CARBOHYDRATE FEATURES OF RECOMBINANT HUMAN GLYCOPROTEINS

Johannis P Kamerling

Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University,
P.O. Box 80.075, NL-3508 TB Utrecht, The Netherlands.
Telephone: +31 30 2533479 Fax: +31 30 2540980. E-mail: kame@boc.chem.ruu.nl

ABSTRACT

In the present approaches of developing pharmaceutically relevant drugs based on recombinant human glycoproteins, the establishment of the glycosylation patterns of these biomolecules is highly actual. In this minireview a short overview of the state of the art with respect to glycoprotein glycan analysis is given. Furthermore, the glycosylation of plasminogen activators and erythropoietins is summarized in more detail.

Key words: recombinant glycoproteins, carbohydrate analysis, plasminogen activator, erythropoietin

Biotecnologia Aplicada 1996;13:167-180

RESUMEN

El establecimiento de los patrones de glicosidación de fármacos basados en glicoproteínas humanas recombinantes constituye un elemento de gran actualidad en las tendencias modernas de desarrollo de esas biomoléculas. El presente trabajo de revisión pasa revista al estado del arte del análisis de glicanos de las glicoproteínas. Además, sumariza en mayor detalle la glicosidación de los activadores del plasminógeno y las eritropoietinas.

Palabras claves: glicoproteínas recombinantes, análisis de carbohidratos, activador de plasminógeno, eritropoietina

Introduction

In recent years the academic and industrial interest in the Carbohydrate part of glycoproteins has grown dramatically. This is mainly due to accumulating evidence concerning the relevance of the glycan chains in the physical and biological behaviour of these biomolecules. N-linked glycoprotein glycans award essential physical properties like proper folding and stabilization of the conformation of glycoproteins, protease resistance, and charge and water binding capacity. Furthermore, they are relevant in biological recognition processes like protein targetting and cell-cell interactions. It has been well established that enzymes of N-linked carbohydrate chains of glycoproteins, especially the range of non-reducing terminal structural elements, are species specific, tissue specific, cell-type specific, and protein specific. The terminal sequences are often differentially expressed during early development, provide masking functions to prevent rapid clearance from circulation, activate host immunological systems, act as immunodeterminants, and influence bioactivity. The importance of O-linked carbohydrate chains has been mainly discussed in terms of charge, water binding properties, and stability. For comprehensive reviews, see Rademacher, et al. (1), Allen and Kisailus (2), Lis and Sharon (3) and Mon treuil, et al. (4, 5).

It is without doubt that the present knowledge on the importance of glycoprotein glycosylation has strongly influenced the discussion with respect to genetically engineered proteins prepared in heterologous cell types, with or without a glycosylation machinery (6-8). In the context of therapeutic administration of recombinant human glycoproteins, an increasing interest of pharmaceutical industries is observed to consider the glycosylation patterns of the engineered proteins, with regard to applicability, tolerance and patent position.

The development of advanced analytical methods to unravel the structures of the ensembles of carbohydrate chains, present at specific amino acid residues in polypeptide backbones, has been the first step in the glycoprotein revival, and formed the basis for what today is called "glyobiology". Although in recent years several approaches based on $^1$H-NMR spectroscopy (9, 10) and mass spectrometry (11-16) have become available, and much research is focused on the development of profiling techniques based on high performance separation procedures (17-19), the primary structure analysis of glycoprotein glycans has still not reached the level of a routine analysis. Even now, completely novel oligosaccharide elements, as part of glycoprotein glycans, are identified. Therefore, it should be stressed that, due to the complexity of the carbohydrate chains, not only in terms of monosaccharide constituents and sequences, but also in terms of non-carbohydrate substituents, the current approaches have to be applied with great care.

In general, N-glycans share a common Asn-linked pentasaccharide core element Manα1-6(Mana1-3) Manβ1-4GlcNAcb1-4GlcNAcb1- (Man, D-mannose), which can be extended in several ways. Based on typical elongations of this core element, three major types of N-glycans can be distinguished, namely, (i) the oligomannose type; (ii) the complex type, comprising the N-acetyllactosamine and/or the N,N'-diacetyllactosaminide type, whereas also β-D-Xylose (Xyl) attached at O2 of βMan can be present, and (iii) the hybrid type, showing characteristics of both the oligomannose and the complex type. As an illustration, in Scheme 1 two composite N-linked structures are presented, covering the major findings for (i) and (ii).

For the mentioned O-glycans, so far nine core elements can be defined, which are depicted in Scheme 2. These cores can be extended in several ways, comprising e.g. Galβ1-4GlcNAcb or Galβ1-3GlcNAcb (Gal, D-galactose) elements, and terminated by monosaccharides like L-fucose (Fuc), Gal, GalNAc or N-acetylneuraminic acid (Neu5Ac) / N-glycolylneuraminic acid (Neu5Gc) (10). Moreover, O-acetyl and sulfate groups can occur, as demonstrated for the N-glycans.

