

(51) International Patent Classification:
A61K 38/46 (2006.01)(21) International Application Number:
PCT/US2013/032171(22) International Filing Date:
15 March 2013 (15.03.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/618,656 30 March 2012 (30.03.2012) US
61/717,190 23 October 2012 (23.10.2012) US

(72) Inventors; and

(71) Applicants : SHIRE HUMAN GENETIC THER-
APIES, INC. [US/US]; 300 Shire Way, Lexington, MA
02421 (US). LIN, Melanie [US/US]; 300 Shire Way, Lex-
ington, MA 02421 (US). SALINAS, Paul [US/US]; 300
Shire Way, Lexington, MA 02421 (US). WU, Shiaw-lin
[US/US]; 300 Shire Way, Cambridge, MA 02142 (US).(74) Agent: LU, Shihua; Lando & Anastasi LLP, Riverfront
Office Park, One Main Street, Suite 1100, Cambridge, MA
02142 (US).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU,
RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ,
TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA,
ZM, ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: METHODS OF ANALYZING AND PREPARING PROTEIN COMPOSITIONS

Figure 1



(57) Abstract: Methods for analyzing protein samples of recombinant human arylsulfatase A(rhASA), having one or more disulfide linkages are described. In some instances, the samples are analyzed using enzyme digestion, e.g., multi-enzyme digestion and liquid chromatography-mass spectrometry (LC-MS). The proteins described herein can be glycosylated (e.g., fully or partially glycosylated) or non-glycosylated. The methods described herein can include one or more of enzymatic digestion, chromatography, and mass spectrometry, e.g., multi-enzyme digestion and/or liquid chromatography-mass spectrometry (LC-MS). In some embodiments, the mass spectrometry utilizes collision induced dissociation (CID) and/or electron transfer dissociation (ETD). For example, the mass spectrometry can utilize CID, ETD, and CID of the isolated charge-reduced ions (MS3), e.g., after ETD.

METHODS OF ANALYZING AND PREPARING PROTEIN COMPOSITIONS

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Application Serial No. 61/618,656, filed on March 30, 2012, and U.S. Application Serial No. 61/717,190, filed on October 23, 2012. The disclosures of the prior applications are considered part of (and are incorporated by reference in) the disclosure of this application.

BACKGROUND

Arylsulfatase A (ASA) is a lysosomal enzyme, which catalyzes the hydrolysis of cerebroside sulfate (galactosylceramide-3-O-sulfate or sulfatide) to cerebroside and sulfate. Deficiency of this enzyme cumulates cerebroside sulfate and leads to destruction of myelin in the central and peripheral nervous systems resulting in a progressive demyelination disease known as metachromatic leukodystrophy (MLD). MLD patients can be treated by enzyme replacement therapy (ERT) using recombinant human arylsulfatase A (rhASA).

SUMMARY

The disclosure relates, at least in part, to methods of analyzing and/or preparing samples (*e.g.*, protein samples) that contain one or more disulfide linkages. For example, the disclosure provides methods that can be used to analyze, control, or monitor the production of proteins, *e.g.*, ASA, *e.g.*, rhASA. The proteins described herein can be glycosylated (*e.g.*, fully or partially glycosylated) or non-glycosylated. The methods described herein can include one or more of enzymatic digestion, chromatography, and mass spectrometry, *e.g.*, multi-enzyme digestion and/or liquid chromatography-mass spectrometry (LC-MS). In some embodiments, the mass spectrometry utilizes collision induced dissociation (CID) and/or electron transfer dissociation (ETD). For example, the mass spectrometry can utilize CID, ETD, and CID of the isolated charge-reduced ions (MS3), *e.g.*, after ETD. Further, compounds and compositions that can be detected or prepared by the methods described herein are also disclosed.

The methods described herein can be used with proteins, *e.g.*, glycoproteins, *e.g.*, ASA, *e.g.*, rhASA. The analysis of proteins, *e.g.*, by enzyme digestion (*e.g.*, multi-enzyme digestion) and/or LC-MS (*e.g.*, LC-MS utilizing CID and/or ETD), can be used to evaluate starting materials, processes, intermediates, and final products in the production of proteins. For example, the methods described herein are useful for analyzing, evaluating, or processing a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation, *e.g.*, to determine whether to accept or reject a batch of the protein preparation, or to guide or control of a step in the production of the protein. Therefore, the methods disclosed herein are useful, *e.g.*, from a process standpoint, *e.g.*, to monitor or ensure batch-to-batch consistency or quality, or to evaluate a sample with regard to a reference, *e.g.*, a preselected value. For example, the presence, distribution, or amount of one or more subject entities, *e.g.*, a structure, species, or fraction described herein, can be used in these evaluations. In some embodiments, the methods disclosed herein can be used where the presence, distribution, or amount, of one or more of the subject entities in the sample may possess or impinge on the biological activity. In some embodiments, the methods are useful from a structure-activity prospective, to evaluate or ensure biological equivalence.

In some cases, the methods described herein are directly applicable. In some other cases, one of ordinary skill in the art will appreciate that modifications may be needed and can institute those as guided by the art and this disclosure.

In one aspect, the disclosure provides a method of evaluating or processing a sample, *e.g.*, a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation. The method includes providing an evaluation of a parameter related to a subject entity, *e.g.*, a structure, species, or fraction described herein, *e.g.*, a subject entity depicted in **Table 1**, **Table 2**, or **FIG. 1**, thereby evaluating or processing the sample. Such parameters can include, or are a function of, the presence, distribution, or amount of a subject entity, *e.g.*, a structure, species, or fraction disclosed herein. Optionally, the method includes providing a determination of whether a test value (*e.g.*, a value correlated with absence or presence) determined for the parameter meets a preselected criteria, *e.g.*, is present as described herein, present in the amount described herein, or distributed as described herein.

In some embodiments, the method includes providing a comparison of the test value determined for a parameter with a reference value, or values, to thereby evaluate or process the sample. In some embodiments, the comparison includes determining if the test value has a preselected relationship with the reference value, *e.g.*, determining if it meets the reference value. The value need not be a numerical value, but can be merely an indication of whether the subject entity is present.

In some embodiments, the method includes determining if a test value is equal to or greater than a reference value, if it is less than or equal to a reference value, or if it falls within a range (either inclusive or exclusive of one or both endpoints). By way of example, the presence, distribution, or amount of a structure, species, or fraction, as depicted in **Table 1**, **Table 2**, or **FIG. 1**, can be determined and, optionally shown to fall within a preselected range.

In some embodiments, the test value, or an indication of whether the preselected relationship is met, can be memorialized, *e.g.*, in a computer readable record.

In some embodiments, a decision or step is taken, *e.g.*, the sample is classified, selected, accepted or discarded, released or withheld, processed into a drug product, shipped, moved to a different location, formulated, labeled, packaged, released into commerce, or sold or offered for sale, depending on whether the preselected relationship is met. For example, based on the result of the determination, or upon comparison to a reference standard, the batch from which the sample is taken can be processed, *e.g.*, as described herein.

The subject entities described herein include, but are not limited to, a structure, a species, or a fraction in a sample, *e.g.*, a protein preparation.

A structure can be, *e.g.*, a particular residue, or group of residues, *e.g.*, an unpaired cysteine, a disulfide linkage (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot), or a combination thereof, existing in a protein, polypeptide, or peptide. In some embodiments, the structure comprises an unpaired cysteine, a disulfide linkage (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot), or a combination thereof, as depicted in **Table 1**, **Table 2**, or **FIG. 1**.

