Analytical Chemistry

Automated assignments of N- and O-site specific glycosylation with extensive glycan heterogeneity of glycoprotein mixtures

John S. Strum[1], Charles C. Nwosu[1], Serenus Hua[1,2], Scott R. Kronewitter[1], Richard R. Seipert[1], Robert J. Bachelor[1], Hyun Joo An[2], and Carlito B. Lebrilla*[,1,3]

1. Department of Chemistry, University of California, Davis, CA 95616, USA.
2. Graduate School of Analytical Science and Technology, Chungnam National University, Daejeon, South Korea.
3. Department of Biochemistry and Molecular Medicine, University of California, Davis, CA 95616, USA.

* To whom correspondence should be addressed:

Carlito B. Lebrilla, Email address: cblebrilla@ucdavis.edu; Tel: +1-530-752-0504; Fax: +1-530-752-8995

Keywords: Glycoproteomics, site-specific glycosylation, tandem mass spectrometry, false-discovery rate, target-decoy approach

Organization of Supplemental Materials

Additional Details of the GP Finder Algorithms

The software analyzes tandem MS data in two distinct phases: identification of potential glycopeptides according to accurate precursor mass and scoring of competing matches for each precursor mass according to tandem MS fragmentation. The first phase is straight forward. All possible peptide components are calculated with the algorithm in **Figure S-20** using the user-provided protein sequences and the maximum experimental mass. The list of theoretical peptides is then used in the algorithm in **Figure S-21** to determine the list of potential glycans present in the sample by subtracting from each experimental mass the mass of each theoretical peptide (generating residual glycan masses). Concurrently, a list of theoretical glycans of specific glycan compositions is generated in the algorithm in **Figure S-22** based on the user-defined parameters of the minimum and maximum numbers of each of the possible monosoccharides. The final part of the first phase is found in the algorithm in **Figure S-23**, where the list of residual glycan masses is compared to the list of the theoretical glycans according to the user-defined mass tolerance. All resulting matches are then further analyzed in the second phase according to their tandem mass spectra. Prior to scoring the tandem MS data, GP Finder uses the algorithm in **Figure S-24** to eliminate tandem MS peak matches that do not follow certain biological rules. This filter simplifies the data analysis by removing many random matches. A similar filter is also used to eliminate false assignments in the first phase, however, here it eliminates unlikely fragments rather than unlikely intact glycopeptide compositions. Representative examples in **Figure S-5, S-6** and **S-26** show the logic behind some of the rules. The acceptable peak matches are scored by the algorithm in **Figure S-25**, which is described in the manuscript by **Eq. 1** through **4** in the Methods section.

Input Parameters for GP Finder

The input protein sequences are UniProt Flat Files. All other parameters were left at the default settings (although they can be specified by the user). The output file is Excel. The results can also be viewed through the program's user-interface. The theoretical glycopeptides were bounded within 20 ppm

of the empirical neutral masses (after mass deconvolution). N- and O-linked theoretical matches were generated separately. N-linked matches were limited to the following monosaccharide compositions: 3-9 Hexose, 2-7 HexNAc, 0-5 deoxyhexose, and 0-4 NeuAc. The O-linked matches were limited to 0-4 Hexose, 1-4 HexNAc, 0-3 deoxyhexose, and 0-3 NeuAc. Results were filtered through the previously published retrosynthetic N-linked library. The tandem data was bounded within 80 ppm of the empirical neutral masses. A set of rules was used to eliminate matches to fragments in the tandem data that are unlikely to be real. The pseudocode for the elimination rules is included in **Figure S-24**. The respective unit-less weights for scoring of the tandem MS fragment types were 5, 5, 4, 3, 2, 1. Additional parameters can be set, such as ignoring data with low absolute or relative abundance or augmenting the score based on extensive chains or ladders of monosaccharide mass differences, however these filters were not applied here.

The software can handle CID data as a CSV file independent of the instrument type. The spectra must be saved in the following format where the key header information includes the retention time, deconvoluted mass-to-charge ratio, and product ion charge state:

Product Ion (*retention time* min) *ionization type fragmentation type* (*deconvoluted precursor mass-to-charge ratio*[z=*charge state*] -> **)
#Point,X(Thompsons),Y(Counts)
line number,deconvoluted product ion mass-to-charge ratio,abundance

Some example data in this format that has been shortened in length is shown below:

```
# Product Ion (20.575 min) ESI CID (1026.9301[z=2] -> **)
#Point,X(Thompsons),Y(Counts)
0,203.0868,36.16447449
1,273.0794,17.00735283
2,291.1014,20.04166603
3,365.119,113.6859741
# Product Ion (20.629 min) ESI CID (723.2941[z=2] -> **)
#Point,X(Thompsons),Y(Counts)
0,203.0918,20.55681801
1,273.0777,251.6022797
2,291.0978,186.9583282
3,324.0412,12
```

The algorithm can technically be used for ETD but it has not yet been tested with these data and we make no claim that it does work with such data.

Discussion of Limitations

The algorithm is based on the strategy used by Nwosu but is more comprehensive in its consideration of evidence, more blind to "expected" results, and is quantitative in respect to its comparison of competing matches. The algorithm does not use glycan self-consistency because the pool of possible glycans is quite large and homogenous (producing extensive glycan families among false matches). A human analyst may see trends that distinguish one set of glycans from another, but the software will consider a much larger pool of glycan families that randomly match because of both the large masses of the theoretical glycans and the similar combinations as a result of having only a few unique monomers. Peptide self-consistency is less susceptible to this problem.

Truly unbiased TDA assumes that the entire span of the decoy and false target distributions are identical. We set the average target and decoy boosts to be the same but do not establish that the entire distribution is the same. Consider a potential scenario where the initial false target and decoy scores are in the range 10–18 and the subsequent target boost is 20 points for 50% of the false target matches while the decoy boost is never more than 10 points. The decoy matches do not score more than 28; however, potentially 50% of the false target matches could be boosted higher than 20 points (or lower, but that is an issue of disguising true matches as false ones). As such, whatever percentage is boosted higher than the average will appear to have an FDR of zero.

TDA is particularly useful for analyzing sets of data files with vastly different quality and content. Such flexibility is afforded by the decoy distribution that is tailored to each dataset. TDA-based FDR is not unlike internal standards used in quantitative chemistry with interfering matrices. Even when the target matches are poor and the target scores completely overlap with the decoy scores, the method

reliably reveals poor target data and has helped us rapidly seek out and find the underlying problems behind target datasets (such as discovering poor S/N or inadequate peak intensity thresholding).

The algorithm can accommodate data from any enzyme; however, the software does not change its algorithm based on the enzyme and does not accommodate variable modifications of amino acids. The algorithm filters data effectively without relying on data-reduction through the use of a specific enzyme.

The paucity of both correct and incorrect assignments for each tandem spectrum can be problematic for making per spectrum statistical inferences. Sufficient false assignments are required if one is to model the incorrect distribution for a single spectrum and assign a likelihood such as the E-value that the top score is not a component of the population of random false assignments.

A histogram of all the scores for matches to a single spectrum shows a top match that is separated from the lower matches (**Figure S-27** and **S-28**). The lower matches are characteristic of random (incorrect) matches whereas the top match is unique with less than a 10% likelihood of belonging to the random score distribution; however, the results for this type of test are highly variable (other examples shown in **Figure S-29**) despite the fact that the distribution of scores that includes only the top score from each spectrum is bimodal (**Figure 1**) and the results were verified with previously published manual analysis. We believe that the poor distinction of top scores for matches to single spectra results from the high similarity among the identified fragments for competing glycopeptides. Another contributing factor is the lack of a sufficient number of false matches to fill out the random distribution, precluding local (per spectrum) statistical inferences, even in cases where the top two scores are substantially separated, as for the data in **Table S-5**.[58-59] Although not employed here, a probability-based approach such as MS-GF[60] could be used for local statistics regardless of the distribution and number of theoretical assignments for each spectrum; however, such an analysis would require modification of MS-GF to handle P and PN ions, glycan losses from precursor, etc.; extensive modifications which have yet to be implemented.

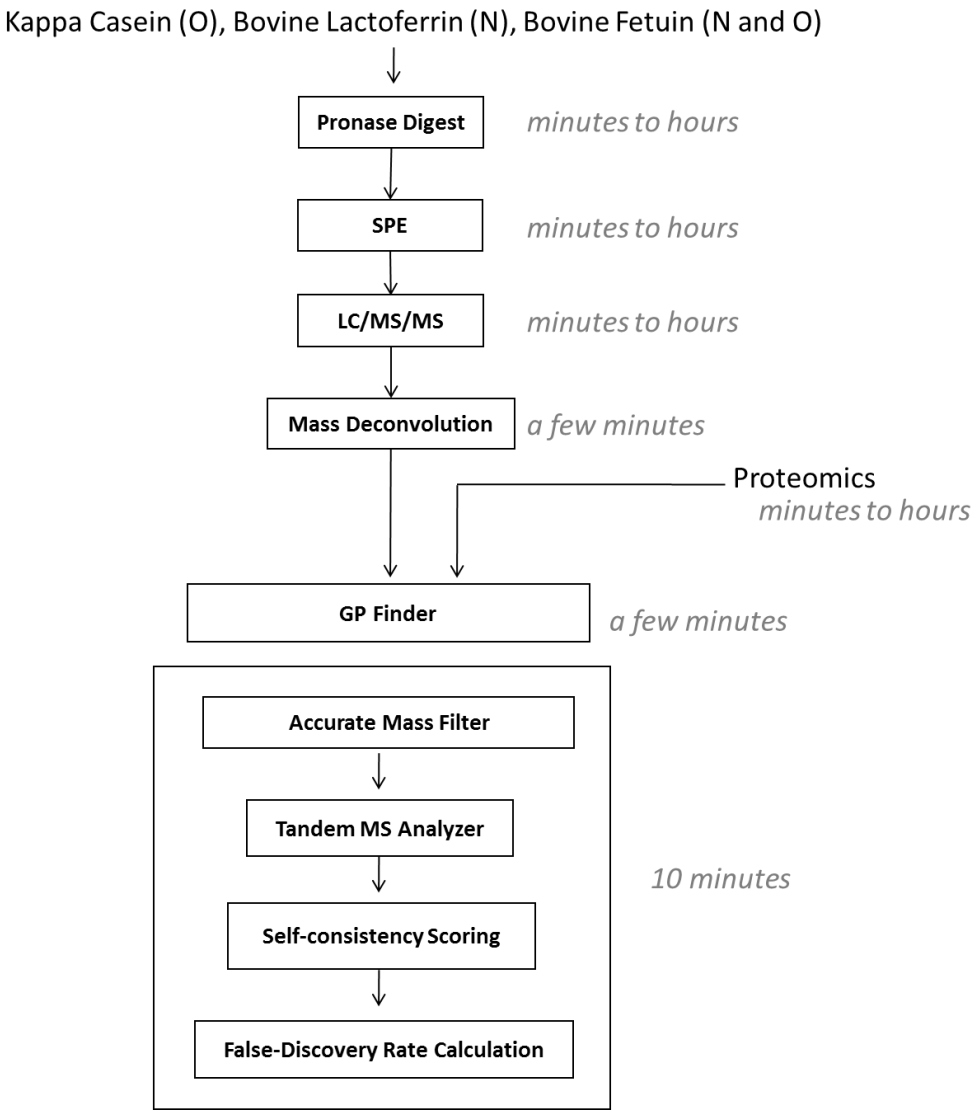


Figure S-1. Workflow. Experimental and data analysis workflow for SSG with non-specific proteolysis and in-house software called GP Finder.

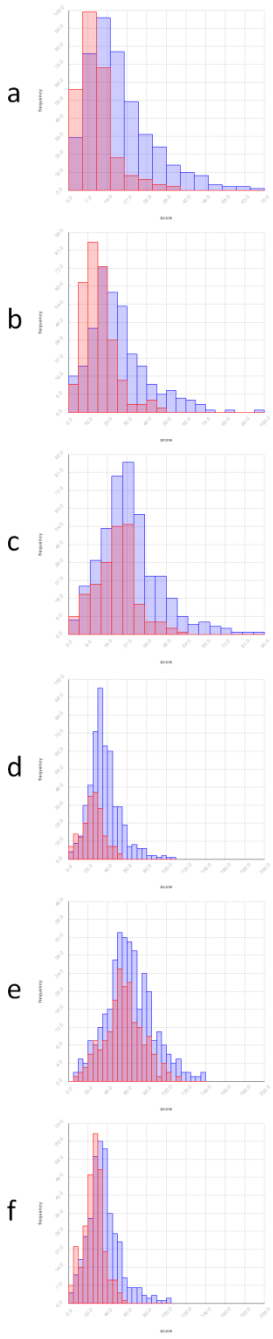


Figure S-2. Analysis of cocktail data with kappa casein protein sequence. A. Base-score only B. Family-size self-consistency boost C. Family-average self-consistency boost ignoring family size D. Family-average self-consistency boost with consideration of family size (standard self-consistency analysis without TDA-compliance) E. Summed family-size and family-average boost F. Family-average self-consistency boost (standard self-consistency analysis with TDA-compliance).

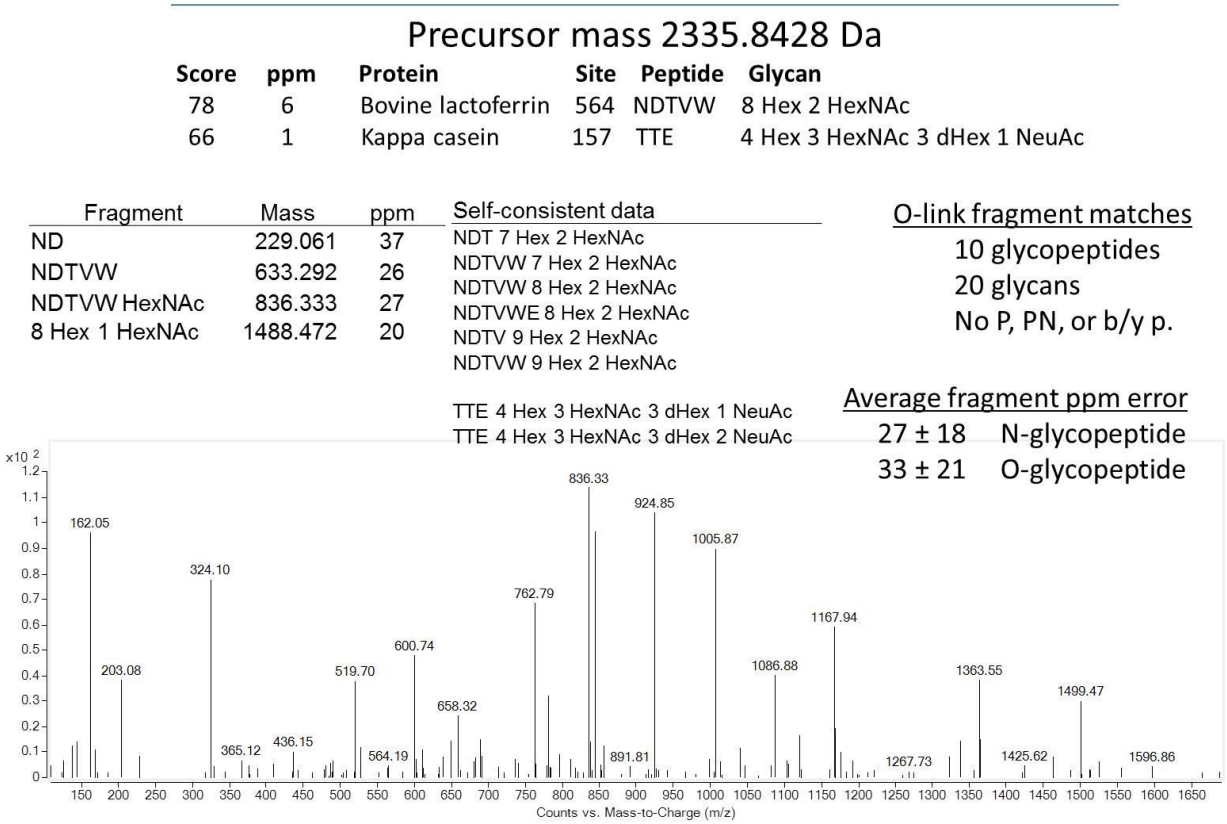


Figure S-3. Example of the evidence available for disambiguating the overlapping assignments above 5% FDR for both N- and O-glycopeptides.

