Supporting Information for Publication

Proteoform profile mapping of the human serum Complement component C9 reveals unexpected new features of *N*-, *O*- and *C*-glycosylation

Vojtech Franc[1,2], Yang Yang[1,2], and Albert J.R. Heck[1,2]*

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

Correspondence: Albert Heck, a.j.r.heck@uu.nl

FIGURE LEGENDS FOR SUPPORTING INFORMATION

Supplementary Figure S1

Low energy HCD MS/MS spectra of the glycopeptides harboring the known canonical *N*-glycosylation sites, derived by proteolytic digestion of C9 by trypsin and AspN. LC MS/MS spectra were acquired for ions with precursor *m/z* of 1099.95 (a) and 1044.71 (b), respectively. Sequential fragmentation of the *N*-glycan part allowed deduction of its glycan composition. "P" = peptide backbone of the glycopeptide.

Supplementary Figure S2

EThcD MS/MS spectra of C9 *N*-terminal tryptic peptides harboring *O*-glycosylation C9. In (a) the non-modified peptide spectrum is shown. In total, five selected LC MS/MS spectra are shown, which were acquired for tryptic *O*-glycopeptides with precursor *m/z* of 863.03 (a), 990.41 (b), 1,081.77 (c), 1,178.80 (d) and 1,300.51 (e), respectively. Fragmentation patterns conclusively confirmed the amino acid sequence of the peptides and composition of the *O*-glycans. However, the precise modification site could not be determined due to a lack of sequence indicative *c* and *z* fragment ions. In the spectra b, c and d, T11 was assigned as most likely modification site based on the presence of long series of non-modified *c* and *z* ions.

Supplementary Figure S3

EThcD MS/MS spectra *C*-mannosylated tryptic peptides originating from TSP domain of C9. LC MS/MS spectra were acquired for ions with precursor *m/z* of 700.95 (a) and 754.96 (b), respectively. In (a) the peptide fragmentation spectrum with one *C*-mannose at W27 is shown. In (b) the fragmentation spectrum reveals occupation of both W (W27 and W30) by *C*-mannoses in the sequence motif WXXW. Fragmentation patterns conclusively confirmed the amino acid sequence of the peptides.

Supplementary Figure S4

Multiple amino acid sequence alignment of the C9 protein from human, mouse, rat, cow, rabbit and horse. The accession numbers provided refer to the protein database UniProtKB. The alignment was constructed using AliView 1.18[1] whereby the *N*-terminal signal peptides were omitted. The *N*-terminus of C9 with the likely *O*-glycosylation site T11 is highlighted in orange

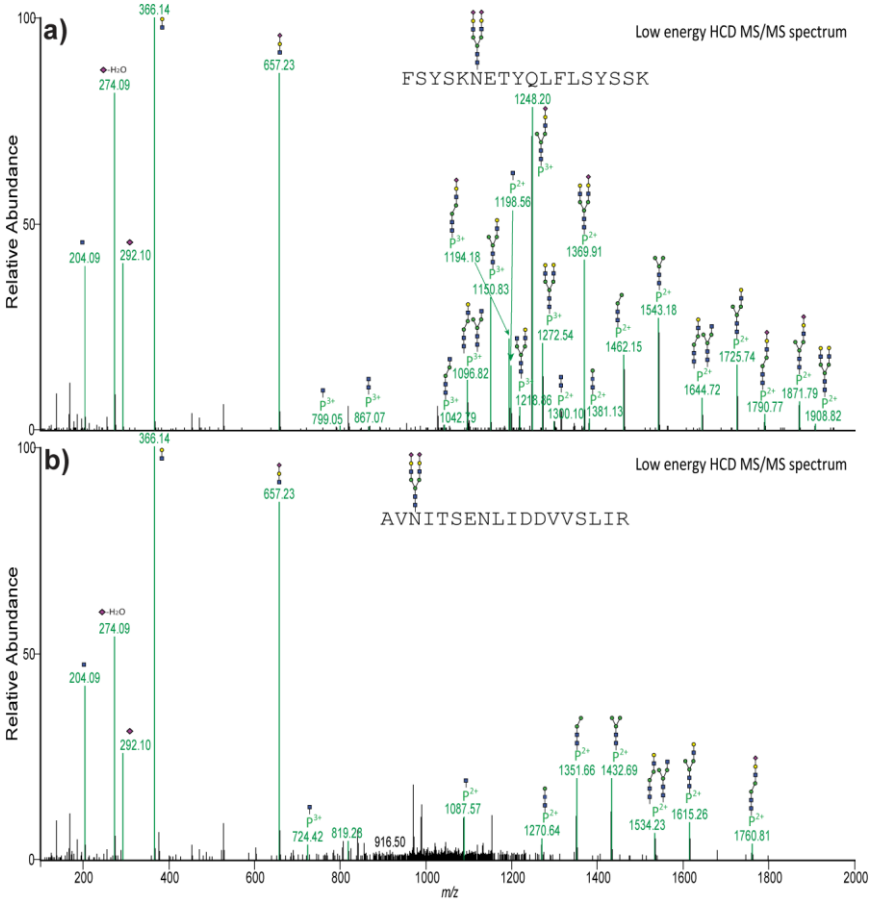
and *C*-mannosylation sites are in green. All *N*-glycosylation sites are highlighted as sequons (N-X-X). Newly discovered *N*-glycosylation site N215 is in purple and the two previously reported sites N256 and N394 are in red. Next, the predicted *N*-glycosylation sites from the selected mammals are shown in magenta.