In some embodiments, the structure comprises one or more of the unpaired cysteine residues at C20, C51, and C276 in ASA. In some embodiments, C51 is

converted to formylglycine. In some embodiments, C51 is not converted to formylglycine. In some embodiments, the structure comprises a disulfide linkage between C282 and C396 in ASA. In some embodiments, the structure comprises one or two of the disulfide linkages between C138 and C154, and between C143 and C150, in ASA. In some embodiments, the structure comprises one, two, or three of the disulfide linkages between C470 and C482, between C471 and C484, and between C475 and C481, in ASA. In some embodiments, the structure comprises one or more, *e.g.*, two, three, four, five, or all, of the disulfide linkages between C282 and C396, between C138 and C154, between C143 and C150, between C470 and C482, between C471 and C484, and between C475 and C481, in ASA. In some embodiments, the structure is detected or determined by a method described herein, *e.g.*, enzyme digestion (*e.g.*, multi-enzyme digestion) and/or LC-MS (*e.g.*, LC-MS utilizing CID and/or ETD, *e.g.*, CID, ETD, and CID-MS3).

A species can be, *e.g.*, a peptide with one or more, *e.g.*, two, three, four, five, six, or more unpaired cysteine residues; a peptide with one or more, *e.g.*, two, three, four, five, six, or more, disulfide linkages (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot); a peptide with one or more, *e.g.*, two, three, four, five, six, or more, unpaired cysteine and one or more, *e.g.*, two, three, four, five, six, or more, disulfide linkages (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot); disulfide-linked peptides with one or more, *e.g.*, two, three, four, five, six, or more, unpaired cysteine residues; disulfide-linked peptides with one or more, *e.g.*, two, three, four, five, six, or more, disulfide linkages (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot); disulfide-linked peptides with one or more, *e.g.*, two, three, four, five, six, or more, unpaired cysteine residues and one or more, *e.g.*, two, three, four, five, six, or more, disulfide linkages (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot). In some embodiments, the disulfide linkage is located within a peptide. In some embodiments, the disulfide linkage is located between two peptides. In some embodiments, the species, *e.g.*, the peptide or disulfide-linked peptides, can contain one or more, *e.g.*, two, three, four or more, formylglycine residues. In some embodiments, the species is a peptide with an unpaired cysteine residue, disulfide-linked peptides with one disulfide linkage, disulfide-linked peptides with two disulfide linkages, or a peptide

with three disulfide linkages, as depicted in **Table 1**. In some embodiments, the species is GCYGHPSSTTPNL (19-31). In some embodiments, the species is YVPVSLC(fgly)TPSRAAL (45-58). In some embodiments, the species is YVPVSLCTPSRAAL (45-58). In some embodiments, the species is RMSRGGCSGL (270-279). In some embodiments, the species is LRCGKGTTYEGGVRE (282-294) and FTQGSASHSDTTADPACHASSSL (381-402) linked with a disulfide linkage between C282 and C396. In some embodiments, the species is CGK (282-284) and AHFFTQGSASHSDTTADPACHASSSLTAHEPPLLYDLSK (378-415) linked with a disulfide linkage between C282 and C396. In some embodiments, the species is DQGPCQ (134-139) and DLTCFPPATPCDGGC (140-154) linked with the disulfide linkages between C138 and C154 and between C143 and C150. In some embodiments, the species is PALQICCHPGCTPRPACCHCPDP with the disulfide linkages between C470 and C482, between C471 and C484, and between C475 and C481. In some embodiments, the species is a peptide or disulfide-linked peptide having the theoretical or observed mass disclosed in **Table 1**. In some embodiments, the species is a peptide or disulfide-linked peptide prepared by a method described herein, *e.g.*, multi-enzyme digestion. In some embodiments, the species is a peptide or disulfide-linked peptide is detected by a method described herein, *e.g.*, LC-MS. In some embodiments, the species is a fragment ion having the theoretical or observed mass disclosed in **Table 2**. In some embodiments, the species, *e.g.*, the peptide or disulfide-linked peptides, comprises a glycosylation site. In some embodiments, the structure is detected or determined by a method described herein, *e.g.*, enzyme digestion (*e.g.*, multi-enzyme digestion) and/or LC-MS (*e.g.*, LC-MS utilizing CID and/or ETD, *e.g.*, CID, ETD, and CID-MS3).

A fraction can be, *e.g.*, a part, portion, or subset of a sample, *e.g.*, protein preparation, *e.g.*, ASA preparation, *e.g.*, rhASA preparation, having one or more of the structures described herein, or comprising one or more of the species described herein.

In some embodiments, the method includes determining the presence, distribution, or amount of at least two or more (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15) of the unpaired cysteine residues and disulfide linkages (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot), *e.g.*, the unpaired cysteine residues and disulfide linkages (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot)

depicted in **Table 1**, in a sample, *e.g.*, a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation. In some embodiments, the sample is evaluated for the presence, distribution, or amount of each of the unpaired cysteine residues and disulfide linkages (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot) depicted in **Table 1**. In some embodiments, one or more structures or species from a subset of the structures or species depicted in **Table 1** are evaluated. In some embodiments, the sample comprises one or more of the compounds depicted in **Table 1**.

In some embodiments, the method includes determining the presence, amount, or distribution of one or more unpaired cysteine residues in a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation, *e.g.*, the presence, amount, or distribution of one or more of unpaired cysteine residues at positions 20, 51, and 276 of ASA, *e.g.*, by peptide digestion, *e.g.*, pepsin digestion, *e.g.*, at a pH less than 8, *e.g.*, at a pH less than 6.8, *e.g.*, at pH 2.

In some embodiments, the method includes determining the presence, amount, or distribution of one or more single disulfides in a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation, *e.g.*, the presence, amount, or distribution of a single disulfide, *e.g.*, between positions 282 and 396 of ASA, *e.g.*, by peptide digestion, *e.g.*, pepsin digestion (*e.g.*, at a pH less than 8, *e.g.*, at a pH less than 6.8, *e.g.*, at pH 2), *e.g.*, trypsin digestion (*e.g.*, at a pH equal to or less than 8, *e.g.*, at pH 6.8), *e.g.*, Lys-C digestion (*e.g.*, at a pH equal to or less than 8, *e.g.*, at pH 6.8), or a combination thereof. In some embodiments, the peptide digestion includes pepsin, *e.g.*, at a pH less than 8, *e.g.*, at a pH less than 6.8, *e.g.*, at pH 2. In some embodiments, the peptide digestion includes trypsin and Lys-C, *e.g.*, at a pH equal to or less than 8, *e.g.*, at pH 6.8.

In some embodiments, the method includes determining the presence, amount, or distribution of one or more nested disulfides in a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation, *e.g.*, the presence, amount, or distribution of one or more of nested disulfides, *e.g.*, including the disulfide bonds between positions 138 and 154 and between positions 143 and 150 of ASA, *e.g.*, by peptide digestion, *e.g.*, PNGase F digestion (*e.g.*, at a pH equal to or less than 8, *e.g.*,

at pH 6.8), *e.g.*, Asp-N digestion (*e.g.*, at a pH equal to or less than 8, *e.g.*, at pH 6.8), *e.g.*, Lys-C digestion (*e.g.*, at a pH equal to or less than 8, *e.g.*, at pH 6.8), *e.g.*, trypsin digestion (*e.g.*, at a pH equal to or less than 8, *e.g.*, at pH 6.8), or a combination of two, three, or four of PNGase F, Asp-N, Lys-C, and trypsin. In some embodiments, the peptide digestion includes PNGase F, Asp-N, Lys-C, and trypsin, *e.g.*, at a pH equal to or less than 8, *e.g.*, at pH 6.8. In some embodiments, the peptide digestion further includes pepsin digestion, *e.g.*, at a pH less than 8, *e.g.*, at a pH less than 6.8, *e.g.*, at pH 2.

