Exploring neoglycoprotein assembly through native chemical ligation using neoglycopeptide thioesters prepared *via* N→S acyl transfer†‡

Jonathan P. Richardson, Chung-Hei Chan, Javier Blanc, Mona Saadi and Derek Macmillan*

Received 1st October 2009, Accepted 17th December 2009
First published as an Advance Article on the web 22nd January 2010
DOI: 10.1039/b920535g

Sugars and simplified oligosaccharide "mimics" can be joined with protein fragments at pre-defined sites using reliable chemical reactions such as thiol alkylation and Cu(I) catalysed azide/acetylene ligation (click chemistry). These fragments have the potential to be assembled into neoglycoprotein therapeutics using native chemical ligation.

Introduction

Native chemical ligation (NCL) is a powerful method for protein synthesis and semi-synthesis that is particularly suited to the preparation of proteins bearing posttranslational modifications.[1] In the case of glycosylated proteins NCL is presently the only method to have demonstrated an ability to deliver proteins bearing multiple forms of defined glycosylation at pre-designated positions employing entirely native linkages.[2] While advances in carbohydrate chemistry have allowed elaborate oligosaccharide syntheses,[3] solid-phase oligosaccharide and glycopeptide synthesis still lacks the simplicity of oligopeptide and oligonucleotide assembly. The importance of oligosaccharides in living systems demands that complex homogeneous replica structures need be prepared to unravel biological functions though there are other potential applications such as improvement of the pharmacokinetic profile of therapeutic proteins, analysis of saccharide-lectin interactions and glycoconjugate targeting where the presentation of *similar* structures may suffice.[4] Furthermore, it is desirable that simpler chemical building blocks are available such that a variety of structures can be rapidly prepared by a non-expert and in a small number of synthetic steps.[5] Neoglycoproteins formed through global or site-specific modification of recombinant (usually bacterially derived) proteins have been extensively investigated though concern as to the possible immunogenic nature of these substances has driven the development of new methods that exert more control on the degree of glycosylation and introduce native sugar-protein linkages.[6] Of the synthetic and semi-synthetic methods available, glycoprotein assembly through NCL[2] and perhaps remodelling of simply glycosylated precursors,[7] which themselves may be more generally accessible through NCL, have emerged as being particularly effective. Significant improvements in NCL methodology combined with the typically mild reaction conditions employed and tolerance to densely functionalised starting materials has also rendered NCL an attractive method for assembling unnatural, and unnaturally modified proteins as research tools and diagnostics in addition to potential therapies.[8] While assembly of a glycoprotein can seem complex and laborious in terms of synthesis, strategic replacement of a few synthetically challenging motifs with alternatives formed through less capricious chemical reactions may greatly simplify the production of analogues such that they can be assembled outwith a handful of specialised laboratories. The advantages of combining chemoselective strategies for glycoconjugate synthesis and assembly of biomolecular multimers have recently been demonstrated[4d,9] and in this manuscript we describe our own progress towards glycoprotein analogues through combination of NCL, cysteine sulfhydryl modification with α-haloacetamide sugars[10] and Cu(I) catalysed azide/acetylene ligation ("click" chemistry).[11] In this case chemoselective couplings are used not only in glycoconjugation but also in the assembly of a simplified oligosaccharide mimic.

Neoglycopeptide thioester synthesis *via* N→S acyl shift

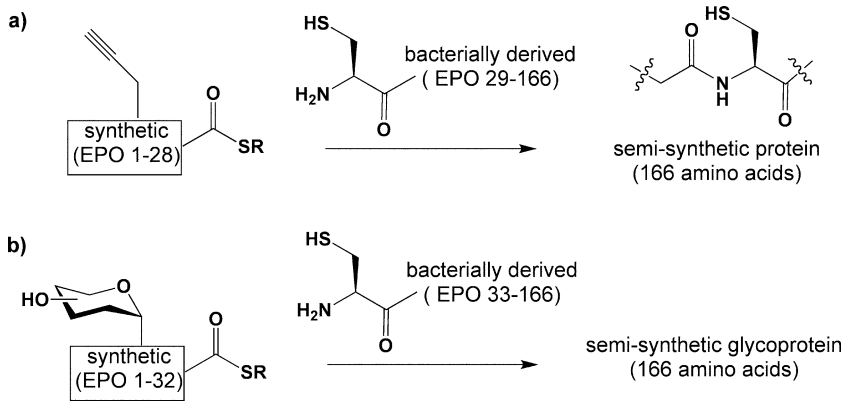
Previously we prepared an acetylene-modified biologically active variant of the 166 amino acid hormone erythropoietin (EPO) through NCL between a bacterially derived N-cysteinyl protein fragment and a synthetic acetylene-containing peptide thioester (Scheme 1a).[12] The thioester component was accessible, albeit in relatively low (5%) yield using the sulfonamide safety-catch linker for solid-phase peptide synthesis.[13]

The main difficulties we encountered were associated with efficient loading of the first amino acid and release of the assembled peptide chain from the solid support. Having recently developed a novel method for the production of peptide thioesters for NCL employing an N→S acyl shift[14] we were keen to apply an optimised procedure[15] in the context of glycopeptides. We believed that this approach might be more economical with regards to precious carbohydrate appendages since only a slight excess of sugar α-haloacetamide would be required to modify free thiol groups within a preassembled precursor peptide in solution (Scheme 2) whereas larger excesses of carbohydrates or glycoaminoacids are usually committed in solid phase peptide synthesis based on resin loading, *before* the peptide is recovered from the solid support.[16] Interestingly, during the course of our experiments we happened upon an unexpected improvement in

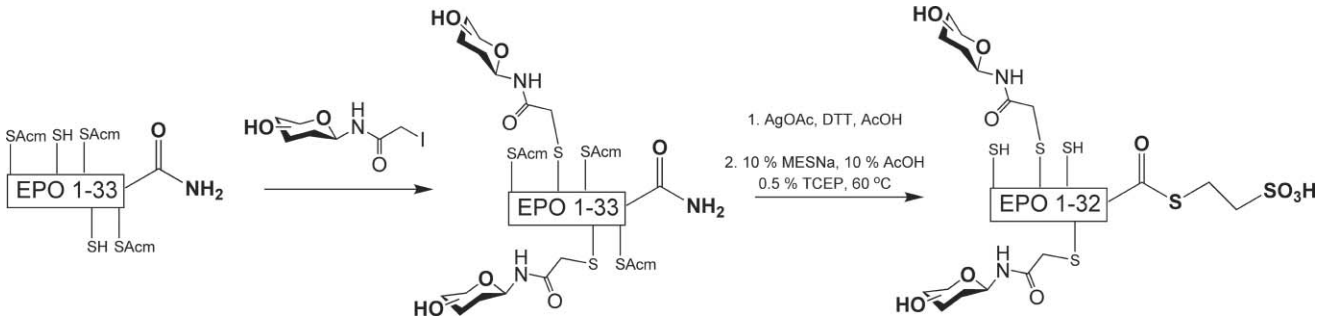
Department of Chemistry, University College London, 20 Gordon Street, London, WC1H 0AJ, UK. E-mail: d.macmillan@ucl.ac.uk

† This paper is part of an *Organic & Biomolecular Chemistry* web theme issue on protein labelling and chemical proteomics. Guest editors: Ed Tate and Matthew Bogyo

‡ Electronic supplementary information (ESI) available: UV-VIS and mass spectra. See DOI: 10.1039/b920535g



Scheme 1 Semi-synthetic approaches to modified EPO analogues employing NCL.[12]



