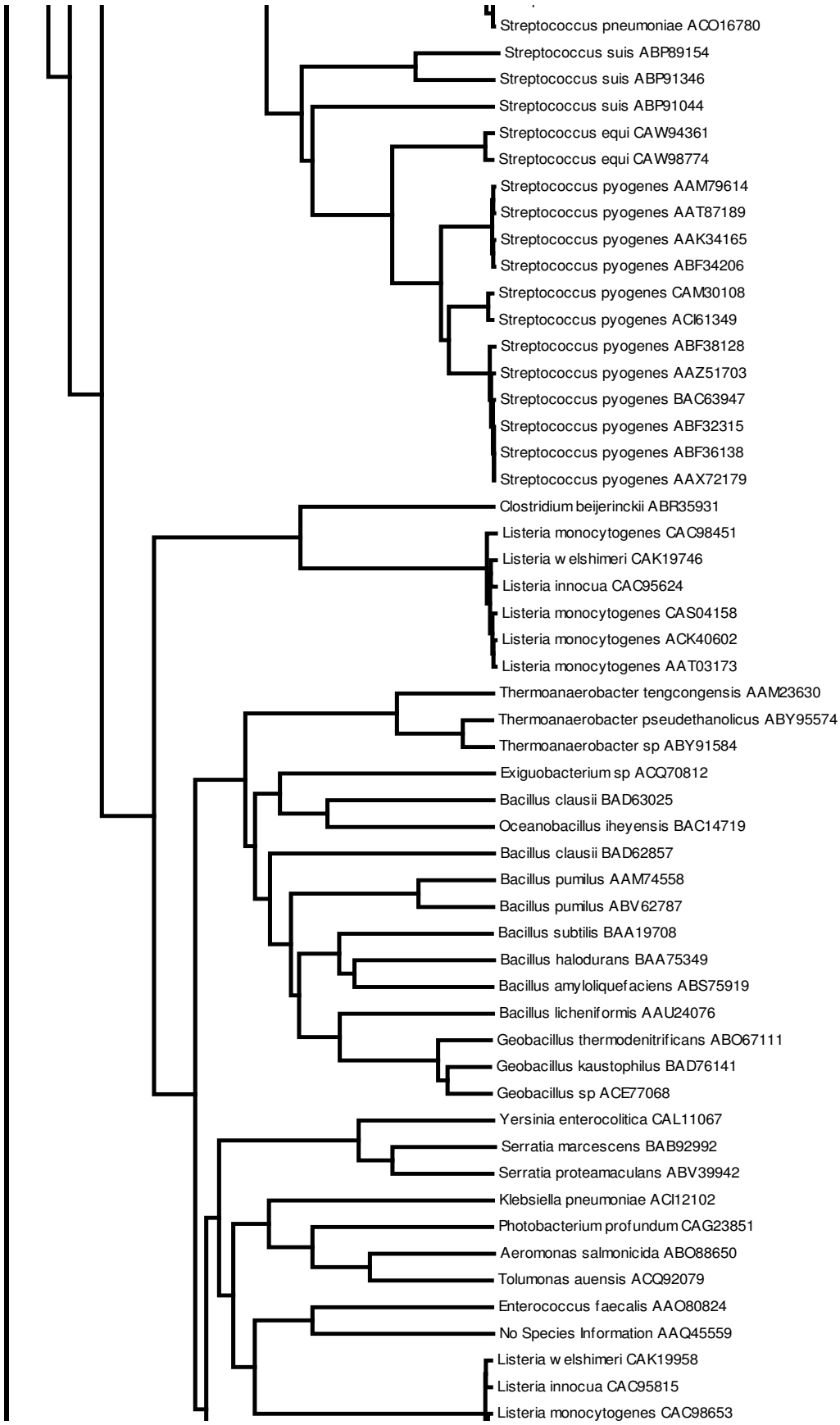
The evolutionary relationships of 1654 archaic, bacterial and eukaryotic GH1 proteins obtained from the CAZy database (http://www.cazy.org/fam/GH1.html),[352] are shown in Figure 127, and is also available to view and download at http://www.cbrg.ox.ac.uk/data.shtml. It is recommended that the file is viewed in Dendroscope.[357] The phylogenetic relationship was inferred using the Quick Neighbour-Joining algorithm as part of the Topali 2 package (http://www.topali.org/)[358]. Phylogenetic visualisation was conducted in MEGA4.[412]

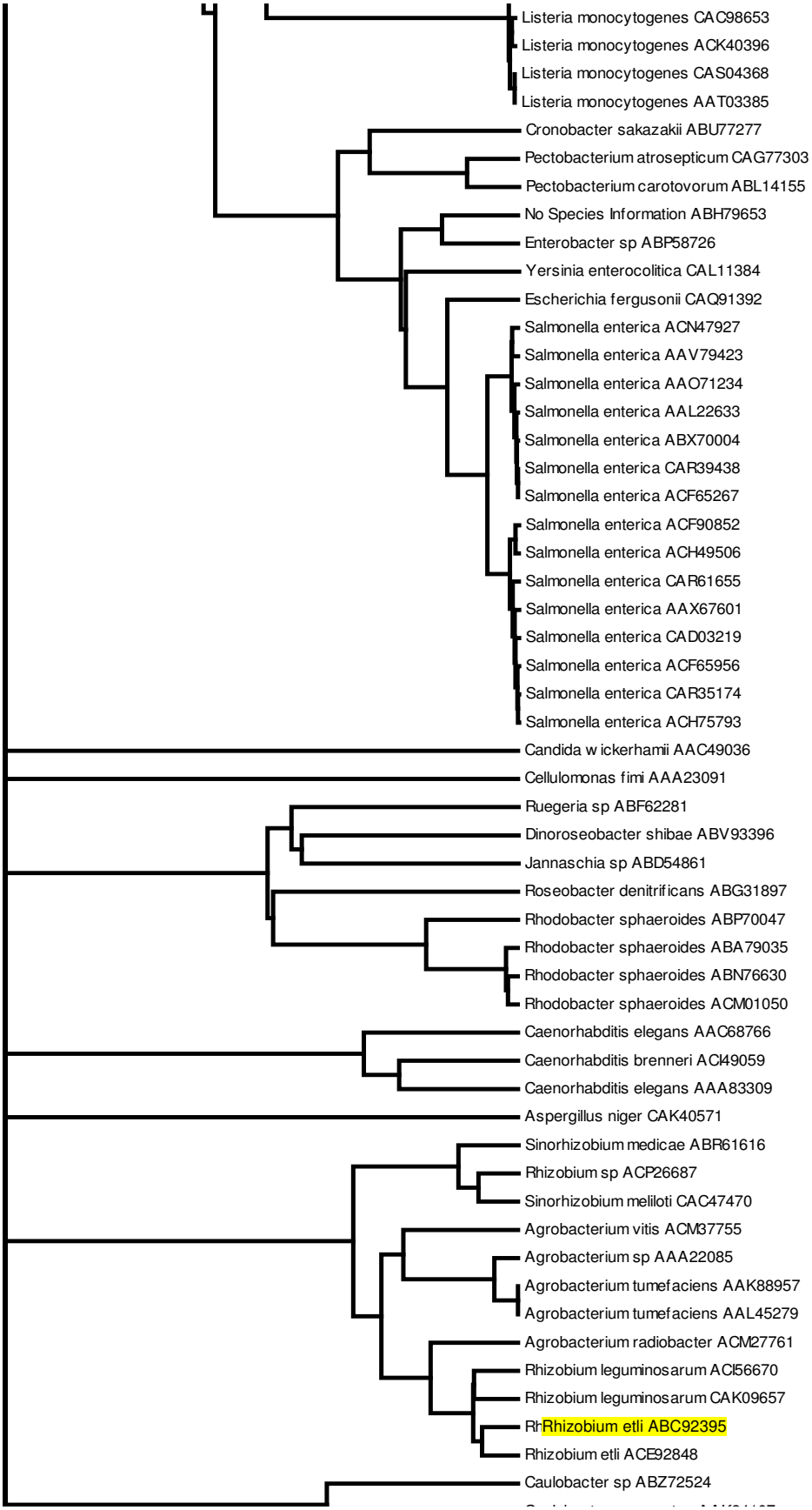

Halothermothrix orenii ACL70805
Bifidobacterium adolescentis BAF38937
Opitutus terrae ACB76900
Sorangium cellulosum CAN93214
Deinococcus geothermalis ABF46302
Phenylobacterium zucineum ACG78481
Geobacter bemidjiensis ACH41010
Medicago truncatula ABE85992
Deinococcus deserti ACO47691
Kosmotoga olearia ACR79134
Dictyoglomus thermophilum ACI18700
Dictyoglomus turgidum ACK41762
Desulfurococcus kamchatkensis ACL10579
No Species Information AAM56784
Staphylothermus marinus ABN70482
Thermococcus sp CAA94187
Thermococcus kodakarensis BAD86016
No Species Information AAM56787
Pyrococcus furiosus AAG28457
Pyrococcus horikoshii BAA29440
Piromyces sp AAP30746
Pantoea agglomerans CAA56282
Klebsiella pneumoniae AAK58464
Corynebacterium jeikeium CAI37094
Sorangium cellulosum CAN94360
Acidobacteria bacterium ABF43014
Caldivirga maquilingensis ABW02360
Sporobolomyces singularis BAD95570
No Species Information AAW99069
Methylacidiphilum infernorum ACD83004
Nostoc punctiforme ACC80873
Corynebacterium urealyticum CAQ03968
Treponema denticola AAS12396
Clostridium phytofermentans ABX42305
Vitis vinifera CAN64433
Sulfolobus tokodaii BAB65785
Thermodesulfovibrio yellowstonii ACI21065
Medicago truncatula AAK92453
Medicago sativa AAZ31067
Rattus norvegicus AAA41539
Gloeobacter violaceus BAC89944
uncultured bacterium ACL37107
Pyrobaculum calidifontis ABO08056
Pyrobaculum arsenaticum ABP50327
Pyrobaculum aerophilum AAL63693
Pyrobaculum islandicum ABL88859
Thermoproteus neutrophilus ACB39666
Striga asiatica ABD98050
Cyanothece sp ACL44382
Nostoc punctiforme ACC79009
Mesoplasma florum AAT75784
Phenylobacterium zucineum ACG78487
Phenylobacterium zucineum ACG80041

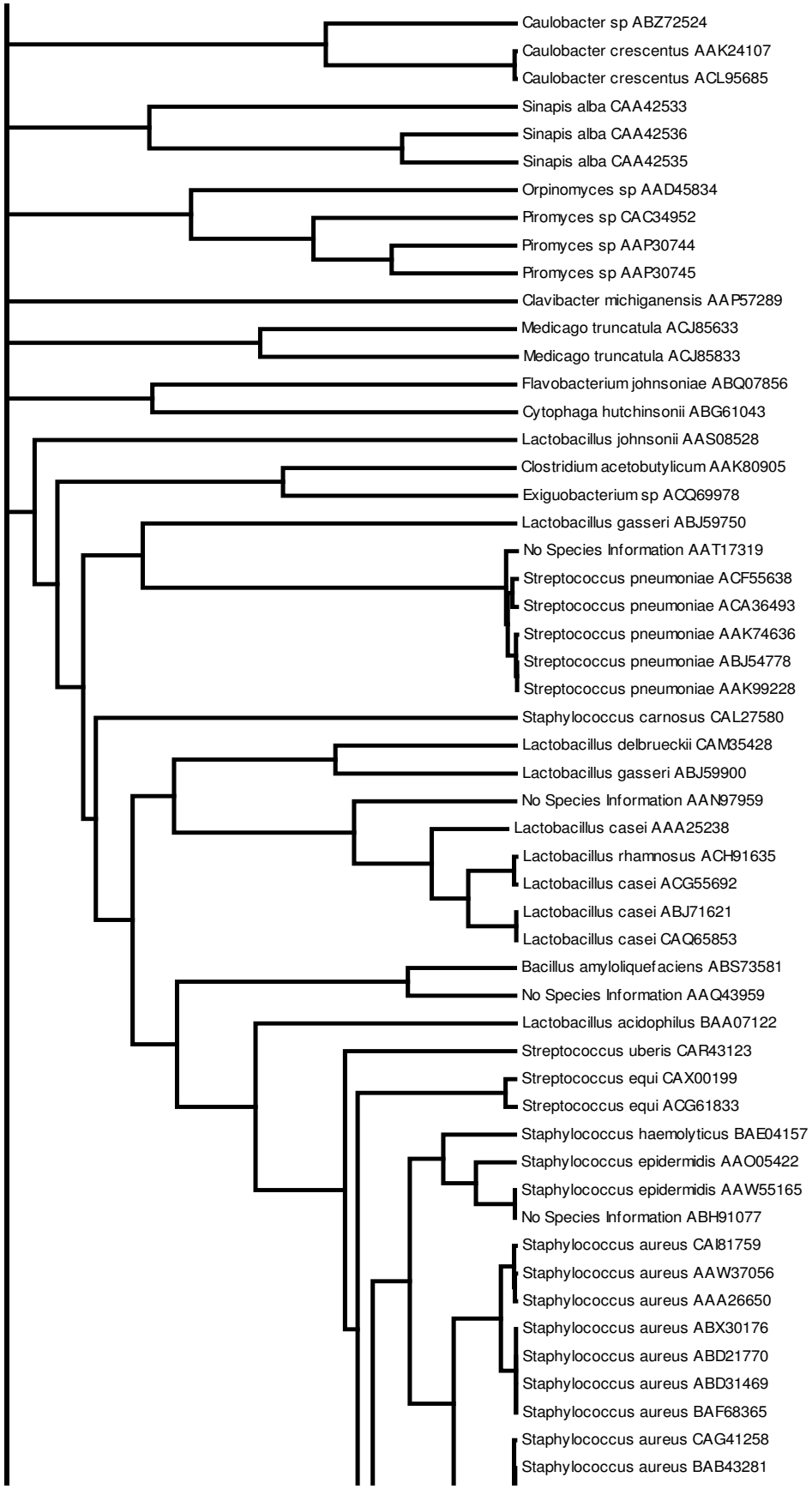

Phenylobacterium zucineum ACG80041
Roseiflexus castenholzii ABU58021
Roseiflexus sp ABQ91331
Thermofilum pendens ABL79070
Thermococcus kodakarensis BAA78713
Pyrococcus abyssi CAB50447
Pyrococcus furiosus AAL80480
Pyrococcus horikoshii BAA29589
uncultured bacterium ABY49715
Pyrococcus furiosus AAL81332
Pyrococcus furiosus AAC44387
Pyrococcus w oesei AAB97862
Pyrococcus abyssi CAB49848
No Species Information AAM56788
No Species Information AAM56793
Thermococcus sp CAH68686
No Species Information AAM56785
Corynebacterium kroppenstedtii ACR18693
Mycobacterium abscessus CAM63269
Nocardia farcinica BAD57466
Saccharopolyspora erythraea CAM03371
Rhodococcus erythropolis BAH34163
Rhodococcus jostii ABG92861
Rhodococcus opacus BAH49006
Novosphingobium aromaticivorans ABD25234
Novosphingobium aromaticivorans ABD24989
Sphingomonas paucimobilis AAG59862
Thermofilum pendens ABL78509
Gloeobacter violaceus BAC91171
Petrotoga mobilis ABX31536
Picrophilus torridus AAT43844
Mycobacterium gilvum ABP44203
Mycobacterium vanbaalenii ABM15802
Arabidopsis thaliana AAF23823
Zea mays ACG27984
Brachypodium distachyon ACF22735
Oryza sativa CAD36515
Sodalis glossinidius BAE74757
Thermotoga lettingae ABV34000
Thermotoga lettingae ABV33892
Fervidobacterium nodosum ABS61401
Thermosipho africanus ACJ75238
Thermoplasma acidophilum CAC12444
Picrophilus torridus AAT44038
Desulfurococcus sp CAH68683
Thermosphaera aggregans AAD43138
Pyrococcus furiosus AAC25555
No Species Information ACM94263
No Species Information ACM94262
Caldivirga maquilingensis ABW01492
Caldivirga maquilingensis ABW01253
Sulfolobus acidocaldarius AAY81155
Sulfolobus solfataricus CAA34074

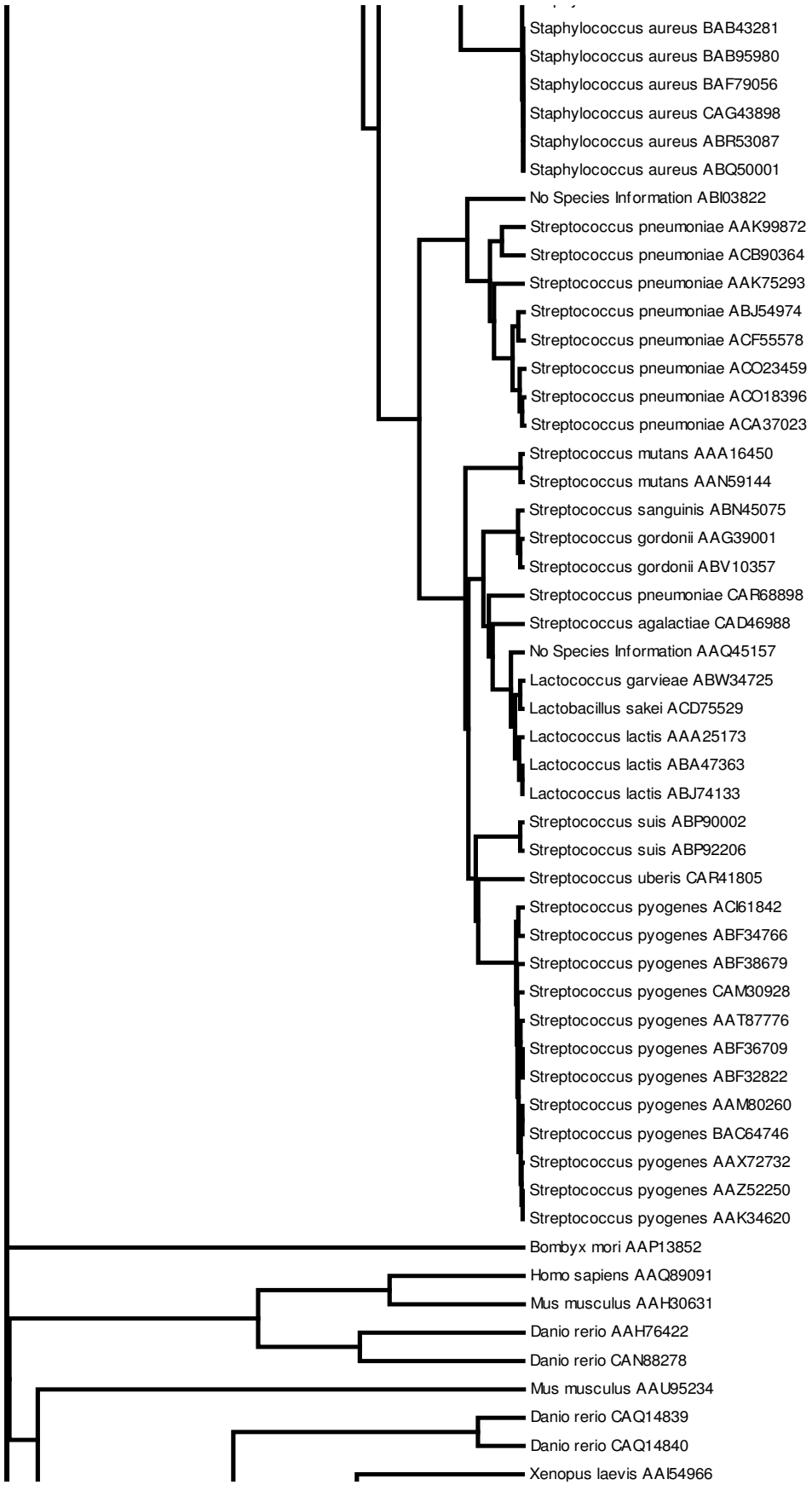

Sulfolobus solfataricus CAA34074
Sulfolobus solfataricus AAD21094
Sulfolobus islandicus ACP39066
Sulfolobus islandicus ABY78030
Sulfolobus shibatae AAA79030
Sulfolobus solfataricus AAA72843
Sulfolobus islandicus ACP56269
Sulfolobus islandicus ACR42944
Sulfolobus islandicus ACP46718
Sulfolobus islandicus ACP36466
Sulfolobus islandicus ACP47593
Caulobacter sp ABZ74400
Caulobacter sp ABZ73130
Solibacter usitatus ABJ83756
Arthrobacter chlorophenolicus ACL38420
Chloroflexus aurantiacus ABY33610
Chloroflexus sp ACM51826
Streptomyces ambofaciens CAJ88232
Frankia sp ABW12660
Frankia sp ABW14028
Streptomyces griseus BAG19255
Streptomyces rochei CAA52344
Bifidobacterium adolescentis BAF40068
Clavibacter michiganensis CAN00121
Kineococcus radiotolerans ABS04001
Burkholderia ambifaria ABI88543
Burkholderia ambifaria ACB65326
Pseudomonas aeruginosa ABR81477
Burkholderia cenocepacia ABF80081
Burkholderia cenocepacia ABK12385
Botryosphaeria rhodina CAJ81226
Aspergillus nidulans EAA61474
Penicillium chrysogenum CAP80972
Clostridium thermocellum ABN52659
Halothermothrix orenii ACL69240
Ostreococcus lucimarinus ABO98624
Ostreococcus tauri CAL55832
Clavibacter michiganensis CAN02319
Clavibacter michiganensis CAQ02518
Aspergillus niger CAK44633
Aspergillus oryzae BAE62705
Streptococcus uberis CAR40665
Bacillus licheniformis AAU25602
Bacillus pumilus ABV64183
Oceanobacillus iheyensis BAC12735
Pectobacterium atrosepticum CAG76544
Cronobacter sakazakii ABU75762
No Species Information ABH77707
Enterobacter sp ABP61945
Klebsiella pneumoniae ACI11094
Klebsiella pneumoniae BAH65027
No Species Information AAR52194
Klebsiella pneumoniae ABR78647
Streptococcus suis ABP93239

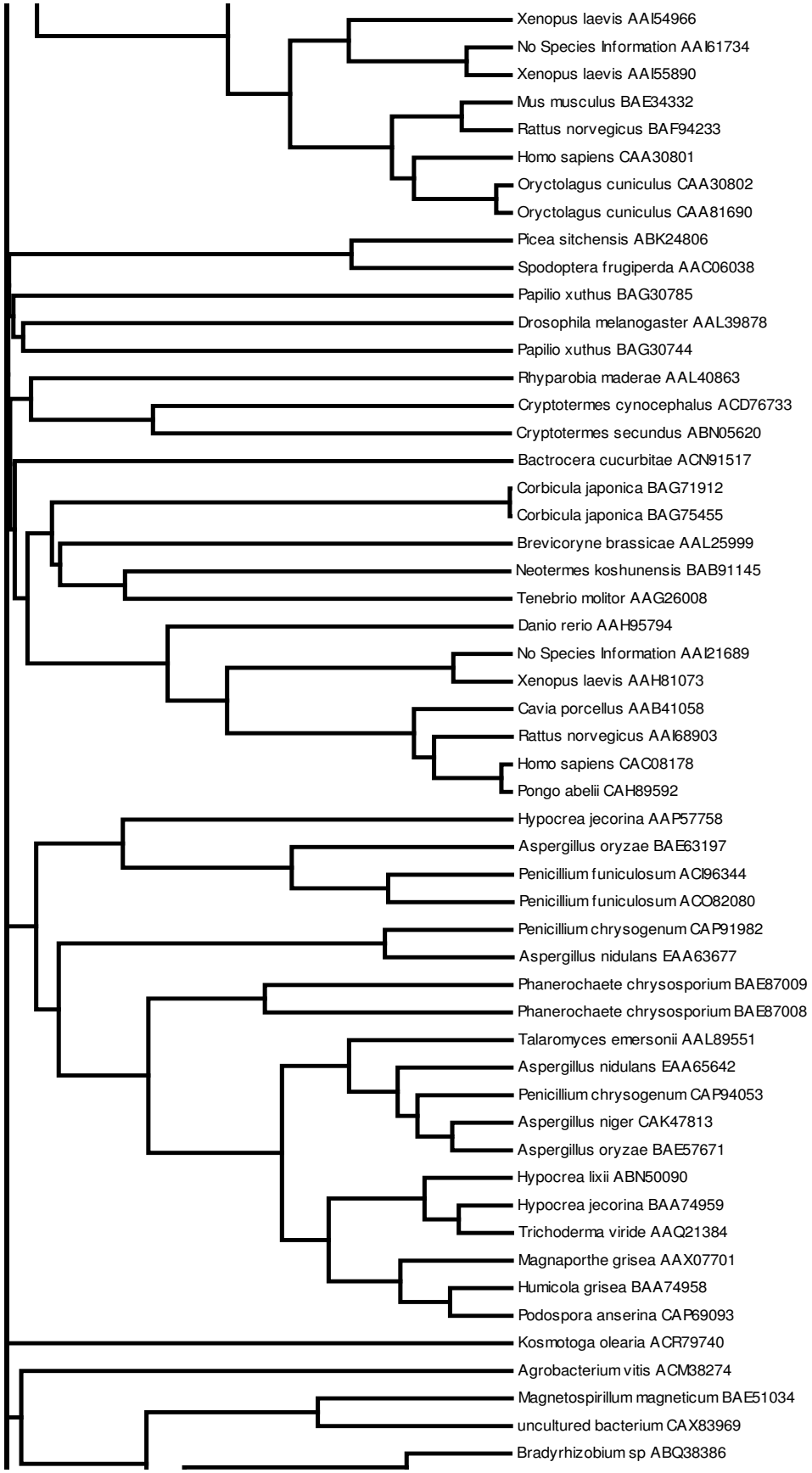

Streptococcus suis ABP93239
Lactococcus lactis CAL97598
Mycoplasma mycoides CAE77454
Yersinia enterocolitica CAL14116
Klebsiella pneumoniae ACI10649
Klebsiella pneumoniae BAH64858
No Species Information AAR50178
Klebsiella pneumoniae ABR78470
Clostridium beijerinckii ABR36675
Clostridium perfringens ABG85541
Clostridium perfringens BAB82015
Clostridium perfringens ABG84939
Streptococcus pneumoniae ACB91239
Streptococcus pneumoniae CAR69790
Streptococcus pneumoniae ACF56690
Streptococcus pneumoniae ACA37143
No Species Information AAT17363
Streptococcus pneumoniae AAK76086
Streptococcus pneumoniae AAL00636
Streptococcus pneumoniae ABJ53669
Lactococcus lactis AAK05587
Lactococcus lactis ABJ73088
Streptococcus uberis CAR43297
Lactococcus lactis ABJ72983
Lactobacillus casei CAQ67883
Lactobacillus plantarum CAD65701
Oenococcus oeni AAS68345
Streptococcus equi CAW94744
Streptococcus equi CAW98372
Streptococcus equi ACG62839
Streptococcus pyogenes ABF38372
Streptococcus pyogenes AAT87466
Streptococcus pyogenes ABF34494
Streptococcus pyogenes ACI61521
Streptococcus pyogenes AAX72464
Streptococcus pyogenes CAM29880
Streptococcus pyogenes AAM79900
Streptococcus pyogenes BAC63663
Streptococcus pyogenes ABF36385
Streptococcus pyogenes ABF32494
Streptococcus pyogenes AAK34377
Streptococcus pyogenes AAZ51931
Lactobacillus sakei CAI55471
Streptococcus gordonii ABV09412
Streptococcus pneumoniae CAR68113
Streptococcus pneumoniae ABJ54846
Streptococcus pneumoniae AAK99048
Streptococcus pneumoniae AAK74443
Streptococcus pneumoniae ACB89527
Streptococcus pneumoniae ACF55115
Streptococcus pneumoniae ACA36979
No Species Information AAT16716
Streptococcus pneumoniae ACO16780