In some embodiments, the method includes determining the presence, amount, or distribution of one or more cystine knots in a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation, *e.g.*, the presence, amount, or distribution of one or more of cystine knots, *e.g.*, including the disulfide bonds between positions 470 and 482, between positions 471 and 484, and between positions 475 and 481 of ASA, *e.g.*, by peptide digestion, *e.g.*, pepsin digestion (*e.g.*, at a pH less than 8, *e.g.*, at a pH less than 6.8, *e.g.*, at pH 2).

The evaluation of the presence, distribution, or amount of a subject entity, *e.g.*, a structure, species, or fraction, can show if the subject entity or a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation, meets a reference standard.

In some embodiments, the methods disclosed herein can be used to determine if a test batch of a protein, *e.g.*, a glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, can be expected to have one or more of the properties of the protein. Such properties can include a property listed on the product insert of a protein, *e.g.*, a glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, a property appearing in a compendium, *e.g.*, the US Pharmacopeia, or a property required by a regulatory agency, *e.g.*, the FDA, for commercial use. A determination made by a method disclosed herein can be a direct or indirect measure of such a property. For example, a direct measurement can be where a desired property is a preselected level of a subject entity, *e.g.*, a structure, species, or fraction, measured. In an indirect measure, the measured subject entity is correlated with, or is a proxy for a desired property, *e.g.*, a property described herein.

Exemplary properties for rhASA can include, but not limited to, a preselected level of specific arylsulfatase activity, *e.g.*, between about 10 and about 500 U/mg, *e.g.*,

between about 50 and about 140 U/mg, between about 50 and about 100 U/mg, or between about 100 and about 140 U/mg; a preselected value for disulfide linkage formation, *e.g.*, at least about 50%, *e.g.*, at least about 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%, of the rhASA molecules in a preparation have one or more of the unpaired cysteine residues, disulfide linkages, and/or cystine knots, as described herein; a preselected value for average molecular weight; a preselected value for glycosylation; and a set of preselected values for molecular weight distribution, *e.g.*, based on glycosylation.

In some embodiments, the sample is processed or evaluated by enzyme digestion, *e.g.*, multi-enzyme digestion. The method includes providing a sample, *e.g.*, a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation; and subjecting the sample to processing or evaluation, *e.g.*, by enzyme digestion, *e.g.*, multi-enzyme digestion. For example, the protein can be digested to peptides, disulfide-linked peptides, or a combination thereof, *e.g.*, by one or more endoproteinases, *e.g.*, in the absence or presence of one or more endoglycosidases. In some embodiments, the peptide comprises at least one unpaired cysteine residue. In some embodiments, the disulfide-linked peptides comprise one or more, *e.g.*, two, three, four, five, or six, disulfide linkages. In some embodiments, the disulfide-linked peptides comprise at least one unpaired cysteine residue. In some embodiments, the peptide comprises a formylglycine residue. In some embodiments, the disulfide-linked peptides comprise a formylglycine residue. In some embodiments, the size of the peptide or disulfide-linked peptides is between about 0.2 kDa and about 20 kDa, *e.g.*, between about 0.5 kDa and about 10 kDa, between about 0.5 kDa and about 2 kDa, between about 0.5 kDa and about 1 kDa, between about 1 kDa and about 2 kDa, and between about 1 kDa and about 5 kDa. In some embodiments, the size of the peptide or disulfide-linked peptide is adjusted, *e.g.*, to ensure sufficiently high-charge states that results in a low mass-to-charge ratio, *e.g.*, $m/z < 900$, *e.g.*, for effective ETD fragmentation.

In some embodiments, the protein is digested by a single enzyme. In some embodiments, the protein is digested by a mixture of two or more, *e.g.*, three, four, five, or six different enzymes. In some embodiments, the protein is digested sequentially by two or more, *e.g.*, three, four, five, or six, different enzymes.

In some embodiments, the enzyme is selected on the basis of the size or sizes of the peptides or disulfide-linked peptides that can be generated by the digestion. In some embodiments, the enzyme is selected on the basis of the structures, *e.g.*, an unpaired cysteine residue, a single disulfide linkage, multiple disulfide linkages, or a cystine knot, of the peptides or disulfide-linked peptides that can be generated by the digestion. In some embodiments, the enzyme is selected to reduce the complexity of mass spectra, *e.g.*, when the protein is glycosylated, *e.g.*, having N-linked glycosylation. In some embodiments, the enzyme is an endoproteinase. In some embodiments, the enzyme is an endoglycosidase. In some embodiments, the enzyme is selected from the group consisting of Lys-C, trypsin, Asp-N, pepsin, and PNGase F. In some embodiments, the enzyme is pepsin. In some embodiments, the mixture of enzymes comprises two or more, *e.g.*, three, four, or all, of Lys-C, trypsin, Asp-N, pepsin, and PNGase F. In some embodiments, the mixture of enzymes comprises Lys-C and trypsin. In some embodiments, the mixture of enzymes comprises Lys-C, trypsin, Asp-N, and PNGase F.

In some embodiments, the digestion pH is optimized to minimize the disulfide scrambling. For example, the disulfide scrambling is reduced by at least about 50%, 60%, 70%, 80%, 90%, 95%, or 99%, *e.g.*, at pH 2, compared to the disulfide scrambling at a digestion pH that is not optimized, *e.g.*, at pH 6.8 or 8. The disulfide scrambling can be measured by a method described herein, *e.g.*, LC-MS. In some embodiments, the sample is digested at a pH between about 2 and about 10, *e.g.*, between about 2 and about 8, between about 2 and about 6, between about 2 and about 4, between about 4 and about 10, between about 6 and about 10, or between about 8 and about 10. In some embodiments, the sample is digested at a pH about 2. In some embodiments, the sample is digested at a pH about 6.8. In some embodiments, the sample is digested at a pH about 8.

In some embodiments, the sample is processed or analyzed by LC-MS, *e.g.*, performed on a digested protein preparation, *e.g.*, a digested rhASA preparation. The method includes providing a sample, *e.g.*, a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation; and subjecting the sample to processing or analysis, *e.g.*, by LC-MS, and optionally, evaluating the

presence, distribution, or amount of a selected subject entity, *e.g.*, a structure, species, or fraction described herein.

In some embodiments, the LC-MS comprises a chromatography column having about 160 Å, 180 Å, 200 Å, 220 Å, or 240 Å pore size, and/or about 3 µm, 5 µm, 7 µm, or 9 µm particle size. In some embodiments, the LC-MS comprises a survey spectrum from m/z 100 to m/z 5000, *e.g.*, from m/z 300 to m/z 2000. In some embodiments, the LC-MS comprises tandem mass spectrometry (MS/MS). In some embodiments, the LC-MS comprises CID, *e.g.*, as CID-MS2. In some embodiments, the LC-MS comprises ETD, *e.g.*, as ETD-MS2. In some embodiments, the LC-MS comprises both CID and ETD. In some embodiments, the CID and ETD are performed on the same precursor ion. In some embodiments, a second CID is performed, *e.g.*, after ETD, *e.g.*, as CID-MS3. In some embodiments, one or more steps of LC-MS are repeated, *e.g.*, to gain additional disulfide linkage information.