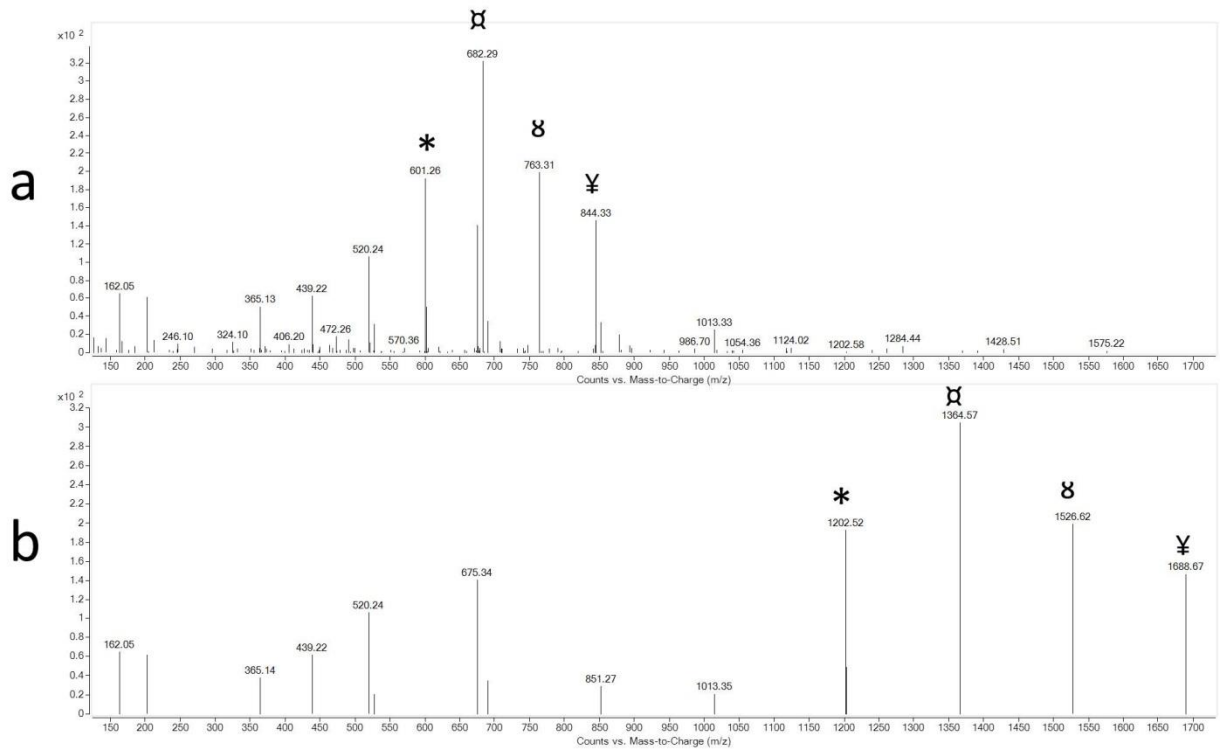


Figure S-4. Difference in N-glycopeptide tandem MS mass deconvolution by changing the molecular composition assumption. A. Mass deconvolution with assumption of organic molecular composition failed to deconvolve four glycopeptide peaks. B. Mass deconvolution with unbiased assumption of molecular composition successfully deconvolved four glycopeptide peaks while failing to deconvolve many smaller glycopeptides and peptides. * IVNQ + 2Hex:2HexNAc ¤ IVNQ + 3Hex:2HexNAc ʁ IVNQ + 4Hex:2HexNAc ¥ IVNQ + 5Hex:2HexNAc

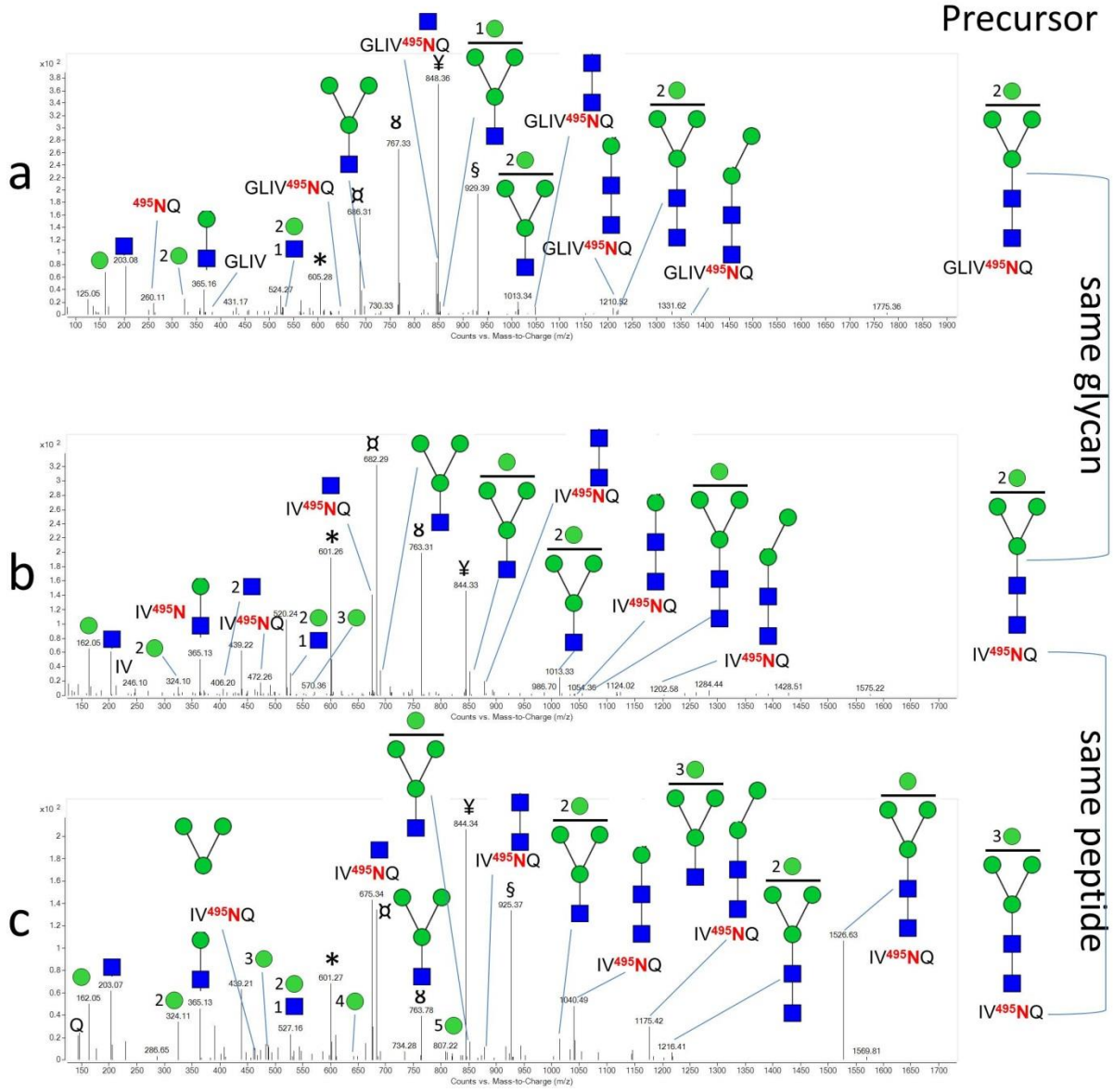


Figure S-5. Tandem mass spectra of site-specific N-glycopeptide analysis at site 495 of bovine lactoferrin. Glycan self-consistency is shown in a and b. Peptide self-consistency is shown in b and c. Blue squares: HexNAc, Green circles: Hex (mannose): * intact peptide + 2Hex:2HexNAc ¤ intact peptide + 3Hex:2HexNAc ϐ intact peptide + 4Hex:2HexNAc ¥ intact peptide + 5Hex:2HexNAc § intact peptide + 6Hex:2HexNAc

Eliminating Unlikely Glycopeptide Fragments

N-linked

1. if Hex > 0 → HexNAc ≥ 2
2. if HexNAc ≥ 4 → Hex ≥ 2
3. HexNAc ≥ dHex
4. if dHex ≥ 2 → Hex ≥ 2 & HexNAc ≥ 3
5. if NeuAc = 1 → Hex ≥ 3 & HexNAc ≥ 3
6. If NeuAc ≥ 2 → Hex ≥ 4 & HexNAc ≥ 4
7. Hex ≥ NeuAc

O-linked

1. HexNAc ≥ 1

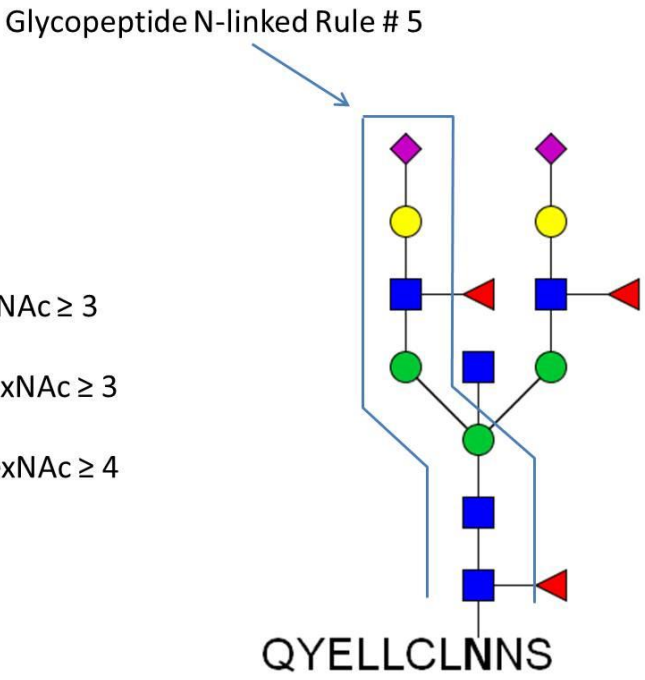


Figure S-6. Rules for fragments of glycopeptides that are allowed by GP Finder.

Composition Rules for Assigning Glycan Only Fragments

N-linked

1. HexNAc ≥ dHex
2. Hex ≥ NeuAc
3. if NeuAc > 1 → Hex ≥ 3 & HexNAc ≥ 2

O-linked

1. HexNAc + Hex ≥ NeuAc
2. HexNAc + Hex ≥ dHex

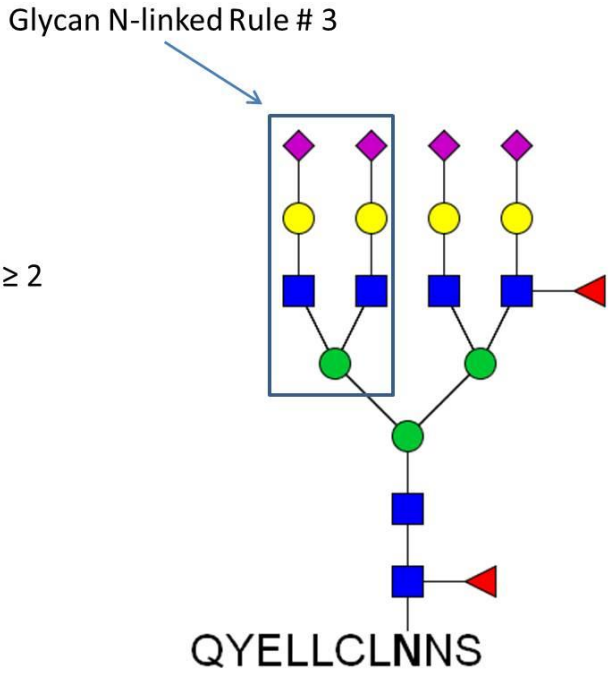


Figure S-7. Rules for fragments of glycans that are allowed by GP Finder.

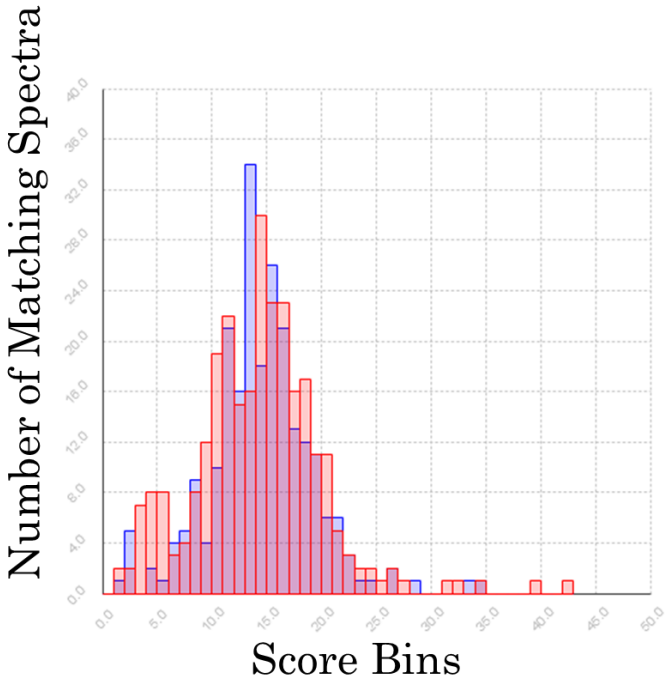


Figure S-8. Decoy Validation. Results from the dataset-bias test for the mixture of bovine fetuin, lactoferrin, and kappa casein. Score distribution of the target and decoy matches compete for top matches to false tandem spectra to reveal bias between the decoy and target scores. The two distributions are shown here to be approximately equivalent.

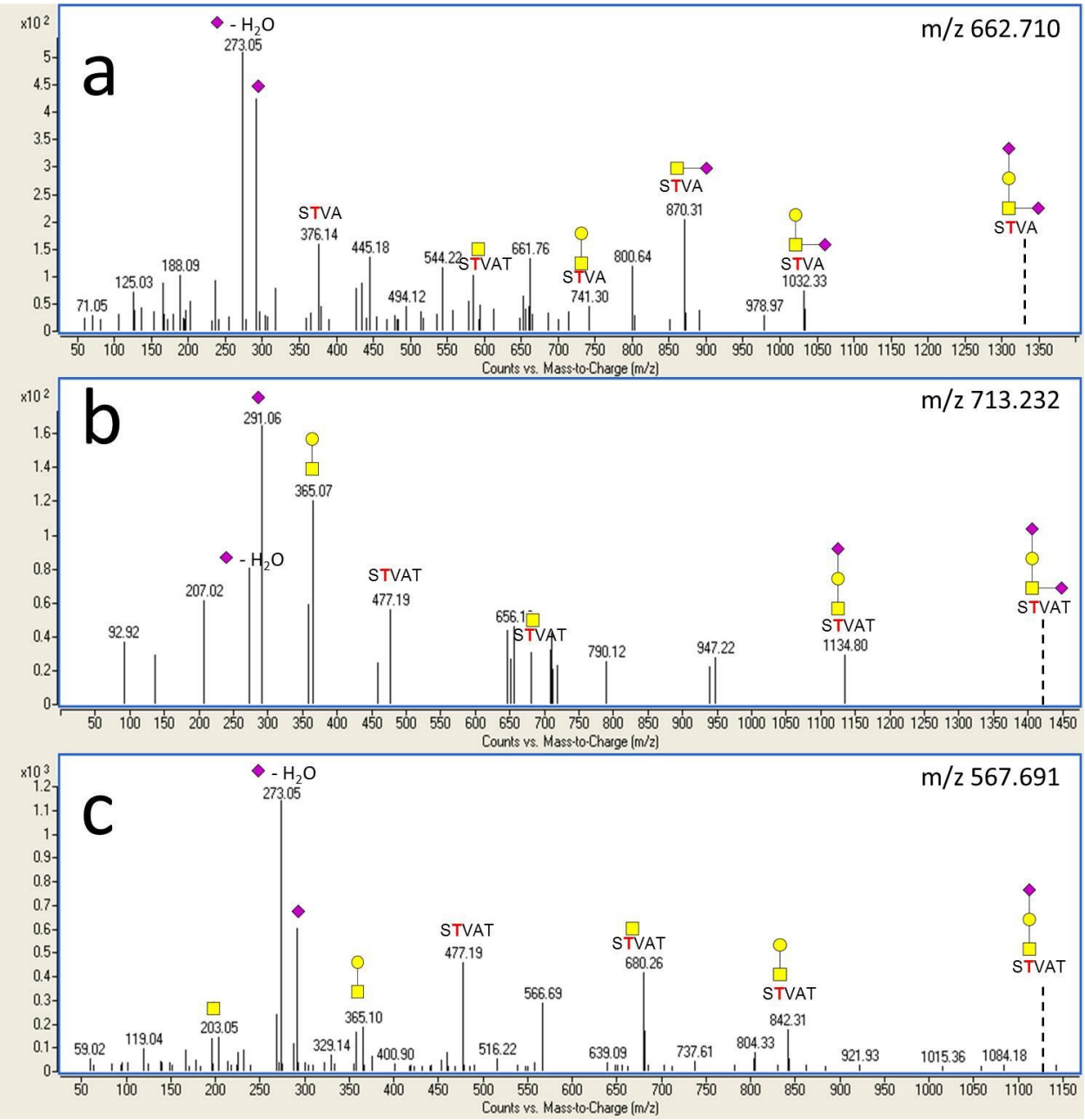


Figure S-9. Example of SSG with O-glycopeptides from kappa casein site 163. Glycan self-consistency is shown with a and b. Peptide self-consistency is shown with b and c.

