Supplementary Information

Elucidation of N-/O-glycosylation and site-specific mapping of sialic acid linkage isomers of SARS-CoV-2 human receptor angiotensin converting enzyme 2

Liming Wei[#[a]], Yuning Chen[#[b,c]], Xiaoxiao Feng[[a]], Jun Yao[[a]], Lei Zhang[[a]], Xinwen Zhou[[a]], Guoquan Yan[[a]], Hong Qiu[[b,c]], Chunhe Wang*[[b,c]], Haojie Lu*[[a]]

[a] Institutes of Biomedical Sciences & Department of Chemistry, Fudan University, 131 Dongan Road, 20032 Shanghai (China)

E-mail: luhaojie@fudan.edu.cn; wangc@simm.ac.cn

[b] Shanghai Institute of Materia Medica, Chinese Academy of Sciences Department, 555 Zuchongzhi Road, 201203 Shanghai(China)

[c] School of Pharmacy, University of Chinese Academy of Sciences, No. 19A Yuquan Road, 100049 Beijing, China

# These authors contributed equally to this work.

Content of SI materials

Supplementary Figures

Supplementary Tables

Supplementary Figures

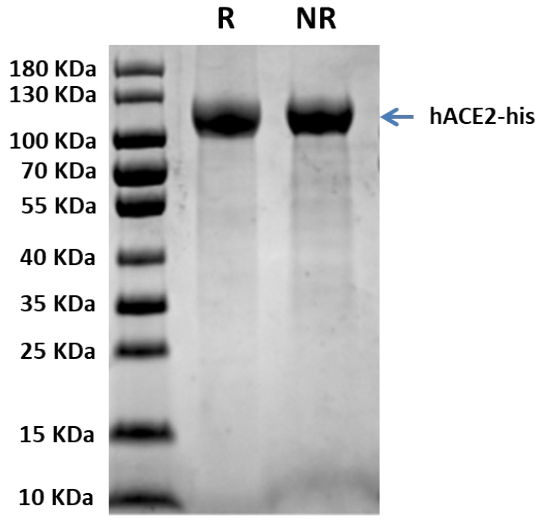
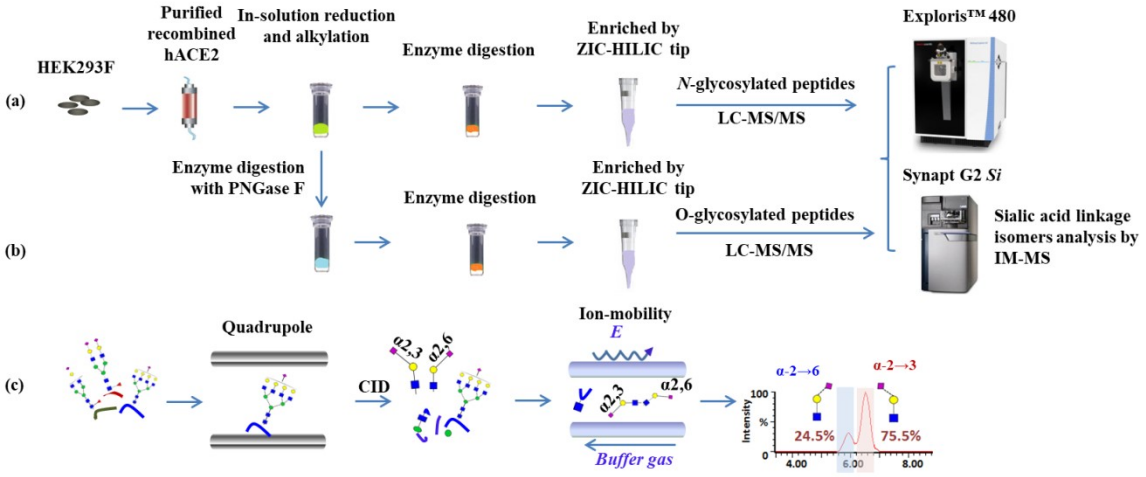


Fig. S1. Gel image of *h*ACE2 protein. SDS-PAGE gel of *h*ACE2 expressed in HEK293F cells. The band of the *h*ACE2 protein is indicated by blue arrow. R and NR marked the *h*ACE2 treated with and without the reducing agent (DTT).



Scheme S1. Workflow of the comprehensive profiling of N-/O- glycosylation and site-specific mapping of sialic acid linkage isomers on *h*ACE2. (a) multiple-enzyme digestion of *h*ACE2 and specific enrichment of N-glycopeptides by ZIC-HILIC tips were used for the identification of N-glycopeptides. (b) multiple-enzyme digestion of *h*ACE2 and specific enrichment of O-glycopeptides by ZIC-HILIC tips were used for the identification of O-glycopeptides. (c) LC-IM-MS was used for the site-specific mapping of sialic acid linkage isomers on *h*ACE2[1].

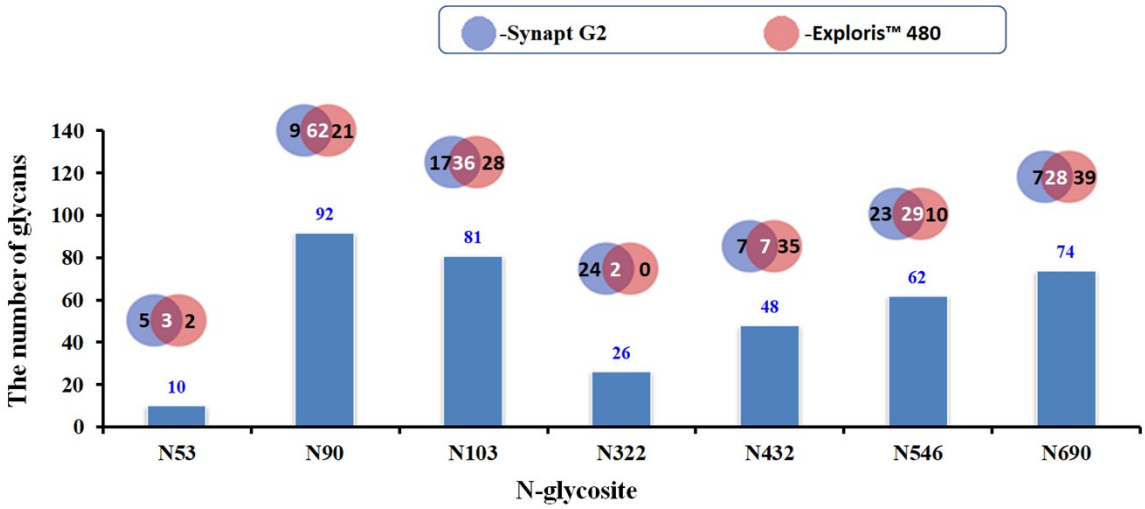


Fig. S2 Site-specific *N*-glycosylation of *h*ACE2 obtained by Synapt G2-*Si* and Exploris™ 480.