Scheme 2 Proposed synthesis of a neoglycopeptide thioester. A preassembled peptide is alkylated at free sulfhydryl groups with sugar α-iodoacetamides (2 equivs per thiol). Next the remaining thiols, involved in disulfide bond formation and required for thioester formation *via* an N→S acyl shift, are exposed and the C-terminal His-Cys motif is converted to a C-terminal histidyl thioester.[15]

our peptide thioesterification reaction in a model system. In our preceding studies all peptides were prepared on Rink amide resin and possessed C-terminal carboxamides (*i.e.* contained a C-terminal Xaa-Cys-$CONH_2$ motif) or were *internal* with respect to the peptide sequence. We had not investigated the reaction in the context of a C-terminal cysteine with a free *carboxyl* group (*i.e.* containing Xaa-Cys-CO_2H) though considered this subtle difference a more accurate model for a recombinant sample that might ultimately be engineered to possess a C-terminal His-Cys motif. We were therefore very much surprised to find that when the terminal cysteine residue possessed a free carboxyl group (*e.g.* **1a**, Fig. 1) the rate of conversion to the corresponding thioesters approximately doubled relative to the carboxamide-terminated counterparts (**1b**, Fig. 1). This raised the possibility that the "+1" amide linkage dramatically influences a peptide's susceptibility to thioester formation in addition to the nature of the Xaa-Cys motif. Conceivably, thioesterification could be catalysed through intramolecular protonation of the scissile amide *N* or carbonyl *O* atoms by the C-terminal carboxyl group. If a similar explanation could be formulated for the role of acidic protons on His (**1**) and Cys (**3**) then this might help account for enhanced thioesterification at these sites. As the reaction was already conducted at low pH (pH 2) we considered this explanation unlikely though incorporated non-acidic variants of His (**2a/b**) and Cys (**4a/b**) to probe the observation in more detail. Importantly **2a/b** and **4a/b** facilitated thioesterification suggesting no essential role for these residues as general acids. Next we investigated whether the rate acceleration attributed to the free carboxyl allowed motifs (such as Ala-Cys) that would not be expected to form thioesters to any appreciable extent at 60 °C,[14] to become more favourable. Indeed the Ala-Cys terminated peptides **6a/b** did undergo thioesterification though more readily as terminal carboxylic acid **6a** than as carboxamide **6b**. Clearly the terminal nature of the motif also serves to increase conversion to thioester, perhaps through greater conformational flexibility which, in part, may account for the ease with which Gly-Cys motifs undergo thioesterification. The Ala-thioesters formed from **6a/b** appeared to undergo significant epimerisation during the course of the reaction. This was confirmed through synthesis and thioesterification of the corresponding D-Ala containing peptides (which also epimerised)[17] while **1a/b**, **2a/b** and **4a/b** showed no signs of epimerisation. Furthermore when the optimal His-Cys motif was transposed to the N-terminus of the peptide resulting in sequence H_2N-**HC**MEEYLKS-$CONH_2$ (**7**, Fig. 1)[18] the rate of thioesterification was slowed a little further than when installed as a C-terminal amide suggesting important stereochemical and electronic requirements for the reaction. The resulting *N*-cysteinyl peptide, itself a potentially useful product if unmasked in the context of a recombinant protein, was isolated and its identity confirmed by LC-MS and NCL reaction with the thioester formed from **6a**.

From this short study we concluded that the C-terminal carboxyl group can enhance the thioesterification reaction at Xaa-Cys motifs such that it can relax the previously observed requirement for Gly, His or Cys at the thioesterification site when the reaction is performed at 60 °C.[14] The effect is likely attributable to both a

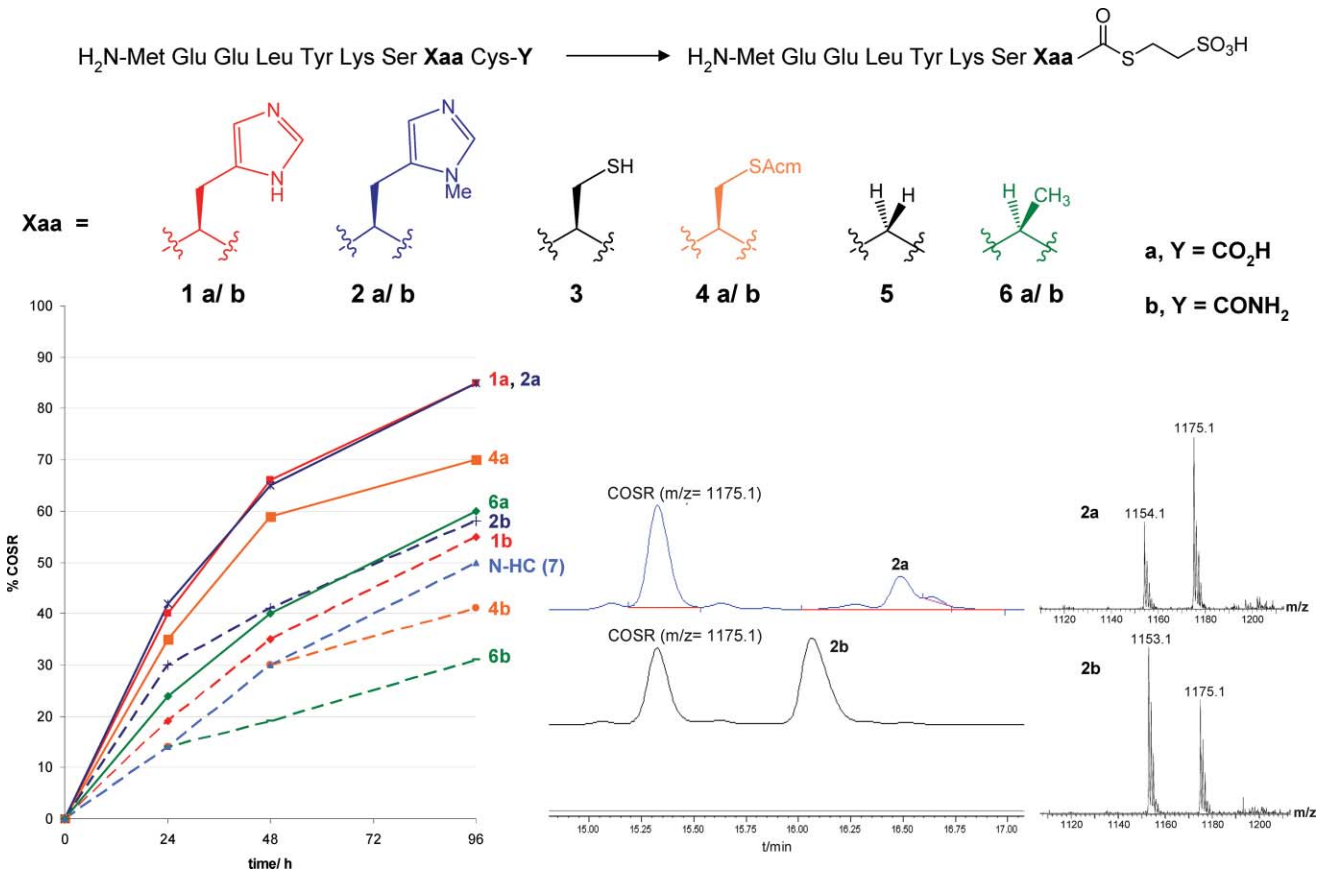


Fig. 1 Model peptides prepared to test the effect of the terminal functional group on thioesterification (each peptide was prepared on 0.02 mmol scale and isolated in 60–65% yield). Formation of thioester is represented with solid lines for C-terminal carboxylic acids and with broken lines for carboxamides. *Reagents and conditions*: a) approx. 1 mM peptide, guanidine.HCl, 10% v/v AcOH, 10% w/v MESNa, 0.5% w/v TCEP.HCl, 60 °C, 96 h. Note: there is generally good agreement between % conversion determined by HPLC peak integration at 230 nm and relative MS peak intensities of the crude reaction mixture (shown above for conversion of **2a** and **2b** to thioester after 48 h).

catalytic role for the carboxyl group and increased conformational flexibility associated with the terminal residue. Even Ala-Cys motifs were transformed (albeit with least efficiency) to thioesters at 60 °C when installed as C-terminal carboxylic acids, though epimerisation (approx. 10–15%) was observed after 48 h. Acting as a general acid catalyst, the high effective concentration of the C-terminal carboxyl may accelerate the breakdown of a carbinolamine intermediate as proposed for the essential histidine residue in the protein splicing mechanism orchestrated by inteins,[19] and in protein thioesterification catalysed by sortase.[20] In contrast to these systems our thioesterification reaction is tolerant to denaturants such as 6 M guanidine hydrochloride.

