The Synthesis of Novel Homogeneous Glycoproteins

Derek Macmillan

The University of Edinburgh

A thesis submitted for the degree of Doctor of Philosophy  April 1999
DECLARATION

This thesis is submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Edinburgh. Unless otherwise stated the work described in this thesis is original and has not been submitted previously in whole or in part for any degree or other qualification at this, or any other university. In accordance with the regulations this thesis does not exceed 70,000 words in length.

Derek Macmillan.
ABSTRACT

A glycoprotein can exist as a spectrum of glycosylated forms, where each protein molecule may be associated with an array of oligosaccharide structures. The overall range of glycoforms can have a variety of different biophysical and biochemical properties. As a result of this 'microheterogeneity' there is a need for the synthesis of homogeneously glycosylated proteins for analytical purposes. The synthesis of novel glycoproteins through the selective reaction of glycosyl iodoacetamides with the thiol groups of cysteine has been developed (Figure 1).

![Figure 1](image)

Through site-directed mutagenesis, whereby the natural asparagine glycosylation sites can be exchanged for cysteine, it was possible to selectively glycosylate proteins at predetermined sites. This provided a general method for the synthesis of homogeneously glycosylated proteins. We chose the glycoprotein hormone erythropoietin as a model system since the $N$-glycans at residues 24, 38 and 83 are essential for \textit{in vivo} biological activity.

Using a modified recombinant erythropoietin gene (prepared in Chapter 2) we optimized protein expression, and over-expressed and purified His$_{10}$-WThEPO, His$_{10}$-Asn$^{24}$Cys, His$_{10}$-Asn$^{38}$Cys, His$_{10}$-Asn$^{83}$Cys and His$_{10}$-Asn$^{38/83}$CysEPO's in yields of 13mgL$^{-1}$ from \textit{E. coli}. This allowed us to probe the structure of rhEPO using techniques such as NMR spectroscopy and electrospray MS (Chapter 3). Having access to larger quantities of EPO, we were also able to develop on-line LC-ESI-MS methods, which allowed us to monitor protein glycosylation reactions with glycosyl $\beta$-$N$-iodoacetamides and map the position of the glycosylation site.

The synthesis of relevant iodoacetamides was attempted and simple glycosyl iodoacetamides and $^{13}$C labeled iodoacetamides were prepared and successfully employed as probes for protein glycosylation.
I would like to take this opportunity to thank my supervisor, Professor Sabine Flitsch, for her support and encouragement throughout my PhD studies. I would also like to thank Dr. Dominic Campopiano for his patience, help and advice with molecular biological techniques.

Doing such a multidisciplinary project, much outside ‘collaborative’ help was required and I would like to acknowledge the support of Mr Brian Wigham (MALDI-TOF spectra), Dr. Emma Beattie and Dr. Julia Richardson (protein NMR), Dr. Adam Gouldsworthy (Q-TOF MS), and Lisa McIver (molecular and cell biology).

Proof reading of this thesis was kindly undertaken by Dr Philip Lowden, Prof. Sabine Flitsch and Krista.

I am grateful for the continued support of the members of Labs 34, 120 and 229.

Finally, I would like to thank those without whom I would never have survived university: Dave, Martin, Jase, Davey Aikman, Sean and Caroline. I would especially like to thank Krista for her complete lack of tolerance on all matters scientific.
EXPLANATORY NOTES

1. Carbohydrate structures in this thesis are numbered according to the convention:

![Carbohydrate Structure]

2. Erythropoietin is the protein under study in this thesis. The wild type protein comprises 166 amino acids (molecular weight=18389). Our expressed His_{10}-tagged species have an additional 21 amino acids fused to the N-terminus of wild type erythropoietin (molecular weight=20917):

$$\text{GHHHHHHHHHHSSGHIEGRHM-}EPO$$

To maintain the native human erythropoietin residue numbering, unless otherwise stated, the polyhistidine tag residues are assigned negative numbers (-21G to -1M).
TABLE OF CONTENTS:

<table>
<thead>
<tr>
<th>Section</th>
<th>Page no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declaration</td>
<td>i</td>
</tr>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Explanatory Note</td>
<td>iv</td>
</tr>
<tr>
<td><strong>CHAPTER 1</strong></td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>General Principles of Protein Glycosylation.</td>
<td>2</td>
</tr>
<tr>
<td>1.1 Biosynthesis of N-linked Glycoproteins.</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Biosynthesis of O-linked Glycoproteins.</td>
<td>5</td>
</tr>
<tr>
<td>1.3 The Biological Roles of Oligosaccharides in Glycoconjugates.</td>
<td>6</td>
</tr>
<tr>
<td>1.4 The Study of the Role of the Oligosaccharides: Glycoproteins and Neoglycoproteins.</td>
<td>10</td>
</tr>
<tr>
<td>1.5 Strategies for the Synthesis and Study of Glycoproteins.</td>
<td>11</td>
</tr>
<tr>
<td>1.5.1 Mutagenic Deletion of the Glycosylation Site.</td>
<td></td>
</tr>
<tr>
<td>1.5.2 Expression of Genes in Carefully Chosen Cell Lines.</td>
<td></td>
</tr>
<tr>
<td>1.5.3 Enzymatic Synthesis.</td>
<td></td>
</tr>
<tr>
<td>1.5.4 In Vitro Translation.</td>
<td></td>
</tr>
<tr>
<td>1.5.5 Chemical Synthesis.</td>
<td></td>
</tr>
<tr>
<td>1.6 Erythropoietin: A Model Protein for Glycosylation Studies.</td>
<td>23</td>
</tr>
<tr>
<td>1.7 Project Aims</td>
<td>31</td>
</tr>
<tr>
<td><strong>CHAPTER 2</strong></td>
<td>33</td>
</tr>
<tr>
<td>Optimisation of a Prokaryotic Expression System for the Expression of Recombinant Human Erythropoietin (rhEPO).</td>
<td></td>
</tr>
<tr>
<td>2.1 The pET System.</td>
<td>35</td>
</tr>
<tr>
<td>2.2 Expression Studies Using pET16b-EPO Constructs.</td>
<td>36</td>
</tr>
<tr>
<td>2.3 Optimisation of Protein Expression.</td>
<td>43</td>
</tr>
<tr>
<td>2.4 NMR Analysis of His$_{10}$-WThEPO.</td>
<td>49</td>
</tr>
<tr>
<td><strong>CHAPTER 3</strong></td>
<td>53</td>
</tr>
<tr>
<td>Electrospray Ionisation Mass Spectrometry (ESI-MS)</td>
<td></td>
</tr>
<tr>
<td>3.1 Off-Line LC-MS Analysis of His$_{10}$-EPO Samples.</td>
<td>56</td>
</tr>
<tr>
<td>3.2 Typical Resolved Mass Spectra.</td>
<td>60</td>
</tr>
<tr>
<td>3.3 Optimisation of Mass Spectrometry Analysis Using On-Line LC-MS.</td>
<td>64</td>
</tr>
</tbody>
</table>
CHAPTER 4

4.1 Oligosaccharide Synthesis. 68
4.2 β-1-4 Mannoside Linkages. 69
4.3 N-Acetyl Groups as Neighbouring Groups for Participation in β-Glycoside Synthesis. 70
4.4 Synthesis of More Complex Naturally Occurring Glycans. 71
4.5 Glycosyl Amines. 71
4.6 Strategic Problems. 73
4.7 Synthesis of β-N-Glycosyl Iodoacetamides. 74
4.8 More Complex Oligosaccharides. 80

CHAPTER 5
Protein Glycosylation and Determination of the Glycosylation Site.

5.1 Identification of the Glycosylation Site: Peptide Mapping. 88
5.2 Enzymatic Methods for Peptide Fragmentation. 88
5.3 Chemical Methods for Peptide Mapping. 89
5.4 Glycosylation of His10-EPO Samples. 91
5.4.1 Analysis of Glycoprotein Products. 91
5.4.2 The Glycosylation Reaction. 91
5.4.3 NTCBA Analysis of Glycoproteins. 93
5.4.4 1H-13C HSQC NMR Spectrum of Partially Glycosylated His10-C38hEPO. 95
5.4.5 Proteolytic Digests of Glycoprotein. 97
5.4.6 Analysing the Digestion Reaction. 98
5.4.7 On-Line LCMS Analysis of on Column Proteolytic Digests. 102
5.4.8 Q-TOF Analysis of Tryptic Digest of Chymotryptic (8KDa) Fragments. 106
5.5 Optimisation of the Glycosylation Reaction. 108

CHAPTER 6
Preparation of the pET16b-C83hEPO(COPT) Vector, Expression and Purification of His10-C83hEPO.

6.1 Preparation of pET16b-C83EPO Vector Using the Megaprimer PCR Method. 114
6.2 Expression and Purification of His10-C83hEPO. 118

General Conclusion. 119
CHAPTER 7
Experimental Details.

6.1 Experimental Details For Chapter 2. 123
6.2 Experimental Details For Chapter 3. 136
6.3 Experimental Details For Chapter 4. 139
6.4 Experimental Details For Chapter 5. 161
6.5 Experimental Details For Chapter 6. 164

References. 167
Abbreviations. 182
CHAPTER 1.

INTRODUCTION,

Principles of Protein Glycosylation and The Role of the Bound Oligosaccharides.
The importance of oligosaccharides is now well appreciated. Cell growth and differentiation, viral and bacterial infection, leukocyte adhesion to inflamed endothelium and tumour cell metastasis are a few examples of the vast array of processes governed by carbohydrate-receptor interactions. It was the ability to accurately sequence the oligosaccharide moieties of glycoproteins that revealed their complexity and diversity in structure. However, despite the large number of glycoproteins known, the types of covalent bond between the protein and carbohydrate show relatively little variation. The two variants which are most prevalent throughout eukaryotes are N-linked, characterised by a β-N-glycosidic linkage between N-acetyl glucosamine and the amide group of certain asparagine residues in the sequence Asn-X-Ser/Thr, and O-linked, characterised by a proximal α-O-glycosidic linkage between N-acetyl galactosamine and the hydroxyl groups of either serine or threonine (Figure 1.1).

**Figure 1.1:** (1) β-N-glycosidic linkage to asparagine in N-linked glycoproteins, (2) and (3) α-O-glycosidic linkages to serine and threonine respectively. R= oligosaccharide chain.
1.1 Biosynthesis of N-Linked Glycoproteins.

N-linked glycoproteins are constructed on the luminal side of the endoplasmic reticulum (Figure 1.2) where a single multimeric enzyme, oligosaccharyl transferase (OST), catalyses the co-translational transfer of a common lipid linked tetradecasaccharide (GlcNAc$_2$Man$_9$Glc$_3$) to an asparagine residue on a nascent polypeptide. How OST does this has been, and still is, a matter of intense debate$^{4,5}$. The consensus sequence: Asn-X-Ser/Thr (where X is any amino acid except proline) required for recognition by the enzyme has been extensively studied in vitro in attempts to establish the specificity of the enzyme. Studies on the model peptide (4) has shown that modification of the asparagine residue itself is almost invariably not tolerated.

![Figure 1.2: Biosynthesis of N-linked glycans.](image)

Glutamine and N-methyl asparagine in place of asparagine are not substrates for OST but thioasparagine (carbonyl oxygen is substituted with sulfur) is. The third amino acid (Ser/Thr) is as invariable and only substitution with cysteine shows low, but measurable, levels of glycosyl acceptor activity. Conformational requirements are also important as the peptide to be glycosylated is believed to form an Asx turn (Figure 1.3) which allows the hydroxyl group of serine or threonine to activate the generally inert amido group of asparagine.
As not all consensus sequences are glycosylated, occupation of the sites may also be influenced by the manner of folding during biosynthesis i.e.: the accessibility of potential sites to OST. After glycosylation, the glycoprotein is translocated to the Golgi apparatus where it is subjected to the host cell's complement of glycosidases and glycosyltransferases. This is where the resulting heterogeneity occurs and is a result of the differing rate constants of particular glycosyltransferases and glycosidases and overall time spent in each post-translational compartment. The major subgroups of mammalian oligosaccharides from N-linked glycoproteins are shown below:

![Figure 1.3: Asparagine amide group activation by the hydroxyl group of serine or threonine](image)

![Figure 1.4: The major subgroups of mammalian N-linked oligosaccharides. A. complex type, B. High mannose type and C. Hybrid type. The structures within the boxes are always conserved, outside the boxes, the structures are variable.](image)
Although all members of these classes share a characteristic core pentasaccharide adjacent to the peptide, the outlying glycoforms vary widely and account for glycoproteins enormous diversity in structure and variation in biophysical and biochemical properties.

1.2 Biosynthesis of O-linked Glycoproteins.

The synthesis of O-linked oligosaccharides occurs entirely by post-translational, sequential addition of nucleotide linked monosaccharides to the protein\textsuperscript{8,9}. Unlike in N-glycoproteins, there is no clear consensus sequence and protein domains likely to be glycosylated are generally rich in serine and threonine but, as with N-glycans, the presence of the glycosylation site is not sufficient for glycosylation and conformational restraints may inhibit glycosylation at specific sites.

The glycans of O-linked glycoproteins comprise five main core structures\textsuperscript{10} and, again, structural diversity comes from the variation in terminal sugars which are generally α-linked N-acetyleneuraminic acid (sialic acid), fucose, galactose, N-acetyl glucosamine and N-acetylgalactosamine.

\[
\text{Sia}(\alpha 2-3)\text{GalNAc} \quad \text{Ser} / \text{Thr}
\]

\[
\begin{align*}
\text{GalNAc}(\alpha 1-3) \quad & \quad \text{Gal}(\beta 1-4)\text{GlcNAc}(\beta 1-3)\text{Gal}(\beta 1-4)\text{GlcNAc}(\beta 1-3)\text{Gal}(\beta 1-3) \\
\text{Fuc}(\alpha 1-2) \quad \text{Sia}(\alpha 2-6) \quad \text{GalNAc} \quad \text{Ser} / \text{Thr} \\
\end{align*}
\]

\[
\text{Fuc}(\alpha 1-2)\text{Gal}(\beta 1-3)\text{GlcNAc}(\beta 1-3)\text{GalNAc} \quad \text{Ser} / \text{Thr}
\]

\[
\begin{align*}
\text{Gal}(\beta 1-4)\text{GlcNAc}(\beta 1-3) \quad & \quad \text{Sia}(\alpha 2-3)\text{Gal}(\beta 1-4) \\
\text{Fuc}(\alpha 1-3) \quad \text{GlcNAc}(\beta 1-6) \quad \text{GalNAc} \quad \text{Ser} / \text{Thr} \\
\end{align*}
\]

Figure 1.5: Core structures of mammalian O-linked glycans\textsuperscript{10}.

The structures of O-linked glycoproteins are diverse and even within a single protein can exhibit considerable microheterogeneity.
1.3 The Biological Roles of Oligosaccharides in Glycoconjugates.

In recent years the biological role of the bound oligosaccharide has been comprehensively reviewed\textsuperscript{1,2}. Despite some observations suggesting important biological roles for the oligosaccharide moiety, a single common theory has not, to the present day, emerged to explain the diversity in oligosaccharide structure, as the importance of the bound oligosaccharide can vary from seemingly trivial to vital for the development, growth, function and survival of the organism. Below are but a few examples of the effects that the presence of sugars can have on glycoproteins\textsuperscript{1}.

1.3.1 Structural, Protective and Stabilising.

As previously mentioned, N-glycosylation relies on the presence of an Asx turn motif (5) for recognition by OST\textsuperscript{4}. It is believed that glycosylation of this motif perturbs the conformation towards a β-turn (6) and the bias towards a β-turn is greatest when the native carbohydrate, chitobiose, is in place\textsuperscript{11}.

![Figure 1.6: Glycosylation of the Asx turn causes a conformational change which results in the formation of a β-turn.](image)

In fact 30% of all glycosylation sites result in sites that fold ultimately into β-turns and it has been suggested that the N-acetyl group of the proximal GlcNAc residue plays a key role sprouting from a steric interaction which hinders rotation about the glycosidic bond. Furthermore, a slightly truncated form of the initial tetradecasaccharide is a ligand for calnexin, a key chaperone protein in the endoplasmic reticulum and may therefore play a direct role in aiding the folding process\textsuperscript{12}. This helps to explain observations that many proteins which have been engineered to lack N-glycans fail to be secreted from the endoplasmic reticulum or
perform their biological functions \textit{in vivo}. If unfolded, misfolded, or aggregated they would most likely be targeted for destruction by intracellular proteases.

\textit{O}-linked glycosylation, though post-translational, has also been shown to have structural implications for proteins. \textit{O}-linked glycosylation reduces the tendency for peptides to form \(\alpha\)-helices and can hinder supercoiling of \(\alpha\)-helices\textsuperscript{13}. A less common form of glycosylation, but one which is gaining great interest is the \textit{O}-linked \(\beta\)-GlcNAc to serine or threonine and this has the tendency to change protein conformations by inducing a turn structure and may play a signalling role, not unsimilar to protein phosphorylation\textsuperscript{14}.

The increase in steric bulk or surface charge imparted by large glycans through a coating of oligosaccharides on glycoproteins may also serve to protect the protein from recognition by proteases or antibodies and prevent aggregation thus increasing circulatory half life\textsuperscript{15,16}.

1.3.2 Receptor and Masking Functions.

This is a highly specific, usually polyvalent, interaction dependent on the identity of the terminal monosaccharide and the stereochemistry about its anomeric linkage to the underlying sugar chain. This type of interaction has been extensively reviewed\textsuperscript{17} and some important processes dictated by it include:

\textit{Bacterial and viral infection}: Bacterial and viral surface lectins, adhesins and hemagglutinins mainly bind to the host cells sialic acids on cell surface glycoproteins and glycolipids. Infection may also be initiated by bacterial or viral lectins interacting with heparin, heparan sulfate, galactose, fucose or \(N\)-acetyl galactosamine in the extracellular matrix or presented on the cell surface\textsuperscript{17}.

\textit{Recognition of antibody coated antigen by macrophages}: Macrophages bind mannose residues on the Fc section of antibodies, but only act polyvalently. Only when antigens are coated with antibodies and many Fc fragments interact with
receptors on the surface is the macrophage activated and an immune response induced\textsuperscript{17,18}.

\textbf{Hyperacute rejection in xenographic transplantation:}

The enzyme $\alpha$-1,3-galactosyltransferase is active in the Golgi apparatus of non-primate mammals and new world monkeys and transfers galactose from UDP-Gal to $N$-acetyl lactosamine (Gal$\beta$1-4GlcNAc-R) on glycoproteins and glycolipids to form the $\alpha$-Gal epitope\textsuperscript{19}. The enzyme has become inactivated in humans and old world monkeys and, in fact, 1\% of all circulating human IgG is raised against the $\alpha$-Gal epitope. The $\alpha$-Gal epitope has become the major immunological obstacle for transplantation of non-primate mammalian organs into human recipients as anti-Gal induces attack by blood monocytes and macrophages and fixes complement, inducing complement mediated lysis of xenograft cells.

\textbf{1.3.3 Modulators of Biological Activity.}

The presence of bound oligosaccharides have a tendency to ‘tune’ the biological activity of the adjacent protein. This effect is most pronounced in glycosylated growth factors where some cell surface receptors for such factors acquire their binding function in a glycosylation dependant manner. In the case of the cytokine, granulocyte macrophage colony stimulating factor (GM-CSF), the fully deglycosylated form is known to be the most active. However, natural GM-CSF, which exists as a range of glycoforms exhibits substantial differences in binding affinity and signal transduction\textsuperscript{1,20}. Many unglycosylated growth factors are expressed from multi-copy genes which are clustered together on chromosomes. It has been argued that this process allows for maximum diversification, providing a spectrum of eg. $\alpha$-interferons which have slightly different structural features which may be able to interact with different targets, leading to different effects. Glycoform variants of single-copy genes may also act through providing charge and size heterogeneity therefore not confining diversity to the protein structure.
1.3.4 Regulatory Functions.

This is an extension of two previously discussed roles for oligosaccharides bound to proteins\textsuperscript{1,17,20}. Many proteins, particularly cytokines, perform their biological functions in a glycosylation dependant manner and have a finite circulatory half-life whether or not they have interacted with their target receptor. 'Aged' glycoproteins are then targeted for clearance from the body by hepatic cells in the liver.

Cytokines are usually highly sialylated on the oligosaccharides non-reducing ends. The presence of branched (bi-tri and tetra antennary) oligosaccharides display many sialic acid residues on the proteins surface. As highly sialylated molecules and cells (e.g. erythrocytes) circulate, they are exposed to sialidases on the surface of endothelial cells and in the blood which trim back the sialic acid and expose terminal galactose residues. The presence of asialoglycoprotein receptors recognise the processed glycoproteins with exposed terminal \( \beta \)-galactoside residues in mammalian plasma proteins. The result is that the glycoprotein can be cleared from the circulation and the oligosaccharides are serving as a timing device which regulates the half-life of proteins in the plasma. They also do this by increasing the overall size of circulating proteins and consequently reduce the rate of glomerular filtration.

1.3.5 Cell-Cell Recognition.

Since cell surfaces are densely coated with sugars it has been predicted that oligosaccharides must be the determinants of cell-cell interactions. We have already met this type or interaction above. 'Aged' erythrocytes become desialylated exposing terminal galactose residues, signalling their demise as they are recognised by receptors on hepatic cells which remove them from circulation. The best characterised cell-cell interactions are those which occur during the inflammatory response\textsuperscript{17}. When tissue is damaged by injury, cytokines are expressed locally which initiate the inflammatory response by inducing the expression of E- and P-selectin (carbohydrate binding proteins) on the surface of endothelial cells\textsuperscript{21}. Neutrophils (the major class of white blood cells) are then attracted to the site of inflammation by a mechanism termed rolling, which describes the sialyl Lewis X (7)(SLe\textsuperscript{X}; expressed on the surface of neutrophils) mediated migration of neutrophils to sites of
inflammation. Additionally, L-Selectin expressed on the surface of neutrophils binds SLe\(^X\) on the surface of endothelial cells. Further, shear resistant, adhesion is also mediated by integrins which are activated by further glycosylation dependant cell adhesion molecules (CAM’s) such as MAdCAM-1 and GlyCAM-1 prior to extravasion\(^{22}\).

1.4 The Study of the Role of the Oligosaccharides: Glycoproteins and Neoglycoproteins.

It has never been more truly said than for the role of the bound oligosaccharides on glycoproteins that for every rule there is an exception. Common features of the above that have fuelled interest into the role of the bound oligosaccharide are that they can mediate specific recognition events or modulate biological processes. Consequently they provide further functional diversity which is essential for the development and survival of cells, tissues, organs and species.

The study of glycoproteins is complicated by the fact that they exist in various glycosylated forms, or ‘glycoforms’, in which each glycosylation site may be associated with one of several possible oligosaccharide sequences. Therefore, a glycoprotein may be a heterogeneous mixture of protein molecules which vary only in the structure of the oligosaccharide. As a result of this microheterogeneity the overall range of glycoforms may have a range of different physical, biochemical and biological properties and their presence may constitute a subtle and sophisticated mechanism of biological control. The importance of glycoforms has been recognised in the case of tissue plasminogen activator, a recombinant fibrinolytic agent in which the glycoforms show differing fibrin-dependant plasminogenolytic activities, and thus the United States patent office has ruled that glycoforms must be specified in patents. As a result of the difficulties in characterising natural glycoproteins there has been increasing interest in the synthesis of homogeneously glycosylated proteins, ie. glycoproteins containing oligosaccharide sidechains of defined structure, for
analytical purposes. The ability to prepare single glycoforms of proteins allows one to probe the structure-activity relationships (SAR's) of glycoproteins, elucidate carbohydrate-lectin interactions, define fine structural requirements and possibly optimise their biological activity\textsuperscript{23,24}.

So, how can the synthesis of homogeneously glycosylated proteins be achieved? Proteins are macromolecules composed of many residues with nucleophilic and electrophilic functional groups therefore the problem is mainly one of obtaining regio and stereoselective glycosylation.

1.5 Strategies for the Synthesis and Study of Glycoproteins.

- 1.5.1 Mutagenic Deletion of the Glycosylation Site.
  A crude method for the study of the biological role of the oligosaccharide moiety of glycoproteins is to isolate the gene which encodes for the protein and delete the glycosylation sites (either completely or sequentially) by recombinant DNA technology and compare the biological activity of the mutant to that of the glycosylated wild type\textsuperscript{25}. This is usually achieved by mutating the asparagine residue (for N-linked glycans) to glutamine or mutating the serine or threonine in the Asn-X-Ser/Thr consensus sequence (or in the case of O-linked glycans) to alanine or glycine. The difficulty in analysing products from this type of approach is establishing whether changes in biological activity or secretion results from a lack of glycosylation, or changes to the protein conformation stemming from the backbone mutation.

- 1.5.2 Expression of Genes in Carefully Chosen Cell Lines.
  Progress has been made towards the use of \textit{in vivo} techniques such as glycosyl engineering. This can exploit the differences in metabolism and glycosylation pathways in different organisms and perturb known pathways using modern methods in genetics such as gene knock out. The use of the baculovirus-insect expression system which has a restricted glycosylation pathway (cannot synthesize complex type \textit{N}-glycans) has been useful in this respect\textsuperscript{26}. Co-expression of specific glycosyl
transferases can effect the synthesis of specific glycoforms. However, the structures of the bound oligosaccharides still tend to be heterogeneous in nature and many glycoforms result.

ldlD is a Chinese hamster ovary (CHO) cell mutant, which is deficient in the 4-epimerase enzyme required to convert glucose into UDP-Galactose and UDP N-acetyl galactosamine. When glucose is provided as the only sugar source in the growth medium, the organism is unable to synthesize galactose and N-acetyl galactosamine and this has an enormous effect on the glycosylation of the expressed protein (Figure 1.7).

The defect can be reversed by the addition of galactose or N-acetyl galactosamine to the growth medium, providing a versatile method for the expression of proteins with defined, truncated glycans, which may then be extended by enzymatic means.

1.5.3 Enzymatic Synthesis.
The ability to construct a homogeneous glycoprotein from the ‘naked’, deglycosylated protein (e.g., expressed in Escherichia coli) is a most attractive approach. The advantages are that glycosylation can be achieved in vitro using glycosyltransferases in a regio and stereoselective manner using unprotected carbohydrates under mild, aqueous conditions.

![Figure 1.7: Representative N- and O-linked oligosaccharides that are obtained as a result of gene expression of recombinant human erythropoietin (EPO) in ldlD cells in the presence or absence of Gal or GalNAc.](image-url)
1.5.3.1 Enzymatic Synthesis of N-linked Glycoproteins.

The problems associated with enzymatic N-linked glycans/glycoproteins are obvious. They are constructed on the lipid bilayer of the ER linked to dolichol phosphate. The substrate for OST is Dol-PP-GlcNAc2Man9Glc3. And the lipid tether is essential for the enzymatic transfer of sugar to protein. The result is that substrates for OST and the glycosyltransferases involved in the tetradecasaccharide assembly are incredibly complex and not simple nucleotide sugar donors of O-linked glycosylation. The tetradecasaccharide is also constructed by membrane bound glycosyltransferases and consequently it has been difficult to isolate, purify and characterise these proteins and few are known. The most common methods for enzymatic synthesis of N-linked homogeneous glycoproteins begin with a mature heterogeneously glycosylated protein

Endo-glycosidases are then employed to cleave the majority of the oligosaccharides leaving only homogeneous regions of the conserved core structure (Figure 1.8).

![Figure 1.8: Wong and co-workers synthesised specific glycoforms of RNAse B by first removing the N-linked glycan with an endoglycanase (Endo H). Sialyl Lewis X was then built up on the truncated glycoprotein.](image)

Specific oligosaccharides can then be built up using selected glycosyltransferases and nucleotide sugars. This method fails however to deal with heterogeneity associated with glycosylation site occupancy. But glycoproteins which differ to such large extents can usually be separated by capillary electrophoresis (CE) or size exclusion chromatography.

1.5.3.2 Enzymatic Synthesis of O-linked Glycoproteins.

As O-linked glycosylation is a post translational process where O-glycosides are formed from sugar-nucleotides this is more straightforward than for N-linked glycoproteins. Until recently, a major limitation was the expense of the sugar
nucleotides but ingenious methods have been developed for their regeneration\textsuperscript{32}. For a fully enzymatic synthesis however, proteins are usually required with the first one or two sugars in place. Fortunately methods for bringing this about are known: trimming back mature glycoprotein glycans with exoglycosidases, expression of glycoproteins in ldlD cells as described above, using inhibitors of $O$-linked glycosylation (8) and partial hydrolysis of heterogeneous $O$-linked glycans (Figure 1.9)\textsuperscript{33,34}:

\begin{figure}
\centering
\includegraphics[width=\textwidth]{Figure_1.9.png}
\caption{Homogeneously glycosylated proteins can be obtained after expression of glycoproteins in the presence of $O$-glycosylation inhibitors or by cleaving the heterogeneous portion of the glycan.}
\end{figure}

- \textbf{1.5.4 In Vitro Translation.}

This method for the synthesis of glycoproteins was developed by Hecht and co-workers. Although this method requires mutated mRNA, tRNA which is recognised sufficiently by the proteins biosynthetic machinery so as to suppress a stop codon and is acylated with the desired amino acid derivative, they have demonstrated the incorporation of several unnatural amino acids and glycoamino acids into proteins at the desired sites\textsuperscript{35}. This methodology has escaped attention due to the complexity of starting materials and the current inability to scale reactions up.
• 1.5.5 Chemical Synthesis.
Chemical synthesis is potentially the most versatile strategy for the synthesis of homogeneous glycoproteins. It allows one to produce both natural and unnatural linkages between the carbohydrate moiety and the protein and explore the effects of altering the oligosaccharide structure. Chemical syntheses of homogeneous glycoproteins generally fall into one of two categories.

