ELSEVIER

Chemistry & Biology 8 (2001) 133–145

Chemistry & Biology

www.elsevier.com/locate/chembiol

Research Paper

Selective in vitro glycosylation of recombinant proteins: semi-synthesis of novel homogeneous glycoforms of human erythropoietin

Derek Macmillan [1], Roslyn M. Bill [2], Karen A. Sage, Dominic Fern, Sabine L. Flitsch*

Department of Chemistry, University of Edinburgh, Kings Buildings, West Mains Road, Edinburgh EH9 3JJ, UK

Received 12 July 2000; revisions requested: 13 October 2000 revisions received 7 December 2000; accepted 8 December 2000; First published online 22 December 2000

Abstract

Background: A natural glycoprotein usually exists 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, although details of structure–function relationships are poorly understood, because of the microheterogeneity of biological samples. Hence, there is clearly a need for synthetic methods that give access to natural and unnatural homogeneously glycosylated proteins. The synthesis of novel glycoproteins through the selective reaction of glycosyl iodoacetamides with the thiol groups of cysteine residues, placed by site-directed mutagenesis at desired glycosylation sites has been developed. This provides a general method for the synthesis of homogeneously glycosylated proteins that carry saccharide side chains at natural or unnatural glycosylation sites. Here, we have shown that the approach can be applied to the glycoprotein hormone erythropoietin, an important therapeutic glycoprotein with three sites of *N*-glycosylation that are essential for in vivo biological activity.

Results: Wild-type recombinant erythropoietin and three mutants in which glycosylation site asparagine residues had been changed to cysteines (His_{10}-**WT**hEPO, His_{10}-Asn**24**Cys, His_{10}-Asn**38**Cys, His_{10}-Asn**83**CyshEPO) were overexpressed and purified in yields of 13 mg l^{-1} from *Escherichia coli*. Chemical glycosylation with glycosyl-β-*N*-iodoacetamides could be monitored by electrospray MS. Both in the wild-type and in the mutant proteins, the potential side reaction of the other four cysteine residues (all involved in disulfide bonds) were not observed. Yield of glycosylation was generally about 50% and purification of glycosylated protein from non-glycosylated protein was readily carried out using lectin affinity chromatography. Dynamic light scattering analysis of the purified glycoproteins suggested that the glycoforms produced were monomeric and folded identically to the wild-type protein.

Conclusions: Erythropoietin expressed in *E. coli* bearing specific Asn → Cys mutations at natural glycosylation sites can be glycosylated using β-*N*-glycosyl iodoacetamides even in the presence of two disulfide bonds. The findings provide the basis for further elaboration of the glycan structures and development of this general methodology for the synthesis of semi-synthetic glycoproteins.

Keywords: Erythropoietin; Human; In vitro glycosylation; Recombinant protein; Semi-synthetic glycoprotein

1. Introduction

Erythropoietin (EPO) is a glycoprotein hormone synthesized in humans by the foetal liver and adult kidney. As a mature glycoprotein, EPO regulates the production of erythrocytes operating under a classical feedback mechanism whereby elevated serum oxygen levels result in lower EPO expression and low oxygen concentrations result in increased EPO production [1–3]. In cases of chronic renal failure, the kidneys are no longer capable of producing EPO resulting in anemia, a condition which can be corrected by intravenous administration of recombinant human EPO (rhEPO), successfully negating the need of regular blood transfusions for patients on dialysis. It is therefore not surprising that rhEPO is among the

[1] Present address: Department of Chemistry, University of California, Berkeley, CA 94720, USA.

[2] Present address: The Lundberg Laboratory, Goeteborg University, P.O. Box 462, 405 30 Goeteborg, Sweden.

* Correspondence: Sabine L. Flitsch; E-mail: s.flitsch@ed.ac.uk

1074-5521/01/$ – see front matter
PII: S1074-5521(00)00065-X

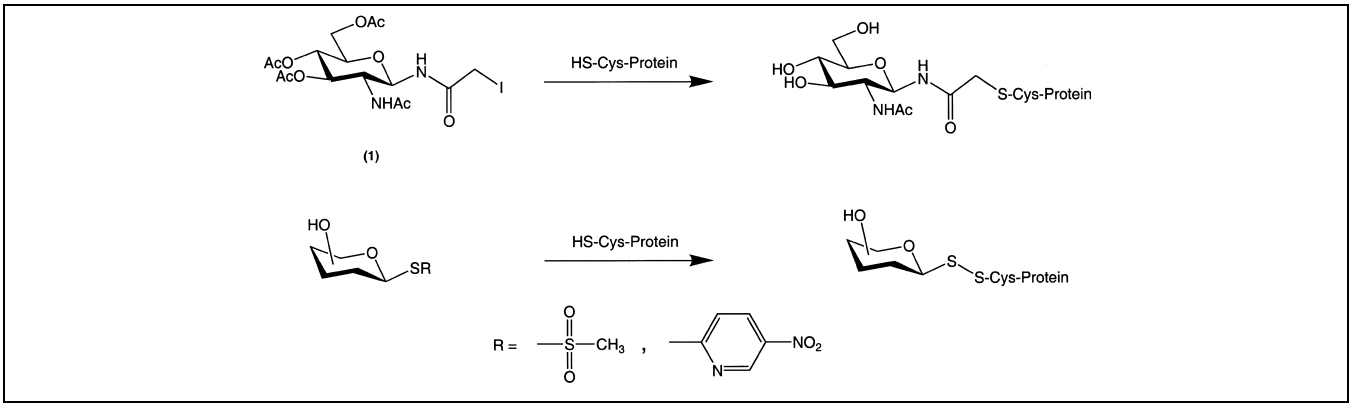


Fig. 1. Selective glycosylating reagents for cysteine side chains [37,41,42].

most successful recombinant therapeutics on the market [4].

EPO is a glycoprotein which exists as a heterogeneous mixture of glycoforms. With as many as 58 different *N*-linked glycans associated with a single sample [5], it has proved difficult to evaluate the effects of glycosylation on the structure and function of rhEPO raising many issues relating to its use as a therapeutic with regards to quality control, antigenicity and patenting. Many groups have reported on the importance of glycosylation for the in vivo biological activity of EPO [6–16]. Using methods such as expression of wild-type and mutant EPO in mammalian cell lines, expression in glycosylation 'deficient' mammalian cell lines, and expression of EPO in *Escherichia coli*, it has become possible to gain some insight as to which glycans are important and why. To summarise the findings briefly, the *O*-linked glycan at serine 138 was found not to be required for biological activity in vitro or in vivo, although it may serve to stabilise the folded structure. More importantly, the *N*-linked glycans at asparagines 24, 38 and 83 are essential for in vivo activity, increasing circulatory half-life through sialylation of tetra-antennary oligosaccharides (which comprise the major *N*-linked oligosaccharides on EPO) and allowing for correct folding and normal secretion (secretion of EPO is reduced to 10% in the absence of *N*-linked glycosylation), presumably through interaction with chaperones in the endoplasmic reticulum [17].

Although EPO is a perfect example of a therapeutic whose remarkable activity has intrigued biologists and chemists alike, employing all the techniques available to the glycobiologist in attempts to evaluate the precise or relative importance of the *N*-linked glycans, it also demonstrates the wider need for the development of methods which would further allow us to explore the role of glycosylation at the molecular level. As protein glycosylation is not under direct genetic control, the preparation of specific glycoforms may benefit from a chemical or chemoenzymatic in vitro approach.

Chemical approaches to the synthesis of homogeneous glycoproteins are complicated by two major factors: (1) Chemical synthesis of large proteins (>100 amino acids (aa)) is still a formidable task as automated solid phase peptide synthesis (SPPS) becomes inefficient for peptides >50 residues [18], a problem which may be soon overcome through novel developments in convergent protein synthesis such as native ligation and expressed protein ligation [18–24]. (2) Although new developments in carbohydrate chemistry have made the synthesis of large structurally defined glycans possible, oligosaccharide synthesis is still far from routine and its union with peptide synthesis less routine still, presenting a considerable challenge for the synthesis of glycopeptides containing as few as 20 amino acid residues which has been met by only a handful of groups [25–30].

While new synthetic methods for glycoprotein assembly continue to bear fruit, so does the synthesis of neoglycopeptides, or glycopeptide mimetics [31–33]. Neoglycopeptides and proteins have proven themselves to be invaluable tools for glycobiology for understanding carbohydrate–lectin interactions, proximal peptide–sugar conformational requirements and for the synthesis of immunostimulating glycoconjugates [34–36]. Continuing research in the group has been directed towards the development of a glycosylation strategy first reported by Flitsch and co-workers using a site-directed mutagenesis glycosylation approach [37–39] whereby free thiol groups of cysteine residues are selectively modified using glycosyl iodoacetamides (Fig. 1). The attractiveness of such an approach has been exploited by other groups [40,41] (Fig. 1), most recently by Jones and co-workers for the synthesis of robust and active glycoforms of the protease subtilisin (which is not naturally glycosylated) [42], in this case using alkane methylthiosulfonates as the thiol selective electrophiles, thus generating disulfide linkages.

