Mapping the N-glycome of Human Von Willebrand Factor

Kevin Canis *,†, Thomas A. J. McKinnon †, Agata Nowak †, Stuart M. Haslam *, Maria Panico *, Howard R. Morris *, Mike A. Laffan † and Anne Dell *,‡

* Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College, London, UK; † Department of Haematology, Faculty of Medicine, Hammersmith Hospital Campus, Imperial College, London, UK.

‡ To whom correspondence should be addressed: Anne Dell, Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London, London, SW7 2AZ, United Kingdom. Tel: +44 (0)20 7594 5219. Fax: 44 (0)20 7225 0458. E-mail address: a.dell@imperial.ac.uk

Running title: The *N*-glycome of the Von Willebrand Factor

Keywords: Von Willebrand factor / *N*-glycosylation / Glycomics / Glycoproteomics

Abstract

Von Willebrand Factor (VWF) is a key component for maintenance of normal haemostasis, acting as the carrier protein of the coagulant Factor VIII and mediating platelet adhesion at sites of vascular injury. There is ample evidence that VWF glycan moieties are crucial determinants of its expression and function. Of particular clinical interest, ABH antigens influence VWF plasma levels according to the blood group of individuals, although the molecular mechanism underlying this phenomenon remains incompletely understood.

This manuscript reports analyses of the human plasma VWF *N*-glycan population using advanced mass spectrometry. Glycomics analyses revealed about 100 distinct *N*-glycan compositions and identified a variety of structural features including lactosaminic extensions, ABH antigens, sulfated antennae as well as bisecting and terminal GlcNAc residues. We estimate that some 300 *N*-glycan structures are carried by human VWF.

Glycoproteomics analyses mapped ten of the consensus sites known to carry *N*-glycans. Glycan populations were found to be distinct although many structural features were shared across all sites. Notably, the H antigen is not restricted to particular *N*-glycosylation sites. Also, the N2635 site, previously designated as unoccupied, was found to be highly glycosylated. The delineation of such varied glycan populations in conjunction with current models explaining VWF activity will facilitate research aimed at providing a better understanding of the influence of glycosylation on VWF function.

Introduction

Von Willebrand Factor (VWF) is a very large multimeric plasma glycoprotein expressed by vascular endothelial cells and megakaryocytes. VWF is a key component for maintenance of normal haemostasis, mediating two essential functions. Firstly, it supports platelet adhesion to damaged surfaces at sites of vascular injury by interacting with both the exposed sub-endothelial matrix proteins and the platelet glycoprotein complexes GpIbα and GpIIbIIIa [1]. Secondly, it acts as the carrier molecule for the pro-coagulant factor VIII, increasing its half-life in blood from 2 to 12 hours in humans and protecting it from activation by Factor Xa and inactivation by protein C [2, 3]. During its biosynthesis VWF undergoes extensive post-translational modification including pro-peptide cleavage, multimerisation and glycosylation, leading to a wide range of very large and structurally complex disulphide-linked multimeric structures of $0.5x10^6$ to over $2x10^7$ Da. The mature VWF monomer macrostructure is conventionally organised in four groups of homologous repeating domains named A, B, C and D joined together by unstructured linkers. A number of functional domains have been characterized along the 2813 amino acid VWF sequence, being involved either in its intrinsic properties (dimerisation, multimerisation, cleavage by the ADAMTS-13 metalloprotease; [4]) or in its interaction with partner proteins such as Factor VIII [5], collagens [6, 7], platelet GpIbα and GpIIbIIIa [8, 9], P-selectin [10] or heparin [11].

The initial determination of the VWF amino acid sequence by Titani and collaborators [12] highlighted the presence of twelve *N*-glycosylation consensus sequons N-X-S/T (amino acids N_{857}, N_{1231}, N_{1515}, N_{1574}, N_{2223}, N_{2290}, N_{2357}, N_{2400}, N_{2546}, N_{2585}, N_{2635} and N_{2790}). All but one of these sites (N_{2635}) was reported to be occupied with *N*-glycans. An additional rare sequon, N-X-C (N_{1147}), was also found to be glycosylated. Early studies of the VWF *N*-glycan population led to the characterisation of about twenty-five *N*-glycan structures [13-15]. Complex type bi-antennary structures were shown to constitute about 80% of the glycans identified, accompanied by tri- and tetra-antennary analogues as well as a minority of high mannose type *N*-glycans. Some structural features such as core fucosylation, antennae sialylation, and presence of lactosaminic elongations and of bisecting GlcNAc residues characterised the overall population. Interestingly, about 13% of the VWF *N*-glycans were shown to carry ABH blood group epitopes, an uncommon feature for a secreted glycoprotein.

Most of the VWF *N*-glycosylation consensus sequons are located in the vicinity of, or within, a functional site and it is therefore not surprising that VWF glycan content affects its properties and functions such as secretion and dimerisation [16, 17], multimerisation [18] and half-life in blood [19]. One of the most striking effects, is the correlation between ABO blood group and VWF plasma level; Significant differences between all ABO groups have been demonstrated by many studies, demonstrating the highest levels for blood group AB individuals, followed by group A, group B and lowest levels for group O individuals, i.e. AB > A ≥ B > O [20, 21]. The mechanism by which ABH antigens affect VWF plasma level remains unclear, although most evidence points to an effect on VWF clearance. ABH antigens also alter VWF susceptibility to proteolysis by ADAMTS-13, a specific protease regulating VWF multimeric size, with a cleavage rate in the rank order O > B ≥ A > AB [22]. VWF sialylation has also been shown to specifically enhance susceptibility to ADAMTS-13 proteolysis [23] whilst protecting it against cleavage by other serine and cysteine proteases such as plasmin, trypsin and chymotrypsin [24]. Interestingly, a recent study by McGrath and collaborators additionally demonstrated that desialylation abolishes the ABH effect on ADAMTS-13 cleavage rate, i.e. asialo-O-VWF and asialo-AB-VWF are cleaved at identical rates by ADAMTS-13 [23].

Although substantial evidence shows a significant influence of the VWF glycan content on its biology and functions, most of the underlying molecular mechanisms remain poorly understood, in part because the knowledge of its glycans is still vastly incomplete. Only a portion of the VWF glycan repertoire has been rigorously defined and almost nothing is known regarding sialylation and potential further modifications. In addition, until now, nothing has been known concerning the specificity of glycosylation at each of the consensus sites. In this paper, we describe our structural definition of the VWF *N*-glycan repertoire using advanced mass spectrometric techniques. We report structural features which had not been addressed previously, such as sulfation, sialylation status and the presence of terminal non bisecting GlcNAc residues. Importantly, we show that ABH blood group-substituted glycans are distributed throughout the VWF molecule.

Materials and methods

Von Willebrand Factor

Plasma-derived VWF was purified from Haemate P (ZLB, Behring, Germany) using a combination of gel filtration and heparin-Sepharose affinity chromatography. Briefly, Haemate P was gel filtered through a Hi-Prep Sephacryl S-400 HR gel filtration column (GE Healthcare, Little Chalfont, UK) using 20 mM Tris-HCl, 100 mM NaCl, pH 7.4. VWF-containing fractions were pooled and passed over a SK-16 chromatography column packed with 30 ml of heparin-sepharose 6 fast flow (GE Healthcare). The column was eluted with 20 mM Tris-HCl, 300 mM NaCl, pH 7.4 and the resulting fraction was dialyzed into 50 mM ammonium bicarbonate, pH 7.4 prior to lyophilisation.

