Supporting Information

N- and O-Glycosylation Analysis of Etanercept using Liquid Chromatography and Quadrupole Time-of-Flight Mass Spectrometry Equipped with Electron Transfer Dissociation Functionality

*Stephane Houel[1], Mark Hillard[2], Ying Qing Yu[1], Niaobh McLoughlin[2], Silvia Millan Martin[2], Pauline M. Rudd[2], Jonathan P. Williams[3], Weibin Chen[1]**

[1]Late Stage Development, Pharmaceutical Life Sciences, Waters Corporation, Milford, MA 01757, USA.

[2]National Institute for Bioprocessing Research and Training, Fosters Avenue, Mount Merrion, Blackrock, Co. Dublin, Ireland

[3]Waters Corporation, Atlas Park, Wythenshawe, Manchester, UK. M22 5PP

*Author for correspondence, Weibin_Chen@Waters.Com

Details of Materials and Methods

Materials

Etanercept (Enbrel, Wyeth, Andover, MA) was purchased as a lyophilized product from Besse Medical (West Chester, OH). Sequencing grade trypsin was obtained from Promega (Madison, WI). A proteolytic enzyme, FabRICATOR (IdeS), was purchased as lyophilized powder from QED Biosciences (San Diego, CA). Peptide glycosidase F (PNGase F) and other exoglycosidases used for N-glycan characterization were purchased from Prozyme (San Leandro, CA). TrisHCl (1M pH 7.6), Guanidine hydrochloride, 8 M urea, 2-aminobenzamide (2AB), Iodoacetamide, 1,3-dicyanobenzene and formic acid (FA) were from Sigma-Aldrich (St. Louis, MO). HPLC grade acetonitrile was from Thermo Fisher Scientific (Fairlawn, NJ). EDEGLY mixed enzyme kit (Sigma, St. Louis, MO) was used to release or partially release the O-glycans from Etanercept.

Sample Preparations

N-glycan release, fluorescent labeling and exoglycosidase digestion

Detailed method for releasing and labeling with 2AB and for the N-linked glycan structure analysis using exoglycosidase arrays and retention time alignment is detailed in the literature.[21] Labeled glycans were digested in 10 µL of 50 mM sodium acetate buffer, pH 5.5 for 18 h at 37 °C using arrays of the following enzymes: ABS, *Arthrobacter ureafaciens* sialidase (EC 3.2.1.18, releases α2-3,6,8 linked non-reducing terminal sialic acid) 0.5 mU/µL; NAN1, *Streptococcus pneumonia* sialidase (EC 3.2.1.18, releases α2-3 linked non-reducing terminal sialic acids); BKF, bovine kidney α-fucosidase (EC 3.2.1.51, releases α1-2,6 linked non-reducing terminal fucose residues more efficiently than α1-3,4 linked fucose, digests core α1-6 fucose) 0.8 mU/µL; BTG, bovine testes β-galactosidase (EC 3.2.1.23, hydrolyses non-reducing terminal β1-4 and β1-3 linked galactose) 40 mU/µL; and SPG,*S. pneumonia β-galactosidase* (EC 3.2.1.23) hydrolyzes nonreducing terminal galactose β(1-4) linkages 2 mU/µL. GUH, hexosaminidase cloned from *Streptococcus pneumoniae* expressed in *E. coli* (EC 3.2.1.30, releases GlcNAc residues but not a bisecting GlcNAc linked to Man) 1 mU/µL. After incubation, enzymes were removed by filtration using Pall spin filters 10 kDa (Pall Corporation). N-glycans were then analyzed by UPLC in hydrophilic interaction liquid chromatography (HILIC) with a fluorescence detector.

O-linked glycan release

The Etanercept sample was re-suspended in 28% $NH_3{\cdot}H_2O$ saturated with $(NH_4)_2CO_3$ and incubated at 60°C for 16 h, and salt was removed with porous graphitic carbon (PGC) (Thermo Scientific Hypersep HyperCarb cartridges). All sample were fluorescently derivatized via reductive amination with mixed with sodium cyanoborohydride in 30% (v/v) acetic acid in DMSO at 65 °C for 2 h. Excess fluorophore was removed with filter paper. Exoglycosidase digestion arrays were performed as described above and all samples were run on UPLC with a fluorescence detector.

