Supporting Information for

"Comprehensive Characterization of Glycosylation and Hydroxylation of Basement Membrane Collagen IV by High-Resolution Mass Spectrometry"

Trayambak Basak^{a, b}, Lorenzo Vega-Montoto^{a,b,#}, Lisa J. Zimmerman^c, David L. Tabb^{c, d}, Billy G. Hudson^{a, b, c} and Roberto M. Vanacore^{a, b,*}

^a Department of Medicine, Division of Nephrology and Hypertension, ^b Center for Matrix Biology, ^cDepartment of Biochemistry, ^dDepartment of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37232

Running Title: Mass spectrometry analysis of collagen IV

To whom correspondence should be addressed: Department of Medicine, Division of Nephrology and Hypertension, Vanderbilt University Medical Center, 1161 21st Avenue South, B-3113 Medical Center North, Nashville, TN 37232, USA, Tel.: (615) 322-8323; Fax: (615) 343-7156; E-mail: roberto.vanacore@vanderbilt.edu

[#]Current address: LVM is at Idaho National Laboratory, Idaho Falls, ID. *Corresponding author

Table of Content												
Supporting information	Page No.											
Supporting table S1	1											
Supporting table S2	2											
Supporting figure S1	3											
Supporting figure S2	4-40											
Supporting figure S3	41-88											
Supporting figure S4	89											
Supporting figure S5	90											
Cited References	91											

Supporting table S1: Detail of samples analyzed with multiple nzymes and different fragmentation method using LTQ-orbitrap Velos and Q-Exactive

Samples	LTQ- orbitrap Velos	Q-Exactive							
1. Purified 7S from PHFR9 cells	Trypsin (CID, ETD)	XX							
2. EHS mouse collagen IV (BD Biosciences)	a. Trypsin b. GluC (CID, ETD)	a. Trypsin b. GluC c. LysC+Trypsin (HCD)							

Supplementary table S2 Comparative analysis of O-glycosylation sites detected in mouse and human col4a1

O-glycosylation sites	Mouse Collagen IV	Human Collagen IV
K ⁴⁵	Glucosyl-galactosyl - hydroxylysine	Not detected
K ⁴⁸	Glucosyl-galactosyl - hydroxylysine	Not detected
K ⁷⁸	Galactosyl / glucosyl-galactosyl-hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
K ⁹⁰	Galactosyl / glucosyl-galactosyl-hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
K ¹⁵⁵	No glycosylation detected	Galactosyl - hydroxylysine
K ²¹⁷	Glucosyl-galactosyl - hydroxylysine	Not detected
K ²²⁸	Glucosyl-galactosyl - hydroxylysine	Not detected
K ²³¹	Glucosyl-galactosyl - hydroxylysine	Not detected
K ²⁹⁵	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ²⁹⁸	Not conserved	Glucosyl-galactosyl - hydroxylysine
K ³²²	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ³⁴³	No glycosylation detected	Glucosyl-galactosyl - hydroxylysine
K ³⁶¹	Glucosyl-galactosyl - hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
K ³⁹³	Glucosyl-galactosyl - hydroxylysine	Not detected
K ⁴⁶⁰	Glucosyl-galactosyl - hydroxylysine	Galactosyl - hydroxylysine
K ⁴⁶³	Galactosyl / glucosyl-galactosyl-hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁴⁹⁷	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁵²⁷	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁵⁷³	Galactosyl / glucosyl-galactosyl-hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁵⁸²	Glucosyl-galactosyl - hydroxylysine	Hydroxylysine
K ⁶¹⁷	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁶³⁵	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁶⁸¹	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁶⁹⁸	Hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁷³¹	No glycosylation detected	Glucosyl-galactosyl - hydroxylysine
K ⁷⁴²	Hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁷⁵⁷	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁸²⁵	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ⁸²⁸	Glucosyl-galactosyl - hydroxylysine	Hydroxylysine
K ¹⁰²⁵	Galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
K ¹⁰⁴³	Galactosyl / glucosyl-galactosyl-hydroxylysine	Hydroxylysine
K ¹⁰⁴⁶	Galactosyl / glucosyl-galactosyl-hydroxylysine	Galactosyl - hydroxylysine
K ¹⁰⁴⁹	Hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
K ¹⁰⁶⁶	Not detected	Galactosyl / glucosyl-galactosyl-hydroxylysine
K ¹⁰⁸¹	Galactosyl / glucosyl-galactosyl-hydroxylysine	Hydroxylysine
K ¹¹¹⁷	Galactosyl - hydroxylysine	Galactosyl - hydroxylysine
K ¹¹²⁰	Glucosyl-galactosyl - hydroxylysine	Hydroxylysine
K ¹¹³²	Galactosyl - hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
K ¹¹⁵⁰	No glycosylation detected	Galactosyl - hydroxylysine
K ¹¹⁰⁵	Galactosyl - hydroxylysine	Hydroxylysine
K ¹¹⁸⁵	Galactosyl – hydroxylysine	Glucosyl-galactosyl – hydroxylysine
K ¹¹⁸⁸	Hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
K ¹²⁰⁰	Glucosyl-galactosyl - hydroxylysine	Hydroxylysine
K ¹²⁰⁸	Galactosyl - hydroxylysine	Hydroxylysine
K ¹²⁰³	Calactosyl - nydroxylysine	nyuloxylysine
K ¹³⁰⁷	Not detected	Galactosyl / glucosyl-galactosyl-hydroxylysine
N ¹³¹⁷ I Z1328	Not detected	Guaedu galactosul hudrowitzing
N ¹³²⁰	Not detected	Glastogyl / glasgyl - hydroxylysine
K1371	Galactosyl - hydroxylycine	No glycosylation detected
K1380	Glucosyl galactosyl bydrowdwing	No glycosylation detected
IX 1407	Glucosyl galactosyl - hydroxylysille	Not detected
	Giucosyi-galaciosyi - nyuloxyiysine	