General chemicals and reagents

Ammonium hydrogen carbonate (NH_4HCO_3), sodium chloride (NaCl), iodoacetic acid (IAA), dithiothreitol (DTT), α-cyano-4-hydroxycinnamic acid, and hexanes were purchased from Sigma-Aldrich Corporation (Poole, Dorset, UK). Glacial acetic acid, acetonitrile (ACN), ammonia (NH_3), chloroform ($CHCl_3$), dimethylsulfoxide (DMSO), methanol (MeOH), propan-1-ol, sodium hydroxide pellets (NaOH) and trifluoroacetic acid (TFA) were from Romil (Waterbeach, Cambridgshire, UK). Gentisic acid (2,5 re-dihydroxybenzoic acid) was from Fluka (Poole, Dorset, UK). Methyl iodide (ICH_3) was obtained from Alpha Aesar (Morecambe, Lancashire, UK). Snakeskin® dialysis tubing (cut-off 7 kDa) was from Pierce (Tattenhall, Cheshire, UK). Tris(hydroxymethyl)aminomethane was from Fisher (Loughborough, Leicestershire, UK). 3,4-diaminobenzophenone (DABP) was from Acros Organics, NJ). Argon and nitrogen gases were supplied by BOC (Guilford, Surrey, UK). All aqueous solutions were prepared using distilled/deionised (Milli-Q) water.

Carboxymethylation

Lyophilized VWF aliquots (typically 100 µg) were dissolved in 200 µl of 2 mg/ml dithiothreitol in degassed 0.6 M Tris, pH 8.5 buffer and reduction was carried out by incubation for 60 min at 37°C. The same volume of 60 mg/ml iodoacetic acid in degassed 0.6 M Tris, pH 8.5 buffer was then added and carboxymethylation was performed by incubation for 2 h at room temperature. The reaction was terminated by dialyzing samples against 50 mM ammonium bicarbonate, pH 7.4 buffer for 36 h at 4°C. After dialysis, the VWF samples were lyophilized.

Tryptic and PNGase F digestions

For glycomics analyses, reduced-carboxymethylated VWF was incubated with L-1-tosylamido-2-phenylethyl chloromethyl ketone (TPCK) treated bovine pancreas trypsin (EC 3.4.21.4; Sigma, Poole, UK) at a 1:50 ratio (w/w) in 50 mM ammonium bicarbonate, pH 8.4 buffer, for 12 h at 37°C. For glycoproteomic analyses, reduced-carboxymethylated VWF was digested using Sequencing Grade Modified Trypsin (Promega) at a 1:50 ratio (w/w) in a 50 mM ammonium bicarbonate, pH 8.4 buffer, for 2 to 14 h at 37°C. The digestion was terminated by addition of one drop of glacial acetic acid and the sample was immediately purified on a reverse-phase C_{18} Sep-Pak cartridge (Waters Corp., Milford, MA, USA) with the propan-1-ol / 5% acetic acid method. The C_{18} cartridge was conditioned successively with methanol, 5% acetic acid, propan-1-ol, and 5% acetic acid. The tryptic digest was loaded onto the cartridge and washed with 20 ml of 5% acetic acid. VWF peptides and glycopeptides were eluted sequentially with each 20%, 40% and 100% (v/v) propan-1-ol in water. 20% and 40% propan-1-ol fractions were pooled and lyophilized. For the PNGase F digestion, the VWF tryptic digest was suspended in 50 mM ammonium bicarbonate, pH 8.4, and digested with three units of the enzyme for 24 h at 37°C. The released *N*-glycans were extracted and purified from the sample using a Sep-Pak C_{18} cartridge with the propan-1-ol / 5% acetic acid method as before, the *N*-glycans being eluted using 5 ml of 5% (v:v) acetic acid. The volume of this fraction was reduced on a Savant Speed-Vac and purified *N*-glycans were finally lyophilised.

Digestion by endo β-galactosidase

Polylactosaminic elongations carried by VWF *N*-glycans were digested using a recombinant endo β-galactosidase from *Escherichia freundii* (EC 3.2.1.103, Seikagaku Corporation). The sample was dissolved in 50 mM ammonium acetate, pH 5.5 buffer. 25 mU of the endo β-galactosidase were added and the sample was incubated for 18 h at 37°C. Another aliquot of the enzyme were added, and the sample was again incubated for 18 h at 37°C. Glycans were purified using a Sep-pak® C_{18} cartridge with the propan-1-ol / 5% acetic acid method prior to lyophilisation.

Galactosylation with $\beta_{1\text{-}4}$ galactosyltransferase

Purified *N*-glycans were dissolved in 100 µl of a 50 mM MOPS, 20 mM $MnCl_2$, pH 7.4 buffer containing 1 mg/ml of UDP-galactose. 1 mU of $\beta_{1\text{-}4}$ galactosyltransferase from bovine milk (EC 2.4.1.90, Merck, Middlesex, UK) was added and the sample was incubated for 24 h at 37°C. Another aliquot of the enzyme solution was then added and the sample was again incubated 24 h at 37°C. After reaction, glycans were purified using a Sep-pak® C_{18} cartridge with the propan-1-ol / 5% acetic acid method prior to lyophilisation.

Digestion by neuraminidase

Desialylation of released glycans or glycopeptides was achieved using a recombinant neuraminidase from *Vibrio cholerae* (EC 3.2.1.18, Sigma). The sample was dissolved in 50 mM ammonium acetate, pH 5.8 buffer. 25 mU of the recombinant neuraminidase were added and the sample was incubated for 18 h at 37°C. Another aliquot of the enzyme was added, and the sample was again incubated for 18 h at 37°C. Glycans were then purified using a Sep-pak® C_{18} cartridge using the propan-1-ol / 5% acetic acid method prior to lyophilisation.

Chemical derivatisation

VWF *N*-glycans were explored using a technique adapted from a recently developed strategy published by Khoo and collaborators [25]. Lyophilised and purified VWF *N*-glycans were dissolved in about 100 µl of the NaOH/DMSO slurry and 100 µl of methyl iodide. The samples were gently mixed and incubated for 3 hours at 4°C. The permethylation was then stopped by addition of a few drops of ice-cold water, followed by pH neutralisation to around pH 6.5 using a 5% (v:v) acetic acid solution. Glycans were then cleaned-up using a Sep-Pak® C_{18} cartridge with the acetonitrile / water protocol. The reverse-phase was conditioned successively with methanol, water, acetonitrile and water. The sample was dissolved in 1:1 (v:v) methanol/water and loaded onto the cartridge. Hydrophilic salts and contaminants were washed off by passing 10 ml of water, then 5 ml of a 2.5% (v:v) acetonitrile solution. Glycans were then sequentially eluted (according to their size and composition) using 5 ml of 10%, 25%, 50% and 75% (v:v) aqueous acetonitrile solution. Organic solvent from each fraction was removed on a Savant Speed-Vac and samples were lyophilised.

Linkage analysis

Permethylated *N*-glycans were hydrolyzed to partially methylated monosaccharides by incubation in 2 M trifluoroacetic acid at 121°C for 2 h. Samples were dried under a stream of nitrogen. Reduction of the monosaccharides and tagging of their carbon 1 was achieved by incubation for 2 h at room temperature in a 10 mg/ml $NaBD_4$ in 2 M NH_4OH solution. The reaction was terminated by the addition of five drops of glacial acetic acid and samples were dried under a stream of nitrogen. Excess borates were removed by co-evaporation with a 10% (v:v) acetic acid in methanol solution followed by drying under a stream of nitrogen. Samples were acetylated by incubation in acetic anhydride for 1 h at 100°C and then dried under a stream of nitrogen. The resulting products were extracted with chloroform in the same way as permethylated samples. The chloroform layer was dried down and the partially methylated alditol acetates (PMAA) were dissolved in hexanes prior to gas chromatography (GC)-MS analyses, performed on a Perkin Elmer Clarus 500 GC-MS (Perkin Elmer Instruments, Shelton, USA) machine fitted with an RTX-5MS column (30 m long, 0.25 mm internal diameter, Thames Restek UK Limited, Saunderton, Bucks, UK). The PMAA were injected at 60°C then the temperature was increased at a rate of 8°C/min to a temperature of 300°C. Data were acquired and processed using GC-MS Turbomass v4.5.0 Instrument Control software.