It has been our aim to undertake a systematic study of the effects of specific glycans on the structure and activity of rhEPO through Asn→Cys mutations at the natural *N*-linked glycosylation sites (N24, N38, and N83), followed by site specific glycosylation after refolding, utilising the

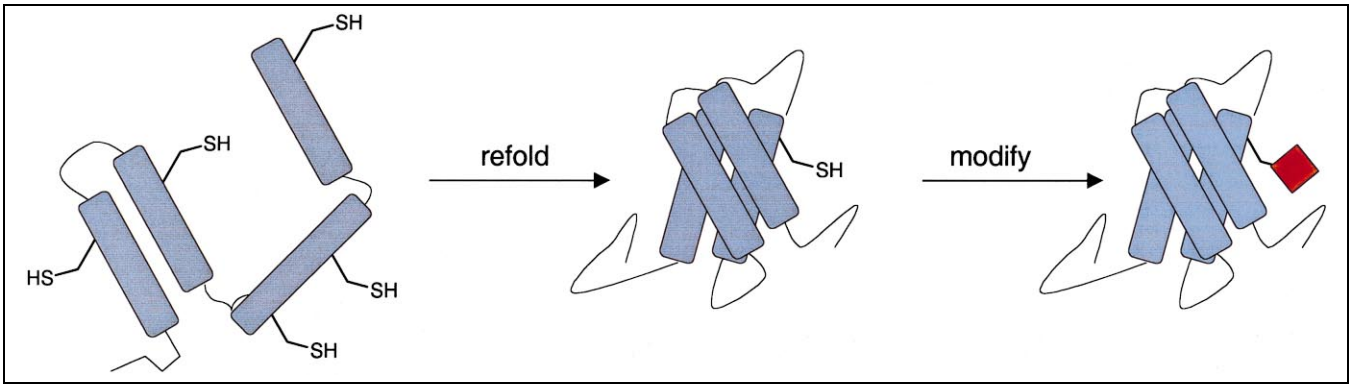


Fig. 2. Cysteine mutants are refolded to expose a single free thiol and the free thiol groups are glycosylated with glycosyl iodoacetamides (shown schematically in red).

strategy outlined in Fig. 2. This is considerably more challenging than in previous reports where thiol modification takes place on proteins bearing fortuitously free cysteine residues or on otherwise cysteine-free proteins which are not naturally glycosylated. Thus, EPO was chosen as a more realistic candidate for in vitro chemical glycosylation because it contains additional cysteine residues involved in disulfide linkages.

2. Results and discussion

2.1. Protein expression

His_{10}-WT EPO and cysteine muteins (N24C, N38C and N83C) were overexpressed and purified from *E. coli* (strain B834(DE3)) using the pET16b expression system as previously described by Boissel and co-workers [17] and essentially according to the manufacturer's instructions. EPO samples expressed in *E. coli* were found to be expressed in a highly aggregated, oxidised form with formed disulfides. In the case of the cysteine mutants the proteins were found to be covalently linked to glutathione so prior to oxidative refolding the samples were reduced.

The decahistidine sequence that had been added to the N-terminus of the recombinant protein allowed facile one step purification by metal chelation affinity chromatography. Surprisingly, rather than the expected single protein product, two proteins of similar molecular weight were shown to co-elute from the Ni^{2+} affinity column (Fig. 3, upper panels). This was also shown to be the case when the protein was overexpressed in *E. coli* strains BL21(DE3) and BL21(DE3)PlysS (data not shown). Both protein species were detected by a monoclonal antibody raised against an N-terminal epitope of rhEPO. MALDI-TOF analysis of the product mixture showed that the lower band of 21 097 Da was most likely to be His_{10}-C38hEPO (the calculated molecular weight for His_{10}-C38hEPO is 20 906 Da) and the upper was shown to be carrying an extending sequence. Analysis of the pET16b multiple cloning site indicated that the extending sequence could arise from an inefficient stop (TGA) codon, an occurrence also reported by MacBeath and coworkers when working with pET vectors [43,44]. The presence of a unique restriction site (*Stu*I) 26 base pairs from the 3′ end of the EPO gene allowed us to introduce a short synthetic oligonucleotide duplex, comprising a new stop codon (TAA) and optimised codons for protein expression in *E. coli*, using cassette mutagenesis (Table 1). Expression of the modified sequence and purification of His_{10}-EPO's from B834(DE3) cells afforded His_{10}-WT and cysteine mutein EPO in yields of 13 mg l^{-1} with no undesirable extending sequences (Fig. 3, lower panels).

2.2. Protein analysis by mass spectrometry

In order to monitor our glycosylation reactions, it was realised that the accuracy of mass spectrometer data had to be improved. The MALDI-TOF spectra were very helpful qualitatively but only provided masses within 50–150 Da of the calculated molecular weights. Increased resolution was achieved using on-line LC-ESI-MS of reduced protein samples. Protein samples were initially eluted from the Ni^{2+} affinity column as highly aggregated species (verified by non-reducing SDS–PAGE). Reduction with 10–50 mM dithiothreitol (DTT) or Tris–carboxyethyl phosphine (TCEP) prior to LC-MS afforded monomeric and highly charged species (Fig. 4), from which good MS data and accurate masses (within 5–20 Da) could be obtained for our protein samples (Table 2).

2.3. Oxidative refolding of protein

Protein samples were reduced after Ni^{2+} affinity purification, dialysed against 6 M guanidinium hydrochloride, 50 mM Tris–HCl; pH 8.0 under a nitrogen atmosphere, diluted 1:50 (to approximately 1 μM) and oxidatively refolded by dialysis against 2% *N*-lauroyl sarcosine, 50 mM Tris–HCl, pH 8.0, 40 μM $CuSO_4$ [17]. Refolded protein was then concentrated by passing the dilute protein solu-

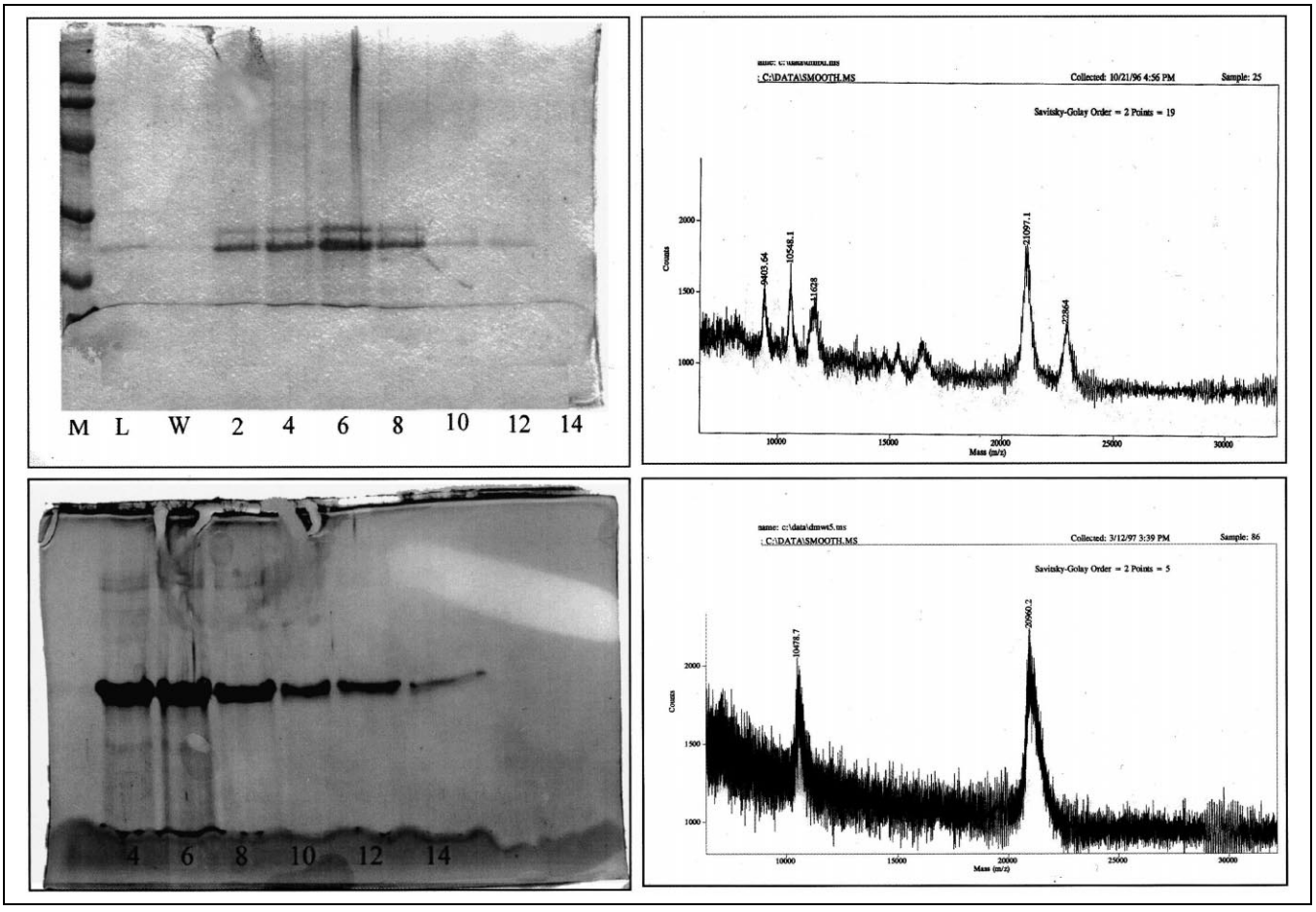


Fig. 3. Expression of the native EPO sequence in pET16b (upper panel, left) afforded two gene products, one carrying a 2 kDa extending sequence (upper panel, right). Expression of EPO, modified with a new stop codon results in a single species (lower panels).

tion, via a syringe filter, over 1.5 ml (bed volume) of Ni^{2+} charged affinity resin. Comparable refolding efficiency was also obtained when the protein samples were refolded using an alternative, 'pulsed', refolding strategy [45] prior to concentration.