Scheme 1. Composite N-glycan structures linked to Asn in the polypeptide via GlcNAc. Elements/units attached via ---- indicate possible chain extensions. Glc, α-glucose; Gal, α-galactose; Fuc, L-fucose; Xyl, α-xylose; Neu5Ac, N-acetylneuraminic acid; Neu5Gc, N-glycolylneuraminic acid; AcO, O-acetyl; P, phosphate; S, sulfate.

GlcNAcα1-P-6---
\[\text{Manα1-2---Manα1-6} \]
\[\text{Manα1-2---Manα1-3} \]
\[\text{Manβ1-4GlcNAcβ1-4GlcNAc} \]
Glcα1-2---Glcα1-3---Glcα1-3---Manα1-2---Manα1-2---Manα1-3
Fucα1-3---
\[\text{Neu5Aca2-3---Galβ1-4GlcNAcβ1-6} \]
\[\text{GlcNAcβ1-4---} \]
\text{Galβ1-4GlcNAcβ1-3---Galβ1-4GlcNAcβ1-2Manα1-6} \]
\text{Galα1-3---} \]
\[\text{Fucα1-2---} \]
\[\text{S-3---} \]
\[\text{Neu5Aca2-3/6---} \]
\[\text{S-4---GalNAcβ1-4---} \]
\text{Neu5Ac/5Gca2-3/6---Galβ1-4---GlcNAcβ1-2Manα1-3} \]
\text{AcO-4/7/8/9...} \]
\[\text{Galβ1-3GlcNAcβ1-4---Xylβ1-2---} \]
\[\text{Neu5Aca2-6---} \]

Scheme 2. Possible O-glycan core structures linked to Ser or Thr in the polypeptide via GalNAc.

<table>
<thead>
<tr>
<th>GalNAc</th>
<th>GalNAc</th>
<th>GlcNAcβ1-3</th>
<th>GalNAc</th>
<th>GalNAc</th>
<th>GlcNAcβ1-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Galβ1-3</td>
<td>GlcNAcβ1-3</td>
<td>GalNAcα1-3</td>
<td>GalNAc</td>
<td>Galβ1-3</td>
<td>GlcNAcβ1-6</td>
</tr>
<tr>
<td>GalNAcα1-6</td>
<td>GlcNAcβ1-6</td>
<td>GalNAc</td>
<td>GalNAc</td>
<td>Galβ1-3</td>
<td>GlcNAcβ1-6</td>
</tr>
<tr>
<td>GalNAc</td>
<td>Galβ1-3</td>
<td>GlcNAcβ1-6</td>
<td>GalNAc</td>
<td>Galβ1-3</td>
<td>GlcNAcβ1-6</td>
</tr>
</tbody>
</table>

In the context of this survey, it should be mentioned that in yeast glycoproteins, N-linked highmannose chains and O-linked (to Ser or Thr) short oligomannose saccharides can occur.

**Structure Analysis of Glycoprotein Glycans**

In view of the present complexity of the analytical problem, it may be evident that the determination of the primary structure of Asn-linked and of Ser- or Thr-linked carbohydrate chains on intact glycoproteins is nearly impossible. Therefore, the preparation of partial structures of the protein, having a single glycosylation site (glycopeptides), or release of the glycan chains from the glycoprotein (oligosaccharides), is a prerequisite. In Scheme 3 the major cleavage procedures for the preparation of mixtures of glycopeptides or oligosaccharides from native glycoproteins are summarized (2, 4, 5, 17-21). For a detailed analysis of the released carbohydrate chains, these mixtures have to be fractionated, and different combinations are in use. Frequently applied combinations are high-voltage paper electrophoresis and Bio-Gel P-4 gel-permeation.


Scheme 3. Major methodologies in the structure analysis of glycoprotein glycans. PNGase, peptide-N\(^{\text{N}}\)-[N-acetyl-\(\beta\)-glucosaminy]-asparagine amidase; PA, 2-pyridylamin; LC, liquid chromatography; GLC-MS, gas-liquid chromatography / mass spectrometry; 1D, one dimensional; 2D, two dimensional; NMR, nuclear magnetic resonance.

