SUPPORTING INFORMATION

Title: "Targeting hydrophilic regions of recombinant proteins by MS using in-solution buffer-free trypsin digestion"

Authors: Lázaro Betancourt[1&*], Luis Ariel Espinosa[2&], Yassel Ramos[2], Mónica Bequet-Romero[3], Elías Nelson Rodríguez[4], Aniel Sánchez[1], Gyorgy Marko-Varga[1], Luis Javier González[2], Vladimir Besada[2*]

Affiliation:

[1]Div. Clinical Protein Science & Imaging, Dept. of Clinical Sciences (Lund) and Dept. of Biomedical Engineering, Lund University, BMC D13, Lund, Sweden.

[2]Mass Spectrometry Laboratory, Department of Proteomics; [3]Pharmaceutics, Biomedical Research; [4]Division for Technological Development, Center for Genetic Engineering and Biotechnology, PO Box 6162, Havana, Cuba.

& Authors with same contribution
* Corresponding authors:

E-mail: lazaro.betancourt@med.lu.se; vladimir.besada@cigb.edu.cu

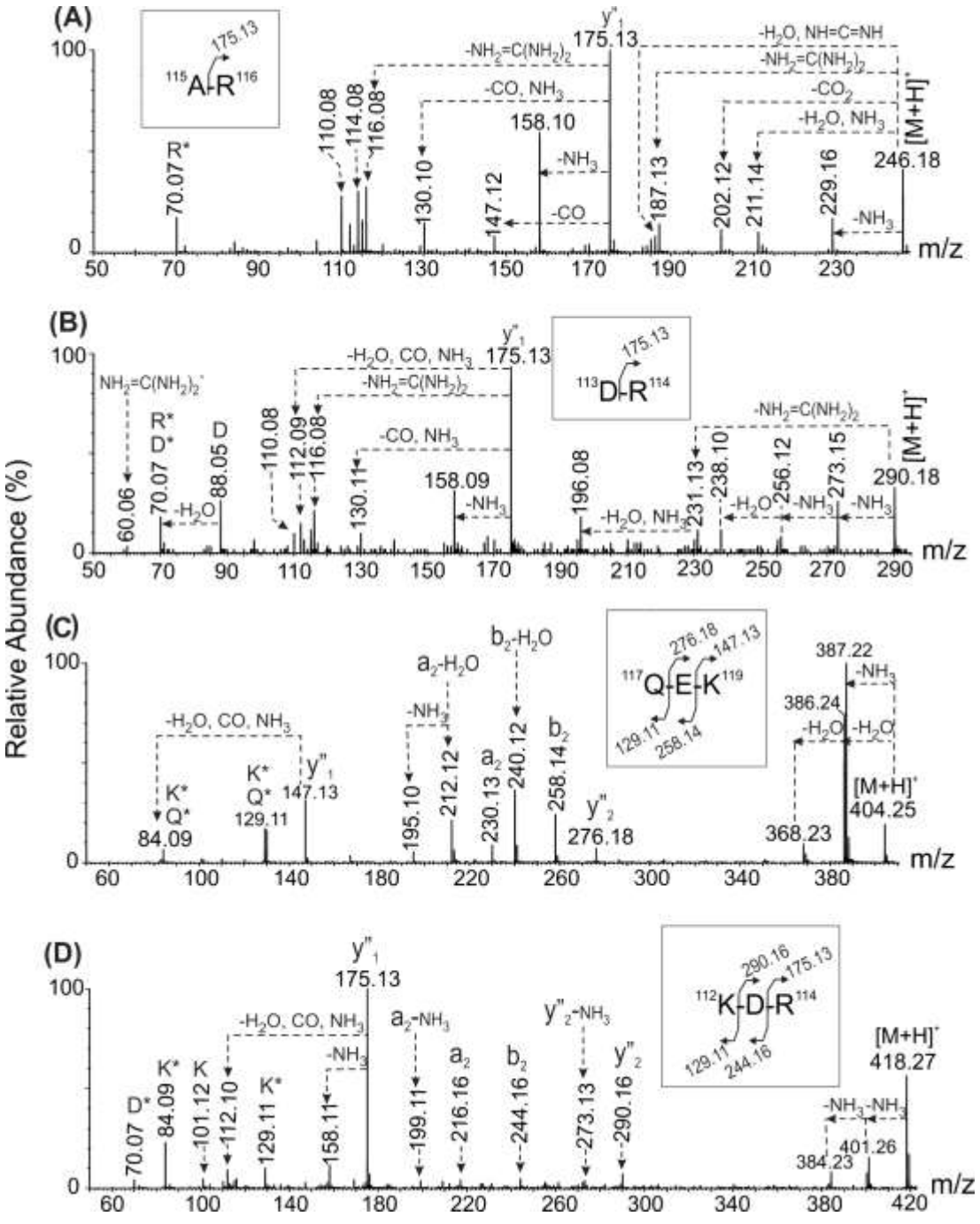

Figure S1. ESI-MS/MS spectra of four short and hydrophilic peptides obtained by the in-solution buffer free tryptic digestion procedure of the fusion VEGF-recombinant protein. (**A**) ^{115}AR^{116} (expected *m/z* 246.16, 1+); **(B)** ^{113}DR^{114} (expected *m/z* 290.15, 1+); (**C**) ^{117}QEK^{119} (expected *m/z* 404.21, 1+); and (**D**) ^{112}KDR^{114} (expected *m/z* 418.24, 1+). Fragment ions observed in (**A**-**D**) that were generated by neutral losses are indicated with broken lines. The assignment of these

fragment ions are in agreement with previous works[1,2,3] that report the fragmentation of protonated α-amino acids and dipeptides containing arginine and lysine. Single-letter amino acids with asterisk represent the immonium-related ions.