The alignment indicated a very little conservation of the *N*-terminus (where *O*-glycosylation was detected on human C9) while *C*-mannosylated sites are in highly conserved TSP domain. The amino acid sequences of selected mammalian species show a relatively low level of conservation of the *N*-glycosylation sites. The most conserved canonical *N*-glycosylation sites are N256 and N394. The N256 is occupied on human and likely also on rabbit and horse. The presence of *N*-glycan at N394 was experimentally confirmed in human and predicted to be modified also on rat, rabbit and horse. Interestingly, the here reported lower occupied non-canonical NAX-site turns out to be conserved. Remarkably, based on the sequence analysis almost all selected species contain a few more putative canonical *N*-glycosylation sites, which are not conserved at all, *e.g.*, the murine C9 protein contains a *N*-glycan motif at N48 located in the TSP domain. This would most likely prevent *C*-mannosylation of this TSP and further influence the repertoire of mouse C9 proteoforms. Bovine C9 seems to be an exception among other species since it contains only one potential canonical *N*-glycosylation site at N430. This site was also predicted to be glycosylated in horse C9, but not in other species. Nevertheless, all latter ones are *N*-glycosylated at the more conserved N394 (NIT/S), suggesting that the presence of a *N*-glycan chain in this C9 region (394-430) may be required for functional purposes. Similarly, murine and rat C9 are missing a *N*-glycosylation sequon at the more conserved N256, but they contain *N*-glycosylation motifs in the non-conserved region between the amino acids 240-248. Rabbit and horse were predicted to be *N*-glycosylated in this region as well, however they contain also exactly the same sequence motif as human (NET) at N256. Therefore, it is likely that the predicted sites N242 (rabbit) and N246 (horse) are not modified unless these species contain two *N*-glycans in this region. Although, these speculations need to be confirmed by experimental data, our alignments hint at that C9 may display species-specific glycosylation patterns. Variation of glycosylation among different animal species has been reported for instance for IgG[2], nevertheless there is a lack of understanding about this phenomenon and it opens questions about the function of site-specific glycosylation, not only on C9, but also on many other plasma proteins.

