

Supporting information for

Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2

Spike Proteins

Yong Zhang^{1,2†}, Wanjun Zhao^{3†}, Yonghong Mao⁴, Yaohui Chen⁴, Shisheng Wang^{1,2}, Yi Zhong¹, Tao Su¹, Meng Gong^{1,2}, Dan Du¹, Xiaofeng Lu^{1,2}, Jingqiu Cheng^{1,2*}, Hao Yang^{1,2*}

¹Key Laboratory of Transplant Engineering and Immunology, MOH; Institutes for Systems Genetics; Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu 610041, China.

²Regenerative Medical Research Center, West China Hospital, Sichuan University, Chengdu 610041, China.

³Department of Thyroid Surgery, West China Hospital, Sichuan University, Chengdu 610041, China.

⁴Department of Thoracic Surgery, West China Hospital, Sichuan University, Chengdu 610041, China.

†These authors contributed equally to this work.

Running title: N-glycosylation of SARS-CoV-2 Spike Protein

***Corresponding Authors:** Hao Yang, PhD, Associate Researcher, Key Lab of Transplant Engineering and Immunology, West China Hospital, Sichuan University

Jingqiu Cheng, PhD, Professor, Key Lab of Transplant Engineering and Immunology, West China Hospital, Sichuan University

Address: No. 1, Keyuan 4th Road, Gaopeng Avenue, Hi-tech Zone, Chengdu 610041, China.
Phone: +86-28-85164031; Fax: +86-28-85164031; E-mail: yanghao@scu.edu.cn

Supplementary Figures and Tables

Supplementary Figure S1. The theoretical of intact N-glycopeptides of S protein derived from the digestion using trypsin (A) or Glu-C (B) alone or in combination (C). Red letter: N-glycosites; Green letter: trypsin cutting sites; Blue letter: Glu-C cutting sites; Underline: theoretical N-glycopeptides without missing cleavage sites.

Supplementary Figure S2. The spectrum of intact N-glycopeptides with the ambiguously assigned N-glycosite (N1134).

Supplementary Figure S3. Comparison of the spectra of the intact N-glycopeptide (N149) before (A) and after (B) enrichment.

Supplementary Figure S4. Comparison of the spectra of intact N-glycopeptides (N709 and N717) after trypsin digestion (A) and Glu-C digestion (B).

Supplementary Figure S5. Comparison of the spectra of intact N-glycopeptides (N709 and N717) (A) and deglycopeptides (B).

Supplementary Figure S6. Comparison of the N-glycosites on the SARS-CoV-2 and SARS-CoV spike proteins.

Supplementary Figure S7. Representative and high-quality spectra of intact N-glycopeptides and deglycosylated peptides.

Supplementary Figure S8. Quantitative analysis of site-specific N-glycosylation of S protein expressed in insect cells (top 5 N-glycans).

Supplementary Figure S9. Quantitative analysis of site-specific N-glycosylation of S protein expressed in human cells (top 5 N-glycans).

Supplementary Figure S10. Microheterogeneity and macroheterogeneity of the N-linked glycopeptides of the S protein.

Supplementary Figure S11. SDS-PAGE analysis of RBDs expressed in insect and human cells.

Table S1. Glycoproteomic identification of the N-glycosites on recombinant SARS-CoV-2 S protein expressed in insect cells.

Table S2. Glycoproteomic identification of the N-glycosites on recombinant SARS-CoV-2 S protein expressed in human cells.

Table S3. Site-specific N-glycosylation characterization of recombinant SARS-CoV-2 spike protein expressed in insect cells.

Table S4. Site-specific N-glycosylation characterization of recombinant SARS-CoV-2 spike protein expressed in human cells.

Table S5. Mutation frequency of SARS-CoV-2 spike protein.

Table S6. Intact N-glycopeptides of the recombinant RBD of the SARS-CoV-2 spike protein.

Table S7. A summary of N-glycans on SARS-CoV-2 spike protein expressed in insect and human cells

Table S8. A detailed comparison of spike protein N-glycosylation among different studies

Table S9. Quantitative analysis of recombinant SARS-CoV-2 spike proteins

Supplementary Figure S1. The theoretical of intact N-glycopeptides of S protein derived from the digestion using trypsin (A) or Glu-C (B) alone or in combination (C). Red letter: N-glycosites; Green letter: trypsin cutting sites; Blue letter: Glu-C cutting sites; Underline: theoretical N-glycopeptides without missing cleavage sites.

A

Theoretical N-glycopeptides of S protein digested using trypsin

VNLTTRTQLPPAYTNSFTRGVVYYPDKVFRSSVLHSTQDLFLPFFSNVTFWHAIHVS~~GT~~NGTKRFDNPVLPFNDGVYFASTEKSNIIIRGWIFGTTL
 DSKTQSL~~LIVN~~NATNVVIKVC~~E~~FQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE~~G~~KQGNFKNLREFVFKNIDGYFKI
 YSKHTPINLVRDLPPQGFSAL~~E~~PLVDLPIGINITRFQTLALHRSYLT~~P~~GDSSSGWTAGAAAYVGYLQPRTFLLKYNE~~NG~~TITDAVDCALDPLSE~~T~~
 KCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN~~DL~~CFTN
 VYADSFVIRGDEV~~R~~QIAPGQTGKIADYNYKLPDDFTGCVIAWNSN~~LD~~SKVGGNYNYLYR~~L~~FRKSNLKPFE~~R~~DISTEIYQAGSTPCNGVEGFNC
 YFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQFG~~R~~DIADTTDAVRDPQT
 LEILDITPCSF~~GG~~VSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTG~~S~~NVFTQTRAGCLIGAEHV~~N~~NSYEC~~D~~PIGAGICASYQTQ
 TNSPRRARSVASQSIAYTMSLGAENSVAYS~~N~~NSIAIPTNFTISVTEILPVSM~~T~~KTSVDCTMYICGDSTEC~~S~~NLLQYGSFCTQLNRAL~~T~~GIAVEQ
 DKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSPKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLG~~D~~IAARDLICAQKFNGLTVLPPLLTDE~~M~~IA
 QYTSALLAGTITSGWTFGAGAALQIPFAMQMAYR~~F~~NGIGVTONVLYENQKLIANQFN~~S~~AI~~G~~KIQDLS~~S~~TASALGK~~L~~QDVVNQNAQALN~~T~~LKQL
 S~~S~~NFGA~~I~~SSV~~L~~NDILS~~R~~LDKVEAEVQIDRLITGR~~L~~QSLQTYVTQQLIRAAEIRASANLAATK~~M~~SECVL~~G~~QSKRVDFCGKGYHLMSPQ~~S~~APHGVV
 FLHVTVYPAQEK~~N~~FTTAPAICH~~D~~GKAHFPRE~~G~~VFVSN~~G~~THWFVTQRNFYEPQIIITDNTFVSGNCDV~~V~~IGIV~~N~~NTVYDPLQPELDSFKEELDKYF
 KNHTSPDVLDGDISGINASVVNIQKEIDRLNEVAKNLES~~L~~IDLQELGKYEQYIKWPAHHHHHHHHHH