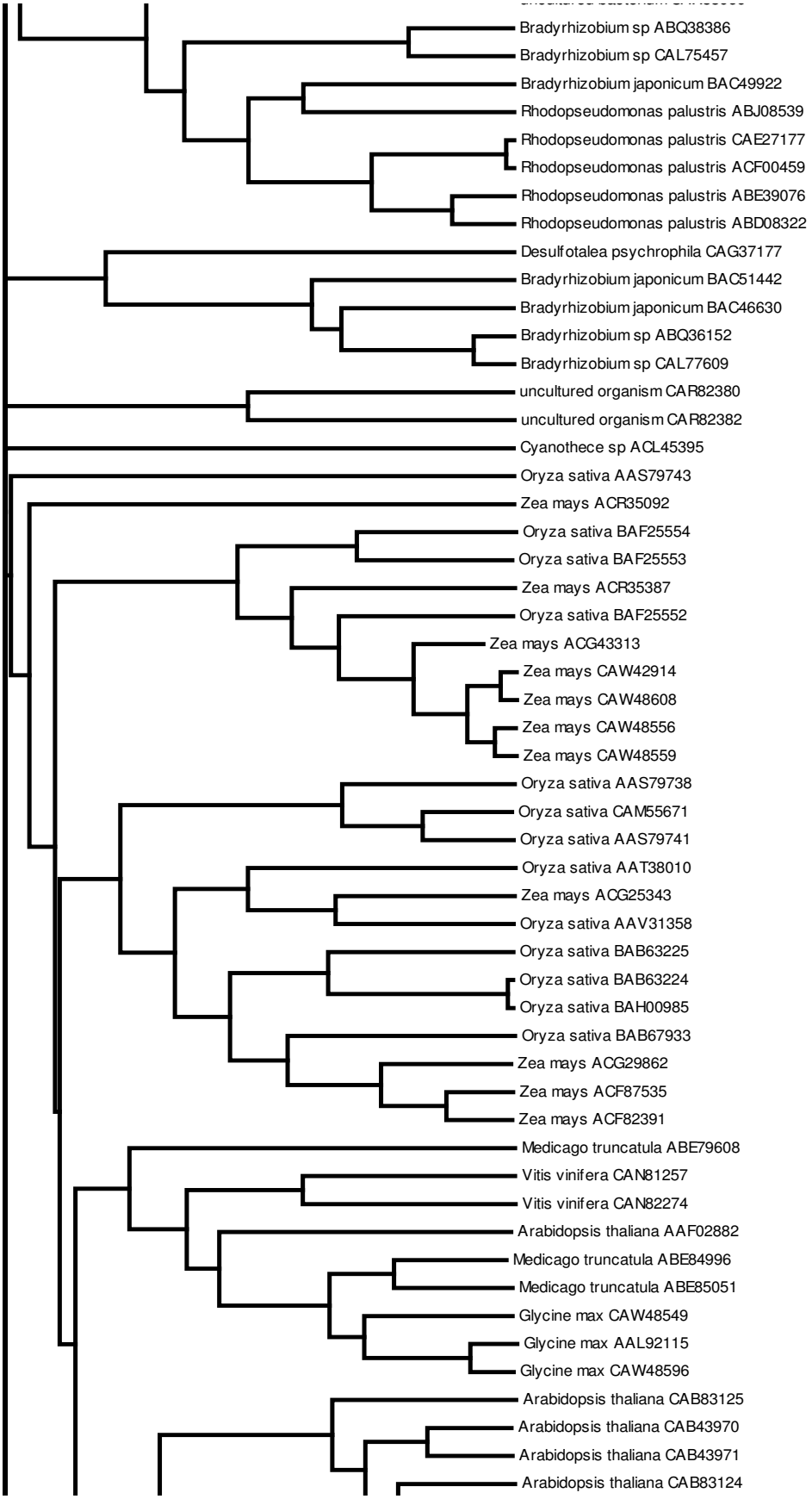

Streptococcus pneumoniae ACO16780
Streptococcus suis ABP89154
Streptococcus suis ABP91346
Streptococcus suis ABP91044
Streptococcus equi CAW94361
Streptococcus equi CAW98774
Streptococcus pyogenes AAM79614
Streptococcus pyogenes AAT87189
Streptococcus pyogenes AAK34165
Streptococcus pyogenes ABF34206
Streptococcus pyogenes CAM30108
Streptococcus pyogenes ACI61349
Streptococcus pyogenes ABF38128
Streptococcus pyogenes AAZ51703
Streptococcus pyogenes BAC63947
Streptococcus pyogenes ABF32315
Streptococcus pyogenes ABF36138
Streptococcus pyogenes AAX72179
Clostridium beijerinckii ABR35931
Listeria monocytogenes CAC98451
Listeria w elshimeri CAK19746
Listeria innocua CAC95624
Listeria monocytogenes CAS04158
Listeria monocytogenes ACK40602
Listeria monocytogenes AAT03173
Thermoanaerobacter tengcongensis AAM23630
Thermoanaerobacter pseudethanolicus ABY95574
Thermoanaerobacter sp ABY91584
Exiguobacterium sp ACQ70812
Bacillus clausii BAD63025
Oceanobacillus iheyensis BAC14719
Bacillus clausii BAD62857
Bacillus pumilus AAM74558
Bacillus pumilus ABV62787
Bacillus subtilis BAA19708
Bacillus halodurans BAA75349
Bacillus amyloliquefaciens ABS75919
Bacillus licheniformis AAU24076
Geobacillus thermodenitrificans ABO67111
Geobacillus kaustophilus BAD76141
Geobacillus sp ACE77068
Yersinia enterocolitica CAL11067
Serratia marcescens BAB92992
Serratia proteamaculans ABV39942
Klebsiella pneumoniae ACI12102
Photobacterium profundum CAG23851
Aeromonas salmonicida ABO88650
Tolumonas auensis ACQ92079
Enterococcus faecalis AAO80824
No Species Information AAQ45559
Listeria w elshimeri CAK19958
Listeria innocua CAC95815
Listeria monocytogenes CAC98653

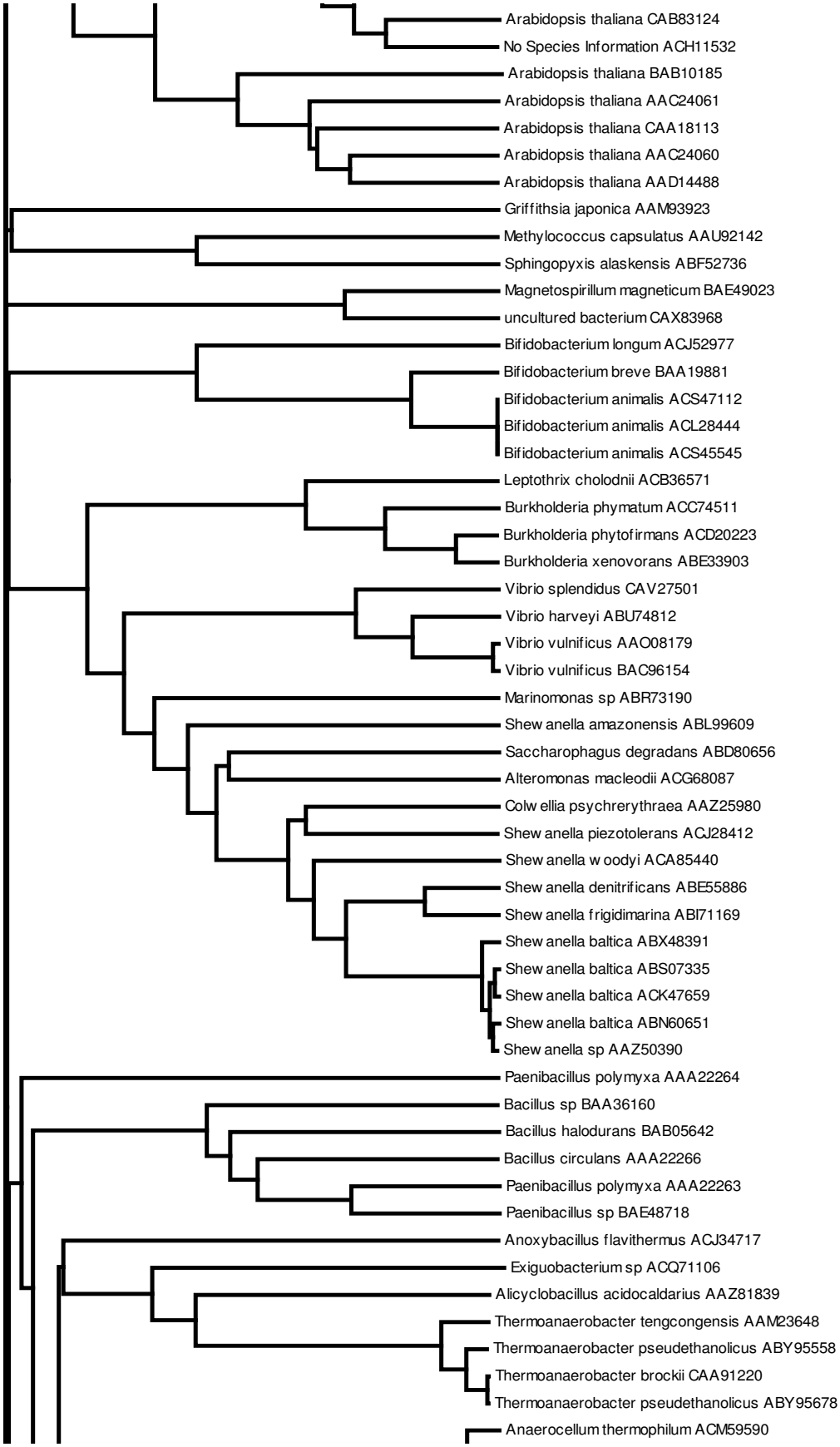

Listeria monocytogenes CAC98653
Listeria monocytogenes ACK40396
Listeria monocytogenes CAS04368
Listeria monocytogenes AAT03385
Cronobacter sakazakii ABU77277
Pectobacterium atrosepticum CAG77303
Pectobacterium carotovorum ABL14155
No Species Information ABH79653
Enterobacter sp ABP58726
Yersinia enterocolitica CAL11384
Escherichia fergusonii CAQ91392
Salmonella enterica ACN47927
Salmonella enterica AAV79423
Salmonella enterica AAO71234
Salmonella enterica AAL22633
Salmonella enterica ABX70004
Salmonella enterica CAR39438
Salmonella enterica ACF65267
Salmonella enterica ACF90852
Salmonella enterica ACH49506
Salmonella enterica CAR61655
Salmonella enterica AAX67601
Salmonella enterica CAD03219
Salmonella enterica ACF65956
Salmonella enterica CAR35174
Salmonella enterica ACH75793
Candida w ickerhamii AAC49036
Cellulomonas fimi AAA23091
Ruegeria sp ABF62281
Dinoroseobacter shibae ABV93396
Jannaschia sp ABD54861
Roseobacter denitrificans ABG31897
Rhodobacter sphaeroides ABP70047
Rhodobacter sphaeroides ABA79035
Rhodobacter sphaeroides ABN76630
Rhodobacter sphaeroides ACM01050
Caenorhabditis elegans AAC68766
Caenorhabditis brenneri ACI49059
Caenorhabditis elegans AAA83309
Aspergillus niger CAK40571
Sinorhizobium medicae ABR61616
Rhizobium sp ACP26687
Sinorhizobium meliloti CAC47470
Agrobacterium vitis ACM37755
Agrobacterium sp AAA22085
Agrobacterium tumefaciens AAK88957
Agrobacterium tumefaciens AAL45279
Agrobacterium radiobacter ACM27761
Rhizobium leguminosarum ACI56670
Rhizobium leguminosarum CAK09657
RhRhizobium etli ABC92395
Rhizobium etli ACE92848
Caulobacter sp ABZ72524

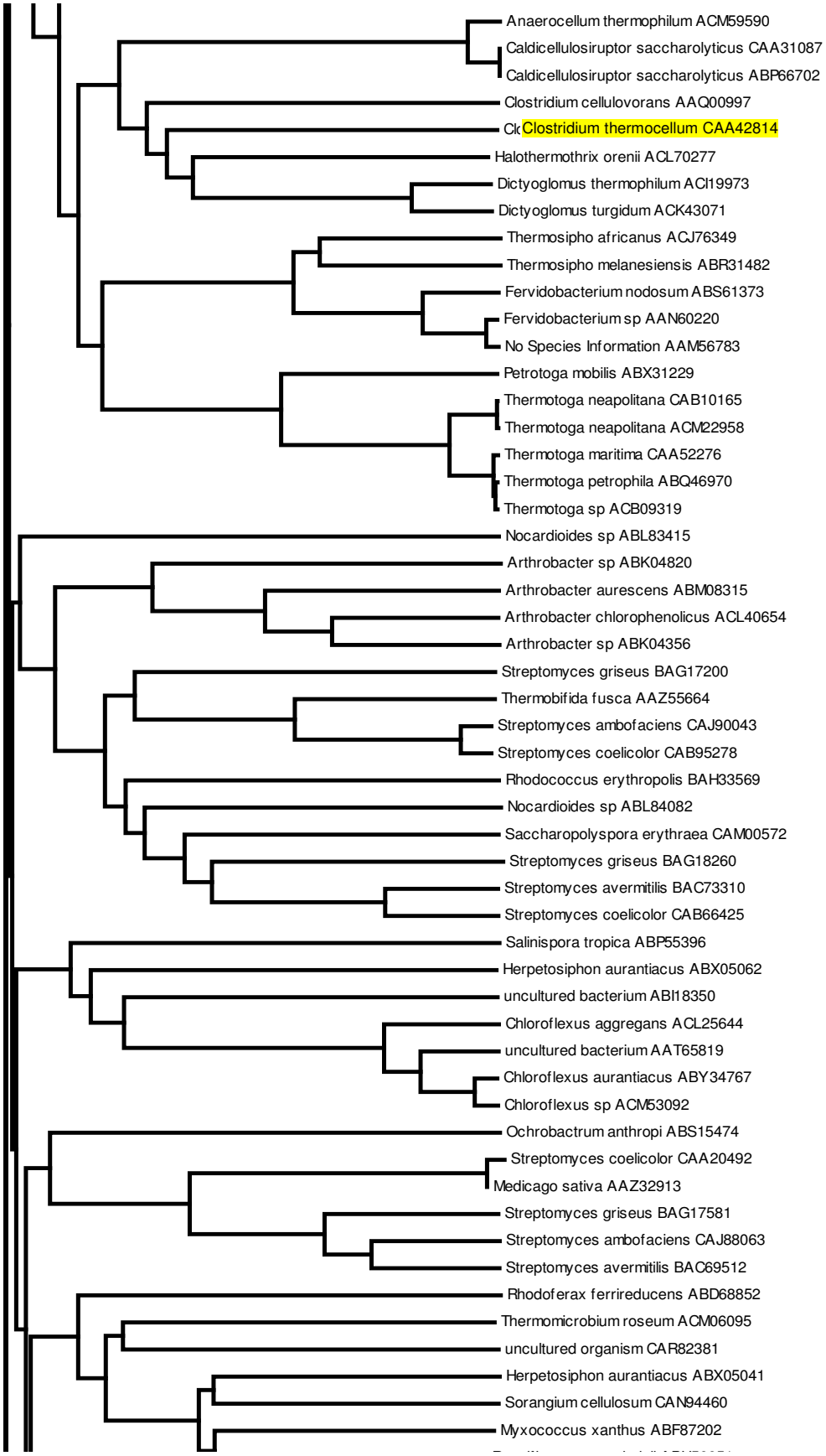

Caulobacter sp ABZ72524
Caulobacter crescentus AAK24107
Caulobacter crescentus ACL95685
Sinapis alba CAA42533
Sinapis alba CAA42536
Sinapis alba CAA42535
Orpinomyces sp AAD45834
Piromyces sp CAC34952
Piromyces sp AAP30744
Piromyces sp AAP30745
Clavibacter michiganensis AAP57289
Medicago truncatula ACJ85633
Medicago truncatula ACJ85833
Flavobacterium johnsoniae ABQ07856
Cytophaga hutchinsonii ABG61043
Lactobacillus johnsonii AAS08528
Clostridium acetobutylicum AAK80905
Exiguobacterium sp ACQ69978
Lactobacillus gasseri ABJ59750
No Species Information AAT17319
Streptococcus pneumoniae ACF55638
Streptococcus pneumoniae ACA36493
Streptococcus pneumoniae AAK74636
Streptococcus pneumoniae ABJ54778
Streptococcus pneumoniae AAK99228
Staphylococcus carnosus CAL27580
Lactobacillus delbrueckii CAM35428
Lactobacillus gasseri ABJ59900
No Species Information AAN97959
Lactobacillus casei AAA25238
Lactobacillus rhamnosus ACH91635
Lactobacillus casei ACG55692
Lactobacillus casei ABJ71621
Lactobacillus casei CAQ65853
Bacillus amyloliquefaciens ABS73581
No Species Information AAQ43959
Lactobacillus acidophilus BAA07122
Streptococcus uberis CAR43123
Streptococcus equi CAX00199
Streptococcus equi ACG61833
Staphylococcus haemolyticus BAE04157
Staphylococcus epidermidis AAO05422
Staphylococcus epidermidis AAW55165
No Species Information ABH91077
Staphylococcus aureus CAI81759
Staphylococcus aureus AAW37056
Staphylococcus aureus AAA26650
Staphylococcus aureus ABX30176
Staphylococcus aureus ABD21770
Staphylococcus aureus ABD31469
Staphylococcus aureus BAF68365
Staphylococcus aureus CAG41258
Staphylococcus aureus BAB43281

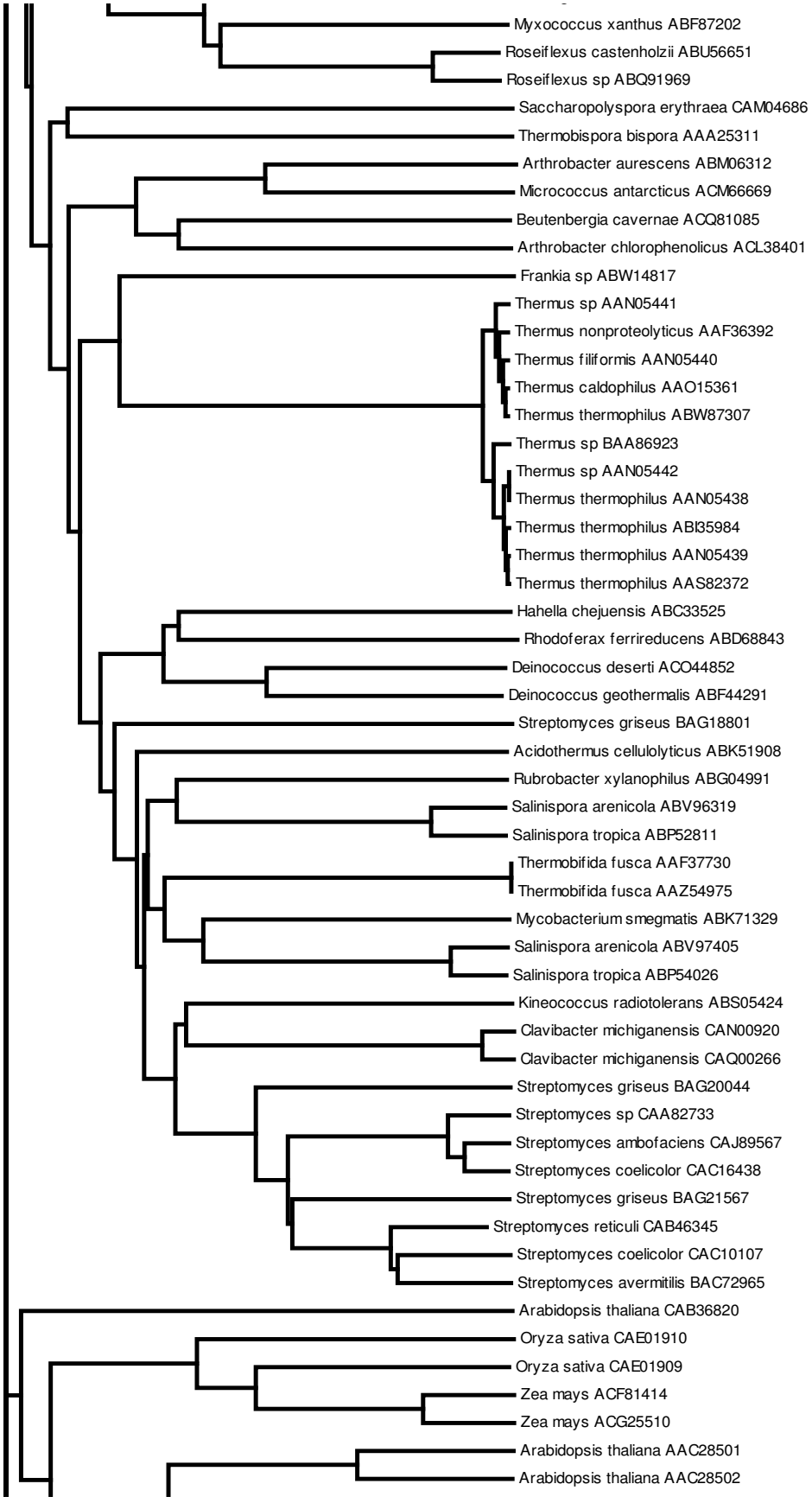

Staphylococcus aureus BAB43281
Staphylococcus aureus BAB95980
Staphylococcus aureus BAF79056
Staphylococcus aureus CAG43898
Staphylococcus aureus ABR53087
Staphylococcus aureus ABQ50001
No Species Information ABI03822
Streptococcus pneumoniae AAK99872
Streptococcus pneumoniae ACB90364
Streptococcus pneumoniae AAK75293
Streptococcus pneumoniae ABJ54974
Streptococcus pneumoniae ACF55578
Streptococcus pneumoniae ACO23459
Streptococcus pneumoniae ACO18396
Streptococcus pneumoniae ACA37023
Streptococcus mutans AAA16450
Streptococcus mutans AAN59144
Streptococcus sanguinis ABN45075
Streptococcus gordonii AAG39001
Streptococcus gordonii ABV10357
Streptococcus pneumoniae CAR68898
Streptococcus agalactiae CAD46988
No Species Information AAQ45157
Lactococcus garvieae ABW34725
Lactobacillus sakei ACD75529
Lactococcus lactis AAA25173
Lactococcus lactis ABA47363
Lactococcus lactis ABJ74133
Streptococcus suis ABP90002
Streptococcus suis ABP92206
Streptococcus uberis CAR41805
Streptococcus pyogenes ACI61842
Streptococcus pyogenes ABF34766
Streptococcus pyogenes ABF38679
Streptococcus pyogenes CAM30928
Streptococcus pyogenes AAT87776
Streptococcus pyogenes ABF36709
Streptococcus pyogenes ABF32822
Streptococcus pyogenes AAM80260
Streptococcus pyogenes BAC64746
Streptococcus pyogenes AAX72732
Streptococcus pyogenes AAZ52250
Streptococcus pyogenes AAK34620
Bombyx mori AAP13852
Homo sapiens AAQ89091
Mus musculus AAH30631
Danio rerio AAH76422
Danio rerio CAN88278
Mus musculus AAU95234
Danio rerio CAQ14839
Danio rerio CAQ14840
Xenopus laevis AAI54966

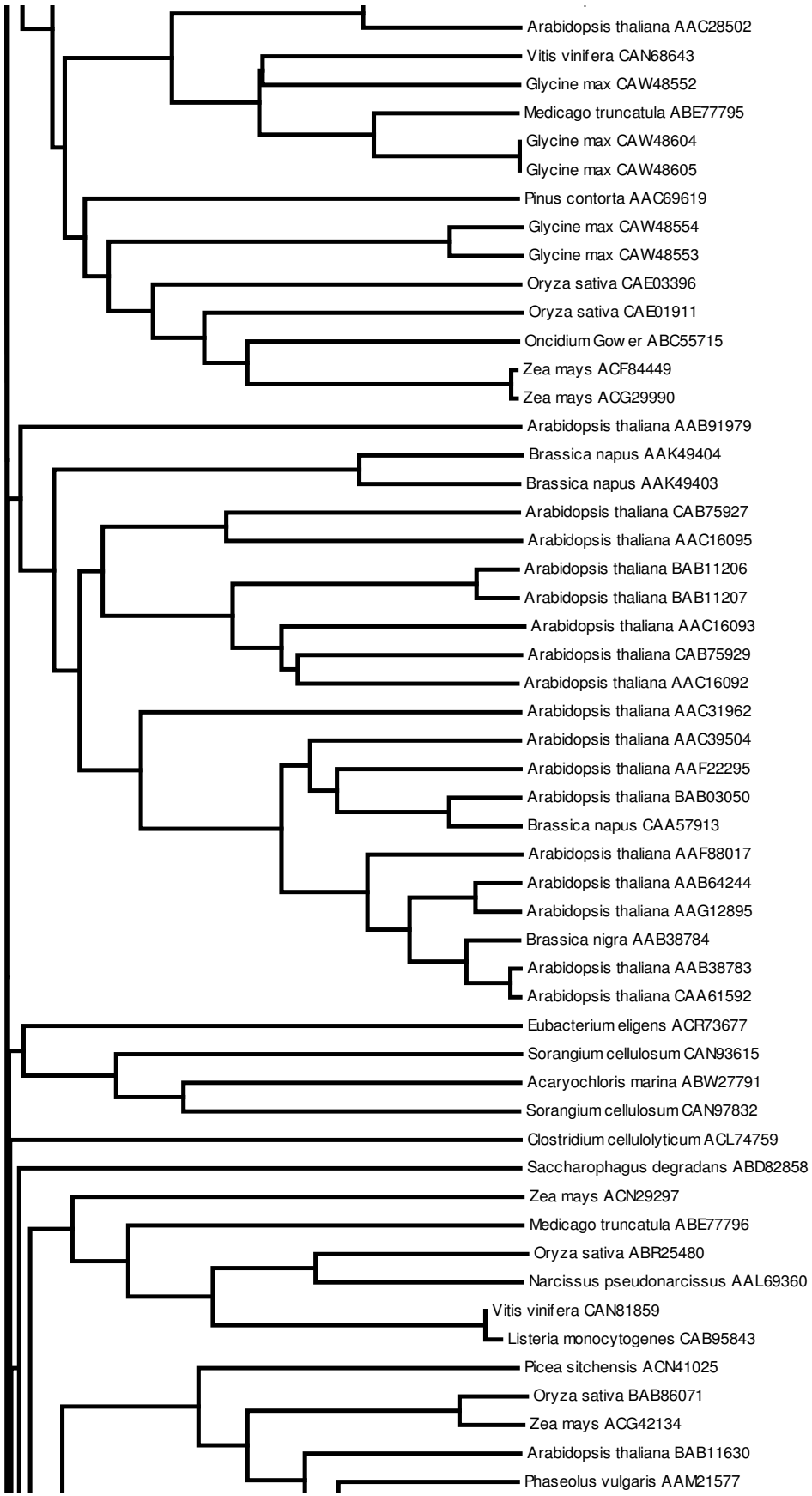

Xenopus laevis AAI54966
No Species Information AAI61734
Xenopus laevis AAI55890
Mus musculus BAE34332
Rattus norvegicus BAF94233
Homo sapiens CAA30801
Oryctolagus cuniculus CAA30802
Oryctolagus cuniculus CAA81690
Picea sitchensis ABK24806
Spodoptera frugiperda AAC06038
Papilio xuthus BAG30785
Drosophila melanogaster AAL39878
Papilio xuthus BAG30744
Rhyparobia maderae AAL40863
Cryptotermes cynocephalus ACD76733
Cryptotermes secundus ABN05620
Bactrocera cucurbitae ACN91517
Corbicula japonica BAG71912
Corbicula japonica BAG75455
Brevicoryne brassicae AAL25999
Neotermes koshunensis BAB91145
Tenebrio molitor AAG26008
Danio rerio AAH95794
No Species Information AAI21689
Xenopus laevis AAH81073
Cavia porcellus AAB41058
Rattus norvegicus AAI68903
Homo sapiens CAC08178
Pongo abelii CAH89592
Hypocrea jecorina AAP57758
Aspergillus oryzae BAE63197
Penicillium funiculosum ACI96344
Penicillium funiculosum ACO82080
Penicillium chrysogenum CAP91982
Aspergillus nidulans EAA63677
Phanerochaete chrysosporium BAE87009
Phanerochaete chrysosporium BAE87008
Talaromyces emersonii AAL89551
Aspergillus nidulans EAA65642
Penicillium chrysogenum CAP94053
Aspergillus niger CAK47813
Aspergillus oryzae BAE57671
Hypocrea lixii ABN50090
Hypocrea jecorina BAA74959
Trichoderma viride AAQ21384
Magnaporthe grisea AAX07701
Humicola grisea BAA74958
Podospora anserina CAP69093
Kosmotoga olearia ACR79740
Agrobacterium vitis ACM38274
Magnetospirillum magneticum BAE51034
uncultured bacterium CAX83969
Bradyrhizobium sp ABQ38386