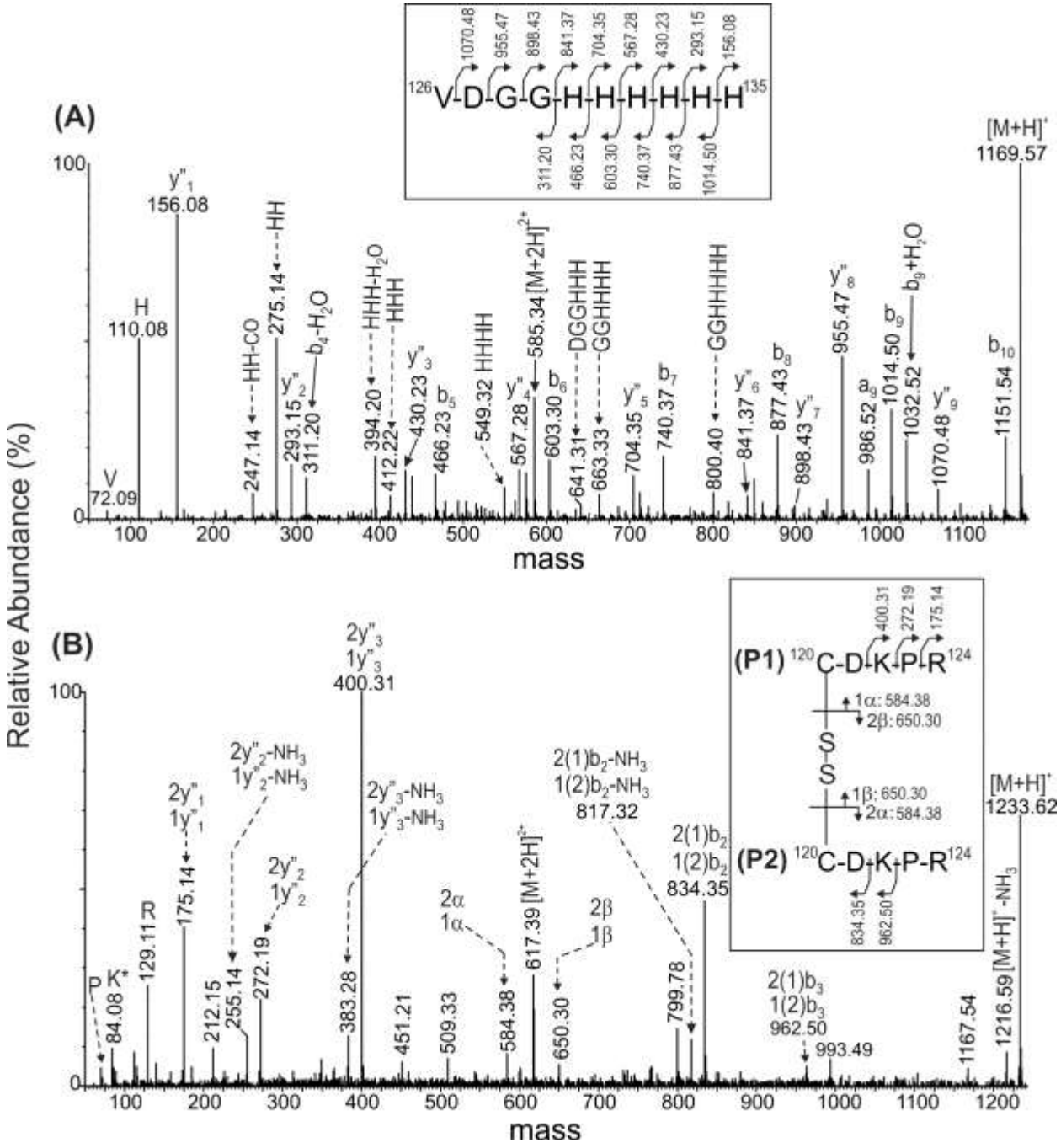

Figure S2. Deconvoluted ESI-MS/MS spectra of two hydrophilic tryptic peptides obtained by the in-solution BFD procedure of the fusion VEGF-recombinant protein. **(A)** ^{126}VDGGHHHHHH^{135} (expected *m/z* 585.26, 2+) corresponding to the C-terminal end and **(B)** peptides linked by an inter-molecular disulfide bond (^{120}CDKPR^{124})-S-S-(^{120}CDKPR^{124}) (expected *m/z* 617.30, 2+). "-S-S-" indicates the sulphur atoms linked by an intermolecular disulfide bond between (P1) and

(P2) peptides. Single-letter amino acids with asterisk represent the immonium-related ions. The nomenclature used for the assignment of fragment ions agrees with previous reports[4].