1.5.5.1 Total Synthesis of Homogeneous Glycoproteins.
The total synthesis of large glycoproteins is a very ambitious goal. In contrast to peptide synthesis, glycopeptide synthesis requires the protection of many more functional groups, reversibly and with maximum selectivity. Furthermore, the glycosidic linkage must be formed stereoselectively and prevented from anomerising throughout the synthesis. For peptide synthesis, high efficiency must be achieved in each of the coupling steps and in solid phase methodology the addition of excess reagent can drive reactions to near quantitative. Advances in carbohydrate chemistry$^{36-43}$ and protecting group strategies$^{44-46}$ has made the synthesis of large oligosaccharide structures possible. Despite this, the synthesis of small glycopeptides (<20 amino acids) still poses a significant challenge but has been met by a handful of groups$^{47-53}$. Developments in solid phase oligosaccharide and glycopeptide chemistry$^{47,54}$ has also improved the efficiency and purity in which glycopeptides can be prepared$^{55-59}$. Although the convergent coupling of tailor made glycans to a peptide scaffold is an appealing strategy, it has only been successful in the construction of N-linked glycopeptides (Figure 1.10). Extension of this to O-linked glycoproteins has been hindered by problems endemic to the formation of sugar-peptide glycosidic bond. O-linked glycans are then usually incorporated as glycoamino acids during sequential solid phase peptide synthesis (SPPS).
Figure 1.10: Natural and unnatural glycopeptide structures which have been synthesised to be incorporated into synthetic peptides. All, except the glycosyl amine, are incorporated into sequential SPPS as α-N-Fmoc protected glycoamino acids.

The molecular weight restrictions of SPPS are a significant issue, only a few groups have synthesised peptides over 100 amino acids in length and SPPS tends to become inefficient for peptides over 50 amino acids in length\textsuperscript{64,65}. Most naturally occurring glycoproteins are substantially larger and the total synthesis of a full-length glycoprotein is yet to be reported.

The most likely solution to this problem will be through convergent methods of protein synthesis such as native ligation or expressed protein ligation\textsuperscript{66-69}. This will ultimately allow the convergent coupling of smaller and more attainable synthetic glycopeptides to synthetic or recombinant peptide fragments.
Small, synthetic glycopeptides are not however without their uses and much of our understanding of the structural consequences of protein glycosylation has come from studies with small synthetic glycopeptide fragments\textsuperscript{11,13,14}.

1.5.5.2 Chemoenzymatic Synthesis of Homogeneous Glycoproteins.

Recombinant DNA technology has allowed for the isolation of ‘naked’, deglycosylated proteins. Expression of the gene encoding for the glycoprotein of interest in non-glycosylating organisms such as \textit{E. coli} (which lack the biological machinery for glycosylation) allows one to selectively glycosylate proteins at specific amino acid residues with activated oligosaccharides. This facilitates the synthesis of neoglycoproteins which are novel glycoproteins designed for analytical purposes. Much of the pioneering work in this field by Lee and co-workers indicated that a protein-carbohydrate linkage formed by coupling activated oligosaccharides to proteins should be specific, covalent and stable under physiological conditions. They should be formed under conditions which do not degrade the oligosaccharide, and should not contain bulky or highly interactive groups which may interfere non-specifically \textit{in vivo}. However, most methods currently available do not adhere to these specifications since they are very non-specific with respect to a single amino acid side chain. Consequently, they may:

- introduce bulky aromatics or long spacer arms which may hinder the necessary protein-carbohydrate interactions.
- produce antigenic, unnatural glycosidic linkages. For example glycosyl amines react with the carboxyl groups of both aspartic and glutamic acid thus \textit{N}-glycosides of glutamine are formed and these linkages are not present in nature.

Some classic and recent neoglycoprotein sugar-protein linkages are shown below:
### Sugar-Peptide linkage

<table>
<thead>
<tr>
<th>Sugar-Peptide linkage</th>
<th>Target amino acid</th>
<th>Reagent</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y, H, K</td>
<td></td>
<td>NHAc</td>
<td>70</td>
</tr>
<tr>
<td>Lys</td>
<td></td>
<td>HO</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td></td>
<td>HO</td>
<td>70</td>
</tr>
<tr>
<td>K</td>
<td></td>
<td>HO</td>
<td>70</td>
</tr>
<tr>
<td>K</td>
<td></td>
<td>HO</td>
<td>70</td>
</tr>
<tr>
<td>K</td>
<td></td>
<td>HO</td>
<td>70</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>NHAc</td>
<td>71,72</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>HO</td>
<td>74,75</td>
</tr>
<tr>
<td>K</td>
<td></td>
<td>HO</td>
<td>76</td>
</tr>
</tbody>
</table>

**Figure 1.11:** Activated carbohydrates which result in the formation of neoglycoproteins with unnatural sugar-protein linkages. These activated saccharides target specific amino acid residues and can be used to modify recombinant or synthetic peptides in the absence of protecting groups.
Some examples, which underscore the power and versatility of the semi-synthetic or chemoenzymatic approach to glycoprotein/ neoglycoprotein synthesis, are discussed below:

One of the most striking uses of neoglycoproteins has been in the investigation of their ability to raise the humoral immune response against tumors displaying the GloboH breast cancer antigen (9). The fully synthetic Globo H antigen was prepared by Danishefsky and co-workers as an allyl glycoside. Ozonolysis of the allylic double bond afforded an aldehyde which was reductively aminated by the ε-amino group of lysine residues on the carrier protein keyhole limpet hemocyanin (with 150 carbohydrate units per protein!).

High loading on the carrier protein allowed polyvalent display of the antigen as would be expected on the cell surface. Neoglycoproteins were used to immunize mice and isolated antibodies were shown to bind to the surface of human Globo H positive tumour cells. This vaccine is now undergoing clinical trials.

An elegant neoglycopeptide synthesis was also reported by Lee and co-workers who, in contrast to the previous example, used oligosaccharides from natural sources and a fully synthetic peptide. Using glycosyl cyanides, they prepared (10) and (11).
Finally, utilising endo-β-N-acetyl-glucosaminidase (Endo-A) and an oligosaccharide isolated from soybean agglutinin, they were able to prepare the C-glycopeptide (12) via transglycosylation.  

\[
\begin{align*}
\text{Man}_1-2\text{Man}_1-6 + \text{Man}_1-6 & \rightarrow \text{Man}_1-2\text{Man}_1-3 + \text{Man}_1-6 \\
\text{Man}_1-2\text{Man}_1-3 + \text{Man}_1-6 & \rightarrow \text{Man}_1-2\text{Man}_1-3 + \text{Man}_1-6 \\
\text{Man}_1-2\text{Man}_1-3 + \text{Man}_1-6 & \rightarrow \text{Man}_1-2\text{Man}_1-3 + \text{Man}_1-6
\end{align*}
\]

This demonstrates that it is possible to introduce synthetically complex high mannose type homogeneous glycans using glycosidases (in reverse) in high concentrations of organic solvents.

Convergent approaches will be the future of synthetic glycoprotein assembly. Convergent methods in protein synthesis are now well established and convergent methods of oligosaccharide elongation have also been investigated (Figure 1.12).  

![Figure 1.12](image)

**Figure 1.12:** Galactose oxidase oxidises the 6-position of Ga1NAc to the C6 aldehyde which can then be coupled to saccharides with hydroxylamine derivatives at the reducing end. The Ga1NAc-Ser/Thr linkage is introduced as a glycoamino acid, preserving the native sugar-protein linkage.
A novel approach for the synthesis of homogeneous glycoproteins was first reported by Davis and Flitsch in 1991\textsuperscript{70} and subsequently by Wong and co-workers\textsuperscript{71}. The use of glycosyl iodoacetamides as the activated carbohydrate moiety affords the selective derivatisation of the thiol groups of free cysteines at pH8 (where thiols are the strongest nucleophiles) (Figure 1.13).

\begin{center}
\includegraphics[width=\textwidth]{figure1_13.png}
\end{center}

Figure 1.13: Selective chemical glycosylation of the free thiol of reduced glutathione using glycosyl iodoacetamides.

The resulting thioether linkage is clearly very similar to the natural asparagine linkage (Figure 1.14). Indeed, the conformation of the bound monosaccharide in both linkages were shown to be identical by \textsuperscript{1}H NMR spectroscopy\textsuperscript{70}. 

\[ \text{Figure 1.14: Natural asparagine linkage.} \]
Figure 1.14: The novel thioether linkage (A) for homogeneous glycoprotein synthesis is identical to the natural sugar-asparagine linkage (B) except it is extended by a CH₂S unit.

Since this method of homogeneous glycoprotein synthesis is very selective, and in nature free cysteine residues are relatively rare, the advantages of this approach are obvious. By mutating the asparagine glycosylation sites to cysteine by site directed mutagenesis, the desired glycoprotein can be prepared providing the mutation does not greatly affect the tertiary structure of the protein. Furthermore, iodoacetamides tend not to react with the 'oxidised' cysteine present in disulfide bridges, expanding the scope of this method to proteins with no free cysteine residues assuming the correct disulfide bonds form during the refolding process. In contrast, most available glycosylation methods (Figure 1.11) afford derivatisation of lysine residues which are a far more common constituents of proteins.

In summary, an ideal homogeneous glycoprotein synthesis would involve the production of a mutated protein backbone (for selective chemical glycosylation), or construction of oligosaccharides on singly glycosylated proteins using glycosyl transferase enzymes. Unfortunately, many of the required enzymes have yet to be isolated, so selective chemical glycosylation (perhaps followed by transglycosylation) is a most attractive and challenging alternative.
Erythropoietin (EPO) is a glycoprotein hormone synthesised by the adult kidney and foetal liver. In adults, mature EPO is secreted as a heavily glycosylated protein where glycosylation comprises 40% of its molecular weight. On secretion, it migrates to the bone marrow where it acts specifically on erythroid progenitor cells, stimulating erythropoiesis through binding to the extracellular domain of the EPO receptor (EPOR), triggering intracellular signalling events including phosphorylation of the receptor and activation of the JAK-STAT, RAS and P13 kinase pathways. Present at concentrations in the picomolar range in human serum, EPO regulates the number of circulating erythrocytes and ultimately controls the transport of oxygen (carried by red blood cells (RBC's)) throughout the body. Serum EPO concentrations are controlled by a classic feedback mechanism whereby elevated serum oxygen level downregulate EPO expression and hypoxia results in elevated EPO expression.

Figure 1.15: On secretion from the kidney, EPO acts on a number of specific cells along the erythroid progenitor cascade: Initial interaction with pluripotential hematopoietic stem cells, burst forming unit-erythroid (BFU-E), and colony forming unit-erythroid (CFU-E) allows further differentiation (aided by further cytokines) into erythrocytes.
In addition to being one of the first recombinant therapeutics, recombinant human EPO (rhEPO) has also become one of the most successful and valuable recombinant therapeutics for the treatment of anaemia resulting from chronic renal failure, negating the need for regular blood transfusions for patients on dialysis. In such disease states the kidneys are no longer capable of synthesising EPO and hence anaemia results from reduced RBC production.

Since its approval in 1989, EPO's use has become more widespread and has been extended for the treatment of anaemia associated with cancer and AIDS. Of course, treatment with rhEPO does rely on the presence of functional erythroid progenitor cells and accordingly has little affect in patients where anaemia is associated with the absence or destruction of these cell types. The clinical use of EPO has sparked huge interest in the cytokine field, examining the roles and potential therapeutic uses of cytokines in virtually every organ system of the body. It is perhaps unsurprising that the majority of approved cytokines fall into the categories of immunological and hematopoietic regulators: interferon-γ for chronic granulomatous disease, Interferon-α-2-b for hepatitis and interferon-β-1-b and interferon-β-1-a for multiple sclerosis amongst others.

rhEPO, marketed by Amgen Inc. (California) as EPOGEN® or PROCRIT® contains Epoetin alfa which exists as a heterogeneous mixture of rigorously characterised glycoforms with as many as 58 different N-linked glycans associated with one sample. Indeed the glycoform composition of EPO has been known to differ between samples from plasma and urine, between samples obtained under different pathophysiological conditions, between samples of purified uhEPO and between samples of rhEPO synthesised in different cell lines. EPO's differing in glycoform composition have been shown to differ in immunological and biological properties.

Although deglycosylated versions of EPO fail to be biologically active in vivo, they retain biological activity in vitro demonstrating that the oligosaccharides, though not essential for receptor binding, perhaps affect the proteins stability, solubility, tissue
specific targeting and *in vivo* biological activity. This has prompted investigations into the roles of the oligosaccharide moiety.

The carbohydrate structures of EPO expressed in Chinese Hamster Ovary (CHO) cells have been thoroughly characterised (Figure 1.16) and have been found to be identical to those found in hEPO.

\[
\begin{align*}
\text{Gal(β1-4) GlcNAc(β1-2) Man(α1-6)} & \\
\text{Gal(β1-4) GlcNAc(β1-2) Man(α1-3)} & \\
\text{Gal(β1-4) GlcNAc(β1-6)} & \\
\text{Gal(β1-4) GlcNAc(β1-2)} & \\
\text{Gal(β1-4) GlcNAc(β1-2) Man(α1-6)} & \\
\text{Gal(β1-4) GlcNAc(β1-4) Man(α1-3)} & \\
\text{Gal(β1-4) GlcNAc(β1-6)} & \\
\text{Gal(β1-4) GlcNAc(β1-2) Man(α1-6)} & \\
\text{Gal(β1-4) GlcNAc(β1-2) Man(α1-3)} & \\
\text{Gal(β1-4) GlcNAc(β1-4) Man(α1-6)} & \\
\end{align*}
\]

\[
\begin{align*}
\text{Fuc(α1-6)} \\
\text{Gal(β1-4) GlcNAc(β1-4) GlcNAc —Asn} & \\
\text{Gal(β1-4) GlcNAc(β1-4) GlcNAc —Asn} & \\
\text{Gal(β1-4) GlcNAc(β1-4) GlcNAc —Asn} & \\
\text{Gal(β1-4) GlcNAc(β1-4) GlcNAc —Asn} & \\
\end{align*}
\]

**Figure 1.16**: Core structures of the *N*-linked glycans found on EPO derived from CHO cells.

Due to the heterogeneity of rhEPO it has (until very recently) evaded complete structural analysis by X-ray crystallography and multidimensional NMR spectroscopy. The major problems associated with structural assignment have been associated with the molecular properties of rhEPO: As a mature glycoprotein it is very heterogeneous and fails to crystallise, as a deglycosylated species it is unstable and only poorly soluble.
Troubles aside, many imaginative approaches have been used to investigate EPO's tertiary structure. The primary structure was solved in 1986 by Goldwasser and co-workers\(^8^7\) who purified urinary human EPO (uhEPO) from patients with aplastic anaemia. From this work they were able to establish that EPO had two disulfide bonds, between C7 and C161 (linking the C and N terminus), and between C29 and C33 as well as determine the protein sequence. Furthermore they identified that EPO was \(N\)-glycosylated at N24, N38 and N83 and \(O\)-glycosylated at S126. CD spectroscopy also intimated that the protein was highly \(\alpha\)-helical. Based on sequence analysis of purified uhEPO it was possible to probe for the human EPO gene; which was later identified, cloned, and over-expressed in CHO cells\(^8^8\). Although little was known about the 3D structure, EPO was presumed to have a 4-anti-parallel \(\alpha\)-helical bundle fold, as a result of weak sequence homology with other cytokines with solved structures such as growth hormone, the interleukins (2-7), G-CSF and GM-CSF. Additionally, this theory was founded on the fact that they all bound to the extracellular domains of receptors with similar structural motifs. This information was used to predict the tertiary structure\(^8^9\). One example is shown below\(^9^0\) (Figure 1.17).

![Figure 1.17: EPO (white) dimerises the extracellular domain of the EPO receptor (EPOR), glycosylation sites are shown in green. Model co-ordinates were kindly provided by Professor Graham Richards.](image-url)
The structure of EPO was also probed with monoclonal antibodies raised against specific EPO epitopes and they were evaluated for their ability to neutralise the activity of rhEPO. These results identified which residues were on the surface of the protein and which residues were involved in receptor binding providing further evidence for the 4-anti-parallel α-helical bundle fold\(^9\). At a later date the active site of EPO was mapped more thoroughly showing that the active site was discontinuous and spread over many residues throughout the protein and that interaction with the receptor may be dependant on electrostatic interactions between exposed basic side chains and the EPOR. Key residues for receptor binding were identified as 11-15, 44-51, 100-108 and 147-151\(^2\). Throughout this analysis one could be excused for believing that the peptide backbone alone constitutes the requirements for biological activity. Of course, this is not the case as EPO is completely inactive \textit{in vivo} in the absence of glycosylation and in cases where glycosylation is prohibited or inhibited the protein fails to undergo normal secretion.

1.6.1 The Role of Disulfide Bonds.

It has been established that the disulfide bonds in EPO help stabilise the structure, although they are not believed to be necessary for establishing the correct fold. Indeed it is the topology of the fold that brings the correct cysteine residues into

**Figure 1.18:** Model structure of hEPO, the disulfide bonds are shown in yellow. The glycosylation sites are also shown.
CHAPTER 1

contact. It may then be considered surprising that the alkylation of the thiol groups after reduction results in the irreversible loss of biological activity. Analysis of alkylated EPO with conformation sensitive antibodies has shown that disruption of the disulfide bonds causes a conformational change. These apparently contrasting observations may indicate that the fold is stabilised during synthesis (by chaperones) but is ultimately unstable and is readily lost. Disruption of the C29-C33 disulfide has less affect on the biological activity of EPO than disruption of the C7-C161 disulfide which results in the loss of biological activity.

1.6.2 Role of Sialylation.
The O- and N-linked glycans of hEPO and rhEPO are mainly bi- and tetra-antennary, capped with sialic acid at the non-reducing end. Removal of sialic acid residues using sialidases exposes terminal galactose residues, causing rapid clearance of EPO from the circulation. The requirement for sialylation is clearly to reduce the rate of clearance and desialylated EPO is biologically inactive in vivo, presumably for this reason as removal of sialic acid does not affect protein conformation and desialylated EPO is fully active in vitro.

1.6.3 Role of O-Linked Glycosylation.
Tetra antennary O-linked glycans predominate in rhEPO and uhEPO. When O-linked glycans were removed enzymatically or EPO was expressed in ldlD cells (lacking O-linked glycans) no difference in biological activity was observed relative to CHO derived EPO. Ser126Gly mutations had severe effects on protein secretion but this was considered more likely to have resulted from structure destabilisation caused by the mutation rather than the absence of O-linked glycans.

1.6.4 Role of N-Linked Glycans.
The N-linked glycans on EPO have been studied most intensively and for a long time were considered key in administering the biological activity of EPO. Numerous studies were conducted and indicated that the highly branched N-linked glycans served as anchors for sialic acid which increases circulatory half-life, allowing EPO to reach its target organs. It is difficult to argue that the N-glycans
play a crucial structural role as the solved NMR structure of *E. coli* derived MKlysEPO (where all N-glycosylation sites were exchanged for lysine) still folds into a 4-antiparallel helical bundle (Figure 1.19). The lysine residues were introduced to increase the solubility of *E. coli* derived EPO. It also became apparent that the receptor binding sites residues: 11-15 and 100-108, and 44-51 and 147-151 were brought together in the tertiary structure to form two biologically active 'patches' which give rise to EPO's bivalency and ability to dimerise it's receptor. Interestingly however, in the absence of *N*-linked glycans, normal EPO secretion is prevented in BHK cells. In fact, while N24 glycosylation is not necessary for normal secretion, the absence of N24 glycosylation actually increases the rate of secretion. In summary, these findings suggest that *N*-glycosylation of EPO may affect the biosynthesis and secretion of EPO in eukaryotic cells, affect tissue specific targeting or influence the ability of EPO to bind it's receptor. Glycans may also prevent EPO from interacting non-specifically with other cytokine receptors. The role of the glycans may become much clearer when the X-ray crystal structure of the co-complex between EPO and the extracellular domain of EPOR is disclosed in July 1999.

EPO is therefore an ideal protein for glycosylation studies with glycosyl iodoacetamides. It is a biologically relevant protein where the *N*-linked glycans at asparagine residues 24, 38 and 83 contribute significantly to its biological activity *in vitro* (where the full pentasaccharide core is required for full biological activity) and *in vivo*. The presence of two disulfide bonds give the methodology a complex setting in which to be evaluated: How will the introduction of a fifth, sixth or seventh cysteine residue affect the structure or the manageability of the protein. For
example, the presence of four cysteine residues allows three possible disulfide bonded structures:

All are not of equal energy and the most stable would be expected to predominate, this is presumably the native folded structure.
The general aim of this project was to develop the cysteine targeting iodoacetamidemethodology to effect the synthesis of homogeneos glycoforms of recombinant human erythropoietin. This involved the over expression, purification and characterisation of wild type (WT) and specific Asn→Cys mutant EPO's from E. coli. Refolding of our samples would then follow, hopefully exposing a thiol group from a single cysteine residue for glycosylation studies using glycosyl iodoacetamides. Once we had established that the correct sugar-protein linkage had been formed (Figure 1.11) we aimed to determine, unambiguously, that the correct site (cysteine residue) had been modified.

Previous work in the group resulted in the construction of many EPO containing vectors, encoding EPO genes with asparagine→cysteine mutations. Specifically, genes encoding N24C-hEPO, N38C-hEPO, and N38/83C-hEPO were prepared (N38/83C was prepared unintentionally: N38C DNA was used as the template DNA in the mutagenesis reaction for the synthesis of N83C-hEPO instead of WThcDNA). A number of expression systems were then evaluated for their ability to over-express WThEPO and cysteine mutants but all failed to provide sufficient quantities of EPO to study protein glycosylation using standard biophysical techniques such as NMR spectroscopy or mass spectrometry. However, the highest yields of EPO (1.0mgL⁻¹) were obtained using the pET expression system, where the proteins were expressed as N-terminal fusions with a decahistidine tag. Previously, EPO samples were expressed, identified by their approximate molecular weights (SDS-PAGE) and the fact that they were detected by a monoclonal antibody raised against a continuous N-terminal epitope. Glycosylation was then detected using horse radish peroxidase conjugated lectins that bind N-acetyl glucosamine in a western blot-type analysis. From this type of analysis however, a number of issues remained unresolved:

1) Was the reaction quantitative?
2) Were any other amino acid side-chains modified?
3) Was the correct cysteine residue modified?
So, the specific aims of this project were:

1) To optimise the expression of His$_{10}$-WT and cysteine mutant EPO’s from *E. coli* using the pET system, particularly, the pET16b constructs provided by predecessors. Failing this, to investigate further expression systems.

2) To refold and characterise, as fully as possible, the *E. coli* derived His$_{10}$-fused proteins.

3) Develop methods for the efficient synthesis of relevant or model glycosyl iodoacetamides.

4) Establish methods that will enable us to monitor and evaluate protein glycosylation.

5) Determine the identification of the glycosylation site in such glycosylated proteins.
CHAPTER 2.

Optimization of a Prokaryotic Expression System for Recombinant Human Erythropoietin (rhEPO).
The expression of erythropoietin in eukaryotic systems such as Chinese hamster ovary (CHO) and baby hamster kidney (BHK) cells has been well documented\textsuperscript{25,88,98,99}. For our purposes however, we required erythropoietin in a deglycosylated form so that we might specifically glycosylate it after expression and elucidate the roles played by specific oligosaccharides in administering \textit{in vivo} biological activity. This was achieved by expressing EPO in \textit{E. coli}, a prokaryotic organism which lacks the biosynthetic machinery for protein glycosylation.

Erythropoietin expression in \textit{E. coli} from cDNA has been achieved as both the native peptide sequence\textsuperscript{100} and as a fusion protein, whereby the gene of interest is inserted immediately downstream of a highly expressed gene encoding a cleavable fusion protein. The fused sequence (which may be on the N or C terminus of the target protein) may serve to protect the target protein from damaging proteases in the host organism, and aid solubility or purification of the protein of interest. To date, EPO has been expressed in \textit{E. coli} as a fusion protein with $\beta$-galactosidase\textsuperscript{101}, $\beta$-lactamase\textsuperscript{102}, glutathione S-transferase\textsuperscript{103}, thioredoxin\textit{A}\textsuperscript{104}, and a decahistidine tag\textsuperscript{89,104}.

Bill and Flitsch first reported on the expression of WThEPO and cysteine mutants as fusions with glutathione-S-transferase using the pGEX-2T expression system. Although purification of the target fusion protein was greatly simplified, the yield of fusion protein was low (Table 2.1) and the target protein was obtained from the fusion in poor yield after preparative SDS-PAGE and electro elution. Boisell and co-workers were then to report on the expression of EPO as a decahistidine tagged fusion protein using the commercially available \textit{pET} expression system, which increased yields of fusion protein and also increased efficacy of purification due to the affinity of the polyhistidine tag for Ni\textsuperscript{2+}. Further investigations\textsuperscript{104} on the expression of WThEPO and cysteine mutants as His\textsubscript{10}-fusion proteins were then
undertaken. Indeed, it appeared that the pET system was the preferred method for efficient expression of rhEPO and cysteine mutants (Table 2.1).

<table>
<thead>
<tr>
<th>Parent Vector</th>
<th>Fusion Protein</th>
<th>Cleavage site</th>
<th>Yield per 1 liter induced culture</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEX-2T</td>
<td>GST-hEPO</td>
<td>Thrombin</td>
<td>hEPO: 25μg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C38hEPO: 25μg</td>
</tr>
<tr>
<td>pET-16b</td>
<td>His$_{10}$-hEPO</td>
<td>Factor Xa</td>
<td>hEPO: 1 mg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C38hEPO: 0.6 mg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C38/83hEPO: 0.4 mg</td>
</tr>
<tr>
<td>pTrxFus</td>
<td>TrxA-hEPO</td>
<td>Enterokinase</td>
<td>N/D</td>
</tr>
<tr>
<td>pET-GST</td>
<td>GST-hEPO</td>
<td>Thrombin</td>
<td>hEPO: 40μg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C38hEPO: 40μg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C38/83hEPO: 20μg</td>
</tr>
</tbody>
</table>

Table 2.1 Summary of previous EPO expression work conducted within the group. TrxA = thioredoxin. The pET system provided the highest yield of EPO.

2.1 The pET System.

In pET vectors$^{105}$ (commercially available from Novagen, Madison, USA), genes to be expressed are brought under the control of strong bacteriophage T7 transcription signals where transcription is induced by providing a source of T7 RNA polymerase in the host cell$^{106}$. This highly active enzyme can transcribe RNA chains five times faster than *E. coli* RNA polymerase and is very selective since the promoter signals are rarely encountered in DNA unrelated to T7 DNA. Consequently the hosts own bacterial RNA polymerase does not induce transcription and hence translation. T7 RNA polymerase is provided by the host cell specifically engineered to carry a chromosomal copy of the λDE3 lysogen. Ultimately, the hosts expression of T7 RNA polymerase is controlled by a *lac*UV5 promoter which is induced by β-isopropyl thiogalactoside (IPTG) and as a result (as is generally the case with genes expressed under the control of the *lac* promoter) some background expression can be observed. This is only problematic however, when the target protein is sufficiently toxic to the host. Protein expression is therefore induced by the addition of IPTG. Novagen have made available a range of vectors which express target proteins with
CHAPTER 2

N or C-terminal tags to aid protein identification and purification. One of the more popular fusion proteins is the enzyme cleavable (thrombin or factor X proteases) His-tag where a polyhistidine tag simplifies the purification of target proteins due to its affinity for Ni\textsuperscript{2+} immobilised on commercially available chelating resins (Figure 2.1).

![Figure 2.1: Decahistidine tagged proteins are immobilized on Ni\textsuperscript{2+} chelating resins through their imidazole side chains.](image)

2.2. Expression Studies Using pET16b-EPO Constructs.

In previous work\textsuperscript{104}, the yield of His\textsubscript{10}-fused protein was still poor (1mg per liter at best). Unfortunately, a contaminant was found to co-purify with the His\textsubscript{10}-EPO fusions of similar molecular weight and once again preparative SDS-PAGE had to be employed to separate the fusion protein from this unwelcome contaminant.

However, the pET16b system was generally considered the most efficient expression system. Optimisation of the expression conditions for His\textsubscript{10}-WT, His\textsubscript{10}-C24, His\textsubscript{10}-C38 and His\textsubscript{10}-C38/83 EPO’s was then undertaken. As previously noted\textsuperscript{104}, expression (monitored by SDS-PAGE of the whole cell protein) was not significantly altered by changing the temperature at which the cells were induced between 30 and 37°C. 1mM IPTG was also considered optimal and in line with previous work and the manufacturers instructions. However, levels of expression of His\textsubscript{10}-EPO’s were found to vary significantly on altering the host strain of E. coli. In previous work, the BL21(DE3) and BL21(DE3)PlysS strains had been employed to facilitate expression of the target proteins but SDS-PAGE was to show that B834(DE3), the
parental strain of BL21(DE3) appeared to give larger expression levels of His_{10}-EPO's per unit volume of induced cell culture (Figure 2.2).

This was a significant improvement on previous results supplying enough material for analytical purposes. Furthermore, it showed that the previously described contaminating higher molecular weight band was in fact detectable by a murine anti-hEPO monoclonal antibody raised against a continuous epitope within the first 24 amino acids from the N-terminus (Figure 2.3).

This implied that both species originated from the same DNA and indicated that the lower molecular weight band was the desired His_{10}-EPO species and the upper band contained some type of extension. An observation contrasting severely with
previous results. Expression of EPO in the presence of PMSF\textsuperscript{107} had no effect on the product distribution (approximately 9:1 lower band/upper band (Figure 2.4)) implying that the product distribution could not be accounted for by proteolytic processing of the initial His\textsubscript{10}-fused species.

![Image of Figure 2.4: Ni\textsuperscript{2+} affinity purification of His\textsubscript{10}-C38hEPO gave rise to two species. M= MWt markers, L=loaded cell free extract, W=50mM imidazole wash, numbers refer to fraction (1.0ml) number.](image)

In order to show that the observation was not an artifact of SDS-PAGE the products were separated by preparative SDS-PAGE (after attempts to separate the species on the column by means of an imidazole gradient had failed) and separated proteins were obtained after electro-elution as described in previous work by Karen Sage\textsuperscript{104}. Comparison of the proteins by SDS-PAGE and western blotting convincingly demonstrated that the two observed bands were distinct chemical species (Figure 2.5).

![Image of Figure 2.5: SDS-PAGE analysis of separated and electro-eluted proteins.](image)
Initial conjecture was verified on obtaining the MALDI-TOF spectrum of the protein mixture prior to preparative SDS-PAGE.