2.4. Protein glycosylation

After purification, refolding and immobilisation of EPO samples on Ni^{2+} charged affinity resin protein samples were eluted from the affinity resin in 200 mM imidazole in 6 M guanidinium hydrochloride and the glycosyl iodoacetamide (**1**, Fig. 1) was added to a final concentration of 10 mM. The reaction was monitored by removing aliquots of the reaction mixture and determining the extent of reaction by LC-ESI-MS (Fig. 5). The reactions did not proceed to completion even with initial concentrations of GlcNAc iodoacetamide as high as 50 mM, but the addition of only a single GlcNAc residue was observed in all cases.

It should be noted that the addition of imidazole was important for achieving selective glycosylation. When refolded protein samples were concentrated by precipitation and redissolved in reaction buffer free of imidazole the reaction was found to be less selective for cysteine thiols and glycosylation was also observed on the N-terminal polypeptide sequence up to residue Ile6, possibly on the N-terminus or on the decahistidine sequence. This was determined using 2-nitro-5-thiocyanato-benzoic acid (NTCBA) mediated cleavage of the peptide backbone [46] followed by analysis using electrospray mass spectrometry (data not shown).

Table 1
Oligonucleotides for cassette mutagenesis

Oligonucleotide	Orientation	Sequence 5′→3′
EPOCOPT 1	sense	CC TGC CTG ACC GGT GAC CGT TAA TGA G
EPOCOPT 2	antisense	GA TCC TCA TTA ACG GTC ACC GGT ACG GCA GG

The TAA stop codon precedes the original TGA stop codon.

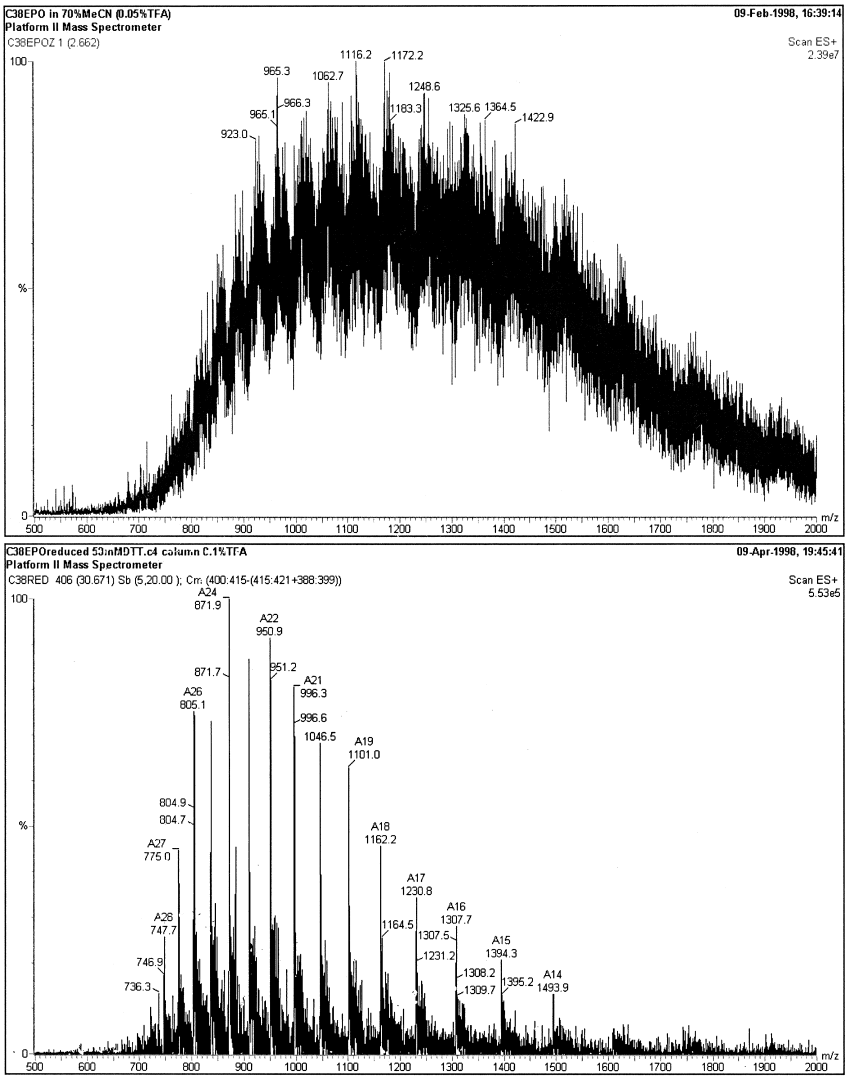


Fig. 4. ESI-MS charge spectra of aggregated (oxidised) His_{10}-C38EPO (upper panel) and reduced (50 mM DTT) His_{10}-C38hEPO (lower panel).

Interestingly, under denaturing conditions, His_{10}-C38EPO was most efficiently glycosylated ($>60\%$), followed by His_{10}-C83EPO and then His_{10}-C24EPO (30% as determined by ESI-MS). Protein glycosylation was less efficient when the protein samples were eluted from the affinity column and glycosylated after refolding.

2.5. Neoglycoprotein purification

Neoglycoproteins were readily purified from unglycosylated material using wheatgerm agglutinin lectin from *Triticum avesticum* immobilized on 4% beaded agarose. The reaction mixture was dialysed against lectin binding buffer (3 M guanidine–HCl, 50 mM Tris–HCl; pH 7.5, 1 M NaCl, 1 mM $MnCl_2$, 1 mM $CaCl_2$) and applied to the column. After washing unbound protein from the column, the glycoprotein was eluted using lectin binding buffer containing 500 mM *N*-acetyl glucosamine (GlcNAc) and glycoprotein fractions were monitored using electrospray LC-MS (Fig. 5).

2.6. Peptide mapping

After confirmation by mass spectrometry that only one

Table 2
Calculated and observed masses of reduced WThEPO, C24, C38 and C83 cysteine mutants

His_{10}-EPO species	Calculated mass (Da)		Observed mass (Da)
	Mono-isotopic	Average	
WT	20 903.8	20 916.9	20 912.6
C24	20 892.8	20 905.9	20 902.1
C38	20 892.8	20 905.9	20 902.8
C83	20 892.8	20 905.9	20 898.0

Mass differences between calculated and observed masses are likely to have resulted from differences in instrument calibration.