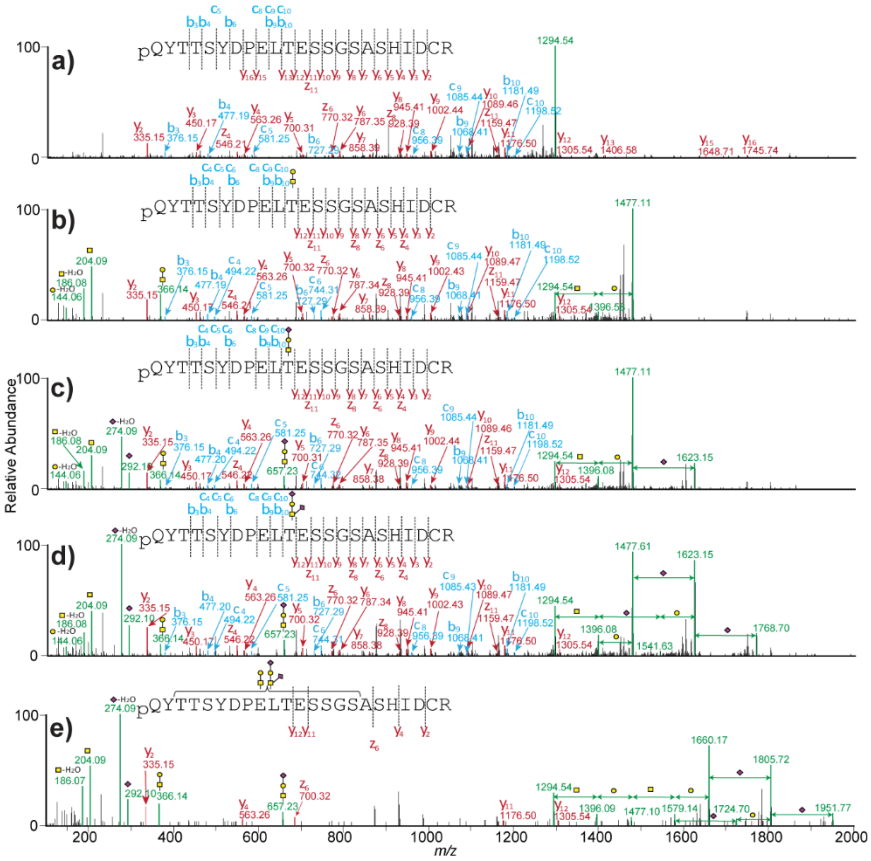
The aligned sequences were processed using ENDscript 3.0[3]. Similarity coloring scheme is a percentage of equivalent residues calculated considering physico-chemical properties.

References

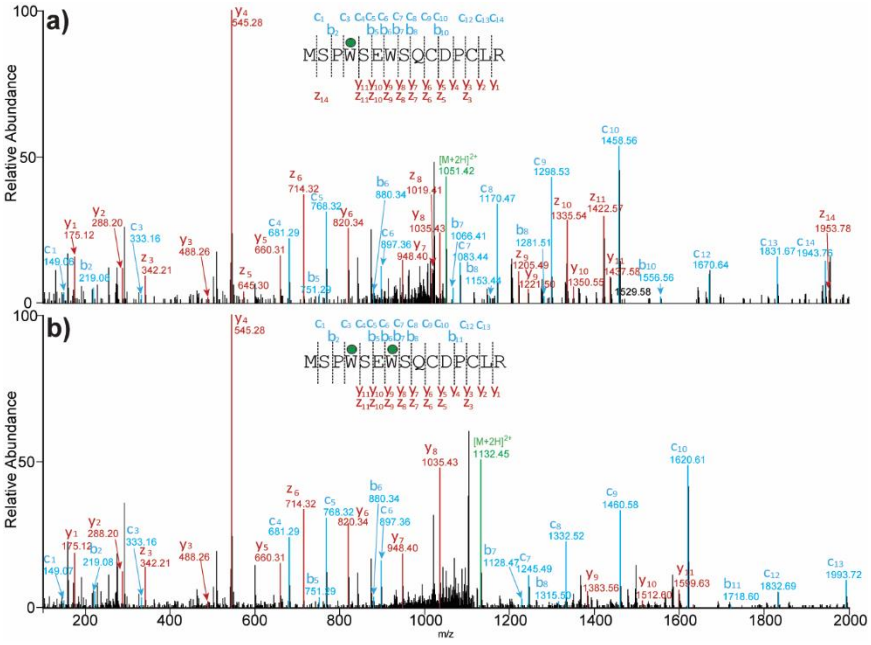
1. Larsson, A. *Bioinformatics* **2014**. *30*, 3276-3278.
2. Raju, T. S.; Briggs, J. B.; Borge, S. M.; Jones, A. J. *Glycobiology* **2000**, *10*, 477-486.
3. Robert, X.; Gouet, P. *Nucleic Acids Res.* **2014**, *42*, W320-324.