<table>
<thead>
<tr>
<th>SPECIES</th>
<th>EXPECTED MWt</th>
<th>OBSERVED MWt</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>20906</td>
<td>21097</td>
</tr>
<tr>
<td>M2</td>
<td>*</td>
<td>22864</td>
</tr>
</tbody>
</table>

* Given that any extension would overwrite a stop codon—what is inserted (if anything) in place of stop is unknown and therefore the mass of the extension is unknown.

**Figure 2.6:** MALDI-TOF mass spectrum of Ni²⁺ affinity purified C38hEPO clearly indicates the presence of an extending peptide sequence.

The MALDI-TOF spectrum (Figure 2.6) clearly depicted two species, one within 200 amu of the expected His₁₀-C38EPO species (M1) and the other some 1767 amu larger (M2), confirming some sort of extension. Analysis of the encoding DNA (Figure 2.7) led us to assume that perhaps the extension was formed as a result of an inefficient stop codon, a phenomenon known to occur as a result of secondary gene structure, allowing protein synthesis to continue until the next stop codon is reached. Additionally, it has been shown that the efficiency of the stop signal can rely on the
identity of the following base. The +1 frame shift is particularly pronounced where a TGA stop codon is followed by G$^{108}$, as is the case in our pET16b-WT and cysteine mutant genes. A pause, or a frame shift are frequent outcomes of such a phenomenon. In fact, calculations were to show that the mass of the extension corresponded approximately to the mass the hypothetical protein species produced if this was indeed the case. The exact mass of any extension would be difficult to predict however since the species which is inserted (if any) in place of the TGA stop codon is unknown.

Figure 2.7: The pET16b multiple cloning site$^{105}$.

<table>
<thead>
<tr>
<th>Potential extending sequence$^{109}$</th>
<th>Calculated mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAANKARKEAELAAATAEQ</td>
<td>1939</td>
</tr>
</tbody>
</table>

A simple experiment that would examine this hypothesis was then devised. The WT and cysteine mutant EPO genes are inserted between the Nde I and BamHI endonuclease sites of the pET16b multiple cloning site (Figure 2.7). The suspected extending sequence was then contained within the following sequence, to the TAA stop codon, just beyond the Esp I site. Enzymatic digestion of the pET16b-EPO plasmid DNA with Esp I and BamHI I followed by religation should delete the majority of the extension. This reaction was performed on the plasmid DNA encoding His$_{10}$-C38 and His$_{10}$-WThEPO as in Figure 2.8, below:
Figure 2.8: Deletion of pET16b extending sequence involved a double endonuclease (BamHI/Bpu1102 I) digest. The non-cohesive ends were ‘filled-in’ using the Klenow fragment of DNA polymerase and blunt ended vector was religated and used to transform competent JM109 cells.

This resulted in a new, shorter, extension sequence:

\[
5' \quad \text{GGA TCT GAG CAA TAA} \quad 3'
\]

\[
\text{Gly Ser Glu Gln Stop}
\]

The resulting products designated pET16b-WTEPO(2) and pET16b-C38EPO(2) were then expressed in B834(DE3) cells and the Ni\(^{2+}\) affinity purified products were compared with those derived from the original plasmids (Figure 2.9).
Figure 2.9: His_{10}\text{-}C38EPO(2) shows no visible signs of an extending sequence by SDS-PAGE when compared with the original His_{10}\text{-}C38EPO (right).

SDS-PAGE showed that the distance between the two protein species (corresponding to molecular weight) had decreased and, as expected, the size of the extension had been decreased. This result was echoed by analysis of the MALDI-TOF spectrum of His_{10}\text{-}C38EPO(2) (Figure 2.10):

Figure 2.10: MALDI-TOF mass spectrum of His_{10}\text{-}C38EPO(2).

Similar results were also obtained with His_{10}\text{-}WTEPO(2).
Although this confirmed that the problem had arisen as a result of the inefficient stop codon it had not solved that problem. The EPO now being expressed could potentially have a four or five amino acid extension and it had lost the ability to be manipulated further with ease as a result of the destruction of the BamH I and Esp I sites (Figure 2.8).

2.3 Optimisation of expression.

A new experiment, involving cassette mutagenesis, was then devised that would serve to solve our extension problem and maintain the integrity of the restriction endonuclease sites. Fortuitously, a unique Stu I restriction site was observed only twenty seven base pairs from the DNA encoding the C terminus of EPO. This allowed us to excise only a small DNA fragment from the EPO gene and replace it with a synthetic 27mer oligo deoxynucleotide.

The experiment (outlined schematically in Figure 2.11) involved the cloning in of the synthetic duplex, replacing the unsatisfactory C- terminus encoding DNA fragment.

The synthetic oligonucleotides were chosen carefully to introduce new optimised codons for expression in \( E. coli \). (replacing human codons, for which \( E. coli \) has not been shown to have a high percentage of corresponding tRNA, with optimal \( E. coli \) codons for the same amino acid) and a new, TAA, stop codon (one more commonly found) followed by the original TGA stop codon:

---

**Figure 2.11:** Introduction of a new C-terminal sequence using cassette mutagenesis.
Original C-terminal sequence:

\[ 5' \text{GAG GCC TGC AGG ACA GGG GAC AGA TGA GGATCC} \]

\[ \text{Glu Ala Cys Arg Thr Gly Asp Arg Stop BamHI} \]

New, optimised sequence: (codon changes in bold type, endonuclease restriction sites in red)

\[ 5' \text{GAG GCC TGC CGT ACC GGT GAC CGT TAA TGA GGATCC} \]

\[ \text{CTC CGG ACG GCA TGG CCA CTG GCA ATT ACT CCTAGG} \]

\[ \text{Glu Ala Cys Arg Thr Gly Asp Arg Stop Stop BamHI} \]

and this therefore required the purchase of two synthetic oligonucleotides which were obtained desalted and deprotected:

\[ 5' \text{CG TGC CGT ACC GGT GAC CGT TAA TGA G} \]

\[ \text{termed EPOCOPT 1} \]

and

\[ 5' \text{GA TCC TCA TTA ACG GTC ACC GGT ACG GCA GG} \]

\[ \text{termed EPOCOPT2} \]

when equimolar quantities of each were ‘melted’ in the conventional manner they afforded the desired fragment

\[ 5' \text{CG TGC CGT ACC GGT GAC CGT TAA TGA G} \]

\[ \text{GG ACG CGT TGG CCA CTG GCA ATT ACT CCTAG} \]

\[ 3' \text{G6 CG TGG CCA CTG GCA ATT ACT (C1 G} \]

which served to optimise \emph{E. coli} expression for this length of DNA, introduce a new stop codon followed by the original stop codon and reformed the restriction endonuclease sites after insertion to allow further genetic manipulations. After the ligation, reaction mixtures were used directly to transform JM109 cells and the resulting colonies were screened. The screening procedure, consisting of two stages, was very straightforward:
First, transformants were cloned and the DNA was purified and digested with \textit{Stu I} and \textit{BamH I}. If the plasmid had just closed upon itself without incorporation of the synthetic duplex the recognition site for \textit{Stu I} would have been lost. The \textit{BamH I} site, however, may have been reformed. The DNA resulting from such an unfavourable process would not be cleaved with \textit{Stu I} but will have been cleaved by \textit{BamH I}.

Second, the incorporation of the synthetic duplex introduced a new \textit{BstE II} site, increasing the number of \textit{BstE II} sites in the plasmid from one to two. Therefore incorporation of the synthetic duplex would result in a fragment being cleaved from the plasmid rather than just linearising when treated with \textit{BstE II}.

Seven colonies were screened initially (Figure 2.12). From the first screen only, one transformant (number 4) fulfilled the initial criteria of being cleaved by \textit{Stu I} and \textit{BamH I} enzymes.

\textbf{Figure 2.12}: \textit{BamH I} digest (left) and \textit{Stu I} digest (right) of transformants. WT=original unmodified pET16b-WT, M=MWt standards, P=transformants where the insert was phosphorylated prior to ligation.
CHAPTER 2

This DNA was then taken on to the next stage and was shown to be cleaved (not just linearised) by *Bst*E II (Figure 2.13). To confirm that the sequence was correct the resulting DNA was sequenced across the inserted region using the pET16b reverse primer (Figure 2.7).

Figure 2.13 Agarose gel shows (from left) *Stu* I cut original pET16b-WThEPO (a standard), molecular weight markers, *Bst*E II cut WThEPO (standard), and *Bst*E II cut pET16b-WThEPO(COPT), followed by markers.

Now that we had the desired gene, termed pET16b-WThEPO(COPT), we were then able to manipulate the remaining genes encoding cysteine mutants. This was particularly simple and all experiments (outlined schematically in figure 2.14) were conducted in parallel: The optimised pET16b-WThEPO(COPT) construct was digested with *Nco* I and *Stu* I to remove the wild type sequence, while the optimised 3' end of the gene (from the synthetic duplex) remained attached to the pET16b-vector. N24C, N38C and N38/83C EPO pET vectors were also cut *Nco* I/ *Stu* I, but this time the fragments corresponding to the mutant encoding cDNA was purified from the endonuclease digestion and these fragments were ligated with the pET vector incorporating the optimised C-terminal sequence. Colonies were screened as previously and now all genes contained the optimised C-terminal sequence.
The four new plasmids designated pET16b-WTEPO(COPT), pET16b-C24EPO(COPT), pET16b-C38EPO(COPT) and pET16b-C38/83EPO(COPT) were then fully sequenced across the His₁₀-EPO encoding region using EPOCOPT 2 (single strand of the synthetic duplex) as primer. The sequences also confirmed that all genes contained the correct optimized sequence. The target proteins were then...
expressed and as expected only a single product was observed after Ni\textsuperscript{2+} affinity chromatography (Figure 2.15): 

**Figure 2.15**: SDS-PAGE analysis of Ni\textsuperscript{2+} affinity purified His\textsubscript{10}-C38hEPO from the optimized pET16b-C38EPO(COPT) vector.

The advantages were two fold:

1) preparative- SDS-PAGE (a cumbersome technique with limitations concerning the quantity of protein that can be loaded on to a gel at any one time) was no longer required to purify our proteins.

2) 100% of the expressed protein has the correct amino acid sequence (no extensions) and the yield of protein after Ni\textsuperscript{2+} chromatography was in excess of 13mgL\textsuperscript{-1} based on extinction coefficient (ε = 22430 at 280nm) and Bradfords assay. All proteins purified from 1L cultures were generally eluted from the column in 9.0ml buffer that when pooled, provided an approximately 60μM stock solution.

The expression of His\textsubscript{10}-EPO's could also be scaled up in a 10 liter fermenter, but with limited reliability. Expression tended to vary from one batch to the next and protein yields were never 10×13mg for a 10 liter culture (7mgL\textsuperscript{-1} at best). Notably, if the initial inoculation of media with overnight starter culture resulted in a culture with OD\textsubscript{600}<0.1 then, on occasion, the culture would lyse early on in the fermentation and never reach induction OD\textsubscript{600}=0.6.

The MALDI-TOF mass spectrum (Figure 2.16) showed that the protein expressed now was a single species:
Although MALDI-TOF mass spectra were qualitatively very useful, in the absence of a delayed extraction MALDI instrument we were not convinced that they would provide the required resolution to allow glycoproteins (with mass differences of two hundred mass units) to be easily observed.

### 2.4 NMR analysis of His\textsubscript{10}-WThEPO.

Having access to much higher quantities of EPO we were interested in looking at EPO by NMR. This work was done in collaboration with Dr Emma Beattie. The three dimensional structure of human erythropoietin has remained unsolved by multidimensional NMR spectroscopy and, until recently\textsuperscript{85}, X-ray crystallography due to its poor solubility and heterogeneity. His\textsubscript{10}-WT was expressed in media

<table>
<thead>
<tr>
<th>Protein</th>
<th>Calculated mass (Da)</th>
<th>Observed mass</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Mono-isotopic</td>
</tr>
<tr>
<td>His\textsubscript{10}-WThEPO</td>
<td>20903.76</td>
<td>20916.85</td>
</tr>
</tbody>
</table>

**Figure 2.16**: MALDI-TOF mass spectrum of His\textsubscript{10}-WThEPO derived from the optimized pET16b-WThEPO(COPT) vector.
containing \((^{15}\text{NH}_4)_{2}\text{SO}_4\) as the only source of nitrogen. The protein was purified from inclusion bodies and the Ni\(^{2+}\) affinity purified material was refolded according to the protocol developed by Boissel and co-workers\(^{39}\).

However, measured concentrations were low in 2% N-lauroyl sarcosine, 50mM tris.HCl; pH8 due to poor solubility. The protein was then precipitated and redissolved in fully deuterated 70% MeCN, 0.1% TFA to see whether we could observe EPO in the \(^{15}\text{N}-^1\text{H}\) HSQC spectrum (Figure 2.17).

![Figure 2.17: \(^1\text{H}-^{15}\text{N}\) HSQC spectrum of His\(_{10}\)-WTI\(\text{hEPO}\). Spectra were acquired in 70% CD\(_3\)CN/D\(_2\)O, 0.1% CF\(_3\)COOD. The three tryptophan residues (10.5ppm in the proton dimension) present in EPO are clearly observed.](image)

From this encouraging spectrum, we went on to look at the \(^1\text{H}-^{15}\text{N}\) HSQC spectrum obtained from a sample refolded against 2% N-lauroyl sarcosine, 50mM tris.HCl; pH 8, 40\(\mu\)M CuSO\(_4\), The paramagnetic Cu\(^{2+}\) ions, employed to catalyse oxidative disulfide bond formation, were removed by dialysis against refolding buffer free of CuSO\(_4\) and then by the addition of chelating resin which could be removed by centrifugation.
The resulting spectrum (Figure 2.18) was very poor, showing very broad signals and little resolution in the $^{15}$N dimension. This indicated that attempts to refold our His$_{10}$-wild type protein according to the protocol of Boissel and co-workers had been unsuccessful and the protein had remained unfolded. An identical spectrum was also obtained when our *E. coli* derived protein was refolded against 35% glycerol, 40 µM CuSO$_4$. We later discovered (Chapter 3, 3.1-3.3) that our *E. coli* derived proteins were expressed in an oxidised form, comprising monomeric and highly aggregated material. This was surprising as we expected the protein to be in a reduced form owing to the reducing conditions (GSH:GSSG >1) in the bacterial cytoplasm. Therefore, the refolding process (in the absence of a prior reductive step) was likely to have had little affect on the expressed protein. Consequently, a further avenue of investigation would be to oxidatively refold the protein after a reductive step and re-examine the solution structure of recombinant human EPO by NMR spectroscopy.

Since our attempts to solve the 3D structure of hEPO by NMR, Cheetham and co-workers (Amgen Inc) have reported on the NMR structure of erythropoietin.
Similarly, they had difficulty in dissolving EPO to the required concentration for NMR and mutations of all N-glycosylation sites from asparagine to lysine were made resulting in MKLysEPO which was soluble at concentrations in excess of 20mgL$^{-1}$. Furthermore, they reported that their MKLysEPO suffered the effects of rapid $^{15}$N and $^{13}$C relaxation rates, giving poor magnetisation transfer through scalar coupling in the NMR experiments used for assignments. This results in broadening of signals and poor signal to noise ratio. Their solved solution structure (Figure 2.19) shows good agreement with previously determined model structures.

![Figure 2.19: Solved solution structures for MKLysEPO (b) and ribbon diagram (a)](image)

The NMR structure of the wild type protein, though undoubtedly similar to that solved for MKLysEPO, remains unsolved.
CHAPTER 3.

Electrospray Ionisation
Mass Spectrometry
(ESI–MS):
Taking a closer look.
Electrospray ionisation mass spectrometry (ESI-MS), a softer ionisation method, has found broad applicability in structural biology, particularly in the analysis of primary structures of biopolymers such as proteins and nucleic acids. ESI-MS has allowed the routine analysis of very large and very labile molecules from small organics to large supramolecular assemblies and protein complexes bound by non-covalent interactions. 

Typically, a sample solution is pumped through a capillary (PEEK or fused silica) whose end is supported by a stainless steel capillary needle. Charged aerosol droplets are electrosprayed by means of an electrical potential (2-5kV) between the needle tip and the cylindrical counter electrode and dispersed into the ionisation region. Desolvation of macro-ions is assisted the use of a counter current ‘drying’ gas and by heating the probe (Figure 3.1).

Macromolecular ions of various charges are formed by a process known as field-ion evaporation, which leads to a rapid decrease in droplet size and desolvation. The mechanism of ion formation in ESI is still much debated and is affected by coloumbic forces in combination with the effects of the ESI interface conditions.
analyte charge, temperature and solvent polarity. The mechanism is shown schematically in Figure 3.2.

![Figure 3.2: Proposed mechanism of macro ion formation in ESI-MS\textsuperscript{110}](image)

Below is shown a typical charge envelope for Myoglobin.

![Figure 3.3: ESI-MS charge envelope for Horse Heart Myoglobin.](image)

What we can assume is that the peaks in the ‘charge envelope’ correspond to a single protein and each peak differs from the next in terms of charge by +1 and that charge is due to protonation. Since the X-axis corresponds to the m/z ratio, the true mass $M_R$, of myoglobin can be determined by multiplying the mass $M_i$ (corresponding to a particular peak) by the charge associated with that species (i), less the mass of the number of protons that provide the charge (i$M_H$):

\[
\begin{align*}
\text{(I)} & \quad M_R = iM_i - iM_H \\
\text{(II)} & \quad M_R = (i+1)M_i - (i+1)M_H
\end{align*}
\]

Where $M_R$ = Molecular weight of protein R  
$M_i$ = m/z of peak with i charges  
$M_H$ = mass of a proton
The charge state corresponding to a particular peak in the envelope can then be determined by rearranging I and II to eliminate $M_R$ and solving the simultaneous equations for charge $i$:

\[
\begin{align*}
I & : \quad M_R = iM_i - iM_{i+1} \\
-II & : \quad -M_R = -(i+1)M_{i+1} + (i+1)M_H
\end{align*}
\]

\[
\begin{align*}
0 &= iM_i - iM_{i+1} + (i+1)M_H \\
0 &= iM_i - iM_{i+1} - M_{i+1} + i + 1 \\
0 &= i(M_i - M_{i+1}) - M_{i+1} + 1 \\
i(M_{i+1} - 1) &= M_{i+1} - 1 \\
i &= \frac{M_{i+1} - 1}{M_i - M_{i+1}}
\end{align*}
\]

Once $i$ is known then each peak can be used to determine a separate estimate of molecular weight $M_R$. Fortunately, the platform II has software (Transform and Maximum entropy) capable of performing this type of analysis on proteins and complex mixtures of peptides almost instantaneously. Thus charge spectra are routinely and rapidly deconvoluted by the mass spectrometer’s data management system (Mass-Lynx v2.3).

### 3.1 Off-line LC-MS analysis of His$_{10}$-EPO samples.

ESI spectra of Ni$^{2+}$ affinity purified His$_{10}$-hEPO samples were initially obtained after the samples had been desalted using C$_4$ or C$_{18}$ reverse phase chromatography. The gradient employed was typically 10-100% acetonitrile over 15 to 20 minutes. Pooled HPLC fractions were lyophilised and re-dissolved in the minimum volume of 70% acetonitrile/water, 0.1% formic acid. These (standard ESI) conditions were very poor and failed to solubilise the EPO samples. A number of organic acids (HCOOH, AcOH, TCA, and TFA) at varying concentrations were then screened along with mixtures of acids (eg. 0.1% AcOH; 0.01% TFA). TFA, at concentrations of 0.05-0.1%, was found to be the most successful at solubilising the EPO samples and
protonating them sufficiently to be observed by ESI-MS. Under these conditions ESI spectra were still very poor (Figure 3.4).

![ESI-MS Spectrum](image)

**Figure 3.4:** Poorly resolved charge envelope for His\(_{10}\)-C38hEPO, acquired in 70% MeCN/H\(_2\)O, 0.05% TFA.

Additionally, ESI spectra of WT and cysteine mutants treated with an excess of sulfhydryl modifying reagents such as NTCBA (12), DTNB (Ellmans reagent (13)), iodoacetamide (14), 5-I-AEDANS (15) showed no evidence that sulfhydryl modification had taken place.

![Chemical Structures](image)
This was surprising since samples of Ni\textsuperscript{2+} affinity purified EPO had not yet been oxidatively refolded and theoretically had 4 or 5 free thiol groups. These findings indicated that perhaps the samples were expressed in an oxidised form (with formed disulfides). Even more surprising was that cysteine mutants, containing 5 cysteine residues (ie. an odd number) still failed to react.

The reasons for this soon became clear. WT EPO samples, though providing poor data, could still be resolved to give masses close to the calculated whereas cysteine mutants had a mass approximately 300 mass units higher than expected (Figure 3.5).

It was hypothesised that this mass anomaly could result from the formation of an adduct with the tripeptide glutathione, (16)(molecular weight= 307.3Da), present in the bacterium. If this was to be the case then it should be possible to reduce it off with β-mercaptoethanol or dithiothreitol (DTT). DTT was shown to be superior in this case as various intermediate EPO-mercaptoethanol species (16b) were frequently observed\textsuperscript{113} when using β-mercaptoethanol as the reducing agent.
DTT tends to be more effective than mercaptoethanol due to the rapid intramolecular cyclisation reaction resulting in a stable 6-membered ring.

The effects of reduction on the ESI spectra were more dramatic than expected. The charge spectrum of the reduced species was far better resolved (Figure 3.7) facilitating accurate mass measurement of all samples.

Figure 3.6: Mechanism of disulfide bond reduction by β-mercaptoethanol and DTT.

Figure 3.7: Charge spectrum of reduced (50mM DTT) His_{10}-C38hEPO.
This was believed to be the result of loss of any structure that the initially expressed protein may have had. Cleaved disulfide bonds resulted in a more open structure which could be protonated more readily\(^\text{113}\). Such an effect was also observed by Przybylski and co-workers\(^\text{110}\) when investigating the effect of disulfide reduction on the structure of hen egg-white lysozyme (HEL). Also, the presence of EPO oligomers or aggregates (observed by non-reducing SDS-PAGE of column purified material) may inhibit sufficient protonation.

Further investigations with all protein samples provided a general rule:

| Reduced protein samples gave good MS data resulting in more accurate masses. |
| Oxidised protein samples gave poor MS data resulting in less accurate masses. |

This was also later shown to be true for oxidatively refolded samples. Unlike their precursors, the refolded protein samples were shown to be monomeric by SDS-PAGE conducted under non-reducing conditions.

### 3.2 Typical Resolved Mass Spectra (shown below for WThEPO)

![Typical resolved (Max Ent) spectrum of reduced His\(^{10}\)-WThEPO shows only one major species over a 5KDa-50KDa mass range.](image-url)

**Figure 3.8:** Typical resolved (Max Ent) spectrum of reduced His\(^{10}\)-WThEPO shows only one major species over a 5KDa-50KDa mass range.
From the acquired data we were able to determine that:

1) Expressed proteins had their N-terminal methionine residues cleaved (this was verified by N-terminal sequencing. Edman degradation gave glycine as the N-terminal amino acid).

2) Proteins were obtained from *E. coli* in an oxidised form—with (either correctly, incorrectly or as a mixture of both) formed disulfides and cysteine mutants were capped with glutathione suggesting that the oxidation occurs inside the cell, possibly as some sort of detoxification mechanism.

Full alkylation was then observed with sulfhydryl modifying reagents such as iodoacetamide (14) after reduction with DTT.

![Figure 3.9: Mass spectrum of His₁₀-C24hEPO (pink) superimposed on the mass spectrum of reduced and iodoacetamide treated His₁₀-C24hEPO (green). The largest peak (in the green spectrum) corresponds to the addition of six (not five as expected) acetamide units.](image-url)
CHAPTER 3

Notably, these reactions were particularly messy and data was poor unless the reducing agent was removed (by protein precipitation or dialysis) prior to the alkylation step. From these preliminary results we could begin to explain our previous findings:

Our NMR spectra (Chapter 2, 2.4) of refolded WThEPO were always broad and looked unfolded. However, if disulfides were formed incorrectly during or soon after protein synthesis (locking it in a particular conformation) then the oxidative refolding step (Boissel et al) would be ineffective in bringing about any change in structure. We cannot, however, explain the differences between our findings (2 species expressed, oxidised) and those of Boissel and co-workers.

Cysteine mutants were simply inert to sulfhydryl modifying reagents because there were no free thiol groups present.

This prompted a new refolding protocol (Figure 3.10) whereby all protein samples were reduced prior to oxidative refolding. It has been reported that the biological activity of EPO is lost irreversibly when reduced and alkylated, but can be regained quantitatively when reduced protein samples are re-oxidised, hinting at the importance of correctly formed disulfides. On introducing an extra cysteine into EPO we were particularly aware of the potential for mis-folding. However, proteins are generally unaffected by single mutations and we continued, utilising established refolding protocols. Protein was reduced with 10mM DTT or tricarboxyethyl phosphine (TCEP) for 3h at room temperature. DTT was then removed by dialysis against 6M guanidine.HCl, 50mM Tris.HCl; pH8 under nitrogen until all DTT had been removed (monitored by Ellmans assay). Proteins were then oxidatively refolded according to the procedure of Boissel and co-workers. Pulse refolding (a newer refolding method) is also being pursued by new members of the group. In order to establish that the disulfide bonds had been correctly reformed we resolved to map the position of the glycosylation site after refolding and glycosylation and this is the subject of chapter 5.
Figure 3.10: Modified refolding protocol for the renaturation of WT and cysteine mutant hEPO's.
CHAPTER 3

From our electrospray data from reduced proteins we could then determine accurate masses for all EPO species:

Masses are calculated for all proteins in their reduced forms.

<table>
<thead>
<tr>
<th>His₁₀EPO species</th>
<th>Calculated mass (Da)</th>
<th>Observed mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mono-isotopic</td>
<td>Average</td>
</tr>
<tr>
<td>WT</td>
<td>20903.76</td>
<td>20916.85</td>
</tr>
<tr>
<td>C24</td>
<td>20892.73</td>
<td>20905.89</td>
</tr>
<tr>
<td>C38</td>
<td>20892.73</td>
<td>20905.89</td>
</tr>
</tbody>
</table>

Table 3.1 Calculated and observed masses of reduced WThEPO and C24 and C38 cysteine mutants. Mass differences between calculated and observed masses are likely to have resulted from the instrument calibration.

3.3 Optimisation of Mass Spectrometry Analysis Using On-Line LC-MS.

Although with reduced protein samples we could obtain good MS data, the procedure: reverse phase HPLC, followed by fraction combining and lyophilisation, then ESI-MS was particularly time consuming and labour intensive and resulted in loss of sample. To minimise handling, an LC-MS method was developed, which combined the ability to use high salt concentrations (necessary to dissolve EPO and conduct modification reactions in) yet provide salt free EPO samples for MS. This
was simply achieved by disconnecting the HPLC from the MS until salt and polar impurities had eluted from the C₄ or C₁₈ reverse phase column. This was determined by monitoring the UV trace (Figure 3.11).

![UV Trace Diagram]

**Figure 3.11:** The UV trace allows us to see when all guanidine and polar reagents have passed through the column and protein elution is observed in the UV trace and in the (total ion chromatograph) TIC.

In summary, the acquired electrospray data had proven itself to be of particular use in helping us establish the state of our proteins as they were presented to us by the bacterium. Furthermore, we were able to gain inferences into the structure of the expressed material and problems that it may have presented to the host organism during expression. With refolded cysteine mutants in hand we could then attempt our glycosylation reactions with glycosyl iodoacetamides.
CHAPTER 4.

The aim of this project was to develop methods for the synthesis of homogeneous glycoforms of recombinant human erythropoietin. Since the absence of the O-linked glycan has no effect on the biological activity of EPO, the focus of the project was on the synthesis of analogues of the N-linked glycans. Although the biological aspect of this project was very time consuming, through a keen interest in gaining further experience in synthetic organic chemistry, synthesis of relevant, useful and more complex carbohydrates was attempted. The core pentasaccharide of EPO is required for full biological activity \textit{in vitro} and the synthesis of this type of structure is a long-term goal. Fortunately, a handful of groups have recently reported on the synthesis of this (Man$_3$GlcNAc$_2$) pentasaccharide core structure$^{48,49,53,55}$ so one could be forgiven for becoming complacent about the difficulties in preparing such a molecule. From a total synthetic viewpoint, a brief discussion of the major issues which arise during the synthesis of (17) and applying known methodologies to the synthesis of the iodoacetamide derivative (18) follows.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig4_1.png}
\caption{Reagents for the convergent methods of glycoprotein and neoglycoprotein synthesis.}
\end{figure}
4.1 Oligosaccharide Synthesis.

Oligosaccharide synthesis is still far from routine and has evaded the luxury of automation, contrasting with the now facile synthesis of oligopeptides and oligonucleotides. This is primarily due to the fact that glycosidation reactions, by which monosaccharides are extended to oligosaccharides are seldom quantitative and tend to result in anomeric mixtures which need to be separated\textsuperscript{37,47,118}.

Most methods of glycoside bond formation can be summarised as belonging to the class below:

![Diagram](image)

Figure 4.2: General mechanism for glycoside bond formation\textsuperscript{118}. The intermediate carbocation is stabilised by the ring oxygen (21).

In the absence of a participating group at C-2 (P=benzyl, P=azide) mixtures of glycosides are prepared in proportions which vary with the reaction conditions. In the presence of a participating group however (P=acetyl, benzoyl, pivaloyl N-phthaloyl, N-acetyl etc.) 1,2-trans glycosides can be prepared in high yields (Figure 4.3):

![Diagram](image)

Figure 4.3: Neighbouring group participation at C-2 facilitates the synthesis of 1,2-trans glycosides\textsuperscript{118,119}. 

68
4.2 $\beta$1-4 Mannoside Linkages.

This oligosaccharide linkage (Figure 4.1, shown in red) has become infamous through years of carbohydrate research. It is known to be one of the most difficult glycosidic linkages to prepare (disfavoured by the anomeric effect and participating groups at C-2). However a number of methods have been developed for its synthesis and wise synthetic methodologies introduce the Man$\beta$1-4GlcNAc linkage at as late a stage in the synthesis as possible. Below are some recent examples of the successful synthesis of the $\beta$-mannoside linkage used in the context of constructing the pentasaccharide core of N-linked glycans.