* Isolation of glycoproteins
* Preparation of carbohydrate chains
  - protease digestion
  - hydrazinolysis for GlcNAc-Asn type
  - PNGase F or A for GlcNAc-Asn type
  - endo-H and other endo-N-acetylglicosaminidases for GlcNAc-Asn type (mainly oligomannose type)
  - alkaline borohydride treatment for GalNAc-Ser/Thr type
* Fractionation and purification, if possible to homogeneity
  - gel-permeation chromatography (Bio-Gel P-4)
  - high-pH anion-exchange chromatography (CarboPac)
  - high-pressure chromatography (e.g. Lichrosorb-NH\(_2\))
* Degradation procedures
  - exo-glycosidases (\(\alpha\)-sialidases, \(\alpha\)-D-galactosidases, \(\beta\)-D-galactosidases, N-acetyl-\(\beta\)-D-hexosaminidases, \(\alpha\)-D-glucosidase, \(\alpha\)-D-mannosidases, \(\beta\)-D-mannosidases, \(\alpha\)-L-fucosidases, \(\beta\)-D-xilosidase)
  - endo-glycosidases (endo-\(\beta\)-D-galactosidases)
* Partial solvolysis
* Periodate oxidation
* Monosaccharide analysis, including absolute configuration determination (GLC-MS)
* Methylation analysis (GLC-MS)
* 1D/2D NMR spectroscopy (\(^1\)H, \(^13\)C, \(^31\)P)
* Mass spectrometry (electron impact, chemical ionization, fast-atom bombardment, electrospray/ionspray, matrix-assisted laser desorption ionization)
* Profiling (LC mapping techniques, high-pH anion-exchange chromatography, capillary electrophoresis, fluorophore-assisted-carbohydrate electrophoresis, reagent-array analyses; combination with exoglycosidases or mass spectrometry)
chromatography (19, 20, 22), or Mono Q/Resource Q, Lichrosorb-NH₂ and CarboPac liquid chromatography (23, 24). For additional information, see also Fukuda and Kobata (17), Hounsell (18, and Lennarz and Hart (19). The structural analysis of isolated carbohydrate chains can be carried out along different routes. Major strategies include methylation analysis (12), exo- and endo-glycosidase studies (2, 17-20), 1H-NMR spectroscopy (9, 10) and mass spectrometry (11, 13, 15, 16). Quite often these approaches are used in combination with each other. But also other approaches have been reported (for general reviews, see Fukuda and Kobata (17), Hounsell (18, and Lennarz and Hart (19)). At present, great effort is placed on the development of profiling procedures, which can be applied to get an impression of the carbohydrate chains present in unknown glycoproteins, or will be of use in controlling possible glycan microheterogeneity shifts in known glycoproteins (batch control). In this respect, more-dimensional HPLC mapping techniques (25), high-PH anion-exchange chromatography (26, 27), capillary electrophoresis (28, 29), fluorophore-assisted-carbohydrate electrophoresis (30), and so-called reagent-array analyses (31) seem to be highly promising.

Over the years, part of our research projects have been focused on the glycan analysis of a wide range of glycoproteins derived from many different biological sources. For an idea of our activities in this field during the last five years, see Bergwerff, et al. (32-35), Damman, et al. (36), De Waard, et al. (37), Gerwig, et al. (38), Hård, et al. (23, 39-41), Hokke, et al. (24, 42, 43), Kubelka, et al. (44), Pfeiffer, et al. (45), Van Dam, et al. (46) and Van Zuylen, et al. (47).

Recombinant Human Glycoproteins

Several reports have appeared in the literature focusing on primary structure analysis of recombinant glycoproteins and on the biological significance of the specific glyclosylation patterns. Typical examples of the structure analysis of recombinant human glycoprotein glycans, as carried out in our research group, comprise the carbohydrate chains of γ-interferon (N-glycoprotein, 48), follistatin (N-glycoprotein, 39), chimeric plasminogen activator (N-glycoprotein, 32) and erythropoietin (N,O-glycoprotein, 43, 49, 50), all produced in Chinese hamster ovary (CHO) cells. Furthermore, detailed studies were carried out on the glycan chains of IgG1 antibodies (N-glycoprotein, 34), produced in murine SP2/0 transfected cell subclones, and of insulin-like growth factor I (O-glycopeptide, 51), produced in Saccharomyces cerevisiae cells. In all these cases, 1H-NMR spectroscopy was applied as the major analytical technique.

Plasminogen Activators

Plasminogen activators (PA) are serine proteases, that catalyze the conversion of inactive plasminogen into active plasmin (52-54). Plasmin is considered to be the primary circulating fibrinolytic enzyme, which can clear thrombi from blood vessels by the degradation of the fibrin network, a process called fibrinolysis. In view of this feature, PAs are of clinical interest for the treatment of thrombolytic disorders. Two major activators have been identified, namely, tissue-type plasminogen activator (t-PA) and urinary-type plasminogen activator (u-PA). Besides the generation of recombinant plasminogen activators, hybrid variants have also been constructed, and in the framework of this minireview one of these chimeric plasminogen activators, K₄u-PA, is included.

Tissue-Type Human Plasminogen Activators

The polypeptide backbone of human t-PA, having a M₀ of about 70 kDa and built up from 527 amino acids (an N-terminal extension of 3 amino acids not included), consists of an N-terminal finger (F) domain, an epidermal growth factor (E) domain, two kringle (K₁ and K₂) domains, and a C-terminal protease (P) domain (54). It exists in a single-chain form and in a two-chain form, and the single-chain form can be converted into the two-chain form by the action of proteases like plasmin (cleavage between Arg-275 and Ile-276). In the two-chain form an A-chain comprises the domains F, E, K₁ and K₂, and a B-chain correlates with the P domain; the two polypeptides are linked through a disulfide bridge. Several reports have discussed the influence of the glycosylation of t-PA on its biochemical / biological properties, and specific information can be obtained from Hansen, et al. (55), Beebe and Aronson (56), Collen, et al. (57), Lucore, et al. (58), Wittwer, et al. (59), Wittwer and Howard (60), Howard, et al. (61), Otter, et al. (62), Berg, et al. (63), and Hajjar and Reynolds (64).

