Mapping Sites of O-Glycosylation and Fringe Elongation on Drosophila Notch

Beth M. Harvey^{#1}, Nadia A. Rana^{#1}, Hillary Moss¹, Jessica Leonardi², Hamed Jafar-Nejad², Robert S. Haltiwanger*¹

From the ¹Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, New York 11794-5215, the ²Program in Developmental Biology, Baylor College of Medicine, Houston, TX 77030, and the ³Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030

*Both authors contributed equally to this work.

*Current address: Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602-

Running title: Mapping O-Glycosylation on Drosophila Notch

To whom correspondence should be addressed: Robert S. Haltiwanger, Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA, Tel.: (706) 542-4151; E-mail: rhalti@uga.edu

Keywords: Notch receptor, Fringe, EGF repeat, O-fucose, O-glucose, O-GlcNAc, mass spectrometry

Supplemental Figures

Supplemental Figure 1. *O*-Fucosylation sites are elongated by Fringe to varying degrees.

MRM analyses showing the relative levels of unmodified (black line), *O*-fucose monosaccharide (red line) and *O*-fucose disaccharide (blue line) of several peptides containing predicted *O*-fucose sites in the absence of Fringe (*A*'-*H*'). List of peptides searched in each EIC are in Table 1, and corresponding spectra are in Fig. S3. *A*, *A*' (EGF3); *B*, *B*' (EGF5); *C*, *C*' (EGF7); *D*, *D*' (EGF8); *E*, *E*' (EGF9); *F*, *F*' (EGF21); *G*, *G*' (EGF28); *H*, *H*' (EGF31). Red triangle represents fucose (dHexose). Blue square is GlcNAc (HexNAc).

Supplemental Figure 2. Sequence alignments of *O*-glycosylated EGF repeats

Alignments of the amino acid sequences within and surrounding the current consensus sequences of O-glycosylation of EGF repeats containing O-fucose (A-B), O-glucose (C) and O-GlcNAc (D) sites shown using WebLogo (1). Sequences of the 22 EGF repeats containing O-fucose sites were organized into either four categories of approximate amounts of O-fucosylation based on EICs generated in the absence of Fringe (A) or four categories of fucose elongation from EICs generated in the presence of Fringe (B) (see Fig. 4). C, Sequences of the 18 EGF repeats containing O-glucose sites were organized into three categories of approximate amounts of glucose elongation based on EICs (see Fig. 5). D, Sequences of the 18 EGF repeats containing O-GlcNAc sites were organized into two categories of either having or not having apparent O-GlcNAc modification based on EICs (see Fig. 6).

Supplemental Figure 3. Mass spectra of O-glycosylated peptides from Drosophila Notch

Drosophila Notch was purified from S2 cells and prepared for mass spectral analyses. For each peptide, top panels show MS spectra at a specific retention time, and red diamonds indicate ions chosen for fragmentation. Parent ions corresponding to specified peptides are fragmented to produce MS2 spectra, shown in the bottom panels. Parent ions of the represented peptide are labeled with the charge state of the peptide in that specific spectrum. Other ions in the MS spectra are from co-eluting material. Blue diamonds in MS2 spectra indicate the position of the parent ion. Spectra representing the most glycosylated form of each peptide found in the MS/MS data are shown. For analyzing O-fucosylation, S2 cells were co-transfected with or without Fringe, and peptides containing an O-fucose consensus sequence were inspected in both conditions (-Fringe or +Fringe). Spectra analyzing O-glucosylation and O-

GlcNAcylation were taken from Notch generated in the absence of Fringe, except the EGF26 O-GlcNAc site. Tables to the right of spectra show the ions searched in EICs analyzing O-fucose, O-glucose and O-GlcNAc glycoforms in Figs. 3, 4A, 5A, 6A and Fig. 8 and MRM in Figs. 3 and S1. Red triangle represents fucose (dHexose). Blue circle is glucose (Hexose). Orange star is xylose (Pentose). Blue square is GlcNAc (HexNAc).

References

1. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190

Fig. S1

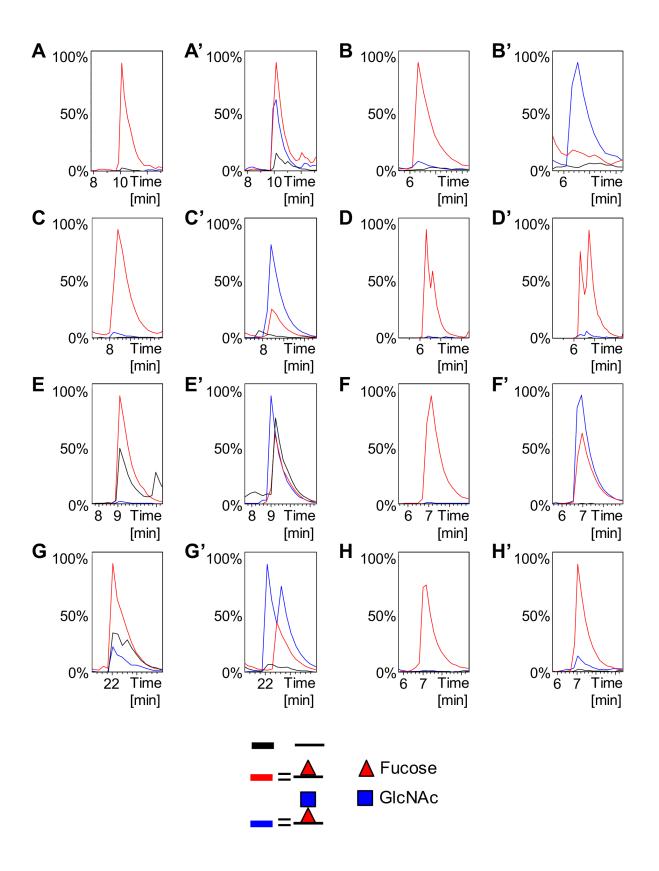


Fig. S2, A

O-Fucose Modification	EGF Repeat	EGF Sequence	Alignment	
<30%	13	DIDECQ <u>S</u> NPCLNDG <u>T</u> CHDKINGFKCSCALGF <u>T</u> GARCQ	4-3-2SIPCINGTCHD	
30%-50%	2	${\tt HRNPCNSMRCQNGGTCQVTFRNGRPGISCKCPLGFDESLCE}$	NEMROLNICOS I DGFNGGN	
3070 0070	26	ndedctessclngg <mark>s</mark> cidgingyncsclagy <u>s</u> gancq	TISSSVQIIWWTVQV	
	3	VPNACDHVTCLNGG <u>T</u> CQLKTLEEYTCACANGY <u>T</u> GERCE		
	4	$ ext{TKNLCASSPCRNGATCTALAGSSSFTCSCPPGFTGDTCS}$		
	8	DVDECAQRDHPVCQNGA <mark>T</mark> CTNTHGSYSCICVNGWAGLDCS		
60-80%	9	$ ext{NTDDCKQAACFYGA}_{ ext{T}} ext{CIDGVGSFYCQCTKGK}_{ ext{T}} ext{GLLCH}$		
	12	$ ext{NINECE}_{\underline{S}}$ HPCQNEG $_{\underline{S}}$ CLDDPGTFRCVCMPGF $_{\underline{T}}$ GTQCE	OF STANDERS AND	
	27	$KLNKCD_{\underline{S}}NPCLNGA_{\underline{T}}CHEQNNEYTCHCPSGF_{\underline{T}}GKQCS$	weldingo lamestry volo	
	32	NKDDCKPGACHNNG <mark>S</mark> CIDRVGGFECVCQPGFVGARCE		
	1	VAASCTSVGCQNGG <mark>T</mark> CVTQLNGKTYCACDSHYVGDYCE		
	5	DIEECQ <u>S</u> NPCKYG <mark>G</mark> TCVNTHGSYQCMCPTGY <u>T</u> GKDCD		
	7	${\tt NYDDCLGHLCQNGG\underline{T}CIDGISDYTCRCPPNF\underline{T}GRFCQ}$		
	17	NVNECH <u>S</u> NPCNNGA <u>T</u> CIDGINSYKCQCVPGF <u>T</u> GQHCE		
≥80%	20	${\tt DIDECS_SNPCQHGG_TCYDKLNAFSCQCMPGY_TGQKCE}$		
	21	${\tt NIDDCVTNPCGNGG\underline{T}CIDKVNGYKCVCKVPFTGRDCE}$		
	23	DIDECSLSSPCRNGA <mark>S</mark> CLNVPGSYRCLCTKGYEGRDCA	I SE	
	24	NTDDCASFPCQNGGTCLDGIGDYSCLCVDGFDGKHCE	wetiliga berneley adu	
	25	$ ext{DINECL}_{ extstyle ext{Q}PCQNGA}_{ ext{T}CSQYVNSYTCTCPLGF}_{ ext{S}GINCQ}$		
	28	YVDWCGQSPCENGA <u>T</u> CSQMKHQFSCKCSAGW <u>T</u> GKLCD		
	30	${\tt EIDECQSQPCQNGGTCRDLIGAYECQCRQGFQGQNCE}$		
	31	NIDDCAPNPCQNGGTCHDRVMNFSCSCPPGTMGIICE		

Fig. S2, B

O-Fucose Elongation	EGF Repeat	EGF Sequence	Alignment
<25%	2	HRNPCNSMRCQNGG <mark>T</mark> CQVTFRNGRPGISCKCPLGFDESLCE	
	13	DIDECQ <u>S</u> NPCLNDG <mark>T</mark> CHDKINGFKCSCALGF <u>T</u> GARCQ	
	26	NDEDCTESSCLNGG <mark>S</mark> CIDGINGYNCSCLAGY <u>S</u> GANCQ	
	31	NIDDCAPNPCQNGG <mark>T</mark> CHDRVMNFSCSCPPGTMGIICE	
	32	NKDDCKPGACHNNG <mark>S</mark> CIDRVGGFECVCQPGFVGARCE	webligs behalp allo
	3	VPNACDHVTCLNGG <mark>T</mark> CQLKTLEEYTCACANGY <u>T</u> GERCE	
	4	$ ext{TKNLCASSPCRNGA}_{ ext{T}} ext{CTALAGSSSFTCSCPPGF}_{ ext{T}} ext{GDTCS}$	4-
25%-50%	9	NTDDCKQAACFYGA $_{ ext{T}}$ CIDGVGSFYCQCTKGK $_{ ext{T}}$ GLLCH	I #2 C.DC. NGATC.Do GSEV
	12	NINECESHPCQNEGSCLDDPGTFRCVCMPGFTGTQCE	
	17	nvnech <u>s</u> npcnnga <u>t</u> cidginsykcocvpgf <u>t</u> goʻhce	N Anticological particular Anticological Ant
	27	KLNKCD <u>S</u> NPCLNGA <u>T</u> CHEQNNEYTCHCPSGF <u>T</u> GKQCS	
	8	DVDECAQRDHPVCQNGA <mark>T</mark> CTNTHGSYSCICVNGWAGLDCS	
	20	DIDECS <u>S</u> NPCQHGG <mark>T</mark> CYDKLNAFSCQCMPGY <u>T</u> GQKCE	4
50%-75%	24	NTDDCASFPCQNGG <mark>T</mark> CLDGIGDYSCLCVDGFDGKHCE	TO GAYS
3070-7370	25	DINECL <u>S</u> QPCQNGA <mark>T</mark> CSQYVNSYTCTCPLGF <u>S</u> GINCQ	
	28	YVDWCGQSPCENGA <mark>T</mark> CSQMKHQFSCKCSAGW <u>T</u> GKLCD	N
	30	EIDECQ <u>S</u> QPCQNGG <mark>T</mark> CRDLIGAYECQCRQGFQGQNCE	
_	1	VAASCTSVGCQNGG <mark>T</mark> CVTQLNGKTYCACDSHYVGDYCE	
≥75%	5	DIEECQ <u>S</u> NPCKYGG <u>T</u> CVNTHGSYQCMCPTGY <u>T</u> GKDCD	
	7	$ ext{NYDDCLGHLCQNGG}_{ ext{T}}^{ ext{CIDGISDYTCRCPPNF}}_{ ext{T}}^{ ext{GR}}$] #2 LSNPI, gNIUII, JR EPRGYT
	21	NIDDCVTNPCGNGG <u>T</u> CIDKVNGYKCVCKVPFTGRDCE	Ĭ ĬŎŹŖĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸ
	23	DIDECSLSSPCRNGA <mark>S</mark> CLNVPGSYRCLCTKGYEGR DCA	weldings benefity with