De-sialylation and tryptic digestion of Etanercept

Ten microliters of Etanercept protein stock solution (50 µg/µL) was mixed with 325 µL of 8M urea and 125 µL of 1M Tris-HCl (pH 7.6), and the protein was subsequently reduced with the addition of 3 µL of 0.5 M DTT. The reduction with DTT was incubated at room temperature for 10 minutes. After DTT reduction step, the sample was cooled to room temperature before it was alkylated with 5 µL of 0.5M Iodoacetamide for 20 minutes at room temperature in the dark. The protein was buffer exchanged using a NAP-5 column (GE Healthcare Life Sciences) to 0.75 ml of 0.1M Tris-HCl (pH 7.6). The estimated protein concentration was about 0.67 µg/µL. Two hundred microliters of the protein solution was incubated with PNGase F and sialidase for 5 hours at 37 ºC. This step was performed to remove all N-linked glycans and the sialic acid residues on O-linked glycans. The deglycosylated protein was incubated with 10 µg of trypsin overnight at 37 ºC to generate tryptic peptides.

Preparation of Fc subunit of Etanercept

Preparation of Fc-fragment of Etanercept fusion protein was achieved using enzyme FabRICATOR (IdeS), and the digestion procedure follows the protocol recommended by the manufacturer. In brief, Etanercept (200 µg) was re-suspended in 50 mM PBS, and 1 unit of FabRICATOR was added for very microgram of sample. The mixture was incubated at 37 °C for 30 min, and the Fc and TNF-α receptor components were fractionated on a UPLC BEH C4 column (Waters, Milford, MA). Samples were manually collected, and N-glycans were released with PNGase F as previously described and subsequently analyzed by UPLC-HILIC-FLR.

Liquid Chromatography and Mass Spectrometry

HILIC-UPLC/FLR

Analysis of the labeled N-glycans was performed on a Waters ACQUITY UPLC BEH Glycan column (2.1×150 mm, 1.7 µm particle).[22]..Instrumentation used included a ACQUITY UPLC with a fluorescence detector (Waters Corporation, Millford, MA, USA) under the control of Empower chromatography workstation software. Mobile phase A is 50 mM ammonium formate buffer (pH 4.4) and mobile phase B is acetonitrile. A linear gradient of 30 – 47% mobile phase A under the flow rate of 561 µL/min was used for the separation. The Fluorescent wavelengths were set at λex = 330 nm and λem = 420 nm.

Experimental conditions for peptide mapping (LC-MS^E methods)

A multiplexed data acquisition method (MS^E) was employed for the mass spectrometric analysis of the tryptic digest of Etanercept.[23] The LC/MS^E data was acquired on a quadrupole time-of-flight mass spectrometer (Synapt G2-S HDMS, Waters Corporation) equipped with electron transfer dissociation (ETD) functionality. An ACQUITY UPLC I-Class system (Waters Corporation) was coupled to the mass spectrometer as an inlet system. Tryptic peptides produced from the digestion of Etanercept were separated on an Acquity BEH UPLC C18

column (2.1 × 150 mm, 1.7 µm). Typically, 0.6 µg of protein digest was injected onto the analytical column for analysis. Mobile phase A was an aqueous solution containing 0.1% formic acid and mobile phase B was acetonitrile with 0.1% formic acid. The peptides were eluted from the column using a gradient from 3 to 35% B over 90 min at a flow rate of 0.2 mL/min and a column temperature of 65°C. At the end of 90-min elution gradient, the column was washed using 80% B for 4 minutes, and re-equilibrated at 3% B for 10 min before the next injection.

For all measurements, the mass spectrometer was operated in positive ESI ion mode with a typical resolving power of 20,000 FWHM. Data were acquired in the continuum mode over a m/z range of 100 - 2000, using a capillary voltage of 3.0 kV, a source temperature of 120 °C, a cone voltage of 40 V and a source offset of 60V. The desolvation temperature was set to 350 °C and the desolvation gas flow rate was 600 L/hour.

The LC/MS^E data was collected by alternating the collision energy of the MS instrument between low energy (MS) and higher energy (MS^E) without precursor selection. The spectral acquisition time at each energy setting was 0.5 s such that one spectrum of MS and MS^E data was acquired every second. In the low energy MS mode the data was acquired with a collision energy of 10 eV and ramped from 20 to 45 eV in the higher energy mode. The collision energy of the transfer cell was held constant at 4 eV. A solution of 50 fmol/uL Glu1-fibrinopeptide B (GFP) in 50% acetonitrile with 0.1% FA was used as a lock-mass solution. The solution was delivered at a flow rate of 5 µL/min using the embedded fluidic system on the mass spectrometer. The data was 'lock-mass' corrected using 785.8421 to correct mass-scale drift. The reference lock mass data was sampled every 30 sec using 0.5 sec scans over the same mass range.