t-PA is N- and O-glycosylated, and has a carbohydrate content of approximately 7%. Two variants of t-PA occur, namely, type I t-PA, which is N-glycosylated at Asn-117 (K₁ domain), Asn-184 (K₂ domain), and Asn-448 (P domain), and type II t-PA, which is glycosylated at Asn-117 and Asn-448. A Fur residue is O-linked to Thr-61 in the E domain. So far, the primary structure of the human serum t-PA N-glycans has not been determined, but studies addressing the elucidation of the glycan chains of non-recombinant t-PA from human colon fibroblast cells (65) and Bowes melanoma cells (65-67), and of recombinant human t-PA from CHO cells (49, 68-70) and C127 mouse epithelial cells (45, 69, 71) have been reported. An overview of the results is presented in Table 1. An evaluation of the various studies shows that each cell type produces t-PA with

Table 1. Survey of structural information concerning major N-glycosylation patterns of human tissue-type plasminogen activators.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Asn-117</th>
<th>Asn-184 (+/-)</th>
<th>Asn-448</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human colon fibroblast cells (65)</td>
<td>predominantly oligomannose type</td>
<td>predominantly N-acetylatedactosamine type; Neu5Ac in α2-6 linkage</td>
<td>predominantly N-acetylatedactosamine type; Neu5Ac in α2-6 linkage</td>
</tr>
<tr>
<td>Bowes melanoma cells (67)</td>
<td>almost exclusively oligomannose type</td>
<td>predominantly N,N'-diacetylactosaminotetraose type; Neu5Ac in α2-6 linkage</td>
<td>predominantly N,N'-diacetylatedactosaminotetraose type; Neu5Ac in α2-6 linkage</td>
</tr>
<tr>
<td>Chinese hamster ovary cells (49, 67, 70)</td>
<td>almost exclusively oligomannose type</td>
<td>predominantly N-acetylatedactosaminotetraose type; Neu5Ac in α2-3 linkage; traces of Neu5Gc</td>
<td>exclusively N-acetylatedactosaminotetraose type; Neu5Ac in α2-3 linkage; traces of Neu5Gc</td>
</tr>
<tr>
<td>C127 mouse epithelial cells (45, 71)</td>
<td>predominantly oligomannose type</td>
<td>N-acetylatedactosaminotetraose type; Neu5Ac in α2-3/6 linkage; Neu5Acα2-3(SO4-6)-Gal; Galα1-3Gal</td>
<td>N-acetylatedactosaminotetraose type; Neu5Ac in α2-3/6 linkage; Neu5Acα2-3(SO4-6)-Gal; Galα1-3Gal</td>
</tr>
</tbody>
</table>

**Urinary-Type Human Plasminogen Activator**

The polypeptide backbone of human u-PA, having a Mr of about 54 kDa and built up from 411 amino acids, consists of an N-terminal epidermal growth factor (E) domain, a kringle (K) domain, and a C-terminal protease (P) domain (54). It is mainly synthesized as a single-chain form, called pro-uPA, but for the greater part converted into a two-chain form, called urokinase, by the action of proteases like plasmin (cleavage between Lys-158 and Ile-159). In urokinase the two polypeptides (A-chain/E-K domain and B-chain/P domain) are linked through a disulfide bridge. u-PA has gained biomedical interest, since it is also involved in physiological and pathological tissue destruction and cell migration processes, such as gland involution and tumor growth (52, 76). Only limited biochemical information on the glycosylation is available, and can be obtained from Zamarron, et al. (77), Sarubbi, et al. (78), Henkin, et al. (79), Li, et al. (80) and Lenich, et al. (81). u-PA is O- and N-glycosylated, and has a carbohydrate content of approximately 2%. A Fuc residue is O-linked to Thr-18 in the E domain. The P domain contains an N-glycosylation site at Asn-302, and recently a study describing the analysis of eleven oligosaccharides has been completed (35, 82). Detailed information has been included in table 2. The main component (30%) is a diantennary chain with two SO4-4GalNAcP1-4GlcNAc antennae. Of further interest is the GalNAcP1-4(Fucα1-3)GlcNAc sequence, being a variant of the well-known Lewis x determinant.

**Human Chimeric Plasminogen Activator K2tu-PA**

In the framework of developing improved PA variants, a hybrid variant of t-PA and u-PA, namely K2tu-PA, has been constructed, wherein sequences coding for the K2 domain of t-PA and for the P domain of u-PA were combined on the gene level, and CHO cells were chosen as expression system (83).
The chimeric PA was designed in order to combine the fibrin specificity of t-PA with the catalytic activity of u-PA. K\textsubscript{fu}-PA is partially glycosylated at Asn-12 (originally Asn-184 of t-PA) and completely glycosylated at Asn-247 (originally Asn-302 of u-PA), and has a carbohydrate content of approximately 6%. A plasmin cleavage site, yielding the A- and B-chains, occurs between Arg-103 and Ile-104.

