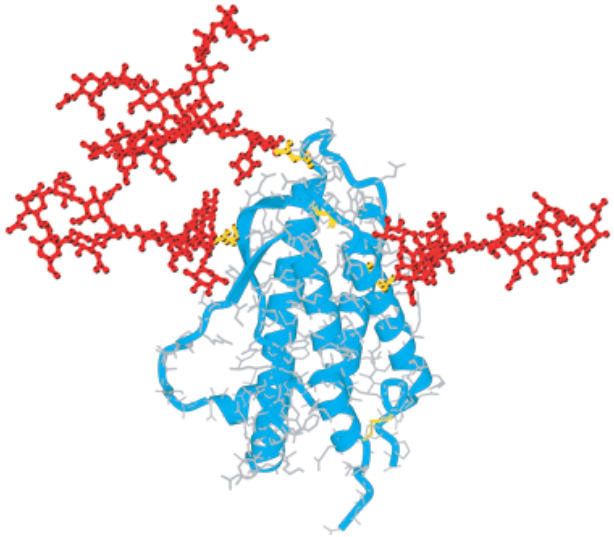
POMPEU FABRA UNIVERSITY
Department of Experimental and Health Sciences

STRUCTURAL ANALYSIS OF ERYTHROPOIETIN GLYCANS



Molecular model of erythropoietin. Courtesy of M.R.Wormald and R.A.Dwek, Oxford Glycobiology Institute, and P.M.Rudd, NIBRT

PhD THESIS

ESTHER LLOP ESCORIHUELA

Neuropsychopharmacology Programme

Municipal Institute of Medical Research (IMIM-Hospital del Mar)

IMIM hospitaldelmar

Barcelona, December 2008

POMPEU FABRA UNIVERSITY
Department of Experimental and Health Sciences

Doctoral Programme: Health and Life Sciences

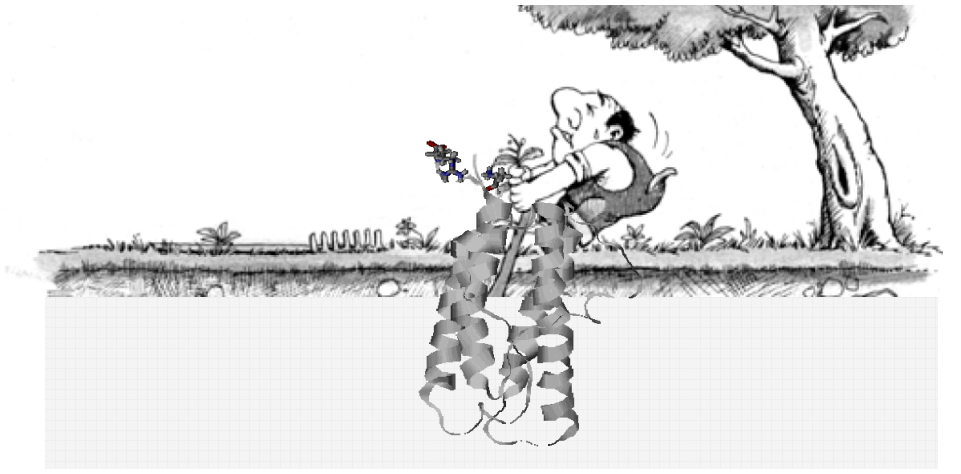
STRUCTURAL ANALYSIS OF ERYTHROPOIETIN GLYCANS

Memoria presentada por ***ESTHER LLOP ESCORIHUELA*** para optar al título de Doctor por la Universidad Pompeu Fabra. Esta tesis doctoral ha sido realizada bajo la codirección del Dr. José Antonio Pascual Esteban y el Dr. Ricardo Gutiérrez Gallego, en el grupo de Investigación en Bioanálisis y Servicios Analíticos, programa de Neuropsicofarmacología del Instituto Municipal de Investigación Médica (IMIM-Hospital del Mar). Programa de Doctorado en Ciencias de la Salud y de la Vida de la Universidad Pompeu Fabra.

Dr. José Antonio Pascual Esteban	Dr. Ricardo Gutiérrez Gallego	Esther Llop Escorihuela
Director de tesis	Director de tesis	Doctorando

Barcelona, Diciembre de 2008

A mis padres, a mi hermano
A Quim



“Las ciencias tienen las raíces amargas, pero muy dulces los frutos”

Aristóteles

Agradecimientos / Acknowledgements

Una sucesión de pequeñas voluntades consigue un gran resultado y… una sucesión de grandes esfuerzos esta tesis. Así pues, quisiera expresar en estas líneas mi gratitud a todas aquellas personas que, de una forma u otra, han participado en la realización de este trabajo.

En primer lugar, me gustaría agradecer al Dr. Ricardo Gutiérrez su implicación en la dirección de esta tesis, el esfuerzo y el tiempo que hemos compartido. Agradecerle también todas sus dulces enseñanzas sobre azúcares, sus consejos y sus críticas, porque gracias a ellas puedo hoy estar escribiendo estas líneas. Gracias por tu ejemplo de superación incansable, por tu espíritu de sacrificio, tu comprensión y confianza y como no, por tu amistad. Porque sin tu apoyo no hubiera sido posible la culminación de esta tesis, la cual constituye la herencia más valiosa que pudiera recibir. Te dedico mi más sincera admiración y respeto. Por lo que ha sido y será…Gracias.

Mi más profundo agradecimiento al Dr. José Antonio Pascual por permitirme participar en este proyecto, por la dirección y corrección de esta tesis, por el apoyo brindado y por los consejos que posibilitaron la conquista de esta meta. Como no, quiero agradecerte también la confianza que has depositado en mi, tu tolerancia, tus palabras de aliento y tus ánimos en tiempos difíciles.

Asimismo, quisiera dar las gracias al Dr. Jordi Segura, quien ha impulsado este trabajo con gran ahínco y diligencia. Gracias por todas tus aportaciones científicas y sobre todo gracias por contar conmigo.

I also extend my appreciation to Professor J.P. Kamerling and Dr. G.J. Gerwig for sharing their expertise on glycobiology and for their constant support, ideas and proposals which provided me important tools for carrying out this experimental work.

Quisiera agradecer también a la Dra. Carmen de Bolós, su colaboración en el proyecto y todas sus aportaciones en nuestras reuniones.

Agradezco al Dr. David Andreu y a los compañeros de “Proteómica de la UPF”, la posibilidad de utilizar sus equipos, sin los cuáles no podría haber llevado a cabo este trabajo, además de sus consejos en compañía de mi querido “MALDI”.

Agradezco al Dr. Jesús Joglar del departamento de química orgánica del CSIC, su soporte y ayuda en los experimentos realizados con el UPLC-ESI-TOF MS.

Como no, muchísimas gracias a todos los compañeros y amigos del departamento de farmacología, a los que aún están y a los que ya se fueron. A Belén y a Viviana que comparten conmigo esa relación de amor-odio por la EPO. Les debo a ellas los consejos y triquiñuelas experimentales. Al comando GH, y me refiero a Gerard y Jaume, cuyas tesis han ido cogidas de la mano con la mía, les agradezco sus discusiones, aunque no siempre científicas, su apoyo y comprensión porque quién mejor que ellos para entenderme. Por supuesto, gracias al otro comando, al “CJ”, Laia, Ali, Neus, Xime, Mariola y a todas sus integrantes porque siempre tuvieron tiempo para mí, dentro y fuera del laboratorio. Tiempo…si,si, tiempo y ganas de escucharme, tiempo y palabras para

animarme, para sonreírme, para aconsejarme, para consolarme, así que no es de extrañar que quiera compartir con todas vosotras esta tesis. A tí Rosalía por tu ayuda, apoyo y comprensión, por los momentos que compartimos, gracias.

Y los que bien me conocen sabrán que esto no ha sido fácil para mí, pero soy consciente que tampoco ha sido fácil para ellos, así que estas líneas son en reconocimiento a mis amigos, a quienes, a pesar de tener lejos he sentido muy cerca….los de Binéfar; Cristina, Laura, Aran, Elena, Esther y Jaime, los de Madrid; Javi y Dani. Gracias por los momentos que hemos pasado juntos, por estar ahí, siempre que os he necesitado. A vosotros, los científicos, los que ya habéis pasado por esto Lourdes, Damià y Cristina por vuestros puntos de vista, por motivarme y obligarme a seguir adelante.

Aquel pudo haber sido un día como todos los demás y… sin embargo, iba a ser diferente. Me refiero con esto al día que entró un tal Quim Mallorquí en el departamento, en el laboratorio, en el grupo de EPO, en el despacho y que más tarde entraría en mi vida y en mi corazón. Así pues, Quim podría incluirse en cualquier de los párrafos anteriores, como compañero, como consejero, como amigo…pero sabiendo que jamás existirá una forma de agradecer todo lo que él ha hecho por mí, lo menos que puedo hacer es dedicarle estas líneas. Porque tú has conseguido que los más profundos problemas no sean problema alguno, porque eres de esa clase de personas que no abundan hoy en día. Porque gracias a tus enseñanzas, a tus consejos, a tu esfuerzo y sobro todo a la confianza que depositaste en mí, lo hemos logrado. Con cariño y admiración una vez más….gracias.

Voldria agrair molt especialment a la Nuri, en Quim i la Núria, el recolzament rebut durant aquest darrers dos anys. Gràcies per la vostra ajuda i confiança. Moltes gràcies per la vostra paciencia i bon humor que han fet que poguès tirar endavant aquesta tesi.

Finalmente, quiero agradecer a mi familia todo su apoyo, su esfuerzo y compresión que hicieron que esta “científica loca” no desistiera. Gracias papá, gracias mamá, porque sin vuestros ánimos, ayuda y esfuerzo esto no habría sido posible. Gracias también al amor, la alegría y la ilusión de mis abuelas. A vosotros ya que jamás podré devolveros lo que habéis hecho por mí os dedico esta tesis. Por todo el tiempo que os he robado pensando en mí, mil gracias.

A mi hermano por su sabiduría, por servirme de guía, de apoyo y sobre todo de terapeuta. Porque entre sus muchas cualidades, destacaría la paciencia de la cuál se dice que es la madre de la ciencia. Por cubrir con tus virtudes mis defectos, Héctor por todo esto y más. Gracias

I would like to thak the world anti-doping agency (WADA) for financial support during all this period.

Gracias a l’Institut Municipal d’Investigació Mèdica (IMIM) por la ayuda recibida para la encuadernación de esta tesis.

Abbreviations

aa	amino acid
AAV	adenovirus-associated vector
ACN	acetonitril
Ala	alanine
Arg	arginine
Asn	asparagine
Asp-N	endoproteinase Asp-N
2-AB	2-amino benzamide
2-DE	two dimensional gel electrophoresis
BHK	Baby Hamster Kidney
CAM	carboxy-amidomethyl
cDNA	complementary deoxyribonucleic acid
CE	capillary electrophoresis
CFG	Consortium for Functional Glycomics
CFU-E	colony-forming unit-erythroid
CHO	Chinese Hamster Ovary
DMB	1,2-diamino-4,5-methylenedioxybenzene
DMSO	dimethylsulfoxide
DTT	dithiothreitol
EPO	erythropoietin
ER	endoplasmic reticulum
ESA	erythropoiesis stimulating agent
ESI	electrospray ionisation
FA	formic acid
Fuc	fucose
GA-EPO	gene-activated erythropoietin
Gal	galactose
GalNAc	*N*-acetylgalactosamine
Glc	glucose
GlcN	glucosamine
GlcNAc	*N*-acetylglucosamine
Glu-C	endoproteinase Glu-C (V8 serine endoprotease)
GPI	glycosylphosphatidylinositol
GS	gas chromatography
GU	glucose unit
Hb	haemoglobin
HCP	hemopoietic cell phosphatase
hEPO	human erythropoietin
HIF-1	hypoxia inducible factor-1
His	histidine
hLys	hydroxylysine
HPLC	high performance liquid chromatography
hPro	hydroxyproline
HRE	hypoxya-responsive element
IAA	iodoacetamide
Ig	immunoglobulin
IEF	isoelectric focusing
INN	International Nonproprietary Names

IOC	International Olympic Committee
IUBMB	International Union of Biochemistry and Molecular Biology
IUPAC	International Union of Pure and Applied Chemistry
JAK2	Janus kinase 2
KEGG	Kyoto Encyclopedia of Genes and Genomes
LacNAc	*N*-acetyllactosamine
Le^x	lewis x
Lys-C	endolysine protease
mAb	monoclonal antibody
MALDI-TOF	matrix assisted laser desorption ionisation time of flight
Man	mannose
MHC	major histocompatibility complex
MS	mass spectrometry
mRNA	messenger ribonucleic acid
m/z	mass-to-charge ratio
NESP	Novel Erythropoiesis-Stimulating Protein (darbepoetin-α)
Neu5Ac	*N*-acetylneuraminic acid
Neu5Gc	*N*-glycolylneuraminic acid
NIBSC	National Institute for Biological Standards & Control
NMR	nuclear magnetic resonance spectroscopy
NP	normal phase
PAGE	polyacrylamide gel electrophoresis
pI	isoelectric point
PI	phosphatidylinositol
PNGase F	peptide-N^4-(acetyl-β-glucosaminyl) asparagine amidase
Pro	proline
PTM	post-translational modification
RAAM	reagent array analysis method
ret	reticulocites
rEPO	recombinant erythropoietin
rEPO-P	recombinant erythropoietin from the European Pharmacopoeia
rEPO-T	recombinant erythropoietin from Teknika
SDS	sodium dodecyl sulphate
Ser	serine
sTfR	soluble transferrin receptor
std	standard
TFA	trifluoroacetic acid
Thr	threonine
Trp	tryptophan
uEPO	urinary erythropoietin
Val	valine
WADA	World Anti-Doping Agency
WAX	weak anion exchange
WGA	wheat germ agglutinin

CONTENTS

1. Introduction

1.1 Glycobiology

1.1.1 General overview

Molecular biology techniques have dominated the analysis of functional gene products for many years, but protein analysis has slowly come back into focus because it is now widely recognised that the investigation of many cellular processes and their regulation is best studied at the protein rather than the nucleotide level. The increase of activities in the field of proteomics promises to improve both the technologies necessary for large-scale protein analysis as well as providing new insights into important cellular processes and the pathology of diseases.

Nowadays, proteomics is present in many fields. The central goal is the identification of proteins that are either involved in a specific cellular process (cell map proteomics) or exhibit an altered expression profile as a result of a change in physiological condition (expression proteomics). Many times, altered protein profiles of tissues or cells are the result of altered protein modification, rather than altered gene expression. Therefore, an in-depth analysis of any proteome should also take into account post-translational modifications (PTMs) [1]. Given the surprisingly limited number of genes in the entire mammalian genome, including humans, post-translational modifications regulating protein function should have a more important role in cell phenotype than assumed thusfar. The central dogma of molecular biology has been revised to include protein PTMs (e.g. acetylation, methylation and alkylation, biotinylation, glutamylation, glycylation, isoprenylation, lipoylation, phosphopantetheinylation, phosphorylation, sulphation and glycosylation) [2]. Glycosylation, or the addition of glycans (carbohydrates) to proteins, is perhaps the most extensive and complex form of protein PTM and it provides the alleged functional diversity to generate many phenotypes from a limited genotype [3].

Over the years, many theories have been put foward regarding the biological role of glycans. However, there is accumulating evidence supporting the participation of glycan in cell growth and development, tumour growth and metastasis [4, 5],

anticoagulation, immune recognition, cell-cell communication and microbial pathogenesis [6]. Although additional paradigms may arise, examples from glycoprotein structure, trafficking, cell adhesion and signalling probably illustrate the major roles that glycans have. However, although the general question about what the function of glycosylation is, has possibly been answered satisfactorily, specific questions about the function of a particular glycan on a specific glycoconjugate is supposed to be a great deal of future work. In fact, the structural characterisation of glycoproteins is of paramount importance for understanding biological processes, diagnosing pathological conditions, or maintaining pharmaceutical quality control [7], just to name a few. Furthermore, this biological and biochemical knowledge has been successfully translated into drug discovery resulting in the creation of carbohydrate-based vaccines [8] and the development of new therapeutic modalities for a wide range of diseases (e.g. NESP for stimulating erythropoiesis [9] or dalteparin, tinzaparin and enoxaparin as anticoagulants [10]). Differences in carbohydrate structures may arise from different gene expression systems and culture condition [11], so glycan analyses can be helpful to discriminate endogenous glycoproteins from recombinant analogues of different origin, demonstrating their suitability to monitor the use of recombinant hormones or even detect their misuse in sports [12].

1.1.2 Glycoprotein heterogeneity

A glycoprotein is a macromolecule composed of a protein (a polymer containing a constant number of linearly arranged amino acids) and one or several glycans (oligosaccharides). The glycan is covalently attached to the protein in a co-translational or post-translational modification. In the study of protein glycosylaton two main aspects should be considered: the macroheterogeneity and the microheterogeneity [13].

1.1.2.1 Macroheterogeneity

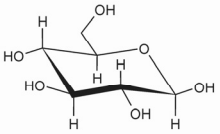
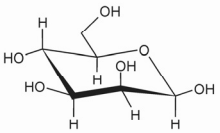
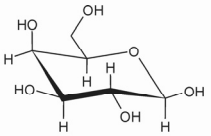
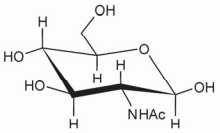
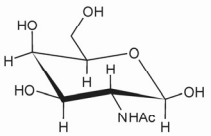
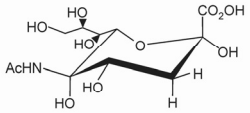
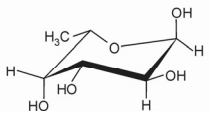
Not all amino acids can be modified by glycosylation (only Ser, Thr, Asn, Trp, hLys, and hPro) and this modification does not occur at every potential glycosylation site, even among different molecules of the same protein. This variation in occupancy with oligosaccharides is referred to as macroheterogeneity, and is a function of the site availability, enzyme kinetics, and substrate concentrations in the endoplasmic reticulum (ER) [14].

1.1.2.2 Microheterogeneity

One of the most fascinating aspects of glycosylation of proteins is the phenomenon of microheterogeneity. This term indicates that at any given glycosylation site on a specific protein synthesised by a particular cell type, a range of variations can be found in the precise structure of the glycan. Even the extent of this heterogeneity can vary considerably from glycosylation site to glycosylation site, from protein to protein, and from cell type to cell type. Glycoproteins with a common polypeptide chain but bearing different glycans are called glycoforms [15]. The occurrence of microheterogeneity is associated with the processing reactions that take place primarily in the Golgi apparatus. The common monosaccharides found in eukaryotic glycoproteins are depicted in Table I and include:

- **Hexoses:** Six-carbon neutral sugars, including glucose (Glc), galactose (Gal), mannose (Man).
- **Hexosamines:** Hexose with an amino group at the C2-position, that can be either free or more commonly, N-acetylated: *N*-acetylglucosamine (GlcNAc) and *N*-acetylgalactosamine (GalNAc)
- **Deoxyhexoses:** six-carbon neutral sugar without the hydroxyl group at the C6-position: fucose (Fuc)

Table I. Structure projections using chair conformation, symbols and abbreviations of the monosaccharides and variants most commonly found in mammalian glycosylation.

MONOSACCHARIDE	STRUCTURE	SYMBOL	ABBREVIATION
β-D-Glucose			Glc
β-D-Mannose			Man
β-D-Galactose			Gal
β-D-*N*-Acetyl-glucosamine			GlcNAc
β-D-N-Acetyl-galactosamine			GalNAc
α-L-N-Acetyl-neuraminic acid			Neu5Ac
α-L-Fucose			Fuc

- **Sialic acids:** Family of nine-carbon acidic sugars. The most common is *N*-acetylneuraminic acid (Neu5Ac). The 5-*N*-acetyl group can also be hydroxylated, giving *N*-glycolylneuraminic acid (Neu5Gc). Both molecules have the potential for additional substitutions at the hydroxyl groups on the 4-, 7-, 8-, and 9-carbons (O-acetyl, O-methyl, O-sulphate, and phosphate groups). Additional complexity arises from the fact that the O-acetyl esters can migrate along the side chain (from the 7- to the 9-position) under physiological conditions [6] (Figure 1).

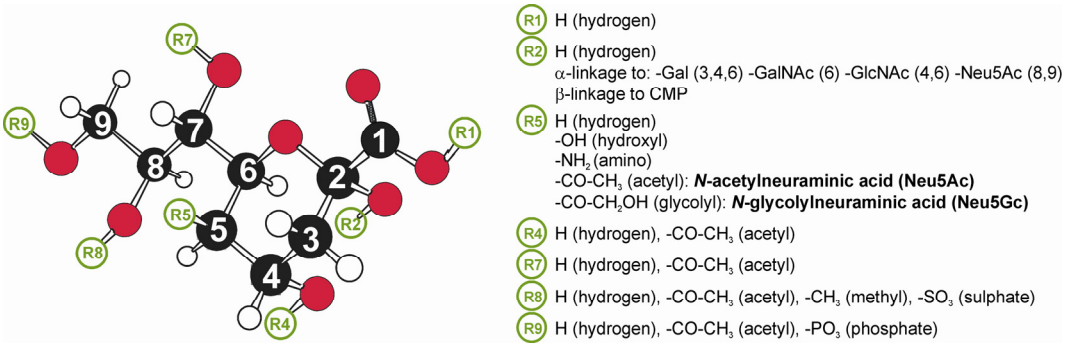


Figure 1. General structure of sialic acids and naturally occurring derivates. Adapted from Varki.

The description of the monosaccharides does not suffice to define glycosylation in a protein. In contrast to nucleotides and proteins which are linear polymers having only one type of linkage, each monosaccharide can theoretically generate an α or a β linkage to any of several labile protons in another monosaccharide or to another type of molecule. The potential structural variety is tremendous, exemplified by the fact that three nucleotide bases or amino acids can only generate six variations whereas three hexoses could produce (depending on which factors are considered) anywhere from 1.056 to 27.648 unique trisaccharides. As the number of units in the polymer increases, this difference in complexity increases exponentially. For mammalian glycans, which have a limited monosaccharide pool, the number of possible hexasaccharide structures has been calculated to be $\sim 10^{12}$ [16]. Actually, not all of these possibilities exist in nature, but it is worth noting that the list of known N-linked glycans contained more than 2000 structures in the year 2003 and this list is constantly growing [17].

1.1.3 Type of glycans in glycoproteins

There are several types of protein glycosylation based on the structure and the mechanism of synthesis. Although the most common glycosylation in naturally occurring proteins are N- and O-glycosylation, other types have also been described.

1.1.3.1 N-glycosylation

In this phenomenon, a pre-formed oligosaccharide moiety is transferred *en bloc* from a long-chain isoprenoid lipid to a potential glycosylation site via N-glycosidic linkage to the amido-group of an asparagine (Asn) residue on a nascent polypeptide. Asparagine residues to which glycans can be attached must be located in a specific sequence context within the primary structure of the protein. Glycosylated asparagine residues are almost invariably found in the sequences Asn-X-Ser or Asn-X-Thr, where X can be any amino acid except proline (Figure 2). Few cases in which the sequence Asn-X-Cys becomes glycosylated have also been described [18]. In animal cells, the sugar linked to an asparagine residue is *N*-acetylglucosamine and the linkage is always in β configuration.

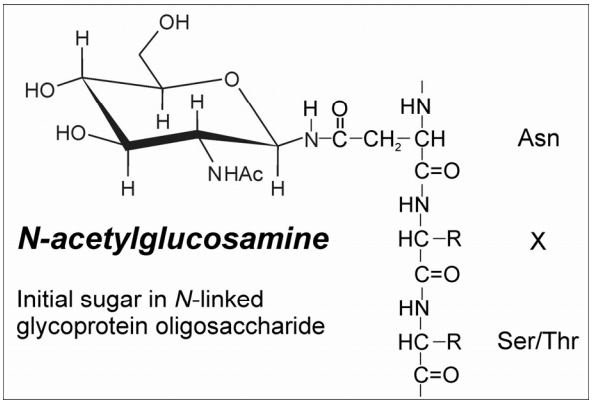


Figure 2. Typical linkage of an *N*-acetylglucosamine to an Asn residue in N-linked glycans.

The assembly of N-linked glycans occurs in a series of complex pathways that can be summarised in three major stages:

1. formation of the lipid-linked precursor oligosaccharide dolicholphosphoryl-$Glc_3Man_9GlcNAc_2$ (Figure 3)

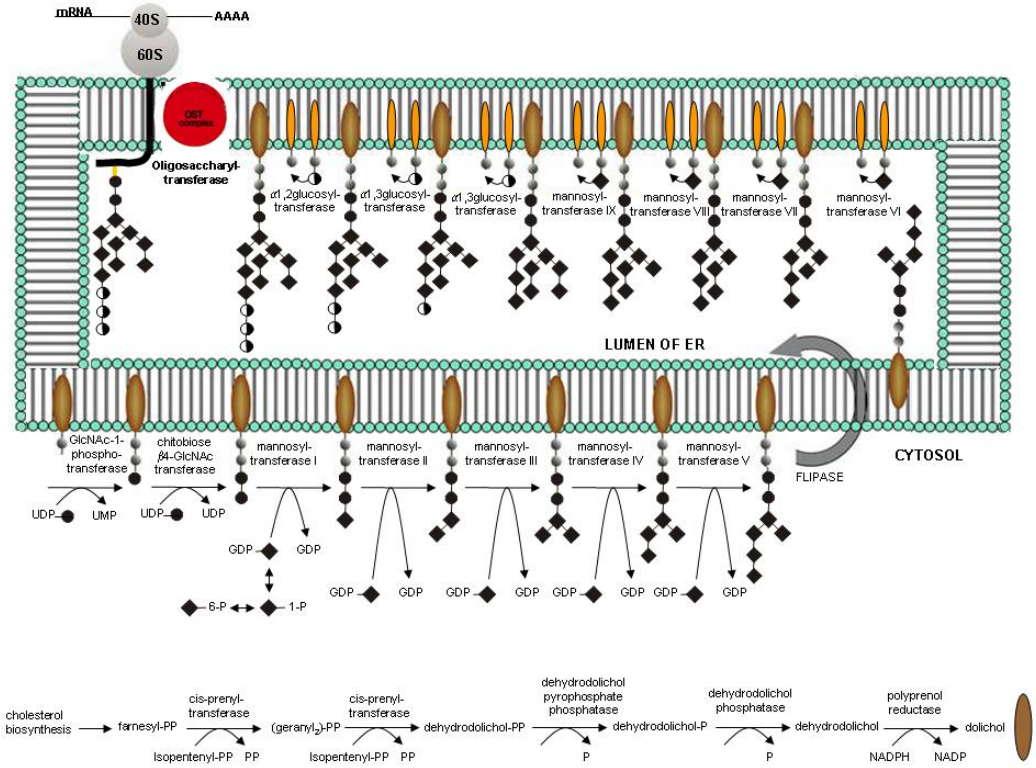


Figure 3. Pathway for generation of the dolichol-linked oligosaccharide donor for protein N-glycosylation.

2. *en bloc* transfer of the tetradecasaccharide to the polypeptide chain which constructs the Asn-X-Ser/Thr sequence (Figure 4)

3. processing and maturation of the oligosaccharide. Processing steps include removal of some of the original sugar residues (trimming) through specific glucosidases and mannosidases, followed by addition of new sugars at the non-reducing termini of the glycan through glycosyl transferases. In this biosynthetic pathway, the structure formed after the action of glucosidases and mannosidases that may remove one to four mannose residues, can remain in this state while the glycoproteins move through various luminal compartiments to the cell surface. Such glycans, containing between five and nine mannose residues, are called

oligomannose type oligosaccharides. Other glycans are processed to more complicated structures as complex type and hybrid type sugar chains. This variation is formed by the number of antennae (formed by adding from one to five GlcNAc residues to the trimannosyl core) and the branch to which the GlcNAc is added by the action of five distinct β-*N*-acetylglucosaminyltransferases (Figure 4). Then, all the nonreducing terminal GlcNAc residues, except for the bisecting GlcNAc, can be extended by addition of monosaccharide units. The number of antennae combined with kind and linkage of residues elongating the different chains will form the large number of existing complex type sugar chains. Within a cell, synthesis of glycoproteins takes place in spatially differentiated steps. Lipid-linked precursor synthesis, *en bloc* transfer to the protein and initial trimming reactions occur in the rough ER, while subsequent processing steps occur as the glycoprotein migrates through the Golgi apparatus.

Perhaps the most striking variation on the core structure is the extension of one branch by alternating Gal and GlcNAc residues. This sequence can be viewed as a series of repeated Gal-GlcNAc disaccharides and is called *N*-acetyllactosamine repeat (LacNAc). The linkages between these monosaccharides can vary, giving rise to type 1 and type 2 LacNAc families. In the most common form (type 2) the linkages are Galβ1-4GlcNAc and GlcNAcβ1-3Gal. In contrast to the type 2 chains, the linkage within each lactosamine unit is Galβ1-3GlcNAc in type 1 chains. The Galβ1-XGlcNAc groups and the Galβ1-XGlcNAc repeats are further elongated by galactosylation, or terminated by fucosylation, sialylation, glucuronylation, and/or sulphation to form the final glycan. Addition of sialic acid residue to the terminal galactose residue would prevent further elongation of the chain, so the length of the final structure results from the balance between the activities of the GlcNAc-transferase and the sialyltransferase.

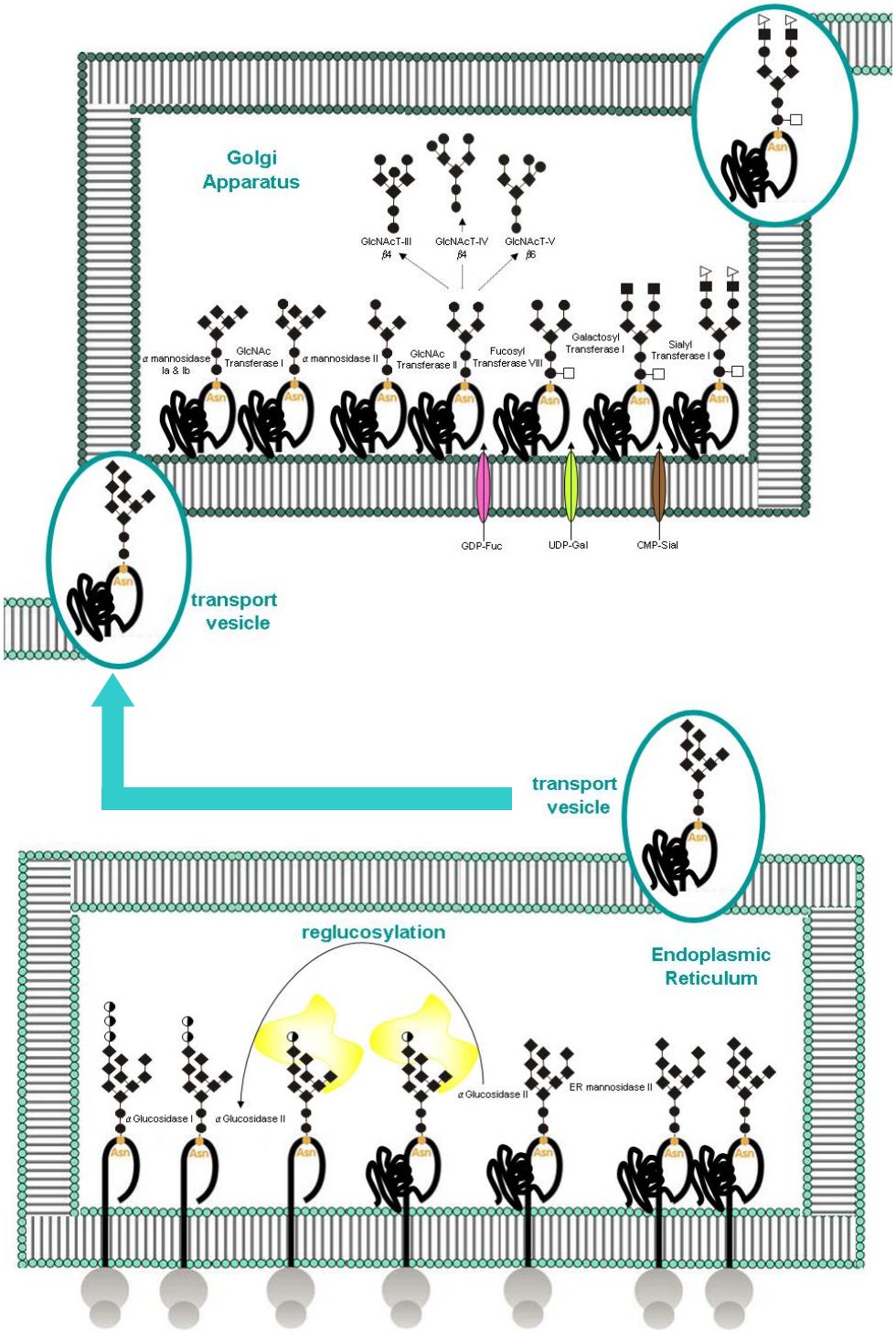


Figure 4. General overview of the pathway for N-glycoprotein biosynthesis and its location within a cell.

According to the IUPAC/IUBMB recommendations, N-glycans can be divided into three classes, all having the same pentasaccharide core Man(α1-3)[Man(α1-6)]Man(β1-4)GlcNAc(β1-4)GlcNAc(β1-N) which is called the trimannosyl core. Based on the structures and locations of the sugar residues added to the trimannosyl core, N-linked sugar chains are classified into three subgroups [19] (Figure 5):

1. **Oligomannose** type, with a variable number of D-Man residues. This term is used for up to nine Man residues ($Man \leq_9 GlcNAc_2$) and the term high-mannose type is used for more than nine Man residues ($Man >_9 GlcNAc_2$) including the terms polymannose or hypermannose used for yeast-glycoprotein-derived N-glycans. The term paucimannose type is also used in the literature to describe the truncated N-glycans ($Man <_9 GlcNAc_2$) with or without fucosylation.

2. **Hybrid** type, contain the characteristic features of both oligomannose- and *N*-acetyllactosamine-type sugar chains.

3. ***N*-acetyllactosamine** type, named also complex type. The outer chains with an *N*-acetylglucosamine residue at their reducing termini are linked to the two α-mannosyl residues of the trimannosyl core and apart from these, contain no other mannose residues. This class is subdivided into
    a. monoantennary
    b. diantennary
    c. triantennary
    d. tetraantennary
    e. pentaantennary

Antennas can present structural variations or substitutions such as:

- Bisecting GlcNAc(β1-4) residue at the β-Man residue
- Fuc(α1-6) or Fuc(α1-3) [in plants] at the Asn-bound GlcNAc residue
- Fuc substitutions in the antennas (Le^x structure connecting to Man)
- occurrence of Xyl(β1-2) at the β-Man residue (in plants)
- occurrence of GalNAc(β1-4)GlcNAc (LacdiNAc; D-GalNAc) besides Gal(β1-4)GlcNAc (LacNAc)

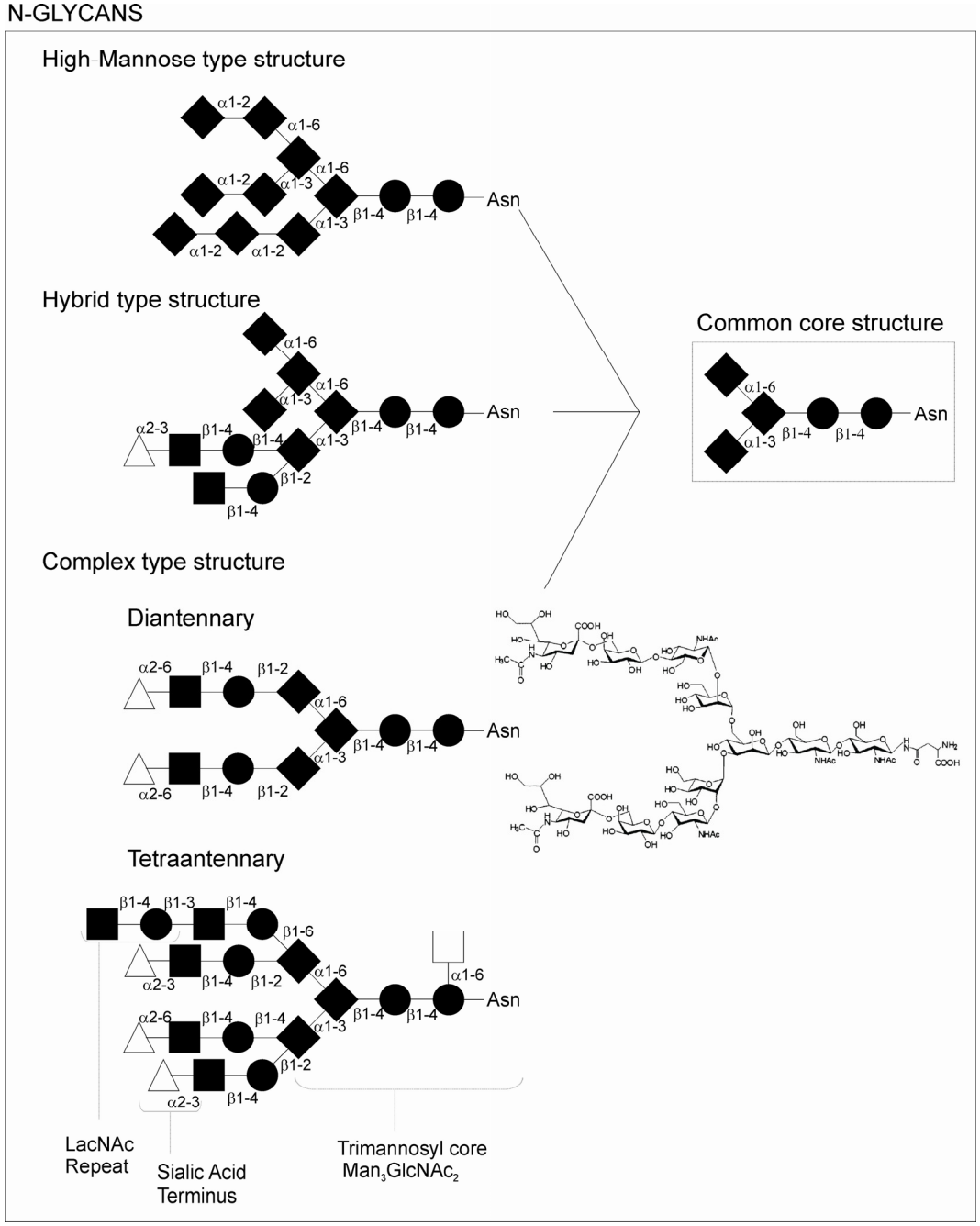


Figure 5. Examples of high mannose type glycans (top), hybrid type glycans (middle) and complex type glycans (bottom). The common core structure for all N-liked glycans is represented in the box. Short-hand notation: (□) fucose, (●) *N*-acetylglucosamine, (◆) mannose, (■) galactose, (∆) sialic acid.

1.1.3.2 O-glycosylation

This type of glycosylation results from the addition, in the Golgi apparatus, of sugar residues to the hydroxyl side chain of serine, threonine, hydroxylysine or hydroxyproline residues. Unlike N-linked glycoproteins, O-linked glycoproteins are synthesised on the protein by the addition of a single sugar residue at a time.