ETD setup and experimental methods

ETD experiments was performed on a hybrid quadrupole / ion mobility / oa-ToF mass spectrometer (Synapt G2-S), Waters Corporation, Manchester, UK) fitted with electron transfer dissociation (ETD) functionality. The instrument has been described in detail elsewhere[24,25] and as such only a brief description will be provided here. The instrument comprises three consecutive, gas filled, travelling wave (T-Wave) RF stacked ring ion guides prior to the ToF mass analyser. For ETD type fragmentation, a sub-ambient pressure (~2 mbar) glow discharge anion source operated at 90µA was used to fill the Trap (first) T-Wave with quadrupole mass selected ETD reagent anions formed from 1,3-dicyanobenzene (m/z 128). Nitrogen 'make-up' gas was used to carry the vapor of 1,3 DCB into the glow discharge region. During an acquisition, the source polarity and quadrupole set mass are switched to allow multiply charged cations formed from ESI of the peptides to interact with stored reagent anions in the Trap T-Wave. This interaction allows an ion-ion type reaction resulting in ETD product ions. For efficient ETD, within the Trap T-Wave, the helium gas was set to a pressure of 0.05 mbar. The Transfer T-Wave was pressurized to 0.005mbar with argon. The Trap T-Wave speed and amplitude which influence the ion-ion interaction time as well as the reaction rate were set to 300m/sec and 0.3V respectively. The source temperature was set at 110 °C, and the sampling cone and the source offset voltage were 25 V and 10 V, respectively.

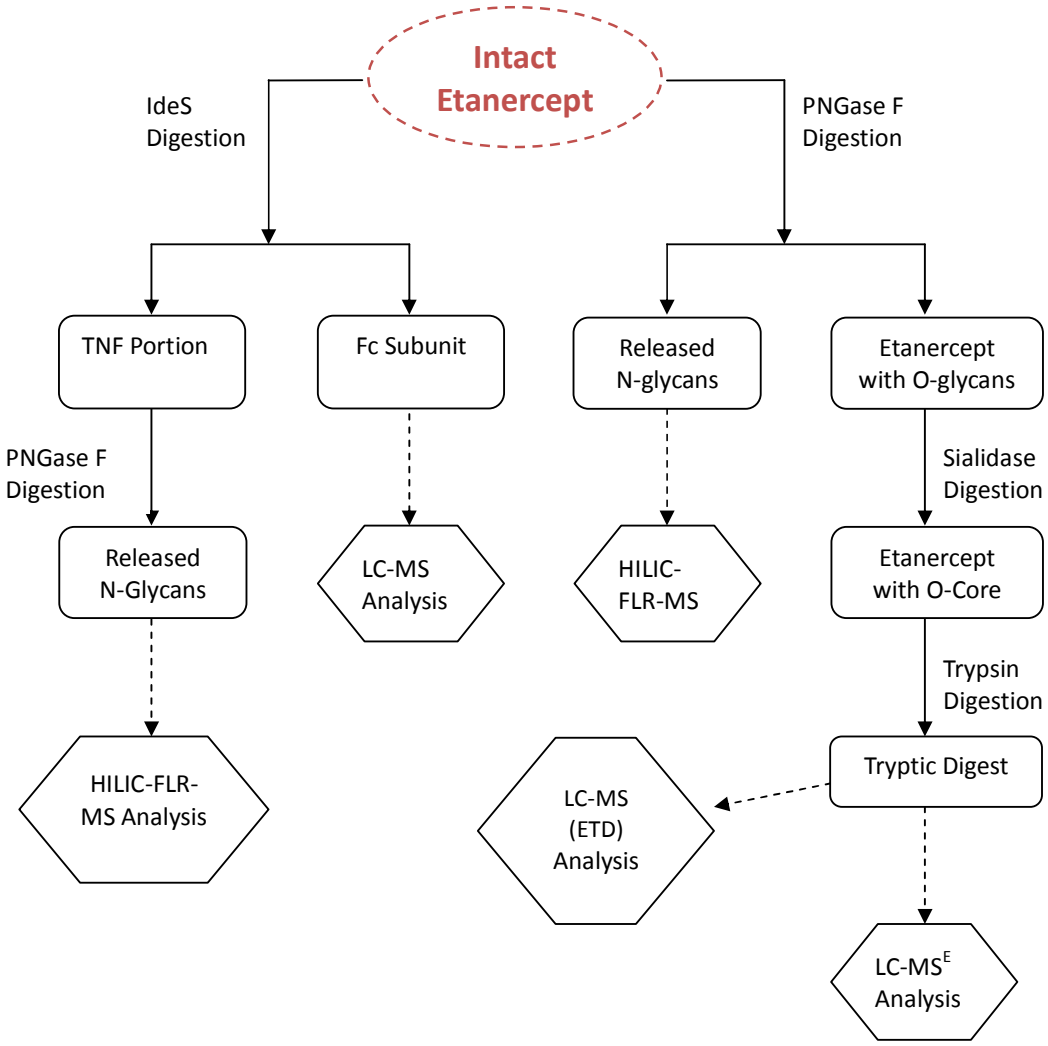
Tryptic peptides were fragmented upon isolation in the quadrupole when they eluted from the C18 BEH column. ETD MS/MS data was acquired over one second scan (signal accumulation) period with an anion refill time of 100 ms between scans. Peptides ions of interest were selected for ETD using the quadrupole with a mass selection window of ± 3.0 - 5.0 *m/z* units. The ion intensity of the radical anions monitored by the glow discharge ionization is typically tuned for around 1×10^6 counts on the instrument. For optimum ETD fragmentation efficiency, the ion intensity of the radical anions is usually tuned to be ~10-100 times higher than that of the precursor ions.