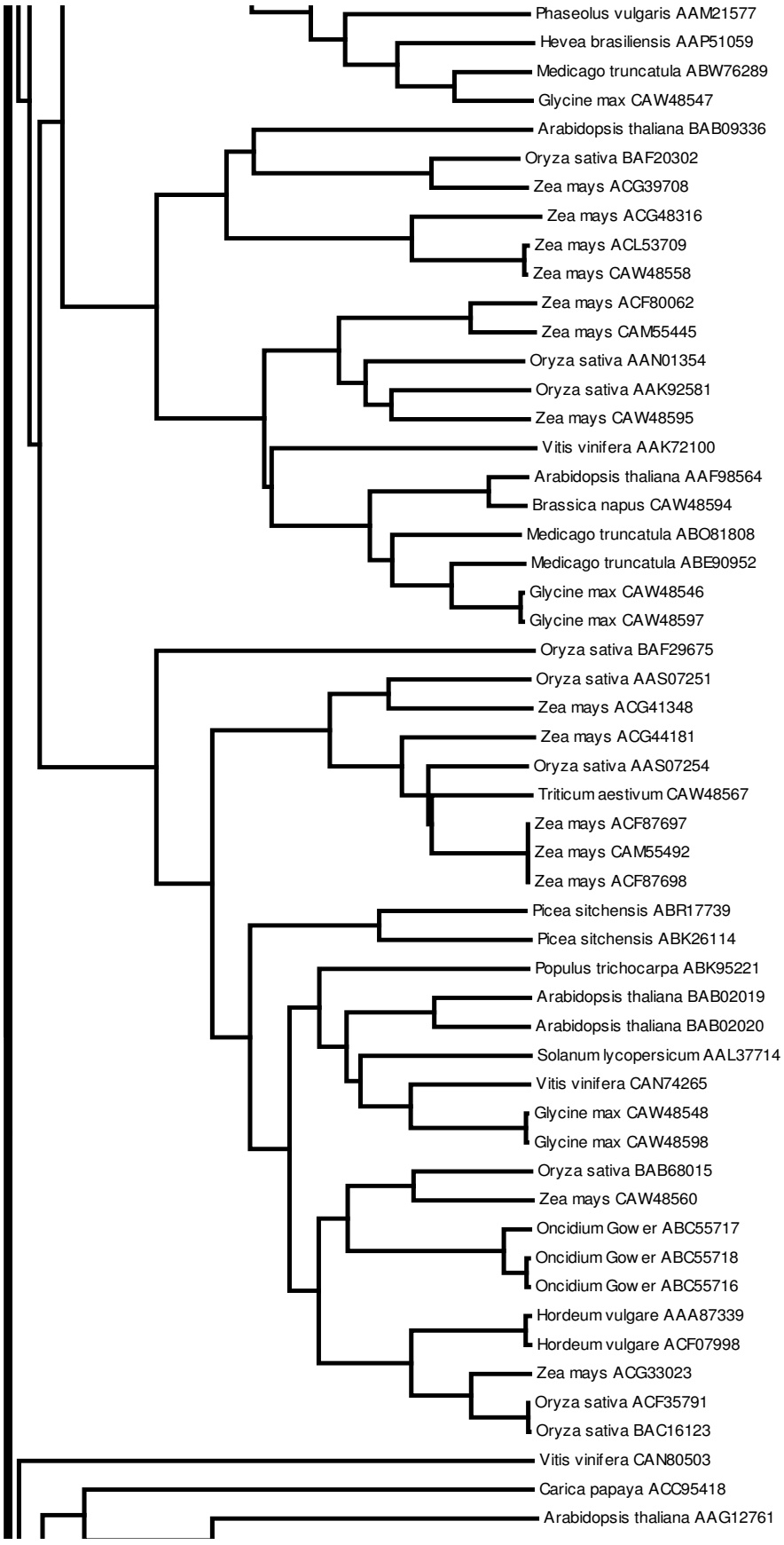

Bradyrhizobium sp ABQ38386
Bradyrhizobium sp CAL75457
Bradyrhizobium japonicum BAC49922
Rhodopseudomonas palustris ABJ08539
Rhodopseudomonas palustris CAE27177
Rhodopseudomonas palustris ACF00459
Rhodopseudomonas palustris ABE39076
Rhodopseudomonas palustris ABD08322
Desulfotalea psychrophila CAG37177
Bradyrhizobium japonicum BAC51442
Bradyrhizobium japonicum BAC46630
Bradyrhizobium sp ABQ36152
Bradyrhizobium sp CAL77609
uncultured organism CAR82380
uncultured organism CAR82382
Cyanothece sp ACL45395
Oryza sativa AAS79743
Zea mays ACR35092
Oryza sativa BAF25554
Oryza sativa BAF25553
Zea mays ACR35387
Oryza sativa BAF25552
Zea mays ACG43313
Zea mays CAW42914
Zea mays CAW48608
Zea mays CAW48556
Zea mays CAW48559
Oryza sativa AAS79738
Oryza sativa CAM55671
Oryza sativa AAS79741
Oryza sativa AAT38010
Zea mays ACG25343
Oryza sativa AAV31358
Oryza sativa BAB63225
Oryza sativa BAB63224
Oryza sativa BAH00985
Oryza sativa BAB67933
Zea mays ACG29862
Zea mays ACF87535
Zea mays ACF82391
Medicago truncatula ABE79608
Vitis vinifera CAN81257
Vitis vinifera CAN82274
Arabidopsis thaliana AAF02882
Medicago truncatula ABE84996
Medicago truncatula ABE85051
Glycine max CAW48549
Glycine max AAL92115
Glycine max CAW48596
Arabidopsis thaliana CAB83125
Arabidopsis thaliana CAB43970
Arabidopsis thaliana CAB43971
Arabidopsis thaliana CAB83124

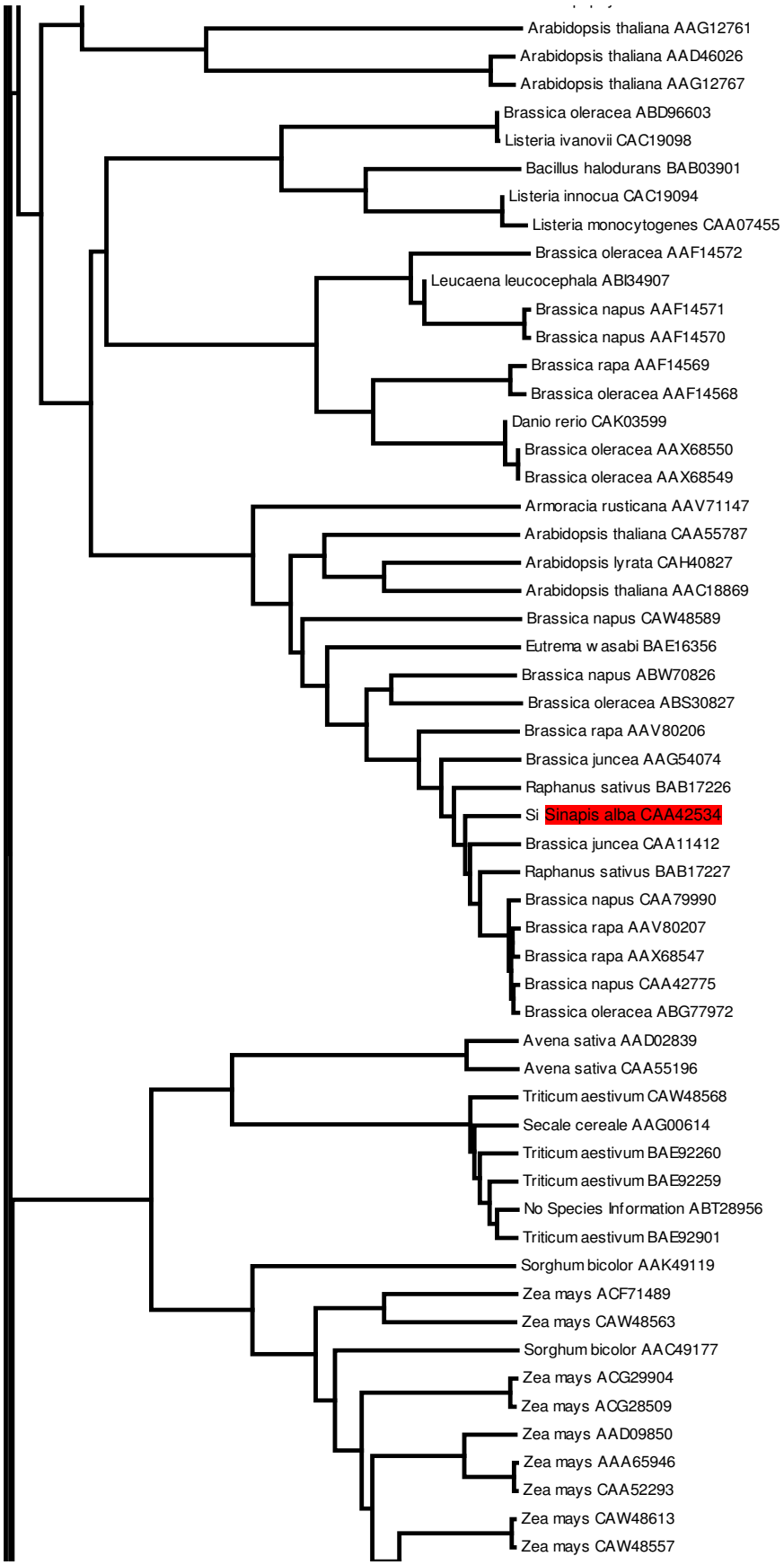

Arabidopsis thaliana CAB83124
No Species Information ACH11532
Arabidopsis thaliana BAB10185
Arabidopsis thaliana AAC24061
Arabidopsis thaliana CAA18113
Arabidopsis thaliana AAC24060
Arabidopsis thaliana AAD14488
Griffithsia japonica AAM93923
Methylococcus capsulatus AAU92142
Sphingopyxis alaskensis ABF52736
Magnetospirillum magneticum BAE49023
uncultured bacterium CAX83968
Bifidobacterium longum ACJ52977
Bifidobacterium breve BAA19881
Bifidobacterium animalis ACS47112
Bifidobacterium animalis ACL28444
Bifidobacterium animalis ACS45545
Leptothrix cholodnii ACB36571
Burkholderia phymatum ACC74511
Burkholderia phytofirmans ACD20223
Burkholderia xenovorans ABE33903
Vibrio splendidus CAV27501
Vibrio harveyi ABU74812
Vibrio vulnificus AAO08179
Vibrio vulnificus BAC96154
Marinomonas sp ABR73190
Shew anella amazonensis ABL99609
Saccharophagus degradans ABD80656
Alteromonas macleodii ACG68087
Colw ellia psychrerythraea AAZ25980
Shew anella piezotolerans ACJ28412
Shew anella w oodyi ACA85440
Shew anella denitrificans ABE55886
Shew anella frigidimarina ABI71169
Shew anella baltica ABX48391
Shew anella baltica ABS07335
Shew anella baltica ACK47659
Shew anella baltica ABN60651
Shew anella sp AAZ50390
Paenibacillus polymyxa AAA22264
Bacillus sp BAA36160
Bacillus halodurans BAB05642
Bacillus circulans AAA22266
Paenibacillus polymyxa AAA22263
Paenibacillus sp BAE48718
Anoxybacillus flavithermus ACJ34717
Exiguobacterium sp ACQ71106
Alicyclobacillus acidocaldarius AAZ81839
Thermoanaerobacter tengcongensis AAM23648
Thermoanaerobacter pseudethanolicus ABY95558
Thermoanaerobacter brockii CAA91220
Thermoanaerobacter pseudethanolicus ABY95678
Anaerocellum thermophilum ACM59590

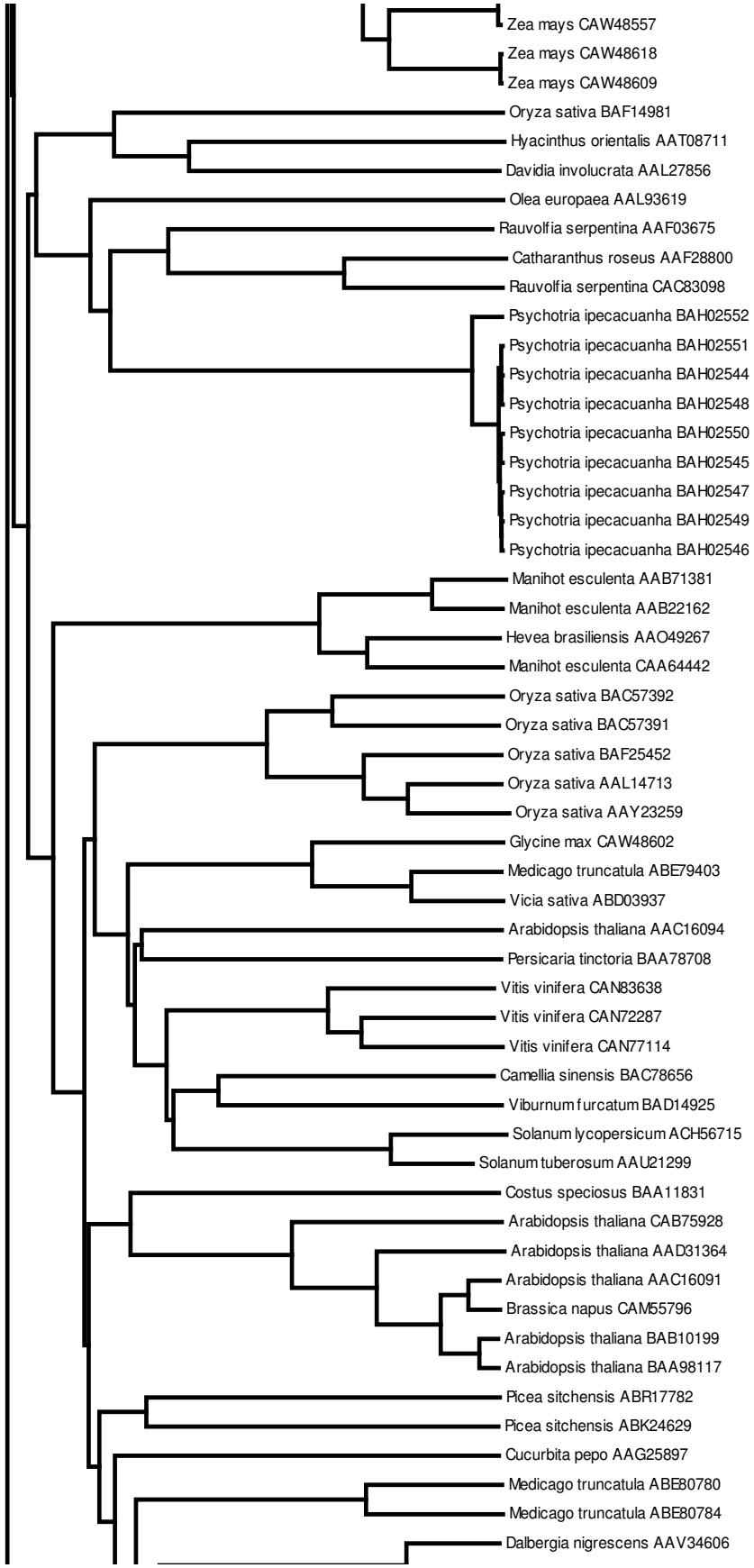

Anaerocellum thermophilum ACM59590
Caldicellulosiruptor saccharolyticus CAA31087
Caldicellulosiruptor saccharolyticus ABP66702
Clostridium cellulovorans AAQ00997
Clostridium thermocellum CAA42814
Halothermothrix orenii ACL70277
Dictyoglomus thermophilum ACI19973
Dictyoglomus turgidum ACK43071
Thermosipho africanus ACJ76349
Thermosipho melanesiensis ABR31482
Fervidobacterium nodosum ABS61373
Fervidobacterium sp AAN60220
No Species Information AAM56783
Petrotoga mobilis ABX31229
Thermotoga neapolitana CAB10165
Thermotoga neapolitana ACM22958
Thermotoga maritima CAA52276
Thermotoga petrophila ABQ46970
Thermotoga sp ACB09319
Nocardioides sp ABL83415
Arthrobacter sp ABK04820
Arthrobacter aurescens ABM08315
Arthrobacter chlorophenolicus ACL40654
Arthrobacter sp ABK04356
Streptomyces griseus BAG17200
Thermobifida fusca AAZ55664
Streptomyces ambofaciens CAJ90043
Streptomyces coelicolor CAB95278
Rhodococcus erythropolis BAH33569
Nocardioides sp ABL84082
Saccharopolyspora erythraea CAM00572
Streptomyces griseus BAG18260
Streptomyces avermitilis BAC73310
Streptomyces coelicolor CAB66425
Salinispora tropica ABP55396
Herpetosiphon aurantiacus ABX05062
uncultured bacterium ABI18350
Chloroflexus aggregans ACL25644
uncultured bacterium AAT65819
Chloroflexus aurantiacus ABY34767
Chloroflexus sp ACM53092
Ochrobactrum anthropi ABS15474
Streptomyces coelicolor CAA20492
Medicago sativa AAZ32913
Streptomyces griseus BAG17581
Streptomyces ambofaciens CAJ88063
Streptomyces avermitilis BAC69512
Rhodoferax ferrireducens ABD68852
Thermomicrobium roseum ACM06095
uncultured organism CAR82381
Herpetosiphon aurantiacus ABX05041
Sorangium cellulosum CAN94460
Myxococcus xanthus ABF87202

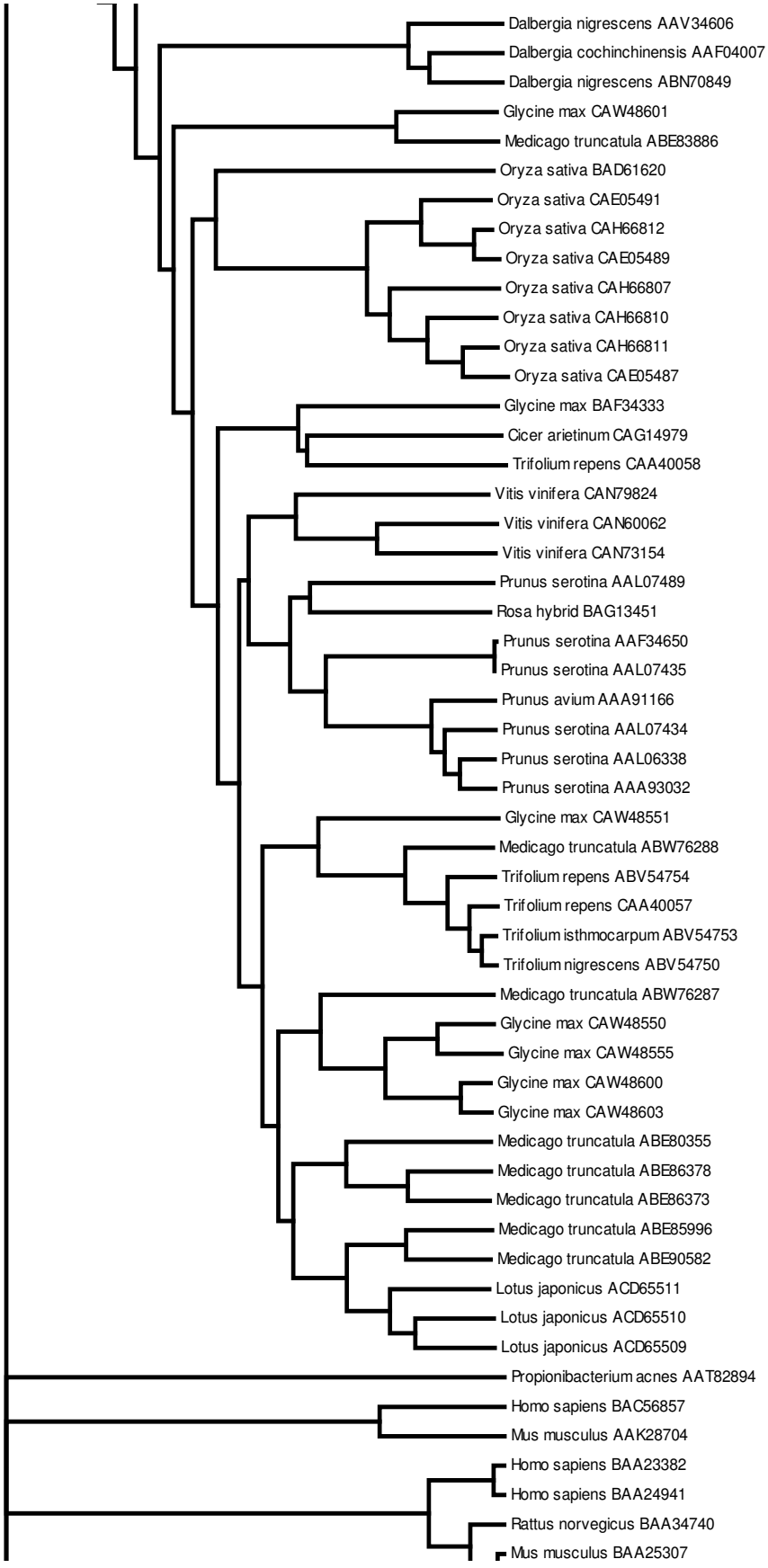

Myxococcus xanthus ABF87202
Roseiflexus castenholzii ABU56651
Roseiflexus sp ABQ91969
Saccharopolyspora erythraea CAM04686
Thermobispora bispora AAA25311
Arthrobacter aurescens ABM06312
Micrococcus antarcticus ACM66669
Beutenbergia cavernae ACQ81085
Arthrobacter chlorophenolicus ACL38401
Frankia sp ABW14817
Thermus sp AAN05441
Thermus nonproteolyticus AAF36392
Thermus filiformis AAN05440
Thermus caldophilus AAO15361
Thermus thermophilus ABW87307
Thermus sp BAA86923
Thermus sp AAN05442
Thermus thermophilus AAN05438
Thermus thermophilus ABI35984
Thermus thermophilus AAN05439
Thermus thermophilus AAS82372
Hahella chejuensis ABC33525
Rhodoferax ferrireducens ABD68843
Deinococcus deserti ACO44852
Deinococcus geothermalis ABF44291
Streptomyces griseus BAG18801
Acidothermus cellulolyticus ABK51908
Rubrobacter xylanophilus ABG04991
Salinispora arenicola ABV96319
Salinispora tropica ABP52811
Thermobifida fusca AAF37730
Thermobifida fusca AAZ54975
Mycobacterium smegmatis ABK71329
Salinispora arenicola ABV97405
Salinispora tropica ABP54026
Kineococcus radiotolerans ABS05424
Clavibacter michiganensis CAN00920
Clavibacter michiganensis CAQ00266
Streptomyces griseus BAG20044
Streptomyces sp CAA82733
Streptomyces ambofaciens CAJ89567
Streptomyces coelicolor CAC16438
Streptomyces griseus BAG21567
Streptomyces reticuli CAB46345
Streptomyces coelicolor CAC10107
Streptomyces avermitilis BAC72965
Arabidopsis thaliana CAB36820
Oryza sativa CAE01910
Oryza sativa CAE01909
Zea mays ACF81414
Zea mays ACG25510
Arabidopsis thaliana AAC28501
Arabidopsis thaliana AAC28502

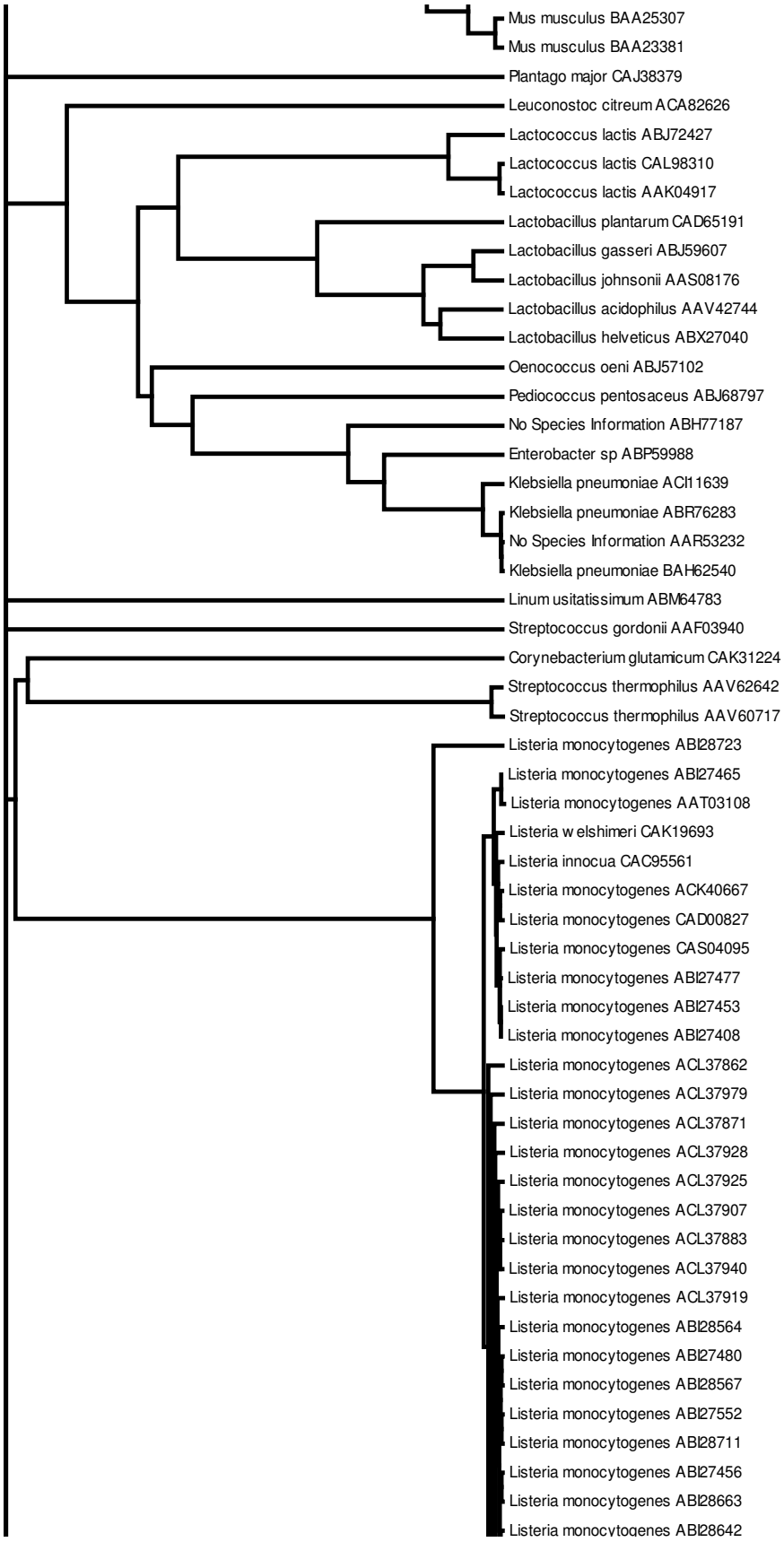

Arabidopsis thaliana AAC28502
Vitis vinifera CAN68643
Glycine max CAW48552
Medicago truncatula ABE77795
Glycine max CAW48604
Glycine max CAW48605
Pinus contorta AAC69619
Glycine max CAW48554
Glycine max CAW48553
Oryza sativa CAE03396
Oryza sativa CAE01911
Oncidium Gower ABC55715
Zea mays ACF84449
Zea mays ACG29990
Arabidopsis thaliana AAB91979
Brassica napus AAK49404
Brassica napus AAK49403
Arabidopsis thaliana CAB75927
Arabidopsis thaliana AAC16095
Arabidopsis thaliana BAB11206
Arabidopsis thaliana BAB11207
Arabidopsis thaliana AAC16093
Arabidopsis thaliana CAB75929
Arabidopsis thaliana AAC16092
Arabidopsis thaliana AAC31962
Arabidopsis thaliana AAC39504
Arabidopsis thaliana AAF22295
Arabidopsis thaliana BAB03050
Brassica napus CAA57913
Arabidopsis thaliana AAF88017
Arabidopsis thaliana AAB64244
Arabidopsis thaliana AAG12895
Brassica nigra AAB38784
Arabidopsis thaliana AAB38783
Arabidopsis thaliana CAA61592
Eubacterium eligens ACR73677
Sorangium cellulosum CAN93615
Acaryochloris marina ABW27791
Sorangium cellulosum CAN97832
Clostridium cellulolyticum ACL74759
Saccharophagus degradans ABD82858
Zea mays ACN29297
Medicago truncatula ABE77796
Oryza sativa ABR25480
Narcissus pseudonarcissus AAL69360
Vitis vinifera CAN81859
Listeria monocytogenes CAB95843
Picea sitchensis ACN41025
Oryza sativa BAB86071
Zea mays ACG42134
Arabidopsis thaliana BAB11630
Phaseolus vulgaris AAM21577


Phaseolus vulgaris AAM21577
Hevea brasiliensis AAP51059
Medicago truncatula ABW76289
Glycine max CAW48547
Arabidopsis thaliana BAB09336
Oryza sativa BAF20302
Zea mays ACG39708
Zea mays ACG48316
Zea mays ACL53709
Zea mays CAW48558
Zea mays ACF80062
Zea mays CAM55445
Oryza sativa AAN01354
Oryza sativa AAK92581
Zea mays CAW48595
Vitis vinifera AAK72100
Arabidopsis thaliana AAF98564
Brassica napus CAW48594
Medicago truncatula ABO81808
Medicago truncatula ABE90952
Glycine max CAW48546
Glycine max CAW48597
Oryza sativa BAF29675
Oryza sativa AAS07251
Zea mays ACG41348
Zea mays ACG44181
Oryza sativa AAS07254
Triticum aestivum CAW48567
Zea mays ACF87697
Zea mays CAM55492
Zea mays ACF87698
Picea sitchensis ABR17739
Picea sitchensis ABK26114
Populus trichocarpa ABK95221
Arabidopsis thaliana BAB02019
Arabidopsis thaliana BAB02020
Solanum lycopersicum AAL37714
Vitis vinifera CAN74265
Glycine max CAW48548
Glycine max CAW48598
Oryza sativa BAB68015
Zea mays CAW48560
Oncidium Gower ABC55717
Oncidium Gower ABC55718
Oncidium Gower ABC55716
Hordeum vulgare AAA87339
Hordeum vulgare ACF07998
Zea mays ACG33023
Oryza sativa ACF35791
Oryza sativa BAC16123
Vitis vinifera CAN80503
Carica papaya ACC95418
Arabidopsis thaliana AAG12761