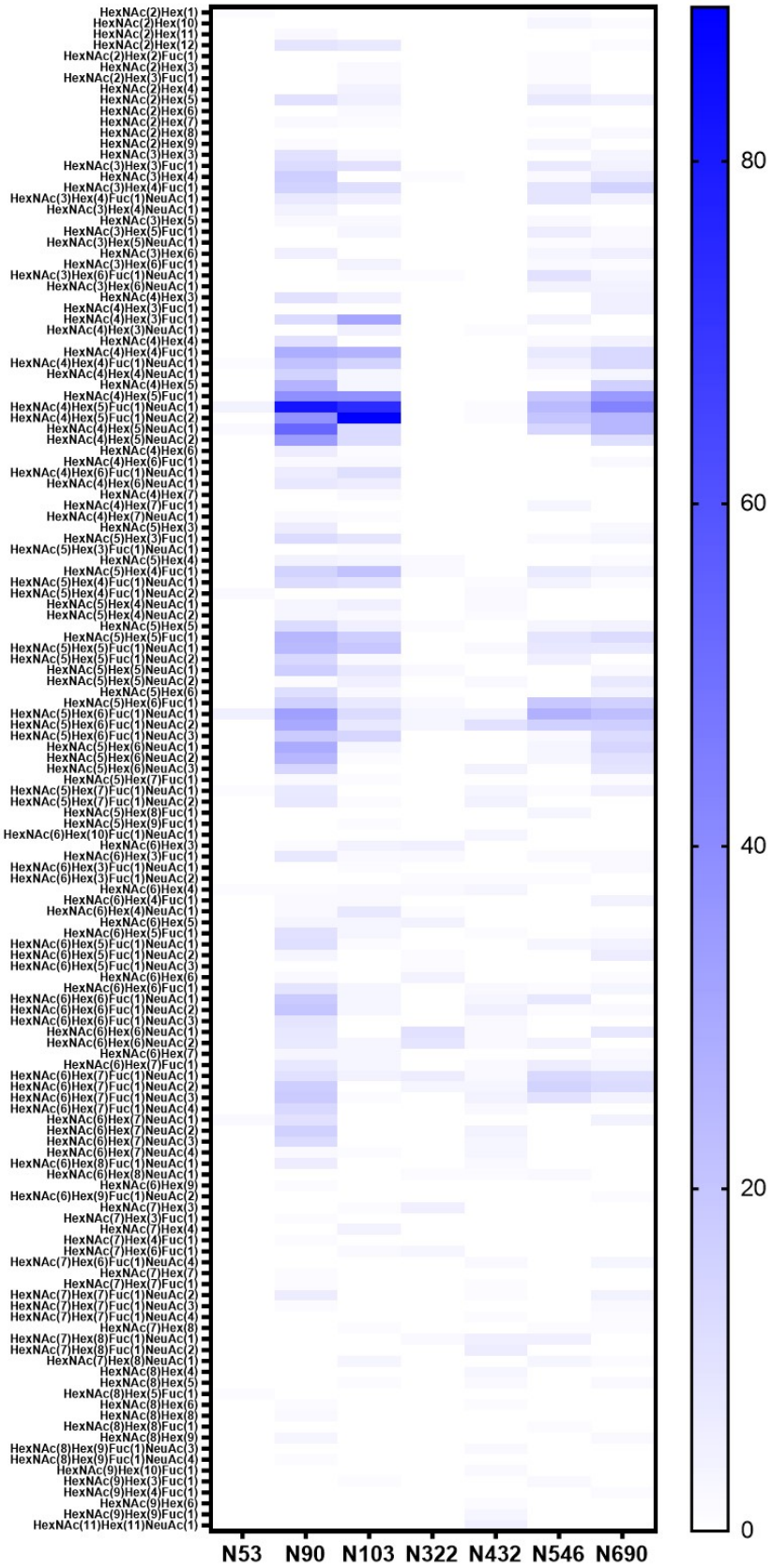


Fig. S3. The distribution and spectra counts of the glycan population present at individual N-linked glycosites.

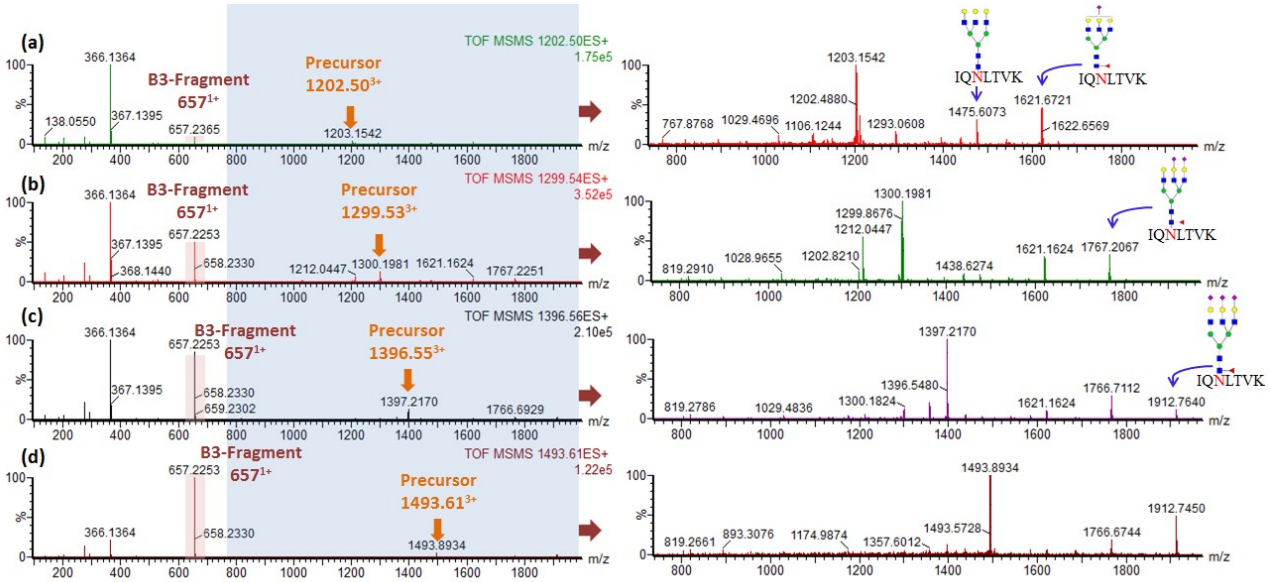


Fig. S4. The CID-IM-MS/MS spectra of the N-glycopeptides from *h*ACE2. N-glycopeptide with the peptide backbone of IQ^{90}**N**$_{\#}$LTVK, (a) N-glycosylated with H7N6F1S1, with the precursor $[M+3H]^{3+}$ at m/z 1202.50 (orange), (b) N-glycosylated with H7N6F1S2, with the precursor $[M+3H]^{3+}$ at m/z 1299.53 (orange), (c) N-glycosylated with H7N6F1S3, with the precursor$[M+3H]^{3+}$at m/z 1396.55 (orange), and (d) N-glycosylated with H7N6F1S4, with the precursor$[M+3H]^{3+}$ at m/z 1493.61 (orange). B_3-trisaccharide fragments (m/z, 657, red) directly cleaved from the intact glycosylated peptides by CID. In the magnification box of MS/MS spectrum, the four spectra have the same pattern with the same masses of the peptides+glycan fragment ions used for the intact glycosylated peptides identification.

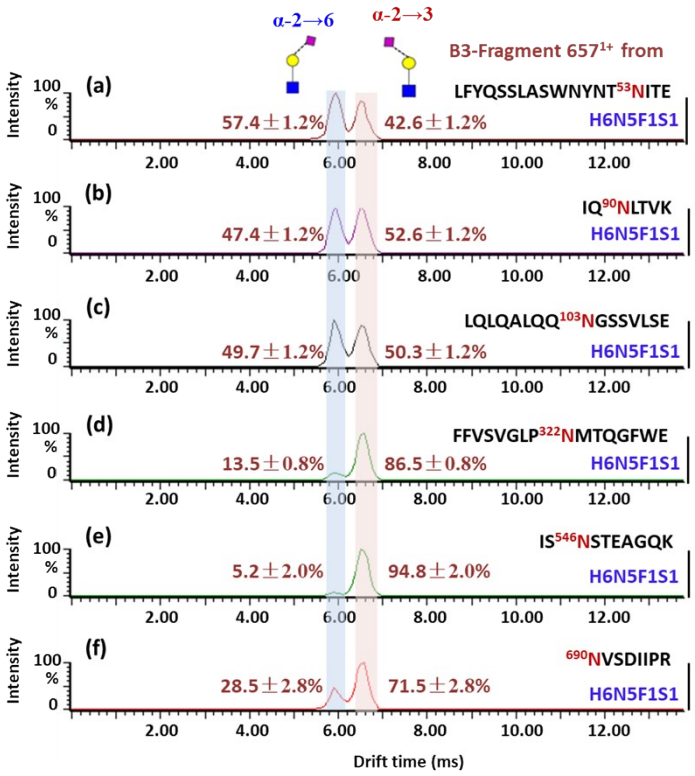


Fig. S5 ATDs of B_3-trisaccharide fragments (m/z 657) from different intact N-glycopeptides of *h*ACE2. These glycosylated peptides own triantennary glycan of H6N5F1S1. Peptides and their glycan compositions are indicated in each panel. Relative percentages of different sialic acid linkages are indicated near each peak.