The ESI-MS/MS spectrum of the His-tag C-terminal peptide (figure S-2A) confirmed the C-terminal sequence shown in figure 2A. Complete set of y''_n ion series as well as six b_n ions (b_4 to b_{10}) were observed. Immonium ions as well as internal fragments detected in the MS/MS spectrum confirmed the assigned sequences. Also, the ESI-MS/MS spectra of ion (*m/z* 617.31, 2+, figure S2 B) further confirmed the inter-molecular disulfide bond peptide (figure 2A) of the protein. The fragment ions detected at *m/z* 584.38, 1+ and 650.30, 1+ were assigned to the asymmetric cleavage of the disulfide bond to yield the peptide (^{120}CDKPR^{124}) with the cysteine transformed as dehydroalanine [1α or 2α] and disulfohydryl cysteine [1β or 2β] residues[4], respectively.

Table S1. Assignment of ESI-MS spectra of the hydrophilic peptides of HBcAg: comparison of in-solution BFD and standard digestion (Std Dig). The experimental and expected m/z values for each peptide that contributed to the 100% sequence verification of HBcAg is summarized in this Table.

Experimental *m/z*			**Expected *m/z***	**z**	**Sequence**	**Detection procedure [b)]**	
02C.IFA.E407 [a)]	**02C.IFA.E505 [a)]**	**02C.IFA.E503 [a)]**				**BFD**	**Std Dig**
638.63	638.64	638.64	638.64	3	^{40}E-R^{56}	x	-
331.23	331.23	331.23	331.22	1	^{151}RR^{152}	x	-
290.15	290.15	290.15	290.15	1	^{153}DR^{154}	x	-
503.28	503.27	503.27	503.27	1	^{153}DRGR^{156}	x	-
359.20	359.21	359.21	359.20	1	^{157}SPR^{159}	x	-
331.23	331.23	331.23	331.22	1	^{160}RR^{161}	x	-
557.31	557.31	557.31	557.30	1	^{162}TPSPR^{166}	x	-
331.23	331.23	331.23	331.22	1	^{167}RR^{168}	x	-
487.29	487.30	487.29	487.32	1	^{167}RRR^{169}	x	-
574.30	574.30	574.30	574.29	1	^{170}SQSPR^{174}	x	-
331.23	331.23	331.23	331.22	1	^{175}RR^{176}	x	-
487.29	487.30	487.29	487.32	1	^{175}RRR^{177}	x	-
477.25	477.24	477.25	477.24	1	^{178}SQSR^{181}	x	-
466.16	466.16	466.17	466.16	1	^{182}ESQC^{185} (C-terminal)	x	-

a) Codes assigned to the three production batches of HBcAg characterized in this study by in-solution BFD.

b) Dashes mean that a given peptide was not detected, while "X" means that the tryptic peptide was detected and its identity was confirmed by ESI-MS/MS analysis. All signals assigned to the HBcAg sequence using the standard digestion procedure were also found in the ESI-MS spectra generated by the in-solution BFD of the three productions batches.