	А.													
	1-130	MGPRLSVWLL	LLFAALLLHE	ERSRAAAKGD	CGGSGCGKCD	CHGVKGQKGE	RGLPGLQGVI	GFPGMQGPEG	PHGPPGQKGD	AGEPGLPGTK	GTRGPPGAAG	YPGNPGLPGI	PGQDGPPGPP	GIPGCNGTKG
	131-260	BRGPLGPPGL	PGFSGNPGPP	GLPGMKGDPG	EILGHVPGTL	LKGERGF PGI	PGMPGSPGLP	GLQGPVGPPG	FTGPPGPPGP	PGPPGEKGQM	GSSFQGPKGD	KGEQGVSGPP	GVPGQAQVKB	KGDFAPTGEK
3	261-390	GQKGE PGF PG	VPGYGEKGE P	GKQGPRGKPG	KDGEKGERGS	PGIPGDSGYP	GLPGRQG PQG	EKGEAGLPGP	PGTVIGTMPL	GEKGDRGYPG	APGLRGEPGP	KGFPGTPGQP	GPPGF PT PGQ	AGA PGF PGER
	391-520	GEKGDQGPPG	VSLPGPSGRD	GAPGPPGPPG	PPGQPGHTNG	IVECQPGPPG	DQGPPGT PGQ	PGLTGEVGQK	GQKGESCLAC	DTEGLRGPPG	PQGPPGEIGF	PGQPGAKGDR	GLPGRDGLEG	LPGPQGSPGL
	521-650	IGQPGAKGEP	GEIFFDMRLK	GDKGDPGFPG	QPGMPGRAGT	PGRDGHPGLP	GPKGSPGSIG	LKGERGPPGG	VGFPGSRGDI	GPPGPPGVGP	IGPVGEKGQA	GF PGG PGS PG	LPGPKGEAGK	VVPLPGPPGA
1	651-780	AGLPGSPGFP	GPQGDRGFPG	TPGRPGIPGE	KGAVGQPGIG	FPGLPGPKGV	DGLPGEIGRP	GSPGRPGFNG	L PGN PG PQGQ	KGEPGIGLPG	LKGQPGLPGI	PGTPGEKGSI	GGPGVPGEQG	LTGPPGLQGI
0	781-910	RGDPGPPGVQ	GPAGPPGVPG	IGPPGAMGPP	GGQGPPGSSG	PPGIKGEKGF	PGFPGLDMPG	PKGDKGSQGL	PGLTGQSGLP	GLPGQQGTPG	VPGFPGSKGE	MGVMGTPGQP	GSPGPAGTPG	LPGEKGDHGL
8	911-1040	PGSSGPRGDP	GFKGDKGDVG	LPGMPGSMEH	VDMGSMKGQK	GDQGEKGQIG	PIGDKGSRGD	PGTPGVPGKD	GQAGHPGQPG	PKGDPGLSGT	PGSPGLPGPK	GSVGGMGLPG	SPGEKGVPGI	PGSQGVPGSP
	1041-1170	GEKGAKGEKG	QSGLPGIGI P	GRPGDKGDQG	LAGFPGSPGE	KGEKGSAGTP	GMPGSPGPRG	SPGNIGHPGS	PGLPGEKGDK	GLPGLDGVPG	VKGEAGLPGT	PGPTGPAGQK	GEPGSDGIPG	SAGEKGEQGV
3	1171-1300	PGRGF PGF PG	SKGDKGSKGE	VGF PGLAGS P	GIPGVKGEQG	FMGPPGPQGQ	PGLPGTPGHP	VEGPKGDRGP	QGQPGLPGHP	GPMGP PGF PG	INGPKGDKGN	QGWPGAPGVP	GPKGDPGFQG	MPGIGGSPGI
2021	1301-1430	TGSKGDMGLP	GVPGFQGQKG	LPGLQGVKGD	QGDQGVPGPK	GLQGPPGPPG	PYDVIKGEPG	LPGPEGPPGL	KGLQGPPGPK	GQQGVTGSVG	LPGPPGVPGF	DGAPGQKGET	GPFGPPGPRG	FPGPPGPDGL
8	1431-1560	PGSMGPPGTP	SVDHGFLVTR	HSQTTDDPLC	PPGTKILYHG	YSLLYVQGNE	RAHGQDLGTA	GS <mark>CLRKFSTM</mark>	PFLFCNINNV	CNFASRNDYS	YWLST PE PMP	MSMA PISGDN	IRPFISRCAV	CEAPAMVMAV
	1561-1669	HSQTIQIPQC	PNGWSSLWIG	YSFVMHTSAG	AEGSGQALAS	PGSCLEEFR S	APFIECHGRG	TCNYYANAYS	FWLATIERSE	MFKKPTPSTL	KAGELRTHVS	RCQVCMRRT		
P)													
).													
1	1-130	MGPRLSVWLL	LLFAALLLHE	ERSRAAAKGD	CGGSGCGKCD	CHGVKGQKGE	RGLPGLQGVI	GF PGMQGPEG	PHGP <mark>P</mark> GQKGD	AGE PGLPGTK	GTRGPPGAAG	YPGN PGLPGI	PGQDGPPGPP	GIPGCNGTKG
1	1-130 131-260	MGPRLSVWLL ERGPLGPPGL	LLFAALLLHE PGFSGNPGPP	ersraaakgd glpgmkgdpg	CGGSGCGKCD BILGHVPGTL	CHGVKGQKGE LKGERGF PGI	PGMPGSPGLP	GP PGMQ GPEG	PHGPPGQKGD	AGEPGLPGTK	GTRGPPGAAG GSSFQGPKGD	Y PGN PGL PGI KGEQGVSGPP	PGQDGPPGPP GVPGQAQVKB	GIPGCNGTKG KGDPAPTGEK
1	1-130 131-260 261-390	MGPRLSVWLL ERGPLGPPGL GQKGEPGFPG	LLPAALLLHE PGP SGN PGP P VPGYGEKGE P	ERSRAAAKGD GLPGMKGDPG GKQGPRGKPG	CGGSGCGKCD BILGHVPGTL KDGEKGERGS	CHGVKGQKGE LKGERGPPGI PGIPGDSGYP	RGLPGLQGVI PGMPGSPGLP GLPGRQGPQG	GP PGMQGPEG GLQGPVGPPG EKGEAGLPGP	PHGPPGQKGD PTGPPGPPGP PGTVIGTMPL	AGE PGLPGTK PGPPGEKGQM GEKGDRGYPG	GTRGP PGAAG GSSFQGPKGD APGLRGE PGP	Y PGN PGL PGI KGEQGVSGPP KGF PGT PGQP	PGQDGPPGPP GVPGQAQVKE GPPGPPTPGQ	GIPGCNGTKG KGDFAPTGEK AGAPGFPGER
1	1-130 131-260 261-390 391-520	NGPRLSVWLL ERGPLGPPGL GQKGEPGPPG GEKGDQGPPG	LLFAALLLHE PGFSGNPGPP VPGYGEKGEP VSLPGPSGRD	ERSRAAAKGD GLPGMKGDPG GKQGPRGKPG GAPGPPGPPG	CGGSGCGKCD RILGHVPGTL KDGEKGERGS	CHGVKGQKGE LKGERGP PGI PGI PGDSGYP IVECQPGPPG	RGLPGLQGVI PGMPGSPGLP GLPGRQGPQG DQGPPG7PGQ	GPPGMQGPEG GLQGPVGPPG EKGEAGLPGP PGLTGEVGQK	PHGPPGQKGD PTGPPGPPGP PGTVIGTMPL GQKGESCLAC	AGE PGLPGTK PGPPGEKGQM GEKGDRGY PG DTEGLRGPPG	GTRGP PGAAG GSSFQGPKGD APGLRGE PGP BOGPPGEIGP	Y PGN PGL PGI KGRQGV3GPP KGF PGT PGQP PGQPGAKGDR	PGQDGPPGPP GVPGQAQVKE GPPGPPTPGQ GLPGRDGLEG	GIPGCNGTKG KGDPAPTGEK AGAPGPPGER LPGPQGSPGL
	1-130 131-260 261-390 391-520 521-650	MGPRLSVWLL ERGPLGPPGL GQKGEPGPPG GEKGDQGPPG IGQPGAKGEP	LLPAALLLHE PGPSGNPGPP VPGYGEKGEP VSLPGPSGRD GEIFFDMRLK	ERSRAAAKGD GLPGMKGDPG GRQGPRGKPG GAPGPPGPPG GDKGDPGPPG	CGGSSGCGKCD RILGHVPGTL KDGEKGERGS == PPGOPGHING OPGNPGRAGT -==	CHGVKGQKGE LKGRRGP PGI PGIPGDSGYP IVECOPGPPG PGRDGHPGLP	RGLPGLQGVI PGM PGS PGL P GLPGRQG PQG DQGP PGT PGQ GPKGS PGS IG	GP PGNQGPEG GLQGPVGPPG EKGEAGLPGP PGLTGEVGQK LKGERGPPGG	PHGPPGQKGD PTGPPGPGPGP PGTVIGTMPL GQKGESCLAC	AGE PGLPGTK PGPPGRKGQM GEKGDRGY PG DTEGLRGPPG GPPGP PGVGP	GTRGP PGAAG GSSFQGPKGD APGLRGE PGP POGPPGEIGF IGPVGEKGQA	Y PGN PGL PGI KGRQGVSGPP KGP PGT PGQP PGQPGAKGDR GP PGGPG8 PG	PGQDGPPGPP GVPGQAQVKE GPPGPPTPQQ GLPGRDGLEG LPGPKGEAGK	GIPGCNGTKG KGDPAPTGRK AGAPGPPGRR LPGPQGSPGL VVPLPGPPGA
	1-130 131-260 261-390 391-520 521-650 651-780	MGPRLSVWLL ERGPLGPPGL GQKGEPGPPG GEKGDQGPPG IGQPGAKGEP AGLPGSPGP	LLPAALLLHE PGPSGNPGPP VPGYGEKGEP VSLPGPSGRD GEIFFDMRLK GPQGDRGFPG	ERSRAAAKGD GLPGMKGDPG GRQGPRGKPG GAPGPPGPPG GDKGDPGPPG TPGRPGIPGE	CGGS3GCGKCD EILGHVPGTL KDGEKGERGS PPGOPGHTNG OPGNPGRAGT 	CHGVKGOKGE LKGERGPPGI PGIPGDSGYP IVECOPGPPG PGRDGHPGLP PPGLPGPKGV	RGLPGLOGVI PGMPGSPGLP GLPGROGPOG DOGPPGTPGO GPKGSPGSIG DGLPGEIGRP	GP PGNQGPEG GLQGPVGPPG EKGEAGLPGP PGLTGEVGOK LKGERGPPGG GS PGRPGENG	PHGP PGQKGD PTGP PG PPGP PGT VIGT MPL GOKGESCLAC VGP PGS RGDI	AGE PGLPGTK PGPPGEKGQM GEKGDRGY PG DTEGLRGPPG GPPGP PGVGP KGGE PGIGLPG	GTRGPPGAAG GSSFQGPKGD APGLRGEPGP POGPPGEIGP IGPVGEKGQA LKGQPGLPGI	YPGN PGLPGI KGEQGVSGPP KGFPGTPGQP PGQPGAKGDR GFPGGPG8PG PGTPGEKGSI	PGQDGPPGPP GVPGQAQVKE GPPGPPTPGQ GLPGRDGLEG LPGPKGEAGK GGPGVPGEQG	GIPGCNGTKG KGDPAPTGEK AGAPGPPGER LPGP0GSPGL VVPLPGPPGA LTGPPGLQGI