Fig. S2, C

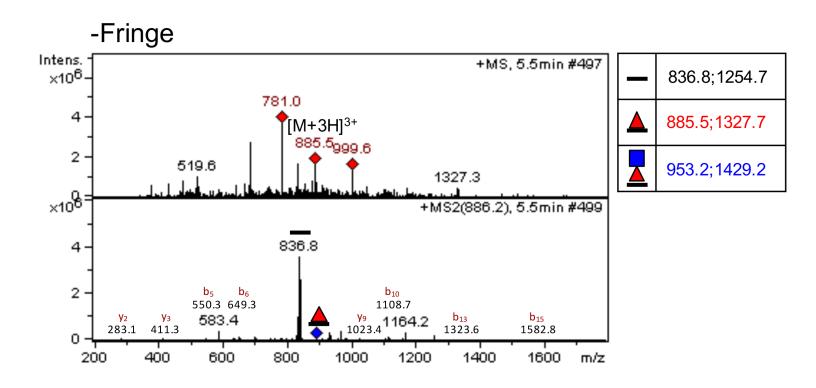
O-Glucose Elongation	EGF Repeat	EGF Sequence	Alignment
	4	TKNLCA <u>S</u> SPCRNGA <u>T</u> CTALAGSSSFTCSCPPGFTGDTCS	
	5	DIEECQ <u>S</u> NPCKYGG <u>T</u> CVNTHGSYQCMCPTGYTGKDCD	
	14	NIDDCQSQPCRNRGICHDSIAGYSCECPPGYTGTSCE	ELPE Q SNDC RIGGT
<u>≤</u> 10%	24	NTDDCASFPCQNGGTCLDGIGDYSCLCVDGFDGKHCE	
	27	KLNKCD <u>S</u> NPCLNGA <u>T</u> CHEQNNEYTCHCPSGFTGKQCS	ĬŢĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸ
	30	EIDECQ <u>S</u> QPCQNGG <u>T</u> CRDLIGAYECQCRQGFQGQNCE	webloga behviley adu
	33	DINECL <u>S</u> NPCSNAGTLDCVQLVNNYHCNCRPGHMGRHCE	
	35	SGQDCDSNPCRVGNCVVADEGFGYRCECPRGTLGEHCE	
	10	LDDACT <u>S</u> NPCHADAICDTSPINGSYACSCATGYKGVDCS	
	12	NINECE <u>S</u> HPCQNEG <u>S</u> CLDDPGTFRCVCMPGFTGTQCE	
10—50%	13	DIDECQ <u>S</u> NPCLNDG <u>T</u> CHDKINGFKCSCALGF <u>T</u> GARCQ	
	16	QINECE <u>S</u> NPCQFDGHCQDRVGSYYCQCQAGTSGKNCE	
	17	nvnech <u>s</u> npcnnga <u>t</u> cidginsykcocvpgftgohce	webiogo behalley edu
	25	DINECL <u>S</u> QPCQNGA <u>T</u> CSQYVNSYTCTCPLGFSGINCQ	
	15	NINDCDSNPCHRGKCIDDVNSFKCLCDPGYTGYICQ	4
>50%	18	NVDECI <u>S</u> SPCANNGVCIDQVNGYKCECPRGFYDAHCL	#2-DIDFI'ASNUI'ANGGC
/30%	19	DVDECASNPCVNEGRCEDGINEFICHCPPGYTGKRCE	
	20	DIDECS <u>S</u> NPCQHGG <u>T</u> CYDKLNAFSCQCMPGYTGQKCE	N C wespells have a control of the c

Fig. S2, D

O-GlcNAc Modification	EGF Repeat	EGF Sequence	Alignment	
	3	VPNACDHVTCLNGG <u>T</u> CQLKTLEEYTCACANGY <u>T</u> GERCE		
	4	TKNLCA <u>S</u> SPCRNGA <u>T</u> CTALAGSSSFTCSCPPGF <u>T</u> GDTCS		
	5	DIEECQ <u>S</u> NPCKYGG <u>T</u> CVNTHGSYQCMCPTGY <u>T</u> GKDCD		
	9	NTDDCKQAACFYGA <mark>T</mark> CIDGVGSFYCQCTKGK <u>T</u> GLLCH		
	13	DIDECQ <u>S</u> NPCLNDG <u>T</u> CHDKINGFKCSCALGF <u>T</u> GARCQ		
No	15	NINDCDSNPCHRGKCIDDVNSFKCLCDPGYTGYICQ		
	16	QINECE <u>S</u> NPCQFDGHCQDRVGSYYCQCQAGT <u>S</u> GKNCE		
	17	nvnech <u>s</u> npcnnga <mark>t</mark> cidginsykcocvpgftgohce		
	19	DVDECASNPCVNEGRCEDGINEFICHCPPGYTGKRCE	weblogs behality ads	
	22	KMDPCASNRCKNEAKCTPSSNFLDFSCTCKLGY <u>T</u> GRYCD		
	25	DINECL <u>S</u> QPCQNGA <mark>T</mark> CSQYVNSYTCTCPLGF <u>S</u> GINCQ		
	27	KLNKCD <u>S</u> NPCLNGA <mark>T</mark> CHEQNNEYTCHCPSGF <u>T</u> GKQCS		
	28	YVDWCGQSPCENGA <mark>T</mark> CSQMKHQFSCKCSAGW <u>T</u> GKLCD		
	4	TKNLCASSPCRNGA <u>T</u> CTALAGSSSFTCSCPPGF <u>T</u> GDTCS		
Yes	11	DIDECDQGSPCEHNGICVNTPGSYRCNCSQGFTGPRCE		
	12	NINECESHPCQNEGSCLDDPGTFRCVCMPGFTGTQCE		
	14	NIDDCQ <u>S</u> QPCRNRGICHDSIAGYSCECPPGY <u>T</u> GTSCE		
	20	DIDECS <u>S</u> NPCQHGG <u>T</u> CYDKLNAFSCQCMPGY <u>T</u> GQKCE	webbys beforky ods	

Fig. S3, A

EGF 1 N61SCTSVGCQNGGTCVTQLNGKTY82



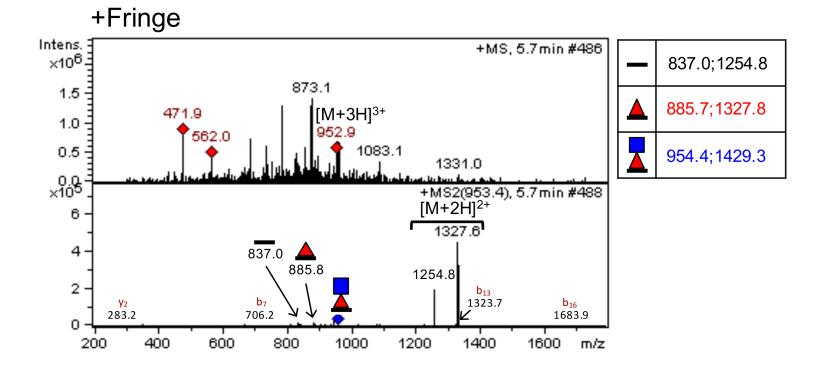
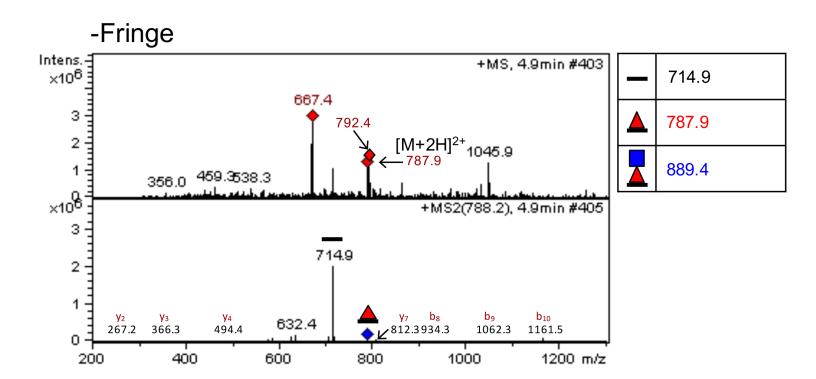