Ultimately, the significant improvement attributed to the terminal carboxyl group (Fig. 1) provided the impetus to prepare our neoglycopeptide thioester precursors on hydroxyl functionalised resins (*e.g.* Novasyn TGA) rather than the Rink amide resins we had employed thus far.[14,15] In anticipation of performing the NCL reaction with a bacterially derived protein fragment comprising residues 33–166 (Scheme 1b) we assembled EPO residues 1–33 containing a C-terminal HC motif for acid-mediated thioester formation (Fig. 2a).[21] Preparation of residues 1–33 allows incorporation of two glycosylation sites (Asn24 and Asn30) found in the enhanced erythropoiesis stimulating agent Aranesp™ and permits NCL to be performed using a native cysteine residue.[22] For optimal compatibility with our thioesterification the synthetic peptide sequence carried two mutations. The first, G28A was introduced to prevent undesirable thioesterification occurring across the Gly[28]-Cys[29] junction, our results from Fig. 1 suggesting that the difference in reactivity between thioesterification at a His-Cys (terminal) motif and an Ala-Cys (internal) motif should confer acceptable selectivity. The second mutation, D8E was introduced to prevent slow hydrolysis at aspartic acid which can occur at elevated temperature under the thioesterification reaction conditions upon prolonged heating, which is completely abolished by substitution with glutamic acid.[23] Furthermore, at positions to be modified with carbohydrates (residues 24 and 30) we introduced orthogonally (trityl) protected cysteine residues to the native cysteines (Cys 7 and Cys 29) and the C-terminal Cys, which were SAcm protected. The desired peptide sequence was therefore assembled in automated fashion on an Applied Biosystems 433A synthesiser and cleaved from the solid support under standard conditions. The crude product was purified by semi-preparative reverse-phase HPLC and isolated in 18% yield (Fig. 2b). Peptide **8** was dissolved in 50 mM NH_4HCO_3; pH 8.0 (2.0 mL) and incubated with the α-iodoacetamide of *N*-acetylglucosamine (approximately 2 equivalents per thiol) at room temperature for

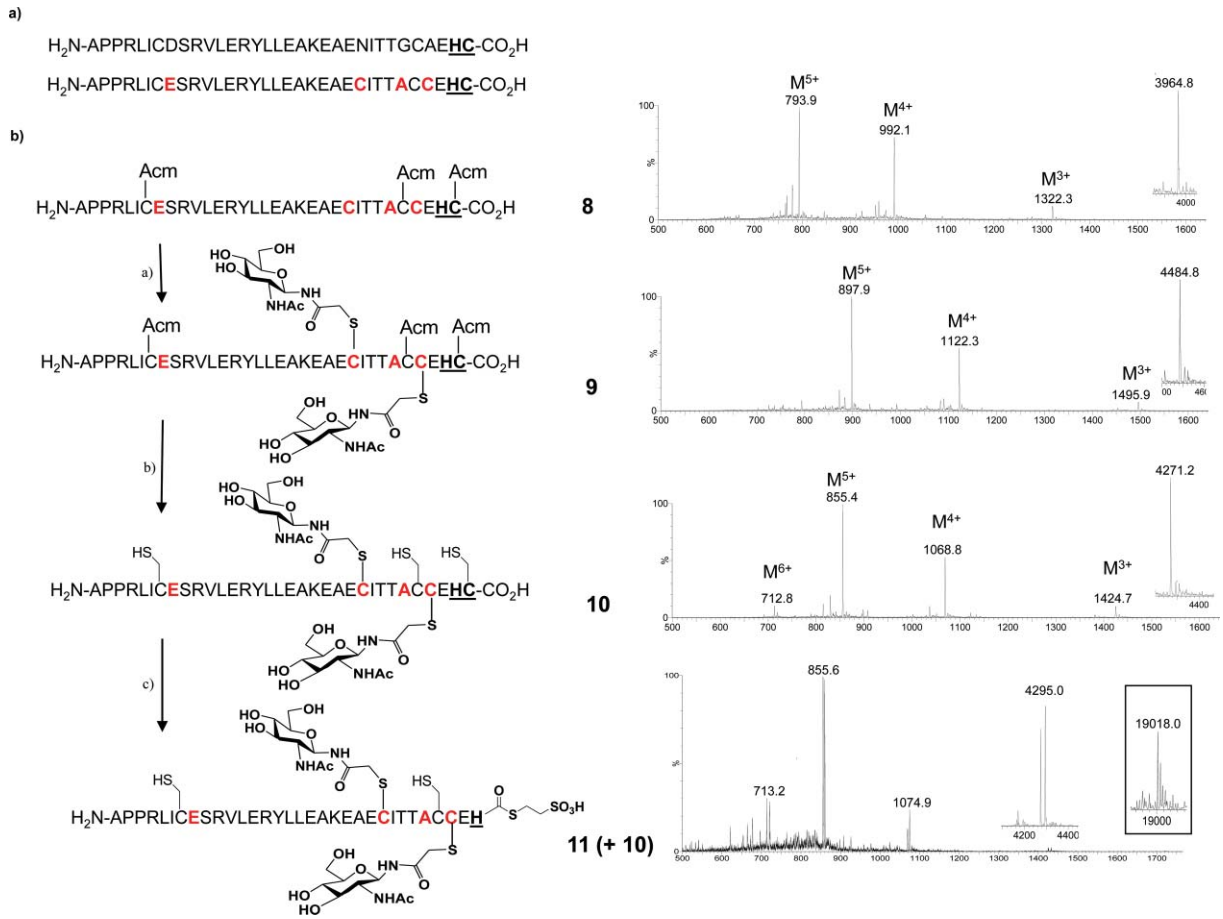


Fig. 2 a) Target EPO peptide sequence for synthesis compared with the wild-type sequence (mutations are highlighted). b) Reactions involved in the transformation from peptide **8** to glycopeptide mimetic thioester **11** with accompanying ESI-MS data for isolated products. *Reagents and conditions*. a) 50 mM NH_4HCO_3, glycosyl iodoacetamide (2 eq per thiol), room temp. 2 h b) 10% v/v aqueous AcOH, AgOAc (10 equiv. per SAcm group), DTT, 30 °C, 16 h. c) Guanidine.HCl, 10% v/v AcOH, 10% w/v MESNa, TCEP. HCl, 60 °C, 48 h.

2 h. *N*-Acetylglucosamine was employed in the first instance to ensure compatibility with the mimetic sugar-protein linkage and thioesterification. Although the reaction appeared quantitative by LC-MS **9** was isolated by HPLC in only 50% yield then subjected to AgOAc/DTT in 10% aqueous acetic acid to cleave acetamidomethyl protecting groups on the remaining cysteine residues. Again the deprotection appeared efficient by HPLC though the product was isolated after HPLC in 40% yield. Thioester precursor **10** was re-dissolved at a concentration of 1.25 mg mL^{-1} in 10% w/v MESNa, 10% v/v AcOH, 0.5% w/v TCEP in 6 M guanidine hydrochloride and heated to 60 °C for 48 h to produce the C-terminal thioester. The reaction afforded two major species in approximately 1 : 1 ratio: the product thioester **11** and unreacted HC-terminated peptide **10**. In agreement with previous results there was no evidence of protein cleavage between the internal Ile[6]-Cys[7] and Ala[28]-Cys[29] motifs. **10** and **11** could not be separated by semi-preparative reverse-phase HPLC though **10** would not participate in NCL and could likely be recovered after ligation.[15] To confirm thioester formation, the partially purified lyophilised product (5 mg) was added, in excess, to the bacterially-derived C-terminal fragment (EPO 33-166, 1 mg)[24] in 6 M guanidine hydrochloride containing 300 mM Na phosphate buffer; pH 7.0, 20 mM 4-mercaptophenylacetic acid (MPAA),[25] and 10 mM TCEP (0.25 mL) for 24 h as previously described.[12] Encouragingly, the desired product was immediately observed by LC-MS (Fig. 3b inset, calculated mass = 19016 Da, observed mass = 19018 Da) suggesting that neoglycopeptide thioester formation *via* the N→S acyl transfer method and subsequent glycoprotein assembly appear feasible. In fact, an early aim had been to compare the efficiency of this thioester synthesis with that of an identical peptide released from the safety-catch linker following solid-phase modification of selectively liberated thiols on solid-support. However, as a consequence of several failed attempts to efficiently unmask both thiols (at cys 24 and cys 30) within the resin-bound 33mer, this route was ultimately abandoned and the comparison no longer possible.