In some embodiments, the method includes determining if a contaminant fraction is present. Examples of contaminants include, *e.g.*, contaminants associated with the manufacturing process, *e.g.*, misfolded proteins, *e.g.*, proteins having one or more undesired unpaired cysteines and disulfide linkages.

In some embodiments, the method includes identifying the distribution of one or more of the fractions, *e.g.*, proteins (*e.g.*, glycoproteins, *e.g.*, ASA, *e.g.*, rhASA) with a desired structure, *e.g.*, a desired unpaired cysteine residue, disulfide linkage, and/or cystine knot, relative to a fraction of proteins with an undesired structure, *e.g.*, an undesired unpaired cysteine residue and/or disulfide linkage (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot), in a preparation.

In some embodiments, the method includes determining if one or more of the fractions, *e.g.*, proteins (*e.g.*, glycoproteins, *e.g.*, ASA, *e.g.*, rhASA) with a desired structure, *e.g.*, a desired unpaired cysteine residue, disulfide linkage, and/or cystine knot, are present at a higher intensity than one or more fractions of proteins with an undesired structure, *e.g.*, an undesired unpaired cysteine residue, disulfide linkage, and/or cystine knot, in a preparation.

In some embodiments, the method includes evaluating a subject entity, *e.g.*, a subject entity described herein, to determine the amount of that entity in a sample. The

amount can be expressed, *e.g.*, in terms of % (*e.g.*, weight % or mole %) of the subject entity in the sample. In some embodiments, the amount of the subject entity is evaluated to determine if it is present in a preselected amount or range, *e.g.*, at least about 50%, *e.g.*, at least about 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%. In some embodiments, a percentage of an unpaired cysteine residue, disulfide linkage, or cystine knot, in a preparation has been determined. In some embodiments, the method includes determining or confirming that the percent of the unpaired cysteine residue or disulfide linkage (*e.g.*, to form a single disulfide, a nested disulfide, or a cystine knot) is in that range.

In some embodiments, the method includes determining the presence, amount, or distribution of one or more formylglycine residues and/or one or more cysteine residues that can be converted to formylglycine form, *e.g.*, the presence, amount, or distribution of Cys 51 and/or formylglycine at position 51 of ASA in a sample, *e.g.*, to determine if formylglycine is present at position 51 of ASA in a preselected amount or range, *e.g.*, at least about 50%, *e.g.*, at least about 60%, at least about 70%, at least about 80%, at least about 90%, *e.g.*, about 50% to about 90%, *e.g.*, about 60% to about 80%, about 65% to about 75%, *e.g.*, about 70%.

In some embodiments, the method further includes classifying, selecting, accepting or discarding, releasing or withholding, processing into drug product, shipping, moving to a different location, formulating, labeling, packaging, releasing into commerce, selling or offering to sell the preparation based, *e.g.*, on the result of the determination or upon comparison to a reference standard.

In another aspect, the disclosure features, a method of evaluating or processing a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation. The method includes: providing an evaluation of a parameter related to a subject entity, *e.g.*, a subject entity described herein, resolved by enzyme digestion, *e.g.*, multi-enzyme digestion; providing an evaluation of a parameter related to a subject entity resolved by LC-MS; and, optionally, providing a determination of whether a test value (*e.g.*, a value correlated to presence or absence, distribution, or amount) determined for the parameter for **Table 1**, or **Table 2**, each meets a preselected criterion for that subject entity, *e.g.*, is present or is present with a certain distribution, or amount as

described herein, *e.g.*, as depicted in **Table 1**, or **Table 2**, thereby evaluating or processing the preparation. In some embodiments, the method includes providing two or more determinations, *e.g.*, for two or more different subject entities depicted in **Table 1**, or **Table 2**.

In some embodiments, the method includes providing a comparison of the value determined for a parameter with a reference value or values, to thereby evaluate the sample. In some embodiments, the comparison includes determining if the test value has a preselected relationship with the reference value, *e.g.*, determining if it meets the reference value.

In another aspect, the disclosure features a method of evaluating or processing a sample, *e.g.*, a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation, that includes making a determination about a sample, *e.g.*, a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation, based upon a method or analysis described herein. In some embodiments, the method further includes classifying, selecting, accepting or discarding, releasing or withholding, processing into drug product, shipping, moving to a different location, formulating, labeling, packaging, releasing into commerce, selling or offering to sell the protein preparation, *e.g.*, the glycoprotein preparation, *e.g.*, the ASA preparation, *e.g.*, the rhASA preparation, based, *e.g.*, on the analysis. In some embodiment, the party making the evaluation does not practice the method or analysis described herein, but merely relies on results which are obtained by a method or analysis described herein.

In another aspect, the disclosure features methods of making a preparation, *e.g.*, a standard preparation of known concentration, by providing a compound described herein, *e.g.*, a compound depicted in **Table 1**, or **Table 2**, and combining it with a solvent. In some embodiments, the standard is at least about 50, 60, 70, 80, 90, 95, 99, 99.5, or 99.9 % of the total amount of the structures or species in the sample. The percentage can be determined, *e.g.*, by dry weight, chain, or molarity.

In one aspect, the disclosure features an isolated, enriched, or purified fraction of a sample, *e.g.*, a protein preparation, *e.g.*, glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation. In some embodiments, the sample is digested,

e.g., by one or more enzymes. In some embodiments, the fraction has one or more structures or species depicted in **Table 1**, or **Table 2**.

In one aspect, the disclosure features a method of analyzing a process, *e.g.*, a manufacturing process, of a protein, *e.g.*, a glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, *e.g.*, rhASA made by a selected process. The method includes: providing a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation; analyzing the protein preparation, using, *e.g.*, a method described herein, *e.g.*, to identify and/or quantify one or more subject entities, *e.g.*, one or more of the subject entities disclosed herein, thereby allowing analysis, *e.g.*, qualitative and/or quantitative analysis, of one or more of the subject entities in the protein preparation. In some embodiments, the method further includes comparing the presence, distribution, or amount of one or more of the subject entities with a reference value, to thereby analyze the process, *e.g.*, the manufacturing process.

In some embodiments, the method further includes maintaining the process, *e.g.*, the manufacturing process, based, at least in part, upon the analysis. In one embodiment, the method further includes altering the process, *e.g.*, the manufacturing process, based, at least in part, upon the analysis. The absence or presence of a subject entity, *e.g.*, a subject entity described herein, can indicate whether the process, *e.g.*, the manufacturing process, needs to be altered. In some embodiments, the party making the evaluation does not practice the method or analysis described herein, but merely relies on results which are obtained by a method or analysis described herein.

In some embodiments, the method includes comparing two or more protein preparations, *e.g.*, glycoprotein preparations, *e.g.*, ASA preparations, *e.g.*, rhASA preparations, *e.g.*, by a method of monitoring or controlling batch-to-batch variation or to compare a preparation to a reference standard. The method includes: providing a first protein preparation, *e.g.*, glycoprotein preparation, *e.g.*, ASA preparation, *e.g.*, rhASA preparation; providing the presence, amount, or distribution of one or more subject entities, *e.g.*, one or more of the subject entities described herein, in the first sample; optionally, providing a second protein preparation, *e.g.*, glycoprotein preparation, *e.g.*, ASA preparation, *e.g.*, rhASA preparation; providing the presence, distribution, or amount of one or more subject entities in the second preparation; and comparing the

presence, distribution, or amount of the one or more subject entities of the first protein preparation with the one or more subject entities of the second protein preparation. In one embodiment, the subject entity is analyzed by a method described herein.