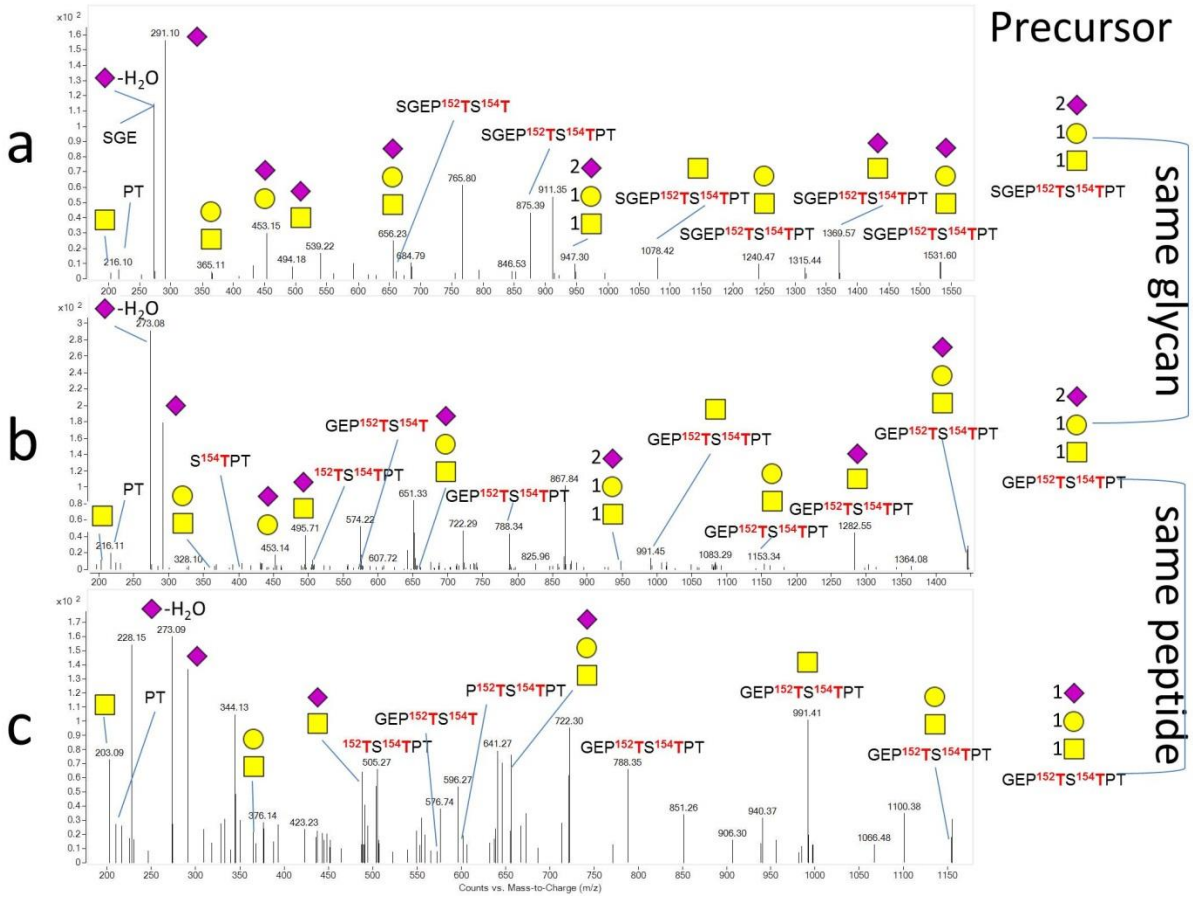


Figure S-10. Self-consistent Data. Tandem mass spectra of site-directed O-glycopeptide analysis at sites 152 and 154 of bovine kappa casein. Glycan self-consistency is shown in a and b. Peptide self-consistency is shown in b and c. Yellow squares: HexNAc, Yellow circles: Hex, Purple diamonds: NeuAc.

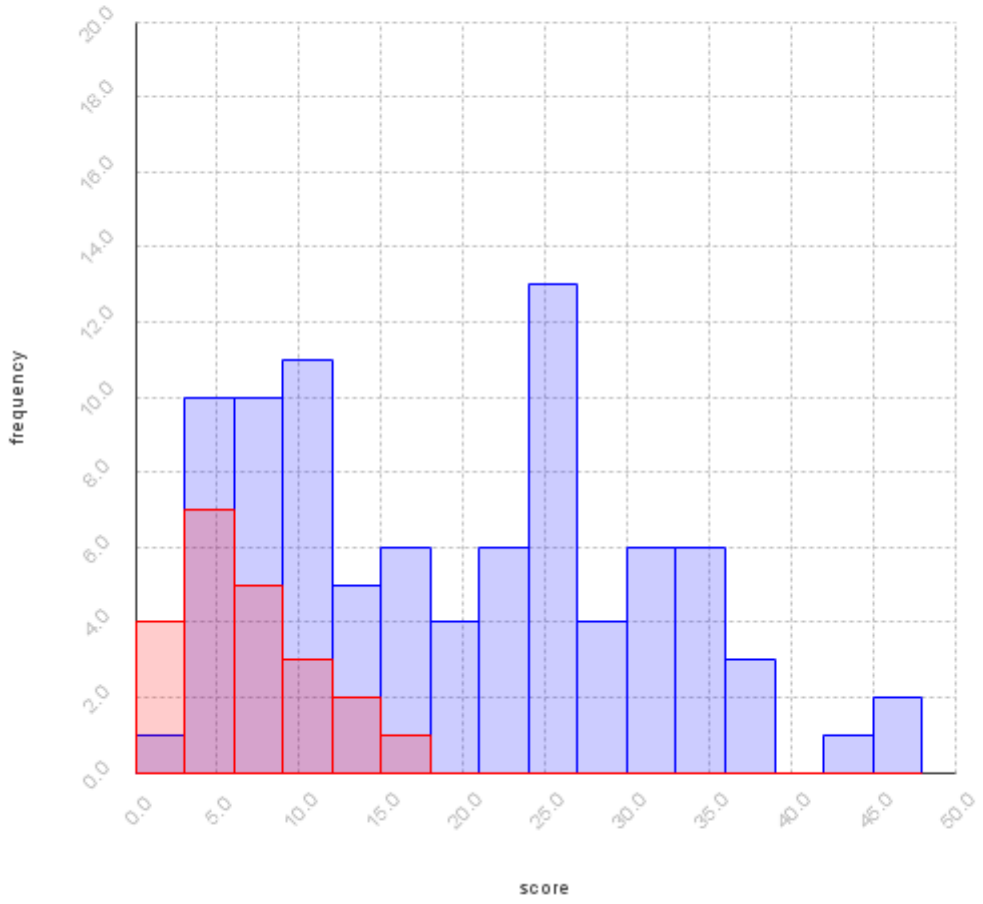


Figure S-11. RNaseB Scores. Histogram of all top-scoring matches for RNaseB. All manually assigned matches scored above the calculated 5% FDR threshold. Due to the limited data-points, the decoy and target distributions are rough estimates.

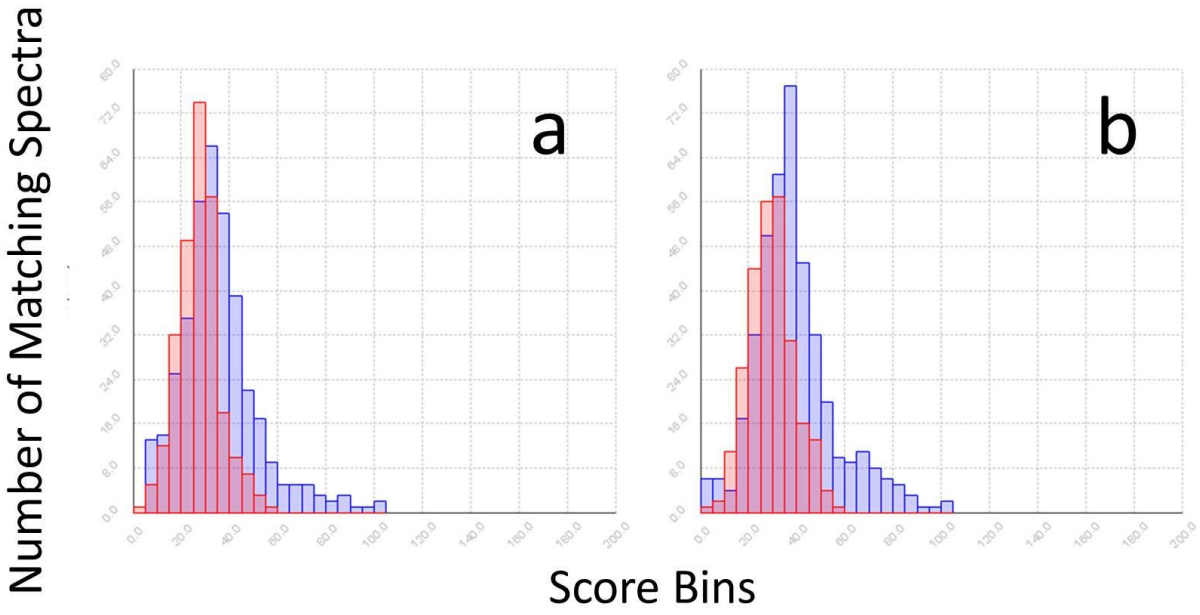


Figure S-12. Example of robust 5%FDR. Many legitimate spectra only match when both of the correct glycoproteins are considered. A. Analysis of cocktail data with sequence of kappa casein generates 36 matches over the score threshold of 55. B Analysis of cocktail data with sequences of kappa casein and bovine fetuin (both O-linked and known to be in sample) generates 61 matches over the score threshold of 55.

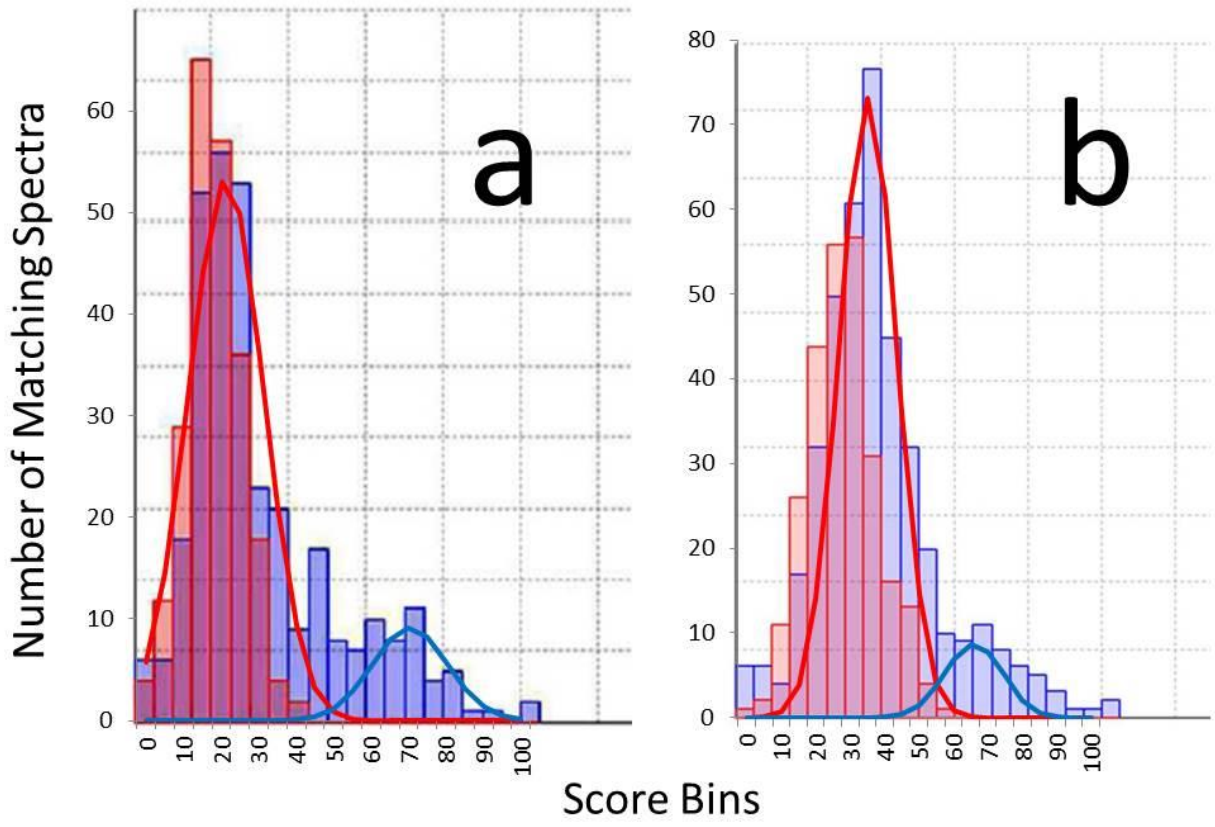


Figure S-13. Example of robust alternate technique for estimating FDR. The score distributions of the target datasets (blue histogram bars) were deconvoluted with commercial software (PeakFit) to estimate the key underlying true positive match (blue line) and negative match (red line) distributions. A. N-linked glycopeptide analysis of three-protein mixture. B. O-linked glycopeptide analysis of three-protein mixture.