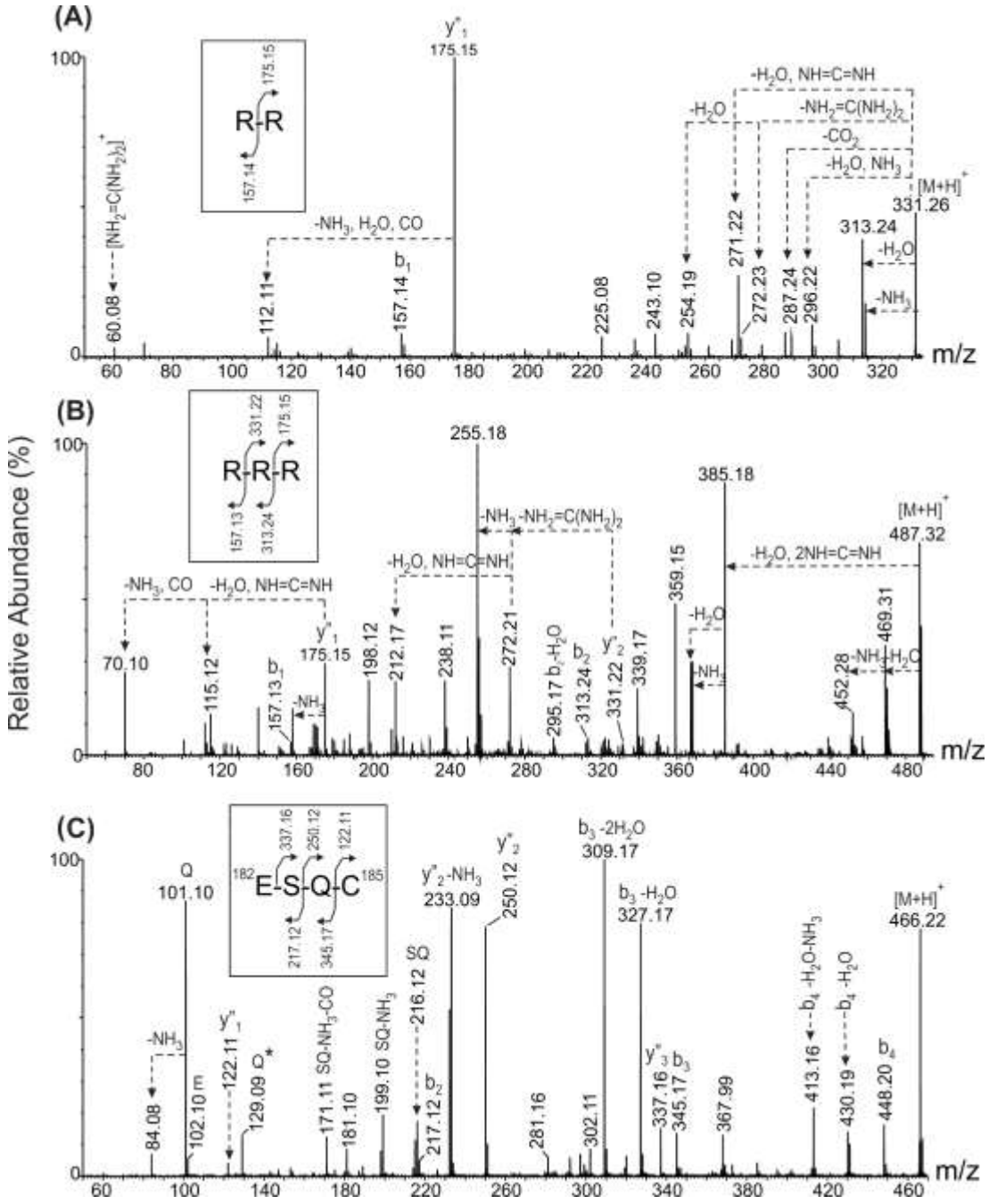

Figure S3. ESI-MS/MS spectra of three hydrophilic peptides obtained by the in-solution BFD procedure of the HBcAg protein: **(A)** RR (expected *m/z* 331.22, 1+; probable amino acid regions 151-152, 160-161, 167-168, 168-169, 175-176 or 176-177); **(B)** RRR (expected *m/z* 487.32, 1+; probable amino acid regions 167-169 or 175-177); and **(C)** ^{182}ESQC^{185} (expected *m/z* 466.22, 1+) corresponding to the C-terminal peptide with free cysteine residue. The most intense fragment ions observed in the MS/MS spectra of arginine peptides shown in **(A)** and **(B)** are generated by the partial and complete loss of the guanidinium group side chain of arginine residues and are in agreement with previous works[1,2,3]. Single-letter amino acids with asterisk represent the immonium-related ions.