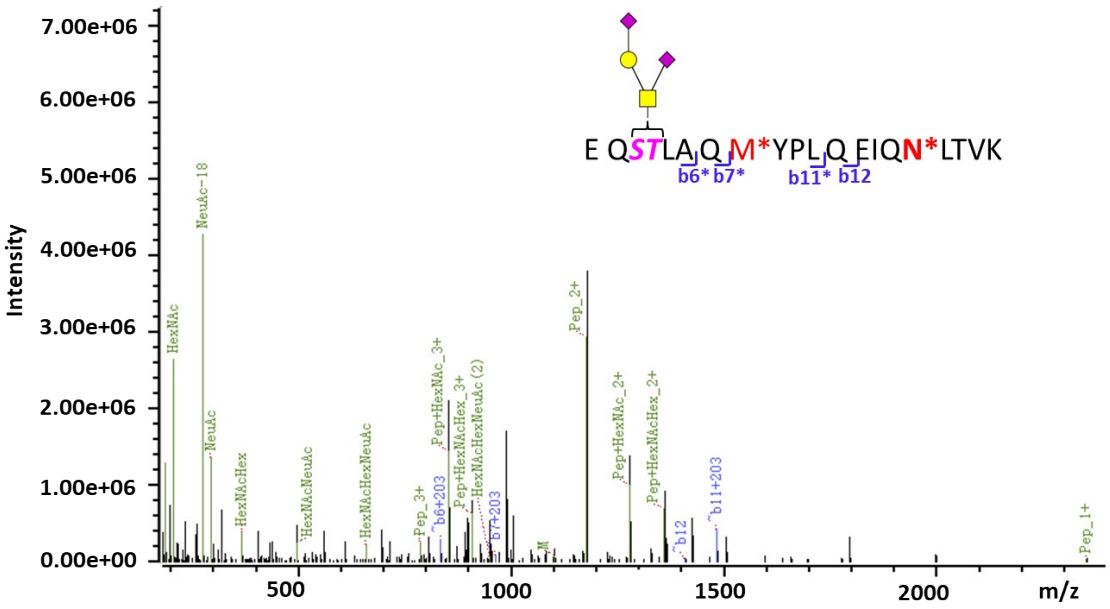


Fig. S6. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of EQ[77]S[78]TLAQM*YPLQEIQN*LTVK from *h*ACE2. O-glycosylated with HexNAc(1)Hex(1)NeuAc(2), with the precursor $[M+2H]^{2+}$ at m/z 1100.176 (scores: 243.5').

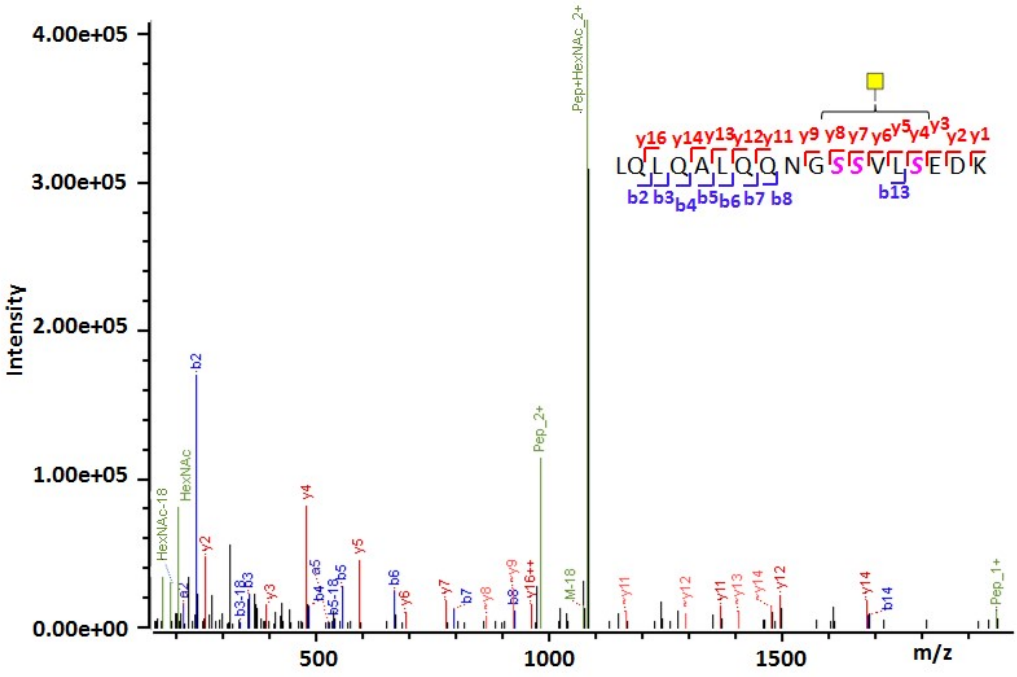


Fig. S7. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of LQLQALQQNG[105]S[106]SVL[109]SEDK from *h*ACE2. O-glycosylated with HexNAc(1), with the precursor $[M+2H]^{2+}$ at m/z 1081.0458 (scores: 740.45').

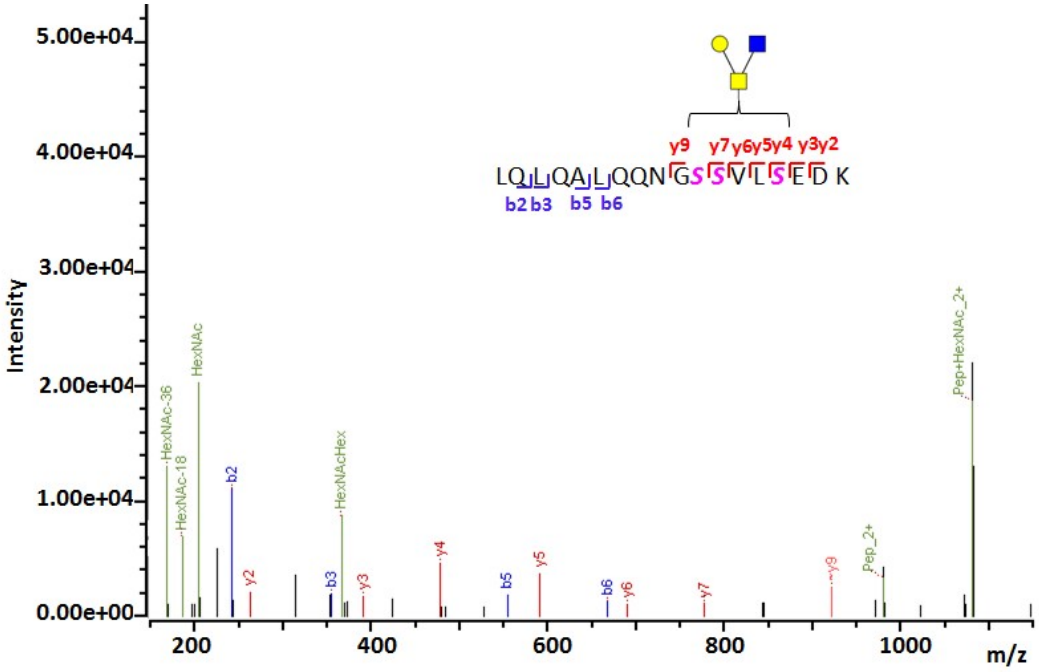


Fig. S8. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of LQLQALQQNG^{105}S^{106}SVL^{109}SEDK from *h*ACE2. O-glycosylated with HexNAc(1)Hex(1), with the precursor $[M+2H]^{2+}$ at m/z 842.744 (scores: 473.9').