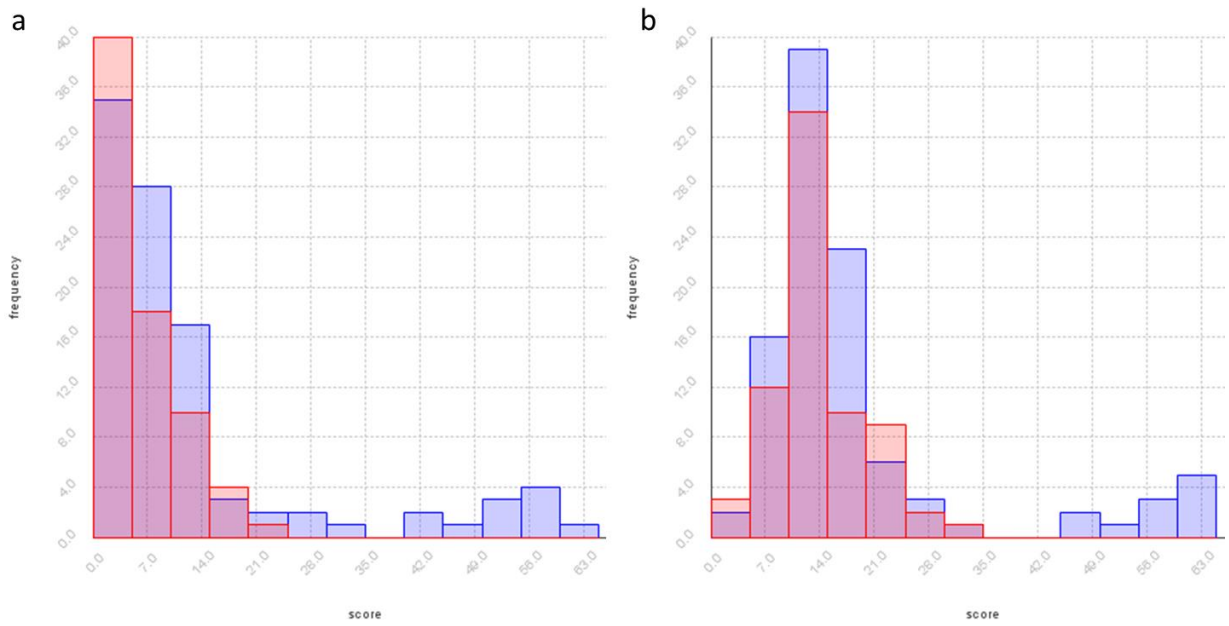


Figure S-14. vLDL Scores. Histograms of the target and decoy score distributions for two O-type glycoproteins (APOE and APOC3) extracted from vLDL nanoparticles. a. Without self-consistency scoring. b. With TDA-compliant self-consistency scoring. Due to the limited data-points, the decoy and target distributions are rough estimates even though the histogram show the actual results.

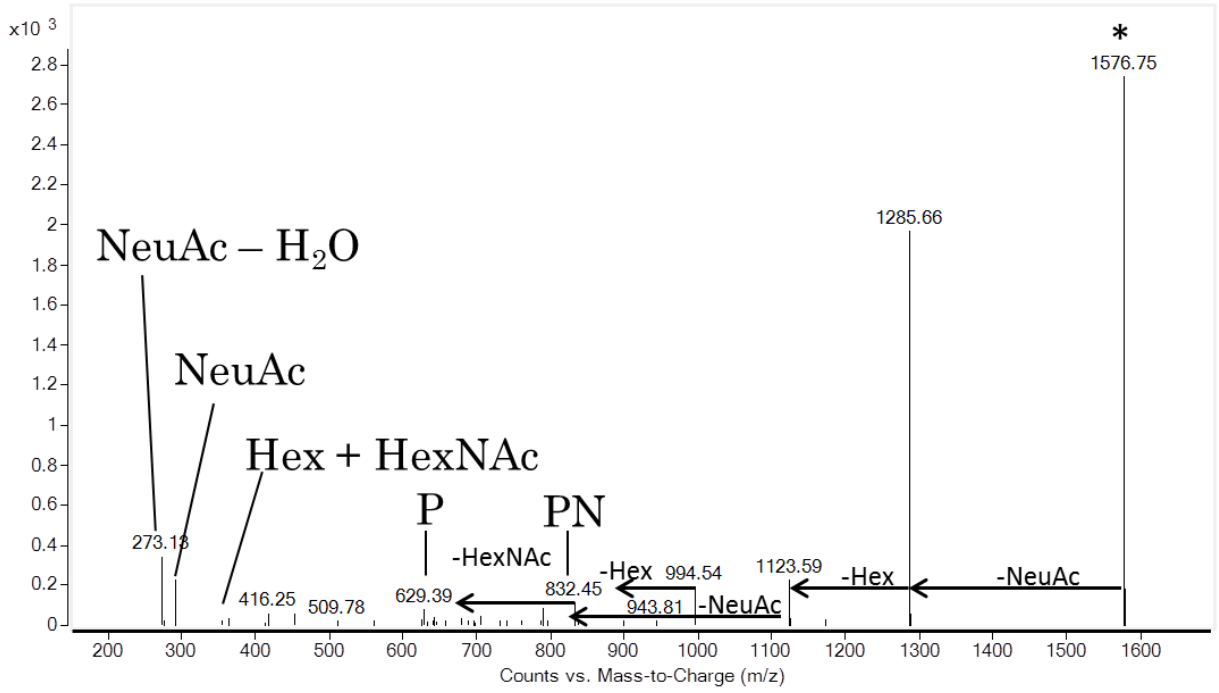
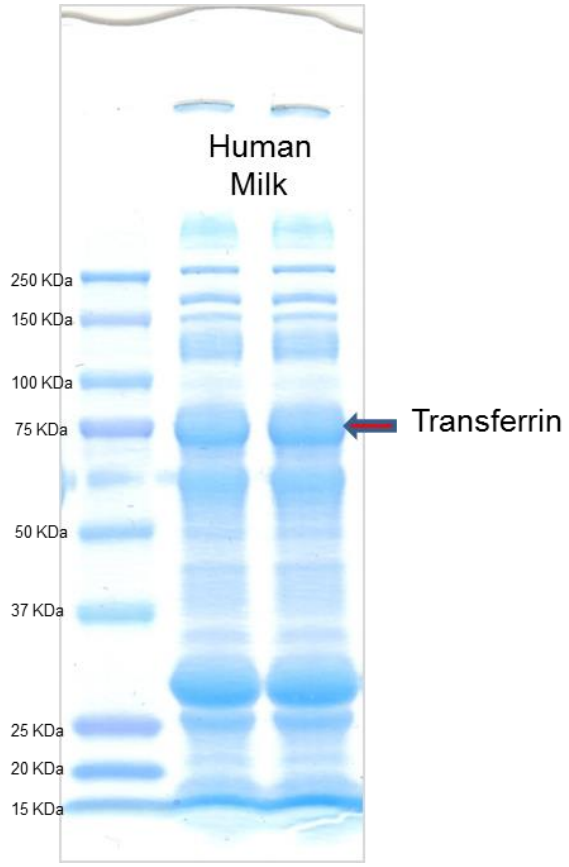


Figure S-15. Structure Analysis. Tandem mass spectrum (neutral mass) of an O-glycopeptide from site 94 of APOC3 that is composed of 1Hex:1HexNAc:2NeuAc + (V)RP**T**SA(V). The combination of P and PN, B- and Y-glycopeptide fragments and diagnostic oxonium ions provide evidence for the assignment.



Rapid INPEG Approach

- 1-D gel analysis of crude human milk
- Excise identified lactoferrin band
- Digest gel pieces with pronase
- Extract glycopeptide digests
- MS and Tandem MS/MS analyses

Figure S-16. INPEG. Scan of gel-separated proteins and glycoproteins from human milk. The newly developed strategy greatly simplifies mixtures. The strategy is called In-gel Nonspecific Proteolysis for Elucidating Glycoproteins (INPEG). Parallel specific and non-specific proteolysis can be performed from components within the same excised band or by excising components for each digestion from nearly identical bands from two parallel lanes.

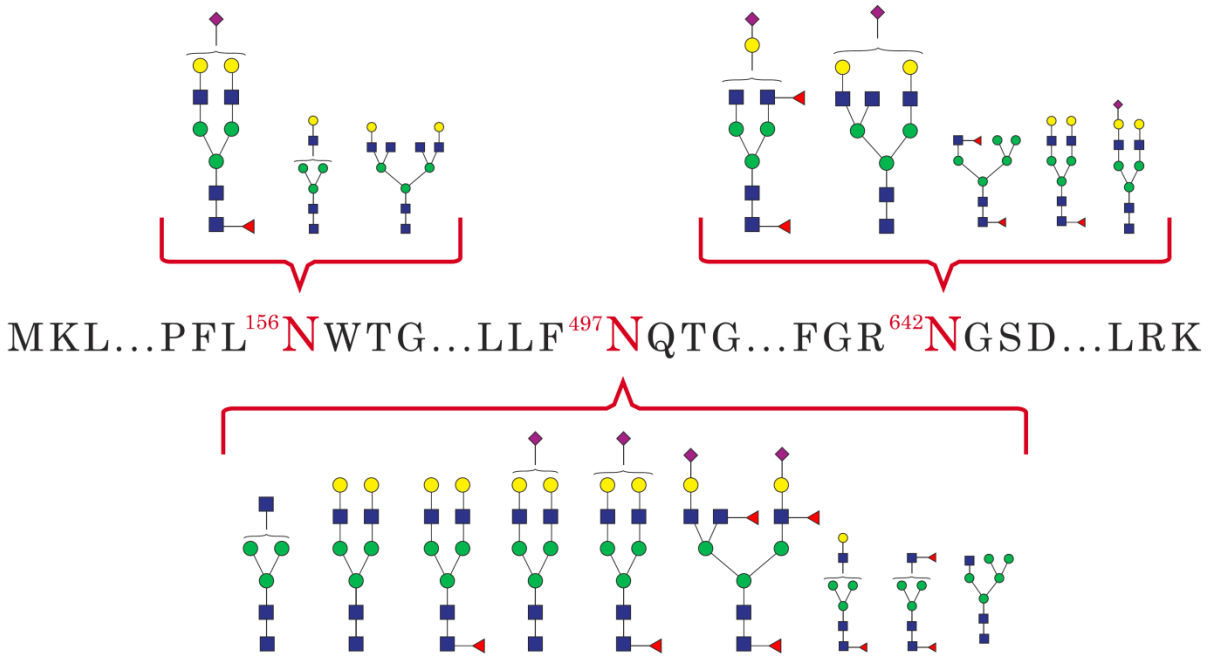


Figure S-17. SSG Map of an N-Glycoprotein. SSG analysis of lactoferrin from human milk. More abundant glycan species are indicated by greater size of glycan cartoon at each site. Blue squares: HexNAc, Yellow circles: Hex (galactose), Green circles: Hex (mannose), Purple diamonds: NeuAc, Red triangles: dHex (fucose).

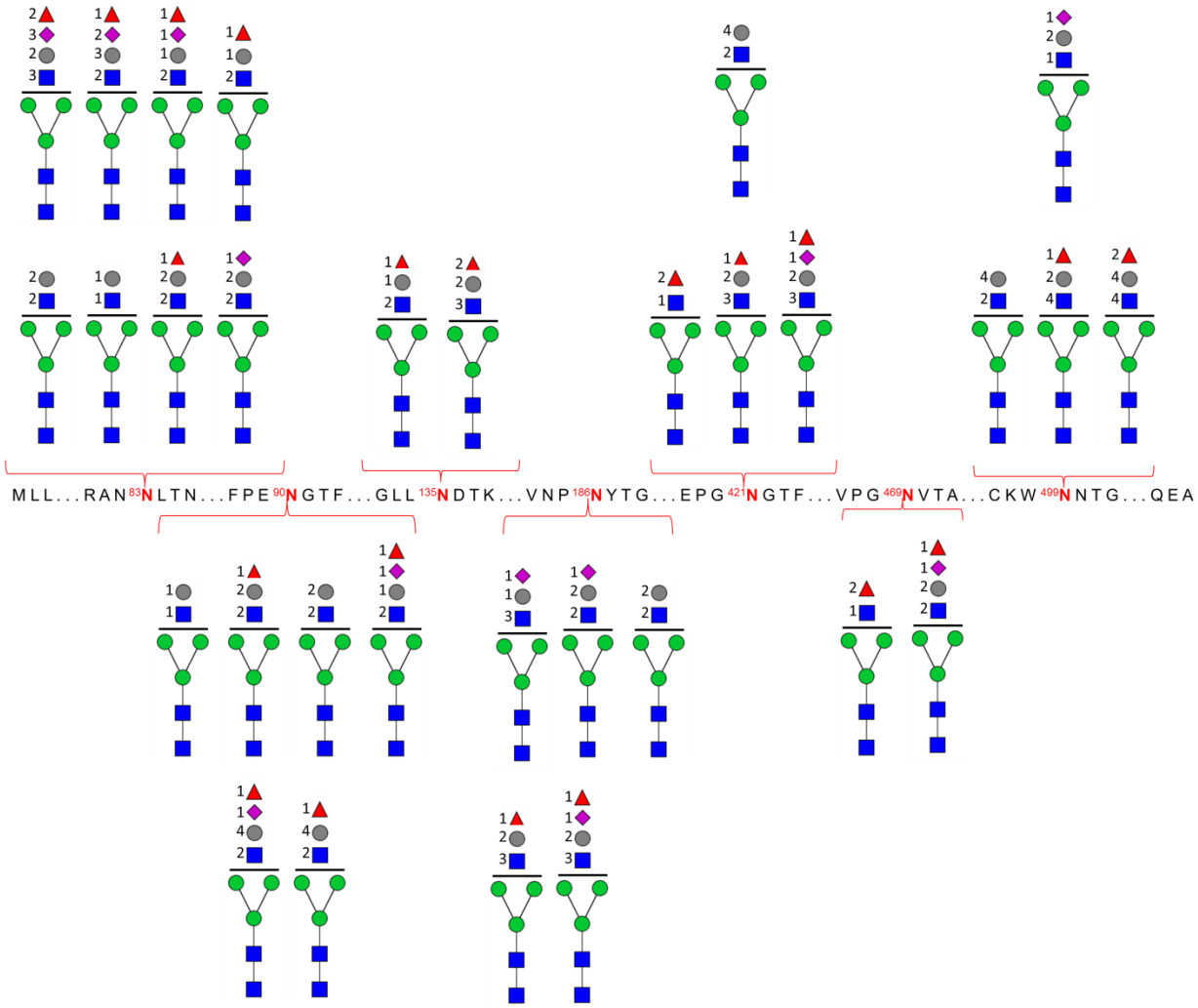


Figure S-18. SSG Map for Polymeric Immunoglobulin Receptor. SSG of polymeric immunoglobulin receptor from human milk, one of two glycoproteins isolated in a gel band after SDS-PAGE of raw human milk. Blue squares: HexNAc, Gray circles: Hex (galactose), Green circles: Hex (mannose), Purple diamonds: NeuAc, Red triangles: dHex (fucose)

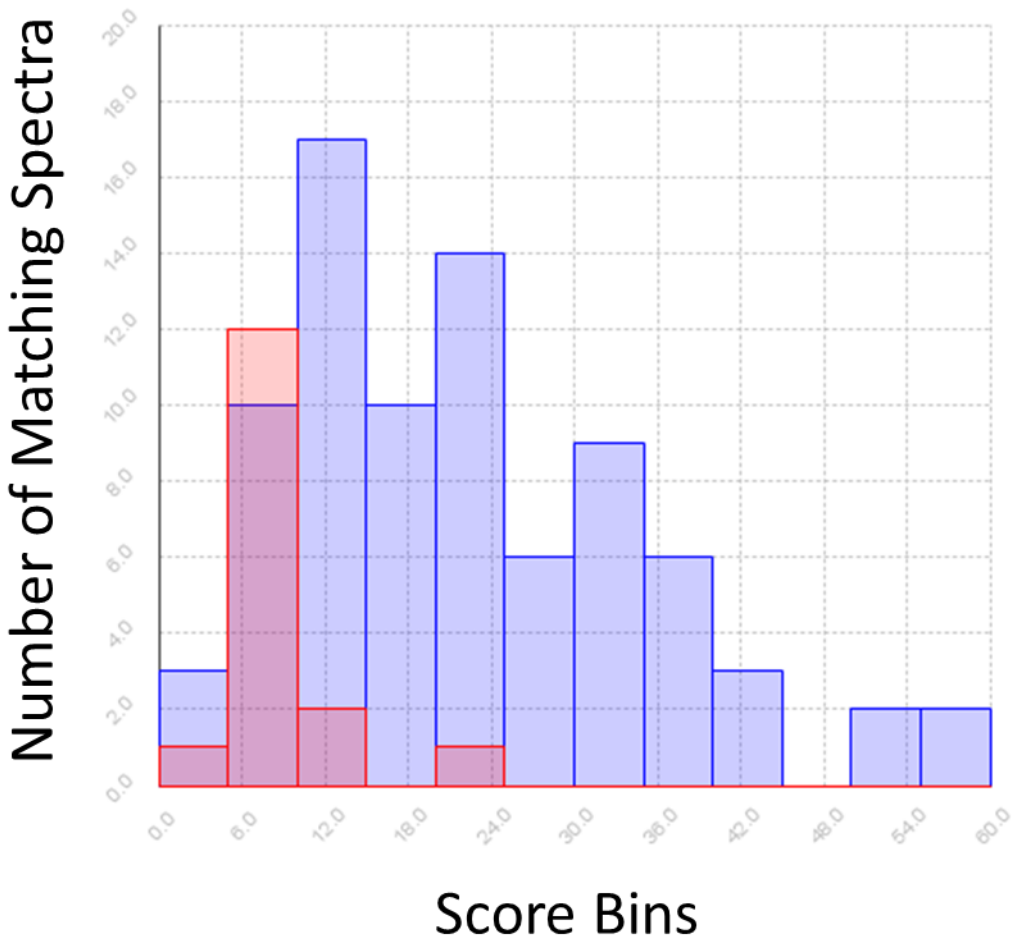


Figure S-19. Polymeric Immunoglobulin Receptor Scores. Histogram of top-scoring matches for human transferrin and polymeric immunoglobulin receptor, both isolated in a gel band after SDS-PAGE of raw human milk. Due to the limited data-points, the decoy and target distributions are rough estimates.