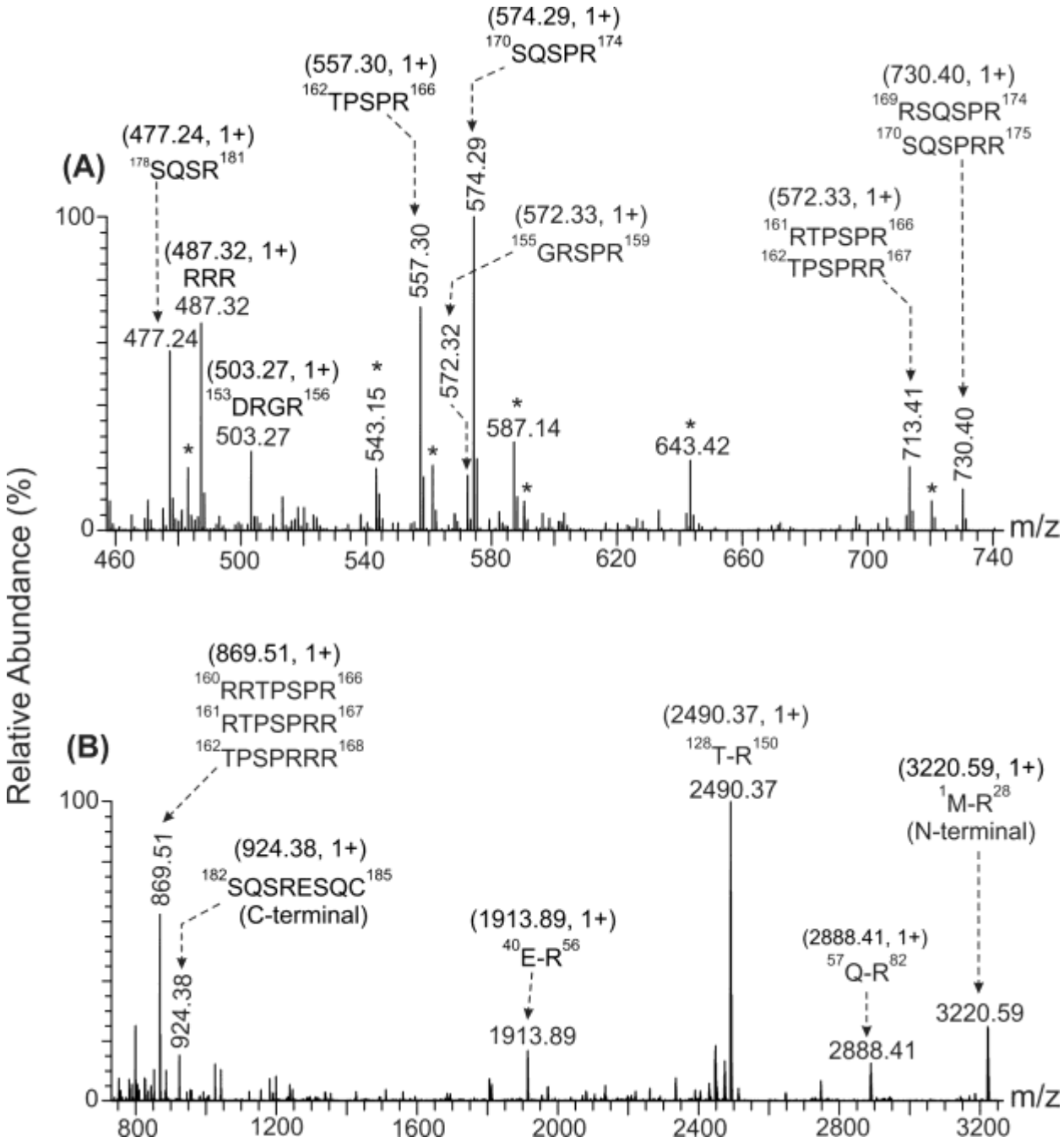

Figure S4. MALDI-MS spectra of the HBcAg tryptic digestion by using the in-solution BFD procedure, **(A)** low mass range and **(B)** high mass range. Signals with asterisks were the unique detected in the matrix spectrum. Peptide sequence RRR (*m/z* 487.32, 1+) could be assigned to amino acid regions 167-169 or 175-177.

Table S2. Summary of the assignment of MALDI-MS spectrum shown in figure S5 obtained from in-solution BFD of HBcAg.

Experimental m/z $(M+H)^+$	**Expected m/z $(M+H)^+$**	**Sequence**
3220.5864	3220.5910	^{1}M-R^{28}
1237.6435	1237.6423	^{29}D-R^{39}
1913.8941	1913.8922	^{40}E-R^{56}
2888.4050	2888.4069	^{57}Q-R^{82}
1579.8165	1579.8148	^{83}D-K^{96}
1990.0857	1990.0843	^{97}I-R^{112}
1720.8987	1720.8992	^{99}Q-R^{112}
1810.9755	1810.9737	^{113}E-R^{127}
2490.3710	2490.3714	^{128}T-R^{150}
2646.4716	2646.4725	^{128}T-R^{151}
503.2683	503.2684	^{153}DRGR^{156}
843.4543	843.4543	^{153}DRGRSPR^{159}
572.3263	572.3263	^{155}GRSPR^{159}
869.5072	869.5064	^{160}RRTPSPR^{166}
1025.6086	1025.6075	^{160}RRTPSPRR^{167}
713.4057	713.4053	^{161}RTPSPR^{166}
869.5072	869.5064	^{161}RTPSPRR^{167}
1025.6086	1025.6075	^{161}RTPSPRRR^{168}
557.3041	557.3042	^{162}TPSPR^{166}
713.4057	713.4053	^{162}TPSPRR^{167}

869.5072	869.5064	^{162}TPSPRRR^{168}
1025.6086	1025.6075	^{162}TPSPRRRR^{169}
487.3210	487.3212	^{167}RRR^{169}
1042.5987	1042.5977	^{167}RRRSQSPR^{174}
886.4966	886.4966	^{168}RRSQSPR^{174}
1042.5987	1042.5977	^{168}RRSQSPRR^{175}
730.3958	730.3954	^{169}RSQSPR^{174}
886.4973	886.4966	^{169}RSQSPRR^{175}
1042.5987	1042.5977	^{169}RSQSPRRR^{176}
574.2943	574.2943	^{170}SQSPR^{174}
730.3958	730.3954	^{170}SQSPRR^{175}
886.4973	886.4966	^{170}SQSPRRR^{176}
1042.5987	1042.5977	^{170}SQSPRRRR^{177}
487.3210	487.3212	^{175}RRR^{177}
945.5457	945.5449	^{175}RRRSQSR^{181}
789.4445	789.4438	^{176}RRSQSR^{181}
633.3429	633.3427	^{177}RSQSR^{181}
477.2415	477.2416	^{178}SQSR^{181}
924.3847	924.3840	^{178}SQSRESQC^{185}