In table 3 the structures of the various glycan chains are presented (32). They clearly represent the glycosylation machinery of CHO cells (Scheme 5). Of interest is the finding of minor amounts of Neu5Gc and Neu5,9Ac\textsubscript{2} residues, besides the major sialic acid Neu5Ac. A discussion on the finding of Neu5Gc and the possible relevance of its appearance in recombinant human glycoproteins of pharmaceutical interest is reported by Hokke, et al. (49).

The glycosylation sites on the A- (Asn-12) and the B- (Asn-247) chain show identical arrays of microheterogeneity. On the A-chain mono- and disialylated compounds have a higher relative abundance than on the B-chain, whereas the reverse holds for the tri- and tetrasiacylated compounds. It should be noted that the relative amounts of the various glycan chains in K\textsubscript{fu}-PA vary subtly with the applied isolation procedures.

The highly sialylated K\textsubscript{fu}-PA does not contain oligomannose- or hybrid-, and N,N'-diacylactosamide-type chains. The degree of sialylation is of interest because of clearance by the hepatic Gal (asialoglycoprotein) receptor (87). The absence of oligomannose-type structures may prolong the half-life of K\textsubscript{fu}-PA relative to that of t-PA, since the uptake from blood circulation by the hepatic Man re-
The absence of terminal SO₄GalNAc (if present, proposed uptake by a hepatic SO₄GalNAc receptor, 88, 89) and terminal GalNAc (if proposed, uptake by the hepatic asialoglycoprotein receptor, 90) elements may increase the half-life of K₃tu-PA relative to that of u-PA. Also the absence of O-linked Fuc in K₃tu-PA may be of importance, since it has been shown that in t-PA this element may mediate binding and degradation by hepatoma cells, and it is suggested that this residue is involved in the uptake of the glycoprotein from the circulation (64). Finally, K₃tu-PA lacks the potential immunogenic determinant GalNAcβ1-4(Fucα1-3)GlcNAc.

Table 2. Survey of the major N-glycosylation pattern of human urinary-type plasminogen activator (35, 82).

<table>
<thead>
<tr>
<th>Compound</th>
<th>Percentage</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>GalNAc</th>
<th>GalNAc</th>
<th>Neu5Acα2-3Gal</th>
<th>β1-4GlcNAcβ1-2Manα1-6</th>
<th>Fucα1-6</th>
<th>Manβ1-4GlcNAcβ1-4GlcNAc</th>
<th>&lt;7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SO₄GalNAc</td>
<td>Neu5Acα2-6GalNAc</td>
<td>SO₄GalNAc</td>
<td>β1-4GlcNAcβ1-2Manα1-3</td>
<td></td>
<td></td>
<td>&lt;7</td>
</tr>
</tbody>
</table>

173 Biotecnologia Aplicada 1996; Vol. 13, No. 3
Table 3. Survey of the major (> 95 %) N-glycosylation pattern of human chimeric plasminogen activator K,5u-PA (32).

| Compound | Percentage*
|----------|----------|
| - Neu5Acα2-3 Galβ1-4GlcNAcβ1-2Manα1-6 | Fuca1-6 Manβ1-4GlcNAcβ1-4GlcNAc | 8 35\(^\text{a,b}\)
| Neu5Acα2-3 Galβ1-4GlcNAcβ1-2Manα1-3 | | |
| - Neu5Acα2-3 Galβ1-4GlcNAcβ1-6 | Fuca1-6 Manβ1-4GlcNAcβ1-4GlcNAc | 1.5 20\(^\text{c,d}\)
| Neu5Acα2-3 Galβ1-4GlcNAcβ1-2Manα1-6 | | |
| - Neu5Acα2-3 Galβ1-4GlcNAcβ1-2Manα1-3 | Fuca1-6 Manβ1-4GlcNAcβ1-4GlcNAc | 0.5 6\(^\text{e,f}\)
| Neu5Acα2-3 Galβ1-4GlcNAcβ1-4 | | |
| Neu5Acα2-3 Galβ1-4GlcNAcβ1-2Manα1-3 | | |
| Neu5Acα2-3 Galβ1-4GlcNAcβ1-4 | | 13\(^\text{g,h}\)

\(^\text{a}\) 1x Neu5Ac + 1x Neu5Gc, 2 %
\(^\text{b}\) 1x Neu5Ac + 1x Neu5,9Ac2, 8 %
\(^\text{c}\) 2x Neu5Ac + 1x Neu5Gc, amount not determined.
\(^\text{d}\) 2x Neu5Ac + 1x Neu5,9Ac2, 1 %
\(^\text{e}\) 2x Neu5Ac + 1x Neu5Gc, amount not determined.
\(^\text{f}\) 2x Neu5Ac + 1x Neu5,9Ac2, 1 %
\(^\text{g}\) 3x Neu5Ac + 1x Neu5Gc, 3 %