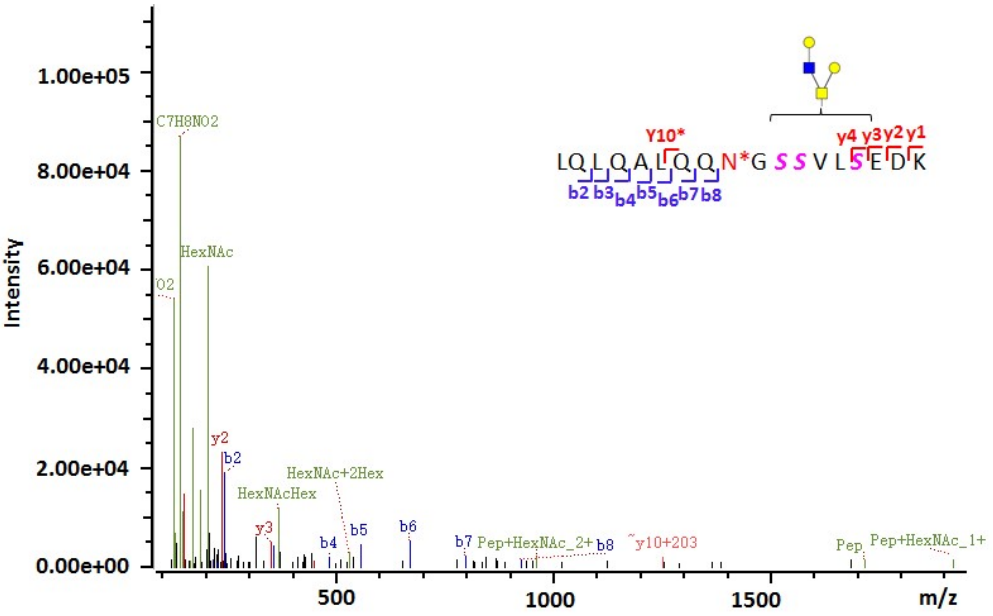


Fig. S9. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of LQLQALQQNG^{105}S^{106}SVL^{109}SEDK from *h*ACE2. O-glycosylated with HexNAc(2)Hex(2), with the precursor $[M+2H]^{2+}$ at m/z 815.7224 (scores: 326.6').

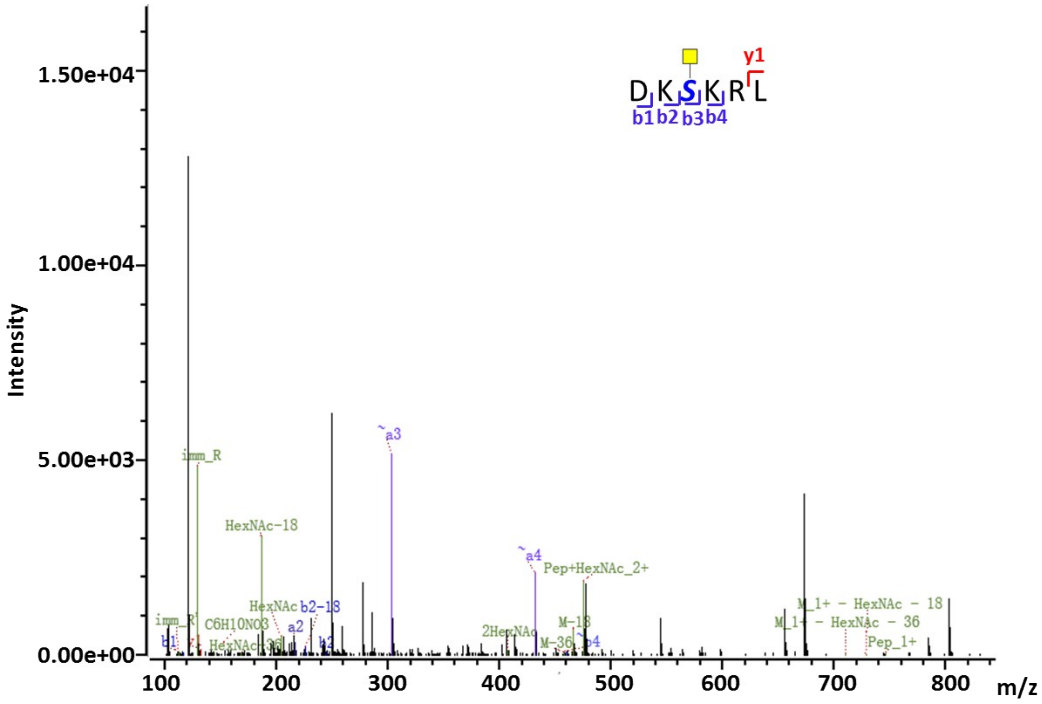


Fig. S10. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of DK[113]SKRL from *h*ACE2. O-glycosylated with Hex(1), with the precursor $[M+2H]^{2+}$ at m/z 475.266 (scores: 157.5').

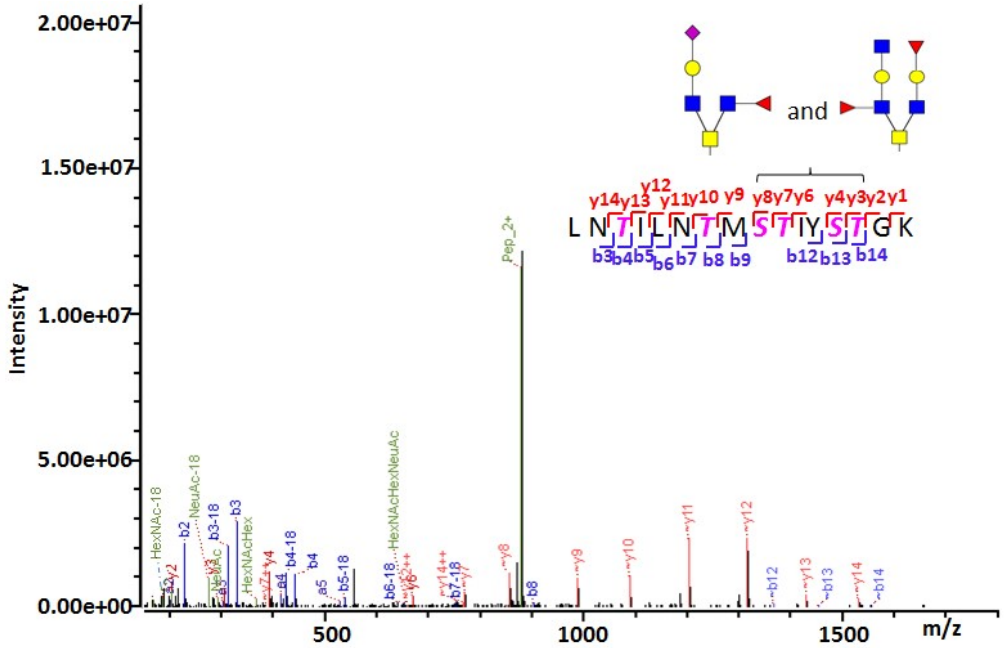


Fig. S11. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of LN[118]TILN[122]TM[124]S[125]TIY[128]S[129]TGK from *h*ACE2. O-glycosylated with HexNAc(3)Hex(1)Fuc(1)NeuAc(1) and HexNAc(4)Hex(2)Fuc(2), with the precursor $[M+5H]^{5+}$ at m/z 879.5829 (scores: 550.5').

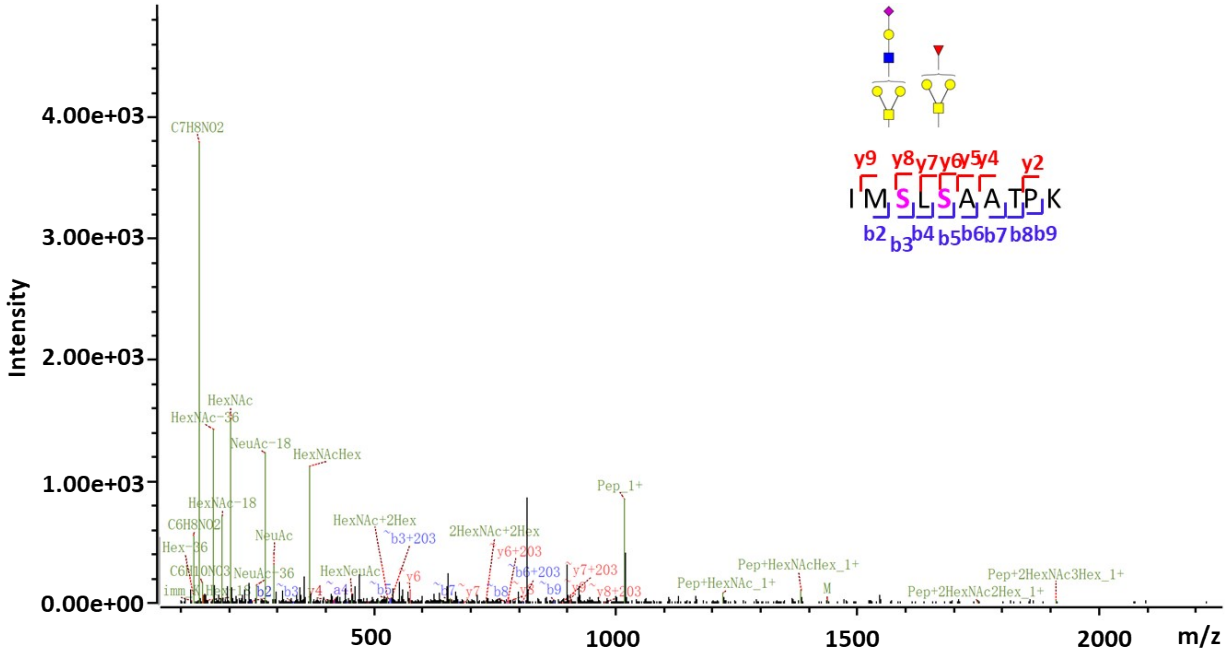


Fig. S12. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of IM[409]SL[411]SAATPK from *h*ACE2. O-glycosylated with HexNAc(3)Hex(5)Fuc(1)NeuAc(1), with the precursor $[M+2H]^{2+}$ at m/z 1438.12 (scores:203.7').