B

Theoretical N-glycopeptides of S protein digested using Glu-C

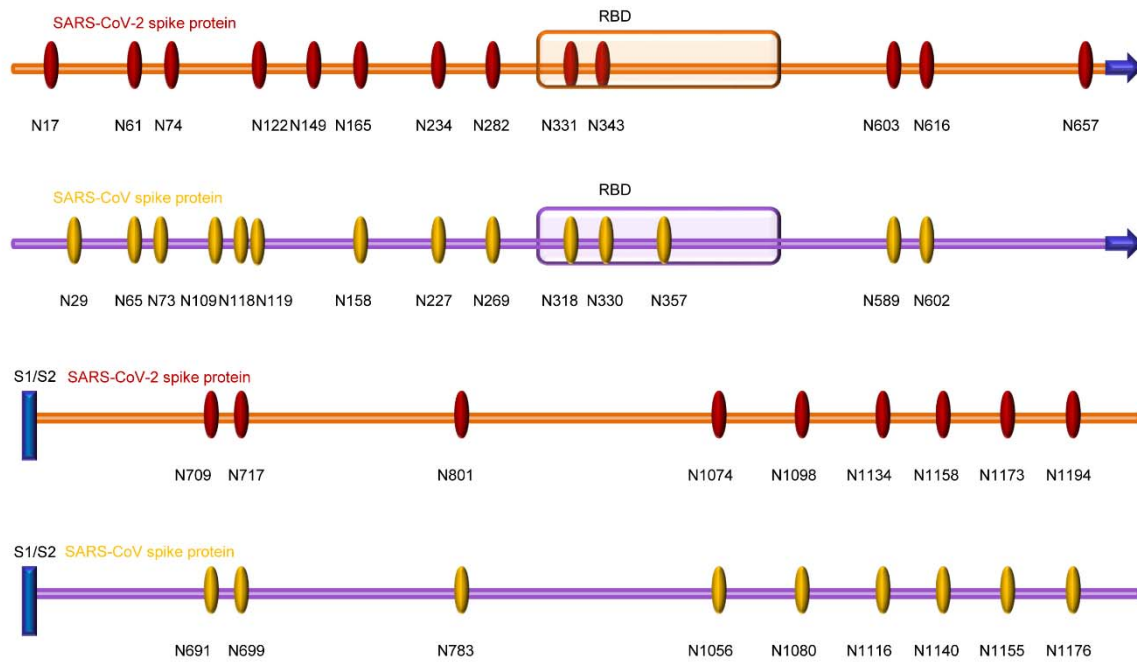
VNLTTRTQLPPAYTNSFTRGVVYYPDKVFRSSVLHSTQDLFLPFFSNVTFWHAIHVS~~GT~~NGTKRFDNPVLPFNDGVYFASTEKSNIIIRGWIFGTTL
 DSKTQSL~~LIVN~~NATNVVIKVC~~E~~FQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE~~G~~KQGNFKNLREFVFKNIDGYFKI
 YSKHTPINLVRDLPPQGFSAL~~E~~PLVDLPIGINITRFQTLALHRSYLT~~P~~GDSSSGWTAGAAAYVGYLQPRTFLLKYNE~~NG~~TITDAVDCALDPLSE~~T~~
 KCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN~~DL~~CFTN
 VYADSFVIRGDEV~~R~~QIAPGQTGKIADYNYKLPDDFTGCVIAWNSN~~LD~~SKVGGNYNYLYR~~L~~FRKSNLKPFE~~R~~DISTEIYQAGSTPCNGVEGFNC
 YFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQFG~~R~~DIADTTDAVRDPQT
 LEILDITPCSF~~GG~~VSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTG~~S~~NVFTQTRAGCLIGAEHV~~N~~NSYEC~~D~~PIGAGICASYQTQ
 TNSPRRARSVASQSIAYTMSLGAENSVAYS~~N~~NSIAIPTNFTISVTEILPVSM~~T~~KTSVDCTMYICGDSTEC~~S~~NLLQYGSFCTQLNRAL~~T~~GIAVEQ
 DKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSPKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLG~~D~~IAARDLICAQKFNGLTVLPPLLTDE~~M~~IA
 QYTSALLAGTITSGWTFGAGAALQIPFAMQMAYR~~F~~NGIGVTONVLYENQKLIANQFN~~S~~AI~~G~~KIQDLS~~S~~TASALGK~~L~~QDVVNQNAQALN~~T~~LKQL
 S~~S~~NFGA~~I~~SSV~~L~~NDILS~~R~~LDKVEAEVQIDRLITGR~~L~~QSLQTYVTQQLIRAAEIRASANLAATK~~M~~SECVL~~G~~QSKRVDFCGKGYHLMSPQ~~S~~APHGVV
 FLHVTVYPAQEK~~N~~FTTAPAICH~~D~~GKAHFPRE~~G~~VFVSN~~G~~THWFVTQRNFYEPQIIITDNTFVSGNCDV~~V~~IGIV~~N~~NTVYDPLQPELDSFKEELDKYF
 KNHTSPDVLDGDISGINASVVNIQKEIDRLNEVAKNLES~~L~~IDLQELGKYEQYIKWPAHHHHHHHHHH

C

Theoretical N-glycopeptides of S protein digested using trypsin and Glu-C

VNLTTRTQLPPAYTNSFTRGVVYYPDKVFRSSVLHSTQDLFLPFFSNVTFWHAIHVS~~GT~~NGTKRFDNPVLPFNDGVYFASTEKSNIIIRGWIFGTTL
 DSKTQSL~~LIVN~~NATNVVIKVC~~E~~FQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE~~G~~KQGNFKNLREFVFKNIDGYFKI
 YSKHTPINLVRDLPPQGFSAL~~E~~PLVDLPIGINITRFQTLALHRSYLT~~P~~GDSSSGWTAGAAAYVGYLQPRTFLLKYNE~~NG~~TITDAVDCALDPLSE~~T~~
 KCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN~~DL~~CFTN
 VYADSFVIRGDEV~~R~~QIAPGQTGKIADYNYKLPDDFTGCVIAWNSN~~LD~~SKVGGNYNYLYR~~L~~FRKSNLKPFE~~R~~DISTEIYQAGSTPCNGVEGFNC
 CYFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQFG~~R~~DIADTTDAVRDPQT
 TLEILDITPCSF~~GG~~VSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTG~~S~~NVFTQTRAGCLIGAEHV~~N~~NSYEC~~D~~PIGAGICASYQ
 TQ~~T~~NSPRRARSVASQSIAYTMSLGAENSVAYS~~N~~NSIAIPTNFTISVTEILPVSM~~T~~KTSVDCTMYICGDSTEC~~S~~NLLQYGSFCTQLNRAL~~T~~GIAVEQ
 EQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSPKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLG~~D~~IAARDLICAQKFNGLTVLPPLLTDE
 MIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYR~~F~~NGIGVTONVLYENQKLIANQFN~~S~~AI~~G~~KIQDLS~~S~~TASALGK~~L~~QDVVNQNAQALN~~T~~LKQL
 S~~S~~NFGA~~I~~SSV~~L~~NDILS~~R~~LDKVEAEVQIDRLITGR~~L~~QSLQTYVTQQLIRAAEIRASANLAATK~~M~~SECVL~~G~~QSKRVDFCGKGYHLMSPQ~~S~~APHGVV
 FLHVTVYPAQEK~~N~~FTTAPAICH~~D~~GKAHFPRE~~G~~VFVSN~~G~~THWFVTQRNFYEPQIIITDNTFVSGNCDV~~V~~IGIV~~N~~NTVYDPLQPELDSFKEELDKYF
 KNHTSPDVLDGDISGINASVVNIQKEIDRLNEVAKNLES~~L~~IDLQELGKYEQYIKWPAHHHHHHHHHH

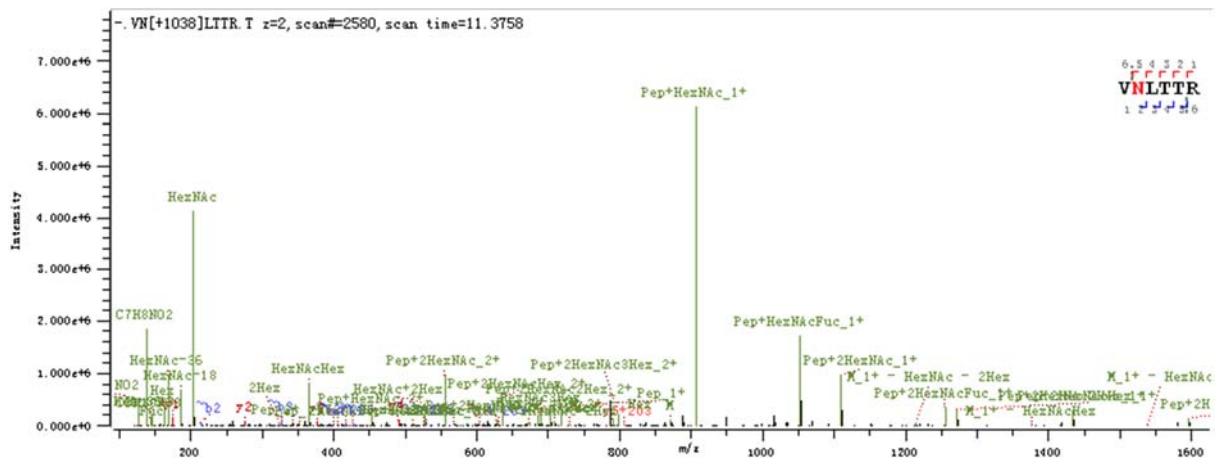
Supplementary Figure S6. Comparison of the N-glycosites on the SARS-CoV-2 and SARS-CoV spike proteins.



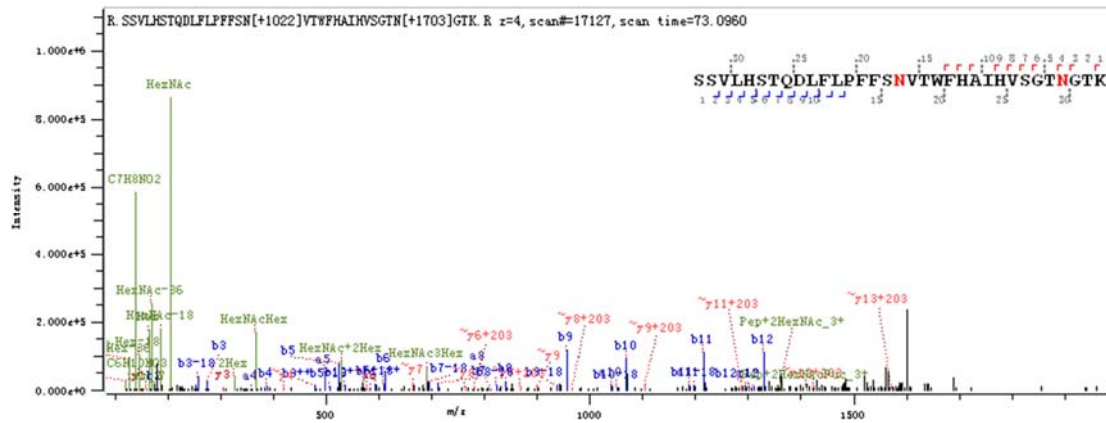
Supplementary Figure S7. Representative and high-quality spectra of intact N-glycopeptides and deglycosylated peptides.