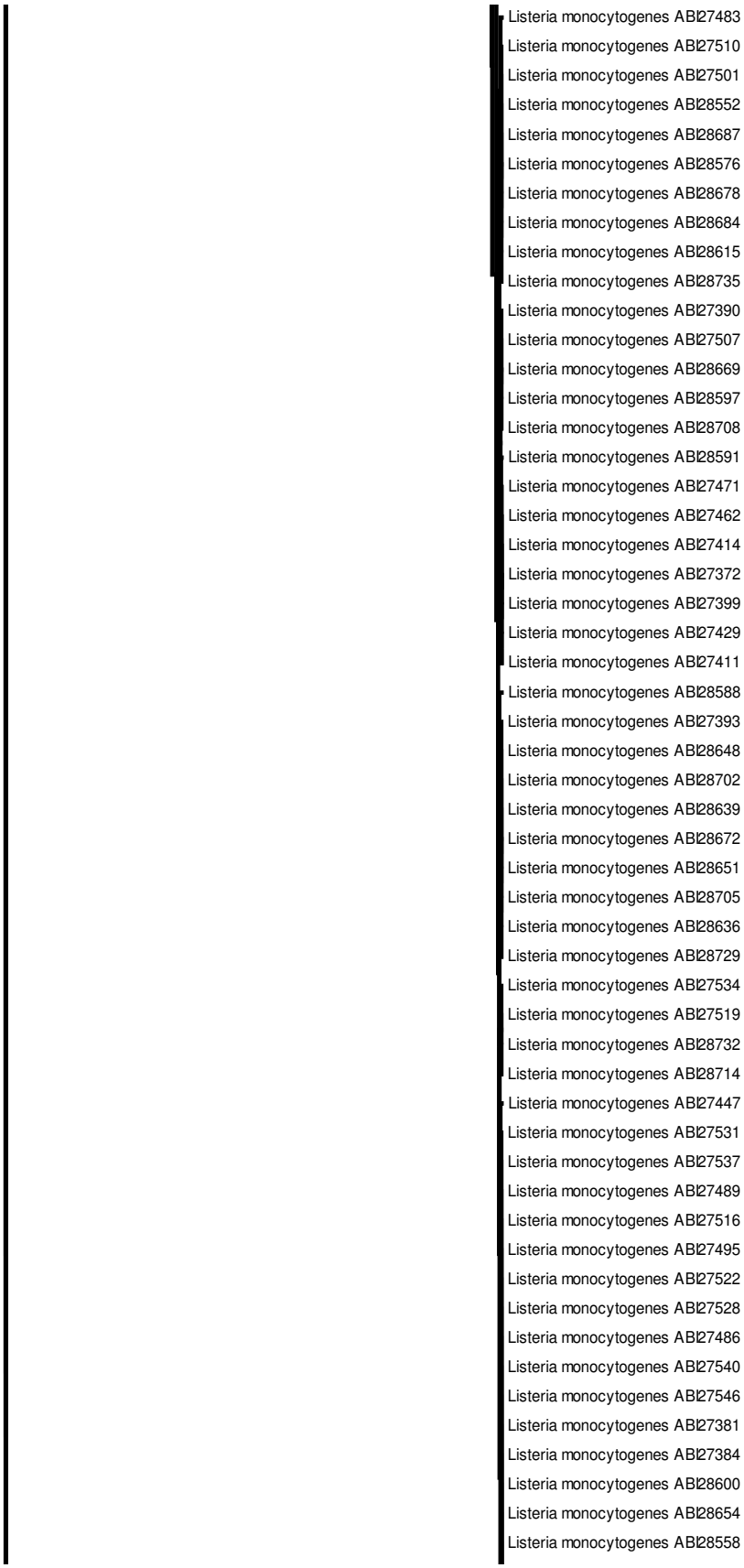

Arabidopsis thaliana AAG12761
Arabidopsis thaliana AAD46026
Arabidopsis thaliana AAG12767
Brassica oleracea ABD96603
Listeria ivanovii CAC19098
Bacillus halodurans BAB03901
Listeria innocua CAC19094
Listeria monocytogenes CAA07455
Brassica oleracea AAF14572
Leucaena leucocephala ABI34907
Brassica napus AAF14571
Brassica napus AAF14570
Brassica rapa AAF14569
Brassica oleracea AAF14568
Danio rerio CAK03599
Brassica oleracea AAX68550
Brassica oleracea AAX68549
Armoracia rusticana AAV71147
Arabidopsis thaliana CAA55787
Arabidopsis lyrata CAH40827
Arabidopsis thaliana AAC18869
Brassica napus CAW48589
Eutrema w asabi BAE16356
Brassica napus ABW70826
Brassica oleracea ABS30827
Brassica rapa AAV80206
Brassica juncea AAG54074
Raphanus sativus BAB17226
Si Sinapis alba CAA42534
Brassica juncea CAA11412
Raphanus sativus BAB17227
Brassica napus CAA79990
Brassica rapa AAV80207
Brassica rapa AAX68547
Brassica napus CAA42775
Brassica oleracea ABG77972
Avena sativa AAD02839
Avena sativa CAA55196
Triticum aestivum CAW48568
Secale cereale AAG00614
Triticum aestivum BAE92260
Triticum aestivum BAE92259
No Species Information ABT28956
Triticum aestivum BAE92901
Sorghum bicolor AAK49119
Zea mays ACF71489
Zea mays CAW48563
Sorghum bicolor AAC49177
Zea mays ACG29904
Zea mays ACG28509
Zea mays AAD09850
Zea mays AAA65946
Zea mays CAA52293
Zea mays CAW48613
Zea mays CAW48557

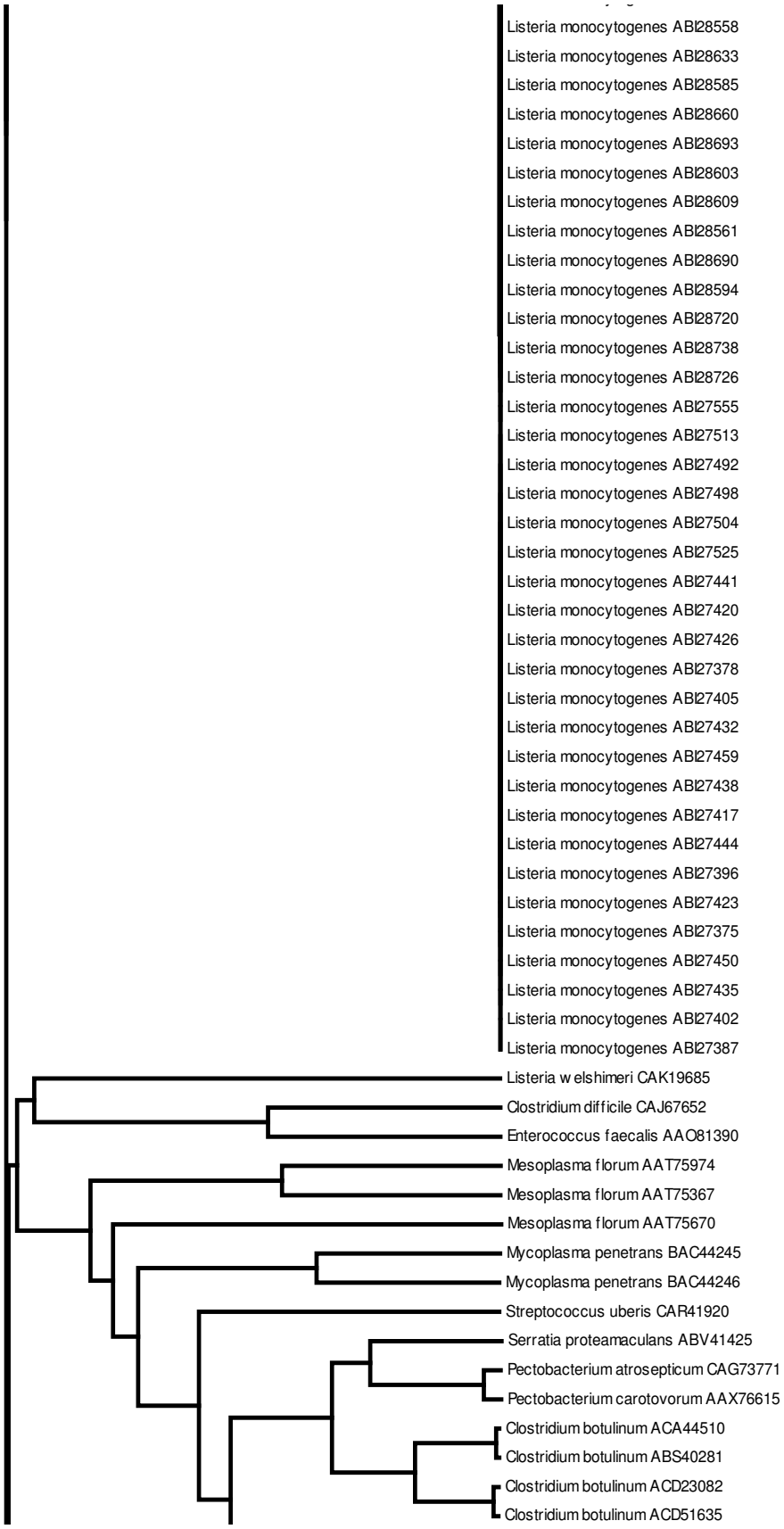

Zea mays CAW48557
Zea mays CAW48618
Zea mays CAW48609
Oryza sativa BAF14981
Hyacinthus orientalis AAT08711
Davidia involucrata AAL27856
Olea europaea AAL93619
Rauvolfia serpentina AAF03675
Catharanthus roseus AAF28800
Rauvolfia serpentina CAC83098
Psychotria ipecacuanha BAH02552
Psychotria ipecacuanha BAH02551
Psychotria ipecacuanha BAH02544
Psychotria ipecacuanha BAH02548
Psychotria ipecacuanha BAH02550
Psychotria ipecacuanha BAH02545
Psychotria ipecacuanha BAH02547
Psychotria ipecacuanha BAH02549
Psychotria ipecacuanha BAH02546
Manihot esculenta AAB71381
Manihot esculenta AAB22162
Hevea brasiliensis AAO49267
Manihot esculenta CAA64442
Oryza sativa BAC57392
Oryza sativa BAC57391
Oryza sativa BAF25452
Oryza sativa AAL14713
Oryza sativa AAY23259
Glycine max CAW48602
Medicago truncatula ABE79403
Vicia sativa ABD03937
Arabidopsis thaliana AAC16094
Persicaria tinctoria BAA78708
Vitis vinifera CAN83638
Vitis vinifera CAN72287
Vitis vinifera CAN77114
Camellia sinensis BAC78656
Viburnum furcatum BAD14925
Solanum lycopersicum ACH56715
Solanum tuberosum AAU21299
Costus speciosus BAA11831
Arabidopsis thaliana CAB75928
Arabidopsis thaliana AAD31364
Arabidopsis thaliana AAC16091
Brassica napus CAM55796
Arabidopsis thaliana BAB10199
Arabidopsis thaliana BAA98117
Picea sitchensis ABR17782
Picea sitchensis ABK24629
Cucurbita pepo AAG25897
Medicago truncatula ABE80780
Medicago truncatula ABE80784
Dalbergia nigrescens AAV34606

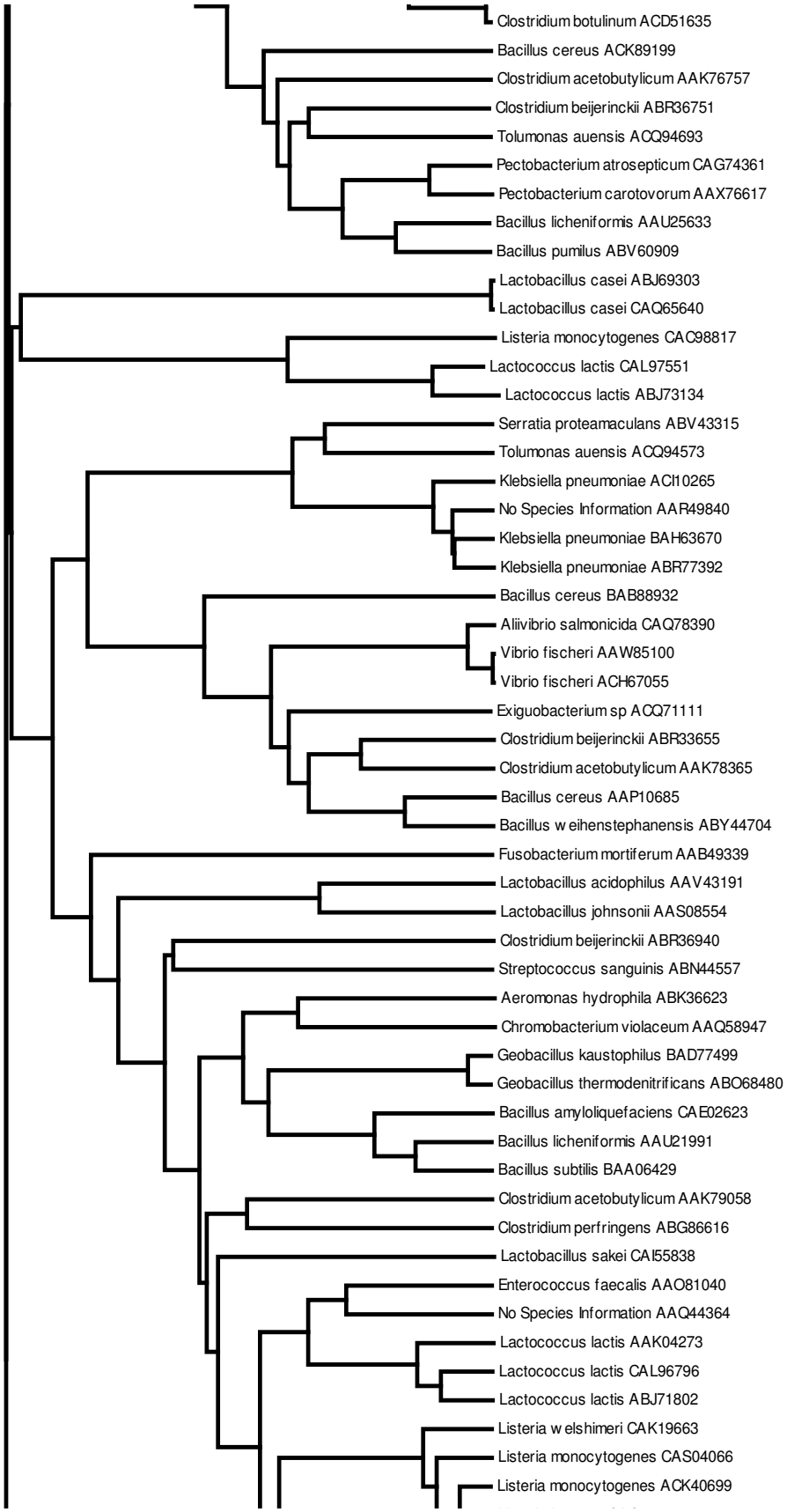

Dalbergia nigrescens AAV34606
Dalbergia cochinchinensis AAF04007
Dalbergia nigrescens ABN70849
Glycine max CAW48601
Medicago truncatula ABE83886
Oryza sativa BAD61620
Oryza sativa CAE05491
Oryza sativa CAH66812
Oryza sativa CAE05489
Oryza sativa CAH66807
Oryza sativa CAH66810
Oryza sativa CAH66811
Oryza sativa CAE05487
Glycine max BAF34333
Cicer arietinum CAG14979
Trifolium repens CAA40058
Vitis vinifera CAN79824
Vitis vinifera CAN60062
Vitis vinifera CAN73154
Prunus serotina AAL07489
Rosa hybrid BAG13451
Prunus serotina AAF34650
Prunus serotina AAL07435
Prunus avium AAA91166
Prunus serotina AAL07434
Prunus serotina AAL06338
Prunus serotina AAA93032
Glycine max CAW48551
Medicago truncatula ABW76288
Trifolium repens ABV54754
Trifolium repens CAA40057
Trifolium isthmocarpum ABV54753
Trifolium nigrescens ABV54750
Medicago truncatula ABW76287
Glycine max CAW48550
Glycine max CAW48555
Glycine max CAW48600
Glycine max CAW48603
Medicago truncatula ABE80355
Medicago truncatula ABE86378
Medicago truncatula ABE86373
Medicago truncatula ABE85996
Medicago truncatula ABE90582
Lotus japonicus ACD65511
Lotus japonicus ACD65510
Lotus japonicus ACD65509
Propionibacterium acnes AAT82894
Homo sapiens BAC56857
Mus musculus AAK28704
Homo sapiens BAA23382
Homo sapiens BAA24941
Rattus norvegicus BAA34740
Mus musculus BAA25307

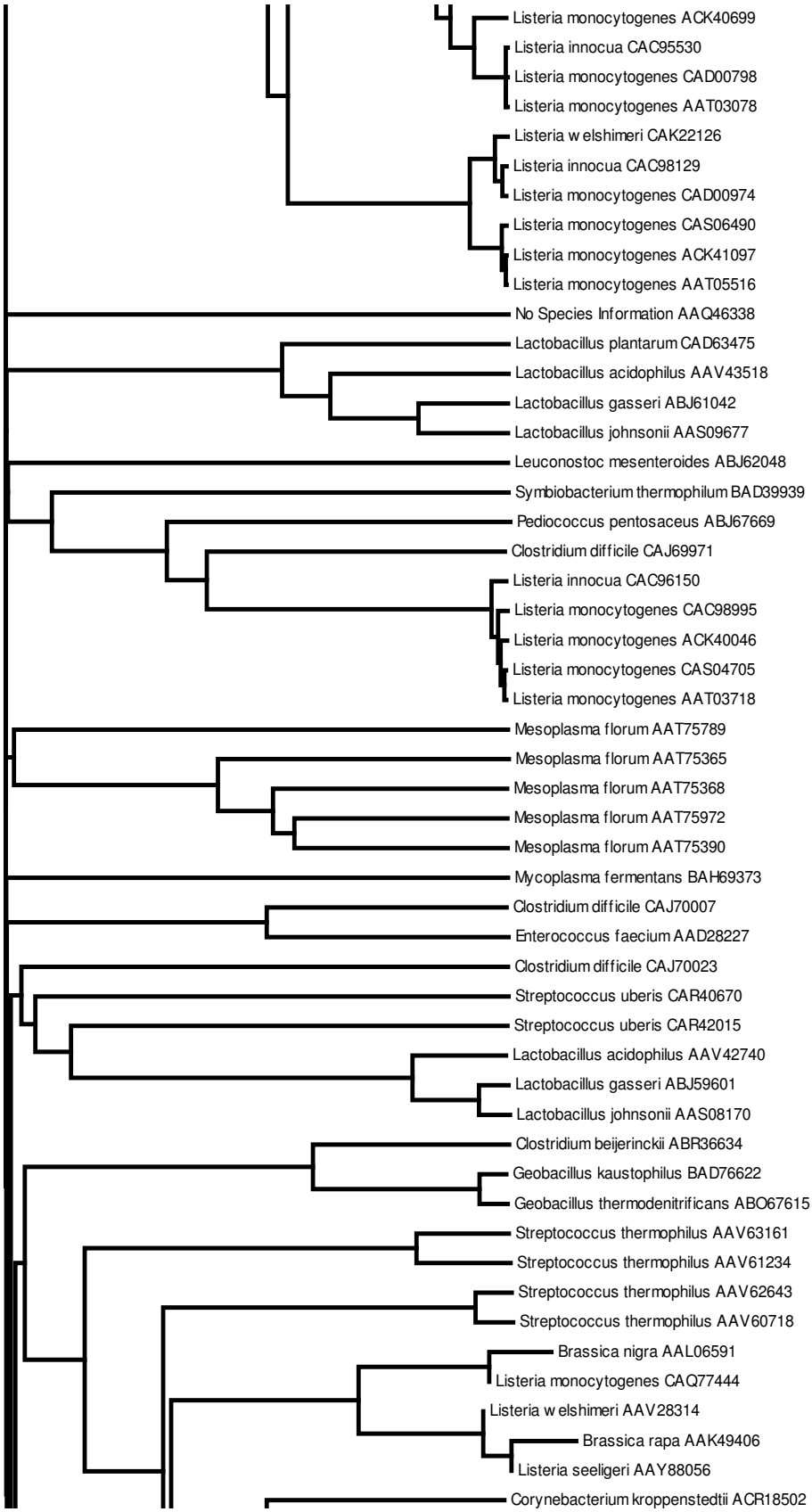

Mus musculus BAA25307
Mus musculus BAA23381
Plantago major CAJ38379
Leuconostoc citreum ACA82626
Lactococcus lactis ABJ72427
Lactococcus lactis CAL98310
Lactococcus lactis AAK04917
Lactobacillus plantarum CAD65191
Lactobacillus gasseri ABJ59607
Lactobacillus johnsonii AAS08176
Lactobacillus acidophilus AAV42744
Lactobacillus helveticus ABX27040
Oenococcus oeni ABJ57102
Pediococcus pentosaceus ABJ68797
No Species Information ABH77187
Enterobacter sp ABP59988
Klebsiella pneumoniae ACI11639
Klebsiella pneumoniae ABR76283
No Species Information AAR53232
Klebsiella pneumoniae BAH62540
Linum usitatissimum ABM64783
Streptococcus gordonii AAF03940
Corynebacterium glutamicum CAK31224
Streptococcus thermophilus AAV62642
Streptococcus thermophilus AAV60717
Listeria monocytogenes ABI28723
Listeria monocytogenes ABI27465
Listeria monocytogenes AAT03108
Listeria welshimeri CAK19693
Listeria innocua CAC95561
Listeria monocytogenes ACK40667
Listeria monocytogenes CAD00827
Listeria monocytogenes CAS04095
Listeria monocytogenes ABI27477
Listeria monocytogenes ABI27453
Listeria monocytogenes ABI27408
Listeria monocytogenes ACL37862
Listeria monocytogenes ACL37979
Listeria monocytogenes ACL37871
Listeria monocytogenes ACL37928
Listeria monocytogenes ACL37925
Listeria monocytogenes ACL37907
Listeria monocytogenes ACL37883
Listeria monocytogenes ACL37940
Listeria monocytogenes ACL37919
Listeria monocytogenes ABI28564
Listeria monocytogenes ABI27480
Listeria monocytogenes ABI28567
Listeria monocytogenes ABI27552
Listeria monocytogenes ABI28711
Listeria monocytogenes ABI27456
Listeria monocytogenes ABI28663
Listeria monocytogenes ABI28642

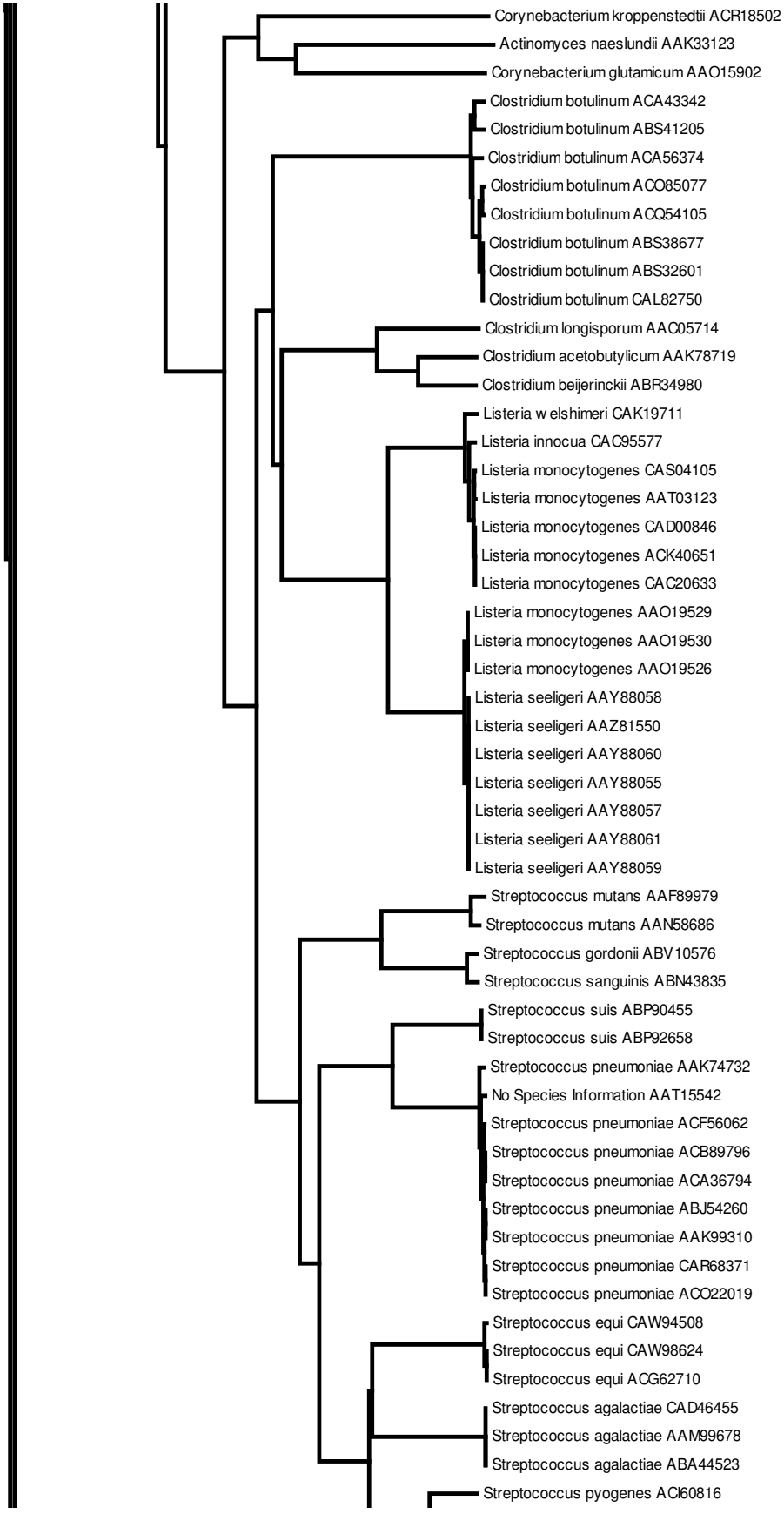

Listeria monocytogenes ABI28642
Listeria monocytogenes ABI28579
Listeria monocytogenes ABI28582
Listeria monocytogenes ABI28606
Listeria monocytogenes ACL37952
Listeria monocytogenes ACL37985
Listeria monocytogenes ACL37865
Listeria monocytogenes ACL37850
Listeria monocytogenes ACL37910
Listeria monocytogenes ACL37853
Listeria monocytogenes ABI28657
Listeria monocytogenes ABI27468
Listeria monocytogenes ABI28630
Listeria monocytogenes ABI28573
Listeria monocytogenes ABI28696
Listeria monocytogenes ABI28675
Listeria monocytogenes ABI28681
Listeria monocytogenes ABI28612
Listeria monocytogenes ABI28666
Listeria monocytogenes ABI28618
Listeria monocytogenes ABI28570
Listeria monocytogenes ABI28555
Listeria monocytogenes ABI28645
Listeria monocytogenes ACL37880
Listeria monocytogenes ACL37844
Listeria monocytogenes ACL37922
Listeria monocytogenes ABI28717
Listeria monocytogenes ACL37973
Listeria monocytogenes ACL37970
Listeria monocytogenes ACL37931
Listeria monocytogenes ACL37937
Listeria monocytogenes ABI27543
Listeria monocytogenes ABI28699
Listeria monocytogenes ABI27549
Listeria monocytogenes ACL37886
Listeria monocytogenes ACL37901
Listeria monocytogenes ACL37847
Listeria monocytogenes ACL37964
Listeria monocytogenes ACL37889
Listeria monocytogenes ACL37946
Listeria monocytogenes ACL37877
Listeria monocytogenes ACL37868
Listeria monocytogenes ACL37976
Listeria monocytogenes ACL37859
Listeria monocytogenes ACL37856
Listeria monocytogenes ACL37955
Listeria monocytogenes ACL37943
Listeria monocytogenes ABI28624
Listeria monocytogenes ABI27474
Listeria monocytogenes ABI28627
Listeria monocytogenes ABI28621
Listeria monocytogenes ABI27483