Affinity chromatography on Ulex europaeus agglutinin (UEA-I) lectin

The *Ulex europaeus* agglutinin I (UEA-I) has been reported to bind efficiently and specifically to fucosylated structures and especially to blood group H containing type 2 chain epitopes (Fuc $\alpha_{1\rightarrow2}$ Gal $\beta_{1\rightarrow4}$ GlcNAc) [26]. 2 ml of agarose bound *Ulex europaeus* agglutinin I (EY Lab, CA, USA) was packed in a 2 ml glass column and equilibrated using 10 column volumes of 10 mM sodium phosphate, 150 mM NaCl, pH 7.4 buffer. The flow rate was set up at 0.1 ml/min for the whole experiment. VWF desialylated glycopeptides aliquots (obtained from treatment by neuraminidase, as described above) were dissolved in 100 µl of PBS buffer and loaded onto the column. The column was washed by 10 column volumes and the subsequent unbound fraction was collected. Bound glycopeptides were then eluted using the same buffer containing 500 mM of L-fucose. Both bound and unbound fractions were immediately purified using a Sep-pak® C_{18} cartridge with the propan-1-ol / 5% acetic acid method to remove the salt and sugar excess prior to LC separation and/or MS analyses.

Nano-LC separation and spotting

VWF tryptic digests were analysed by *off-line* liquid chromatography / matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (*off-line* LC-MALDI-TOF MS). Peptides and glycopeptides were separated using an Ultimate 3000 LC system (Dionex, Sunnyvale, CA), fitted with a analytical Pepmap C_{18} nanocapillary column (15 cm length, 75 mm internal diameter; Dionex). The flow rate was set up at 0.3 µl/min. The sample, diluted in 0.1% (v:v) TFA, was loaded onto the column. Sequential elution was carried out using solvent A (0.1% (v:v) TFA in 2% (v:v) acetonitrile and solvent B (0.1% (v:v) TFA in 90% (v:v) acetonitrile), with a gradient of 0 to 60% solvent B (0 to 36 or 60 min long). The Ultimate 3000 LC system was connected to a Probot MALDI Spotter (LC-Packings). Eluting fractions (0.15 µl) were mixed with an equal volume of α-cyano-hydroxy cinnamic acid matrix (3 mg/ml in 50/50 (v:v) 0.1% TFA / acetonitrile), spotted onto a metal MALDI-TOF/TOF target plate and left to dry at room temperature prior to MS analyses.

Matrix assisted laser desorption ionisation – Time of flight (MALDI-TOF) MS analyses

Derivatised VWF *N*-glycans were dissolved in 20 µl of 1:1 (v:v) methanol:water or VWF peptides were dissolved in 50 µl of 0.1% (v:v) TFA. 1 µl of sample was mixed with 1 µl of the matrix (for glycans in positive ion mode, 10 mg/ml of 2,5-dihydroxybenzoic acid in 50:50 (v:v) methanol:water; for glycans in negative ion mode, 10 mg/ml 3,4-diaminobenzophenone in 50:50 (v:v) acetonitrile/0.1% (v:v) TFA; for peptides, 5 mg/ml in 50/50 (v:v) 0.1% TFA / acetonitrile). 1 µl of the subsequent mixture was spotted onto a 100-well sample plate and dried under vacuum. MALDI-MS analysis was performed using a Voyager DE-STR™ (Applied Biosystems, Warrington, Cheshire, UK) mass spectrometer in the reflectron mode with delayed extraction The instrument was calibrated externally using the Calmix 4700 calibration standard, containing des-Arg1-Bradykinin (Molecular weight 904.46 Da), angiotensin I (Mw 1296.68 Da), human [Glu^1]-fibrinopeptide B (Mw 1570.67 Da), ACTH fragment 1-17 (Mw 2093.08 Da), ACTH fragment 18-39 (Mw 2465.19 Da), and ACTH fragment 7-38 (Mw 3657.92 Da). Data were acquired using Voyager 5 Instrument Control Software and were processed using Data Explorer MS processing software.

Matrix assisted laser desorption ionisation – Time of flight/Time of flight (MALDI-TOF/TOF) MS/MS analyses

Sample-matrix mixtures were prepared as before, spotted onto a MALDI TOF/TOF sample plate and dried under vacuum. MS/MS data were acquired using a 4800 MALDI TOF/TOF™ (Applied Biosystems, Damstadt, Germany) mass spectrometer. The collision energy was set to 1 kV and argon was used as collision gas. The instrument was calibrated using [Glu^1]-fibrinopeptide B human (Sigma-Aldrich, UK) as external calibrant for the MS/MS mode and the 4700 calibration standard, calmix, was used as the external calibrant for the MS mode. Data were acquired using 4000 Series Explorer Instrument Control Software and were processed using Data Explorer MS processing software.

Results

Human plasma VWF was purified from Haemate P, a commercial plasma concentrate manufactured for patients with haemophilia or von Willebrand disease (VWD), using size exclusion and affinity chromatography techniques. The strategy undertaken to achieve VWF glycomics and site-by-site glycosylation analyses is summarised in Figure 1. For glycomics analyses, VWF was reduced, carboxymethylated and digested with trypsin. *N*-glycans were released by PNGase F digestion, purified using a reverse phase C_{18} cartridge then permethylated prior to MALDI-TOF and TOF/TOF analyses. Glycomics experiments were supplemented by enzymatic treatment (neuraminidase, β-galactosyltransferase, endo-β-galactosidase). For glycoproteomics analyses, VWF tryptic digests were desialylated using sialidase A and, in some experiments, enriched by UEA-I affinity chromatography, prior to analyses using an *off-line* C_{18} nano-LC MALDI-TOF/TOF system.

Data from the permethylated VWF *N*-glycans MALDI-TOF MS analysis are shown in Figure 2 and Supplementary Table 1. For glycomics data, all molecular ions are $[M+Na]^+$ and panels are normalised to 100% intensity unless otherwise stated. Structural assignments are based on monosaccharide composition (MALDI-TOF MS, GC-MS), fragmentation analyses (MS/MS), enzymatic digest sequencing and knowledge of the glycan biosynthetic pathways. Bearing in mind that a single molecular ion is often composed of several isoforms (as exemplified in Figure 3), the proposed structures either correspond to the most abundant isoform or are representative of a group carrying a common structural motif. The profile is dominated by sialylated and fucosylated bi-antennary structures (*m/z 2431.2, 2605.3, 2779.4* and *2966.5*). A remarkably high variety of tri- and tetra-antennary *N*-glycans are observed in smaller abundance (*m/z 3402.7, 3793.9, 3922.9, 4314.1, etc*). Minor signals consistent with high mannose structures Man_5 to Man_8 are also detected (*m/z 1579.8, 1783.9, 1988.0* and *2192.1*).

N-glycan fucosylation

A high proportion of molecular ions are consistent with complex type multi-fucosylated *N*-glycans suggesting fucosylation of both core and antennae (*m/z 2418.2, 3041.5, 4025.9 etc.*), in accordance with previous publications reporting the occurrence of ABH blood group antigens on VWF *N*-glycans [15]. The presence of fucose residues was confirmed by gas chromatography (GC)-MS as shown in Table 1 and Supplementary Figure 1. Terminal fucose residues were firmly identified (retention time 17.00 min; EI spectrum in Supplementary Figure 2.A) and core fucosylation was revealed by the detection of 4,6-linked GlcNAc residues (retention time 24.47 min; Supplementary Figure 2.B). Core fucosylation was also demonstrated by CID fragmentation, as shown in Figures 3.A & 3.B, MS/MS spectra of the *m/z 2633.3* and *3054.5* molecular ions providing fragment ions diagnostic of core-fucosylated structures (*m/z 474, 2212* and *2604*).