Data Analysis

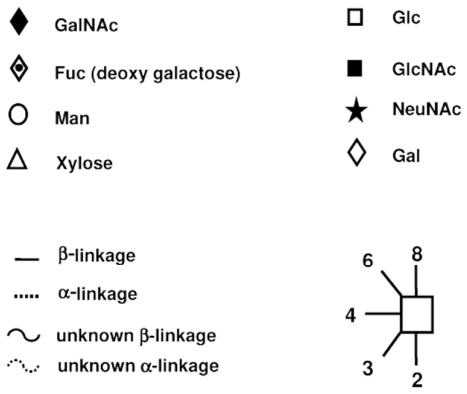
The LC/MS^E data was processed using BiopharmaLynx 1.3.3 (Waters Corporation) for sequence confirmation and glycopeptide identification. Cysteine carbamidomethylation (+57.02 Da) was set as a fixed modification, whereas N-deamidation (+0.98 Da) as variable modification. In order to identify O-glycopeptides in the Etanercept digest, a modifier that represents Core 1 O-glycan subtype (Galactose[β1–3]GalNAc, C14H23O10N, +365.13 Da) was included as variable modifications in the BiopharmaLynx data search. Up to eight Core 1 structures on serine and/or threonine residues on a peptide were permitted during the search. The mass tolerance was set at 10 ppm and 20 ppm for precursor and fragment ions respectively. The identified peptides were confirmed by MS^E spectra with at least five b/y fragment ions (on average) from triplicate analysis.

Data interpretation of ETD mass spectra of glycopeptides was done using tools inside MassLynx 4.1(Waters Corporation, Milford, MA). MaxEnt 3 software was used to deconvolute the raw data into singly charged monoisotopic spectra for easy interpretation of the protonated fragment ions. BioLynx was used to verify the manually assigned sites of glycosylation by *in silico* fragmentation of the proposed glycopeptide.

Scheme S-I. Overview of the experimental strategy developed for the glycosylation structure analysis of N- and O-glycosylation of Etanercept. The flowchart illustrates the workflow and techniques employed in the study.



Scheme S-II. Glycan symbol, linkage and linkage notation used in the manuscript.



All *N*-glycans have two core GlcNAc residues; F at the start of the abbreviation indicates a core α(1-6) fucose linked to the inner GlcNAc; Mx, number (x) of mannose on core GlcNAcs; Ax, number of antenna (GlcNAc) on trimannosyl core; A2, biantennary with both GlcNAcs as β(1-2) linked; A3, triantennary with a GlcNAc linked β1-2 to both mannose and a third GlcNAc linked β(1-4) to the α(1-3) linked mannose; A4, GlcNAcs linked as A3 with additional GlcNAc β(1-6) linked to α(1-6) mannose; B, bisecting GlcNAc linked β(1-4) to β(1-3) mannose; Gx, number (x) of β1-4 linked galactose on the antenna;; Sx, number (x) of sialic acids linked to galactose; the number 3 or 6 in parentheses after S indicates whether the sialic acid is in an α(2-3) or α(2-6) linkage.