Danishefsky’s Method:
Uses an oxidation (24)$\rightarrow$(25)-reduction (25) $\rightarrow$ (26) sequence on the full pentasaccharide$^{48}$:

Ogawa’s Method:
Tethers the 2-position of mannose to the 4-position of N-phthalimido GlcNAc through a $p$-methoxybenzylidine acetal (27), setting up the stereochemistry for reaction on the $\beta$-face of mannose$^{120}$.
Unverzagt's Method\textsuperscript{121} (uses the method developed by Kunz\textsuperscript{122})

The β-mannoside was introduced from the β-glucoside through an intramolecular S\textsubscript{N}2 reaction where the carbonyl oxygen of an N-phenyl urethane at C-3 (28) serves as nucleophile, displacing the chloroacetyl group.

Other recent developments include direct\textsuperscript{123} and ultrasound assisted\textsuperscript{124} S\textsubscript{N}2 displacement of triflates and chloroacetyl groups at C-2 and enzymatic methods\textsuperscript{125,126}.

4.3 N-Acetyl Groups as Neighbouring Groups for Participation in β-glycoside Synthesis.

These groups can be very problematic in syntheses as they have a removable proton, which can lead to stable reaction intermediates under acidic or basic conditions\textsuperscript{119}.

Figure 4.4: N-acetyl groups can form stable 1,3-oxazolines.

This problem is generally solved using N-protecting groups which lack this proton and many are known and have been recently reviewed\textsuperscript{127,128}:
4.4 Synthesis of More Complex Naturally Occurring Glycans.

For oligosaccharide extension, beyond the core pentasaccharide, orthogonal protection is required on C-2 of the two terminal mannose residues to allow extension of oligosaccharide to more complex natural products\textsuperscript{48}. This is not an issue here however our preliminary target would be the core pentasaccharide.

4.5 Glycosyl Amines.

Glycosyl amines have long been at the center of investigations into N-linked glycoprotein chemistry as they have many desirable properties. They have a cyclic structure and tend to adopt the β-anomeric configuration, which is found in the natural sugar-protein linkage. A good synthesis of glycosyl amines has consequently been the focus of much research over the past two decades but generally, one of two methods are used.

4.5.1 Glycosyl amines from naturally occurring mono and oligosaccharides.

In the absence of protecting groups, saccharides from natural sources can be converted to the corresponding glycosyl amine by direct condensation of the aldehyde form of the reducing monosaccharide with ammonia\textsuperscript{50,57,129} (formed through the decomposition of ammonium bicarbonate, NH\textsubscript{4}HCO\textsubscript{3} in saturated aqueous solutions):
These reactions proceed in near quantitative yield and the products are usually used without further purification as the starting material does not react significantly in the following step (coupling to synthetic peptide). The major problem with this reaction is that all traces of ammonium bicarbonate must be removed (lyophilised) as it may also take part in the following reactions.

4.5.2 Glycosyl Amines from Reduction of Glycosyl Azides

Glycosyl azides (37) have proved very useful in the synthesis of glycosyl amines, they are more stable and can be incorporated at early stages in the synthesis and remain intact throughout. They are generally introduced by direct $S_N2$ displacement of $\alpha$-halides (36) with inversion at the anomeric centre under phase transfer catalysis conditions\textsuperscript{130}. Alternatively, through Lewis acid mediated cleavage of $\beta$-acetates in the presence of homogeneous sources of azide (trimethyl silyl azide)$^{131}$.

Figure 4.6: Common syntheses of glycosyl azides (37) include $S_N2$ displacement of glycosyl halides (36) under phase transfer catalysis conditions and exchange of anomeric acetates with azide ion in the presence of lewis acids.

Azides can then be reduced using a number of methods, catalytic hydrogenation ($H_2$ and Pd/C, PtO\textsubscript{2} or Raney nickel), thioacetic acid, thiols$^{48,55,131}$. Problems are that the
synthesis of α-chlorides is poor and even worse for large saccharides, as is the synthesis of beta acetates.

One recently reported example, which demonstrates the versatility of the glycosyl azide and latent functionality, prepared the glycoamino acid (39) in high yield, directly from the glycosyl azide, a useful reaction for glycopeptide chemistry.\(^\text{132}\)

4.6 Strategic Problems.

Synthetic glycosyl amines are usually fully protected when coupled to fully protected synthetic peptides with activated aspartate residues-global deprotection then follows. Our glycosyl iodoacetamides are usually deprotected and water soluble prior to reaction with unprotected recombinant proteins.

Potential caveats that should be noted are:

1. Glycosyl amines are generally unstable and the amino group is readily eliminated.
   
   If we chose to deprotect the oligosaccharide fully and then treat with iodoacetic anhydride, the amine, or precursor (usually azide) must withstand the deprotection conditions.

2. If the fully protected oligosaccharide is coupled to activated iodoacetic acid (using standard peptide coupling reagents) the resulting iodoacetamide must withstand the deprotection conditions.

The issues detailed above cause significant problems when attempting the synthesis of this class of compounds and must be considered prior to attempting their synthesis.

For our preliminary glycosylation studies we would be considering simpler synthetic targets such as the iodoacetamide derivatives of N-acetyl glucosamine (GlcNAc, (40)) or chitobiase (41), the first two residues of the pentasaccharide core.
Retrosynthetic analysis of these compounds is very simple, resulting in glycosyl amines and activated versions of iodoacetic acid (anhydrides, acid chlorides, active esters etc.)

4.7 Synthesis of β-N-Glycosyl Iodoacetamides.

The direct synthesis of the glycosylamine (34), R=H, from the parent hemi-acetal (33), R=H, was achieved by dissolving N-acetyl glucosamine in the minimum volume of saturated aqueous ammonium bicarbonate, the reaction mixture was then stirred at room temperature for five days. Further NH$_4$HCO$_3$ was added throughout to ensure that the solution remained saturated. The excess NH$_4$HCO$_3$ was removed by lyophilisation of the reaction mixture to afford the crude product as a pale yellow solid. No attempt was made to purify the product further. The $^1$H NMR spectrum of this amine contained other non-integral anomeric resonances corresponding to unreacted starting material but generally the product is used directly as the starting material should not react in subsequent steps. The product was then dissolved in the minimum volume of 1M NaHCO$_3$ and portions of iodoacetic anhydride were added at regular intervals until TLC indicated that the starting material had been consumed (Figure 4.7).

![Figure 4.7: Selective iodoacetylation of the glycosyl amine in aqueous solution.](image)

The product was purified by passing the reaction mixture over a column of acidic ion exchange resin, then basic ion-exchange resin and lyophilised. This reaction was
also used to prepare the glycosyl iodoacetamides of lactose (42), maltose (43) and chitobiose (41):

We hoped to use this synthesis of glycosyl amines as a route to more complex sugars. The amino group could be trapped with FmocOSu\(^{134}\) and purified in yields of 80% but the Fmoc group was generally unstable to further reactions.

We then chose to attempt the synthesis of larger saccharide-\(\beta\)-iodoacetamides using glycosyl azides. In the simplest case, for N-acetyl glucosamine, the iodoacetamide (40) could be prepared in 5 steps (Figure 4.8):

\begin{align*}
\text{AcCl, 16h, RT, 50\%, (II) NaN}_3, \text{TBAHS, DCM, Sat. aq. NaHCO}_3, \text{RT, 1h, 97\%, (III) H}_2, \text{PtO}_2, \text{THF, RT, 3h, 74\%, (IV) pyridine, DCM, -40\degree C, 0.5h, 95\%, (V) NaOMe, MeOH, RT, 2h, quantitative.}
\end{align*}
The procedure included the use of Horton's synthesis of the per-acetylated-α-chloride\textsuperscript{137}. This reaction proceeded in only 50% yield with the other major product being the fully acetylated GlcNAc. The halide was then displaced with azide ion under phase transfer catalysis conditions as previously described (Figure 4.6) in excellent yield. After reduction of the azide, we chose to use the acid chloride (45) of iodoacetic acid since attempts at coupling the per-acetylated glycosyl amine (44) with iodoacetic acid using peptide coupling reagents (pyBOP, EEDQ, DCC)\textsuperscript{135} were low yielding (0%, 20% and 40% respectively). The acid chloride was prepared in near quantitative yield using a slight excess of thionyl chloride:

\[
\begin{align*}
&\text{I} \quad \text{O} \quad \text{OH} \quad \xrightarrow{\text{SOCl}_2} \quad \text{I} \quad \text{O} \quad \text{Cl} \\
&\text{O} \quad \text{I} \quad \text{O} \quad \text{OAc} \quad \xrightarrow{\text{I}} \quad \text{O} \quad \text{I} \quad \text{O} \quad \text{N}_3
\end{align*}
\]

The synthesis of the glycosyl amine (44) was later simplified and optimised further using the commercially available β-acetate (38)\textsuperscript{131}.

\[
\begin{align*}
&\text{38} \quad \xrightarrow{\text{I}} \quad \text{37} \quad \xrightarrow{\text{II}} \quad \text{44} \\
&\text{AcO} \quad \text{O} \quad \text{OAc} \quad \xrightarrow{\text{I}} \quad \text{AcO} \quad \text{O} \quad \text{N}_3 \quad \xrightarrow{\text{II}} \quad \text{AcO} \quad \text{O} \quad \text{NH}_2
\end{align*}
\]

\textbf{Figure 4.9:} (I)TMS-N\textsubscript{3}, SnCl\textsubscript{4}, CH\textsubscript{2}Cl\textsubscript{2}, RT, 16h, 97%, (II), H\textsubscript{2}, 5%Pd/C, MeOH, 3h, quantitative.

This meant that the low yielding synthesis of the α-chloride (36)\textsuperscript{136-138} was no longer required and the protected glycosyl amine (44) was prepared in 97% yield over the two steps\textsuperscript{131} compared with 36% over three steps (Figure 4.8). With our N-acetyl glucosamine derived glycosyl iodoacetamide in hand we conducted some test reactions with a suitably protected cysteine residue\textsuperscript{139,140} (Figure 4.10)
Figure 4.10: (I) Et$_3$N, (Ac)$_2$O, CHCl$_3$, RT, 1 h, 88%, (II) Zn, HCl, MeOH, 0°C, 2.5 h, 42%, (III) 50 mM (NH$_4$)$_2$CO$_3$ (pH 8), RT, 1 h, 98%.

And then the same reaction was conducted with the tripeptide glutathione:
Throughout these experiments we also noted that the chemical shift of the $\alpha$-methylene carbon (of the iodoacetamide, (40)) changed significantly from $-3.58$ to $+35.2$ in (50) and (52)(Table 4.1):

<table>
<thead>
<tr>
<th>Compound</th>
<th>$^{13}$C Chemical shift of $\alpha$-methylene carbon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iodoacetic acid</td>
<td>-6.59</td>
</tr>
<tr>
<td>Iodoacetic acid chloride (45)</td>
<td>+4.12</td>
</tr>
<tr>
<td>Deprotected glycosyl iodoacetamide (40)</td>
<td>-3.58</td>
</tr>
<tr>
<td>Cysteine derivative (50)</td>
<td>35.17</td>
</tr>
<tr>
<td>Glutathione derivative (52)</td>
<td>35.18</td>
</tr>
</tbody>
</table>

Table 4.1: $^{13}$C chemical shifts of the $\alpha$-methylene carbon in glycosyl iodoacetamides and reaction products.

We then considered whether a 100% $^{13}$C label in this position may function as a useful indicator of protein glycosylation. The $^{13}$C labelled GlcNAc iodoacetamide (53) was then prepared from $^{13}$C labelled iodoacetic acid chloride as in Figure 4.9 as only a slight excess of the acid chloride would be required. The use of $^{13}$C labelled iodoacetic anhydride in a reaction with the deprotected glycosyl amine (Figure 4.8) would result in loss of 50% (as iodoacetic acid) of the extremely expensive $^{13}$C labelled material and this was therefore not considered viable.

![Diagram](image)

We used our $^{13}$C labelled iodoacetamide in a test reaction with glutathione and glycosylation on cysteine was clearly observed (Figure 4.11) as we saw the shift in $^{13}$C label from $-3.58$ to $+35.2$ppm.
Figure 4.11: (a) shows the $^{13}$C NMR spectrum of the $^{13}$C labelled iodoacetamide (53) and (b) shows the $^{13}$C NMR spectrum of the corresponding glutathione derivative (52).

We then hoped to use this technique to analyse protein glycosylation at a later stage.
4.8 More Complex Oligosaccharides.

Investigations into the synthesis of the pentasaccharide core tend to favour the retrosynthetic analysis below (Figure 4.12)\textsuperscript{48,120,121}.

![Figure 4.12: Common retrosynthetic analysis of the pentasaccharide of N-linked glycans, P=Ac, P'=Bn, R=SEt.](image)

Key features are the syntheses of the building blocks (57), (58), (59) and (60) or analogs thereof. For example, (57) is ultimately transformed into (58) (R=SEt) by the action of dimethyl dioxirane (DMDO) and ethane thiol\textsuperscript{48} and finally serves as an acceptor for two mannose residues. Building block (57) was chosen as a key intermediate since it is capable of being transformed in GlcNAc, mannose, or glucose\textsuperscript{39,47}, maximising the possibilities for diversity, which may be useful in structure-function glycoprotein studies.

80
Synthesis of Building Blocks: 4-p-Methoxybenzyl-D-Glucal (57).

This work was carried out according to procedures of Danishefsky and co-workers\(^\text{48}\) (Figure 4.13). 3,4,6-tri-\textit{O}-acetyl-D-glucal (61) was deacetylated with sodium methoxide in methanol. The crude product was treated with \textit{p}-anisaldehyde and anhydrous \textit{ZnCl}_2 to afford the 4,6-\textit{p}-methoxybenzylidene acetal (62) in 20% yield. The 'Butyldimethylsilyl group (TBDMS) was efficiently installed using standard conditions and then the benzylidene acetal was regioselectively opened onto the 4-position of the glycal using diisobutyl aluminium hydride (DIBAL-H). The TBDMS group was then removed using tetrabutyl ammonium fluoride in high yield to afford (57). The standard method for introducing benzylidene acetals (trans acetalation with \textit{p}-methoxybenzaldehyde dimethyl acetal and \textit{p}-toluene sulfonic acid) was not employed here due to the acid sensitivity of the enol-ether functionality of glycals.

![Chemical reaction diagram](image)

Figure 4.13 (I) NaOMe, MeOH, pMeO\((C_6H_4)CHO, \text{ZnCl}_2, 16\text{h}, 20\%, (II) TBDMS-Cl, imidazole, DCM, 1.5\text{h}, 75\%, (III) DIBAL-H, DCM, -20^\circ\text{C}, 3\text{h}, 80\%, (IV) TBAF, THF, 16\text{h}, 100\%)

Further building blocks (59) and (60) (Figure 4.12) were also prepared using a similar strategy, generally involving the use of benzylidene or \textit{p}-methoxybenzylidene acetals (prepared from the aryl aldehydes dimethyl acetals (65)) which could be regioselectively opened, but this time onto the 6-position\(^\text{141-143}\).
Synthesis of an Analog of Building Block (57) (Figure 4.14).

The peracetylated glycosyl azide (37) was deacetylated and the 4,6-p-methoxybenzylidine acetal was introduced using p-methoxybenzaldehyde dimethyl acetal and p-toluene sulfonic acid (TsOH) in 80% yield\(^\text{141}\). The remaining C-3 hydroxyl group was then benzylated\(^\text{141}\). The low yield of this reaction was accounted for by partial hydrolysis of the benzylidine acetal under the reaction conditions. The p-methoxybenzylidine acetal was then reductively opened onto C-6 in 85% yield based on recovered starting material (actual yield was 58%).

Figure 4.14: (I) NaOMe, MeOH, 2h, then \(p\text{MeO(C}_6\text{H}_4\text{)}\text{CH(OMe)}_2\), pTsOH, DMF, 2h, 50°C, 80%, (II) BnBr, Ba(OH)\(_2\), 3A mol sieves, DMF, 83%, (III) Na(CN)BH\(_3\), TFA, DMF, 16h, 85%

Synthesis of Building Block (59) (Figure 4.15).

For this building block, we chose N-phthalamido protection on C-2 to see whether the yield of reactions (particularly benzylaion at C-3 and reductive acetal opening onto C-6) could be improved. Benzylidene acetals (not p-methoxy) which are cleaved under more strongly reducing conditions can be used in this case as the thioglycoside is stable to reducing conditions (unlike azide) and, if need be, this building block can also be transformed into a glycosyl azide or further glycosyl donors\(^\text{41,123,144,145}\):
The N-phthalimido group was introduced using a well established procedure and the thioethyl glycoside (71) was prepared in good yield (73%) via the Lewis acid mediated exchange of the anomeric acetyl group with ethane thiol. Introduction of the benzylidene acetal was carried out essentially as before (Figure 4.15).

Unfortunately there was no time to complete the synthesis of the pentasaccharide core but we had managed to prepare smaller model compounds (such as GlcNAc iodoacetamide) and the GlcNAc derived $^{13}$C labelled iodoacetamide. These were taken forward to attempt protein glycosylation and the results are disclosed in the following chapter. We expect that there should be no particular problem in assembling the pentasaccharide core considering that many routes have now been published. The most likely progression is described below.
CHAPTER 5.

Protein Glycosylation and Determination of the Glycosylation Site.
Protein Glycosylation and Determination of the Glycosylation Site

With refolded proteins in hand, and the ability to observe reactions at the molecular level, we went on to investigate whether or not it would be possible to alkylate the free thiol group. For the first time, we hoped to be able to address some issues in more detail than previously possible:

1) The extent of glycosylation: Is the glycosylation or alkylation reaction quantitative?
2) The specificity of the iodoacetamide group: Are any other amino acid side chains modified?
3) The identity of the glycosylation site: Is the correct cysteine residue modified? (this also tells us whether the correct disulfides were formed during the oxidative refolding step).

In order to achieve these goals a general method for the analysis of glycosylated or alkylated proteins was devised and this is described schematically in Figure 5.1. The strategy involves peptide mapping via peptide mapping using LC-MS.

**Figure 5.1: Schematic representation of our glycosylation and peptide mapping strategy. Refolded protein is glycosylated (green triangle), reduced and then cleaved into smaller peptide fragments. Analysis of the peptides reveals the position of the glycosylation site as only the modified peptide will be shifted in molecular weight and have an increased mass corresponding to the alkylating agent.**
Given that we could determine that glycosylation had taken place, how would we establish the position of the glycosylation site? In order to determine the nature of the glycosylation site we needed to employ reliable methods for the cleavage of the peptide backbone. Fragmentation of the glycoprotein into peptide fragments would allow us to map the position of the modified amino acid residue. Classical methods for the chemical and enzymatic cleavage of proteins have been extensively reviewed\textsuperscript{149-153}. In the following section, only the most frequently employed enzymatic methods for protein cleavage (associated with peptide mapping) will be discussed. Additionally, only the most relevant chemical methods for peptide cleavage (at amino acid residues containing sulfur) will be discussed.

5.2 Enzymatic Methods for Peptide Fragmentation.
The most commonly employed proteolytic enzymes in peptide mapping are trypsin, α-chymotrypsin and V8 protease from \textit{Staphylococcus aureus}\textsuperscript{152}. All three are serine proteases and cleave peptide bonds using a conserved active site motif known as the catalytic triad (Figure 5.2), named as such because the peptide hydrolysis is catalysed by conserved aspartate, histidine and serine side chains. The second fragment is released as the acyl enzyme intermediate is hydrolysed\textsuperscript{119}.

![Mechanism of hydrolysis of peptides by chymotrypsin](image)

\textbf{Figure 5.2: Mechanism of hydrolysis of peptides by chymotrypsin}\textsuperscript{119}. 

88
CHAPTER 5

These proteases however tend not to hydrolyse the peptide indiscriminately and instead cleave the peptide backbone adjacent to specific amino acid residues. This selectivity is dictated by the 3D shape, size, and surface charge within the active site:

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Selective for</th>
<th>Structural requirement</th>
<th>Residues cleaved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsin</td>
<td>Basic side-chains</td>
<td>Protonated amino group (positive charge)</td>
<td>Arg/ Lys</td>
</tr>
<tr>
<td>α-Chymotrypsin</td>
<td>Hydrophobic side-chains</td>
<td>Aromatics</td>
<td>Trp/ Tyr/ Phe</td>
</tr>
<tr>
<td>V8 (Glu-C)</td>
<td>Acidic side-chains</td>
<td>Deprotonated acids (negative charge)</td>
<td>Glu/ Asp</td>
</tr>
</tbody>
</table>

Table 5.1: Common serine proteases and their specificities.

Generally, the protease is used in 10:1-100:1 protein/protease ratios at 37°C

5.3 Chemical Methods for Peptide Mapping.

The nucleophilicity of sulfur, which results from it’s ability to be easily polarised through it’s weakly held outer electrons, makes it an ideal target for selective chemistry amongst the plethora of functional groups presented by proteins. Performed usually at pH8, where amines are protonated and less nucleophillic, it is relatively simple to conduct selective chemical reactions at amino acids containing sulfur. Since cysteine and methionine residues are also relatively rare, the number of fragments is low.

5.3.1 2-Nitro-5-thiocyanatobenzoic Acid (NTCBA).

NTCBA (12, see chapter 3) works by providing a source of CN and once protein cyanlation occurs (selectively on cysteine residues). Cyanylation of the thiol group induces an intramolecular cyclisation reaction under basic conditions (pH 10) aided by neighbouring group participation resulting in cleavage of the peptide backbone at cysteine residues.
Figure 5.3: Proposed mechanism of NTCBA induced peptide backbone cleavage.

The C-terminal fragment is modified at its N-terminus where the nitrile group is incorporated as part of the iminothiazolidine (ITZ) heterocycle.

5.3.2 Cyanogen Bromide (CNBr): Cleavage at Methionine.

Cyanogen bromide has for a long time been considered the most reliable chemical method protein fragmentation with cleavage occurring quantitatively at methionine residues. As above, cleavage occurs through neighbouring group participation (Figure 5.4). As methionine is a relatively rarely occurring amino acid, it is effective in providing few peptide fragments and has been utilised in some cases to remove histidine tags from recombinant proteins, a procedure normally employing the use of enzymes (thrombin or factor Xa).

Figure 5.4: Neighbouring group participation in the cleavage of the peptide backbone at methionine by CNBr.
5.4 Glycosylation of His₁₀-EPO Samples

With refolded EPO samples in hand we went on to study site specific glycosylation.

5.4.1 Analysis of Glycoprotein Products.

As a result of developing an on-line LC-MS method for glycoprotein analysis, we aimed to use this to monitor the progress of the glycosylation reaction by removing aliquots at timed intervals. As with previous samples, good MS data was only acquired when the protein samples were reduced prior to on-line LC-MS, and complimenting previous reactions with alkylating reagents, the reaction products were only 'clean' when the alkylating reagent was removed prior to reduction.

For analytical purposes, aliquots were precipitated by the addition of 10 volumes of 1:1 acetone/methanol (where buffer salts and alkylating reagents remain in solution) and stored at -20°C for at least ten minutes. Precipitated protein was then collected by centrifugation and redissolved in 10-20μL of 6M guanidine.HCl, 50mM tris.HCl; pH8 and treated with DTT to a final concentration of 10mM. Reduced samples were then analysed using on-line LC-MS as described in chapter 3.

5.4.2 The Glycosylation Reaction.

In an effort to rapidly assess the refolding process (described in chapter 3) for it's ability to reform the disulfide bonds the first glycosylation reactions were crude and designed purely to assess the number of free thiol groups (predicted to be one). Refolded protein samples were concentrated by passing the dilute refolded protein solution over a 1.0ml bed of His-bind resin (Novagen). Immobilised protein samples were eluted in the minimum volume (3.0ml) of refolding buffer containing 500mM imidazole (3.0ml). The protein was precipitated using acetone/methanol as
previously described and redissolved in 100μL of alkylation buffer (6M guanidine.HCl, 500mM Tris; pH 8, 20mM EDTA). GlicNAc iodoacetamide was then added from a 100mM stock solution (in water, alkylation buffer, or DMF) to a final concentration of 10mM. Aliquots were removed from the reaction mixture at 24h intervals and analysed by on-line LC-MS (Figure 5.5) as described in chapter 3.

**Figure 5.5:** Glycosylation of refolded His10-C38hEPO at t=1h (a) and t=24h (b). At t=1h, the charge envelope is typical of the purified protein. At t=24h the ESI spectrum has changed and each peak associated with a particular charge state is a doublet, indicating that the reaction is proceeding as two species are now visible.

After 24h, the deconvoluted charge envelopes showed clearly (Figure 5.6) how the glycosylation reaction had proceeded. The results are summarised below in Table 5.2:

<table>
<thead>
<tr>
<th>Peak</th>
<th>Calculated Average Mass</th>
<th>Observed mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>(His10-C38EPO) 20905.89</td>
<td>20903.5</td>
</tr>
<tr>
<td>M2</td>
<td>(M1+ GlicNAc) 21166.89</td>
<td>21163.0</td>
</tr>
<tr>
<td>M3</td>
<td>(M2+ 2GlicNAc) 21427.89</td>
<td>21424.5</td>
</tr>
</tbody>
</table>

**Table 5.2:** Calculated and observed masses for His10-C38EPO and glycosylated.

**Figure 5.6:** deconvoluted charge spectrum.
The analysis of the MS data showed that the reaction proceeded in approximately 30% yield. Surprisingly, it was also clear that a second GlcNAc residue was binding. Logically, we would have predicted the addition of a single GlcNAc residue resulting from one free thiol (and 2 formed disulfides) or the addition of three GlcNAc residues resulting from three free thiols (and one formed disulfide) and so forth, so the addition of two GlcNAc residues was quite unpredicted (compare Figure 3.9 however). Furthermore, the longer the reaction proceeded (72h), the more diglycosylation was observed. These results were highly reproducible on C38 and C24 EPO mutants but glycosylation reactions conducted on His₁₀-C38EPO were generally higher yielding (more glycosylated product was observed) than those conducted on the C24 mutant.

5.4.3 NTCBA Analysis of Glycoproteins.
The addition of two GlcNAc residues was clearly puzzling and we resolved to analyse the reduced, glycosylated protein with NTCBA. We analysed all proteins (non-glycosylated and glycosylated) and found that they were cleaved quantitatively to give MS data in good agreement with calculated values. This also proved a good general method (non-molecular biological) for proving the identity of the protein or position of the mutation by simply observing the sizes of the fragments.

<table>
<thead>
<tr>
<th>Fragment masses for His₁₀-C38</th>
<th>NTCBA Fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G(-21)-I6</td>
</tr>
<tr>
<td>Calculated</td>
<td>3186.5</td>
</tr>
<tr>
<td>Observed</td>
<td>3186.0</td>
</tr>
</tbody>
</table>

Table 5.3: Calculated and observed masses for the peptide fragments resulting from NTCBA cleavage of His₁₀-C38hEPO. Masses under 500Da were not observed as they lay outside the acquisition range (500-2000Da). n/o = not observed

The partially glycosylated C38-mutant was reduced and gave fragments in good agreement with the calculated values (Table 5.3). If the protein had however been glycosylated on C38, the resulting sulfide would not have been cleaved allowing a larger peptide (C33-A160+GlcNAc, Mwt=14412.3Da) fragment to be observed and
this was the case (Figure 5.7). This fragment (14249.0 Da) did not correspond to the mass of our glycosylated fragment (expected molecular weight of 14412.3) and was in fact some 163 mass units smaller. Neither did it correspond to peptides resulting from incomplete cleavage which would be expected at 14282.56Da (from missed cleavage at C161), and 14140.36Da (missed cleavage at C33).

We considered a number of potential mechanisms whereby the glycoprotein linkage may be modified or destroyed. Although the larger fragment was not observed when NTCBA treated unglycosylated protein was analysed, it was observed when iodoacetamide treated C38 was NTCBA treated and analysed.

Alternatively, it may be possible that the sulfide was cyanylated and eliminated under the basic reaction conditions (below). Such β-elimination would not be possible for homocysteine or methionine due to the increased chain length and even if they were to become cyanylated under the reaction conditions, the cyano group is likely to be hydrolysed and lost in the following step (pH10; 37°C). The observed fragment size may well turn out to be independent of the glycosylating reagent under these conditions.
From NTCBA analysis of the partially glycosylated C38 mutant we did however see the mass corresponding to the addition of \( \text{GlcNAc} \), but to the N terminal region of the protein (from the beginning of the his-tag (G-21) to EPO residue 6). This was unexpected and may have arisen from modification of the \( \alpha \)-amino group (N-terminus) or the polyhistidine tag. Additionally, when the reaction was carried out in the presence of 5mM L-histidine there was no observable difference in the product distribution. Although glycosylation of the his-tag was undesirable in that it made it difficult to assess the degree of cysteine modification, this was not considered to be too problematic as our long term goal was to remove the entire his-tag at the end of the synthesis.

5.4.4 \( ^1\text{H}^{13}\text{C} \) HSQC NMR Spectrum of Partially Glycosylated His\(_{10}\)-C38hEPO. The carbon-13 labelled glycosyl iodoacetamide (53) prepared in the previous chapter was then employed to shed light on the glycosylation reaction.

Figure 5.8: N-terminal modification of the C38 mutant by GlcNAc iodoacetamide.

Figure 5.9: The \( ^1\text{H}^{13}\text{C} \) HSQC spectrum shows glycosylation on both nitrogen and sulfur.
This spectrum clearly demonstrated that glycosylation was also occurring on cysteine (\(^{13}\)C chemical shift of approximately 35ppm).