Antennae fucosylation was demonstrated via the detection of 2-linked galactose residues amongst the PMAA population (retention time = 19.88 min; Supplementary Figure 2.C). MS/MS analyses confirmed the presence of H antigens through the detection of diagnostic fragment ions *m/z* 433 and 660 (Fuc_1Hex_1 and $Fuc_1Hex_1HexNAc_1$, respectively; Figure 3.A). A diagnostic fragment ion at *m/z 905* also demonstrates blood group A antigens within this sub-population of molecular ions. In the same way, blood group B antigens were observed on other VWF *N*-glycans via diagnostic fragment ions at *m/z 619* and *864* (Fuc_1Hex_2 and $Fuc_1Hex_2HexNAc_1$, respectively; data not shown). Finally, no evidence for Lewis or

sialyl Lewis epitopes was found, suggesting that on VWF *N*-glycans, antennae fucosylation is restricted to the position 2 of galactose residues.

Presence of lactosamine extensions

Some high mass molecular ions suggest the presence of lacNAc extensions because the total of lacNAc units is higher than the four needed for a tetra-antennary glycan (Figure 2.D*; m/z 4314.1, 4488.2* and *4675.2*). In addition, CID fragmentation of many lower mass molecular ions such as *m/z 3054.5* yielded fragment ions characteristic of lacNAc elongations (Figure 3.B; *m/z 1296*, *1781* and *2142*), highlighting that this feature is also present in bi- and tri-antennary structures. The presence of lacNAc elongations was also assessed by digesting native VWF *N*-glycans with an endo β-galactosidase prior to permethylation and MS analyses. The low mass region of the subsequent MS spectrum (Supplementary Figure 3.A) contains a variety of molecular ions consistent with lactosaminic elongations, with or without ABH blood group epitopes or a sialic acid residue (*m/z 722.4, 896.5, 1083.6, 1100.6, 1141.6* and *1171.6*). The structure of these elongations was confirmed by MS/MS analysis as shown in Supplementary Figure 3.B & 3.C.

Presence of both bisecting and terminal GlcNAc residues

A number of glycan compositions determined by MS analysis suggest the presence of HexNAc residues additional to those present in the N-glycan core plus the lacNAc building blocks of the antennae (Figure 2; *m/z 2315.2, 2663.3, 3211.6, etc.*). These extra HexNAcs could theoretically be attributed to GalNAc residues forming the A antigen in the periphery, to Sda epitopes (NeuAc $_{\alpha 2,3}$ (GalNAc $_{\beta 1\text{-}4}$) Gal $_{\beta 1\text{-}4}$ GlcNAc), to bisecting GlcNAc residues or to non-galactosylated antennae.

The A antigen was characterised by MS/MS analysis of a number of molecular ions of this sub-population, as shown in Figure 3.A. Global fragmentation data suggest that 15-20 % of the extra HexNAcs are GalNAc involved as an A antigen. The Sda epitope was firmly ruled out by the overall MS/MS data which showed no evidence for the predicted diagnostic fragment ion at *m/z 1092* ($NeuAc_1Hex_1HexNAc_2$). The nature of some of these HexNAc residues was then uncovered by GC-MS analyses, which revealed the presence of 3,4,6-linked mannose within the population of *N*-glycan-derived PMAA (retention time = 21.43 min; Supplementary Figure 2.D). This residue is characteristic of bisected structures in which the 3,6-substituted β-Man is additionally substituted by the so-called bisecting GlcNAc in a $\beta_{1\text{-}4}$ linkage. Finally, in order to confirm or exclude the presence of terminal non-bisected GlcNAc residues (non-galactosylated, or truncated antennae), native *N*-glycans were incubated with $\beta_{1,4}$-galactosyltransferase in the presence of UDP-Gal. This enzyme catalyses the addition of a galactose residue to exposed GlcNAc residues, but, due to steric hindrance, does not have access to bisecting GlcNAc residues. The resulting MALDI-TOF spectra of VWF permethylated *N*-glycans before and after incubation are shown in Supplementary Figure 4**.** Both spectra are normalised to fully capped *N*-glycans which cannot be further modified by the enzyme (*m/z 2779.4* and *2792.4*). Molecular ions consistent with fully galactosylated glycans (*m/z 2070.0, 2244.1, 2417.2, 2431.2, 2605.3* and *2693.5*) are detected at higher intensity after treatment by $\beta_{1,4}$-galactosyltransferase whereas molecular ions consistent with non-fully galactosylated species decrease in intensity (*m/z 2285.2* and *2850.5*) after reaction, demonstrating that a portion of VWF *N*-glycans carry a terminal GlcNAc residue in a non-bisected terminal position. Such truncated antennae are particularly uncommon for blood glycoproteins, on which *N*-glycans tend to be fully processed.

Characterisation of sulphated epitopes

An early report suggested that VWF carry a population of sulphated oligosaccharides [27]. The analysis of sulphated glycans by mass spectrometry is a difficult task, due to their general low abundance, the lability of sulphate groups and their poor ionisation with standard MS techniques. Nevertheless, an optimised sulphoglycomics methodology has been recently reported by Khoo and colleagues [25], and its adaptation to the VWF model allowed us to characterise a family of sulphated *N*-glycans within the oligosaccharide population. The negative ion mode MALDI-TOF MS spectrum of the permethylated VWF *N*-glycans shown in Figure 4 exhibits a number of $[M-H]^-$ molecular ions consistent with mono-sulphated complex type structures (*m/z* 2647.3, 2821.4, 3008.5, 3096.5, 3457.7, 3631.8 and 3818.9). These *N*-glycans are broadly consistent with the positive ion mode profile established earlier (Figure 2.A). A related population of molecular ions was observed (*m/z* 2567.3, 2741.4, 2928.5 and 3377.7), corresponding to glycans having lost sulphite (SO_3) during or after the ionisation process. Negative ion mode MS/MS provided diagnostic fragment ions confirming the presence of sulphate groups. For instance, MS/MS of *m/z* 2647.3 yielded a sulphate ion at *m/z* 97 plus fragment ions consistent with a core-fucosylated bi-antennary mono-sialylated glycan carrying a sulphate on the lacNAc moiety of the sialylated antenna (Supplementary Figure 5; *m/z* 514.0, 700.3, 1808.3 and 2272.0).

Glycoproteomics analyses

Following the characterisation of the VWF *N*-glycans, off-line LC-MS/MS glycoproteomics approaches were applied to the study of individual glycosylation sites. *In silico* tryptic digestion of mature VWF leads to 136 tryptic peptides (Mw > 500 Da), among which 13 contain a single *N*-glycosylation consensus sequon (Table 2; peptide-mass software; www.expasy.org). VWF was reduced, alkylated, trypsinized, and digested by neuraminidase (glycoproteomics data) or PNGase F (proteomics data) prior to analysis using a LC Pepmap C_{18} nanocapillary column connected to a Probot MALDI spotter and an Applied Biosystems 4800 MALDI TOF/TOF instrument. Automated interpretation of the proteomics data using the MASCOT search engine (www.matrixscience.com) confirmed a high purity level with only trace levels of contaminating peptides. Overall, about 60% of the VWF amino sequence (UniProtKB P04275) was successfully detected and sequenced, amongst which were ten tryptic peptides possessing a *N*-glycosylation consensus sequon (Table 2; Supplementary Figure 6). Glycoproteomics analyses were performed manually and led to the observation of various glycosylated forms of each of these ten peptides. With the exception of N_{1147}, all these sites appeared to be fully occupied since their non-glycosylated counterparts were only detected after PNGase F digestion. Comprehensive data are shown for a single N-glycosylation site (Table 2; P4; I_{1509}GEADFNR_{1516}) while other individual site glycoproteomics data are summarised in Supplementary Tables 2 to 11.