Spectra of intact N-glycopeptides from recombinant SARS-CoV-2 spike proteins expressed in insect cells

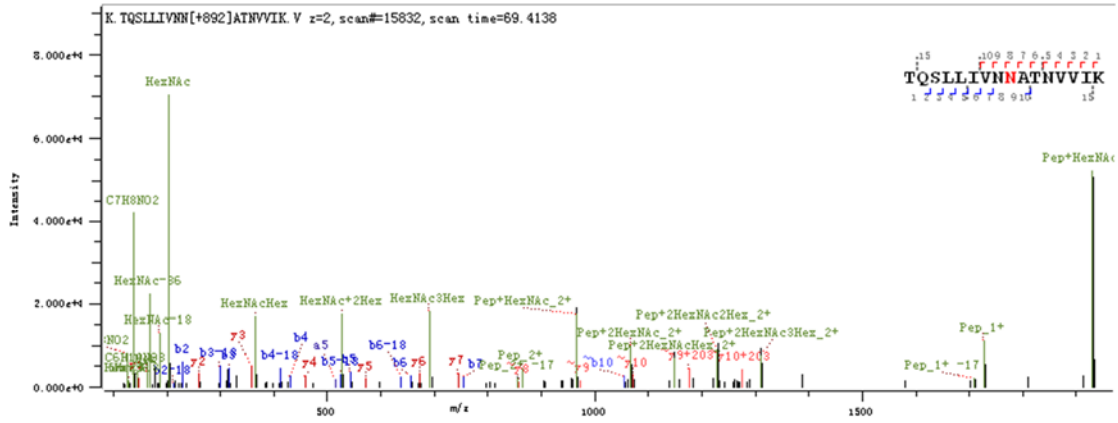
N17



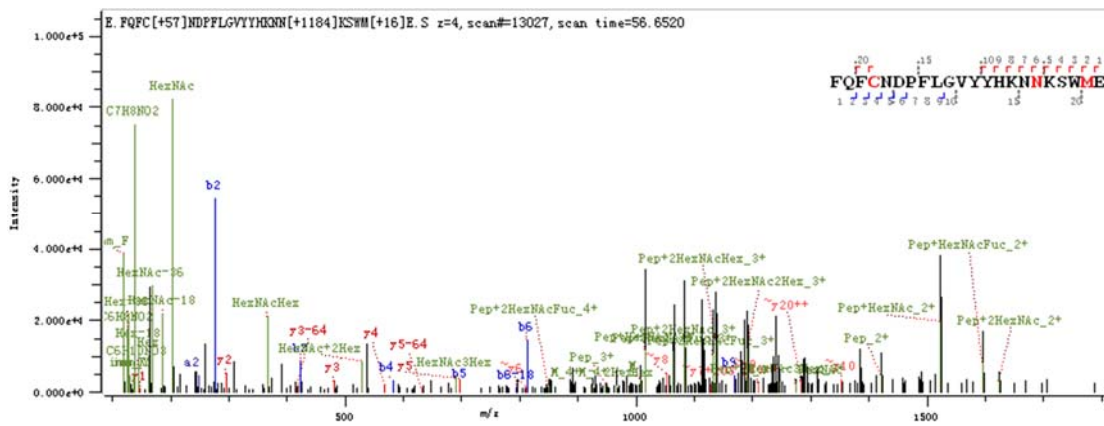
N61 & N74



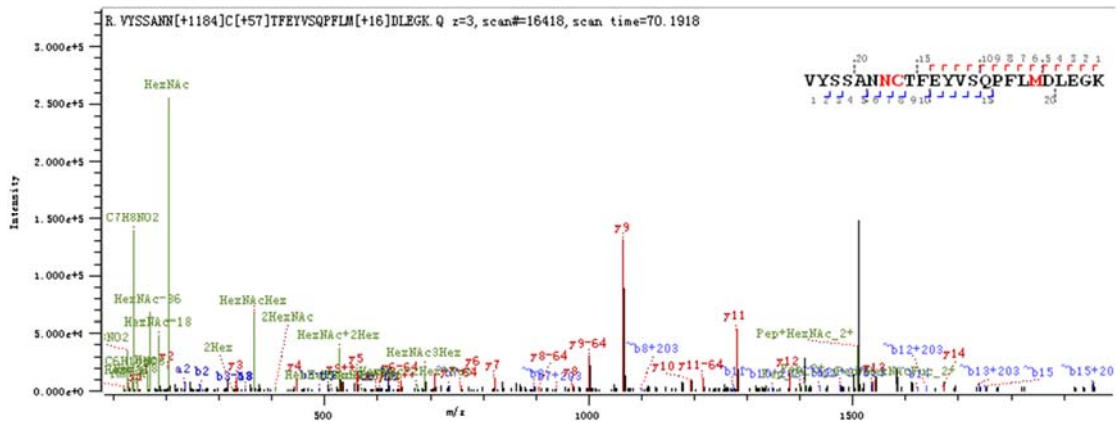
N122



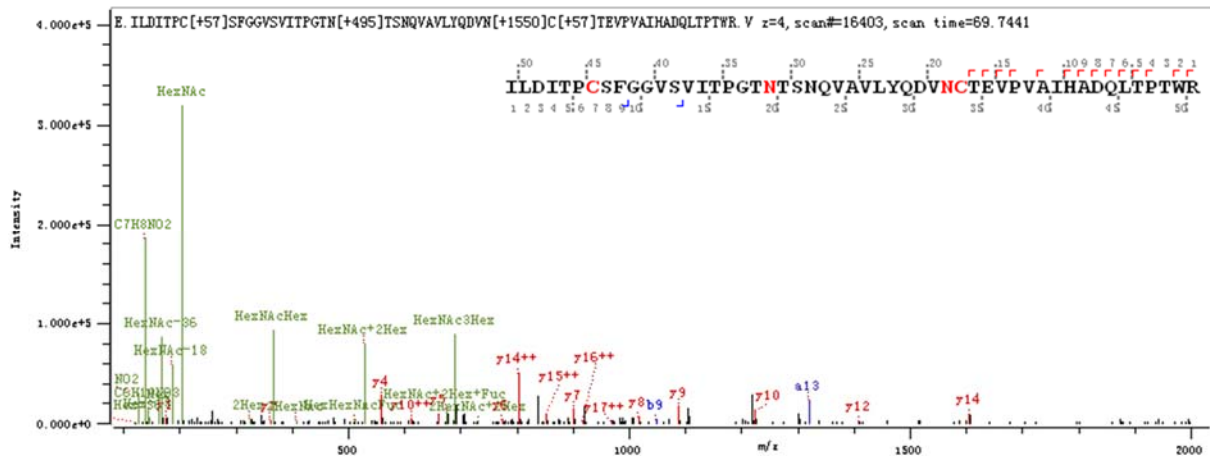
N149



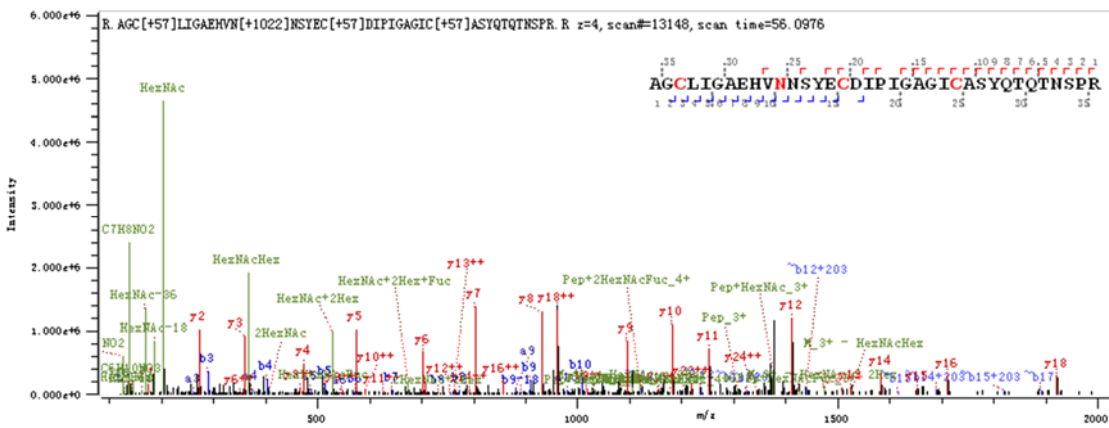
N165



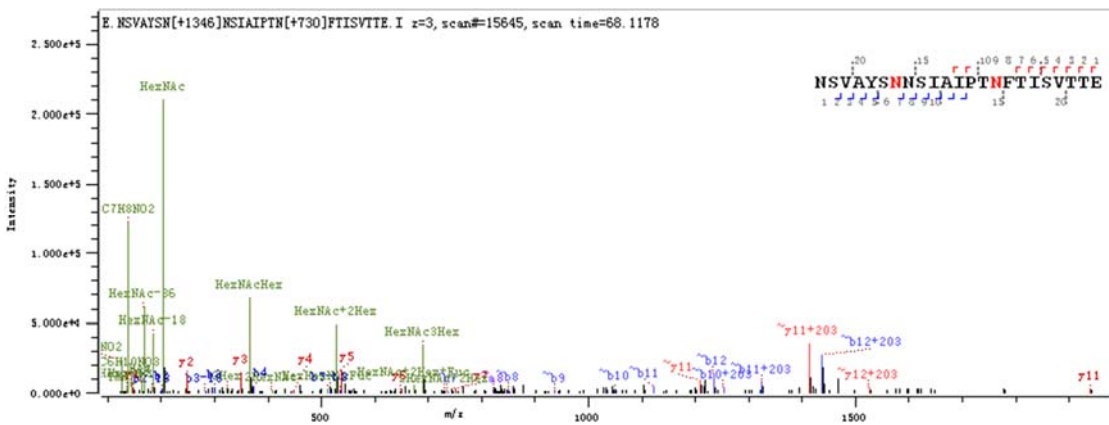
N234

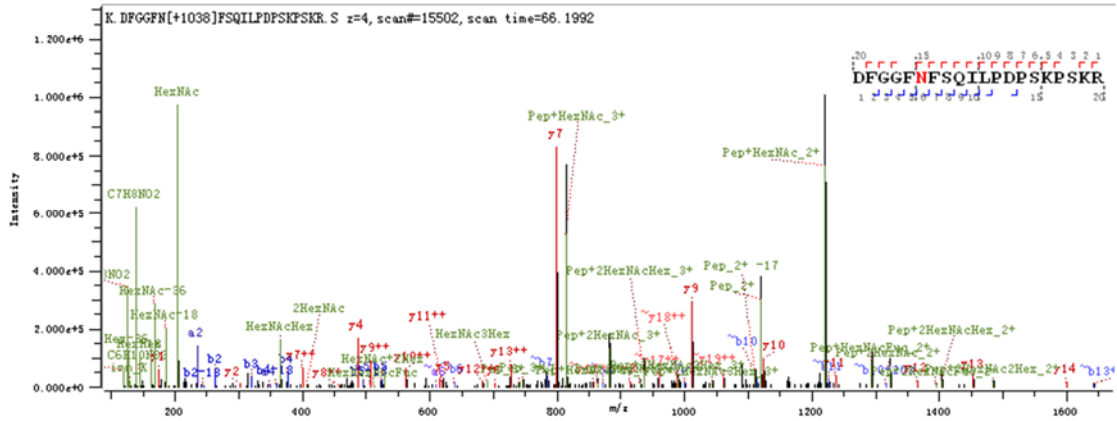