In one embodiment, the method can further include making a decision, *e.g.*, to classify, select, accept or discard, release or withhold, process into drug product, move to a different location, ship, formulate, label, package, release into commerce, sell or offer to sell the preparation, *e.g.*, the protein preparation, *e.g.*, the glycoprotein preparation, *e.g.*, the ASA preparation, *e.g.*, the rhASA preparation, based, at least in part, upon the determination, and optionally, carrying out the decision.

In one aspect, the disclosure features a method of making a batch of protein, *e.g.*, glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, having a preselected property, *e.g.*, meeting a release specification, label requirement, or compendial requirement, *e.g.*, a property described herein. The method includes: providing a test protein preparation, *e.g.*, glycoprotein preparation, *e.g.*, ASA preparation, *e.g.*, rhASA preparation; analyzing the test preparation according to a method described herein; determining if the test preparation includes the presence, distribution, or amount of one or more of the structures provided in **Table 1**, or **Table 2**; and selecting the test preparation to make the protein, thereby making a batch of protein.

In one aspect, the disclosure features a method of predicting or ensuring that a batch of protein, *e.g.*, glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, will have a preselected property, *e.g.*, that it will meet a release specification, label requirement, or compendial requirement, *e.g.*, a property described herein. The method includes: providing a test preparation, *e.g.*, glycoprotein preparation, *e.g.*, ASA preparation, *e.g.*, rhASA preparation; analyzing the test protein preparation according to a method described herein, wherein satisfaction of the preselected reference, *e.g.*, one or more references disclosed herein, by the test protein preparation, is predictive of or ensures that a batch of protein made from the test protein preparation will have a preselected property, *e.g.*, it will meet a release specification, label requirement, or compendial requirement, *e.g.*, a property described herein.

In one aspect, the disclosure features a method of making one or more batches of a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an

rhASA preparation, having one or more disulfide linkages, wherein one or more subject entities of the batches varies less than a preselected range or has some preselected relationship with a reference standard. For example, it is present at a lower, higher, or equivalent level as a standard or is within (or outside) a range of values. In some embodiments, the method includes determining one or more subject entity (*e.g.*, one or more structures or fractions) of one or more batches of a product, and selecting a batch as a result of the determination. In some embodiments, the method can also include comparing the results of the determination to preselected values, *e.g.*, a reference standard. In some embodiments, evaluation of the value, *e.g.*, the presence of a subject entity, is made by a method described herein. In some embodiments, the method further includes classifying or selecting one or more batches having a structural property that varies less than the preselected range, *e.g.*, a range described herein. In other embodiments, the method can further include adjusting the dose of the batch to be administered, *e.g.*, based on the result of the determination of the subject entity.

In another aspect, the disclosure features a method of determining a reference value for a protein composition, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation, and determining the presence, amount, or distribution of one or more subject entities depicted in **Table 1**, or **Table 2**. In some embodiments, evaluation of the value, *e.g.*, the presence, distribution, or amount of the one or more of the subject entities, is made by a method described herein.

In another aspect, the disclosure features a method for determining bioequivalence. The method includes some or all of the following: providing or determining a value for the presence, amount, or distribution of one or more subject entities, *e.g.*, one or more of the subject entities described herein, in a first protein preparation, *e.g.* a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation; providing or determining the bioavailability of the preparation; providing a reference value, *e.g.*, by providing or determining presence, amount, or distribution of one or more subject entities, *e.g.*, one or more of the subject entities described herein, in a second protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation; and comparing the presence, amount, or distribution of one or more of the subject entities of the first preparation with the reference value, *e.g.*, from the

second protein preparation. In some embodiments, evaluation of the one or more structures or fractions is made by a method described herein.

In some embodiments, the method further comprises monitoring for presence, tissue distribution, spatial distribution, temporal distribution or retention time, in a cell or a subject, *e.g.*, an experimental animal. In some embodiments, the method includes determining the presence, amount, or distribution of one or more subject entities described herein of one or more batches of a product. In some embodiments, the method further includes selecting a batch as a result of the determination. In some embodiments, the method further includes comparing the results of the determination to preselected values, *e.g.*, a reference standard.

In another aspect, the invention provides a method for determining the safety or suitability of a protein preparation, *e.g.*, a glycoprotein preparation, *e.g.*, an ASA preparation, *e.g.*, an rhASA preparation, for use in a particular indication. The method includes one or more, typically all, of the following: determining the presence, amount, or distribution of one or more subject entities, *e.g.*, one or more of the subject entities described herein, in the protein; providing a reference value or sample; determining if the protein is acceptable, *e.g.*, by comparing a value for the presence, amount, or distribution of one or more subject entities of the protein with the reference value or with a value determined from the sample. For example, when the protein is rhASA, one or more of the structures, species, or fractions described herein can be used as a reference. When a preselected index of similarity is met, the protein can be determined to be safe or suitable. In some embodiments, the reference sample is associated with one or more undesired effects. In some embodiments, the reference sample is associated with one or more desired effects. In some embodiments, evaluation of the presence, amount, or size distribution of the one or more structures or fractions, *e.g.*, one or more structures or fractions described herein, in the protein is made by a method described herein. In some embodiments, the indication is MLD.

In another aspect, the disclosure features a method of one or more of: providing a report to a report receiving entity; evaluating a sample of protein, *e.g.*, glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, for compliance with a reference standard, *e.g.*, an FDA requirement; seeking indication from another party that a sample of protein, *e.g.*, glycoprotein, *e.g.*,

ASA, *e.g.*, rhASA, meets some predefined requirement; and submitting information about a sample of protein, *e.g.*, glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, to another party. Exemplary receiving entities or other parties include a government, *e.g.*, the U.S. federal government, *e.g.*, a government agency, *e.g.*, the FDA.

The method includes one or more (and typically all) of the following: performing one or more steps in making and/or testing a batch of protein, *e.g.*, glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, in a first country, typically the United States; sending at least an aliquot of the sample outside the first country, *e.g.*, sending it outside the United States, to a second country; preparing, or receiving, a report which includes data about the structure of the protein sample, *e.g.*, data related to a structure, species, or fraction described herein, *e.g.*, data generated by one or more of the methods described herein; and providing said report to a report recipient entity.

In some embodiments, the report receiving entity can determine if a predetermined requirement or reference value is met by said data and optionally, a response from the report receiving entity is received, *e.g.*, by a manufacturer, distributor, or seller of the protein preparation, *e.g.*, the rhASA preparation. In some embodiments, upon receipt of approval from said report recipient entity, protein, *e.g.*, rhASA, from said batch is selected, packaged, or placed into commerce.

In one aspect, the disclosure features a method of evaluating a sample of protein, *e.g.*, glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, that includes receiving data with regard to the presence or level of a structure or fraction described herein in a sample of protein, *e.g.*, glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, *e.g.*, wherein the data was prepared by one or more methods described herein; providing a record which includes said data and optionally includes an identifier for a batch of protein, *e.g.*, glycoprotein, *e.g.*, ASA, *e.g.*, rhASA; submitting said record to a decision-maker, *e.g.*, a government agency, *e.g.*, the FDA; optionally, receiving a communication from said decision maker; optionally, deciding whether to release market the batch of protein, *e.g.*, glycoprotein, *e.g.*, ASA, *e.g.*, rhASA, based on the communication from the decision maker. In one embodiment, the method further includes releasing the sample.