Fig. S3, B

EGF 2 104RCQNGGTCQVTF115



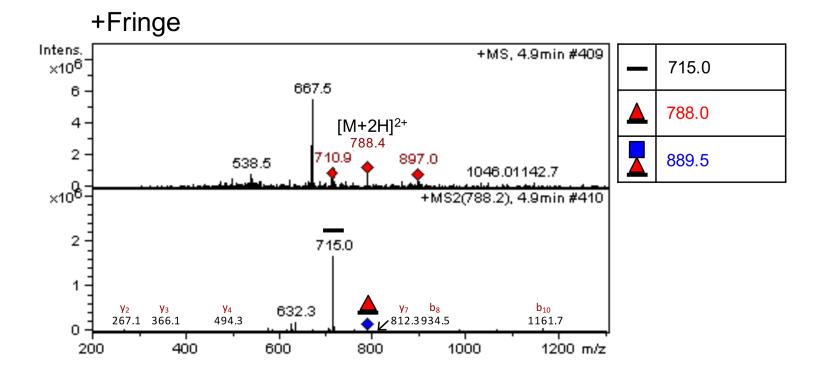


Fig. S3, C

EGF 3 126 CPLGFDESLCEIAVPNACDHVTCLNGGTCQLK157

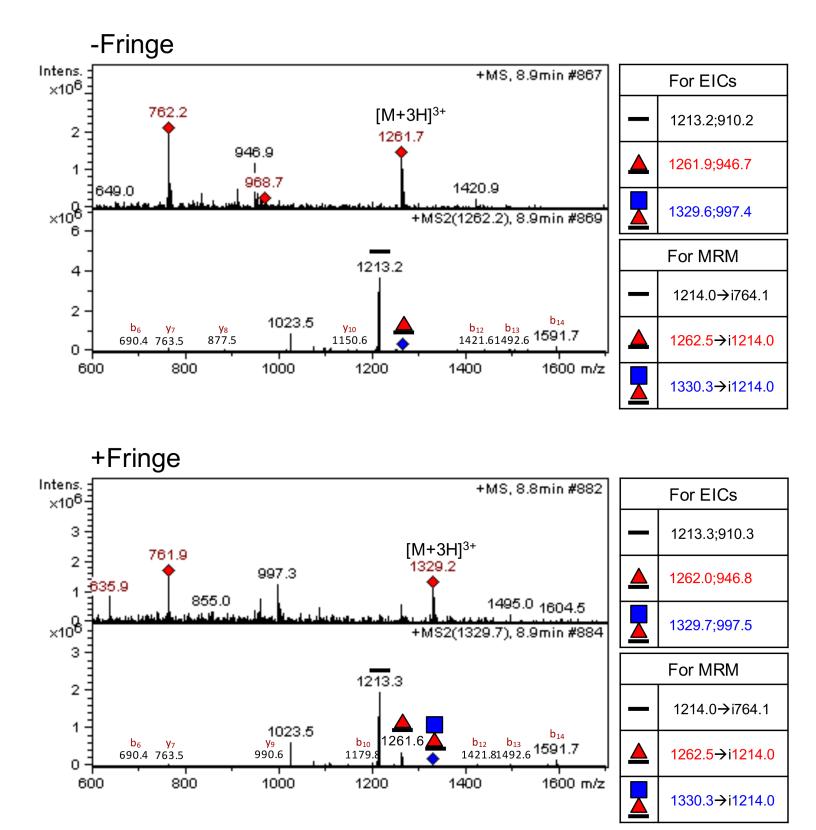


Fig. S3, D EGF 3

¹⁵⁸TLEEYTCACANGYTGER¹⁷⁴

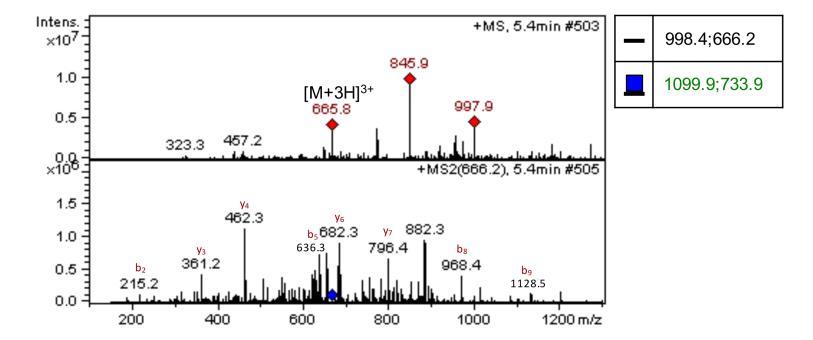


Fig. S3, E

¹⁷⁹NLCASSPCR¹⁸⁷

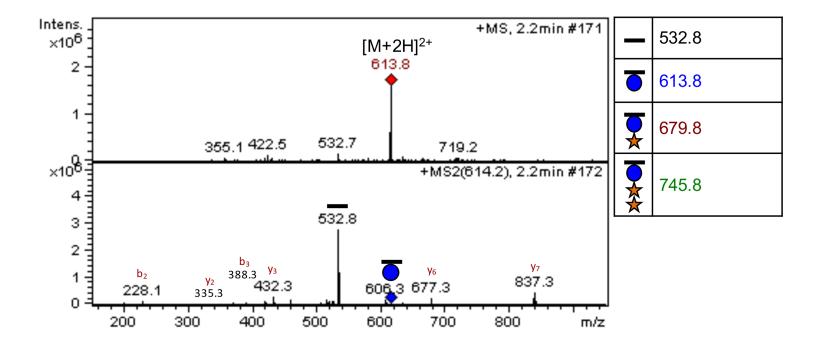
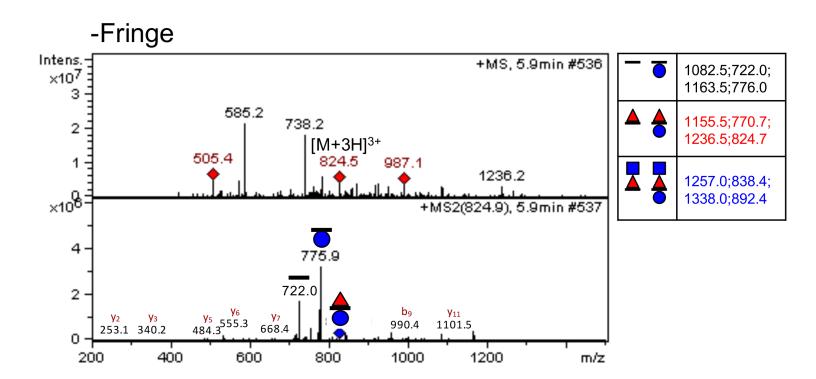


Fig. S3, F

¹⁸¹CASSPCRNGATCTALAGSSSF²⁰¹



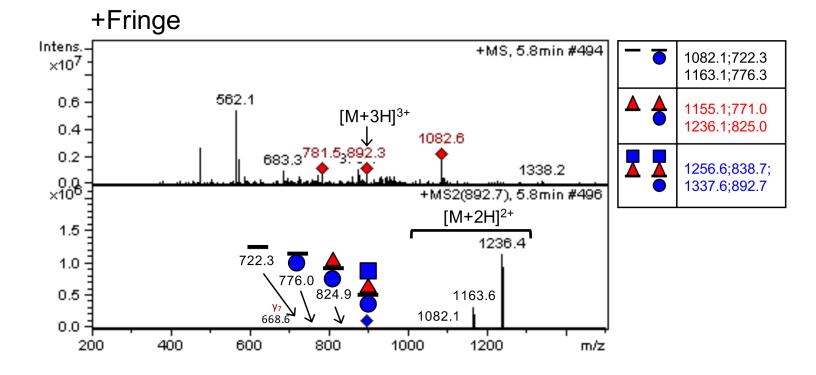


Fig. S3, G

EGF 4

202TCSCPPGFTGDTCSY216

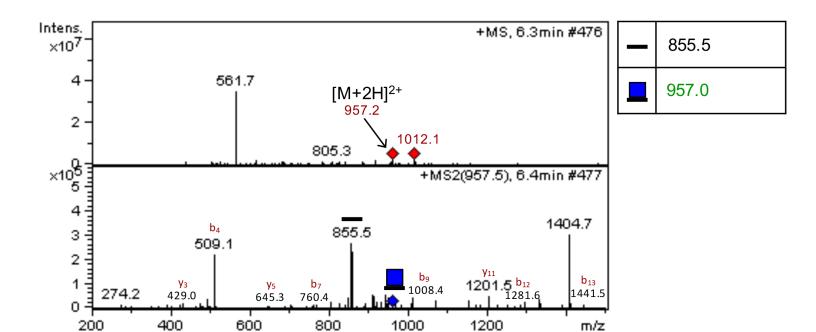


Fig. S3, H

EGF 5 217DIEECQSNPCKY²²⁸

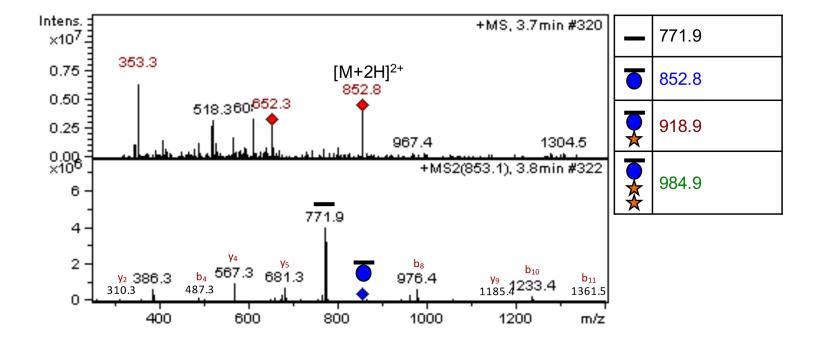
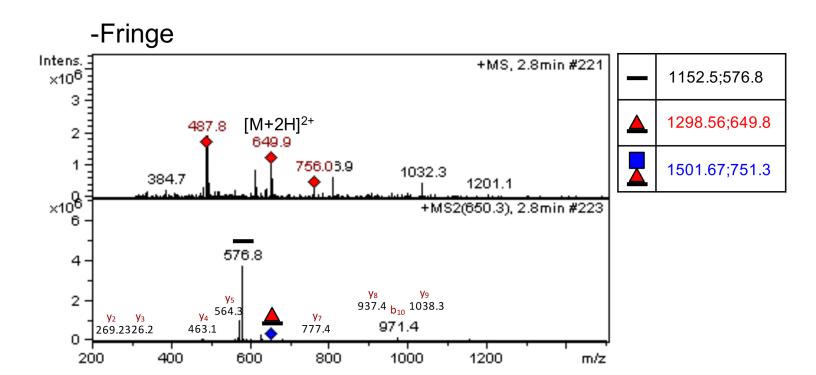