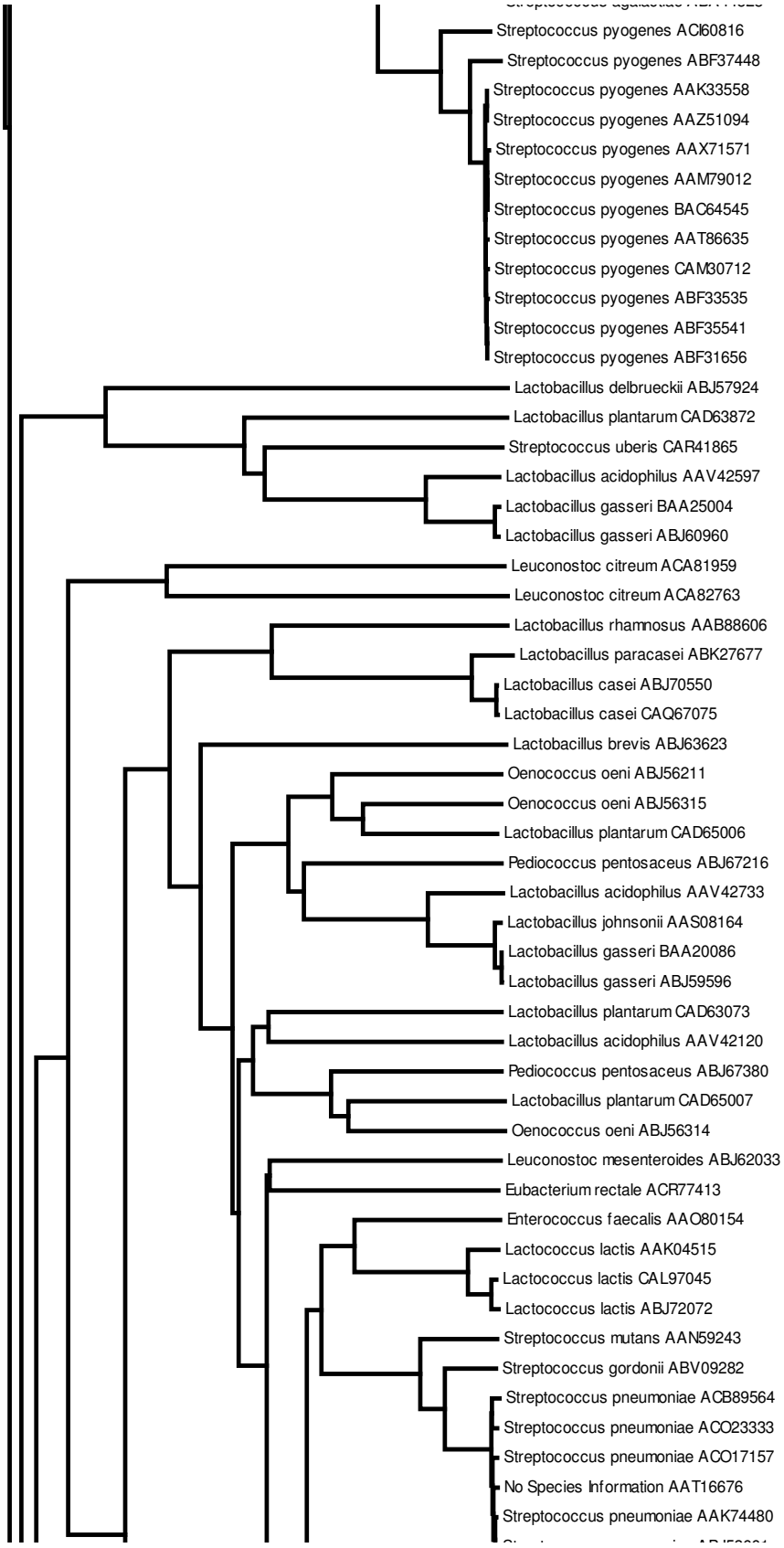

Listeria monocytogenes ABI27483
Listeria monocytogenes ABI27510
Listeria monocytogenes ABI27501
Listeria monocytogenes ABI28552
Listeria monocytogenes ABI28687
Listeria monocytogenes ABI28576
Listeria monocytogenes ABI28678
Listeria monocytogenes ABI28684
Listeria monocytogenes ABI28615
Listeria monocytogenes ABI28735
Listeria monocytogenes ABI27390
Listeria monocytogenes ABI27507
Listeria monocytogenes ABI28669
Listeria monocytogenes ABI28597
Listeria monocytogenes ABI28708
Listeria monocytogenes ABI28591
Listeria monocytogenes ABI27471
Listeria monocytogenes ABI27462
Listeria monocytogenes ABI27414
Listeria monocytogenes ABI27372
Listeria monocytogenes ABI27399
Listeria monocytogenes ABI27429
Listeria monocytogenes ABI27411
Listeria monocytogenes ABI28588
Listeria monocytogenes ABI27393
Listeria monocytogenes ABI28648
Listeria monocytogenes ABI28702
Listeria monocytogenes ABI28639
Listeria monocytogenes ABI28672
Listeria monocytogenes ABI28651
Listeria monocytogenes ABI28705
Listeria monocytogenes ABI28636
Listeria monocytogenes ABI28729
Listeria monocytogenes ABI27534
Listeria monocytogenes ABI27519
Listeria monocytogenes ABI28732
Listeria monocytogenes ABI28714
Listeria monocytogenes ABI27447
Listeria monocytogenes ABI27531
Listeria monocytogenes ABI27537
Listeria monocytogenes ABI27489
Listeria monocytogenes ABI27516
Listeria monocytogenes ABI27495
Listeria monocytogenes ABI27522
Listeria monocytogenes ABI27528
Listeria monocytogenes ABI27486
Listeria monocytogenes ABI27540
Listeria monocytogenes ABI27546
Listeria monocytogenes ABI27381
Listeria monocytogenes ABI27384
Listeria monocytogenes ABI28600
Listeria monocytogenes ABI28654
Listeria monocytogenes ABI28558

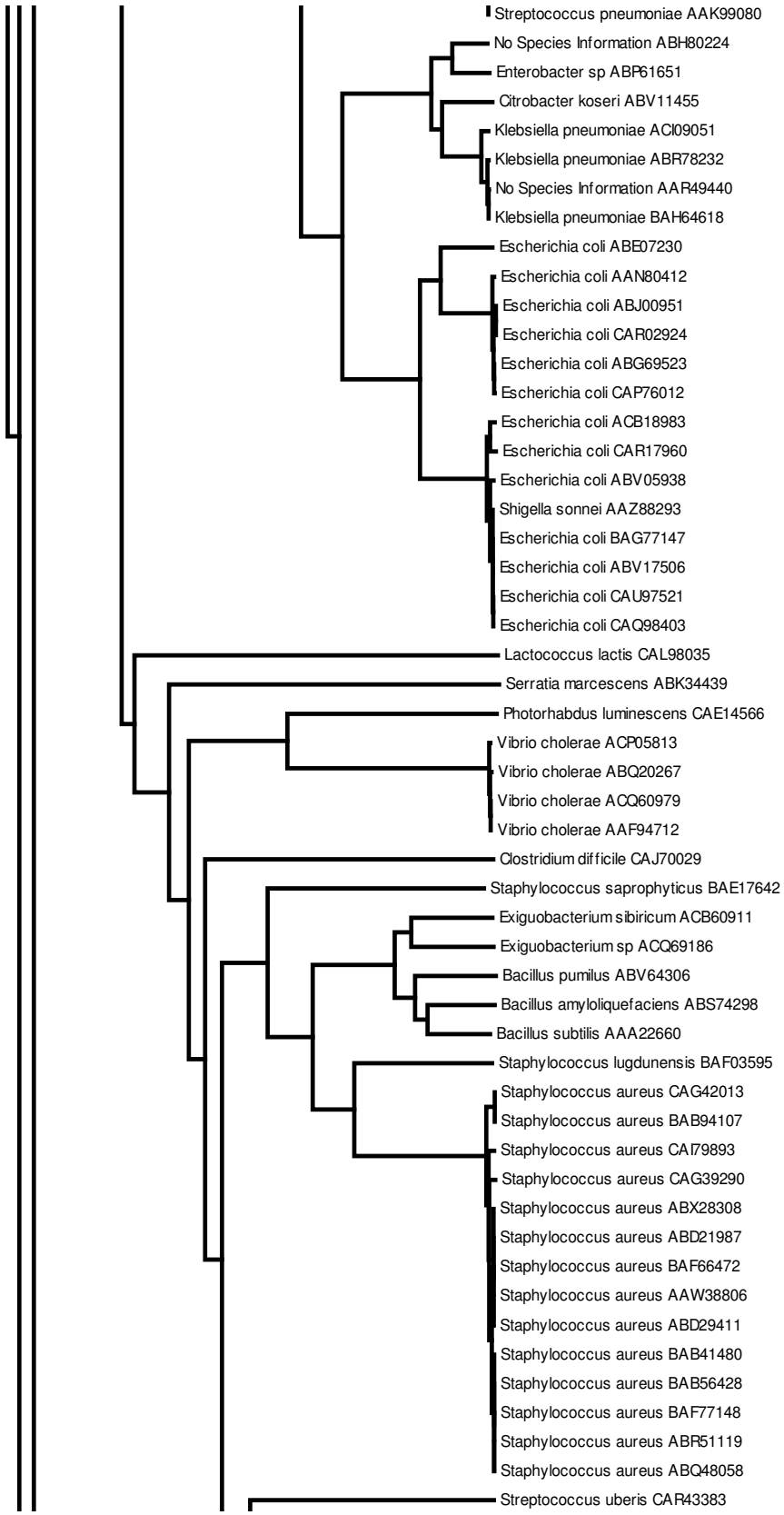

Listeria monocytogenes ABI28558
Listeria monocytogenes ABI28633
Listeria monocytogenes ABI28585
Listeria monocytogenes ABI28660
Listeria monocytogenes ABI28693
Listeria monocytogenes ABI28603
Listeria monocytogenes ABI28609
Listeria monocytogenes ABI28561
Listeria monocytogenes ABI28690
Listeria monocytogenes ABI28594
Listeria monocytogenes ABI28720
Listeria monocytogenes ABI28738
Listeria monocytogenes ABI28726
Listeria monocytogenes ABI27555
Listeria monocytogenes ABI27513
Listeria monocytogenes ABI27492
Listeria monocytogenes ABI27498
Listeria monocytogenes ABI27504
Listeria monocytogenes ABI27525
Listeria monocytogenes ABI27441
Listeria monocytogenes ABI27420
Listeria monocytogenes ABI27426
Listeria monocytogenes ABI27378
Listeria monocytogenes ABI27405
Listeria monocytogenes ABI27432
Listeria monocytogenes ABI27459
Listeria monocytogenes ABI27438
Listeria monocytogenes ABI27417
Listeria monocytogenes ABI27444
Listeria monocytogenes ABI27396
Listeria monocytogenes ABI27423
Listeria monocytogenes ABI27375
Listeria monocytogenes ABI27450
Listeria monocytogenes ABI27435
Listeria monocytogenes ABI27402
Listeria monocytogenes ABI27387
Listeria welshimeri CAK19685
Clostridium difficile CAJ67652
Enterococcus faecalis AAO81390
Mesoplasma florum AAT75974
Mesoplasma florum AAT75367
Mesoplasma florum AAT75670
Mycoplasma penetrans BAC44245
Mycoplasma penetrans BAC44246
Streptococcus uberis CAR41920
Serratia proteamaculans ABV41425
Pectobacterium atrosepticum CAG73771
Pectobacterium carotovorum AAX76615
Clostridium botulinum ACA44510
Clostridium botulinum ABS40281
Clostridium botulinum ACD23082
Clostridium botulinum ACD51635

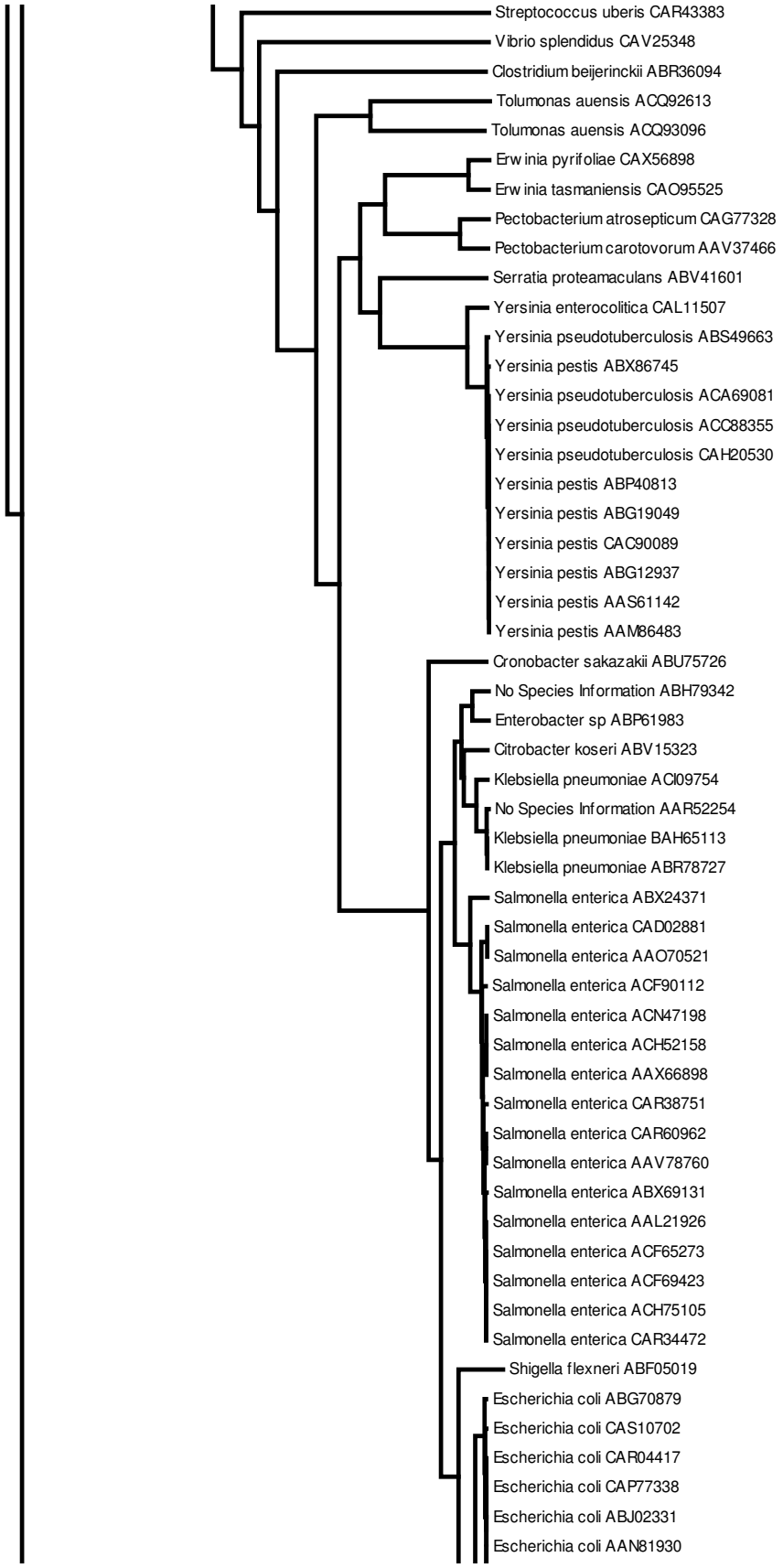

Clostridium botulinum ACD51635
Bacillus cereus ACK89199
Clostridium acetobutylicum AAK76757
Clostridium beijerinckii ABR36751
Tolumonas auensis ACQ94693
Pectobacterium atrosepticum CAG74361
Pectobacterium carotovorum AAX76617
Bacillus licheniformis AAU25633
Bacillus pumilus ABV60909
Lactobacillus casei ABJ69303
Lactobacillus casei CAQ65640
Listeria monocytogenes CAC98817
Lactococcus lactis CAL97551
Lactococcus lactis ABJ73134
Serratia proteamaculans ABV43315
Tolumonas auensis ACQ94573
Klebsiella pneumoniae ACI10265
No Species Information AAR49840
Klebsiella pneumoniae BAH63670
Klebsiella pneumoniae ABR77392
Bacillus cereus BAB88932
Aliivibrio salmonicida CAQ78390
Vibrio fischeri AAW85100
Vibrio fischeri ACH67055
Exiguobacterium sp ACQ71111
Clostridium beijerinckii ABR33655
Clostridium acetobutylicum AAK78365
Bacillus cereus AAP10685
Bacillus weihenstephanensis ABY44704
Fusobacterium mortiferum AAB49339
Lactobacillus acidophilus AAV43191
Lactobacillus johnsonii AAS08554
Clostridium beijerinckii ABR36940
Streptococcus sanguinis ABN44557
Aeromonas hydrophila ABK36623
Chromobacterium violaceum AAQ58947
Geobacillus kaustophilus BAD77499
Geobacillus thermodenitrificans ABO68480
Bacillus amyloliquefaciens CAE02623
Bacillus licheniformis AAU21991
Bacillus subtilis BAA06429
Clostridium acetobutylicum AAK79058
Clostridium perfringens ABG86616
Lactobacillus sakei CAI55838
Enterococcus faecalis AAO81040
No Species Information AAQ44364
Lactococcus lactis AAK04273
Lactococcus lactis CAL96796
Lactococcus lactis ABJ71802
Listeria welshimeri CAK19663
Listeria monocytogenes CAS04066
Listeria monocytogenes ACK40699

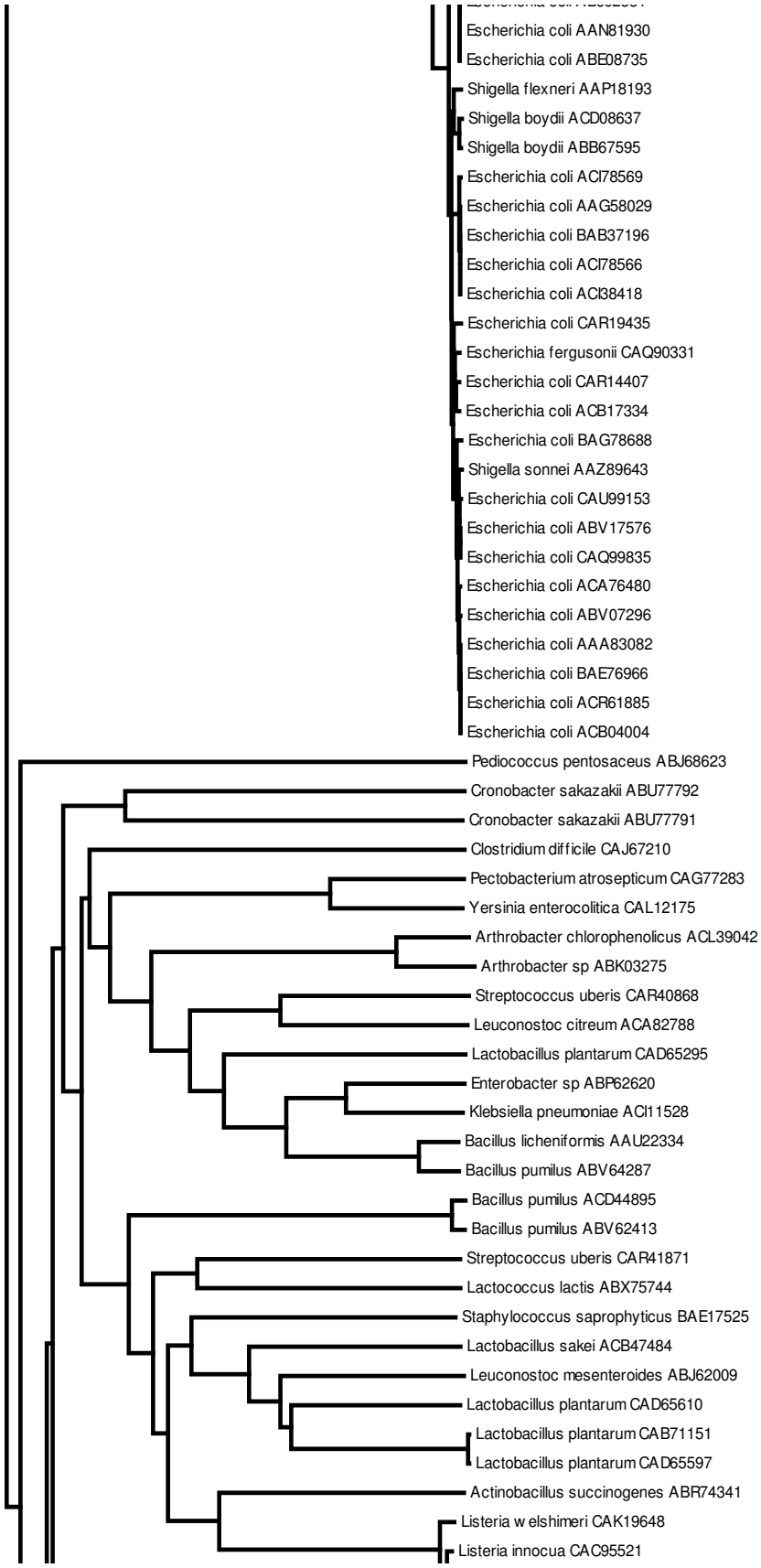

Listeria monocytogenes ACK40699
Listeria innocua CAC95530
Listeria monocytogenes CAD00798
Listeria monocytogenes AAT03078
Listeria welshimeri CAK22126
Listeria innocua CAC98129
Listeria monocytogenes CAD00974
Listeria monocytogenes CAS06490
Listeria monocytogenes ACK41097
Listeria monocytogenes AAT05516
No Species Information AAQ46338
Lactobacillus plantarum CAD63475
Lactobacillus acidophilus AAV43518
Lactobacillus gasseri ABJ61042
Lactobacillus johnsonii AAS09677
Leuconostoc mesenteroides ABJ62048
Symbiobacterium thermophilum BAD39939
Pediococcus pentosaceus ABJ67669
Clostridium difficile CAJ69971
Listeria innocua CAC96150
Listeria monocytogenes CAC98995
Listeria monocytogenes ACK40046
Listeria monocytogenes CAS04705
Listeria monocytogenes AAT03718
Mesoplasma florum AAT75789
Mesoplasma florum AAT75365
Mesoplasma florum AAT75368
Mesoplasma florum AAT75972
Mesoplasma florum AAT75390
Mycoplasma fermentans BAH69373
Clostridium difficile CAJ70007
Enterococcus faecium AAD28227
Clostridium difficile CAJ70023
Streptococcus uberis CAR40670
Streptococcus uberis CAR42015
Lactobacillus acidophilus AAV42740
Lactobacillus gasseri ABJ59601
Lactobacillus johnsonii AAS08170
Clostridium beijerinckii ABR36634
Geobacillus kaustophilus BAD76622
Geobacillus thermodenitrificans ABO67615
Streptococcus thermophilus AAV63161
Streptococcus thermophilus AAV61234
Streptococcus thermophilus AAV62643
Streptococcus thermophilus AAV60718
Brassica nigra AAL06591
Listeria monocytogenes CAQ77444
Listeria welshimeri AAV28314
Brassica rapa AAK49406
Listeria seeligeri AAY88056
Corynebacterium kroppenstedtii ACR18502

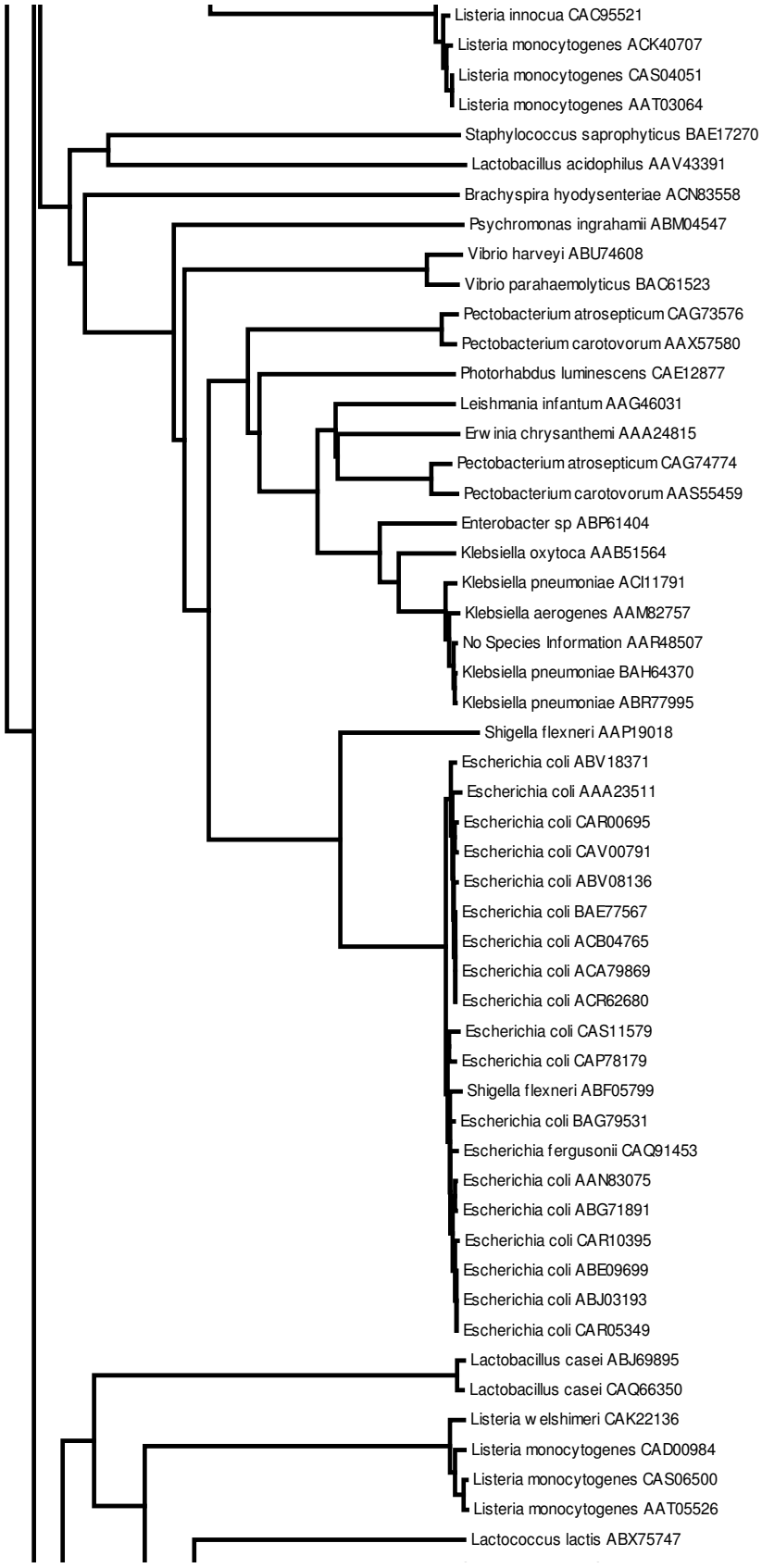

Corynebacterium kroppenstedtii ACR18502
Actinomyces naeslundii AAK33123
Corynebacterium glutamicum AAO15902
Clostridium botulinum ACA43342
Clostridium botulinum ABS41205
Clostridium botulinum ACA56374
Clostridium botulinum ACO85077
Clostridium botulinum ACQ54105
Clostridium botulinum ABS38677
Clostridium botulinum ABS32601
Clostridium botulinum CAL82750
Clostridium longisporum AAC05714
Clostridium acetobutylicum AAK78719
Clostridium beijerinckii ABR34980
Listeria welshimeri CAK19711
Listeria innocua CAC95577
Listeria monocytogenes CAS04105
Listeria monocytogenes AAT03123
Listeria monocytogenes CAD00846
Listeria monocytogenes ACK40651
Listeria monocytogenes CAC20633
Listeria monocytogenes AAO19529
Listeria monocytogenes AAO19530
Listeria monocytogenes AAO19526
Listeria seeligeri AAY88058
Listeria seeligeri AAZ81550
Listeria seeligeri AAY88060
Listeria seeligeri AAY88055
Listeria seeligeri AAY88057
Listeria seeligeri AAY88061
Listeria seeligeri AAY88059
Streptococcus mutans AAF89979
Streptococcus mutans AAN58686
Streptococcus gordonii ABV10576
Streptococcus sanguinis ABN43835
Streptococcus suis ABP90455
Streptococcus suis ABP92658
Streptococcus pneumoniae AAK74732
No Species Information AAT15542
Streptococcus pneumoniae ACF56062
Streptococcus pneumoniae ACB89796
Streptococcus pneumoniae ACA36794
Streptococcus pneumoniae ABJ54260
Streptococcus pneumoniae AAK99310
Streptococcus pneumoniae CAR68371
Streptococcus pneumoniae ACO22019
Streptococcus equi CAW94508
Streptococcus equi CAW98624
Streptococcus equi ACG62710
Streptococcus agalactiae CAD46455
Streptococcus agalactiae AAM99678
Streptococcus agalactiae ABA44523
Streptococcus pyogenes ACI60816