From NTCBA analysis of the C38 mutant (glycosylation of the C24 mutant was so poor that we thought it of no use in NTCBA analysis) we could see that perhaps glycosylation of C38 was occurring but glycosylation of the N-terminus or polyhistidine tag was occurring also. We could not assign the mass of the larger peptide observed when the partially glycosylated C38 mutant was digested with NTCBA and this method may have caused destruction of the sugar-peptide linkage. NTCBA, and consequently CNBr analysis were considered too harsh for our glycoproteins. We could clearly observe that the reaction was also not proceeding to completion and hypothesized that perhaps rapid protein dimerisation (resulting from the relatively high protein concentration in the reaction) may be competing with slow (relative to acetamidation) glycosylation (Figure 5.10):

![Figure 5.10: Protein dimerisation may possibly compete with sluggish protein glycosylation.](image)

This would help explain our observations as dimerisation rapidly quenches thiol groups leaving them unreactive to GlcNAc iodoacetamide and leaving only N-terminal amino groups free to react. Of course, dimers were not observed in the mass spectrum as a result of sample reduction prior to LC-MS to allow us to acquire the best data. Dimers and higher oligomers could however be observed by non-reducing SDS-PAGE. We decided to repeat this reaction but this time with a high initial concentration

![Figure 5.11: incubation of C38EPO with 50mM GlcNAc results in increased glycosylation.](image)
of the iodoacetamide, adding GlcNAc iodoacetamide to a final concentration of 50mM (Figure 5.11). On this occasion we saw far greater consumption of the starting material but also di and tri glycosylation. This result was inconclusive since although more EPO had been consumed we could not be sure that this increased glycosylation was on cysteine with concomitant increased glycosylation of the N-terminus or on the N-terminus alone. Glycosylation of the N-terminus results in an amine, potentially more reactive than the offending N-terminus which could react further with excess iodoacetamide to afford a quaternary ammonium species (Figure 5.12)

![Figure 5.12: Potentially, three GlcNAc residues could add to the amino terminus.](image)

5.4.5 Proteolytic Digests of Glycoprotein.
In order to investigate the glycosylation further and under mild conditions we then digested the glycosylated proteins with trypsin, α-chymotrypsin and V8 protease. A complication was that concentrated EPO samples were only soluble in guanidine.HCl at concentrations no lower than 3M and one concern was that these conditions would be too denaturing to allow the proteases to function properly. Glycoprotein samples were precipitated from reaction mixtures and redissolved in 25μL of 6M guanidine.HCl, 100mM tris; pH8.0 and then diluted to 50μL with water. The
enzymes were added to concentrations of 10% from 10mgml\(^{-1}\) stock solutions and incubated at 37°C for 1-16h.

5.4.6 Analysing the Digestion Reaction.

i) MALDI-TOF Analysis of Protease Digestion.

The simplest method for observing digestion was to desalt the reaction using C4 reverse phase HPLC and analyse by MALDI-TOF MS since MALDI allows us to see all fragments as singly charged species on the same spectrum. 10 minute fractions (at 50\(\mu\)Lmin\(^{-1}\)=0.5ml) were collected, lyophilised, and redissolved in 70% acetonitrile/water, 0.1% TFA and analysed. All fractions, except the first five minutes, were analysed since the first five minutes allowed for the elution of the salts and could not be fed into the mass spectrometer. Test reactions were carried out on unglycosylated protein samples. Trypsin was best at cutting EPO (Figure 5.13) followed by chymotrypsin (leaving much of the sample uncut) and finally V8 protease which failed to cut the protein in 3M guanidine.HCl.

![MALDI-TOF mass spectrum of trypsin cleaved His\(_{10}\)-WThEPO. These peptides are all in one lyophilised fraction.](image)

**Figure 5.13:** MALDI-TOF mass spectrum of trypsin cleaved His\(_{10}\)-WThEPO. These peptides are all in one lyophilised fraction.
<table>
<thead>
<tr>
<th>Peak</th>
<th>Mass (calc)</th>
<th>Mass (obs)</th>
<th>Peptide sequence</th>
<th>Peak</th>
<th>Mass (calc)</th>
<th>Mass (obs)</th>
<th>Peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1314.73</td>
<td>1315.53</td>
<td>162-171</td>
<td>8</td>
<td>3729.32</td>
<td>3728.77</td>
<td>98-131</td>
</tr>
<tr>
<td>2</td>
<td>1442.83</td>
<td>1444.29</td>
<td>161-171</td>
<td>9</td>
<td>4099.75</td>
<td>4100.71</td>
<td>153-187</td>
</tr>
<tr>
<td>3</td>
<td>2349.74</td>
<td>2351.22</td>
<td>153-171</td>
<td>10</td>
<td>4270.97</td>
<td>4272.38</td>
<td>125-164</td>
</tr>
<tr>
<td>4</td>
<td>2820.26</td>
<td>2821.96</td>
<td>125-152</td>
<td>11</td>
<td>4366.03</td>
<td>4367.16</td>
<td>132-171</td>
</tr>
<tr>
<td>5</td>
<td>2942.59</td>
<td>2944.19</td>
<td>98-124</td>
<td>12</td>
<td>6116.04</td>
<td>6118.82</td>
<td>132-171</td>
</tr>
<tr>
<td>6</td>
<td>3191.73</td>
<td>3195.73</td>
<td>161-187</td>
<td>13</td>
<td>6900.99</td>
<td>6899.09</td>
<td>125-187</td>
</tr>
<tr>
<td>7</td>
<td>3483.91</td>
<td>3486.53</td>
<td>132-164</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Calculated and observed masses of tryptically digested WThEPO. The numbering of the peptide sequences includes amino acids 1 to 21 of the his-tag.

From the results for trypsin above, only half of the protein could be observed (residues 98 onwards, EPO residues 77-end) and at no stage with either trypsin or chymotrypsin could the N-terminal half of the protein be observed.

**ii) LC-ESI Analysis of Protease Digestion.**

Using the on-line LC-MS method developed in chapter 3 we also analysed the trypsin and chymotrypsin digested protein but again only fragments corresponding to the C-terminal end of the gene could be observed.

From the results with His₁₀-WThEPO and the partially glycosylated C38 mutant we could also see that glycosylation was not occurring on this portion of the protein. *No glycosylation was observed on C7 or C161.* We assumed that, due to the high polarity (Figure 5.14) of the N-terminus and the small size of the proteolytic fragments (proteolytic sites tend to be clustered around this end of the protein), the N-terminal fragments may have been eluting from the C₁₈ reverse phase column with the salt and any highly polar impurities (within the first five minutes) and were not run through the mass spectrometer.
iii) On Column Protease Digestion.

Since we could not observe the N-terminal region of the protein by direct proteolysis of the reduced protein we decided to sequester this region by first immobilising the protein on a metal affinity column. Reduced protein samples were separated from the reducing agent by precipitation and redissolved in 6M guanidine and loaded onto the column. The immobilised protein was then equilibrated with 3M guanidine, 50mM tris; pH8.0 for tryptic digests (pH7.3 for chymotryptic digests) and exposed to solutions (0.1mg-ml\(^{-1}\)-1.0mgml\(^{-1}\)) of the required protease for 1h. Elution of the protease solution was prevented by plugging the end of the column. After one hour the plug was removed and unbound fragments were washed from the column and bound fragments were eluted as in Figure 5.15.
His$_{10}$-tagged protein sample is immobilised on His-bind resin

1 mg/ml protease (RT, 1 h)

serine protease cleavage

$\bullet$ = Trypsin

allow to drain and wash unbound fragments

elute bound fragments with imidazole

Analyse mixture of fragments Using LC-ESI-MS or MALDI-TOF-MS

Figure 5.15: General protocol for on column protease digestion of His$_{10}$-EPO samples.
CHAPTER 5

Eluted fractions were lyophilised and taken up in DCM (500µL) to dissolve imidazole and the precipitated peptide was collected by centrifugation. This process was repeated and the peptide was dissolved in 1:1 acetonitrile/water prior to on-line LC-MS.

5.4.7 On-Line LCMS Analysis of on Column ProteolyticDigests.

i) Trypsin.

With trypsin as the protease, and His₁₀-WThEPO as the test protein, we could see that only one fragment had remained bound to the resin, the his-tag (sequence=GHHHHHHHHHHSSGHIEGR, Figure 5.16). Even when the enzyme solution (0.1mgml⁻¹) was allowed to flow through the column uninhibited, this was the only isolated fragment and perhaps indicates that this arginine residue (-2R) is relatively exposed.

\[
\text{DHCHWTT } 48 \text{ (3.306) Sm (Mn, 2x0.80); Sb (5.30.00); Cm (45:50-(50:68+38:44)) 1.88e5}
\]

Figure 5.16: The charge envelope of the only isolable species from an on column tryptic digest shows the 2+, 3+ and 4+ ions of the his-tag (Mwt=2272).

ii) Chymotrypsin.

The on-column cleavage was then attempted with chymotrypsin as this had been less efficient at cutting EPO samples in solution. This time two fragments were shown to remain bound to the column after washing (Figure 5.17). As above, the first fragment (retention time=24mins) was resolved and was the N-terminal fragment once again.

Figure 5.17: TIC of on-resin-proteolysis fragments.
cleaving at the first possible cleavage site from the N-terminus (Peak 1, -21G-Y15). The second fragment (Peak 2, -21G-Y49) was larger and was cleaved at the next tyrosine residue along, containing cysteine residues 29, 33 and 38 (Figure 5.18). This isolable fragment was ideal for further study:

Gly His His His His His His Ser Ser Gly His Ile Glu Gly Arg His Met Ala Pro Pro Arg Leu Ile Cys Asp Ser Arg Val Leu Glu Arg Tyr, Leu Leu Glu Ala Lys Glu Ala Glu Asn Ile Thr Thr Gly Cys Ala Glu His Cys Ser Leu Asn Glu Asn Ile Thr Val Pro Asp Thr Lys Val Asn Phe Tyr

\[ M_{wt_{PK1}} = 4308.77 \quad M_{wt_{PK2}} = 8058.95 \]

Figure 5.18: (a) the charge spectrum of peak 1 (Fig. 5.17) and (b) charge spectrum of peak 2.

As we could now observe and purify the N-terminal region we went on conduct the same procedure on the glycosylated C38 mutant. Again we could see that this section of the protein had been mono and di-glycosylated (Figure 5.19).
Figure 5.19: Glycosylation of the sequestered N-terminal region.

Furthermore, these N-terminal fragments were significantly more soluble than the intact protein and could be dissolved in 50mM NH$_4$HCO$_3$; pH8 and digested further with trypsin or V8 protease to cause further fragmentation (Figure 5.20).

**Table 5.5:** Calculated and observed masses for partially glycosylated His$_{10}$-C38EPO, digested on Ni$^{2+}$ affinity resin with chymotrypsin.

<table>
<thead>
<tr>
<th>Peak</th>
<th>Average Mass</th>
<th>Observed m/z</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>8047.9</td>
<td>8042.6</td>
<td>5.3</td>
</tr>
<tr>
<td>M2</td>
<td>8308.9</td>
<td>8301.3</td>
<td>7.6</td>
</tr>
<tr>
<td>M3</td>
<td>8569.9</td>
<td>8560.5</td>
<td>9.4</td>
</tr>
</tbody>
</table>

[(M1+GlcNAc)]

**Figure 5.20:** Evidence for mono, and a small portion of di-glycosylation on the N-terminal fragment of His$_{10}$-C38hEPO after further digestion with V8 protease.
The fragment mixture was then digested further with V8 protease and we could see
the N-terminus and could also observe the addition of as many as two GlcNAc
residues (Figure 5.20) to the N-terminal fragment. Additionally we could observe
further fragments (Figure 5.21) and see that glycosylation was also occurring on the
fragment which incorporated C38.

<table>
<thead>
<tr>
<th>Peak</th>
<th>Observed m/z</th>
<th>Corresponding sequence</th>
<th>Calculated average m/z</th>
<th>Calculated based on observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2112.3</td>
<td>NITTGCAEHCSLNECITVPD</td>
<td>2121.38</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>2316.6</td>
<td>AENITTGCAEHCSLNECITVPD</td>
<td>2321.57</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>2374.0</td>
<td>1 + GlcNAc</td>
<td>2382.38</td>
<td>2373.3</td>
</tr>
<tr>
<td>4</td>
<td>2577.0</td>
<td>2 + GlcNAc</td>
<td>2582.57</td>
<td>2577.6</td>
</tr>
</tbody>
</table>

* The smaller equidistant peaks are +23 Na adducts.

Figure 5.21: Glycosylated C38EPO fragments. Peaks 3 and 4 correspond to glycosylated 1 and 2 respectively

Unfortunately however, the fragments also contained two further cysteine residues
from the native C29-C33 disulfide bond (red). The enzyme had failed to cleave at
the glutamate residues (underlined) between cysteine residues. **This proved that
glycosylation was occurring on the correct fragment.**
A further approach was to digest the immobilised 8KDa chymotryptic fragments (Figure 5.19) with trypsin. This gave smaller fragments and, as with the V8 protease digestion, the fragment of interest was also home to the two cysteine residues of the native disulfide bond. We could then attempt to sequence this peptide in the collision cell of the Q-TOF mass spectrometer. This work was carried out in collaboration with Dr Adam Gouldsworthy.

From the digested mixture we could see that the proteolysis had occurred (Figure 5.22). Through MS-MS we were able to transmit a single ion into the collision cell. Ions transmitted (892.66 and 979.68) were the 3+ ions corresponding to the peptide below and its glycosylated peptide analog respectively:

Glu Ala Glu Asn Ile Thr Thr Gly Cys Ala Glu His Cys Ser Leu Asn Glu Cys Ile Thr Val Pro Asp Thr Lys

We then aimed to sequence this fragment by collision induced fragmentation in the collision cell. Although the original 3+ ion peaks were small, it did not imply that
they were only fractionally present. More intense peaks were also sequenced and were shown to be from the same peptide indicating that this peptide may only ionise poorly.

When both ions were fragmented they could be partially sequenced.

![Figure 5.23: Collision induced fragmentation of the glycosylated and non-glycosylated small tryptic peptides.](image)

Considering that the fragmentation was induced from two ions thought to have the same sequence it was not surprising, that on fragmentation, many of the same ions were observed (Figure 5.23). Unfortunately either peptide could only be sequenced as far along as the isoleucine residue adjacent to the cysteine residue of interest (sequence obtained=Ile Thr Val Pro Asp Thr Lys). In the case of the glycosylated peptide, it showed a m/z peak at 204.08Da characteristic of an N-acetyl hexosamine (Figure 5.23):
Interestingly, there was also a mass difference observed when transmitting masses into the collision cell. The observed masses differed from the theoretical isotope distribution by two to three mass units (Figure 5.24) indicating that perhaps a disulfide bond had been reformed during the proteolytic digest (perhaps the native C29-C33).

Figure 5.24: Theoretical and observed isotope profiles for the 980(+3) ion for the glycosylated tryptic fragment.

5.5 Optimisation of the Glycosylation Reaction.

Throughout our previous crude glycosylation reactions we noted that the protein was not being glycosylated quantitatively and considered that it was unlikely that dimerisation may be occurring solely as a result of the high protein concentrations.

Believing that the precipitation step prior to previous glycosylation reactions may be having a detrimental effect on the reaction we devised a new protocol (Figure 5.25) which avoided this step.
Oxidative refolding of dilute, reduced protein

Concentrate refolded protein on His-Bind resin (1.0 ml)

Elute from column with 6M guanidine.HCl, 50mM tris; pH 8, 500mM imidazole (3.0 ml), followed with 6M guanidine.HCl, 50mM tris; pH 8 (2.0 ml)

Add GlcNAc iodoacetamide to 10mM and incubate at room temperature (24-48h) on blood rotator.

Figure 5.25: Optimised protocol for protein glycosylation.
CHAPTER 5

This dramatically changed the glycosylation profile (Figure 5.26) with mono-glycosylation proceeding to approximately 60%. Surprisingly there was little di-glycosylation observed after as much as 48h.

Reactions conducted under these conditions were again highly reproducible. Unfortunately no considerable improvement in the yield of the reaction of refolded His-10C24EPO with GlcNAc iodoacetamide was observed under these conditions. Non-reducing SDS-PAGE showed that after the refolding of His10-C24hEPO some soluble aggregates were still present but it was difficult to establish whether reactions proceeded in poor yield due to aggregation as a result of mis-folding, as a result of competitive dimerisation or whether the thiol group was less accessible. Interestingly, in the case of His10-C38hEPO, when the protein was eluted from the column in refolding buffer (2% N-lauroyl sarcosine, 50mM Tris; pH8.0, 40μM CuSO4, 400mM imidazole) instead of 6M guanidine hydrochloride and treated with GlcNAc iodoacetamide, the degree of glycosylation was very poor. Glycosylation was also not observed when a 10mM solution of GlcNAc iodoacetamide was passed over the affinity column immobilised His10-C38hEPO.

This clearly demonstrated that sample preparation played a key role in the efficiency of the subsequent glycosylation and that the more dilute the protein sample was, the less likely it was to aggregate and the higher the yield of the glycosylation.

Following on from this, it would be logical to assume that conducting the reaction at even lower protein concentrations may result in even higher yields of glycosylated product and preliminary results show that this is the case. At lower protein
concentrations however, it becomes more difficult to recover the protein from the reaction mixture for monitoring as it tends not to precipitate and LC-MS data is presently poor. One drawback from this is that more glycosyl iodoacetamide is required to achieve the same concentration and this may not be trivial for more complex saccharides.

It was also possible to react the glycosyl iodoacetamide derived from maltose with refolded His$_{10}$-C38hEPO (Figure 5.27). Glycosylation of the C38 mutant afforded a product with a molecular weight of 21282Da (+380) which corresponded to the addition of maltose. Additionally, a peak at 20958 was observed. This may have resulted from reaction of a small percentage of iodoacetic acid or iodoacetamide (emphasising the need for complete removal of ammonium bicarbonate prior to treatment with iodoacetic anhydride).

Clearly there are still significant obstacles to overcome if we are to unambiguously assign the position of the glycosylation site when working with the C24 and C38 mutants. In order to fully determine the position of the GlcNAc residue on the glycosylated fragment. These preliminary experiments should be scaled up so that we can obtain sufficient quantities of smaller peptide fragments (from on-column chymotryptic digests or after further digests with V8 protease or trypsin). If we could isolate sufficient quantities of each fragment using metal affinity, lectin affinity, or reverse phase chromatography then we may be able to sequence them or look at them by NMR spectroscopy. This will allow unambiguous assignment of the glycosylation site.

Perhaps, for initial studies, the pET16b-N83ChEPO construct should be prepared. Standard tryptic digests of the whole glycoprotein in solutions of 3M guanidine have
always been capable of revealing peptides relating to this end of the protein (Figure 5.13, Table 5.4) and glycopeptides resulting from glycosylation on C83 should be far simpler to observe than those derived from the C24 or C38hEPO mutants by online LC-MS or MALDI-TOF MS.
CHAPTER 6.

Preparation of the pET16b-C83hEPO(COPT) Vector, Expression and Purification of His$_{10}$-C83hEPO.
6.1 Preparation of pET16b-C83EPO, Using the Megaprimer PCR Method.

Due to difficulties with the unambiguous determination of the C24 and C38 EPO mutant glycosylation site we chose to synthesise the C83 mutant hoping that glycosylation at this position would be more readily observed. The C83 mutant was prepared using a Polymerase chain reaction (PCR) based method.

The Polymerase Chain Reaction was invented by K.B. Mullis\(^{155}\) and it revolutionised genetics by providing a simple method for the amplification of predetermined DNA sequences using thermostable DNA polymerases isolated from organisms thriving in the harsh conditions of hot-springs (\textit{Thermus aquaticus}) and deep hydrothermal vents (\textit{Pyroccoccus} species for example)\(^{156,157}\).

Since its invention, many variants have been reported which allow for the introduction of mutagenic sequences at predetermined sites (site directed mutagenesis) as well as DNA amplification. For the synthesis of the pET16b-C83EPO mutant we chose the 'megaprimer' PCR method (Figure 6.1)\(^{158-159}\). Using this technique the C83 mutant could be prepared in two (or three) rounds of PCR. A third step is only usually employed if the yield of the desired PCR product is low and it needs to be amplified further.

![Figure 6.1: The megaprimer PCR method. A mutagenic primer and flanking reverse primer are initially employed to amplify part of the desired gene. Next, the megaprimer (bearing the mutation) and another flanking primer synthesise the remainder of the gene (PCR2). The whole sequence may then be amplified using both flanking primers.](image-url)
First, pET16b-WThEPO(COPT) prepared in chapter 2 was cut NcoI/HinDIII and ligated into NcoI/ HinDIII cut pUC18-RBS-BioB (a high copy number plasmid) which was kindly provided by Dr. Dominic Campopiano (University of Edinburgh). This new plasmid termed pUC18-RBS-WThEPO would be our template for the PCR reactions and was cloned and the DNA was purified.

A synthetic mutagenic primer (termed N83C), containing a TGC (Cys) mis-matched codon in place of AAC (Asn) was then obtained desalted and deprotected:

\[
\text{GCC CTG TTG GTC TGC TCT TCC CAG CCG T (N83C)}
\]

This primer was chosen as it had already proved successful in the previous (though unintentional) synthesis of the C38/83EPO double mutant\(^\text{103}\).

The three rounds of PCR were conducted using both Taq DNA polymerase (Pharmacia) and deep vent DNA polymerase (New England Biolabs). Using Taq polymerase, the 1\(^\text{st}\) PCR step (synthesis of the megaprimer) was successful but subsequent steps were messy. With Deep Vent polymerase, all steps worked extremely well. This may have been due to the higher fidelity of Deep Vent polymerase, it is known to have proof reading capabilities and does not label the 3\(^\prime\) end of genes with ‘A’ as Taq frequently does. Consequently, only the Deep Vent derived results are discussed here. In the first round of PCR (Figure 6.2), the megaprimer was prepared using N83C as the forward primer and M13 (pUC reverse sequencing primer) as the reverse primer. The PCR product was purified by horizontal agarose electrophoresis (Figure 6.2) which gave and an intense band at approximately 500base pairs (expected size was 550 base pairs).

**Figure 6.2:** Megaprimer synthesis using N83C and M13 primers.
This megaprimer was then purified from the excised gel slice and used as the reverse primer in the second round of PCR. Again pUC18-RBS-WThEPO was used as the template DNA and NR (pUC forward sequencing primer) was employed as the forward primer. In this step we would prepare the remainder of the gene (Figure 6.1).

The second PCR reaction afforded another intense band (Figure 6.3) at approximately 1000 base pairs in size (another, smaller and less intense band was also observed at ca. 750 base pairs). The intense band was again excised and purified from the agarose gel. Next, 5% of this PCR product was used in the third round of PCR (Figure 6.1) using NR as the forward primer and M13 as the reverse to amplify the target DNA further. The remainder was cut with NcoI and HindIII so that it could be re-ligated back into NcoI and HindIII cut pUC18-RBS-BioB (Figure 6.4).

After the overnight endonuclease digestion with NcoI and HindIII a clear and intense band was observed between 750 and 1000 base pairs (Figure 6.4, left). We expected the size of our fragment to be approximately 810 base pairs. Additionally, amplification of the gene (product from 2nd round of PCR) using NR and M13 primers also gave an intense band (Figure 6.4, right) which was purified.
The purified NcoI/ HinDIII cut gene (Figure 6.4, left) was then ligated into pUC18-RBS and transformed into competent TOP10 (Pharmacia) cells and selected on LB-ampicillin plates. Six single transformants were then picked and grown overnight at 37°C. The plasmid DNA from each was then purified and screened to see whether or not the PCR fragment had been incorporated into pUC18-RBS. **Positives were determined using a simple procedure.** Plasmid DNA was cut with *HinD* III (to linearise), and with *NcoI*/*HinDIII* (to cut our PCR derived insert back out) and *HinC* II (should not cut as the mutation had destroyed this unique site). All six colonies were screened and all six were shown to be positives by restriction endonuclease digestion (Interestingly, when *Taq* DNA polymerase was used, only 2 of the six colonies screened were positives). Deep vent derived DNA samples 1 and 2 were then taken forward and sequenced using the M13/NR primers. Both could be fully sequenced and both contained the desired AAC→TGC mutation and no PCR derived errors.

Genes encoding His<sub>10</sub>-C83hEPO were then cut *NcoI/BamHI* out of pUC18-RBS (Figure 6.5) and ligated into *NcoI/BamHI* cut pET16b. The ligation mixture was used to transform competent TOP10 cells as before and positives were determined by cutting the product with *NcoI* and *BamHI* to excise the gene from the plasmid.

**Figure 6.5:** Agarose gel shows (from left) *NcoI/BamHI* cut pUC18-RBS-C83hEPO sample 1, sample 2 (fully sequenced), markers, sample 3, *NcoI/BamHI* cut pET16b and further markers.
6.2 Expression and Purification of His$_{10}$-C83hEPO.
With the newly synthesised pET16b-C83hEPO(COPT) vector in hand we went on to express and purify the gene product from B834(DE3) cells as before (chapter 2). The gene product was purified on a Ni$^{2+}$ affinity column as a single band (Figure 6.6).

![Figure 6.6: Column fractions from Ni$^{2+}$ affinity purified His$_{10}$-C83hEPO.](image)

We hope that this protein may prove particularly useful in further developing the cysteine mutant/glycosyl iodoacetamide neoglycoprotein synthesis methodology.
In B834(DE3) cells, the pET expression system facilitated the efficient expression and purification of His\textsubscript{10}-WThEPO, His\textsubscript{10}-C24hEPO, His\textsubscript{10}-C38hEPO, and His\textsubscript{10}-C38/83hEPO in similar yields of approximately 13mgL\textsuperscript{-1}. In addition to providing us with sufficient quantities of protein, pure enough for further study, there are properties of the His-tag which are particularly suited to our needs. The His-tag provided a handle with which we could manipulate EPO with ease and it did not undergo proteolytic processing allowing us to look at proteolytic digests on resin bound EPO samples. Consequently, it should remain in place until the glycoproteins have been fully characterised.

MALDI-TOF and ESI-MS were invaluable (and essential) techniques for protein characterisation and the study of protein glycosylation allowing us, for the first time, to formulate and test hypotheses regarding protein structure and glycosylation at the molecular level. Through ESI-MS and non-reducing SDS-PAGE we were able to establish that proteins were presented to us by B834(DE3) cells in an oxidised form with formed disulfides. Cysteine mutants, containing an additional thiol, were also obtained as glutathione adducts and were unreactive towards sulphydryl selective reagents. While His\textsubscript{10}-WThEPO was highly monomeric after expression, the cysteine mutants were highly aggregated, forming aggregates with masses greater than 100KDa. These aggregates could however be broken down by the action of reducing agents such as DTT or TCEP, giving us the opportunity to refold our samples under conditions known to re-establish the correct tertiary structure. After refolding, non-reducing SDS-PAGE indicated that all proteins were mainly monomeric with a small percentage of soluble aggregated material remaining. Protein was recovered from the refolding step in >80% yield.

We chose to assess the efficacy of the refolding step after protein glycosylation (through peptide mapping) inferring that the correctly folded structure was obtained if the correct cysteine residue was glycosylated. It was encouraging that per-glycosylation was rarely observed and, if so, only to a small extent.
Protein glycosylation was readily monitored by ESI-MS and was never quantitative. Initially we believed that this resulted from competitive quenching of thiol groups through fast protein dimerisation relative to slow protein glycosylation. It is equally likely that soluble aggregates resulting from a small percentage of mis-folded (and incapable of being glycosylated) samples were carried through and this material was broken down by sample reduction prior to on-line LC-MS and displayed as unglycosylated material. This explained why the addition of huge excesses of glycosylating reagent still failed to consume all the unglycosylated EPO. Consequently, it may be true that size exclusion chromatography will be as useful in separating glycosylated EPO from non-glycosylated EPO as techniques such as lectin-affinity chromatography.

A persistent problem was glycosylation of the N-terminal region of our His$_{10}$-tagged proteins. Ultimately, we aimed to remove the His-tag after our glycoproteins were prepared so the fact that it was occurring was actually less problematic than the fact that it made it difficult to quantify the degree of glycosylation on cysteine. N-terminal modification could however be partially inhibited under our optimised protein glycosylation conditions. The reasons for this are unclear.

Generally, glycosylation of the C38 mutant was far more successful and proceeded in higher yield (approx. 60%) than on the C24 mutant (approx. 20%). The correct fragment was clearly glycosylated (cysteine glycosylation was also intimated by NMR spectroscopy) but unfortunately it also housed the cysteine residues of the C29-C33 disulfide thus unquestionable assignment of the glycosylation site was hindered. From MS data we could also determine that glycosylation had not occurred on C7 or C161 of EPO (the cysteine residues which make up the disulfide which connect the C and N termini of EPO).

As a general rule: **The more dilute the protein samples were prior to the glycosylation reaction, the higher the yield of glycosylation.** Glycosylation reactions carried out at very high dilutions (1µM protein) appeared to proceed quantitatively or near quantitatively by on-line LC-MS analysis. Unfortunately, the
lower the protein concentration was, the harder it was to recover it from the reaction mixture (for reaction monitoring) by precipitation and the mass spectral data were poor. Other methods of sample concentration for analysis may need to be employed here. The drawback of this is method is that more glycosyl iodoacetamide is required to establish the same concentration and may be more costly for larger glycans.

We were also able to glycosylate C38hEPO with maltosyl iodoacetamide and the glycosyl iodoacetamide of chitobiose (results not shown). When the glycosyl iodoacetamides were prepared from the amines obtained from the saturated ammonium bicarbonate reaction, the complete removal of the salt was key to it’s success.

As more reliable methods for the total synthesis of N-linked glycans become available, the more attractive total synthesis becomes as a realistic method for neoglycoprotein preparation. The products are usually crystalline and easier to characterise and provide further scope for conducting structure activity relationships as unnatural saccharides may be incorporated.

Importantly, the protein glycosylation studies are preliminary and there is much scope for further optimisation. One factor which may need to be addressed further is the protein refolding step. It tends to require huge volumes of buffer (this problem may be solved by pulse refolding) and its efficacy is difficult to assess as the refolding buffers are not compatible with NMR or CD spectroscopy. All steps should be scaled up so that we may be able to isolate large enough quantities of glycosylated proteins and peptide fragments for further sequence analysis or NMR spectroscopy. Then it may be possible to unambiguously assign the position of the glycosylation site. This may also be achieved through the synthesis of the N83C mutant which has a potential glycosylation site where enzymatic proteolysis tends to produce fragments which are more readily observed by on-line LC-MS and the glycosylation site is located far from other cysteine residues.
CHAPTER 7.

Experimental Details.
7.1 Experimental Details For Chapter 2.

General Experimental Details:

pET16b-WThEPO, pET16b-C24hEPO, pET16b-C38hEPO, and pET16b-C38/83hEPO were the gifts of Dr. Karen Sage (formerly of the Dyson Perrins Laboratory, University of Oxford, UK). All E. coli bacterial strains for cloning (JM109, JM101) and expression (B834(DE3), BL21(DE3) and BL21(DE3)PlysS) were the gift of Dr. Dominic Campopiano (University of Edinburgh).