Glycomimetic synthesis

Having demonstrated the union of *N*-acetyl glucosamine with a peptide thioester precursor, we envisaged that, in the simplest case, synthesis of a tripeptide "scaffold" displaying saccharides (exemplified by **12**, Scheme 3) may serve as a mimic for branched oligosaccharides of the type typically observed extending from the pentasaccharide core structure of *N*-linked glycoproteins.[4a] Following synthesis, routine enzymatic sialylation of the

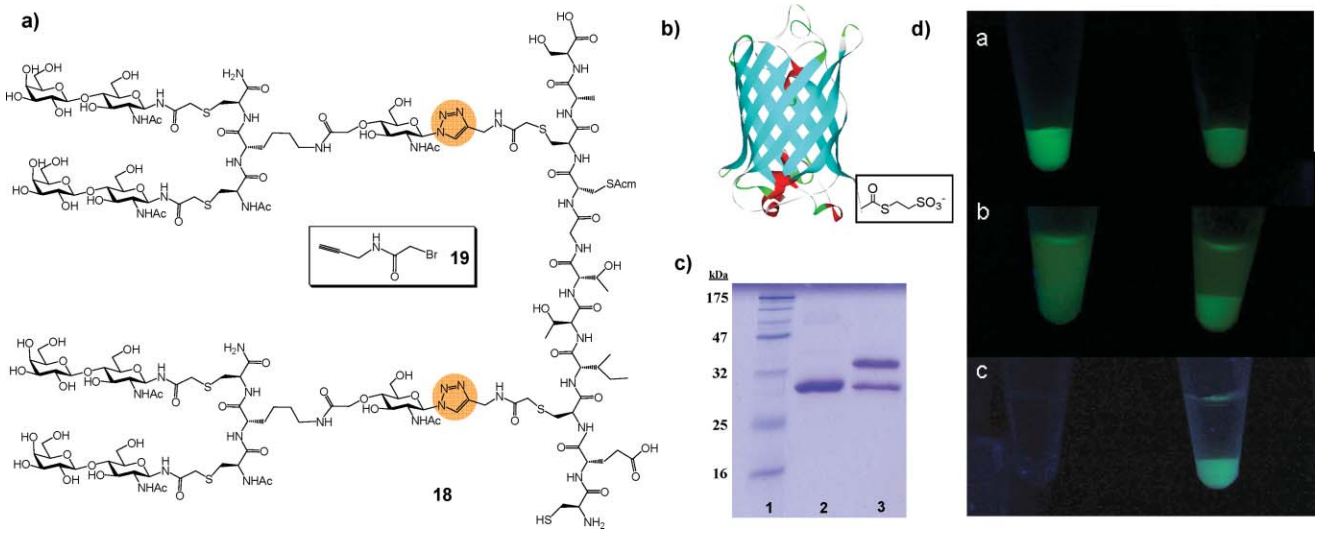
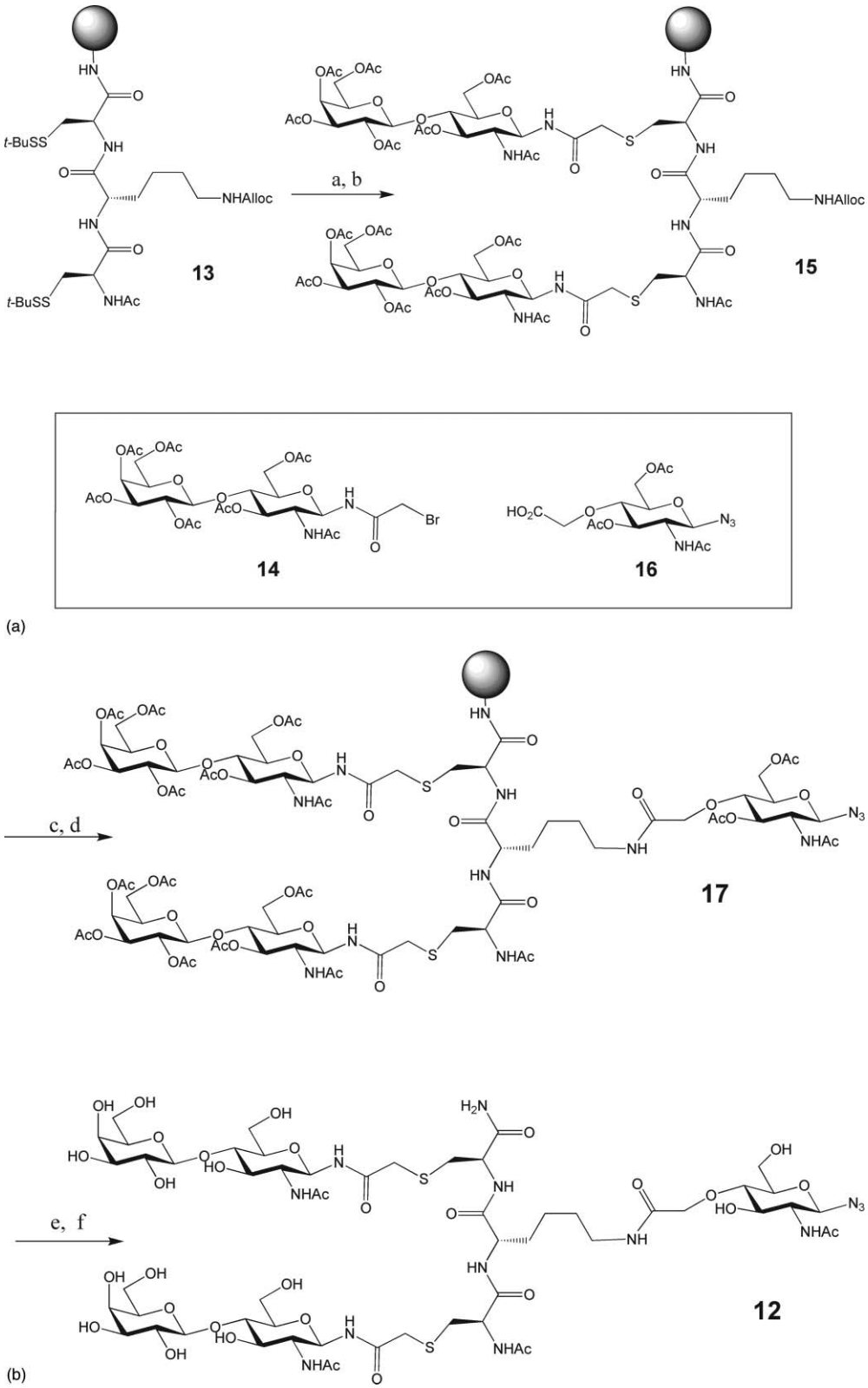


Fig. 3 a) An erythropoietin derived peptide bearing two oligosaccharide mimics. The positions of attachment of **12** to the acetylene-linked peptide are highlighted. b) Schematic representation of green fluorescent protein (GFP) C-terminal thioester (modified from PDB entry 1w7s). c) NCL reaction between GFP thioester and **18** after 48 h. Lane 1, molecular weight markers; lane 2, GFP (approx. mass = 28 kDa); lane 3, ligation reaction after 48 h. d) Lectin binding assay: a) A solution of GFP (left) and crude GFP linked-**18** (right); b) after addition of *Ricinus communis* RCA_{120} conjugated agarose beads; c) after washing with lectin binding buffer.