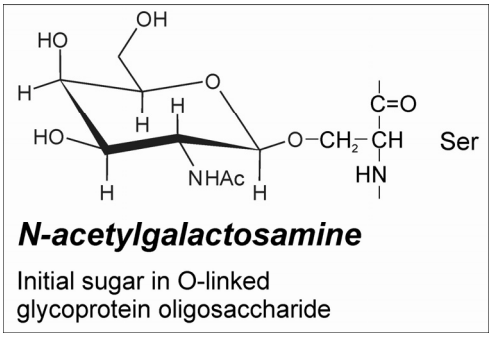


Figure 6. Typical linkage of an *N*-acetylgalactosamine to a Ser residue in O-linked glycans.

Currently, the O-linked sugar chains are grouped into the eight classes according to their core structures (Figure 7). Core 1 and 2 are most commonly found in mucins and other glycoproteins. In contrast, cores 3 and 4 are rather restricted to mucins. Core 5 was detected in few materials such as meconium and human adenocarcinoma, and core 6 only in human glycoproteins. So far, core 7 has been detected in bovine submaxillary mucin only. The rare core 8 glycan α-D-Gal-(1-3)-D-GalNAc was identified in a rather simple trisaccharide form amongst the complex oligosaccharides isolated from human respiratory mucins, collected from a patient suffering from chronic bronchitis [20]. The core structures can be further substituted with fucose and/or sialic acid, or elongated with linear or branched sequences of Galβ1-4GlcNAc and/or Galβ1-3GlcNAc groups, similar to the outer chain moieties of the complex-type N-linked sugar chains. The isozymes that catalyse the *N*-acetylgalactosaminylation (ppGaNTase) contain a ricin-like lectin motif, and many of them preferably use *N*-acetylgalactosaminylated peptide as substrate. This substrate specificity

produces clustered O-linked sugar chains, which are widely observed in many glycoproteins.

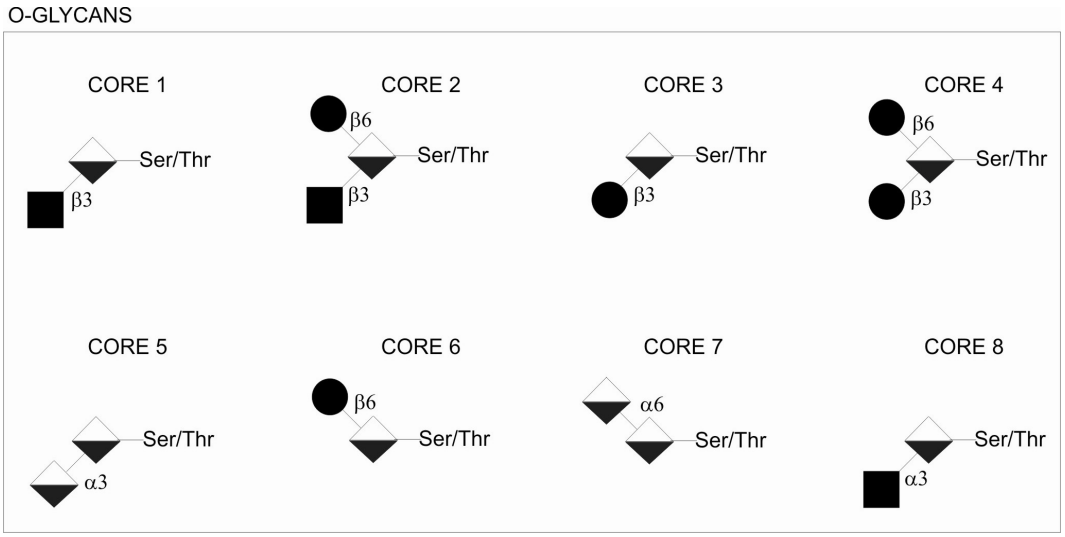


Figure 7. Core structures of mucin-type O-glycans. Short-hand notation: (◆) *N*-acetylgalactosamine, (●) *N*-acetylglucosamine, (■) galactose.

1.1.3.3 C-mannosylation

This novel type of protein glycosylation discovered by Hofsteenge *et al*. [21] is unique in that an α-mannose is directly bound to the indole C2 carbon atom of a Trp residue through a C-C bond to produce C-mannosylated Trp (CMW). It occurs at the first Trp in the consensus amino acid sequence Trp-X-X-Trp in proteins (where X can be any amino acid) [22]. There are several examples of proteins containing this type of glycosylation, including RNAse2, interleukin-12, complement proteins, properdin, thrombospondin, F-spondin, erythropoietin receptor and some mucins [23].

1.1.3.4 P-glycosylation

The attachment of sugar to protein via a phosphodiester bond represents another, quite distinct, type of glycopeptide linkage in which GlcNAc, Man, Xyl and Fuc have been found to be involved [24]. The GlcNAc-α-1-P-Ser linkage has

been found in various proteins from *Dictyostelium* including proteinase-1 [25]. Man-α1-P-Ser has been observed in several major proteins of *Leishmania*, and Xyl-1-P-Ser has been found in *T. cruzy*. Furthermore, evidence for the presence of Fuc-β-1-P-Ser in *Dictyostelium* has also been obtained [26].

1.1.3.5 Glypiation

The process of adding glycosylphosphatidylinositol (GPI) to proteins has been observed in eukaryotes ranging from yeast to mammals. GPI anchored proteins are membrane bound proteins linked at their carboxy-terminus through a phosphodiester of phosphoethanolamine, to a trimannosyl glucosamine (Man3-GlcN) core [27]. The reducing end of GlcN is linked to phosphatidylinositol (PI) that is subsequently anchored through another phosphodiester linkage to the cell membrane in its hydrophobic region. Furthermore, the Man3-GlcN oligosaccharide core may undergo various modifications during secretion from the cell.

1.1.3.6 Non-enzymatic glycosylation

Non-enzymatic glycosylation is denominated glycation and creates glycoproteins through the chemical addition of sugars to polypeptides [28]. Because of the nature of this type of glycosylation, the factors that control glycation are simply time and the concentration of sugar with respect to protein. Older proteins are more glycated [29], and people with higher circulating levels of glucose experience higher levels of non-enzymatic glycosylation, even though glycation also may occur with galactose and even fructose. This is the basis of the glycosylated haemoglobin diagnostic test used for the monitoring and long-term maintenance of blood sugar levels in diabetics [30].

1.1.4 Functional properties of glycans

A comparison among proteins serves as a good starting point to understand how glycans carry out useful biological functions. In spite of the diverse biological

roles, proteins share two common features that unify the study of their properties: each protein is synthesised as identical copy by translation of a messenger RNA template that is encoded in the genome, and the activity of a protein results from the formation of a precisely folded three-dimensional structure. As described above, glycans are assembled without a template through a series of individually catalysed reactions. The resulting structures are not unique, because many different proteins are modified with a common set of glycan structures and different copies of a single polypeptide backbone can be modified with distinct glycans. Furthermore, glycans lack a discrete, folded structure. All of these features make it difficult to establish structure-function relationships for glycans and some novel principles must be defined to describe the ways in which glycans function. One explanation for the lack of simple rules why specific glycans are attached to specific proteins is that the functions of the protein and glycan portions of many glycoproteins can be independent of each other. That is, all copies of a particular protein perform the same function regardless of what glycans are attached and all copies of a particular glycan perform the same function, although they are attached to different proteins.

Owing to the ubiquitous presence of glycans at the cell-extracellular interface, glycans are located in an environment of many proteins of very different and complex nature such as enzymes, receptors, molecules of immune-system including antibodies (immunoglobulins), cytokines and molecules of the major histocompatibility complex (MHC), structural proteins like collagen, components of the zona pellucida, which surrounds the oocyte essential for sperm-egg interaction, several hormones as follicle stimulating hormone, luteinising hormone, thyroid stimulating hormone, human chorionic gonadotropin, alfa-fetoprotein and erythropoietin [3].

The addition of carbohydrates to a peptide chain will change the shape, size, polarity, etc. of the protein structure. These glycans will affect many of the physiological and biochemical properties of the glycoprotein to which they are attached including folding, viscosity, stability, solubility, biological activity,

antigenicity, *in vivo* clearance rate, protection of polypeptide chains against proteolytic enzymes, secretion, recognition and interaction with other proteins or non-proteins components of the cell. Consequently, the roles of glycans are diverse, participating in normal biological development, tumour growth and metastasis, anticoagulation, cell-cell communication, immune defence against pathogens via the identification of exogenous carbohydrates and microbial pathogenesis. Conversely, many bacterial and viral pathogens (such as *Escherichia coli* or *Haemophilus influenza*) initially adhere to host tissues by binding specifically to carbohydrates on the host's cell surfaces. Thus, there is an interest in developing therapeutic agents that can interfere with, modulate, or exploit carbohydrate-based host-pathogen interactions [31].

1.1.5 Strategies for the analysis of glycoprotein glycans

The analysis of the covalently attached glycans to glycoprotein (glycan sequencing) is a more daunting task than the sequencing of proteins and nucleic acids. One difficulty lies in the branched nature of the structures and another in the extensive heterogeneity of the glycans. When there are 20 or more different glycans linked to a particular asparagine residue in a glycoprotein, complete sequencing of the glycans requires separation and analysis of each of the components, many of which will be present in small amounts. In the end, a profile of glycans must be established rather than a single structure. Although glycan studies are complex, they are important to perform as they often have particular relevance in many areas of advanced clinical and biomedical research. Indeed, it is becoming increasingly clear that many "biomarkers" of health and disease are in fact glycoproteins. Such biomarkers now encompass many major diseases ranging from diabetes, rheumatoid arthritis or cardiovascular pathologies to various cancers. Recent advances in analytical instrumentation and methods are increasingly allowing high-sensitivity glycoprotein analysis to be performed on a routine basis although the precise approach varies according to the nature of

each particular glycoprotein. Structural glycoproteomics is a multifaceted problem requiring the definition of several important features of glycoproteins.

- analysis of entire glycoprotein
- protein sequence (peptide mapping)
- disulfide bonding
- glycopeptides after protease digestion
- sites of glycosylation
- glycans attached to each glycosylation site after either, a treatment with anhydrous hydrazine or by digestion with PNGase F
- linkage, anomeric configuration, branching and sequence of each glycan
- presence of non-carbohydrate modifications

Although no single method has emerged as the preferred approach to glycan structure determination, the approaches used in throughout the thesis are summarised in this chapter giving a flavour of the power and the limitations of the techniques that are available at present.

1.1.5.1 Intact glycoprotein characterisation

As stated, detailed structural glycoproteomics requires several experimental approaches. Initial experiments begin with isolation of the glycoprotein that is generated by either recombinant technology, or extracted from a biological matrix such as an organ, tissue, cell or body fluids. This usually involves a variety of chromatography-based techniques. Solid-phase chromatography, lectin-based glycan capture, immunopurification, and electrophoresis are among the most routinely used technique for this task (See appendix I). The generation of information about glycoprotein glycosylation can be carried out at different levels of structural detail. The choice of the level depends largely on the research request and the sample amounts available.

Isoelectric focusing and 2-DE gel electrophoresis are often employed as the first step for isolation or analysis of glycoproteins [32]. Glycoprotein bands observed are broad due to the heterogeneous glycosylation pattern, thus making a complete separation of different glycoforms difficult. Furthermore, in 2-DE gels, glycoproteins produce characteristic “trains” of protein bands reflecting different isoelectric points and/or molecular masses of the glycoforms. However, a significant drawback of this technique, not particularly related to the glycosylation, is the frequent under-representation of membrane glycoproteins in common 2-DE gel electrophoresis due to a low solubilising power of the non-ionic and zwitterionic detergents generally used in IEF for hydrophobic proteins. In addition, many of these glycoproteins are only weakly stained by conventional dyes due to the high carbohydrate content. A sensitive detection and characterisation of the glycosylation pattern of electro-blotted proteins may be achieved by carbohydrate-specific lectins. Using a panel of lectins with different specificities, glycoproteins can be probed for defined oligosaccharide epitopes [33]. This strategy is usually applied when analysing structures that resemble others already characterised [34]. Additional information on the type of glycans attached can be obtained by combining gel electrophoretic separation and/or lectin probing in combination with treatment by specific exo- and endo-glycosidases. Using this approach, an initial assessment of the glycosylation properties of a given glycoprotein may be achieved [35].

Capillary electrophoresis (CE) has been also described for analysis of the heterogeneity in intact glycoproteins. Resolution of single glycoforms has been described, for instance, for ribonuclease B, ovalbumin, α1-acid glycoprotein and recombinant human erythropoietin [36]. Whereas the high resolving power of CE is undoubtedly very useful for comparative monitoring the glycoform populations of, for example, recombinant glycoproteins [37], it does not allow any solid conclusion on the nature of glycan chains attached.

Matrix-assisted laser desorption ionization-time of flight-mass spectrometry (MALDI-TOF MS) is extensively used in glycomics because of its robustness, low

cost, simplicity, ease of operation and data interpretation. However, when a large glycoprotein with multiple glycosylation sites is directly analysed by this technique, only an average molecular mass can be calculated because the TOF analyser is not able to resolve the molecular ions of the individual glycoforms. Despite the widespread use of MALDI-TOF MS, studies recently published have demonstrated that glycoproteins undergo fragmentation of the more labile carbohydrate (sialic acids) in the ion source or during acceleration, obtaining less reliable average molecular mass values [38]. The choice of adequate matrices and ionisation parameters are critical and influence drastically in the mass accurancy and resolution.

Electrospray ionization (ESI) MS is better suited to the analysis of intact glycoprotein populations than MALDI-TOF MS because its higher resolution permits the identification of individual glycoforms [39]. Even though this technique presents problems because of the less efficient desolvation and data complexity (multiple adduct formation), recent improvements resulted in successful analyses of highly glycosylated glycoproteins [40-42].

1.1.5.2 Peptide and glycopeptide characterisation

Once the accurate molecular mass of the intact glycoprotein is achieved, more detailed methods are performed in order to predict first, and then identify N- and O-glycosylation sites. These analyses include the characterisation of the peptides and glycopeptides by CE, HPLC, LC-ESI MS and MALDI-TOF MS after proteolytic degradation [43]. Trypsin, chymotrypsin, Asp-N, Glu-C and Lys-C are most commonly used to obtain glycopeptides carrying single glycosylation sites. Mass spectrometric identification of glycopeptides within a complex protein digest is still a challenging task due to following reasons; they often constitute only a minor portion of the total peptide mixture, the signal intensities of glycopeptides are often relatively low compared to those of non-glycosylated peptides (heterogeneity in glycan population) and glycopeptide signals are often suppressed in the presence of other peptides, especially if the glycans are

terminated with a negatively charged sialic acid residue [44]. The proteomic analyses provide important information on site occupancy and heterogeneity. However, to complete the picture, including linkages definition, branching patterns, stereochemistry, etc. glycans must be released from the glycoprotein and subjected to the glycomic analyses.

1.1.5.3 Glycan analysis

In order to obtain structural data of individual glycans, sugar moieties have to be separated from the protein. For the generation of free glycans, two general procedures are commonly employed, enzymatic digests with endoglycosidases or chemical degradation protocols. The former method is often employed for N-glycans, since PNGase F cleaves N-glycans from the polypeptide backbone. Exceptions are glycans with (α1-3) bound fucose at the reducing *N*-acetylglucosamine, as found in plant glycoporteins. The digest yields glycans with intact reducing termini and the method is applicable to glycoproteins separated and contained within an SDS-gel as well as on blotting membranes. The PNGase F digestion modifies Asn containing N-glycan to Asp. The subsequent mass shift (0.984 Da) may then be detected via ESI-MS/MS and MALDI-TOF MS sequencing of the peptide. Further use is made of glycosidases like endo F, endo H and PNGase A, which cleave at different sites within the trimannosyl core structure of N-glycans.

Since there is no enzyme available for the release of intact sialylated O-glycans from glycoproteins, chemical cleavage (β-elimination) is often used during the analysis of this glycosylation type. To identify O-glycosylation sites within a peptide sequence, a β-elimination procedure based on the use of NH_4OH results in the incorporation of NH_3 to the amino acid formerly carrying the glycan and the mass difference may be observed within the amino acid sequence by ESI-MS/MS [45]. Another interesting, mass spectrometry-based, method for the identification of sites modified by O-GlcNAc is based on a mild β-elimination followed by

Michael addition of dithiothreitol (BEMAD) or biotin pentylamine to tag O-GlcNAc sites [46].

Mass spectrometry and nuclear magnetic resonance spectroscopy (NMR) [47] are the most important methods for analysing glycans that have not previously been described or when enzymes are not available to define certain linkages. Although the amounts of material required for NMR analyses are substantially larger than for mass spectrometry, material is not destroyed during analysis. In many circumstances sample quantities are limiting, being in these cases MS the method of choice [48]. The accurate masses that can be determined by mass spectrometry can provide important compositional information to be used in combination with enzymatic digestions. Following the generation of glycan pools, both in solution and from gel-separated glycoproteins, a common approach for glycan sequencing is the application exo- and endoglycosidases of known specificity [49]. The effect of the enzymatic treatments is monitored by MS and by chromatography of the digestion products that are previously labelled at the reducing end with a fluorescent dye [50]. This method is commonly known as RAAM (reagent array analysis method) [51]. Finally, the nature and number of the constituent monosaccharides can also be analysed. Nowadays, approaches based on GC and LC procedures after hydrolysis or methanolysis are used to determine both, the monosaccharide and the sialic acids composition of carbohydrate chains [52].

1.1.5.4 Database and informatics for glycobiology and glycomics

A great deal has been learnt about the roles of glycans, but there are also many indications suggesting that glycans have functions that we do not yet understand. A range of approaches have been described and as our understanding of glycobiology broadens to encompass more examples, new routes to the application of this knowledge in the diagnosis and treatment of disease will undoubtedly emerge. For this reason, it is essential for the glycobiology community to pool and share available knowledge and resources. Indeed, this is

already taking place in different regions of the world. C. von der Lieth (Heidelberg, Germany) discussed some of the more prominent open-access consortia with respect to their methods of operation and particular strengths [53]; these include the Kyoto Encyclopedia of Genes and Genomes (KEGG) in Japan, the Consortium for Functional Glycomics (CFG) in the USA and EurocarbDB (including glycoSCIENCES.de in Germany). Strong interactions between these three, as well as further integration of other groups and resources are actively being pursued. These efforts are resulting in the development of novel resources and technologies for glycomics and also are likely to be the vanguard for future advances in integrated glycoproteomics [54] (Table II).

Table II. Summary of the main carbohydrate databases.

Carbohydrate databases		
NAME	CENTRE	ACTIVITY
Glycosciences.de	DKFZ-Heidelberg	Structures, NMR, MS, X-ray, References
KEGG-Glycan	KEGG Kyoto Encyclopedia of Genes and Genomes	Structures, References, Pathways
CFG-Glycan Database	Consortium Functional Glycomics	Structures, References, Occurrence
GlycoSuite	Proteome Systems Ltd	Structures, References, Occurrence, Disease
Bacterial Carbohydrate Structure DataBase	N.D. Zelinsky Institute of Organic Chemistry	Structural, bibliographic and related information on bacterial carbohydrate structures
E.coli O-antigen Database	University of Stockholm	

Related carbohydrate information in gene or protein databases		
NAME	CENTRE	ACTIVITY
CAZy	Architecture et Fonction des Macromolécules Biologiques (AFMB), Marseille	Carbohydrate active enzymes
GGDB	AIST	Human glycogens
KEGG Orthology	KEGG Kyoto Encyclopedia of Genes and Genomes	Glycosyltransferases
KEGG Pathway	KEGG Kyoto Encyclopedia of Genes and Genomes	Carbohydrate Metabolism

GPI Anchor Biosynthesis Report	Research Institute of Molecular Pathology	Enzymes for GPI anchor biosynthesis
Glycosylation Pathways	Consortium Functional Glycomics	
NC-IUBMB		Glycosyltransferases
SphinGOMAP		Pathways of sphingolipid biosynthesis
BRENDA	University of Cologne	Enzymes functional data
ExPaSY ENZYME		Enzyme nomenclature DB
Lectines	CNRS-CERMAV, Grenoble	3D structures of Lectins
LectinesDB	INDIAN INST. OF SCIENCE	Plant lectins
CancerLectinDB	INDIAN INST. OF SCIENCE	Lectins in Cancer
CTDL	Imperial College London	Animal lectins
PDB2LINUCS	DKFZ-Heidelberg	Online Acess to carbohydrate relate information in PDB
DOUGAL	Birkbeck College, London	DB of Glycoprotein Structures
Glycoconjugate.DB	Hokaido University	Glycan structures contained in the PDB
GlycoEpitope DB	Kawasaki laboratory	Antibodies for GlycoEpitopes
CFG-GBP-DB	Consortium Functional Glycomics	Glycan Binding Protein
CFG- Microarray Data	Consortium Functional Glycomics	Glyco-geme Chip
Pathogen Sugar-Binding Database	The MITRE Corportation	Carbohydrate sequences to which pathogenic organism specifically adhere
GlycoPep DB		A glycopeptide database

Prediction / analysis of Glycosylation positions in proteins		
NAME	CENTRE	ACTIVITY
GlySeq	DKFZ-Heidelberg	Analysis of Glycoprotein Sequences
NetNGlyc	Center for Biological Sequence Analysis	N-glycosylation
NetOGlyc	Center for Biological Sequence Analysis	O-glycosylation
YinOYang	Center for Biological Sequence Analysis	Glyco-, Phosphorylation
Big-PIPredictor	Center for Biological Sequence Analysis	GPI- anchor prediction
DGPI		GPI- anchor prediction
GlyProt	Research Institute of Molecular Pathology	*In silico* glycosylation

Graphical represtnetations and nomenclature		
NAME	CENTRE	ACTIVITY
GlycanBuilder	EUROCarbDB	Visual editor of glycan structures
LINUCS	DKFZ-Heidelberg	Linear encoding of sugars
LiGraph	DKFZ-Heidelberg	Graphical representation
IUPAC		Nomenclature
CarbosML	Mitsui Knowledge Industry	Carbohydrate sequence markup language
GLYDE	University of Georgia	Glycan Data Exchange Standard
CFG		Symbolic Nomenclature for Glycans
MonoSaccharideDB	DKFZ-Heidelberg	Systematic names for monosaccharides

3D structures		
NAME	CENTRE	ACTIVITY
SWEET II	DKFZ-Heidelberg	Generation of 3D structure
GLYCAM	CCRS	Building of carbohydrate structures for MD simulations using AMBER
Disaccharides	CERMAV	Conformational maps
GlycoMapsDB	DKFZ-Heidelberg	Conformational maps
GlyVicinity	DKFZ-Heidelberg	Analysis of Protein-Carbohydrate Interactions
GlyTorsion	DKFZ-Heidelberg	Analysis of Carbohydrate Torsion Angles found in PDB
CARP	DKFZ-Heidelberg	Carbohydrate Ramachandran Plot
Pdb-care	DKFZ-Heidelberg	PDB Carbohydrate Residue check
Dynamic Molecules	DKFZ-Heidelberg	Molecular dynamics of glycans

1.2 Erythropoietin

1.2.1 Function and production sites

Erythropoietin (EPO) is a glycoprotein hormone that regulates cellular proliferation and differentiation along the erythroid lineage. In humans, EPO is mainly produced by hepatocytes in the fetus and in the peritubular cells in the kidneys of the adult. It is secreted in response to a decreased oxygen content of the blood and targets the bone marrow (Figure 8).

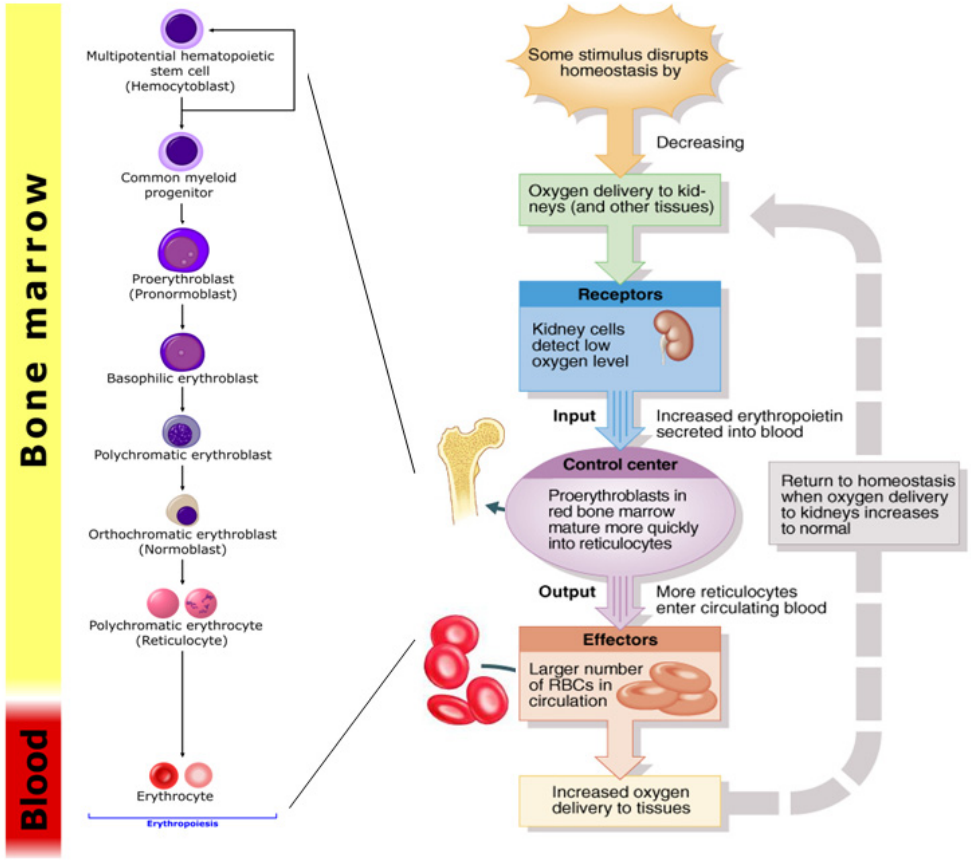


Figure 8. Scheme of human erythopoiesis (left). Feed-back mechanism for regulating the erythropoietin production (right).

Until recently, EPO was thought to act exclusively on erythrocytic progenitors. However, recent studies have shown that EPO is a pleiotropic hormone (several functions outside of hematopoiesis). For example, EPO receptor mRNA and/or protein have been shown to be present in endothelial cells, epicardium and pericardium, renal mesangial and epithelial cells, pancreatic islets, placenta and defined areas of brain [55]. Based on these findings it has been proposed that erythropoietin plays a dual role in vascular protection by preserving endothelial cell integrity and by promoting new capillary formation from pre-existing vessels into an avascular area, a process known as angiogenesis. EPO also can decrease blood-brain barrier permeability during injury [56] and maintain the establishment of cell to cell junctions. However, the physiological role of the EPO/EPO receptor system in non-erythrocytic tissues requires further clarification [57].

1.2.2 Physiology (mechanism of action)

As with EPO, the EPO receptor was earlier thought to act exclusively on erythrocytic progenitors. It is primarily expressed on the erythroid cells between the CFU-E and the pronormoblast stage of erythroid cell development. The number of EPO receptors per cell gradually decreases during erythroid cell development, and are absent in reticulocyte and mature erythrocyte.

The production and secretion of EPO, as well as the expression of EPO receptors are regulated by the tissue oxygen supply. In most tissues, including kidney, liver, uterus and additional organs like brain, the expression of both, EPO and the EPO receptor is regulated primarily by hypoxia-inducible factor 1 (HIF-1), an α, β-heterodimeric protein that is activated by a variety of stressors, including hypoxia. Hypoxia is not the only factor involved in the regulation of EPO and its receptor; other metabolic disturbances, a variety of cytokines, insulin release or anemic stress can also lead to increased expression of EPO and its receptor [58]. Once EPO binds to the receptor, a conformational change is induced resulting in a tighter connection of the two receptor molecules [57, 59] (Figure 9).

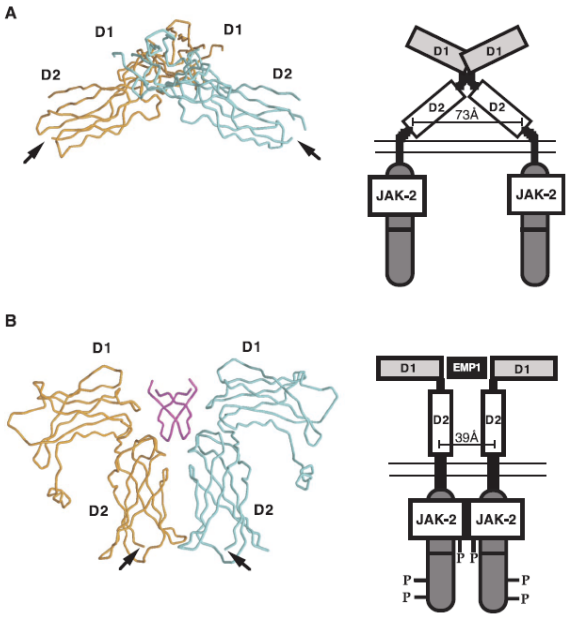


Figure 9. (A) Schematic representation of the quaternary structure of the native EPO receptor dimer (without ligand). The individual domain (D1 and D2) configuration of the two EPO receptor molecules keeps the intracellular ends far enough apart to prevent autophosphorylation of JAK2. (B) The quaternary structure of the EPO-EPO receptors complex. The EPO molecule induces a close dimer association of both the D1 and D2 domains so their intracellular regions become substrates for phosphorylation with two JAK2 molecules. From Livnah, 1999 [59].

As a result, two Janus kinase 2 (JAK2) tyrosine kinase molecules, which are in contact with the cytoplasmic region of the EPO receptor molecules, are activated. Thereupon, several tyrosine residues of the EPO receptor are phosphorylated and exhibit docking sites for signalling proteins containing SRC homology 2 (SH2) domains. Then, several signal transduction pathways are channelled, including phosphatidyl-inositol 3-kinase (PI-3K/Akt), JAK2, STAT5, MAP kinase and protein kinase C. The signalling cascade results in survival, proliferation and differentiation of erythrocytic progenitors. However, the specific roles of the various enzymes and transcriptional cofactors in these functions are only partially understood. The EPO/EPO receptor complex is internalised and degraded. In addition, the action of EPO is terminated by HCP (hemopoietic cell phosphatase) which catalyses the dephosphorylation of JAK2 (Figure 10).

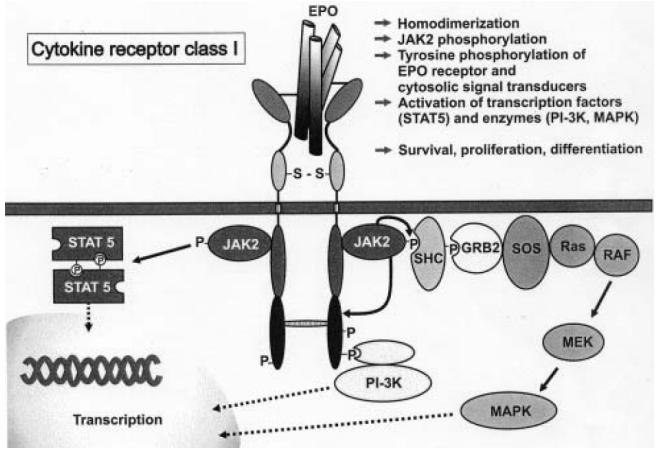


Figure 10. Simplified scheme of EPO signalling pathways. The signalling cascade results in survival, proliferation and differentiation of erythrocytic progenitors. From Jelkmann, 2004 [55].

1.2.3 Metabolism

Although it is recognised that erythropoietin is cleared from circulation with a fairly short half-life, surprisingly little is known about the metabolism and elimination of the hormone in humans. The volume of distribution of EPO is similar to, or slightly greater than, the plasma volume. This suggests that the hormone is distributed largely intravascularly with any extravascular component being of minor importance. However, there has been much debate about which organs may be important in the metabolism of erythropoietin; kidneys, liver, and bone marrow have all been suggested as playing a possible role [60] (Figure 11). Clinical and laboratory studies have led to the concept that catabolism of intact EPO by EPO receptor mediated uptake and degradation in erythropoietic tissues could account for most of the loss of the hormone from the circulation (dominant pathway). To a lower degree, EPO may be cleared by the kidneys following glomerular filtration (by means of the galactose receptor), once it is desialylated by action of tissue and blood sialidases in the liver. Evidences against a possible role of the bone marrow is provided by data suggesting that there is no difference in erythropoietin kinetics in condition of bone marrow hypoplasia, hyperplasia, or ablation.

However, in these studies, high doses of EPO were administered, producing non-linear pharmacokinetics.

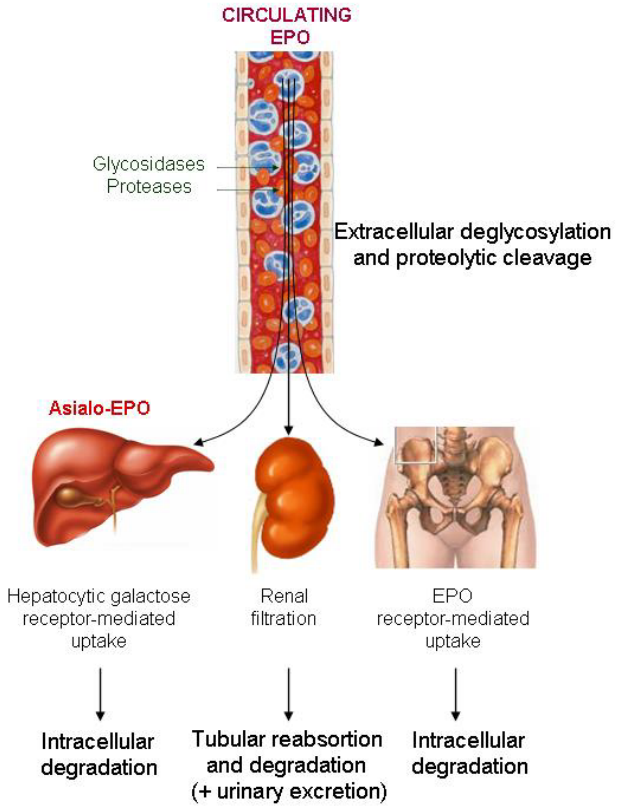


Figure 11. Sites and mechanism possibly participating in the degradation of circulating EPO. Adapted from Jelkman [61].

1.2.4 Protein structure and nomenclature

The gene encoding for erythropoietin is located on chromosome 7 of the human genome and encodes a protein of 193 amino acids including a 27 aa signal peptide. The intact protein consists of a 165 (des Arg166) AA sequence after polypeptide processing (as a result of post-translational modifications, the carboxy-terminal arginine is removed). The 166-amino acid protein has a molecular weight of 18398 Da [62]. The amino and carboxy termini of the protein are linked by a long disulfide bond between (Cys7-Cys161) and the other

disulfide bond occurs between (Cys29-Cys33). Post-translational carbohydrate addition leads to the final glycoprotein containing three N-linked (Asn-24, 38, 83) and one O-linked (Ser-126) oligosaccharide chains representing about 38% of the total mass (~30.000 Da) (Figure 12).

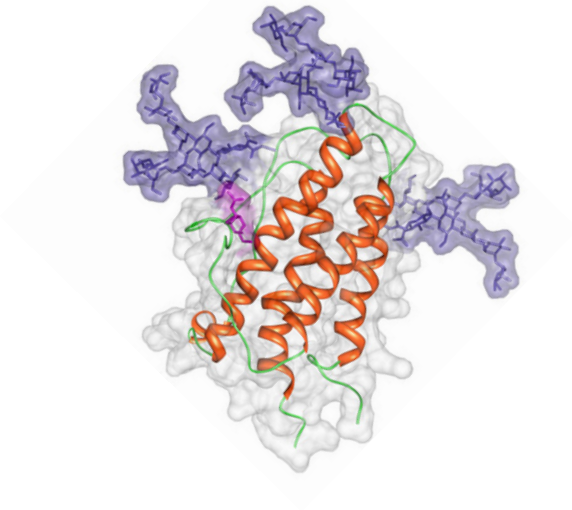


Figure 12. Model of the three dimensional structure of erythropoietin. The four α-helices are in orange, loops between helices are depicted in green. The 3 N- and 1 O-glycosylation sites are indicated by violet and pink respectively. From www.glycam.com

As explained before, carbohydrate moieties of different glycoproteins may have many diverse functions. In particular, for EPO it has been shown that the addition of carbohydrate is required for secretion from the cell, to increase the solubility of the molecule, etc. More specifically, sialic acid residues play an important role for biological activity *in vivo.* When these residues are removed from EPO, the resulting molecules have an increased activity *in vitro*, but less activity *in vivo*, presumably due to removal from circulation by the asialoglycoprotein receptor in the liver. The same evidence is obtained when EPO is expressed in *E. Coli* where only the EPO polypeptide is produced. EPO molecules are active *in vitro*, but have very low *in vivo* activity [63]. The cloning and expression of the erythropoietin gene allowed the development of recombinant human

erythropoietin as a pharmaceutical. It has become one of the most important biotechnology products because it has provided new therapeutic solutions for a variety of disease and oxygen-deficiency states. Administration of rEPO is very frequent in cases of anaemia produced by chronic renal failure, zidovudine treatment in HIV patients, non-myeloid malignancies, prematurity, autologous blood donation, chronic inflammation, bone marrow transplantation, surgical blood loss, thalassemia, and in case of chemotherapy treatments [64]. Initial synthesis using *E.Coli* was rapidly replaced by Baby hamster Kidney (BHK) cells and Chinese Hamster Ovary (CHO) cells once the key role of glycosylation on the *in vivo* activity was demonstrated. Recombinant erythropoietins used in pharmaceutical preparations have the same amino acid sequence. Nevertheless, different preparations can vary in their degree of glycosylation and in their glycan composition and/or structure due to differences in the mammalian cell lines used to express the protein, differences in downstream processing and purification strategies used. For these reasons, the World Health Organization has recommended the following International Nonproprietary Names (INN) for recombinant EPOs sharing the same amino acid sequence but differing in their glycosylation (Table III).

Table III. International Nonproprietary Names (INN) working document 05.179. World Health Organisation, 2006.

General policies for glycosylated compounds: For glycoproteins
Identification of the group with a stem, e.g. for erythropoietin: *-poetin*, indication of differences in the amino acid chain by using a random prefix and indication of differences in the glycosylation pattern by another designator, expressed by a Greek letter spelt in full and added as second word to the name (e.g. *epoetin alfa*). The Greek letters are used in the Greek alfabetical order.
Erythropoietin type blood factors
The common stem for erythropoietin type blood factors is *-poetin*.
In the case of erythropoietins, it was decided to select epoetin together with a Greek letter to differentiate between compounds of the same amino acid sequence as human erythropoietin which vary in the glycosylation pattern (see general policies for naming glycoproteins).
INNs with different amino acid sequence will be named using the *-poetin* stem and a random prefix.
darbepoetin alfa, *epoetin alfa*, *epoetin beta*, *epoetin gamma*, *epoetin delta*, *epoetin epsilon*, *epoetin zeta*, *epoetin theta*, *epoetin iota* (selected during the 42nd Consultation), *epoetin omega*.

An interesting challenge resides in the fast development of the huge market for treatment of the anaemia from chronic renal failure and malignancies. Different formulations of recombinant EPO have been developed both in academia [65, 66] and by pharmaceutical industries. Some examples of epoetin are listed in Table IV:

Table IV. Different types of rEPO, the main commercial products and manufacturing companies.