Any of the methods described herein can further include determining and/or providing an analysis regarding one or more biological activities or properties of the

preparation or sample. For example, the biological activity can be one or more of arylsulfatase activity, molecular weight distribution, and average molecular weight. The methods can further include comparing any of arylsulfatase activity, molecular weight distribution, and average molecular weight to a reference standard, *e.g.*, a reference standard described herein, for the protein, *e.g.*, the glycoprotein, *e.g.*, ASA, *e.g.*, rhASA. In some embodiments, the reference standard for arylsulfatase activity is about 20 to about 250 U/mg, *e.g.*, about 50 to 140 about U/mg.

In one aspect, the disclosure features an enriched, isolated, or purified preparation of a compound from **Table 1**, or **Table 2**.

In one aspect, the disclosure features a set of standard preparations. The set includes a plurality of standards each having a different concentration of a compound of **Table 1**, or **Table 2**. In some embodiments, the standard preparation is free of other subject entities, *e.g.*, other structures, species, or fractions described herein. The standard preparations can be used to determine the identity of the subject entity. The standard preparations can also be used to evaluate the concentration of the subject entity. For example, concentrations can be evaluated in terms of weight/weight, weight/volume, or molarity. In some embodiments, the compound is provided in a solvent. The set of standards can be used in the evaluation of one or more samples, *e.g.*, one can assay for a subject entity and compare the assay result with a value obtained from one or more of the standards. For example, one can determine the absorbance or other parameters and compare that with a standard curve for the relevant parameter derived from the set of standard preparations and determine the concentration of the subject entity. In some embodiments, the standard in a set is individually enriched, isolated, or purified. In some embodiment, the set includes at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 standards.

The details of one or more embodiments of the disclosure are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts the primary structure of rhASA with disulfide linkages and unpaired cysteines. The sites of disulfide linkages and unpaired cysteines are indicated and the N-glycosylation motifs are underlined.

FIG. 2A depicts the mass and charge of the pepsin-digested peptide with an unmodified (free) Cys20.

FIG. 2B depicts the CID-MS2 spectrum of the precursor from **FIG. 2A**. The sequence and theoretical mass of the peptide are indicated in the insert of **FIG. 2A**.

FIG. 3A depicts the mass and charge of the pepsin-digested peptide with a modified (fgly) Cys51.

FIG. 3B depicts the CID-MS2 spectrum of the precursor from **FIG. 3A**. The sequence and theoretical mass of the peptide are indicated in the insert of **FIG. 3A**.

FIG. 4A depicts the mass and charge of the pepsin-digested peptide with an unmodified (free) Cys51.

FIG. 4B depicts the CID-MS2 spectrum of the precursor from **FIG. 4A**.

FIG. 5A depicts the mass and charge of the pepsin-digested peptide with an unmodified (free) Cys276.

FIG. 5B depicts the CID-MS2 spectrum of the precursor from **FIG. 5A**.

FIG. 6A depicts the mass and charge of the Lys-C plus trypsin-digested peptide with a single disulfide (Cys282 with Cys396).

FIG. 6B depicts the CID-MS2 spectrum of the precursor from **FIG. 6A**.

FIG. 6C depicts the ETD-MS2 spectrum of the precursor from **FIG. 6A**.

FIG. 7 depicts the nested disulfides in a tryptic peptide.

FIG. 8 depicts the digestion strategy for nested disulfides.

FIG. 9 depicts the digestion strategy for nested disulfides using Lys-C+Trypsin+PNGase F+Asp-N.

FIG. 10A depicts the mass and charge of the Lys-C plus trypsin plus Asp-N plus PNGaseF-digested peptide with two disulfides (Cys138 with Cys154 and Cys143 with Cys150). The sequence and theoretical mass of the peptide are indicated in the insert.

FIG. 10B depicts the CID-MS2 spectrum of the 2+ and 3+ charged precursor from **FIG. 10A**.

FIG. 10C depicts the ETD-MS2 spectrum of the 2+ and 3+ charged precursor from **FIG. 10A**.

FIG. 11 depicts the precursor ion mass (2+) and its corresponding CID-MS2 spectra of the scrambled disulfides.

FIG. 12 depicts the precursor ion mass (4+) and its corresponding CID-MS2 spectra of the nested disulfide-linked peptide (derived from pepsin digestion).

FIG. 13A depicts the mass and charge of the pepsin-digested peptide with three disulfides (Cys470 with Cys482, Cys471 with Cys484 and Cys475 with Cys481). The sequence and theoretical mass of the peptide are indicated in the insert.

FIG. 13B depicts the CID-MS2 spectrum (using the LTQ) of the 3+ charged precursor from **FIG. 13A**.

FIG. 13C depicts the ETD-MS2 spectrum (using the Orbitrap) of the 4+ charged precursor from **FIG. 13A**.

FIG. 13D depicts the ETD-MS2 spectrum (using the Orbitrap) of the 5+ charged precursor from **FIG. 13A**.

FIG. 13E depicts the CID-MS3 spectrum (using the Orbitrap) of m/z 1312.6 from **FIG. 13C**.

FIG. 13F depicts the CID-MS3 spectrum (using the LTQ) of m/z 1312.6 from **FIG. 13C**.

DETAILED DESCRIPTION

The disclosure relates, at least in part, to methods of analyzing and/or preparing samples (*e.g.*, protein samples) that contain one or more disulfide linkages. The methods described herein can include one or more of enzymatic digestion, chromatography, and mass spectrometry, *e.g.*, multi-enzyme digestion and/or liquid chromatography-mass spectrometry (LC-MS). Characterization of proteins such as rhASA during or after biopharmaceutical manufacturing can increase manufacturers' ability to maintain the drug function and to control the stability or batch-to-batch variability.

Definitions

The term “protein preparation” as used herein refers to both protein drug substance preparations and protein drug product preparations. The term “rhASA preparation” as used herein refers to both rhASA drug product preparations and rhASA drug substance preparations. The term “drug product preparation” refers to a preparation having the purity required for and being formulated for pharmaceutical use. The term “drug substance preparation” refers to a preparation for pharmaceutical use but is not necessarily in its final formulation and/or comprises one or more non-product contaminant (*e.g.*, one or more inorganic product such as sulfate, chloride, acetate and phosphates, protein contaminant, process by-product such as benzyl alcohol and benethonium).

The term “enriched preparation” or “enriched fraction” as used herein refers to a preparation or fraction which is significantly enriched for a subject entity, *e.g.*, a subject entity described herein. Significant enrichment can, by way of example, be based on weight/weight, weight/volume, or molarity. Enrichment can be with respect to a naturally occurring material, *e.g.*, protein, *e.g.*, glycoprotein, *e.g.*, ASA. In some embodiments, in the case of a subject entity which is present in the naturally occurring protein, the subject entity is present in the enriched preparation at least 2, 5, 10, 50 or 100 times the concentration (as determined, *e.g.*, by weight/weight, weight/volume, or molarity) that is found in the protein that has not been enriched. In some embodiments, in the case of a subject entity which is present in rhASA, the subject entity is present in the enriched preparation at least 2, 5, 10, 50 or 100 times the concentration (as determined, *e.g.*, by weight/weight, weight/volume, or molarity) that is found in rhASA that has not been enriched. In some embodiments, the subject entity can be accompanied by a solvent, diluent, or carrier. In other embodiments, the subject entity is substantially free of a solvent, diluent, or carrier. In some embodiments, the subject entity can be accompanied by a medium, *e.g.*, a buffer, matrix, or other material used to effect separation and/or eluent, used in its enrichment. In other embodiments, the preparation is substantially free of such elements. In a preferred embodiment, the preparation is provided in an enclosure which is substantially free of contaminants.