Fig. S3, I

EGF 5

²²⁹GGTCVNTHGSY²³⁹



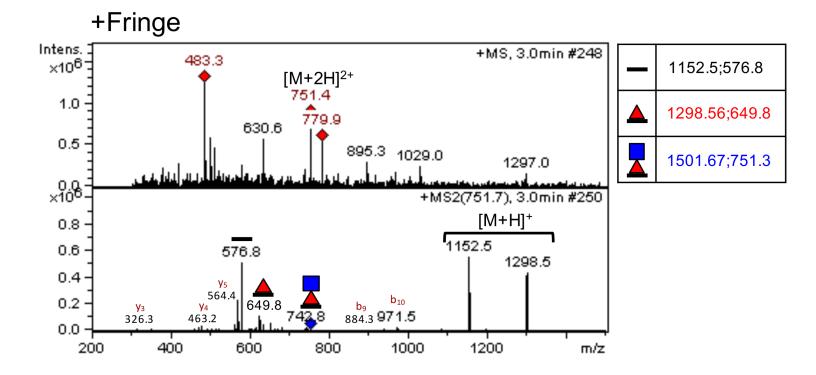


Fig. S3, J

²²⁸YGGTCVNTHGSYQCMCPTGYTGK²⁵⁰

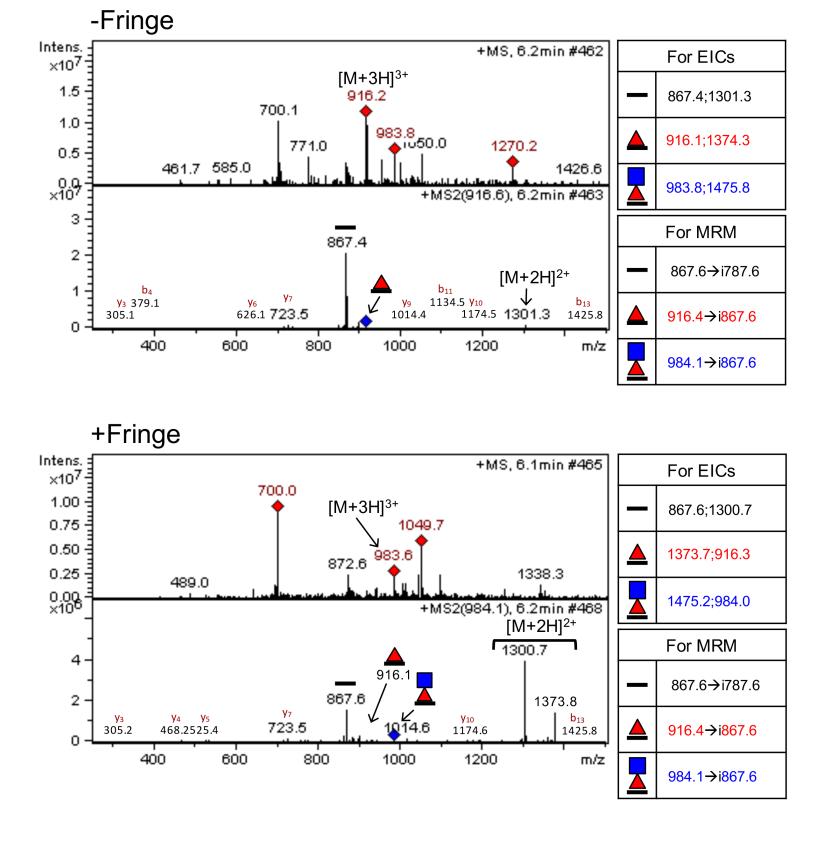


Fig. S3, K EGF 5

²⁴⁸TGKDCDTKYKPCSPSPCQNGGICRSNGLSY²⁷⁷

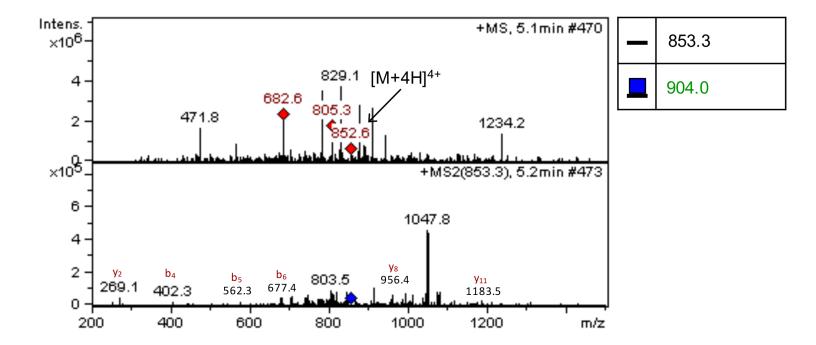


Fig. S3, L

EGF 7 289NCEQNYDDCLGHLCQNGGTCIDGISDYTCR318

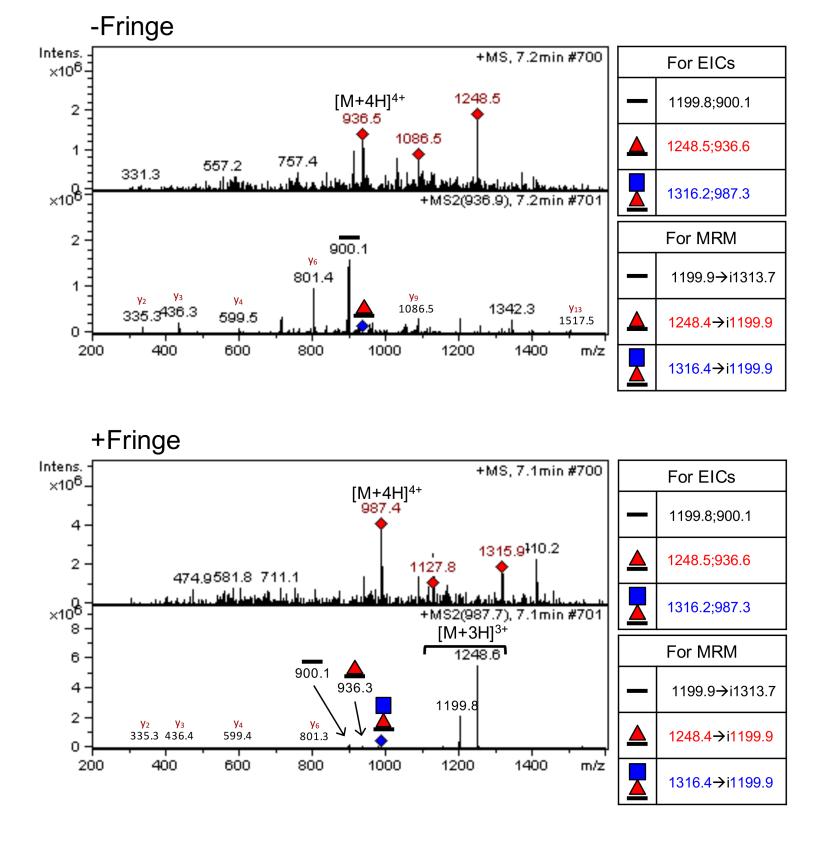
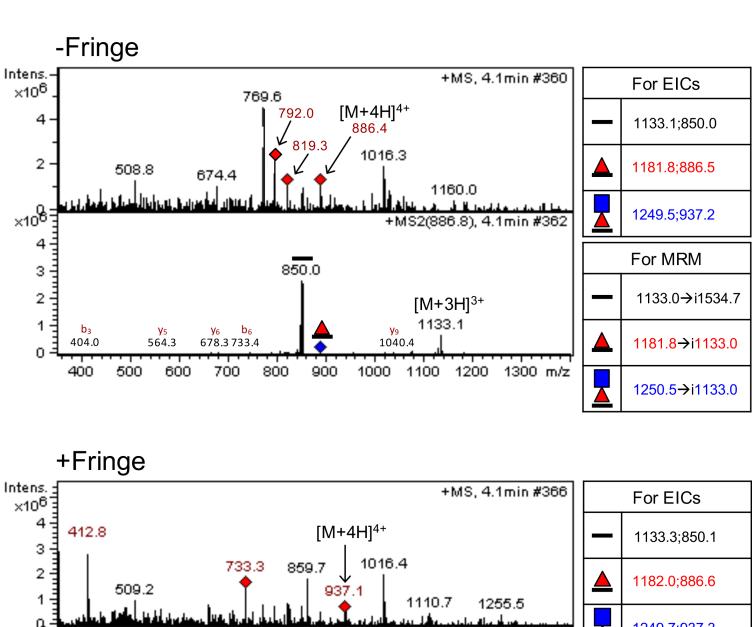


Fig. **S3**, **M**

EGF 8 328CQDDVDECAQRDHPVCQNGATCTNTHGSY356



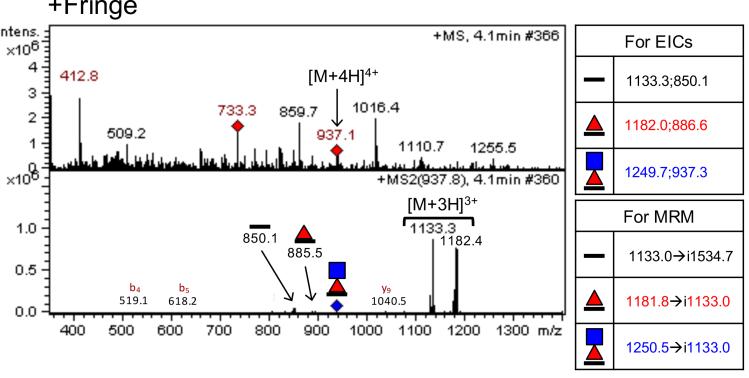
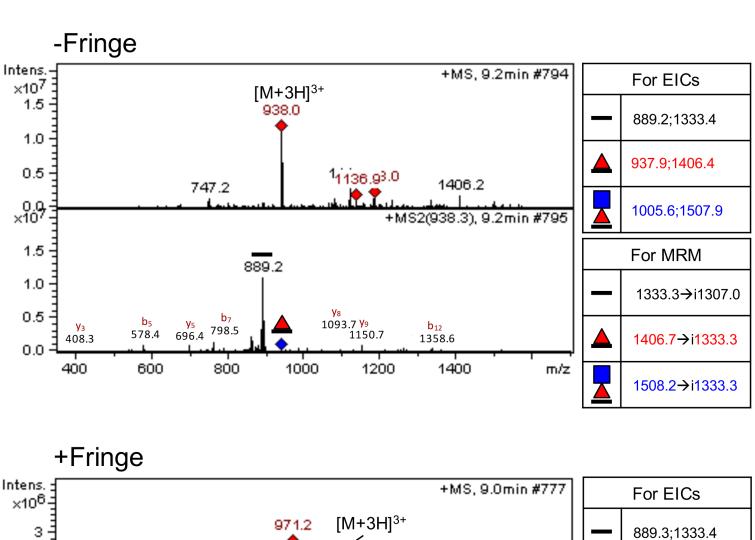


Fig. S3, N

EGF 9 378QAACFYGATCIDGVGSFYCQCTK400



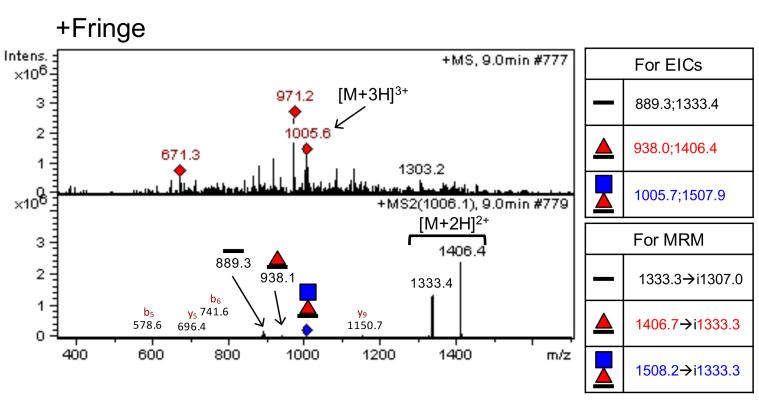


Fig. S3, O

EGF 9 396CQCTKGKTGLL406

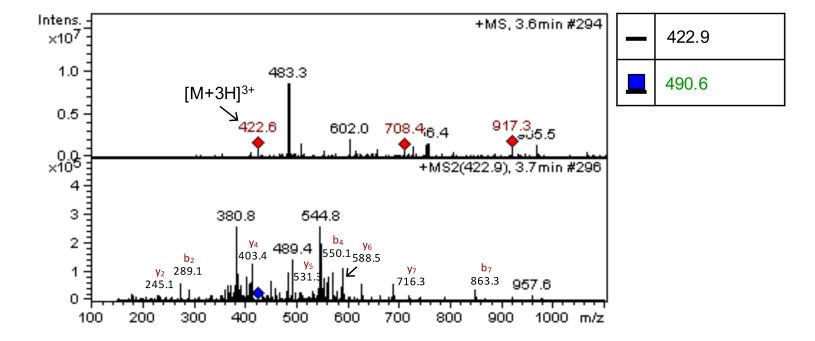


Fig. S3, P

EGF 9 and EGF 10

⁴⁰³TGLLCHLDDACTSNPCHADAICDTSPINGSYACSCATGYK⁴⁴²

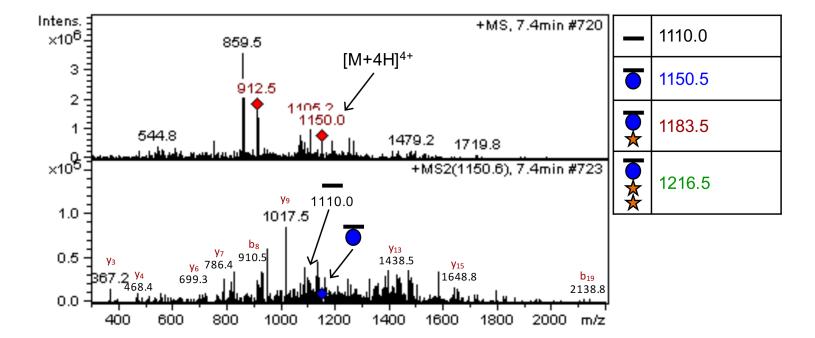


Fig. S3, Q

⁴⁷⁴CNCSQGF<u>T</u>GPR⁴⁸⁴

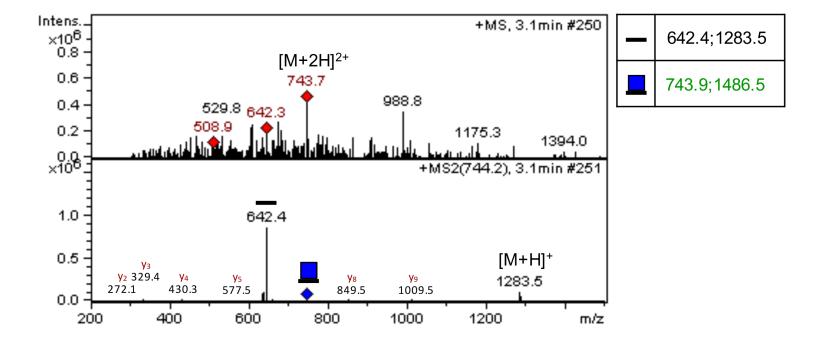


Fig. S3, R

⁴⁸⁵CETNINECESHPCQNEGSCLDDPGTFR⁵¹¹

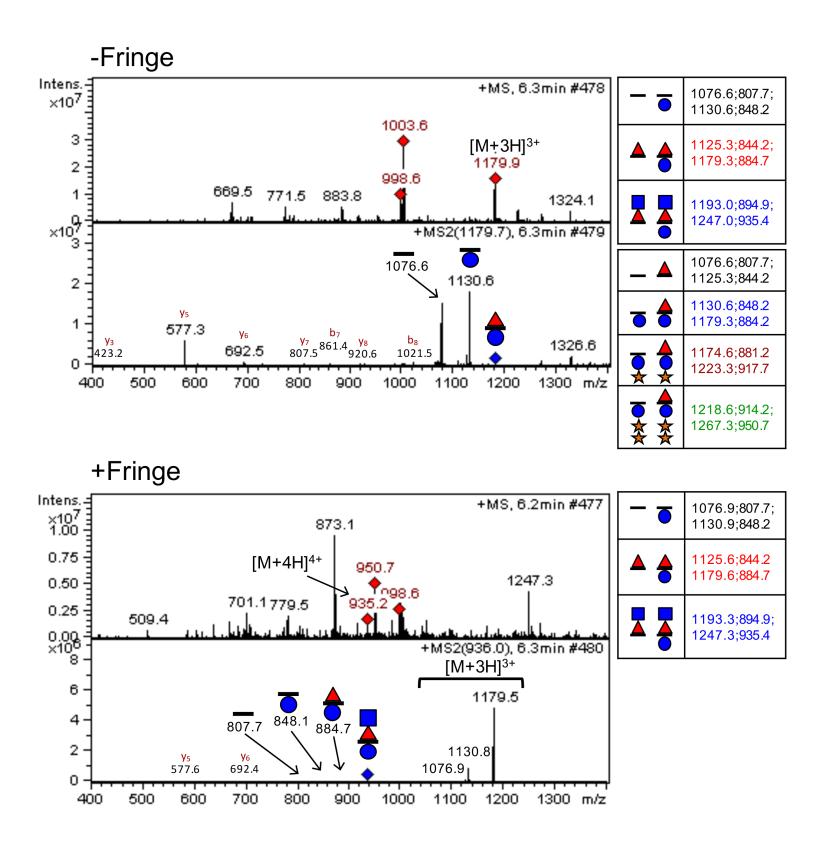
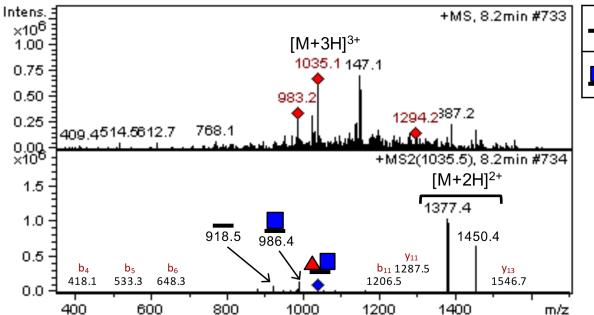