Supplementary Figure S1

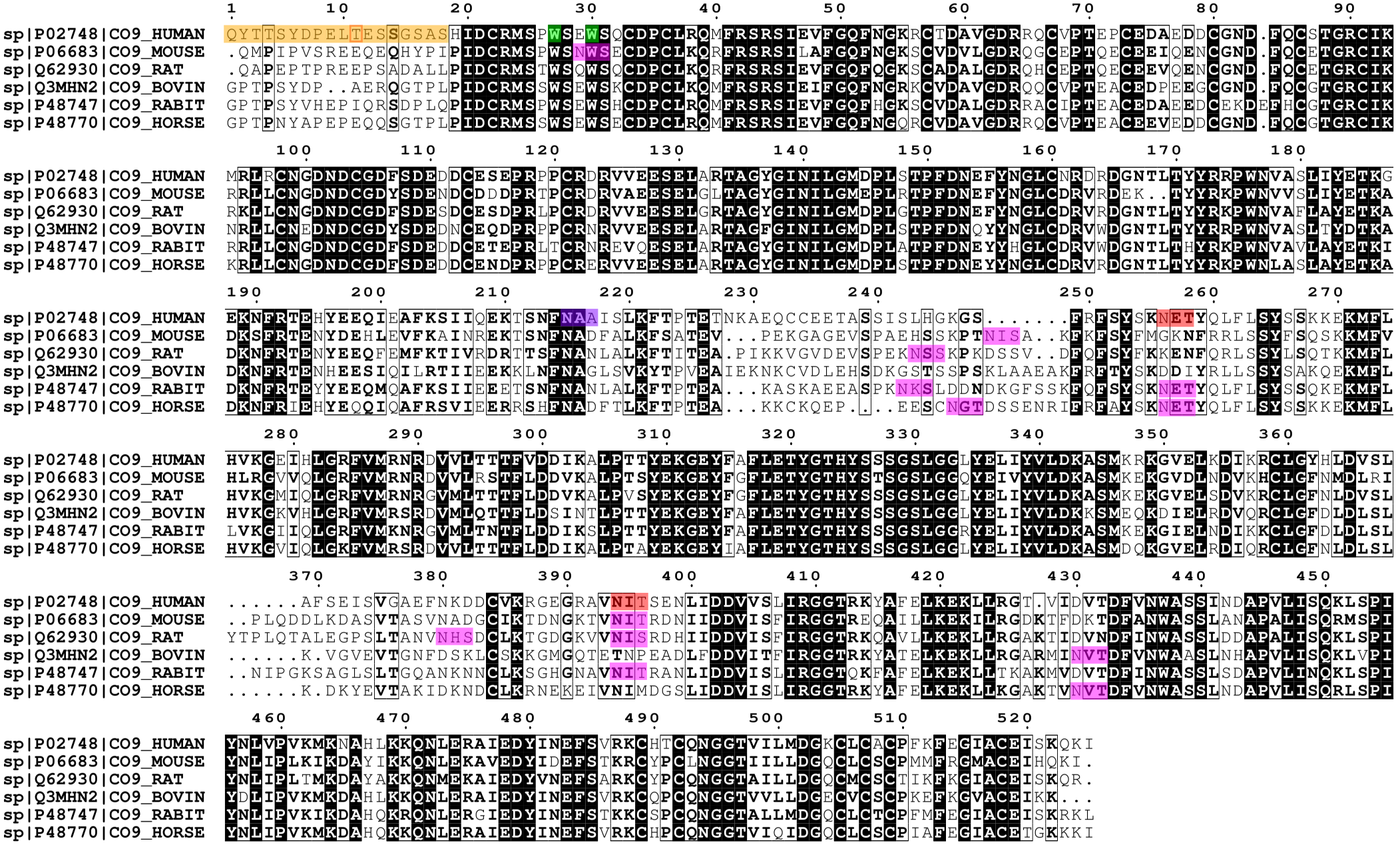


Supplementary Figure S2



Supplementary Figure S3

S4 – supplementary figure – Multiple amino acid sequence alignment of the C9 protein from human, mouse, rat, cow, rabbit and horse