All restriction endonucleases (except Bpu1102 I) and DNA modifying enzymes were obtained from New England Biolabs and were used according to the suppliers instructions. Bpu1102 was obtained from Stratagene. Murine anti-human EPO monoclonal antibody was obtained from Genzyme diagnostics. HRP-conjugated rabbit IgG to mouse IgG was from DAKO. IPTG was from Europa. Guanidine.HCl was obtained from Fluka and 6M solutions were filtered through a 0.45µm filter (from Millipore) prior to use. All other routine laboratory reagents were from Sigma.

Centrifugation of small samples in Eppendorf tubes was carried out at maximum speed in an Eppendorf centrifuge 5410. Larger samples (up to 15.0ml) were centrifuged in a Sanyo MISTRAL 2000R. Large scale (20.0ml - 0.3L) samples were centrifuged in a Sorvall RC 5C plus centrifuge.

Shake flask fermentations were carried out in a New Brunswick Scientific INNOVA 4000 incubator shaker (37°C, 250rpm). Large scale fermentations (10L) were carried out in a New Brunswick Scientific BioFlow 1000 fermenter at 37°C, stirred at 250rpm with an air flow of 2Lmin⁻¹.

Sonication was performed using a Sanyo Soniprep 150 at 50% of maximum power.
Cloned plasmid DNA was purified from 10.0ml overnight cultures using WIZARD™ mini-prep kits (Promega) according to the manufacturer’s instructions.

DNA was purified from excised agarose gel slices using a Prep-A-Gene DNA purification kit (Bio-Rad) according to the manufacturer’s instructions.

PCR reactions were carried out on a Techne cyclogene thermal cycler.

Horizontal agarose electrophoresis was carried out in ANACHEM model H3-SET apparatus at 100V according to the manufacturer’s instructions. 100ml Agarose gels were generally 1% agarose in TAE (0.04M Tris-Acetate, 0.001M EDTA; pH8) buffer containing 4µL of 5mgml⁻¹ ethidium bromide.

Protein electrophoresis was carried out using Bio-Rad PROTEAN II (preparative) and Mini-PROTEAN II (analytical) vertical gel electrophoresis apparatus. Acrylamide resolving gels (15-18%) were generally prepared from 30% acrylamide/ Bis acrylamide stocks and prepared according to the manufacturers instructions. Samples were loaded and gels were ran for approximately 1.5h at 200V, 60mA. SDS-PAGE gels were stained (0.1% w/v Coomassie blue R-250, 40% v/v methanol, 10% v/v acetic acid) for 0.5-1h and destained with several washings of 40% v/v methanol, 10% v/v acetic acid. Destained gels were stored in water prior to drying.

Protein purified via elution from excised slices of SDS-Polyacrylamide gel, were eluted using a BIO-RAD model 422 Electro-Eluter according to the manufacturers instructions.
Bacterial Growth Media.

**LB (Luria - Bertani) medium**

Per Litre:

- Bacto tryptone: 10.5g
- Bacto yeast: 5g
- Bacto yeast extract: 10g
- NaCl: 10g

Adjust pH to 7.5 with NaOH

AUTOCLAVE

**M9 (minimal) medium A**

Per Litre:

- Na$_2$HPO$_4$: 6g
- KH$_2$PO$_4$: 3g
- NaCl: 0.5g
- NH$_4$Cl: 1g

Adjust pH to 7.4

AUTOCLAVE and add:

- 1M MgSO$_4$: 2ml
- 20% glucose: 10ml
- 1M CaCl$_2$: 0.1ml

For $^{15}$N labelling Minimal media was prepared as follows:

**M9 (minimal) medium B**

Per Litre:

- Na$_2$HPO$_4$: 6g
- KH$_2$PO$_4$: 3g
- NaCl: 0.5g

Adjust pH to 7.4

AUTOCLAVE
GENERAL MICROBIOLOGY

Standard techniques were performed as described in Molecular Cloning – A Laboratory Manual (Sambrook, Fritsch and Maniatis, 2nd Edition, 1989, Cold Spring Harbour Press, USA). Standard sterile technique was maintained during all microbiological routine work. Media, glassware and equipment were sterilised by autoclaving at 121°C for 15mins. in an Astell Scientific autoclave. Thermolabile solutions were sterilised by filtration through Sartorious minisart 0.22μm filters. Plates were prepared and media innoculated either on ethanol swabbed surfaces or in a Class II environmental cabinet. Antibiotic stock solutions were stored at -20°C and were used at the following concentrations: 100μgml⁻¹ ampicillin, 50μgml⁻¹ carbenicillin, or 34μgml⁻¹ chloramphenicol.

Preparation of E. coli Competent Cells.

LB medium (10.0ml) was seeded with 10μl of the bacterial host cell line. This was grown at 37°C overnight with shaking (250rpm). An aliquot (100μl) of the overnight culture was used to seed 10ml of fresh LB medium containing 200μl of 1M MgCl₂. This was grown as above until A₆₀₀ = 0.2. The cells were then collected by centrifugation for 15min. at 3000rpm (2700G) at 4°C. The supernatant was then discarded and the cells were resuspended in 4.0ml of fresh, ice-cold transformation buffer (50mM CaCl₂, 10mM Tris; pH8.0). The cells were left on ice for 30min with occasional agitation. The cells were then collected by centrifugation for 15min. at 3000rpm (2700G) at 4°C. The supernatant was then discarded and the cells were resuspended in 400μl of ice-cold transformation buffer (50mM CaCl₂, 10mM Tris; pH8.0) and kept on ice until required.
CHAPTER 7

Transformation of DNA into Competent Cells.
A 100μl aliquot of competent cells was transferred to a sterile 1.5ml eppendorf tube and left on ice for 0.5h. 5μl of the DNA sample was then added, the tube was tapped gently, and then left on ice for a further 0.5h. The sample was then heat shocked for 5min at 37°C. L.B. Medium (1.0ml) was then added and the sample was incubated at 37°C for 1h. The cells were then collected by centrifugation for 15min. at 3000rpm (2700G) at 4°C. The supernatant was then discarded and the cells were resuspended in 100μl of fresh L.B. medium. The cells were then spread to dryness on selective (containing 100μg/ml ampicillin or 50μg/ml carbenicillin) L-agar plates and the plates were grown inverted at 37°C overnight.

Cloning of Plasmid DNA.
A single colony, transformed with one of the pET16b-WT or cysteine mutant containing plasmids was used to seed 10.0ml of fresh LB medium containing the required antibiotic. The culture was then incubated at 37°C with shaking overnight. The cells were then collected by centrifugation at 3000rpm (2700G) at 4°C. Cloned plasmid DNA was then purified using Wizard (Promega) or Quiagen spin column DNA purification kits according to the manufacturers instructions.

MOLECULAR BIOLOGY.

Endonuclease Digestions.

General analytical restriction endonuclease digestion.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid DNA *</td>
<td>5μL</td>
</tr>
<tr>
<td>Restriction enzyme</td>
<td>0.5μL</td>
</tr>
<tr>
<td>10×NEB buffer</td>
<td>2μL</td>
</tr>
<tr>
<td>dH₂O</td>
<td>12.5μL</td>
</tr>
</tbody>
</table>

* plasmid DNA prepared from 10.0ml overnight culture.
Reactions were then incubated at the required temperature (variable depending on enzyme) for 1h and linearized DNA was observed by horizontal agarose electrophoresis.

**Preparative restriction endonuclease digestion.**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid DNA *</td>
<td>30µL</td>
</tr>
<tr>
<td>Restriction enzyme</td>
<td>1µL</td>
</tr>
<tr>
<td>10×NEB buffer</td>
<td>5µL</td>
</tr>
<tr>
<td>dH₂O</td>
<td>14µL (to a final volume of 50µL)</td>
</tr>
</tbody>
</table>

Reactions were then incubated at the required temperature (variable depending on enzyme) for 2-3h and linearized DNA was separated from restriction endonuclease by horizontal agarose electrophoresis. The band containing the required DNA was excised and the DNA was purified from the gel slice using a BIO-RAD Prep-A-Gene kit according to the manufacturer’s instructions.

**Optimisation of Expression by the Incorporation of a New C-Terminal Sequence.**

Oligonucleotides (below) were prepared by GIBCO BRL and were obtained fully deprotected and desalted.

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Orientation</th>
<th>Sequence 5' → 3'</th>
<th>Restriction Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPOCOP 1</td>
<td>Sense</td>
<td>CC TGC CGT ACC GGT GAC CGT TAA TGA G</td>
<td>BamHI / StuI</td>
</tr>
<tr>
<td>EPOCOP 2</td>
<td>Antisense</td>
<td>GA TCC TCA TTA ACG GTC ACC GGT ACG GCA GG</td>
<td>BamHI / StuI</td>
</tr>
</tbody>
</table>

**Annealing Oligonucleotides.**

Equimolar amounts of oligonucleotide EPOCOP 1 and EPOCOP 2 were mixed and diluted to 200µL with sterile water (final concentration = 19.1µM) The mixture was then heated in a boiling water bath for 5min. and allowed to cool to room temperature
**Ligation of annealed oligonucleotides into linearised plasmid DNA.**

Ligation reactions were prepared as follows:

<table>
<thead>
<tr>
<th></th>
<th>Cut DNA</th>
<th>Synthetic Duplex</th>
<th>10×ligase buffer</th>
<th>T4 ligase</th>
<th>dH₂O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction</td>
<td>5µL</td>
<td>11µL</td>
<td>2µL</td>
<td>2µL</td>
<td>-</td>
</tr>
<tr>
<td>Control</td>
<td>5µL</td>
<td>-</td>
<td>2µL</td>
<td>2µL</td>
<td>11µL</td>
</tr>
</tbody>
</table>

The reaction mixtures were allowed to stand at room temperature overnight and then used to transform competent JM109 cells. The transformed cells were plated out onto selective LB-agar plates containing 100µg/ml ampicillin. The purified plasmid DNA from individual colonies was screened to confirm the presence of the oligonucleotide insert and positives were sequenced across the insert region using the pET system reverse primer (NOVAGEN).

**DNA Sequencing.**

Sequencing of DNA was accomplished using the ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit (with AmpliTaq®DNA polymerase, FS) according to the manufacturers instructions.

The sequencing mix used was as follows:

- **Terminator ready Reaction mix**: 8µL
- **Template (double stranded)**: 4µL
- **Primer**: 2µL
- **dH₂O**: 6µL (to a final volume of 20µL)
The reaction mixture was overlain with 50µL of oil and subject to the following sequencing conditions:

96°C for 5min (1 cycle)

96°C for 30 seconds
45°C for 15 seconds
60°C for 4 minutes (30 cycles)

4°C hold

The sequencing reaction mixture was then submitted to the automated fluorescent sequencing service (Nicola Preston, ICMB, University of Edinburgh) for analysis.

**PROTEIN EXPRESSION.**

**Production of His$_{10}$ Tagged Erythropoietin.**

Expression of His$_{10}$-tagged proteins was carried out according to manufacturer’s instructions (pET systems Manual, Novagen). A single colony of *E. coli* (Strain B834(DE3)) transformed with the cDNA encoding WT or mutant EPO (BamH I / Nde I inserted into pET16b) was used to inoculate a 10ml culture of LB medium containing either carbenicillin (50µgml$^{-1}$) or ampicillin (100µgml$^{-1}$). The culture was incubated at 37°C with shaking (250rpm) overnight. This culture was then used to inoculate 4×250ml of fresh LB medium containing the required antibiotic and incubated at 37°C with shaking until $A_{600}=0.6$. IPTG was then added to a concentration of 1mM and the incubation was continued with shaking at 37°C for a further 3h. The cells were then harvested by centrifugation at 8000rpm (10000×G) at 4°C for 10min. The pellet could then be stored at −20°C prior to purification.
CHAPTER 7

Preparation of a Cell-Free Extract.
[Quantities given are for a pellet derived from 500ml induced culture]

The pellet was thawed on ice and resuspended in 25.0ml binding buffer (5mM imidazole, 500mM NaCl, 20mM Tris-HCl; pH8) containing 0.005% PMSF. The cells were then sonicated on ice for 10 periods of 30s separated by 30s intervals. Inclusion bodies and cellular debris was collected by centrifugation at 10000rpm (12000×G) for 15min at 4°C. The resulting pellet was then suspended in binding buffer (20.0ml) containing 0.005% PMSF. MgCl₂ and DNAse I were then added to final concentrations of 10mM and 20μg/ml respectively and the suspension incubated at 20°C for 20min. The sample was then sonicated on ice for 5 periods of 30s separated by 30s intervals and centrifuged at 10000rpm (12000×G) for 15min at 4°C. The resulting pellet was then resuspended in 20.0ml of binding buffer containing 6M guanidinium hydrochloride and 0.005% PMSF. The sample was then shaken vigorously and stirred magnetically on ice for 0.5h to completely solubilise the protein. Any remaining insoluble material was collected by centrifugation at 20000rpm (48000×G). The supernatant was then stored at −20°C prior to affinity chromatography.

Column Chromatography of His₁₀-EPO’s Under Denaturing Conditions.
[All buffers contain 6M guanidinium hydrochloride and 0.005% PMSF]

The solubilised inclusion bodies in binding buffer were purified under denaturing conditions using Ni²⁺ immobilised on a metal chelation resin as described in the pET system manual (Novagen) The column (5.0ml column bed volume) was charged with 15.0ml aqueous NiSO₄ and equilibrated with 25.0ml binding buffer. The cell extract was then loaded onto the column and unbound proteins were washed from the column with 50.0ml binding buffer and then 25.0ml wash buffer (50mM imidazole, 500mM NaCl, 20mM Tris-HCl; pH8). The target protein was then eluted in 1.0ml fractions using elution buffer (500mM imidazole, 500mM NaCl, 20mM Tris-HCl; pH8). The
fractions were then analysed by SDS-PAGE: 10µL aliquots were precipitated by the addition of 100µL of 1:1 acetone / methanol and placed in the freezer (-20°C) for 5 min. The protein was collected by centrifugation at maximum speed in a microcentrifuge and taken up 15µL of SDS loading buffer.

**Western Blotting Analysis of Purified EPO Samples.**

[For western blots, 1/50 - 1/100 of the protein loaded for coomassie blue staining was sufficient to give a good response with the following detection method]

Following SDS-PAGE, resolved proteins were transferred to PVDF membrane in a BIO-RAD Trans-Blot cell at 40V for 2.5h in transfer buffer (25mM Tris.HCl; pH 8, 192mM glycine, 15% v/v methanol). The membrane was then incubated overnight at room temperature with shaking in blocking buffer (20mM Tris.HCl; pH 8, 37mM NaCl, 0.1% v/v Tween-20, 5% w/v non-fat milk powder) and then incubated with the primary antibody (Monoclonal murine anti-hEPO) at a dilution of 1:2000 in blocking buffer for 1h. The membrane was then washed with fresh blocking buffer (3×15minutes) and then incubated for a further 1h with the secondary antibody (Horse radish peroxidase labelled rabbit anti-mouse IgG) at a dilution of 1:1000. The membrane was then washed with blocking buffer (3×15minutes) and then with water and the detection solution was added.

**Detection Solution:**

- o-dianisidine 10.0mg
- dH₂O 19.0ml
- 0.1M imidazole 1.0ml
- 30% H₂O₂ 0.2ml
Membranes were incubated in this solution until clear strong bands were observed (5-15 mins). When stained bands were observed the reaction was quenched by washing the membrane with water (3×15 mins).

**Large Scale Production of His10-Tagged Erythropoietin.**

A single colony of *E. coli* transformed with one of the pET-16b EPO clones was used to inoculate 5.0-10.0ml of LB medium containing the required antibiotic. This culture was then grown at 37°C with shaking for approx. 8h. This culture was then used to inoculate 500ml of fresh LB medium containing the required antibiotic and this culture was grown at 37°C with shaking overnight. The following day this culture was used to inoculate 8.0L of fresh LB medium and this was then grown in the fermenter and induced as described above for small scale work (see also general experimental details for fermenter settings).

**15N Labelling of His10-WThEPO (1L scale).**

A single colony of *E. coli* transformed with one of the pET-16b EPO clones was used to inoculate 40.0ml of LB medium containing the required antibiotic. This culture was then grown at 37°C with shaking overnight. The cells were then collected by centrifugation at 3000rpm (2700G) and the supernatant was discarded. The cells were then resuspended in 20.0ml minimal media B (see general experimental details). This solution was then used to inoculate 1L of minimal media B containing MgSO4 (2.0ml from a 1M stock solution), glucose (10ml from a 20% w/v stock solution), CaCl2 (0.1ml from a 1M stock solution), (15NH4)2SO4 (0.66g, a final concentration of 5mM) and L-methionine (added to a final concentration of 50µg/ml1; B834(DE3) is a methionine auxotroph. The culture was then incubated at 37°C with shaking and induced as described above for non 15N labelled material.
Renaturation of His_{10}-EPO Samples (PROTOCOL I).

EPO samples expressed in *E. coli* were found to be expressed in an oxidised form with formed disulfides. In the case of the cysteine mutants the proteins were found to contain an additional glutathione so prior to oxidative refolding the samples were reduced.

Ni^{2+} affinity purified protein samples (30-60 µM) were reduced with 10 mM DTT for 3 h at 37°C to afford a pale orange / brown solution. The reducing agent was then removed by dialysis against 6 M guanidine.HCl, 50 mM tris.HCl; pH 8 under Nitrogen. The dialysis procedure was monitored using Ellman's reagent: A 1.0 ml sample of the dialysis buffer was treated with 10 µL 0.1 M Ellman's reagent and A_{412nm} was measured. Dialysis was complete when A_{412nm} = 0.

The resulting colourless solution was then diluted 1:50 (0.6-1.2 µM) and dialysed overnight at 4°C against refolding buffer (2% w/v N-lauroyl sarcosine, 50 mM Tris.HCl; pH 8, 40 µM CuSO_{4}) and then against 2% w/v N-lauroyl sarcosine, 50 mM Tris.HCl; pH 8 for a further 5 h. The protein samples could then be concentrated by loading (via a 0.45 µm syringe filter) onto a Ni^{2+} affinity column (equilibrated with 5 column volumes of refolding buffer) and eluted with refolding buffer containing 500 mM imidazole. Yields for the refolding step were determined by UV A_{280} and found to be in excess of 80%.

Renaturation of His_{10}-EPO Samples (PROTOCOL II: Pulsed Refolding Technique).

EPO samples expressed in *E. coli* were found to be expressed in an oxidised form with formed disulfides. In the case of the cysteine mutants the proteins were found to contain an additional glutathione so prior to oxidative refolding the samples were reduced.
Ni^{2+} affinity purified protein samples (4.0ml, 30-60μM) were reduced with 1mM Tricarboxyethyl phosphine (TCEP) for 1h at 37°C to afford a pale yellow solution. The reducing agent was then removed by dialysis against 6M guanidine.HCl, 50mM tris.HCl; pH8 under Nitrogen. The dialysis procedure was monitored using Ellmans reagent: A 1.0ml sample of the dialysis buffer was treated with 10μL 0.1M ellmans reagent and $A_{412nm}$ was measured. Dialysis was complete when $A_{412nm} = 0$.

The resulting colourless solution was then added at a rate of 500μL per hour to 400ml of refolding buffer (2% w/v N-lauroyl sarcosine, 50mM Tris.HCl; pH 8, 40μM CuSO₄) with stirring and then stirred overnight at 4°C. The protein solution was then concentrated by loading (via a 0.45μm syringe filter) onto a Ni^{2+} affinity column (equilibrated with 5 column volumes of refolding buffer) and eluted with refolding buffer containing 500mM imidazole. Yields for the refolding step were determined by UV $A_{280}$ and found again to be in excess of 80%.
6.2 Experimental Details For Chapter 3.

**Protein Electrospray Mass Spectrometry.**

General Experimental Details:

Mass spectrometry of protein samples and modified protein samples was carried out on a Micro-Mass Platform ESI mass spectrometer. Data was analysed using mass lynx v2.3 software.

**For off-line LC-MS,** HPLC was carried out on a WATERS HPLC system (Waters 600 controller connected to a Waters™ tuneable absorbance detector set at 280nm. Flow rate = 1.0ml/min⁻¹.

Samples were lyophilised on a savant speed vac plus SC110A

**For on-line LC-MS,** HPLC was carried out on a Waters 2690 microbore separations module connected to a Waters™ tuneable absorbance detector set at 280nm. Flow rate = 0.05ml/min⁻¹.

Solvents for HPLC were filtered before use.

**Off-Line LC-MS**

*Sample preparation:*

1) Samples to be analysed were precipitated out of solutions by the addition of 10 volumes of 1:1 acetone / methanol. The mixture was then placed in the freezer (-20°C) for 5 mins and the precipitate was then collected by centrifugation. The precipitate was then dissolved in 70% MeCN, 0.05%TFA (50μL).
2) All samples were then purified using a PHENOMENEX ATLANTIS C₄ (4.6 × 250mm) reverse phase HPLC column to remove trace amounts of salt and small molecules prior to injection into the mass spectrometer.

**HPLC conditions**:  
A = 100% H₂O, 0.05% TFA  
B = 100% MeCN, 0.05% TFA  

- 0-5 mins. 90% A, 10% B - 70% A, 30% B  
- 5-20 mins. 70%A, 30%B – 100%B  
- 20-30 min 100%B  
- 30-35 min 100%B - 90%A, 10%B  
- 35-40 min 90%A, 10%B

3) 1.0ml fractions were collected and those containing protein were lyophilised. The lyophilised protein samples were then dissolved in 75% MeCN, 0.1% AcOH, 0.025 – 0.05% TFA or 75% iPrOH, 0.5% AcOH, 0.025 – 0.05% TFA (100μL) and injected into the mass spectrometer by means of an automatic infusion pump (15μLmin⁻¹).

4) ESI spectra were recorded at a cone voltage of 70V (full experimental details are in appendix I)

**On-Line LC-MS**

Alternatively, and to minimize handling samples were analysed by on-line LC-MS:

**Analysis of Intact or Unmodified WT, C24, C38 and C38/83 EPO Samples.**

Unmodified His₁₀-WT and Mutant EPO’s were analysed directly after Ni²⁺ affinity chromatography after reduction with 10-50mM DTT at 37°C for 0.5h (although LC-MS can be carried out on unreduced samples, unreduced samples give very poor, often unresolvable charge spectra)
Samples were then purified using a PHENOMENEX Jupiter, microbore, C4 (1 × 150mm) reverse phase HPLC column (to remove trace amounts of salt and small molecules) coupled to ESI-MS.

**Acetamidation and Glycosylation of Fully Reduced Protein Samples.**

Column purified protein samples (typically 100µL; 60µM, ~0.1mg) were reduced by the addition of DTT to a final concentration of 10mM. The samples were then incubated at 37°C for 0.5h. The fully reduced protein samples were then separated from excess reducing agent by precipitation, using 10 volumes of 1:1 Acetone/ Methanol as previously described. The resulting precipitate was then redissolved in alkylation buffer (6M guanidinium hydrochloride, 0.5M Tris.HCl; pH 8, 20mM EDTA). Iodoacetamide or glycosyl iodoacetamide was then added to a final concentration of 10mM and the reaction mixture was allowed to stand at room temperature for 16h with the exclusion of light. The reaction was monitored by removing 10µL aliquots from the reaction mixture and the modified protein samples were analysed directly using On-line LC-MS.
Experimental Details For Chapter 4.

General Experimental Details.
Infrared spectra were recorded for Nujol suspensions or chloroform solutions using a Bio-Rad FTS 7 Interferometer coupled to a Bio-Rad SPC 3200 Processor, and a Perkin-Elmer FT Spectrophotometer. Ir. bands were strong and sharp unless specified as w (weak), br (broad) or vs (very strong).

$^1$H N.M.R. spectra were measured from the stated solvents on Brucker WP 200 SY, WP 250 SY, WP 360 SY and Varian 200MHzGemini2000 spectrometers. Signals were sharp unless specified as br(broad); and s = singlet, d = doublet, t = triplet, q = quartet and m = multiplet. $^{13}$C NMR spectra were obtained at 62.9MHz (250MHz spectrometer) and 90.6MHz (360MHz spectrometer).

Mass spectral data was obtained using A.E.I. MS-902 and Kratos MS-50TC instruments. Fast atom bombardment (FAB) spectra were measured from matrices in thioglycerol - acetonitrile or nitrobenzyl alcohol – acetonitrile.

C, H, N, Analyses were obtained using a Carlo-Erba Elemental analyser model 1106. Unless otherwise stated melting points of samples were determined on Griffin melting point apparatus and are uncorrected.

Where necessary, atmospheric moisture was excluded by connecting a drying tube filled with fused calcium chloride to a water condenser.

All organic extracts were dried over anhydrous magnesium sulfate or sodium sulfate prior to evaporation under reduced pressure. Solvents used were of technical grade unless otherwise specified.
CHAPTER 7

Wet column flash-chromatography was carried out over silica (Merck Grade 60). TLC was carried out using Polygram SIL F/ UV254 pre-coated aluminium sheets.

**General Synthesis of Glycopyranosyl Amines (19)**

![Chemical structure](image)

A solution of $N$-acetyl-$\alpha$-D-glucosamine in the minimum volume of saturated aqueous NH$_4$HCO$_3$ and was stirred at room temperature for 5 days. Further additions of NH$_4$HCO$_3$ were made throughout this period to ensure that the solution remained saturated.

The reaction mixture was then diluted with water and frozen in a bath of liquid nitrogen while spinning on a rotary evaporator (to give the maximum surface area) and freeze dried. This procedure was repeated (usually three times) until the flask reached constant weight affording a pale yellow solid which was used without further purification, Rf=0.22 (ethyl acetate: methanol 1:1) m/z=222.1196 (M+2$^2+$, 100%), C$_8$H$_{18}$N$_2$O$_5$ requires 222.1216.

**General Synthesis of Glycosyl-$\beta$-N-Iodoacetamides.**

To a solution of the glycosyl amine (15mg, 6.8×10$^{-5}$) in 1M NaHCO$_3$ (300μL) was added iodoacetic anhydride (120mg, 3.4×10$^{-4}$). The reaction was then stirred at room temperature in the absence of light. After 2h TLC indicated that the reaction was complete. The reaction mixture was then diluted with water (1cm$^3$) and then passed over a short column of AG50W-X8 resin (hydrogen form) and the eluent was then passed through a short column of AG 1-X2 resin (hydroxide form). The eluent was then lyophilised to obtain the product (24mg, 91%) as a colourless solid.
N-Acetyl glucosamine (5.00g; 22.6mmol) was added to stirring acetyl chloride (10.0cm³) and the resulting suspension was stirred magnetically for 16h. Chloroform (40.0cm³) was then added to the amber solution and the resulting solution was poured into ice (40.0g) and water (10.0cm³) with stirring. The organic layer was immediately separated and run into saturated aqueous NaHCO₃ (40.0cm³) and ice with stirring, the neutralisation being completed in a separating funnel. The organic layer was then separated and dried for 10mins. The organic layer was then filtered with suction and the residue washed thoroughly with dichloromethane. The resulting orange solution was concentrated to approximately 10cm³ under vacuum. Anhydrous ether (50.0cm³) was then added and the product crystallised out immediately. The product was stoppered and left to stand at room temperature for 16h. The pale yellow solid was then filtered with suction, washed with anhydrous ether (2x 15.0cm³) and allowed to dry for 5 minutes to afford the crude product which was subsequently purified by flash chromatography over silica (100% ethylacetate) to afford the product (4.17g; 50%) as a colourless crystalline solid, mpt. 124-125°C (lit. 127-128°C), \( \nu_{\text{max}} \) 3239(NH), 1742 and 1642(CO) cm\(^{-1}\), \( \delta_{\text{H}} \) (250MHz; CDCl₃) 6.17(1H, d, J=3.8Hz, H-1), 5.86(1H, d, J=8.7Hz, NH), 5.31(1H, dd, J=9.8Hz, H-3), 5.19(1H, dd, J=9.6Hz, H-4), 4.57-4.46(1H, m, H-2), 4.31-4.06(3H, m, H-6b, H-6a and H-5), 2.09(3H, s, COCH₃), 2.04(6H, s, 2xCOCH₃) and 1.97(3H, s, COCH₃), Found: m/z (FAB-MS), 366(M⁺, 90%), 330(M⁺-Cl, 100%), C₁₄H₂₀NO₈Cl requires m/z, 366 (M⁺).
To a solution of the glucosyl chloride (36) (1.00g; 2.7mmol), TBAHS (0.92g; 2.7mmol) and sodium azide (0.53g; 8.1 mmol) in dichloromethane (10.0cm³) was added saturated aqueous NaHCO₃ (10.0cm³). The resulting biphasic solution was stirred vigorously at room temperature for 1h.

Ethyl acetate (100.0cm³) was then added and the organic layer was separated and washed with saturated aqueous NaHCO₃ (1×20.0cm³), water (2×20.0cm³) and saturated aqueous NaCl (1×10.0cm³). The organic phase was then dried over Na₂SO₄ and the solvent removed under vacuum to afford the pure (by TLC) azide (0.91g; 91%) as a white solid mpt. 158-161°C (lit. 166-167°C; νmax 3622 (NH), 2118 (N₃), 1747 and 1642 (CO) cm⁻¹, δH (250MHz; CDCl₃) 5.84(1H, d, J=8.9Hz , NH ), 5.24(1H, dd, J=9.7Hz, H-3), 5.08(1H, dd, J=9.7Hz, H-4), 4.76(1H, d, J=9.3Hz, H-1), 4.26(1H, dd,J₆a-H₆b =12.5Hz J₆a-H₅ =4.8Hz, H-6b), 4.14(1H, dd, J₆a-H₅ =2.4Hz, H-6a), 3.90(1H, m, H-2), 3.82-3.75(1H, m, H-5) and 2.09, 2.02, 2.01, and 1.96(12H, 4s, COCH₃) ppm, Found: m/z (FAB-MS), 373 (MH⁺), C₁₄H₂₀N₄O₈ requires m/z, 372.3 (M⁺).
To a solution of the per-acetate (1.0g, 2.57mmol), in freshly distilled DCM (15.0cm³) containing SnCl₄ (1.0M in anhydrous DCM, 0.93mmol, 932µL) under nitrogen was added azido trimethylsilane (800µL, 6.04mmol) and the reaction mixture was stirred at room temperature.