The term “isolated preparation” or “isolated fraction” as used herein refers to a preparation or fraction which is significantly isolated for a subject entity, *e.g.*, a subject entity described herein. Significant isolation can, by way of example, be based on weight/weight, weight/volume, or molarity. Isolation can be with respect to a naturally occurring material, *e.g.*, protein, *e.g.*, glycoprotein, *e.g.*, ASA. In some embodiments, in the case of a subject entity which is present in the naturally occurring protein, the subject entity is present in the isolated preparation at least 2, 5, 10, 50 or 100 times the concentration (as determined, *e.g.*, by weight/weight, weight/volume, or molarity) that is found in the protein that has not been isolated. In some embodiments, in the case of a subject entity which is present in rhASA, the subject entity is present in the isolated preparation at least 2, 5, 10, 50 or 100 times the concentration (as determined, *e.g.*, by weight/weight, weight/volume, or molarity) that is found in rhASA that has not been isolated. In some embodiments, the subject entity can be accompanied by a solvent, diluent, or carrier. In other embodiments, the subject entity is substantially free of a solvent, diluent, or carrier. In some embodiments, the subject entity can be accompanied by a medium, *e.g.*, a buffer, matrix, or other material used to effect separation and/or eluent, used in its isolation. In other embodiments, the preparation is substantially free of such elements. In a preferred embodiment, the preparation is provided in an enclosure which is substantially free of contaminants.

The term “purified preparation” or “purified fraction” as used herein refers to a preparation or fraction which is significantly purified for a subject entity, *e.g.*, a subject entity described herein. Significant purification can, by way of example, be based on weight/weight, weight/volume, or molarity. Purification can be with respect to a naturally occurring material, *e.g.*, protein, *e.g.*, glycoprotein, *e.g.*, ASA. In some embodiments, in the case of a subject entity which is present in the naturally occurring protein, the subject entity is present in the purified preparation at least 2, 5, 10, 50 or 100 times the concentration (as determined, *e.g.*, by weight/weight, weight/volume, or molarity) that is found in the protein that has not been purified. In some embodiments, in the case of a subject entity which is present in rhASA, the subject entity is present in the purified preparation at least 2, 5, 10, 50 or 100 times the concentration (as determined, *e.g.*, by weight/weight, weight/volume, or molarity) that is found in rhASA that has not

been purified. In some embodiments, the subject entity can be accompanied by a solvent, diluent, or carrier. In other embodiments, the subject entity is substantially free of a solvent, diluent, or carrier. In some embodiments, the subject entity can be accompanied by a medium, *e.g.*, a buffer, matrix, or other material used to effect separation and/or eluent, used in its purification. In other embodiments, the preparation is substantially free of such elements. In a preferred embodiment, the preparation is provided in an enclosure which is substantially free of contaminants.

Arylsulfatase A

Arylsulfatase A (or cerebroside-sulfatase) is an enzyme that breaks down cerebroside 3-sulfate (or sulfatide) into cerebroside and sulfate. Specifically, galactosyl sulfatide is normally metabolized by the hydrolysis of 3-O-sulphate linkage to form galactocerebroside through the combined action of the lysosomal enzyme arylsulfatase A (EC 3.1.6.8) (Austin *et al. Biochem J.* 1964, 93, 15C-17C) and a sphingolipid activator protein called saposin B. A deficiency of arylsulfatase A occurs in all tissues from patients with the late infantile, juvenile, and adult forms of metachromatic leukodystrophy (MLD). As used herein, the arylsulfatase A protein will be termed “ASA” and the saposin B will be termed “Sap-B”.

ASA is an acidic glucoprotein with a low isoelectric point. Above pH 6.5, the enzyme exists as a monomer with a molecular weight of approximately 100 kDa. ASA undergoes a pH-dependent polymerisation forming a dimer at pH 4.5. In human urine, the enzyme consists of two nonidentical subunits of 63 and 54 kDa (Laidler PM *et al. Biochim Biophys Acta.* 1985, 827, 73-83). ASA purified from human liver, placenta, and fibroblasts also consist of two subunits of slightly different sizes varying between 55 and 64 kDa (Draper RK *et al. Arch Biochemica Biophys.* 1976, 177, 525-538, Waheed A *et al. Hoppe Seylers Z Physiol Chem.* 1982, 363, 425-430, Fujii T *et al. Biochim Biophys Acta.* 1992, 15 1122, 93-98). As in the case of other lysosomal enzymes, ASA is synthesized on membrane-bound ribosomes as a glycosylated precursor. It then passes through the endoplasmic reticulum and Golgi, where its N-linked oligosaccharides are processed with the formation of phosphorylated and sulfated oligosaccharide of the complex type (Waheed A *et al. Biochim Biophys Acta.* 1985, 847, 53-61, Braulke T *et al.*

Biochem Biophys Res Commun. 1987, 143, 178-185). In normal cultured fibroblasts, a precursor polypeptide of 62 kDa is produced, which translocates via mannose-6-phosphate receptor binding (Braulke T *et al. J Biol Chem.* 1990, 265, 6650-6655) to an acidic prelysosomal endosome (Kelly BM *et al. Eur J Cell Biol.* 1989, 48, 71-78).

The methods described herein can be used to purify ASA from any source, *e.g.*, from tissues, or cultured cells (*e.g.*, human cells (*e.g.*, fibroblasts) that recombinantly produce ASA).

The length (18 amino acids) of the human ASA signal peptide is based on the consensus sequence and a specific processing site for a signal sequence. Hence, from the deduced human ASA cDNA (EMBL GenBank accession numbers J04593 and X521151) the cleavage of the signal peptide occurs in all cells after residue number 18 (Ala), resulting in the mature form of the human ASA.

Multiple forms of ASA have been demonstrated on electrophoresis and isoelectric focusing of enzyme preparations from human urine (Luijten JAFM *et al. J Mol Med.* 1978, 3, 213), leukocytes (Dubois *et al. Biomedicine.* 1975, 23, 116-119, Manowitz P *et al. Biochem Med Metab Biol.* 1988, 39, 117-120), platelets (Poretz *et al. Biochem J.* 1992, 287, 979-983), cultured fibroblasts (Waheed A *et al. Hoppe Seylers Z Physiol Chem.* 1982, 363, 425-430, Stevens RL *et al. Biochim Biophys Acta.* 1976, 445, 661-671, Farrell DF *et al. Neurology.* 1979, 29, 16-20) and liver (Stevens RL *et al. Biochim Biophys Acta.* 1976, 445, 661-671, Farrell DF *et al. Neurology.* 1979, 29, 16-20, Sarafian TA *et al. Biochem Med.* 1985, 33, 372-380). Treatment with endoglycosidase H, sialidase, and alkaline phosphatase reduces the molecular size and complexity of the electrophoretic pattern, which suggests that much of the charge heterogeneity of ASA is due to variations in the carbohydrate content of the enzyme.