INN	*Origin*	*Commercial Products and Manufacturing Companies*
epoetin alfa	hamster cells CHO	Epoade (Sankyo), Epogen (Amgen), Epopen (Esteve), Epoxitin (Janssen-Cilag), Eprex (Janssen), Erypo (Janssen-Cilag), Espo (Kirin Brewery Company, Pharmaceutical Division), Globuren (Cilag), Procrit (Ortho Biotech).
epoetin beta	hamster cells CHO	Epoch (Chugai), Epogin (Chugai), Erytrogen (Roche), Neorecormon (Roche), Erantin (Boehringer Mannheim), Marogen (Chugai), Recormon (Boehringer Mannheim).
epoetin gamma	murine cells (C-127)	not available
epoetin epsilon	hamster cells BHK	not available
epoetin omega	hamster cells BHK	Hemax and Epomax (originally Elanex Pharmaceuticals; now owned by Baxter).
epoetin delta	human cells (HT-1080)	Dynepo (co-developed by Transkaryotic Therapies and Aventis).
epoetin zeta	human cells	Retacrit (Hospira Enterprises B.V., Hoofddorp, The Netherlands), Silapo (STADA Arzneimittel AG, Bad Vilbel, Germany).

The expiration of patents for epoetins has prompted the development of new versions of these products, which may increase competition and reduce price. For traditional pharmaceutical products, generics are easily developed and gain market demonstrating physicochemical similarity and bioequivalence by pharmacokinetic and pharmacodynamic studies in normal volunteers. This concept of a generic drug cannot be applied directly to biopharmaceuticals due to the difficulty in developing equivalent agents for glycoproteins like epoetin.

Because of their complexity, biopharmaceuticals cannot be adequately characterised using current analytical methods. Biological, physiological, and overall clinical qualities of epoetins are highly dependent on specific processes of production, purification and formulation. Thus, the use of the same cell line and gene, and similar production, purification and formulation processes does not guarantee the product will be equivalent to the erythropoietin reference standard, as even batch-to-batch variation have been observed. For this reason, the study of glycosylation from different erythropoietin preparations is necessary [67].

1.2.5 Endogenous human erythropoietin

The isolation and purification of naturally occurring EPO, like any other protein present at similar concentrations in biological fluids, has always been a very difficult task. Nowadays it continues being a great problem due to the large amount of starting material needed and the optimisation of the assay required. From early studies, it was thought that EPO was not stored in great quantities in any organ of the body, so there were no clusters of EPO-producing cells that could be isolated readily from which substantial amounts of EPO could be purified [68]. Potential sources of naturally occurring EPO included the urine or plasma of large anaemic animals, including humans [69], organs such as the kidney, and cell lines derived from tumours such as renal tumours that spontaneously produced EPO [70].

1.2.5.1 Early data about EPO

After several years of work to purify EPO from the plasma of anemic sheep, Goldwasser and Kung reported the isolation of sub-milligram amounts of apparently pure material with a potency of 7450-8250 U/mg [71]. Because the EPO amount available was so small, data obtained were limited, being only reported the apparent Mw by SDS-PAGE (46 kDa) and sedimentation coefficient (4.6 s) [72]. An interesting approach to the isolation of EPO was reported by Spivak *et al.* in 1977 [73]. They discovered that wheat germ agglutinin (WGA)

and phytohemagglutinin bound EPO significantly in crude urinary preparations. Immobilising WGA on agarose allowed an 8- to 100-fold purification of urinary EPO with recoveries of 40% but no homogeneous replicates were achieved. Even without pure material, some information on EPO's biochemical properties was obtained in the 70s, that still today is very useful. Some examples are that EPO was found to be sensitive to tryptic digestion, and that treatment with sialidase markedly shortened its *in vivo* half life [74]. This effect demonstrated that terminal sialic acid residues were present on the EPO's oligosaccharides. Furthermore, the presence of one or more disulfide bonds critical for EPO biological activity was proposed [75].

1.2.5.2 Difficulties in the purification of human urinary and serum EPO

In 1977, Miyake *et al.* described a seven step procedure that yielded highly purified human EPO [69]. Remarkably, the starting material was approximately 2550 litters of urine collected in Japan from patients with aplastic anemia. The urinary protein was isolated and lyophilised. A lot of work was needed but finally, a purification scheme was developed. It resulted in an EPO preparation with a potency of 70400 U/mg with a 21% yield. Interestingly, two pure fractions were obtained that exhibited slightly different motilities when subjected to gel electrophoresis at pH 9. An asialo EPO form was also identified. In this case, the apparent Mw of native EPO determined by SDS-PAGE was 39 kDa. A second purification method for human urinary EPO was reported by Sasaki *et al.* [76] and consisted of preparation of an immunoaffinity column by coupling a monoclonal antibody against EPO to agarose [77, 78]. Approximately 6 mg of EPO were isolated from around 700 litters of human urine. They reported a specific activity of 81600 U/mg. Some heterogeneity was observed by SDS-PAGE and western blot, being presumably due to partial deglycosylation. They also reported an N-terminal amino acid sequence of 30 aa, which differed in three positions from the N-terminal ovine EPO sequence disclosed by Goldwasser *et al.* [79]. Despite the publication of two purification methods, pure human urinary EPO remains difficult

to obtain. Therefore, several years passed before any significant biochemical characterisation of the material was published.

1.2.5.3 Biochemical properties reported for human urinary and serum EPO

The human EPO gene was cloned independently by two groups and in 1985 both published the human EPO gene sequence. Lin *et al.* [80] isolated the human EPO gene from a genomic phage library using mixed oligonucleotide probes derived from primary amino acid sequence information. Jacobs *et al.* [81], using a similar approach, obtained a human cDNA from a human fetal liver library as well as the gene itself. Also in 1985, Dordal *et al.* reported the first data about carbohydrate composition of human urinary EPO [82]. In their work it was shown that EPO digestion with PNGase F removed most of the carbohydrates from the protein indicating that the oligosaccharides were attached to asparagines residues. Furthermore it was mistakenly suggested that there was an apparent lack of O-linked sugar indicated by the absence of *N*-acetylgalactosamine residues. Removal of carbohydrates resulted in complete loss of *in vivo* biological activity in mice but activity was retained when assayed *in vitro* in bone marrow cell culture. These findings suggested that the oligosaccharide portion of EPO, although required for action *in vivo*, is not required for interaction with the target cells of the erythrocytic progenitors. On year later, Lai *et al.* reported the primary structure of human urinary EPO as well as the apparent glycosylation sites, disulfide bonds and a preliminary secondary structural characterisation by circular dichroism [83]. The authors reported that EPO has a polypeptide chain of 166 aa and contains four cysteine residues at positions 7, 29, 33, and 161. Two disulfide bonds between Cys7 and Cys161 and between Cys29 and Cys33 were also demonstrated, confirming the prediction of the presence of intrachain disulfides [84]. The calculated Mw of the de-N-glycosylated protein was reported as 18398 Da. The consensus sequences for N-linked glycosylation sites, Asn-X-Thr/Ser, at positions 24, 38 and 83 were also found. Sequence analysis of peptides indicated a serine at position 120 and no identifiable phenylthiohydantoin for

position 126. However, amino acid composition analysis revealed the presence of two serine residues in this fragment and analysis of the DNA sequence indicated that a serine is present at position 126. From these data, the authors deduced that position 126 was a serine that was O-glycosylated [83], in contrast to previous work from the same group one year earlier [82]. The circular dichroism spectrum obtained indicated antiparallel and parallel beta sheet content of 0%, a helix content of 50% and "turns and others" of 50%. Just after the human EPO gene was cloned several studies were published comparing glycosylation from urinary EPO and recombinant preparations. In 1987 Sasaki *et al.* reported that the carbohydrate moiety of uEPO was indistinguishable from rEPO synthesised in CHO except for a slight difference in sialylation [85]. Both EPOs were found to contain one O-glycan per mol of protein, an its major component was elucidated to be Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAc-Ser. The N-glycans, analysed after desialylation, for rEPO (CHO) were found to consist of diantennary (1.4%), triantennary (10%), triantennary with one LacNAc (3.5%), tetraantennary (31.8%) with one (32.1%), two (16.5%) or three (4.7%) LacNAc repeats. Around 75% of N-glycans were found to be core-fucosylated. By means of specific sialidase digestions it was demonstrated that all sialic acids contained in the rEPO N-glycans were attached through α2-3 linkage. One year later Takeuchi *et al.* carried out a new comparative study of the N-linked oligosaccharides of rEPO synthesised in CHO cells with those of uEPO [86]. They found that glycans were similar except for the sialyl linkage. All glycans from rEPO contained only Neu5Acα2-3Gal linkage while uEPO presented 60% of this kind of linkage but also a 40% of Neu5Acα2-6Gal linkage. They reported that the major glycan chains were fucosylated tetraantennary complex type with and without *N*-acetyllactosamine repeating units in their outer chain and small amounts of 2,4- and 2,6-branched triantennary and diantennary glycans. At the same time, Tsuda *et al.* published a comparative structural study of N-linked oligosaccharides of urinary (immunopurified from urine of two arbitrary groups of aplastic anemic patients) and rEPO synthesised in BHK cells [87]. They performed EPO

digestions with different exoglycosidases and isolated the resulting proteins demonstrating that removal of sialic acids yielded EPO with a higher affinity for the receptor and an increased *in vitro* activity. Furthermore, it was also found that removal of the oligosaccharides decreased thermal stability of the protein. At the glycan level, mass spectrometric analyses were performed after desialylation because sialylated glycans undergo prompt and metastable fragmentation. A significant degree of variety was observed when samples of uEPO isolated from different individuals were compared. Despite the variability observed, they reported some differences between glycans contained in uEPO preparations and those present in rEPO synthesised in BHK. Although the eight types of oligosaccharides identified in uEPO were also present in the rEPO (BHK), important differences in abundance were described. The tetraantennary structure with *N*-acetyllactosamine repeats were abundant in rEPO but was not detected in uEPO. The monosaccharide analysis resulted in higher sialic acid content in rEPO (BHK) than in uEPO. Oligosaccharides without core fucosylation were detected in both samples although in lower percentage than the core-fucosylated. Moreover, oligosaccharides with unusual structures containing the Gal(β1-3)GlcNAc group were reported in both, u- and rEPO (BHK) [88].

The confirmation that removal of N-linked glycosylation sites increased affinity for the receptor and that *in vitro* activity was retained while *in vivo* activity was nearly absent arose from the work by Yamaguchi *et al.* [89]. They prepared seven mutant genomic clones lacking the N-glycosylation sites in all possible combinations and expressed them in BHK cells. Expression of EPO in which all three N-glycosylation sites were mutated was only 10% of the wild type EPO. Delorme *et al.* showed that removal of any of the N-glycosylation sited reduced *in vivo* but not *in vitro* biological activity [90]. These authors specifically addressed the question of EPO O-glycosylation and found that mutation of serine 126 to valine resulted in a molecule that was only slightly less active *in vivo* than was wild type EPO, and concluded that O-linked glycosylation of Serine 126 is not essential for activity. In a study on the role of N-linked glycans on EPO's activity,

Higuchi *et al.* used enzymatic digestion to partially or completely remove N-linked or O-linked glycans. They observed that as de-N-glycosylation progressed, the *in vivo* activity decreased markedly whereas *in vitro* activity increased and that this activity was correlated with the number of sialic acid residues on the molecule. The completely de-N-glycosylated EPO has a 3-fold higher specific activity than the fully glycosylated hormone. On the other hand, removal of the O-linked carbohydrate chain by replacing the serine with glutamine, valine, histidine, or alanine allowed the authors to conclude that O-glycan had no effect on either *in vivo* or *in vitro* activity. Narhi *et al.* published a new function of EPO glycosylation as structural stabiliser of the protein [91]. They studied protein unfolding, assessed by circular dichroism, and compared stability between fully glycosylated EPO, asialo EPO and non-glycosylated EPO. The oligosaccharides increased EPO stability to denaturing agents such as heat and guanidinium hydrochloride. With respect to asialo EPO, it was similar to that of fully glycosylated EPO at acidic pH, suggesting that although glycosylation was important for tertiary structural stability, the sialic acid residues were not. The stability of non-glycosylated EPO was much less than either of the two other species.

In 1997 the site-specific glycan heterogeneity of human uEPO was addressed by Nielsen *et al.* [92]. The site-specific glycoprofile was determined on two levels, a qualitative profile showing which types of glycans are present, and a quantitative profile showing the relative abundances of the different forms. Again, the presence of sialic acid residues could not be unambiguously determined, since sialylated glycans tend to fragment. For this reason sialylated peptides were reported as estimation and are not completely reliable. The data demonstrated that all of the three potential N-glycosylation sites of uEPO are glycosylated. The types of glycan structures observed for uEPO were the same as those reported previously for uEPO and rEPO. The dominating N-glycans were core-fucosylated complex type containing a variable number of sialic acids. The estimated distribution of di-, tri- and tetraantennary glycans varied significantly between the three N-glycosylation sites being the Asn24 the most heterogenous, and the only

one containing diantennary glycans. The glycans at Asn38 and Asn83 were mainly tri- and tetraantennary. All the N-glinked sites were found to contain small amounts (<10%) of tetraantennary glycans with LacNAc repeats. From the comparison of the different sets of data published for uEPO some differences regarding to the percentages of each type of structures (major discrepancies in diantennary glycans) are evident. Sialic acids were disregarded in the comparison because analyses reported for uEPO prior to Nielsen were performed on desialylated glycans. It must be taken into consideration that the glycoprofiles can be influenced by different sources of uEPO and by the purification procedures employed, but a certain correlation can be considered in all articles about uEPO, especially the reported dominance of core fucosylated tetraantenary N-glycans. From the late nineties to now, further studies using different methodologies provided greater detail about oligosaccharide structures of rEPO [36, 37, 93-96]. As expected, some differences were observed between the N-linked structures on BHK and on CHO-cell-produced EPO, e.g. (60% of the rEPO protein from BHK cells was O-glycosylated while ~100% of rEPO from CHO contained O-glycan at Ser-126). However, in some aspects reported information resulted contradictory [88].

In 2001 Skibeli *et al.* reported the isolation of EPO from the serum of anemic human donors using an immunoaffinity method [97]. The existence of several glycoforms with different pI in human and murine sera had been previously suggested, however, only the urinary species of endogenous human EPO had been isolated and characterised. The main purpose of the study was to compare the glycan structures of naturally occurring EPO with those found in various forms of recombinant human EPO. The authors confirmed the differences in glycosylation for human serum EPO and rEPO detected by gel-electrophoretic methods and established that both are expressed as a population of many different glycoforms. They used charge analysis of oligosaccharides released from EPO and found that human serum EPO contained only mono-, di-, and tri-acidic oligosaccharides, lacking the tetra acidic oligosaccharides present in the

glycans of recombinant EPO. These observations corresponded with the lower molecular weight observed for fully glycosylated human serum EPO compared with rEPO. Differences between human serum EPO and rEPO must be investigated but could be explained by the fact that different cell systems display unique repertoires of oligosaccharides. The glycosylation pattern obtained for endogenous serum EPO was also different to the previously reported for also endogenous urinary EPO. Several reports on endogenous EPO indicate that circulating EPO contains fewer acidic glycoforms than urinary EPO. Wide *et al.* also showed that rEPO secreted in the urine is more acidic than rEPO analysed from serum of the same subject and conclude that the charge difference could be attributed to a difference in renal handling of the various glycoforms of endogenous EPO, indicating reabsorption of certain glycoforms [98]. Interestingly, they found that a substantial fraction of human serum EPO lacks O-glycosylation on serine 126. This may explain the apparently contradictory results reported by Dordal *et al.* [82] and Lai *et al.* [83]. Several studies reported similar Mw for purified urinary EPO and rEPOs. However, more recent papers reported urinary EPO to be smaller [99, 100]. Also for human serum EPO, Skibeli *et al.* reported a significantly lower Mw than rEPO, probably because of less extensive sialylation and other possible alterations in the glycan structures [97]. Taken together, these findings could indicate structural differences between the serum and the urinary forms of EPO caused by differential renal reabsorption or post-secretion processing of the glycans. However, further investigations are required in order to elucidate the structure of endogenous EPO glycans and demonstrate whether glycan content of EPO from patients with aplastic anemia is identical or not to healthy individuals. Sugar profiling of glycoproteins with regard to pathologic conditions has lately been a subject of much interest because detection of altered glycosylation pattern of specific glycoproteins can be used as diagnostic marker of some diseases. Sugar profiling of EPO is a challenging task because of its versatility. It may be useful in the diagnosis of medical conditions, in pharmaceutical quality control an in evaluating misuse of recombinant EPO in

sports. However, glycan analysis entails some difficulties when compared to protein analysis. The key to unambiguous protein identification in proteomics lies in the general availability of redundant information about the protein under investigation. For peptide mass mapping, any set of peptides covering a significant portion of a given protein holds sufficient information to warrant identification, but in glycosylation analysis, redundant information is obviously no longer available. In this case absolute purification of the glycoprotein is required to ensure that extracted glycans analysed derive only from the glycoprotein of interest. Currently, several groups are investigating methods for purification low-abundant proteins, including EPO, from urine and serum as well as finding alternative approaches, such as the assay currently in development by the Swedish biotechnology company MAIIA [http://www.maiiadiagnostics.com].

At present, efforts on the development of an approach for efficient endogenous EPO purification are also being made within our group. Some preliminary studies, showed isoform discrimination when urinary EPO was immunopurified. Either the monoclonal anti-human EPO antibody (R&D, AE7A5) retained preferently the more basic isoforms of EPOs or elution solution (pH 2.2) is not capable of eluting the more acidic bands. A poster including these results [101] is included in the Appendix I.

The optimisation of immunoaffinity purification protocols was considered essential from the beginning, not only for the analysis of real serum or urine samples at a later stage, but from the fact that the international urinary EPO standard is an enriched extract from human urine rather than a pure substance, and thus subject to all sorts of interferences. By using an IAC purification method developed in our laboratory the urinary EPO standard from NIBSC was further purified. Since we are interested in studying particularly those endogenous isoforms that may be confused with exogenous by the IEF method, the selectivity of the monoclonal antibody resulted favourable to our interests. In a first attempt to understand the IEF profiles from uEPO and rEPOs, analyses of glycans from immunopurified uEPO standard from NIBSC were performed. Results obtained

indicated small structures, mainly diantennary N-glycans, and absence of additional charges different from sialic acids, as was expected based on uEPO IEF profile [102] (Appendix II). Glycan differences between the immunopurified urine EPO standard preparation and the recombinant do not only arise from number of antennas, but also branching, sialic acids content and absence of Neu5Gc in urinary EPO preparation. These findings are both, contradictory with previous works reporting similar structures for both erythropoietins [85, 86] and in disagreement with the more acidic IEF profile for uEPO. Given that erythropoietin is normally present in urine at a very low concentration (10 ng/L) in comparison with the total protein concentration (>50 mg/L) [103] and other abundant glycoproteins as for example Tamm-Horsfall glycoprotein which is excreted in ~15 mg/L in humans [104]; the glycan pool obtained from immunopurified uEPO sample could not be considered pure enough or attributable to uEPO. The glycans from non-specific retained glycoproteins different from EPO probably masked those belonging to EPO.

Together with the optimisation of IAC columns in terms of improving purification efficiency, recovery and eliminating all possible non-specific binding of other glycoproteins different from EPO, a novel plate-immunopurification method has been developed in our group (Mallorqui *et al.* [105]). Although EPO isoform discrimination was solved, the low sample capacity makes this technique very useful for IEF and western blot purposes but rather inefficient for immunopurifying the amount required for glycan characterisation. Despite these efforts, the natures of additional charges proposed in urinary erythropoietin have not yet been identified.

1.2.6 New alternatives to EPO: analogues and mimetics

As with many other therapeutic areas in modern-day medicine, scientific advances in drug development (using such techniques as recombinant DNA technology, site-directed mutagenesis, pegylation of molecules, peptide library screening, and gene transfer) have resulted in the development of potential new agents and strategies for stimulating erythropoiesis. Although other strategies for increasing oxygen delivery to tissues are being developed (e.g., respirocytes or PBI-1402 created by ProMetic) in this chapter, only the therapeutic molecules that exert their activity via the EPO mechanism of action are reviewed.

1.2.6.1 Novel Erythropoiesis Stimulating Protein (NESP)

Despite the undoubted therapeutic efficacy of epoetin, a major limitation of this very often chronic treatment is that it usually has to be administered by injection two or three times a week. Efforts directed at producing longer-acting erythropoietin analogues that should retain the biological properties of erythropoietin but require less frequent dosing gave a novel erythropoiesis stimuating protein. The new erythropoietin analogue NESP, pharmaceutically named darbepoetin alfa, was synthesised using the latest techniques of DNA technology [63]. It was released in 2001 and marketed by Amgen. NESP was engineered to contain five N-linked oligosaccharide chains, whereas native erythropoietin contains only three (Figure 13). The amino acid sequence of this molecule differs from that of native human EPO at five postitions (Ala30Asn, His32Thr, Pro87Val, Trp88Asn and Pro90Thr), which allows attachment of two additional oligosaccharides at asparagine residue positions 30 and 88. This increased the Mw from ~30 kDa to ~37 kDa, the average carbohydrate content of the molecule from 38% to 51% and the maximum number of sialic acids from 14 to 22. These changes resulted in an increase of its *in vivo* half-life of three-fold (25.3 h for NESP vs 8.5 h for epoetin alfa after intravenous injection), meaning less frequent administration [106].

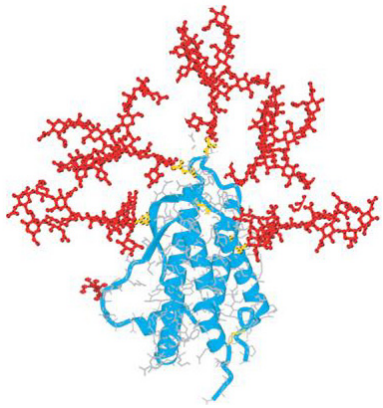


Figure 13. Model of the three dimensional structure of NESP. The four α-helices are in blue and the 5 N- and 1 O-glycosylation sites are indicated in red. Adapted from M.R.Wormald, R.A.Dwek (Oxford Glycobiology Institute) and P.M.Rudd (NIBRT).

1.2.6.2 Continuous Erythropoietin Receptor Activator (CERA)

CERA is a methoxy polyethylene glycol-epoetin beta complex that differs from erythropoietin through the integration of amide bonds between the N-terminal ε-amino group of lysine (predominantly lysine-52 or 45), using a single succinimidyl butanoic acid linker [107]. The resulting molecule integrates a single 30 kDa polymer chain into the erythropoietin molecule, thus increasing the molecular weight to twice that of epoetin at about ~60 kDa, and considerably increasing its elimination half-life (Figure 14). Due to its higher Mw, CERA is not expected to be significantly secreted in urine for which its detection will probably be done easier in serum. The application of SDS-PAGE enables a clear differentiation of CERA and rEPO β and their IEF profile show six bands in both but those corresponding to CERA are located in a more basic region [108]. Preclinical studies have shown that CERA has a long half-life than epoetin (134 h vs 8.5 h after intravenous injection). Furthermore, the binding characteristics with the erythropoietin receptor are different, mainly because CERA association to the receptor is relatively slow. These observations, combined with the pharmacokinetic data, suggest that extend administration intervals may be feasible with CERA [109].

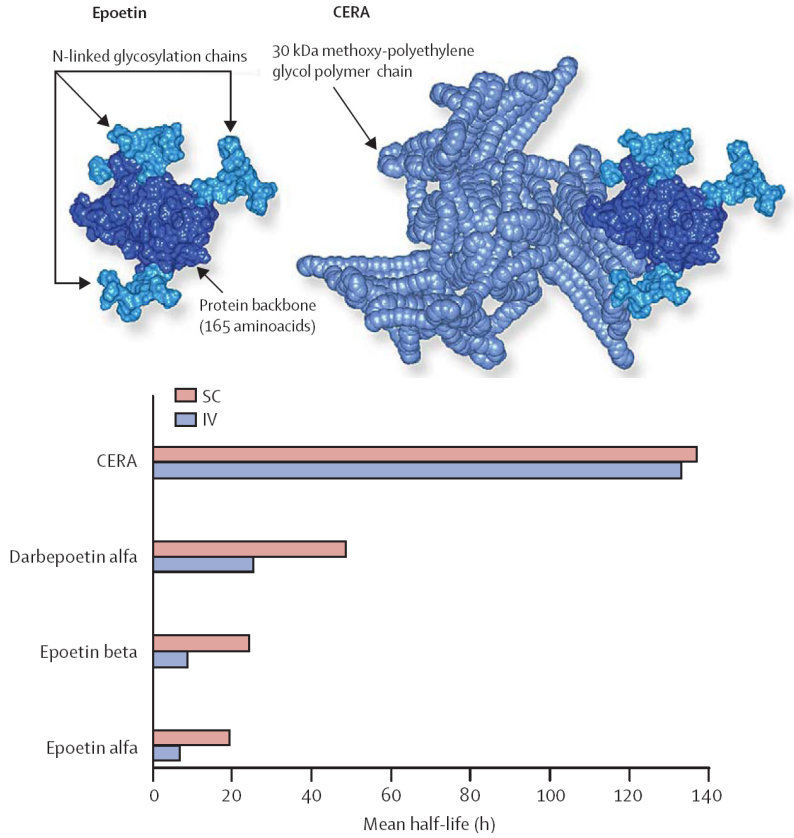


Figure 14. Comparison of epoetin and CERA structures (top) and representation of mean half-lives of CERA, darbepoetin alfa (NESP), epoetin beta and epoetin alfa (bottom). From Macdougall, 2006, [9].

1.2.6.3 Synthetic Erythropoiesis Protein (SEP)

SEP is another erythropoietin polymer that exploits recent advances in the chemical synthesis of proteins and polymers. Using solid phase peptide synthesis and branched precision polymer constructs, a 51-kDa protein-polymer construct was made containing a 166-amino-acid polypeptide chain similar to the sequence of EPO but differs significantly in the number and type of attached polymers. Instead N-glycans, SEP contains two branched polymer moieties bearing a total of eight negative charges (Figure 15). The polymer stimulates erythropoiesis through activation of the erythropoietin receptor.

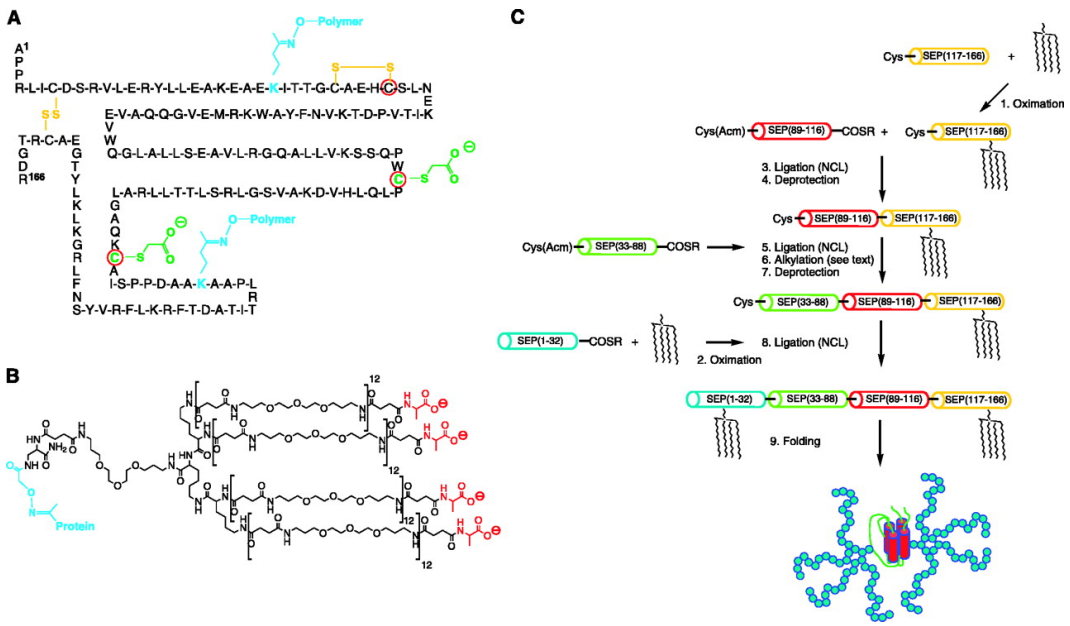


Figure 15. Molecular structure of SEP. (A) Primary amino acid sequence and disulfide bonds. The sites of polymer attachment are coded in blue, and the carboxymethyl-modified cysteine residues in green. The three ligation sites are circled in red. (B) Structure of the branched, negatively charged polymer moiety. The aminooxyacetyl chemoselective site of protein attachment is coded in blue and the charge control center in red. (C) Scheme for the synthesis of SEP by chemical ligation. From Kochendoerfer, 2003, [110].

In SDS-PAGE, SEP migrates as a single sharp band with an apparent molecular mass of ~73 kDa, but mass spectrometry revealed a molecular weight of ~50.8 kDa. Its isoelectic point is ~5 in accord with a single band present in IEF. It has been reported that SEP had comparable specific activity *in vitro*, but superior duration of action *in vivo* (longer circulation lifetime) than EPO [110] (Figure 16).

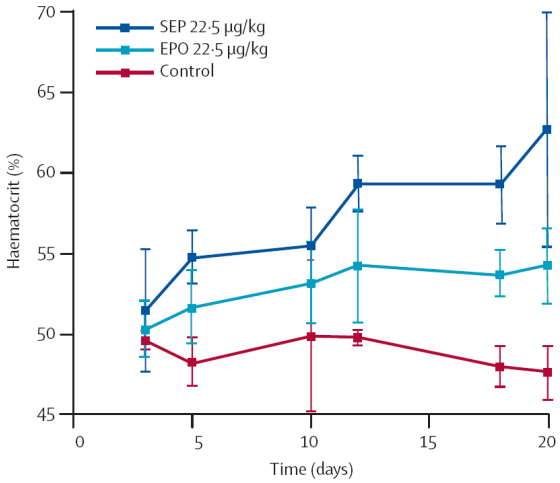


Figure 16. Haemopoietic activity SEP and EPO given once weekly to normal mice. From Macdougall, 2006, [9].

1.2.6.4 Erythropoietin oligomers and fusion proteins

Another interesting possibility for the anemia treatment is derived from a cDNA encoding a fusion protein of two complete human EPO molecules linked through flexible chemical linkers (17-aa peptide) or with the Fc portion of immunoglobulin (Ig). The Fc portion of Ig imparts the prolonged *in vivo* half-life characteristic of Ig, due, at least in part, to the oligosaccharide attached to this portion of the molecule (e.g. CTNO 528; Centocor®) [111]. All fusion proteins have been designed to increase the half-life in circulation and extend the biological effects of EPO *in vivo*.

1.2.6.5 Erythropoeitin mimetic agents

Using random phage display libraries and receptor affinity purifications, it was possible to isolate a group of small peptides with no structural homology to the endogenous EPO molecule that act as full agonists of the EPO receptor. EPO mimetics and hypoxia inducible factor-1 (HIF-1) agonists constitute additional efforts to mimic the actions of EPO in a clinical situation. Perhaps the most important reason to pursue of agents mimicking the action of EPO, especially

small molecules, is the possibility of alternative means of administration including oral and inhalation. On the other hand, these small molecules possess much shorter *in vivo* half-lifes than recombinant EPO. Such limitation can be overcome by combining the polypeptide to a polymer [e.g. Hematide created by Affimax is synthetic PEGylated peptide-based erythropoiesis-stimulating agent (ESA)]. Attempts to design orally active erythropoiesis-stimulating agents have led to small non-peptide molecules capable of supporting the *in vitro* proliferation of EPO responsive cells. Several compounds have been created, but their EPO receptor affinity and biological activity were much lower than those of EPO. Although they did not materialise into clinical application, such findings show that small, orally active molecules with interesting potential exist [112].

1.2.6.6 Erythropoietin gene therapy

A variety of methods have been reported for experimental gene therapy, replacing rEPO injections. The use of genetically modified fibroblasts secreting EPO resulted in Hb normalisation in EPO-deficient mice [113]. Another method could be the injection of a plasmid-encoded EPO into the skeletal muscle followed by electroporation (gene incorporation into nuclear DNA using an intensified electric current). A hypoxia-dependent gene therapy with an adenovirus-associated vector (AAV) carrying EPO gene, driven by hypoxia-responsive element (HRE), has been tested *in vivo*. Injection of this AAV into the skeletal muscle of EPO-deficient mice resulted in an increase in haematocrit from 20 to 50%. Interestingly, the haematocrit level stabilised without over-shooting the normal ranges as the promoter was switched off by the oxygenation of peripheral blood [114]. This finding could have important clinical relevance as normalisation of Hb without running any risk of developing secondary polycythaemia. Gene therapy includes also implantable capsules containing cells expressing the EPO gene either constitutively or in a regulated form (e.g. EPO expression in mice and rats only dependent on orally administered mifepristone) [115]. In 2002, a British pharmaceutical company (Oxford BioMedica) developed

Repoxygen as a treatment for severe anemia. Repoxygen is not merely another EPO analogue drug, as those explained above in this chapter, it is an experimental virus designed to deliver a therapeutic gene and insert it into a person's DNA. Repoxygen is the tradename for a type of gene therapy that induces controlled release of erythropoietin (EPO) in response to low oxygen concentration. It has been developed in mice, is still in preclinical development and has not been extensively tested in humans. In 2004, Takacs *et al*. [116] made use of the ability of antigen-specific lymphocytes to undergo repeated cycles of antigen-driven clonal expansion and contraction as an elaborate tool for EPO gene delivery. They introduced genes encoding EPO into a small number of antigen-specific B lymphocytes of transgenic mice. The mice were then immunised with a specific antigen, phycoerythrin, giving rise to a pool of B cells that would proliferate and express EPO when stimulated by exposure to phycoerythrin. The mice were subsequently sacrificed, and lymphocytes were harvested from their spleen and lymph nodes and injected into mice with anaemia due to EPO deficiency. Serum EPO levels and haematocrit values increased in the recipient mice in response to the administration of phycoerythrin, and the time course of this response was consistent with the known kinetics of the clonal expansion of B cells in response to an antigen. Furthermore, the effect was maintained during several cycles of antigen challenge. None of these approaches has yet gone beyond the preclinical stage and it is likely that this approach to EPO therapy will require more investigations [117].

1.2.7 Erythropoietin and sport

The ability to carry sufficient oxygen and nutritional substances to the muscles represents the major limit to intensity and length of physical effort. Despite the very effective homeostatic mechanisms humans possess, oxygen resources are rapidly consumed during intense and extended physical activities resulting in a decrease of muscular function. As the largest part of the oxygen of the blood is normally carried by red blood cells, substantial rises in erythrocytes induce

greater power and resistance [118]. The International Olympic Committee (IOC) officially banned blood doping after the 1984 Olympics. Following the cloning of the EPO gene in 1985, the rEPO became commercially available for therapeutic purposes (1987) and its potential usefulness in enhancing athletic performance became clear. Beside a few reports proving the benefits on athletic performance induced by the administration of rEPO, the theoretical advantages of this practice have enormously favoured its diffusion. The advantages offered by the administration of rEPO are basically identical to those following the practice of blood doping. Additionally, some traditional adverse effects of blood transfusion, such as allergic reaction, haemolytic crisis and transmission of pathogens are virtually absent. Furthermore, rEPO disappears completely from the blood stream within a week from the last injection but its effect on the red blood cells persists for months. The affair involving the Festina team during the 1998 Tour de France revealed a widespread abuse of EPO in cycling and led to a drastic change in the IOC's approach to doping in sports and ultimately to the creation of the World Anti-Doping Agency (WADA) in 1999. Subsequently, rEPO was included in the list of prohibited substances shortly after its commercial availability.

Table V. The official list of the Prohibited Substances (S2 group). WADA Prohibited List 2008 [119].

S2. HORMONES AND RELATED SUBSTANCES

The following substances and their releasing factors are prohibited:

1. Erythropoietin (EPO);
2. Growth Hormone (hGH), Insulin-like Growth Factors (e.g. IGF-1), Mechano Growth Factors (MGFs);
3. Gonadotrophins (e.g. LH, hCG), prohibited in males only;
4. Insulins;
5. Corticotrophins;

and other substances with similar chemical structure or similar biological effect(s).

Throughout the years the doping control program has produced various rumours and scandals related to rEPO abuse. The withdrawal of six Chinese female track-and-field athletes from the Sydney 2000 Olympics coincided with the introduction of the first EPO blood tests, fuelling suspicions of EPO abuse as the cause for

their previous exceptional performances. During the Salt Lake City 2002 Winter Olympics four athletes were stripped of their medals after being found positive for NESP. In 2006 the "Operación Puerto" investigation of a medical clinic in Spain, linked several elite cyclists with blood-doping methods, and uncovered multiple doses of EPO and bags of frozen blood for reinfusion [105]. Many examples of EPO abuse by elite athletes have been and are still being reported. The excessive use of EPO is associated with serious adverse side-effects, including hypertension, headaches, and an increased rate of thrombotic events as a result of an EPO-induced rise in the hematocrit and thickening of the blood [120]. In addition, EPO withdrawal could be implicated in neocytolysis, that is, the hemolysis of young red blood cells in the presence of increased hematocrit. Ultimately, EPO abuse could cause death [121].

Methods for detection of doping with EPO include the combination of direct and indirect approaches. The indirect methods incorporate changes of hematological parameters of erythropoiesis, such as haemoglobin, percentage of reticulocytes, and serum concentrations of EPO and soluble transferrin receptors [122, 123]. Some of these parameters could be altered up to 4 weeks after rEPO use, thereby increasing the chances of detection. In addition, such methods might offer the advantage of detecting other manipulations aside from the use of rEPO. Algorithms have been developed that are capable of detecting the use of rEPO either during the administration (ON-model: haemoglobin, EPO, and soluble transferring receptors) or after having discontinued administration (OFF-model: haemoglobin and reticulocytes) [124]. The former model detects up to 100% of rEPO-containing samples during the period of use, but has a short time window of detection (approximately 48 h) following injection of relatively low doses of EPO. In contrast, the OFF-model allows detection for up to 2 weeks after EPO withdrawal. The direct method, currently used in WADA-accredited antidoping laboratories is based on differences in the pattern and extent of glycosylation of rEPO as compared to the endogenously produced protein. The glycosylation pattern of rEPO preparations is determined by several factors, including the cell

line from which they are recombinantly expressed, the media employed for cell culturing, and the methods of protein purification [125]. The different arrangements of sugar residues found in rEPOs result in differences in their isoelectric points that are detected by a method combining isoelectric focusing (IEF) and double immunoblotting [103]. The situation is more complicated for the detection of EPO variants produced in human cells, such as epoetin delta.

During the Sydney 2000 Olympics, a combination of the direct (urine) and indirect (blood) methods for detection of EPO was implemented. However, since 2004, only the urine-based IEF method has been used for the detection of EPO abuse. Nevertheless, the indirect EPO methods will be integrated in the model foreseen for the creation of the hematological module of the athlete's biological passport (longitudinal follow-up over time of the athlete's biological parameters). These concerns have emphasised the need for alternative strategies that should either be complementary to the IEF method or eventually replace the approach. In this thesis, the structural elements contained in single IEF bands of rEPO standards are elucidated, in an attempt to better understand the underlying reasons for their differential IEF behaviour. This is a challenging task due to the important problems that arise from the analytical point of view. The main difficulties are:

- EPO concentration in human serum/plasma and urine are very low (usually in the fM range).
- Difficulties in EPO purification from biological matrixes.
- Limitations in the sensitivity of the analytical equipment for this purpose.
- rEPO and endogenous EPO differ only slightly in the composition of their glycoforms.