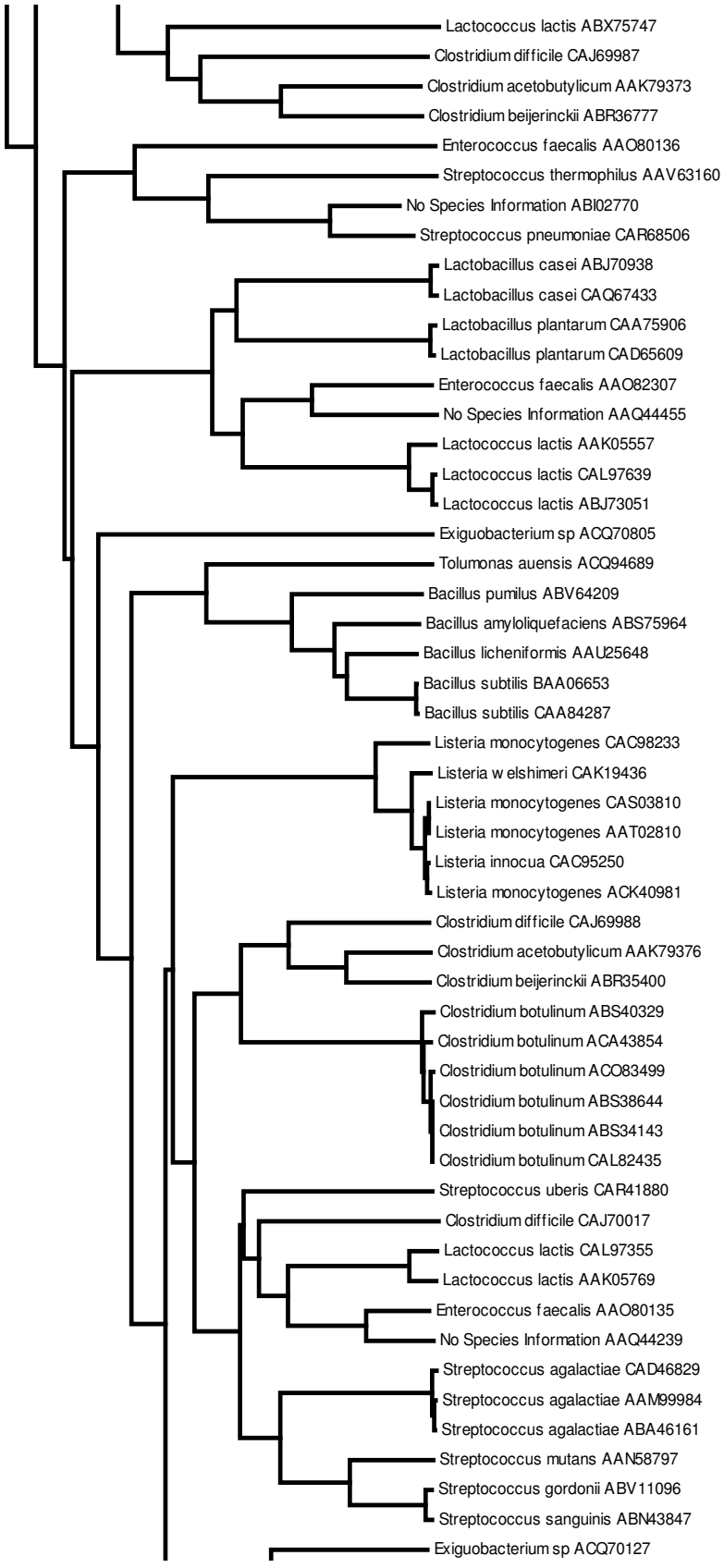

Streptococcus pyogenes ACI60816
Streptococcus pyogenes ABF37448
Streptococcus pyogenes AAK33558
Streptococcus pyogenes AAZ51094
Streptococcus pyogenes AAX71571
Streptococcus pyogenes AAM79012
Streptococcus pyogenes BAC64545
Streptococcus pyogenes AAT86635
Streptococcus pyogenes CAM30712
Streptococcus pyogenes ABF33535
Streptococcus pyogenes ABF35541
Streptococcus pyogenes ABF31656
Lactobacillus delbrueckii ABJ57924
Lactobacillus plantarum CAD63872
Streptococcus uberis CAR41865
Lactobacillus acidophilus AAV42597
Lactobacillus gasseri BAA25004
Lactobacillus gasseri ABJ60960
Leuconostoc citreum ACA81959
Leuconostoc citreum ACA82763
Lactobacillus rhamnosus AAB88606
Lactobacillus paracasei ABK27677
Lactobacillus casei ABJ70550
Lactobacillus casei CAQ67075
Lactobacillus brevis ABJ63623
Oenococcus oeni ABJ56211
Oenococcus oeni ABJ56315
Lactobacillus plantarum CAD65006
Pediococcus pentosaceus ABJ67216
Lactobacillus acidophilus AAV42733
Lactobacillus johnsonii AAS08164
Lactobacillus gasseri BAA20086
Lactobacillus gasseri ABJ59596
Lactobacillus plantarum CAD63073
Lactobacillus acidophilus AAV42120
Pediococcus pentosaceus ABJ67380
Lactobacillus plantarum CAD65007
Oenococcus oeni ABJ56314
Leuconostoc mesenteroides ABJ62033
Eubacterium rectale ACR77413
Enterococcus faecalis AAO80154
Lactococcus lactis AAK04515
Lactococcus lactis CAL97045
Lactococcus lactis ABJ72072
Streptococcus mutans AAN59243
Streptococcus gordonii ABV09282
Streptococcus pneumoniae ACB89564
Streptococcus pneumoniae ACO23333
Streptococcus pneumoniae ACO17157
No Species Information AAT16676
Streptococcus pneumoniae AAK74480


Streptococcus pneumoniae AAK99080
No Species Information ABH80224
Enterobacter sp ABP61651
Citrobacter koseri ABV11455
Klebsiella pneumoniae ACI09051
Klebsiella pneumoniae ABR78232
No Species Information AAR49440
Klebsiella pneumoniae BAH64618
Escherichia coli ABE07230
Escherichia coli AAN80412
Escherichia coli ABJ00951
Escherichia coli CAR02924
Escherichia coli ABG69523
Escherichia coli CAP76012
Escherichia coli ACB18983
Escherichia coli CAR17960
Escherichia coli ABV05938
Shigella sonnei AAZ88293
Escherichia coli BAG77147
Escherichia coli ABV17506
Escherichia coli CAU97521
Escherichia coli CAQ98403
Lactococcus lactis CAL98035
Serratia marcescens ABK34439
Photorhabdus luminescens CAE14566
Vibrio cholerae ACP05813
Vibrio cholerae ABQ20267
Vibrio cholerae ACQ60979
Vibrio cholerae AAF94712
Clostridium difficile CAJ70029
Staphylococcus saprophyticus BAE17642
Exiguobacterium sibiricum ACB60911
Exiguobacterium sp ACQ69186
Bacillus pumilus ABV64306
Bacillus amyloliquefaciens ABS74298
Bacillus subtilis AAA22660
Staphylococcus lugdunensis BAF03595
Staphylococcus aureus CAG42013
Staphylococcus aureus BAB94107
Staphylococcus aureus CAI79893
Staphylococcus aureus CAG39290
Staphylococcus aureus ABX28308
Staphylococcus aureus ABD21987
Staphylococcus aureus BAF66472
Staphylococcus aureus AAW38806
Staphylococcus aureus ABD29411
Staphylococcus aureus BAB41480
Staphylococcus aureus BAB56428
Staphylococcus aureus BAF77148
Staphylococcus aureus ABR51119
Staphylococcus aureus ABQ48058
Streptococcus uberis CAR43383


Streptococcus uberis CAR43383
Vibrio splendidus CAV25348
Clostridium beijerinckii ABR36094
Tolumonas auensis ACQ92613
Tolumonas auensis ACQ93096
Erw inia pyrifoliae CAX56898
Erw inia tasmaniensis CAO95525
Pectobacterium atrosepticum CAG77328
Pectobacterium carotovorum AAV37466
Serratia proteamaculans ABV41601
Yersinia enterocolitica CAL11507
Yersinia pseudotuberculosis ABS49663
Yersinia pestis ABX86745
Yersinia pseudotuberculosis ACA69081
Yersinia pseudotuberculosis ACC88355
Yersinia pseudotuberculosis CAH20530
Yersinia pestis ABP40813
Yersinia pestis ABG19049
Yersinia pestis CAC90089
Yersinia pestis ABG12937
Yersinia pestis AAS61142
Yersinia pestis AAM86483
Cronobacter sakazakii ABU75726
No Species Information ABH79342
Enterobacter sp ABP61983
Citrobacter koseri ABV15323
Klebsiella pneumoniae ACI09754
No Species Information AAR52254
Klebsiella pneumoniae BAH65113
Klebsiella pneumoniae ABR78727
Salmonella enterica ABX24371
Salmonella enterica CAD02881
Salmonella enterica AAO70521
Salmonella enterica ACF90112
Salmonella enterica ACN47198
Salmonella enterica ACH52158
Salmonella enterica AAX66898
Salmonella enterica CAR38751
Salmonella enterica CAR60962
Salmonella enterica AAV78760
Salmonella enterica ABX69131
Salmonella enterica AAL21926
Salmonella enterica ACF65273
Salmonella enterica ACF69423
Salmonella enterica ACH75105
Salmonella enterica CAR34472
Shigella flexneri ABF05019
Escherichia coli ABG70879
Escherichia coli CAS10702
Escherichia coli CAR04417
Escherichia coli CAP77338
Escherichia coli ABJ02331
Escherichia coli AAN81930


Escherichia coli AAN81930
Escherichia coli ABE08735
Shigella flexneri AAP18193
Shigella boydii ACD08637
Shigella boydii ABB67595
Escherichia coli ACI78569
Escherichia coli AAG58029
Escherichia coli BAB37196
Escherichia coli ACI78566
Escherichia coli ACI38418
Escherichia coli CAR19435
Escherichia fergusonii CAQ90331
Escherichia coli CAR14407
Escherichia coli ACB17334
Escherichia coli BAG78688
Shigella sonnei AAZ89643
Escherichia coli CAU99153
Escherichia coli ABV17576
Escherichia coli CAQ99835
Escherichia coli ACA76480
Escherichia coli ABV07296
Escherichia coli AAA83082
Escherichia coli BAE76966
Escherichia coli ACR61885
Escherichia coli ACB04004
Pediococcus pentosaceus ABJ68623
Cronobacter sakazakii ABU77792
Cronobacter sakazakii ABU77791
Clostridium difficile CAJ67210
Pectobacterium atrosepticum CAG77283
Yersinia enterocolitica CAL12175
Arthrobacter chlorophenolicus ACL39042
Arthrobacter sp ABK03275
Streptococcus uberis CAR40868
Leuconostoc citreum ACA82788
Lactobacillus plantarum CAD65295
Enterobacter sp ABP62620
Klebsiella pneumoniae ACI11528
Bacillus licheniformis AAU22334
Bacillus pumilus ABV64287
Bacillus pumilus ACD44895
Bacillus pumilus ABV62413
Streptococcus uberis CAR41871
Lactococcus lactis ABX75744
Staphylococcus saprophyticus BAE17525
Lactobacillus sakei ACB47484
Leuconostoc mesenteroides ABJ62009
Lactobacillus plantarum CAD65610
Lactobacillus plantarum CAB71151
Lactobacillus plantarum CAD65597
Actinobacillus succinogenes ABR74341
Listeria w elshimeri CAK19648
Listeria innocua CAC95521


Listeria innocua CAC95521
Listeria monocytogenes ACK40707
Listeria monocytogenes CAS04051
Listeria monocytogenes AAT03064
Staphylococcus saprophyticus BAE17270
Lactobacillus acidophilus AAV43391
Brachyspira hyodysenteriae ACN83558
Psychromonas ingrahamii ABM04547
Vibrio harveyi ABU74608
Vibrio parahaemolyticus BAC61523
Pectobacterium atrosepticum CAG73576
Pectobacterium carotovorum AAX57580
Photorhabdus luminescens CAE12877
Leishmania infantum AAG46031
Erwinia chrysanthemi AAA24815
Pectobacterium atrosepticum CAG74774
Pectobacterium carotovorum AAS55459
Enterobacter sp ABP61404
Klebsiella oxytoca AAB51564
Klebsiella pneumoniae ACI11791
Klebsiella aerogenes AAM82757
No Species Information AAR48507
Klebsiella pneumoniae BAH64370
Klebsiella pneumoniae ABR77995
Shigella flexneri AAP19018
Escherichia coli ABV18371
Escherichia coli AAA23511
Escherichia coli CAR00695
Escherichia coli CAV00791
Escherichia coli ABV08136
Escherichia coli BAE77567
Escherichia coli ACB04765
Escherichia coli ACA79869
Escherichia coli ACR62680
Escherichia coli CAS11579
Escherichia coli CAP78179
Shigella flexneri ABF05799
Escherichia coli BAG79531
Escherichia fergusonii CAQ91453
Escherichia coli AAN83075
Escherichia coli ABG71891
Escherichia coli CAR10395
Escherichia coli ABE09699
Escherichia coli ABJ03193
Escherichia coli CAR05349
Lactobacillus casei ABJ69895
Lactobacillus casei CAQ66350
Listeria welshimeri CAK22136
Listeria monocytogenes CAD00984
Listeria monocytogenes CAS06500
Listeria monocytogenes AAT05526
Lactococcus lactis ABX75747


Lactococcus lactis ABX75747
Clostridium difficile CAJ69987
Clostridium acetobutylicum AAK79373
Clostridium beijerinckii ABR36777
Enterococcus faecalis AAO80136
Streptococcus thermophilus AAV63160
No Species Information ABI02770
Streptococcus pneumoniae CAR68506
Lactobacillus casei ABJ70938
Lactobacillus casei CAQ67433
Lactobacillus plantarum CAA75906
Lactobacillus plantarum CAD65609
Enterococcus faecalis AAO82307
No Species Information AAQ44455
Lactococcus lactis AAK05557
Lactococcus lactis CAL97639
Lactococcus lactis ABJ73051
Exiguobacterium sp ACQ70805
Tolumonas auensis ACQ94689
Bacillus pumilus ABV64209
Bacillus amyloliquefaciens ABS75964
Bacillus licheniformis AAU25648
Bacillus subtilis BAA06653
Bacillus subtilis CAA84287
Listeria monocytogenes CAC98233
Listeria w elshimeri CAK19436
Listeria monocytogenes CAS03810
Listeria monocytogenes AAT02810
Listeria innocua CAC95250
Listeria monocytogenes ACK40981
Clostridium difficile CAJ69988
Clostridium acetobutylicum AAK79376
Clostridium beijerinckii ABR35400
Clostridium botulinum ABS40329
Clostridium botulinum ACA43854
Clostridium botulinum ACO83499
Clostridium botulinum ABS38644
Clostridium botulinum ABS34143
Clostridium botulinum CAL82435
Streptococcus uberis CAR41880
Clostridium difficile CAJ70017
Lactococcus lactis CAL97355
Lactococcus lactis AAK05769
Enterococcus faecalis AAO80135
No Species Information AAQ44239
Streptococcus agalactiae CAD46829
Streptococcus agalactiae AAM99984
Streptococcus agalactiae ABA46161
Streptococcus mutans AAN58797
Streptococcus gordonii ABV11096
Streptococcus sanguinis ABN43847
Exiguobacterium sp ACQ70127

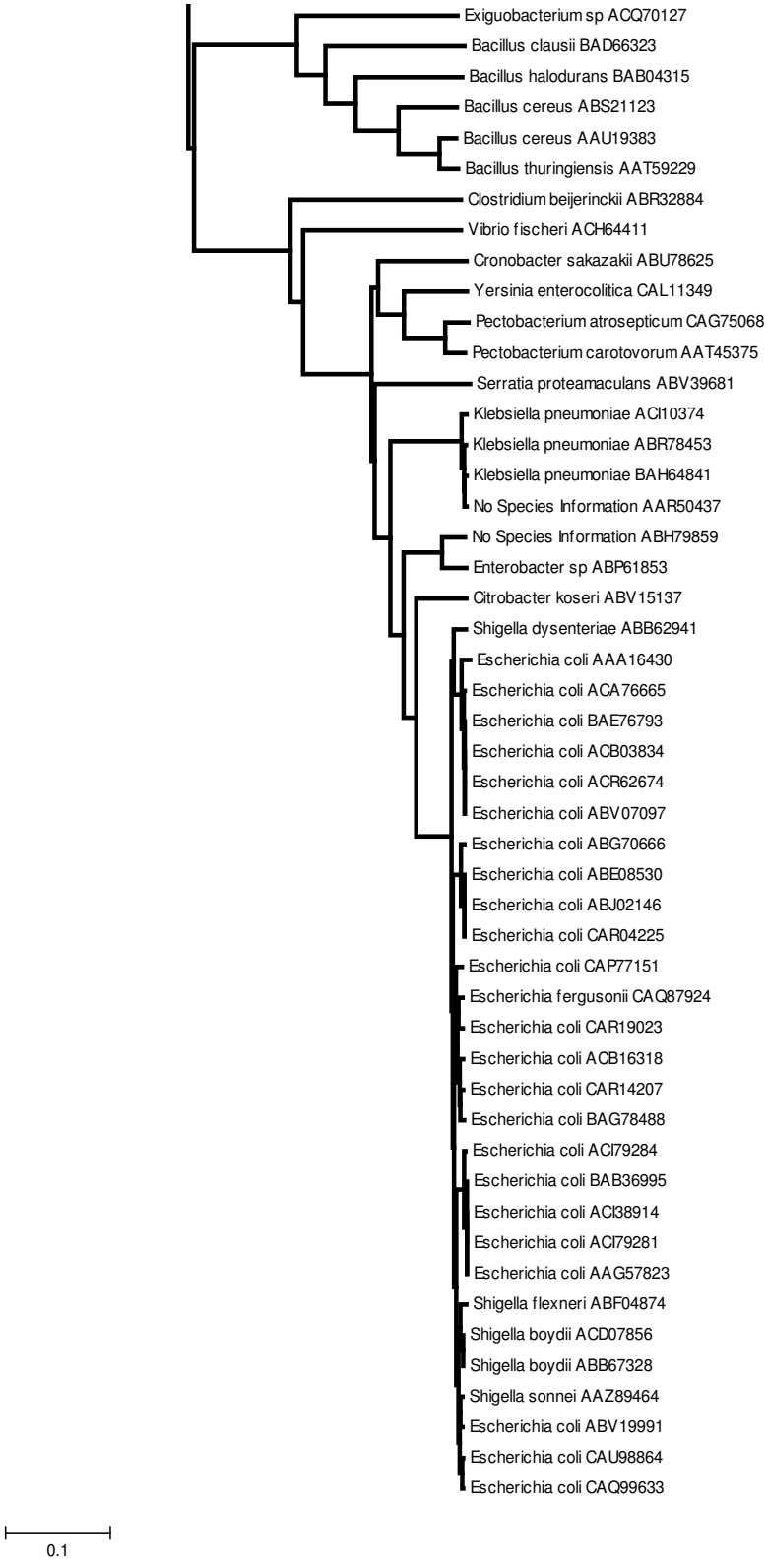



**Figure 127.** Phylogenetic analysis of 1654 archaic, bacterial and eukaryotic GH1 proteins ( = cleaving proteins; = slow-cleaving proteins; = non-cleaving proteins)

## 22. Cleaved-tag affinity purification

### 22.1. Cleaved-tag affinity purification

**T0-*Ss*βG** - T0-*Ss*βG (1 mL, 11.8 mg/mL, 50 mM sodium phosphate, pH 6.5) was added to a BRAND UV cuvette and was irradiated with UV light at room temperature for 4 hours. The reaction mixture was purified using an AKTA FPLC system and HisTrap FF 1 mL column. Protein was injected and eluted using a gradient from 100% binding buffer (20 mM Tris-HCl, 5 mM imidazole, 0.5 M NaCl, pH 7.8) to 100% elution buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, pH 7.8) (Figure 83). All eluent collected from the purification was analysed by SDS-PAGE (Figure 83).

**GFP-*Ss*βG** - GFP-*Ss*βG (2.5 mL, 1.9 mg/mL, 4.75 mg) was added to a quartz flask and was irradiated with UV light for 2 hours. Ni-NTA agarose resin (Qiagen) (15 mL) and 50 mM sodium phosphate buffer (pH 6.5) (7.5 mL) were added to the irradiated protein solution. The slurry was stirred for 4 hours at 4$^{o}$C, then 50 mM sodium phosphate buffer (pH 6.5) (10 mL) and GnHCl (final conc. 7 M) were added. The slurry was stirred overnight at 4$^{o}$C and was then purified manually; The slurry was loaded into an empty stoppered (cotton wool) syringe. Protein was eluted and the column was washed using binding buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, 7 M GnHCl, pH 7.8) followed by elution buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, pH 7.8). All eluent collected from the purification was concentrated and analysed by SDS-PAGE (Figure 93). Samples for SDS-PAGE were prepared using the Pierce Sample Prep Kit to remove GnHCl prior to analysis.

## 22.2. Protein denaturation

**Heating -** WT *Ss*βG (1 mL, 1.8 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a BRAND UV cuvette and was irradiated with UV light at room temperature for 2 hours. The reaction mixture was transferred to a microcentrifuge tube and was heated at 100$^{o}$C for 30 minutes.

**SDS and DTT -** WT *Ss*βG (1 mL, 1.8 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a BRAND UV cuvette and was irradiated with UV light at room temperature for 2 hours. The reaction mixture was transferred to a microcentrifuge tube and 2% w/v SDS (20 mg) and 200 mM DTT (30 mg) were added. The reaction mixture was heated at 100$^{o}$C for 15 minutes and was then purified using an AKTA FPLC system and HisTrap FF 1 mL column. Protein was injected and eluted using a gradient from 100% binding buffer (20 mM Tris-HCl, 5 mM imidazole, 0.5 M NaCl, pH 7.8) to 100% elution buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, pH 7.8). All eluent collected from the purification was analysed by SDS-PAGE.

**8 M urea -** WT *Ss*βG (1 mL, 1.8 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a BRAND UV cuvette and was irradiated with UV light at room temperature for 2 hours. Urea (480 mg) was added to a final concentration of 8 M and the reaction mixture was purified using an AKTA FPLC system and HisTrap FF 1 mL column. Protein was injected and eluted using a gradient from 100% binding buffer (20 mM Tris-HCl, 5 mM imidazole, 0.5 M NaCl, 8 M urea, pH 7.8) to 100% elution buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, 8 M urea, pH 7.8). All eluent collected from the purification was analysed by SDS-PAGE.

**Hydrophobic interaction chromatography -** WT *Ss*βG (1 mL, 1.8 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a BRAND UV cuvette and was

irradiated with UV light at room temperature for 2 hours. The reaction mixture was transferred to a microcentrifuge tube and ammonium sulfate to a final concentration of 2 M was added. The reaction mixture was purified using an AKTA FPLC system and HiTrap Phenyl HP 1 mL column. Protein was injected and eluted using a gradient from 100% buffer A (50 mM TEA, 2 M ammonium sulphate, pH 7.5) to 100% buffer B (50 mM TEA, pH 7.5). All eluent collected from the purification was analysed by SDS-PAGE.

**7 M Guanidine hydrochloride (GnHCl) -** WT *Ss*βG (1 mL, 1.8 mg/mL, 50 mM sodium phosphate buffer, pH 6.5) was added to a BRAND UV cuvette and was irradiated with UV light at room temperature for 2 hours. Ni-NTA agarose resin (1 mL) was added and the reaction mixture was incubated for 3 hours at 4$^{o}$C, before GnHCl (670 mg) to a final concentration of 7 M was added. The reaction mixture was incubated overnight at 4$^{o}$C and was purified manually; The slurry was loaded into an empty stoppered (cotton wool) syringe. Protein was eluted and the column was washed using binding buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, 7 M GnHCl, pH 7.8) followed by elution buffer (20 mM Tris-HCl, 300 mM imidazole, 0.5 M NaCl, pH 7.8). All eluent collected from the purification was analysed by SDS-PAGE. SDS-PAGE samples were prepared using the Pierce SDS-PAGE Sample Prep Kit to remove GnHCl prior to analysis.

### 22.3. GFP-*Ss*βG$_{tag}$ refolding

**Direct dilution/dialysis** - GFP-*Ss*βG$_{tag}$ was refolded by rapid dilution into an iFOLD Protein Refolding System 2 buffer matrix; An iFOLD Protein Refolding Plate 2 (Novagen) was thawed at 40$^{o}$C for 15 minutes. GFP-*Ss*βG$_{tag}$ (550 μL, 1 mg/mL) was diluted with $H_2O$ (450 μL) to give a 0.5 mg/mL solution. GFP-*Ss*βG$_{tag}$ (10 μL, 0.5

mg/mL) was added to each well and the refolding screen was incubated overnight at room temperature with gentle shaking. Each refolding reaction (100 µL) was diluted 3x with 50 mM Tris-HCl (pH 8.0) and was analysed for protein precipitation ($A_{340}$) and fluorescence (excitation 390 nm, emission 510 nm) using a spectrophotomer (SpectraMax Plus, Molecular Devices) and fluorimeter (POLARstar OPTIMA), respectively (Figure 94). A larger scale refolding of GFP-*Ss*β$G_{tag}$ was performed by dialysis: GFP-*Ss*β$G_{tag}$ (90 µL, 1 mg/mL) was added to a Slide-A-Lyzer MINI (7 KDa MWCO) dialysis unit (Pierce) and was dialysed into refolding buffer (50 mM CHES, 1 M NDSB 256 [dimethylbenzylammonium-propanesulfonate], 1 mM EDTA, 9 mM glutathione [GSH], 1 mM glutathione disulfide [GSSG], 240 mM NaCl and 10 mM KCl [pH 9.0]) (10 mL) overnight, the buffer was changed (10 mL) and the protein sample dialysed for a further 4 hours. Denatured GFP-*Ss*β$G_{tag}$, refolded GFP-*Ss*β$G_{tag}$ and the dialysis buffer were analysed for fluorescence (excitation 390 nm, emission 510 nm) and protein precipitation ($A_{340}$) (Figure 95).

**Hydrophobic interaction chromatography -** Refolding of GFP-*Ss*β$G_{tag}$ was carried out using an AKTA FPLC system and HiTrap Phenyl HP 1 mL column (GE Healthcare). The column was washed with $H_2O$ (2 CV). Ammonium sulphate was added to the protein to a final concentration of 2 M, and was injected and eluted using a gradient from 100% buffer A (50 mM TEA, 2 M ammonium sulphate, pH 7.5) to 100% buffer B (50 mM TEA, pH 7.5) (Figure 101). All eluent collected from the purification was analysed by SDS-PAGE, and for refolding by fluorescence (excitation 390 nm, emission 510 nm). Samples for SDS-PAGE were prepared using the Pierce Sample Prep Kit to remove GnHCl prior to analysis.

### 22.4. GnHCl concentration

GFP-*Ss*βG (1.8 mg/mL, 1 mL, 1.8 mg) was added to a BRAND UV cuvette and was irradiated with UV light for 2 hours. Then Ni-NTA agarose resin (Qiagen) (10 mL) and 50 mM sodium phosphate buffer (pH 6.5) (4 mL) were added. The slurry was incubated with shaking for 4 hours at 4$^{o}$C. The supernatant (unbound protein) was removed and to the remaining resin (0.5 g x 6) was added 0, 0.5, 1, 3, 5 or 7 M GnHCl (500 uL). The slurries were incubated with shaking overnight at 4$^{o}$C. The beads (45-165 μm) were imaged using apparatus similar to that previously reported [http://pubs.acs.org/doi/abs/10.1021/ja808128s]. Briefly, samples were viewed with a 60x objective (Plan Apo TIRF; Nikon Instruments) excited with an argon ion laser (Spectra Physik, 488nm, 4 mW). Emitted fluorescence was collected through the same objective, transmitted through dichroic (Q495LP) and band-pass (HQ525/50) filters (Chroma Technology Corp, Rockingham, VT), and imaged using a 128 x 128 pixel CCD detector (iXon DU-860, Andor Technology PLC, Belfast, UK) with 40 ms exposure. The reaction mixtures were additionally viewed using UV light. The reaction mixtures were added to spin cups (cellulose acetate filter, Pierce) and centrifuged (1 minute), and the supernatant was collected. Elution buffer (1 mL) was added to the resin. The samples were vortexed to mix and were centrifuged (1 minute). The supernatant was collected. All eluent was analysed by SDS-PAGE. Samples for SDS-PAGE were prepared using the Pierce Sample Prep Kit to remove GnHCl before being analysed.

**Chapter 7: References**


1. Schmidt, R. R., *Angew. Chem.* 1986, 98, 213-36.
2. Laine, R. A.; Pamidimukkala, K. M.; French, A. D.; Hall, R. W.; Abbas, S. A.; Jain, R. K.; Matta, K. L., *J. Am. Chem. Soc.* 1988, 110, 6931-9.
3. Apweiler, R.; Hermjakob, H.; Sharon, N., *Biochim. Biophys. Acta* 1999, 1473, 4-8.
4. Parodi, A. J., *Annu. Rev. Biochem.* 2000, 69, 69-93.
5. Helenius, A., *Mol. Biol. Cell* 1994, 5, 253-65.
6. Opdenakker, G.; Rudd, P. M.; Ponting, C. P.; Dwek, R. A., *FASEB J.* 1993, 7, 1330-7.
7. Imperiali, B.; O'Connor, S. E., *Curr. Opin. Chem. Biol.* 1999, 3, 643-49.
8. Shental-Bechor, D.; Levy, Y., *Proc. Natl. Acad. Sci. U.S.A.* 2008, 105, 8256-61.
9. Sriram, V.; Willard, C. A.; Liu, J.; Brutkiewicz, R. R., *Immunology* 2008, 123, 272-81.
10. Sola, R. J.; Griebenow, K., *J. Pharm. Sci.* 2009, 98, 1223-45.
11. Skropeta, D., *Bioorg. Med. Chem.* 2009, 17, 2645-53.
12. Dwek, R. A., *Chem. Rev.* 1996, 96, 683-720.
13. Doores, K. J.; Gamblin, D. P.; Davis, B. G., *Chem. Eur. J.* 2006, 12, 656-65.
14. Sears, P.; Wong, C. H., *Cell. Mol. Life Sci.* 1998, 54, 223-52.
15. Lowe, J. B.; Marth, J. D., *Annu. Rev. Biochem.* 2003, 72, 643-91.
16. Varki, A., *Glycobiology* 1993, 3, 97-130.
17. Davis, B. G., *Chem. Rev.* 2002, 102, 579-601.
18. Gamblin, D. P.; Scanlan, E. M.; Davis, B. G., *Chem. Rev.* 2009, 109, 131-63.
19. Helenius, A.; Aebi, M., *Science* 2001, 291, 2364-69.

20. Kannagi, R., *Curr. Opin. Struct. Biol.* 2002, 12, 599-608.

21. Haltiwanger, R. S., *Curr. Opin. Struct. Biol.* 2002, 12, 593-98.

22. Nishiwaki, K.; Kubota, Y.; Chigira, Y.; Roy, S. K.; Suzuki, M.; Schvarzstein, M.; Jigami, Y.; Hisamoto, N.; Matsumoto, K., *Nat. Cell Biol.* 2004, 6, 31-37.

23. Lowe, J. B., *Cell* 2001, 104, 809-12.

24. Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A., *Nature* 1998, 393, 648-59.

25. *National Institute of Allergy and Infectious Diseases (NIAID).*

26. Rademacher, T. W.; Parekh, R. B.; Dwek, R. A., *Annu. Rev. Biochem.* 1988, 57, 785-838.

27. Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A., *Biochemistry* 1994, 33, 17-22.