	1-130 131-260 261-390 391-520 521-650 651-780 781-910	NGPRLSVWLL ERGPLGPPGL GOKGEPGFPG GEKGDOGFPG IGOPGAKGEP AGLPGSPGFP RGDPGPGVQ	LLPAALLLHE PGPSGNPGPP VPGYGEKGEP VSLPGPSGRD GEIPFDMRLK GPQGDRGFPG GPAGPPGVPG	ERSRAAAKGD GLPGMKGDPG GRQGPRGKPG GAPGPPGPPG GDKGDPGPPG TPGRPGIPGE IGPPGAMGPP	CGGSSGCGKCD RILGHVPGTL KDGEKGERGS == PPGOPGHTNG QPGNPGRAGT -== KGAVGOPGIG GGOGPPGSSG	CHGVKGOKGE LKGERGP PGI PGI PGDSGYP IVECOPGPPG PGRDGHPGLP PPGLPGPKGV PPGLKGEKGF	RGLPGLQGVI PGMPGSPGLP GLPGRQGPQG DQGPPGTPGQ GPKGSPGSIG DGLPGEIGRP PGPPGLDNPG	GP PGMQGPEG GLQGPVGPPG EKGEAGLPGP PGLTGEVGQK LKGERGPPGG GSPGRPGENG PKGDKGSQGL	PHGPPGQKGD PTGPPGPPGP PGTVIGTMPL GQKGESCLAC VGPPGSRGDI LPGNPGPQGQ PGLTQQSGLP	AGE PGLPGTK PGPPGEKGQM GEKGDRGY PG DTEGLRGPPG GPPGP PGVGP - GLPGQQGT PG	GTRGPPGAAG GSSPQGPKGD APGLRGEPGP DOGPPGEIGP IGPVGEKGQA LKGQPGLPGI VPGEPGSKGE	YPGN PGLPGI KGEQGVSGPP KGP PGT PGQP PGQPGAKGDR GF PGGPG8 PG PGT PGEKGSI MGVMGT PGQP	PGQDGPPGPP GVPGQAQVKE GPPGFPTPGQ GLPGRDGLEG 	GIPGCNGTKG KGDPAPTGEK AGAPGPPGER LPGP9GS9GL VVPLPGP9GA LTGP9GLQGI LPGERGDHGL
	1-130 131-260 261-390 391-520 521-650 651-780 781-910 911-1040	NGPRLSVWLL RRGPLGPPGL GQKGEPGPPG GEKGDQGPPG IGQPGAKGEP AGLPGSPGPP RGDPGPPGVQ - PGSSGPRGDP	LLPAALLLHE PGPSGNPGPP VPGYGEKGEP VSLPGPSGRD GEIPFDMRLK GPQGDRGFPG GPAGPPGVPG GPKGDKGDVG	ERSRAAAKGD GLPGMKGDPG GKQGPRGKPG GAPGPPGPPG GDKGDPGPPG TPGRPGIPGE IGPPGAMGPP LPGMPGSMEH	CGGSGCGKCD RILGHVPGTL KDGEKGERGS PPGOPGHTNG OPGNPGRAGT CGQGPPGSGG GGQGPPGSSG VDMGSMKGQK	CHGVKGQKGE LKGRRGP PGI PGI PGDSGYP IVECOPG PPG PGRDGHPGLP PPGLRGPKGV GDQGEKGQIG	RGLPGLOGVI PGMPGSPGLP GLPGROGPOG DOGPPGTPGO GPKGSPGSIG DGLPGEIGRP PGFPGLDMPG PTGCKGSPG	GPPGMQGPEG GLQGPVGPPG EKGEAGLPGP PGLTGEVGQK LKGERGPPGG GSPGRPGENG PKGDKGSQGL 	PHGP PGQKGD PTGP PGQ PPGP GQKGESCLAC VGP PGS RGDI LPGN PG PQGQ PGLTGQ SGLP GQAGH PGQPG	AGE PGLPGTK PGPPGRKGQM GEKGDRGY PG DTEGLRGPPG GPPGP PGVGP KGB PGIGLPG - GLPGQQGT PG PKGDPGLSGT	GTRGP PGAAG GSSFQ GPKGD APGLRGE PGP POGP PGEIGP IGP VGE KGQA LKGQP GLPGI VPGF PGSKGE PGSPGLPGPK	YPGN PGLPGI KGEQGVSGPP KGFPGTPGQP PGOPGAKGDR GFPGGPGSPG PGTPGEKGSI NGVNGTPGQP	PGQD GPPGPP GVPGQAQVKE GPPGP PT PGQ GLPGRDGLEG GLPGPKGEAGK GGPGVPGEQG GSPGPAGT PG SPGEKGVPGI	GIPGCNGTKG KGDPAPTGRK AGAPGPPGER LPGPQGSPGL VVPLPGPPGA LTGPPGLQGI LPGEKGDHGL PGSQGVPGSP
	1-130 131-260 261-390 391-520 521-650 651-780 781-910 911-1040 1041-1170	NGPRLSVWLL ERGPLGPPGL GOKGEPGPPG GEKGDQGPPG IGOPGAKGEP AGLPGSPGPP RGDPGPPGVQ - PGSSGPRGDP GEKGARGEKG	LLPAALLLHE PGP SGN PGP P VPGYGEKGE P SEIPFDARLK GEIPFDMRLK GPQGDRGF PG GPAGPRGVPG GPKGDKGDVG	ERSRAAAKGD GLPGMKGDPG GKQGPRGKPG GAPGPPGPPG GDKGDPGPPG TPGRPGIPGE IGPPGANGPP LPGNPGSMEH GRPGDKGDQG	CGGSGCGKCD RILGHVPGTL KDGEKGERGS 	CHGVKGOKGE LKGERGP PGI PGI PGDSGYP IVECOPG PPG PGRDGHPGLP PGLPGPKGV GDQGEKGQIG KGEKGSAGTP	RGLPGLQGVI PGMPGSPGLP GLPGRQGPQG DOGPPGTPGQ GPRGSPGSIG DGLPGEIGRP PGP FGLDMPG PTGDKGSRGD 	GP PGNQGPEG GLQGPVGPPG EKGEAGLPGP PGLTGEVGOK LKGERGPPGG GSPGRPGENG PKGDRGSQGL PGT PGV PGKD SPGNIGHPGS	PHGP PGQKGD PTGP PG PPGP GOKGES CLAC VGP PGS RGDI LPGN PG PGGQ GQAGH PG QPG PGL PGQ PG PG PGL PG EKGDK	AGE PGLPGTK PGPPGERKGQM GEKGDRGY PG DTEGLRGPPG GPPGP PGVGP - GLPGQQGT PG PKGDPGLSGT GLPGLDGVPG	GTRGP PGAAG GSSFQGPKGD APGLRGE PGP POGPPGEIGP IGPVGEKGQA LKGQPGLPGI VPGE PGSKGE PGSPGLPGPK	YPGN PGLPGI KGEQGVSGPP FGCPGTPGQP PGOPGAKGDR GFPGGPGSPG GFPGGPGSPG MGVMGTPGQP GSVGGMGLPG PGPTGPAGQK	PGOD GPPGPP GVPGQAQVKE GPPGP PT PGQ GLPGRDGLEG LPGPKGEAGK GSPGPKGEAGK GSPGPAGEPG SPGEKGVPGI GEPGSDGI PG	GIPGCNGTKG KGDPAPTGEK AGAPGPPGER LPGPGGSPGL VVPLPGPPGA LTGPPGLQGI LPGEKGDHGL PGSQGVPGSP GAGEKGEQGV
) 1-130 131-260 261-390 391-520 521-650 651-780 781-910 911-1040 911-1040 1041-1170 1171-1300	NGPRLSVWLL RRGPLGPPGL GOKGEPGPPG GEKGDQGPPG IGQPGAKGEP AGLPGSPGPP RGDPGPPGVQ - PGSSGPRGDP GEKGAKGEKG PGRGPPGPPGP	LLPAALLLHE PGPSGNPGPP VPGYGEKGEP GEIPFDMRLK GPQGDRGPPG GPAGPPGVPG GPKGDKGDVG QSGLPGIGIP	ERSRAAAKGD GLPGMKGDPG GADGPPGPPG GDKGDPGPPG TPGRPGIPGE LPGMPGGMEH GRPGDKGDQG	CGGSGCGKCD RILGHVPGTL KDGEKGERGS OPGNPGRAGT OPGNPGRAGT C GGQGPPGSSG C UDMGSMKGQK LAGPPGSPGE GIPGVKGEQG	CHGVKGOKGE LKGERGPPGI PGIPGDSGYP IVECOPGPPG PGRDGHPGLP PPGLKGEKGP GDQGERGQIG KGEKGSAGTP EMGEPGPQGQ	RGLPGLQGVI PGMPGSPGLP GLPGRQGPQG GDGPPGTPGQ GPKGSPGSIG DGLPGEIGRP PGPPGLDMPG PTGDKGGRGD GMPGSPGPRG PGLPGTPGHP	GP PGNQGPEG GLQGPVGPPG EKGEAGLPGP PGLTGEVGQK LKGERGPPGG GSPGRPGPNG PKGDKGSQGL 	PHGP PGQKGD PTGP PG PPGP GOKGES CLAC VGP PGS RGDI LPGN PG PGG GQAGH PGQ PG PGL PGC PGG PGL PGC PGG QGQ PGL PGHP	AGE PGLPGTK PGPPGERGQM GERGDRGY PG GERGDRGY PG GPPGP PGVGP - GLPGQQGT PG - GLPGQQGT PG - GLPGLDGV PG GPMGP PGE PG	GTRGP PGAAG GSSFQGPKGD APGLRGE PGP POGPPGEIGF IGPVGEKGQA LKGQPGLPGI VPGE PGSKGE PGSPGLPGPK VKGEAGLPGT	YPGN PGLPGI KGEQGVSGPP KGP PGT PGQP PGOPGAKGDR GP PGGPG8 PG PGT PGGPG8 PG GSVGGMGL PG PGT PGEKG8I GSVGGMGL PG PGPT GPAGQK	PGQDGPPGPP GVPGQAQVKE GPPGPPTPGQ GLPGRDGLEG LPGPKGEAGK GSPGPAGEQG GSPGPAGEQG GSPGPAGEQG GEPGSDGIPG GPKGDPGEQG	GIPGCNGTKG KGDPAPTGEK AGAPGPPGER LPGPQGSPGL VVPLPGPPGA LTGPPGLQGT LPGEKGDHGL PGSQGVPGSP SAGEKGEQGV