N657

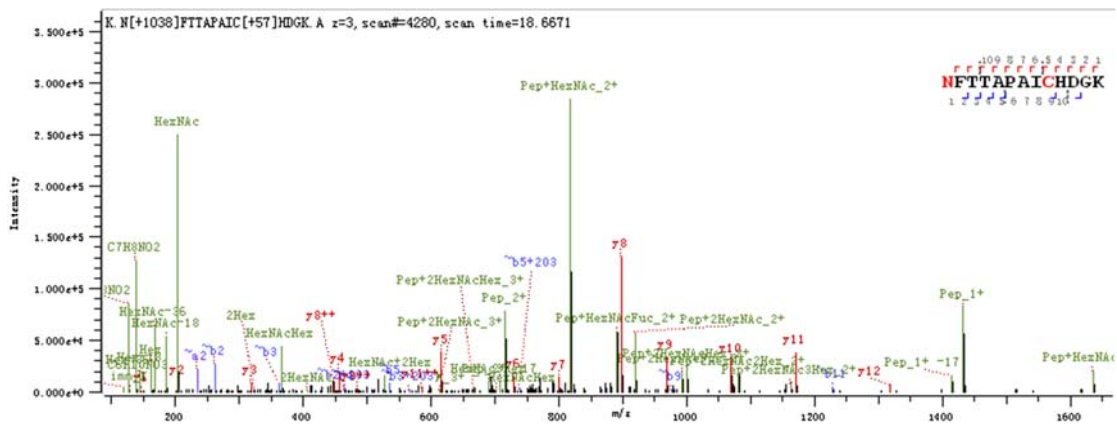


N709 & N717

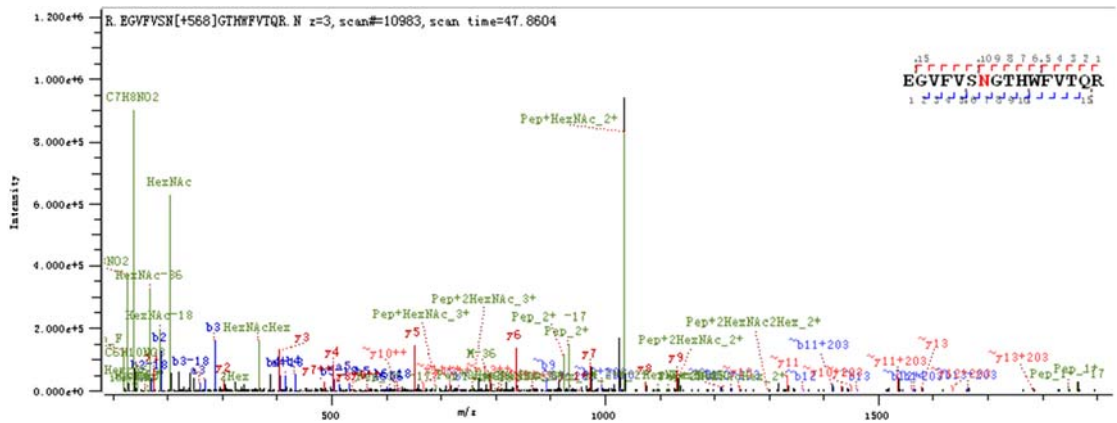




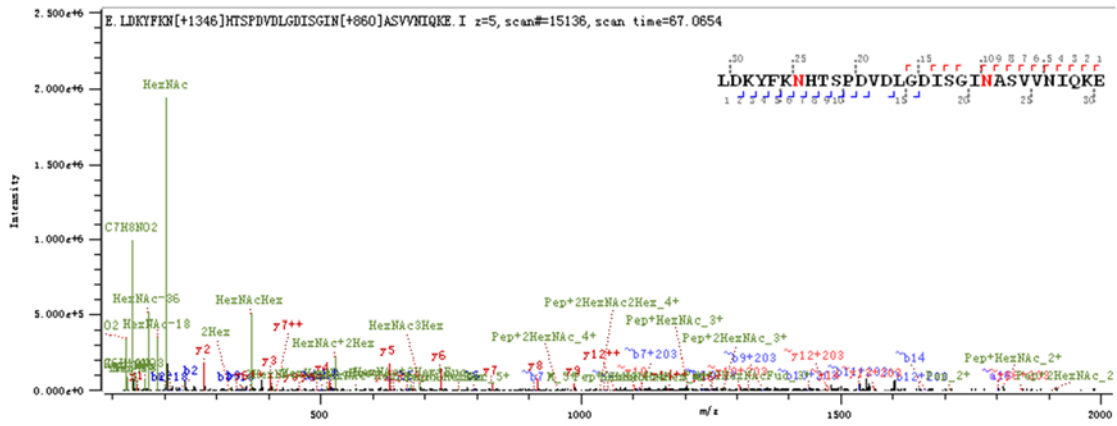
N1074



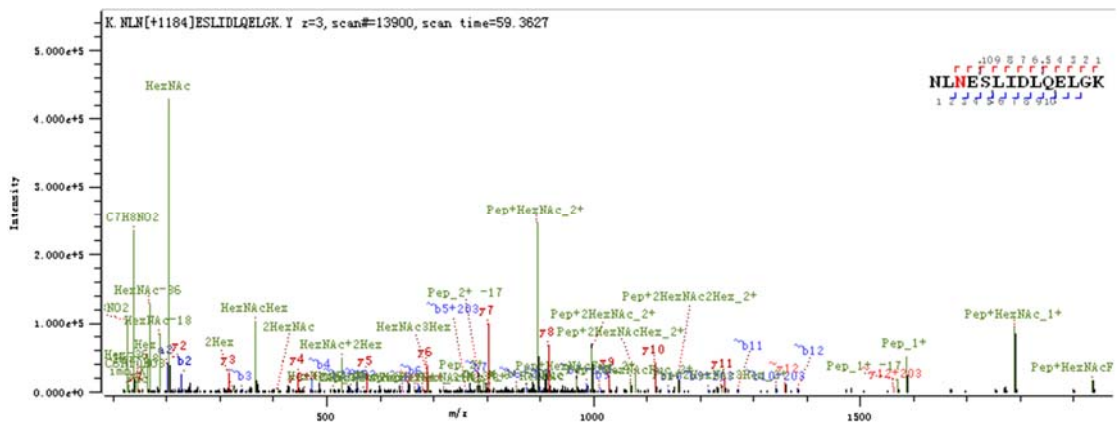
N1098



N1158 & N1173

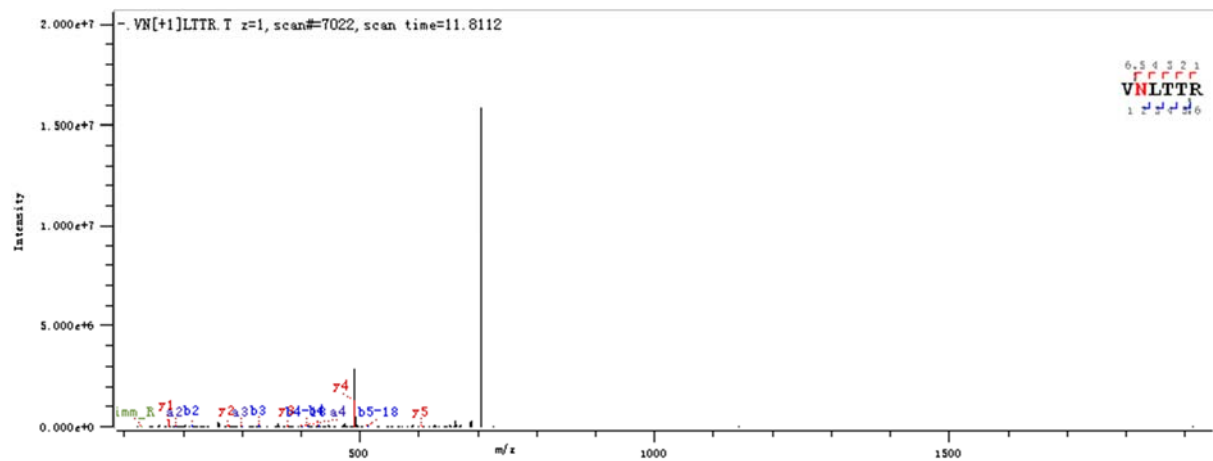


N1194

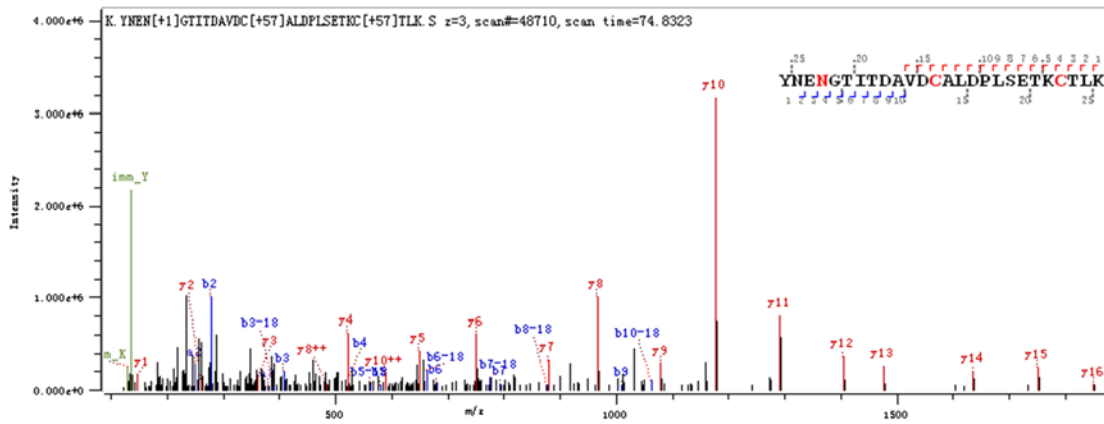


Spectra of deglycosylated peptides from recombinant SARS-CoV-2 spike proteins expressed in insect cells