Three MALDI-TOF MS spectra obtained from LC fractions eluting between 20 and 22 min are shown in Supplementary Figure 7. They exhibit a considerable number of $[M+H]^+$ ions consistent with various glycosylated forms of the IGEADFNR peptide. MS analyses were complemented by CID fragmentation data such as that shown in Figure 5.A for *m/z 3420.8*. High mass fragment ions are consistent with loss of antennae (*m/z 2528.0, 2909.4* and *3055.5*) whilst the corresponding antennae fragments are observed in the low mass range (*m/z 204.0, 366.1* and *512.2*). The mid-mass range of the spectrum is highly informative and includes the *m/z 921.3*

fragment ion corresponding to the $[M+H]^+$ ion of the peptide itself, accompanied by two major signals at *m/z 904.3* and *1004.4* resulting from asparagine side-chain and GlcNAc cross-ring cleavages, respectively. The three further signals at *m/z 1124.5, 1270.5* and *1327.6* show the IGEADFNR peptide substituted by the *N*-glycan core monosaccharides. To compensate for the lack of information on the amino acid sequence, the VWF tryptic digest was additionally digested by PNGase F prior to nano-LC C_{18} injection. The observation of a molecular ion at *m/z 922.3* consistent with the deglycosylated form of the peptide and its sequencing by CID fragmentation (Supplementary Figures 8.A and 8.B) confirmed the initial IGEADFNR sequence. The N_{1515} glycosylation site was thus demonstrated to carry a heterogeneous population of complex type bi-, tri- and tetra-antennary structures including multi-fucosylated (Supplementary Figure 7; *m/z 2836.3, 3404.5, 3566.6*) as well as "extra HexNAc" species (*m/z 2747.3, 3404.5, 3769.5*). The global description of the IGEADFNR glycopeptide population is shown in Supplementary Table 4.

The importance of localising the blood group antigens along the VWF amino acid sequence led us to perform affinity chromatography using agarose bound fucose-specific Ulex europaeus agglutinin (UEA-I) in order to obtain glycopeptide fractions enriched in the H antigen. UEA-I bound and unbound fractions were analysed on the same *off-line* nano-LC C_{18} MALDI-TOF/TOF system. The UEA-I enrichment step allowed the identification of many glycopeptides containing at least one H antigen, which were masked by more abundant alternative structures in the experiments described earlier. The MS spectrum shown in Figure 5.B was generated by summing the relevant consecutive spectra and therefore shows the overall population of H antigen containing *N*-glycans found at the N_{1515} glycosylation site. Mono- to tetra-fucosylated N-glycans were observed, as well as a number of "extra HexNAc" and lactosamine extension containing species at relatively high abundance. The presence of H antigen at N_{1515} was finally confirmed by CID fragmentation, as shown in the exemplar MS/MS spectrum in Figure 5.C (*m/z 512.2, 2837.3* and *3348.6*).

Unexpectedly, we found that, with the single exception of N_{1147}, the H antigen is present at all detected VWF N-glycosylation sites. Based on the relative glycopeptide proportions established for all sites (Supplementary Tables 2 to 11), we evaluated the average number of complex type antennae and occurrence of antennae fucosylation per glycosylation site (Supplementary Table 12). Interestingly, the highest density of fucosylation was found at N_{2635} (15.9 % of the antennae being fucosylated), while most of the other sites carry *N*-glycans with about 10 % of fucosylated antennae. The three sites N_{857}, N_{2290} and N_{2585} show a much lower degree of antennae fucosylation (1 to 3 %). In addition, the level of branching was found to be reasonably constant, with 2.2-2.3 antennae per glycan on average, with the exceptions of N_{857}/N_{1147}, carrying undersized structures (1.6-2.0 antennae per glycan) and of N_{1515}/N_{1574}, showing a greater proportion of bulky *N*-glycans (2.6 antennae per glycan).

Discussion

The glycomics data presented in this report show the great complexity of the VWF *N*-glycans as a population of approximately 300 distinct structures. Complex type *N*-glycans, mainly bi-antennary, dominate the profile, although traces of high-mannose and truncated structures are also detected (Figure 2 and Supplementary Table 1). This profile is broadly in agreement with the study performed by Matsui and collaborators, although these workers only identified the most abundant *N*-glycans after desialylation, comprising about 25 structures [15]. Our analyses highlight a number of epitopes previously unknown on VWF and reveal the nature of low abundance components that were previously missed. The level of complexity of the

VWF *N*-glycan population is schematised in Figure 6.A and salient structural features are discussed in the next sections.

Core fucosylation and capping of the terminal galactose, with either a $\alpha_{2,6}$ sialic acid or a $\alpha_{1,2}$ fucose residue, are predominant features. Human VWF *N*-glycans show a high sialylation status with about 50% of the overall complex type antennae being capped with a $\alpha_{2,6}$ neuraminic acid residue, while PMAA analyses revealed that about 15% of the VWF *N*-glycan antennae are capped with a $\alpha_{1,2}$ fucose residue (Table 1). The presence of Lewis epitopes was ruled out by the absence of 3,4-linked GlcNAc residues in the linkage data and diagnostic fragment ions in tandem MS analyses. ABH antigens were found in similar abundance on all the families of complex type *N*-glycans, suggesting that $\alpha_{1,2}$ fucosylation of galactose residues is not a structure-specific process (Figure 2 and Table 1). Evidence for antennae containing both a neuraminic acid and a fucose residue was not found, in keeping with evidence indicating that fucosylation of sialylated antennae is restricted to $\alpha_{2,3}$ sialylation in mammals [28]. However, 35% of the total N-glycan compositions identified carry both sialic acid(s) and ABH antigen(s), but on distinct antennae. Finally, in contrast to the VWF *O*-linked glycans reported recently by our group, no evidence for di-sialic acid motifs has been found on *N*-glycans [29]. The $\alpha_{2,6}$ neuraminic acid residues identified here have been shown to have a protective effect against cleavage by serine and cysteine proteases but specifically enhance susceptibility to ADAMTS-13 proteolysis [23, 24]. As reported earlier, ADAMTS-13 activity is affected by ABH antigens and it is interesting to note that VWF desialylation abolishes the ABH blood group effect on the ADAMTS-13 proteolysis rate. Terminal $\alpha_{2,6}$ neuraminic residues are also recognised by the sialic acid binding Ig-like lectin family, the Siglecs [30, 31]. Although the individual affinity studies are still in progress, the Consortium for Functional Glycomics data suggest that Siglecs 2, 5, 7, 9 and 10 present a significant affinity for $\alpha_{2,6}$ NeuAc residues. Interestingly, Lenting and collaborators recently reported preliminary data suggesting that Siglecs 5, 7 and 9 are able to interact with VWF [32].

VWF has been reported once to carry sulfated *N*-glycans at the N_{1147} and/or N_{1231} glycosylation sites, although until now, this early observation has not been corroborated with structural evidence [27]. Using a modern MS-based methodology, a pool of about 10 mono-sulfated *N*-glycans was firmly identified here (Figure 4), comprising bi- and tri-antennary structures containing at least one fucose and one NeuAc residue. From our MS/MS data, we conclude that sulfation occurs mainly on the sialylated LacNAc antennae (Supplementary Figure 5). Antennae sulfation may have important implications, in particular regarding recognition by specific glycan binding proteins. Human B cells expressing Siglec 2 have also been shown to interact with $\alpha_{2,6}$sialylated 6-sulfo-LacNAc epitopes [33]. Although the nature of the recognition remains unclear, VWF is known to interact with P-selectin and it is believed that this interaction is key to anchoring high molecular weight VWF to the surface of stimulated endothelial cells, thus facilitating their cleavage by ADAMTS-13 [34]. Interaction between P-selectin and its major ligand (P-selectin glycoprotein ligand-1) is known to be mediated by both a sialyl $Lewis^X$ epitope and an adjacent sulfated tyrosine residue. Although $Lewis^X$ is absent on VWF *N*-glycans, sulfation of an antenna could conceivably, in conjunction with fucosylation of another nearby antenna, constitute a motif recognised by P-selectin.