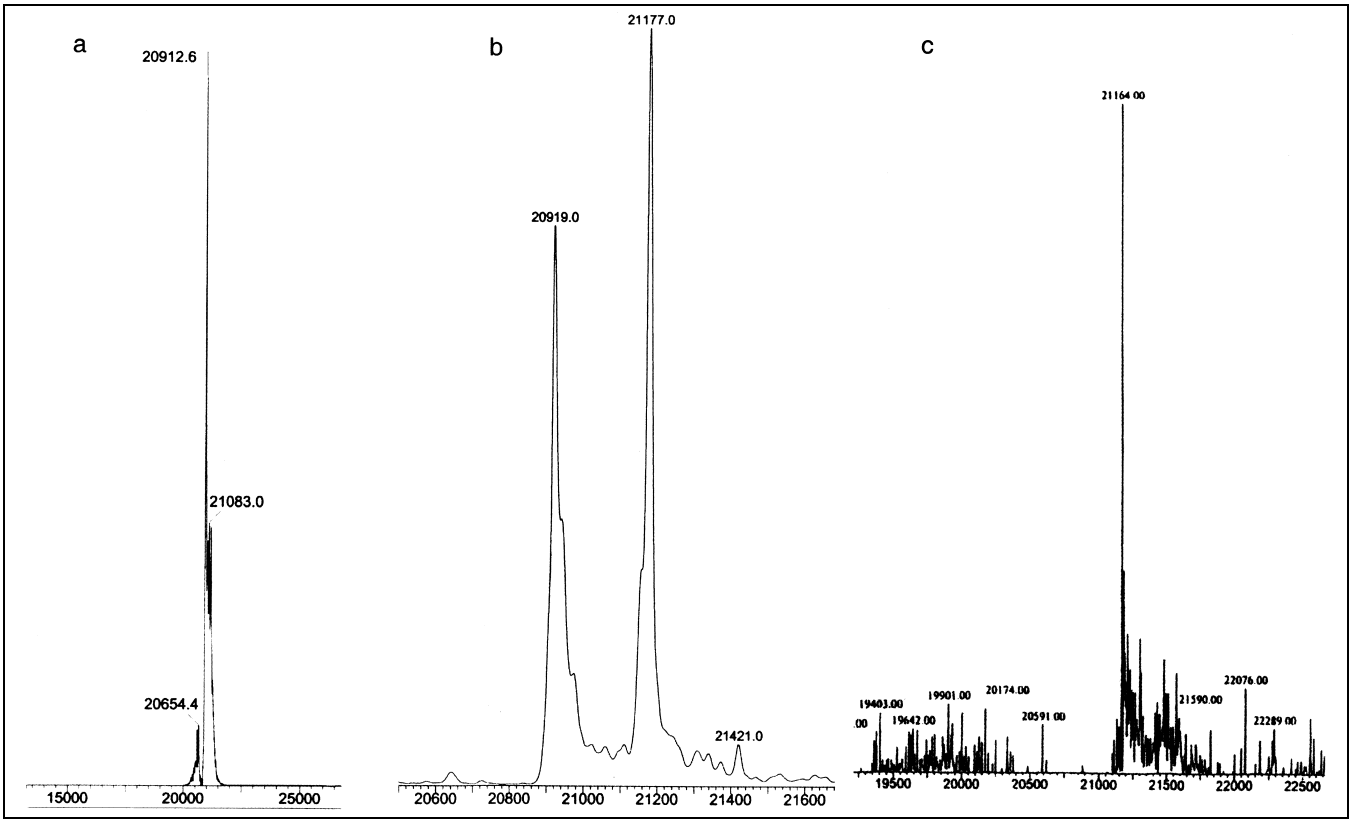


Fig. 5. (a) ESI-MS of reduced protein samples can provide accurate masses such that addition of sulfhydryl modifying reagents can be observed. (b) Optimised glycosylation conditions (plus imidazole in the reaction buffer) result in increased glycosylation of His_{10}-C38hEPO with almost no His-tag modification. (c) Lectin affinity purified His_{10}-C83(GlcNAc)hEPO.

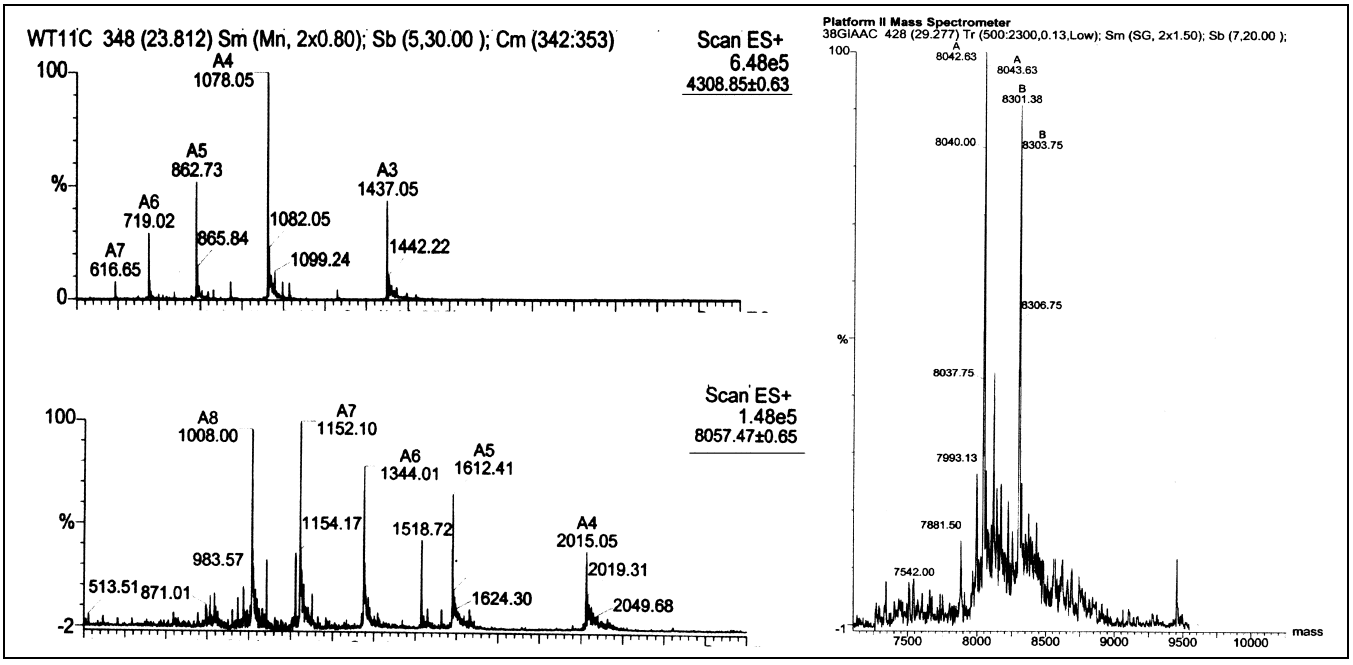


Fig. 6. Digestion of immobilised WThEPO with chymotrypsin resulted in two major fragments: −21 M to 15Y (upper left) and −21 M to 49Y (lower left). Digestion of the glycosylated C38 mutant with chymotrypsin also gave these fragments; less 11 Da for Asn→Cys mutation. In this case (right) the larger 8 kDa fragment is also glycosylated.