Fig. S3, S

⁵⁰¹GSCLDDPGTFRCVCMPGFTGTQCE⁵²⁴



 918.5;1377.4; 1450.4;967.2
1478.9;986.2; 1551.9;1034.9

Fig. S3, T

EGF 12 and EGF 13

⁵¹²CVCMPGFTGTQCEIDIDECQSNPCLNDGTCHDK⁵⁴⁴

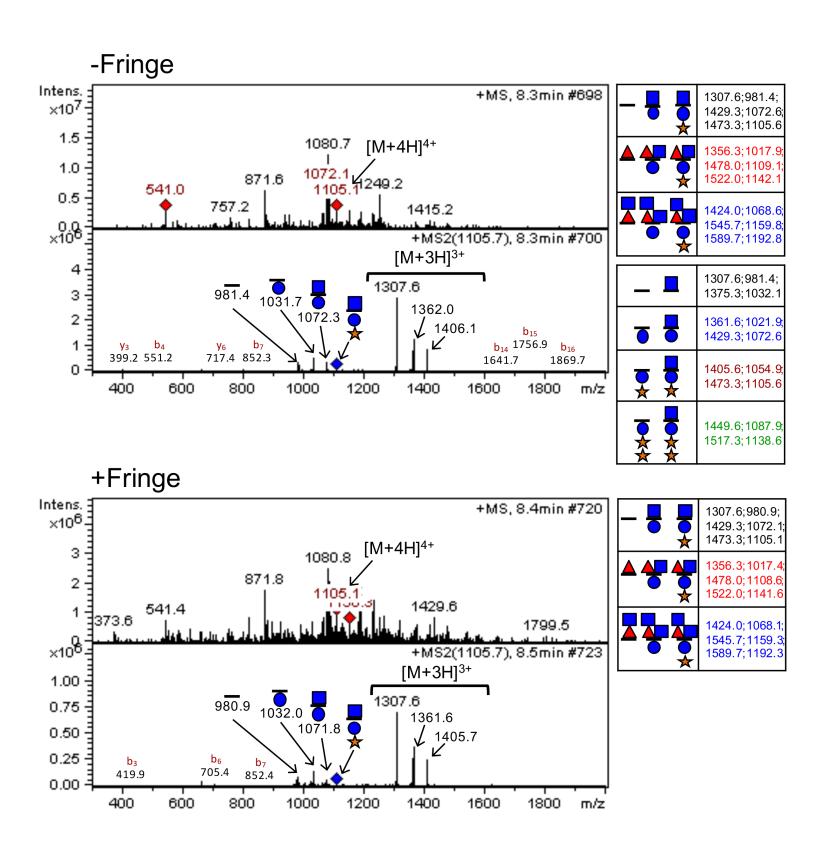
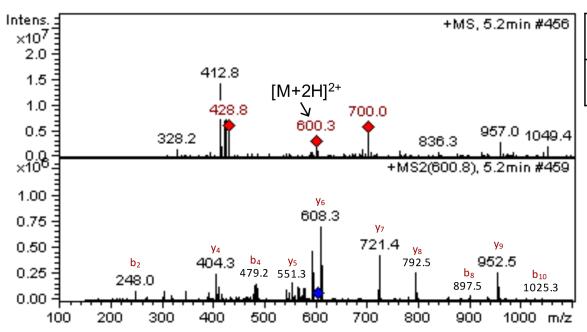


Fig. S3, U

EGF 13

⁵⁵⁰CSCALGFTGAR⁵⁶⁰



600.8
702.3

Fig. S3, V

EGF 13 and EGF 14

⁵⁵⁷**T**GARCQINIDDCQ**S**QPCRNRGICHDSIAGY⁵⁸⁶

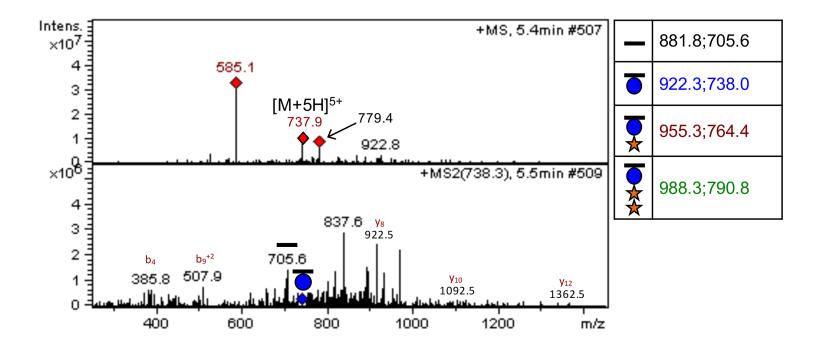


Fig. S3, W

EGF 14 and 15

⁵⁷⁷GICHDSIAGYSCECPPGYTGTSCEININDCDSNPCHR⁶¹³

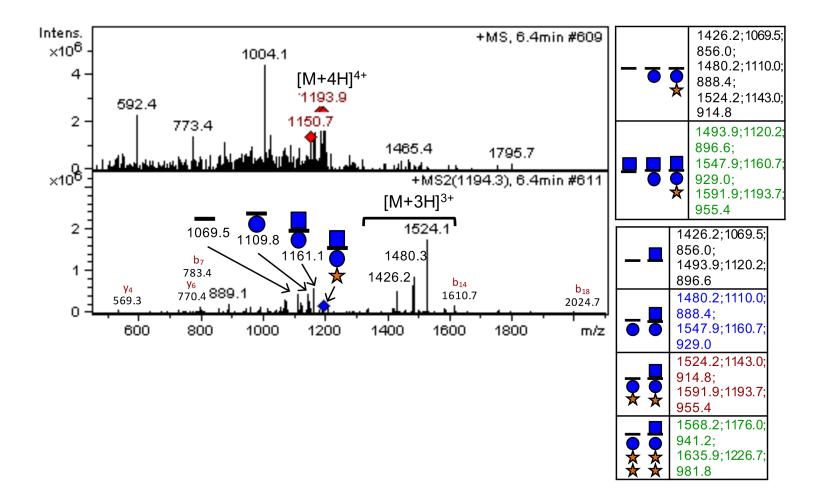


Fig. S3, X

EGF 15 625CLCDPGYTGYICQK638

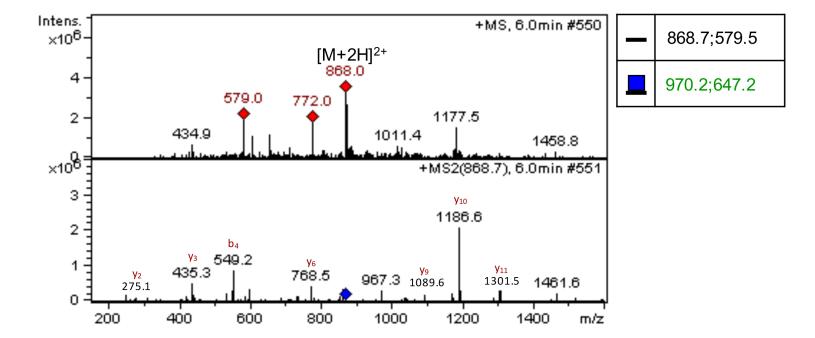


Fig. S3, Y EGF 16



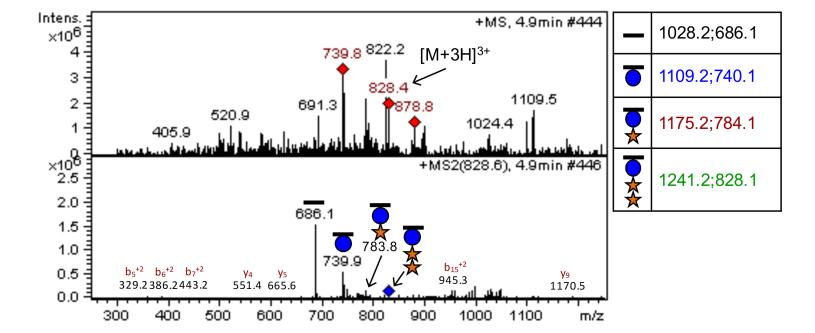


Fig. S3, Z

EGF 16 658VGSYYCQCQAGTSGK672

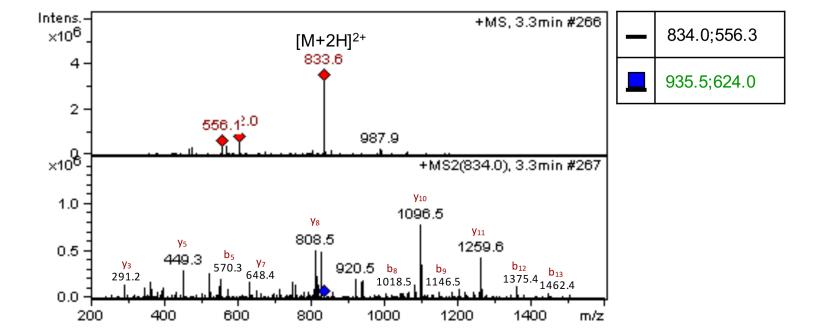


Fig. S3, A'

673NCEVNVNECHSNPCNNGATCIDGINSYK700

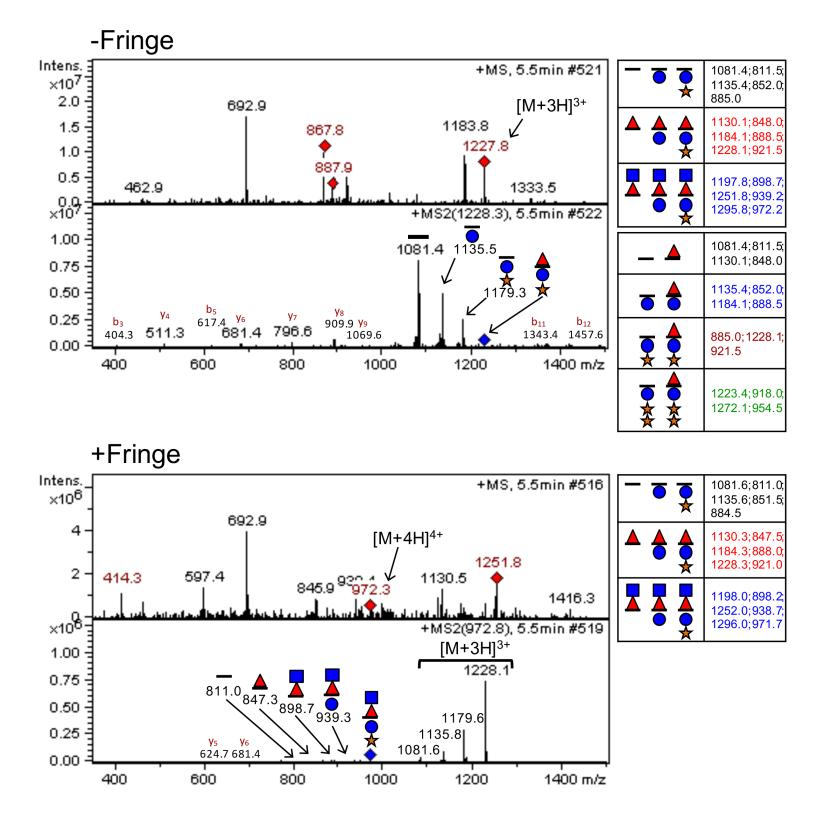


Fig. S3, B'

EGF 17 701CQCVPGFTGQHCEK714

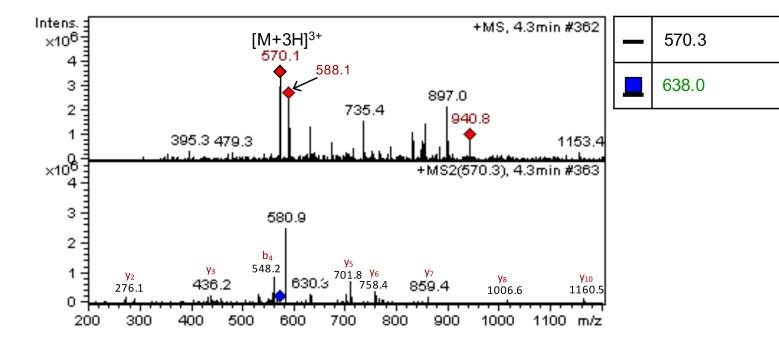


Fig. S3, C'