Our glycoproteomics data showed that *N*-glycans substituted with blood group H-antigen are present on all but one of the observed glycosylation sites although its abundance is not uniform as represented in Figure 6.B. Glycans of the N-terminal domain (N_{857} and N_{1147}) lack significant antennae fucosylation. N_{2290} and N_{2585} are rather poorly fucosylated relative to other sites, of which about 10% of the antennae are substituted with a $\alpha_{1,2}$ fucose (Supplementary Table 12). The most heavily fucosylated site is N_{2635}, on which N-glycans carry the H antigen on about 16% of its antennae which is surprising and important, as this later glycosylation sequon was

until now reported as not glycosylated. The average number of antennae per glycans at individual sites appears to be reasonably constant, with 2.2-2.3 per glycan at most of the sites, with the exception of N_{857} and N_{1147} carrying smaller structures and of N_{1515} and N_{1574} whose *N*-glycans are somewhat more bulky (an average of 2.6 antennae per glycan). These two glycosylation sites flank the ADAMTS-13 cleavage site, and our data suggest that this region is richer in the H-antigen than the rest of the molecule. Whether the H antigen density affects cleavage remains to be established. The trend of structural characteristics observed on individual glycosylation sites of desialylated VWF is summarised in Figure 6.B and detailed in Supplementary Tables 2 to 11.

No satisfactory understanding of the correlation of ABH antigen expression to VWF cleavage by ADAMTS-13 has been provided to date. Possible explanations include steric hindrance affecting ADAMTS-13 cleavage, potential recognition of ABH antigens by ADAMTS-13, or an effect of the ABH status on the overall glycosylation. A recent report provided interesting evidence that desialylation of VWF abolishes the ABH blood group effect on the ADAMTS-13 cleavage rate, suggesting a relationship between the blood group status and the level of sialylation [22, 23]. It has also been suggested that blood group O-VWF could be more extensively sialylated than A- and B-VWF. It is conceivable that bulky A and B epitopes on one or more antennae sterically hinder sialic acid addition to other antennae on the same or nearby glycans. To address this issue, lectin purification strategies should be developed in the future to allow the enrichment of sialylated forms of blood group-substituted glycopeptides.

To our knowledge, no other glycoprotein of this complexity has had its glycome and glycoproteome characterised in the depth achieved in the current study. This work highlights the ability of modern mass spectrometry-based techniques to provide highly detailed information concerning the diversity of glycoforms of extremely complex glycoproteins which should prove relevant for understanding their correspondingly complex functions. The well-known association between ABO blood group and plasma concentration of VWF is well documented and although most evidence suggests that it is mediated by an effect on clearance, the responsible receptors have not been fully identified. The Ashwell (asialoglycoprotein) receptor appears to make some contribution [35] but additional receptors remain to be identified and knowing that their ligands must exist among the glycan structures described in this report should aid their identification. Notably, a recent genome-wide association study identified the CLEC4M lectin locus as a novel determinant of VWF plasma level [36]. Proteolysis is known to be an important mechanism for regulation of VWF function and may also play a role in clearance. We have already demonstrated that N-linked glycans can modulate VWF interaction with the regulatory enzyme ADAMTS-13 [37] and that a complex balance between ABO antigen and sialic acid content determines the overall susceptibility of VWF to proteolytic degradation [23]. More recent data have indicated that the VWF interacts with the members of the galectin family via its carbohydrate moieties and that this has an effect on VWF function [38]. The significance of these interactions and their dependence on specific glycan structures at specific sites in the VWF molecule are not yet elucidated but will be facilitated by the data presented here. Finally, VWF interacts with many other molecules, both before and after release from the endothelial cell. These include the numerous contents of the Weibel Palade bodies which are largely co-located by binding to VWF and we anticipate these too will be subject to regulation by the VWF glycans.

Acknowledgments

This work was supported by the British Heart Foundation [FS/06/069/21490] and the Biotechnology and Biological Sciences Research Council [SF19107, B19088, and BBF008309/1]. We are also grateful for support from the NIHR Biomedical Research Centre Funding Scheme.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Contribution: K.C., T.A.M., A.N., S.M.H., M.P., H.R.M., M.A.L. & A.D. designed the research; K.C., T.A.M. & M.P. performed the experiments; K.C., S.M.H., M.P., H.R.M., M.A.L. & A.D. analyzed the data; K.C., S.M.H., M.P., H.R.M., M.A.L. & A.D. wrote the paper.

References

1 Sakariassen, K. S., Bolhuis, P. A. and Sixma, J. J. (1979) Human blood platelet adhesion to artery subendothelium is mediated by factor VIII-Von Willebrand factor bound to the subendothelium. Nature **279**, 636-638

2 Fay, P. J. (2004) Activation of factor VIII and mechanisms of cofactor action. Blood Rev **18**, 1-15

3 Fay, P. J., Coumans, J. V. and Walker, F. J. (1991) von Willebrand factor mediates protection of factor VIII from activated protein C-catalyzed inactivation. J Biol Chem **266**, 2172-2177

4 Reininger, A. J. (2008) Function of von Willebrand factor in haemostasis and thrombosis. Haemophilia **14 Suppl 5**, 11-26

5 Foster, P. A., Fulcher, C. A., Marti, T., Titani, K. and Zimmerman, T. S. (1987) A major factor VIII binding domain resides within the amino-terminal 272 amino acid residues of von Willebrand factor. J Biol Chem **262**, 8443-8446

6 Mazzucato, M., Spessotto, P., Masotti, A., De Appollonia, L., Cozzi, M. R., Yoshioka, A., Perris, R., Colombatti, A. and De Marco, L. (1999) Identification of domains responsible for von Willebrand factor type VI collagen interaction mediating platelet adhesion under high flow. J Biol Chem **274**, 3033-3041

7 Lankhof, H., van Hoeij, M., Schiphorst, M. E., Bracke, M., Wu, Y. P., Ijsseldijk, M. J., Vink, T., de Groot, P. G. and Sixma, J. J. (1996) A3 domain is essential for interaction of von Willebrand factor with collagen type III. Thromb Haemost **75**, 950-958

8 De Marco, L., Girolami, A., Zimmerman, T. S. and Ruggeri, Z. M. (1985) Interaction of purified type IIB von Willebrand factor with the platelet membrane glycoprotein Ib induces fibrinogen binding to the glycoprotein IIb/IIIa complex and initiates aggregation. Proc Natl Acad Sci U S A **82**, 7424-7428

9 Beacham, D. A., Wise, R. J., Turci, S. M. and Handin, R. I. (1992) Selective inactivation of the Arg-Gly-Asp-Ser (RGDS) binding site in von Willebrand factor by site-directed mutagenesis. J Biol Chem **267**, 3409-3415

10 Michaux, G., Pullen, T. J., Haberichter, S. L. and Cutler, D. F. (2006) P-selectin binds to the D'-D3 domains of von Willebrand factor in Weibel-Palade bodies. Blood **107**, 3922-3924

11 Fujimura, Y., Titani, K., Holland, L. Z., Roberts, J. R., Kostel, P., Ruggeri, Z. M. and Zimmerman, T. S. (1987) A heparin-binding domain of human von Willebrand factor. Characterization and localization to a tryptic fragment extending from amino acid residue Val-449 to Lys-728. J Biol Chem **262**, 1734-1739