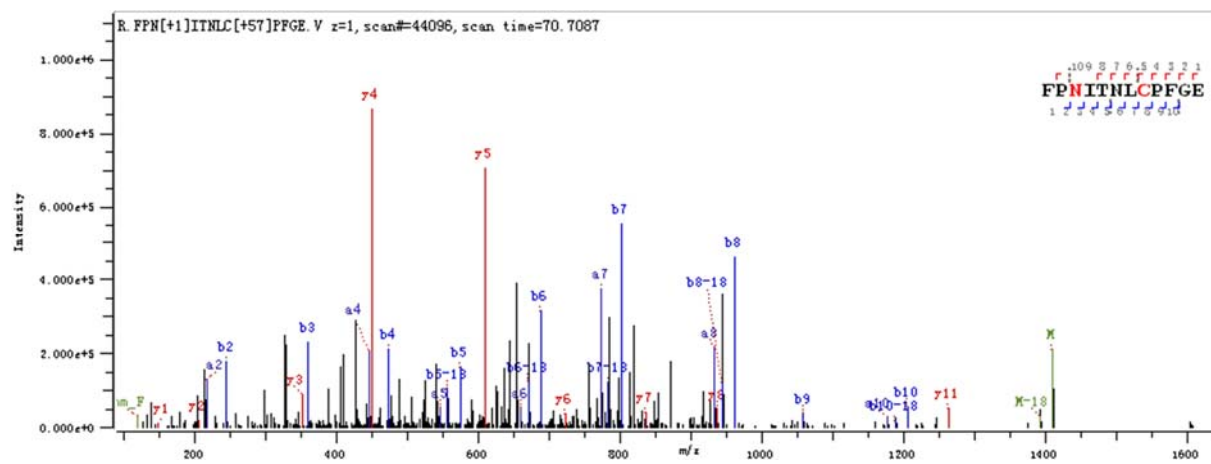
N17



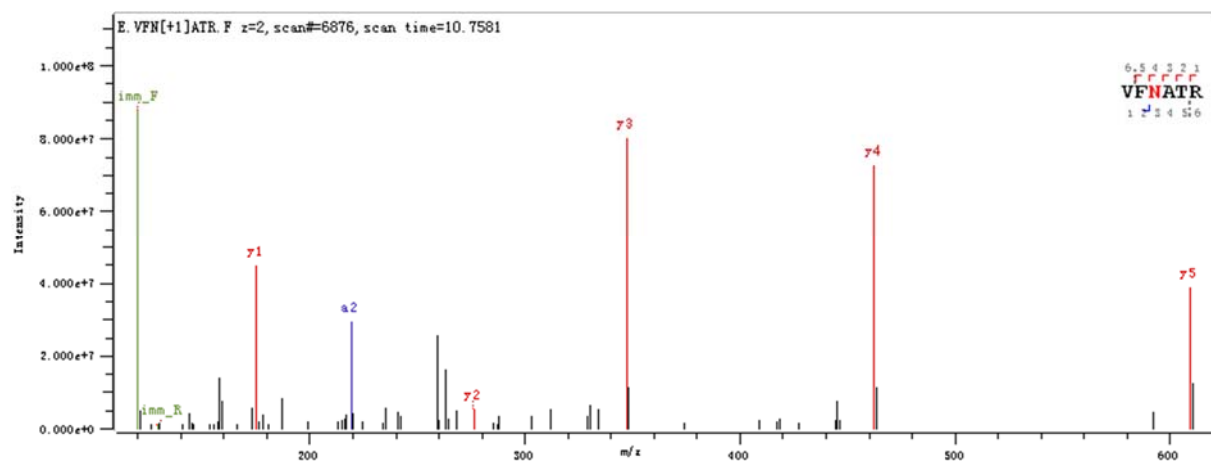
N122



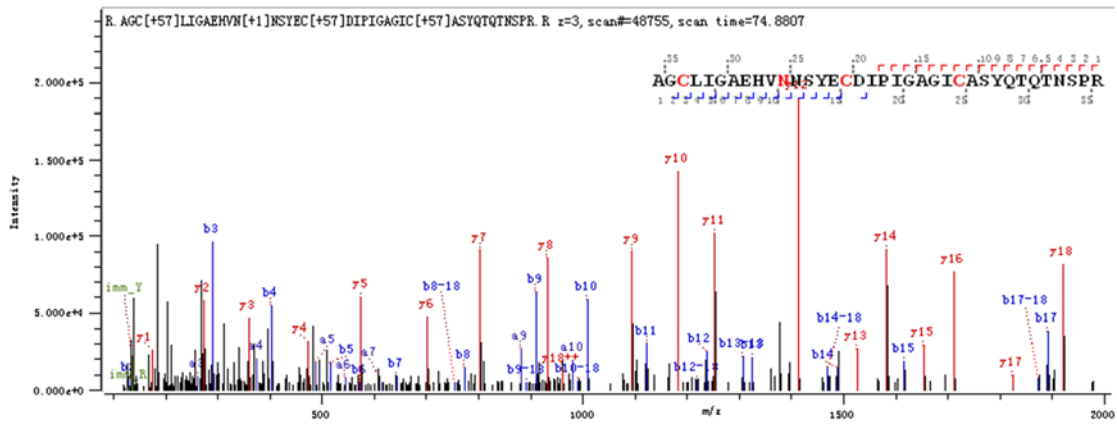
N331



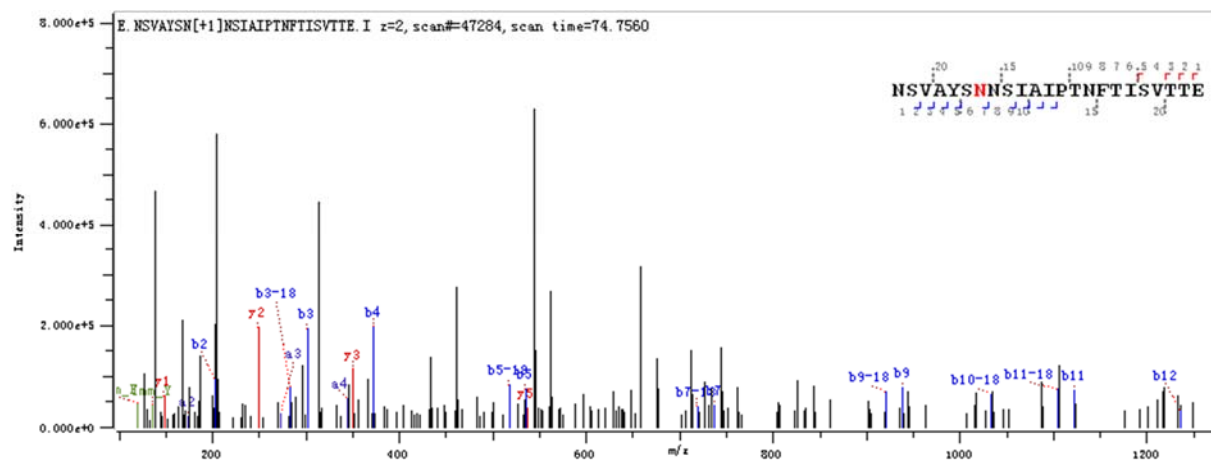
N343



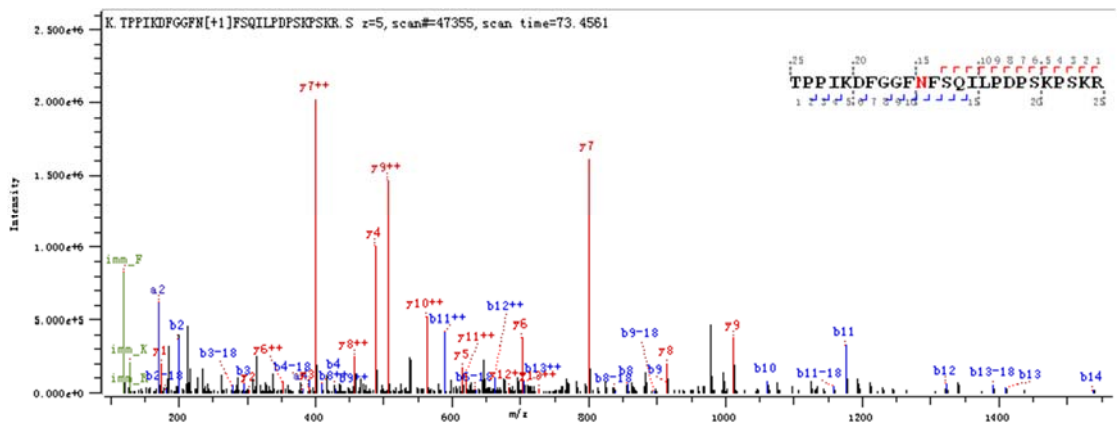
N657



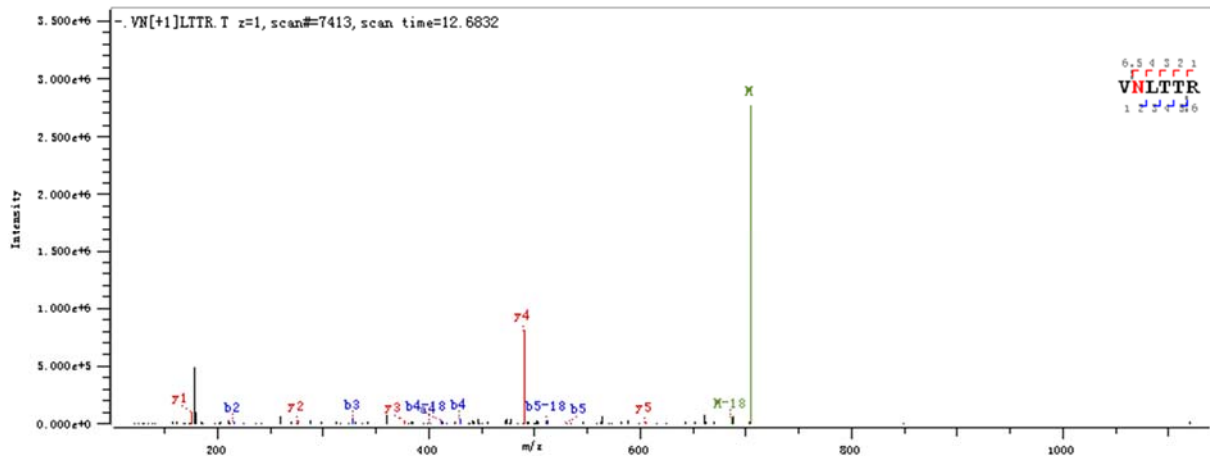
N709