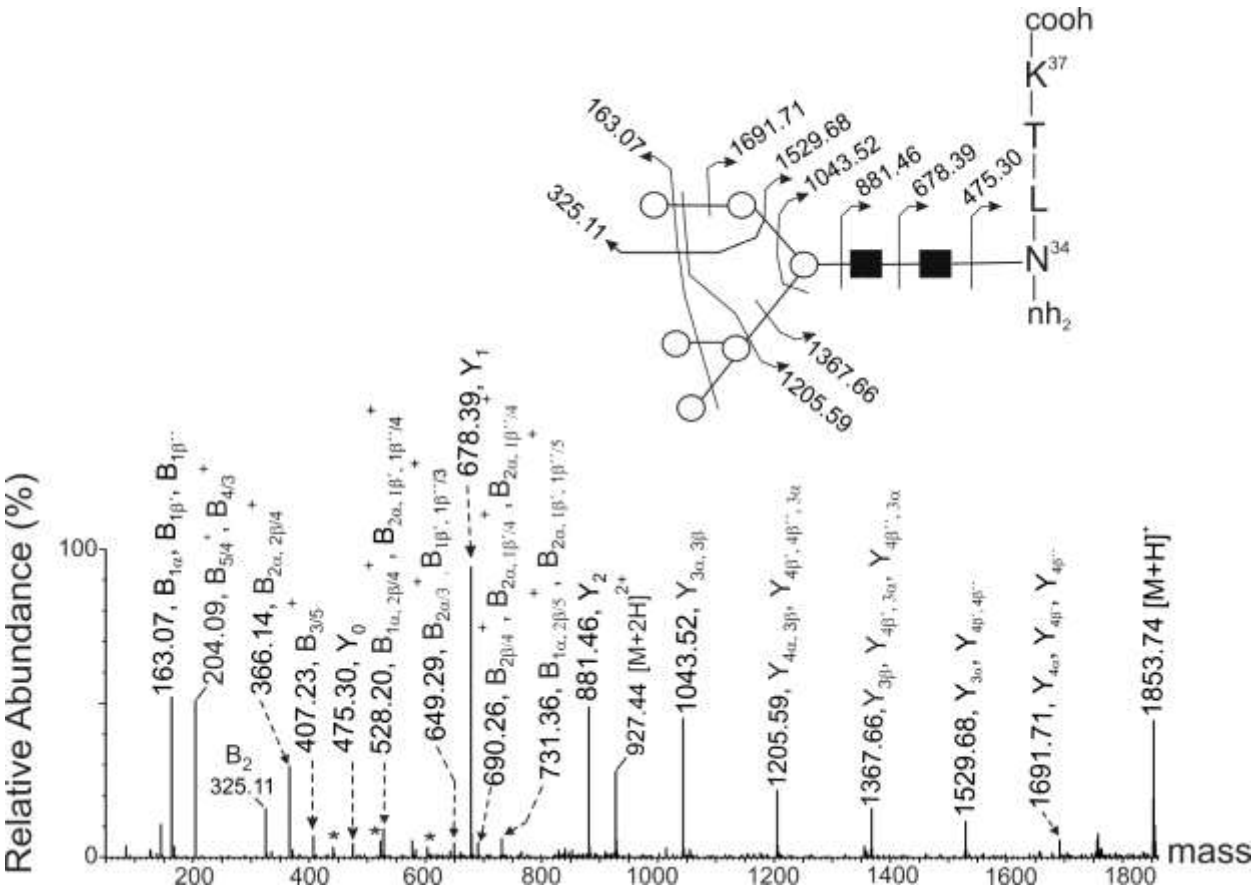

Figure S5. Deconvoluted ESI-MS/MS spectrum of a hydrophilic N-glycopeptide (^{34}NLTK^{37})-$GlcNAc_2Man_6$ (expected *m/z* 927.39, 2+) obtained by the in-solution BFD procedure of the reduced and S-alkylated RNase B. The N-glycan structure is represented according to the nomenclature proposed by Harvey and coworkers[5]. The nomenclature used for the fragment ions is in agreement with the one proposed by Domon and Costello[6]. Signals labeled with asterisks are artefacts of the MaxEntropy software and correspond to multiply-charged ions not completely transformed into singly-charged ions in the deconvolution process.

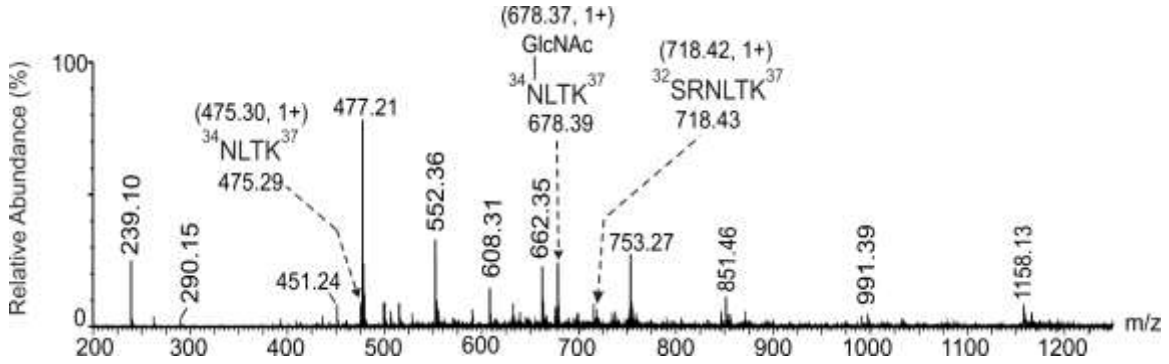

Figure S6. ESI-MS spectrum (*m/z* 200-1250) of the in-solution BFD procedure of RNase B deglycosylated with Endo H (see experimental section). The arrows indicate signals assigned to hydrophilic peptides confirming the partial occupancy of the Asn^{34} N-glycosylation site. Man and GlcNAc represent mannose and N-acetylglucosamine residues, respectively. Between parentheses the expected *m/z* values are indicated.