12 Titani, K., Kumar, S., Takio, K., Ericsson, L. H., Wade, R. D., Ashida, K., Walsh, K. A., Chopek, M. W., Sadler, J. E. and Fujikawa, K. (1986) Amino acid sequence of human von Willebrand factor. Biochemistry **25**, 3171-3184
13 Debeire, P., Montreuil, J., Samor, B., Mazurier, C., Goudemand, M., van Halbeek, H. and Vliegenthart, J. F. (1983) Structure determination of the major asparagine-linked sugar chain of human factor VIII--von Willebrand factor. FEBS Lett **151**, 22-26
14 Samor, B., Michalski, J. C., Debray, H., Mazurier, C., Goudemand, M., Van Halbeek, H., Vliegenthart, J. F. and Montreuil, J. (1986) Primary structure of a new tetraantennary glycan of the N-acetyllactosaminic type isolated from human factor VIII/von Willebrand factor. Eur J Biochem **158**, 295-298
15 Matsui, T., Titani, K. and Mizuochi, T. (1992) Structures of the asparagine-linked oligosaccharide chains of human von Willebrand factor. Occurrence of blood group A, B, and H(O) structures. J Biol Chem **267**, 8723-8731
16 Wagner, D. D., Mayadas, T. and Marder, V. J. (1986) Initial glycosylation and acidic pH in the Golgi apparatus are required for multimerization of von Willebrand factor. J Cell Biol **102**, 1320-1324
17 McKinnon, T. A., Goode, E. C., Birdsey, G. M., Nowak, A. A., Chan, A. C., Lane, D. A. and Laffan, M. A. Specific N-linked glycosylation sites modulate synthesis and secretion of von Willebrand factor. Blood **116**, 640-648
18 Gralnick, H. R., Williams, S. B. and Rick, M. E. (1983) Role of carbohydrate in multimeric structure of factor VIII/von Willebrand factor protein. Proc Natl Acad Sci U S A **80**, 2771-2774
19 Sodetz, J. M., Pizzo, S. V. and McKee, P. A. (1977) Relationship of sialic acid to function and in vivo survival of human factor VIII/von Willebrand factor protein. J Biol Chem **252**, 5538-5546
20 Gill, J. C., Endres-Brooks, J., Bauer, P. J., Marks, W. J., Jr. and Montgomery, R. R. (1987) The effect of ABO blood group on the diagnosis of von Willebrand disease. Blood **69**, 1691-1695
21 Miller, C. H., Haff, E., Platt, S. J., Rawlins, P., Drews, C. D., Dilley, A. B. and Evatt, B. (2003) Measurement of von Willebrand factor activity: relative effects of ABO blood type and race. J Thromb Haemost **1**, 2191-2197
22 Bowen, D. J. (2003) An influence of ABO blood group on the rate of proteolysis of von Willebrand factor by ADAMTS13. J Thromb Haemost **1**, 33-40
23 McGrath, R. T., McKinnon, T. A., Byrne, B., O'Kennedy, R., Terraube, V., McRae, E., Preston, R. J., Laffan, M. A. and O'Donnell, J. S. (2010) Expression of terminal alpha2-6-linked sialic acid on von Willebrand factor specifically enhances proteolysis by ADAMTS13. Blood **115**, 2666-2673
24 Berkowitz, S. D. and Federici, A. B. (1988) Sialic acid prevents loss of large von Willebrand factor multimers by protecting against amino-terminal proteolytic cleavage. Blood **72**, 1790-1796
25 Yu, S. Y., Wu, S. W., Hsiao, H. H. and Khoo, K. H. (2009) Enabling techniques and strategic workflow for sulfoglycomics based on mass spectrometry mapping and sequencing of permethylated sulfated glycans. Glycobiology **19**, 1136-1149
26 Matsui, T., Hamako, J., Ozeki, Y. and Titani, K. (2001) Comparative study of blood group-recognizing lectins toward ABO blood group antigens on neoglycoproteins, glycoproteins and complex-type oligosaccharides. Biochim Biophys Acta **1525**, 50-57
27 Carew, J. A., Browning, P. J. and Lynch, D. C. (1990) Sulfation of von Willebrand factor. Blood **76**, 2530-2539
28 Chandrasekaran, E. V., Xue, J., Xia, J., Chawda, R., Piskorz, C., Locke, R. D., Neelamegham, S. and Matta, K. L. (2005) Analysis of the specificity of sialyltransferases toward mucin core 2, globo, and related structures. identification of the sialylation sequence and the effects of sulfate, fucose,

methyl, and fluoro substituents of the carbohydrate chain in the biosynthesis of selectin and siglec ligands, and novel sialylation by cloned alpha2,3(O)sialyltransferase. Biochemistry **44**, 15619-15635

29 Canis, K., McKinnon, T. A., Nowak, A., Panico, M., Morris, H. R., Laffan, M. and Dell, A. (2010) The plasma von Willebrand factor O-glycome comprises a surprising variety of structures including ABH antigens and disialosyl motifs. J Thromb Haemost **8**, 137-145

30 Crocker, P. R. and Redelinghuys, P. (2008) Siglecs as positive and negative regulators of the immune system. Biochem Soc Trans **36**, 1467-1471

31 McMillan, S. J. and Crocker, P. R. (2008) CD33-related sialic-acid-binding immunoglobulin-like lectins in health and disease. Carbohydr Res **343**, 2050-2056

32 Lenting, P. J., Pegon, J. N., Christophe, O. D. and Denis, C. V. (2010) Factor VIII and von Willebrand factor--too sweet for their own good. Haemophilia **16 Suppl 5**, 194-199

33 Kannagi, R., Ohmori, K. and Kimura, N. (2009) Anti-oligosaccharide antibodies as tools for studying sulfated sialoglycoconjugate ligands for siglecs and selectins. Glycoconj J **26**, 923-928

34 Padilla, A., Moake, J. L., Bernardo, A., Ball, C., Wang, Y., Arya, M., Nolasco, L., Turner, N., Berndt, M. C., Anvari, B., Lopez, J. A. and Dong, J. F. (2004) P-selectin anchors newly released ultralarge von Willebrand factor multimers to the endothelial cell surface. Blood **103**, 2150-2156

35 Grewal, P. K., Uchiyama, S., Ditto, D., Varki, N., Le, D. T., Nizet, V. and Marth, J. D. (2008) The Ashwell receptor mitigates the lethal coagulopathy of sepsis. Nat Med **14**, 648-655

36 Smith, N. L., Chen, M. H., Dehghan, A., Strachan, D. P., Basu, S., Soranzo, N., Hayward, C., Rudan, I., Sabater-Lleal, M., Bis, J. C., de Maat, M. P., Rumley, A., Kong, X., Yang, Q., Williams, F. M., Vitart, V., Campbell, H., Malarstig, A., Wiggins, K. L., Van Duijn, C. M., McArdle, W. L., Pankow, J. S., Johnson, A. D., Silveira, A., McKnight, B., Uitterlinden, A. G., Aleksic, N., Meigs, J. B., Peters, A., Koenig, W., Cushman, M., Kathiresan, S., Rotter, J. I., Bovill, E. G., Hofman, A., Boerwinkle, E., Tofler, G. H., Peden, J. F., Psaty, B. M., Leebeek, F., Folsom, A. R., Larson, M. G., Spector, T. D., Wright, A. F., Wilson, J. F., Hamsten, A., Lumley, T., Witteman, J. C., Tang, W. and O'Donnell, C. J. (2010) Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. Circulation **121**, 1382-1392

37 McKinnon, T. A., Chion, A. C., Millington, A. J., Lane, D. A. and Laffan, M. A. (2008) N-linked glycosylation of VWF modulates its interaction with ADAMTS13. Blood **111**, 3042-3049

38 Saint-Lu, N., Oortwijn, B. D., Pegon, J. N., Odouard, S., Christophe, O. D., de Groot, P. G., Denis, C. V. and Lenting, P. J. (2012) Identification of galectin-1 and galectin-3 as novel partners for von Willebrand factor. Arterioscler Thromb Vasc Biol **32**, 894-901

Table 1. GC-MS linkage analysis of partially methylated alditol acetates derived from VWF *N*-glycans.

The elution time is in minutes and the relative abundance is normalised to 1.00, abundance of the major component, 2-linked mannose.