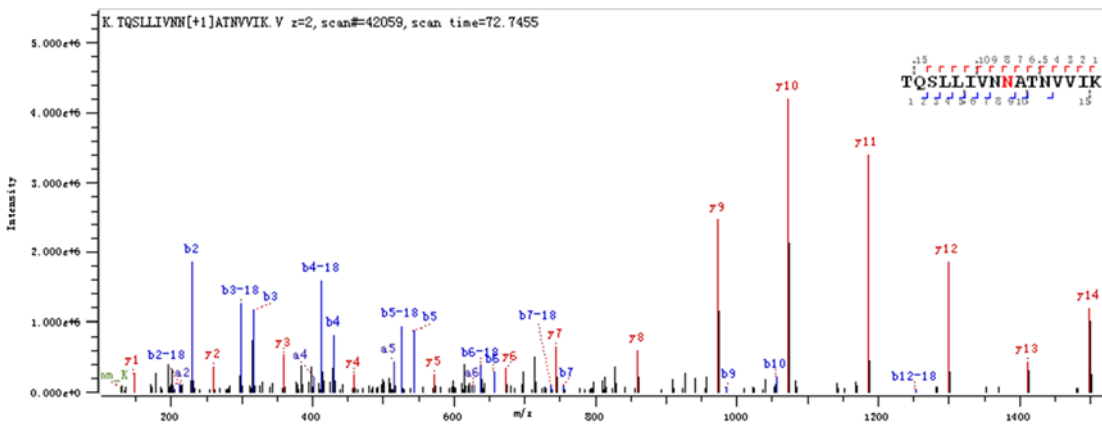
N801



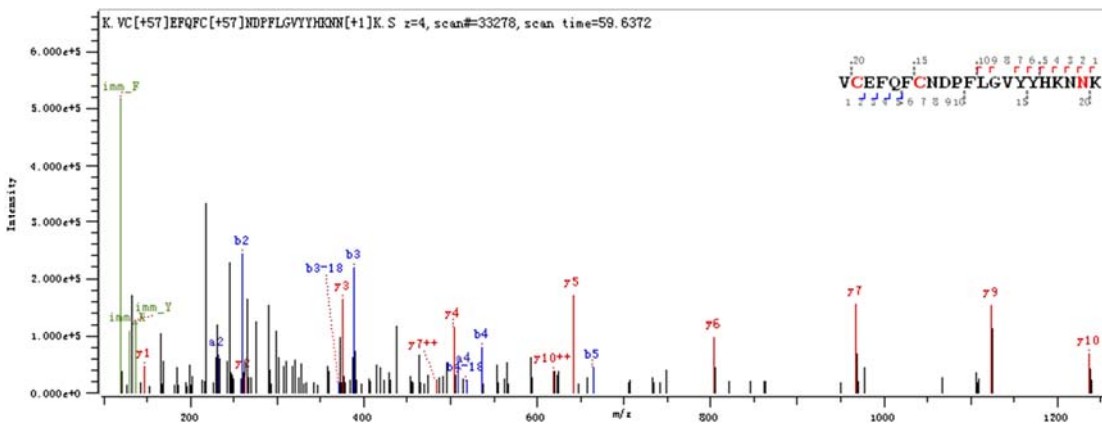
N1074



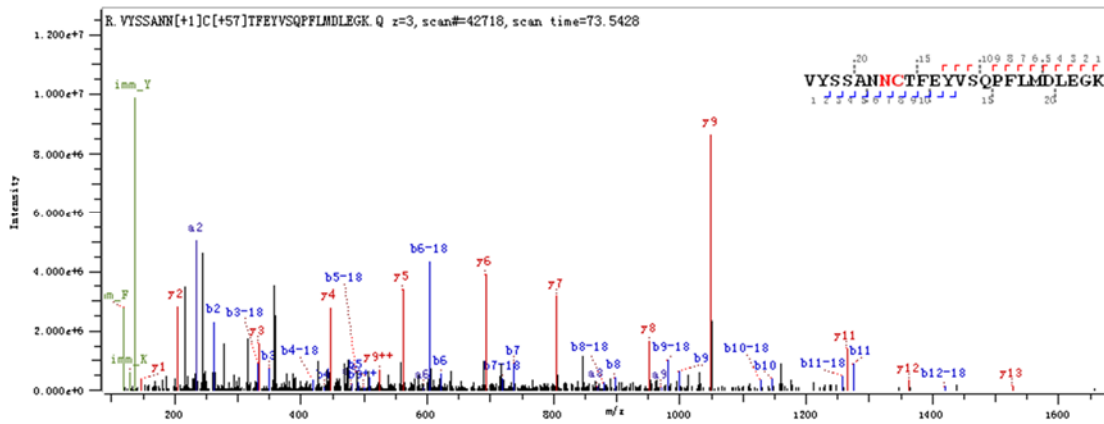
N122



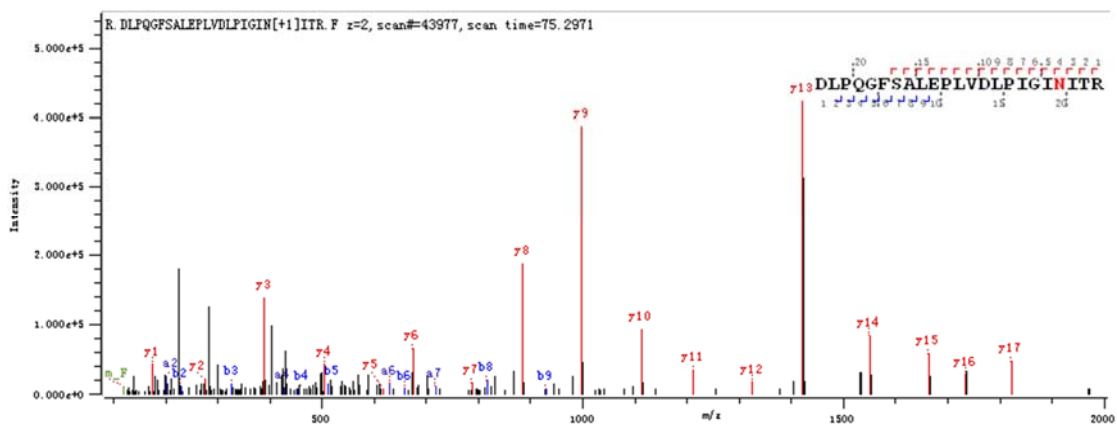
N149



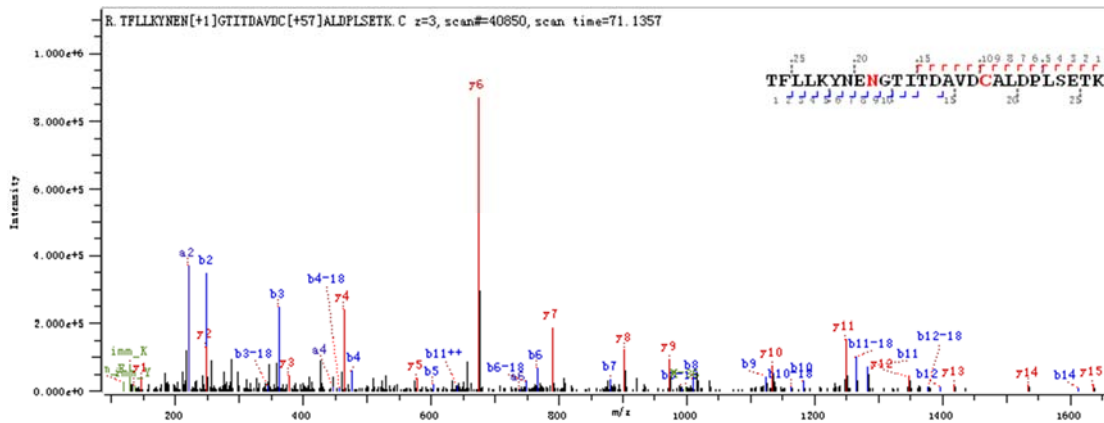
N165



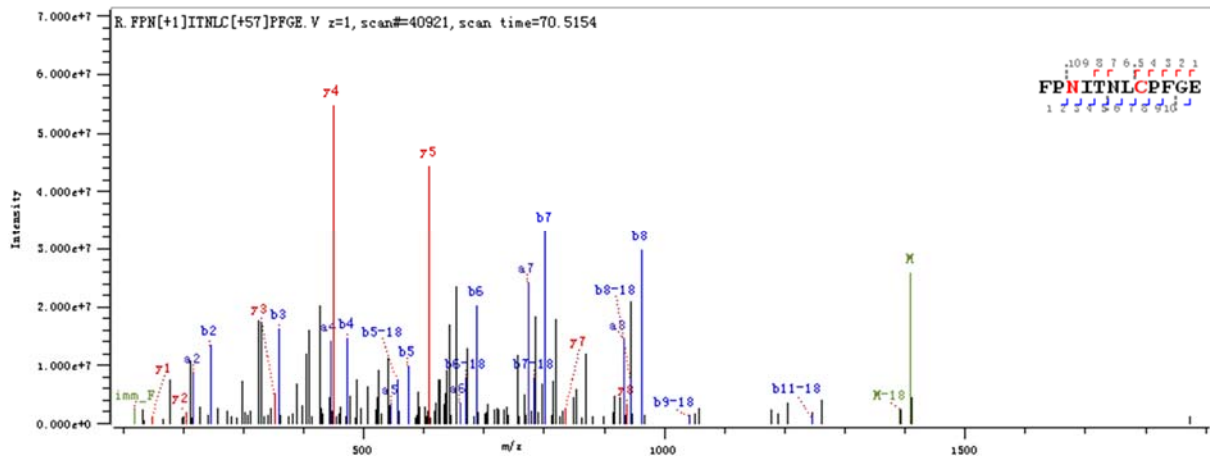
N234