sp|P02748|CO9_HUMAN
sp|P06683|CO9_MOUSE
sp|Q62930|CO9_RAT
sp|Q3MHN2|CO9_BOVIN
sp|P48747|CO9_RABIT
sp|P48770|CO9_HORSE

S5 – supplementary document - the certificate of analysis of the C9 sample

CERTIFICATE OF ANALYSIS

Complement Technology, Inc.
4801 Troup Hwy, Suite 701
Tyler, Texas 75703, USA

Product: **C9 Protein**
Catalog # **A126 Lot # 9c**
Exp. Date 7/29/2018

Description: C9 Purified Human Complement Protein

Specifications	Limits	Results
PROTEIN CONCENTRATION	0.50 – 1.2 mg/mL C9 C9 has an extinction coefficient of $E^{1\%}/_{280nm}$ = 9.88	1.03 mg/ml
FILL VOLUME	0.250 – 0.275 mL	0.270 mL
BUFFER	Phosphate buffered saline, pH 7.2	Phosphate buffered saline, pH 7.2
PRESERVATIVE	None, filtered through a 0.22 µm pore size filter.	None, filtered through a 0.22 µm pore size filter.
PURITY	> 85% by SDS PAGE	>95% by SDS PAGE
FUNCTIONAL ACTIVITY	Titer >150,000 C9H50/mg	255,000 Units/mg
	>70% of C9 activity in NHS on a mg/mg basis.	>100%
IMMUNOCHEMISTRY	≤ trace amounts of IgG, IgA, IgM, albumin, C3, C4, C5, C6, C7 or C8	None Detected
HUMAN SERUM/PLASMA STARTING MATERIAL		
ANTI-HIV-I ANTIBODIES	Negative	Negative
ANTI-HIV-II ANTIBODIES	Negative	Negative
ANTI-HCV ANTIBODIES	Negative	Negative
HBsAg	Negative	Negative
HIV-1 p24 Ag	Negative	Negative

Store at -70°C or below.
Avoid Repeated Freeze/Thaw

FOR RESEARCHUSE ONLY
NOT FOR HUMAN OR DRUG USE

Connie Wood
Signature of Analyst

7/30/14
Date of Analysis

Phone: 1-903-581-8284 FAX: 1-903-581-0491 Email: contactCTI@aol.com
Web Site: www.ComplementTech.com