1.2.8 Recombinant erythropoietin and analogues: a challenge for doping control

In this section it is included a review published by our group on substances enhancing oxygen delivery to muscles and methods developed by doping control laboratories to evidence their misuse. The paper regarding those subjects was published in ***Therapeutic Drug Monitoring, 2004; 26(2): 175-179.***

Pascual JA, Belalcazar V, de Bolos C, Gutiérrez R, Llop E, Segura J.
Recombinant erythropoietin and analogues: a challenge for doping control.
Ther Drug Monit. 2004 Apr;26(2):175-9.

Bibliography

[1] B. Kuster, T. N. Krogh, E. Mortz and D. J. Harvey, *Glycosylation analysis of gel-separated proteins*. Proteomics *1,* 350-361 (2001).

[2] G. T. Cantin and J. R. Yates, III, *Strategies for shotgun identification of post-translational modifications by mass spectrometry*. J Chromatogr A *1053,* 7-14 (2004).

[3] R. Raman, S. Raguram, G. Venkataraman, J. C. Paulson and R. Sasisekharan, *Glycomics: an integrated systems approach to structure-function relationships of glycans*. Nat Methods *2,* 817-824 (2005).

[4] R. Peracaula, L. Royle, G. Tabares, G. Mallorqui-Fernandez, S. Barrabes, D. J. Harvey, R. A. Dwek, P. M. Rudd and R. de Llorens, *Glycosylation of human pancreatic ribonuclease: differences between normal and tumor states*. Glycobiology *13,* 227-244 (2003).

[5] R. Peracaula, G. Tabares, L. Royle, D. J. Harvey, R. A. Dwek, P. M. Rudd and R. de Llorens, *Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins*. Glycobiology *13,* 457-470 (2003).

[6] M. Taylor and K. Drickamer, *Introduction to Glycobiology* , pp 1-202. Ed: M. Taylor and K. Drickamer, Oxford University Press (2003).

[7] Z. Shriver, S. Raguram and R. Sasisekharan, *Glycomics: a pathway to a class of new and improved therapeutics*. Nat Rev Drug Discov *3,* 863-873 (2004).

[8] L. Schofield, M. C. Hewitt, K. Evans, M. A. Siomos and P. H. Seeberger, *Synthetic GPI as a candidate anti-toxic vaccine in a model of malaria*. Nature *418,* 785-789 (2002).

[9] I. C. Macdougall and K. U. Eckardt, *Novel strategies for stimulating erythropoiesis and potential new treatments for anaemia*. Lancet *368,* 947-953 (2006).

[10] M. Sundaram, Y. Qi, Z. Shriver, D. Liu, G. Zhao, G. Venkataraman, R. Langer and R. Sasisekharan, *Rational design of low-molecular weight heparins with improved in vivo activity*. Proc Natl Acad Sci USA *100,* 651-656 (2003).

[11] N. Jenkins, R. B. Parekh and D. C. James, *Getting the glycosylation right: implications for the biotechnology industry*. Nat Biotechnol *14,* 975-981 (1996).

[12] O. Barroso, I. Mazzoni and O. Rabin, *Hormone abuse in sports: the antidoping perspective*. Asian J Androl *10,* 391-402 (2008).

[13] P. B. Mills, K. Mills, A. W. Johnson, P. T. Clayton and B. G. Winchester, *Analysis by matrix assisted laser desorption/ionisation-time of flight mass spectrometry of the post-translational modifications of alpha 1-antitrypsin isoforms separated by two-dimensional polyacrylamide gel electrophoresis*. Proteomics *1,* 778-786 (2001).

[14] J. Jones, S. S. Krag and M. J. Betenbaugh, *Controlling N-linked glycan site occupancy*. Biochim Biophys Acta *1726,* 121-137 (2005).

[15] P. M. Rudd and R. A. Dwek, *Glycosylation: heterogeneity and the 3D structure of proteins*. Crit Rev Biochem Mol Biol *32,* 1-100 (1997).

[16] R. A. Laine, *A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems*. Glycobiology *4,* 759-767 (1994).

[17] K. T. Pilobello and L. K. Mahal, *Deciphering the glycocode: the complexity and analytical challenge of glycomics*. Curr Opin Chem Biol *11,* 300-305 (2007).

[18] Y. Satomi, Y. Shimonishi and T. Takao, *N-glycosylation at Asn(491) in the Asn-Xaa-Cys motif of human transferrin*. FEBS Lett *576,* 51-56 (2004).

[19] J. P. Kamerling, *Comprehensive glycoscience*; Introduction to glycoscience synthesis of carbohydrates, Volume 1, pp 31-33. Ed. J. P. Kamerling, Elsevier (2008).

[20] M. Maemura, A. Ohgaki, Y. Nakahara, H. Hojo and Y. Nakahara, *Solid-phase synthesis of core 8 O-glycan-linked MUC5AC glycopeptide*. Biosci Biotechnol Biochem *69,* 1575-1583 (2005).

[21] J. Hofsteenge, D. R. Muller, T. de Beer, A. Loffler, W. J. Richter and J. F. Vliegenthart, *New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us*. Biochemistry *33,* 13524-13530 (1994).

[22] R. G. Spiro, *Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds*. Glycobiology *12,* 43R-56R (2002).

[23] A. Furmanek and J. Hofsteenge, *Protein C-mannosylation: facts and questions*. Acta Biochim Pol *47,* 781-789 (2000).

[24] P. A. Haynes, *Phosphoglycosylation: a new structural class of glycosylation?* Glycobiology *8,* 1-5 (1998).

[25] D. P. Mehta, M. Ichikawa, P. V. Salimath, J. R. Etchison, R. Haak, A. Manzi and H. H. Freeze, *A lysosomal cysteine proteinase from Dictyostelium discoideum contains N-acetylglucosamine-1-phosphate bound to serine but not mannose-6-phosphate on N-linked oligosaccharides*. J Biol Chem *271,* 10897-10903 (1996).

[26] G. Srikrishna, L. Wang and H. H. Freeze, *Fucosebeta-1-P-Ser is a new type of glycosylation: using antibodies to identify a novel structure in Dictyostelium discoideum and study multiple types of fucosylation during growth and development*. Glycobiology *8,* 799-811 (1998).

[27] M. A. Ferguson, *The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research*. J Cell Sci *112,* 2799-2809 (1999).

[28] J. W. Hartog, A. A. Voors, S. J. Bakker, A. J. Smit and D. J. van Veldhuisen, *Advanced glycation end-products (AGEs) and heart failure: pathophysiology and clinical implications*. Eur J Heart Fail *9,* 1146-1155 (2007).

[29] N. Nass, B. Bartling, S. A. Navarrete, R. J. Scheubel, J. Borgermann, R. E. Silber and A. Simm, *Advanced glycation end products, diabetes and ageing*. Z Gerontol Geriatr *40,* 349-356 (2007).

[30] H. Unoki and S. Yamagishi, *Advanced glycation end products and insulin resistance*. Curr Pharm Des *14,* 987-989 (2008).

[31] M. L. DeMarco and R. J. Woods, *Structural glycobiology: a game of snakes and ladders*. Glycobiology *18,* 426-440 (2008).

[32] P. Kleinert, T. Kuster, D. Arnold, J. Jaeken, C. W. Heizmann and H. Troxler, *Effect of glycosylation on the protein pattern in 2-D-gel electrophoresis*. Proteomics *7,* 15-22 (2007).

[33] R. Rosenfeld, H. Bangio, G. J. Gerwig, R. Rosenberg, R. Aloni, Y. Cohen, Y. Amor, I. Plaschkes, J. P. Kamerling and R. B. Maya, *A lectin array-based methodology for the analysis of protein glycosylation*. J Biochem Biophys Methods *70,* 415-426 (2007).

[34] M. Nagano, G. Stubiger, M. Marchetti, G. Gmeiner, G. Allmaier and C. Reichel, *Detection of isoforms of recombinant human erythropoietin by various plant lectins after isoelectric focusing*. Electrophoresis *26,* 1633-1645 (2005).

[35] H. Geyer and R. Geyer, *Strategies for analysis of glycoprotein glycosylation*. Biochim Biophys Acta *1764,* 1853-1869 (2006).

[36] E. Balaguer, U. Demelbauer, M. Pelzing, V. Sanz-Nebot, J. Barbosa and C. Neususs, *Glycoform characterization of erythropoietin combining glycan and intact protein analysis by capillary electrophoresis - electrospray - time-of-flight mass spectrometry*. Electrophoresis *27,* 2638-2650 (2006).

[37] C. Neususs, U. Demelbauer and M. Pelzing, *Glycoform characterization of intact erythropoietin by capillary electrophoresis-electrospray-time of flight-mass spectrometry*. Electrophoresis *26,* 1442-1450 (2005).

[38] E. Gimenez, F. Benavente, J. Barbosa and V. Sanz-Nebot, *Towards a reliable molecular mass determination of intact glycoproteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*. Rapid Commun Mass Spectrom *21,* 2555-2563 (2007).

[39] D. S. Ashton, C. R. Beddell, D. J. Cooper, S. J. Craig, A. C. Lines, R. W. Oliver and M. A. Smith, *Mass spectrometry of the humanized monoclonal antibody CAMPATH 1H*. Anal Chem *67,* 835-842 (1995).

[40] S. F. Wheeler, P. M. Rudd, S. J. Davis, R. A. Dwek and D. J. Harvey, *Comparison of the N-linked glycans from soluble and GPI-anchored CD59 expressed in CHO cells*. Glycobiology *12,* 261-271 (2002).

[41] M. Schirm, I. C. Schoenhofen, S. M. Logan, K. C. Waldron and P. Thibault, *Identification of unusual bacterial glycosylation by tandem mass spectrometry analyses of intact proteins*. Anal Chem *77,* 7774-7782 (2005).

[42] K. Nagy, K. Vekey, T. Imre, K. Ludanyi, M. P. Barrow and P. J. Derrick, *Electrospray ionization fourier transform ion cyclotron resonance mass spectrometry of human alpha-1-acid glycoprotein*. Anal Chem *76,* 4998-5005 (2004).

[43] J. Reinders, U. Lewandrowski, J. Moebius, Y. Wagner and A. Sickmann, *Challenges in mass spectrometry-based proteomics*. Proteomics *4,* 3686-3703 (2004).

[44] T. M. Annesley, *Ion suppression in mass spectrometry*. Clin Chem *49,* 1041-1044 (2003).

[45] G. J. Rademaker, S. A. Pergantis, L. Blok-Tip, J. I. Langridge, A. Kleen and J. E. Thomas-Oates, *Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity*. Anal Biochem *257,* 149-160 (1998).

[46] L. Wells, K. Vosseller, R. N. Cole, J. M. Cronshaw, M. J. Matunis and G. W. Hart, *Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications*. Mol Cell Proteomics *1,* 791-804 (2002).

[47] J. Duus, C. H. Gotfredsen and K. Bock, *Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations*. Chem Rev *100,* 4589-4614 (2000).

[48] A. Dell and H. R. Morris, *Glycoprotein structure determination by mass spectrometry*. Science *291,* 2351-2356 (2001).

[49] P. M. Rudd, C. Colominas, L. Royle, N. Murphy, E. Hart, A. H. Merry, H. F. Hebestreit and R. A. Dwek, *A high-performance liquid chromatography based strategy for rapid, sensitive sequencing of N-linked oligosaccharide modifications to proteins in sodium dodecyl sulphate polyacrylamide electrophoresis gel bands*. Proteomics *1,* 285-294 (2001).

[50] K. R. Anumula, *Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates*. Anal Biochem *350,* 1-23 (2006).

[51] S. Prime and T. Merry, *Exoglycosidase sequencing of N-linked glycans by the reagent array analysis method (RAAM)*. Methods Mol Biol *76,* 53-69 (1998).

[52] J. P. Kamerling and G.J. Gerwig, *Structural analysis of naturally occurring sialic acids*. Glycobiology Protocols; Methods in molecular biology, vol 347, pp 69-91. I. Brockhausen, Humana Press Inc., Totowa, NJ (2006).

[53] C. W. der Lieth, A. Bohne-Lang, K. K. Lohmann and M. Frank, *Bioinformatics for glycomics: status, methods, requirements and perspectives*. Brief Bioinform *5,* 164-178 (2004).

[54] R. T. Lee, G. Lauc and Y. C. Lee, *Glycoproteomics: protein modifications for versatile functions. Meeting on glycoproteomics*. EMBO Rep *6,* 1018-1022 (2005).

[55] W. Jelkmann, *Molecular biology of erythropoietin*. Intern Med *43,* 649-659 (2004).

[56] J. Rossert, *Erythropoietin-induced, antibody-mediated pure red cell aplasia*. Eur J Clin Invest *35 Suppl 3,* 95-99 (2005).

[57] K. Maiese, F. Li and Z. Z. Chong, *New avenues of exploration for erythropoietin*. JAMA *293,* 90-95 (2005).

[58] K. Maiese, F. Li and Z. Z. Chong, *Erythropoietin in the brain: can the promise to protect be fulfilled?* Trends Pharmacol Sci *25,* 577-583 (2004).

[59] O. Livnah, E. A. Stura, S. A. Middleton, D. L. Johnson, L. K. Jolliffe and I. A. Wilson, *Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation*. Science *283,* 987-990 (1999).

[60] I. C. Macdougall, *Optimizing the use of erythropoietic agents-- pharmacokinetic and pharmacodynamic considerations*. Nephrol Dial Transplant *17 Suppl 5,* 66-70 (2002).

[61] W. Jelkmann, *The enigma of the metabolic fate of circulating erythropoietin (Epo) in view of the pharmacokinetics of the recombinant drugs rhEpo and NESP*. Eur J Haematol *69,* 265-274 (2002).

[62] J. R. Toledo, O. Sanchez, S. R. Montesino, G. Y. Fernandez, M. P. Rodriguez and J. A. Cremata, *Differential in vitro and in vivo glycosylation of human erythropoietin expressed in adenovirally transduced mouse mammary epithelial cells*. Biochim Biophys Acta *1726,* 48-56 (2005).

[63] J. C. Egrie and J. K. Browne, *Development and characterization of novel erythropoiesis stimulating protein (NESP)*. Br J Cancer *84 Suppl 1,* 3-10 (2001).

[64] A. J. Sytkowski, *Does erythropoietin have a dark side? Epo signaling and cancer cells*. Sci STKE *2007,* 38 (2007).

[65] V. Restelli, M. D. Wang, N. Huzel, M. Ethier, H. Perreault and M. Butler, *The effect of dissolved oxygen on the production and the glycosylation profile of recombinant human erythropoietin produced from CHO cells*. Biotechnol Bioeng *94,* 481-494 (2006).

[66] R. Montesino, J. R. Toledo, O. Sanchez, A. Sanchez, D. J. Harvey, L. Royle, R. A. Dwek, P. M. Rudd, G. J. Gerwig, J. P. Kamerling and J. A. Cremata, *Monosialylated biantennary N-glycoforms containing GalNAc-GlcNAc antennae predominate when human EPO is expressed in goat milk*. Arch Biochem Biophys *470,* 163-175 (2008).

[67] H. Schellekens, *Biosimilar epoetins: how similar are they?,* European Association of Hospital Pharmacists, Scientific Section 3*,* 243-247 (2004).

[68] J. P. Lewis, D. A. Alford, J. H. Rathjen, Jr. and R. D. Lange, *Preparation of human urine concentrate for erythropoietin studies*. J Lab Clin Med *66,* 987-993 (1965).

[69] T. Miyake, C. K. Kung and E. Goldwasser, *Purification of human erythropoietin*. J Biol Chem *252,* 5558-5564 (1977).

[70] P. M. Cotes and D. R. Bangham, *The international reference preparation of erythropoietin*. Bull World Health Organ *35,* 751-760 (1966).

[71] E. Goldwasser and C. K. Kung, *Purification of erythropoietin*. Proc Natl Acad Sci U S A *68,* 697-698 (1971).

[72] E. Goldwasser and C. K. Kung, *The molecular weight of sheep plasma erythropoietin*. J Biol Chem *247,* 5159-5160 (1972).

[73] J. L. Spivak, D. Small and M. D. Hollenberg, *Erythropoietin: isolation by affinity chromatography with lectin-agarose derivatives*. Proc Natl Acad Sci U S A *74,* 4633-4635 (1977).

[74] D. W. Briggs, J. W. Fisher and W. J. George, *Hepatic clearance of intact and desialylated erythropoietin*. Am J Physiol *227,* 1385-1388 (1974).

[75] A. J. Sytkowski, *Denaturation and renaturation of human erythropoietin*. Biochem Biophys Res Commun *96,* 143-149 (1980).

[76] R. Sasaki, S. Yanagawa and H. Chiba, *Isolation of erythropoietin by monoclonal antibody*. Biomed Biochim Acta *42,* 202-206 (1983).

[77] S. Yanagawa, K. Hirade, H. Ohnota, R. Sasaki, H. Chiba, M. Ueda and M. Goto, *Isolation of human erythropoietin with monoclonal antibodies*. J Biol Chem *259,* 2707-2710 (1984).

[78] S. Yanagawa, S. Yokoyama, K. Hirade, R. Sasaki, H. Chiba, M. Ueda and M. Goto, *Hybridomas for production of monoclonal antibodies to human erythropoietin*. Blood *64,* 357-364 (1984).

[79] C. W. Distelhorst, D. S. Wagner, E. Goldwasser and J. W. Adamson, *Autosomal Dominant familial erythrocytosis due to autonomous erythropoietin production*. Blood *58,* 1155-1158 (1981).

[80] F. K. Lin, S. Suggs, C. H. Lin, J. K. Browne, R. Smalling, J. C. Egrie, K. K. Chen, G. M. Fox, F. Martin and Z. Stabinsky, *Cloning and expression of the human erythropoietin gene*. Proc Natl Acad Sci USA *82,* 7580-7584 (1985).

[81] K. Jacobs, C. Shoemaker, R. Rudersdorf, S. D. Neill, R. J. Kaufman, A. Mufson, J. Seehra, S. S. Jones, R. Hewick, E. F. Fritsch, *Isolation and characterization of genomic and cDNA clones of human erythropoietin*. Nature *313,* 806-810 (1985).

[82] M. S. Dordal, F. F. Wang and E. Goldwasser, *The role of carbohydrate in erythropoietin action*. Endocrinology *116,* 2293-2299 (1985).

[83] P. H. Lai, R. Everett, F. F. Wang, T. Arakawa and E. Goldwasser, *Structural characterization of human erythropoietin*. J Biol Chem *261,* 3116-3121 (1986).

[84] A. J. Sytkowski, L. Feldman and D. J. Zurbuch, *Biological activity and structural stability of N-deglycosylated recombinant human erythropoietin*. Biochem Biophys Res Commun *176,* 698-704 (1991).

[85] H. Sasaki, B. Bothner, A. Dell and M. Fukuda, *Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA*. J Biol Chem *262,* 12059-12076 (1987).

[86] M. Takeuchi, S. Takasaki, H. Miyazaki, T. Kato, S. Hoshi, N. Kochibe and A. Kobata, *Comparative study of the asparagine-linked sugar chains of human erythropoietins purified from urine and the culture medium of recombinant Chinese hamster ovary cells*. J Biol Chem *263,* 3657-3663 (1988).

[87] E. Tsuda, M. Goto, A. Murakami, K. Akai, M. Ueda, G. Kawanishi, N. Takahashi, R. Sasaki, H. Chiba and H. Ishihara, *Comparative structural study of N-linked oligosaccharides of urinary and recombinant erythropoietins*. Biochemistry *27,* 5646-5654 (1988).

[88] J. P. Kamerling. *Carbohydrate features of recombinant human glycoproteins*. Biotecnología Aplicada *13,* 167-180 (1996).

[89] K. Yamaguchi, K. Akai, G. Kawanishi, M. Ueda, S. Masuda and R. Sasaki, *Effects of site-directed removal of N-glycosylation sites in human erythropoietin on its production and biological properties*. J Biol Chem *266,* 20434-20439 (1991).

[90] E. Delorme, T. Lorenzini, J. Giffin, F. Martin, F. Jacobsen, T. Boone and S. Elliott, *Role of glycosylation on the secretion and biological activity of erythropoietin*. Biochemistry *31,* 9871-9876 (1992).

[91] L. O. Narhi, T. Arakawa, K. H. Aoki, R. Elmore, M. F. Rohde, T. Boone and T. W. Strickland, *The effect of carbohydrate on the structure and stability of erythropoietin*. J Biol Chem *266,* 23022-23026 (1991).

[92] H. Rahbek-Nielsen, P. Roepstorff, H. Reischl, M. Wozny, H. Koll and A. Haselbeck, *Glycopeptide profiling of human urinary erythropoietin by matrix-assisted laser desorption/ionization mass spectrometry*. J Mass Spectrom *32,* 948-958 (1997).

[93] P. E. Groleau, P. Desharnais, L. Cote and C. Ayotte, *Low LC-MS/MS detection of glycopeptides released from pmol levels of recombinant erythropoietin using nanoflow HPLC-chip electrospray ionization*. J Mass Spectrom *43,* 924-935 (2008).

[94] V. Sanz-Nebot, F. Benavente, E. Gimenez and J. Barbosa, *Capillary electrophoresis and matrix-assisted laser desorption/ionization-time of flight-mass spectrometry for analysis of the novel erythropoiesis-stimulating protein (NESP)*. Electrophoresis *26,* 1451-1456 (2005).

[95] G. Stubiger, M. Marchetti, M. Nagano, C. Reichel, G. Gmeiner and G. Allmaier, *Characterisation of intact recombinant human erythropoietins applied in doping by means of planar gel electrophoretic techniques and matrix-assisted laser desorption/ionisation linear time-of-flight mass spectrometry*. Rapid Commun Mass Spectrom *19,* 728-742 (2005).

[96] G. Stubiger, M. Marchetti, M. Nagano, R. Grimm, G. Gmeiner, C. Reichel and G. Allmaier, *Characterization of N- and O-glycopeptides of recombinant human erythropoietins as potential biomarkers for doping analysis by means of microscale sample purification combined with MALDI-TOF and quadrupole IT/RTOF mass spectrometry*. J Sep Sci *28,* 1764-1778 (2005).

[97] V. Skibeli, G. Nissen-Lie and P. Torjesen, *Sugar profiling proves that human serum erythropoietin differs from recombinant human erythropoietin*. Blood *98,* 3626-3634 (2001).

[98] L. Wide, C. Bengtsson, B. Berglund and B. Ekblom, *Detection in blood and urine of recombinant erythropoietin administered to healthy men*. Med Sci Sports Exerc *27,* 1569-1576 (1995).

[99] A. Khan, J. Grinyer, S. T. Truong, E. J. Breen and N. H. Packer, *New urinary EPO drug testing method using two-dimensional gel electrophoresis*. Clin Chim Acta *358,* 119-130 (2005).

[100] M. Kohler, C. Ayotte, P. Desharnais, U. Flenker, S. Ludke, M. Thevis, E. Volker-Schanzer and W. Schanzer, *Discrimination of recombinant and endogenous urinary erythropoietin by calculating relative mobility values from SDS gels*. Int J Sports Med *29,* 1-6 (2008).

[101] V. Belalcazar, J. Segura, C. de Bolos, S. H. Peng, R. Gutiérrez-Gallego, E. Llop and J. Pascual. *Potential isoform discrimination during immunoaffinity purification of epoetin and darbepoetin.* Euromedlab, IFCC-FECSCC. 22nd European congress of clinical chemistry and laboratory medicine [Spain society of clinical biochemistry and molecular pathology]. Barcelona, Spain (2003).

[102] E. Llop, R. Gutiérrez-Gallego, V. Belalcazar, G. J. Gerwig, J. P. Kamerling, J. Segura and J. Pascual. *Structural features of EPO and analogues: glycosylation profiles*. Schanzer, W., Geyer, H., Gotzmann, A, and Mareck, U. Recent Advances in Doping Analysis. Proceedings of the Manfred Donike Workshop, *12*, 297-308 (2004).

[103] F. Lasne, L. Martin, N. Crepin and J. de Ceaurriz, *Detection of isoelectric profiles of erythropoietin in urine: differentiation of natural and administered recombinant hormones*. Anal Biochem *311,* 119-126 (2002).

[104] F. Serafini-Cessi, N. Malagolini and D. Cavallone, *Tamm-Horsfall glycoprotein: biology and clinical relevance*. Am J Kidney Dis *42,* 658-676 (2003).

[105] J. Mallorqui, J. Segura, C. de Bolos, R. Gutiérrez-Gallego and J. A. Pascual, *Recombinant erythropoietin found in seized blood bags from sportsmen*. Haematologica *93,* 313-314 (2008).

[106] J. C. Egrie, E. Dwyer, J. K. Browne, A. Hitz and M. A. Lykos, *Darbepoetin alfa has a longer circulating half-life and greater in vivo potency than recombinant human erythropoietin*. Exp Hematol *31,* 290-299 (2003).

[107] I. C. Macdougall, *CERA (Continuous Erythropoietin Receptor Activator): a new erythropoiesis-stimulating agent for the treatment of anemia*. Curr Hematol Rep *4,* 436-440 (2005).

[108] V. Belalcazar, R. Ventura, J. Segura and J. A. Pascual, *Clarification on the detection of epoetin delta and epoetin omega using isoelectric focusing*. Am J Hematol (2008).

[109] S. Fishbane, A. Pannier, X. Liogier, P. Jordan, F. C. Dougherty and B. Reigner, *Pharmacokinetic and pharmacodynamic properties of methoxy polyethylene glycol-epoetin beta are unaffected by the site of subcutaneous administration*. J Clin Pharmacol *47,* 1390-1397 (2007).

[110] 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. Kent and J. A. Bradburne, *Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein*. Science *299,* 884-887 (2003).

[111] K. Schriebl, E. Trummer, C. Lattenmayer, R. Weik, R. Kunert, D. Muller, H. Katinger and K. Vorauer-Uhl, *Biochemical characterization of rhEpo-Fc fusion protein expressed in CHO cells*. Protein Expr Purif *49,* 265-275 (2006).

[112] H. F. Bunn, *New agents that stimulate erythropoiesis*. Blood *109,* 868-873 (2007).

[113] C. Rinsch, P. Dupraz, B. L. Schneider, N. Deglon, P. H. Maxwell, P. J. Ratcliffe and P. Aebischer, *Delivery of erythropoietin by encapsulated myoblasts in a genetic model of severe anemia*. Kidney Int *62,* 1395-1401 (2002).

[114] K. Binley, Z. Askham, S. Iqball, H. Spearman, L. Martin, M. de Alwis, A. J. Thrasher, R. R. Ali, P. H. Maxwell, S. Kingsman and S. Naylor, *Long-term reversal of chronic anemia using a hypoxia-regulated erythropoietin gene therapy*. Blood *100,* 2406-2413 (2002).

[115] J. L. Nordstrom, *The antiprogestin-dependent GeneSwitch system for regulated gene therapy*. Steroids *68,* 1085-1094 (2003).

[116] K. Takacs, R. C. Du, S. Nabarro, N. Dillon, J. H. McVey, Z. Webster, A. Macneil, I. Bartok, C. Higgins, D. Gray, M. Merkenschlager and A. G. Fisher, *The regulated long-term delivery of therapeutic proteins by using antigen-specific B lymphocytes*. Proc Natl Acad Sci USA *101,* 16298-16303 (2004).

[117] A. Mikhail, A. Covic and D. Goldsmith, *Stimulating Erythropoiesis: Future Perspectives*. Kidney Blood Press Res *31,* 234-246 (2008).

[118] G. Lippi and G. Guidi, *Laboratory screening for erythropoietin abuse in sport: an emerging challenge*. Clin Chem Lab Med *38,* 13-19 (2000).

[119] www.wada-ama.org/rtecontent/document/2008_List_En.pdf. *The official list of the Prohibited Substances (S2 group)*. WADA Prohibited List (2008).

[120] F. Locatelli and L. Del Vecchio, *Pure red cell aplasia secondary to treatment with erythropoietin*. J Nephrol *16,* 461-466 (2003).

[121] T. R. Lappin, A. P. Maxwell and P. G. Johnston, *EPO's alter ego: erythropoietin has multiple actions*. Stem Cells *20,* 485-492 (2002).

[122] R. Parisotto, C. J. Gore, K. R. Emslie, M. J. Ashenden, C. Brugnara, C. Howe, D. T. Martin, G. J. Trout and A. G. Hahn, *A novel method utilising markers of altered erythropoiesis for the detection of recombinant human erythropoietin abuse in athletes*. Haematologica *85,* 564-572 (2000).

[123] J. R. Delanghe, M. Bollen and M. Beullens, *Testing for recombinant erythropoietin*. Am J Hematol *83,* 237-241 (2008).

[124] J. Segura, J. A. Pascual and R. Gutiérrez-Gallego, *Procedures for monitoring recombinant erythropoietin and analogues in doping control*. Anal Bioanal Chem *388,* 1521-1529 (2007).

[125] W. Jelkmann, *Control of erythropoietin gene expression and its use in medicine*. Methods Enzymol *435,* 179-197 (2007).

2. Objectives

The inception of the present research work resides in the differences observed when endogenous erythropoietin and its recombinant analogues are separated in an isoelectic focusing gel, according to their pI values, and the lack of understanding on the origin of the distinct IEF profiles. Recombinant erythropoietins are produced in mammalian cells (CHO and BHK) transfected with the authentic human EPO gene, resulting in the same polypeptide chain as the endogenous counterpart. The chronic use of EPO for some diseases and the development of some health problems, including autoimmune responses and enhancement of tumor progression, render these subtle differences very interesting. Ultimately, the study of differences between endogenous and exogenous EPO, is of utmost importance for the development of an alternative method for detection of rEPO misuse in athletes. The working hypothesis was that the observed differences are due to post-translational modifications, most probably, glycosylation.

The general objective of this thesis was the structural characterisation of N-glycans from recombinant erythropoietins (epoetin α/β) and its analogue (darbepoetin α) establishing the structural features responsible for the differences in isoelectric properties (isoforms distribution), with respect to endogenous erythropoietin. Once glycan profiles from these samples were obtained, they were also compared to glycans contained in a novel rEPO (epoetin δ) that is engineered in the human fibrosarcoma cell line HT-1080 by homologous recombination and “gene activation”.

This general objective was divided into the following specific objectives:

- Development of an appropriate separation method for glycoproteins (2-DE procedure) that allows releasing intact N-glycans and subsequently peptide mapping.
- Optimisation of glycoproteomic techniques for detailed structural analysis of N-glycans released from intact glycoproteins (total pool)

and from 2-DE gel bands (individual isoforms) using MS and HPLC combined with exoglycosidase array digestions.

- Improvement of methodologies for sialic acids analysis (rEPO Std, pharmaceutical preparations and biological samples)

This research project was performed within the framework of the following projects funded by the World Anti-Doping Agency (WADA):

- "Unequivocal confirmation of recombinant erythropoietin (rEPO) in human urine through structural evidences of specific glycosylation"

- "Rapid screening (and confirmatory) method for rhEPO and NESP based on immunorecognition of its exogenous *N*-glycolylneuraminic acid content"

3. Results

Chapter 3.1

Erythropoietin peptide mapping.

3.1.1 Introduction

Recombinant human erythropoietins (rEPO α/β and rEPO δ) and darbepoetin alfa (NESP) are glycoprotein pharmaceuticals for the treatment of anemia in humans by stimulating erythrocyte production [1]. However these agents are misused in the context of human and equine sports due to their potential to enhance performance [2]. In both contexts, the development of methods that allow the characterisation, identification and differentiation between rEPOs, NESP, and endogenous EPO is very important. In the framework of therapeutic proteins, because of a better understanding of the structure-functional relationship of different erythropoietins with their characteristic glycosylation has facilitated the development and approval of several second- and third-generation products, engineered specifically through alteration of their carbohydrate content to improve the product's plasma half-life [3]. In the last two years the reliability of the exogenous EPO-detection methods has been questioned by several authors [4-7]. The most worrisome problem is that the method is based on the IEF profile of EPO samples [8] and this test is not always easily interpreted [9]. For this reason, the development of new complementary techniques based on structural differences, preferably using mass spectrometry, is necessary [10]. Darbepoetin alfa (NESP) is a hyperglycosylated EPO analogue designed for prolonged survival in the circulation and thus extended biological activity [11]. Like epoetin alfa and beta, darbepoetin alfa is produced in CHO cells. However, the amino acid sequence of NESP differs from that of endogenous and recombinant EPO at five positions, allowing incorporation of two additional N-linked oligosaccharides to the asparagines residues at position 30 and 88, without alteration of the conformation. Differentiation between rEPOs and NESP can be based on the difference in amino acid sequence A30N, H32T, P87V, W88N, and P90T, and mass spectrometry is the technique of choice for specific identification of these differences [12]. Epoetin delta, also called gene-activated erythropoietin or Dynepo^{TM}, is produced in a human fibrosarcoma cell line HT-1080 by homologous recombination. These cells are genetically engineered to transcribe

and translate the dormant native human EPO gene under the control of an active cytomegaly virus (CMV) promoter, for which the peptide sequence is supposed to be identical to that of endogenous EPO [13]. As glycosylation is cell-dependent, minor differences are expected when epoetin delta and endogenous EPO are compared.

Although it was supposed identical polypeptide chains for different rEPOs and endogenous EPO, the peptide mapping is necessary in order to assess the overall peptide sequence, corroborate the assumption, and facilitate decision whether or not to investigate glycan structure in more detail. The basic principle of mass fingerprinting is the comparison of the measured peptide masses with calculated peptide masses from database entries. Every protein results in a unique set of peptide masses after cleavage with a specific protease. Depending on the mass accurancy and mass resolution of the instrument, only a few peptide masses are required for reliable protein identification [14]. Our strategy for investigating site-specific glycosylation, combines the glycan-specific endoglycosidase PNGase F with trypsin proteolysis. Deglycosylation with PNGase F releases intact N-linked glycans and converts the asparagines residues of the glycoprotein or glycopeptide into aspartic acid. This change in mass can be detected by MS corroborating the existence of N-glycan at that residue. The proportion of the amino acid sequence detected after proteolysis alone varies depending on the mass spectrometric technique used, and a complete sequence has never been observed without deglycosylation. However, when rEPO was deglycosylated prior to proteolysis there was an increase in sequence coverage and all three N-linked glycosylation sites could be monitored for occupancy [15]. In the present study is demonstrated indeed no differences exist between polypeptide chains between different rEPOs, and that peptide mapping allows O-glycopeptide analysis evidencing the discriminatory power of O-glycoforms. The final aim of this work is describing and easy mass spectrometric approach for discriminate between rEPOs and its structural analogue NESP. It is based on fingerprint of the glycoproteins detecting the two

differential tryptic peptides to provide a powerful tool in the fight against blood doping with rEPO and NESP in the horse racing industry [10].

3.1.2 Materials and methods

Materials

Recombinant human EPO (rEPO produced in CHO cells) was obtained from European Pharmacopoeia Commission (rEPO α/β), epoetin delta (rEPO produced in human fibrosarcoma HT-1080 cells) was obtained from Shire Pharmaceuticals (rEPO δ, Dynepo) and NESP (darbepoetin α) was acquired from Amgen. Recombinant peptide-N^4-(acetyl-β-glucosaminyl)-asparagine amidase F (PNGase F, EC 3.1.27.5) was purchased from Roche. Calbiosorb™ and endoproteinase Glu-C (EC 3.4.21.19) were from Calbiochem. Dithiothreitol (DTT) was from Amersham Biosciences. Trypsin, bovine serum albumin (BSA), bovine fetuin, phosphate-buffered saline (PBS), iodoacetamide (IAA), EDTA and 3.5-dimethoxy-4-hydroxycinnamic acid were purchased from Sigma. α-Cyano-4-hydroxycinnamic acid was from Fluka. Poros 20 R2 resin was from Applied Biosystems. GELoader Tips were purchased from Eppendorf. All other chemicals were of highest purity commercially available.

De-N-glycosylation

In order to release N-linked oligosaccharides, erythropoietin samples were directly dissolved in 50 mM sodium phosphate buffer (pH: 7.3). Then 1 IU of PNGase F was added and the mixture incubated 16 h at 37 ºC [16]. As this protocol resulted in incomplete deglycosylation of NESP, two different procedures for the reduction and alkylation of disulfide bonds were explored in order to facilitate complete release of the N-glycans.

-1- Dried glycoprotein was dissolved in 200 µl of denaturing buffer containing 50 mM Tris-HCl (pH 7.4), 50 mM EDTA; 3% (w/v) SDS, 1% 2-mercaptoethanol. The solution was mixed, boiled for 4 minutes at 100 ºC and cooled to room temperature. Decanoyl-*N*-methyl glutamide (MEGA-10) 6% (w/v) was added to

the mixture, vortexed and allowed to stand for 30 min. Subsequently, PNGase F (3 IU) was added, and the solution incubated for 18 h at 37 ºC. Samples were boiled for 2 min, and a fresh batch of PNGase F (2 IU) was added. The incubation was continued for 6 h at 37 ºC. Finally, sample was boiled for 2 min. To neutralise the detergent present in the mixture, a proportional amount of Calbiosorb^{TM} was added to each sample and then they were incubated for 10 min at room temperature with occasional gentle agitation. Then, the resin was settled, and the supernatant free of detergent used for purification on SPE columns.

-2- Glycoproteins were reduced with 50 μl of 10 mM DTT in 50 mM phosphate buffer (pH 7.3) for 30 min at 56 ºC. Following DTT treatment, proteins were alkylated with 50 µl of 50 mM IAA in 50 mM phosphate buffer (pH 7.3) for 30 min in the dark. Then, samples were centrifuged with 10 kDa filter 10 min at 6900 g, sonicated for 3 min and the retentate dissolved in 150 µl 50 mM sodium phosphate buffer (pH 7.3). PNGase F (1 IU) was added and the mixture incubated for 16 h at 37 ºC.

Protein purification prior to trypsin digestion

Both, de-N-glycosylated proteins and peptides after proteolysis were desalted and separated from the N-glycans by means of micro-column purification; 0.5 µl reversed-phase (Poros 20 R2) in acetonitrile (ACN) was packed in a long, narrow pipette tip (GELoader, eppendorf) [17]. Prior to use, a washing step was performed with 10 µl ACN, followed by an equilibration step with 10 µl 0.1% TFA. The protein sample was acidified with acqueous FA 5%, loaded on the column and slowly passed through the medium. After loading the sample, the flow-through fraction containing the N-glycans was collected. Then, a washing step was performed with 10 µl of 0.1% TFA and mixed with non-retained fraction. The column was run completely dry by pressing air through it for a few seconds. Finally a 10 µl volume of ACN:H_2O:TFA (80:20:0.1 v/v/v) was loaded, slowly passed through the column and the protein eluted. Next, samples were dried in a vacuum centrifuge for tryptic and/or endoproteinase digestion.

Proteolysis

Native and de-N-glycosylated EPO and NESP samples (aliquots of 1 μg) were digested in 100 mM ammonium bicarbonate (pH 7.9) containing 20 ng of trypsin (ratio Enzyme:Substrate 1:50). Incubations were performed at 37 ºC for 16 h. Peptide mixtures generated after proteolysis were again desalted and cleaned by Poros R2 micro-column. A 1μl-aliquot was separated for MALDI-TOF MS analysis and the rest was lyophilised for UPLC-ESI-TOF MS experiments.