⁷¹⁵NVDECI<u>S</u>SPCANNGVCIDQVNGYK⁷³⁸

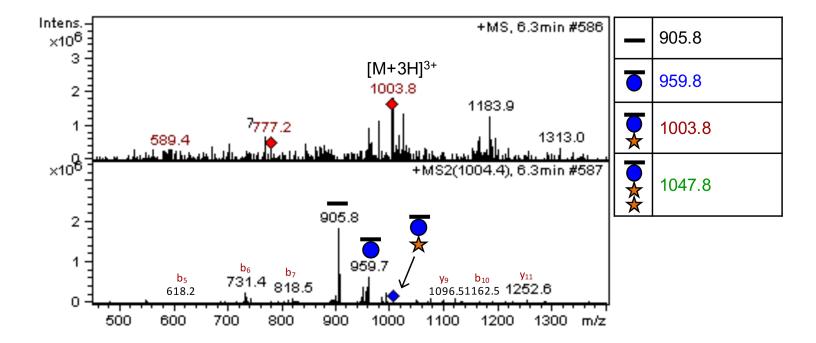


Fig. S3, D'

744GFYDAHCLSDVDECASNPCVNEGR767

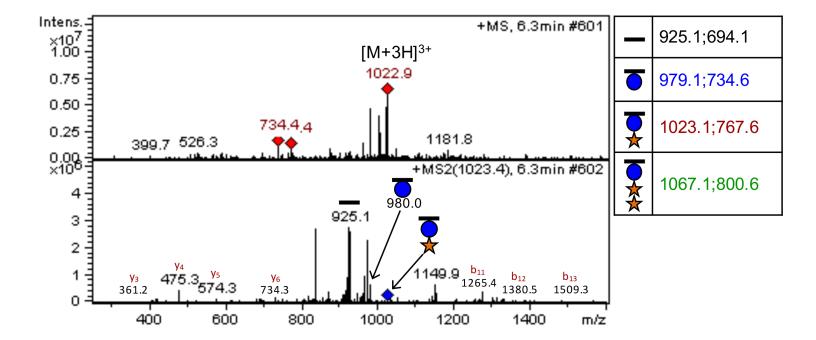


Fig. S3, E'

EGF 19

⁷⁶⁸CEDGINEFICHCPPGYTGK⁷⁸⁶

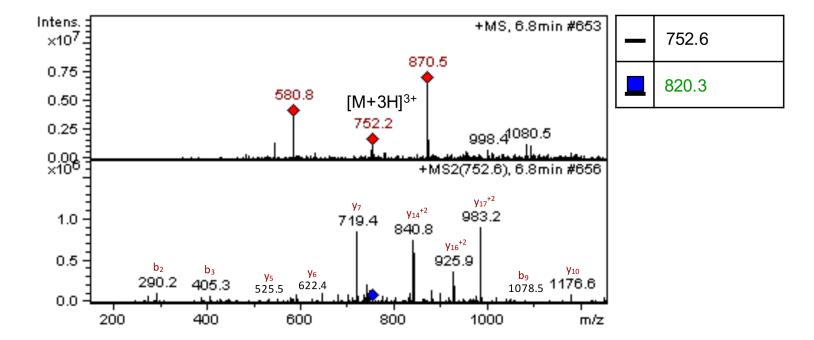


Fig. S3, F'

EGF 20

787RCELDIDECSSNPCQHGGTCYDK809

Note: [M+H]³⁺ is observed in spectra, but is not in EICs

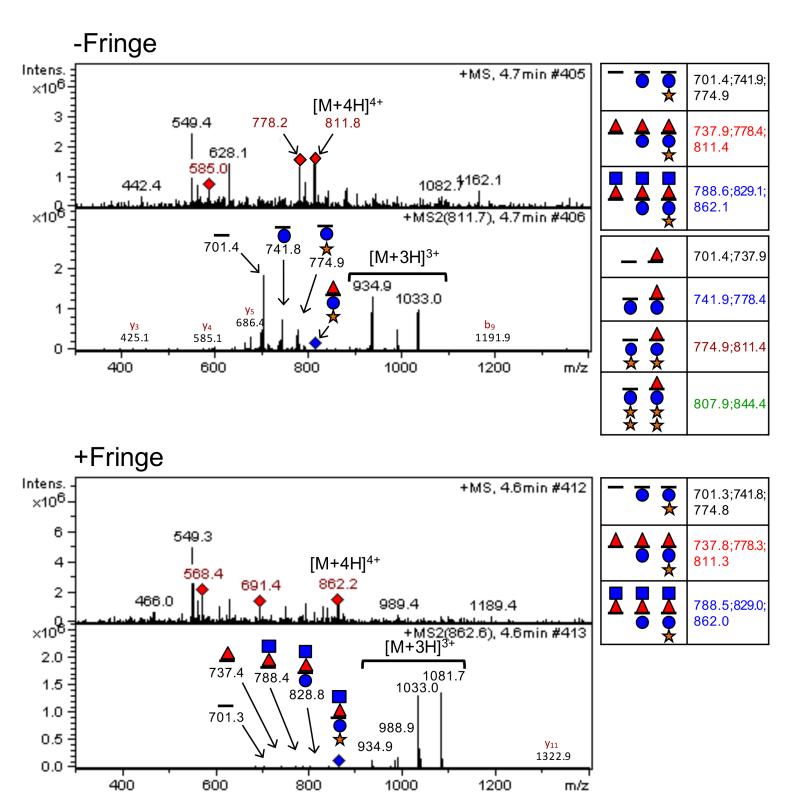


Fig. S3, G'

 810 LNAFSCQCMPGY $\underline{\mathsf{T}}$ GQK 825

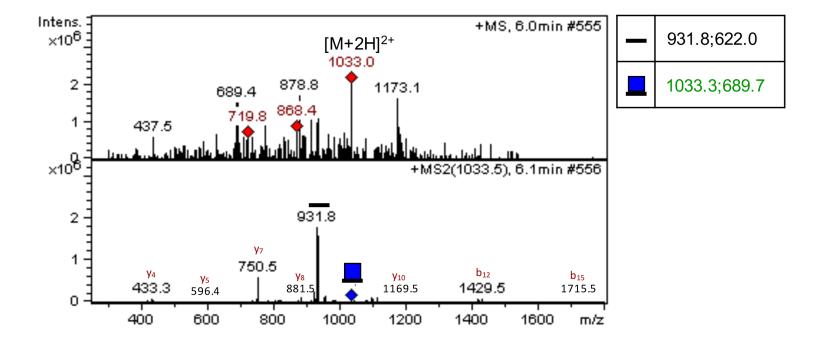


Fig. S3, H'

826CETNIDDCVTNPCGNGGTCIDKVNGYK852

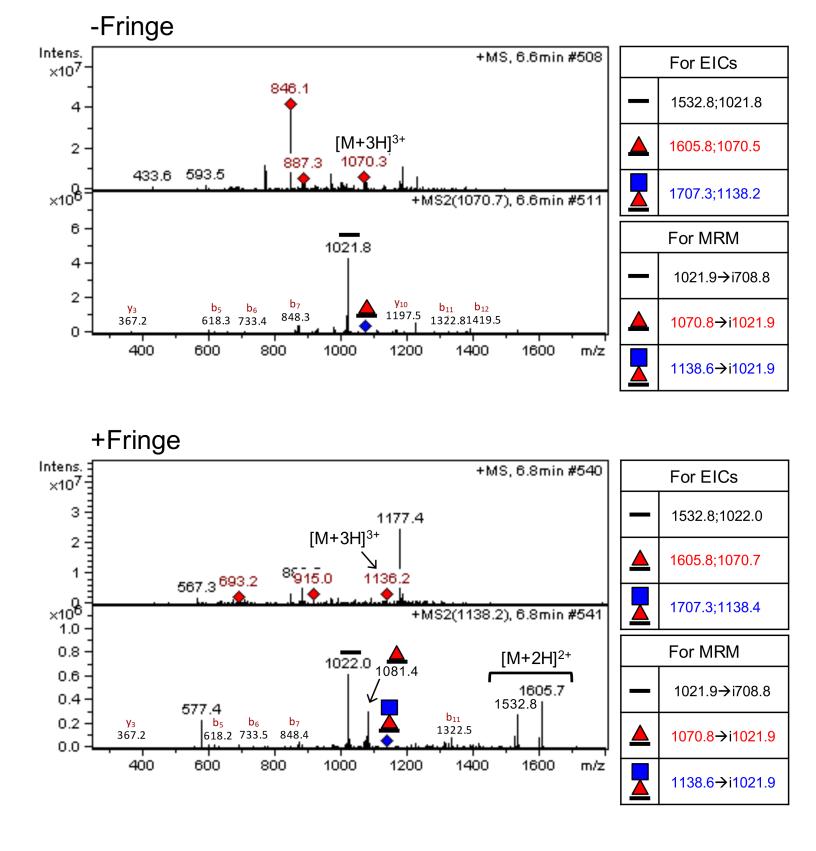
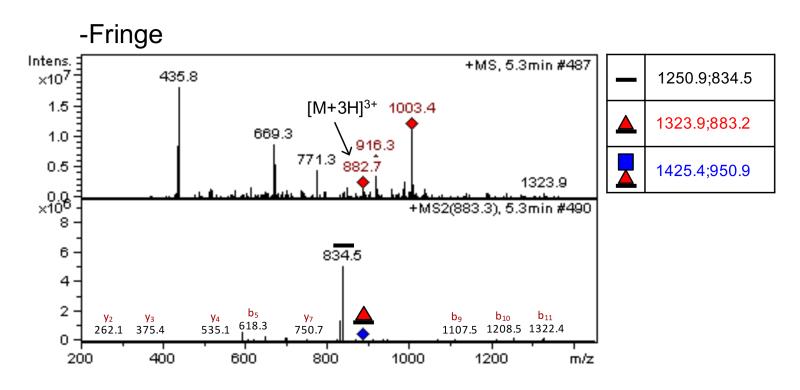


Fig. S3, I'

826CETNIDDCVTNPCGNGGTCIDK847

Used for Fig. 8A; Note: Shows same results as previous peptide



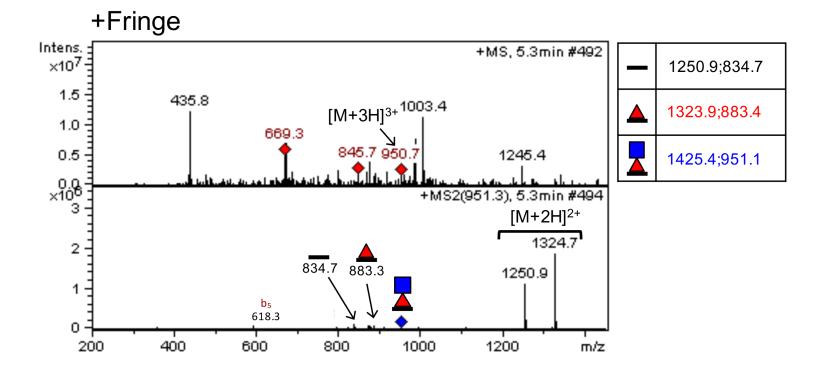
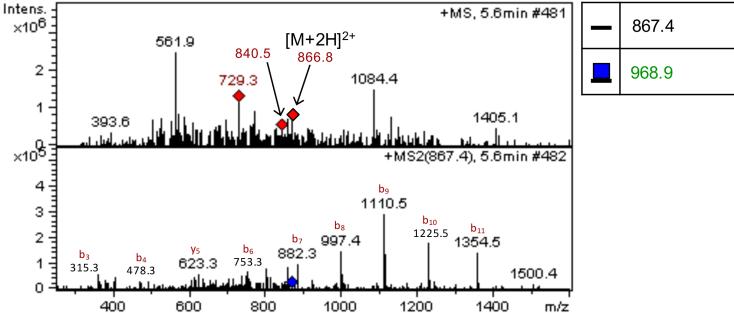


Fig. S3, J'