Figure S-20. *in silico* Non-Specific Digest Algorithm. Algorithm for the generation of theoretical peptide tags around glycosites with non-specific protease

```
//iterate through list of glycosites
for (int j = nextSiteIndex; j < glycositeNum; j++) {
    int n = 0;
    double extendedFragmentMass = 0;
    double fragmentMass = 0.0;
    int residueShiftNum = 0;
    int glycosite = (Integer) glycosites.get(j);
    String fragmentResidue = "";
    String extendedFragmentResidues = "";
    do {
        int s = 0;//variable s determines the number of amino acids to the left of the current center
        //current center starts with glycosite and moves to the right with "residueShiftNum++"
        do {
            if (extendedFragmentMass == 0) {//first
                fragmentMass = nextRightMass(glycosite, residueShiftNum, annotatedSequenceList);
                if (fragmentMass >= highGlycopeptideMass) {
                    break;
                }
                if (residueShiftNum <= rightResidueLimit) {//conditional statement allows the user to limit the number residues searched to right of glycosite
```

```
        fragmentResidue = nextRightResidue(glycosite, residueShiftNum, annotatedSequenceList);
        extendedFragmentMass = fragmentMass;
        extendedFragmentResidues = fragmentResidue;
    } else {
        break;
    }
}
ArrayList data = new ArrayList();
//add current peptide information to a self-contained list that will be added to the list of theoretical peptides
data.add(glycosite + 1);//glycosites
data.add(extendedFragmentMass);//peptide mass (theoretical used also as experimental)
data.add(extendedFragmentResidues);//amino acid sequence of peptide backbone
data.add(sequenceFileName);//name of protein sequence file name (FASTA formated file)
data.add(s);//position of site on peptide
peptideCalc.add(k, data);
s++;//current distance on left of center (reference) amino acid, iterates toward the left
k++;//count of generated peptides
if ((glycosite - s) >= 0) {
    if (s <= leftResidueLimit) {//conditional statement allows the user to limit the number residues searched to left of glycosite
        extendedFragmentMass += nextLeftMass(glycosite, s, annotatedSequenceList);
```

```
                extendedFragmentResidues = nextLeftResidue(glycosite, s, annotatedSequenceList) + extendedFragmentResidues;
            } else {
                break;
            }
        } else {
        }
    //stop searching to left if current peptide mass is larger than largest experimental glycopeptide mass or if left side of protein sequence is reached
    } while (extendedFragmentMass < highGlycopeptideMass && (glycosite >= s));
if (((aminoAcidSequenceLen - 1) - glycosite) > residueShiftNum) {//deal with glycosites close to the right end of the protein sequence
        residueShiftNum++;//iterate to the right of the glycosite
        if (residueShiftNum <= rightResidueLimit) {//conditional statement allows the user to limit the number residues searched to right of glycosite
            fragmentMass += nextRightMass(glycosite, residueShiftNum, annotatedSequenceList);
            fragmentResidue += nextRightResidue(glycosite, residueShiftNum, annotatedSequenceList);
            extendedFragmentMass = fragmentMass;
            extendedFragmentResidues = fragmentResidue;
        } else {
            break;
        }
    } else {
        residueShiftNum++;
```

```
        }
    } while ((fragmentMass < highGlycopeptideMass) && (((aminoAcidSequenceLen) - glycosite) > residueShiftNum));
}
int extendedFragmentMassIndex = 1;
Collections.sort(peptideCalc, new ListComparator(extendedFragmentMassIndex, true));
//provide the mass of the next amino acid to the right
public double nextRightMass(int glycosite, int residueShiftNum, ArrayList annotatedSequenceList) {
    double residueMass = (Double) ((ArrayList) annotatedSequenceList.get(glycosite + residueShiftNum)).get(1);
    return residueMass;
}
//provide the letter of the next amino acid to the right
public String nextRightResidue(int glycosite, int residueShiftNum, ArrayList annotatedSequenceList) {
    String residue = (String) ((ArrayList) annotatedSequenceList.get(glycosite + residueShiftNum)).get(0);
    return residue;
}
//provide the mass of the next amino acid to the left
public double nextLeftMass(int glycosite, int m, ArrayList annotatedSequenceList) {
    double residueMass = (Double) ((ArrayList) annotatedSequenceList.get(glycosite - m)).get(1);
    return residueMass;
}
```

```
//provide the letter of the next amino acid to the right
public String nextLeftResidue(int glycosite, int m, ArrayList annotatedSequenceList) {
    String residue = (String) ((ArrayList) annotatedSequenceList.get(glycosite - m)).get(0);
    return residue;
}
```

Figure S-21.Determination of Experimental Glycan Mass Algorithm. Algorithm for the calculation of all potential glycan masses by subtracting theoretical peptide masses from experimental glycopeptide masses

```
for (int i = nextMassIndex; i < massListLen; i++) {
   j = 0;
   double mass = ((Double) ((ArrayList) massList.get(i)).get(0)) * charge;
   if (mass > ((Double) ((ArrayList) peptideCalc.get(j)).get(GlobalConstants.peptideMassIndex))) {
      do {
         //Add on the OH so that the sugar part forms a complete aldehyde
         GlycanMass = mass - (((Double) ((ArrayList) peptideCalc.get(j)).get(GlobalConstants.peptideMassIndex)));
         ArrayList data = new ArrayList();
         data.add(mass);//experimentally observed glycopeptide mass
         try {
            intensity = (Double) ((ArrayList) massList.get(i)).get(1);
         } catch (Exception e) {
            intensity = -1;
         }
         data.add(intensity);//intensity
         data.add(GlycanMass);//residual mass (will compare with glycan masses)
         data.add((((ArrayList) peptideCalc.get(j)).get(3)));//sequenceFileName
         data.add((((ArrayList) peptideCalc.get(j)).get(0)));//glycosite
         data.add((((ArrayList) peptideCalc.get(j)).get(1)));//peptide mass
```

```
            data.add((((ArrayList) peptideCalc.get(j)).get(2)));//peptide sequence
            data.add((((ArrayList) peptideCalc.get(j)).get(4)));//site index
            if (includeRetentionTime) {
               try {
                  data.add((Double) ((ArrayList) massList.get(i)).get(2));//data.add(Double.parseDouble((String)((ArrayList)massList.get(i)).get(2)));
               } catch (Exception e) {
                  data.add(-1);
               }
            } else {
               data.add(-1);
            }
            peptideExp.add(k, data);
            k += 1;
            j += 1;
         } while ((j < (peptideCalc.size())) && (mass > ((Double) ((ArrayList) peptideCalc.get(j)).get(GlobalConstants.peptideMassIndex))));
      } else {
      }
   }
   Collections.sort(peptideExp, new ListComparator(GlobalConstants.glycanMassIndex, true));
```

Figure S-22. Exhaustive Glycan Library Algorithm. Algorithm for the exhaustive generation of possible glycans

```
Double thisMass;
for (i = iFrom; i <= iTo; i++) {
   for (j = jFrom; j <= jTo; j++) {
      for (k = kFrom; k <= kTo; k++) {
         for (m = mFrom; m <= mTo; m++) {
            for (n = nFrom; n <= nTo; n++) {
               for (o = oFrom; o <= oTo; o++) {
                  for (p = pFrom; p <= pTo; p++) {
                     for (q = qFrom; q <= qTo; q++) {
                        for (r = rFrom; r <= rTo; r++) {
                           thisMass = (Double) ((ArrayList) ResidueParameters.get(0)).get(4) * ((double) i)
                                                                    + (Double) ((ArrayList) ResidueParameters.get(1)).get(4) * ((double) j)
                                                                    + (Double) ((ArrayList) ResidueParameters.get(2)).get(4) * ((double) k)
                                                                    + (Double) ((ArrayList) ResidueParameters.get(5)).get(4) * ((double) m)
                                                                    + (Double) ((ArrayList) ResidueParameters.get(3)).get(4) * ((double) n)
                                                                    + (Double) ((ArrayList) ResidueParameters.get(4)).get(4) * ((double) o)
                                                                    + (Double) ((ArrayList) ResidueParameters.get(6)).get(4) * ((double) p)
                                                                    + (Double) ((ArrayList) ResidueParameters.get(7)).get(4) * ((double) q)
                                                                    + (Double) ((ArrayList) ResidueParameters.get(8)).get(4) * ((double) r);
```

```
        if (thisMass == 0 || ((highGlycanMass > 0.001) && (thisMass.compareTo(highGlycanMass + errorLimit) > 0))) {
        }
        else {
            ArrayList data = new ArrayList();
            data.add(thisMass);//theoretical glycan mass
            data.add(i);//Hex
            data.add(j);//HexNAc
            data.add(k);//DxyHex
            data.add(n);//NeuAc
            data.add(o);//NeuGC
            data.add(m);//Pent
            data.add(p);//KDN
            data.add(q);//HexA
            data.add(r);//user defined A
            data.add(glycanIDNum++);
            glycanCalc.add(L, data);
            L++;
        }
    }
}
```

```
                    }
                }
            }
        }
    }
  }
}
```

Collections.sort(glycanCalc, new ListComparator(0, true));//sortObj.quickSort(glycanCalc, 0);

Figure S-23. Matching Theoretical and Experimental Possibilities Algorithm. Algorithm for the comparison of the theoretical glycan masses and the residual mass of the experimental data after subtracting the theoretical peptides

```
if (errorType.equals("ppm")) {
    errorFactor = errorLimit / 1000000;
    errorIsPPM = true;
} else {
    epsilonStandard = errorLimit;
    errorIsPPM = false;
}
for (i = 0; i < tempLongListLen; i++) {
    residualGlycan = (Double) ((ArrayList) peptideExp.get(i)).get(2);
    for (k = lowest; k < tempGlycanListLen; k++) {
        realGlycan = (Double) ((ArrayList) glycanCalc.get(k)).get(0);
        realGlycanWithCharges = realGlycan * charge;
        if (errorType.equals("ppm")) {
            epsilonStandard = errorFactor * (realGlycanWithCharges + (Double) ((ArrayList) peptideExp.get(i)).get(5));//get GPMass
        }
        if ((realGlycanWithCharges - epsilonStandard) > residualGlycan){
            break;
        }
        if ((realGlycanWithCharges + epsilonStandard) < residualGlycan){
```

```
            lowest = k + 1;
            continue;
        }
        if (((realGlycanWithCharges - epsilonStandard) <= residualGlycan) && ((realGlycanWithCharges + epsilonStandard) >= residualGlycan)){
            ArrayList data = new ArrayList();
            data.add((((ArrayList) peptideExp.get(i)).get(0)));//0 experimentally observed glycopeptide mass
            data.add((((ArrayList) peptideExp.get(i)).get(1)));//1 intensity
            data.add((((ArrayList) peptideExp.get(i)).get(2)));//2 residual glycan mass (will compare with glycan masses)
            data.add(realGlycan);//theoretically calculated glycan mass
            if (errorIsPPM) {
                data.add(Math.abs(((realGlycanWithCharges - residualGlycan) / (realGlycanWithCharges + (Double) ((ArrayList) peptideExp.get(i)).get(5))) *
1000000));
            }//ppm error
            else {
                data.add(Math.abs(realGlycanWithCharges - residualGlycan));
            }//Dalton error
            data.add(((ArrayList) glycanCalc.get(k)).get(1));//4 ChargeCarrier
            data.add(((ArrayList) glycanCalc.get(k)).get(2));//5 Hex
            data.add(((ArrayList) glycanCalc.get(k)).get(3));//6 HexNAc
            data.add(((ArrayList) glycanCalc.get(k)).get(4));//7 DxyHex
            data.add(((ArrayList) glycanCalc.get(k)).get(5));//8 NeuAc
```

```
data.add(((ArrayList) glycanCalc.get(k)).get(6));//9 NeuGC
data.add(((ArrayList) glycanCalc.get(k)).get(7));//10 Pent
data.add(((ArrayList) glycanCalc.get(k)).get(8));//11 KDN
data.add(((ArrayList) glycanCalc.get(k)).get(9));//12 HexA
int userDef = ((Integer) ((ArrayList) glycanCalc.get(k)).get(10)).intValue();
data.add(userDef);//13 user defined A
String protName = (String) (((ArrayList) peptideExp.get(i)).get(3));
if (userDef >= 1 && runAutomatedDecoy == true) {
   protName = "DECOY_" + protName;
}
data.add(protName);//14 sequenceFileName
data.add((((ArrayList) peptideExp.get(i)).get(4)));//15 glycosite
data.add((((ArrayList) peptideExp.get(i)).get(5)));//16 observed peptide mass (same as theoretical peptide mass)
data.add((((ArrayList) peptideExp.get(i)).get(6)));//17 peptide sequence
data.add(((ArrayList) glycanCalc.get(k)).get(11));//18
data.add((((ArrayList) peptideExp.get(i)).get(7)));
FindGlycopeptides.serialNumber = serialNumber;
data.add((((ArrayList) peptideExp.get(i)).get(8)));//20 retention time
glycanExp.add(j, data);
j++;
```

```
        }
    }
}
Collections.sort(glycanExp, new ListComparator(0, true));
```

Figure S-24. Algorithms for removing glycan and glycopeptide fragments that are unlikely candidates in light of biological glycosylation profiles.

```
class isAcceptableGlycan
    //returns false unless a specified condition is met
    PossibilitiesObject posObj = (PossibilitiesObject) ((ArrayList) tanPar.getPO()).get(0);

    if isNLinked

      if NeuAc or NeuGc <=Hexose
        if Deoxyhexose<=HexNac
          if NeuAc or NeuGc > 1
            if Hex >= 3
              if HexNAc >= 2
                return true;
          else
            return true;
      else
        if Hex = 0
          return true;

    if isOLinked
      if (NeuAc + NeuGc) = 1 and all others = 0
        return true
      if dHex = 1 and all others = 0
        return true
      if NeuAc + NeuGc + dHex <= Hexose + HexNAc
        return true

    return false

class isAcceptableGlycopeptide
    //returns true unless a specified case is met
    if isNLinked
      //rules for NeuAc
      switch NeuAc
```

```
   case 0:
      break
   case 1:
      if !(Hex >= 3)
         return false
      if !(HexNAc >= 3)
         return false
      break
   default:
      if !(Hex >= 4)
         return false
      if !(HexNAc >= 4)
         return false
      if NeuAc > Hex
         return false

      break

//rules for NeuGC
switch NeuGc
   case 0:
      break;
   case 1:
      if !(Hex >= 3)
         return false
      if !(HexNAc >= 3)
         return false
      break
   default:
      if !(Hex >= 4)
         return false
      if !(HexNAc >= 4)
         return false
      if NeuGC > Hex
```

```
            return false

          break

    if !(Hex >= NeuAc + NeuGc)
       return false

    if dHex >=2
       if Hex < 2
          return false
       if HexNAc < 3
          return false

    if !(HexNAc > dHex)
       return false

    if HexNAc >= 3
       if !(Hex >= 1)
          return false

    if HexNAc >= 4
       if !(Hex > 1)
          return false

    if Hex >= 1
       if !(HexNAc >= 2)
          return false

if isOLinked
    if !((Hex + HexNAc) >= NeuAc + NeuGc + dHex)
       return false
    if !(HexNAc > 0)
       return false
```

```
return true
```