Mass spectrometric analysis

MALDI-TOF MS

Samples (proteins, peptides) were dissolved in water at varying concentrations. An aliquot was mixed with the corresponding matrix and <1 μl of this preparation applied to the polished stainless steel target and allowed to dry at room temperature. Sinapinic acid (10 mg/ml in ACN:H_2O:TFA, 50:50:0.1 v/v/v) was chosen for protein analysis and a solution of α-cyano-4-hydroxycinnamic acid (20 mg/ml in ACN:H_2O:TFA, 70:30:0.1 v/v/v) was used for peptide samples. Experiments were carried out on a Voyager-DE™ STR Biospectrometry workstation (Applied Biosystems), equipped with a N_2 laser (337 nm). Two different analytical strategies were applied in which samples were measured both in linear and reflectron mode. The former, being more sensitive but with less resolution for glycoprotein analyses, provided information on the mean molecular mass. The latter method was used for identification of peptide fingerprinting based on precise mass measurements. The array detector was set to the high resolution position and mass scans were accumulated in the mass range (between 550-3500 Da for peptides and between 10000 and 50000 Da for proteins depending on the sample). For data acquisition, the laser intensity was 1800-2000 and the laser beam was moved manually over the sample in order to compensate for sample depletion under the laser beam. External calibration of the spectrometer was performed using albumin for glycoprotein analyses and in case of peptide spectra, calibration was carried out with Sequazyme™ Peptide

Mass Standards Kit of the desired range (PerSeptive Biosystems). Recorded data were processed with Data Explorer™ Software (Applied Biosystems).

UPLC-ESI-TOF MS

Peptide analyses were performed using an Acquity™ Ultra Performance Liquid Chromatograph (WATERS) coupled to a LCT premier XE ESI-TOF instrument (Micromass). Trypsinised samples were dissolved in 100 µl of H_2O:ACN:FA (50:50:0.1 v/v/v) and 2 µl injected on a Acquity UPLC BEH C_{18} column (2.1*100 mm, 1.7 µm). Peptides were eluted with a linear gradient from 96% A [H_2O: ACN:FA (95:5:0.1 v/v/v)] to 38% A over 20 min, then the running buffer was changed to 95% B [H_2O: ACN:FA (5:95:0.1 v/v/v)] in 0.5 min. This condition was maintained for 1.5 min, then the buffer was returned to the initial conditions in 0.5 min, stabilised for 1 min before next injection. The flow rate was 0.2 ml/min. Mass spectra were acquired in both positive and negative mode over a range of *m/z* 50-1500. The capillary voltage was set at 3000 V (positive) and 2800 V (negative), the desolvation temperature was 350 ºC and the desolvation gas-flow set to 350-400 L/h. For all analyses the TOF tube voltage was kept at 7200 (reflectron at 1800 V) with a pusher setting of 900 (pusher offset at 0.93). Recorded data were processes using MassLynx version 4.1 (Waters Inc.) The calibration of the mass spectrometer was performed using Leucine-enkephalin (*m/z* 556 for positive and *m/z* 554 for negative mode).

3.1.3 Results and discussion

Optimisation of de-N-glycosylation protocol

In order to verify the purity of all preparations and compare samples at level of average molecular mass and glycosylation content, MALDI-TOF MS analyses were performed, both before and after de-N-glycosylation (Figure 1). The molecular weights of rEPOs were determined by MALDI-TOF MS, taking into account that the laser intensity value influenced the mass centroid of the measured *m/z* values [18]. Using optimised laser conditions good detection

sensitivity was achieved and in-source fragmentation could be minimised, rendering an average molecular mass of 29.397 Da for rEPO α/β and 29.753 for rEPO δ. NESP was analysed following the same experimental procedure as for rEPOs. As NESP bears two additional carbohydrate moieties, the molecular weight obtained for the intact glycoprotein was 36.690 which is slightly smaller than the theoretical 37.100 Da given by Egrie [19]. Furthermore, as NESP presented higher carbohydrate content than rEPOs, the influence of laser intensity on its fragmentation during the MALDI process could be more pronounced. Due to the high heterogeneity of sugar moieties, values given should be only taken as approximate average masses. When the three glycoproteins were compared in terms of the broadening of the obtained unresolved peaks, NESP showed the higher heterogeneity, followed by rEPO α/β and surprisingly rEPO δ yielded a narrower peak. The higher peak-width in NESP can not be directly compared to that showed by rEPOs because the former contains two additional N-glycans with their respective variability. However, differences between rEPOs (both having the same site occupancy) may indicate lower glycan heterogeneity. Interestingly, this phenomenon also occurred when the two rEPOs are separated in a SDS-PAGE gel in which the rEPO δ appeared as a much sharper band, suggesting more homogeneous structure than rEPO α/β [20].

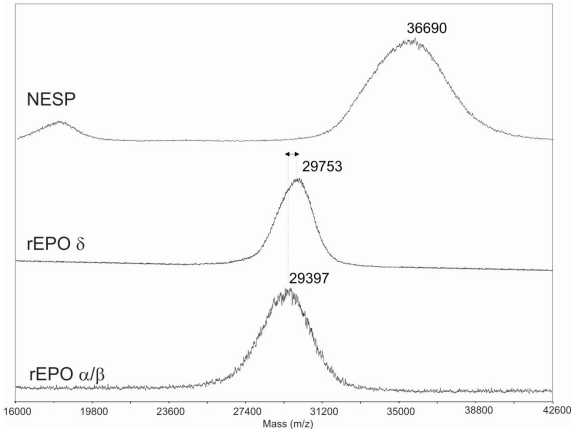


Figure 1. MALDI-TOF MS positive ion spectra of NESP (top), rEPO δ (middle) and rEPO α/β (bottom).

To achieve complete peptide mapping and addressing the carbohydrate content, representative of all N-glycans, a complete enzymatic de-N-glycosylation was required. Furthermore, the complete removal of all N-glycans from the rEPOs led to molecules containing only one O-glycan allowing the study of this type of glycosylation. During our investigations we observed different kinetics of de-N-glycosylation process of rEPOs and NESP. From the data observed in Figure 2 it was concluded that the simplified protocol consisting of the sample incubation with PNGase F in 50 mM sodium phosphate buffer pH 7.3, results in a complete de-glycosylation of rEPO α/β and rEPO δ. However, such straight-forward de-glycosylation protocol was not completely efficient for NESP (Figure 2A) possibly because the introduction of the two additional N-glycans render one of the five N-glycans less accessible for the enzyme as evidenced by the observation of a peak at ~22 kDa. A mass difference of ~3 kDa compared with the fully de-N-glycosylated protein indicated that one N-glycan structure, most likely tetraantennary, was still attached to the polypeptide chain. Even after an extended incubation time (72 h) and excess enzyme the 22 kDa peak persisted (Figure 2B). A steric hindrance for PNGase F is supposed to be responsible for this phenomenon in native structure. In order to solve this problem, digestions were performed in a denaturing buffer containing SDS. Although this protocol permitted an efficient digestion, the addition of detergent suppresses ionisation and contributes to the chemical noise [21]. Posterior purification steps to remove detergent prior the analyses, were accompanied by significant loss of material. This drawback was overcome by a denaturing pre-treatment of the glycoprotein with DTT followed by blocking of the resulting sulphidryl groups with IAA. Total NESP de-N-glycosylation was achieved with better protein recovery using the latter protocol. Posterior analyses were conducted using N-glycans released by the latter procedure described here. This phenomenon was previously described by Stübiger et *al.* [22] who hypothesised that sterical features of the native glycoprotein, particularly disulfide bonding, may be responsible of PNGase F resistant N-glycosylation site. Given that the resistant asparagine residue could

actually not be elucidated on the intact protein, further peptide analyses are needed to solve this problem. The mass spectra from the fully de-N-glycosylated rEPO samples yielded values coherent with O-glycosylated molecules were the O-glycosylation corresponded to the T, sialyl-T, disialyl-T antigen (Figure 2C, D, E for NESP, rEPO δ and rEPO α/β respectively). From the mass values obtained before and after de-N-glycosylation an estimation of the carbohydrate content could be made; ~38% for rEPOs, ~49% for NESP. Signals at *m/z* 18.249 revealed that a small proportion of non-O-glycosylated glycoproteins exist in all three samples. Concerning the C-terminus of rEPOs a comparison between theoretical and observed masses permitted concluding that the sequence of mature glycoprotein consist of 165 rather than 166 aa predicted from cDNA sequence lacking the C-terminal Arg residue (Des-Arg 166) [23].

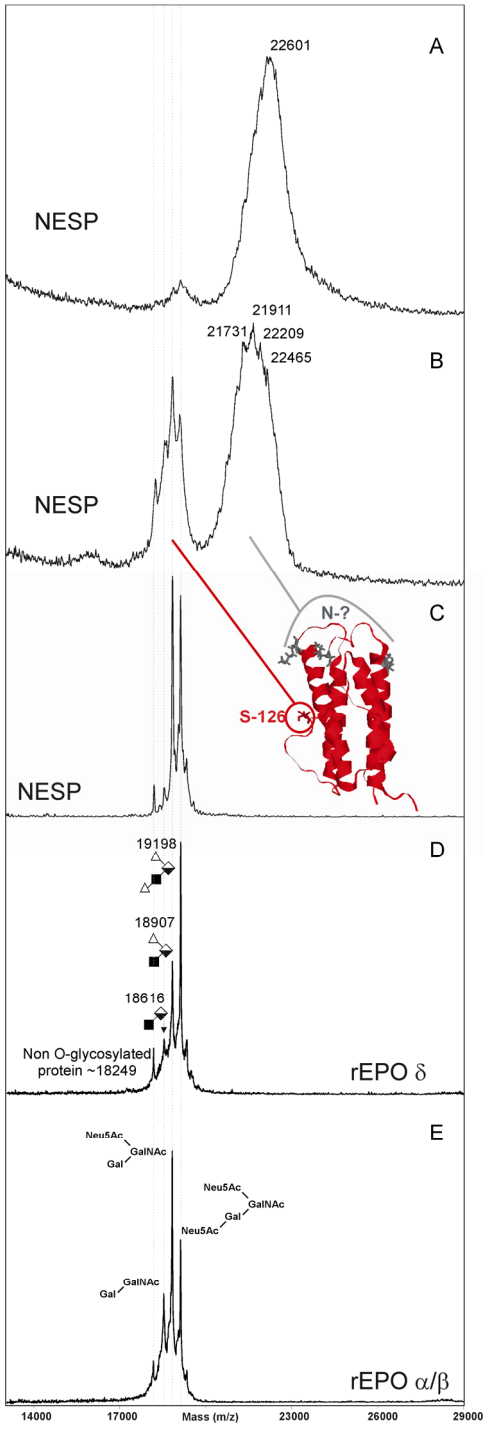


Figure 2. MALDI-TOF MS positive mass spectrum of partially de-N-glycosylated NESP after PNGase F digestion (A: 16 h); (B: 72 h). C, D, E: Complete de-N-glycosylated NESP (after reduction and alkylation), rEPO δ and rEPO α/β respectively.

Peptide mapping

Although rEPO and NESP are analogous, there is a 3% difference in their amino acid sequence. The five variant amino acids in NESP are located in the two regions that relate to two tryptic peptides (21-45) and (77-97). The first corresponds to sequence EAENITTGCAEHCSLNENITVPDTK in rEPOs and EAENITTGCNETCSLNENITVPDTK in NESP; in both a disulfide bond between Cys29 and Cys33 is present. The second corresponds to sequence GQALLVNSSQPWEPLQLHVDK in rEPOs and GQALLVNSSQVNETLQLHVDK in NESP. Theoretically, these two tryptic peptides could be employed for differentiation between rEPO and NESP but, they are glycopeptides with N-glycans at Asn24, Asn38 and Asn83 in rEPOs and two additonal glycans in NESP (located at Asn30 and Asn88) which poorly ionise under positive ionisation conditions, and thus, cannot be detected in a digestion mixture. Tryptic digestions of the three glycoprotein rEPO α/β, rEPO δ and NESP were performed in order to elicit their peptide and glycopeptide fingerprint. Peptide mapping of N-glycosylated samples resulted in the observation of peptides only and absence of the glycopeptides containing the differences. These analyses showed identical peptides for rEPOs and NESP and differentiation between both could not be achieved (data not shown). To overcome this problem and given that only (O-glyco)-peptides could be detected in the mass spectrum, de-N-glycosylation prior to protease digestions was required. De-N-glycosylated samples were digested with trypsin obtaining less complex mixtures that permitted greater sensitivity and total peptide mapping. Peptide analyses were performed by both, MALDI-TOF MS and UPLC-ESI-TOF MS.

In order to address the partial de-N-glycosylation problem described above for NESP, peptide analyses were carried out first, after de-N-glycosylation but without reduction and alkylation before PNGase F digestion and second after total de-N-glycosylated protein. In the first case, peptide mapping showed peptide 77-97 but peptide 21-45 was not observed (Figure 3 top). Conversely, peptide mass spectrum from reduced, alkylated and de-N-glycosylated NESP yielded

complete sequence coverage (Figure 3 bottom), indicating that the Asn resistant to PNGase F digestion was located in that peptide. However, this glycopeptide contains 3 N-glycosylation sites and the specific position could not be identified. From the 3 possible N-glycosylation sites, two (Asn24 and Asn38) are also present in rEPOs which do not showed partial de-N-glycosylation and the additional site in Asn30 is located between Cys29 and Cys33 that are connected by a disulfide bound forming a small loop in the 3D structure of the native molecule. For these reasons it is speculated that the glycosylation site responsible of partial de-N-glycosylation in NESP is Asn30, but more investigations should be done to corroborate this hypothesis.

Further peptide mapping analyses were performed following reduction and alkylation of NESP before de-N-glycosylation in order to obtain higher coverage and to be able to analyse the two differential peptides between NESP and rEPOs. Conversely, in these experiments, neither reduction of disulfide bonds nor alkylation of cysteine residues from rEPO samples were performed prior PNGase F digestion.

LC-ESI-TOF-MS

Peptides from rEPOs and NESP were analysed by MALDI-TOF MS and UPLC-ESI-TOF MS. A search against the Mascot database showed a sequence coverage higher than 90% in both techniques. Here, only results from UPLC-ESI TOF analyses are presented. Since the amino acid sequences of EPO and NESP are known, simulation of tryptic digestion of the proteins was conducted, and the resulting proteolytic peptides are summarised in Table I. As the spectral range was limited to 1500 Da in the TOF detector, in these analyses double and triply charged ions were used for peptide identification.

Table I. Observed peptides for rEPO and NESP using LC-ESI-TOF measurements. CAM indicates carbamidomethylated residues and MC is miss cleavages.

	Mass	$[M+2H]^{2+}$	$[M+3H]^{3+}$	Frag.	Modifcation	Peptide Sequence
T1	440.2616			1-4		APPR
T2	1184.6204			1-10	MC(1)/Cys-CAM	APPRLICDSR
T2	706.3552	353.6815		5-10		LICDSR
T2	763.3767			5-10	Cys-CAM	LICDSR
T3	516.3140	258.6609		11-14		VLER
T4	736.4239	368.7159		15-20		YLLEAK
E5	2689.1614	1345.1076	897.0745	21-45	Cys S-S Cys	EAEDITTGCAEHCSLNEDITVPDTK
E5	2805.2189			21-45	Cys-CAM	EAEDITTGCAEHCSLNEDITVPDTK
N5	2699.1546			21-45	Cys S-S Cys	EAEDITTGCDETCSLNEDITVPDTK
N5	2813.1975	1407.0987	938.4040	21-45	Cys-CAM	EAEDITTGCDETCSLNEDITVPDTK
T6	927.4723	464.2401	309.8293	46-52		VNFYAWK
T6	1083.5734	542.2903		46-53		VNFYAWKR
T7	175.1189			53-53		R
T8	2682.4395	1341.7233	894.8180	53-76	MC(1)	RMEVGQQAVEVWQGLALLSEAVLR
T8	2526.3384	1263.6726	842.7847	54-76		MEVGQQAVEVWQGLALLSEAVLR
E9	2360.2244	1180.6156	787.4126	77-97		GQALLVDSSQPWEPLQLHVDK
N9	2295.1826	1148.0949	765.7324	77-97		GQALLVDSSQVDETLQLHVDK
N9	2878.5268	1349.7670	960.1804	77-103	MC (1)	GQALLVDSSQVDETLQLHVDKAVSGLR
T10	602.3620	301.6849		98-103		AVSGLR
T11	803.4985	402.2532		104-110		SLTTLLR
T12	587.3511	294.1795		111-116		ALGAQK
T13	2034.0978			111-131	MC(1)	ALGAQKEAISPPDAASAAPLR
T13	1465.7645	733.3862	489.2601	117-131		EAISPPDAASAAPLR
T13	1830.8967			117-131	Hex-HexNAc	EAISPPDAASAAPLR
T14	924.4785	462.7432	308.8314	132-139		TITADTFR
T14	1052.5735			132-140	MC(1)	TITADTFRK
T15	147.1128			140-140		K
T16	435.2714			141-143		LFR
T16	563.3664			140-143	MC(1)	KLFR
T17	898.4781	449.7427	300.1642	144-150		VYSNFLRGK
T17	1083.5945			144-152	MC(1)	VYSNFLRGKGK
T18	204.1342			151-152		GK
T19	260.1968			153-154		LK
T20	912.4244	456.7158	304.8130	155-162		LYTGEACR
T20	969.4458	485.2265	323.8201	155-162	Cys-CAM	LYTGEACR
T20	1185.5204			155-165	MC(1)	LYTGEACRTGD
T20	1242.5419			155-165	MC(1)/Cys-CAM	LYTGEACRTGD
T21	292.1139			163-165		TGD
T2-T20	1615.7651	808.3862	539.2599	5-10 + 155-162		LICDSR--LYTGEACR

Peptides of the three glycoproteins were obtained from trypsinolysis after de-N-glycoslation and allowed sequence coverage higher than 97% as depicted in figure 3 and 4 for rEPOs and NESP, respectively. As rEPO α/β and rEPO δ

displayed identical peptides, mass spectra showed must be consider representative of both rEPOs.

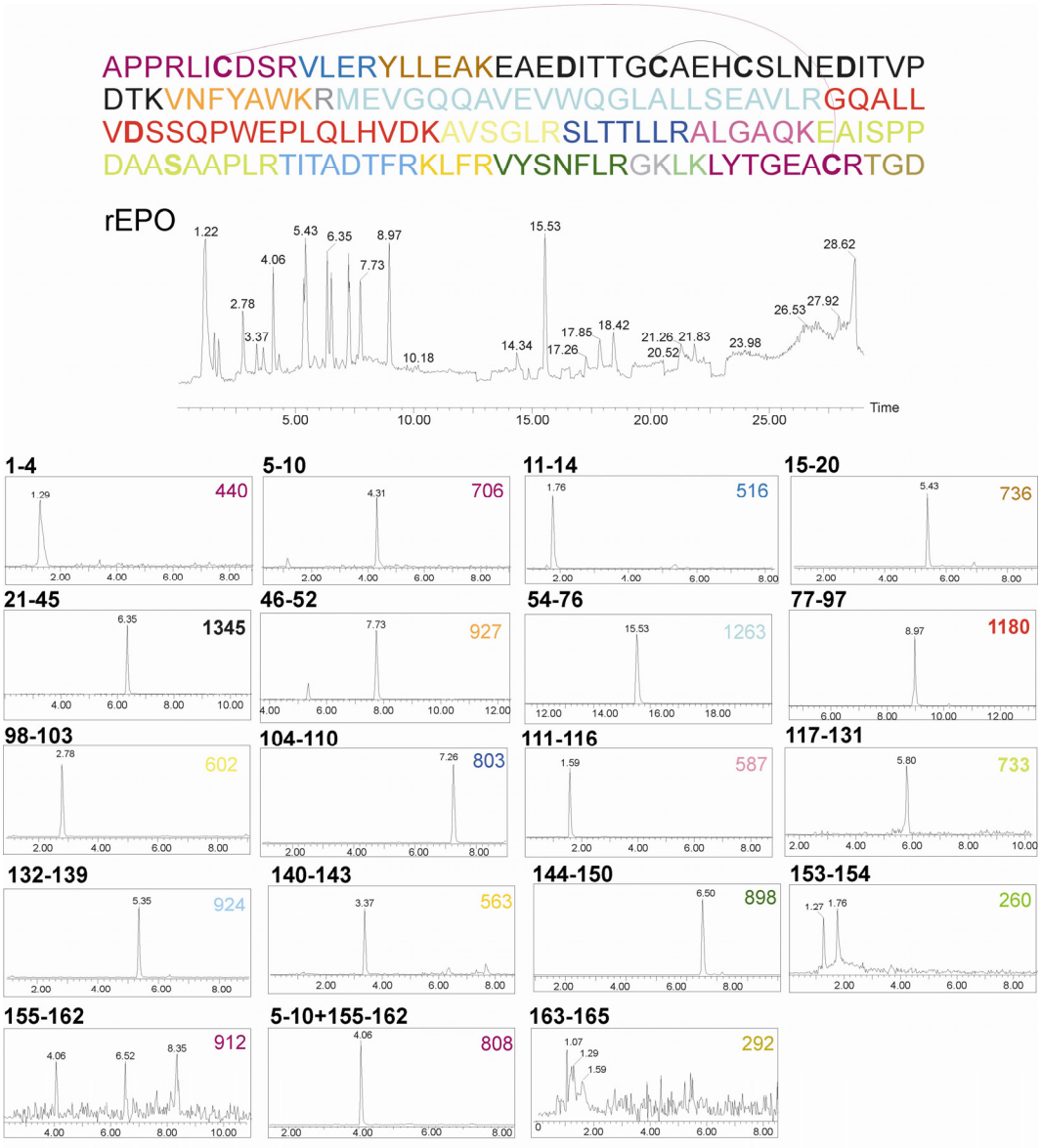


Figure 3. UPLC-ESI-TOF MS total ion chromatogram of de-N-glycosylated rEPOs, with in the lower segments the extracted ion chromatograms for the single peptides.

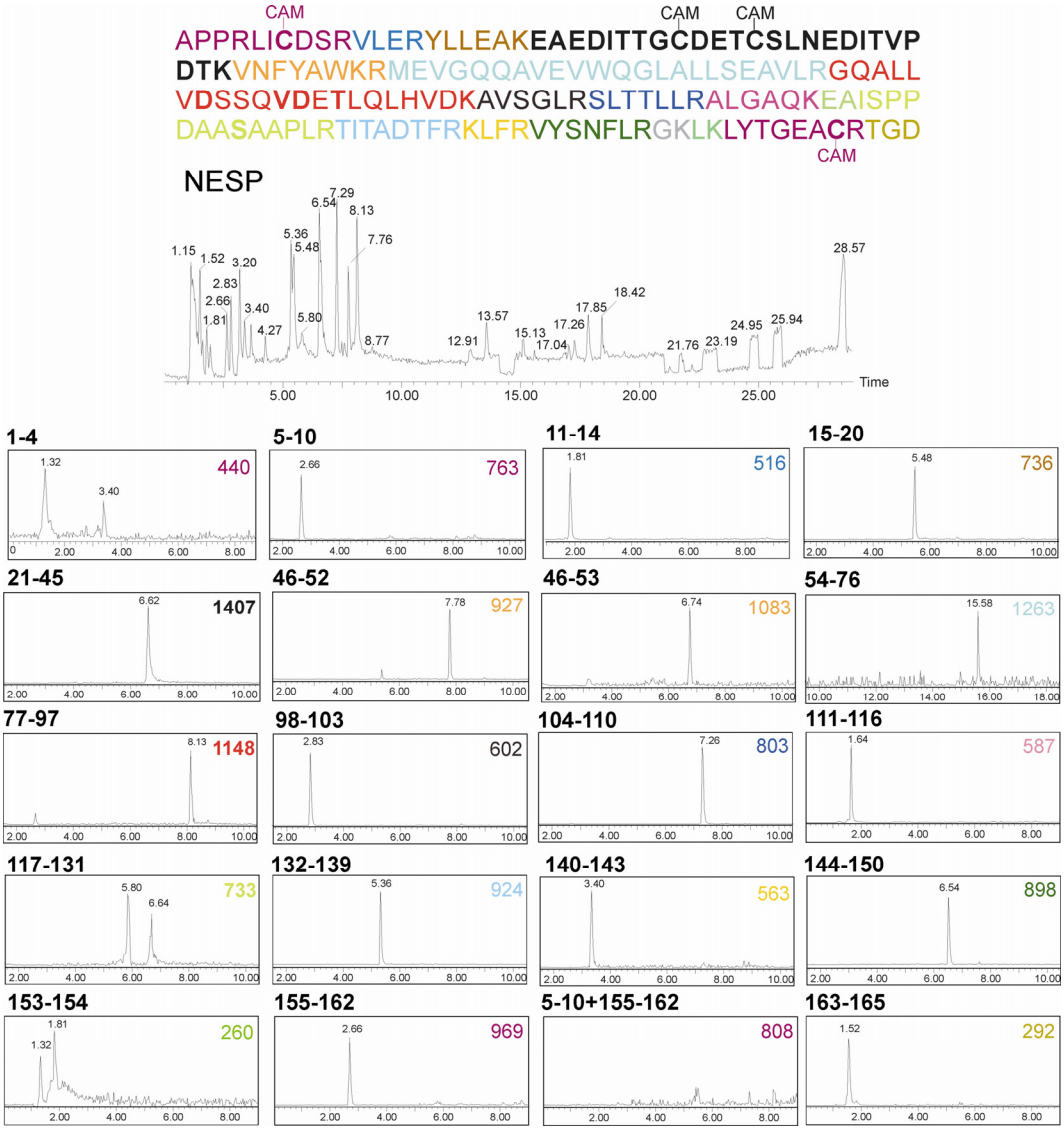


Figure 4. UPLC-ESI-TOF MS total ion chromatogram of de-N-glycosylated NESP, with in the lower segments the extracted ion chromatograms for the single peptides.

Given that rEPOs and NESP share 97% similarity in amino acid sequence, the majority of proteolytic peptides are common for both (Table I). In this comparative study, one of the differential peptides arose from the difference in sample preparation. The disulfide bonds of the two rEPOs were not reduced nor alkylated so peptides T2 and T20 are not considered diagnostic peptides. Releasing of the peptides T2 and T20 from the T2-T20 cluster (linked together by the disulfide

bond between Cys7 and Cys161) would not add further information for differentiate the two analogues. As expected, in rEPOs samples, the doubly (808) and triply charged (539) ions corresponding to T2+T20 appeared to predominate. Ions representing separated peptides T2 and T20 were also detected in lower abundance but also these were different from the analogues in NESP as in this case the Cys residues are not carboxyamido-methylated. Conversely, reduction and alkylation treatments were needed for NESP peptide mapping in order to achieve complete de-N-glycosylation. According to this treatment, peptides T2 and T20 were detected with their cystein residues carboxyamido-methylated. This modification is reflected by a mass increment of 57 Da in each peptide containing a Cys residue. Neither the T2-T20 peptide nor T2 and T20 without CAM were detected (Figure 4).

The most relevant fragments were the E5/N5 (21-45) and E9/N9 (77-97) because these two tryptic peptides could be employed for differentiating rEPOs from NESP (Figure 5). Interestingly, the peptide 21-45 contains both a disulfide bond and several N-glycosylation sites; two in rEPOs and three in NESP. In case of rEPOs, analyses were performed under non-reductive conditions, so the disulfide bond remained intact reducing the mass by two protons with respect to the linear sequence, whereas the de-N-glycosylation converted asparagine into aspartic acid incrementing the mass by two mass units. Thus, the resulting peptide (E5) is identical in mass to a non-glycosylated peptide obtained after reduction and alkylation and could only be differentiated through MS-MS experiments. For identification purposes, the sequence 21-45 is of interest as in NESP this peptide contains two mutations (A30-N and H32-T) yielding a theoretical mass difference of + 7.00 Da. In practice, however one other aspect must be considered. The mutation in A30 to N30 inserts a new glycosylation site that after PNGase F digestion is deamidated to aspartic acid. This change is reflected by mass increment of 0.98 Da. It should be also noted that before de-N-glycosylation NESP sample was reduced and alkylated, so in this case Cys29 and Cys33 are carboxyamidomethylated. These two modifications mean a mass increment of

114.05 over the non alkylated peptide. Ions 1407 and 938 corresponding to singly and double charged of fragment N5 were identified in NESP sample only. N5 included 21-45 aa sequence in which both Cys are CAM and the three N are converted to Asp (D). Alternatively, the peptide N9 (77-97) could be also employed for discriminatory purposes as this one incorporates the other three amino acid modifications (P87-V, W88-N, and P90-T) giving rise to a theoretical mass difference of 66.03 Da. Data showed that this difference was 65.04 as N88 is a glycosylation site that was converted to Asp (D) after PNGase F digestion. In figure 5, the characteristic ions for the differential peptide 77-97 are also shown. Ions corresponding to rEPOs 77-97 fragment (1180 and 767; doubly and triply charged respectively) were not found in NESP preparation. Conversely, characteristic ions belonging to the same fragment in NESP (1148 and 765 doubly and triply charged respectively) were absent in rEPOs.

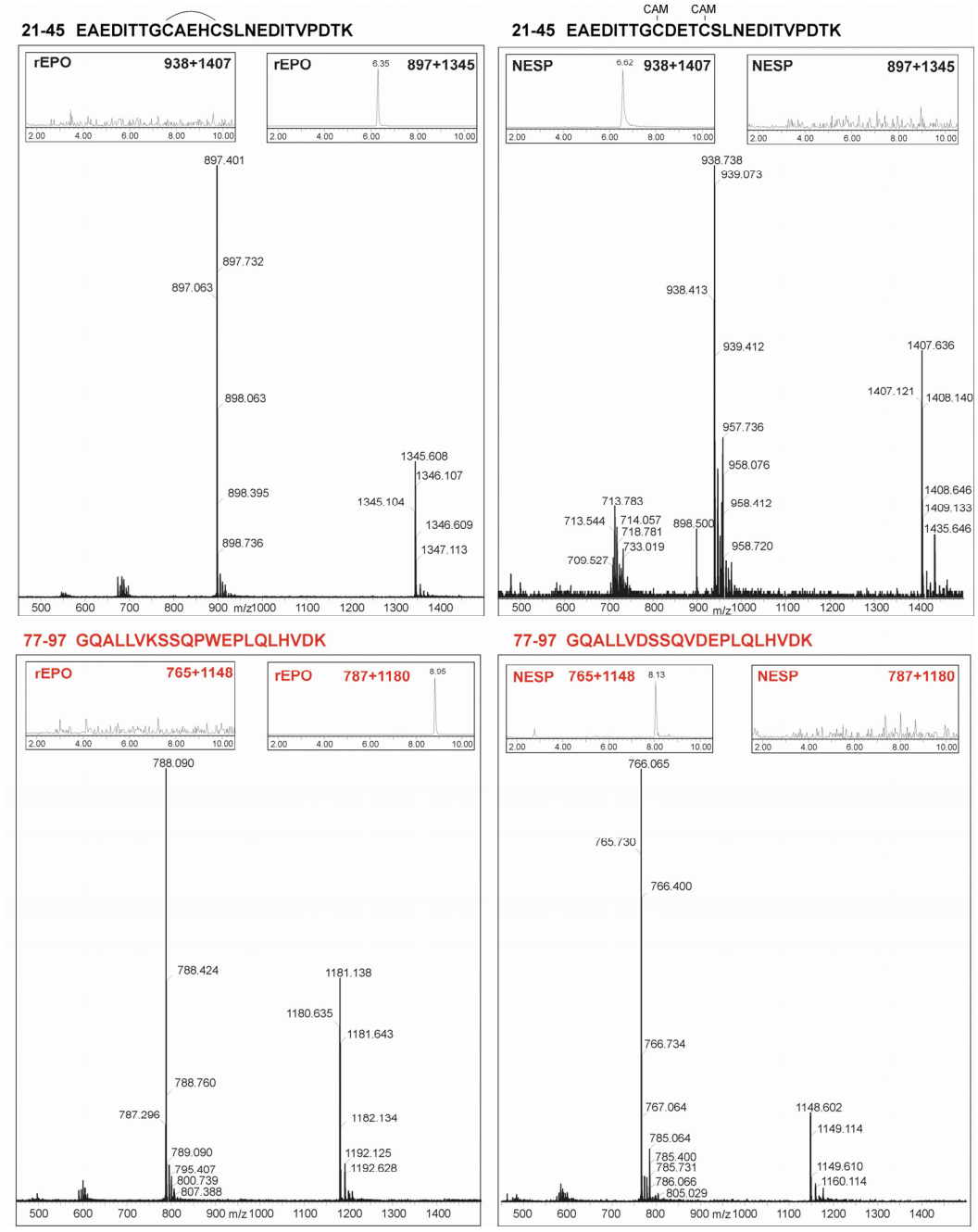


Figure 5. Example of diagnostic peptides for rEPOs (left) and NESP (right) able to distinguish between both. Extracted-ion UPLC-ESI-TOF MS chromatograms for peptides 21-45 (top) and 77-97 (bottom). Both peptides are truly diagnostic for each substance as demonstrate by absence of 1407 and 1148 ions in rEPOs and absence of 1345 and 1180 in NESP. The corresponding mass spectra are depicted with doubly and triply protonated species for each peptide.

Despite that the unique diagnostic peptides for unambiguous discrimination between rEPOs and NESP are segments 21-45 and 77-97, additional differences arose from the O-glycopeptide. These differences were not only between rEPO and NESP but also between rEPO α/β and rEPO δ (Figure 6). First of all it must be highlighted that UPLC-ESI-TOF MS allowed the analyses of the intact O-glycopeptides maintaining all its glycoforms [25]. In previous analyses, conducted with MALDI-TOF MS only the non-O-glycosylated peptide 117-131 and Tn antigen could be observed in the peptide mixture. Contrary to MALDI-TOF MS, UPLC-ESI-TOF MS was capable of detecting peptides containing the Tn, T, sialyl-T and disialyl-T antigen at $[M+2H]^{2+}$ of 834.92, 915.93, 1061.50, and 1207.56, respectively. Furthermore, the O-glycopeptide 117-131 was also observed devoid of O-glycan $[M+2H]^{2+}$ at *m/z* 733.38.

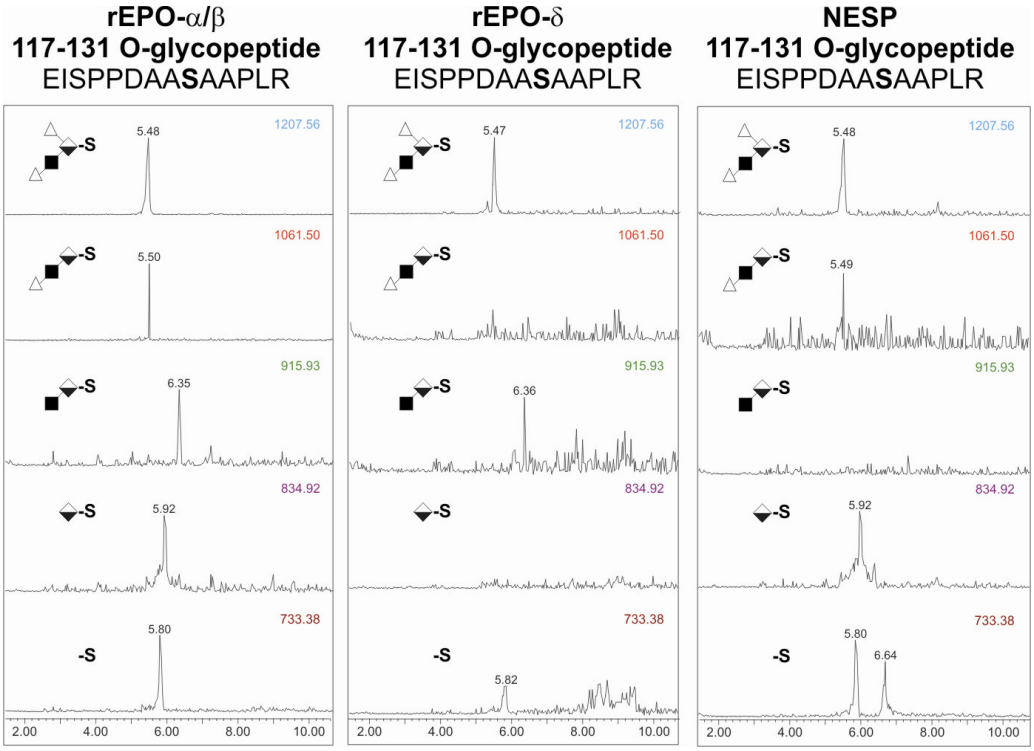


Figure 6. O-glycopeptide structures of rEPO α/β (left), rEPO δ (middle) and NESP (right) detected by UPLC-ESI-TOF MS analysis after trypsinolysis. Short-hand notation: (⬙) *N*-acetylgalactosamine, (■) galactose, (Δ) sialic acid.

The glycopeptide mapping allowed the determination of the carbohydrate structure of O-linked glycans without separation from the peptide backbone. This is important because until now, there is no commercially available enzyme, capable of releasing intact O-glycans. Chemical methods employed for this purpose may alter the original structure of O-glycans by losing labile residues like sialic acids. A comparison of the O-glycopeptide signals revealed a considerable variability in the three samples. rEPO δ showed the most homogeneous structure, being predominant the disialyl-T form, small proportion of T antigen and completely absence of sialyl-T and Tn antigen. The peptide without O-glycan could also be identified but in low amounts. In rEPO α/β sample, all forms, including non-O-glycosylated peptide, Tn, T, sialyl-T and disialyl-T were observed, and in NESP all mentioned structures except for T antigen were detected. Thus, the analysis of O-glycoforms could be used to differentiate two rEPOs that have 100% sequence homology.

Data obtained for O-glycan isoforms from O-glycopeptides analyses do not seem to be in agreement with results obtained by MALDI-TOF MS of the de-N-glycosylated protein. Some authors described loss of sialic acids when glycoproteins are analysed by MALDI-TOF MS [24] and recommended a soft ionisation technique (ESI) for more reliable analyses [26]. Further experiments must be done in order to understand the reasons for these differences, and corroborate the real O-glycoforms present in different EPO samples. Furthermore, detection of each individual O-glycoform appears as a very promising method for differentiation between rEPOs and the quantification of these biomarker candidates could lead to the development of an antidoping test [22]. Since mass spectrometry is the technique of choice for specific identification of banned substances in doping control analyses [27] and LC-MS is the preferred technique for identification of protein-based drugs such as rEPO and NESP, we propose the exploitation of the results for, once sensitivity required is achieved, developing a confirmatory method capable of detecting the exogenous EPO analogues present in biological fluids.

3.1.4 Bibliography

[1] S. Elliott, *Erythropoiesis-stimulating agents and other methods to enhance oxygen transport*. Br J Pharmacol *154,* 529-541 (2008).

[2] C. E. Cooper, *The biochemistry of drugs and doping methods used to enhance aerobic sport performance*. Essays Biochem *44,* 63-83 (2008).

[3] G. Walsh and R. Jefferis, *Post-translational modifications in the context of therapeutic proteins*. Nat Biotechnol *24,* 1241-1252 (2006).

[4] A. Khan, J. Grinyer, S. T. Truong, E. J. Breen and N. H. Packer, *New urinary EPO drug testing method using two-dimensional gel electrophoresis*. Clin Chim Acta *358,* 119-130 (2005).

[5] W. W. Franke and H. Heid, *Pitfalls, errors and risks of false-positive results in urinary EPO drug tests*. Clin Chim Acta *373,* 189-190 (2006).

[6] A. Khan and M. S. Baker, *Non-specific binding of monoclonal human erythropoietin antibody AE7A5 to Escherichia coli and Saccharomyces cerevisiae proteins*. Clin Chim Acta *379,* 173-175 (2007).