) 1-130 131-260 261-390 391-520 521-650 651-780 781-910 911-1040 1041-1170 1041-1170 1171-1300 1301-1430	NGPRLSVWLL ERGPLGPPGL GOKGEPGFPG GEKGDOGFPG AGLPGSPGFPG PGSSGPRGDP GEKGARGEKG PGGEFPGFPG TGSKGDMGLP	LLPAALLLHE PGPSGNPGPP VPGYGEKGEP GEIFPDMRLK GPQGDRGPPG GPKGDKGDVG QSGLPGIGIP SKGDKGSKGE	ERSRAAAKGD GLPGMKGDPG GAPGPPGPPG GDKGDPGPPG TPGRPGIPGE IGPPGAMGPP LPGMPGSMEH GRPGDKGDQG VGEPGLAGS P LPGLQGVKGD	CGGSGCGKCD RILGHVPGTL KDGEKGERGS OPGNPGRAGT COPGNPGRAGT COPGNPGRAGT COPGNPGRAGT COPGNPGRAGT COPGNPGSSG COPGNGSMKGQK LAGE PGSPGE GIPGVKGEQG COPGNGSPGEK	CHGVKGOKGE LKGERGPPGI PGIPGDSGYP IVECOPGPPG PGRDGHPGLP PPGLKGERGF GDQGEKGQIG KGEKGSAGTP FMGPPGPQGQ GLQGPPGPPG	RGLPGLQGVI PGMPGSPGLP GLPGRQGPQG GPKGSPGSIG DGLPGEIGRP PGEPGLDNPG PTGDRGSRGD GMPGSPGPRG PGLPGTPGHP PYDVIKGEPG	GP PGNQGPEG GLQGPVGPPG EKGEAGLPGP PGLTGEVGOK LKGERGPPGG GSPGRPGPNG GSPGRPGPNG PGTPGVPGKD SPGNIGHPGS VEGPKGDRGP	PHGP PGQKGD FTGP PG PPGP GOKGESCLAC VGP PGS RGDI LPGN PG PGGQ GQAGH PGQ PG GQAGH PGQ PG PGL PGE KGDK PGL PGE KGDK KGL QGP PGFK	AGE PGLPGTK PGPPGRKGQM GEKGDRGY PG GPPGP PGVGP GPPGP PGVGP GLPGLGLPG - GLPGQGGT PG GLPGLDGVPG GLPGLDGVPG GPQGVTGSVG	GTRGP PGAAG GSSFQGPKGD PQGPPGEIGP IGPVGEKGQA LKGQPGLPGI VPGF PGSKGE PGSPGLPGPK VKGEAGLPGT INGPKGDKGN	Y PEN PGL PGI KGEQGV3GPP FGOPGAKGDR GP PGGPG3PG PGT PGEKGSI NGVNGT PGQP GSVGCNGL PG PGPT GPAGQK QGW PGAPGVP DGAPGQKET	PGQDGPPGPP GVPGQAQVKE GPPGPPTPGQ GLPGRDGLEG LPGPKGEAGK GGPGVPGEQG GSPGPAGEPG GSPGPAGEPG GEPGSDGIPG GPFGPBGPRG GPFGPBGPRG	GIPGCNGTKG KGDPAPTGEK AGAPGPPGER LPGPQGSPGL LTGPPGLQGI LPGEKGDHGL PGSQGVPGSP SAGEKGEQGV MPGIGGSPGI FPGPPGPGGL
) 1-130 131-260 261-390 391-520 521-650 651-780 911-1040 911-1040 1041-1170 1171-1300 1301-1430 1431-1560	NGPRLSVWLL RRGPLGPPGL GQKGEPGPPG GEKGDQGPPG AGLPGSPGPP RGDPGPPGVQ - PGSSGPRGDP GEKGAKGEKG PGRGPPGPPG TGSKGDHGLP PGSMGPPGPP	LLPAALLLHE PGPSGNPGPP VPGYGEKGEP VSLPGPSGRD GEIPFDMRLK GPQGDRGFPG GPAGPPGVPG GPKGDKGDVG QSGLPGIGIP SKGDKGSKGE GVPGPQGQKG 	ERSRAAAKGD GL PGMKGDPG GKQGPRGKPG GDKGDPGPPG TPGRPGIPGE IGPPGANGPP LPGMPGSMEH GR PGDKGDQG VGF PGLAGS P LPGLQGVKGD	CGGSGCGKCD RILGHVPGTL KDGEKGERGS PPGOPGHTNG OPGNPGRAGT COPGNPGRAGT GGQGPPGSSG COPGNKGQK GIPGVKGEQG GIPGVKGEQG COPGSVEGPK PPGTKILYHG	CHGVKGOKGE LKGERGP PGI PGI PGDSGYP IVECOPG PPG PGRDGHPGLP PPGLRGERGP GDQGEKGQIG KGERGSAGTP FMGPPGPGQQ GLQGPPGPPG YSLLYVQGNE	RGLPGLOGVI PGMPGSPGLP GLPGROGPOG DOGPPGTPGO GPKGSPGSIG PGCPGLDMPG PTGDKGSPGPRG PGLPGTPGHP PYDVIKGEPG RAHGQDLGTA	GPPGMQGPEG GLQGPVGPPG EKGEAGLPGP PGLTGEVGQK LKGERGPPGG GSPGRPGENG PKGDKGSQGL 	PHGP PGQKGD PTGP PGQ PGQ GQKGESCLAC VGP PGS RGDI LPGN PG PQGQ GQAGH PGQPG PGL PGQ PGQ GQAGH PGQPG PGL PGC PGPK QGQ PGL PGH P KGLQG P FGPK	AGE PGLPGTK PGPPGRKGQM GEKGDRGY PG DTEGLRGPPG GPPGP PGVGP GLPGLGGPG GLPGQQGTPG GLPGLDGVPG GPMGP PGF PG GQQGVTGSVG CNPASRNDYS	GTRGP PGAAG GSSFQGPKGD APGLRGE PGP POGPPGEIGF IGPVGEKGQA LKGQPGLPGI VPGF PGSRGE PGSPGLPGPK VKGEAGLPGT INGPK GDKGN	YPGN PGLPGI KGEQGVSGPP PGOPGAKGDR GFPGGPGSPG PGTPGEKGSI NGVNGTPGQP GSVGGNGLPG PGPTGPAGQK QGWPGAPGVP DGAPGQKGET NSMA PISGDN	PGOD GPPGPP GVPGQAQVKE GPPGP PT PGQ GLPGRDGLEG GSPGPAGEQG GSPGPAGEQG GSPGPAGEQG GSPGPAGEQG GPFGPGPGFQG GPFGPPGPRG IRPFISRCAV	GI PGCNGTKG KGDPAPTGRK AGA PGP PGER LPG POGSPGL LTG PPGLQGI LTG PPGLQGI LPG EKGDHGL PGSQGVPGSP SAGEKGEQGV MPGIGGSPGI F PGPPGPDGL CEA PAMWMAV
) 1-130 131-260 261-390 391-520 521-650 651-780 781-910 911-1040 1041-1170 1171-1300 1301-1430 1431-1560 1561-1669	NGPRLSVWLL ERGPLGPPGL GOKGEPGPPG GEKGDOGPPG AGLPGSPGPP RGDPGPPGVQ - PGSSGPRGDP GEKGARGERG PGRGEPGEPG TGSKGDMGLP RGSMGPPGTP	LLPAALLHE PGFSGNPGPP VPGYGEKGEP GEIPFDMRLK GPQGDRGFPG GPRGDRGPVG GSGLPGIGIP SKGDKGSKGE GVFGFQGQKG 	ERSRAAAKGD GLPGMKGDPG GKQGPRGKPG GAPGPPGPPG GDKGDPGPPG TPGRPGIPGE IGPPGAMGPP LPGMPGSMEH GRPGDKGDQG VGEPGLAGSP LPGLQGVKGD HSQTTDDPLC	CGGSGCGKCD RILGHVPGTL KDGEKGERGS PPGOPGHTNG OPGNPGRAGT GPGVPGPGIG GGOGPPGSSG GGOGPPGSSG GIPGVKGEQG GIPGVKGEQG GIPGVKGEQG GIPGVKGEQG GIPGVKGEQG AEGSGQALAS	CHGVKGOKGE LKGERGP PGI PGI PGDS GY P IVECOPG PPG PGRDGHPGLP PGRDGHPGLP PGRDGHPGLP GDQGEKGQIG KGENGSAGT P GDQGEKGQIG GLQGPPG PPG GLQGPPG PPG YSLLYVQGNE PGSCLEEFRS	RGLPGLQGVI PGMPGSPGLP GLPGROGPQG GDGDPG7PGO GDGDPG7PG0 GPKGSPGSIG PGLPGEIGRP PTGDKGSRGD 	GP PGNQGPEG GLQGPVGPPG EKGEAGLPGP PGL7GEVGQK LKGERGPPGE GSPGRPGENG PKGDKGSQGL PGT PGVPGKD SPGNIGHPGS VEGPKGDRGP LPGPEGPPGL LPGPEGPPGL CSCLRKFSTM	PHGP PGQKGD PTGP PG PPGP GOKGES CLAC VGP PGS RGDI LPGN PG PGGQ GQAGH PG QPG GQAGH PG QPG PGL PGE KGDK CGQ QC QC PG PG PK CGL QG QP FG PK CGL QG QP FG PK CGL QG QP FG PK CGL QG PFG PK CGL QG PK CGL QG PK CGL QG PK CGL QG PK CGL QG PK CGL QGL QGL QGL QGL QGL QGL QGL QGL QGL Q	AGE PGLPGTK PGPPGRKGQM GEKGDRGY PG DTEGLRGPPG GUPGLPGPG GLPGQQGTPG GLPGLDGVPG GLPGLDGVPG GPMGPPGPPGP GUPGPGPGPG GPMGPPGPPG MFKRPTPSTL	GTRGP PGAAG GSSFQ GPKGD APGLRGE PGP POGPPGEIGF IGPVGEKGQA LKGQPGLPGI VPGF PGSKGE PGSPGLPGPK VKGEAGLPGT INGPKGDKGN LPGPPGVPGF YELST PE PMP KAGELRTHVS	YPGN PGLPGI KGEQGVSGPP PGOPGAKGDR GPPGPGSPG PGOPGAKGDR GGVGGRGLPG GSVGGRGLPG PGPTGPAGQK QGWPGAPGVP DGAPGQKGET NSMA PISGDN	PGQD GPPGPP GVPGQAQVKE GPPGP PT PGQ GLPGRDGLEG GSPGPAGEAGK GSPGPAGEAGK GSPGPAGEPG GSPGBAGEPG GPFGPGPGFQG GPFGPFGPFQFQ IRPFISRCAV	GI PGCNGTKG KGDPAPTGRK AGA PGP PGER LPG PGGSPGL VVPL PGPSGSPGL LTG PPGLQGT LPG PGG QGV PGSP SAGEKGEQGV MPGI GGS PGI FPG PFG PGGL CEA PAMVMAV