Figure S-25. Scoring algorithm for tandem MS fragments.

```
class ScoreGenerator

    ArrayList decoyFamSizeListRandom
    ArrayList averFamDecoyScoreListRandom
    ArrayList randomTargetFamSizeList
    ArrayList averFamRandomTargetScoreList
    ArrayList targetFamSizeList
    ArrayList averFamTargetScoreList

    for each mass
      for each retention time
        for each theoretical glycopeptide
            score = (completePeptideCount * score4CompletePeptide)
                + completePeptidePlusHexNAcCount * score4CompletePeptidePlusHexNAc)
                * (completePeptideCount + completePeptidePlusHexNAcCount)
                + (peptideFragmentCount * score4PeptideFragment)
                + (intactPeptideWithFragmentedGlycanCount * score4IntactPeptideWithFragmentedGlycanCount)
                + (fragmentedPeptideWithFragmentedGlycanCount * score4FragmentedPeptideWithFragmentedGlycanCount)
                + (glycanFragmentCount * score4GlycanFragment)
            baseScore = score.getValue
            if useSelfConsistency()
              peptideFamilyCount = getPeptideFamilyCount(curRH);
              if peptideFamilyCount >= minimumFamilyCount
                score += peptideFamilyCount * perMemberScore
            if useNeutralLoss
              score += neutralLossCount * 10
            if useDiagnosticIonBoost(
              score += getDiagnosticOxoniumScore
            if useWaterLoss
              score += waterLossCount * 10;
            if removeData
              if score >= scoreThreshold
```

```
                keepDataObj4CurrentTheoreticalGlycopeptide(true)
            if isDecoy
               decoyFamSizeList.add(peptideFamilySize)
               averFamDecoyScoreList.add(score)

int decoyFamSizeSum = 0;
for (int i = 0; i < decoyFamSizeList.size(); i++)
   decoyFamSizeSum += decoyFamSizeList.get(i)
if decoyFamSizeList.size() > 0
   decoyFamSize = decoyFamSizeSum / decoyFamSizeList.size()
double decoyScoreSum = 0.0
for (int i = 0; i < averFamDecoyScoreList.size(); i++)
   decoyScoreSum += averFamDecoyScoreList.get(i)
if (averFamDecoyScoreList.size() > 0)
   averFamDecoyScore = decoyScoreSum / averFamDecoyScoreList.size()

if applyAverageScore
   for each mass
      for each retention time
         for each theoretical glycopeptide
            if useDiagnosticIonBoost
               score += getDiagnosticOxoniumScore
            if peptideFamilyCount >= minimumFamilyCount
               score += peptideFamilyAverage
            if removeData
               if score >= scoreThreshold
                  keepDataObj4TheoreticalGlycopeptide(true)

   for each mass
      for each retention time
         for each theoretical glycopeptide
            if !isDecoy
               targetFamSizeList.add(peptideFamilyCount)
               averFamTargetScoreList.add(getBaseScore);
```

```
            if peptideFamilyCount == decoyFamSize
               randomTargetFamSizeList.add(peptideFamilyCount)
               averFamRandomTargetScoreList.add(getBaseScore)
         else
            decoyFamSizeListAver.add(peptideFamilyCount);
            averFamDecoyScoreListAver.add(getBaseScore)

int decoyFamSizeSum = 0;
for (int i = 0; i < decoyFamSizeList.size(); i++)
   decoyFamSizeSum += decoyFamSizeList.get(i)
if decoyFamSizeList.size() > 0
   decoyFamSize = decoyFamSizeSum / decoyFamSizeList.size()
double decoyScoreSum = 0.0
for (int i = 0; i < averFamDecoyScoreList.size(); i++)
   decoyScoreSum += averFamDecoyScoreList.get(i)
if (averFamDecoyScoreList.size() > 0)
   averFamDecoyScore = decoyScoreSum / averFamDecoyScoreList.size()


int randomTargetFamSizeSum = 0;
for (int i = 0; i < randomTargetFamSizeList.size(); i++)
   randomTargetFamSizeSum += randomTargetFamSizeList.get(i)
if randomTargetFamSizeList.size() > 0
   targetFamSizeRandom = randomTargetFamSizeSum / randomTargetFamSizeList.size()
double randomTargetScoreSum = 0.0
for (int i = 0; i < averFamRandomTargetScoreList.size(); i++)
   randomTargetScoreSum += averFamRandomTargetScoreList.get(i)
if (averFamRandomTargetScoreList.size() > 0)
   averFamRandomTargetScore = randomTargetScoreSum / averFamRandomTargetScoreList.size()


int targetFamSizeSum = 0;
for (int i = 0; i < targetFamSizeList.size(); i++)
   targetFamSizeSum += targetFamSizeList.get(i)
```

```
if targetFamSizeList.size() > 0
   targetFamSizeRandom = targetFamSizeSum / targetFamSizeList.size()
double targetScoreSum = 0.0
for (int i = 0; i < averFamTargetScoreList.size(); i++)
   targetScoreSum += averFamTargetScoreList.get(i)
if (averFamTargetScoreList.size() > 0)
   averFamTargetScore = targetScoreSum / averFamTargetScoreList.size()

double bias = averFamTargetScoreRandom - averFamDecoyScoreAver;

if applyAverageScore && tdaCompliant
   for each mass
      for each retention time
         for each theoretical glycopeptide
            if !isDecoy
               score -= bias;
```

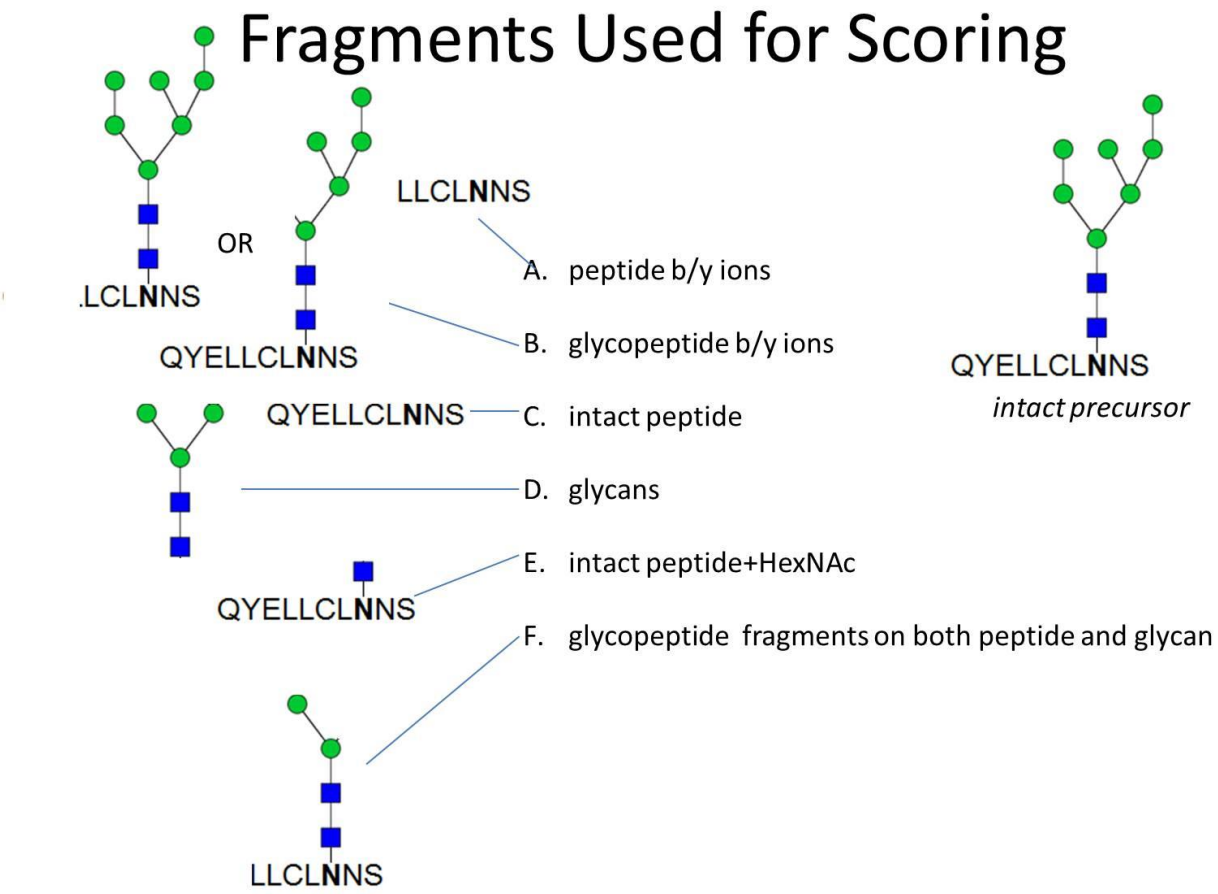


Figure S-26. Fragment types that are scored by GP Finder.

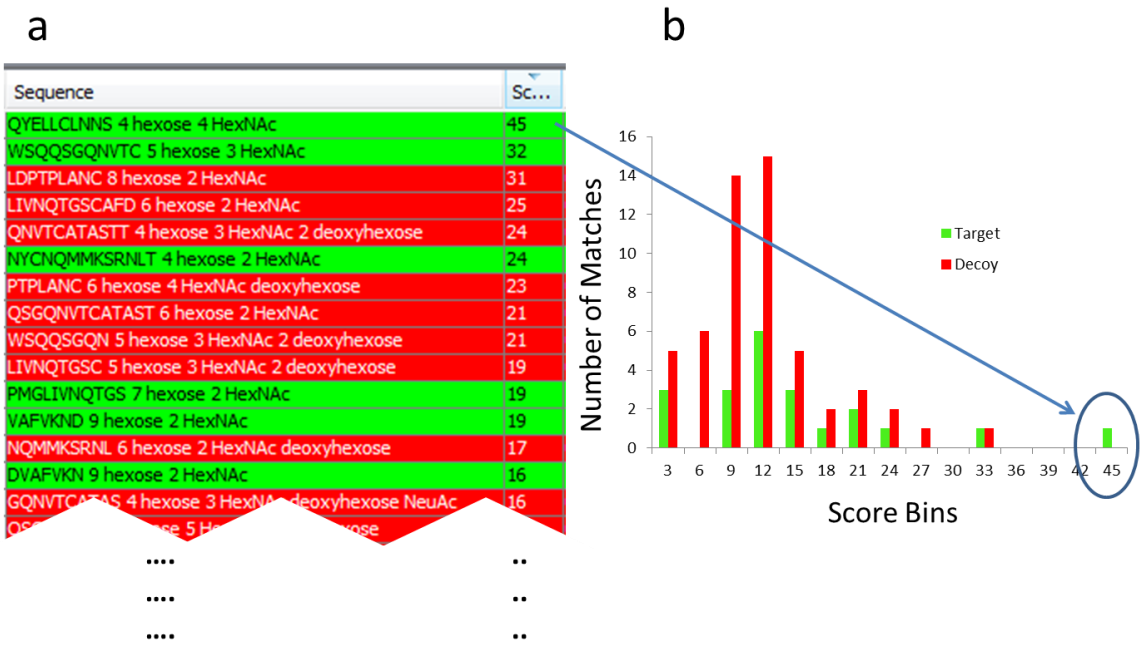


Figure S-27. Matching Scores to a Single Spectrum. The scores matching to a single tandem mass spectrum are shown here from one of the data points in Figure 1a. a. Abbreviated list of theoretical matches to a precursor mass. b. Histogram of all scores matching the precursor mass.

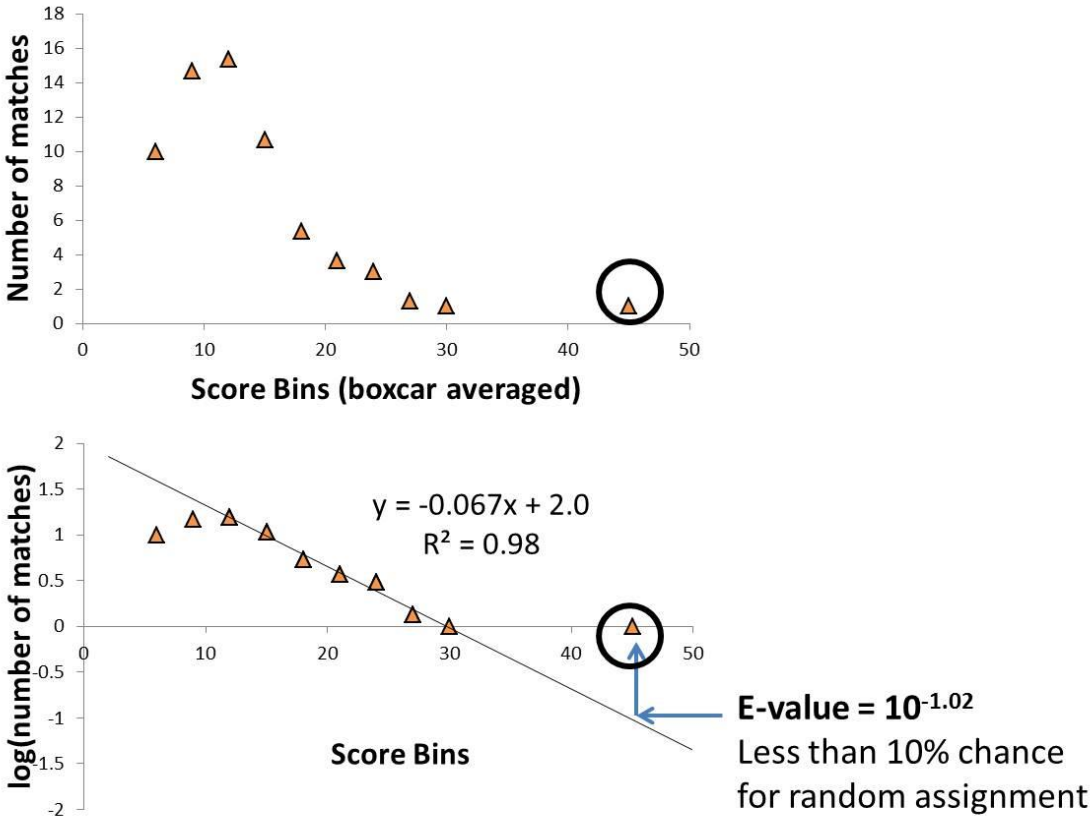


Figure S-28. E-value of Top Score. Demonstration of E-value calculation for the histogram in Figure S-27. a. Normal scale shows the exponentially decreasing tail. b. log-scale allows calculation of linear fit and prediction of the likelihood of observing a random match at the top score, i.e., the likelihood that the top scoring match for a particular spectrum is a part of the distribution of random matches.