Elution time	Diagnostic ions	Assignment	Relative abundance	% of the population
17.00	102, 115, 118, 131, 162, 175	Terminal Fucose	0.15	5.2
18.48	102, 118, 129, 145, 161, 162, 205	Terminal Mannose	trace	< 1.0
18.76	102, 118, 129, 145, 161, 205	Terminal Galactose	0.29	10.2
19,66	129, 130, 145, 161, 190, 205, 234	2-linked Mannose	1.00	35.2
19.88	129, 130, 145, 161, 190, 234	2-linked Galactose	0.12	4.2
19.92	118, 129, 161, 202, 203, 234, 277	3-linked Mannose	0.08	2.8
20.45	102, 118, 129, 162, 189, 233	6-linked Galactose	0.38	13.3
20.82	113, 130, 173 190, 233, 274	2,4-linked Mannose	0.10	5.3
21.20	129, 130, 189, 190, 233, 234	2,6-linked Mannose	0.03	1.0
21,38	101, 118, 129, 189, 202, 234, 305	3,6-linked Mannose	0.33	11.6
21.83	118, 129, 139, 202, 259, 333	3,4,6-linked Mannose	0.05	1.7
22,31	117, 129, 143, 145,159, 203, 273	Terminal GlcNAc	0.03	1.0
22,31	117, 129, 143, 145,159, 203, 273	Terminal GalNAc		
23.20	113, 117, 143, 159, 203, 233	4-linked GlcNAc	0.32	11.2
24.47	117, 127, 143, 159, 261, 301	4,6-linked GlcNAc	0.05	1.7

Table 2. Properties of VWF *N*-glycosylation consensus sequons.

The numbering of the VWF tryptic peptides that contain a N-glycosylation consensus sequon does not include the propeptide moieties and therefore begins at amino acid 764.

Peptide number	Predicted N-glycosylated sites	Tryptic peptide sequences	$[M+H]^+$	*Mapped*
P1	N_{857}	W_{856}**N**CTDHVCDATCSTIGMAHYLTFDGLK$_{882}$	3176.3	√
P2	N_{1147}	Y_{1146}**N**SCAPACQVTCQHPEPLACPVQCVEGCHA HCPPGK$_{1181}$	4183.6	√
P3	N_{1231},	V_{1212}TLNPSDPEHCQICHCDVV**N**LTCEACQEPGG VVPPTDAPVSPTTLYVEDISEPPLHDFYCS R_{1274}	7198.2	No
P4	N_{1515}	I_{1509}GEADF**N**R_{1516}	921.4	√
P5	N_{1574}	Y_{1570}QGG**N**R_{1575}	694.3	√
P6	N_{2223}	H_{2219}CDG**N**VSSCGDHPSEGCFCPPDK$_{2241}$	2622.9	√
P7	N_{2290}	V_{2289}**N**CTTQPCPTAK$_{2300}$	1378.6	√
P8	N_{2357}	G_{2342}LQPTLTNPGECRP**N**FTCACR$_{2363}$	2452.1	√
P9	N_{2400}	T_{2386}QCCDEYECACNCV**N**STVSCPLGYLASTATN DCGCTTTTCLPDK$_{2429}$	5123.9	No
P10	N_{2546}	**N**$_{2546}$VSCPQLEVPVCPSGFQLSCK$_{2566}$	2409.1	√
P11	N_{2585}	M_{2579}EACML**N**GTVIGPGK$_{2593}$	1578.7	√
P12	N_{2635}	E_{2633}E**N**NTGECCGR$_{2643}$	1327.5	√
P13	N_{2790}	T_{2779}EPMQVALHCT**N**GSVVYHEVLNAMECK$_{2805}$	3119.4	No

Figure legends

Figure 1. Overview of structural glycomics and glycoproteomics strategies employed in this study. * = *β_{1-4} galactosyltransferase, endo β-galactosidase or neuraminidase*

Figure 2. MALDI-TOF-MS profile of the permethylated N-linked glycans from VWF

The full scale spectrum (**A**) and subsequent magnified portions (**B** to **D**; m/z 1500-2500, 2500-3500 and 3500-5000, respectively) show the high complexity of the VWF glycan population. Structural assignments were based on compositions assigned from molecular weights, complemented by MS/MS information and the results of glycosidase digests. As discussed in the text, individual signals often correspond to a number of isomers, but for clarity only the most abundant isoform or one representative of a family carrying a common structural motif is represented. All molecular ions are $[M+Na]^+$ and panels are normalised to 100% intensity. Structural assignments are based on monosaccharide composition, fragmentation analyses, and knowledge of the glycan biosynthetic pathways. The sugar symbols are those employed by the Consortium for Functional Glycomics (www.functionalglycomics.org).

Figure 3. MALDI-TOF-TOF fragment ion spectra of the parent ions at m/z 2663 and 3054.

These MS/MS spectra obtained after collisional activation reveal diagnostic fragmentation patterns consistent with the structures shown in the inset. The horizontal arrows depict the loss of the designated glycan moieties from the $[M+Na]^+$ precursor ion. Monosaccharide symbols are as shown in Figure 2. (**A**) The fragmentation pattern of the parent ion *m/z* 2663 reveals the presence of two distinct isomeric glycans differing in the position of a terminal HexNAc residue. Fragment ions at *m/z* 660 and 905 are consistent with terminal blood group A, while the fragment ion at *m/z* 433 confirms the presence of blood group H epitope. In the lower structure representation, the terminal non-reducing GlcNAc residue could also be a bisecting GlcNAc. (**B**) Collisional activation of the m/z 3054 parent ion provides evidence for numerous isoforms as well as the presence of lactosaminic elongations (diagnostic fragment ions at *m/z* 1296, 1781 and 2142). The three core fucosylated structures that are shown are assigned as the major components based on fragment ion abundances. However, the antenna fragment ion at *m/z* 660 indicates the presence of other components carrying a fucosylated *N*-acetyl lactosamine antenna and therefore lacking core fucosylation (not represented). Taking into account different antennae and core fucosylation combinations, at least 12 different isoforms can be attributed to this single molecular ion.

Figure 4. The VWF *N*-glycan repertoire contains sulphated oligosaccharides.

Analyses of the permethylated VWF *N*-glycans in the negative ion mode afforded a number of $[M-H]^-$ molecular ions consistent with mono-sulphated complex type *N*-glycans (*m/z* 2647, 2821, 3008, etc.; sulphate indicated by a S in a red circle). A related family of molecular ions lacking an 80 Da moiety is also detected (*m/z* 2567, 2741, 2928 etc., indicated with a blue star). This series is assigned to in-source loss of sulphite from each of the sulphated glycans.

Figure 5. Glycoproteomic analyses of the N_{1515} glycosylation site (glycopeptide P4).

MS and MS/MS spectra were obtained by analysis of a desialylated tryptic digest of VWF. (**A**) MS/MS analysis of parent ion *m/z* 3420 shows fragments resulting from the loss of antennae whose compositions are shown on the horizontal arrows (*m/z* 2832, 2909 and 3055). Antennae fragment ions are observed in the low mass range (*m/z* 204, 366 and 512) while mid-mass fragment ions provide evidence for the identity of the peptide and the glycan core (*m/z* 904, 921, 1004). The fucose is found on either the core GlcNAc (*m/z* 1270) or the antennae (*m/z* 512 and 1124). (**B**) The MALDI-TOF profile of UEA-I bound desialylated P4 glycopeptide (see Table 2 for the sequence of the P4 peptide). A remarkable variety of H-antigen containing N-glycans are observed. (**C**) MS/MS analysis of the parent ion *m/z 3859* shows the presence of several H antigens on a single glycan (*m/z* 512 and 2837), one being potentially presented on a lactosamine extension (*m/z* 2674).

Figure 6. The VWF glycome and site-specificity.

(**A**) Representation of the human VWF *N*-glycan structural features. This figure highlights the set of epitopes observed along this study, comprising ABH blood groups, LacNAc extensions, bisecting and terminal GlcNAc and sulphated sialylated LacNAc antennae. (**B**) The functional domains of VWF are shown together with information on binding partners, the ADAMTS13 cleavage site and the dimerization region. The N-glycosylation sites that have been characterised in this study are shown as red circles and the structures of the most abundant glycans at each site are shown in the annotations. X indicates N-glycosylation sites whose glycosylation has not yet been defined.