[7] M. Beullens, J. R. Delanghe and M. Bollen, *False-positive detection of recombinant human erythropoietin in urine following strenuous physical exercise*. Blood *107,* 4711-4713 (2006).

[8] F. Lasne, L. Martin, N. Crepin and J. de Ceaurriz, *Detection of isoelectric profiles of erythropoietin in urine: differentiation of natural and administered recombinant hormones*. Anal Biochem *311,* 119-126 (2002).

[9] J. R. Delanghe, M. Bollen and M. Beullens, *Testing for recombinant erythropoietin*. Am J Hematol *83,* 237-241 (2008).

[10] F. Guan, C. E. Uboh, L. R. Soma, E. Birks, J. Chen, Y. You, J. Rudy and X. Li, *Differentiation and identification of recombinant human erythropoietin and darbepoetin Alfa in equine plasma by LC-MS/MS for doping control*. Anal Chem *80,* 3811-3817 (2008).

[11] J. C. Egrie and J. K. Browne, *Development and characterization of novel erythropoiesis stimulating protein (NESP)*. Nephrol Dial Transplant *16 Suppl 3,* 3-13 (2001).

[12] R. Deicher and W. H. Horl, *Differentiating factors between erythropoiesis-stimulating agents: a guide to selection for anaemia of chronic kidney disease*. Drugs *64,* 499-509 (2004).

[13] W. Jelkmann, *Developments in the therapeutic use of erythropoiesis stimulating agents*. Br J Haematol *141,* 287-297 (2008).

[14] J. Reinders, U. Lewandrowski, J. Moebius, Y. Wagner and A. Sickmann, *Challenges in mass spectrometry-based proteomics*. Proteomics *4,* 3686-3703 (2004).

[15] P. B. Mills, K. Mills, A. W. Johnson, P. T. Clayton and B. G. Winchester, *Analysis by matrix assisted laser desorption/ionisation-time of flight mass spectrometry of the post-translational modifications of alpha 1-antitrypsin isoforms separated by two-dimensional polyacrylamide gel electrophoresis*. Proteomics *1,* 778-786 (2001).

[16] E. Llop, R. Gutiérrez-Gallego, V. Belalcazar, G. J. Gerwig, J. P. Kamerling, J. Segura and J. A. Pascual, *Evaluation of protein N-glycosylation in 2-DE: Erythropoietin as a study case*. Proteomics *7,* 4278-4291 (2007).

[17] J. Gobom, E. Nordhoff, E. Mirgorodskaya, R. Ekman and P. Roepstorff, *Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry*. J Mass Spectrom *34,* 105-116 (1999).

[18] E. Gimenez, F. Benavente, J. Barbosa and V. Sanz-Nebot, *Towards a reliable molecular mass determination of intact glycoproteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*. Rapid Commun Mass Spectrom *21,* 2555-2563 (2007).

[19] J. C. Egrie and J. K. Browne, *Development and characterization of novel erythropoiesis stimulating protein (NESP)*. Br J Cancer *84 Suppl 1,* 3-10 (2001).

[20] M. Kohler, C. Ayotte, P. Desharnais, U. Flenker, S. Ludke, M. Thevis, E. Volker-Schanzer and W. Schanzer, *Discrimination of recombinant and endogenous urinary erythropoietin by calculating relative mobility values from SDS gels*. Int J Sports Med *29,* 1-6 (2008).

[21] S. M. Stanley and A. Poljak, *Matrix-assisted laser-desorption time-of flight ionisation and high-performance liquid chromatography-electrospray ionisation mass spectral analyses of two glycosylated recombinant epoetins*. J Chromatogr B Analyt Technol Biomed Life Sci *785,* 205-218 (2003).

[22] G. Stubiger, M. Marchetti, M. Nagano, R. Grimm, G. Gmeiner, C. Reichel and G. Allmaier, *Characterization of N- and O-glycopeptides of recombinant human erythropoietins as potential biomarkers for doping analysis by means of microscale sample purification combined with MALDI-TOF and quadrupole IT/RTOF mass spectrometry*. J Sep Sci *28,* 1764-1778 (2005).

[23] M. A. Recny, H. A. Scoble and Y. Kim, *Structural characterization of natural human urinary and recombinant DNA-derived erythropoietin. Identification of des-arginine 166 erythropoietin*. J Biol Chem *262,* 17156-17163 (1987).

[24] K. B. Linsley, S. Y. Chan, S. Chan, B. B. Reinhold, P. J. Lisi and V. N. Reinhold, *Applications of electrospray mass spectrometry to erythropoietin N- and O-linked glycans*. Anal Biochem *219,* 207-217 (1994).

[25] E. Mortz, T. Sareneva, I. Julkunen and P. Roepstorff, *Does matrix-assisted laser desorption/ionization mass spectrometry allow analysis of carbohydrate heterogeneity in glycoproteins? A study of natural human interferon-gamma*. J Mass Spectrom *31,* 1109-1118 (1996).

[26] D. J. Harvey, L. Royle, C. M. Radcliffe, P. M. Rudd and R. A. Dwek, *Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry*. Anal Biochem *376,* 44-60 (2008).

[27] M. Thevis and W. Schanzer, *Mass spectrometry in sports drug testing: Structure characterization and analytical assays*. Mass Spectrom Rev *26,* 79-107 (2007).

Chapter 3.2

Study of charges responsible for differential IEF profiles of EPOs.

Belalcazar V, Gutiérrez Gallego R, Llop E, Segura J, Pascual JA.
Assessing the instability of the isoelectric focusing patterns of erythropoietin in urine.
Electrophoresis. 2006 Nov;27(22):4387-95.

Chapter 3.3

Study of erythropoietin glycosylation.

Llop E, Gallego RG, Belalcazar V, Gerwig GJ, Kamerling JP, Segura J, Pascual JA.
Evaluation of protein N-glycosylation in 2-DE: Erythropoietin as a study case.
Proteomics. 2007 Dec;7(23):4278-91.

Chapter 3.4

Glycan profiling for discrimination between rEPO δ (dynepo™, produced in human cells) and rEPO α/β (BRP std, produced in CHO cells).

Llop E, Gutiérrez-Gallego R, Segura J, Mallorquí J, Pascual JA.
Structural analysis of the glycosylation of gene-activated erythropoietin(epoetin delta, Dynepo).
Anal Biochem. 2008 Dec 15;383(2):243-54. Epub 2008 Sep 3.

Chapter 3.5

Sialic acid profiling: characterisation of sialic acids released from rEPO preparations and from biological samples.

3.5.1 Introduction

In recent years, interest in the diversity of sialic acids has been renewed, because these monosaccharides play important biological functions. Sialic acids are a family of nine-carbon carboxylated sugars, widely distributed in the mammalian glycoconjugates such as glycoproteins and glycolipids, and located predominantly at the outermost ends. Sialic acids present an extreme diversity, and more than 60 different compounds were identified differing in the presence or absence of an amino group in position 5 (sialic acids or 3-deoxy-D-glycero-D-galacto-nonulosonic acid [Kdn]), different acylations of the NH_2 group at position 5 (acetyl, glycolyl), and various substituents of the different hydroxyl groups (acetyl, lactyl, methyl, sulphate, phosphate, etc.) [1]. Although the free forms have mainly the β-anomeric ring conformation, sialic acids bound to glycoconjugates occur specifically in the α-anomeric form. The diverse roles of sialic acids include the microbe binding that leads to infections [2, 3], the regulation of the immune response, the progression and spread of human malignancies and in certain aspects of human evolution [4]. Sialic acids are important not only as part of the oligosaccharides characterisation in glycoproteins but also have been reported as markers of some diseases [5]. Given that sialic acids are directly related to half-life of pharmaceuticals [6], analyses are relevant for monitoring glycosylation in marketed protein drugs for quality control [7].

The most common mammalian sialic acids are *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc). Celular Neu5Gc is generated by hydroxylation of the sugar nucleotide donor CMP-Neu5Ac to CMP-Neu5Gc, catalysed by CMP-Neu5Ac hydroxylase (CMAH). Although Neu5Gc is a major sialic acid in most mammals (including our closest evolutionary relatives, the great apes), it is absent in healthy adult humans [8]. Despite this, presence of Neu5Gc was demonstrated in human carcinomas [9, 10] and fetal tissues [11]. An improved affinity antibody detected not only Neu5Gc expression in tumour cells, but also traces in normal human tissues [12], despite the fact that humans

are unable of its synthesis. This paradox is now explained by the fact that humans may absorb small quantities of Neu5Gc from dietary sources (principally red meat and milk) and metabolically incorporate it into certain cell types. A similar “contamination” by Neu5Gc apparently could occur also in the biotechnology industry, arising from the use of animal cells, animal sera and other animal products during manufacture. Presence of Neu5Gc is likely widespread in many biopharmaceuticals and cells, and could potentially play a part in immune responses against such agents [13]. This scenario of quality control of therapeutics together with the possibility of detecting the abuse of some recombinant glycoproteins by athletes (rEPO, rhCG, rLH) [14] demands high resolution separation techniques for sialic acids and high sensitivity for detection this minor compound in biological fluids like serum, plasma, and/or urine. Although several methods based on colorimetric assays, lectins and antibodies exist for sialic acid detection [15], the structural analysis of naturally occurring sialic acis is currently carried out by the following two main approaches: analysis of volatile sialic acids derivatives by gas-chromatography (GC) combined with electron-impact mass spectrometry (EI/MS), and analysis of fluorescently labelled sialic acids by high-performance liquid chromatography (HPLC) [16] eventually coupled to electrospray mass spectrometry (ESI/MS) [17]. In this chapter, a highly sensitive HPLC method with fluorescence detection, for the determination of different sialic acids in plasma and urine is described.

3.5.2 Materials and methods

Materials

Recombinant human EPO (rEPO produced in CHO cells) was obtained from the European Pharmacopoeia Commission (rEPO α/β) and epoetin delta (rEPO produced in human fibrosarcoma HT-1080 cells) from Shire Pharmaceuticals (rEPO δ; Dynepo™). Darbepoetin α (NESP) was purchased as pharmaceutical preparation from Amgen. Monoclonal anti-human erythropoietin antibody (clone 9C21D11 was from R&D systems. 96-well microtiter plates were from Costar.

Sialic acid reference panel was from Glyco. Phospate-buffered saline (PBS), polyvinylpirrolidone (PVP), dimethyl sulfoxide (DMSO), 1,2-diamino-4,5-methylenedioxy-benzene (DMB) and 2-mercaptoethanol were purchased from Sigma. Sodium cyanoborohydride ($NaBH_3CN$) was from Fluka. All other chemicals were of highest purity commercially available.

Immunopurification of plasma samples

Due to the high protein content in plasma, a preliminary step to extract EPO by immunoaffinity was necessary prior to sialic acid hydrolysis. An immunopurification method was recently developed by our group [18] and employed for this purpose. Briefly, a 10 µg/mL solution of monoclonal anti-human erythropoietin antibody (clone 9C21D11 from R&D systems) in 50 mM ammonium bicarbonate pH 9.0 was coupled to microtiter plates for 1 h at 37 ºC. After coupling, plates were washed three times with 10 mM phospate-buffered saline pH 7.4 (PBS) and blocked with 1% polyvinylpirrolidone in PBS overnight at 4 ºC. Once the plates were ready, plasma samples were applied to the plates (100 µl/well) and incubated for 16 h at 4 ºC with shaking. After incubation, blocking reagent was aspirated and wells washed once with PBS. Next, plates were washed twice with PBS and finally erythropoietin was eluted with a 0.7% acetic acid solution pH 2.0. Eluates were immediately frozen and lyophilised before sialic acid hydrolysis.

Sialic acid analyses

For the study of the sialic acid heterogeneity, sialic acids were released from the carbohydrate chains and derivatised with 1,2-diamino-4,5-methylene dioxybenzene (DMB) [19]. Briefly, 10 µg of each standard sample was hydrolysed with 10 µl of 2 M aqueous acetic acid solution for 3 h at 80 ºC. Next, a 7 mM DMB solution (10 µl) in 1.4 M aqueous acetic acid containing 18 mM sodium hydrosulphite and 1 M β-mercaptoethanol was added. The mixture was kept for 2 h at 50 ºC. Fluorescently labelled residues were diluted twice in water

and analysed directly by UPLC ESI-TOF mass spectrometry using an Acquity TM Ultra Performance Liquid Chomatograph (WATERS) coupled to a LCT premier XE ESI-TOF instrument (Micromass). LC analyses were performed at a flow-rate of 0.2 ml/min using an Acquity UPLC BEH C_{18} column (100 x 2.1 mm, 1.7 µm). Derivatised sialic acids were eluted isocratically employing MeOH:ACN:H_2O (7:9:84) for 5 minutes, then with a 1 min gradient the running buffer was changed to MeOH:ACN:H_2O (7:25:68), maintained for 1 min, instantaneously returned to initial conditions, and stabilised for 1 min before the next injection [20]. Mass spectra were acquired in negative ion mode over a mass range-to-charge (*m/z*) ratio of 50-650 Da. The capillary voltage was set at 2800 V (negative), the desolvation temperature to 250-300 ºC and the desolvation gas-flow set to 300 L/h. The TOF tube voltage was kept at 7200 (reflectron at 1800 V) with a pusher setting of 900 V (pusher offset at 0.93). Recorded data were processed using MassLynx version 4.1 (Waters Inc.). DMB derivatives of sialic acids released from 1 µg of recombinant EPO samples and from immunopurified plasma samples were also analysed by reversed phase chromatography on a capillary column Zorbax SB-C_{18} (150 x 0.3 mm, 3.5 µm) using ACN:H_2O (20:80) as eluent (mobile phase) and a flow rate of 4 µl/min. The sialic acid separation was completed within 35 min. Chromatographic analyses were performed on an Agilent 1100 series capillary instrument equipped with a JASCO micro21FP capillary fluorescence detector (λ_{ex}=373 nm and an λ_{em}=448 nm).

3.5.3 Results and discussion

Sialic acid profiling of recombinant preparations

In the framework of this work, sialic acids released from different recombinant EPO preparations were analysed by using two different approaches. The first method used was based on ultra-performance liquid chromatography coupled with electrospray ionisation mass spectrometry (ESI-TOF MS) and was applied to sialic acids released from rEPO α/β (BRP standard), NESP (both synthesised in CHO cells) and rEPO δ produced in human cells. The resulting chromatograms are depicted in Figure 1. The major constituent in the three recombinant EPO

preparations was ion *m/z* 424.1 corresponding to Neu5Ac. Although it was the only sialic acid present in rEPO δ, other species appeared in rEPO α/β and NESP. The presence of *N*-glycolylneuraminic acid (Neu5Gc) was confirmed in rEPO α/β and NESP by monitoring the molecular ion *m/z* 440.1 which due to its more hydrophilic composition, was eluted at the shortest elution time (2.51 min). O-acetylated derivatives were also observed by monitoring molecular ions at *m/z* 466.1 and *m/z* 508.1. The former corresponds to di-acetylated species Neu5,7(8)Ac_2 and Neu5,9Ac_2 that despite their identical molecular mass, were identified by comparison to the different retention times of commercial sialic acids panel. The later, observed at *m/z* 508 corresponded to the tri-acetylated variant Neu5,7,9Ac_3. Data obtained for rEPO α/β and NESP were in perfect agreement with literature reporting the presence of ~1% Neu5Gc in recombinant preparations, not only in rEPO [20, 21] but also in other recombinant glycoproteins expressed in normal CHO cells such as tissue plasminogen activator, chorionic gonadotropin [14], etc. Interestingly, no Neu5Gc was present in rEPO δ despite being a recombinant preparation (Figure 1). Although this rEPO preparation is produced in human cells that are not able to produce Neu5Gc, it has been reported that human cells could potentially take up this compound from the media (containing animal product) which rEPO preparation is cultured in [22]. This fact is relevant for rEPO δ clinical applications because the absence of Neu5Gc could avoid immunogenic response in patients [23] and in the framework of anti-doping policies due to the presence of Neu5Gc could be suitable for detecting the abuse of some rEPOs preparations but not the rEPO δ [21].

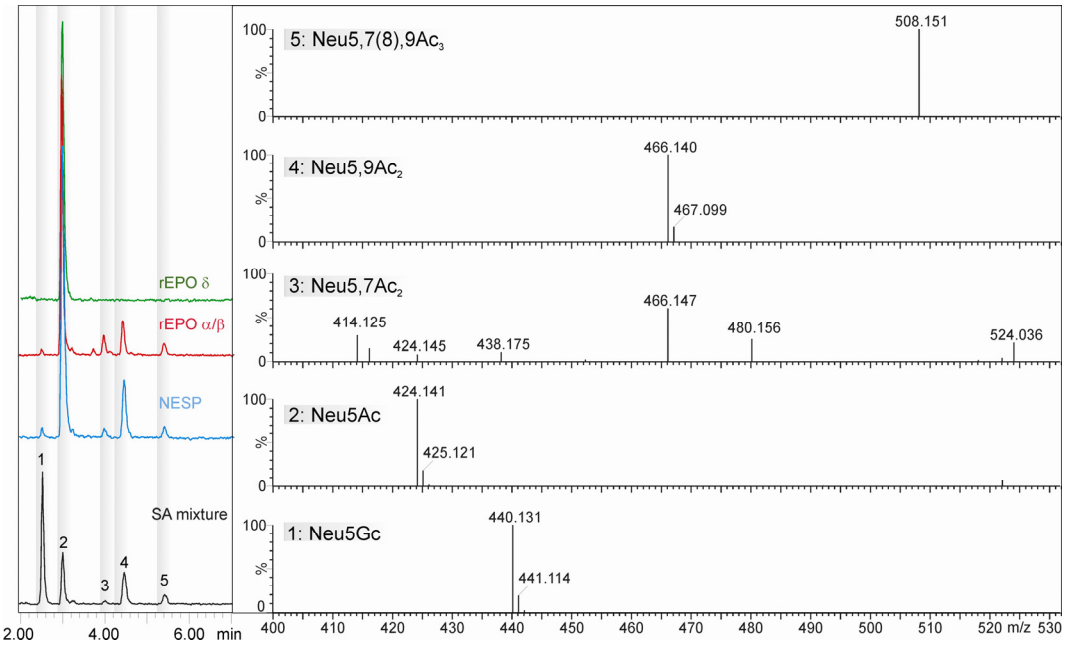


Figure 1. On the left, the extracted ion chromatograms of LC-MS analysis of the DMB-derivatised sialic acid residues derived from rEPO δ (in green); rEPO α/β (in red); NESP (in blue) compared with the standard sialic acid (SA) mixture. On the right, the ESI-MS spectra of major sialic acid species found in rEPO α/β and NESP preparations.

The method described above, with a limit of detection of 6 pmol (~2 ng) for Neu5Gc and Neu5Ac DMB derivatives, resulted very useful for characterising recombinant EPO pharmaceuticals and standards for which sample amount of μg's are available. However the main goal of the present work consists of detecting the non-human sialic acid Neu5Gc in human biological fluids (plasma and urine) from athletes abusing of recombinant preparations containing this compound. Given that the erythropoietin is normally present in urine and plasma at a very low concentration (about 10 ng/L and 100 ng/L, respectively) [24] the LC-MS method could not be applied to these samples. For this reason, we adapted a highly sensitive capillary HPLC method for the determination of sialic acids in immunopurified erythropoietin from plasma. Although mass spectrometry is the technique of choice for specific identification of banned substances in doping control analyses [25], in the method developed, fluorescence detection

was preferred for its high sensitivity and the possibility of quantitative determination (quantitative determination of sialic acids from rEPOs published by our group as part of their glycan analyses [21, 26]). Another advantage is that the derivatisation reaction is specific for alfa-keto acids (as are sialic acids) [27], so very specifically one single fluorescent group is attached to every sialic acid minimising matrix interferences and allowing an accurate quantification under the conditions reported [28, 29]. Although the method has not been validated yet, the limit of detection of capillary HPLC for DMB-Neu5Ac and Neu5Gc was 6 fmol (~1.9 pg) each (s/n=3). In this method, assignment was based on the retention times as compared with the reference compounds (Glyco sialic acid reference panel).

Characterisation of sialic acids released from biological samples.

The presence of Neu5Gc was monitored in genuine plasma samples from individuals treated with rEPO. Plasma samples were immunopurified according to the procedure developed in our group [18], filtered through 50 kDa in order to remove possible PVP eluted from the well plates and then lyophilised. The purified extracts were derivatised with DMB immediately before analysis. Samples chosen as showcases were from 15,2 mL of the following plasma's: from a negative individual (meaning plasma containing only endogenous EPO), negative plasma spiked with 30 ng rEPO α/β and positive plasma (meaning plasma with an recombinant EPO IEF profile.

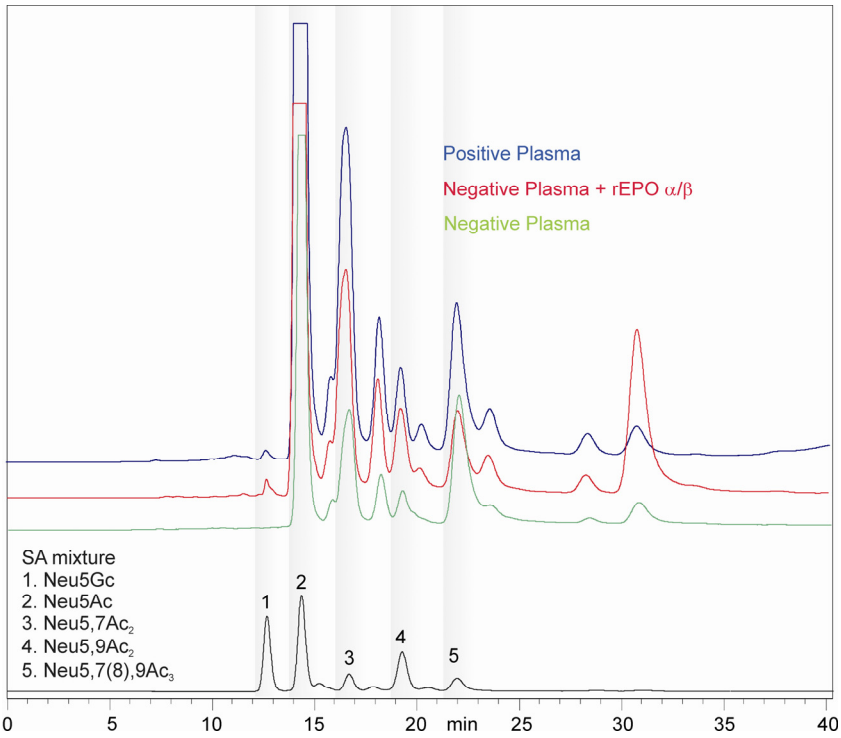


Figure 2. HPLC-FLD analysis of the DMB-derivatised sialic acid residues derived from EPO immunopurified from negative plasma (in green); negative plasma spiked with rEPO α/β (in red) and positive plasma (rEPO IEF profile) compared with the standard sialic acid (SA) mixture.

Sialic acid analyses of immunopurified samples indicated, that blank plasma used as a negative control does not contain Neu5Gc. However, when this plasma was spiked with 30 ng of rEPO before the purification, the resulting chromatogram showed the presence of Neu5Gc (Figure 2). In order to corroborate the unequivocal presence of Neu5Gc, and as a control of its chromatographic behaviour; pure Neu5Gc was added to the sialic acids extract from negative plasma to monitor the appearance of its peak. The same procedure was followed with the positive plasma confirming by co-elution the exact retention time and identification of the NeuGc peak found in the plasma.

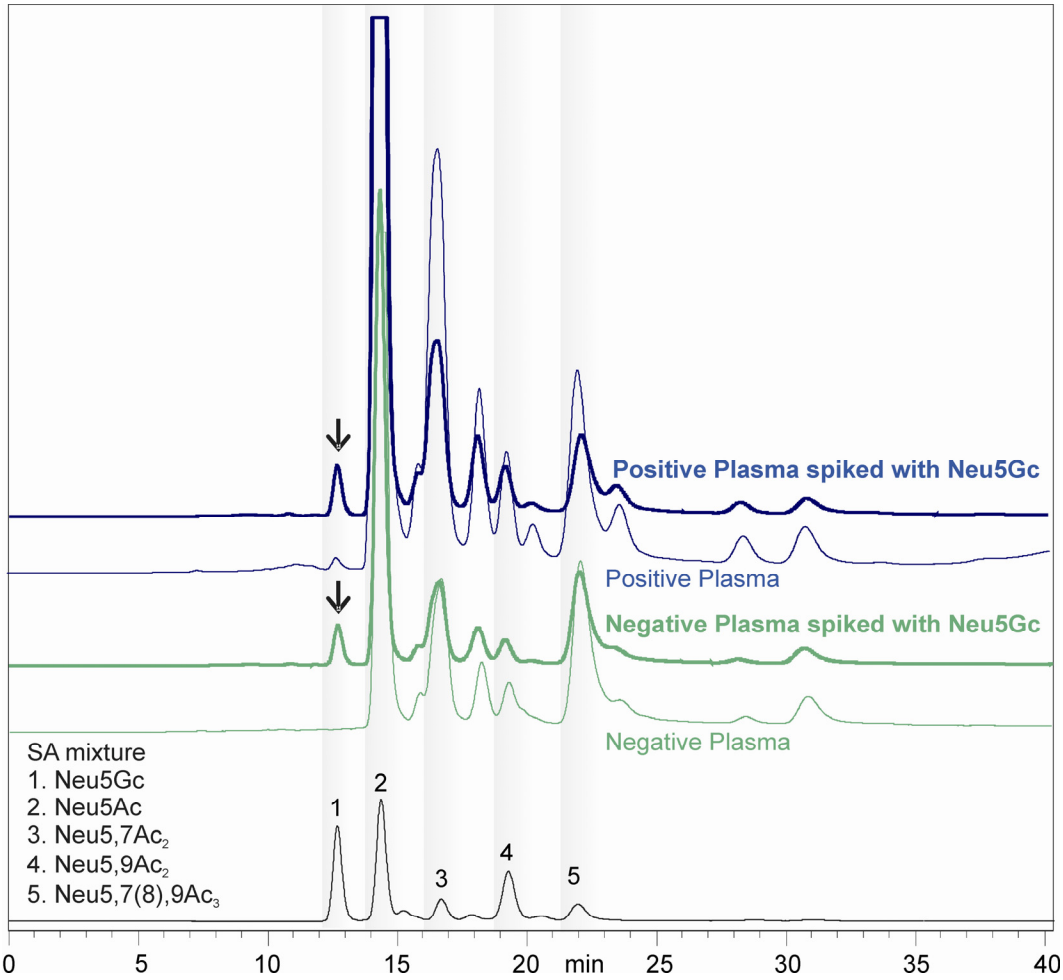


Figure 3. HPLC-FLD analysis of the DMB-derivatised sialic acid residues derived from EPO immunopurified from negative plasma (green thin line); negative plasma spiked with Neu5Gc (green thick line); positive plasma (blue thin line) and positive plasma spiked with Neu5Gc (blue thick line) compared with the standard sialic acid (SA) mixture.

Finally, a plasma sample from an athlete suspected to be administered with rEPO was immunopurified and analysed by the regular IEF method developed by Lasne *et al.* for urine samples [24]. The IEF profile of this sample showed similar to the rEPO standard so was subsequently assayed for its sialic acids content (Figure 4).

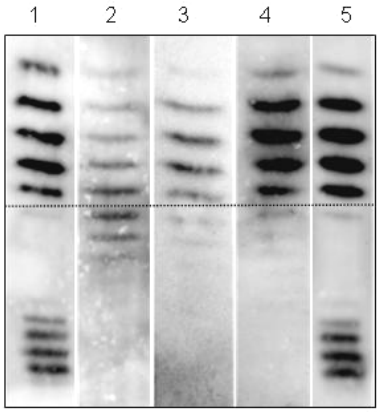


Figure 4. Analysis of erythropoietin in plasma samples by isoelectric focusing. (1,5) Mixture of rEPO α/β standard plus its analogue NESP; (2) Negative plasma (from a subject known to contain endogenous EPO only); (3) Negative plasma spiked with rEPO α/β; (4) Positive plasma.

As can be seen in the chromatogram (Figure 3), the presence Neu5Gc could be established. Even though the immunopurification method should not be considered 100% specific as the amounts of EPO are very low, in comparison to other proteins, the fact that the negative plasma sample did not show any Neu5Gc is very promising. The suggestion that Neu5Gc (ingested as part of the diet) may be incorporated to other proteins does not seem be a real impediment for the developed method because it does not affect EPO. Although this method was successful in confirming the presence of rEPO α/β, this observation has to be further validated with a larger amount of plasma samples from different origins and possibly try to adapt the method to a mass spectrometric detection. However, the sensitivity obtained using capillary fluorescence detection is already at the limit of what can be achieved using nanoLC/MS. The limitations of this methods consist of the matrix and volume required (~30 ml of blood), the fact that urine is preferred in doping control analyses [30], and its usefulness for unequivocal discrimination between recombinant and endogenous EPO when the recombinant has been synthesised in human like the second-generation product rEPO δ. We consider that this method, after validation in both urine and plasma samples, should be exploited as screening and/or confirmation method for

detecting not only the administration of recombinant erythropoietin but also other recombinant drugs synthesised in CHO cells. Given that Neu5Gc can be immunogenic in humans; this method could be also used in quality control of these pharmaceutical selecting only preparations with minimal Neu5Gc content [31].

3.5.4 Bibliography

[1] N. M. Varki and A. Varki, *Diversity in cell surface sialic acid presentations: implications for biology and disease*. Lab Invest *87,* 851-857 (2007).

[2] F. N. Lamari and N. K. Karamanos, *Separation methods for sialic acids and critical evaluation of their biologic relevance*. J Chromatogr B Analyt Technol Biomed Life Sci *781,* 3-19 (2002).

[3] F. Lehmann, E. Tiralongo and J. Tiralongo, *Sialic acid-specific lectins: occurrence, specificity and function*. Cell Mol Life Sci *63,* 1331-1354 (2006).

[4] N. Sharon, *Carbohydrates as future anti-adhesion drugs for infectious diseases*. Biochim Biophys Acta *1760,* 527-537 (2006).

[5] P. Sillanaukee, M. Ponnio and I. P. Jaaskelainen, *Occurrence of sialic acids in healthy humans and different disorders*. Eur J Clin Invest *29,* 413-425 (1999).

[6] J. C. Egrie, E. Dwyer, J. K. Browne, A. Hitz and M. A. Lykos, *Darbepoetin alfa has a longer circulating half-life and greater in vivo potency than recombinant human erythropoietin*. Exp Hematol *31,* 290-299 (2003).

[7] H. Schellekens, *Recombinant human erythropoietins, biosimilars and immunogenicity*. J Nephrol *21,* 497-502 (2008).

[8] H. H. Chou, T. Hayakawa, S. Diaz, M. Krings, E. Indriati, M. Leakey, S. Paabo, Y. Satta, N. Takahata and A. Varki, *Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution*. Proc Natl Acad Sci U S A *99,* 11736-11741 (2002).

[9] H. Higashi, Y. Hirabayashi, Y. Fukui, M. Naiki, M. Matsumoto, S. Ueda and S. Kato, *Characterization of N-glycolylneuraminic acid-containing gangliosides as tumor-associated Hanganutziu-Deicher antigen in human colon cancer*. Cancer Res *45,* 3796-3802 (1985).

[10] S. Kawachi and T. Saida, *Analysis of the expression of Hanganutziu-Deicher (HD) antigen in human malignant melanoma*. J Dermatol *19,* 827-830 (1992).

[11] M. J. Martin, A. Muotri, F. Gage and A. Varki, *Human embryonic stem cells express an immunogenic nonhuman sialic acid*. Nat Med *11,* 228-232 (2005).

[12] V. Padler-Karavani, H. Yu, H. Cao, H. Chokhawala, F. Karp, N. Varki, X. Chen and A. Varki, *Diversity in Specificity, Abundance and Composition of Anti-Neu5Gc Antibodies in Normal Humans: Potential Implications for Disease*. Glycobiology *18 (10),* 818-830 (2008).

[13] N. Hashii, N. Kawasaki, Y. Nakajima, M. Toyoda, Y. Katagiri, S. Itoh, A. Harazono, A. Umezawa and T. Yamaguchi, *Study on the quality control of cell therapy products. Determination of N-glycolylneuraminic acid incorporated into human cells by nano-flow liquid chromatography/Fourier transformation ion cyclotron mass spectrometry*. J Chromatogr A *1160,* 263-269 (2007).

[14] R. Ramirez-Llanelis, E. Llop, R. Ventura, J. Segura and R. Gutiérrez-Gallego, *Can glycans unveil the origin of glycoprotein hormones? - human chorionic gonadotrophin as an example* -. J Mass Spectrom *43,* 936-948 (2008).

[15] H. J. Kim, S. J. Lee and H. J. Kim, *Antibody-based enzyme-linked lectin assay (ABELLA) for the sialylated recombinant human erythropoietin present in culture supernatant*. J Pharm Biomed Anal (2008). PMID: 18722069

[16] J. P. Kamerling and G. J. Gerwig, *Structural analysis of naturally occurring sialic acids*. Methods Mol Biol *347,* 69-91 (2006).

[17] N. Morimoto, M. Nakano, M. Kinoshita, A. Kawabata, M. Morita, Y. Oda, R. Kuroda and K. Kakehi, *Specific distribution of sialic acids in animal tissues as examined by LC-ESI-MS after derivatization with 1,2-diamino-4,5-methylenedioxybenzene*. Anal Chem *73,* 5422-5428 (2001).

[18] J. Mallorqui, J. Segura, C. de Bolos, R. Gutiérrez-Gallego and J. A. Pascual, *Recombinant erythropoietin found in seized blood bags from sportsmen*. Haematologica *93,* 313-314 (2008).

[19] J. C. Bigge, T. P. Patel, J. A. Bruce, P. N. Goulding, S. M. Charles and R. B. Parekh, *Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid*. Anal Biochem *230,* 229-238 (1995).

[20] C. H. Hokke, A. A. Bergwerff, G. W. van Dedem, J. van Oostrum, J. P. Kamerling and J. F. Vliegenthart, *Sialylated carbohydrate chains of recombinant human glycoproteins expressed in Chinese hamster ovary cells contain traces of N-glycolylneuraminic acid*. FEBS Lett *275,* 9-14 (1990).

[21] E. Llop, R. Gutiérrez-Gallego, J. Segura, J. Mallorqui and J. A. Pascual, *Structural analysis of the glycosylation of gene-activated erythropoietin (epoetin delta, Dynepo)*. Anal Biochem (2008). PMID: 18804089

[22] P. Tangvoranuntakul, P. Gagneux, S. Diaz, M. Bardor, N. Varki, A. Varki and E. Muchmore, *Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid*. Proc Natl Acad Sci U S A *100,* 12045-12050 (2003).

[23] W. Jelkmann, *Developments in the therapeutic use of erythropoiesis stimulating agents*. Br J Haematol *141,* 287-297 (2008).

[24] F. Lasne, L. Martin, N. Crepin and J. de Ceaurriz, *Detection of isoelectric profiles of erythropoietin in urine: differentiation of natural and administered recombinant hormones*. Anal Biochem *311,* 119-126 (2002).

[25] M. Thevis and W. Schanzer, *Mass spectrometry in sports drug testing: Structure characterization and analytical assays*. Mass Spectrom Rev *26,* 79-107 (2007).

[26] E. Llop, R. Gutiérrez-Gallego, V. Belalcazar, G. J. Gerwig, J. P. Kamerling, J. Segura and J. Pascual. *Structural features of EPO and analogues: glycosylation profiles*. Schanzer, W., Geyer, H., Gotzmann, A, and Mareck, U. Recent Advances in Doping Analysis. Proceedings of the Manfred Donike Workshop, *12*, 297-308 (2004).

[27] K. R. Anumula, *Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates*. Anal Biochem *350,* 1-23 (2006).

[28] S. Hara, Y. Takemori, M. Yamaguchi, M. Nakamura and Y. Ohkura, *Fluorometric high-performance liquid chromatography of N-acetyl- and N-glycolylneuraminic acids and its application to their microdetermination in human and animal sera, glycoproteins, and glycolipids*. Anal Biochem *164,* 138-145 (1987).

[29] S. Hara, M. Yamaguchi, Y. Takemori, M. Nakamura and Y. Ohkura, *Highly sensitive determination of N-acetyl- and N-glycolylneuraminic acids in human serum and urine and rat serum by reversed-phase liquid chromatography with fluorescence detection*. J Chromatogr *377,* 111-119 (1986).

[30] J. Mallorqui, J. Segura, C. de Bolos, R. Gutiérrez-Gallego and J. A. Pascual, *Reply to: [Comment to: Recombinant erythropoietin found in seized blood bags from sportsmen. Haematologica 2008;93:313-4*. Haematologica *93,* e56 (2008).

[31] C. Sheridan, *Commercial interest grows in glycan analysis*. Nat Biotechnol *25,* 145-146 (2007).

4. Discussion

In the development of this thesis the structural characterisation of the glycans of four different, but very closely related erythropoietin preparations, have been addressed. First, the variables affecting the IEF profiles of recombinant erythropoietin produced in CHO cells (rEPO α/β) endogenous erythropoietin obtained from urine samples (uEPO) and the hyper-glycosylated recombinant erythropoietin analogue named NESP (also produced in CHO cells) were studied. Given the fact that rEPO is purified to maximise sialic acid content [1], the bands corresponding to rEPO should be close to the maximum of 14 residues. A similar evaluation was done for NESP because it is also a therapeutic drug and enriched for the highest number of sialic acids, the maximum being 22 [2]. Given the terminal location of sialic acids in glycans, the maximum number of these residues reported for erythropoietin is 14, meaning that the IEF bands for uEPO cannot be explained by sialic acids alone [3]. From the location of uEPO bands in IEF gels, additional negative charges are hypothesised. This lack of insight, added to the emergence of profiles shifted towards less acidic pI values (so called atypical profiles) generated the necessity of studying factors that could potentially affect the erythropoietin isoform distribution. After a careful evaluation of the possible charge-attributing residues in proteins a strategy was devised to unravel the origin of charges contained in the individual IEF bands. Digestions were performed with several exo-enzymes including phosphatases, glucuronidases, sulphatases, sialidases and endo-enzymes such as PNGase F. In addition, and wherever possible chemical treatments were performed for confirmation. From those experiments, it could be concluded that the glycans of recombinant molecules contain exclusively sialic acids as charge-conferring entities and that the endogenous material contains additional charged structures. The differences are located in the glycans as de-N-glycosylation renders equivalent IEF profiles for both endogenous and exogenous material but no further speciation could be indicated as specific enzymes targeting glycan-modifications (such as Man-phosphatase or a GlcNAc-sulphatase) were not available. An intriguing finding resulted from degradation of rEPO and NESP with

a sulphatase from *Helix Pomatia*. With this sulphatase, that addresses sulphates attached to aromatic ring structures, only recombinant molecules gradually loose part of the charge in a process, resembling a partial desialylation. As other sulphatase preparations did not caused this effect, it could be postulated that an enzyme activity similar to sialidase present in commercial preparation of sulphatase from *Helix Pomatia* could be responsible for the altered IEF profiles in biological samples. In the absence of any efficient sialidase inhibitor, an effective remedy based on the addition of a competing substrate (sialyl-lactose or *p*-nitrocathecol sulphate) to samples was proposed to solve the instability of urine samples in doping control.