Figure S-1. Comparison of WAX *N*-linked fractions overlaid from Etanercept **(A).** Total released profile **(B).** UPLC analysis of each of the WAX fractions, neutral, mono- and di-sialylated glycans. Demonstrates that the neutral, mono- and di-sialyated fractions can be overlaid to show the same profile as the total glycan pool.

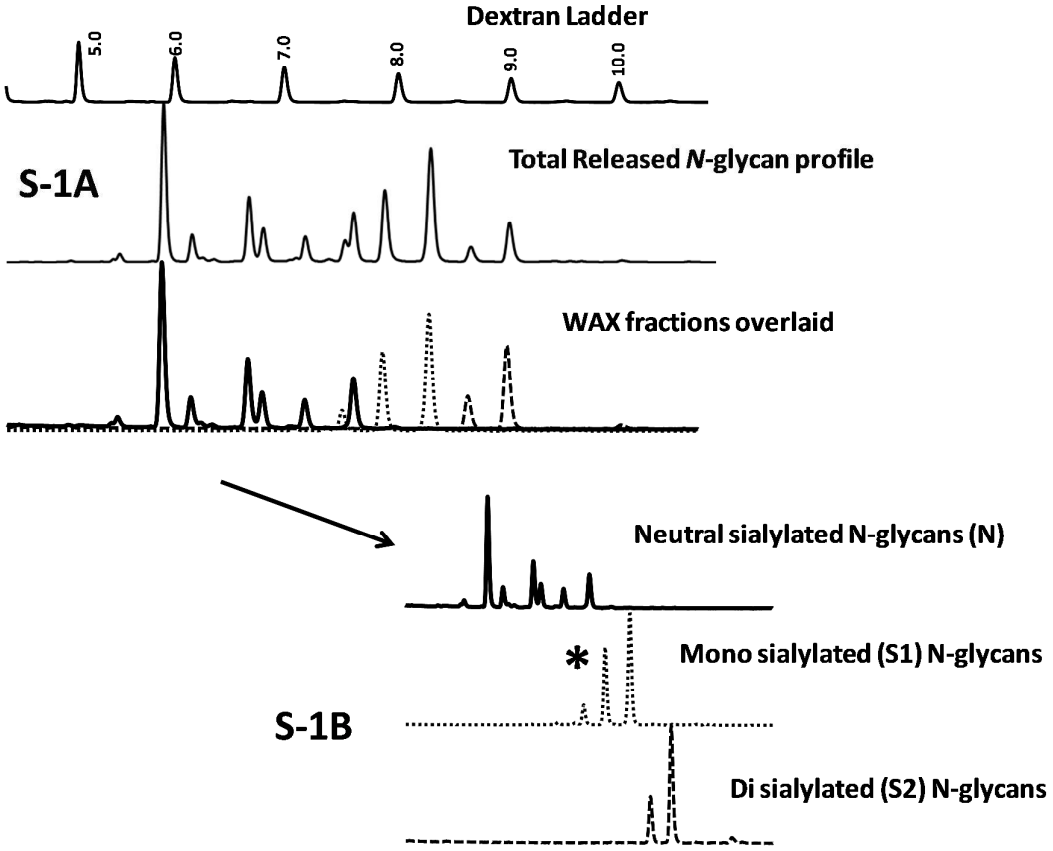


Figure S-2. Deconvoluted mass spectrum of the Fc subunit obtained by the FabRICATOR enzyme digestion of Etanercept. The glycoforms are assigned based on the summation of the masses of amino acid backbone of Fc and individual glycan. The structures displayed were adopted from the release N-glycan data summarized in supplemental Table 1.