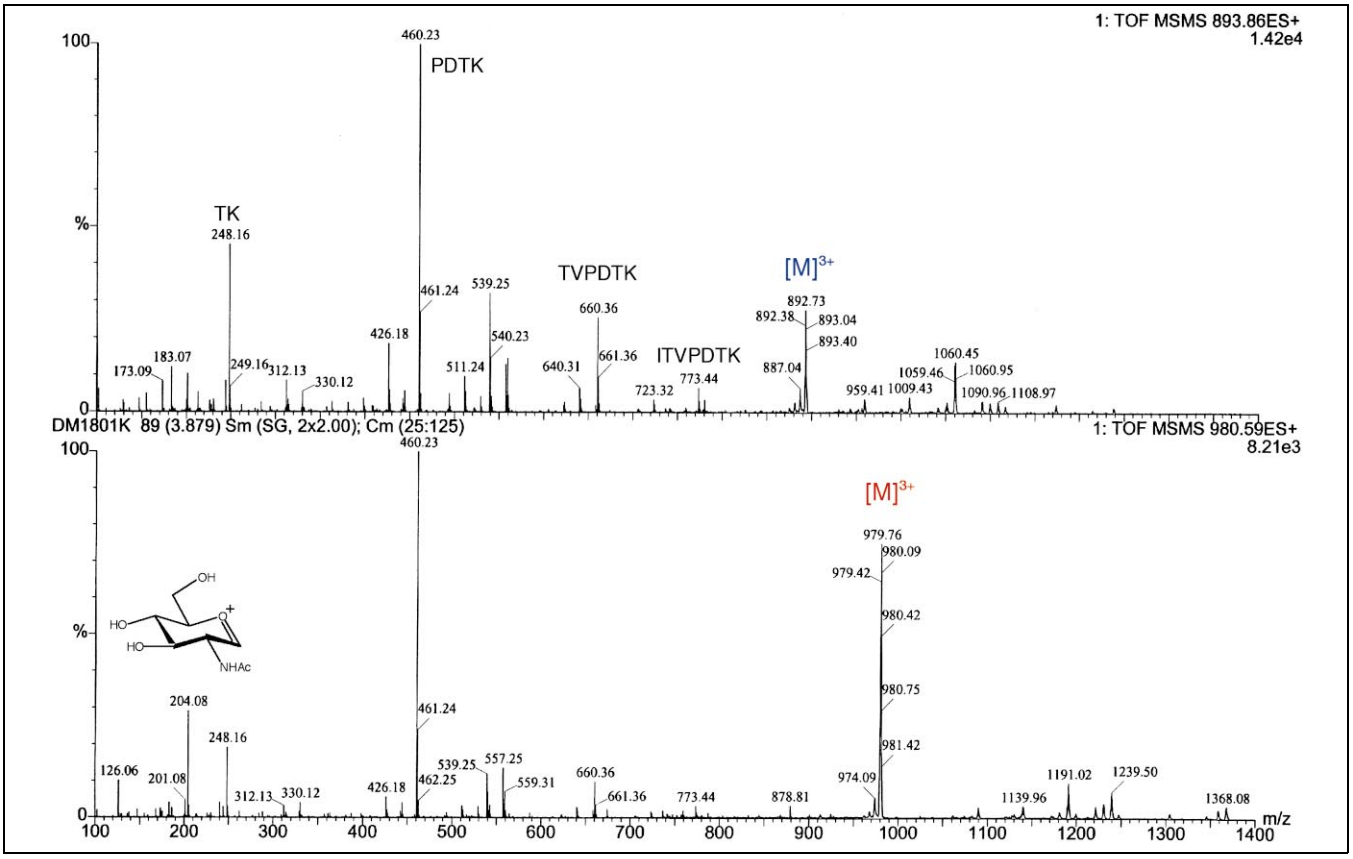


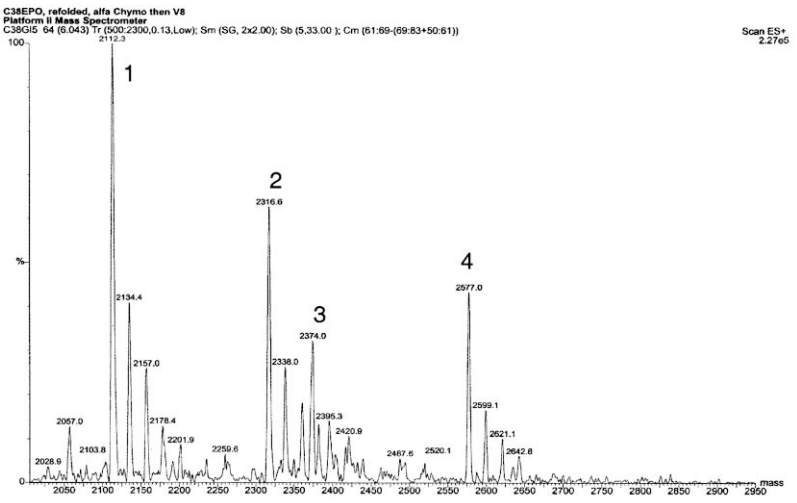
Fig. 7. MS-MS verified that the correct fragment was glycosylated. The [EAENITTG[29]**C**AEH[33]**C**SLN E[38]**C**ITVPDTK]$^{3+}$ ions corresponding to the unglycosylated (indicated in blue) and glycosylated (red) tryptic fragments were analysed. Although the sugar peptide linkage was not stable to the sequencing conditions (collision induced fragmentation of parent ion 979.76), the appearance of the characteristic molecular ion (*N*-acetyl hexosamines) at 204.08 indicated that it was in fact present. MS-MS sequencing also showed that both peptides contained similar daughter ions after collision induced dissociation and the peptide sequences were identical.

carbohydrate residue had been attached to each protein it was important to rigorously establish the site of chemical glycosylation on the EPO polypeptide chain. In this context it was encouraging to find that wild-type EPO expression in the same vectors as the cysteine mutants (i.e. differing only by the single Asn to Cys mutation) did not react with iodoacetamide **1** as confirmed by mass spectrometry (data not shown).

For more detailed analysis, several selective hydrolysis protocols, both chemical and enzymatic, were investigated in order to establish the glycosylation site by peptide mapping.

Although it was possible to separate glycosylated from unglycosylated protein using lectin affinity chromatography, the whole reaction mixture (glycosylated and unglycosylated protein) was analysed, since unglycosylated material functioned as a useful internal calibrant for mass spectrometry measurements. Initially, well-established proteolysis protocols for tryptic digests were studied. Thus, peptides were dissolved at concentrations of approximately 60 μM in 3 M guanidinium hydrochloride and digested with trypsin or chymotrypsin for 1–16 h. Analysis was complicated by the fact that proteolytic cleavage sites were clustered about the regions of interest affording small polar peptide fragments which failed to bind reverse phase columns and eluted with the buffer salts (for C24 and C38 mutants), diminishing their ability to be detected by the mass spectrometer.

Since the areas of interest were close to the N-terminal decahistidine sequence for mutants C24 and C38, reduced glycoprotein samples were immobilised on charged Ni^{2+} affinity resin and subjected to solutions of protease on resin. After washing, soluble protein fragments were eluted from the column with imidazole. Chymotrypsin was found to be most effective, reproducibly affording two soluble fragments with molecular weights of approximately 4.3 kDa and 8 kDa (Fig. 6). These soluble fragments could then be purified and subjected to further proteolytic cleavage by trypsin or V8 protease in volatile buffers (50 mM $NaHCO_3$, pH 8.0) compatible with LC-ESI-MS. Further analysis of these soluble peptides indicated that glycosylation had occurred on the desired fragment (between residues 21 and 45 as determined by tryptic digest; Fig. 7, and between residues 24 and 43 as determined by V8 proteolysis; Fig. 8). The glucosamine residue could not be completely assigned to the C38 (for His_{10}-C38hEPO) or C24 (for His_{10}-C24hEPO) as the observed peptide fragments also housed the cysteine residues of the native



Peak	Observed m/z	Corresponding sequence	Calculated average m/z	Calculated based on observed
1	2112.3	NITTG**C**AE̲H**C**SLNE̲**C**ITVPD	2121.4	-
2	2316.6	AENITTG**C**AE̲H**C**SLNE̲**C**ITVPD	2321.6	-
3	2374.0	1 + GlcNAc	2382.4	2373.3
4	2577.0	2+ GlcNAc	2582.6	2577.6

Fig. 8. Glycosylated His_{10}-C38EPO fragments produced by V8 proteolysis of the soluble 8 kDa chymotryptic fragment (see Fig. 6, right). Peaks 3 and 4 correspond to glycosylated 1 and 2 respectively. The smaller equidistant peaks are +23 Na^+ adducts.

C29–C33 disulfide bond. Cleavage with V8 protease repeatedly failed to cleave the peptide backbone at the glutamate residues within this short peptide sequence (Fig. 8). Isotopic analysis of the molecular ions in (Fig. 7) showed that this 3+ ion corresponds to a mass which was two mass units smaller than calculated indicating that a disulfide bond was present in this short peptide. Encouragingly, glycosylation was never observed on C7 or C161 (the cysteine residues which form a disulfide bond linking the N- and C-termini of the native folded structure). This was also true for C24hEPO. In the case of the C83hEPO mutant, the glycosylation site was on a larger peptide fragment, near no other cysteine residues, allowing the glycosylated fragment to be observed more readily (calculated molecular weight = 5868, observed molecular weight = 5866, Fig. 9) and allowing unambiguous assignment of the C83 glycosylation site.

As an alternative protein cleavage method, the chemical reagent NTCBA (2-nitro-5-thiocyano-benzoic acid) was explored, which is known to selectively cleave peptide

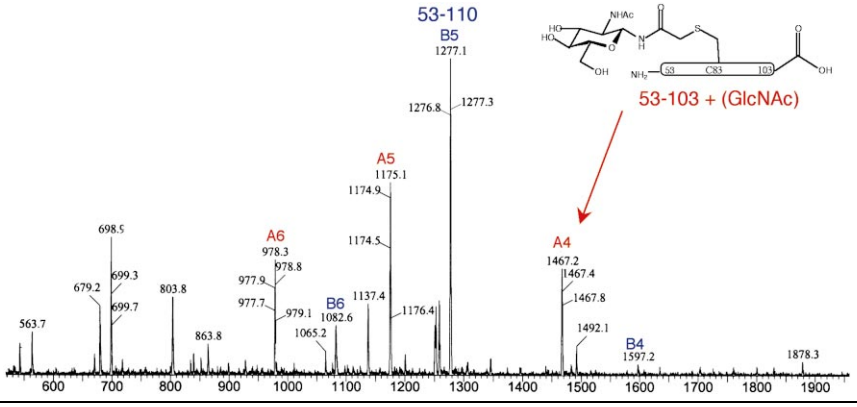


Fig. 9. Tryptic digest of glycosylated His_{10}-C83hEPO, shows glycosylation on C83.

Table 3
Calculated and observed masses for the peptide fragments resulting from NTCBA cleavage of His_{10}-C38hEPO

Fragment masses for His_{10}-C38hEPO	NTCBA fragments					
	Gly(-21)-Ile6	Cys7-Gly28	Cys29-His32	Cys33-Glu37	Cys38-Ala160	Cys161-End
Calculated	3186.5	2535.8	483.5	589.6	13 593.8	731.8
Observed	3186.0	2534.5	–	n/o	13592.5	n/o

Masses under 500 Da were not observed as they lay out with the acquisition range (500–2000 Da). n/o = not observed.

bonds adjacent to free cysteine residues. NTCBA treatment of the reduced wild-type EPO protein indeed produced fragments corresponding to cleavage at cysteine residues 7, 29, 33 and 161, and similar results were observed for cysteine mutants (Table 3). As would be predicted, glycosylation at specific sites of the cysteine mutants appeared to prevent NTCBA cleavage of the peptide backbone. Unfortunately it was not possible to observe the expected larger fragment of molecular weight = 14 412.3 Da, corresponding to glycosylation on C38 (fragment = C33-(C38(GlcNAc))-A160) because of apparent elimination of the carbohydrate from the glycosylated peptide fragment during NTCBA treatment resulting in a peptide with mass = 14 249 Da (163 Da smaller than predicted). Glycosylation was therefore monitored by proteolytic rather than chemical (NTCBA) cleavage.