Supporting figure S1: A representative diagram showing improved sequence coverage of mouse col4a1 by optimizing the number of allowed posttranslational modifications and miscleavages in the database search using MyriMatch. A. A MS data file from a trypsin digested EHS collagen sample was searched with a standard database search using carbamidomethylation as static modification, oxidation of methionine as dynamic modification and allowing for only 2 dynamic modifications per peptide. B. The same MS data file searched with MyriMatch, but increasing the number of miscleavages (up to 4) and dynamic modifications (up to 10) using the hydroxylation and O-linked glycosylation motifs described in Table 1.

Supporting figure S2: Peptide spectrum matches (PSMs) of 35 O-glycosylation,1 hydroxyproline site at Xaa position of Gly-HyP-Gln motif and 7 3-hydroxyproline sites in mouse col4a1 chain. PSMs were annotated using pLABEL¹ or IDPicker utility software Ion Matcher².

m/z = 780.3089 charge = +4

28
KJGJDJC₊₅₇GGSJGC₊₅₇GKJC₊₅₇DC₊₅₇HGVK₊₃₄₀GQK₊₃₄₀GE⁵⁰





m/z= 91 charge=-	5.5043, +5	PSM showing the identification of Galactosyl-HyK ^{217,228,231}													Mouse col4a1 chai				hai					
¹⁹⁷ G	¹⁹⁷ G P P ₊₁₆ G F T G P P ₊₁₆ G P P ₊₁₆ G P P G P P ₊₁₆ G E K ₊₃₄₀ G Q M G S S F Q C												G P	×+34	_o Gl) <mark>K</mark> +	- ₃₄₀ G	E ²³³						
																Oı	rbitra	ap-V	elos	5, ET	D Fi	ragme	ntatic	วท
See+002	o2 - Cr1-H83+++ 211.25 zr2-9461uccco1taal actcox/1[1]++ 211.25 = - cd-H83+++ 243.47 =	←		$\frac{1}{2} = \frac{1}{2} + \frac{1}$	$\frac{1}{1-2} \frac{1}{2} $	→ ²⁰ + ² [fi] -3-61iuosyijalačiosy1[1] ⁺⁺⁺⁺ 20.43 c23-61uosy1galactosy1[1] ⁺⁺⁺ 720.43 z14-2•61uosy1galactosy1[1] ⁺⁺ 720.43 <u>− − − − − − − − − − − − − − − − − − − </u>	$= \frac{-24}{2} + 72.06 \cdot 65$ $= \frac{-24}{2} + 70.06 \cdot 65$ $= \frac{-24}{2} - 720.64$	27. 62.0.5.16.16.0.601.601.601.00001[1]+++++ 649.71 c25-3+61100007[6]1actoox/1[1]++ 649.71	06	A	1100 2010 2010 2010 2010 2010 2010 2010		21	221-2461ucosy18,18,18ctcosy118,1+ 1245, 84 211-21-2011, 1257, 872, 89619-14, 1281, 281 211-2511ucosy ² (14, 1287, 252, 2-661ucosy18,11actcosy18,11a 101-101-101-101-101-101-101-101-101-101		67		<pre></pre>	8	22 - 224-1204-1 1866. 39 - 224-1204-1204-1204-1214-1214-1214-1214-1	===229-61ucos/1galactos/1[r] ++ 1762.00	8 ====================================	20 ► 213-1101 1928.57 = 213-1102 1928.56 [1] -246incoso/islatictoso/islatictoso/islatictoso/islatictoso/islatictoso/islatictoso/isla	00

7

in





PSM showing the identification of Glucosylgalactosyl-HyK^{322,343}







m/z= 956.9242 charge=+4							PSM showing the identification of Glucosylgalactosyl-HyK ^{460,463}											Mouse col4a1 chain			
43	⁴ C ₊₅₇ (G P	P ₊₁₆	G D	Q G P	P ₊₁	₆ G T	PG		ſG		ĠΕ	V	GC	K	+34	10 G	ίο κ	+340	G E ⁴⁶⁵
1.4	e+005 吕⊣		289.10						88 y16-NH3++-		syl[K]+++ 90C			(Q E>	kact	ive	, HCE) Frag	ment	ation
Relative Intensity (%)	0 10 20 30 40 50 50 50 50 50 50 50 50 50 50 50 50 50	90 ⁻ 170 ⁻ + 5 ⁰ 200	100	от по	00	B Image: 10 model Image: 10 model Image: 10 model 20 Image: 10 model Image: 10 model Image: 10 model 20 20 Image: 10 model Image: 10 model Image: 10 model 20 20 Image: 10 model Image: 10 model Image: 10 model 20 20 Image: 10 model Image: 10 model Image: 10 model 20 20 Image: 10 model Image: 10 model Image: 10 model 20 20 Image: 10 model Image: 10 model Image: 10 model 20 20 Image: 10 model Image: 10 model Image: 10 model 20 20 Image: 10 model Image: 10 model Image: 10 model 20 20 Image: 10 model Image: 10 model Image: 10 model 20 20 Image: 10 model Image: 10 model Image: 10 model 20 20 Image: 10 model Image: 10 model 10 model 20 20 Image: 10 model Image: 10 model 10 model 20 20 <td>8 </td> <td></td> <td>24 2461ucosylealactosyl[K]++ 758,88 y16-H20+++ 758, 2461ucosylealactosyl[K]++ 758,88 y16-H20+++ 758, 2461ucosylealactosyl[K]++ 758,88 y16-H20+++ 758, 2461ucosylealactosylea</td> <td>- y-1.100037 Idea refronse 1 (a) ++ 7.95, 10 - weak <t< td=""><td>0 * 24-60 corr) (k) +++ 682, 42 * 21-80 +++ * 23-50 for corr) (k) +++ 682, 42 * 23-50 for corr) (k) +++ 612, 45 * * 22-26 for corr 1 (k) +++ 612, 45 * 22-50 for corr 1 (k) +++ 612, 45 * 23-50 for corr 1 (k) ++ 612, 45 * * 23-50 for corr 1 (k) ++ 612, 45 * 21-80 ++ * 200, 55 * 23-50 for corr 1 (k) ++ 612, 45</td><td>=====================================</td><td>bi0+ 100, 40 k31-246hcosy1galactosy1[K]+++ 1036, 00 k31-246hcosy1galactosy1[K]+++ 1030, 40 k31-246hcosy1galactosy1[K]+++ 1000, 40 k32+#34++ 1045, 48 k32+#34+++ 1045, 48 k32+#34+++ 1045, 48 k32+#34+++ 1045, 48 k32+#34+++ 1045, 48 k32++++++++++++++++++++++++++++++++++++</td><td>0.2</td><td></td><td>1210-01-02-121-010-002-1681ect-023/181++1187.55</td><td></td><td></td><td>05.080.154 ±120+ 1360.56 bit =bit4+ 1378.54</td><td>0 </td><td>1500-y15-2461acosy1ealactors/1K1+ 1516.72</td></t<></td>	8 		24 2461ucosylealactosyl[K]++ 758,88 y16-H20+++ 758, 2461ucosylealactosyl[K]++ 758,88 y16-H20+++ 758, 2461ucosylealactosyl[K]++ 758,88 y16-H20+++ 758, 2461ucosylealactosylea	- y-1.100037 Idea refronse 1 (a) ++ 7.95, 10 - weak - weak <t< td=""><td>0 * 24-60 corr) (k) +++ 682, 42 * 21-80 +++ * 23-50 for corr) (k) +++ 682, 42 * 23-50 for corr) (k) +++ 612, 45 * * 22-26 for corr 1 (k) +++ 612, 45 * 22-50 for corr 1 (k) +++ 612, 45 * 23-50 for corr 1 (k) ++ 612, 45 * * 23-50 for corr 1 (k) ++ 612, 45 * 21-80 ++ * 200, 55 * 23-50 for corr 1 (k) ++ 612, 45</td><td>=====================================</td><td>bi0+ 100, 40 k31-246hcosy1galactosy1[K]+++ 1036, 00 k31-246hcosy1galactosy1[K]+++ 1030, 40 k31-246hcosy1galactosy1[K]+++ 1000, 40 k32+#34++ 1045, 48 k32+#34+++ 1045, 48 k32+#34+++ 1045, 48 k32+#34+++ 1045, 48 k32+#34+++ 1045, 48 k32++++++++++++++++++++++++++++++++++++</td><td>0.2</td><td></td><td>1210-01-02-121-010-002-1681ect-023/181++1187.55</td><td></td><td></td><td>05.080.154 ±120+ 1360.56 bit =bit4+ 1378.54</td><td>0 </td><td>1500-y15-2461acosy1ealactors/1K1+ 1516.72</td></t<>	0 * 24-60 corr) (k) +++ 682, 42 * 21-80 +++ * 23-50 for corr) (k) +++ 682, 42 * 23-50 for corr) (k) +++ 612, 45 * * 22-26 for corr 1 (k) +++ 612, 45 * 22-50 for corr 1 (k) +++ 612, 45 * 23-50 for corr 1 (k) ++ 612, 45 * * 23-50 for corr 1 (k) ++ 612, 45 * 21-80 ++ * 200, 55 * 23-50 for corr 1 (k) ++ 612, 45	=====================================	bi0+ 100, 40 k31-246hcosy1galactosy1[K]+++ 1036, 00 k31-246hcosy1galactosy1[K]+++ 1030, 40 k31-246hcosy1galactosy1[K]+++ 1000, 40 k32+#34++ 1045, 48 k32+#34+++ 1045, 48 k32+#34+++ 1045, 48 k32+#34+++ 1045, 48 k32+#34+++ 1045, 48 k32++++++++++++++++++++++++++++++++++++	0.2		1210-01-02-121-010-002-1681ect-023/181++1187.55			05.080.154 ±120+ 1360.56 bit =bit4+ 1378.54	0 	1500-y15-2461acosy1ealactors/1K1+ 1516.72



m/z= 782.9878 charge=+5

474GLRG	P ₊ P ₊	₁₆ G	PQ	GΡ	P ₊₁₆	GE	ÍG	FP+	16 G	QP	+16		+340	G C		ι L P₊	-16 G	RIDO	5 L E ⁵⁰⁹
Relative Intensity (%)		8	60 00	000	00 10 10 10 10 10 10 10 10 10 10 10 10 10 1	8	6	200 	Mill Mill <th< td=""><td>6 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td><td>0 0 0 0 0 0 0 0 0 0 0 0 0 0</td><td>001</td><td>1200</td><td>g</td><td>Exac</td><td>tive, F</td><td>HCD F</td><td>ragme</td><td>ntation</td></th<>	6 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0	001	1200	g	Exac	tive, F	HCD F	ragme	ntation
																	n	n/z	





PSM showing the identification of Galactosyl-HyK⁵⁷³





m/z= 866.9266charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK⁶⁸¹



m/z= 1010.4997 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK⁷⁵⁷

Mouse col4a1 chain







m/z= 627.9313 charge=+6

¹⁰¹⁵G M G L P G S P G E
$$K_{+178}$$
 G V P₊₁₆G I P G S Q G V P₊₁₆G S P G E K G A K_{+340} G E K_{+16}

Orbitrap-Velos, CID Fragmentation







PSM showing the identification of Galactosyl-HyK 1081

Q Exactive, HCD Fragmentation



m/z= 793.6757 charge=+6





m/z= 569.0098 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{1120,1132}



m/z= 997.968 charge=+4







m/z= 1014.0001 charge=+2

PSM showing the identification of Glucosylgalactosyl-HyK 1206





¹²⁶⁶GD K₊₁₇₈ GNQ GWPGAPGVPGPK₊₁₇₈ GD P₊₁₆ GFQGM₊₁₆P₊₁₆ GIGGSP₊₁₆ GI T G S K¹³⁰⁴









m/z= 1141.5 (+2)

PSM showing the identification of 3-HyP⁴⁴⁴

Mouse col4a1 chain

$$^{434}C_{+57} Q P G P_{+16} P G D Q G P_{+16} P_{+16} G T P_{+16} G Q P_{+16} G L T G E^{456}$$

Q Exactive, HCD Fragmentation


m/z= 767.34 (+3)

PSM showing the identification of 3-HyP⁴⁷⁸

Mouse col4a1 chain

Q Exactive, HCD Fragmentation



m/z= 682.33 (+2)