terminal galactose residues is all that would be required to complete the biantennary sialyl *N*-acetyllactosamine motif.[26] While simplifying synthesis considerably, notably by removal of the core β-mannoside, we retained the branched nature of the *N*-acetyllactosamine antennae in a format we hoped would still be sufficiently "biological" to permit clearance *via* the hepatic asialoglycoprotein receptor providing a potential advantage over polyethylene glycol (PEG)-linked proteins as they are not efficiently degraded and accumulate in kidney vacuoles.[27] **12** was assembled on Rink amide resin, initially employing commercially available amino acids and standard peptide coupling methodology using HBTU/HOBt as coupling reagents in the presence of *N*,*N*- diisopropylethyl amine (DIPEA). **13** was then exposed to dithiothreitol (DTT) to remove the *S*-*tert*-butylthio groups, followed by *N*-acetyllactosamine bromoacetamide (**14**) to afford **15**. The *N*-allyloxycarbonyl (alloc) group was removed with $Pd(PPh_3)_4$/1,3-dimethylbarbituric acid and the glycosyl azide **16** was coupled to the free ε-amino group of lysine using standard peptide coupling methods.[28] It was hoped that the azide functionality would next facilitate the direct site-specific modification of acetylenic peptides or proteins, circumventing the often troublesome Lansbury aspartylation of glycosyl amines.[29] Following cleavage from the solid support and acetyl ester deprotection, **12** was obtained in 47% overall yield (based on initial resin loading). To demonstrate that the "biantennary" saccharide mimic **12** could be recognised as an oligosaccharide we wished to evaluate it in a more biological context and so developed a lectin binding experiment employing the commercially available lectin *Ricinus communis* RCA_{120} immobilised on agarose beads. This lectin was chosen since it binds to the monosaccharide D-galactose and has been used to isolate glycoproteins from natural sources. To this end we appended two saccharide mimics (containing four terminal Gal residues) to a short peptide sequence derived from erythropoietin residues 22–32 adorned with an N-terminal cysteine residue (**18**, Fig. 3a). The oligopeptide sequence of **18** was assembled on Novasyn-TGT resin. At positions where oligosaccharide mimic **12** was to be installed *SS*-*t*Bu protected cysteine residues were introduced and, following reduction on solid-phase, the two free thiol groups were capped with 2-bromoacetyl propargylamide (**19**).[28] After cleavage from the solid-support we introduced two oligosaccharide mimics employing two equivalents of **12** per acetylene. Briefly **12** was incubated with the acetylene-linked peptide in phosphate buffered saline (0.1 M, pH 8.0) containing the Sharpless tris-benzyltriazolylmethylamine Cu(I) ligand,[11] $CuSO_4$, and sodium ascorbate for 24 h at room temperature, then **18** was purified from the reaction mixture using semi-preparative reverse-phase HPLC in 69% yield. **18** could also be prepared by first reacting **12** with **19** and uniting this thiol-selective oligosaccharide mimic with the same peptide sequence displaying free cysteine residues (as described in Fig. 2b employing *N*-acetyl glucosamine) in solution with very similar efficiency. **18** was then ligated to Green Fluorescent Protein (GFP) thioester (Fig. 3b) which was easily obtained by expression of the GFP gene from the commercially available DNA vector pTYB-1.[30] Use of GFP in this way provided a more glycoprotein-like context for the oligosaccharide mimic and allowed protein immobilisation to be easily observed. The native chemical ligation reaction was conducted in 200 mM sodium phosphate buffer; pH 8.0, containing 100 mM NaCl, 2% w/v 2-mercaptoethanesulfonic acid (MESNa), 5 mM tris-carboxyethyl phosphine (TCEP) and excess **18**. SDS-PAGE analysis of the ligation reaction after 48 h (Fig. 3b) showed that the reaction had proceeded to over 50% completion with the ligated product (molecular weight of approximately 32 kDa) clearly visible. The ligation reaction was then concentrated and ultimately obtained in 200 μl of Dulbeccos PBS solution; pH 7.4. This solution was tested for interaction with the galactose-binding lectin. When compared with GFP alone, which demonstrated no interaction with the immobilised



Scheme 3 Solid-phase synthesis of a branched oligosaccharide mimics. *Reagents/conditions*: a) 10% w/v DTT, DMF, 2.5% v/v DIPEA, 48 h. b) **14**, (2 equiv. per thiol), DMF, 2.5% v/v Et_3N, 16 h, c) 1,3-dimethylbarbituric acid, $Pd(PPh_3)_4$, DMF–$CHCl_3$, 2 h, quant. d) **16** (2.5 equiv.), PyBOP/HOBt, DIPEA, DMF, 16 h. e) 95% v/v TFA, 2.5% v/v ethanedithiol, 2.5% v/v H_2O, 3 h. f) 5% v/v hydrazine hydrate, 10 : 1 H_2O–MeOH, 24 h. Overall yield = 47%.

lectin, the **18**-linked GFP was clearly retained by the beads after extensive washing with Dulbeccos PBS solution suggesting an interaction between the lectin and the terminal galactose residues of **18** (Fig. 3d). Furthermore the fluorescence could be washed from the beads by incubation with a solution of D-galactose (100 mM) in Dulbeccos PBS solution, confirming the specificity of the interaction.

Conclusion

Significant progress has been made in a NCL-based route to neoglycoproteins using thioesters derived from an N→S acyl shift-triggered reaction. This allowed an EPO neoglycoform, decorated with simple sugar candidates for enzymatic remodelling, to be assembled. The next step is to replace simple sugars with oligosaccharide mimetics such as **12** which was shown to be recognised and bound by a lectin. This simplified saccharide-display scaffold was prepared in good yield with minimal purification and can be fused with protein fragments at pre-defined sites using reliable thiol targeting or azide/acetylene chemistry. Following native chemical ligation **12** might ultimately form part of useful protein therapeutics.

Experimental details

General experimental details

^{1}H NMR spectra were recorded at 500, 350 and 300 MHz, ^{13}C NMR spectra were recorded at 125, 63 and 75 MHz respectively on a Bruker instrument. Electrospray mass spectrometry was carried out on a Waters Acquity UPLC-SQD MS system with an applied voltage of 60 V. Analytical HPLC was performed on a Dionex Ulimate 3000 HPLC system equipped with a C_{18} reverse phase column (Phenomenex Spherclone ODS (4.6 mm × 250 mm)) employing a gradient of 5–95% acetonitrile containing 0.1% TFA over 45 min (flow rate of 1.0 mL min^{-1}). Semi-preparative HPLC was performed using a Phenomenex Jupiter proteo column (10 mm × 250 mm) and a gradient of 5–60% acetonitrile containing 0.1% TFA over 45 min (flow rate of 4.0 mL min^{-1}). All reagents and solvents (excluding HPLC solvents) were standard laboratory grade and used as supplied unless otherwise stated. Where a solvent was described as dry it was purchased as anhydrous grade. All resins and Fmoc amino acids for peptide synthesis were purchased from Merck Biosciences with the exception of FmocHis(Me)-OH which was obtained from Bachem.

Peptide thioester synthesis

Synthesis and thioesterification of model peptides 1(a/b), 2(a/b), 4(a, b), 6(a, b) and 7(a, b). Synthesis of these compounds was performed as previously described.[15] 10 μL aliquots of each 1.0 mL reaction were taken at 24, 48, and 96 h intervals and analysed by HPLC. Conversion to product was determined through comparison of peak area at 230 nm. The identity of the starting materials and products were confirmed at $t = 0$ and $t =$ 96 h by LC-MS.