EGF 22 900TGRYCDEDIDECSL913



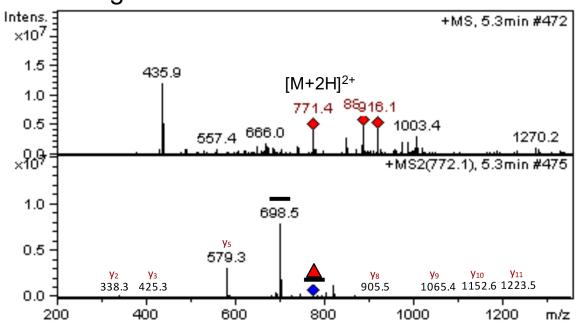
_	867.4
	968.9

Fig. S3, K'

EGF 23

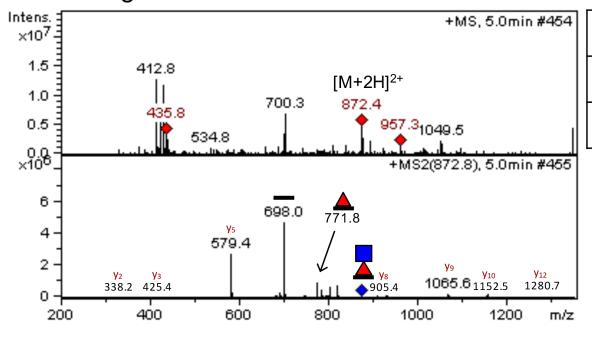
919NGASCLNVPGSYR931





1	698.5
_	771.5
	873.0

+Fringe



_	698.0
	771.0
	872.5

Fig. S3, L'

1

0

 b_7

800

802.6916.51017.4

1000

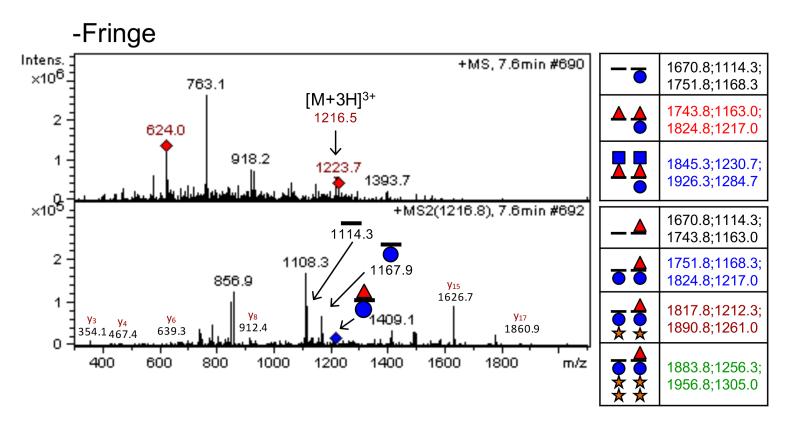
1200

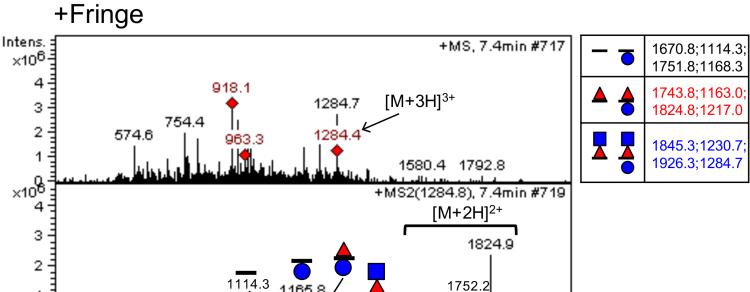
524.2

600

400

939EGRDCAINTDDCASFPCQNGGTCLDGIGDY968





1400

1671.2

1800

m/z

1600

Fig. S3, M'

972CVDGFDGKHCETDINECLSQPCQNGATCSQY1002

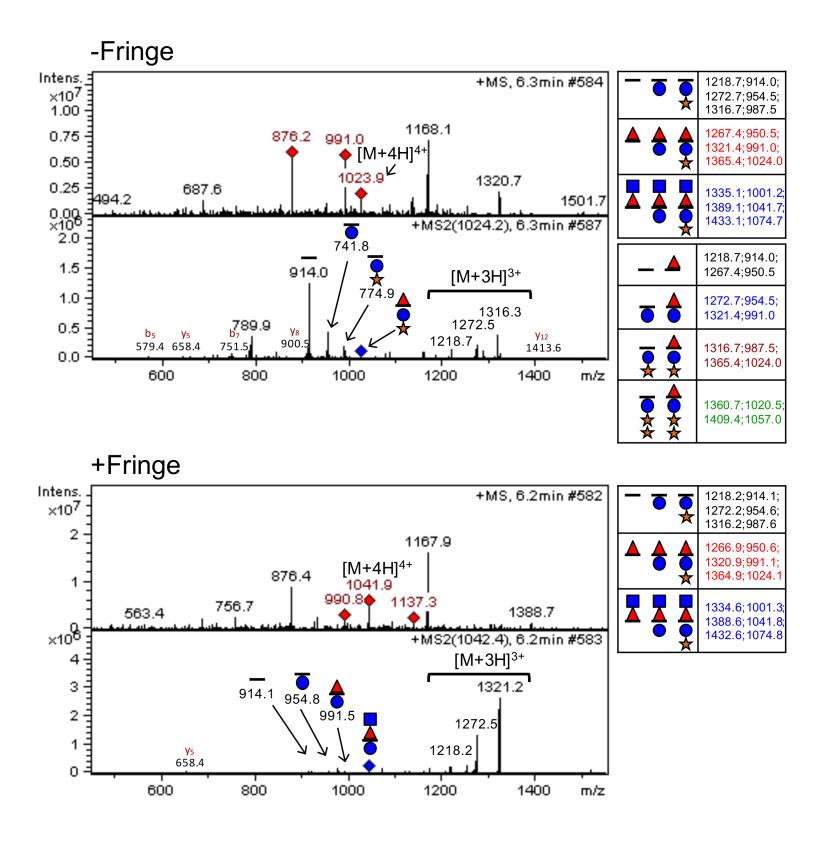


Fig. S3, N'

¹⁰¹⁵**S**GINCQTNDEDCTESSCL¹⁰³²

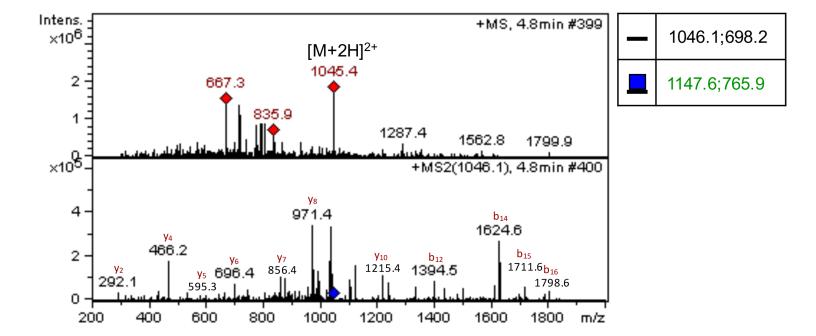
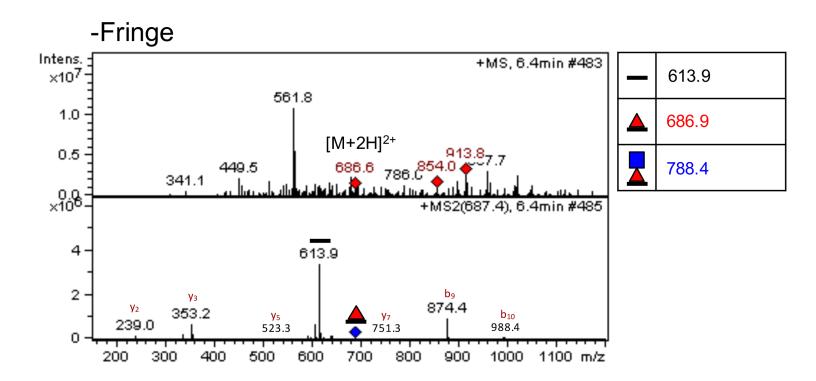


Fig. S3, O'

EGF 26

¹⁰³³NGGSCIDGINGY¹⁰⁴⁴



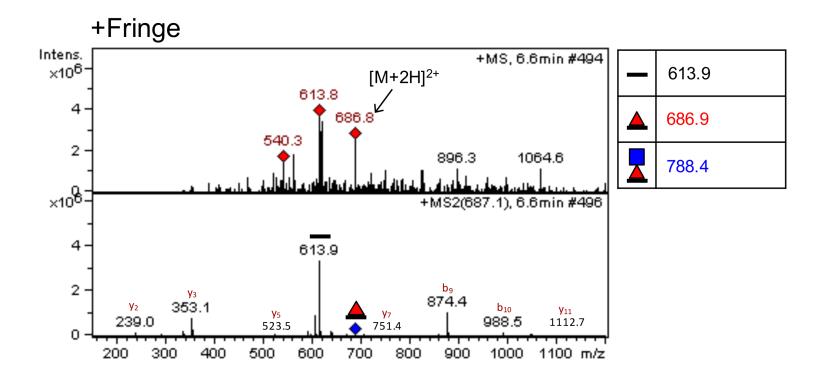


Fig. S3, P'

EGF 26 1050 AGY SGANCQY 1059

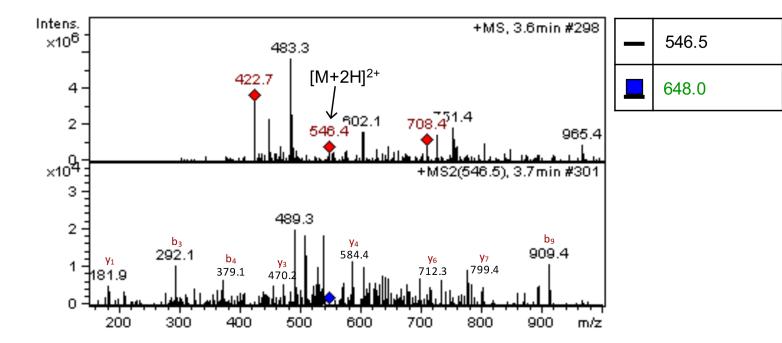
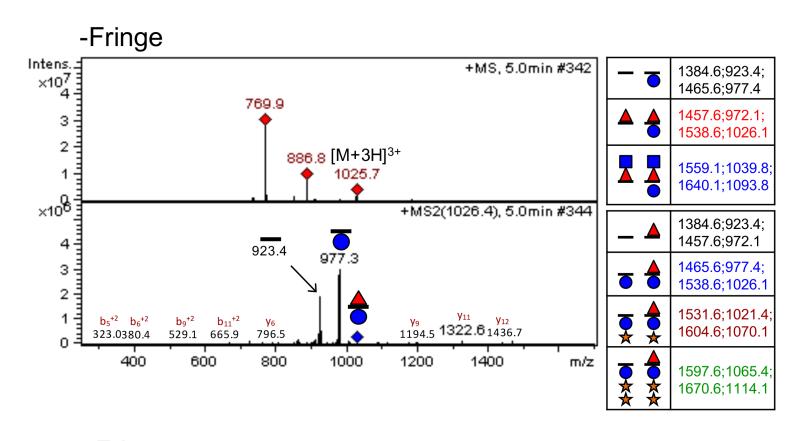
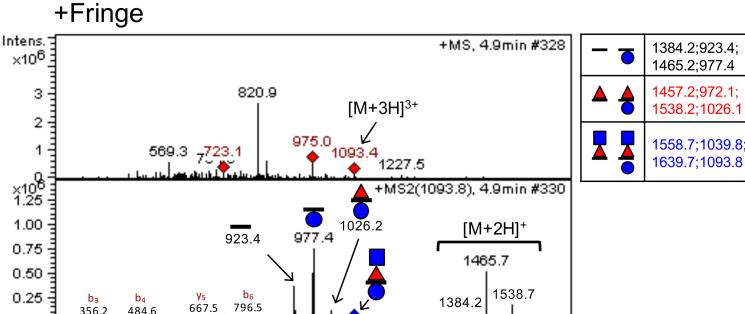


Fig. S3, Q'

¹⁰⁶⁰KLNKCDSNPCLNGATCHEQNNEY¹⁰⁸²





1200

796.5

800

1000

484.6

600

356.2

400

0.00

1384.2

1400

m/z

Fig. S3, R'