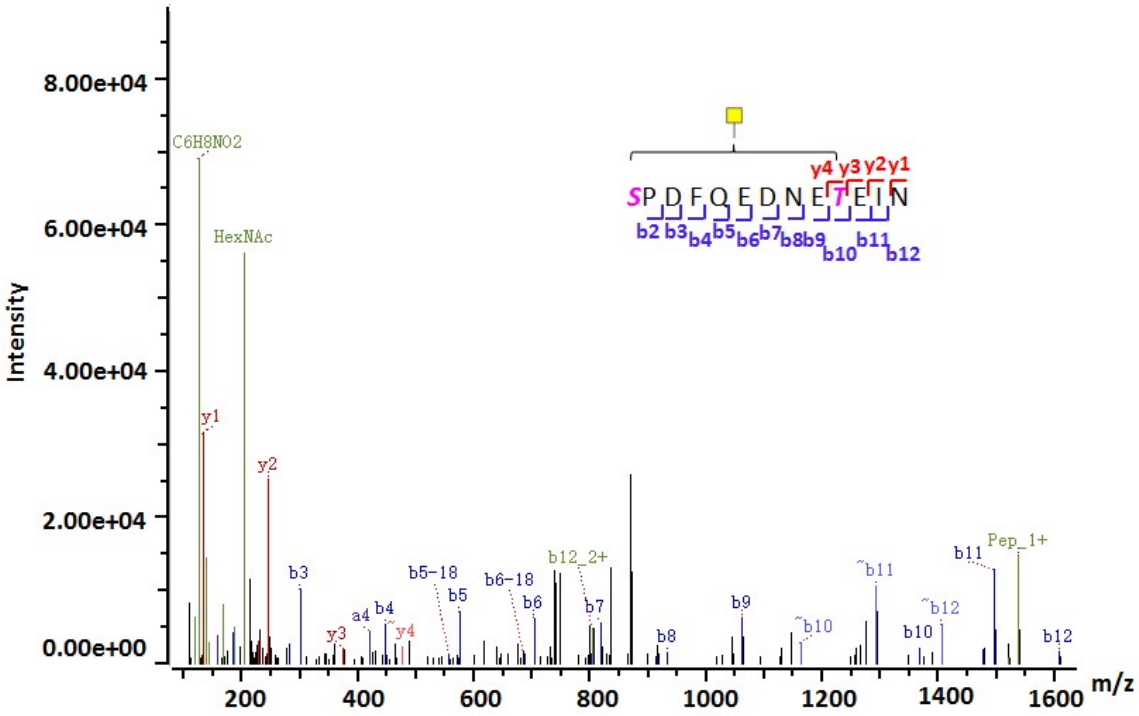


Fig. S13. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of [425]SPDFQEDNE[434]TEIN from *h*ACE2. O-glycosylated with HexNAc(1), with the precursor $[M+2H]^{2+}$ at m/z 870.8521 (scores:451.1').

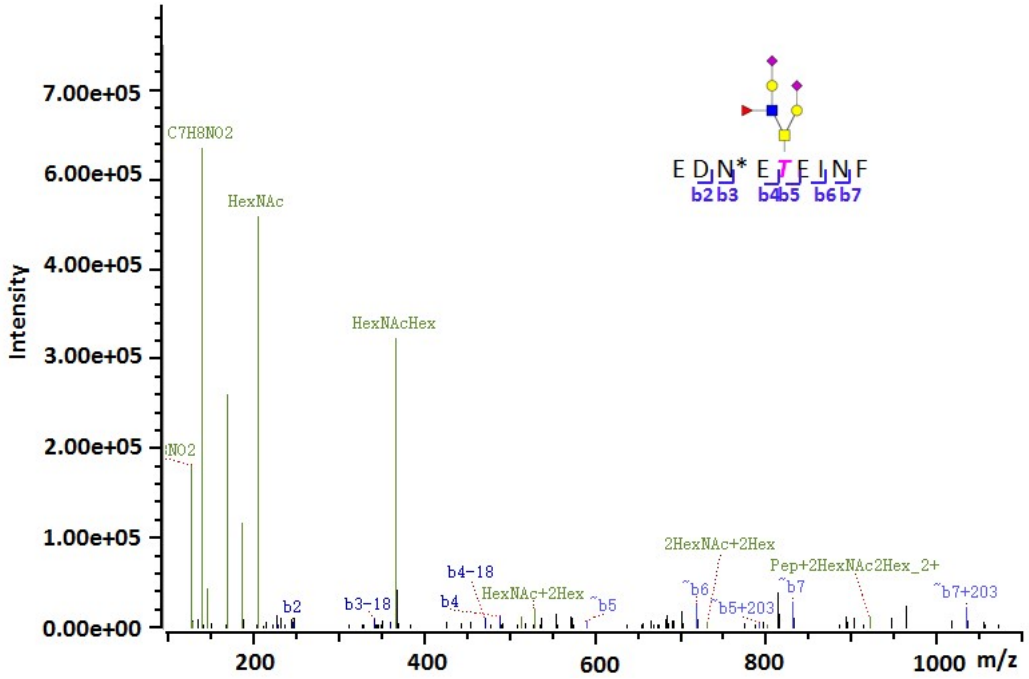


Fig. S14. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of EDNE^{434}TEINF from *h*ACE2. O-glycosylated with HexNAc(2)Hex(2)Fuc(1)NeuAc(2), with the precursor $[M+2H]^{2+}$ at m/z 1285.4823 (scores: 178.5').

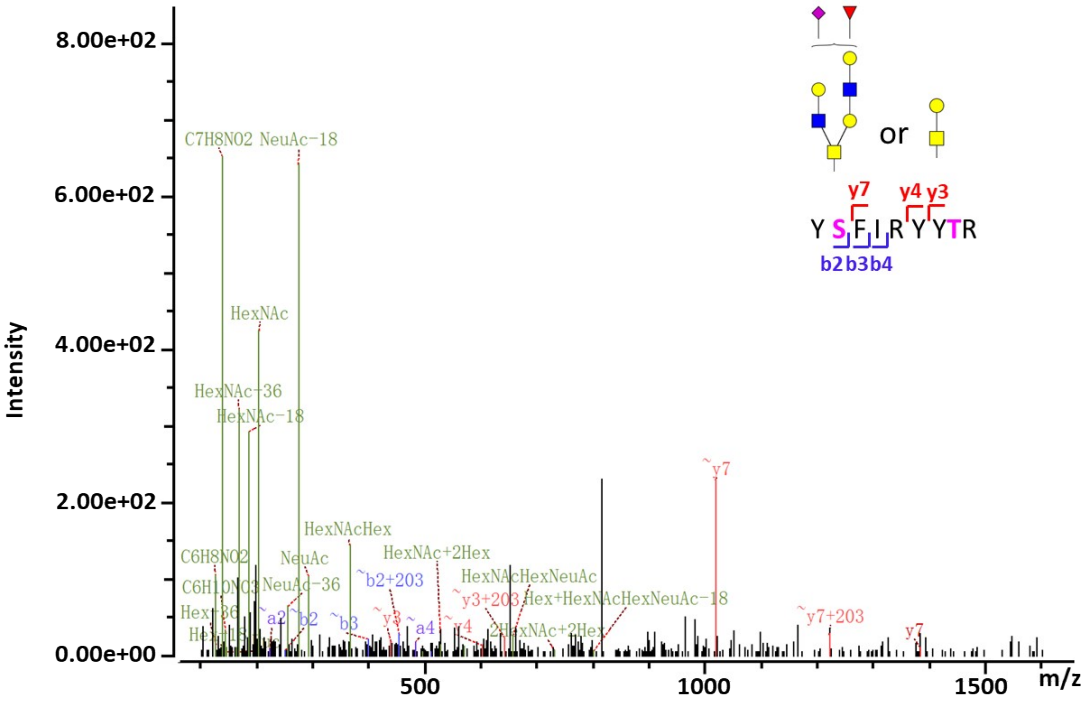


Fig. S15. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of Y^{511}SFIRYY^{517}TR from *h*ACE2. O-glycosylated with HexNAc(4)Hex(4)Fuc(1)NeuAc(1), with the precursor $[M+2H]^{2+}$ at m/z 1583.676 (scores: 153.5').