Synthesis of EPO residues 1–33 (8). The peptide containing a C-terminal HC-motif was prepared on 0.05 mmol scale using standard Fmoc SPPS procedures. Briefly, commercially available Nova-Syn TGA resin (Merck Biosciences) was loaded with Fmoc-Cys(Acm)-OH by double coupling with HBTU/HOBt in the presence of DIPEA and then extended using HBTU/HOBt (Fastmoc protocol) in automated fashion using an Applied Biosystems 433A peptide synthesiser. The coupling time was 0.5 h. Target sequence: APPR(Pmc)L***I***C(Acm)**E**(O*t*Bu)S(*t*Bu)R(Pmc)V̲L̲E̲(Q*t*B̲u̲)R(Pmc)Y(*t*Bu)L̲L̲E̲(Q*t*B̲u̲)AK̲(Boc)E̲-(Q*t*B̲u̲)AE̲(Q*t*B̲u̲)**C(Trt)*I***T(*t*Bu)T(*t*Bu)**A**C(Acm)**C(Trt)**E(O*t*Bu)-H(Trt)C(Acm)-resin. Underlined residues were double coupled, and residues shown in bold are substitutions from the wild-type sequence. The dry resin was then transferred to a glass vial and exposed to a solution comprised of 95% TFA, 2.5% EDT, 2.5% H_2O (6.0 mL), with stirring for 5 h. The crude product was then collected by centrifugation at 3000 rpm following precipitation from cold diethyl ether (50.0 mL). The fully deprotected and precipitated product was redissolved in 25% aqueous MeCN and purified by semi-preparative HPLC. The major peak (retention time = 32.7 min, 35 mg, 18%) was analysed by ESI-MS and was found to correspond to the desired product. ESI-MS (ES+) observed mass = 3964.8 Da calculated mass = 3964.6 Da.

Alkylation with sugar iodoacetamides (9). The purified peptide (30.0 mg 7.57 μmol) was dissolved in 50 mM NH_4HCO_3; pH 8.0 (3.0 mL) and *N*-acetylglucosamine iodoacetamide (15 mg, 39 μmol) was added. The reaction mixture was incubated at room temperature with shaking for 2 h, neutralised with AcOH then loaded directly onto a semi-preparative HPLC column to afford the product (retention time = 30.6 min, 17 mg, 3.78 μmol) as a white fluffy solid following lyophilisation. ESI-MS (ES+) observed mass = 4484.8 Da, calculated mass = 4485.1 Da.

Acm deprotection (10). The purified glycopeptide (11 mg, 2.77 μmol) was redissolved in 10% v/v aqueous acetic acid (2.0 ml) and AgOAc (13 mg, 83.2 μmol) was added. After shaking at room temperature for 30 min DTT was added to a final concentration of 10% w/v and shaking was continued at room temperature overnight. The thick yellow precipitate was removed by centrifugation and the supernatant was loaded directly onto a semi-preparative HPLC column to afford the product (5 mg, 1.1 μmol) as a white fluffy solid following lyophilisation. ESI-MS (ES+) observed mass = 4271.8 Da calculated mass = 4271.8 Da.

Thioester formation (11). Peptide **10** (5 mg, 1.1 μmol) was dissolved in 6 M guanidine.HCl (3.2 mL) and acetic acid 400 μL followed by MESNa (400 mg) and TCEP.HCl (5 mg). The reaction mixture was then incubated at 60 °C with shaking at 700 rpm in an Eppendorf thermomixer for 48 h. After 48 h the conversion was estimated as approximately 50% by LC-MS and the reaction mixture was loaded directly onto a semi-preparative HPLC column and thioester **11** was isolated along with unreacted **10** (5 mg) as a white fluffy solid following lyophilisation.

Native chemical ligation. A recombinant EPO fragment comprising residues 33–166 (approx 1.0 mg)[17,24] was dissolved in 6 M guanidine hydrochloride containing 300 mM Na phosphate buffer, pH 7.0 (0.25 mL), containing 20 mM TCEP and 50 mM MPAA and added to thioester **11** (5 mg and still containing about 50% **10**). The reaction was shaken at room temperature for 24 h. A sample (3.0 μL) was taken for LC-MS analysis after which time the reaction appeared complete. Further TCEP (10 mM) was added and the reaction shaken at room temperature for a further 1 h.

The neoglycoprotein was then partially purified by dilution with water (10 volumes) followed by incubation on ice for 0.5 h. The precipitated protein was collected by centrifugation at maximum speed on a microcentrifuge for 5 min. (the synthetic fragment is highly water soluble whereas the protein is not). The protein pellet was resolubilised in 6 M guanidine hydrochloride, treated with dithiothreitol to a final concentration of 10 mM then incubated at 37 °C for 0.5 h. A single repetition of the precipitation and resolubilisation process (as above) afforded a protein sample that was largely free of all synthetic components (**10**, or **11**) though the process can be repeated further (see supporting information for more details). Since the aim of the experiment was merely to verify thioester formation through NCL the product was not purified further nor refolded.

Synthesis of oligosaccharide mimic 12

Solid-phase peptide synthesis. The synthesis of compound **12** was conducted on 0.10 mmol scale on Rink Amide MBHA resin. The Fmoc group was removed by treatment with 20% (v/v) piperidine in DMF (2 × 5 min), and similarly between each coupling step. Coupling reactions with Fmoc amino acids were conducted using 0.50 mmol (5 equiv.) of each amino acid, HBTU/HOBt and DIPEA for 3 h. The reaction progress was monitored using the Kaiser ninhydrin test. After each coupling and deprotection the resin was washed with DCM and DMF (5 min each).

Cys-StBu deprotection. DTT (200 mg) was dissolved in dry DMF (1.8 mL) and 2.5% v/v DIPEA was added. After stirring for 5 min the solution was transferred to a peptide synthesis vessel containing resin-bound SS^{t}Bu protected peptide (**13**). After 16 h the resin was filtered off and washed exhaustively with DMF, then DMF–H_2O (1 : 1), DMF and finally DCM. This procedure was then repeated.

***N*-Acetyl lactosaminyl bromoacetamide couplings.** Bromoacetamide **14** (2 eq. *per* thiol: 0.10 mmol × 2 thiols × 2 eq. = 302 mg, 0.40 mmol) was dissolved in DMF (1.0 mL) and NEt_3 (84 μL, 0.6 mmol) and transferred to a peptide synthesis vessel containing resin-bound SS^{t}Bu deprotected peptide. The reaction was allowed to proceed for 24 h. After this time, the resin was filtered and washed exhaustively with DMF and then DCM.

Alloc deprotection. $Pd(PPh_3)_4$ (58 mg, 0.05 mmol) was dissolved in a solution of 1 : 1 DMF–$CHCl_3$ (1.5 mL) and 1,3-dimethylbarbituric acid (156 mg, 1 mmol, 10 eq) was added. The yellow solution was transferred to the reaction vessel containing the resin and shaken at room temperature for 2 h. The resin (now red in colour) was then washed sequentially with the following solutions: 0.5% v/v DIPEA in DMF (4 × 2.0 mL); DMF (6 × 2.00 mL); 0.5% w/v sodium diethyldithiocarbamate trihydrate in DMF (4 × 2.0 mL); followed by a final wash with DMF (6 × 2.0 mL). The washes returned the resin to its original colour.

Coupling of 2-acetamido-2-deoxy-3,6-di-*O*-acetyl-4-carboxymethyl-β-D-glucopyranosyl azide. The resin was washed with DMF (2.0 mL). Coupling reactions with 2-acetamido-2-deoxy-3,6-di-*O*-acetyl-4-carboxymethyl-β-D-glucopyranosyl azide, **16** (97 mg, 0.25 mmol, 2.5 equiv.), 2.5 equiv. PyBOP/HOBt and 5 equiv. DIPEA for 16 h. The process was repeated. Afterwards the resin was exhaustively washed with DCM and DMF.

Cleavage from the solid support. **17** was cleaved from the resin by treatment with the cleavage cocktail (95% TFA, 2.5% v/v water and 2.5% v/v EDT) for 3 h. After 3 h the resin was filtered off and treated once again with fresh cleavage cocktail for another 3 h. The combined TFA filtrates containing peracetylated **12** were precipitated with cold diethyl ether and collected by centrifugation at 3000 rpm for 10 min. The precipitate was dried by flushing with nitrogen for 30 min. The precipitate was resuspended in a water–methanol solution (10 : 1) containing 5% v/v hydrazine hydrate (6.0 ml) and stirred at room temperature for 24 h. **12** was then was purified by reverse phase semi preparative HPLC (water–acetonitrile 5–95% over 40 min, 0.1% TFA, t_R = 14.2 min) and obtained as a fluffy white solid (36 mg, 47% yield based on initial resin loading) after lyophilisation. Analytical data of compound **12**: $C_{56}H_{93}N_{13}O_{32}S_2$. MW: 1524.79. Found: (M + 1) 1525.8; (M + 2) 763.5.