28. Prien, J. M.; Ashline, D. J.; Lapadula, A. J.; Zhang, H.; Reinhold, V. N., *J. Am. Soc. Mass Spectrom.* 2009, 20, 539-56.

29. Edwards, E. H.; Sprague, E. A.; Kelley, J. L.; Kerbacher, J. J.; Schwartz, C. J.; Elbein, A. D., *Biochemistry* 1989, 28, 7679-87.

30. Bertozzi, C. R.; Freeze, H. H.; Varki, A.; Esko, J. D., *Essentials of Glycobiology*, 2nd ed.; 2009; pp 719-32.

31. Rudd, P. M.; Dwek, R. A., *Crit. Rev. Biochem. Mol. Biol.* 1997, 32, 1-100.

32. Rich, J. R.; Withers, S. G., *Nat. Chem. Biol.* 2009, 5, 206-15.

33. Davis, B. G.; Lloyd, R. C.; Jones, J. B., *J. Org. Chem.* 1998, 63, 9614-15.

34. Davis, B. G.; Maughan, M. A. T.; Green, M. P.; Ullman, A.; Jones, J. B., *Tetrahedron Asymmetry* 2000, 11, 245-62.

35. Lloyd, R. C.; Davis, B. G.; Jones, J. B., *Bioorg. Med. Chem.* 2000, 8, 1537-44.

36. Liu, H.; Wang, L.; Brock, A.; Wong, C.-H.; Schultz, P. G., *J. Am. Chem. Soc.* 2003, 125, 1702-03.

37. Macmillan, D.; Bill, R. M.; Sage, K. A.; Fern, D.; Flitsch, S. L., *Chem. Biol.* 2001, 8, 133-45.

38. Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R., *J. Am. Chem. Soc.* 1999, 121, 11684-89.

39. Witte, K.; Sears, P.; Wong, C.-H., *J. Am. Chem. Soc.* 1997, 119, 2114-18.

40. Piontek, C.; Silva, D. V.; Heinlein, C.; Pohner, C.; Mezzato, S.; Ring, P.; Martin, A.; Schmid, F. X.; Unverzagt, C., *Angew. Chem., Int. Ed.* 2009, 48, 1941-45.

41. Piontek, C.; Ring, P.; Harjes, O.; Heinlein, C.; Mezzato, S.; Lombana, N.; Pohner, C.; Puttner, M.; Silva, D. V.; Martin, A.; Schmid, F. X.; Unverzagt, C., *Angew. Chem., Int. Ed.* 2009, 48, 1936-40.

42. Wolfgang, J., *Eur. J. Haematol.* 2007, 78, 183-205.

43. Taylor, C. M., *Tetrahedron* 1998, 54, 11317-62.

44. Krieg, J.; Glasner, W.; Vicentini, A.; Doucey, M.-A.; Loffler, A.; Hess, D.; Hofsteenge, J., *J. Biol. Chem.* 1997, 272, 26687-92.

45. Mehta, D. P.; Ichikawa, M.; Salimath, P. V.; Etchison, J. R.; Haak, R.; Manzi, A.; Freeze, H. H., *J. Biol. Chem.* 1996, 271, 10897-903.

46. Haynes, P. A., *Glycobiology* 1998, 8, 1-5.

47. Spiro, R. G., *Adv. Protein Chem.* 1973, 27, 349-467.

48. Smythe, C.; Cohen, P., *Eur. J. Biochem.* 1991, 200, 625-31.

49. Strahm, A.; Amado, R.; Neukom, H., *Phytochemistry* 1981, 20, 1061-3.

50. Colley, K. J.; Baenziger, J. U., *J. Biol. Chem.* 1987, 262, 10290-5.

51. Spiro, R. G., *Glycobiology* 2002, 12, 43R-56R.

52. Horton, D.; Wander, J. D., *Carbohydrates: Chemistry and Biochemistry*, 2nd ed.; 1990; Vol. 4B p799.

53. Lote, C. J.; Weiss, J. B., *FEBS Lett.* 1971, 16, 81-5.

54. Lote, C. J.; Weiss, J. B., *Biochem. J.* 1971, 123, 25p.

55. Weiss, J. B.; Lote, C. J.; Bobinski, H., *Nature* 1971, 234, 25-6.

56. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H., *Science* 1994, 266, 776-9.

57. Muir, T. W.; Sondhi, D.; Cole, P. A., *Proc. Natl. Acad. Sci. U.S.A.* 1998, 95, 6705-10.

58. Barrai, I.; Volinia, S.; Scapoli, C., *Int. J. Pept. Protein Res.* 1995, 45, 326-31.

59. Davis, B. G., *Pure Appl. Chem.* 2009, 81, 285-98.

60. Lee, Y. C.; Stowell, C. P.; Krantz, M. J., *Biochemistry* 1976, 15, 3956-63.

61. Gray, G. R., *Arch. Biochem. Biophys.* 1974, 163, 426-28.

62. Gildersleeve, J. C.; Oyelaran, O.; Simpson, J. T.; Allred, B., *Bioconjug. Chem.* 2008, 18, 1485-90.

63. Pearce, O. M. T.; Fisher, K. D.; Humphries, J.; Seymour, L. W.; Smith, A.; Davis, B. G., *Angew. Chem., Int. Ed.* 2005, 44, 1057-61.

64. Robinson, M. A.; Charlton, S. T.; Garnier, P.; Wang, X.-t.; Davis, S. S.; Perkins, A. C.; Frier, M.; Duncan, R.; Savage, T. J.; Wyatt, D. A.; Watson, S. A.; Davis, B. G., *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 14527-32.

65. Danishefsky, S. J.; Allen, J. R., *Angew. Chem., Int. Ed.* 2000, 39, 836-63.

66. Pan, Y.; Chefalo, P.; Nagy, N.; Harding, C.; Guo, Z., *J. Med. Chem.* 2005, 48, 875-83.

67. Chefalo, P.; Pan, Y.; Nagy, N.; Harding, C.; Guo, Z., *Glycoconj. J.* 2004, 20, 407-14.

68. Bernstein, M. A.; Hall, L. D., *Carbohydr. Res.* 1980, 78, C1.

69. Ragupathi, G.; Koganty, R. R.; Qiu, D.; Lloyd, K. O.; Livingston, P. O., *Glycoconj. J.* 1998, 15, 217-21.

70. Hou, S.-j.; Saksena, R.; Kovác, P., *Carbohydr. Res.* 2008, 343, 196-210.

71. Chernyak, A.; Karavanov*, A.; Ogawa, Y.; Kovác*, P., *Carbohydr. Res.* 2001, 330, 479-86.

72. Johansson, S. M. C.; Arnberg, N.; Elofsson, M.; Wadell, G.; Kihlberg, J., *ChemBioChem* 2005, 6, 358-64.

73. Tietze, L. F.; Schröter, C.; Gabius, S.; Brinck, U.; Goerlach-Graw, A.; Gabius, H. J., *Bioconjug. Chem.* 1991, 2, 148-53.

74. Kamath, V. P.; Diedrich, P.; Hindsgaul, O., *Glycoconj. J.* 1996, 13, 315-19.

75. Meeks, M. D.; Saksena, R.; Ma, X.; Wade, T. K.; Taylor, R. K.; Kovac, P.; Wade, W. F., *Infect. Immun.* 2004, 72, 4090-101.

76. Saksena, R.; Ma, X.; Kovac, P., *Carbohydr. Res.* 2003, 338, 2591-603.

77. Saksena, R.; Zhang, J.; Kovac, P., *Tetrahedron Asymmetry* 2005, 16, 187-97.

78. Saksena, R.; Ma, X.; Wade, T. K.; Kovac, P.; Wade, W. F., *FEMS Immunol. Med. Microbiol.* 2006, 47, 116-28.

79. Bundle, D. R.; Rich, J. R.; Jacques, S.; Yu, H. N.; Nitz, M.; Ling, C.-C., *Angew. Chem., Int. Ed.* 2005, 44, 7725-29.

80. Pozsgay, V.; Vieira, N. E.; Yergey, A., *Org. Lett.* 2002, 4, 3191-94.

81. Berkin, A.; Coxon, B.; Pozsgay, V., *Chem. Eur. J.* 2002, 8, 4424-33.

82. Kubler-Kielb, J.; Pozsgay, V., *J. Org. Chem.* 2005, 70, 6987-90.

83. Sen Gupta, S.; Kuzelka, J.; Singh, P.; Lewis, W. G.; Manchester, M.; Finn, M. G., *Bioconjug. Chem.* 2005, 16, 1572-79.

84. Schofield, L.; Hewitt, M. C.; Evans, K.; Siomos, M.-A.; Seeberger, P. H., *Nature* 2002, 418, 785-89.

85. Ghosh, S. S.; Kao, P. M.; McCue, A. W.; Chappelle, H. L., *Bioconjug. Chem.* 1990, 1, 71-76.

86. Unverzagt, C.; Andre, S.; Seifert, J.; Kojima, S.; Fink, C.; Srikrishna, G.; Freeze, H.; Kayser, K.; Gabius, H.-J., *J. Med. Chem.* 2002, 45, 478-91.

87. Macmillan, D.; Daines, A. M.; Bayrhuber, M.; Flitsch, S. L., *Org. Lett.* 2002, 4, 1467-70.

88. Yamamoto, N.; Sakakibara, T.; Kajihara, Y., *Tetrahedron Lett.* 2004, 45, 3287-90.

89. Ito, Y.; Hagihara, S.; Matsuo, I.; Totani, K., *Curr. Opin. Struct. Biol.* 2005, 15, 481-89.

90. Shin, I.; Jung, H. j.; Lee, M. r., *Tetrahedron Lett.* 2001, 42, 1325-28.

91. Ni, J.; Singh, S.; Wang, L.-X., *Bioconjug. Chem.* 2003, 14, 232-38.

92. Davis, N. J.; Flitsch, S. L., *Tetrahedron Lett.* 1991, 32, 6793-6.

93. Wong, S. Y. C.; Guile, G. R.; Dwek, R. A.; Arsequell, G., *Biochem. J.* 1994, 300, 843-50.

94. Swanwick, R. S.; Daines, A. M.; Tey, L.-H.; Flitsch, S. L.; Allemann, R. K., *ChemBioChem* 2005, 6, 1338-40.

95. Marcaurelle, L. A.; Bertozzi, C. R., *J. Am. Chem. Soc.* 2001, 123, 1587-95.

96. Macindoe, W. M.; van Oijen, A. H.; Boons, G.-J., *Chem. Commun.* 1998, 847-48.

97. Watt, G. M.; Boons, G.-J., *Carbohydr. Res.* 2004, 339, 181-93.

98. Rodriguez, E. C.; Winans, K. A.; King, D. S.; Bertozzi, C. R., *J. Am. Chem. Soc.* 1997, 119, 9905-06.

99. Shimaoka, H.; Kuramoto, H.; Furukawa, J.-i.; Miura, Y.; Kurogochi, M.; Kita, Y.; Hinou, H.; Shinohara, Y.; Nishimura, S.-I., *Chem. Eur. J.* 2007, 13, 1664-73.

100. Orr, G. A.; Rando, R. R., *Nature* 1978, 272, 722.

101. Grigalevicius, S.; Chierici, S.; Renaudet, O.; Lo-Man, R.; Deriaud, E.; Leclerc, C.; Dumy, P., *Bioconjug. Chem.* 2005, 16, 1149-59.

102. Totani, K.; Matsuo, I.; Ito, Y., *Bioorg. Med. Chem. Lett.* 2004, 14, 2285-89.

103. Totani, K.; Ito, Y., *Trends Glycosci. Glycotechnol.* 2005, 17, 121-30.

104. Davis, B. G., *Chem. Commun.* 2001, 351-52.

105. Gamblin, D. P.; Garnier, P.; Ward, S. J.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G., *Org. Biomol. Chem.* 2003, 1, 3642-44.

106. Prante, O.; Einsiedel, J.; Haubner, R.; Gmeiner, P.; Wester, H.-J.; Kuwert, T.; Maschauer, S., *Bioconjug. Chem.* 2007, 18, 254-62.

107. Zhong, L.; Arner, E. S. J.; Holmgren, A., *Proc. Natl. Acad. Sci. U.S.A.* 2000, 97, 5854-59.

108. Gamblin, D. P.; Garnier, P.; van Kasteren, S.; Oldham, N. J.; Fairbanks, A. J.; David, B. G., *Angew. Chem., Int. Ed.* 2004, 43, 828-33.

109. Bernardes, G. J. L.; Grayson, E. J.; Thompson, S.; Chalker, J. M.; Errey, J. C.; ElOualid, F.; Claridge, T. D. W.; Davis, B. G., *Angew. Chem., Int. Ed.* 2008, 47, 2244-47.

110. Watt, G. M.; Lund, J.; Levens, M.; Kolli, V. S. K.; Jefferis, R.; Boons, G.-J., *Chem. Biol.* 2003, 10, 807-14.

111. Mullegger, J.; Chen, H. M.; Warren, R. A. J.; Withers, S. G., *Angew. Chem., Int. Ed.* 2006, 45, 2585-88.

112. Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G., *Science* 1989, 244, 182-88.

113. Cornish, V. W.; Benson, D. R.; Altenbach, C. A.; Hideg, K.; Hubbell, W. L.; Schultz, P. G., *Proc. Natl. Acad. Sci. U.S.A.* 1994, 91, 2910-14.

114. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., *Science* 2001, 292, 498-500.

115. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G., *Science* 2003, 301, 964-67.

116. Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G., *J. Am. Chem. Soc.* 2003, 125, 11782-83.

117. Anderson, J. C.; Schultz, P. G., *Biochemistry* 2003, 42, 9598-608.

118. Tian, F.; Tsao, M.-L.; Schultz, P. G., *J. Am. Chem. Soc.* 2004, 126, 15962-63.

119. Ryu, Y.; Schultz, P. G., *Nat. Methods* 2006, 3, 263-65.

120. van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A., *J. Am. Chem. Soc.* 2000, 122, 1282-88.

121. Kiick, K. L.; Tirrell, D. A., *Tetrahedron* 2000, 56, 9487-93.

122. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A., *FEBS Lett.* 2001, 502, 25-30.

123. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R., *Proc. Natl. Acad. Sci. U.S.A.* 2002, 99, 19-24.

124. Link, A. J.; Tirrell, D. A., *J. Am. Chem. Soc.* 2003, 125, 11164-65.

125. Link, A. J.; Vink, M. K. S.; Tirrell, D. A., *J. Am. Chem. Soc.* 2004, 126, 10598-602.

126. Link, A. J.; Tirrell, D. A., *Methods* 2005, 36, 291-98.

127. Link, A. J.; Vink, M. K. S.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A., *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 10180-85.

128. Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G., *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 7566-71.

129. Ganoza, M. C.; Buckingham, K.; Hader, P.; Neilson, T., *J. Biol. Chem.* 1984, 259, 14101-4.

130. Ilegems E, P. H., Vogel H., *Protein Eng. Des. Sel.* 2004, 17, 821-27.

131. Kresge, N.; Simoni, R. D.; Hill, R. L., *J. Biol. Chem.* 2009, 284, e5-e7.

132. Pastrnak, M.; Magliery, T. J.; Schultz, P. G., *Helv. Chim. Acta* 2000, 83, 2277-86.

133. Cowie, D. B.; Cohen, G. N., *Biochim. Biophys. Acta* 1957, 26, 252-61.

134. McNaught, A. D.; Wilkinson, A., *IUPAC Compendium of Chemical Terminology*, 2nd ed.; 1997; Vol. 64, p 147.

135. van Kasteren, S. I.; Kramer, H. B.; Jensen, H. H.; Campbell, S. J.; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G., *Nature* 2007, 446, 1105-09.

136. van Kasteren, S. I.; Kramer, H. B.; Gamblin, D. P.; Davis, B. G., *Nat. Protoc.* 2007, 2, 3185-94.

137. Floyd, N.; Vijayakrishnan, B.; Koeppe, J. R.; Davis, B. G., *Angew. Chem., Int. Ed.* 2009, 48, 7798-802.

138. Budisa, N., *Angew. Chem., Int. Ed.* 2004, 43, 6426-63.

139. Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A., *Angew. Chem., Int. Ed.* 2000, 39, 2148-52

140. Schoffelen, S.; Lambermon, M. H. L.; van Eldijk, M. B.; van Hest, J. C. M., *Bioconjug. Chem.* 2008, 19, 1127-31.

141. Cornish, V. W.; Hahn, K. M.; Schultz, P. G., *J. Am. Chem. Soc.* 1996, 118, 8150-51.

142. Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G., *Proc. Natl. Acad. Sci. U.S.A.* 2003, 100, 56-61.

143. Zhang, Z.; Gildersleeve, J.; Yang, Y.-Y.; Xu, R.; Loo, J. A.; Uryu, S.; Wong, C.-H.; Schultz, P. G., *Science* 2004, 303, 371-73.

144. Xu, R.; Hanson, S. R.; Zhang, Z.; Yang, Y.-Y.; Schultz, P. G.; Wong, C.-H., *J. Am. Chem. Soc.* 2004, 126, 15654-55.

145. Fahmi, N. E.; Dedkova, L.; Wang, B.; Golovine, S.; Hecht, S. M., *J. Am. Chem. Soc.* 2007, 129, 3586-97.

146. Tornoe, C. W.; Christensen, C.; Meldal, M., *J. Org. Chem.* 2002, 67, 3057-64.

147. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., *Angew. Chem., Int. Ed.* 2002, 41, 2596-99.

148. Merkel, L.; Beckmann, H. S. G.; Wittmann, V.; Budisa, N., *ChemBioChem* 2008, 9, 1220-24.

149. Wan, Q.; Chen, J.; Chen, G.; Danishefsky, S. J., *J. Org. Chem.* 2006, 71, 8244-49.

150. Seebeck, F. P.; Szostak, J. W., *J. Am. Chem. Soc.* 2006, 128, 7150-51.

151. Wang, J.; Schiller, S. M.; Schultz, P. G., *Angew. Chem., Int. Ed.* 2007, 46, 6849-51.

152. Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G., *J. Am. Chem. Soc.* 2008, 130, 5052-53.

153. Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. J. L.; Davis, B. G., *J. Am. Chem. Soc.* 2008, 130, 9642-43.

154. Posner, T., *Ber. Dtsch. Chem. Ges.* 1905, 38, 646-57.

155. Hoyle, C. E., *Polym. Prepr.* 2008, 49, 155-56.

156. Hoyle, C. E.; Lee, T. Y.; Roper, T., *J. Polym. Sci.* 2004, 42A, 5301-38.

157. Rim, C.; Lahey, L. J.; Patel, V. G.; Zhang, H.; Son, D. Y., *Tetrahedron Lett.* 2009, 50, 745-47.

158. Griesbaum, K., *Angew. Chem., Int. Ed.* 1970, 9, 273-87.

159. Rissing, C.; Son, D. Y., *Organometallics* 2008, 27, 5394-97.

160. Rissing, C.; Son, D. Y., *Organometallics* 2009, 28, 3167-72.

161. Triola, G.; Brunsveld, L.; Waldmann, H., *J. Org. Chem.* 2008, 73, 3646-49.

162. Dondoni, A., *Angew. Chem., Int. Ed.* 2008, 47, 8995-97.

163. Fiore, M.; Marra, A.; Dondoni, A., *J. Org. Chem.* 2009, 74, 4422-25.

164. Fiore, M.; Chambery, A.; Marra, A.; Dondoni, A., *Org. Biomol. Chem.* 2009, 7, 3910-13.

165. Dondoni, A.; Massi, A.; Nanni, P.; Roda, A., In *unpublished work*, 2009.

166. Wittrock, S.; Becker, T.; Kunz, H., *Angew. Chem., Int. Ed.* 2007, 46, 5226-30.

167. Jonkheijm, P.; Weinrich, D.; Koehn, M.; Engelkamp, H.; Christianen, P. C. M.; Kuhlmann, J.; Maan, J. C.; Nuesse, D.; Schroeder, H.; Wacker, R.; Breinbauer, R.; Niemeyer, C. M.; Waldmann, H., *Angew. Chem., Int. Ed.* 2008, 47, 4421-24.

168. Hecht, M.-L.; Stallforth, P.; Silva, D. V.; Adibekian, A.; Seeberger, P. H., *Curr. Opin. Chem. Biol.* 2009, 13, 354-59.

169. Buskas, T.; Thompson, P.; Boons, G.-J., *Chem. Commun.* 2009, 5335-49.

170. Buskas, T.; Li, Y.; Boons, G.-J., *Chem. Eur. J.* 2004, 10, 3517-24.

171. Pollard, A. J.; Perrett, K. P.; Beverley, P. C., *Nat. Rev. Immunol.* 2009, 9, 213-20.

172. *World Health Organisation: Fact Sheet #294*, 2005.

173. Fernandez Santana, V.; Peña Icart, L.; Beurret, M.; Costa, L.; Verez Bencomo, V.; Minoru, F., *Meth. Enzymol.* 2006, 415, 153-63.

174. Verez-Bencomo, V.; Fernandez-Santana, V.; Hardy, E.; Toledo, M. E.; Rodriguez, M. C.; Heynngnezz, L.; Rodriguez, A.; Baly, A.; Herrera, L.; Izquierdo, M.; Villar, A.; Valdes, Y.; Cosme, K.; Deler, M. L.; Montane, M.; Garcia, E.; Ramos, A.; Aguilar, A.; Medina, E.; Torano, G.; Sosa, I.; Hernandez, I.; Martinez, R.; Muzachio, A.; Carmenates, A.; Costa, L.; Cardoso, F.; Campa, C.; Diaz, M.; Roy, R., *Science* 2004, 305, 522-25.

175. Hoft, D. F., *The Lancet* 2008, 372, 164-75.

176. Behr, M. A.; Small, P. M., *Nature* 1997, 389, 133-34.

177. Kallenius, G.; Pawlowski, A.; Hamasur, B.; Svenson, S. B., *Trends Microbiol.* 2008, 16, 456-62.

178. Hamasur, B.; Källenius, G.; Svenson, S. B., *Vaccine* 1999, 17, 2853-61.

179. Hamasur, B.; Haile, M.; Pawlowski, A.; Schröder, U.; Williams, A.; Hatch, G.; Hall, G.; Marsh, P.; Källenius, G.; Svenson, S. B., *Vaccine* 2003, 21, 4081-93.

180. Holemann, A.; Seeberger, P. H., *Protein-Carbohydr. Interact. Infect. Dis.* 2006, 151-74.

181. Phalipon, A.; Tanguy, M.; Grandjean, C.; Guerreiro, C.; Belot, F.; Cohen, D.; Sansonetti, P. J.; Mulard, L. A., *J. Immunol.* 2009, 182, 2241-47.

182. Werz, D. B.; Seeberger, P. H., *Angew. Chem., Int. Ed.* 2005, 44, 6315-18.

183. Liu, X.; Siegrist, S.; Amacker, M.; Zurbriggen, R.; Pluschke, G.; Seeberger, P. H., *ACS Chem. Biol.* 2006, 1, 161-64.

184. *World Health Organisation: Report on the Global AIDS Epidemic*, 2008.

185. Appay, V., *Eur. J. Immunol.* 2009, 39, 1999-2003.

186. Walker, L. M.; Phogat, S. K.; Chan-Hui, P.-Y.; Wagner, D.; Phung, P.; Goss, J. L.; Wrin, T.; Simek, M. D.; Fling, S.; Mitcham, J. L.; Lehrman, J. K.; Priddy,

F. H.; Olsen, O. A.; Frey, S. M.; Hammond, P. W.; Protocol, G. P. I.; Kaminsky, S.; Zamb, T.; Moyle, M.; Koff, W. C.; Poignard, P.; Burton, D. R., *Science* 2009, 1178746.

187. Calarese, D. A.; Scanlan, C. N.; Zwick, M. B.; Deechongkit, S.; Mimura, Y.; Kunert, R.; Zhu, P.; Wormald, M. R.; Stanfield, R. L.; Roux, K. H.; Kelly, J. W.; Rudd, P. M.; Dwek, R. A.; Katinger, H.; Burton, D. R.; Wilson, I. A., *Science* 2003, 300, 2065-71.

188. Mandal, M.; Dudkin, V. Y.; Geng, X.; Danishefsky, S. J., *Angew. Chem., Int. Ed.* 2004, 43, 2557-61.

189. Geng, X.; Dudkin, V. Y.; Mandal, M.; Danishefsky, S. J., *Angew. Chem., Int. Ed.* 2004, 43, 2562-65.

190. Wang, J.; Li, H.; Zou, G.; Wang, L.-X., *Org. Biomol. Chem.* 2007, 5, 1529-40.

191. Astronomo, R. D.; Lee, H.-K.; Scanlan, C. N.; Pantophlet, R.; Huang, C.-Y.; Wilson, I. A.; Blixt, O.; Dwek, R. A.; Wong, C.-H.; Burton, D. R., *J. Virol.* 2008, 82, 6359-68.

192. Ni, J.; Song, H.; Wang, Y.; Stamatos, N. M.; Wang, L.-X., *Bioconjug. Chem.* 2006, 17, 493-500.

193. Krauss, I. J.; Joyce, J. G.; Finnefrock, A. C.; Song, H. C.; Dudkin, V. Y.; Geng, X.; Warren, J. D.; Chastain, M.; Shiver, J. W.; Danishefsky, S. J., *J. Am. Chem. Soc.* 2007, 129, 11042-44.

194. Wang, S.-K.; Liang, P.-H.; Astronomo, R. D.; Hsu, T.-L.; Hsieh, S.-L.; Burton, D. R.; Wong, C.-H., *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105, 3690-95.

195. Joyce, J. G.; Krauss, I. J.; Song, H. C.; Opalka, D. W.; Grimm, K. M.; Nahas, D. D.; Esser, M. T.; Hrin, R.; Feng, M.; Dudkin, V. Y.; Chastain, M.; Shiver, J. W.; Danishefsky, S. J., *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105, 15684-89.

196. Kim, Y. J.; Varki, A., *Glycoconjugate J.* 1997, 14, 569-76.

197. Dennis, J. W.; Laferte, S.; Waghorne, C.; Breitman, M. L.; Kerbel, R. S., *Science* 1987, 236, 582-85.

198. Dube, D. H.; Bertozzi, C. R., *Nat. Rev. Drug. Discov.* 2005, 4, 477-88.

199. Nagorny, P.; Kim, W. H.; Wan, Q.; Lee, D.; Danishefsky, S. J., *J. Org. Chem.* 2009, 74, 5157-62.

200. Ragupathi, G.; Koide, F.; Livingston, P. O.; Cho, Y. S.; Endo, A.; Wan, Q.; Spassova, M. K.; Keding, S. J.; Allen, J.; Ouerfelli, O.; Wilson, R. M.; Danishefsky, S. J., *J. Am. Chem. Soc.* 2006, 128, 2715-25.

201. Cho, Y. S.; Wan, Q.; Danishefsky, S. J., *Bioorg. Med. Chem.* 2005, 13, 5259-66.

202. Williams, L. J.; Harris, C. R.; Glunz, P. W.; Danishefsky, S. J., *Tetrahedron Lett.* 2000, 41, 9505-08.

203. Ragupathi, G.; Park, T. K.; Zhang, S.; Kim, I. J.; Graber, L.; Adluri, S.; Lloyd, K. O.; Danishefsky, S. J.; Livingston, P. O., *Angew. Chem., Int. Ed. Engl.* 1997, 36, 125-28.

204. Zhu, J.; Wan, Q.; Ragupathi, G.; George, C. M.; Livingston, P. O.; Danishefsky, S. J., *J. Am. Chem. Soc.* 2009, 131, 4151-58.

205. van Hest, J. C. M.; Tirrell, D. A., *FEBS Lett.* 1998, 428, 68-70.

206. Lacombe, J. M.; Rakotomanomana, N.; Pavia, A. A., *Tetrahedron Lett.* 1988, 29, 4293-96.

207. Schips, C.; Ziegler, T., *J. Carbohydr. Chem.* 2005, 24, 773-88.

208. Knapp, S.; Myers, D. S., *J. Org. Chem.* 2002, 67, 2995-99.

209. Sletten, E. M.; Bertozzi, C. R., *Angew. Chem., Int. Ed.* 2009, 48, 2-27.

210. Floyd, N.; Oldham, N. J.; Eyles, C. J.; Brouard, M.; Davis, B. G. Abstracts of Papers, 237th ACS National Meeting, Salt Lake City, UT, United States, March 22-26, 2009, ORGN 260, #1229118.

211. Alessandro, D.; Alessandro, M.; Paolo, N.; Aldo, R., *Chem. Eur. J.* 2009, In Press.

212. Smith, P. W. G.; Drinkwater, D. J., *J. Chem. Soc., C,* 1971, 1305-7.

213. Kennewell, P. D.; Matharu, S. S.; Taylor, J. B.; Westwood, R., *J. Chem. Soc., Perkin Trans. 1* 1982, 2485-500.

214. Biagini, S. C. G.; Gibson neé Thomas, S. E.; Keen, S. P., *J. Chem. Soc., Perkin Trans. 1* 1998, 16, 2485-500.

215. Kramer, D. H. University of Oxford, 2006.

216. Wolf, L. B.; Sonke, T.; Tjen, K. C. M. F.; Kaptein, B.; Broxterman, Q. B.; Schoemaker, H. E.; Rutjes, F. P. J. T., *Adv. Synth. Catal.* 2001, 343, 662-74.