PSM showing the identification of HyP⁴⁸¹

Mouse col4a1 chain



Orbitrap-Velos, CID Fragmentation





38



⁵⁹⁸G D I G P₊₁₆ P₊₁₆ G P₊₁₆ G V G P I G P V G E K₊₃₄₀ G Q A G F P₊₁₆ G G P₊₁₆ G S P₊₁₆ G L P G P K₊₃₄₀ G E A G K⁶⁴⁰

Mouse col4a1 chain

m/z= 1148.28 (+4)

PSM showing the identification of 3-HyP⁶⁴⁷ m/z= 830.744 (+3) Mouse col4a1 chain $^{641}VVPLP_{+16}GP_{+16}PGAAGLP_{+16}GSP_{+16}GFP_{+16}GPQGDR^{666}$ 40 Orbitrap-Velos, CID Fragmentation y13+2 30.744 659.06 precursor) 30 Intensity (10^3) 20 у7 b13 y22+2 4.67 1041.79 y10+2 530.40 10 10 y7+2 371.95 y13 1316.67 b19+2 b16 b18 1601.76 258.17 b20 0 500 1000 1500 20 0 m/z

PSM showing the identification of 3-HyP¹⁴²⁴, 3-HyP¹⁴³⁶

Mouse col4a1 chain

1420
GFP₊₁₆ G^{P}_{+16} P_{+16}^{-16} $G^{P}_{P}_{+16}^{-16}$ $G^{P}_{+16}^{-16}$ $G^{P}_{+16}^$

Q Exactive, HCD Fragmentation



Supporting figure S3: Peptide spectrum matches (PSMs) of 39 O-glycosylation, 14 3-hydroxyproline and 3 hydroxyproline sites at Xaa position of Gly-HyP-Gln motif in human col4a1 chain. PSMs were annotated using pLABEL¹ or IDPicker utility software Ion Matcher².

m/z= 1200.8137; charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{78,90}



m/z= 957.6457; charge=+5

PSM showing the identification of Galactosyl-HyK^{78,90}



m/z= 1001.8837; charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK¹⁵⁶





PSM showing the identification of Glucosylgalactosyl-HyK^{295,298}



m/z= 920.4349; charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{321,343}



m/z= 1247.923; charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK³⁶¹

		356 (GEP	+16 G	PK.	+ 340	GF	P ₊₁	6 GL	P ₊₁₆	GC	2P ₊₁	₆ G	PP+	16 G	_P ₊₁₆ V	PIG		AP	+16	G F P	+16 G	E R ³⁹⁰
Relative Intensity (%)	20 30 40 50 80 70 80 90	356 (A4+ 474.1	+16 ^G +16 ^G	actosv1[K]++ F _{y13+} ²³ 636.7	+ 340 + 362.2	GF	lucosylgalactosyl[K]+ 802.2	erori'r: ***** *****		tosyl[K]++ '1024.[j]+++ 1019.8	Heese bill-Glucosylgalactosyl[K]+ 1073.8 ++	+ 1129.75]+++ 1121.3 [M]-Glucosylgalactosyl[K]+++ 1140.4 v24++ 1167.3	<u>vlsalactosyl [K] + 1186.0</u> sylsalactosyl [K] + 1198.7 1235.9	alactosyl[K]++ 1278.8 <u>b28</u> -6/10cosyl2alactosyl[K]++ 1314.3	Plaalactosyl [K]++ 1370.3 sylgalactosyl [K]++ 1320.3 tcovul [K]++ 1399.4	Tripe accult[K]++ 1472.9	alactosyl [K]++ 1531.4 sylsalactosyl [K]+ 1553.9	1 ctosyl[K]+ 1650.8	-Glucosylgalactosyl [K]++ 1709.5	€ F	+16JG	E R ³⁹⁰
	0 			bi7-Glucosylgalact	6000 FTO S	Bild-Glucosylgalac	v15++ 743.2	008	00 00 00 00 00 00 00 00 00 00	v2011.5	ĕ ► b21-Glucosyl≋alact	10 100 1000 1000 1000 1000 1000 1000 1	22/mm/23/14/1125/314/1	b12-61ucosy	b27-61ucosy166 b10+ 1296.3	1400 1401 140 140	150	6 5 5 5 5 5 5 5 5 5 5 5 5 5	<pre>b31++ 1635.1</pre>	1003.6 1028 y31++ 1603.6 1028 y31++ 1720.2	8'82/21 ++88 ⁶ 1800	1001	• • • • • • • • •

m/z = 942.94; charge = +4

PSM showing the identification of Galactosyl-HyK³⁶¹



m/z= 926.9843 charge=+6

PSM showing the identification of Galactosyl-HyK⁴⁶⁰



m/z= 881.4156; charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK⁴⁹⁷



m/z= 1229.9143; charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK⁵²⁷



m/z= 566.2649; charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK⁵⁷³

Human col4a1 chain

Ċ	612 (GPIG	DJK	+340	G		F P ₊	-16GG	P+	16 G	SP+16	GL	orgrb	۲ K ₊₃₄	^و G_E	P G K	640
5.5	5e+002	2									387.1						
	-1 -1 -1						2.0				718++						
	- 8						650	5 01 02 03	021.3		P.3						
							1 [K] +-	+ 728 55.2 801.2	++	88							
	8-						ctosy	*1 [K]- ++ 	osyı Lr	K] ++	9						
	2				22.5		/lgala	actos yl[K] syl[K]	act.	osyl [1000						
ity (%)					വ് +		lucos	sylgal Ad 1 lactos alacto	82 20 20	galact	[K] ++1						
Intens	9			-1	y] [K]		8-2*G 2	Sylgal sylgal	++ [X]	80sy11 927.2	<u>983.8</u> tosyl 514.3						
lative	8-			9 449	0 531.0	20 20 20 20 20 20 20 20 20 20 20 20 20 2	2,44 2,44	0.6000 841181 0.6000 2*6100	ctosyl	3000 3000 3000	alac salac tertosyl	04.2 32.2					
Re	8 -			[K]++ 461.(498. [] ¹ +- ¹	- 583 ++		769.0 769.0	lgalac	[K]+ ⁺ osy1[] 955.	aon aon acosyl acosyl 1027 1027 Antriv galac	+ + 11 11					
				+[X]+	[K] ++	([[K]]++ [X]++ [X]++ [X]+(V])		svlg? K + vsot	ucosy 3.7	tosyl alact [K]++	M]-G1(K]+H	891 [K] 81 [K]			8.4		
	8-			løalad ctosyl	tosyl alacto	ctosy] cosy][lacto tosyl	145.20 tosyl fealact	-18-61 + 87 87	galac galac 31.3 ¹ € .6syl	ctosy -2*G11	lactos lactos	σ		1+ 136		
	- 3			An n Increw Igalae	lgalac cosylg	salact salact salact	sol of sol of solver galac	0alac	unuoKy ■ v15+	+ 955	vigala Mchiñ Mchiñ	<mark>089.8</mark> sylga sylga]	1206		pl		
	-	01.1		AD-G1 ucosy	ucosy] 2*Gluc	555.1 lucosy icosyla fillicosy	icosyl Loosyl			والتي التي التي التي التي التي التي التي	lucosy Forth	Gluco Gluco	221.3				
	-9	3÷ ∎	L.	3-2*G1	10-GI	8++ *12-6 13-61u	7 <mark>18-2</mark> 15-61 15-61	6101- 613++		64 0 67+	V22-G	625-23	10+ 1				
	- - +	, p. , d. p										له به آباد ا		₩₩₩ ₩ <u>₩</u>	, n.t. <mark>1</mark>41. <u>t</u>. , n.t.	• • •,• ••	
		500	400		500	000	/00	000			n/z	1100	1200	1300	1400	1500	1000

m/z= 862.9186 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK⁶⁸¹

Human col4a1 chain

667	G F	FP ₊₁₆ GTP ₊₁₆ GRP ₊₁₆ GLP ₊₁₆ GEK ₊₃₄₀ GAVGQP ₊₁₆ GIGFP ₊₁₆ GPP ₊₁₆ GPK ₊₁₆ ⁶⁹⁸	
27	e+00		
2.1	Ę⊣		
	-		
	8-		
	- - 		
	2 - 2 - 2		
Intensity (%)		647.5 685.1 +++ 741.9 *y1[K]+++ 974.6	
Relative	- 20	را الال الحالي الحالي الحالي الحالي المرابع الحالي المرابع الحالي محالي الحالي الحالي الحالي محالي الحالي الحالي محالي الحالي الحالي محالي محالي الحالي محالي محالي الحالي محالي م محالي محالي محالي محالي محالي	
	- 4 - 4	Image: Second	
	- 3 - 3 - 3	92.6 92.6 10 ctosyl [] 10 cosyl [] 10 ctosyl [] 10 ctosyl	
	- 3	+ 349.1 + 349.1 + 349.1 76.1 77.1 77.1 76.1 77.	
	- 9	v8++ v8++ v8+++ v8++ v8+++ v8+++ v8++++ v8+++ v8++++ v8+++ v8++++ v8++++ v8++++ v8++++ v8++++ v8++++ v8++++ v8++++ v8+++++ v8+++++ v8+++++ v8+++++ v8+++++ v8+++++ v8++++++ v8+++++++ v8++++++++++++++++++++++++++++++++++++	
	•+		<u>P</u>
		m/z	