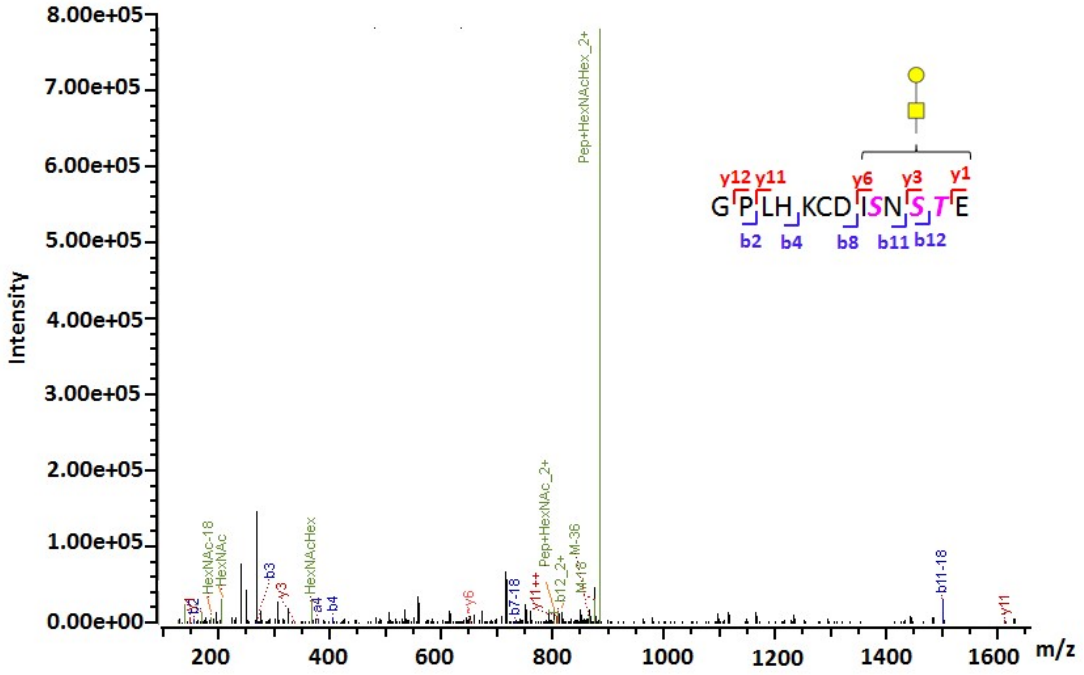


Fig. S16. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of GPLHKCDI[545]SN[547]S[548]TE from *h*ACE2. O-glycosylated with HexNAc(2)Hex(2)Fuc(1)NeuAc(2), with the precursor $[M+2H]^{2+}$ at m/z 1285.4823 (scores: 178.5').

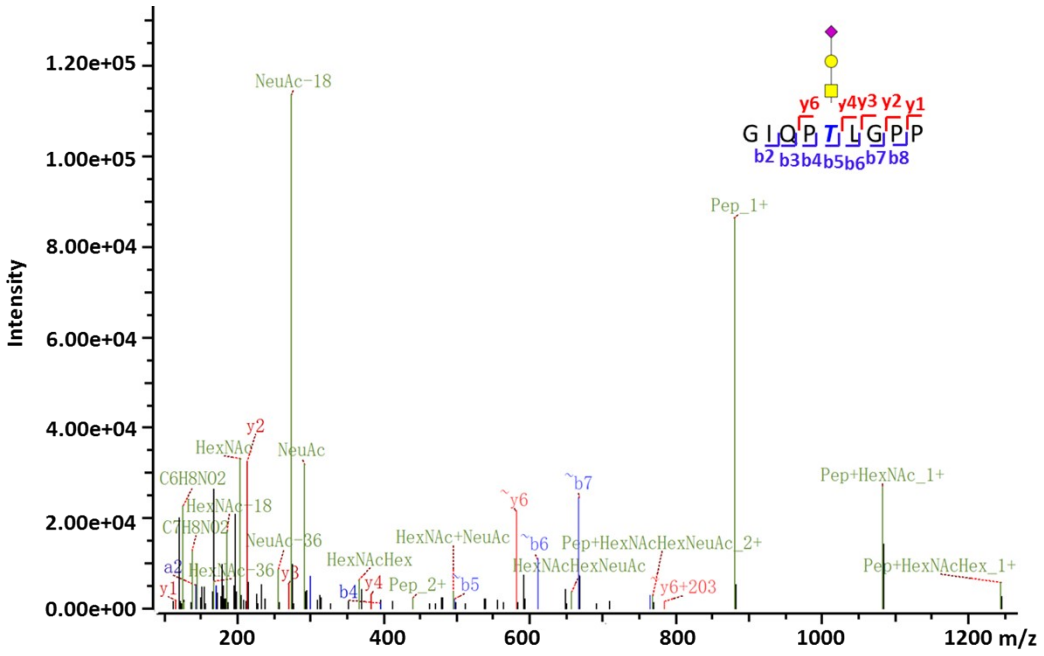


Fig. S17. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of GIQP[730]TLGPP from *h*ACE2. O-glycosylated with HexNAc(1)Hex(1)NeuAc(1), with the precursor $[M+2H]^{2+}$ at m/z 768.3578, scores: 516.44').

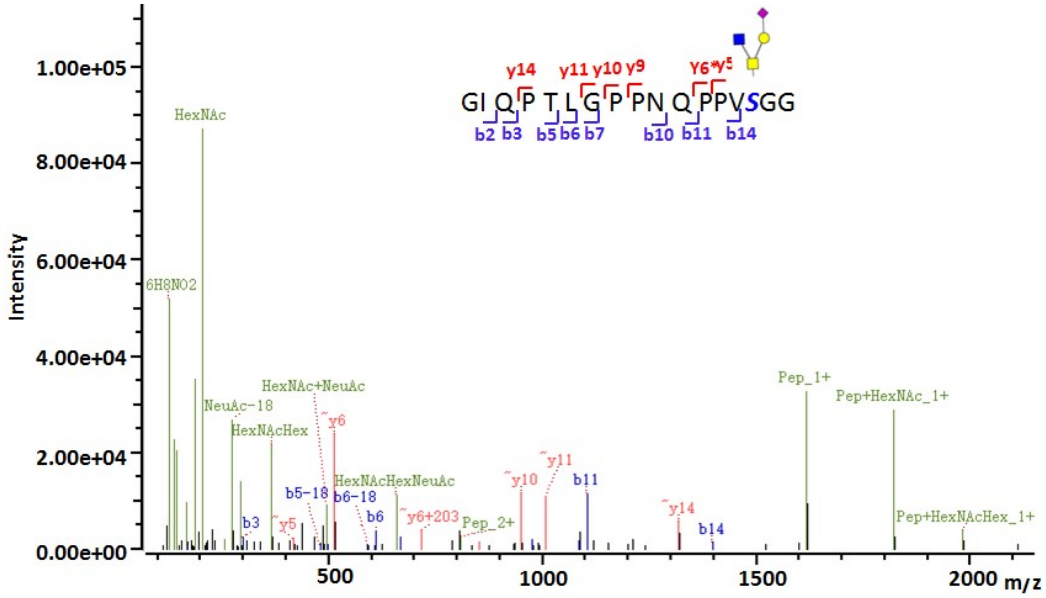


Fig. S18. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of GIQPTLGPPNQPPV[740]SGG from *h*ACE2. O-glycosylated with HexNAc(2)Hex(1)NeuAc(1), with the precursor $[M+2H]^{2+}$ at m/z 1238.0707, scores: 579.9').