However, NTCBA cleavage proved to be useful for determining the nature of the side product of the glycosylation reaction that was initially observed when treatment with iodoacetamide **1** was performed in the absence of imidazole. After NTCBA cleavage a glycosylated N-terminal fragment up to residue Ile6 was identified as containing the second glycosylation site. Since this glycosylated peptide fragment was stable to NCTBA cleavage conditions, this suggests that this second glycosylation site was due to a nucleophilic residue other than the thiol of cysteine.

2.7. Effects of glycosylation on EPO stability

Recombinant EPO expressed in *E. coli* easily aggregates and for nearly a decade had evaded complete structural analysis using standard biophysical techniques. Only recently have the crystal structures been solved of the EPO/EPO receptor complex [3] and NMR structure of *E. coli* derived EPO containing hydrophilic amino acid substitutions and additions [47].

Dynamic light scattering (DLS) [48] was considered ideal for initial studies of how glycosylation affects hEPO structure, as it was relatively insensitive to the high salt concentrations needed to solubilise EPO. Using a DynaPro-801 light scattering unit we could measure light scattered at an angle of 90° from a protein solution as it passes through an avalanche photodiode into an autocorrelator. Measuring the changes in light scattering as a function of time one can obtain the translational diffusion coefficient (D_T) and ultimately the hydrodynamic radius (R_H) using the Stokes–Einstein equation:

$$R_H = k_B T / 6\pi\eta D_T$$

Where k_B is the Boltzmann constant, T is the absolute temperature and η is the solvent viscosity. Using DLS we hoped to provide further evidence that our refolded glycoproteins were monomeric and provide some insight to their stability relative to unmodified protein. Non-reducing SDS–PAGE analysis had indicated that our refolded proteins were monomeric and not aggregated (data not shown). These results were validated by DLS (Fig. 10). By measuring light scattering as a function of increasing concentration of guanidine hydrochloride the resulting 'unfolding curve' showed a smooth transition to a 'random coil' state of the protein with a mid-point of about 2 M guanidine hydrochloride for both the His_{10}-WThEPO and the purified glycosylated His_{10}-C83hEPO. The data compare well with previous findings [16] where circular dichroism spectroscopy was employed to determine a mid-point for unfolding at 1.75 M guanidine hy-

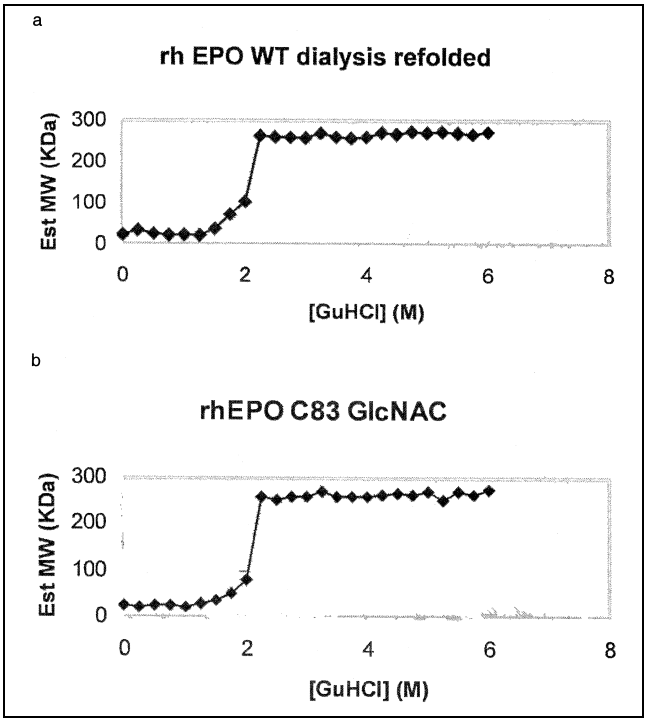


Fig. 10. Dynamic light scattering in the presence of increasing concentrations of guanidinium hydrochloride show little difference in stability of His_{10}-C83(GlcNAc)hEPO relative to His_{10}-WThEPO.

drochloride for Chinese hamster ovary (CHO) cell derived (glycosylated) EPO and 1.2 M for *E. coli* derived EPO. Our results suggest that there is no significant destabilisation resulting from the introduction of the novel glycoamino acid.

3. Significance

The *N*-glycosylation of proteins is an important co-translational modification greatly implicated in protein folding, stability and function [18,34,49]. As a result of not being under direct genetic control, new methods for the synthesis of glycoproteins and neoglycoproteins should prove valuable. Through site-directed mutagenesis it was possible to successfully introduce single Asn → Cys mutations at the natural *N*-linked glycosylation sites of erythropoietin. An N-terminal decahistidine tag allowed efficient expression and purification of protein samples. The mutants were refolded according to literature procedures with no significant loss of material. The refolded protein samples were then treated with an excess of the glycosyl iodoacetamide **1**, which afforded only singly glycosylated products. Rigorous ESI-MS analysis of the resulting glycoproteins showed glycosylation on the desired region of the protein. Lectin affinity chromatography was then employed to separate glycosylated protein from unglycosylated protein and the stability of purified glycoforms could be evaluated using dynamic light scattering. This showed that the proximal *N*-acetyl glucosamine residue did not considerably affect the stability of the muteins relative to the wild-type erythropoietin. To our knowledge this is the first example where a natural glycosylation site has been targeted for chemical glycosylation using the present methodology. It is also the first report in which the method of using engineered cysteine glycosylation sites has been used in the presence of other cysteine side chains involved in disulfide linkages. The results suggest that the method of cysteine mutagenesis combined with chemical glycosylation is an attractive and general method for generating neoglycoproteins with structural properties very similar to natural glycoproteins. This method also allows introduction of larger and more structurally complex glycans which will be the focus of future research.

4. Materials and methods

4.1. General experimental details

Genes encoding WThEPO and cysteine mutants [38] were inserted (*Nde*I/*Bam*HI) into pET16b (Novagen). Proteins were expressed and purified from *E. coli* essentially according to the manufacturer's instructions and as described by Boissell and coworkers. His-bind resin was from Novagen, All restriction endonucleases and DNA modifying enzymes were obtained from New England Biolabs and were used according to the suppliers instructions. 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 6 M solutions were filtered through a 0.45 μm filter (from Millipore) prior to use. All other routine laboratory reagents were from Sigma. Mass spectrometry of protein samples and modified protein samples was carried out on a Micro-Mass Platform ESI mass spectrometer. ESI spectra were recorded at a cone voltage of 70 V. Data was analysed using mass lynx v2.3 software. **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 280 nm. Flow rate = 0.05 ml min^{-1}. Samples were purified using a PHENOMENEX Jupiter, microbore, C4 (1 × 150 mm) reverse phase HPLC column (to remove trace amounts of salt and small molecules) coupled to ESI-MS. Solvents for HPLC were filtered before use. *HPLC conditions*: 10–100% acetonitrile over 20 min, 100% acetonitrile (10 min) then 100–10% acetonitrile over 10 min.

4.2. Optimisation of EPO expression by the incorporation of a new C-terminal sequence

Oligonucleotides (Table 1) were prepared by Gibco BRL and were obtained fully deprotected and desalted. Equimolar quantities of oligonucleotides EPOCOPT 1 and EPOCOPT 2 were mixed and diluted to 200 μl with sterile water (final concentration = 19.1 μM). The mixture was then heated in a boiling water bath for 5 min and allowed to cool to room temperature. The synthetic duplex was then ligated into *Bam*HI/*Stu*I cut pET16b-WThEPO. After positives were identified, genes encoding the cysteine mutants were subcloned (*Nde*I/*Stu*I) into this optimised vector. All optimised constructs were fully sequenced.

4.3. Renaturation of His_{10}-EPO samples

Ni^{2+} affinity purified protein samples (30–60 μM (determined using an estimated extinction coefficient of 22 430 M^{-1} cm^{-1})) 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 Ellmans reagent: A 1.0 ml sample of the dialysis buffer was treated with 10 μl 0.1 M Ellman's reagent and absorption at 412 nm was measured. Dialysis was complete when $A_{412\,nm} = 0$. 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 five column volumes of refolding buffer) and eluted with refolding buffer containing 500 mM imidazole. Yields for the refolding step were determined by absorption at 280 nm and found to be in excess of 80%.