217. Johannesson, P.; Lindeberg, G.; Tong, W.; Gogoll, A.; Synnergren, B.; Hyberg, F.; Karlen, A.; Hallberg, A., *J. Med. Chem.* 1999, 42, 4524-37.

218. Gorohovsky, S.; Meir, S.; Shkoulev, V.; Byk, G.; Gellerman, G., *Synlett* 2003, 10, 1411-14.

219. An aliquot was removed from the reaction mixture and was submitted to microwave irradiation for 10 minutes at 50$^{o}$C, 10 minutes at 80$^{o}$C and 10 minutes at 100$^{o}$C. The composition of the reaction mixture remained unchanged by t.l.c.

220. Adamczyk, M.; Johnson, D. D.; Reddy, R. E., *Tetrahedron Asymmetry* 1999, 10, 775-81.

221. Kokotos, G.; Padron, J. M.; Martın, T.; Gibbons, W. A.; Martın, V. S., *J. Org. Chem.* 1998, 63, 3741-44.

222. Mehta, A.; Jaouhari, R.; Benson, T. J.; Douglas, K. T., *Tetrahedron Lett.* 1992, 33, 5441-44.

223. Kennewell, P. D.; Matharu, S. S.; Taylor, J. B.; Westwood, R.; Sammes, P. G., *J. Chem. Soc., Perkin Trans. 1* 1982, 2553-62.

224. Padrón, J. M.; Kokotos, G.; Martín, T.; Markidis, T.; Gibbons, W. A.; Martín, V. S., *Tetrahedron Asymmetry* 1998, 9, 3381-94.

225. Bhat, R. G.; Porhiel, E.; Saravanan, V.; Chandrasekaran, S., *Tetrahedron Lett.* 2003, 44, 5251-53.

226. Pine, S. H.; Pettit, R. J.; Geib, G. D.; Cruz, S. G.; Gallego, C. H.; Tijerina, T.; Pine, R. D., *J. Org. Chem.* 1985, 50, 1212-16.

227. Tebbe, F. N.; Parshall, G. W.; Reddy, G. S., *J. Am. Chem. Soc.* 1978, 100, 3611-13.

228. Kurokawa, N.; Ohfune, Y., *Tetrahedron* 1993, 49, 6195-222.

229. Evans, D. A.; Britton, T. C.; Ellman, J. A., *Tetrahedron Lett.* 1987, 28, 6141-44.

230. Goundry, W. R. F.; Baldwin, J. E.; Lee, V., *Tetrahedron* 2003, 59, 1719-29.

231. Callant, P.; D'Haenens, L.; Vandewalle, M., *Synth. Commun.* 1984, 14, 155-61.

232. Ohira, S., *Synth. Commun.* 1989, 19, 561-4.

233. Muller, S.; Liepold, B.; Roth, G. J.; Bestmann, H. J., *Synlett* 1996, 521-22.

234. Lindlar, H., *Helv. Chim. Acta* 1952, 35, 446-50.

235. Hawker, S.; Bhatti, M. A.; Griffin, K. G., *Chim. Oggi* 1992, 10, 49-51.

236. Greene, T. W.; Wuts, P. G. M., *Protective Groups in Organic Synthesis*, 3rd ed.; 1999.

237. North, M., *Principles and Applications of Stereochemistry*, 1998; pp 115-16.

238. Wenzel, T. J. W., J. D., *Chirality* 2003, 15, 256-70.

239. Rothchild, R., *Enantiomer* 2000, 5, 457-71.

240. Pirkle, W. H. H., D. J.;, *Topics in Stereochemistry*, 1982; Vol. 13, pp 263-331.

241. Andersen, K. K.; Gash, D. M.; Robertson, J. D., *Asymmetric Synthesis: Analytical Methods.*, 1983; Vol. 1, pp 45-58.

242. Myers, A. G.; Gleason, J. L.; Yoon, T.; Kung, D. W., *J. Am. Chem. Soc.* 1997, 119, 656-73.

243. Morrison, J. D., *Asymmetric Synthesis*, 1983; Vol. 1.

244. Sanger, F., *Biochem. J.* 1945, 39, 507-10.

245. Dale, J. A.; Dull, D. L.; Mosher, H. S., *J. Org. Chem.* 1969, 34, 2543-49.

246. Dale, J. A.; Mosher, H. S., *J. Am. Chem. Soc.* 1973, 95, 512-19.

247. Audic, S.; Lescot, M.; Claverie, J.-M.; Scholz, H. C., *BMC Genomics* 2009, 10, No pp given.

248. Calderon-Copete, S. P.; Wigger, G.; Wunderlin, C.; Schmidheini, T.; Frey, J.; Quail, M. A.; Falquet, L., *BMC Bioinformatics* 2009, 10.

249. Cherng, Y. J., *Tetrahedron* 2000, 56, 8287-89.

250. Zhai, W., *Tetrahedron* 1988, 5425-30.

251. Miller, J. R.; Pulley, S. R.; Hegedus, L. S.; DeLombaert, S., *J. Am. Chem. Soc.* 1992, 114, 5602-07.

252. Arnold, U.; Hinderaker, M. P.; Koeditz, J.; Golbik, R.; Ulbrich-Hofmann, R.; Raines, R. T., *J. Am. Chem. Soc.* 2003, 125, 7500-01.

253. Heidlas, J. E.; Lees, W. J.; Pale, P.; Whitesides, G. M., *J. Org. Chem.* 1992, 57, 146-51.

254. Sen, A. K.; Banerji, N., *Indian J. Chem.* 1989, 28B, 818-23.

255. Bonner, W. A.; Kahn, J. E., *J. Am. Chem. Soc.* 1951, 73, 2241-5.

256. Johnston, B. D.; Pinto, B. M., *J. Org. Chem.* 2000, 65, 4607-17.

257. Zemplen, G.; Kuntz, A., *Brit.* 1923, 56B, 1705-10.

258. Ibatullin, F. M.; Selivanov, S. I., *Tetrahedron Lett.* 2002, 43, 9577-80.

259. Blanc-Muesser, M.; Defaye, J.; Driguez, H., *Carbohydr. Res.* 1978, 67, 305-28.

260. Blanc-Muesser, M.; Defaye, J.; Driguez, H., *J. Chem. Soc., Perkin Trans. 1* 1982, 15-18.

261. Gadelle, A.; Defaye, J.; Pedersen, C., *Carbohydr. Res.* 1990, 200, 497-8.

262. Schneider, W.; Gille, R.; Eisfeld, K., *Ber. Dtsch. Chem. Ges.* 1928, 61B, 1244-59.

263. Schneider, W.; Leonhardt, H., *Ber. Dtsch. Chem. Ges.* 1929, 62B, 1384-9.

264. Larsson, R.; Ramström, O., *Eur. J. Org. Chem.* 2006, 285-91.

265. Hammond, G. S.; Neuman, R. C., Jr., *J. Am. Chem. Soc.* 1963, 85, 1501-8.

266. Gupta, S. N.; Nandi, U. S., *J. Polym. Sci.* 1968, 6B, 457-9.

267. Chen, C. C.; Milburn, M. V.; Overberger, C. G., *J. Polym. Sci.* 1986, 24C, 627-30.

268. Fan, S.; Boey, F. Y. C.; Abadie, M. J. M., *Eur. Polymer J.* 2008, 44, 2123-29.

269. Szmant, H. H.; Mata, A. J.; Namis, A. J.; Panthananickal, A. M., *Tetrahedron* 1976, 32, 2665-80.

270. Fava, A.; Reichenbach, G.; Peron, U., *J. Am. Chem. Soc.* 1967, 89, 6696-700.

271. Kuhne, G.; Diesen, J. S.; Klemm, E., *Angew. Makromol. Chem.* 1996, 242, 139-45.

272. Kharasch, M. S.; Nudenberg, W.; Mantell, G. J., *J. Org. Chem.* 1951, 16, 524-32.

273. Kharasch, M. S.; Nudenberg, W.; Meltzer, T. H., *J. Org. Chem.* 1953, 18, 1233-6.

274. Deprele, S.; Montchamp, J.-L., *J. Org. Chem.* 2001, 66, 6745-55.

275. Hideki, Y.; Hiroshi, S.; Koichiro, O., *Bull. Chem. Soc. Jpn.* 2001, 74, 225-35.

276. Perchyonok, V. T.; Lykakis, I. N.; Tuck, K. L., *Green Chem.* 2008, 10, 153-63.

277. Aguilar, C. F.; Sanderson, I.; Moracci, M.; Ciaramella, M.; Nucci, R.; Rossi, M.; Pearl, L. H., *J. Mol. Biol.* 1997, 271, 789-802.

278. Vetting, M. W.; Hegd, S. S.; Hazleton, K. Z.; Blanchard, J. S., *Protein Sci.* 2007, 16, 755-60.

279. Strable, E.; Prasuhn, D. E.; Udit, A. K.; Brown, S.; Link, A. J.; Ngo, J. T.; Lander, G.; Quispe, J.; Potter, C. S.; Carragher, B.; Tirrell, D. A.; Finn, M. G., *Bioconjug. Chem.* 2008, 19, 866-75.

280. Kozlovska, T. M.; Cielens, I.; Dreilinna, D.; Dislers, A.; Baumanis, V.; Ose, V.; Pumpens, P., *Gene* 1993, 137, 133-37.

281. Strable, E.; Finn, M. G., *Curr. Top. Microbiol. Immunol.* 2009, 327, 1-21.

282. Sambrook, J.; Russell, D., *Molecular Cloning: A Laboratory Manual*, 3rd ed.; 2001.

283. Reinhold, B. B.; Reinhold, V. N., *J. Am. Soc. Mass Spectrom.* 1992, 3, 207-15.

284. Mann, M.; Meng, C. K.; Fenn, J. B., *Anal. Chem.* 2002, 61, 1702-08.

285. Gadgil, H. S.; Pipes, G. D.; Dillon, T. M.; Treuheit, M. J.; Bondarenko, P. V., *J. Am. Soc. Mass Spectrom.* 2006, 17, 867-72.

286. Wu, L.-Z. S., Yue-Biao; Xie, Jin-Bin; Wang, Wei *J. Mol. Struct.* 2008, 882, 101-06.

287. Permyakov, E. A.; Permyakov, S. E.; Deikus, G. Y.; Morozova-Roche, L. A.; Grishchenko, V. M.; Kalinichenko, L. P.; Uversky, V. N., *Proteins: Struct. Funct. Genet.* 2003, 51, 498-503.

288. Borkman, R. F.; McLaughlin, J., *Photochem. Photobiol.* 1995, 62, 1046-51.

289. Clymer, D. J.; Geren, C. R.; Ebner, K. E., *Biochemistry* 1976, 15, 1093-7.

290. Artyukhov, V. G.; Gusinskaya, V. V.; Dvurekova, E. A.; Rubtsov, M. P., *Biofizika* 2007, 52, 1001-07.

291. Floyd, N.; Oldham, N. J.; Eyles, C. J.; Taylor, S.; Filatov, D. A.; Brouard, M.; Davis, B. G., *J. Am. Chem. Soc.* 2009, 131, 12518-19.

292. Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., *Macromolecules* 1994, 27, 5471-75.

293. Biemann, K., *Meth. Enzymol.* 1990, 193, 886-7.

294. Geoghegan, K. F.; Dixon, H. B. F.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; McColl, A. S.; Kamath, A. V.; Stroh, J. G., *Anal. Biochem.* 1999, 267, 169-84.

295. Kaltgrad, E.; O'Reilly, M. K.; Liao, L.; Han, S.; Paulson, J. C.; Finn, M. G., *J. Am. Chem. Soc.* 2008, 130, 4578.

296. Vallegaard, K.; Fridborg, K., *Acta Crystallogr.* 1994, 50D, 105-09.

297. I. Vasiljeva, T. K., I. Cielens, A. Strelnikova, A. Kazaks, V.; Ose, P. P., *FEBS Lett.* 1998, 431, 7.

298. Chackerian, B., *Expert Rev. Vaccines* 2007, 6, 381.

299. Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C., *Anal. Biochem.* 1985, 150, 76-85.

300. Bernardes, G. J. L.; Gamblin, D. P.; Davis, B. G., *Angew. Chem., Int. Ed.* 2006, 45, 4007-11.

301. Ando R, H. H., Yamamoto-Hino M, Mizuno H, Miyawaki A., *Proc. Natl. Acad. Sci. U.S.A.* 2002, 99, 12651-56.

302. Wiedenmann, J.; Ivanchenko, S.; Oswald, F.; Schmitt, F.; Roecker, C.; Salih, A.; Spindler, K.-D.; Nienhaus, G. U., *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 15905-10.

303. Hayashi, I. M., H.; Tong, K. I.; Furuta, T.; Tanaka, F.; Yoshimura, M.; Miyawaki, A.; Ikura, M., *J. Mol. Biol.* 2007, 372, 918-26.

304. Mizuno, H.; Mal, T. K.; Tong, K. I.; Ando, R.; Furuta, T.; Ikura, M.; Miyawaki, A., *Mol. Cell* 2003, 12, 1051-58.

305. Nienhaus, K.; Nienhaus, G. U.; Wiedenmann, J.; Nar, H., *Proc. Natl. Acad. Sci. U.S.A.* 2005, 102, 9156-9.

306. Jones, B. J.; Vergne, M. J.; Bunk, D. M.; Locascio, L. E.; Hayes, M. A., *Anal. Chem.* 2007, 79, 1327-32.

307. Duff, M. R., Jr.; Kumar, C. V., *Angew. Chem., Int. Ed.* 2006, 45, 137-39.

308. Suzuki, A.; Hasegawa, M.; Ishii, M.; Matsumura, S.; Toshima, K., *Bioorg. Med. Chem. Lett.* 2005, 15, 4624-27.

309. Reiko Miyake, J. T. O., Dadong Xu, William M. Jackson, and Claude F. Meares, *J. Am. Chem. Soc.* 1999, 121, 7453-54.

310. Zein, N.; Casazza, A. M.; Doyle, T. W.; Leet, J. E.; Schroeder, D. R.; Solomon, W.; Nadler, S. G., *Proc. Natl. Acad. Sci. U.S.A.* 1993, 90, 8009-12.

311. Plourde, G., 2nd; El-Shafey, A.; Fouad, F. S.; Purohit, A. S.; Jones, G. B., *Bioorg. Med. Chem. Lett.* 2002, 12, 2985-8.

312. Kumar, C. V.; Thota, J., *Inorg. Chem.* 2005, 44, 825-27.

313. Buranaprapuk, A.; Kumar, C. V.; Jockusch, S.; Turro, N. J., *Tetrahedron* 2000, 56, 7019-25.

314. Kumar, C. V.; Buranaprapuk, A.; Opiteck, G. J.; Moyer, M. B.; Jockusch, S.; Turro, N. J., *Proc. Natl. Acad. Sci. U.S.A.* 1998, 95, 10361-6.

315. Kumar, C. V. B., A.;, *J. Am. Chem. Soc.* 1999, 121, 4262-70.

316. Kumar, C. V.; Buranaprapuk, A., *Angew. Chem., Int. Ed.* 1997, 36, 2085-87.

317. Grammer, J. C.; Loo, J. A.; Edmonds, C. G.; Cremo, C. R.; Yount, R. G., *Biochemistry* 1996, 35, 15582-92.

318. Peters, F. B.; Brock, A.; Wang, J.; Schultz, P. G., *Chem. Biol.* 2009, 16, 148-52.

319. Endo, M. M., *Angew. Chem., Int. Ed.*

2004, 43, 5643.

320. England, P. M.; Lester, H. A.; Davidson, N.; Dougherty, D. A., *Proc. Natl. Acad. Sci. U.S.A.* 1997, 94, 11025-30.

321. Arnau, J.; Lauritzen, C.; Petersen, G. E.; Pedersen, J., *Protein Expr. Purif.* 2006, 48, 1-13.

322. Gevaert, K.; Vandekerckhove, J., *Electrophoresis* 2000, 21, 1145-54.

323. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A., *Nucleic Acids Res.* 2003, 31, 3784-8.

324. Edman, P., *Acta Chem. Scand.* 1950, 4, 283-93.

325. Residues are numbered according to the FASTA sequence associated with PDB ID: 1GOW.

326. Kelly, S. M.; Jess, T. J.; Price, N. C., *Biochim. Biophys. Acta* 2005, 1751, 119-39.

327. Frank Sobott; Stephen J. Watt; Julia Smith; Mariola J. Edelmann; Kramer, H. B.; Kessler, B. M., *J. Am. Soc. Mass Spectrom.* 2009.

328. Takai, K.; Sasai, Y.; Morimoto, H.; Yamazaki, H.; Yoshii, H.; Inoue, S., *J. Biol. Chem.* 1984, 259, 4452-7.

329. Soloshonok, V. A.; Ueki, H.; Ellis, T. K.; Yamada, T.; Ohfune, Y., *Tetrahedron Lett.* 2005, 46, 1107-10.

330. Brettle, R.; Mosedale, A. J., *J. Chem. Soc., Perkin Trans. 1* 1988, 2185-95.

331. Andreani, A.; Rambaldi, M.; Locatelli, A.; Pifferi, G., *Eur. J. Med. Chem.* 1994, 29, 903-6.

332. Soloshonok, V. A.; Ueki, H.; Ellis, T. K.; Yamada, T.; Ohfune, Y., *Tetrahedron Lett.* 2005, 46, 1107-10.

333. Rasnik, I.; McKinney, S. A.; Ha, T., *Nat. Methods* 2006, 3, 891-93.

334. Merkel, P. B.; Nilsson, R.; Kearns, D. R., *J. Am. Chem. Soc.* 1972, 94, 1030-1.

335. Epe, B., *Chem. Biol. Interact.* 1991, 80, 239-60.

336. Snyder, J. W.; Lambert, J. D. C.; Ogilby, P. R., *Photochem. Photobiol.* 2006, 82, 177-84.

337. Karotki, A.; Kruk, M.; Drobizhev, M.; Rebane, A.; Nickel, E.; Spangler, C. W., *IEEE J. Sel. Top. Quantum. Electron.* 2001, 7, 971-75.

338. Frederiksen, P. K.; McIlroy, S. P.; Nielsen, C. B.; Nikolajsen, L.; Skovsen, E.; Jorgensen, M.; Mikkelsen, K. V.; Ogilby, P. R., *J. Am. Chem. Soc.* 2005, 127, 255-69.

339. Kropp, P. J.; Fravel, H. G.; Fields, T. R., *J. Am. Chem. Soc.* 1976, 98, 840-41.

340. Bent, D. V.; Hayon, E., *J. Am. Chem. Soc.* 1975, 97, 2612-19.

341. Kleywegt, G. J., *J. Mol. Biol.* 1999, 285, 1887-97.

342. Chi, Y. I.; Martinez-Cruz, L. A.; Jancarik, J.; Swanson, R. V.; Robertson, D. E.; Kim, S. H., *FEBS Lett.* 1999, 445, 375-83.

343. Beil, S.; Kehrli, H.; James, P.; Staudenmann, W.; Cook, A. M.; Leisinger, T.; Kertesz, M. A., *Eur. J. Biochem.* 1995, 229, 385-94.

344. Bomati, E. K.; Austin, M. B.; Bowman, M. E.; Dixon, R. A.; Noel, J. P., *J. Biol. Chem.* 2005, 280, 30496-503.

345. Perugino, G.; Trincone, A.; Giordano, A.; Van der Oost, J.; Kaper, T.; Rossi, M.; Moracci, M., *Biochemistry* 2003, 42, 8484-93.

346. Chi, Y. I.; Martinez-Cruz, L. A.; Jancarik, J.; Swanson, R. V.; Robertson, D. E.; Kim, S. H., *FEBS Lett.* 1999, 445, 375-83.

347. Boltes, I.; Czapinska, H.; Kahnert, A.; von Bulow, R.; Dierks, T.; Schmidt, B.; von Figura, K.; Kertesz, M. A.; Uson, I., *Structure* 2001, 9, 483-91.

348. Dierks, T.; Miech, C.; Hummerjohann, J.; Schmidt, B.; Kertesz, M. A.; Von Figura, K., *J. Biol. Chem.* 1998, 273, 25560-64.

349. Eichhorn, E.; Van Der Ploeg, J. R.; Kertesz, M. A.; Leisinger, T., *J. Biol. Chem.* 1997, 272, 23031-36.

350. Bailey, T. L.; Elkan, C., *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1994, 2, 28-36.

351. Edgar, R. C., *Nucleic Acids Res.* 2004, 32, 1792-97.

352. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., *Nucleic Acids Res.* 2009, 37, D233-D38.

353. Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J., *Bioinformatics* 2009, 25, 1189-91.

354. Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E., *Genome Res.* 2004, 14, 1188-90.

355. Schneider, T. D.; Stephens, R. M., *Nucleic Acids Res.* 1990, 18, 6097-100.

356. Eddy, S. R., *Bioinformatics* 1998, 14, 755-63.

357. Huson, D. H.; Richter, D. C.; Rausch, C.; Dezulian, T.; Franz, M.; Rupp, R., *BMC Bioinformatics* 2007, 8, 460.

358. Milne, I.; Lindner, D.; Bayer, M.; Husmeier, D.; McGuire, G.; Marshall, D. F.; Wright, F., *Bioinformatics* 2009, 25, 126-7.

359. Catanzano, F.; Graziano, G.; De Paola, B.; Barone, G.; D'Auria, S.; Rossi, M.; Nucci, R., *Biochemistry* 1998, 37, 14484-90.

360. D'Auria, S.; Rossi, M.; Barone, G.; Catanzano, F.; Del Vecchio, P.; Graziano, G.; Nucci, R., *J. Biochem.* 1996, 120, 292-300.

361. D'Auria, S.; Barone, R.; Rossi, M.; Nucci, R.; Barone, G.; Fessas, D.; Bertoli, E.; Tanfani, F., *Biochem. J.* 1997, 323 833-40.

362. D'Auria, S.; Nucci, R.; Rossi, M.; Bertoli, E.; Tanfani, F.; Gryczynski, I.; Malak, H.; Lakowicz, J. R., *J. Biochem.* 1999, 126, 545-52.

363. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C., *Science* 1994, 263, 802-5.

364. Sawin, K. E., *Methods Cell Biol.* 1999, 58, 123-38.

365. Gerisch, G.; Mueller-Taubenberger, A., *Meth. Enzymol.* 2003, 361, 320-37.

366. Wahlfors, J.; Loimas, S.; Pasanen, T.; Hakkarainen, T., *Histochem. Cell Biol.* 2001, 115, 59-65.

367. Robart, F. D.; Ward, W. W., *Photochem. Photobiol.* 1990, S92.

368. Ward, W. W.; Bokman, S. H., *Biochemistry* 1982, 21, 4535-40.

369. Ward, W. W.; Prentice, H. J.; Roth, A. F.; Cody, C. W.; Reeves, S. C., *Photochem. Photobiol.* 1982, 35, 803-8.

370. Roth, A. F.; Ward, W. W., *Photochem. Photobiol.* 1983, 37, S71.

371. Fisher, H. E.; Mintz, C. S., *J. Ind. Microbiol. Biotechnol.* 2000, 24, 323-26.

372. Chen, Y.; Leong, S. S. J., *J. Chromatogr.* 2009, 1216A, 4877-86.

373. Arslan, P. E.; Chakrabartty, A., *Biochem. Cell Biol.* 2009, 87, 631-39.

374. Vallejo, L. F.; Rinas, U., *Microb. Cell Fact.* 2004, 3, No pp given.

375. Cody, C. W.; Prasher, D. C.; Westler, W. M.; Prendergast, F. G.; Ward, W. W., *Biochemistry* 1993, 32, 1212-8.

376. Vincentelli, R.; Canaan, S.; Campanacci, V.; Valencia, C.; Maurin, D.; Frassinetti, F.; Scappucini-Calvo, L.; Bourne, Y.; Cambillau, C.; Bignon, C., *Protein Sci.* 2004, 13, 2782-92.

377. Leland, P.; Grabski, A., iFOLD Protein Refolding System 1. *Novagen: inNovations* 2006, pp 8-9.

378. Xie, J.-B.; Zhou, J.-M., *Biochemistry* 2008, 47, 348-57.

379. Fukuda, H.; Arai, M.; Kuwajima, K., *Biochemistry* 2000, 39, 12025-32.

380. Tsumoto, K.; Umetsu, M.; Kumagai, I.; Ejima, D.; Arakawa, T., *Biochem. Biophys. Res. Comm.* 2003, 312, 1383-86.

381. Bradford, M. M., *Anal. Biochem.* 1976, 72, 248-54.

382. Weisiger, J. R., *J. Biol. Chem.* 1950, 186, 591-602.

383. Meek, J. S.; Rowe, J. W., *J. Am. Chem. Soc.* 1955, 77, 6675-7.

384. Gu, X.; Ndungu, J. M.; Qiu, W.; Ying, J.; Carducci, M. D.; Wooden, H.; Hruby, V. J., *Tetrahedron* 2004, 60, 8233-43.

385. Keglevic, D.; Horvat, J.; Plavisc, F., *Carbohydr. Res.* 1976, 47, 49-61.

386. McNaught, A. D., *Carbohydr. Res.* 1996, 297, 1-92.

387. D. P. Gamblin, *D. Phil. Thesis, University of Oxford*, 2005.

388. Latham, H. G., Jr.; May, E. L.; Mosettig, E., *J. Org. Chem.* 1950, 15, 884-9.

389. Bredereck, H.; Hoschele, G., *Chem. Ber.* 1953, 86, 47-54.

390. Beyer, H.; Schultz, U., *Chem. Ber.* 1954, 87, 78-81.

391. Cerny, M.; Pacak, J., *Coll. Czech. Chem. Comm.* 1961, 26, 2084-6.

392. Ferrier, R. J.; Furneaux, R. H., *Carbohydr. Res.* 1977, 57, 73-83.

393. Berman, E.; Daman, M. E.; Dill, K., *Carbohydr. Res.* 1983, 116, 144-9.

394. Defaye, J.; Gadelle, A.; Pedersen, C., *Carbohydr. Res.* 1991, 217, 51-8.

395. Horton, D.; Wolfrom, M. L., *J. Org. Chem.* 1962, 27, 1794-800.

396. Wrede, F., *Hoppe-Seyler's Z. Physiol. Chem.* 1922, 119, 46-50.

397. Prey, V.; Grundschober, F., *Monatsh. Chem.* 1960, 91, 358-62.

398. Schlubach, H. H., *Ber. Dtsch. Chem. Ges.* 1926, 59B, 840-4.

399. Sakata, M.; Haga, M.; Tejima, S.; Akagi, M., *Chem. Pharm. Bull.* 1964, 12, 652-6.

400. Cohen, S. B.; Halcomb, R. L., *J. Am. Chem. Soc.* 2002, 124, 2534-43.

401. Leaback, D. H.; Walker, P. G., *J. Chem. Soc.* 1957, 4754-60.

402. Zhdanov, Y. A.; Levitan, G. E., *Russ. J. Gen. Chem.* 1976, 46, 937-8.

403. zu Reckendorf, W. M.; Bonner, W. A., *J. Org. Chem.* 1961, 26, 4596-9.

404. Fujihira, T.; Chida, M.; Kamijo, H.; Takido, T.; Seno, M., *J. Carbohydr. Chem.* 2002, 21, 287-92.

405. Cerny, M.; Stanek, J.; Pacak, J., *Monatsh. Chem.* 1963, 94, 290-4.

406. Micheel, F.; Micheel, H., *Ber. Dtsch. Chem. Ges.* 1930, 63B, 386-93.

407. Tejima, S.; Maki, T.; Akagi, M., *Chem. Pharm. Bull.* 1964, 12, 528-32.

408. Neuhoff, V.; Stamm, R.; Pardowitz, I.; Arold, N.; Ehrhardt, W.; Taube, D., *Electrophoresis* 1990, 11, 101-17.

409. Fishbein, W. N., *Anal. Biochem.* 1972, 46, 388-401.

410. Saitou, N.; Nei, M., *Mol. Biol. Evol.* 1987, 4, 406-25.

411. Schwarz, R.; Dayhoff, M., *Matrices for Detecting Distant Relationships.*, 1979; pp 353-58.

412. Tamura, K.; Dudley, J.; Nei, M.; Kumar, S., *Mol. Biol. Evol.* 2007, 24, 1596-99.

## Licence agreement

This is a License Agreement between Nicola Floyd ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by The American Association for the Advancement of Science, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

| | |
|---|---|
| License Number | 2278330656310 |
| License date | Sep 29, 2009 |
| Licensed content publisher | The American Association for the Advancement of Science |
| Licensed content publication | Science |
| Licensed content title | Antibody Domain Exchange Is an Immunological Solution to Carbohydrate Cluster Recognition |
| Licensed content author | Daniel A. Calarese, Christopher N. Scanlan, Michael B. Zwick, Songpon Deechongkit, Yusuke Mimura, Renate Kunert, Ping Zhu, Mark R. Wormald, Robyn L. Stanfield, Kenneth H. Roux, Jeffery W. Kelly, Pauline M. Rudd, Raymond A. Dwek, Hermann Katinger, Dennis R. Burton, Ian A. Wilson |
| Licensed content date | Jun 27, 2003 |
| Type of Use | Thesis |
| Requestor type | Other Individual |
| Format | Print |
| Portion | Text Excerpt |
| Number of pages requested | 1 |
| Order reference number | |
| Title of your thesis / dissertation | A Novel Glycoconjugation Method and Unexpected Photo-Induced Site-Selective Protein Cleavage Reaction |
| Expected completion date | Sep 2009 |
| Estimated size(pages) | 400 |
| Total | 0.00 GBP |
| Terms and Conditions | |