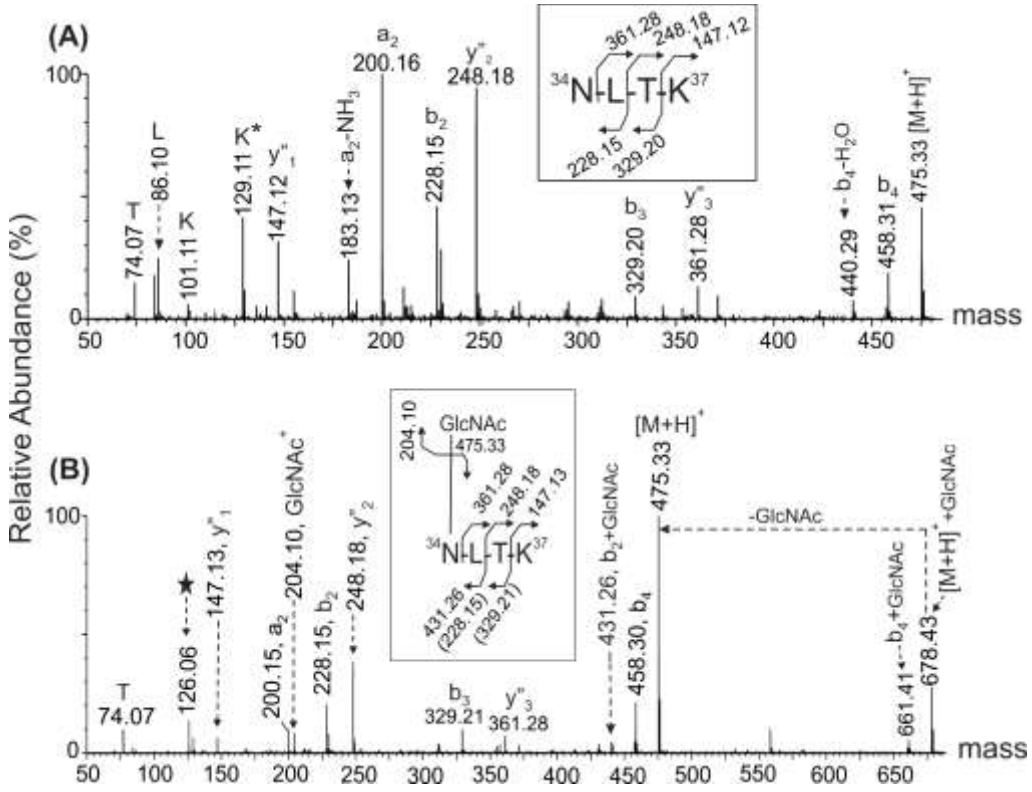

Figure S7. Deconvoluted ESI-MS/MS spectra of two hydrophilic peptides obtained by the in-solution BFD procedure of the RNaseB deglycosylated with Endo H: **(A)** ^{34}NLTK^{37} (expected *m/z* 475.30, 1+) and **(B)** (^{34}NLTK^{37})-GlcNAc (expected *m/z* 678.37, 1+). Both peptides contain the potential N-glycosylation located at Asn^{34}. The ion at *m/z* 204.10 corresponds to the oxonium ion of GlnNAc residue that remains linked to Asn^{34} after Endo H treatment. Star indicates a satellite ion at m/z 126.06 corresponding to a fragment of a GlcNAc residue in agreement with previous work[7].