PSM showing the identification of Glucosylgalactosyl-HyK⁶⁹⁸

m/z= 1089.1052 charge=+5

55

m/z= 1001.8720 charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK⁷⁴²



PSM showing the identification of Glucosylgalactosyl-HyK⁷⁵⁷

Human col4a1 chain





Human col4a1 chain

m/z

PSM showing the identification of Glucosylgalactosyl-HyK¹⁰²⁵

Patron Induction Participants (Control of Control of Co	1.	8e+004.						37.51 37.51	+10				+ 170		+ 10	
Image: Description of the second of								8.37 7] [K] ++++ 9								
Addition Intermedy (5) Image: second statistic cose (10) ++++ 304.87 Image: second statistic cose (10) ++++ 304.87 Image: second statistic cose (10) ++++ 304.87 Image: second statistic cose (10) ++++ 304.87 Image: second statistic cose (10) ++++ 304.87 Image: second statistic cose (10) ++++ 304.87 Image: second statistic cose (10) ++++ 304.87 Image: second statistic cose (10) ++++ 700.40 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) ++ 783.44 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) ++ 783.44 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) ++ 783.44 Image: second statistic cose (10) +++ 700.40 Image: second statistic cose (10) ++ 100.40 Image: second statistic cose (10) +++ 1100.60 Image: second statistic cose (10) ++ 110.70 Image: second statistic cose (10) +++ 1100.60 Image: second statis cose (10) ++ 110.70 Image: second statistic c		- 8-						867.34 - y28+++ 91 ylgalactos3	23							
Relation linearly (A) ¹⁰		- - 2 - 1						y26+++ 908.67 b39-G1ucos]++++ 1007		.57					
Relation	Intensity (%)	8-					+ 738.43	+ 852.51 osvlfKl+++	rt 900.44 alactosyl[K	55 1068.48 +++ 1115.61	[K] +++ 1150					
2 20 20 20 40 2 2 2 2 4 3 4 2 4 4 4 2 2 4 4 4 5 4 2 4 4 4 4 6 4 4 4 4 4 4 4 4 6 4	Relative	- 2 <mark>2</mark> - 2						++^«*][K]++ ++ 860,12 0881galact(926.72	actosyılA] ⁺ -Glućosylga ++ 1038 59	2+++ 1055.P yl[K]++++ ctosyl[K]++	galactosyl + 1170.08	1277.92	+++ 1354.78			
0 10 20 30 0 10 20 30 0 10 20 30 0 10 20 30 0 10 20 30 0 10 20 30 0 10 20 30 0 10 20 31 0 10 10 10 10 0 10 10 10 10 10 0 10 10 10 10 10 10 0 10 10 10 10 10 10 10 0 10		- 1	- 374.87				+ 710.40 746.41 :tosyl[K]++ +++ 791.44	стаки] «1 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 1 2 1 2	1ucosylgal 유유 : 20 + 1008.212	+ 1069.40s 1086.11 110F 26 ucosylgala	b-đlưcosyl tosyl [K]+++	tosyl [K] ++	actosyl[K]+ 1386.58		1497.38	
0 10 20 0 10 10 <td< th=""><th></th><th>-33</th><th>sy][K]+++</th><th></th><th>g</th><th></th><th>10.40 10.40 748.51 ++++ ucosylgalac * 780 561++</th><th>84.71 woorleada woorleada woorleada woorleada woorleada woorleada</th><th>+ 900 022-0 </th><th>1046.45*** .osyl [K] +++ .osyl [K] +++ .osyl [K] +++ </th><th>ь ннн 115суз cosylgalac нн 1197.22</th><th>ıcosylgalac</th><th>lucosylgal(osyl[K]+++</th><th></th><th>ctosyl[K]+</th><th></th></td<>		-33	sy][K]+++		g		10.40 10.40 748.51 ++++ ucosylgalac * 780 561++	84.71 woorleada woorleada woorleada woorleada woorleada woorleada	+ 900 022-0 	1046.45*** .osyl [K] +++ .osyl [K] +++ .osyl [K] +++ 	ь ннн 115суз cosylgalac нн 1197.22	ıcosylgalac	lucosylgal(osyl[K]+++		ctosyl[K]+	
		20	osylgalacto		y10++ 537.8 558.09	608,36	678.31 711.70°=1°2 729.38 48.b30++++ 7.378.b26-61 781.v32+++	30.765+++ 8 + 871.08	+ 900.023+++ 082 <u>837+++</u> 102.15 11+++ 1022.6 1021.05	75 63 64 act	11.75.67-61. 11.75.67-61. 11.75.67-61. 12.15.43	ĤÎ261.44 1277 b29−61	328.56 1358.7143-6 :osylgalact	1439.78	ucosylgalad	
300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1800			b17-Gluc	1	y22++++ !	b17+++	b27++++ v21+++ b21+++ b22+++ 7 v15++ 76	v16++ 8		**************************************	464144 464144 4464144 4464144 442444 1442 1444 1444	b384_y12 y36+++	y27++ 1 b3ġ+î∓∵] b44-G1uc	P539++	b17-G1	
		0 11	300 400	500))	600	700 800	900	1000	1100	1200	130	0 140	0	1500	1600



PSM showing the identification of Galactosyl-HyK¹⁰⁴⁶

Human col4a1 chain







m/z= 659.3373 charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK¹⁰⁴⁹

ucosylgalactosyl[% 9.9e+002 ₽-8. ē 538,5 8 -Glucosylgalactosyl[K]++ 2 Relative Intensity (%) 8. 8,088 <mark>ନ୍ତ</mark> -655.5 ++ 667.1 bil -Glucosylgalactosyl[K]++ y9+ 907.2 595, 5, <mark>8</mark> 801.1 ++ 817.1 465.1 436.5 401.0 503.5 무. 929.2 b11-Glucosylgalactosyl[K] + 1076.3 872.2 bl4-Glucosylgalactosyl[K]+ 979.2 ucosylgalactosyl[K]+ 80 -Glucosylgalactosyl[K]++ 827 2 b10-Glucosylgalactosyl[K]++ 599,191 galactosyl[k]++ ucosylgalactosyl[K] -Glucosylgalactosyl[K]+ .0++ 489.6 ucosylgalactosyl[K]+ b10-Glucosylgalactosyl[K]+ 09-Glucosylgalactosyl[K]+ 8b17++ 352.1 8-É 757.2 b8+ 1125.2 Ο b10+ 1253.2 276.1 무-0 -600 500 900 1000 1100 300 400 700 1200 200 800 1300

m/z

62

Human col4a1 chain

b11+ 1400.4

1400

PSM showing the identification of Galactosyl-HyK¹⁰⁶⁶

Human col4a1 chain

m/z= 1085.5333 charge=+4

PSM showing the identification of Galactosyl-HyK¹¹¹⁷

m/z= 909.2668 charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK¹¹³²

m/z= 464.8250 charge=+5

PSM showing the identification of Galactosyl-HyK¹¹⁸⁸

m/z= 913.2631 charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK^{1185,1188}

Human col4a1 chain

¹¹⁸³ G ^C D	K ₊₃₄₀ G S ⁽ K ₊	- ₃₄₀ G E	VGFP	+16 G L		P ₊₁₆ G	P ₊₁₆ (GSK	GEQ	G F M	GPP	GPQ	GQPG	L ¹²²³
9.	0e+003 톁ျ	///]++ 1254 ₃				
						+ 804,48				b20-NH3				
	87 - - - -					syl [K] +++			.04.05					
	8-					<i>y</i> lgalacto:			·y35+++ 11					
	- - -					-2*61 ucos) 398. 45			- 1097 89					
nsity (%)	- - - - - - -		+++ 507.83			9+++ 804, 4y37)+++ 842, 11 K] +++ 881, 85 K] +++ 881, 85	927.45 ++ 970.73		197 - v35-MH3+++					
Relative Inte	- 		lactosyl[K]+	120+++ 636,29 63.50	7.13	bl b20 b21 b20		8	.82 8 35- H20+++ 10	210.71			_	
	- 4 - -		-G1 ucosyl ge]++ 62 <mark>y20-</mark>] syl[K]+++ E	(*** 771.51	++ 805.43	actôsýl[K]. actôsýl[K]. ** orr r <u>a</u> *"Glucosylg	46. 1022.49	40+++ 1078 05.25 8.06 .79	yl[K]+++ 1; + 1248 10	310.19 .85		97 · 0041 ++1	
	- - -]+++ 403.33	3+++ 507. (b18	.18 galactosyl[K cosylgalactos	galactosyl[K]	actosyl[K]++ ++++ 823.56]+++ 823.56]+++ 859.49 ⁵ ara ara ara ara ara ara ara ara ara	11++++ 942.57 ++++ 942.57 ++++ 942.57 ++++ 942.57 +++ 940.57	A 1026 1036	HI3+++ 1073- + 1*y23++ 117 w1[K]+++ 114 1[K]+++ 1162	osylgafactos . 1 <u>eul (M1++</u> 1262,3359.91	, b21-NH3++ 1; 1[K]+++ 1355		galactosyll.	
	- 3	galactosyl[K	+ 507.(b9-ME	b6-NH3++ 619 [b10-61 ucosyl 3. Eb21 - 2*61 u	6.29 720.10 121-61 úcosyl: 121-61 úcosyl:	-++ 806.26gal	المالية المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحم المحمد المحمد المحمد المحم محمد المحمد المحم المحمد المحمد المحمد المحمد المحم المحمد المحم	81 1036. y33-	saractory34-M alactacy121 27,77 27,77 281 galactos sylgalactosy. .65	لمبادر المراجع ا 1253 - 1253 - 1254 - 12533 - 12533 - 12533 - 12533 - 12533 - 12533 - 12533 - 12533 -	1-H20++ 1310 svlgalactosv		10-2*41 acos	
		■b12-2*G1ucosy	^{−++02H-50++}	10,12,11,2,11,2,13,13,13,13,13,13,13,13,13,13,13,13,13,	930++++ 70 930+++ 737.38 923+++ 737.38 922++++ 757.3	b27+		v21.470.41.100 b26-#204.1.100 v33+++ 1041.	1,000 01 000001 0 1,000 01 000001 0 1,000 01100 0110 1,000 01100 01100 1,000 01100 01100 1,000 01100 01100	- b6-H20+ 1235.4		- b23++ 1390.98	•	
	₀ <mark>╡╷╶<mark>┑</mark>╸┡ŧ╊╪╸┹╤╍╌ᡒ┉╇ 300</mark>	арій (,	500 e	41.4,149.4,141 000						1200	1300	1400	j. s. ya. is pini p. 149. a t. i n a 1500	Ļ
								m/z	2					

PSM showing the identification of Glucosylgalactosyl-HyK¹²⁶⁵

Human col4a1 chain

m/z= 1065.5275 charge=+3

m/z= 1688.9413 charge=+5

PSM showing the identification of Galactosyl-HyK¹³⁰⁴

m/z = 992.8703 charge = +5 PSM showing the identification of Glucosylgalactosyl-HyK^{1304,1328} and Galactosyl-HyK¹³¹⁹ Human col4a1 chain

m/z= 575.8043 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK¹³⁴⁰

Human col4a1 chain



72

m/z= 957.6498(+5) PSM showing the identification of HyP⁷¹ and 3-HyP⁷⁴ Human col4a1 chain 52 G L P₊₁₆ G L Q G V I G F P₊₁₆ G M Q G P E G P₊₁₆ Q G P₊₁₆ P₊₁₆ G Q K G D T G E P₊₁₆ G L P G T K₊₃₄₀ G T R G P P G⁹⁷



m/z= 957.63(+6)