**Erythropoietin**

Human erythropoietin (EPO), a glycoprotein that is synthesized mainly by the kidney, stimulates the proliferation and differentiation of erythroid progenitor cells (91). The polypeptide backbone of active EPO (Mr 34 kDa) is built up from 165 aminoacids, and has three N-glycosylation sites at Asn-24, Asn-36, and Asn-83, respectively, and one O-glycosylation site at Ser-126. Its average carbohydrate content is about 40 % (92-94). Urinary EPO is generally isolated from pools of urine of aplastic anemic patients. EPO has been expressed in various heterologous cell systems and the recombinant human glycoprotein is an important therapeutic agent for the treatment of anemia associated with renal failure (93, 95, 96). From several studies, it has become clear that the glycosylation is of great importance for the biological functioning of EPO. It could be demonstrated that removal or modification of the glycan chains, or prevention of the glycosylation at specific sites by site-directed mutagenesis, results in altered in vivo and in vitro activity (95, 97-105).

So far, studies have mainly been reported, describing the generation of recombinant human EPO in cells having a glycosylation machinery, such as CHO cells and baby hamster kidney (BHK) cells. With respect to the primary structure analysis of the carbohydrate chains of urinary EPO and recombinant EPO expressed in CHO and BHK cells, several investigations have been published (Urinary EPO: see 106-109; CHO EPO: see 43, 106, 108-114; BHK EPO: see 107, 115-117). It should be emphasized that the various research groups studied different batches of EPO supplied by different companies. Also the use of other cell lines has been evaluated, and typical examples comprise \(\psi\)2 cells derived from NIH/3T3 (118), insect cells (119), and cultured tobacco cells (120).

In Table 4, the monosaccharide analysis data of urinary and CHO EPO's, reported in the literature (43, 106), are compared (see also 107). At a first glance these data are quite similar, although a tendency may exist in that CHO EPO may contain slightly higher Gal and/or GlcNAc values. However, because different ensembles of carbohydrate chains can give rise to similar monosaccharide analyses, it is dangerous to draw detailed conclusions in terms of identity, similarity or non-identity of EPOs directly from such an analysis (see below).

The most detailed study so far on the N,O-glycosylation of EPO has been reported by Hokke, et al.

Scheme 5. Survey of structural aspects of Chinese hamster ovary cell derived recombinant glycoproteins.

Carbohydrate chains of recombinant human glycoproteins expressed in CHO cells

* can contain
- (phosphorylated) oligomannose-type structures
- hybrid-type structures
- N-acetyllactosamine-type structures with mono-, di-, tri-, tetra-, and tetra-antennae, including oligo-(N-acetyllactosamine) extensions
- terminal sialic acid (Neu5Ac, Neu5,9Ac2 and Neu5Gc), but only in α2-3 linkage to Galβ
* have not shown terminations
- with sialic acid in α2-6 linkage to Galβ
- with intersecting GlcNAc (GlcNAcβ1-4Manβ1-4 element)
- with Galα1-3Galβ
- via GalNAcβ1-4GlcNAc (N,N'-diacetyllactosediame type)

a Neu5Ac in α2-6 linkage has only been reported by Davidson and Castellino (84).
b This termination has only been described by Ashford, et al. (85).
c A GalNAc-transferase activity has been detected in CHO cells by Dharmesh, et al. (86).

(43, 49, 50) for CHO EPO. It includes the identification of over 35 different N-linked and two O-linked oligosaccharides, representing at least 95 mol/100 mol carbohydrate chains, and the results of this N- and O-glycosylation study are presented in Tables 5 and 6, respectively. Although the literature results of the various studies on the N,O-glycosylation of urinary, CHO and BHK EPOs are to some extent similar, discrepancies are observed on a more detailed level. The discrepancies hold even for probes of a similar biological origin, investigated by different research groups. It may be evident that the structural data published for EPO support the statement in the Introduction that primary structure analysis of glycoprotein glycans has still not reached the level of a routine analysis, which is also of importance for the growing interest in profiling techniques.

Comparison of the structures of the N-glycans of urinary, CHO and BHK EPOs shows qualitatively the same branching patterns (di-, tri-, tetra-) for the N-acetyllactosamine-type of chains (Table 5). The major portion of the identified carbohydrate chains comprise sialylated tetraantennary oligosaccharides with 0-2 Galβ1-4GlcNAcβ1-3 (N-acetyllactosamine) repeats. However, it should be noted that large differences have been observed between different pools/individuals when comparing the amounts of di- and tetraantennary structures for urinary EPO (107, see also 106). A similar finding has been reported for specific CHO EPO probes (111). Based on the analysis of the ensembles of N-glycans in more than 30 different batches of BHK EPO, Nimtz, et al. (116) indicated that occasionally di- and tri/tri'antennary oligosaccharides were barely detectable. From a biological point of view, a high content of tetraantennary carbohydrate chains is important since there is a positive correlation between the in vivo activity of recombinant EPO and the ratio of tetra- to diantennary oligosaccharides (111). Also, a high degree of sialylation (see Table 5) is of great importance since terminal Gal residues are recognized by the heparin Gal (asialo glycoprotein) receptor resulting in a rapid clearance of poorly sialylated EPO from the blood stream, having a negative effect on the in vivo activity (95).