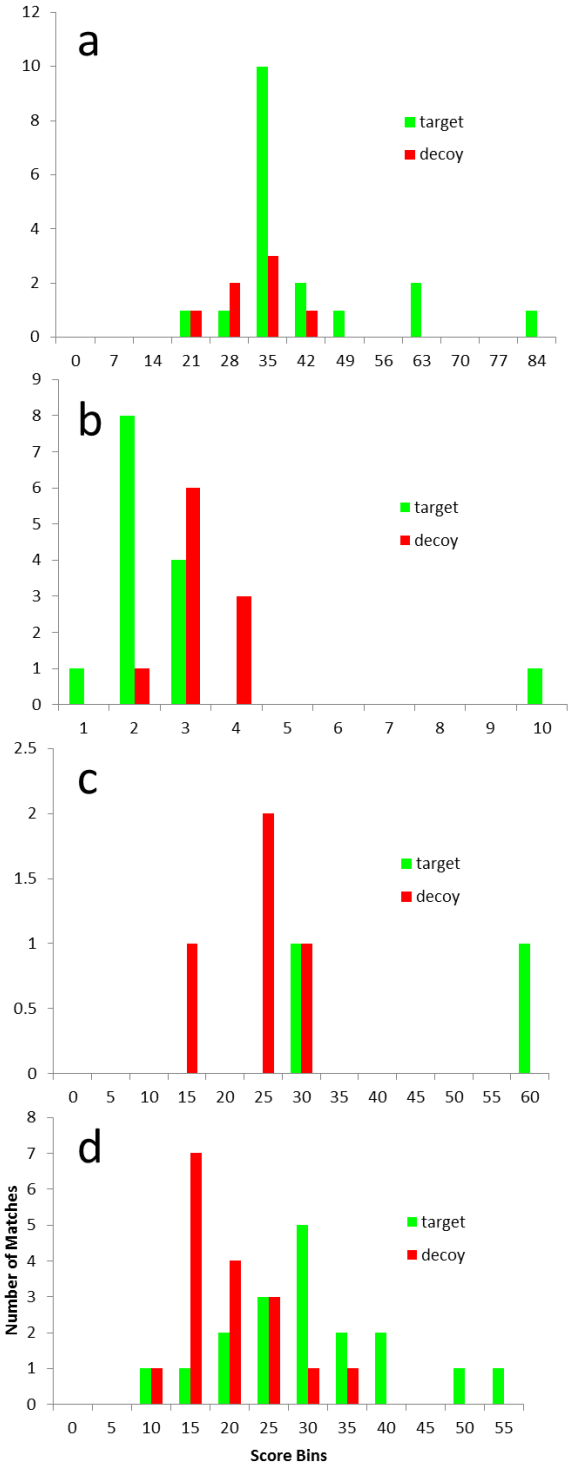


Figure S-29. Matches to Individual Spectra. Representative score distributions of the matches for four unrelated tandem spectra from four data points in Figure 1c. The top score was distinguished from the second highest score by the presence of P+PN and by a large peptide family with high-scoring members. Some top-assignments are ambiguous, as shown in d.

Table S-1. Self-Consistent Data. Self-consistency matches confirming a glycoform (Hex:HexNAc:2NeuAc) for CASK_BOVIN on site 154.

GP Mass	Hex	HexNAc	NeuAc	Sequence	RT	Mass Error	Score
1652.62655	1	1	2	STPTTEA	22.9725	7	74
1682.67515	1	1	2	TSTPTTE	26.51533	11	75
1735.66495	1	1	2	GEPTSTPT	30.6365	7	95
1822.69935	1	1	2	SGEPTSTPT	24.28325	11	104
1880.73315	1	1	2	STPTTEAVE	27.85567	16	76
1965.75615	1	1	2	GEPTSTPTTE	27.59767	12	89
2052.78715	1	1	2	SGEPTSTPTTE	28.913	11	85

Table S-2. RNaseB Results. Theoretical glycopeptide matches to RNaseB digest data. The TDA-based FDR of 5% was calculated at a score threshold of 15. Manual analysis confirmed an actual 5% FDR at a score threshold of 20.

GP Mass	Hex	HexNAc	dHex	NeuAc	Peptide Site	Sequence	Retention Time	Protein site	Score	Family Count
1504.57295	5	2	0	0	2	RN	10.0025	60	46	11
1617.65675	5	2	0	0	2	RNL	11.3930	60	45	11
1461.55475	5	2	0	0	1	NL	10.7785	60	43	11
1779.70975	6	2	0	0	2	RNL	10.6653	60	37	8
1348.47155	5	2	0	0	1	N	9.5180	60	36	11
1718.70635	5	2	0	0	2	RNLT	11.3773	60	36	11
1690.69855	5	2	0	0	1	NLTK	9.6195	60	35	11
1666.62635	6	2	0	0	2	RN	9.2120	60	35	8
2043.81475	6	3	1	0	3	LAN	10.2550	99	35	2
1990.73335	8	2	0	0	2	RN	9.2290	60	35	5
2048.76455	8	2	0	0	1	NLT	9.8805	60	35	5
1617.65675	5	2	0	0	2	RNL	11.6695	60	33	11
1947.71475	8	2	0	0	1	NL	9.7625	60	33	5
1805.73555	5	2	0	0	3	SRNLT	12.3975	60	32	11
1805.73555	5	2	0	0	1	NLTKD	12.3975	60	32	11
1504.57295	5	2	0	0	2	RN	10.2350	60	31	11
2103.81975	8	2	0	0	2	RNL	10.5270	60	31	5
1724.65755	6	2	0	0	1	NLT	9.9080	60	30	8
1852.74935	6	2	0	0	1	NLTK	8.9415	60	30	8
1461.55475	5	2	0	0	1	NL	11.0705	60	27	11
1562.60475	5	2	0	0	1	NLT	10.7573	60	27	11
1510.52255	6	2	0	0	1	N	8.6620	60	27	8
1828.68315	7	2	0	0	2	RN	9.2520	60	27	6
1617.65675	5	2	0	0	2	RNL	11.8640	60	26	11
1718.70635	5	2	0	0	2	RNLT	11.6575	60	26	11
1880.75935	6	2	0	0	2	RNLT	10.4377	60	26	8
1880.75935	6	2	0	0	2	RNLT	10.6780	60	26	8
1880.75935	6	2	0	0	2	RNLT	11.7600	60	26	8
1967.79215	6	2	0	0	3	SRNLT	11.1775	60	25	8
1967.79215	6	2	0	0	1	NLTKD	11.1775	60	25	8
2204.86855	8	2	0	0	2	RNLT	10.4950	60	25	5

1591.61835	5	2	0	0	3	SRN	16.1155	60	24	11
1704.69355	5	2	0	0	3	SRNL	12.4580	60	24	11
1805.73555	5	2	0	0	3	SRNLT	12.1265	60	24	11
1805.73555	5	2	0	0	1	NLTKD	12.1265	60	24	11
1933.83775	5	2	0	0	3	SRNLTK	11.6610	60	24	11
1933.83775	5	2	0	0	4	KSRNLT	11.6610	60	24	11
1623.60875	6	2	0	0	1	NL	9.8997	60	24	8
1666.62635	6	2	0	0	2	RN	9.4150	60	24	8
2042.81535	7	2	0	0	2	RNLT	10.5260	60	23	6
2210.82395	9	2	0	0	1	NLT	9.6725	60	23	4
2366.91975	9	2	0	0	2	RNLT	10.1710	60	23	4
2109.76595	9	2	0	0	1	NL	9.6620	60	22	4
2338.91615	9	2	0	0	1	NLTK	8.7500	60	22	4
2153.79315	4	5	0	0	1	NDSR	9.0780	156	21	2
2042.81535	7	2	0	0	2	RNLT	10.2500	60	21	6
2716.138425	3	3	0	0	12	KCQQWSQQSGQNVT	35.3060	387	20	6
1692.648525	5	2	0	0	2	KNDT	28.8880	564	20	4
1967.79215	6	2	0	0	3	SRNLT	11.4535	60	20	8
1967.79215	6	2	0	0	1	NLTKD	11.4535	60	20	8
2715.134025	5	2	0	0	1	NCSVRQQTQHAVE	35.0740	99	18	3
2014.80915	7	2	0	0	1	NLTK	8.9975	60	16	6
1338.53655	3	2	0	0	2	LNNS	20.8550	252	15	3
1552.66055	3	2	1	0	5	TPLAN	19.9557	99	15	2
1746.686925	4	2	0	0	5	AFVKND	33.4230	564	15	3
2270.92495	6	2	0	0	6	VAFVKNDT	11.3645	564	15	3
2270.92495	6	2	0	0	5	AFVKNDTV	11.3645	564	15	3
1672.57075	7	2	0	0	1	N	8.8000	60	15	6
2292.90655	4	3	1	0	3	LANCSVRQ	11.1290	99	14	7
1635.61775	5	2	0	0	3	LANC	23.6590	99	14	3
2129.85355	7	2	0	0	3	SRNLT	11.4380	60	14	6
2129.85355	7	2	0	0	1	NLTKD	11.4380	60	14	6
1314.51215	4	2	0	0	2	KN	14.4380	564	13	3
1548.58055	5	2	0	0	1	NVT	23.2030	387	12	3
1353.56955	3	2	0	0	4	LCLN	25.6150	252	11	3
1481.62155	3	2	0	0	3	SRNLT	25.8120	60	11	5
1481.62155	3	2	0	0	1	NLTKD	25.8120	60	11	5

1481.62155	3	2	0	0	3	SRNLT	26.9700	60	11	5
1481.62155	3	2	0	0	1	NLTKD	26.9700	60	11	5
1481.62155	3	2	0	0	3	SRNLT	28.2160	60	11	5
1481.62155	3	2	0	0	1	NLTKD	28.2160	60	11	5
1526.61595	3	2	0	0	5	MKSRN	16.3080	60	11	5
1899.809325	3	2	0	0	8	NQMMKSRN	27.0920	60	11	5
1606.61675	4	3	1	0	2	AN	1.0930	99	11	7
1917.785925	4	3	1	0	5	TPLAN	30.9640	99	11	7
1729.67055	3	2	1	0	3	GQNVTCA	27.0980	387	10	3
1930.85735	3	2	1	0	1	NGSYLQLV	15.9340	176	10	3
2621.05915	3	2	1	0	7	SQQSGQNVTCATASTT	25.5350	387	10	3
1412.51955	3	3	0	0	3	GQN	7.0910	387	10	6
1729.67055	3	3	0	0	2	QNVTCA	27.0980	387	10	6
1719.60415	3	3	0	1	1	NNS	7.3280	252	10	5
2623.02375	3	3	0	1	8	WSQQSGQNVTC	26.4930	387	9	2
2331.977625	4	2	1	0	6	ELLCLNNSRA	21.5830	252	9	2
1352.56455	3	2	0	0	2	QNVT	24.8870	387	8	3
1352.56455	3	2	0	0	2	QNVT	25.8010	387	8	3
1352.56455	3	2	0	0	2	QNVT	26.0775	387	8	3
1366.60735	3	2	0	0	1	NLTK	41.4270	60	8	5
1368.49035	3	2	0	0	4	AESNG	8.2560	176	8	2
3493.377525	3	5	0	1	5	QSGQNVTCATASTTDDC	33.6880	387	8	2
1311.51235	3	2	0	0	3	LANC	24.2060	99	7	2
1848.77375	3	3	2	0	4	LCLN	11.9690	252	7	3
2625.03275	3	4	2	0	6	QQSGQNVTCA	26.6340	387	7	4
1881.65715	4	3	0	1	1	NNS	7.1970	252	7	3
1755.59855	5	2	1	0	2	ANCS	23.1910	99	7	2
1788.749925	3	2	0	0	1	NDSRVVHA	30.3790	156	6	4
1788.749925	3	2	0	0	1	NDSRVVHA	31.3480	156	6	4
2183.00195	3	2	0	0	5	LAPLNDSRVVHA	20.0505	156	6	4
1514.61555	3	2	1	0	2	KNDT	28.2400	564	5	4
2157.925425	3	3	2	0	2	LNNSRAP	19.4467	252	4	3
2740.183125	3	4	1	1	2	RNLTKDRC	33.2190	60	4	2
1716.68175	4	2	0	0	4	LCLNNS	17.4735	252	4	1
2184.00215	4	2	1	0	5	LAPLNDSRV	19.8570	156	4	2
1527.60035	3	2	1	0	1	NNSR	17.3445	252	3	1

1527.63455	3	2	1	0	1	NNSR	16.4900	252	3	1
2181.00255	3	4	1	0	7	PLLAPLN	16.8660	156	3	1
2048.91895	4	2	1	0	5	MKSRNLT	39.7060	60	0	1

Table S-3. Protein Mixture Results. All N and O glycopeptide matches for CASK_BOVIN, FETUA_BOVIN, and TRFL_BOVING that scored at 5% FDR, sorted by protein, site, and glycan. Red.=redundant with one or more other possible assignments at that site, NvO=conflicting protein assignments (mostly N v. O), neutral=neutral precursor mass, A.=amino acid number on peptide, RT=retention time in minutes, Site=protein glycosylation site, M.E.=mass error (ppm)

Red, NvO	m/z	neutral	Hex	HexNAc	dHex	NeuAc	Protein	A.	RT	Site	M.E.	Score	Sequence
	772.8454	1543.6751	1	1	0	1	CASK_BOVIN	5	25.16475	142	15	101	TEIPTINT
	880.3832	1758.7507	1	1	0	1	CASK_BOVIN	1	27.07033	142	15	63	TINTIASGEPT
	734.2918	1466.5679	2	2	1	0	CASK_BOVIN	4	28.40025	152	15	57	GEPTST
	531.2195	1060.4233	1	1	0	1	CASK_BOVIN	3	15.744	154	7	63	TSTP
	531.2195	1060.4233	1	1	0	1	CASK_BOVIN	3	14.43	154	7	58	TSTP
	646.2574	1290.4991	1	1	0	1	CASK_BOVIN	2	16.94967	154	15	79	STPTTE
	646.2574	1290.4991	1	1	0	1	CASK_BOVIN	2	16.488	154	15	62	STPTTE
	723.2941	1444.5725	1	1	0	1	CASK_BOVIN	6	19.8	154	15	70	GEPTSTPT
0,1	723.2941	1444.5725	1	1	0	1	CASK_BOVIN	6	16.7725	154	15	69	GEPTSTPT
	827.3211	1652.6265	1	1	0	2	CASK_BOVIN	2	22.9725	154	7	74	STPTTEA
	842.3454	1682.6751	1	1	0	2	CASK_BOVIN	3	26.51533	154	7	75	TSTPTTE
	868.8403	1735.6649	1	1	0	2	CASK_BOVIN	6	30.6365	154	15	95	GEPTSTPT
	868.8403	1735.6649	1	1	0	2	CASK_BOVIN	6	31.3345	154	15	81	GEPTSTPT
	912.3575	1822.6993	1	1	0	2	CASK_BOVIN	7	24.28325	154	3	104	SGEPTSTPT
	912.3575	1822.6993	1	1	0	2	CASK_BOVIN	7	31.3545	154	3	92	SGEPTSTPT
	912.3575	1822.6993	1	1	0	2	CASK_BOVIN	7	32.247	154	3	86	SGEPTSTPT
	941.3744	1880.7331	1	1	0	2	CASK_BOVIN	2	27.85567	154	0	76	STPTTEAVE
	983.8859	1965.7561	1	1	0	2	CASK_BOVIN	6	27.59767	154	19	89	GEPTSTPTTE
	1027.4014	2052.7871	1	1	0	2	CASK_BOVIN	7	28.913	154	7	85	SGEPTSTPTTE
0,1	1006.8773	2011.7389	4	3	1	1	CASK_BOVIN	1	24.0215	154	15	76	TPT
0,1	1089.9696	2177.9235	4	4	0	0	CASK_BOVIN	1	16.2945	154	15	57	TPTTEAV
	844.3572	1686.6987	2	2	1	1	CASK_BOVIN	2	20.9595	157	7	67	TTEAV
0,1	1168.9292	2335.8427	4	3	3	1	CASK_BOVIN	2	23.34525	157	15	66	TTE
0,1	1314.4767	2626.9377	4	3	3	2	CASK_BOVIN	2	26.01975	157	15	62	TTE
	422.1881	842.3605	1	1	0	0	CASK_BOVIN	2	9.786	163	15	55	STVAT
	517.2228	1032.4299	1	1	0	1	CASK_BOVIN	2	15.0435	163	7	58	STVA
	812.3502	1622.6847	1	1	0	2	CASK_BOVIN	5	24.47533	163	19	71	AVESTVA
	862.8593	1723.7029	1	1	0	2	CASK_BOVIN	5	24.9325	163	16	74	AVESTVAT
0,1	1068.9363	2135.8569	4	4	0	0	CASK_BOVIN	5	14.796	163	19	57	AVESTVA
0,1	1176.0044	2349.9931	4	4	0	0	CASK_BOVIN	5	16.618	163	19	69	AVESTVATL