Synthesis of compound 18

The acetylene-modified peptide (3.60 mg, 2.67×10^{-3} mmol) and **12** (14.30 mg, 9.38×10^{-3} mmol) were dissolved in 0.1 M PBS buffer solution; pH 8 (1.90 mL). Sodium ascorbate (13.00 mg, 6.66×10^{-2} mmol) and tris-benzyltriazolylmethylamine (TBTA) (35.0 mg, 6.66×10^{-2} mmol) were added to the solution. The final suspension was sonicated for 5 min in a cold water bath. Then, from a stock solution in PBS, $CuSO_4 \cdot 5H_2O$ (0.67 mg, 2.67×10^{-3} mmol, 26.7 mM, 100 μL) was added. The reaction was finally shaken for 24 h at 20 °C and 1200 rpm in an Eppendorf thermomixer. The reaction was next centrifuged at 14000 rpm for 5 min, and filtered through celite plug in a Pasteur pipette. The filtrate was purified by reverse phase semi-preparative HPLC (water–acetonitrile 5%–95% over 40 min, 0.1% TFA). Fractions containing the product were lyophilised to afford a fluffy white solid **18** (t_R = 16.1 min, 8.10 mg, 69%), and a white solid **12** (t_R = 14.2 min) of recovered starting material (5 mg). $C_{164}H_{268}N_{40}O_{84}S_8$. MW: 4400.6. Found: (M + 3) 1468.6, (M + 4 (+K)) 1111.5, (M + 4) 1101.7, (M + 5) 881.7. Deconvoluted mass = 4402.8 Da.

Ligation between green fluorescent protein (GFP)- and EPO (residues 22–32) containing two oligosaccharide mimetics

The GFP gene was PCR amplified from the commercially available cloning vector pcDNA3.1/NT-GFP-TOPO and subsequently cloned into expression vector pTYB-1 using the NdeI/EcoRI (after silencing of the internal NdeI restriction site) restriction sites. Positive colonies were fully sequenced across the coding region. The resulting plasmid was expressed as described as below.

Translated sequence

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGD-ATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFSRYP-DHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAE-VKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSH-NVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQN-TPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLE-FVTAAGITHGMDELYK**EFLEGSS** (bold residues are vector

derived amino acids) Calculated molecular weight = 27616.9 Da.

Expression of GFP-Sce intein–chitin binding domain (CBD) fusion protein. *Escherichia coli* strain BL21(DE3) was transformed with pTYB1-GFP. A single transformed colony was inoculated into 10.0 ml of LB medium supplemented with ampicillin to a final concentration of 100 μg ml^{-1} and grown to saturation at 37 °C with shaking. This culture was used to inoculate 2 × 500 ml of fresh LB ampicillin medium. Cultures were grown at 37 °C with shaking until A_{600} reached 0.6. Expression of the GFP-Sce intein-CBD fusion was induced by addition of isopropyl-thio-β-D-galactopyranoside (IPTG) to each culture at a final concentration of 1 mM. Induction was allowed to proceed for 5 h at 30 °C with shaking. Bacteria were harvested by centrifugation at 8000 rpm in a JA 10.500 rotor (Beckman) for 15 min at 4 °C and stored at –20 °C until required.

Purification of GFP-Sce intein-CBD fusion protein. Pelleted cells were resuspended in 10.0 ml of column buffer (20 mM Tris-HCl; pH 8.0, 500 mM NaCl, 1 mM EDTA, 1 mM PMSF) and lysed by sonication on ice (20 cycles of 30 s sonication separated by 30 s cooling on ice). The crude lysate was cleared by centrifugation at 12000 rpm in a JA 25.50 rotor (Beckman) for 15 min at 4 °C. The supernatant was loaded onto a column containing 1.5 ml of chitin beads (NEB) pre-equilibrated with 20 ml of column buffer and allowed to drain by gravity flow. Unbound proteins were washed from the beads with 50 ml of column buffer prior to native chemical ligation. For control experiments, on-column cleavage of GFP was performed by addition of 10 ml of cleavage buffer (column buffer containing 50 mM DTT). Cleavage buffer was allowed to drain after which the column was immediately sealed and incubated at 4 °C for 18 h. Cleaved GFP was eluted in 15 ml of cleavage buffer containing 1 mM PMSF.

Native chemical ligation of GFP to unnatural glycopeptide. For native chemical ligation with **18**, 5 ml of ligation buffer (200 mM sodium phosphate pH 8.0, 100 mM NaCl, 5 mM TCEP, 2% (w/v) MESNa) was added to the chitin bead-bound GFP and allowed to drain by gravity flow after which the column was immediately sealed. **18** (4.2 mg) was resuspended in 100 μl of ligation buffer and added to the chitin beads. This reaction was mixed at intervals at room temperature for 48 h. Proteins were eluted from the column in 2 ml of ligation buffer (without MESNa). Ligation buffer was removed from eluted proteins and replaced with Dulbecco's phosphate buffered saline (PBS) pH 7.4 by centrifugation through a 30 kDa molecular weight cut-off filter (VivaScience) during which procedure any residual (unligated) glycopeptide was removed from the ligation product. Following buffer exchange, the GFP-glycopeptide conjugate was resuspended in 200 μl of Dulbecco's PBS.

Lectin interaction studies. To demonstrate carbohydrate–lectin interaction, 100 μl of cross-linked 4% beaded agarose displaying lectin from *Ricinus communis* (Sigma) was pre-equilibrated in Dulbecco's PBS and added to 50 μl of GFP–glycopeptide conjugate or DTT-hydrolysed GFP alone as negative control. Solutions were gently mixed for 1 h at room temperature. Beads were washed 5 times with 1 ml of Dulbecco's PBS to remove unbound conjugate. Residual buffer was carefully removed after the final wash and beads were further incubated with an additional 50 μl of conjugate or cleaved GFP at 4 °C for 48 h. Beads were washed again as described and bound GFP-glycopeptide visualised by exposure to U.V. radiation, using a 4 W UV lamp, model UVGL-25 Mineralight® Lamp at 254 nm in a darkened cabinet. To demonstrate the carbohydrate-specific nature of the observed lectin interaction, bound glycoprotein–GFP conjugate was competed off the lectin beads by mixing with a 100 mM solution of D-(+)-galactose in Dulbecco's PBS, after which beads were washed and visualised as described above.

SDS PAGE analysis. For GFP ligation studies, samples were mixed with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) loading buffer, boiled for 5 min and resolved by SDS PAGE on pre-cast 12% polyacrylamide gels using the Protean III system (BioRad). Preparations of erythropoietin were precipitated by addition of 20 sample volumes of methanol and acetone solution (1 : 1 v/v), incubated at –20 °C for 30 min and collected by centrifugation. Samples were resolubilised in 20 μl of 8 M urea and boiled with loading buffer as described. Proteins were resolved by SDS PAGE on pre-cast 18% polyacrylamide gels as described. Electrophoresed proteins were visualised by Coomassie staining.

Acknowledgements

The authors acknowledge financial support from BBSRC, The Wellcome Trust and The Royal Society.

Notes and references

1 (*a*) P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. Kent, *Science*, 1994, 266; (*b*) D. Macmillan, *Angew. Chem., Int. Ed.*, 2006, **45**, 7668; (*c*) D. Rauh and H. Waldmann, *Angew. Chem., Int. Ed.*, 2007, **46**, 826; (*d*) J.-P. Pellois and T. W. Muir, *Curr. Opin. Chem. Biol.*, 2006, **10**, 487; (*e*) V. Muralidharan and T. W. Muir, *Nat. Methods*, 2006, **3**, 429; (*f*) C. Chatterjee, R. K. McGinty, J.-P. Pellois and T. W. Muir, *Angew. Chem., Int. Ed.*, 2007, **46**, 2814; (*g*) S. B. H. Kent, *Chem. Soc. Rev.*, 2009, **38**, 338–351; (*h*) R. J. Payne and C.-H. Wong, *Chem. Commun.*, 2010, **46**, 21–43.