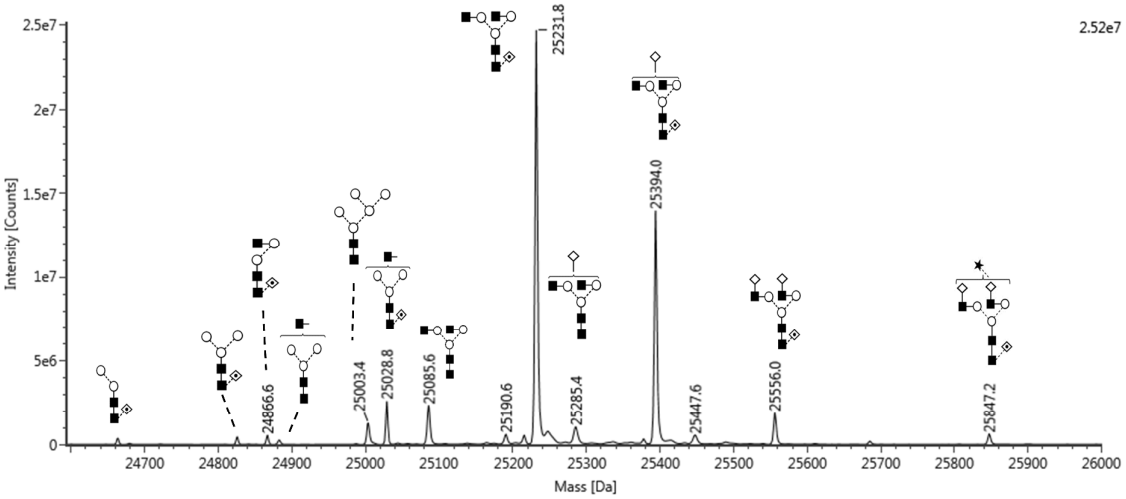
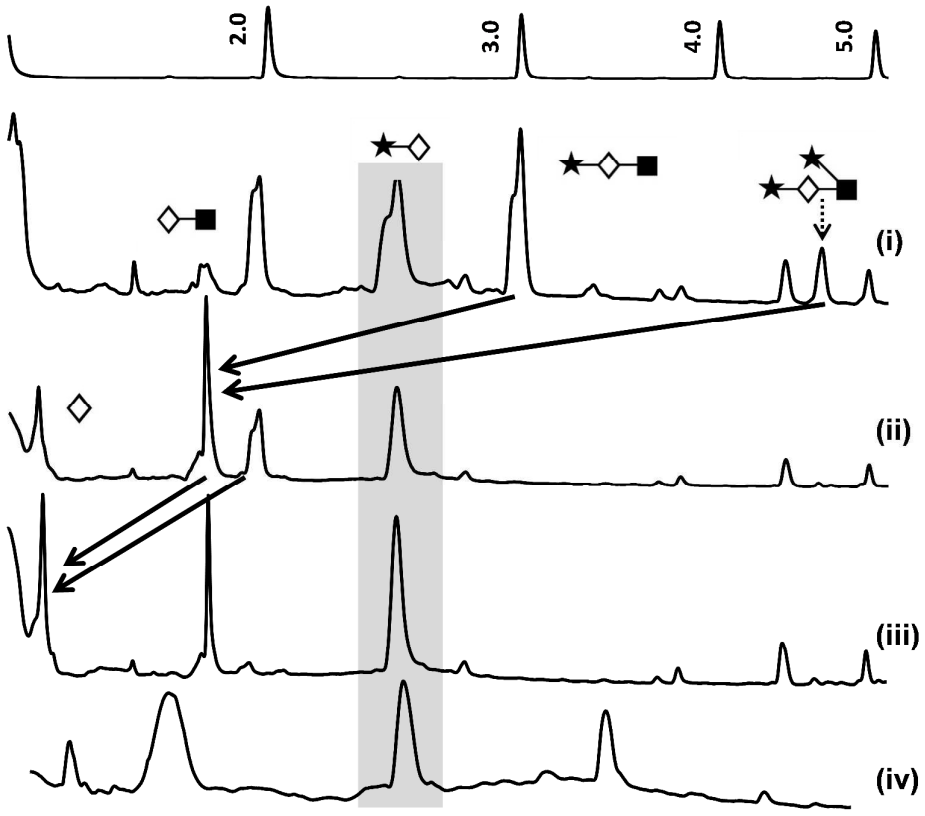


Figure S-3. UPLC-FLR analysis of total O-glycan pool released from Etanercept. (i) Undigested total released O-linked glycans; (ii) ABS exoglycosidase digestion to determine sialylated O-glycans. Peaks 3 and 4 were identified as mono- and di-sialylated core 1 O-glycans, respectively; (iii) ABS and BTG digestion was used to determine the neutral glycan structures; (iv) blank injection(control). The grey highlighted peak is from a reagent containment. The percentage area and GU value associated with each O-glycan identified are listed below the chromatograms.



Digest		UND			ABS			ABS+BTG	
Glycan	No	% Area	GU		% Area	GU		% Area	GU
Gal	0			0	22.82	1.13	0	44.68	1.15
Neutral	1	7.31	1.77	1	42.77	1.77	1	49.78	1.78
Peeling	2	33.56	1.97	2	34.41	1.97	2	5.54	1.93
Mono SA	3	48.99	3						
Di SA	4	10.14	4.63						

Number peaks referred to in figure legend

Figure S-4. Extracted ion chromatograms for T15 O-glycopeptides containing six or seven Core 1 tags. Duplicate runs were displayed here to show the reproducibility of the separation. The identity of T15 isoforms were labelled at the peak top.