After 20h the reaction mixture was diluted to 50cm³ with DCM and washed with 10% K₂CO₃ (3×20.0cm³) and brine (2×20.0cm³). The organic layer was then dried and the solvent was removed under vacuum to afford the product (homogeneous by TLC) as a white solid (954mg, 99%). Analytical data as above.

2-Acetamido-2-Deoxy-3,4,6-Tri-O-Acetyl-β-D-Glucopyranosyl Amine (44)

\[
\begin{array}{ccc}
\text{OAc} & \text{N} & \text{OAc} \\
\text{AcO} & \text{O} & \text{NHAc} \\
\text{NHAc} & \text{N} & \text{AcO} \\
\end{array}
\xrightarrow{\text{H₂, PtO₂}}
\begin{array}{ccc}
\text{OAc} & \text{N} & \text{OAc} \\
\text{AcO} & \text{O} & \text{NHAc} \\
\text{NHAc} & \text{N} & \text{AcO} \\
\end{array}
\]

A solution of the glycosyl azide (17) (1.50g; 4.0mmol) in anhydrous THF (27.0cm³) was catalytically hydrogenated at atmospheric pressure with PtO₂ for 1h.

The catalyst was then removed by filtration through Celite and the filtrate was evaporated to dryness under vacuum to afford the pure (by TLC) amine (1.03g; 74%) as a pale grey solid m.p.=159°C , ν max 1744 and 1682 (CO) cm⁻¹ , δ n (250MHz; CDCl₃) 5.72(1H, d, J=8.9Hz ,NH), 5.10-5.00(2H, m,H-3 and H-4), 4.20(1H, dd, JHH₆b =12.3Hz JH₆b-H₅ = 4.8Hz, H-6b), 4.13-3.97(2H, m, H-6 and H-2), 3.99(1H, d, J=9.5, H-1) 3.66-3.60(1H, m, H-5) and 2.08, 2.02, 2.01, and 1.96(12H, 4s, 4×COCH₃), Found: m/z (FAB-MS), 347(MH⁺ , 44%), 330(M-NH₂, 66%), C₁₄H₂₂N₂O₈ requires m/z 346.3 (M⁺ ).
CHAPTER 7

2-Iodo-Acetyl Chloride (45).
Thionyl chloride (0.1cm\(^3\); 1.30mmol) was added dropwise over 5 mins. with stirring to iodoacetic acid (200mg; 1.08mmol) and the resulting pink suspension was stirred at room temperature with the exclusion of atmospheric moisture and light.

After 24h the excess thionyl chloride was removed under vacuum to afford the acid chloride (214 mg, 97.2%) as a pink oil. \(\delta_R\) (250MHz; CDCl\(_3\)) 4.23(s, CH\(_2\)I), \(\delta_C\) (CDCl\(_3\)) 167.12(COCl) and 4.12(CH\(_2\)I), Found: m/z (EI-MS), 203.9 (M\(^+\)), \(C_9H_5OC\)I requires m/z, 203.8 (M\(^+\)).

2-Acetamido-2-Deoxy-3,4,6-Tri-O-Acetyl-1-N-[1-(2-Iodo)Acetyl]-\(\beta\)-D-Glucopyranose (46).\(^7\)

![Chemical Structure]

A solution of the protected glucopyranosylamine (44) (187mg; 0.54mmol) in anhydrous DCM (4.0cm\(^3\)) was treated with pyridine (0.044cm\(^3\); 0.54mmol) and cooled to -40°C with the exclusion of light under Nitrogen. This solution was then treated dropwise with stirring with 2-iodo-acetyl-chloride (110mg; 0.54mmol) and the reaction mixture was stirred at -40°C.

After 0.5h the reaction was quenched with chloroform (40.0cm\(^3\)) and the resulting yellow solution was washed with saturated aqueous NaHCO\(_3\) (2×15.0cm\(^3\)) water (1×15.0cm\(^3\)). The organic phase was then dried, filtered, concentrated under vacuum to approx. 2cm\(^3\) and then co-evaporated with toluene (2×5.0cm\(^3\)) to afford the iodoacetamide (264mg; 95%) as a yellow solid mpt.=200-202°C(decomp.), \(\nu_{\text{max}}\) 1745...
and 1684 (CO) cm⁻¹, δₜ (250 MHz; CDCl₃) 7.65 (1H, d, J=8.1 Hz, NH), 6.57 (1H, d, J=8.4 Hz, NH), 5.15-4.98 (3H, m, H-1, H-3 and H-4), 4.29 (1H, dd, J₆₆₉-H₆₆₈ = 12.5 Hz, J₆₆₆-H₅₆₅ = 4.3 Hz, H-6), 4.22-4.12 (1H, m, H-2), 4.08 (1H, dd, J₆₆₅-H₆₆₄ = 2.1 Hz, H-6), 3.79-3.73 (1H, m, H-5), 3.62 (2H, s, CH₂I) and 2.07, 2.06, 2.03 and 1.99 (12H, 4x, 4x CH₃CO), Found: 36.5% C; 4.3% H; 5.2% N, m/z (FAB-MS), 515 (MH⁺, 63%), 389 (M-I, 41%) C₁₀H₁₇N₂O₉I requires 37.4% C; 4.5% H; 5.5% N, m/z, 515 (MH⁺).

2-Acetamido-2-Deoxy-1-N-[1-(2-Iodo)Acetyl]-β-D-Glucopyranose (40).⁷¹

The iodoacetamide (46) (65.0 mg; 0.13 mmol) as a solution in anhydrous methanol (10.0 cm³) was treated with a solution of sodium methoxide in methanol (0.5 M; 0.078 cm³) under nitrogen with the exclusion of light and the resulting solution was stirred at room temperature for 3 h.

The solution was then treated with a solution of NH₄Cl (0.5 M; 0.078 cm³) and the solvent was removed under vacuum. The residue was dissolved in water (14.0 cm³), extracted with ethyl acetate (2×3.5 cm³) and the aqueous layer was evaporated to dryness to afford the pure (by TLC) deprotected product (56.4 mg; 100%) as an orange solid. Rₑ=0.68 (CHCl₃: MeOH:H₂O, 12:12:3), δₜ (250 MHz; D₂O) 5.33 (1H, d, J=9.7 Hz, H-1), 3.82 (1H, dd, J₆₆₆-H₆₆₅ = 12.3 Hz, J₆₆₆-H₅₆₅ = 2.1 Hz, H-6), 3.76 (1H, dd, J₂₃ = 9.8 Hz, H-2), 3.69 (1H, dd, J₆₆₅ = 4.8 Hz, H-6), 3.67 (2H, q, J₆₆₅ = 10.0 Hz, CH₂I), 3.55 (1H, dd, J₃₄ = 9.5 Hz, H-3), 3.45 (1H, m, H-5), 3.42 (1H, dd, J₄₅ = 10.1 Hz, H-4) and 1.98 (3H, s, COCH₃), δC (D₂O) 174.7 and 173.1 (CO), 78.6 (C(1)H), 77.7 (C(5)H), 74.1 (C(3)H), 69.4 (C(4)H), 60.4 (C(6)H₂), 54.3 (C(2)H), 22.4 (COCH₃) and -3.6 (CH₂I), Found: m/z (FAB-MS), 389.019 (MH⁺), C₁₀H₁₇N₂O₉I requires m/z, 389.020 (MH⁺).
**N-N-Diacetyl-L-Cystine Dimethyl Ester (48)**

Triethylamine (0.58g; 5.8mmol) was added to L-cystine dimethyl ester dihydrochioride (47) (1.0g; 2.9mmol) as a suspension in chloroform (15.0cm³). Acetic anhydride (3.0cm³) was then added and the resulting solution was stirred at room temperature for 1h.

The reaction mixture was then washed with water (2×5.0cm³) and then separated and the organic layer was dried and the solvent was removed under vacuum to afford the product (0.89g; 87.6%) as a white crystalline solid mpt. 118-122°C, $\delta_H$ (250MHz; CDCl₃) 6.62(2H, d, $J=7.4Hz$, 2×NH), 4.88-4.80(2H, m , 2×CH), 3.75(6H, s, 2×CO₂CH₃), 3.18(4H, dd, $J=5.1Hz$ $J=2.9Hz$, 2×CH₂S) and $\delta_C$ (CDCl₃) 170.8 and 170.1(CO), 52.6(αCH), 51.5 (CO₂CH₃), 40.4 (CH₂S) and 22.9(COCH₃).

**N-Acetyl-L-Cysteine Methyl Ester (49)**

A solution of N-N-diacetyl-L-cystine dimethyl ester (48) (200mg; 0.57mmol) in methanol (10.0cm³) was stirred and zinc powder (0.37g; 5.7mmol) was added. The resulting suspension was cooled to 0°C and treated dropwise with concentrated HCl until TLC indicated no starting material was present (0.1cm³ of HCl added in total).

After 2.5h the reaction mixture was diluted with DCM (30.0cm³) and the colourless solution was washed with water (1×10.0cm³), dilute NaHCO₃ (1×10.0cm³) and then water (1×10.0cm³). The solution was then dried and the solvent was removed under vacuum to afford the product (84.5mg, 42%) as white crystalline solid mpt. 76-80°C, $\nu_{max}$ 3432 (NH), 1741 and 1672 (CO) cm⁻¹, $\delta_H$ (200MHz; CDCl₃) 6.80(1H , s(br), NH ), 4.86-4.80(1H, m , CH), 3.72(3H, s, CO₂CH₃ ), 2.93(2H, dd, $J=9.2Hz$ $J=4.4Hz$, CH₂S), 2.01(3H, s, COCH₃), and 1.39(1H, t, $J=9.0Hz$, SH), $\delta_C$ (CDCl₃) 170.5 and 169.7 (CO) ,
53.4(CH), 52.7 (CO$_2$CH$_3$), 26.7 (CH$_2$S) and 23.0(COH$_3$), Found: m/z (FAB-MS), 178(MH$^+$, 100%), C$_{16}$H$_{27}$N$_3$O$_9$S requires m/z, 178 (MH$^+$).

2-Acetamido-2-Deoxy-1-N-[1-(2-S-(N-Acetyl-a-Methyl-L-Cysteinyl)Acetyl]-β-D-Glucopyranose (50).

A solution of N-acetyl-L-cysteine-methyl ester (20mg; 0.11mmol) and the deprotected iodoacetamide (40)(43mg; 0.11mmol) in ammonium carbonate (50mM, 20.0cm$^3$) was stirred at room temperature with the exclusion of light.

After 1h the solvent was removed under vacuum to afford the crude product as a yellow solid which was taken up in water (3.0cm$^3$), co-concentrated with methanol (3×3.0cm$^3$) and lyophilised to again obtain a yellow solid. This solid was then dissolved in methanol, filtered and the solvent was removed under vacuum to afford the crude product which was purified by flash chromatography over silica (EtOAc:MeOH, 1:1) to obtain the pure cysteine derivative (47.1mg, 97.9%) as a pale yellow solid mpt. 59-65°C, δ$_H$ (360MHz; D$_2$O) 5.13(1H, d, J=9.8Hz, H-1), 4.71(1H, dd, J=8.1Hz J=5.0Hz, CHCH$_2$S), 3.94(1H, dd, J$_{H6a-H5b}$ =12.3Hz J$_{H6b-H5a}$=2.0Hz, H-6b), 3.90(1H, dd, J$_{23}$ =9.9Hz, H-2), 3.83(3H, s, CO$_2$CH$_3$), 3.80(1H, dd, J$_{H6a-H5s}$ =4.8Hz, H-6a), 3.69(1H, dd, J$_{1}$ =10.1, J$_{2}$=8.5Hz, H-3), 3.59-3.57(1H, m, H-5), 3.54(1H, dd, J=9.8Hz, H-4), 3.38(2H, q, J$_{AB}$ =15.3Hz, COCH$_3$), 3.15(1H, dd, J$_{H6a-H5b}$ =14.1Hz J$_{Hb-Ch}$=5.0Hz, CHCH$_2$S), 3.00(1H, dd, J$_{H6-Ch}$=8.2Hz, CHCH$_2$S) and 2.11 and 2.07(6H, 2×s, 2×COCH$_3$)ppm. δ$_C$ (D$_2$O) 174.6, 174.2, 172.8 and 172.3(CO), 78.6(C(1)H), 77.6(C(5)H), 73.9(C(3)H), 69.4(C(4)H), 60.5(C(6)H$_2$), 54.2(C(2)H), 53.2(CO$_2$CH$_3$), 52.4(CHCO$_2$H), 35.2(COCH$_2$S), 32.9(CH$_2$S) and 22.0 and 21.6(2×COCH$_3$), Found: m/z (FAB-MS), 438(MH$^+$, 73%) and 460(MNa$^+$, 100%), C$_{16}$H$_{27}$N$_3$O$_9$S requires m/z, 438 (MH$^+$).
Glutathione Derivative (52)\textsuperscript{71}.

The iodoacetamide (40) (43mg; 0.11mmol) was added to a solution of glutathione (34.4mg; 0.11mmol) in 50mM ammonium carbonate (50.0cm\textsuperscript{3}, pH8). The resulting solution was stirred at room temperature with the exclusion of light and monitored using Ellmans reagent to detect the free thiol groups. Ellmans reagent (0.01M; 0.01cm\textsuperscript{3}) was added to an aliquot (0.01cm\textsuperscript{3}) in ammonium carbonate solution (50mM, 1.0cm\textsuperscript{3}) and the absorption peak at 412nm was measured using UV spectroscopy. The reference cell contained the initial glutathione solution (0.01cm\textsuperscript{3}) in ammonium carbonate solution (50mM; 1.0cm\textsuperscript{3}).

After 2.5h the the reaction mixture was lyophilised and the crude product was dissolved in water (2.0cm\textsuperscript{3}) and co-concentrated with MeOH (3×10.0cm\textsuperscript{3}) and the water was removed under vacuum to afford the product (65mg; 104.2%) as a pale yellow solid (1 spot by TLC),\textsuperscript{5} \( \delta \) (360MHz; D\textsubscript{2}O) 5.15(1H, d, J=9.7Hz, H-1), 4.65(1H, dd, J\textsubscript{1} =8.6Hz J\textsubscript{2} =5.0Hz, \( \alpha \)CHCH\textsubscript{2}S), 3.94(1H, dd, J\textsubscript{H\textsubscript{6a}-H\textsubscript{6b}} =10.6Hz J\textsubscript{H\textsubscript{5b}-H\textsubscript{5a}} =1.8Hz, H-6\textsubscript{b}), 3.90(1H, t, J\textsubscript{23} =9.9Hz, H-2), 3.85-3.78 (4H, m, H-6\textsubscript{a}, CH\textsubscript{2}CO\textsubscript{2}H and \( \alpha \)CH), 3.70(1H, t, J=9.3Hz, H-3), 3.62-3.58(1H, m, H-5), 3.54(1H, t, J=9.1Hz, H-4), 3.39(2H, s, COCH\textsubscript{2}S), 3.13(1H, dd, J\textsubscript{H\textsubscript{a}-H\textsubscript{b}} =14.0Hz J\textsubscript{H\textsubscript{b}-CH} =5.0Hz, CHCH\textsubscript{2}S), 2.96(1H, dd, J\textsubscript{H\textsubscript{a}-CH} =8.8Hz, CHCH\textsubscript{2}S), 2.63-2.54(2H, m, COCH\textsubscript{2}CH\textsubscript{2}), 2.24-2.18(2H, m, CH\textsubscript{2}CH) and 2.05(3H, s, COCH\textsubscript{3}), \( \delta \)\textsubscript{c} (D\textsubscript{2}O) 174.6, 174.2, 172.8 and 172.3(CO), 78.6(C(1)H), 77.6(C(5)H), 74.0(C(3)H), 69.4(C(4)H), 60.5(C(6)H\textsubscript{2}), 54.2(C(2)H), 54.0(CHCH\textsubscript{2}), 52.9(CHCH\textsubscript{2}S), 43.3(CH\textsubscript{2}CO\textsubscript{2}H), 35.2(COCH\textsubscript{2}S), 33.5(CH\textsubscript{2}S), 31.3(CH\textsubscript{2}CONH), 26.1(\( \alpha \)CHCH\textsubscript{2}CH\textsubscript{2}) and 22.1(COCH\textsubscript{3}).

Found: m/z (FAB-MS), 568 (MH\textsuperscript{+}), C\textsubscript{20}H\textsubscript{33}N\textsubscript{5}O\textsubscript{12}S requires m/z, 567 (MH\textsuperscript{+}).

Control Experiment for the Derivatisation of Glutathione.

This experiment was run exactly as above except that no glycosyl iodoacetamide was added to the glutathione in solution. The number of free thiol groups were monitored over three hours.
2-Acetamido-2-Deoxy-1-N-[1-(2,2-Iodo, $^{13}$C)Acetyl]-$\beta$-D-Glucopyranose (53).

The $^{13}$C labelled, deprotected glycosyl iodoacetamide (53) was prepared as above for 2-acetamido-2-deoxy-1-N-[1-(2-iodo)acetyl]-$\beta$-D-glucopyranose from 2-acetamido-2-deoxy-3,4,6-tri-O-acetyl-$\beta$-D-glucopyranosyl amine and the $^{13}$C labelled 2-iodo acetyl chloride in 86% yield over all three steps from the $^{13}$C labelled iodoacetic acid. The $^{13}$C labelled iodoacetamide was obtained as a yellow solid. The $^{13}$C NMR spectrum of the labelled material affords an intense $^{13}$C peak at -3.14ppm, Found: m/z (FAB-MS), 390 (MH$^+$, 11%).

$^{13}$C Labelled Glutathione Derivative (52).

This experiment run exactly as above for (52) except that the $^{13}$C labelled glycosyl iodoacetamide (53) was used.

After 5.25h the solvent was removed under vacuum to afford the crude product as a yellow solid which was dissolved in water (5.0cm$^3$) and co-concentrated with MeOH (5×5.0cm$^3$). On the addition of methanol (10.0cm$^3$) the product precipitated out of solution and the white glutathione derivative (38mg; 61%) was collected by filtration and dried under vacuum. The $^{13}$C NMR spectrum of the glutathione derivative afforded a very intense peak at 35.18ppm, Found: m/z (FAB-MS), 569.19475(MH$^+$, 73%), C$_{20}$H$_{34}$N$_5$O$_{12}$S requires m/z, 569.19528 (MH$^+$).

4,6-O-p-Methoxybenzylidine-D-Glucal (62)$^{48}$.

![Reaction Scheme]

To a solution of tri-O-acetyl-D-glucal (61) (2.0g) in anhydrous methanol (10.0cm$^3$), under nitrogen, was added a solution of NaOMe in methanol (0.5M, 200µl, 1×10$^{-4}$mol).
CHAPTER 7

After 1.5h the reaction was quenched by the addition of NH₄Cl (0.5M, 200μL, 1×10⁻⁴). The solvent was removed under vacuum and the deacetylated product (1.0g, 6.8mmol), and anhydrous ZnCl₂ (0.6g, 4.4mmol) were taken up in p-anisaldehyde (2.0cm³, 16.4mmol) and stirred at room temperature.

After 16h the green reaction mixture was treated with a solution of Na₂CO₃ (10% w/v, 10.0cm³) with stirring. The precipitated Zn(OH)₂ was then filtered off with suction and the precipitate was washed with water (20.0cm³) and ether (50.0cm³). The organic phase was separated and the aqueous phase was extracted with ether (5×50.0cm³). The combined organic extracts were dried [Na₂SO₄], filtered, and the solvent was removed under vacuum to afford the crude product as an orange oil which was purified by flash chromatography over silica (100% CHCl₃, then 3:2 EtOAc / Pet. Ether 40-60°C) to afford the pure product (0.33g, 20%) as a white solid. Rf=0.53 (3:2 EtOAc / Pet. Ether 40-60°C), δH (200MHz; CDCl₃) 7.42(2H, d, J=8.9Hz, 2×Arff), 6.89(2H, d, J=8.9Hz, 2×ArH), 6.32(1H, d, J=6.2Hz, H-1), 5.54(1H, s, ArCH), 4.76(1H, dd, J₁,₂=6.2Hz, J₂,₃=2.0Hz, H-2), 4.51-4.46(1H, m, CH), 4.37-4.31(1H, m, CH), 3.95-3.73(5H, m, OCH₃ and CH and CH) and 2.31(1H, d, J=5.3Hz, OH)ppm. δC (CDCl₃) 160.2(qC, Q(OMe)), 144.0(CH), 129.4(qC), 127.4(2C×H, m-ArCH), 113.6 (2×CH, o-ArCH), 103.4(CH, CHAr), 110.7(CH,C-1), 80.6(CH), 68.2(CH), 68.1(CH₂, C-6), 66.5(CH), and 55.2(CH₃, OCH₃).

4,6-O-p-Methoxybenzylidine-3-Butyldimethylsilyl-D-Glucal (63)₄⁸.
To a stirred solution of the glycal (62) (0.25g, 0.95mmol) in anhydrous DCM (5.0cm³) was added imidazole (0.16g, 2.4mmol) and TBDMS-chloride (0.18g, 1.2mmol). The resulting suspension was stirred at room temperature.

After 1.5h water (30.0cm³) was added and the organic phase was separated. The aqueous phase was then extracted with CH₂Cl₂ (3×10.0cm³) and the combined organic extracts were dried [Na₂SO₄], filtered, and the solvent was removed under vacuum to afford the crude product as a yellow oil. The oil was purified by flash chromatography over silica (95:5 Pet. Ether 40-60°C / EtOAc) to afford the pure product (267mg, 75%) as a colourless oil. R=0.47 (9:1 Pet. Ether 40-60°C / EtOAc), δₚ (250MHz; CDCl₃) 7.42(2H, d, J=8.9Hz, 2×ArH), 6.89(2H, d, J=8.9Hz, 2×ArH), 6.29(1H, d, J=6.2Hz, H-1), 5.54(1H, s, ArCH), 4.67-4.64(1H, m, CH), 4.51-4.47(1H, m, CH), 4.34-4.31(1H, m, CH), 3.88-3.73(5H, m, OCH₃ and CH and CH), 0.89 (9H, s, 'Bu) and 0.09 and 0.07(6H, 2×s, 2×CH₃)ppm. δₚ (CDCl₃) 159.8(qC, C(OMe)), 143.2(CH), 129.8(qC), 127.1(2C×H, m-ArCH), 113.4(2×CH, o-ArCH), 105.3(CH, CHAr), 110.1(CH, C-1), 80.4(CH), 68.7(CH), 68.2(CH₂, C-6), 67.2(CH), 55.(CH₃, OCH₃) 25.7(3×CH₃, tBu), 18.1(qC, tBu) and -4.5 and -4.9(2×CH₃, TBDMS). C₂₀H₃₀O₅Si requires m/z=378.54, found m/z (ESI-MS), 378.5(M⁺, 100%).

4-p-Methoxybenzyl-3-tButyldimethylsilyl-D-Glucal (64).

A solution of the protected glycal (63) (200mg, 5.28×10⁻⁴mol) in anhydrous DCM (5.0cm³) cooled to -20°C was stirred under nitrogen. A solution of DIBAL-H (1M in toluene, 2.11cm³; 2.11×10⁻³mol) was then added dropwise.
CHAPTER 7

After 3h the reaction was quenched by the addition of methanol (0.25cm$^3$). The reaction mixture was then diluted with EtOAc (5.0cm$^3$) and then a solution of K-Na tartrate (20%w/v, 2.5cm$^3$) was added. The resulting solution was stirred at room temperature for 1h. The mixture was separated and the aqueous phase was extracted with EtOAc (3×5.0cm$^3$). The combined organic extracts were dried [Na$_2$SO$_4$], filtered, and the solvent was removed under vacuum to afford the crude product as a colourless oil. The oil was purified by flash chromatography over silica (Pet. Ether 40-60°C / EtOAc) to afford the pure product (162mg, 81%) as a colourless oil. R$_f$=0.5 (70:30 Pet. Ether 40-60°C / EtOAc), $\delta_H$ (200MHz; CDCl$_3$) 7.27(2H, d, J=8.9Hz, 2×ArH), 6.87(2H, d, J=8.9Hz, 2×ArH), 6.31(1H, d, J=6.1Hz, H-1), 4.68(2H, q, J$_{AB}$=11.3Hz, ArCH$_2$), 4.66(1H, dd, J$_{12}$=6.1Hz, J$_{23}$=2.9Hz, H2), 4.35-4.31(1H, m, CH), 3.96-3.91 (1H, m, CH), 3.89-3.79 (5H, m, OCH$_3$ and CH and CH), 3.63-3.57, (1H, m, CH), 2.03(1H, s(br), OH), 0.90(9H, s, 3Bu) and 0.1(6H, s, 2×CH$_3$) ppm. $\delta_C$ (CDCl$_3$) 159.2(qC, C(OMe)), 143.4(CH), 129.9(qC), 129.5(2C×H, m-ArCH), 113.8(2×CH, o-ArCH), 103.4(CH), 76.2(CH), 73.4(CH$_2$), 68.5(CH), 61.8(CH$_2$, C-6), 55.2(CH$_3$, OCH$_3$) 25.7(3×CH$_3$, tBu), 17.9(qC, tBu) and -4.5 and -4.7(2×CH$_3$, TBDMS). C$_{20}$H$_{33}$O$_5$Si requires m/z=380.56, found m/z (FAB-MS), 379.2174(M$, 100\%$).

4-p-Methoxybenzyl-D-Glucal (57)$^{48}$.

TBAF (1M in THF, 0.55ml; 5.52×10$^{-4}$mol) was added to a solution of the glycal (64) (140mg, 3.68×10$^{-4}$mol) in anhydrous THF (4.0cm$^3$). The resulting pale yellow solution was stirred at room temperature.
After 16h the solvent was removed under vacuum and the product was purified from the pale yellow residue via flash chromatography over silica (Pet. Ether 40-60°C / EtOAc) to afford the pure product (98mg, 100%) as a waxy white solid. Rf=0.43 (80:20 Pet. Ether 40-60°C / EtOAc), mpt. 90-92°C, δH (200MHz; CDCl3) 7.39(2H, d, J=8.8Hz, 2×ArH), 6.97(2H, d, J=8.8Hz, 2×ArH), 6.45(1H, d, J=6.0Hz, H-1), 4.95-4.71(m, br, 3H), 4.39-4.35(1H, m, CH), 3.95-3.83(5H, m, OCH3 and CH and CH), 3.70-3.63(1H, m, CH) and 3.40(1H, s, br, H) ppm. δC (CDCl3) 151.3(qC, Q(OMe)), 135.4(CH), 122.4(qC), 121.3(2×CH, m-ArCH), 105.2(2×CH, o-ArCH), 95.3(CH), 69.9(CH), 68.3(CH), 65.2(CH2), 60.8(CH), 52.5(CH2), 46.2(CH3, OCH3).

4-p-Methoxybenzaldehyde dimethyl acetal\textsuperscript{141}.

A solution of p-anisaldehyde (30cm\textsuperscript{3}, 0.25mol), trimethyl orthoformate (30.2cm\textsuperscript{3}, 0.275mol) and toluene p-sulfonic acid monohydrate (0.24g, 1.25mmol) was stirred at room temperature for 18h.

After 18h, sodium carbonate (0.265g, 2.5mmol) was added and the mixture was stirred for a further 1h. The residue was filtered and the methyl formate was removed on a rotary evaporator. The remaining pink solution was distilled to give the acetal (32.28g, 71%) as colourless oil, bpt=49°C at 0.03mmHg (100°C at 7mmHg\textsuperscript{141}), δH (200MHz; CDCl3) 7.37(2H, d, J=8.8Hz, 2×ArH), 6.89(2H, d, J=8.8Hz, 2×ArH), 5.36(1H, s, CH), 3.81(3H, s, p-CH3O) and 3.31(6H, s, 2×CH3O) ppm.

2-Acetamido-2-Deoxy-4,6-O-p-Methoxybenzylidine-β-D-Glucopyranosyl Azide(66)\textsuperscript{141}.

![Chemical structure of 2-Acetamido-2-Deoxy-4,6-O-p-Methoxybenzylidine-β-D-Glucopyranosyl Azide(66)](image)

\textsuperscript{141} See references for further details.
The per-acetylated glycosyl azide was dissolved in anhydrous methanol (10.0 cm$^3$) and NaOMe (200 μL, 0.5 M solution in anhydrous methanol) was added. After 2 h the reaction was treated with NH$_4$Cl and the solvent was removed under vacuum. The residue (125 mg, 5.03×10$^{-4}$ mol) was dried under high vacuum and dissolved in anhydrous DMF (5.0 cm$^3$). p-Methoxy benzaldehyde dimethyl acetal (275 mg, 1.5×10$^{-3}$ mol) and p-tosic acid (20 mg, 1.21×10$^{-4}$ mol) were then added and the reaction mixture was stirred at room temperature. After 2 h the reaction mixture was connected to a rotary evaporator and the reaction was continued at 30°C and 30 mbar for a further 7 h.

The reaction mixture was concentrated under high-vacuum and then poured into an ice-cold mixture of DCM (5.0 cm$^3$) and sat. aqueous NaHCO$_3$ (5.0 cm$^3$) with stirring. The precipitate was filtered off with suction and washed with water and DCM. The solid was dried under vacuum and crystallised from ethyl acetate to afford the pure p-methoxy benzylidene acetal (147 mg, 80%) as white crystals R$_f$=0.38 (100% EtOAc).

The product was very insoluble and a portion was acetylated in an attempt to increase its solubility for analysis. The product was dissolved in pyridine (2.0 cm$^3$). Acetic anhydride (1.0 cm$^3$) was added and the reaction was stirred at room temperature. After 5 h the solution was evaporated under high vacuum. The solid was taken up in water (10.0 cm$^3$) and DCM (20 cm$^3$). The organic layer was separated and washed with saturated aqueous NaHCO$_3$ (1×10 cm$^3$) and again with water (1×10 cm$^3$), dried and the solvent was removed under vacuum to afford the acetylated product in quantitative yield. R$_f$=0.59 (100% EtOAc), δ$_d$ (250 MHz; CDCl$_3$/ CD$_3$OD) 7.24 (2H, d, J=8.9 Hz, 2×ArH), 6.77 (2H, d, J=8.9 Hz, 2×ArH), 5.37 (1H, s, ArCH), 5.08 (1H, dd, J$_{3,4}$=9.3 Hz, H3), 4.55 (1H, d, J$_{1,2}$=9.3 Hz, H1), 4.21 (1H, dd, J$_{H6a-H6b}$=10.4, J$_{H6b-H6}$=4.8 Hz, H6b), 3.92 (1H, dd, J$_{2,3}$=10.2 Hz, H2), 3.69 (1H, dd, H6a), 3.68 (3H, s, OCH$_3$), 3.60 (1H, dd, J$_{3,4}$=9.3 Hz, H4), 3.50 (1H, m, H5) and 1.95 and 1.85 (6H, 2×COCH$_3$) ppm. δ$_c$, 171.4, 170.9, 159.8 and 128.8 (qC), 127.1 (2×ArCH), 113.2 (2×ArCH), 101.2 (CH), 88.8 (CH), 78.1 (CH), 71.4 (CH), 68.1 (CH), 67.9 (CH$_2$), 54.9 (OCH$_3$), 53.3 (CH) and 22.2 and
20.3(COCH₃), Found: m/z (FAB-MS), 407(MH⁺, 74%), C₁₉H₂₂N₄O₇ requires m/z, 407.4 (MH⁺).