148

149 **Supplementary Table S1**

150 List of identified and validated C9 peptides from tryptic digest

Peptide Modified Sequence	Precursor (m/z)	Product Charge	Mass Error (ppm)	Total Area	Retention Time (min)
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R	863.0345	3	-2.8	459621984	33.64
QYTTSYDPELTESSGSASHIDC[+57]R; [+947.3]	1776.2229	2	-5.2	21697820	32.52
QYTTSYDPELTESSGSASHIDC[+57]R; [+947.3]	1184.4843	3	-4.2	6088687616	32.52
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+947.3]	1178.8088	3	-4.7	30135201792	32.44
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+947.3]	884.3584	4	-5.0	1015169024	32.44
QYTTSYDPELTESSGSASHIDC[+57]R; [+656.2]	1087.4525	3	-4.6	5874841600	32.63
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+656.2]	1081.7770	3	-4.2	31467601920	32.63
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+656.2]	811.5846	4	-3.5	273742464	32.63
QYTTSYDPELTESSGSASHIDC[+57]R; [+365.1]	990.4207	3	-3.3	786598784	32.23
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+365.1]	984.7452	3	-4.1	2876220160	32.23
QYTTSYDPELTESSGSASHIDC[+57]R; [+947.3], [+365.1]	1306.1951	3	-4.6	630539584	30.92
QYTTSYDPELTESSGSASHIDC[+57]R; [+947.3], [+365.1]	979.8981	4	-1.9	68517296	30.92
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+947.3], [+365.1]	1300.5196	3	-2.7	1488793728	30.94
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+947.3], [+365.1]	975.6415	4	-3.0	218757088	30.94
MSPW[+162.1]SEWSQC[+57]DPC[+57]LR	1050.9266	2	-4.2	15285429248	36.77
MSPW[+162.1]SEWSQC[+57]DPC[+57]LR	700.9535	3	-3.7	4703996928	36.77
M[+16]SPW[+162.1]SEWSQC[+57]DPC[+57]LR	1058.9241	2	-3.3	1.4677E+11	35.26
M[+16]SPW[+162.1]SEWSQC[+57]DPC[+57]LR	706.2851	3	-2.7	25765066752	35.26
MSPW[+162.1]SEW[+162.1]SQC[+57]DPC[+57]LR	1131.9530	2	-4.8	3710511360	33.03
MSPW[+162.1]SEW[+162.1]SQC[+57]DPC[+57]LR	754.9711	3	-4.5	1836591872	33.03
M[+16]SPW[+162.1]SEW[+162.1]SQC[+57]DPC[+57]LR	1139.9505	2	-4.5	34263304192	31.73
M[+16]SPW[+162.1]SEW[+162.1]SQC[+57]DPC[+57]LR	760.3028	3	-3.4	16343121920	31.73
TSNFNAAISLK	583.3142	2	-3.2	1.68502E+11	34.31
TSN[+2204.8]FNAAISLK	1685.7004	2	-4.0	94825072	33.32
TSN[+2204.8]FNAAISLK	1124.1361	3	-2.7	1007365504	33.32
TSN[+2204.8]FNAAISLK	843.3539	4	-3.6	17227846	33.32
AVN[+2204.8]ITSENLIDDVVSLIR	1392.6209	3	-4.8	1146552704	11.58
AVN[+2204.8]ITSENLIDDVVSLIR	1044.7175	4	-2.8	474522624	11.58
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKKEK	1743.7657	3	-1.6	900461184	36.13
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKKEK	1308.0761	4	-4.1	11739473920	36.13
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKKEK	1046.6623	5	-2.3	41915990016	36.13
GSFRFSYSKN[+1913.7]ETYQLFLSYSSKKEK	1235.3023	4	-4.9	307079104	35.53
GSFRFSYSKN[+1913.7]ETYQLFLSYSSKKEK	988.4433	5	-4.0	714075200	35.53
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKK	1658.0532	3	-5.4	3501403392	37.32
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKK	1243.7917	4	-4.6	13250072576	37.32
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKK	995.2348	5	-3.4	17270624256	37.32
GSFRFSYSKN[+1913.7]ETYQLFLSYSSKK	1561.0214	3	-3.6	32079828	36.53
GSFRFSYSKN[+1913.7]ETYQLFLSYSSKK	1171.0179	4	-3.4	386805568	36.53
GSFRFSYSKN[+1913.7]ETYQLFLSYSSKK	937.0157	5	-3.2	702033536	36.53
FSYSKN[+2204.8]ETYQLFLSYSSK	1466.2805	3	-4.5	1916245760	38.22

FSYSKN[+2204.8]ETYQLFLSYSSK	1099.9622	4	-4.2	1386641664	38.22
FSYSKN[+1913.7]ETYQLFLSYSSK	1369.2487	3	-3.1	25390424	37.28