4.4. LC-MS analysis of intact or unmodified WT, C24, C38 and C83 EPO samples

Unmodified His_{10}-WT and mutant EPO's were analysed directly by LC-MS spectrometry after Ni^{2+} affinity chromatography and reduction with 10–50 mM DTT at 37°C for 0.5 h (although LC-MS can be carried out on unreduced samples, unreduced samples gave very poor, often unresolvable charge spectra).

4.5. General method for refolded protein glycosylation with glycosyl iodoacetamides

Ni^{2+} affinity purified EPO samples (3.0 ml, 60 μM, approximately 3.7 mg) were reduced and refolded as previously described. The dilute refolded protein was then passed through a 0.45 μm syringe filter (50.0 ml syringe) onto a column of charged His-bind resin (1.0 ml bed volume). After the refolding buffer had drained the immobilised protein was eluted (6 M guanidine–HCl, 50 mM Tris, pH 8, 400 mM imidazole, 2.5 ml) and the column was washed with further 6 M guanidine–HCl, 50 mM Tris, pH 8 (2.5 ml). The protein solution (approximately 20 μM as measured by UV) was then treated with the glycosyl iodoacetamide to a final concentration of 10 mM (500× excess). The reaction was then mixed on a blood rotator for 24 h with the exclusion of light. The reaction was monitored by the removal of 0.5 ml aliquots, which were precipitated using 10 volumes of 1:1 acetone/methanol. Precipitation was completed by incubation on ice for 10 min and the precipitate was collected by centrifugation. The precipitate was dissolved in 6 M guanidine–HCl, 50 mM Tris, pH 8 (20 μl), reduced by the addition of DTT to a final concentration of 10 mM and heating at 37°C (0.5 h), and analysed using on-line LC-MS as described previously. The reactions were complete when no further protein modification was observed in the ESI-MS spectrum.

4.6. General method for 2-nitro-5-thiocyanato-benzoic acid (NTCBA) analysis of His_{10}-WT and cysteine mutant EPO's

Ni^{2+} affinity column purified protein samples (typically 200 μl; 50 μM, 200 μg) were precipitated using 10 volumes of 1:1 acetone/methanol as previously described and redissolved in 100 μl alkylation buffer (6 M guanidinium–hydrochloride, 0.5 M Tris–HCl, pH 8, 20 mM EDTA). DTT was then added to a final concentration of 10 mM and the samples were incubated at 37°C for at least 3 h. NTCBA was then added to a final concentration of 30 mM and the resulting yellow/orange solution was allowed to stand at room temperature for 0.5 h. The solution was acidified to pH 4 with 30% TCA (60 μl) and the modified protein was precipitated from the now colourless solution as above. The precipitate was taken up in 50 μl cleavage buffer (3–6 M guanidinium–hydrochloride, 50 mM Tris–HCl, pH 10) and incubated at 37°C for 16 h. The reaction mixture was then analysed directly using LC-MS.

4.7. Tryptic digest analysis of EPO samples

50 μl of denatured protein samples in (Ni^{2+} affinity column) elution buffer at concentrations of typically 55 μM (1.2 mg ml^{-1}) were diluted with 50 μl dH_2O, treated with 0.4 μl of 15 mg ml^{-1} trypsin (final protein:trypsin ratio of 20:1) and the resulting solution was incubated at 37°C for 16 h. After 16 h the cleavage was observed by LCMS.

4.8. Tryptic digest analysis of reduced EPO samples

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

4.9. Tryptic/chymotryptic digest analysis of reduced and immobilised EPO samples

EPO samples were initially reduced with 10 mM DTT for 1 h 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 6 M guanidine–HCl or the minimum volume of 3 M guanidine–HCl and loaded onto a Ni^{2+} charged His-bind column (0.5 ml column bed). Any unbound material was removed by washing with 4 column volumes (2.0 ml) of 3 M guanidine–HCl, 50 mM Tris–HCl, pH 7.3. A solution of chymotrypsin (0.1–1.0 mg ml^{-1}) in 3 M guanidine–HCl, 50 mM Tris–HCl, pH 7.3 (3.0 ml) was applied to the column and 1.0 ml 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 1 h. The protease solution was then allowed to drain from the column and the resin was washed with 3 M guanidine–HCl, 50 mM Tris–HCl, pH 8 (10.0 ml), 6 M guanidine (1.0 ml) and water (2.0 ml). Hydrophilic fragments were eluted from the column with 400 mM imidazole and 0.25 ml fractions were collected. 2.5 μl 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 (0.5 ml) to dissolve the imidazole and the peptide fragments were collected by centrifugation. The supernatant was discarded and the remaining material was dissolved in 10 μl of 1:1 acetonitrile/water, 0.1% TFA for LC-ESI-MS analysis or dissolved in 10 mM NH_4HCO_3 for further proteolysis.

4.10. Lectin affinity chromatography

Glycosylated His_{10}-EPO samples were purified using wheat-germ agglutinin lectin from *T. avesticum* immobilized on 4% beaded agarose (1.0 ml packed column bed volume). The column was initially equilibrated with lectin binding buffer (LBB): 3 M guanidine–hydrochloride, 50 mM Tris–HCl, pH 7.5, 1 M NaCl, 1 mM $MnCl_2$, 1 mM $CaCl_2$. The sample to be purified was dialysed

against LBB overnight at 4°C (or precipitated with acetone/methanol as previously described and then redissolved in LBB). This solution was applied to the column using a syringe fitted with a 0.45 μm filter. The unbound material was eluted and the column with 10 column volumes of LBB to remove unglycosylated protein. The glycoprotein was then eluted using the same buffer containing 500 mM *N*-acetyl glucosamine. Fractions could then be analysed by SDS–PAGE or ESI-MS after concentration.

4.11. Dynamic light scattering (DLS)

Protein samples were concentrated to approximately 2 mg ml^{-1} using a centricon (3 kDa molecular weight cutoff). The concentrated stock solution was then injected through a 0.1 μm filter into the light scattering cell until the lower cell was full (approximately 13 μl). The remaining stock solution was split into 300 μl aliquots. Guanidine–hydrochloride was added to each aliquot to form final concentrations of 0.25 M to 6 M in 0.25 M increments. Each aliquot was subjected to DLS to obtain an 'unfolding curve'.

Acknowledgements

We would like to thank Dr Dominic Campopiano and Dr Adam Gouldsworthy (Q-TOF) for helpful discussions and the BBSRC and the Edinburgh Centre for Protein Technology for financial support.