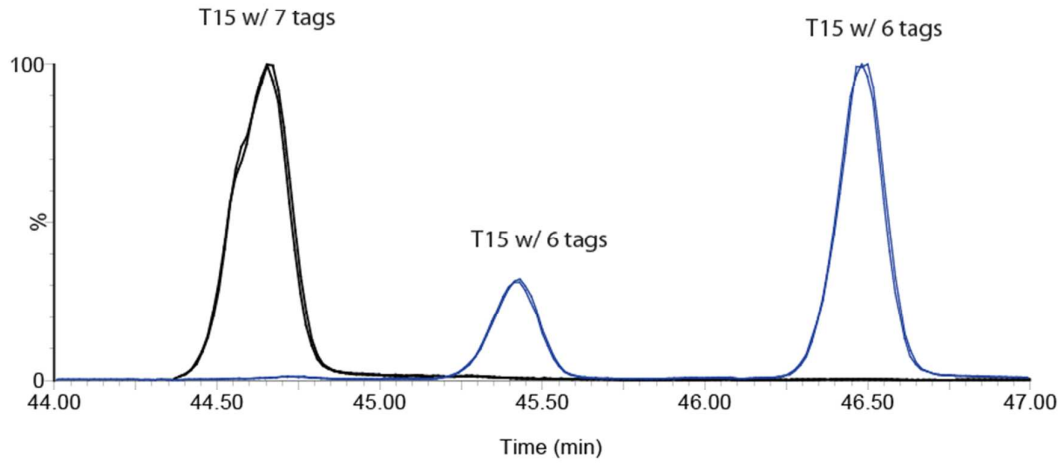
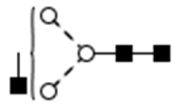
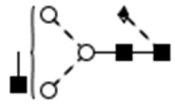
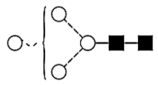
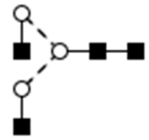
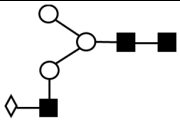
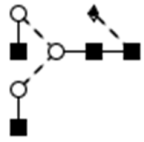
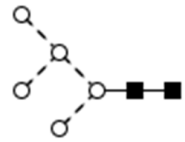
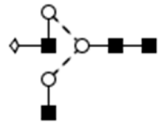
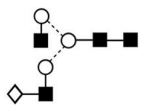
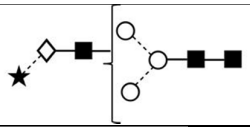
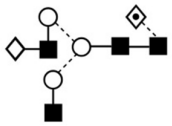
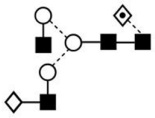
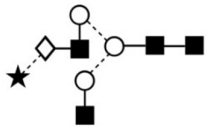
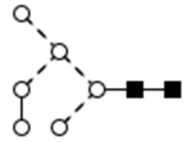
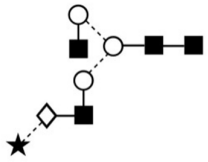
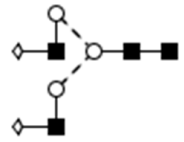
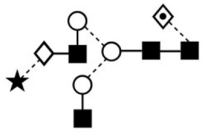
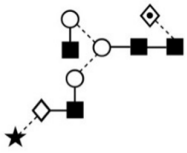
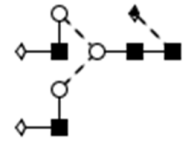
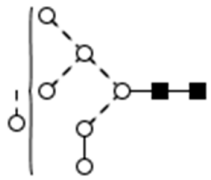
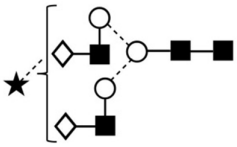
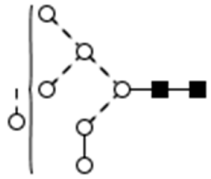
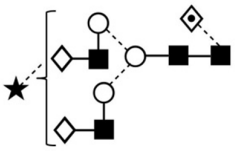
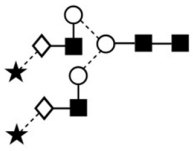
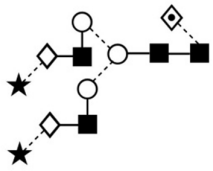
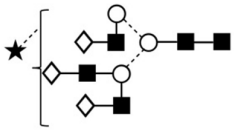
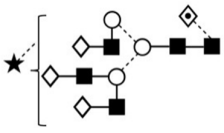
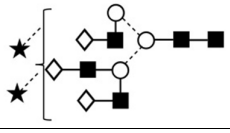
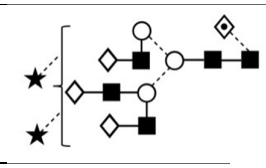
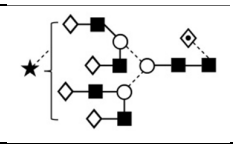
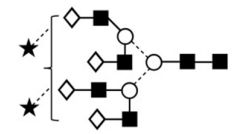
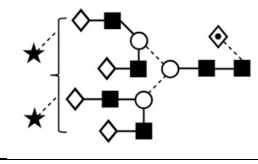