151

152 List of identified and validated C9 peptides from trypsin + AspN digest

Peptide Modified Sequence	Precursor (m/z)	Product Charge	Mass Error (ppm)	Total Area	Retention Time (min)
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R	863.0345	3	-2.4	678611456	33.59
QYTTSYDPELTESSGSASHIDC[+57]R; [+947.3]	1776.2229	2	-10.7	141675808	32.15
QYTTSYDPELTESSGSASHIDC[+57]R; [+947.3]	1184.4843	3	-9.4	903599424	32.15
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+947.3]	1178.8088	3	-4.5	36548460544	32.49
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+947.3]	884.3584	4	-4.4	1401541120	32.49
QYTTSYDPELTESSGSASHIDC[+57]R; [+656.2]	1087.4525	3	-5.1	10548512768	32.66
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+656.2]	1081.7770	3	-4.2	78340882432	32.61
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+656.2]	811.5846	4	-3.4	788462400	32.61
QYTTSYDPELTESSGSASHIDC[+57]R; [+365.1]	990.4207	3	-3.2	2936786432	32.13
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+365.1]	984.7452	3	-4.1	15281887232	32.13
QYTTSYDPELTESSGSASHIDC[+57]R; [+947.3], [+365.1]	1306.1951	3	-6.3	1116945664	30.91
QYTTSYDPELTESSGSASHIDC[+57]R; [+947.3], [+365.1]	979.8981	4	-4.7	137224160	30.91
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+947.3], [+365.1]	1300.5196	3	-3.2	2119178752	30.76
Q[-17]YTTSYDPELTESSGSASHIDC[+57]R; [+947.3], [+365.1]	975.6415	4	-5.2	289106560	30.76
MSPW[+162.1]SEWSQC[+57]DPC[+57]LR	1050.9266	2	-4.4	1.67946E+11	36.67
MSPW[+162.1]SEWSQC[+57]DPC[+57]LR	700.9535	3	-4.1	19312404480	36.67
M[+16]SPW[+162.1]SEWSQC[+57]DPC[+57]LR	1058.9241	2	-3.7	2.8184E+11	35.13
M[+16]SPW[+162.1]SEWSQC[+57]DPC[+57]LR	706.2851	3	-3.2	44113780736	35.13
MSPW[+162.1]SEW[+162.1]SQC[+57]DPC[+57]LR	1131.9530	2	-4.7	33620144128	32.98
MSPW[+162.1]SEW[+162.1]SQC[+57]DPC[+57]LR	754.9711	3	-4.3	9210617856	32.98
M[+16]SPW[+162.1]SEW[+162.1]SQC[+57]DPC[+57]LR	1139.9505	2	-3.7	94306058240	31.59
M[+16]SPW[+162.1]SEW[+162.1]SQC[+57]DPC[+57]LR	760.3028	3	-3.8	19439058944	31.59
TSNFNAAISLK	583.3142	2	-3.2	3.9697E+11	34.14
TSN[+2204.8]FNAAISLK	1685.7004	2	-3.9	451645856	33.45
TSN[+2204.8]FNAAISLK	1124.1361	3	-3.0	3637175040	33.45
TSN[+2204.8]FNAAISLK	843.3539	4	-4.8	84116424	33.45
AVN[+2204.8]ITSENLIDDVVSLIR	1392.6209	3	-1.4	688354496	11.58
AVN[+2204.8]ITSENLIDDVVSLIR	1044.7175	4	-3.1	363000064	11.58
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKKEK	1743.7657	3	-2.0	2310239232	36.12
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKKEK	1308.0761	4	-3.9	27235827712	36.12
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKKEK	1046.6623	5	-2.5	73779724288	36.12
GSFRFSYSKN[+1913.7]ETYQLFLSYSSKKEK	1646.7339	3	-5.0	568055296	35.55
GSFRFSYSKN[+1913.7]ETYQLFLSYSSKKEK	1235.3023	4	-4.1	2871592960	35.55
GSFRFSYSKN[+1913.7]ETYQLFLSYSSKKEK	988.4433	5	-3.8	6345625600	35.55
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKK	1658.0532	3	-3.1	2621754368	37.28
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKK	1243.7917	4	-4.1	17482567680	37.28
GSFRFSYSKN[+2204.8]ETYQLFLSYSSKK	995.2348	5	-3.1	18819860480	37.28
GSFRFSYSKN[+1913.7]ETYQLFLSYSSKK	1561.0214	3	-4.8	90132496	37.32
GSFRFSYSKN[+1913.7]ETYQLFLSYSSKK	1171.0179	4	-3.6	1967233024	37.32