EGF 27 1091 TGKQCSEYVDW1101

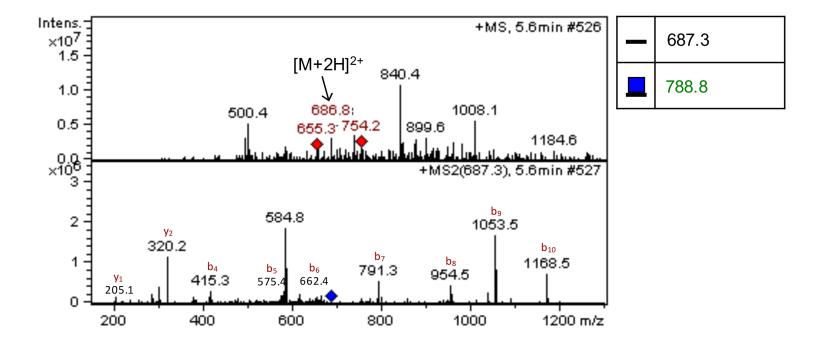


Fig. S3, S'

¹⁰⁹⁴QCSEYVDWCGQSPCENGATCSQMK¹¹¹⁷

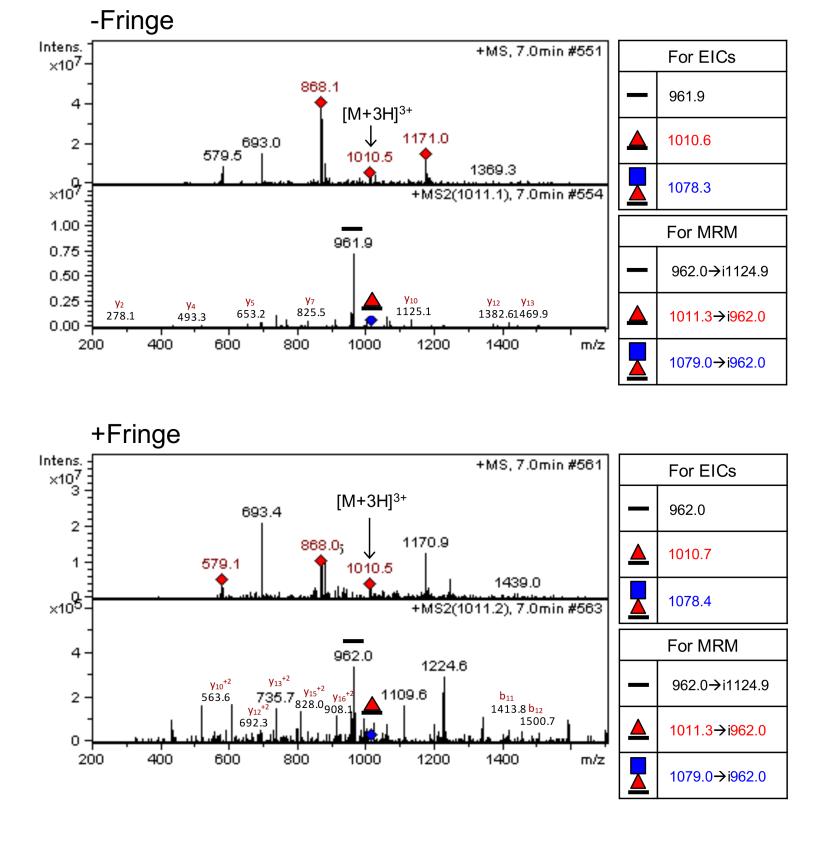


Fig. S3, T' EGF 28

¹¹²⁹**T**GKLCDVQTISCQDAADRKGL¹¹⁴⁹

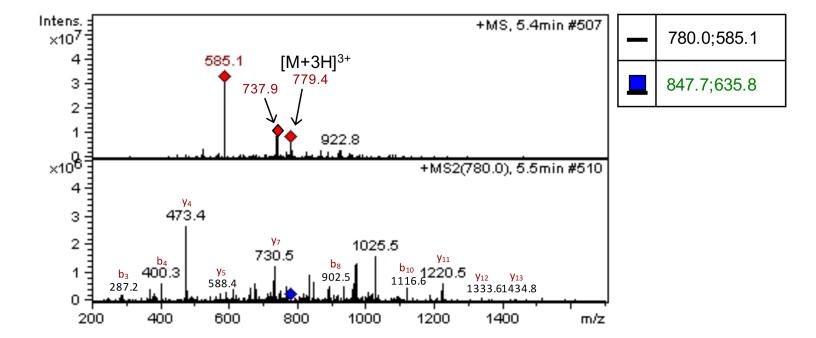
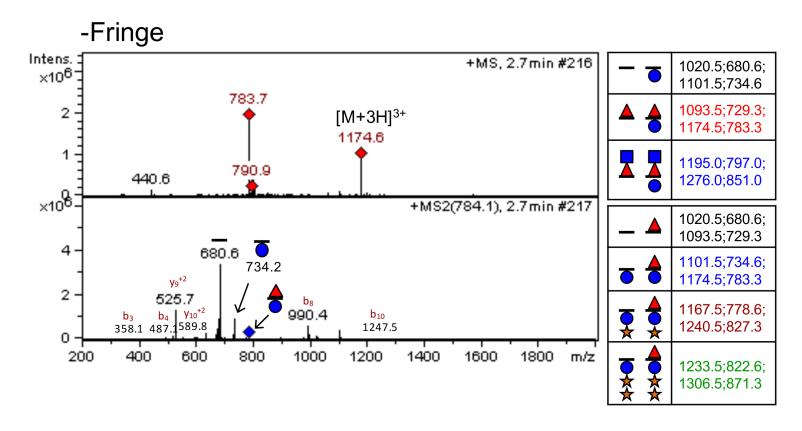


Fig. S3, U'

¹¹⁸³EIDECQSQPCQNGGTCR¹¹⁹⁹



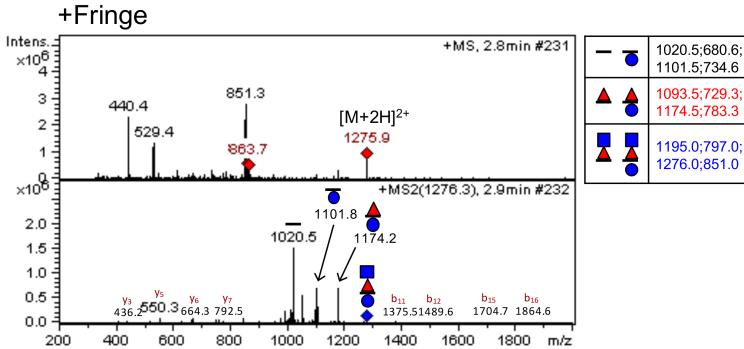
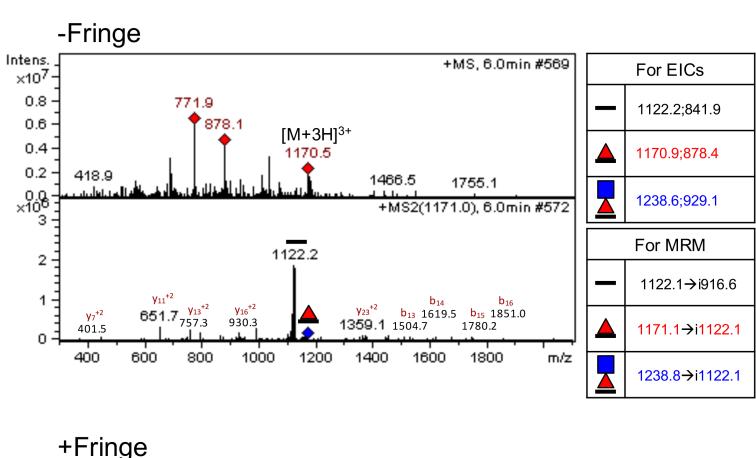


Fig. S3, V'

EGF 31

¹²¹¹QGFQGQNCELNIDDCAPNPCQNGG<u>T</u>CHDR¹²³⁹



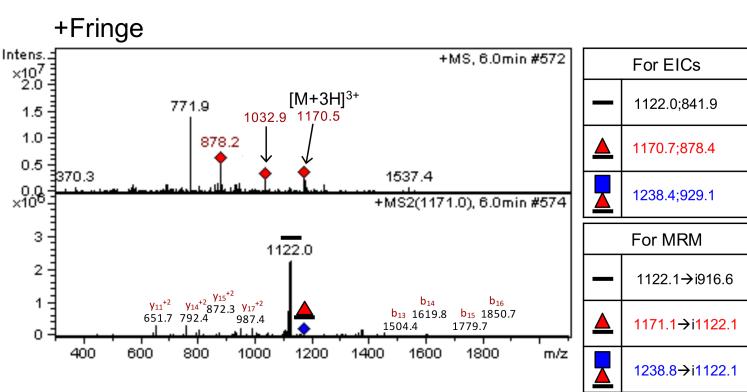
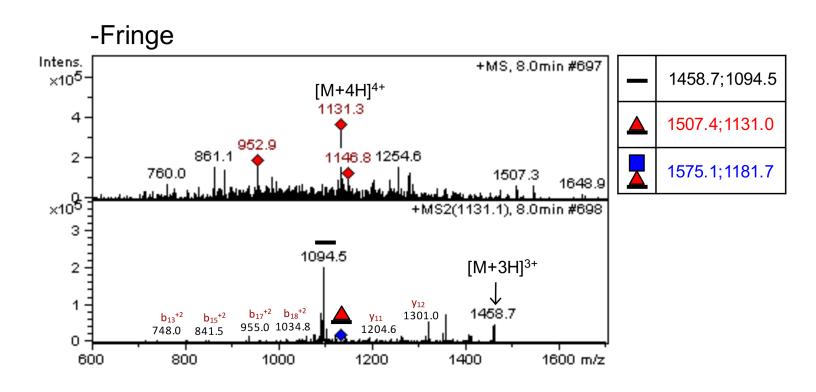


Fig. S3, W'

¹²³⁹RVMNFSCSCPPGTMGIICEINKDDCKPGACHNNG<u>S</u>CID¹²⁷⁶



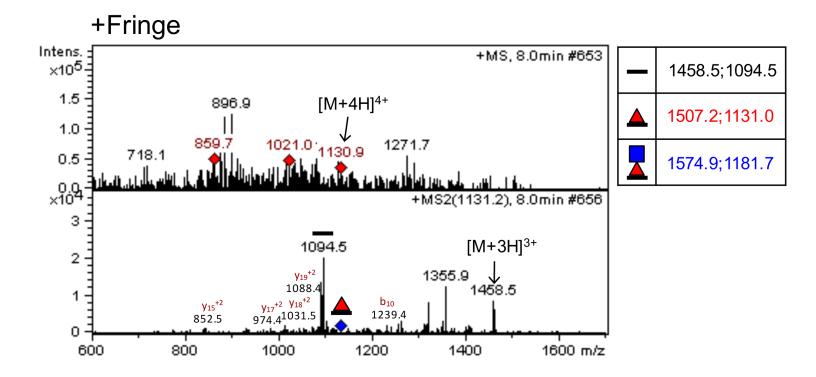


Fig. S3, X'

1303SNPCSNAGTLDCVQL1317

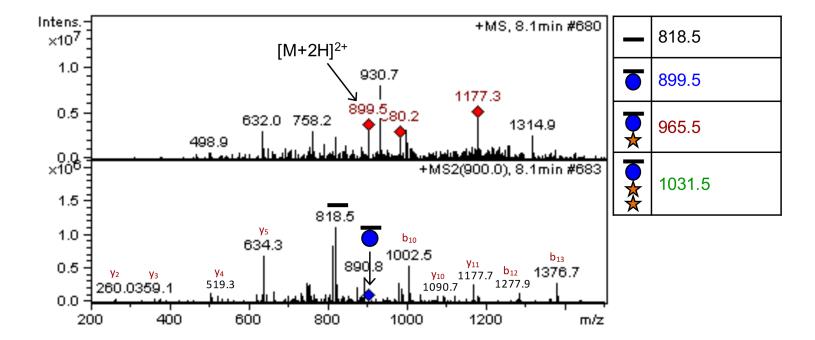


Fig. S3, Y'

1371NCELSGQDCDSNPCR1385