2-Acetamido-2-Deoxy-3-Benzyl-4,6-O-p-Methoxybenzylidine-β-D-Glucopyranosyl Azide (67)

The alcohol (125mg, 3.43×10⁻⁴mol), Ba(OH)₂·8H₂O (250mg, 7.89×10⁻⁴mol) and 3Å molecular sieves were stirred in anhydrous DMF (3.0cm³) under nitrogen. Benzyl bromide (147mg, 8.58×10⁻⁴mol, 102μL) was then added dropwise and stirring was continued at room temperature.

After 4h the reaction mixture was diluted with DCM (5.0cm³) and filtered. The precipitate was washed with further DCM (10.0cm³). The filtrate was washed with water (2×10.0cm³) and the organic phase was dried [Na₂SO₄], filtered, and evaporated to afford the product (129mg, 83%) as a white solid, mp=175-178(decomp), δₜ (250MHz; CDCl₃/CD₃OD) 7.35(2H, d, J=8.9Hz, 2×ArH), 7.25(5H, s(br), 5×ArH), 6.86(2H, d, J=8.9Hz, 2×ArH), 5.48(1H, s, ArCH), 4.86(1H, d, J=9.1Hz, H-1), 4.78(2H, q, JAB =11.8Hz, CH₂), 4.39(1H, dd, JH₆a-H₆b=10.3, JH₆b-H₅=4.8Hz, H₆b), 3.98-3.80(7H, m, CH₃, CH₂, 2×CH), 3.97-3.71(1H, m, CH), 3.67-3.55(1H, m, CH) and 1.96(3H, s, COCH₃)ppm, δc, 170.0, 158.3, 136.4 and 127.8(qC), 126.5(2×ArCH), 126.3(2×ArCH), 125.9(ArCH), 125.6(2×ArCH), 111.8(2×ArCH), 99.5 (CH), 87.1(CH), 80.1(CH), 75.7(CH), 72.6(CH₂), 66.6(CH₂), 66.4(CH), 53.6(CH), 53.5(OCH₃) and 20.9(COCH₃), Found: m/z (FAB-MS), 455.(MH⁺), C₂₃H₂₆N₄O₆ requires m/z, 455.5(MH⁺).
2-Acetamido-2-Deoxy-3-Benzyl-6-p-Methoxybenzyl-β-D-Glucopyranosyl
Azide(68)\(^{141}\).

A solution of the glycosyl azide (480mg, 1.1×10^{-3}mol) and Na(CN)BH\(_3\)(332mg, 5.28×10^{-3}mol) in anhydrous DMF (10cm\(^3\)) containing 3A molecular sieves was stirred in an ice-bath under nitrogen. An ice cold solution of trifluoroacetic acid (817μL, 10.6×10^{-3}mol) in anhydrous DMF (5.0cm\(^3\)) was then added dropwise and then the reaction mixture was allowed to come to room temperature by removal of the ice bath and stirred at room temperature for a further 16h under nitrogen.

The reaction mixture was then filtered with suction and poured into ice-cold saturated aqueous NaHCO\(_3\) (30.Oml) with stirring. The aqueous phase was then extracted chloroform (3×60.0ml), and the combined organic extracts were dried (Na\(_2\)SO\(_4\)) and the solvent was removed under vacuum to afford the crude product which was purified by flash column chromatography over silica to regain unreacted starting material (129mg, 27%) and afford the pure product(291mg, 58%) as a white solid. \(\delta\)\(_{n}\) (250MHz; CDCl\(_3\))

7.37-7.21(5H, m, ArH), 7.25(2H, d, J=8.8Hz, 2×ArH), 6.87(2H, d, J=8.8Hz, 2×ArH),
5.54(1H, d, J=7.8Hz, NH), 4.90(1H, d, J=9.1Hz, H-1), 4.73(2H, q, J\(_{AB}\) =11.7Hz, CH\(_2\)),
4.50(2H, q, J\(_{AB}\) =11.6Hz, CH\(_2\)), 3.82(1H, dd, J=10.2, J=8.6Hz, H3), 3.79(3H, s, OCH\(_3\)),
3.76-3.64(3H, m, H-4 and 2×H6), 3.58-3.80(1H, m, H5), 3.43-3.32(1H, m, H2),
2.94(1H, s(br), OH) and 188(3H, s, COCH\(_3\)), \(\delta\)\(_c\) , 170.0, 159.0, and 138.1 (qC),
129.4(2×ArCH), 128.5(2×ArCH), 128.1(2×ArCH), 127.9(ArCH), 113.8(2×ArCH),
87.7(CH), 79.9(CH), 75.7(CH), 74.1(CH\(_2\)), 73.3(CH\(_2\)), 72.8(CH), 69.6(CH\(_2\)), 55.8(CH),
55.2(OCH\(_3\)) and 23.4(COCH\(_3\)), Found: m/z (FAB-MS), 457.(MH\(^+\)), C\(_{23}\)H\(_{22}\)N\(_4\)O\(_6\) requires m/z, 456.5(MH\(^+\)).
2-Phthalamido-2-Deoxy-1,3,4,6-Tetra-O-Acetyl-\(\beta\)-D-Glucopyranose(71)\(^{127}\).

\[
\text{HO} \quad \text{HO} \quad \text{OH} \quad \text{NH}_2 \quad \text{HN}\]

1. NaOMe/MeOH
2. Phthalic anhydride/ Et\(_3\)N
3. Ac\(_2\)O/ pyridine

To a stirred solution of sodium methoxide (3.0g; 46.3mmol) in anhydrous methanol (75.0cm\(^3\)) was added glucosamine hydrochloride (10.0g; 46.3mmol). The reaction mixture was stirred for 10min at room temperature and then filtered with suction. Phthalic anhydride (3.5g; 23mmol) was then added to the filtrate and stirring was continued for a further 20mins. A further portion of phthalic anhydride (3.5g; 23mmol) was then added followed by triethylamine (7.6ml; 55.6mmol). The reaction mixture was stirred at room temperature for 10 mins and then at 50°C for 0.5h. The resulting mixture (containing a thick white precipitate) was cooled for 1h in an ice bath and then filtered with suction. The precipitate was washed with cold methanol (2x20.0cm\(^3\)) and dried under high vacuum. The dry white solid was then suspended in acetic anhydride (44.5cm\(^3\)), cooled to ice bath temperature and then pyridine (22.7cm\(^3\)) was added carefully with stirring. The reaction was then stirred at room temperature for 16h.

After 16h the reaction mixture was poured into ice/water (200cm\(^3\)) and extracted with chloroform (3x200cm\(^3\)). The combined organic extracts were washed with 5% HCl (1x120cm\(^3\)), saturated NaHCO\(_3\) (1x120cm\(^3\)), water (1x120cm\(^3\)) and brine (1x100cm\(^3\)). The organic phase was then dried (Na\(_2\)SO\(_4\)), filtered and the solvent was removed under vacuum to afford the crude product as an orange oil which was purified by flash column chromatography over silica (hexane/Ethyl acetate 1:1) to afford the product (8.65g, 40%) as a white foam. \(\delta_H\) (250MHz; CDCl\(_3\)) 7.86-7.75(4H, m, 4xArH), 6.50(1H, d, J\(_{1,2}\)=10.0Hz, H-1), 5.86(1H, dd, J\(_{3,4}\)=9.5Hz, H-3), 5.20(1H, dd, J\(_{4,5}\)=9.9Hz, H-4),
4.45(1H, dd, J\textsubscript{2,3}=10Hz, H-2), 4.37(1H, dd, J\textsubscript{H6b-H6a}=11.4Hz, J\textsubscript{H6b-H5}=4.1Hz, H-6\textsubscript{b}), 4.12(1H, dd, J\textsubscript{H6a-H6b}=2.9Hz, H-6\textsubscript{a}), 4.03(1H, ddd, H-5), 2.11, 2.05 and 2.00 and 1.87(12H, 4\times s, 4\times COCH\textsubscript{3})ppm.

2-Phthalamido-2-Deoxy-3,4,6-tri-O-Acetyl-1-Thioethyl-\beta-D-glucopyranoside(71)\textsuperscript{144}.

To a stirred solution of 2-phthalamido-2-deoxy-1, 3, 4, 6-tetra-O-acetyl-\beta-D-glucopyranoside (1.0g, 2.09mmol), ethane thiol (263\mu L, 3.56mmol) and 3A molecular sieves in anhydrous chloroform (2.0ml) under nitrogen was added boron trifluoride diethyl etherate (464\mu L, 3.77mmol) and the reaction mixture was stirred at room temperature.

After 16h the reaction mixture was diluted with chloroform (20.0ml) and washed with sat. aqueous NaHCO\textsubscript{3} (2\times20.0ml) and water (1\times20.0ml). The organic phase was dried [Na\textsubscript{2}SO\textsubscript{4}], filtered and the solvent was removed under vacuum to afford the crude product as a pale yellow foam which was subsequently purified by flash chromatography over silica (Pet. Ether 40-60°C/ EtOAc, 3:2) to afford the pure product (716mg, 71\% ) as a white foam. R\textsubscript{f}=0.26(Pet.ether/EtOAc, 3:2), \delta\textsubscript{H} (250MHZ; CDCl\textsubscript{3}) 7.86-7.71(4H, m, 4\times Ar\textsubscript{H}), 5.81(1H, dd, J\textsubscript{3,4}=9.1Hz, H-3), 5.47(1H, d, J\textsubscript{1,2}=10.6Hz, H-1), 5.16(1H, dd, J\textsubscript{4,5}=10.2Hz, H-4), 4.38(1H, dd, J\textsubscript{2,3}=10.3Hz, H-2), 4.29(1H, dd, J\textsubscript{H6b-H6a}=12.4Hz, J\textsubscript{H6b-H5}=4.9Hz, H-6\textsubscript{b}), 4.15(1H, dd, J\textsubscript{H6a-H6b}=2.3Hz, H-6\textsubscript{a}), 3.91-3.85(1H, m, H-5), 2.70 and 2.62(2H, 2\times dq, J\textsubscript{AB}=15Hz, J\textsubscript{CH2-CH3}=7.5Hz), 2.09, 2.02 and 1.84(9H, 3\times s, 3\times COCH\textsubscript{3}) and 1.20 (3H, t, J=7.4Hz, CH\textsubscript{3})ppm, \delta\textsubscript{C} (CDCl\textsubscript{3}) 170.6, 169.9, 169.4, 167.6 and 167.0 (6\times qC, 6\times CO), 134.3 and 134.2 (2\times CH, 2\times ArCH), 131.5 and 131.0 (2\times qC) 123.6 (2\times CH, 2\times ArCH), 81.1(CH), 75.8(CH), 71.4(CH), 68.7(CH), 62.2(CH\textsubscript{2}),
53.5(CH), 24.2(CH=CH), 20.6, 20.5 and 20.3(3×CO-CH₃) and 14.8(CH₂-CH₃) ppm. Found: m/z (FAB-MS), 480.1349(100%, MH⁺), C₂₂H₂₅NO₉S requires m/z, 480.1349 (MH⁺).

2-Phthalamido-2-Deoxy-4,6-O-p-Methoxybenzylidine-1-Thioethyl-β-D-glucopyranoside(72)

Sodium methoxide (0.5M, 0.2cm³, 1×10⁻⁴) was added to a stirred solution of 2-phthalamido-2-deoxy-3, 4, 6-tri-O-acetyl-1-thioethyl-β-D-glucopyranoside (0.5g, 1.04mmol) in anhydrous methanol (10.0cm³) under nitrogen and the reaction mixture was stirred at room temperature.

After 1h the reaction was quenched by the addition of glacial acetic acid (10μL). The solvent was removed under vacuum and toluene (10.0cm³) was evaporated from the residue.

The residue was then dissolved in anhydrous DMF (5.0cm³) and p-methoxy benzaldehyde dimethyl acetal (436mg, 2.39mmol) and p-toluene sulfonic acid (45mg, 0.25mmol) were added. The reaction mixture was then stirred at 50°C.

After 1h the reaction mixture was diluted with toluene (20.0ml) and washed with sat. aqueous NaHCO₃ (1×20.0ml). The aqueous phase was separated and extracted with further toluene (2×20.0ml). The combined organic extracts were washed with water (1×20.0ml), dried [Na₂SO₄], filtered and treated with silica gel (0.5g) for 5min. The silica was filtered off and the solvent was removed under vacuum to afford the crude
product as a yellow oil which was purified by flash chromatography over silica to afford the pure product (354mg, 72%) as a white solid. Rf=0.41 (3:2 Pet. Ether 40-60°C/ EtOAc), δ1H (250MHz; CDCl3) 7.85-7.66(4H, m, 4×ArH), 7.39(2H, d, J=8.6Hz, 2×m-CH), 6.87(2H, d, J=8.6Hz, 2×o-CH), 5.39(1H, s, CH), 5.37(1H, dd, J1,2=10.6Hz, H-1), 4.61(1H, dt, J3,4=12.2Hz, JH3-OH=3.5Hz, H-3), 4.35(1H, dd, JH6b-H6a=9.9Hz, JH6b-H5=4.2Hz, H-6b), 4.28(1H, dd, J2,3=10.1Hz, H-2), 3.81-3.76(1H, m, H-6a), 3.78(3H, s(br), OCH3), 3.64(1H, ddd, H-5), 3.54(1H, dd, J4,5=12.2Hz, H-4) 2.83(1H, d, J=3.5Hz, OH), 2.74-2.59(2H, m, CH2CH3), and 1.71 (3H, t, J=7.5Hz, CH3)ppm. δc (CDCl3) 168.1 and 167.6(2×qC, 2×CO), 160.1(qC), 134.0(2×CH, 2×ArCH), 131.5 and 131.4 (2×qC, ArC), 129.2(qC), 127.5(2×CH, 2×ArCH), 123.7 and 123.2(2×CH, 2×ArCH), 113.6 (2×CH, 2×ArCH), 101.7(CH, CH(O)2), 81.9(CH), 81.7(CH), 70.2(CH), 69.3(CH), 68.4(CH2, C6), 55.3(CH), 55.2(CH3, OCH3), 24.0(SCH2) and 14.7(CH2CH3). C24H25N07S requires m/z=471.5, found m/z (FAB-MS), 472(MH+, 59%).
Experimental Details For Chapter 5.

General Method for Refolded Protein Glycosylation with Glycosyl Iodoacetamides.
Ni\textsuperscript{2+} affinity purified EPO samples (3.0ml, 60\mu M, approximately 3.7mg) were reduced and refolded as previously described. The dilute refolded protein was then passed through a 0.45\mu m syringe filter (50.0ml syringe) onto a column of charged His-bind resin (1.0ml bed volume). After the refolding buffer had drained the immobilised protein was eluted (6M guanidine.\text{HCl}, 50mM Tris; pH8,400mM imidazole, 2.5ml) and the column was washed with further 6M guanidine.\text{HCl}, 50mM Tris; pH8 (2.5ml). The protein solution (approximately 20\mu M by UV) was then treated with the glycosyl iodoacetamide to a final concentration of 10mM (500\times excess). The reaction was then mixed on a blood rotator for 24h with the exclusion of light. The reaction was monitored by the removal of 0.5ml aliquots, which were precipitated using 10 volumes of 1:1 Acetone/ Methanol as previously described. The precipitate was dissolved in 6M guanidine.\text{HCl}, 50mM Tris; pH8 (20\mu L), reduced by the addition of DTT to a final concentration of 10mM and heating at 37°C (0.5h), and analysed using on-line LC-MS as described previously. The reactions when were complete when no further protein modification was observed in the ESI-MS spectrum.

General Method for 2-Nitro-5-Thiocyanato-Benzoic Acid (NTCBA) Analysis of \text{His}\textsubscript{10}-WT and Cysteine Mutant EPO's.
Column purified protein samples (typically 200\mu L; 50\mu M, 200\mu g) were precipitated using 10 volumes of 1:1 Acetone/ Methanol as previously described and redissolved in 100\mu L alkylation buffer (6M guanidinium hydrochloride, 0.5M Tris.\text{HCl}; pH 8, 20mM EDTA). DTT was then added to a final concentration of 10mM and the samples were incubated at 37°C for at least 3h. NTCBA was then added to a final concentration of 30mM and the resulting bright yellow/orange solution was allowed to stand at room temperature for 0.5h. The solution was then acidified to pH 4 with 30% TCA (60\mu L)
and the modified protein was precipitated from the now colourless solution as above. The precipitate was then taken up in 50μL cleavage buffer (3-6M guanidinium hydrochloride, 50mM Tris.HCl; pH10) and incubated at 37°C for 16h. The reaction mixture was then analysed directly using on-line LC-MS.

**Tryptic Digest Analysis of WThEPO.**

50μL of denatured protein samples in (Ni²⁺ affinity column) elution buffer at concentrations of typically 55μM (1.2mgml⁻¹) were diluted with 50μL dH₂O and then treated with 0.4μL of 15mgml⁻¹ trypsin (final protein:trypsin ratio of 20:1) and the resulting solution was incubated at 37°C for 16h. After 16h the cleavage was observed by LCMS.

Alternatively, fractions were collected from the LC-MS microbore HPLC system (every 10mins), lyophilised and the peptide mixtures were submitted for MALDI-TOF analysis.

**Tryptic Digest Analysis of Reduced WThEPO.**

As above except that the EPO sample was reduced with DTT to a final concentration of 10mM for 0.5 h and then the reducing agent was removed by precipitation of the protein sample, prior to incubation with trypsin.

**Tryptic/Chymotryptic Digest Analysis of Reduced and Immobilised EPO Samples.**

EPO samples were initially reduced with 10mM DTT for 1h at room temperature and the protein was separated from the reducing agent by protein precipitation as previously described. The protein samples were then redissolved in 6M guanidine.HCl or the minimum volume of 3M Guanidine.HCl and loaded onto a Ni²⁺ charged His-bind column (0.5 ml column bed). Any unbound material was removed by washing with 4 column volumes (2.0ml) of 3M guanidine.HCl, 50mM TrisHCl; pH8 (pH 7.3 for chymotrypsin). A solution of TCPK treated trypsin/chymotrypsin (0.1-1.0mgml⁻¹) in 3M guanidine.HCl, 50mM TrisHCl; pH8/pH7.3 (3.0ml) was applied to the column and 1.0ml of this solution was allowed to drain from the column. The column was then
plugged and the immobilised protein samples were exposed to the protease solution for a further 1h. The protease solution was then allowed to drain from the column and the resin was washed with 3M guanidine.HCl, 50mM TrisHCl; pH8 (10.0ml), 6M guanidine (1.0ml) and water (2.0ml). Hydrophillic fragments were then eluted from the column with 400mM imidazole and 0.25ml fractions were collected. 2.5µl of each fraction was spotted onto nitrocellulose membrane for western ‘dot-blotting’. Once dry the fractions were probed for N-terminal fragments using monoclonal murine anti human EPO antibody. Positive fractions were lyophilised. The lyophilised sampled were taken up in DCM (0.5ml) to dissolve the imidazole and the peptide fragments were collected by centrifugation. The supernatant was discarded and the remaining material was dissolved in 10µL of 1:1 acetonitrile/water, 0.1% TFA and 5mM DTT and submitted for MALDI-TOF analysis.

Neoglycoprotein Purification.

*Lectin affinity chromatography:*
Glycosylated His₁₀-EPO samples were purified using *Bandeiraea simplicifolia* II lectin (*Griffonia simplicifolia* II lectin; BS I) immobilised on 4% beaded agarose (0.5ml packed column bed volume). The column was initially equilibrated with 50mM sodium phosphate buffer; pH 7.2, 150mM NaCl. The sample to be purified was dialysed out of the glycosylating reaction mixture and into 50mM sodium phosphate buffer; pH 7.2, 150mM NaCl overnight at 4°C. This solution was then applied to the column using a syringe fitted with a 0.45µM filter and was allowed to stand at room temperature for 2h. The unbound material was then eluted and the column was washed with 50mM sodium phosphate buffer; pH 7.2, 150mM NaCl (3×1.0ml) to remove unglycosylated protein. The glycoprotein was then eluted using the same phosphate buffer containing 500mM N-acetyl glucosamine. Fractions could then be analysed by SDS-PAGE or ESI-MS after concentration.
CHAPTER 7

Experimental Details For Chapter 6.

General Experimental Details as for Chapter 2 except:

Deep Vent™ DNA polymerase was obtained from New England Biolabs.

1st round of PCR (synthesis of the megaprimer).

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template (double stranded pUC18 WThEPO)</td>
<td>1 µL</td>
</tr>
<tr>
<td>Primer (N83C)</td>
<td>5 µL</td>
</tr>
<tr>
<td>Primer (M13)</td>
<td>5 µL</td>
</tr>
<tr>
<td>10mM dNTP’s</td>
<td>5 µL</td>
</tr>
<tr>
<td>Thermopol™ (10× buffer)</td>
<td>5 µL</td>
</tr>
<tr>
<td>Deep Vent DNA Polymerase</td>
<td>1 µL</td>
</tr>
<tr>
<td>dH2O</td>
<td>28 µL  (to a final volume of 50 µL)</td>
</tr>
</tbody>
</table>

The reaction mixture was overlain with 50 µL of oil and subject to the following PCR cycling conditions:

96°C for 5 min (1 cycle)

96°C for 30 seconds
54°C for 15 seconds
60°C for 1 minute, 15 seconds (30 cycles)

4°C hold

The PCR product (ca 500 base pairs) was purified by horizontal agarose electrophoresis, further purified from the excised gel slice using the Quia-Quick™ DNA purification kit and used as a megaprimer for the 2nd round of PCR.
### 2nd Round of PCR.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template double stranded pUC18 WThEPO</td>
<td>1µL</td>
</tr>
<tr>
<td>Primer (N83C megaprimer)</td>
<td>13µL</td>
</tr>
<tr>
<td>Primer (NR)</td>
<td>5µL</td>
</tr>
<tr>
<td>10mM dNTP's</td>
<td>5µL</td>
</tr>
<tr>
<td>Thermopol™ (10× buffer)</td>
<td>5µL</td>
</tr>
<tr>
<td>Deep Vent DNA Polymerase</td>
<td>1µL</td>
</tr>
<tr>
<td>dH₂O</td>
<td>20µL (to a final volume of 50µL)</td>
</tr>
</tbody>
</table>

The PCR was then carried out as for the 1st round of PCR.

The PCR product (ca 1000 base pairs) was purified by horizontal agarose electrophoresis and further purified from the excised gel slice using the Quiagen Quick DNA purification kit (Quiagen) as before.

Deep Vent DNA polymerase gave enough of this product to use directly but the PCR product (from the 2nd round of PCR) could be amplified further as follows:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template (double stranded)</td>
<td>1µL</td>
</tr>
<tr>
<td>Primer (NR)</td>
<td>5µL</td>
</tr>
<tr>
<td>Primer (M13)</td>
<td>5µL</td>
</tr>
<tr>
<td>10mM dNTP's</td>
<td>5µL</td>
</tr>
<tr>
<td>Thermopol™ (10× buffer)</td>
<td>5µL</td>
</tr>
<tr>
<td>Deep Vent DNA Polymerase</td>
<td>1µL</td>
</tr>
<tr>
<td>dH₂O</td>
<td>28µL (to a final volume of 50µL)</td>
</tr>
</tbody>
</table>

The PCR products from the second or third round of PCR were digested with Nco I and HindIII to afford a sharp and intense band (ca 750 base pairs) corresponding to our
amplified fragment. This fragment was purified as before and ligated (as described in experimental details for chapter 2) into Neo I/ HinDIII cut pUC18-RBS.

Single transformants were chosen and screened HinD III (to linearise), Neo I and HinDIII (to excise the fragment) and HinC II (to prove mutation was incorporated) and positive clones were fully sequenced as described in experimental details for chapter 2.

This DNA was then excised form pUC18-RBS with Neo I and BamH I and ligated Neo I/ BamH I cut pET 16b to reintroduce our new C83 mutant into the expression vector.

This new construct, pET16b-C83hEPO(COPT), was then expressed and purified from E. coli as previously described (Chapter 2).
References:


13) Mehta, S., Meldal, M., Ferro, V., Duus, J.O., and Bock, K., Internally Quenched Fluorogenic, α-Helical Dimeric Peptides and Glycopeptides for the Evaluation of


CHAPTER 1


144) Ferrier, R.J., and Furneaux, R.H., Synthesis of 1,2-trans-Related 1-


146) Seifert, J., and Unverzagt, C., Synthesis of Three Biantennary N-Glycans


148) Tsarbopoulos, A., karas, M., Strupat, K., Pramanik, B.N., Nagabhushan, T.L.,
   Hillenkamp, F., Comparative Mapping of Recombinant Proteins and
   Glycoproteins by Plasma Desorption and Matrix-Assisted Laser
   (b) Billeci, T.M., and Stults, J.T., Tryptic Mapping of Recombinant Proteins by
   (c) Huberty, M.C., Vath, J.E., Yu, W., and Martin, S.A.,
   Site-Specific Carbohydrate Identification in Recombinant Proteins Using


   **182**(1990), 602-613.


154) Jacobson, G.R., Schaffer, M.H., Stark, G.R., and Vanaman, T.C., Specific
   Chemical Cleavage in High Yield at the Amino Peptide Bonds of Cysteine and


### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ac</td>
<td>Acetate</td>
</tr>
<tr>
<td>Bn</td>
<td>Benzyl</td>
</tr>
<tr>
<td>Bpt.</td>
<td>boiling point</td>
</tr>
<tr>
<td>Boc</td>
<td>t-butoxy carbonyl</td>
</tr>
<tr>
<td>DBMP</td>
<td>2,6-Ditert-butyl-4-methyl pyridine</td>
</tr>
<tr>
<td>DCC</td>
<td>Dicyclohexyl carbodiimide</td>
</tr>
<tr>
<td>DCM</td>
<td>Dichloromethane</td>
</tr>
<tr>
<td>DDQ</td>
<td>2,3-Dichloro-5,6-dicyano 1,4-benzoquinone</td>
</tr>
<tr>
<td>DEPT</td>
<td>Distortionless Enhancement by Polarisation Transfer</td>
</tr>
<tr>
<td>DMF</td>
<td>Dimethyl formamide</td>
</tr>
<tr>
<td>dNTP's</td>
<td>Deoxynucleotide triphosphates</td>
</tr>
<tr>
<td>Dol</td>
<td>Dolichol</td>
</tr>
<tr>
<td>EEDQ</td>
<td>2-ethoxy-1-N-ethoxycarbonyl-1,2-dihydroxyquinoline</td>
</tr>
<tr>
<td>EI-MS</td>
<td>Electron Impact mass spectrometry</td>
</tr>
<tr>
<td>EPO</td>
<td>Erythropoietin</td>
</tr>
<tr>
<td>ESI-MS</td>
<td>Electrospray ionisation mass spectrometry</td>
</tr>
<tr>
<td>FAB-MS</td>
<td>Fast atom bombardment mass spectrometry</td>
</tr>
<tr>
<td>Fc</td>
<td>Fragment (conserved region) of immunoglobulins.</td>
</tr>
<tr>
<td>Fmoc</td>
<td>9-Fluorenlymethoxy carbonyl</td>
</tr>
<tr>
<td>Fuc</td>
<td>Fucose</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>GalNAc</td>
<td>N-acetyl galactosamine</td>
</tr>
<tr>
<td>Glc</td>
<td>Glucose</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetyl glucosamine</td>
</tr>
<tr>
<td>GSH</td>
<td>Glutathione (reduced)</td>
</tr>
<tr>
<td>GSSG</td>
<td>Glutathione (oxidised)</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear Single quantum coherence</td>
</tr>
<tr>
<td>5-I-AEDANS</td>
<td>5-Iodoacetamido-ethylenediamino naphthalene sulfonic acid</td>
</tr>
<tr>
<td>IR</td>
<td>Infra-Red</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix Assisted Laser Desorption Ionisation-Time of Flight</td>
</tr>
<tr>
<td>Man</td>
<td>Mannose</td>
</tr>
<tr>
<td>MeOTf</td>
<td>Methyl triflate</td>
</tr>
<tr>
<td>Mpt.</td>
<td>Melting point</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonace</td>
</tr>
<tr>
<td>NTCBA</td>
<td>2-nitro-5-thiocyanatobenzoic acid</td>
</tr>
<tr>
<td>Pet.</td>
<td>Petroleum</td>
</tr>
<tr>
<td>Phth</td>
<td>Phthalimido</td>
</tr>
<tr>
<td>PMB</td>
<td>p-Methoxy benzyl</td>
</tr>
<tr>
<td>Py</td>
<td>Pyridine</td>
</tr>
<tr>
<td>PyBOP</td>
<td>Benzotriazolyloxy-tris(pyrrolidino)-phosphonium hexafluorophosphate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecylsulfate Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Sia</td>
<td>Sialic acid</td>
</tr>
<tr>
<td>TBAHS</td>
<td>Tetrabutylammonium hydrogen sulfate</td>
</tr>
<tr>
<td>TBAF</td>
<td>Tetrabutylammonium fluoride</td>
</tr>
<tr>
<td>TBDMS</td>
<td>Tertiary butyldimethyl silyl</td>
</tr>
<tr>
<td>THF</td>
<td>Tetrahydrofuran</td>
</tr>
<tr>
<td>TIC</td>
<td>Total Ion Chromatogram</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin Layer Chromatography</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>UV/Vis</td>
<td>Ultraviolet/ Visible</td>
</tr>
</tbody>
</table>