Once the differences between recombinant and endogenous erythropoietin were evidenced, proteins were analysed in more detail. First of all, peptide mass fingerprints were performed in order to evaluate all possible post-translational modifications located in the polypeptide chains that could not be addressed by enzymatic treatments. After this, systematic structural elucidation of EPO-glycans was conducted, allowing the complete characterisation of the glycoprotein. Given the low concentration of EPO found in biological fluids, as well as the low degree of purity of the endogenous international standard from NIBSC, purification of urinary erythropoietin for structural analyses resulted unviable. In the first attempt to characterise uEPO glycans, unreliable data were obtained as glycans coming from other urinary glycoproteins contained in the international standard contaminated the results. Then, serious efforts were put into the development of purification strategies in order to isolate urinary erythropoietin. Exploration of immunoaffinity methods using different monoclonal and polyclonal antibodies, different immobilisation strategies as well as alternative purification platforms resulted in encouraging data. However, more efforts are required to up-scale this technique in order to obtain sufficient high purity uEPO for complete characterisation. The unique structural data available for urinary erythropoietin glycosylation arise from 30-year old papers where hundreds of liters of urine were purified. Glycan structures described for uEPO were similar to those reported

later on for rEPOs, however these data seem not to fully agree with the present findings of lower molecular weight and higher acidity of uEPO as compared to rEPO. In general, individuals from which urinary erythropoietin was purified for glycan analyses suffered from aplastic anemia, possibly resulting in altered glycosylation patterns that should not be necessarily representative of human EPO glycan profiles. In this thesis, as a part of the ongoing research on EPO, methodology for erythropoietin glycoforms characterisation was developed, optimised, and subsequently applied to structural analyses of recombinant variants. Peptide mass fingerprint of rEPO α/β, the newly marketed recombinant EPO produced in human cells rEPO δ (dynepo^{TM}) and the analogue NESP in which 5 aa have been modified, allowed concluding that both rEPO preparations are identical from the peptide point of view. As expected, they comprise a 165 aa polypeptide chain with two disulfide bridges, three N-glycans and one O-glycan. Additional glycosylation sites were discarded as well as other post-translational modifications occurring in specific aa; as for example methylation of Cys; phosphorylation (Tyr, Thr, Ser) and sulphation (Cys, Tyr). These data corroborated those obtained from earlier enzymatic studies (*vide supra*). The peptide analysis of NESP confirmed the 5 mutated amino acids and the two additional glycosylation sites. Site occupancy could be verified through the transformation of the asparagine residue from which the glycan is removed, into aspartic acid after de-N-glycosylation. This conversion is easily detected by its mass difference of 0.98 Da in the mass spectrum. One important observation derived from peptide and protein analyses was the identification of a PNGase F resistant N-glycosylation site of NESP. Without prior reduction and alkylation, trypsinolysis only rendered peptide T77-97 containing the N-glycosylation sites at Asn83 and Asn88. It was therefore supposed that the PNGase F resistant glycans should be located at one of the three other possible N-glycosylation sites. Then a reduction and alkylation protocol was optimised, and tryptic digestion of the reduced/alkylated and de-N-glycosylated NESP revealed a new peptide corresponding to T21-45 bearing the N-glycosylations sites Asn24, Asn30 and

Asn38. As this phenomenon of partial de-N-glycosylation was not observed in rEPO samples, it was speculated that resistant N-glycosylation site was Asn30. Moreover, this glycosylation site is located between Cys29 and Cys33 that is interlaced by a disulfide bond to form a small loop in the 3-D structure of the native molecule. Despite this evidence, peptide sequencing should be performed to unambiguously demonstrate that Asn30 is, in fact, the PNGase F resistant site in NESP. Comparison of the rEPOs and NESP peptide analyses after de-N-glycosylation revealed only two different tryptic peptides T21-45 and T77-97 containing the 5 aa mutated and the two additional N-glycosylation sites. These findings may be used for the development of a mass spectrometric method for detection of rEPO and NESP in biological fluids, necessary tools for antidoping purposes. During our investigations, peptide analysis turned out to be very useful for the structural characterisation of the single O-glycan contained in the rEPOs and NESP since no enzymatic release of intact O-glycan is possible. A detailed structural analysis of the O-glycoforms disclosed important differences between the recombinant products. The di-sialylated O-glycan is the only one present in rEPO δ whereas in NESP and rEPO α/β a different proportion of non-sialylated and mono-sialylated O-glycoforms were shown. Analysis of O-glycans had been previously performed using MALDI-TOF MS of the de-N-glycosylated proteins. Despite the poor spectral resolution obtained for the entire O-glycoprotein, this approach revealed three O-glycoforms in the three samples. Although differences between relative intensity of sialylation degree of rEPO α/β, rEPO δ and NESP could be observed, results from MALDI-TOF MS were not in total agreement with LC-ESI MS data obtained for O-glycoproteins and O-glycopeptides. The evaluations of both techniques allowed concluding that more reliable results arise from LC-ESI MS because this soft ionisation minimises sample de-sialylation. Although it is stated that the O-glycan appears to lack importance with respect to the *in vitro* and *in vivo* activity of EPO, these analyses could be exploited for differentiating recombinant preparations and possibly the endogenous EPO in which more variability is expected. Interestingly, Dordal *et al.* [4] reported than O-

glycan was absent in urinary EPO and Skibeli *et al*. [5] found that a substantial fraction of human serum EPO lacks O-glycosylation on Ser126.

Once the recombinant erythropoietins were characterised at peptide level, and following the starting hypothesis, N-glycosylation was investigated. The initial work-plan aimed at the development of methodologies for sugar profiling and mass spectrometric characterisation. Then, the application of those protocols to total glycan pool from rEPO α/β, rEPO δ and NESP, and finally the elucidation of the glycan composition of each individual glycoform of the three samples after separation by isoelectric focusing. Regarding N-glycosylation analyses, several approaches combining SDS-PAGE, IEF and 2-DE with HPLC-FLD, MALDI-TOF MS and UPLC-ESI-TOF MS were performed [6]. Average molecular weight was calculated by MALDI-TOF MS although an approximate estimation was also done by SDS-PAGE. While NESP showed the highest molecular weight by both techniques due to the additional N-glycans, mass spectra obtained from rEPO α/β yielded a slightly lower average mass than rEPO δ. Conversely, in SDS the latter migrated faster indicating lower apparent Mw. The laser intensity during the MALDI ionisation process showed to be extremely important as mass variations were produced by fragmentation of labile residues such as sialic acids [7, 8]. Another important factor to be considered in Mw determination is the possibility of sodium and potassium adduct formation during ionisation and/or desorption process in MALDI-TOF MS because they could partially modify the average mass [9]. Despite the particular instrumental variations, other authors reported similar effects when rEPO α/β and rEPO δ were compared by SDS-PAGE and MALDI-TOF MS [10]. This difference added to the lower Mw reported previously for endogenous urinary EPO [11] is currently being exploited by doping control laboratories as a screening procedure for differentiating endogenous EPO from the recombinant analogues. However, while NESP can be clearly differentiated due to its higher molecular weight, more efforts must be dedicated to improve the SDS resolution allowing unequivocal differentiation between rEPOs and uEPO. The microheterogeneity of different EPOs could be also inferred from SDS-PAGE

and MALDI-TOF MS. As expected, NESP showed the broadest peak in mass spectra and the most diffuse band gel in SDS. Interestingly, rEPO δ yielded the narrowest peak and the sharpest band in gel electrophoresis [10]. More homogeneous carbohydrate structures were supposed for rEPO δ but surprisingly, its IEF profile showed three additional bands, absent in its homologuous rEPO α/β.

In order to enable an unambiguous discrimination between different EPO preparations, unique structural features of each product needed to be identified. In this thesis, first the overall carbohydrate content and then the constituents released from intact glycoproteins were analysed. Finally, attention was paid to the structural elucidation of each individual glycoform after separation by 2-DE. Comparison of the sugar profiles from rEPO α/β, rEPO δ and NESP revealed significant differences. The most characteristic feature of NESP was its most extensive acetylation with up to two O-acetyl groups per sialic acid residue, while in case of rEPO α/β this modification was less significant and in rEPO δ it was completely absent. O-acetylation of sialic acids glycan structures is well known to play a regulatory role in several metabolic processes, particularly inhibition of neuraminidase activity, leading to a reduced hepatic uptake and subsequent increase in glycoprotein circulation time. This phenomenon will certainly contribute to the long half-life of NESP and may have been introduced deliberately. O-acetylation is easily detected by MALDI-TOF MS through the mass increment of 42 Da, and might serve as potential biomarker for EPO α/β and NESP administration. Several studies have shown that different erythropoietin isoforms have different biological activity, being the less acidic isoforms those with a shorter half-life and lower activity.

In order to evaluate the charged residues in EPO are concerned, WAX analyses were performed [12]. Results revealed similar charges contained in N-glycans from rEPO α/β and rEPO δ, but significant charge differences in the O-glycans, being higher for rEPO δ. NESP N-glycans resulted to contain higher sialic acid

content than rEPOs whereas its O-glycan contributes more than in rEPO α/β but less than in rEPO δ to the total glycan charge.

Differences in microheterogeneity were also found, despite the fact that all this biopharmaceuticals are specifically purified for the maximum sialic acid content. Comparison of sugar profiles indicated more homogeneous structures for rEPO δ containing predominantly tetraantennary structures, some triantennary N-glycans and traces of LacNac repeats. Next in heterogeneity was NESP containing a glyco repertoire with a high proportion of LacNac repeats and triantennary N-glycans. Finally rEPO α/β showed the highest heterogeneity, containing, on top of the structures mentioned for its analogues, diantennary structures and higher proportion of LacNac repeats (up to 3 units per tetraantennary structure). Carbohydrates are known to play an important role in protecting the peptide core of glycoproteins from the action of proteases and in determining biological activity. In particular, the highly branched N-glycans stabilise the conformation of glycoproteins, and prevent EPO from renal clearance [13]. It has been reported that truncated rEPO with di- instead of tetraantennary glycans is secreted into the urine in greater amounts [14]. Considering the possibility that EPO species with an increased molecular weight might have a reduced rate of clearance, Sytkowski *et al.* [15] showed that chemically linked EPO dimers or recombinant fusion proteins containing two EPO molecules exhibited a prolonged plasma half-life and stronger erythropoiesis-stimulating activity than monomers [16]. Further work is required, however, to identify the pathways of metabolism and elimination of this glycoprotein hormone. Data obtained from excretion studies performed in our laboratory, showed that urinary EPO profiles of individuals administered with rEPO are identical to the rEPO drug when endogenous EPO production has been suppressed. Contrary to what happens in rEPO, urinary erythropoietin is thought to represent a special fraction of the hormone that has escaped from circulation. A different IEF profile has been reported for endogenous EPO from urine and plasma, being the former more acidic [5, 17, 18]. However more acidic glycoforms seems to represent a higher half-life [19]. Although the

pharmacokinetic properties of rEPOs have been extensively studied [20], important discrepancies exist about the main site and mechanism of the removal of EPO from circulation. The lack of knowledge concerning the metabolic fate of EPO hampers the understanding of endogenous EPO differences so more investigations on this topic are needed.

The next step in elucidating rEPOs glycan composition was the application of the reagent array analysis method (RAAM). These sequential digestions with different exoglycosidases allowed arriving to the conclusion that no linkage differences exist between the glycans of all three glycoproteins. The fucosylated trimannosyl core $Man_3GlcNAc_2$ is shared by all N-glycans belonging to rEPO α/β, rEPO δ and NESP but there are differences in the sialic acid content and the branching level. In contrast to human cells, CHO cells do not express sialyl-α2-6 transferase [14, 21], so the sialic acids contained in rEPOs produced in these cells are exclusively added to galactose via α2-3 linkage [22]. However, several papers have been published addressing the issue in which CHO cells have been engineered to resemble the human glycan profile by transfecting the appropriate sialyl-transferase [23]. Our charge analyses after specific α2-3 sialidase confirmed that all N-glycans from the two drugs synthesised in CHO cells contained exclusively sialic acids linked in that position. CHO cells are also incapable of producing Le^x oligosaccharides (lack of α1-3/4 fucosyl transferase) and bisecting *N*-acetylglucosamine transferase [24]. Several reports on endogenous urinary EPO have suggested the presence of typical carbohydrate features, absent in rEPOs produced in CHO cells, as for example the occurrence of a Galβ1-3GlcNAcβ1-4Manα1-3 motif instead of Galβ1-4GlcNAcβ1-4Manα1-3 [25] or the presence of Neu5Acα2-6Gal linkage [26]. Inconsistencies exist with respect to the type of sialic acids linkage in uEPO glycans; Sasaki *et al.* found evidence for the exclusive existence of α2-3 linkages [27], whereas Takeuchi demonstrated the occurrence of both α2-3 and α2-6 linkages [26]. It was reported that absence of α2-6 linkages may enhance the clearance of a protein from the

bloodstream so special attention should be paid to this issue. With respect to the α1-6 fucosylation of the Asn-bound GlcNAc residues in uEPO, different percentages have been reported [24], but we found quantitative fucosylation in all rEPOs. The occurrence of sulphate groups in urinary EPO has been suggested by several groups including ours but structural studies have not been published thus far. It has been reported that CHO-cell-derived rEPO also possesses sulphated oligosaccharides [28], although it is stated elsewhere that these cells are not capable of such sulphation [29]. At branching level, our results show absence of LacNAc repeating units in the EPO produced in human cells. This seems to be in agreement with studies published by Tsuda *et al*. [25] and Takeuchi *et al.* [26] where a clear tendency towards higher LacNAc repeats content is displayed in CHO and BHK rEPOs. This might be related with the fact that uEPO loaded in a SDS-gel runs faster than its recombinant analogues [10]. Glycans belonging to rEPO δ did not show charges different from sialic acids Further work is required to understand the IEF profile of endogenous EPO, and to identify the metabolic pathways and pharmacodynamics of this glycoprotein hormone. Several reports on endogenous EPO have shown that circulating EPO contains fewer acidic glycoforms than urinary EPO, speculating that the charge difference could be attributed to a difference in renal handling of the various glycoforms or post-secretion processing of the glycans [14]. However glycan structures responsible for those differences have never been reported.

The current work addresses also the characterisation of glycans present in individual, IEF separated, isoforms of these recombinant erythropoietins. Comparison of glycan structures from different samples confirmed that co-migrating isoforms contained identical number of sialic acids. Nevertheless, each band includes the microheterogeneity in the branching (number of antennas, LacNAc repeats, and acetyl-subforms) also observed for the total pool. The results further indicated that the more acidic the the pI of the IEF band, the higher the level of sialylation. A clear tendency towards increasing tetraantennary tetrasialylated structures with increasing acidity of the bands was shared by the

three drugs. This effect was more easily observed in rEPO δ due to its more homogenous glycan structures. The extensive O-acetylation and the presence of triantennary structures of NESP or LacNac repeats and diantennary structures in rEPO α/β made these analyses less evident. Mass spectrometry should be the basis for identifying any mass differences produced by unusual monosaccharide constituents and to unravel the ultimate monosaccharide composition of glycans. MALDI-TOF MS was the better choice both for total glycans pool and individual IEF separated isoforms. This technique is relatively tolerant to the presence of contaminants and gives a quantitative response with negligible fragmentation for neutral N-glycans of varying structures. Unfortunately, rEPO glycans are highly (negatively) charged, making a quantitative analysis very difficult. In glycan analyses of highly sialylated structures, a considerable fragmentation, both in source and post-source has been reported, resulting in the loss of sialic acids [7, 8, 30]. Given that most of the heterogeneity in rEPO glycans resides in the sialic acid content, intact characterisation of intact structures was needed. Thus, efforts were made for optimising instrument parameters to minimise desialylation. Comparison of glycans released from different rEPOs and glycans contained in separated isoforms of the same rEPO sample showed sialylation differences indicating that either sialic acids loss did not occur or this was minimal [30]. Quantities of individual glycan structures present in a single band were estimated using peak height measured, smoothing to remove isotopic resolution or summation of the peak heights of all isotopes. From percentages of structures present in a single band, the charge contents were calculated and compared with the values obtained from WAX analyses. The average number of sialic acids contained in N-glycans seemed to be two units higher when this data is calculated by integrating peaks obtained by weak anion exchange profiling. In this technique, N-glycans are derivatised before analysis. As one single fluorescent tag is linked to each N-glycan, relative quantification is far more robust than when using mass spectrometry. The labelling protocol to introduce the fluorophore at the reducing-end sugar is carried out in slightly acidic

conditions. Bigge *et al.* [31] claim that 2 h at 65ºC yields an optimum between labelling and desialylation (only 2%). Nevertheless Watanabe *et al.* [32] reported a significant loss of sialic acids and proposed lowering the temperature to 37 ºC and prolonging the incubation to 16 h in order keep the labelling yield. In our hands, and employing a mixture of tetraantennary, multisialylated structures, both protocols rendered similar labelling yields and structure preservation so the first was used in our analyses paying special attention to preservation of sialic acids. However, some degree of desialylation cannot be discarded causing slight underestimation of sialic acids in WAX analyses. For all these reasons, MALDI-TOF MS analyses were only performed for identification purposes and quantification of sialic acids was performed using peak integration after WAX analyses. Surprisingly, the sialic acid average contained in N-glycans resulted to be similar for rEPO δ (9.73) and rEPO α/β (9.83). Hence, summing up O- and N-glycan contribution of sialic acids rEPO δ contains 11.73 as a mean of the glycoprotein and rEPO α/β contains 11.39. Despite the two additional acidic isoforms in rEPO δ, its mean sialic acid content is only slightly higher. This can be explained by the relative proportion of these bands, and from the balancing effect of the more basic isoforms (band 6 and 7) that appear more abundant in rEPO δ. The mean number of sialic acids obtained from both rEPOs is in agreement with European Pharmacopoeia requirements, consisting of values exceeding 10 mol/mol. Sialic acids calculation in the NESP total glycans resulted in 18.02 for the N-glycans plus 1.75 for its O-glycans, giving a mean of 19.78 sialic acids. When rEPO α/β from the European Pharmacopoeia is focused in IEF gel, 7 bands are clearly visible. Considering that the theoretical maximum number of sialic acids in rEPO is 14, and the calculated mean value is 11.39, the most acidic band should contain close to 14 sialic acids. For rEPOs all possible sialic acids combination from 3 N-glycans and 1 O-glycan containing from 0 to 4 and from 0 to 2 sialic acids respectively were calculated. Data resulted in one unique combination for obtaining a total of 14 sialic acids and 4, 10, 19, 31, 43 and 52 combinations to obtain 13, 12, 11, 10, 9 and 8 sialic acids respectively. Following

theoretical predictions, the most acidic band should contain tetraantennary tetrasialylated N-glycans only. Experimental data, however, revealed different composition with presence of triantennary structures in the most acidic band of rEPO. In rEPO δ where two more acidic bands are shown, only the first one (band α) could be analysed by MALDI-TOF MS. This analysis showed only tetraantennary N-glycans but containing from 2 to 4 sialic acids. Certain degree of desialylation could not be discarded. However a confirmatory WAX analysis could not be performed because of the insufficient amount of material recovered from single bands. Assuming 14 sialic acid for α-band, then β-band should contain additional charges other than sialic acids similarly as for uEPO. By the observation of uEPO profile in IEF gel (bands located between rEPO α/β and NESP), our starting hypothesis argumenting additional charges is fully substantiated. Further analyses must be done in order to ascertain charges present in uEPO for better understanding of erythropoietin function and metabolism as well as for unequivocal discrimination between rEPOs and endogenous EPO in doping control [33].

Once, entire glycans were analysed, particular attention was paid to the particular sialic acid species present in the different rEPOs samples as these carbohydrate moieties influence the biological and physical properties of biopharmaceutical proteins [34]. Despite the more than 60 natural analogues of sialic acid described, *N*-acetylneuraminic acid containing an acetyl substitution at C5 is one of the most common sialic acid derivatives [35]. Regarding sialic acid analyses of rEPOs, *N*-glycolylneuraminic acid must be also considered since the majority of rEPOs are synthesised by CHO or BHK cells and therefore might also contain this species. This sialic acid occurs frequently in animal cells, but is absent in humans because of an internal frame-shift mutation in the gene that encodes CMP-Neu5Ac hydroxylase. The Neu5Gc is essentially foreign and, hence, monitoring its content in biopharmaceutical samples could be very important. The primary epitope on this moiety, which is referred to as the Hanganutziu-Deicher antigen, might stimulate an immune response in the host. In 2005, Martin *et al.*

[36] demonstrated that when human embryonic stem cells are cultured in a specialised medium that is derived from animal sources, the exogenous Neu5Gc residues might be introduced into the host. The consumption of animal products, such as red meat and milk, could also introduce traces of this residue into human proteins. As the addition of sialic acids has been described to enhance the therapeutic properties of EPO and a variety of biopharmaceutical proteins [37], the versatility of sialic acids, coupled with developments in sialyltransferase engineering, should enable more glycoproteins to be hypersialylated and, subsequently, evaluated for clinical studies. The optimisation of protocols for the large-scale production of hypersialylated proteins in mammalian cell lines, as for example the erythropoietin analogue NESP [38], should be considered as a mechanism for generating large amounts of highly effective pharmaceutical preparations. However, the regulation of the Neu5Gc content needs to be addressed with reference to the potential immunogenicity associated with the introduction of multiple glycan chains that contain this sialic acid residue. Sialic acids contained in rEPO α/β, rEPO δ and NESP were evaluated using the method previously described by Hokke *et al.* [39]. This method consists of a sialic acids hydrolysis prior to derivatisation with a fluorescence group. Detection based on fluorescence offers the best choice for chromatographic determination of carbohydrates at high sensitivity [40]. Despite the good results obtained, downscaling of this methodology using a capillary HPLC coupled to a capillary fluorescence detector was necessary to improve the limit of detection in one order of magnitude. In this thesis, we propose a general strategy for simultaneous determination of sialic acids using both, HPLC-FLD and UPLC-ESI-MS. Although mass values yield useful information for unequivocal discrimination between Neu5Gc and Neu5Ac in rEPOs, this method can only be applied to biological samples with supraphisiological levels. The three rEPOs evaluated showed Neu5Ac as the major constituent but significant differences in the degree of O-acetylations could be also identified. These modifications resulted to be minimal in rEPO δ, more abundant in rEPO α/β and extensive in NESP.

Nevertheless, the identification of the exact locations of the O-acetyl linkage is somewhat cumbersome due to the ability of these substituents to migrate from one hydroxyl to another along the alkyl structure. Despite the care taken in sample handling to avoid migration of O-acetyl groups, the heterogeneity observed in replicates could still be affected by the specific batch of rEPO and/or sample treatment prior to analysis. As expected, rEPO δ produced in human cells (devoid of the corresponding hydroxylase) did not contain Neu5Gc while rEPO α/β and NESP yielded 1.3 mol% and 1.5 mol%, respectively. As the sensitivity of our analytical method was still not good enough for detecting the expected concentrations of Neu5Gc in 20 mL of urine samples from athletes, plasma samples from patients treated with rEPO were employed. These samples were immunopurified prior to sialic acid analyses and results revealed absence of Neu5Gc in blank plasma used as a negative control. In contrast, when this plasma sample was spiked with rEPO, its chromatogram showed this compound. In a real sample showing a IEF profile consistent with the presence of rEPO, the variant Neu5Gc could also be detected. The presence of Neu5Gc in immunopurified biological samples would indicate the presence of exogenous erythropoietin. However this methodology would not be applicable to rEPOs synthesised in human cells like rEPO δ and further analyses are necessary to fully discard false positives produced by incorporation of Neu5Gc from diet to newly synthesised glycoproteins.

Taken together, these findings may indicate that these analyses (peptide mapping + N-glycans + O-glycan + sialic acids analyses) should be used to monitor the production process of recombinant EPOs. This is justified because rEPOs, even if engineered in the same cell line (e.g. CHO cells) will never be identical among different manufacturers. In fact, differences in IEF behaviour and in physicochemical properties were detected when 11 samples of CHO cell-derived rEPOs from eight manufacturers outside the EU and the USA were investigated [41]. Differences between batches of the products from the same manufacturer were also evidenced. In contrast to other drugs, chemically

synthesised, rEPOs are complex therapeutic proteins produced by living organisms and cannot be adequately characterised using current analytical methods. In our opinion complete characterisation of these biopharmaceuticals could help in detection of damaged preparations (e.g. oxidation, deamidation, missfolding, aggregation or most importantly contamination with bacterial polysaccharides and proteins other than rEPOs coming from host cells) that may produce immunogenic responses in patients [42]. According to the rEPO-specific annex to the EMEA guidelines (Committee for Medicinal Products for Human Use), *"biosimilar manufacturers must apply a single reference product and conduct tests to demonstrate biophysical similarity, non-clinical pharmacodynamics and toxicological data, and data showing the clinical similarity to the reference product".* Biosimilar rEPOs have to be tested in at least two efficacy studies in chronic kidney disease patients, including a correction phase study in erythropoietin-naive patients (s.c. administration of the drug) and a 6-month maintenance phase study in patients on dialysis who are already treated with an erythropoietin (i.v. administration). In addition, it is likely that, even if new biosimilar products were immunogeneic, this would probably not be detected during the short-term clinical safety studies required because ESA-induced antibody formation and pure red cell aplasia (PRCA) become obvious only after a median treatment period of 9 months. Biopharmaceutical innovators accumulate experience in the production and purification process to select the isoforms with high biological activity or lower immunogenic potential [43]. Furthermore, in the light of the medical and economic success, structural data concerning rEPO glycosylation published by scientific community promote the creation of biopharmaceutical congeners and synthetic erythropoiesis stimulating compounds [44]. At this moment, only sialic acids analyses, IEF profiles and Mw determinations by western blot can be applied to biological samples being insufficient for structural characterisation. The sensitivity limitations of analytical instruments and the low amounts of EPO present in biological fluids (~10 ng/L in

urine and ~100 ng/L in serum) [45] are the main reasons for failing in revisiting structural investigations on endogenous EPO.

EPO is now considered to have applicability in a variety of nervous system disorders that can overlap with vascular disease, metabolic impairments, and immune system function [46]. As a result, EPO may offer palliative relieve for a broad number of disorders including Alzheimer's disease, cardiac insufficiency, stroke, trauma, and diabetic complications. In a number of clinical conditions, EPO has shown to prevent neuronal and vascular degeneration, and reduce inflammatory cell activation [47]. Recent works have elucidated a number of novel cellular pathways governed by EPO that can open new avenues to avert deleterious effects of this agent and offer previously unrecognised perspectives for therapeutic strategies. In this context, structural studies as those reported in this thesis should be related with the role of EPO in different cellular targets. Recently, studies have revealed more about the complexities of differential post-translational modifications of human proteins and the potential implications on different functions in the human body. New findings suggest that differential glycosylation patterns of human proteins that occur in the blood (so-called "serum-type" modifications) are different from the ones that occur in the cerebrospinal fluid in the brain ("brain-type" modifications). This difference may be of paramount importance for the design of effective therapeutics. In the case of erythropoietin, one can also refer to a serum-type EPO (or a "renal-type" or a "urinary-type" EPO) for the protein that is produced in the kidney and that circulates in the blood, as compared to EPO that is been produced by other tissues such as the brain. The characterisation of individual glycoforms will help to decide host cell lines and genetic engineering needed to obtain glycoproducts with novel or improved *in vivo* properties. One interesting example is rEPO because only a highly sialylated subfraction (based on the IEF pattern) representing ~20% of the total recombinant glycoprotein is manufactured for medical treatment. These preparations usually obtained from CHO, BHK and HT-1080 cells are very useful in the treatment of chronic kidney disease in which

long half-life is required. However, glycoforms contained in these drugs are relatively useless in the treatment of disorders related to the central or peripheral nervous system as well as in the treatment of afflictions related to ischemia/reperfusion induced disorders. Furthermore, it has been found, that glycoforms having on average a lower sialic acid residue are more effective for the treatment of non-hematopoietic diseases due to their higher receptor affinity and high dosage can be used to limit the adverse side effects of erythropoiesis because these glycoforms are rapidly cleared from the blood. Other aspects of the EPO glycoforms exhibiting anti-inflammatory and immunosuppressive activities are hyperfucosylation, and, hence, most likely decorated extensively with Le^x epitopes.

Human cells named PER.C6™ have been proposed as a suitable but not unique expression system for recombinant erythropoietin containing above mentioned glycoforms. By using methodology described in articles presented previously, we are now able to efficiently characterise and select the glycoforms of a rEPO pharmaceutical based not only in their sialic acid content but also LacNac repeats and fucosylation degree. However, we need to understand in much more detail basic glycosylation structure of different types of endogenous erythropoietin (urinary, plasmatic, neural etc.) and the relationship between the structure and the function of these erythropoietin variants.

Bibliography

[1] W. Schlags, B. Lachmann, M. Walther, M. Kratzel and C. R. Noe, *Two-dimensional electrophoresis of recombinant human erythropoietin: a future method for the European Pharmacopoeia?* Proteomics *2,* 679-682 (2002).

[2] J. C. Egrie and J. K. Browne, *Development and characterization of novel erythropoiesis stimulating protein (NESP)*. Nephrol Dial Transplant *16 Suppl 3,* 3-13 (2001).

[3] V. Belalcazar, R. Gutiérrez-Gallego, E. Llop, J. Segura and J. A. Pascual, *Assessing the instability of the isoelectric focusing patterns of erythropoietin in urine*. Electrophoresis *27,* 4387-4395 (2006).

[4] M. S. Dordal, F. F. Wang and E. Goldwasser, *The role of carbohydrate in erythropoietin action*. Endocrinology *116,* 2293-2299 (1985).

[5] V. Skibeli, G. Nissen-Lie and P. Torjesen, *Sugar profiling proves that human serum erythropoietin differs from recombinant human erythropoietin*. Blood *98,* 3626-3634 (2001).

[6] E. Llop, R. Gutiérrez-Gallego, V. Belalcazar, G. J. Gerwig, J. P. Kamerling, J. Segura and J. A. Pascual, *Evaluation of protein N-glycosylation in 2-DE: Erythropoietin as a study case*. Proteomics *7,* 4278-4291 (2007).

[7] E. Gimenez, F. Benavente, J. Barbosa and V. Sanz-Nebot, *Towards a reliable molecular mass determination of intact glycoproteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*. Rapid Commun Mass Spectrom *21,* 2555-2563 (2007).

[8] D. J. Harvey, *Proteomic analysis of glycosylation: structural determination of N- and O-linked glycans by mass spectrometry*. Expert Rev Proteomics *2,* 87-101 (2005).

[9] G. Stubiger, M. Marchetti, M. Nagano, C. Reichel, G. Gmeiner and G. Allmaier, *Characterisation of intact recombinant human erythropoietins applied in doping by means of planar gel electrophoretic techniques and matrix-assisted laser desorption/ionisation linear time-of-flight mass spectrometry*. Rapid Commun Mass Spectrom *19,* 728-742 (2005).

[10] M. Kohler, C. Ayotte, P. Desharnais, U. Flenker, S. Ludke, M. Thevis, E. Volker-Schanzer and W. Schanzer, *Discrimination of recombinant and endogenous urinary erythropoietin by calculating relative mobility values from SDS gels*. Int J Sports Med *29,* 1-6 (2008).

[11] A. Khan, J. Grinyer, S. T. Truong, E. J. Breen and N. H. Packer, *New urinary EPO drug testing method using two-dimensional gel electrophoresis*. Clin Chim Acta *358,* 119-130 (2005).

[12] E. Llop, R. Gutiérrez-Gallego, J. Segura, J. Mallorqui and J. A. Pascual, *Structural analysis of the glycosylation of gene-activated erythropoietin (epoetin delta, Dynepo)*. Anal Biochem (2008). PMID: 18804089

[13] W. Jelkmann, *The enigma of the metabolic fate of circulating erythropoietin (Epo) in view of the pharmacokinetics of the recombinant drugs rhEpo and NESP*. Eur J Haematol *69,* 265-274 (2002).

[14] T. Misaizu, S. Matsuki, T. W. Strickland, M. Takeuchi, A. Kobata and S. Takasaki, *Role of antennary structure of N-linked sugar chains in renal handling of recombinant human erythropoietin*. Blood *86,* 4097-4104 (1995).

[15] A. J. Sytkowski, E. D. Lunn, K. L. Davis, L. Feldman and S. Siekman, *Human erythropoietin dimers with markedly enhanced in vivo activity*. Proc Natl Acad Sci U S A *95,* 1184-1188 (1998).

[16] A. J. Sytkowski, E. D. Lunn, M. A. Risinger and K. L. Davis, *An erythropoietin fusion protein comprised of identical repeating domains exhibits enhanced biological properties*. J Biol Chem *274,* 24773-24778 (1999).

[17] F. Lasne, L. Martin, J. A. Martin and J. de Ceaurriz, *Isoelectric profiles of human erythropoietin are different in serum and urine*. Int J Biol Macromol *41,* 354-357 (2007).

[18] J. Mallorqui, J. Segura, C. de Bolos, R. Gutiérrez-Gallego and J. A. Pascual, *Recombinant erythropoietin found in seized blood bags from sportsmen*. Haematologica *93,* 313-314 (2008).

[19] M. N. Fukuda, H. Sasaki, L. Lopez and M. Fukuda, *Survival of recombinant erythropoietin in the circulation: the role of carbohydrates*. Blood *73,* 84-89 (1989).

[20] E. Varlet-Marie, A. Gaudard, M. Audran and F. Bressolle, *Pharmacokinetics/pharmacodynamics of recombinant human erythropoietins in doping control*. Sports Med *33,* 301-315 (2003).

[21] E. C. Svensson, B. Soreghan and J. C. Paulson, *Organization of the beta-galactoside alpha 2,6-sialyltransferase gene. Evidence for the transcriptional regulation of terminal glycosylation*. J Biol Chem *265,* 20863-20868 (1990).

[22] H. J. Kim, S. J. Lee and H. J. Kim, *Antibody-based enzyme-linked lectin assay (ABELLA) for the sialylated recombinant human erythropoietin present in culture supernatant*. J Pharm Biomed Anal (2008). PMID: 18722069

[23] E. Grabenhorst, P. Schlenke, S. Pohl, M. Nimtz and H. S. Conradt, *Genetic engineering of recombinant glycoproteins and the glycosylation pathway in mammalian host cells*. Glycoconj J *16,* 81-97 (1999).

[24] C. H. Hokke, A. A. Bergwerff, G. W. van Dedem, J. P. Kamerling and J. F. Vliegenthart, *Structural analysis of the sialylated N- and O-linked carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. Sialylation patterns and branch location of dimeric N-acetyllactosamine units*. Eur J Biochem *228,* 981-1008 (1995).

[25] E. Tsuda, M. Goto, A. Murakami, K. Akai, M. Ueda, G. Kawanishi, N. Takahashi, R. Sasaki, H. Chiba, H. Ishihara and ., *Comparative structural study of N-linked oligosaccharides of urinary and recombinant erythropoietins*. Biochemistry *27,* 5646-5654 (1988).

[26] M. Takeuchi, S. Takasaki, H. Miyazaki, T. Kato, S. Hoshi, N. Kochibe and A. Kobata, *Comparative study of the asparagine-linked sugar chains of human erythropoietins purified from urine and the culture medium of recombinant Chinese hamster ovary cells*. J Biol Chem *263,* 3657-3663 (1988).

[27] H. Sasaki, B. Bothner, A. Dell and M. Fukuda, *Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA*. J Biol Chem *262,* 12059-12076 (1987).

[28] N. Kawasaki, Y. Haishima, M. Ohta, S. Itoh, M. Hyuga, S. Hyuga and T. Hayakawa, *Structural analysis of sulfated N-linked oligosaccharides in erythropoietin*. Glycobiology *11,* 1043-1049 (2001).

[29] C. F. Goochee, M. J. Gramer, D. C. Andersen, J. B. Bahr and J. R. Rasmussen, *The oligosaccharides of glycoproteins: bioprocess factors affecting oligosaccharide structure and their effect on glycoprotein properties*. Biotechnology (N Y) *9,* 1347-1355 (1991).

[30] E. Mortz, T. Sareneva, I. Julkunen and P. Roepstorff, *Does matrix-assisted laser desorption/ionization mass spectrometry allow analysis of carbohydrate heterogeneity in glycoproteins? A study of natural human interferon-gamma*. J Mass Spectrom *31,* 1109-1118 (1996).

[31] J. C. Bigge, T. P. Patel, J. A. Bruce, P. N. Goulding, S. M. Charles and R. B. Parekh, *Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid*. Anal Biochem *230,* 229-238 (1995).

[32] T. Watanabe, N. Inoue, T. Kutsukake, S. Matsuki and M. Takeuchi, *Labeling conditions using a 2-aminobenzamide reagent for quantitative analysis of sialo-oligosaccharides*. Biol Pharm Bull *23,* 269-273 (2000).

[33] J. Segura, J. A. Pascual and R. Gutiérrez-Gallego, *Procedures for monitoring recombinant erythropoietin and analogues in doping control*. Anal Bioanal Chem *388,* 1521-1529 (2007).

[34] T. Dingermann, *Recombinant therapeutic proteins: production platforms and challenges*. Biotechnol J *3,* 90-97 (2008).

[35] J. P. Kamerling and G. J. Gerwig, *Structural analysis of naturally occurring sialic acids*. Methods Mol Biol *347,* 69-91 (2006).

[36] M. J. Martin, A. Muotri, F. Gage and A. Varki, *Human embryonic stem cells express an immunogenic nonhuman sialic acid*. Nat Med *11,* 228-232 (2005).

[37] I. C. Macdougall and K. U. Eckardt, *Novel strategies for stimulating erythropoiesis and potential new treatments for anaemia*. Lancet *368,* 947-953 (2006).

[38] J. C. Egrie and J. K. Browne, *Development and characterization of novel erythropoiesis stimulating protein (NESP)*. Br J Cancer *84 Suppl 1,* 3-10 (2001).

[39] C. H. Hokke, A. A. Bergwerff, G. W. van Dedem, J. van Oostrum, J. P. Kamerling and J. F. Vliegenthart, *Sialylated carbohydrate chains of recombinant human glycoproteins expressed in Chinese hamster ovary cells contain traces of N-glycolylneuraminic acid*. FEBS Lett *275,* 9-14 (1990).

[40] K. R. Anumula, *Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates*. Anal Biochem *350,* 1-23 (2006).

[41] H. Schellekens, *Biosimilar epoetins: how similar are they?* European Association of Hospital Pharmacists, Scientific Section *3,* 243-247 (2004).

[42] W. Jelkmann, *Recombinant EPO production--points the nephrologist should know*. Nephrol Dial Transplant *22,* 2749-2753 (2007).

[43] C. Sheridan, *Commercial interest grows in glycan analysis*. Nat Biotechnol *25,* 145-146 (2007).

[44] H. F. Bunn, *New agents that stimulate erythropoiesis*. Blood *109,* 868-873 (2007).

[45] F. Lasne, L. Martin, N. Crepin and J. de Ceaurriz, *Detection of isoelectric profiles of erythropoietin in urine: differentiation of natural and administered recombinant hormones*. Anal Biochem *311,* 119-126 (2002).

[46] K. Maiese, F. Li and Z. Z. Chong, *New avenues of exploration for erythropoietin*. JAMA *293,* 90-95 (2005).

[47] K. Maiese, F. Li and Z. Z. Chong, *Erythropoietin in the brain: can the promise to protect be fulfilled?* Trends Pharmacol Sci *25,* 577-583 (2004).