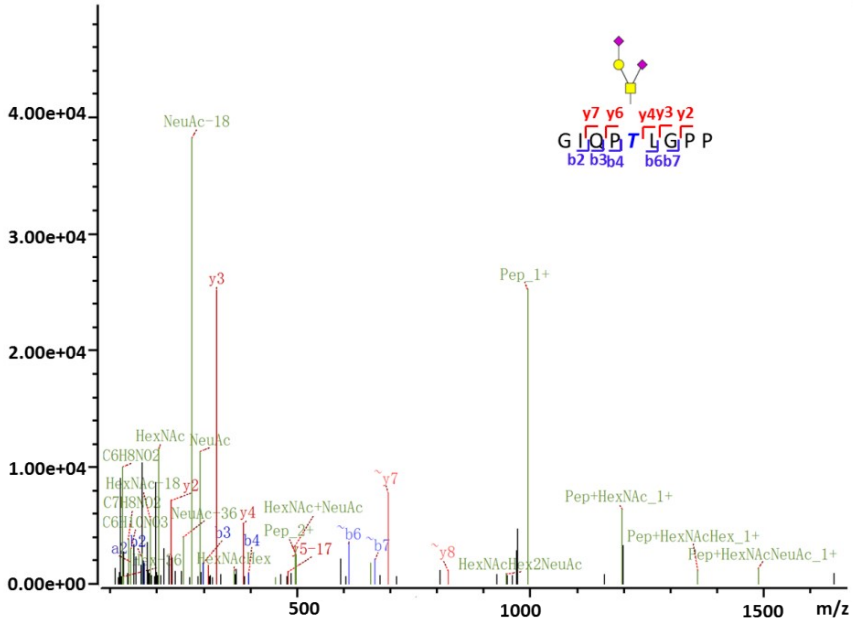


Fig. S19. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of GIQP[730]TLGPP from *h*ACE2. O-glycosylated with HexNAc(1)Hex(1)NeuAc(2), with the precursor $[M+H]^{+}$ at m/z 970.9258, scores: 444.95').

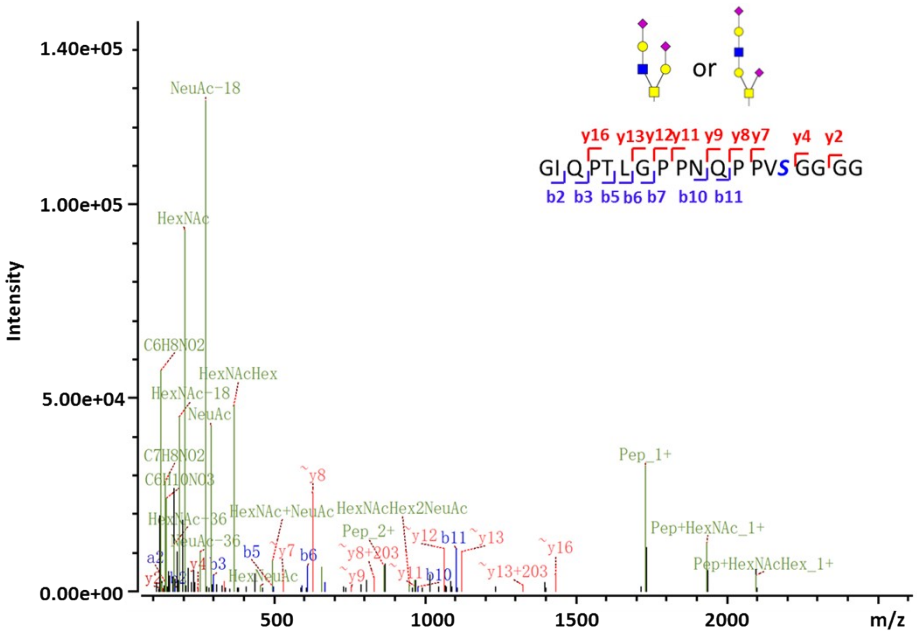


Fig. S20. The MS/MS spectrum of the O-glycopeptides with the peptide backbone of GIQPTLGPPNQPPV^{740}SGG from *h*ACE2. O-glycosylated with HexNAc(2)Hex(2)NeuAc(2), with the precursor $[M+3H]^{3+}$ at m/z 1014.7794, scores: 590.84').

>sp|Q9BYF1|ACE2_HUMAN Angiotensin-converting enzyme 2 OS=Homo sapiens OX=9606 GN=ACE2 PE=1 SV=2

^{18}QSTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNT^{53}NITEENVQNMNNAGDKWSAFLKEQ^{77}S^{78}TLAQ MYPLQEIQ^{90}NLTVKLQLQALQQ^{103}NG^{105}S^{106}SVL^{109}SEDK^{113}SKKRLN^{118}TILN^{122}TM^{124}S^{125}TIY^{128}S^{129}T GKVCNPDNPQECLLLEPGLNEIMANSLDYNERLWAWESWRSEVGKQLRPLYEEYVVLKNEMARANHYED YGDYWRGDYEVNGVDGYDYSRGQLIEDVEHTFEEIKPLYEHLHAYVRAKLMNAYPSYISPIGCLPAHLLGD MWGRFWTNLYSLTVPFGQKPNIDVTDAMVDQAWDAQRIFKEAEKFFVSVGLP^{322}NMTQGFWENSMLT DPGNVQKAVCHPTAWDLGKGDFRILMCTKVTMDDFL^{371}TAHHEMGHIQYDMAYAAQPFLLRNGANEG FHEAVGEIM^{409}SL^{411}SAA^{414}TPKHLKSIGLL^{425}SPDFQED^{432}NE^{434}TEINFLLKQALTIVGTLPFTYMLEKWR WMVFKGEIPKDQWMKKWWEMKREIVGVVEPVPHDETYCDPASLFHVSNDY^{511}SFIRYY^{517}TRTLYQFQF QEALCQAAKHEGPLHKCDI^{545}S^{546}N^{547}S^{548}TEAGQKLFNMLRLGKSEPWTLALENVVGAKNMNVRPLLN YFEPLFTWLKDQNKNSFVGWSTDWSPYADQSIKVRISLKSALGDKAYEWNDNEMYLFRSSVAYAMRQYFL KVKNQMILFGEEDVRVANLKPRISFNFFVTAPK^{690}NVSDIIPRTEVEKAIRMSRSRINDAFRLNDNSLEFLGI QP^{730}TLGPPNQPPV^{740}SGGGGSHHHHHH

Fig. S21. Site-specific N- and O-glycosylation characterization of *h*ACE2. High-confidence O-glycosites identified using typsin/Glu-C (T/G) or proteinase K in three replicates. Mapping of unambiguously identified O-glycosites (S113, T371, T 434, T730 and T740) highlighted with blue, others O-glycosites (S77/T78, S105/S106/S109, T118/T122/S124/T125/S129/T130, S409/S411/T414, S425, S511/T517 and S545/S547/T548) highlighted with pink.

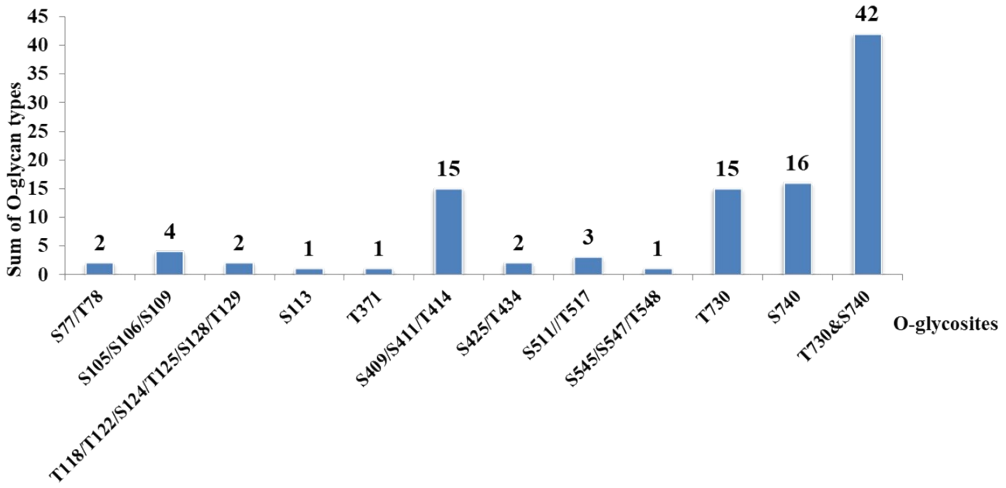


Fig. S22. Site-specific O-linked glycans of *h*ACE2 obtained by LC-MS/MS.