PSM showing the identification of 3-HyP⁷⁴





m/z= 1200.8142 (+4)

PSM showing the identification of 3-HyP⁹⁵

$${}^{52}\text{GLP}_{+16}\text{GL Q GV I G F PG M}_{+16} \text{Q GP}_{+16}\text{E G PQGP}_{+16}\text{PGQK}_{+340} \text{GDT GE P}_{+16} \text{GL P}_{+16} \text{G TKGTRGP}_{+16}\text{P}_{+16}\text{G}^{97}$$



PSM showing the identification of 3-HyP³⁷²





m/z= 889.1826(+4)

PSM showing the identification of 3-HyP⁴²¹



m/z= 886.748 (+3)

PSM showing the identification of 3-HyP⁴⁷⁸



PSM showing the identification of 3-HyP^{478,484}

$$^{477}GP_{+16}P_{+16}GPQGP_{+16}P_{+16}GEIGFP_{+16}GQP_{+16}GAK_{+340}GDR^{500}$$



m/z= 922.687 (+4)

PSM showing the identification of HyP⁵¹⁴



m/z

PSM showing the identification of 3-HyP⁵⁸⁷

Human col4a1 chain



81

PSM showing the identification of 3-HyP⁶⁰²

⁵⁹⁸GDTG P₊₁₆P₊₁₆G P P₊₁₆G Y G P A G P I G D K⁶¹⁷



⁵⁹⁸G D T G P_{+16} P_{+16} G P P_{+16} G Y G P A G P I G D K⁶¹⁷



PSM showing the identification of 3-HyP⁶⁴⁷



m/z= 849.429 (+3)

PSM showing the identification of HyP⁶⁶²

Human col4a1 chain

⁶⁴¹IVPLP₊₁₆GPP₊₁₆GAEGLP₊₁₆GSP₊₁₆GFPGP₊₁₆QGDR⁶⁶⁶



PSM showing the identification of 3-HyP^{786,803}

 $^{782}GEP_{+16}GP_{+16}P_{+16}GLP_{+16}GSVGSP_{+16}GVP_{+16}GIGP_{+16}P_{+16}GAR^{807}$



m/z= 782.3984 (+3)

PSM showing the identification of 3-HyP¹³⁶⁷



m/z= 782.8649 (+4)

PSM showing the identification of 3-HyP^{1424,1436}





Supporting figure S4: Comparison of 3-HyP sites identified in col4a1 from mouse, human and bovine. The horizontal box represents the full length col4a1 sequence and the vertical red lines indicate the corresponding 3-HyP sites. The information for 3 HyP sites from bovine lens capsule (C) were summarized from Pokidysheva et al³ and Hudson et al⁴. The 3-HyP sites marked with green represents the conserved sites among mouse, human and bovine.

1	MFSFVDLRLL	LLLGATALLT	HGQEDIPEVS	CIHNGLRVPN	GETWKPEVCL	ICICHNGTAV
61	CDDVQCNEEL	DCPNPQRREG	ECCAFCPEEY	VSPNSEDVGV	EGPKGDPGPQ	GPRGPVGPPG
121	RDGIPGQPGL	PGPPGPPGPP	GPPGLGG NFA	SQMSYGYDEK	SAGVSVPGPM	GPSGPRGLPG
181	PPGAPGPQGF	QGPPGEPGEP	GGSGPMGPRG	PPGPPGKNGD	DGEAGKPGRP	GERGPPGPQG
241	ARGLPGTAGL	1 PGM <mark>K</mark> GHRGFS	Z GLDGA <mark>K</mark> GDAG	3 PAGPKGEPGS	PGENGAPGQM	GPRGLPGERG
301	RPGPPGTAGA	RGNDGAVGAA	GPPGPTGPTG	PPGFPGAVGA	K GEAGPQGAR	GSEGPQGVRG
361	EPGPPGPAGA	AGPAGNPGAD	5 GQPGAKGANG	APGIAGAPGF	PGARGPSGPQ	6 GPSGPPGPKG
421	NSGEPGAPGN	7 8 KGDTGAKGEP	GATGVQGPPG	PAGEEGKRGA	RGEPGPSGLP	GPPGERGGPG
481	SRGFPGADGV	9 AGPKGPSGER	10 Gapgpagpkg	SPGEAGRPGE	11 AGLPGAKGLT	GSPGSPGPDG
541	KTGPPGPAGQ	DGRPGPAGPP	GARGQAGVMG	FPGPKGTAGE	PGKAGERGLP	GPPGAVGPAG
601	KDGEAGAQGA	PGPAGPAGER	GEQGPAGSPG	FQGLPGPAGP	PGEAGKPGEQ	GVPGDLGAPG
661	PSGARGERGF	PGERGVQGPP	GPAGPRGNNG	12 APGNDGAKGD	TGAPGAPGSQ	GAPGLQGMPG
721	ERGAAGLPGP	13 14 KGDRGDAGPK	GADGSPGKDG	ARGLTGPIGP	15 PGPAGAPGD <mark>K</mark>	GEAGPSGPPG
781	PTGARGAPGD	RGEAGPPGPA	GFAGPPGADG	16 QPGAKGEPGD	17 TGVKGDAGPP	GPAGPAGPPG
841	PIGNVGAPGP	18 KGPRGAAGPP	GATGFPGAAG	RVG <mark>P</mark> PGPSGN	AGPPGPPGPV	19 GKEGGKGPRG
901	ETGPAGRPGE	VGPPGPPGPA	20 GEKGSPGADG	PAGSPGTPGP	QGIAGQRGVV	GLPGQRGERG
961	FPGLPGPSGE	PGKQGPSGSS	GERGPPGPMG	PPGLAGPPGE	SGREGSPGAE	GSPGRDGAPG
1021	21 AKGDRGETGP	AGPPGAPGAP	GAPGPVGPAG	KNGDRGETGP	AGPAGPIGPA	GARGPAGPQG
1081	22 PRGDKGETGE	23 QGDRGIKGHR	GFSGLQGPPG	SPGSPGEQGP	SGASGPAGPR	GPPGSAGSPG
1141	KDGLNGLPGP	IG P PGPRGRT	GDSGPAG <mark>P</mark> PG	PPGPPGPPGP	PSGGYDFSFL	PQPPQEKSQD
1201	GGRYYRADDA	NVVRDRDLEV	DTTLKSLSQQ	IENIRSPEGS	RKNPARTCRD	LKMCHSDWKS
1261	GEYWIDPNQG	CNLDAIKVYC	NMETGQTCVF	PTQPSVPQKN	WYISPNPKEK	KHVWFGESMT
1321	DGFPFEYGSE	GSDPADVAIQ	LTFLRLMSTE	ASQNITYHCK	NSVAYMDQQT	GNLKKALLLQ
1381	GSNEIELRGE	GNSRFTYSTL	VDGCTSHTGT	WGKTVIEYKT	TKTSRLPIID	VAPLDIGAPD
1441	QEFGLDIGPA					

Supporting figure S5: Summary of prolyl 3-hydroxylation and glycosylation sites in mouse col1a1⁴⁻⁷. The matured col1a1 sequences are shown in black. The 23 potential lysine sites for O-glycosylation are numbered. Bold '**K**' indicates O-glycosylated residue in the Yaa position of Gly-Xaa-Yaa motif. Bold '**P**' indicates 3 hydroxyproline in the Xaa position of Gly-Xaa-HyP.

References:

- Sun, R.X.; Dong, M.Q.; Song, C.Q.; Chi, H.; Yang, B.; Xiu, L.Y.; Tao, L.; Jing, Z.Y.; Liu, C.; Wang, L.H.; Fu, Y.; He, S.M. Improved peptide identification for proteomic analysis based on comprehensive characterization of electron transfer dissociation spectra. J Proteome Res 2010, 9, 6354–6367.
- Ma, Z. Q.; Chambers, M. C.; Ham, A.J.; Cheek, K.L.; Whitewell, C.W.; Aerni, H.R.; Schilling, B.; Miller, A.W.; Caprioli, R.M.; Tabb, D. L. ScanRanker: Quality assessment of tandem mass spectra via sequence tagging. J Proteome Res 2011, 10, 2896-2904.
- Pokidysheva, E.; Boudko, S.; Vranka, J.; Zientek, K.; Maddox, K.; Moser, M.; Fassler, R.; Ware, J.; Bachinger, H. P.: Biological role of prolyl 3-hydroxylation in type IV collagen. Proc Natl Acad Sci U S A 2014, 111, 161-6.
- 4. Hudson, D. M.; Joeng, K. S.; Werther, R.; Rajagopal, A.; Weis, M.; Lee, B. H.; Eyre, D. R. Posttranslationally abnormal collagens of prolyl 3-hydroxylase-2 null mice offer a pathobiological mechanism for the high myopia linked to human LEPREL1 mutations. J Biol Chem 2015, 290, 8613-8622.
- Pokidysheva, E.; Zientek, K. D.; Ishikawa, Y.; Mizuno, K.; Vranka, J. A.; Montgomery, N. T.; Keene, D. R.; Kawaguchi, T.; Okuyama, K.; Bachinger, H. P. Posttranslational modifications in type I collagen from different tissues extracted from wild type and prolyl 3-hydroxylase 1 null mice. J Biol Chem 2013, 288, 24742-24752.
- Cabral, W. A.; Perdivara, I.; Weis, M.; Terajima, M.; Blissett, A. R.; Chang, W.; Perosky, J. E.; Makareeva, E. N.; Mertz, E. L.; Leikin, S.; Tomer, K. B.; Kozloff, K. M.; Eyre, D. R.; Yamauchi, M.; Marini, J. C.: Abnormal type I collagen post-translational modification and crosslinking in a cyclophilin B KO mouse model of recessive osteogenesis imperfecta. PLoS genetics 2014, 10, e1004465.
- 7. Perdivara, I.; Yamauchi, M.; Tomer, K. B. Molecular Characterization of Collagen Hydroxylysine -Glycosylation by Mass Spectrometry: Current Status. Aust J Chem 2013, 66, 760-769.