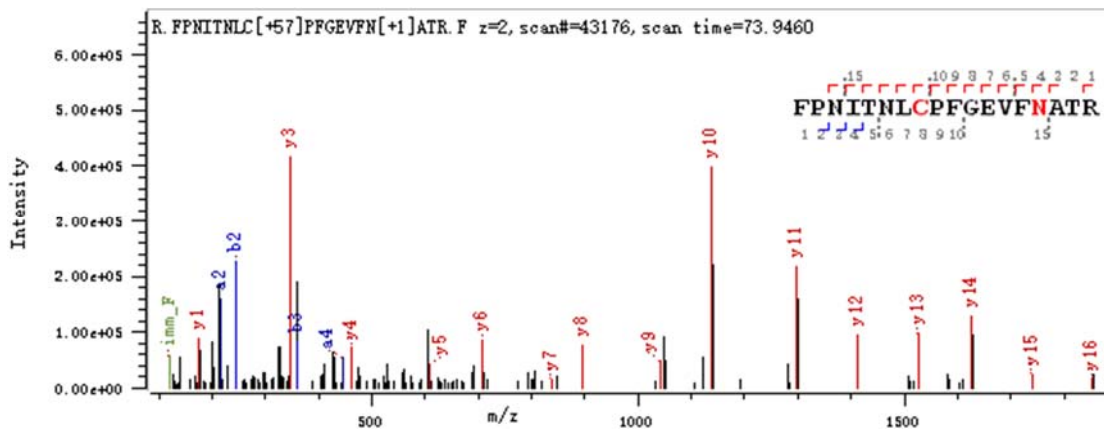
N282



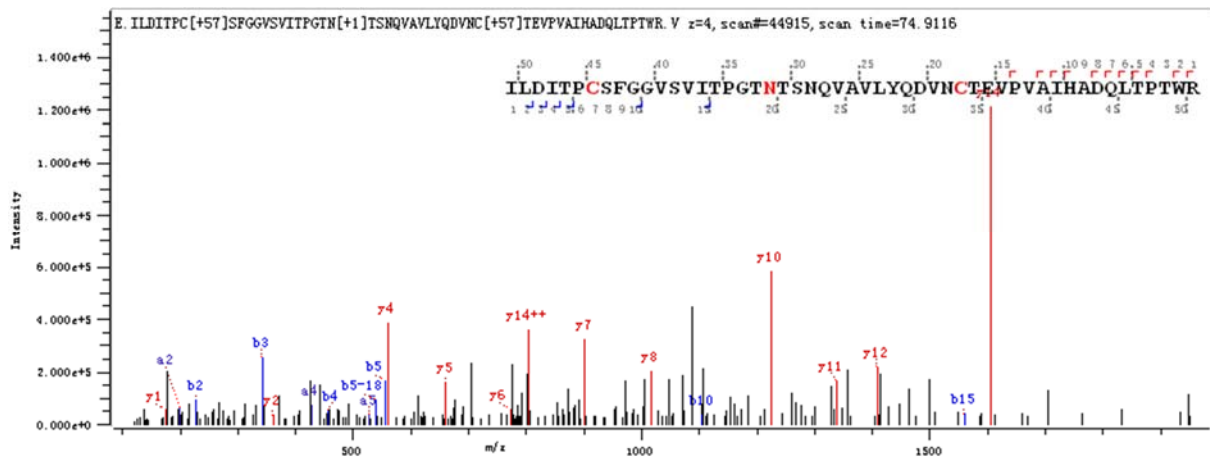
N331



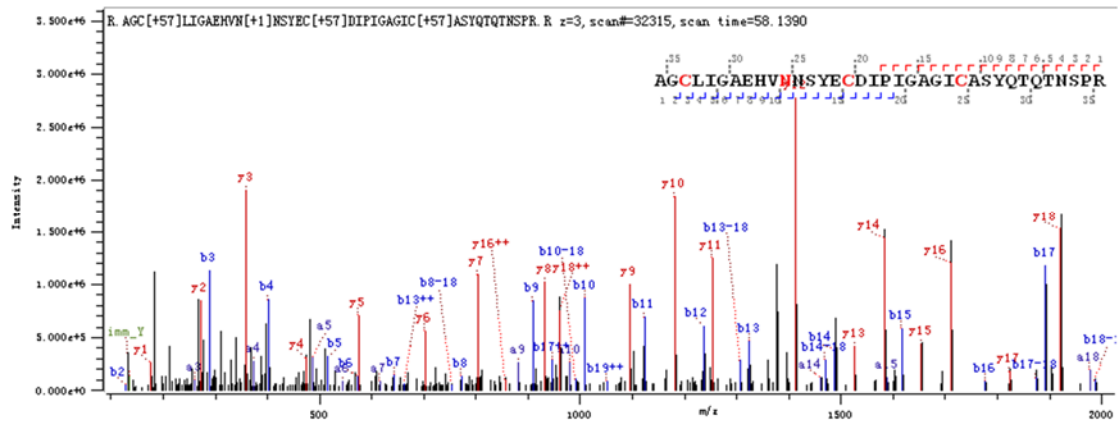
N343



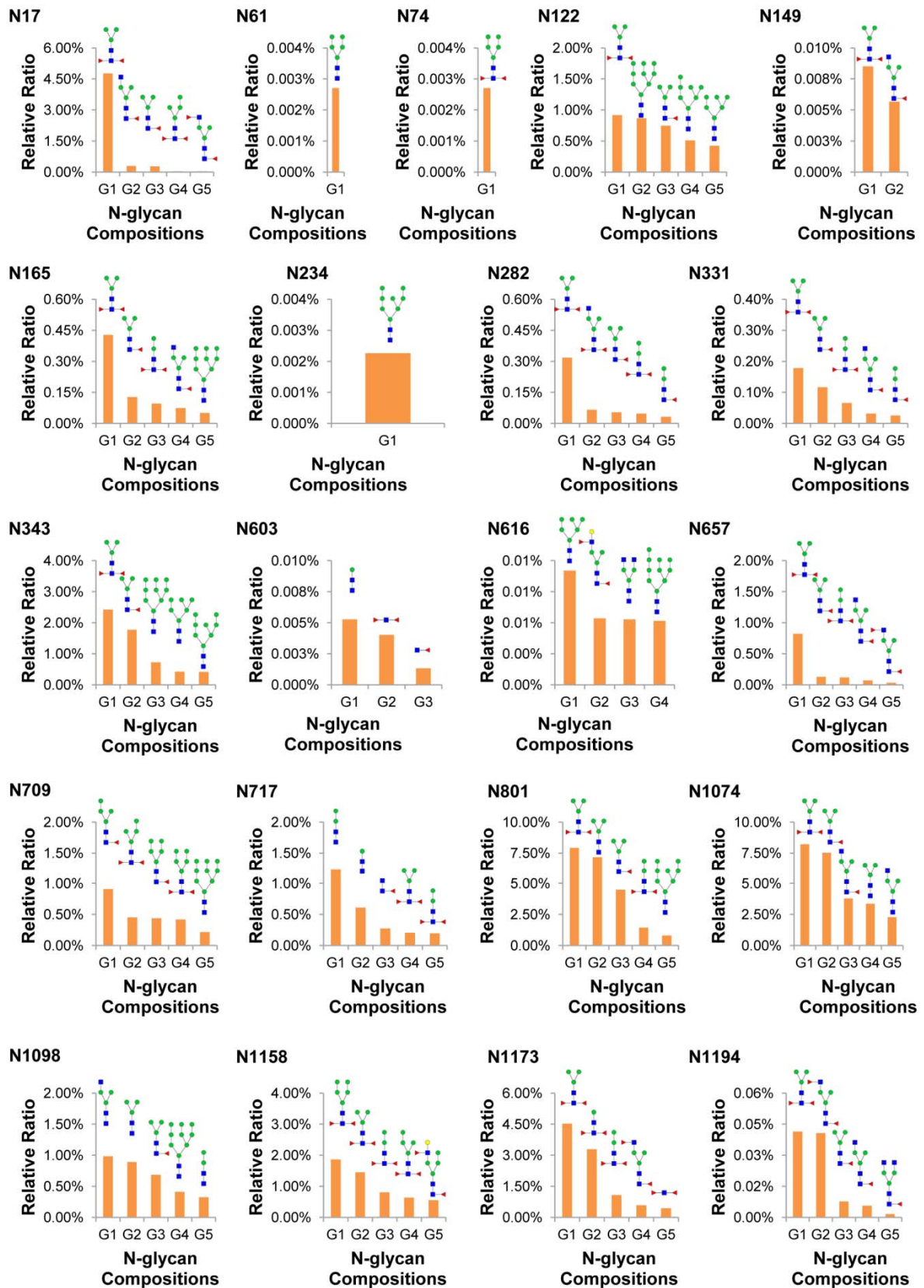
N603



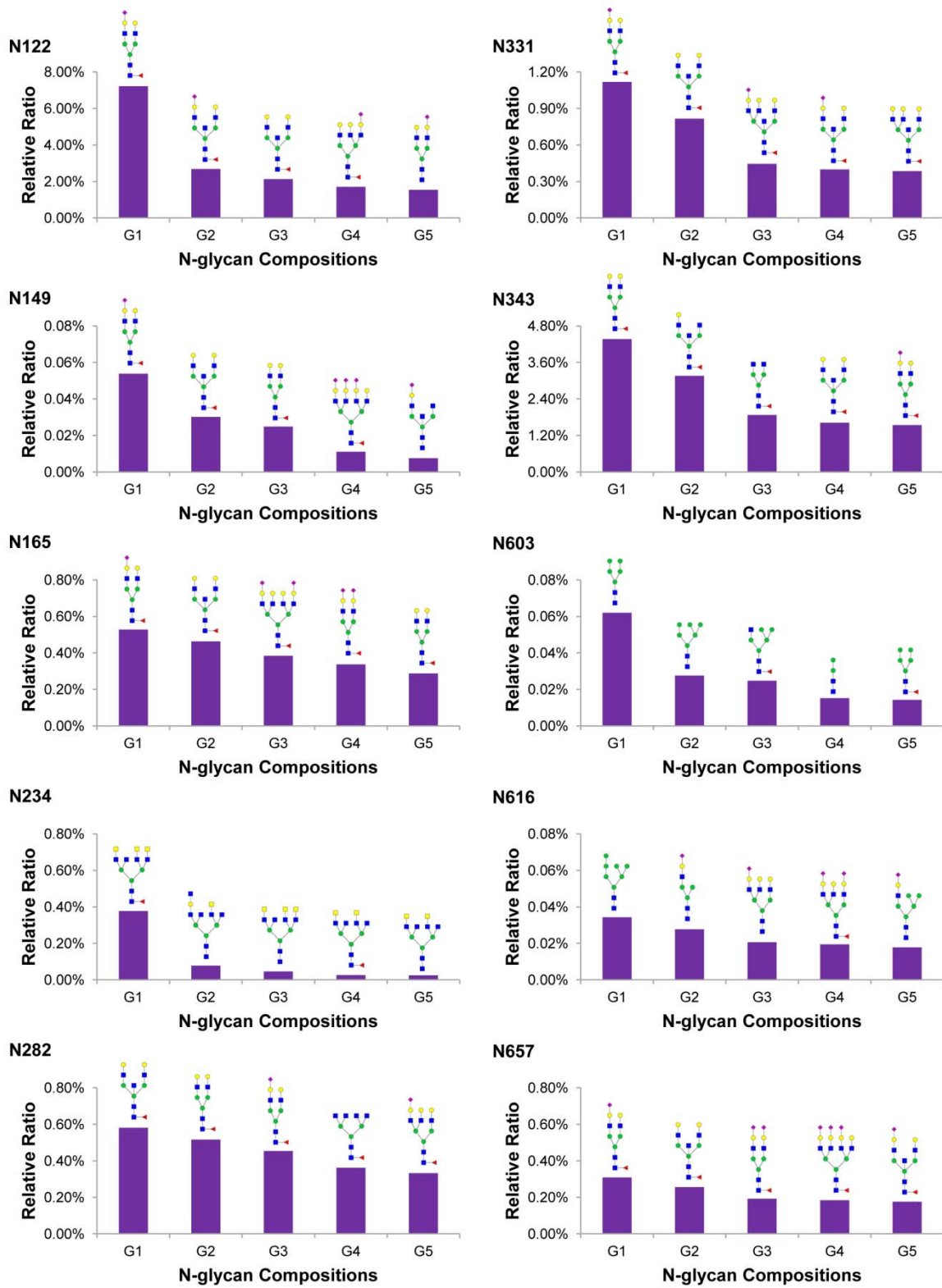
N657



Supplementary Figure S8. Quantitative analysis of site-specific N-glycosylation of S protein expressed in insect cells (top 5 N-glycans).



Supplementary Figure S9. Quantitative analysis of site-specific N-glycosylation of S protein expressed in human cells (top 5 N-glycans).



Supplementary Figure S11. SDS-PAGE analysis of RBDs expressed in insect and human cells.