5. Conclusions

The results obtained in this thesis, which are included in all the scientific publications included in this manuscript, allow the following general conclusions to be extracted:

1. The charge profiles observed in IEF for rEPO α/β and NESP are due to the absolute number of sialic acid residues present whereas uEPO shows additional charges, different from sialic acids, confined in the carbohydrate moiety.

2. Combining the 2-DE gel electrophoresis with mass spectrometric and weak anion exchange analyses with fluorescence detection the percentages of N-glycans containing different sialic acids number that are present in individual IEF bands can be identified and quantified. Although both techniques allow calculating percentages of sialylated structures present in a single band, the later is preferred for more accurate quantification.

3. The additional charges contained in urinary EPO, inferred from its IEF profile have been proven to be different from sialic acids and localised within the carbohydrate moiety of the hormone. However, the hypothesis that these additional charges are due to sulphate substituents at galactose or *N*-acetylglucosamine residues could not be demonstrated.

4. Peptide mapping after de-N-glycosylation allows characterisation of the structures contained in the single O-glycosylation site contained in different rEPO preparations. The structural analysis of O-glycopeptides revealed the same glycan type (core 1) but significant differences in the sialic acid content.

5. In order to obtain a reliable glycan profiling of a glycoprotein complete deglycosylation is required. Hence, we concluded that sterical features of the native molecules, particularly disulfide bonding, may be responsible of incomplete deglycosylation of NESP. A reduction and alkylation protocol

is needed for complete de-N-glycosylation of the recombinant erythropoietin analogue NESP.

6. The characterisation of N-glycans by a combination of different approaches (MALDI-TOF-MS, HPLC-FLD, and RAAM) has proven to be capable of distinguishing between three very similar glycoproteins as rEPO α/β, rEPO δ and NESP. These analyses showed lower microheterogeneity for rEPO δ containing almost exclusively tetraantennary N-glycans with different degree of sialylation (from 1 to 4). The absence of diantennary structures and LacNAc repeats are proposed as potential markers for rEPO δ differentiation. This homogeneity seems to be responsible for the absence of diagonality in its IEF profile.

7. Important differences in sialic acid acetylations were found between rEPOs and NESP. Mass spectra yielded up to 8 O-acetylation in tetraantennary NESP N-glycans while this modification was insignificant in rEPOs.

8. A very sensitive capillary HPLC separation method for sialic acids was developed in order to specifically detect Neu5Gc in biological samples. This method, once validated, may be useful as a confirmatory method to detect rEPO misuse as well as for quality control of rEPO pharmaceuticals manufacture.

9. Sialic acid analysis highlighted a key difference between rEPO δ, synthesised in human cells, and other recombinant EPO preparations produced in CHO cells. The former lacks *N*-glycolylneuraminic acid, a non-human sialic acid species that could produce immune response in patient treated with recombinant preparation containing it. Contrary, rEPO α/β and NESP were found to contain around 1% of this sialic acid.

6. Appendix

POTENTIAL ISOFORM DISCRIMINATION DURING IMMUNOAFFINITY PURIFICATION OF EPOETIN AND DARBEPOETIN

V. Belalcazar[1], J. Segura[1,2], C. de Bolós[1], S.H Peng[1], R. Gutiérrez[1,2], E. Llop[1] and J. A. Pascual[1,2]

1. Pharmacology Research Unit, Institut Municipal d'Investigació Mèdica (IMIM), Barcelona (Spain)
2. Dept. Experimental Sciences and Health, Universitat Pompeu Fabra, Barcelona (Spain)

INTRODUCTION

Immunoaffinity chromatography (IAC) has been extensively used for the purification of human erythropoietin (hEPO) from biological samples. Product identification by SDS-PAGE reflects the average molecular weight but not the isoform distribution. However, microheterogeneity of the of the glycoprotein, particularly the sialic acid content has proven to be essential for its biological activity. Moreover, manufacturers of the recombinant protein (rhEPO) select and purified the more acidic isoforms for its commercial preparations. Isoform distribution is also essential in differentiating between endogenous urinary hEPO from rhEPO. This is the basis for the isoelectric focusing (IEF) method currently used in doping control.
The aim of the present work was to study the potential of different antibodies to discriminate between hEPO and rhEPO isoforms used in IAC and monitored by IEF.

EXPERIMENTAL

Immunoaffinity Columns:

- Monoclonal Immunoaffinity column: 250µg of monoclonal anti-human EPO antibody (R&D Systems) immobilized on CNBr- activated Sepharose (Pharmacia)
- Policlonal Immunoaffinity column: 2mg of anti-human erythropoietin policlonal antibody (Sigma) immobilized on CNBr- activated Sepharose (Pharmacia)

Samples preparation:

Blank urines (OB) were spiked with the follow standard solution:
- 500 mIU of rEPO standard (alpha and beta isoforms, erythropoietin BRP Batch Nº1; European Pharmacopoeia Commission)
- 500-1000mIU of Human urinary EPO standard (endogenous isoforms, 2nd International reference preparation; WHO International Laboratory for Biological Standards) of human urinary EPO.
- Darbepoetin: 10ng of NESP (novel erythropoiesis stimulating protein) Aranesp, Amgen.

Samples analysis:

A. SAMPLE TREATMENT :

_1. OB + standard solution (
_2. Centrifugation
_3. Filtration and concentration by Centricom® Plus- 20 (Millipore)

B. PURIFICATION BY MONOCLONAL AND POLICLONAL IAC :

_4. Equilibration: Tris 50mM, 1% BSA, pH 7.4, flow rate: 1ml/min
_5. Sample loading: flow rate: 10 µL/min
_6. Washing: Tris 50mM, 1% BSA, pH 7.4, flow rate: 1ml/min
_7. Elution: Glicine- HCl 0.1 M, pH 2.2, flow rate: 1 ml/min
Eluents + Tris 1M, 1% BSA, pH 9 (neutralizing buffer).
_8. Washing: Tris 50mM, 1% BSA, pH 7.4, flow rate: 1ml/min
_9. Stored: 10 % Sodium azide

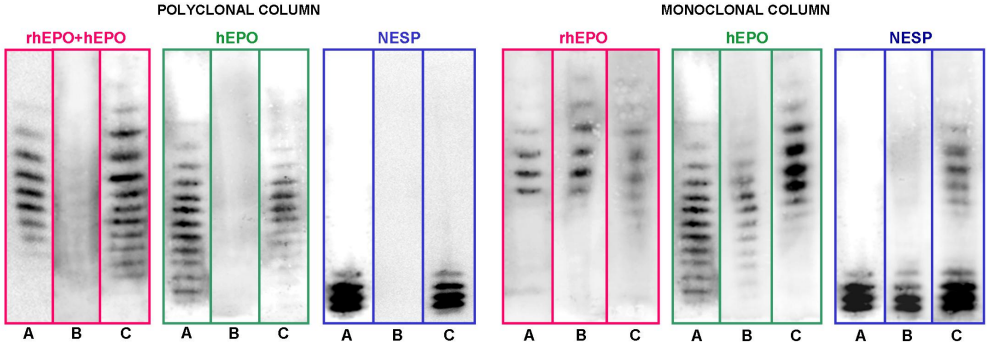
C. IEF, DOUBLE-BLOTTING AND DETECTION :

_10. Polyacrylamide gel, 5% T, 3% C, 2-6 pH gradient, 7M Urea. 200V, 3600 Vh, 1W/ cm gel length.
_11. Immunostaining with MAB anti-EPO (R&D Systems).
_12. Double-blotting
_13. Immunostaining with secondary antibody: biotinylated goat anti-mouse
_14. Incubated with Streptavidin-peroxidase complex.
_15. Detection by chemiluminescent used a Luminescent image analyser with a CCD camera (LAS-100, Fujifilm).

FRACTIONS ANALISED BY IEF:

FRACTIONS	SAMPLES	COLUMS
A. Sample before purification	rhEPO	Polyclonal IAC
B. Sample load flow-through	hEPO	&
C. Eluent	NESP	Monoclonal IAC

RESULTS



CONCLUSIONS

The monoclonal antibody retained preferently the more basic isoforms of hEPO and rhEPO. In the case of NESP all isoforms were recognised.
The polyclonal antibody showed a marked affinity for hEPO and NESP and less for rhEPO.
The capacity of the antibodies of purified some isoforms using IAC techniques acquires special relevance in the characterisation between different EPO related products.

REFERENCES

Wojchowski, D.M. et al. Biochim Biophys. Acta, 1987, pg: 170-78.
Lasne, F. J Immunol.Methods, 2001, pg :125-131.
Lasne, F. et al. Anal. Biochem, 2002, pg:119-126.
Pascual, J.A. et al. Recent advances in doping analysis (10), pg 135-144.
Belalcazar, V. et al. Recent advances in doping analysis (11).

ACKNOWLEDGEMENTS
The authors would like to thank all the staff from the Laboratory especially Belén Cano and David Díaz.

E. Llop[1,2], R. Gutiérrez Gallego[1,2], V. Belalcázar[1,2], G.J. Gerwig[3], J.P. Kamerling[3], J.Segura[1,2] and J.A. Pascual[1,2]

STRUCTURAL FEATURES OF EPO AND ANALOGUES: GLYCOSYLATION PROFILES

(1) Pharmacology Research Unit, IMIM, Barcelona (Spain)
(2) Department of Experimental and Health Sciences, UPF, Barcelona (Spain)
(3) Bijvoet Center for Biomolecular Research, Department of Bio-Organic Chemistry, Section of Glycoscience and Biocatalysis, Utrecht University (The Netherlands)

INTRODUCTION

Erythropoietin (EPO) is a glycoprotein hormone that is secreted primarily by adult kidneys in response to tissue hypoxia, and it is involved in the regulation of the level of red blood cells. EPO molecule contains 3 N-linked (Asn-24, 38, 83) and 1 O-linked (Ser-126) oligosaccharide chains representing about 40% of the total mass (29.4 kDa). The recombinant analogue (rEPO), available since 1986 has found widespread use in the treatment of anaemia, renal failure, cancer, AIDS, etc. Furthermore, rEPO is illicitly used by athletes to boost the delivery of oxygen to the tissues and enhance their performance in endurance sports. For this reason, the IOC banned the use of rEPO as ergogenic agent in 1990. In 2001 a novel erythropoiesis-stimulating protein (NESP)[1] was synthesised. NESP possesses two additional N-glycans (Asn 30, 88) compared to endogenous or recombinant EPO. The number and composition of the N-glycans is clearly important in the metabolism of this glycoprotein because the carbohydrate content determines the ½-life time. Current tests to differentiate between urinary endogenous (uEPO) and recombinant analogues (rEPO, NESP) are based on differences in their isoelectric focussing (IEF) profiles. It is well known that both, the natural and recombinant EPO have identical peptide backbones. However, differences arising from post-translational modifications have been reported and we hypothesised that the observed IEF profiles[2] are originated by cell-type specific glycosylation. Here we describe the meticulous structural characterisation of both endogenous and exogenous EPO's and analogues. The particular structural features of each glycoprotein encountered may pave the way to the unambiguous detection of rEPO/NESP abuse and the developed methodology may be employed in a clinical-diagnostic setting.

EXPERIMENTAL

Materials - Human urinary EPO (2nd international reference preparation 1971) was purchased from National Institute for Biological Standards and Control (NIBSC, Hertfordshire, United Kingdom). Recombinant human EPO was obtained from European Pharmacopoeia Commission (Strasburg/France). NESP (darbepoetin α) was acquired from Amgen. Monoclonal anti-human EPO antibody (clone: AE75A) was from R&D Systems. CNBr-activated Sepharose 4B was from Amersham Biosciences. Centricon plus-20™ and YM-30™ centrifugal filters were purchased from Millipore. Dextran T70 was from Pharmacia Biotech. Recombinant peptide-N^4-(acetyl-β-glucosaminyl)-asparagine amidase F (PNGase F, EC 3.1.27.5), recombinant β-1,4-galactosidase (EC 3.2.1.23), recombinant endo-α-N-acetylglucosaminidase (EC 3.2.1.97), and recombinant α-2,3,6,8-neuraminidase (EC 3.2.1.18) were purchased from Calbiochem. Carbograph graphitised carbon ultra-clean columns (150 mg in 4 ml) were purchased from Alltech. 2,5-dihydroxybenzoic acid (DHB), bovine serum albumin (BSA), bovine fetuin and phosphate-buffered saline (PBS) were purchased from Sigma. 2-aminobenzamide (2AB), sodium cyanoborohydride ($NaCNBH_3$), dimethylsulfoxide (DMSO) were from Fluka. GELoader Tips were purchased from Eppendorf. Normal phase TSK gel Amide-80 column (4.6 x 250 mm) that was obtained from TOSOH BIOSEP. Weak anion exchange VYDAC 301 VHP column (7.5 mm ID x 50 mm) was from VYDAC. All other chemicals were of highest purity commercially available.

Sample preparation - Human urinary EPO was subjected to affinity-chromatography. Aliquots of ~ 3.5 ng were loaded column containing monoclonal anti-human EPO antibody covalently linked to CNBr-activated Sepharose 4B, previously equilibrated with 2 ml Tris-HCl 50 mM (pH 7.4) containing 1% BSA at a flow rate: 1 ml/min. Flow-through were re-applied onto the column at a flow rate of 50 µl/min. Then, the column was washed with 2 ml of equilibration buffer. Bound uEPO was eluted with 0.1 M Glycine-HCl (pH 2.2) and collected in 400 µl 1M Tris-HCl (pH 9.0) containing 1% BSA (uEPO-F1). Two subsequent elution fractions were collected (uEPO-F2 and uEPO-F3). Immuno-purified uEPO samples, NESP and rEPO dissolved in water, were desalted and concentrated by filtration using 30 kDa filters and lyophilised.

(Applied Biosystems), equipped with a N_2 laser (337nm). Typically, spectra were acquired in linear mode for negative polarity and in reflectron mode for positive polarity. Mass scans were accumulated in the range between 900-5000 Da. External calibration of the spectrometer was performed using a mixture of glucose oligomers in positive ion mode and 2AB-derivatised fetuin *N*-glycans in negative mode. Recorded data were processed with Data Explorer™ Software (Applied Biosystems).

RESULTS AND DISCUSSION

Monosaccharide analysis- Monosaccharide analysis of fraction I of rEPO purified with Bio-Gel P-2 (Figure 1) revealed the presence of Fuc, Man, Gal, GalNAc, GlcNAc, Neu5Ac in the molar ratio of 1.0 : 3.0 : 4.7 : 0.2 : 0.2 : 6.1 : 4.9 with respect to the internal standard and setting the molar value of Man at 3. Neu5Ac represents the total sialic acid content.

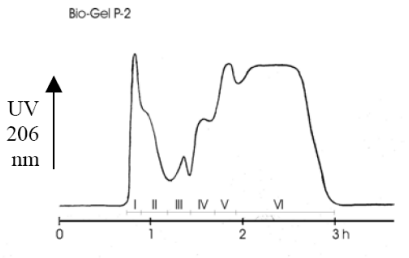


Figure 1: Elution profile of rEPO on Bio-Gel P-2. I – VI indicate the different fractions collected. Only fraction I and II contained glycoprotein.

Table I shows a normalised compilation of monosaccharide values for rEPO. Although at first glance these data are similar, significant structural feature can be deduced: i.e. the ratio 3:1 (Man:Fuc) in rEPO indicates a complete core fucosylation of the N-linked glycans as CHO cells are known to lack the α-1,3/4-fucosyl transferase. Similarly, the 3:6 ratio for Man:GlcNAc in combination with the ratio of 3:5 for Man and NeuAc points towards the presence of predominantly tetra-antennary complex-type N-glycans with a high degree of sialylation as well as the presence of *N*-acetyl lactosamine repeats. These trends are slightly different for uEPO indicating possible structural glycosylation differences.

Table I Monosaccharide composition of erythropoietin from different origins. Explanation of the abbreviations: rEPO, recombinant erythropoietin from Chinese hamster ovary (CHO) cells; rEPO-T, recombinant erythropoietin from Chinese hamster ovary (CHO) cells as produced by Teknika; rEPO BHK, recombinant erythropoietin from baby hamster kidney cells; uEPO, human urinary erythropoietin.

MOLAR RATIO (mannose taken as 3)	**rEPO BRP Stnd***	**uEPO[10]**	**rEPO-T[5]**	**uEPO[11]**	**rEPO BHK[11]**
Fucose (Fuc)	1.0	0.9	1.0	1.0	1.0
Mannose (Man)	**3.0**	**3.0**	**3.0**	**3.0**	**3.0**
Galactose (Gal)	4.7	4.2	4.3	4.9	4.7
***N*-acetyl galactosamine (GalNAc)**	0.2	0.3	traces	8.2	8.6
***N*-acetyl glucosamine (GlcNAc)**	6.1	5..3	6.0		
***N*-acetyl neuraminic acid (NeuAc)**	4.9	3.4	4.1	3.9	5.6

* Monossacharide analysis performed as a part of this study.

Sialic Acid analysis - Analysis of the sialic acid speciation of the different erythropoietin samples resulted in chromatograms depicted in figure 2A-C. The profile corresponding to the uEPO displayed , in addition to excess reagent peaks, a single sialic acid species that when compared to the reference compound (not shown) corresponded to Neu5Ac. The recombinant glycoproteins showed the presence of this sialic acid residue as the most abundant, yet not only, species.

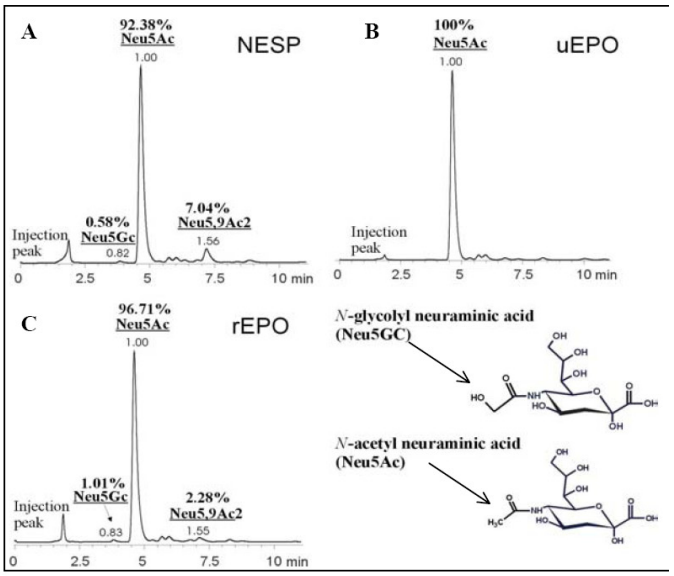
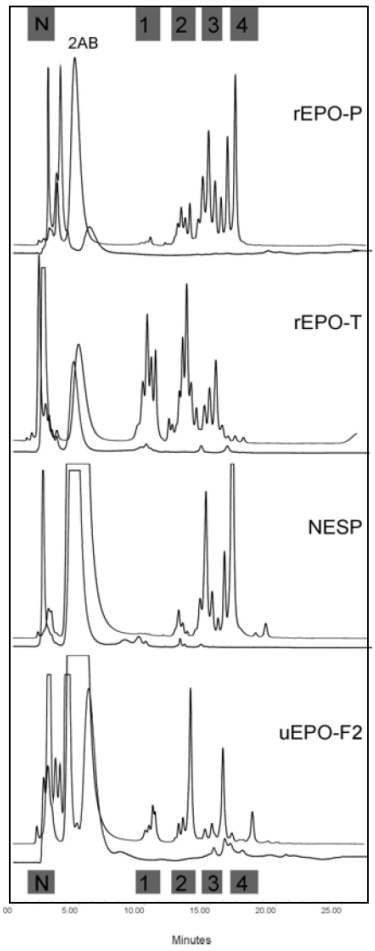


Figure 2: Reversed phase HPLC analysis of fluorescently labelled (DMB) sialic acid residues from: A, NESP; B, uEPO; C: rEPO.

In addition, mono-O-acetylated Neu5,9Ac$_2$ was observed for NESP (~ 7%) and rEPO (~ 2%) and importantly, *N*-glycolyl neuraminic acid was found to be present in both (0.6 and 1.0%, respectively). This residue cannot be present on human glycoproteins as humans lack CMP-*N*-acetylneuraminic acid hydroxylase (CMAH) that converts Neu5Ac into Neu5Gc. Therefore, this observation provide a handle for the unequivocal discrimination between endogenous and recombinant EPO.

Weak Anion Exchange (WAX) HPLC profiles (figure 3)- For the urinary preparation, four (flow-through, F1, F2, and F3) fractions were profiled and peaks of significance were only observed in the flow-through (FT) and F2. From a comparison between the profiles from FT (not shown) and elution (F2, figure 3) derived from immuno-purification of uEPO it could be observed that neutral oligosaccharides are almost absent in F2. In contrast, they were very abundant in FT. This can be explained because FT fraction contains, in addition to non-retained EPO other glycoproteins. The charge profiles for both rEPOs from the European Pharmacopoeia (rEPO-P) and Teknika (rEPO-T), respectively clearly demonstrated the post-expression purification of rEPO-P towards higher sialylated oligosaccharides with the complete absence of neutral oligosaccharides in the former and their presence in the latter. It should be noted that the chromatogram observed for rEPO (rEPO-P) shows a similar profile as previously reported for epoetin beta by Skibeli et al[12]. This similarity is corroborated in the normal phase HPLC profiles (*vide infra*).



In figure 3, the profiles from rEPO-P and NESP correspond in terms of structures, although significant differences are observed for the absolute amounts of particular structures. Also for NESP a complete absence of neutral structures is visualised. Significant differences can be appreciated between the charge profiles of NESP, rEPO-P and uEPO-F2 indicating absolute differences in glycosylation. Where the profiles of the recombinant molecules display increasing intensities for structures containing from 1 to 4 charges, respectively, and significant heterogeneity within uniformly charged structures, the charge distribution for uEPO-F2 appears to be more homogeneous with the most intense peak corresponding to di-charged structures. Nevertheless, the retention times of the latest eluting structures (higher charged) structures in uEPO-F2 appear to be longer than those corresponding to the latest eluting structures from the recombinant variants.

Figure 3: WAX HPLC profiles of different erythropoietins before (upper trace) and after sialidase digestions (lower trace). rEPO-P: from European pharmacopoeia (NIBSC); rEPO-T from Teknika; NESP from Amgen.; uEPO-F2: elution fraction from the IAC purification of endogenous uEPO. Explanation: N, neutral structures; 1, mono-charged; 2, di-charged; 3, tri-charged;

From the WAX profiles after de-sialylation it becomes evident that the charges present in the N-linked carbohydrate chains from both rEPO and NESP correspond solely to sialic acids as the profiles show no peaks of significance in the region of charged structures. However, the uEPO-F2 profile after the same treatment still shows charged structures. As all digestions were performed on similar amounts of material and the enzyme used has a broad specificity, it must be concluded that the endogenous uEPO contains additional charges that reside in part, or in whole, in the glycosylation and are different from sialic acids. As such we hypothesised that these could be sulfate substituents at galactose or *N*-acetyl glucosamine residues.[13]

Normal Phase HPLC profiles - The same N-linked samples were analysed by normal phase HPLC (figure 4A-D). The overall glycosylation profiles (1st profile of each panel) represent the structural heterogeneity contained in the N-glycosylation.

From a comparison between the different EPO´s it became evident that both rEPO-P and NESP contain relatively large structures, possibly indicating the presence of *N*-acetyl lactosamine repeating units. In contrast the uEPO-F2 profile showed peaks with shorter retention times indicating less structural heterogeneity.

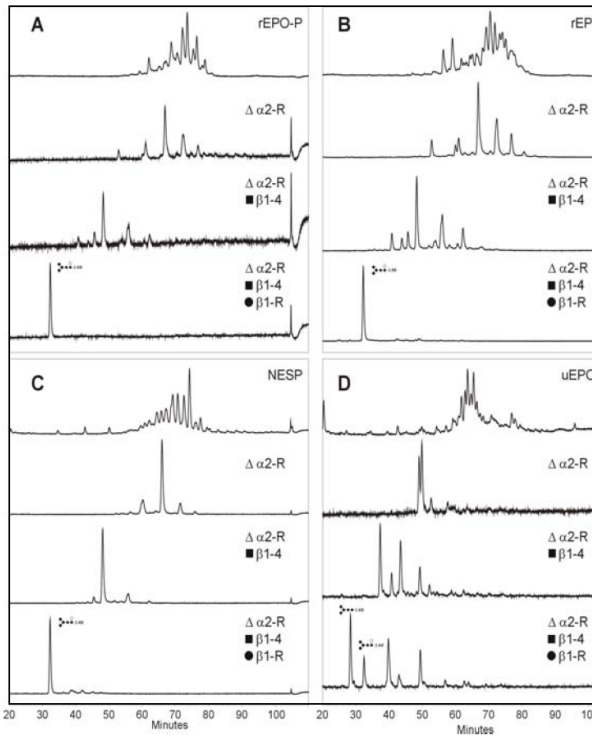


Figure 4: Normal phase HPLC profiles of different erythropoietins before (top panels) and after RAAM. **A**: rEPO-P from European pharmacopoeia (NIBSC); **B**: rEPO-T from Teknika; **C**: NESP from Amgen.; **D**: uEPO-F2 elution fraction from the IAC purification of endogenous uEPO. Enzymatic digestions: Δα2-R, sialidase; ■β1-4, galactosidase; ●β1-R, N-acetyl glucosaminidase.

In order to complete the structural analysis of the N-glycans a reagent array analysis method (RAAM) was carried out which consisted in successive and sequential exoglycosidase digestions. For standard complex-type N-glycans RAAM should result in the single penta- or hexasaccharide depicted in figure 5. Indeed the profiles corresponding to the recombinant molecules yielded a single peak corresponding to a single structure as demonstrated by mass spectrometry (*vide infra*). As such, the differences in glycosylation profiles observed for the recombinant molecules can be tracked down to the level of branching (i.e. bi-, tri -, and tetra-antennary complex type N-glycans) and presence of a different number of *N*-acetyl lactosamine repeating units. Strikingly, the comparative profiles of uEPO with respect to any of the recombinant analogues were completely different after each single exoglycosidase digestion. For this erythropoietin the triple-enzyme combination does not lead to a single structure but results in five major and even so many minor compounds. One of these co-elutes with the resulting structure in the recombinant analogues for which structural identity was assumed at this point. A single structure elutes at shorter time and assigned the same structure

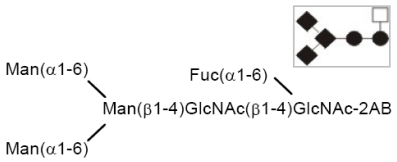
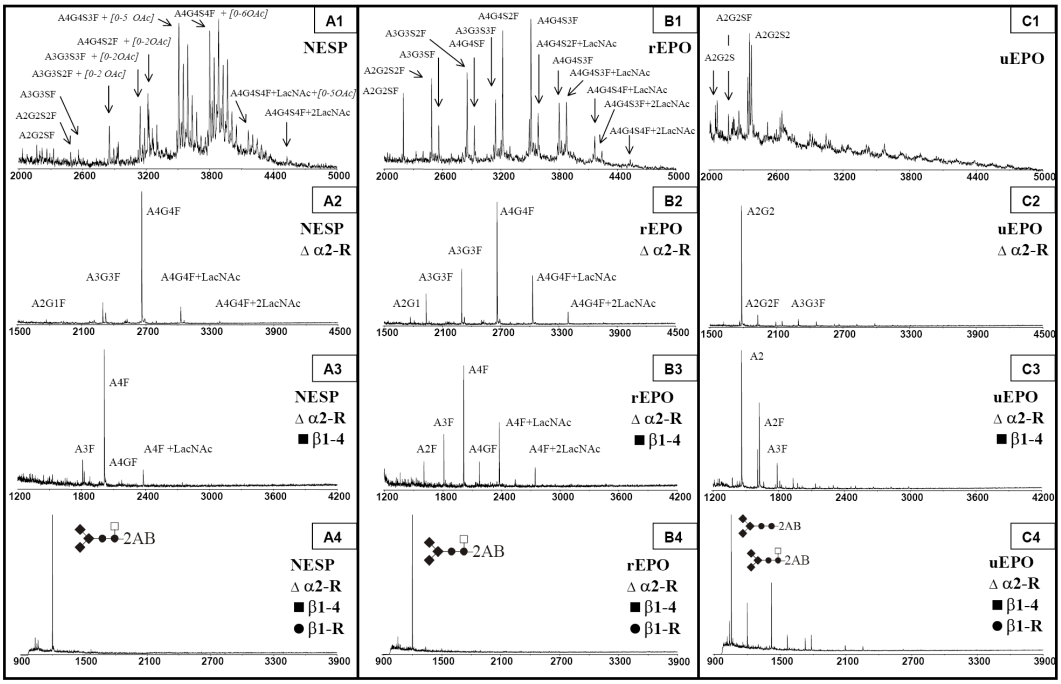


Figure 5: N-glycan core structure containing 3 Man (◆) 2 GlcNAc (●) and 1 Fuc (□). The latter is partially absent in uEPO.

devoid of the core-fucose. These observations were corroborated by mass spectrometric analyses.

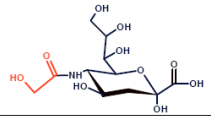
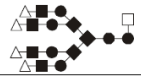
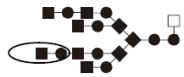
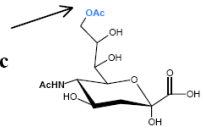
Matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF-MS)
The mass spectrum of the entire 2AB-labelled glycans pool obtained from rEPO showed around 20 different structures in the mass range between 2000 and 4500 Da (Figure 6-B1). Predominantly, tetra-antennary type N-glycans with a high degree of sialylation, complete core fucosylation and one or two *N*-acetyl lactosamine repeating units could be assigned. This observation is in complete agreement with the selective purification of this pharmaceutical. The equivalent mass spectrum of NESP (Figure 6-A1) presented similar structural heterogeneity, albeit a distinctive mass profile due to peak-clusters spaced by 42 Da. These clusters could be identified as O-acetylated *[OAc]* variants of a single structure. In case of IAC purified-uEPO the overall spectrum mass (Figure 6-C1) proved more difficult to interpret. Even though peaks were encountered throughout the entire mass range the most abundant signals corresponded to fairly simple biantennary type N-glycans, both with and without core fucosylation. Mass spectrometric analyses after RAAM at different stages (Figures 6-A-C panels 2-4) were performed and corroborated the earlier findings from HPLC. The structural heterogeneity of rEPO and NESP are solely due to different sialic acids present in both species as post-sialidiase mass spectra resulted nearly identical, both in mass values and peak intensities. As expected after the triple-enzyme treatment, the N-glycans of NESP and rEPO are reduced to a single structure corresponding in mass to the core-fucosylated (α1-6 linked to the first *N*-acetyl-glucosamine) hexasaccharide. However, this structure only accounts for approximately 30% in uEPO, the remaining 70% being non-fucosylated. This difference in core-fucosylation could be another potent tool in discriminating between urinary EPO and recombinant preparations.

Legend to figure 6 (next page): Mass spectra of different erythropoietins. **A**: NESP from Amgen.; **B**: rEPO-P from European pharmacopoeia (NIBSC); **C**: uEPO-F2 elution fraction from the IAC purification of endogenous uEPO. **A1-C1:** Negative ion MALDI mass spectra of the entire 2AB-labelled glycan pools. **A2-C4:** Positive ion MALDI mass spectra after different enzymatic digestions: $\Delta\alpha$2-R: sialidase; ■ β1-4: galactosidase; ● β1-R: N-acetyl glucosaminidase. Structure compositions are indicated with the following symbols:
e.g **A4G4S2F+LacNac+[0-5*OAc*]**
A: nº of antenna (4)
G: galactose (4)
S: sialic acid (2)
F: fucose (1)
LacNAc: *N*-acetyl-lactosamine repeats (1)
OAc: O-acetylation in sialic acid (from 0 to 5)


A1
NESP
A4G4S3F + [0-3 OAc]
A4G4S4F + [0-6OAc]
A4G4S2F + [0-2OAc]
A3G3S3F + [0-2OAc]
A3G3S2F + [0-2 OAc]
A3G3SF
A2G2S2F
A2G2SF
A4G4S4F+LacNAc+[0-3OAc]
A4G4S4F+2LacNAc
2000 2600 3200 3800 4400 5000
B1
rEPO
A4G4S2F
A4G4S3F
A3G3S2F
A3G3S3F
A4G4SF
A3G3SF
A4G4S2F+LacNAc
A2G2S2F
A2G2SF
A4G4S3F
A4G4S3F+LacNAc
A4G4S4F+LacNAc
A4G4S3F+2LacNAc
A4G4S4F+2LacNAc
2000 2600 3200 3800 4400 5000
C1
uEPO
A2G2SF
A2G2S2
A2G2S
2000 2600 3200 3800 4400 5000
A2
NESP
Δ α2-R
A4G4F
A3G3F
A4G4F+LacNAc
A2G1F
A4G4F+2LacNAc
1500 2100 2700 3300 3900 4500
B2
rEPO
Δ α2-R
A4G4F
A3G3F
A4G4F+LacNAc
A2G1
A4G4F+2LacNAc
1500 2100 2700 3300 3900 4500
C2
uEPO
Δ α2-R
A2G2
A2G2F
A3G3F
1500 2100 2700 3300 3900 4500
A3
NESP
Δ α2-R
■ β1-4
A4F
A3F
A4F +LacNAc
A4GF
1200 1800 2400 3000 3600 4200
B3
rEPO
Δ α2-R
■ β1-4
A4F
A3F
A4F+LacNAc
A2F
A4GF
A4F+2LacNAc
1200 1800 2400 3000 3600 4200
C3
uEPO
Δ α2-R
■ β1-4
A2
A2F
A3F
1200 1800 2400 3000 3600 4200
A4
NESP
Δ α2-R
■ β1-4
● β1-R
2AB
900 1500 2100 2700 3300 3900
B4
rEPO
Δ α2-R
■ β1-4
● β1-R
2AB
900 1500 2100 2700 3300 3900
C4
uEPO
Δ α2-R
■ β1-4
● β1-R
2AB
2AB
900 1500 2100 2700 3300 3900

CONCLUSIONS

A meticulous structural analysis of the N-linked carbohydrate chains of erythropoietin molecules from different origins has been carried out. In the case of endogenous EPO immuno affinity purification preceded structural analysis whereas this was not necessary for the recombinant variant. The most promising features are summarised in table II. At the monosaccharide level structural differences include the presence of Neu5Gc in a 1 mol % was in all recombinant EPO preparations. This sialic acid species cannot be produced by human, so its presence on erythropoietin provides an absolute evidence of the exogenous origin. Furthermore, different ratios between fucose and mannose were found. This was corroborated by both HPLC profiling of the N-glycans and mass spectrometry after full RAAM, with the observation of complete fucosylation of the core in rEPO and NESP, present only in ~ 30 % in uEPO. From WAX analyses, both before and after sialidase digestion, the presence of charged residues, other than sialic acids in uEPO became evident. These residues are apparently absent in recombinant species. This feature could be used for the development of a confirmatory method in EPO doping control. Even though it remains to be proven structurally, we hypothesise that these additional charges are due to sulfate groups present on galactose or *N*-acetyl glucosamine. Noteworthy is the presence of more complex and branched structures with presence of LacNAc repeats in recombinant EPOs as well as the existence of multiple O-acetylation of the sialic acid residues in NESP. As far as we know, this phenomenon has not been described for the pharmaceutical preparations and experiments are being conducted to clarify this point. At the moment, further studies are being performed to elucidate the structure of the different glycoforms of rEPO, uEPO and NESP, contained in uniformly charged isoforms after isoelectric focusing in order to enable an unambiguous discrimination between endogenous and exogenous EPO preparations.

TABLE II	NESP	rEPO	uEPO
Neu5Gc 	**~0.6%**	**~1.0%**	**0%**
Charges 	SIALIC ACIDS	SIALIC ACIDS	Sialic Acids **Additional Charges (SO_4?)**
Non-fucosylated Core 	**~0%**	**~0%**	**~70%**
LacNAc repeats 	1 and 2 repeats	1 and 2 repeats	None
Most abundant structures	Tetra-antennary	Tetra-antennary	Bi-antennary
NeuOAc 	**1 to 6***	-	-

* this number of O-acetyl groups is present in one N-glycan containing up to four sialic acid residues.

Acknowledgments

The authors are thankful to the members of the Pharmacology Research Unit at the Municipal Institute of Medical Research (IMIM). A research grant from the World Anti-Doping Agency (WADA) is gratefully acknowledged.

REFERENCES

1. Egrie,J.C. & Browne,J.K. Development and characterization of novel erythropoiesis stimulating protein (NESP). *Br. J Cancer* **84 Suppl 1**, 3-10 (2001).

2. Lasne,F., Martin,L., Crepin,N., & de Ceaurriz,J. Detection of isoelectric profiles of erythropoietin in urine: differentiation of natural and administered recombinant hormones. *Anal. Biochem.* **311**, 119-126 (2002).

3. Kamerling,J.P. Carbohydrate features of recombinant human glycoproteins. *Biotecnologia Aplicada* **13**, 167-180 (1996).

4. Hara,S., Yamaguchi,M., Takemori,Y., Furuhata,K., Ogura,H., & Nakamura,M. Determination of mono-O-acetylated N-acetylneuraminic acids in human and rat sera by fluorometric high-performance liquid chromatography. *Anal. Biochem.* **179**, 162-166 (1989).

5. Hokke,C.H., Bergwerff,A.A., van Dedem,G.W., van Oostrum,J., Kamerling,J.P., & Vliegenthart,J.F. Sialylated carbohydrate chains of recombinant human glycoproteins expressed in Chinese hamster ovary cells contain traces of N- glycolylneuraminic acid. *FEBS Lett.* **275**, 9-14 (1990).

6. Packer,N.H., Lawson,M.A., Jardine,D.R., & Redmond,J.W. A general approach to desalting oligosaccharides released from glycoproteins. *Glycoconj. J* **15**, 737-747 (1998).

7. Bigge,J.C., Patel,T.P., Bruce,J.A., Goulding,P.N., Charles,S.M., & Parekh,R.B. Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Anal. Biochem.* **230**, 229-238 (1995).

8. Gutierrez Gallego R., Haseley,S.R., Van Miegem,V.F., Vliegenthart,J.F., & Kamerling,J.P. Identification of carbohydrates binding to lectins by using surface plasmon resonance in combination with HPLC profiling. *Glycobiology* **14**, 373-386 (2004).

9. Guile,G.R., Rudd,P.M., Wing,D.R., Prime,S.B., & Dwek,R.A. A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analyzing oligosaccharide profiles. *Anal. Biochem.* **240**, 210-226 (1996).

10. Sasaki,H., Bothner,B., Dell,A., & Fukuda,M. Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA. *J Biol. Chem.* **262**, 12059-12076 (1987).

11. Tsuda,E., Goto,M., Murakami,A., Akai,K., Ueda,M., Kawanishi,G., Takahashi,N., Sasaki,R., Chiba,H., Ishihara,H., & . Comparative structural study of N-linked oligosaccharides of urinary and recombinant erythropoietins. *Biochemistry* **27**, 5646-5654 (1988).

12. Skibeli,V., Nissen-Lie,G., & Torjesen,P. Sugar profiling proves that human serum erythropoietin differs from recombinant human erythropoietin. *Blood* **98**, 3626-3634 (2001).

13. Kawasaki,N., Haishima,Y., Ohta,M., Itoh,S., Hyuga,M., Hyuga,S., & Hayakawa,T. Structural analysis of sulfated N-linked oligosaccharides in erythropoietin. *Glycobiology* **11**, 1043-1049 (2001).