	567.7408	1133.4659	1	1	0	1	CASK_BOVIN	1	13.1615	186	7	57	TSTAV
	696.3080	1390.6003	1	1	0	1	CASK_BOVIN	5	16.603	186	7	67	TVQVTST
	841.8529	1681.6901	1	1	0	2	CASK_BOVIN	5	24.70575	186	16	65	TVQVTST
	841.8529	1681.6901	1	1	0	2	CASK_BOVIN	5	24.899	186	16	62	TVQVTST
0,1	723.2941	1444.5725	3	3	1	0	FETUA_BOVIN	2	16.7725	99	12	40	AN
0,1	1176.0044	2349.9931	4	4	0	0	FETUA_BOVIN	3	16.618	99	6	61	LANCSVRQ
	957.8650	1913.7143	5	3	0	1	FETUA_BOVIN	2	12.9935	99	12	45	AN
	721.2992	1440.5827	1	1	0	1	FETUA_BOVIN	4	21.87933	271	15	74	EAPSAVPD
	721.2992	1440.5827	1	1	0	1	FETUA_BOVIN	4	22.292	271	15	68	EAPSAVPD
	756.8139	1511.6121	1	1	0	1	FETUA_BOVIN	5	24.20833	271	19	62	AEAPSAVPD
1,0	756.8139	1511.6121	1	1	0	1	FETUA_BOVIN	4	24.48	271	19	57	EAPSAVPDA
1,0	756.8139	1511.6121	1	1	0	1	FETUA_BOVIN	5	24.48	271	19	57	AEAPSAVPD
	756.8228	1511.6299	1	1	0	1	FETUA_BOVIN	5	23.73125	271	19	79	AEAPSAVPD
	821.3357	1640.6557	1	1	0	1	FETUA_BOVIN	6	26.87325	271	15	80	EAEAPSAVPD
	939.3729	1876.7301	2	2	0	1	FETUA_BOVIN	5	27.34	271	15	74	AEAPSAVPD
	1057.4580	2112.9003	2	2	0	1	FETUA_BOVIN	5	21.51675	280	7	71	AAGPTPSAAGPPV
	956.8727	1911.7297	2	2	0	2	FETUA_BOVIN	5	27.959	280	15	84	AAGPTPS
1,0	956.8727	1911.7297	2	2	0	2	FETUA_BOVIN	1	28.1805	280	5	83	TPSAAGP
1,0	956.8727	1911.7297	2	2	0	2	FETUA_BOVIN	4	28.1805	280	15	83	AGPTPSA
1,0	956.8727	1911.7297	2	2	0	2	FETUA_BOVIN	5	28.1805	280	15	83	AAGPTPS
0,1	992.3895	1982.7633	2	2	0	2	FETUA_BOVIN	5	28.04367	280	15	80	AAGPTPSA
	1153.4726	2304.9295	2	2	0	2	FETUA_BOVIN	5	28.59225	280	8	59	AAGPTPSAAGPP
	1203.0090	2404.0023	2	2	0	2	FETUA_BOVIN	5	33.60975	280	8	76	AAGPTPSAAGPPV
1,0	1203.0090	2404.0023	2	2	0	2	FETUA_BOVIN	4	32.982	280	13	66	AGPTPSAAGPPVA
1,0	1203.0090	2404.0023	2	2	0	2	FETUA_BOVIN	5	32.982	280	8	66	AAGPTPSAAGPPV
	664.2791	1326.5425	1	1	0	1	FETUA_BOVIN	7	15.08567	282	19	60	AAGPTPSA
	774.3096	1546.6035	1	1	0	2	FETUA_BOVIN	7	20.56033	282	5	73	AAGPTPS
	774.3096	1546.6035	1	1	0	2	FETUA_BOVIN	7	20.765	282	5	66	AAGPTPS
	845.3450	1688.6743	1	1	0	2	FETUA_BOVIN	7	21.58667	282	15	67	AAGPTPSAA
	845.3450	1688.6743	1	1	0	2	FETUA_BOVIN	7	21.825	282	15	67	AAGPTPSAA
	765.3050	1528.5943	2	1	1	2	FETUA_BOVIN	2	17.29575	282	5	55	PSA
1,0	888.9302	1775.8447	1	1	0	1	FETUA_BOVIN	3	31.871	341	16	62	QPSIPGGPVRL
1,0	888.9302	1775.8447	1	1	0	1	FETUA_BOVIN	6	31.871	341	3	62	IVGQPSIPGGPV
	888.9302	1775.8447	1	1	0	1	FETUA_BOVIN	6	31.56033	341	3	61	IVGQPSIPGGPV
	937.8315	1873.6473	7	2	0	0	TRFL_BOVIN	1	11.938	252	11	74	NNS
	945.3354	1888.6551	7	2	0	0	TRFL_BOVIN	3	12.603	252	19	58	CLN

	975.3637	1948.7117	8	2	0	0	TRFL_BOVIN	1	11.592	252	19	64	NN
	1018.8616	2035.7075	8	2	0	0	TRFL_BOVIN	1	11.69267	252	19	79	NNS
	1026.3601	2050.7045	8	2	0	0	TRFL_BOVIN	3	12.3875	252	19	76	CLN
	1026.3601	2050.7045	8	2	0	0	TRFL_BOVIN	3	12.5865	252	19	68	CLN
	1075.4034	2148.7911	8	2	0	0	TRFL_BOVIN	2	12.822	252	6	51	LNNS
	1096.9115	2191.8073	8	2	0	0	TRFL_BOVIN	1	11.606	252	6	42	NNSR
	1056.3737	2110.7317	9	2	0	0	TRFL_BOVIN	1	11.682	252	18	81	NN
	1099.8884	2197.7611	9	2	0	0	TRFL_BOVIN	1	11.773	252	18	91	NNS
	1107.3908	2212.7659	9	2	0	0	TRFL_BOVIN	3	12.241	252	18	104	CLN
	1156.4272	2310.8387	9	2	0	0	TRFL_BOVIN	2	12.857	252	8	102	LNNS
	1177.9353	2353.8549	9	2	0	0	TRFL_BOVIN	1	11.64733	252	6	83	NNSR
	775.3002	1548.5847	5	2	0	0	TRFL_BOVIN	1	13.482	387	19	70	NVT
	811.3010	1620.5863	5	2	0	0	TRFL_BOVIN	4	13.3295	387	12	64	SGQN
	860.8346	1719.6535	5	2	0	0	TRFL_BOVIN	4	14.952	387	18	51	SGQNV
	892.3286	1782.6415	6	2	0	0	TRFL_BOVIN	4	13.613	387	12	44	SGQN
	941.8587	1881.7017	6	2	0	0	TRFL_BOVIN	4	13.94675	387	12	72	SGQNV
	937.3508	1872.6859	7	2	0	0	TRFL_BOVIN	1	12.519	387	18	73	NVT
	973.3502	1944.6847	7	2	0	0	TRFL_BOVIN	4	12.527	387	12	58	SGQN
	1022.8848	2043.7539	7	2	0	0	TRFL_BOVIN	4	13.6585	387	8	68	SGQNV
	1022.8848	2043.7539	7	2	0	0	TRFL_BOVIN	4	14.002	387	8	49	SGQNV
	1073.4095	2144.8033	7	2	0	0	TRFL_BOVIN	4	14.068	387	6	45	SGQNVT
	967.8510	1933.6863	8	2	0	0	TRFL_BOVIN	1	12.247	387	18	61	NV
	967.8510	1933.6863	8	2	0	0	TRFL_BOVIN	1	12.675	387	18	49	NV
	1018.3784	2034.7411	8	2	0	0	TRFL_BOVIN	1	12.61733	387	8	79	NVT
	1054.3737	2106.7317	8	2	0	0	TRFL_BOVIN	4	12.313	387	18	46	SGQN
	1054.3737	2106.7317	8	2	0	0	TRFL_BOVIN	4	12.626	387	18	46	SGQN
	1103.9115	2205.8073	8	2	0	0	TRFL_BOVIN	4	13.784	387	12	64	SGQNV
	1048.8828	2095.7499	9	2	0	0	TRFL_BOVIN	1	12.255	387	8	66	NV
	1135.3984	2268.7811	9	2	0	0	TRFL_BOVIN	4	12.354	387	18	60	SGQN
	1184.9384	2367.8611	9	2	0	0	TRFL_BOVIN	4	13.545	387	4	60	SGQNV
0,1	1006.8773	2011.7389	4	3	1	0	TRFL_BOVIN	1	24.0215	495	6	55	NQTGSC
	911.8471	1821.6785	4	4	0	0	TRFL_BOVIN	1	11.9045	495	18	48	NQT
	988.4189	1974.8221	4	4	0	0	TRFL_BOVIN	5	15.61267	495	18	60	GLIVN
	1023.9377	2045.8597	4	4	0	0	TRFL_BOVIN	4	15.613	495	8	64	LIVNQ
	1052.4550	2102.8943	4	4	0	0	TRFL_BOVIN	5	16.371	495	18	50	GLIVNQ
	1074.4626	2146.9095	4	4	0	0	TRFL_BOVIN	4	15.89467	495	10	46	LIVNQT

	1102.9757	2203.9357	4	4	0	0	TRFL_BOVIN	5	16.421	495	8	67	GLIVNQT
0,1	1068.9363	2135.8569	4	5	0	0	TRFL_BOVIN	3	14.796	495	2	67	IVNQ
0,1	1089.9696	2177.9235	4	5	0	0	TRFL_BOVIN	5	16.2945	495	5	74	GLIVN
	1119.4611	2236.9065	4	5	0	0	TRFL_BOVIN	3	14.87767	495	1	53	IVNQT
	1125.4766	2248.9375	4	5	0	0	TRFL_BOVIN	4	16.4515	495	10	41	LIVNQ
	1155.4881	2308.9605	4	5	0	0	TRFL_BOVIN	6	16.17833	495	14	40	MGLIVN
	1155.4881	2308.9605	4	5	0	0	TRFL_BOVIN	6	16.678	495	14	40	MGLIVN
1,0	1204.5116	2407.0075	4	5	0	0	TRFL_BOVIN	4	17.042	495	14	47	LIVNQTG
1,0	1204.5116	2407.0075	4	5	0	0	TRFL_BOVIN	5	17.042	495	6	47	GLIVNQT
	845.3478	1688.6799	5	2	0	0	TRFL_BOVIN	3	15.8145	495	18	72	IVNQ
	866.3683	1730.7209	5	2	0	0	TRFL_BOVIN	5	17.455	495	19	80	GLIVN
	901.8887	1801.7617	5	2	0	0	TRFL_BOVIN	4	17.673	495	19	48	LIVNQ
	930.4006	1858.7855	5	2	0	0	TRFL_BOVIN	5	18.27475	495	19	74	GLIVNQ
	952.4100	1902.8043	5	2	0	0	TRFL_BOVIN	4	17.79375	495	18	55	LIVNQT
1,0	980.9229	1959.8301	5	2	0	0	TRFL_BOVIN	4	18.415	495	18	46	LIVNQTG
1,0	980.9229	1959.8301	5	2	0	0	TRFL_BOVIN	5	18.415	495	10	46	GLIVNQT
	946.8855	1891.7553	5	3	0	0	TRFL_BOVIN	3	14.33	495	6	42	IVNQ
	1033.4191	2064.8225	5	3	0	0	TRFL_BOVIN	6	18.218	495	0	45	MGLIVN
	1053.9506	2105.8855	5	3	0	0	TRFL_BOVIN	4	16.029	495	17	45	LIVNQT
	1082.4617	2162.9077	5	3	0	0	TRFL_BOVIN	5	16.5685	495	17	66	GLIVNQT
	805.8037	1609.5917	6	2	0	0	TRFL_BOVIN	2	12.62	495	19	70	VN
	870.8206	1739.6255	6	2	0	0	TRFL_BOVIN	1	13.3035	495	12	72	NQT
	920.3583	1838.7009	6	2	0	0	TRFL_BOVIN	2	14.301	495	18	60	VNQT
	926.3737	1850.7317	6	2	0	0	TRFL_BOVIN	3	16.21725	495	5	87	IVNQ
	947.3943	1892.7729	6	2	0	0	TRFL_BOVIN	5	17.99	495	12	48	GLIVN
	976.9021	1951.7885	6	2	0	0	TRFL_BOVIN	3	16.35833	495	18	54	IVNQT
	982.9104	1963.8051	6	2	0	0	TRFL_BOVIN	4	18.1035	495	18	80	LIVNQ
	992.3805	1982.7453	6	2	0	0	TRFL_BOVIN	2	13.937	495	6	58	VNQTGS
	992.3805	1982.7453	6	2	0	0	TRFL_BOVIN	2	15.347	495	6	51	VNQTGS
0,1	992.3895	1982.7633	6	2	0	0	TRFL_BOVIN	2	28.04367	495	6	47	VNQTGS
	1011.4250	2020.8343	6	2	0	0	TRFL_BOVIN	5	18.73325	495	18	74	GLIVNQ
	1033.4377	2064.8597	6	2	0	0	TRFL_BOVIN	4	18.26875	495	6	66	LIVNQT
1,0	1061.9412	2121.8667	6	2	0	0	TRFL_BOVIN	4	18.823	495	17	46	LIVNQTG
1,0	1061.9412	2121.8667	6	2	0	0	TRFL_BOVIN	5	18.823	495	4	46	GLIVNQT
1,0	1061.9478	2121.8799	6	2	0	0	TRFL_BOVIN	4	18.9165	495	8	70	LIVNQTG
1,0	1061.9478	2121.8799	6	2	0	0	TRFL_BOVIN	5	18.9165	495	6	70	GLIVNQT

	1061.9478	2121.8799	6	2	0	0	TRFL_BOVIN	5	16.774	495	6	57	GLIVNQT
	945.3325	1888.6493	7	2	0	0	TRFL_BOVIN	1	12.5355	564	19	65	NDT
	1087.9048	2173.7939	7	2	0	0	TRFL_BOVIN	1	23.535	564	6	54	NDTVW
	1087.9048	2173.7939	7	2	0	0	TRFL_BOVIN	1	23.87775	564	6	46	NDTVW
0,1	1168.9292	2335.8427	8	2	0	0	TRFL_BOVIN	1	23.34525	564	6	78	NDTVW
	1233.4439	2464.8721	8	2	0	0	TRFL_BOVIN	1	26.5625	564	0	40	NDTVWE
	1156.9196	2311.8235	9	2	0	0	TRFL_BOVIN	1	12.894	564	4	54	NDTV
	1249.9580	2497.9003	9	2	0	0	TRFL_BOVIN	1	23.23275	564	6	83	NDTVW
0,1	1314.4767	2626.9377	9	2	0	0	TRFL_BOVIN	1	26.01975	564	6	56	NDTVWE

Table S-4. Protein Mixture Summary. Comprehensive SSG for bovine lactoferrin, fetuin, and kappa casein with TDA-compliant results above 5% FDR. Three sites did not contain matches above the 5% FDR threshold.

Protein	Site	Type	Glycans	Glycopeptides
Bovine Lactoferrin	252	N	3	12
	300	N		
	387	N	5	17
	495	N	6	37
	564	N	3	7
Bovine Fetuin	99	N	3	3
	156	N		
	176	N		
	271	O	2	6
	280	O	2	7
	282	O	3	5
	341	O	1	2
Kappa Casein	142	O	1	2
	152	O	1	1
	154	O	4	12
	157	O	3	3
	163	O	4	6
	186	O	2	3

Table S-5. Increased Top Score Distinction. Top five target matches for precursor mass 2310.8388 Da. The top scoring match was substantially more separated from all other matches after applying the self-consistency algorithm.

Score							
w/o self-consistency	w/self-consistency and TDA-compliance	Peptide	Hex	HexNAc	dHex	NeuAc	Protein
56	102	LNNS	9	2	0	0	BLF
44	59	NQTGSC	8	2	0	0	BLF
35	44	CLNN	8	2	1	0	BLF
22	28	SNGSY	6	3	0	0	BF
18	22	NDTVWE	4	3	0	1	BLF