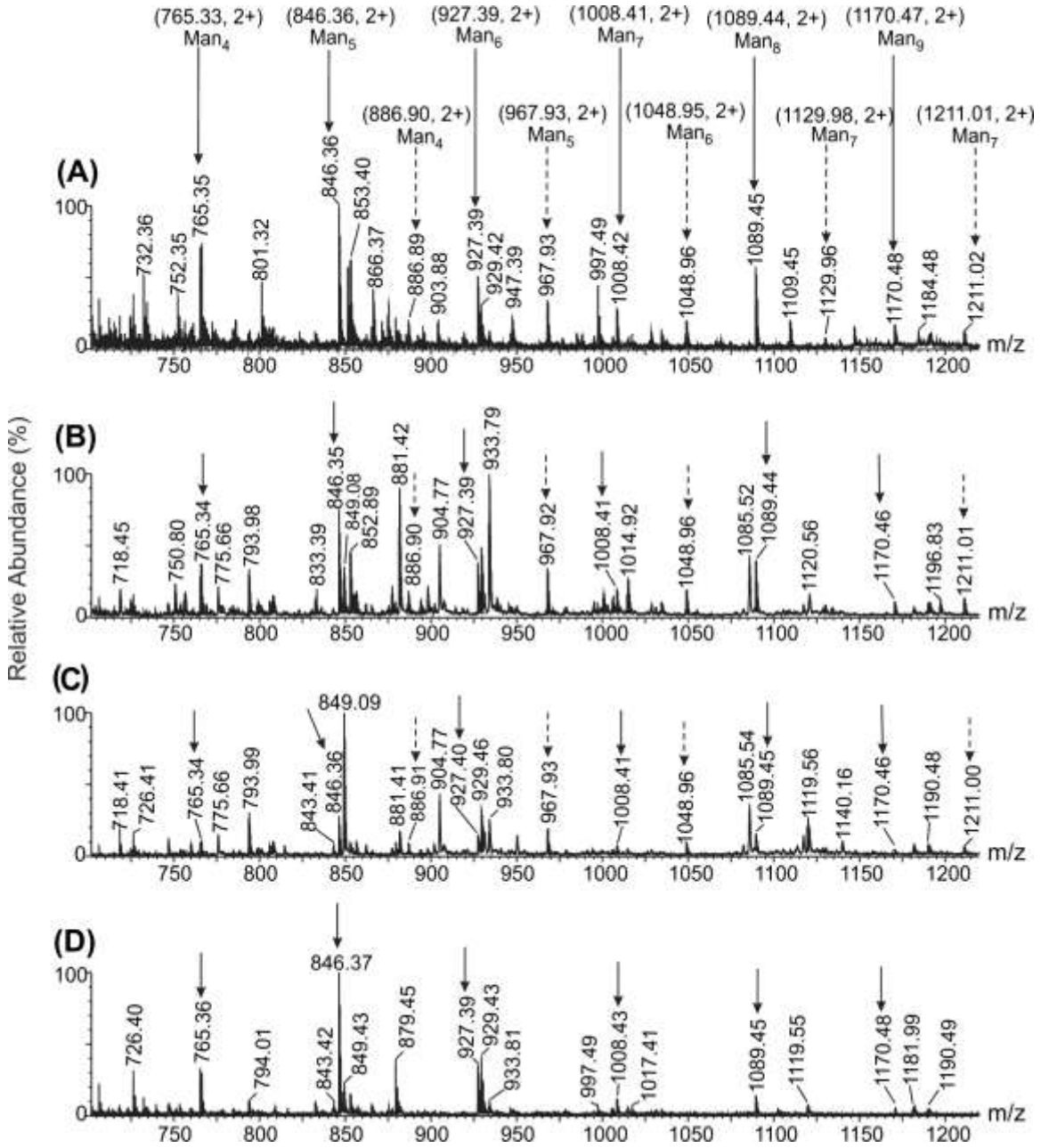

Figure S8. ESI-MS spectra (m/z 700-1220) of the tryptic digestion of S-alkylated and non-deglycosylated RNase B by using the in-solution BFD procedure during 16 h at 37 ^{o}C in presence of **(A)** aqueous solution, **(B)** 20% acetonitrile and **(C)** 50% acetonitrile, except for **(D)** performed in 80% acetonitrile and kept for 1h at room temperature (22 ^{o}C). The arrows shown correspond to signals exclusively assigned to hydrophilic high-mannose N-glycopeptides. Continuous and broken arrows correspond to the tetra- (^{34}NLTK^{37}) and hexa-peptide (^{32}SRNLTK^{37}), respectively linked to GlcNAc and Man residues. Man and GlcNAc represent mannose and N-acetylglucosamine residues and the expected values are indicated within parentheses.

References

[1] N. N. Dookeran, T. Yalcin, and A. G. Harrison. Fragmentation Reactions of Protonated a-Amino Acids. J. Mass Spectrom., 31(5):500–508, 1996. https://doi.org/10.1002/(SICI)1096-9888(199605)31:5<500::AID-JMS327>3.0.CO;2-Q.

[2] M. W. Forbes, R. A. Jockusch, A. B. Young, and A. G. Harrison. Fragmentation of protonated dipeptides containing arginine. Effect of activation method. J. Am. Soc. Mass. Spectrom., 18(11):1959–1966, 2007. https://doi.org/10.1016/j.jasms.2007.08.003.

[3] P. Y. I. Shek, J. Zhao, Y. Ke, K. W. M. Siu, and A. C. Hopkinson. Fragmentations of Protonated Arginine, Lysine and Their Methylated Derivatives: Concomitant Losses of Carbon Monoxide or Carbon Dioxide and an Amine. J. Phys. Chem. A, 110(27):8282–8296, 2006. https://doi.org/10.1021/jp055426k.

[4] M. Mormann, J. Eble, C. Schwöppe, R. M. Mesters, W. E. Berdel, J. Peter-Katalinić, and G. Pohlentz. Fragmentation of intra-peptide and inter-peptide disulfide bonds of proteolytic peptides by nanoESI collision-induced dissociation. Anal. Bioanal. Chem., 392(5):831–838, 2008. https://doi.org/10.1007/s00216-008-2258-7.

[5] D. J. Harvey, A. H. Merry, L. Royle, M. P. Campbell, R. A. Dwek, and P. M. Rudd. Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. Proteomics, 9(15):3796–3801, 2009. https://doi.org/10.1002/pmic.200900096.

[6] B. Domon and C. E. Costello. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. Glycoconjugate J., 5(4):397–409, 1988. https://doi.org/10.1007/BF01049915.

[7] J. P. Williams, S. Pringle, K. Richardson, L. Gethings, J. P. C. Vissers, M. De-Cecco, S. Houel, A. B. Chakraborty, Y. Q. Yu, W. Chen, and J. M. Brown. Characterisation of glycoproteins using a quadrupole time-of-flight mass spectrometer configured for electron transfer dissociation. Rapid Commun. Mass Spectrom., 27(21):2383–2390, 2013. https://doi.org/10.1002/rcm.6684.