References

[1] T.B. Krantz, Erythropoietin, Blood 77 (1991) 419–434.
[2] F.K. Lin, S. Suggs, C.H. Lin, J.K. Browne, R. Smalling, J.C. Egrie, K. Chen, G.M. Fox, F. Martin, Z. Stabinsky, S.M. Badrawi, P.H. Lai, E. Goldwasser, Cloning and expression of the human erythropoietin gene, Proc. Natl. Acad. Sci. USA 82 (1985) 7580–7584.
[3] R.S. Syed, S.W. Reid, C. Li, J.C. Cheetham, K.H. Aoki, B. Liu, H. Zhan, T.D. Osslund, A.J. Chirino, J. Zhang, J. Finer-Moore, S. Elliott, K. Sitney, B.A. Katz, D.J. Matthews, J.J. Wendoloski, J. Egrie, R.M. Stroud, Efficiency of signalling through cytokine receptors depends critically on receptor orientation, Nature 395 (1998) 511–516.
[4] S. Gillis, D.E. Williams, Cytokine therapy: lessons learned and future challenges, Curr. Opin. Immunol. 10 (1998) 501–503.
[5] R.S. Rush, P.L. Derby, D.M. Smith, C. Merry, G. Rogers, M.F. Rohde, V. Katta, Microheterogeneity of erythropoietin carbohydrate structure, Anal. Chem. 67 (1995) 1442–1452.
[6] P.L. Storring, R.J. Tiplady, R.E.G. Das, B. Rafferty, G. Misty, Lectin-binding assays for the isoforms of human erythropoietin: comparison of urinary and four recombinant erythropoietins, J. Endocrinol. 150 (1996) 401–412.
[7] H. Sasaki, B. Bothner, A. Dell, M. Fukuda, Carbohydrate structure of erythropoietin expressed in chinese hamster ovary cells by a human erythropoietin cDNA, J. Biol. Chem. 262 (1987) 12059–12076.
[8] K. Yamaguchi, K. Akai, G. Kawanishi, M. Ueda, S. Masuda, R. Sasaki, Effects of site directed removal of *N*-glycosylation sites in human erythropoietin on it's production and biological properties, J. Biol. Chem. 266 (1991) 20434–20439.
[9] L.C. Wasley, G. Timony, P. Murtha, J. Stoudemire, A.J. Dorner, J. Caro, M. Krieger, R.J. Kaufman, The importance of *N*- and *O*-linked oligosaccharides for the biosynthesis and in vitro and in vivo biologic activities of erythropoietin, Blood 77 (1991) 2624–2632.
[10] S. Elliott, D. Chang, E. Delorme, C. Dunn, J. Egrie, J. Giffin, T. Lorenzini, C. Talbot, L. Hesterberg, Isolation and characterisation of conformation sensitive antierythropoietin monoclonal antibodies: effect of disulfide bonds and carbohydrate on recombinant human erythropoietin structure, Blood 87 (1996) 2714–2722.
[11] T. Misaizu, S. Matsuki, T.W. Strickland, M. Takeuchi, A. Kobata, S. Takasaki, Role of antennary structure of *N*-linked sugar chains in renal handling of recombinant human erythropoietin, Blood 86 (1995) 4097–4104.
[12] M.R. Fibi, P. Hermentin, J.U. Pauly, L. Lauffer, G. Zettlmeissl, *N*- and *O*-glycosylation muteins of recombinant human erythropoietin secreted from BHK-21 cells, Blood 85 (1995) 1229–1236.
[13] M. Higuchi, M. Oh-eda, H. Kuboniwa, K. Tomonoh, Y. Shimonaka, N. Ochi, Role of sugar chains in the expression of the biological activity of human erythropoietin, J. Biol. Chem. 267 (1992) 7703–7709.
[14] M. Takeuchi, N. Inoue, T.W. Strickland, M. Kubota, M. Wada, R. Shimizu, S. Hoshi, H. Kozutsumi, S. Takasaki, A. Kobata, Relationship between sugar chain structure and biological activity of recombinant human erythropoietin produced in chinese hamster ovary cells, Proc. Natl. Acad. Sci. USA 86 (1989) 7819–7822.
[15] L.O. Narhi, T. Arakawa, K.-H. Aoki, R. Elmore, M. Rohde, T. Boone, T.W. Strickland, The effect of carbohydrate on the structure and stability of erythropoietin, J. Biol. Chem. 266 (1991) 23022–23026.
[16] J.-P. Boissel, W.-R. Lee, S.R. Presnell, F.E. Cohen, H.F. Bun, Erythropoietin structure-function relationships, mutant proteins that test a model of tertiary structure, J. Biol. Chem. 268 (1993) 15983–15993.
[17] E.S. Trombetta, A. Helenius, Lectins as chaperones in protein folding, Curr. Opin. Struct. Biol. 8 (1998) 587–592.
[18] P.E. Dawson, S.B. Kent, Synthesis of native proteins by chemical ligation, Annu. Rev. Biochem. 69 (2000) 923–960.
[19] M. Holford, T.W. Muir, Adding splice to protein engineering, Structure 6 (1998) 951–956.
[20] X. Li, L. Zhang, S.E. Hall, J.P. Tam, A new ligation method for N-terminal tryptophan-containing peptides using the Pictet-Spengler reaction, Tetrahedron Lett. 41 (2000) 4069–4073.
[21] F.S. Gimble, Putting protein splicing to work, Chem. Biol. 5 (1998) R251–R256.
[22] P. Wang, R. Layfield, M. Landon, R.J. Mayer, R. Ramage, Transfer active ester condensation: a novel technique for peptide segment coupling, Tetrahedron Lett. 39 (1998) 8711–8714.
[23] Y.S. Shin, K.A. Winans, B.J. Backes, S.B.H. Kent, J.A. Ellman, C.R. Bertozzi, Fmoc-based synthesis of peptide-α thioesters: application to the total synthesis of a glycoprotein by native chemical ligation, J. Am. Chem. Soc. 121 (1999) 11684–11689.
[24] C.-H. Wong, T.J. Tolbert, Intein-mediated synthesis of peptides containing carbohydrates and other molecular probes, J. Am. Chem. Soc. 122 (2000) 5421–5428.
[25] Large, D.G., Warren, C.D. (Eds.), Glycopeptides and Related Compounds, Synthesis, Analysis and Applications, Marcel Dekker, New York, 1997.
[26] S.J. Danishefsky, S. Hu, P.F. Cirillo, M. Eckhardt, P.H. Seeberger, A highly convergent total synthetic route to glycopeptides carrying a high mannose core pentasaccharide domain *N*-linked to a natural peptide motif, Chem. Eur. J. 3 (1997) 1617–1628.
[27] C. Unverzagt, A modular system for the synthesis of complex *N*-glycans, Angew. Chem. Int. Ed. Engl. 36 (1997) 1989–1991.
[28] Z.-W. Guo, Y. Nakahara, T. Ogawa, A practical and efficient synthesis of complex-type biantennary heptasaccharide-asparagine conjugate, a key building block for the synthesis of complex *N*-linked glycopeptides, Tetrahedron Lett. 38 (1997) 4799–4802.
[29] E. Meinjohanns, M. Meldal, H. Paulsen, R.A. Dwek, K. Bock, Novel sequential solid-phase synthesis of *N*-linked glycopeptides from natural sources, J. Chem. Soc. Perkin Trans. 1 (1998) 549–560.

[30] E.C. Rodriguez, K.A. Winans, D.S. King, C.R. Bertozzi, A strategy for the chemoselective synthesis of *O*-linked glycopeptides with native sugar-peptide linkages, J. Am. Chem. Soc. 119 (1997) 9905–9906.
[31] C.-H. Wong, P. Sears, Carbohydrate mimetics: a new strategy for tackling the problem of carbohydrate-mediated biological recognition, Angew. Chem. Int. Ed. Engl. 38 (1999) 2300–2324.
[32] L.A. Marcaurelle, C.R. Bertozzi, New directions in the synthesis of glycopeptide mimetics, Chem. Eur. J. 5 (1999) 1384.
[33] B.G. Davis, Recent developments in glycoconjugates, J. Chem. Soc. Perkin Trans. 1 22 (1999) 3215–3237.
[34] B. Imperiali, S.E. O'Connor, Effect of *N*-linked glycosylation on glycopeptide and glycoprotein structure, Curr. Opin. Chem. Biol. 3 (1999) 643–649.
[35] S.J. Danishefsky, J.A. Allen, From the laboratory to the clinic: A retrospective on fully synthetic carbohydrate-based anticancer vaccines, Angew. Chem. Int. Ed. Engl. 39 (2000) 836–863.
[36] W. Dullenkopf, G. Ritter, S.R. Fortunato, L.J. Old, R.R. Schmidt, Synthesis of structurally defined antigen-immunostimulant combination for use in cancer vaccines, Chem. Eur. J. 5 (1999) 2432–2438.
[37] N.J. Davis, S.L. Flitsch, A novel method for the specific glycosylation of proteins, Tetrahedron Lett. 32 (1991) 6793–6796.
[38] R.M. Bill, P.C. Winter, C.M. McHale, V.M. Hodges, G.E. Elder, J. Caley, S.L. Flitsch, R. Bicknell, T.R.J. Lappin, Expression and mutagenesis of recombinant human and murine erythropoietins in *Escherichia coli*, Biochim. Biophys. Acta 1261 (1995) 35–43.
[39] Sage, K., The synthesis Of glycopeptides and glycoproteins, PhD Thesis, University of Oxford, 1996.
[40] S.Y.C. Wong, G.R. Guile, R.A. Dwek, G. Arsequell, Synthetic glycosylation of proteins using *N*-(beta-saccharide) iodoacetamides - applications in site specific glycosylation and solid phase enzymatic oligosaccharide synthesis, Biochem. J. 300 (1994) 843–850.
[41] Macindoe, W.M., Van-Oijen, A.H., Boons, G.J., A unique and highly facile method for synthesising disulfide linked neoglycoconjugates: a new approach for remodelling of peptides and proteins, J. Chem. Soc. Chem. Commun. (1998) 847–848.
[42] B.G. Davis, R.C. Lloyd, J.B. Jones, Controlled site selective glycosylation of proteins by combined site directed mutagenesis and chemical modification approach, J. Org. Chem. 63 (1998) 9614–9615.
[43] E.S. Poole, C.M. Brown, W.P. Tate, The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*, EMBO J. 14 (1995) 151–158.
[44] G. MacBeath, P. Kast, UGA read-through artifacts - when popular gene expression systems need a pATCH, Biotechniques 24 (1998) 789–794.
[45] H. Lilie, E. Schwarz, R. Rudolph, Advances in refolding of proteins produced in *E. coli*, Curr. Opin. Biotechnol. 9 (1998) 497–501.
[46] G.R. Jacobson, M.H. Schaffer, G.R. Stark, T.C. Vanaman, Specific chemical cleavage in high yield at the amino peptide bonds of cysteine and cystine residues, J. Biol. Chem. 248 (1973) 6583–6591.
[47] J.C. Cheetham, D.M. Smith, K.H. Aoki, J.L. Stevenson, T.J. Hoeffel, R.S. Syed, J. Egrie, T.S. Harvey, NMR structure of human erythropoietin and a comparison with its receptor bound conformation, Nat. Struct. Biol. 5 (1998) 861–866.
[48] Protein Solutions (ed.), Dynapro-801 operators manual.
[49] A. Varki, Biological roles of oligosaccharides: all of the theories are correct, Glycobiology 3 (1993) 97–130.