Table S-1. Analysis of released N-glycans from Etanercept. The N-glycan pool from Etanercept was subjected to arrays of exoglycosidase enzymes as described in "Experimental Methods" to derive the structures shown. The relative abundance was calculated from the fluorescence intensities of the peaks from the HILIC-UPLC analysis. The site heterogeneity was acquired by intact mass analysis of the Fc subunit and/or the analysis of released glycans from either the TNF-α receptor or Fc component. Please refer to Scheme S-II for glycan notation and UOXF drawings.

Peak	Glycan	UOXF Symbol	Mean GU	SD	Relative % area	Site of modification	Identified by MS1
1	A1		4.87	0.01	0.12	Fc/TNF-α	
2	FA1		5.23	0.01	0.03	Fc	Yes
3	M4		5.37	0.01	0.26	Fc/TNF-α	
4	A2		5.44	0.01	1	Fc/TNF-α	Yes
5	A1G(4)1		5.79	0.01	<1%	Fc/TNF-α	
6	F(6)A2		5.91	0.01	21.82	Fc/TNF-α	Yes
7	M5		6.20	0.01	3.29	Fc/TNF-α	Yes
8	A2[6]G(4)1		6.30	0.01	0.31	TNF-α	Yes
9	A2[3]G(4)1		6.41	0.01	0.23	TNF-α	Yes
10			6.61	0.01	<1%	TNF-α	

	A1G(4)1S(3)1						
11	F(6)A2[3]G(4)1		6.73	0.01	9.38	Fc/TNF-α	Yes
12	F(6)A2[3]G(4)1		6.86	0.01	4.77	Fc/TNF-α	Yes
13	A2[6]G(4)1S(3)1		6.98	0.01	<1%	TNF-α	
14	M6 D3		7.15	0.01	0.53	TNF-α	Yes
15	A2[3]G(4)1S(3)1		7.15	0.01	<1%	TNF-α	Yes
16	A2G(4)2		7.23	0.01	3.87	TNF-α	Yes
17	F(6)A2[6]G(4)1S(3)1		7.43	0.01	0.32	TNF-α	Yes
18	F(6)A2[3]G(4)1S(3)1		7.57	0.01	2.87	TNF-α	Yes
19	F(6)A2G(4)2		7.65	0.01	7.78	Fc/TNF-α	Yes

20	M7 D3		7.65	0.01	<1%	Not confirmed	
21	A2G(4)2S(3)1		7.91	0.01	12.51	TNF-α	Yes
22	M7 D3		7.91	0.01	<1%	Not confirmed	
23	F(6)A2G(4)2S(3)1		8.32	0.01	20.36	Fc/TNF-α	Yes
24	A2G(4)2S(3,3)2		8.67	0.01	2.6	TNF-α	Yes
25	F(6)A2G(4)2S(3,3)2		9.03	0.01	6.98	TNF-α	Yes
26	A3G(4)3S(3)1		9.03	0.01	<1%	TNF-α	Yes
27	F(6)A3G(4)3S(3)1		9.47	0.01	0.1	TNF-α	Yes
28	A3G(4)3S(3,3)2		9.73	0.01	<1%	TNF-α	

29	F(6)A3G(4)3S(3,3)2		10.14	0.01	0.35	TNF-α	
30	F(6)A4G(4)4S(3)1		10.84	0.01	0.14	TNF-α	
31	A4G(4)4S(3,3)2		11.31	0.02	0.01	TNF-α	
32	F(6)A4G(4)4S(3,3)2		11.92	0.02	0	TNF-α	