The studies reported by Tsuda, et al. (107) and Takeuchi, et al. (108) showed a clear tendency that CHO and BHK EPO's contain a higher content of N-linked carbohydrate chains with Galβ1-4GlcNAcβ1-3 (N-acetyllactosamine) repeats than the urinary form. In recombinant EPO's the repeating N-acetyllactosamine units were found to be, at any case, in the branches arising from the α1-6 linked Man residue: Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-6Manα1-6 and/or Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-6

Table 4. Monosaccharide analysis data of urinary and CHO erythropoietins (43, 106). Mannose is taken as 3.0.

<table>
<thead>
<tr>
<th></th>
<th>Fuc</th>
<th>Man</th>
<th>Gal</th>
<th>GlcNAc</th>
<th>GalNAc</th>
<th>Sialic acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urinary</td>
<td>0.9</td>
<td>3.0</td>
<td>4.2</td>
<td>5.3</td>
<td>0.3</td>
<td>3.4</td>
</tr>
<tr>
<td>EPO</td>
<td>0.8-1.4</td>
<td>3.0</td>
<td>4.3-5.7</td>
<td>5.9-6.5</td>
<td>trace-0.5</td>
<td>3.1-4.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neu5Acα2-3</td>
<td>19b,c</td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td>3</td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neu5Acα2-3Galβ1-4GlcNacβ1-3</td>
<td>1</td>
</tr>
<tr>
<td>Neu5Acα2-3Galβ1-4GlcNacβ1-3</td>
<td>5d</td>
</tr>
<tr>
<td>Neu5Acα2-3Galβ1-4GlcNacβ1-3</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Neu5Acα2-3Galβ1-4GlcNacβ1-3</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>

---

2x Neu5Ac + 1x Neu5,9Ac2, <1 %.
3x Neu5Ac + 1x Neu5,9Ac2, 2 %.
3x Neu5Ac + 1x Neu5Gc, 2 %.
3x Neu5Ac + 1x Neu5Gc, <1 %.
3x Neu5Ac + 1x Neu5Gc, <1 %.
Table 5. Survey of the major (>95%) N-glycosylation pattern of recombinant human erythropoietin expressed in Chinese hamster ovary cells (43).

<table>
<thead>
<tr>
<th>Compound</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>&lt;1 &lt;1 7</td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>&lt;1 &lt;1 2</td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>2 1 1 9^o</td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>3 2</td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>&lt;1 3 5 1</td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
<tr>
<td>Neu5Acα2-3</td>
<td></td>
</tr>
</tbody>
</table>

Cont.
Table 6. Survey of identified O-glycosylation patterns of urinary and recombinant erythropoietin samples.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Urinary EPO</th>
<th>CHO EPO</th>
<th>BHK EPO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(106)</td>
<td>(109)</td>
<td>(106)</td>
</tr>
<tr>
<td>GalNAc</td>
<td>-</td>
<td>70</td>
<td>-</td>
</tr>
<tr>
<td>Galβ1-3GalNAc</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Neu5Acα2-6GalNAc</td>
<td>-</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td>Galβ1-3GalNAc</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Neu5Acα2-3Galβ1-3</td>
<td>-</td>
<td>+</td>
<td>32</td>
</tr>
<tr>
<td>Neu5Acα2-6</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Neu5Acα2-3Galβ1-3</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Neu5Acα2-HexHexNAc</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Relative amounts (mol %)

(The only extensions reported by Hokke, et al. (43), Takeuchi, et al. (108), Rice, et al. (112)). For a survey of the N-acetyllactosamine extensions, as found by Hokke, et al. (43), see Table 5.

Focusing on tetraantennary glycans of CHO EPO, also in the sequence Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-4Manα1-3 which has been identified (106, 113). Small amounts of tetraantennary carbohydrate chains with three repeats have been reported in Sasaki, et al. (106), Watson, et al. (113), and Linsley, et al. (114). In this case the third extension can comprise the Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-4Manα1-3 element, but also the Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-3 sequence has been identified (106). Also the occurrence of trimeric N-acetyllactosamine units has been mentioned to occur (106). Focusing on tetraantennary glycans of BHK EPO, also with Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-3 have been established (107, 116). The presence of repeats in di- and triantennary glycans of CHO EPO has been described by Hokke, et al. (43), and the presence of a repeat, Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-4Manα1-3, in triantennary glycans of CHO EPO by Sasaki, et al. (106) (see also 116). Linsley, et al. (114) reported the existence of repeats for both the tri- and the triantennary structures. For urinary and BHK EPO, the occurrence of a Galβ1-3GlcNAcβ1-4Manα1-3 instead of a Galβ1-4GlcNAcβ1-4Manα1-3 sequence in tri- and tetraantennary glycans has been proposed by Tsuda, et al. (107). In a few BHK EPO batches small amounts of N,N'-diacetyltetraosaminidic acid or hybrid type of chains were observed (116).