2 (*a*) D. Macmillan and C. R. Bertozzi, *Tetrahedron*, 2000, **56**, 9515; (*b*) T. J. W. Tolbert and C.-H. Wong, *J. Am. Chem. Soc.*, 2000, **122**, 5421; (*c*) D. Macmillan and C. R. Bertozzi, *Angew. Chem., Int. Ed.*, 2004, **43**, 1355; (*d*) B. Wu, J. Chen, J. D. Warren, G. Chen, Z. Hua and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2006, **45**, 4116; (*e*) Y.-Y. Yang, S. Ficht, A. Brik and C.-H. Wong, *J. Am. Chem. Soc.*, 2007, **129**, 7690; (*f*) N. Yamamoto, Y. Tanabe, R. Okamoto, P. E. Dawson and Y. Kajihara, *J. Am. Chem. Soc.*, 2008, **130**, 501; (*g*) C. Piontek, D. V. Silva, C. Heinlein, C. Pöhner, S. Mezzato, P. Ring, A. Martin, F. X. Schmid and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2009, **48**, 1941; (*h*) C. Piontek, P. Ring, O. Harjes, C. Heinlein, S. Mezzato, N. Lombana, C. Pöhner, M. Püttner, D. V. Silva, A. Martin, F. X. Schmid and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2009, **48**, 1936; (*i*) P. Nagorny, B. Fasching, X. Li, G. Chen, B. Aussedat and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2009, **131**, 5792.

3 (*a*) P. H. Seeberger, *Chem. Soc. Rev.*, 2008, **37**, 19; (*b*) X. Zhu and R. R. Schmidt, *Angew. Chem., Int. Ed.*, 2009, **48**, 1900.

4 (*a*) P. M. St Hilaire and M. Meldal, *Angew. Chem., Int. Ed.*, 2000, **39**, 1162; (*b*) L. L. Kiessling, J. E. Gestwicki and L. E. Strong, *Angew. Chem., Int. Ed.*, 2006, **45**, 2348; (*c*) S.-Y. Chen, S. Cressman, F. Mao, H. Shao, D. W. Low, H. S. Beilan, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, L. Savatski, J. W. Adamson, C. E. Bozzini, A. Kung, S. B. H. Kent, J. A. Bradburne and G. G. Kochendoerfer, *Chem. Biol.*, 2005, **12**, 371; (*d*) S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony and B. G. Davis, *Nature*, 2007, **446**, 1105; (*e*) S. I. v. Kasteren, S. J. Campbell, S. Serres, D. C. Anthony, N. R. Sibsonc and B. G. Davis, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 18.

5 (*a*) M. R. Pratt and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2003, **125**, 6149; (*b*) I. S. Carrico, B. L. Carlson and C. R. Bertozzi, *Nat. Chem. Biol.*, 2007, **3**, 321; (*c*) E. M. Sletten and C. R. Bertozzi, *Angew. Chem., Int. Ed.*, 2009, **48**, 6974.
6 (*a*) R. M. Bill and S. L. Flitsch, *Chem. Biol.*, 1996, **3**, 145; (*b*) T. Buskas, S. Ingale and G.-J. Boons, *Glycobiology*, 2006, **16**, 113R; (*c*) H. Hojo and Y. Nakahara, *Peptide Sci*, 2007, **88**, 308; (*d*) D. P. Gamblin, E. M. Scanlan and B. G. Davis, *Chem. Rev.*, 2009, **109**, 131.
7 (*a*) B. Li, H. Song, S. Hauser and L. X. Wang, *Org. Lett.*, 2006, **8**, 3081; (*b*) W. Huang, H. Ochiai, X. Zhang and L.-X. Wang, *Carbohydr. Res.*, 2008, **343**, 2903; (*c*) H. Ochiai, W. Huang and L.-X. Wang, *J. Am. Chem. Soc.*, 2008, **130**, 13790; (*d*) T. B. Parsons, J. W. B. Moir and A. J. Fairbanks, *Org. Biomol. Chem.*, 2009, **7**, 3128.
8 (*a*) B. L. Pentelute, Z. P. Gates, V. Tereshko, J. L. Dashnau, J. M. Vanderkooi, A. A. Kossiakoff and S. B. H. Kent, *J. Am. Chem. Soc.*, 2008, **130**, 9695; (*b*) K. Mandal, B. L. Pentelute, V. Tereshko, A. A. Kossiakoff and S. B. H. Kent, *J. Am. Chem. Soc.*, 2009, **131**, 1362; (*c*) G. G. Kochendoerfer, S.-Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent and J. A. Bradburne, *Science*, 2003, **299**, 884.
9 J. Xiao and T. J. Tolbert, *Org. Lett.*, 2009, **11**, 4144.
10 (*a*) N. J. Davis and S. L. Flitsch, *Tetrahedron Lett.*, 1991, **32**, 6793; (*b*) D. Macmillan, A. M. Daines, M. Bayrhuber and S. L. Flitsch, *Org. Lett.*, 2002, **4**, 1467.
11 T. R. Chan, R. Hilgraf, K. B. Sharpless and V. V. Fokin, *Org. Lett.*, 2004, **6**, 2853.
12 J. P. Richardson and D. Macmillan, *Org. Biomol. Chem.*, 2008, **6**, 3977.
13 Y. Shin, K. A. Winans, B. J. Backes, S. B. H. Kent, J. A. Ellman and C. R. Bertozzi, *J. Am. Chem. Soc.*, 1999, **121**, 11684.
14 J. Kang, J. P. Richardson and D. Macmillan, *Chem. Commun.*, 2009, 407.
15 J. Kang and D. Macmillan, *Org. Biomol. Chem.*, 2009, **7**, 4918.
16 (*a*) T. Murase, T. Tsuji and Y. Kajihara, *Carbohydr. Res.*, 2009, **344**, 762; (*b*) K. Hirano, D. Macmillan, K. Tezuka, T. Tsuji and Y. Kajihara, *Angew. Chem., Int. Ed.*, 2009, **48**, 9557–9560.
17 See supporting information for further details.
18 Note that **7** (Fig. 1) is not a thioester though conversion is plotted on the same graph for comparison.
19 (*a*) Z. Du, P. T. Shemella, Y. Liu, S. A. McCallum, B. Pereira, S. K. Nayak, G. Belfort, M. Belfort and C. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 11581; (*b*) C. Ludwig, D. Schwarzer and H. D. Mootz, *J. Biol. Chem.*, 2008, **283**, 25264.
20 N. Suree, C. K. Liew, V. A. Villareal, W. Thieu, E. A. Fadeev, J. J. Clemens, M. E. Jung and R. T. Clubb, *J. Biol. Chem.*, 2009, **284**, 24465.
21 Coupling of cysteine to hydroxy-functionalised resins employing HBTU/HOBt may well lead to racemisation of the C-terminal cysteine though this should be of little consequence since the cysteine is excised. Cysteine racemisation can also be minimised by using a trityl resin, *e.g.* NovaSYN TGT.
22 S. Elliott, D. Chang, E. Delorme, T. Eris and T. Lorenzini, *J. Biol. Chem.*, 2004, **279**, 16854.
23 The D8E mutation is non-essential as hydrolysis at Asp is slow relative to thioesterification but the conservative mutation is unlikely to affect bioactivity as this region does not bind the EPO receptor.
24 D. Macmillan and L. Arham, *J. Am. Chem. Soc.*, 2004, **126**, 9530.
25 E. C. B. Johnson and S. B. H. Kent, *J. Am. Chem. Soc.*, 2006, **128**, 6640.
26 C. Unverzagt and J. Seifert, *Tetrahedron Lett.*, 2000, **41**, 4549.
27 A. Bendele, J. Seely, C. Richey, G. Sennello and G. Shopp, *Toxicol. Sci.*, 1998, **42**, 152.
28 D. Macmillan, J. Blanc, *J. PCT*, *Int. Appl.*, 2008, WO2008001109.
29 (*a*) S. T. Anisfeld and P. T. Lansbury, *J. Org. Chem.*, 1990, **55**, 5560; (*b*) J. S. Miller, V. Y. Dudkin, G. J. Lyon, T. W. Muir and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2003, **42**, 431.
30 www.neb.com.