GSFRFSYSKN[+1913.7]ETYQLFLSYSSKK	937.0157	5	-3.6	748020992	37.32
FSYSKN[+2204.8]ETYQLFLSYSSK	1466.2805	3	-2.8	43814805504	38.1
FSYSKN[+2204.8]ETYQLFLSYSSK	1099.9622	4	-3.6	29031387136	38.1
FSYSKN[+1913.7]ETYQLFLSYSSK	1369.2487	3	-3.9	3948540416	37.32
FSYSKN[+1913.7]ETYQLFLSYSSK	1027.1884	4	-3.2	1117805312	37.32

Site-specific quantification of PTMs on C9 based on peptide data

Glycosylation site	Glycan composition	Relative abundance (%)
N-term	-	0.51
N-term	HexNAc1Hex1Sia2	35.78
N-term	HexNAc1Hex1Sia1	52.75
N-term	HexNAc1Hex1	8.27
N-term	HexNAc2Hex2Sia2	2.69
Cman1	Man	77.02
Cman2	Man2	22.98
N236(0)	-	99.15
N236	HexNAc4Hex5Sia2	0.85
N277(2)	HexNAc4Hex5Sia2	95.05
N277(1)	HexNAc4Hex5Sia2	4.95
N415	HexNAc4Hex5Sia2	100.00

Supplementary Table S2

List of validated proteoforms

Proteoform	relative abundance (%)	N-glycan (total composition)	O-glycan	C-Man	Calculated mass (m/z)	Observed mass (m/z)	Calculated deconvoluted mass (Da)	Observed deconvoluted mass (Da)	Standard deviation (±Da)
1	20.58	HexNAc8Hex10Sia4	HexNAc1Hex1	Man	4396.44	4396.43	65933.59	65933.45	0.07
2	20.58	HexNAc8Hex10Sia2	HexNAc1Hex1Sia1	Man	4396.44	4396.43	65933.59	65933.45	0.07
3	6.14	HexNAc8Hex10Sia4	HexNAc1Hex1	Man2	4407.25	4407.22	66095.73	66095.30	0.22
4	6.14	HexNAc8Hex10Sia2	HexNAc1Hex1Sia1	Man2	4407.25	4407.22	66095.73	66095.30	0.22
5	100.00	HexNAc8Hex10Sia4	HexNAc1Hex1Sia1	Man	4415.86	4415.89	66224.85	66225.35	0.25
6	100.00	HexNAc8Hex10Sia2	HexNAc1Hex1Sia2	Man	4415.86	4415.89	66224.85	66225.35	0.25
7	29.83	HexNAc8Hex10Sia4	HexNAc1Hex1Sia1	Man2	4426.67	4426.66	66386.99	66386.90	0.05
8	29.83	HexNAc8Hex10Sia2	HexNAc1Hex1Sia2	Man2	4426.67	4426.66	66386.99	66386.90	0.05
9	68.64	HexNAc8Hex10Sia4	HexNAc1Hex1Sia2	Man	4435.27	4435.28	66516.11	66516.20	0.05
10	20.48	HexNAc8Hex10Sia4	HexNAc1Hex1Sia2	Man2	4446.08	4446.08	66678.25	66678.20	0.02
11	3.31	HexNAc8Hex10Sia4	HexNAc2Hex2Sia2	Man	4459.63	4459.61	66881.44	66881.15	0.15
12	0.99	HexNAc8Hex10Sia4	HexNAc2Hex2Sia2	Man2	4470.44	4470.62	67043.59	67046.30	1.36
13	1.26	HexNAc12Hex15Sia6	HexNAc1Hex1Sia1	Man	4562.90	4562.93	68430.49	68430.95	0.23
14	1.26	HexNAc12Hex15Sia4	HexNAc1Hex1Sia2	Man	4562.90	4562.93	68430.49	68430.95	0.23
15	0.86	HexNAc12Hex15Sia6	HexNAc1Hex1Sia2	Man	4582.32	4582.33	68721.75	68721.95	0.10