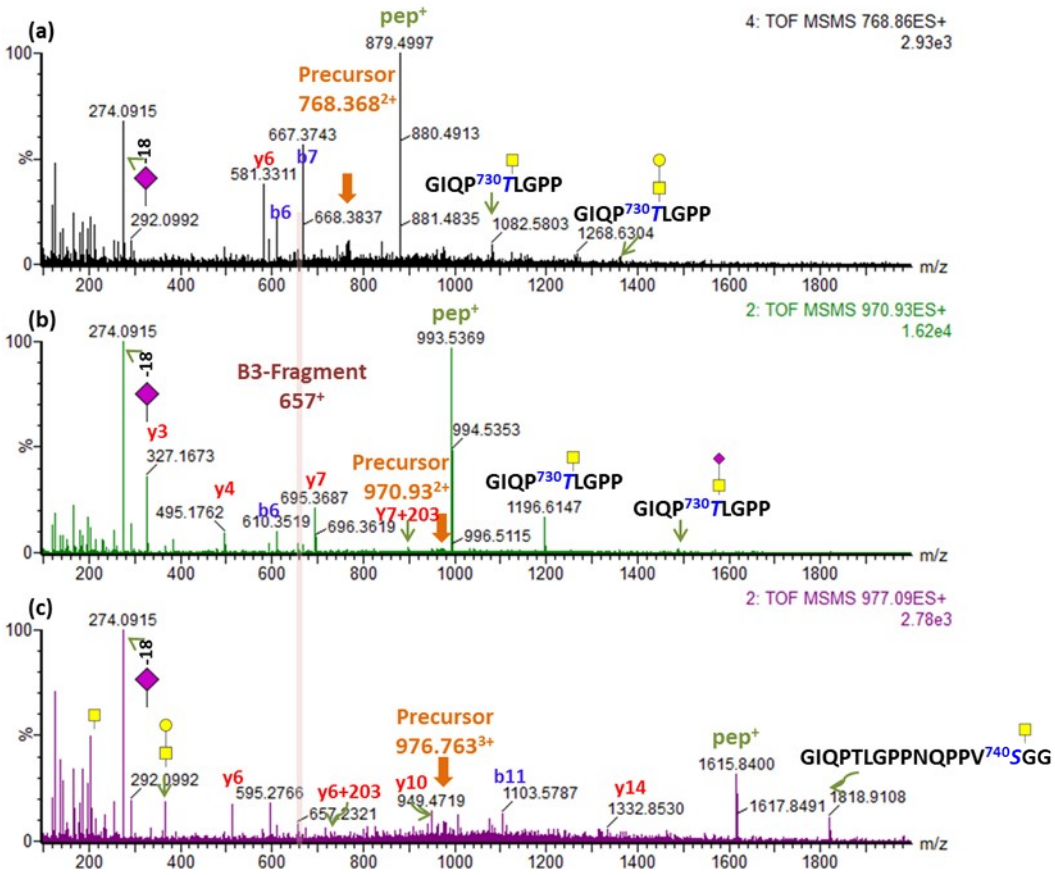


Fig. S23. The CID-IM-MS/MS spectra of the O-glycopeptides from *h*ACE2. O-glycopeptide with the peptide backbone of GIQP[730]TLGPP, (a) O-glycosylated with H1N1S1, with the precursor $[M+2H]^{2+}$ at m/z 768.368 (orange), (b) O-glycosylated with H1N1S2, with the precursor $[M+2H]^{2+}$ at m/z 970.93 (orange). O-glycopeptide with the peptide backbone of GIQPTLGPPNQPPV[740]SGGGG (c) O-glycosylated with H2N2S2, with the precursor $[M+3H]^{3+}$ at m/z 976.76 (orange). B_3-trisaccharide fragments (m/z, 657.24) directly cleaved from the intact glycosylated peptides by CID.

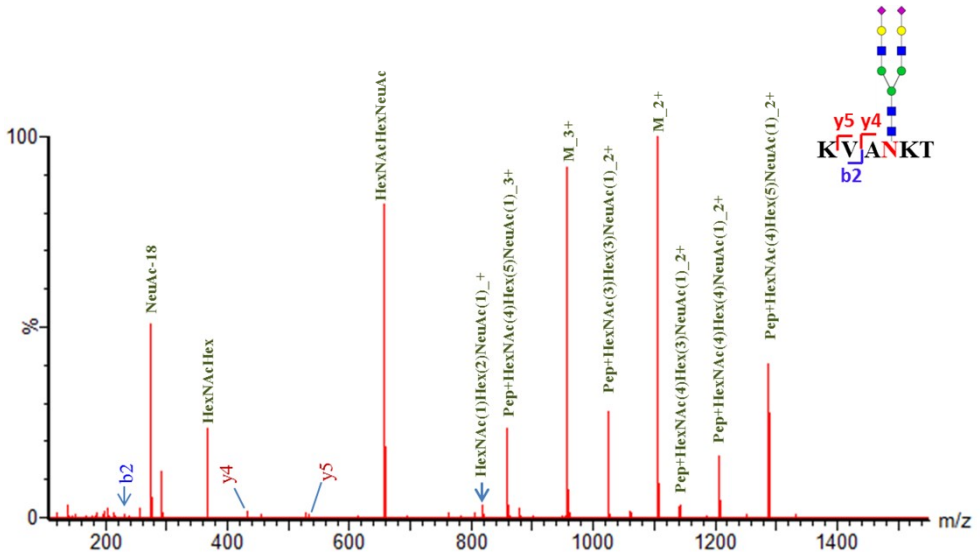


Fig. S24. The mass spectrum of the standard N-glycopeptides with the peptide backbone of KVAN#KT modified with the N-glycans of H5N4S2, with the precursor $[M+3H]^{3+}$ at m/z 955.7249).

Table S1. Identified N-glycopeptides and N-linked glycans on *h*ACE2. **Table S1a**, intact N-glycopeptides identified by Orbitrap Exploris™ 480. **Table S1b**, intact N-glycopeptides identified by Synapt G2-*Si*.

Table S2. Glycoform abundances observed across *h*ACE2 protein. The relative abundances of all N-linked glycans detected at each site are displayed by the spectra counts, in which were showed in Fig. S3. The likely N-linked glycan structures were followed in the Table S2.

Table S3. Glycoform abundances observed across *h*ACE2 protein. The upper table shows the categorized glycan compositions at each N-linked glycan site. The total averages are shown in the right-hand table. The lower table further categorizes the glycan compositions into high-mannose-, hybrid-, and complex-type as well as the percentage of glycan compositions containing at least one fucose or one sialic acid residue. See also Figure 2.

Table S4. Pattern and relative quantification of sialic acid linkage isomer of N-glycopeptides from *h*ACEs by LC-CID-IM-MS analysis.

Table S5. Identified O-glycopeptides and O-linked glycans of *h*ACE2. **Table S5a**, intact O-glycopeptides identified by Orbitrap Exploris™ 480. **Table S5b**, intact O-glycopeptides identified by Synapt G2-Si.

Table S6. The O-linked glycan composition assigned to the 12 O-glycosites. The relative abundances of all O-linked glycans detected at each site are displayed by the spectra counts. The likely O-linked glycan structures were followed in the Table S6.

Table S7. The distribution and relative abundance of O-glycopeptides and O-linked glycans at T730, S740 and T730/S740, of which were showed in Figure 6.

Author Contributions

All authors have accepted responsibility for the entire content of this manuscript and approved its submission. L.W., Y. C., H. Q., C. W. and H. L. designed research; L.W., Y. C., X. F., J. Y., G. Y. and J. Z. performed research; L.W., L. Z., H. Q., C. W. and H. L. analyzed data; L. W., Y. C., X. F., J. Y., L. Z., G. Y., J. Z., H. Q., C. W., and H. L. wrote the paper.

1. X. Feng, H. Shu, S. Zhang, Y. Peng, L. Zhang, X. Cao, L. Wei and H. Lu, *Analytical chemistry*, 2021, **93**, 15617-15625.