In all studies focused on CHO EPO's the linkage between sialic acid and galactose in the N-glycans was demonstrated to be exclusively α2-3 (Table 5), whereby Hokke, et al. (43, 49) showed that besides Neu5Ac as the major sialic acid (95%) also Neu5Gc (2%) and Neu5,9Ac2 (3%) do occur. Linsley, et al. (114) described for each detected NeuAc-containing compound a satellite wherein one Neu5Ac residue was replaced by one Neu5Gc residue (see also Table 5). The same type of linkage with Neu5Ac was also established for BHK EPOs, and in this case Neu5Gc has been mentioned as a possible minor sialic acid constituant (116). Inconsistencies exist with respect to the type of linkage in urinary EPO glycan chains. Based on methylation analysis data, Sasaki, et al. (106) found evidence for the exclusive existence of α2-3 linkages, whereas Takeuchi, et al. (108) demonstrated the occurrence of both α2-3 (major) and α2-6 (occasionally) linkages.

The occurrence of sulphate groups in urinary and CHO EPO has been suggested (108, 121), but structural studies have not been published thus far. With respect to the α1-6 fucosylation of the Asn-bound GlcNAc residues in CHO and BHK EPO's, amounts between at least 80% (106, 108, 113) and nearly 100% (43, 116) have been reported. In BHK EPO the presence of a mannose-6-phosphate group has been demonstrated (117).

In Table 6 a survey of identified O-glycosylation patterns of urinary, CHO and BHK EPO's is pre...
sented. Comparing the results of the seven studies on recombinant EPOs (43, 106, 109, 113-116) it is clear that the major carbohydrate chains are the trisaccharide Neu5Acα2-3Galβ1-3GalNAc and the tetrasaccharide Neu5Acα2-3Galβ1-3( Neu5Acα2-6)GalNAc. Additional O-linked chains comprise in general desialylated forms of these oligosaccharides. Interestingly, Linsley, et al. (114) also reported the desaccharide Neu5Acα2-6GalNAc and its desialylated form, the same saccharides as found by Inoue, et al. (109) in urinary EPO, and claimed to be unique for the urinary EPO. In the case of BHK EPO, Nintz, et al. (116) published an O-glycosylation percentage of only 60%. A similar finding of partial O-glycosylation was also reported for CHO EPO by Linsley, et al. (114).

The variations observed in the degree of sialylation, as found in the literature for recombinant EPOs, may have been caused by the use of different purification methods to obtain recombinant EPO (distribution of glycoforms which make up the final product). Other causes for the variation in the degree of sialylation and fucosylation could be the use of various cell clones, different culturing conditions, which can lead to differences in the activities of the α2-3 sialyltransferase or α1-6 fucosyltransferase, and the presence of sialidase activity for CHO cells recently reported in the culture fluid (122). In principle, the inconsistencies in sialylation may also have been caused partly by the use of different methods for the release and the purification of the carbohydrate chains. It should be noted that the non-random distribution of sialic acids in tetra- and diantennary glycans, as found by Hokke, et al. (43), is in good agreement with the branch specificity of human placental α2-3 sialyltransferase (123). With respect to the differences observed in the relative amounts of triantennary and triantennary oligosaccharides in various recombinant probes, it should be stated that these different ratios cannot result from artefacts caused by analytical problems. They may be explained by differences in the activities of the involved glycosyltransferases for the various cell lines and culture conditions.

Finally, the site-specific N-glycosylation in CHO EPO has been studied by Sasaki, et al. (110) and Linsley, et al. (114). Di-, tri-α- and tetraantennary (major) glycans were detected for Asn-24, whereas Asn-38 and Asn-83 contained mainly tetraantennary chains. With respect to the site-specific N-glycosylation pattern in BHK EPO, it was established that Asn-24 was glycosylated with di-, tri-α- and tetraantennary chains, and hybrid-type chains, whereas Asn-38 and Asn-83 contained almost exclusively the tetraantennary chains (116). In addition, recently, a minor phosphorylated oligomannose-type carbohydrate chain, GlcNAcα1-P-6Manα1-2Manα1-3[Manα1-6(Manα1-3)Manβ1-6 Manβ1-4GlcNAcβ1-4GlcNAc, was found at Asn-24 (117).

Acknowledgements

The author gratefully acknowledge the contributions of Drs. AA Bergwerff, JBL Damm, P De Waard, GJ Gerwig, K Här, CH Hokke, JHGM Mutsaers, and JFG Vliegenthart from the Bijvoet Center for Biomolecular Research at Utrecht University for their contributions to the different topics presented in this paper. These investigations were supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO), and the Netherlands Program for innovation Oriented Carbohydrate Research (OPK-k) with financial aid from the Ministry of Economic Affairs and the Ministry of Agriculture, Nature Management and Fisheries.