The active site of ASA contains an essential histidine residue (Lee GD and Van Etten RL, *Arch Biochem Biophys.* 1975, 171, 424-434) and two or more arginine residues (James GT, *Arch Biochem Biophys.* 1979, 97, 57-62). Many anions are inhibitors of the enzyme at concentrations in the millimolar range or lower.

A protein modification has been identified in two eukaryotic sulfatases (ASA and arylsulfatase B (ASB)) and for one from the green alga *Volvox carteri* (Schmidt B *et al. Cell.* 1995, 82, 271-278, Selmer T *et al. Eur J Biochem.* 1996, 238, 341-345). This

modification leads to the conversion of a cysteine residue, which is conserved among the known sulfatases, into a 2-amino-3-oxopropionic acid residue (Schmidt B *et al. Cell.* 1995, 82, 271-278). The novel amino acid derivative is also recognized as C-formylglycine (FGly). In ASA and ASB derived from MSD cells, the Cys-69 residue is retained. Consequently, it is proposed that the conversion of the Cys-69 to FGly-69 is required for generating catalytically active ASA and ASB, and that deficiency of this protein modification is the cause of MSD. Cys-69 is referred to the precursor ASA which has an 18 residue signal peptide. In the mASA the mentioned cysteine residue is Cys-51.

Further investigations have shown that a linear sequence of 16 residues surrounding the Cys-51 in the mASA is sufficient to direct the conversion and that the protein modification occurs after or at a late stage of co-translational protein translocation into the endoplasmic reticulum when the polypeptide is not yet folded to its native structure (Dierks T *et al. Proc Natl Acad Sci.* 1997, 94, 11963-11966, Wittke, D. *et al. (2004), Acta Neuropathol. (Berl.)*, 108, 261-271).

The human ASA gene structure has been described. The ASA gene is located near the end of the long arm of chromosome 22 (22q13.31-qter), it spans 3.2 kb (Kreysing *et al. Eur J Biochem.* 1990, 191, 627-631) and consists of eight exons specifying the 507 amino acid enzyme unit (Stein *et al. J Biol Chem.* 1989, 264, 1252-1259). Messenger RNAs of 2.1, 3.7, and 4.8 kb have been detected in fibroblast cells, with the 2.1-kb message apparently responsible for the bulk of the active ASA generated by the cell (Kreysing *et al. Eur J Biochem.* 1990, 191, 627-631). The ASA sequence has been deposited at the EMBL GenBank with the accession number X521150. Differences between the published cDNA and the coding part of the ASA were described by Kreysing *et al. (Eur J Biochem.* 1990, 191, 627-631). The cDNA sequence originally described by Stein *et al. (J Biol Chem.* 1989, 264, 1252-1259) and the cDNA sequence described by Kreysing *et al. (Eur J Biochem.* 1990, 191, 627-631) have been deposited at the EMBL GenBank with the following accession numbers J04593 and X521151, respectively.

Several polymorphisms and more than 40 disease-related mutations have been identified in the ASA gene (Gieselmann *et al. Hum Mutat.* 1994, 4, 233-242, Barth *et al. Hum Mutat.* 1995, 6, 170-176, Draghia *et al. Hum Mutat.* 1997, 9, 234-242). The disease-related mutations in the ASA gene can be categorised in two broad groups that

correlate fairly well with the clinical phenotype of MLD. One group (I) produces no active enzyme, no immunoreactive protein, and expresses no ASA activity when introduced into cultured animal cell lines. The other group (A) generates small amounts of cross-reactive material and low levels of functional enzyme in cultured cells. Individuals homozygous for a group (I) mutation, or having two different mutations from this group, express late infantile MLD. Most individuals with one group (I)-type and one group (A)-type mutation develop the juvenile-onset form, whereas those with two group (A)-type mutations generally manifest adult MLD. Some of the mutations have been found relatively frequently, whereas others have been detected only in single families. It is possible to trace specific mutations through members of many families, however general carrier screening is not yet feasible.

In addition to the disease-related mutations described above, several polymorphisms have been identified in the ASA gene. Extremely low ASA activity has been found in some clinically normal parents of MLD patients and also in the general population. This so-called pseudodeficiency ASA has been associated with a common polymorphism of the ASA gene (Gieselmann *et al. Dev Neurosci.* 1991, 13, 222-227).

The crystal structure of human ASA (without glycosylation) shows that there are a total of 15 cysteines, six disulfide linkages, and three free cysteines including one that is posttranslationally modified to formylglycine (Lukatela *et al., Biochem.* 1998, 37, 3654-3664). A cystine knot is formed at the C-terminal end of the molecule that consists of three disulfide linkages: Cys470 with Cys482, Cys471 with Cys484, and Cys475 with Cys481. The term "cystine knot" as used herein refers to a protein structural motif where three disulfides (6 cysteine residues in close proximity in a protein backbone), with one of the disulfide passing through a ring, formed by the other two disulfide bonds (Le Nguyen *Biochimie.* 1990, 72(6-7): 431-5). Cystine knots are known to enhance protein structural stability, and they can be found in many proteins with a wide range of biological functions, such as inhibition, growth stimulation, and cyclization (Alvarez *Reprod. Biol. Endocrinol.* 2009, 7: 90; Daly *et al., Curr. Opin. Chem. Biol.* 2011, 15: 362-8). However, the cysteine-knot family shows low sequence homology, and it is therefore hard to predict cysteine-knot signatures by sequence alignment. Furthermore, there are 15 ways to form three disulfides into a cystine knot, and the

determination of the correct assignment of the disulfide bonds is a challenge. The disruption of the cystine knot by the mutation of Cys 470 with Arg from patients with MLD has been reported (Coulter-Mackie *et al.*, *Mol. Genet. Metab.* 2003, 79, 91-98), and that indicates that the cystine knot is associated with protein function. The cystine knot is also known to stabilize the protein where recent studies have shown that a partially or fully reduced cystine knot makes the protein susceptible to chemical or proteolytic degradation (Kolmar *FEBS J.* 2008, 275, 2684-2690).

The conformation of ASA is also pH-dependent. ASA forms a homo-dimeric protein at neutral pH but becomes a homo-octamer at acidic pH, such as in the lysosome. The stability of the enzyme seems to relate to the dimer-to-octamer transition in the lysosomal milieu, in which the octamerization process has been shown to be disrupted by a phenylalanine replacement mutation at Cys282 (Marcao *et al.*, *Biochem. Biophys. Res. Commun.*, 2003, 306, 293-297).

Enzyme Digestion

A sample described herein (*e.g.*, an rhASA preparation) can be lyophilized and/or dried in a vacuum oven, *e.g.*, at about 40°C, 43°C, 46°C, 49°C, 52°C or 55°C, for about 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 or 24 hours. For example, the sample can be lyophilized and/or dried under one of the following conditions: 40°C for 12 hours; 46°C for 8 hours; 49°C for 6 hours; 52°C for 4 hours. A lyophilized or dried sample can be reconstituted in water or a suitable buffer at a concentration of about 1, 2, 5, 10, 15, 20, 50 mg/mL.

The sample, *e.g.*, an ASA preparation, can be buffer exchanged with, *e.g.*, 100 mM ammonium bicarbonate (pH 8), 50 mM Tris-HCl buffer (pH 6.8), or 10 mM HCl (pH 2), *e.g.*, using a 10kD molecular weight cutoff filter, and concentrated to, *e.g.*, 1 mg/mL. Enzyme can be added to the protein solution at 1: 5, 1:10, 1:20, 1:30, 1:40, 1:50, 1:60, 1:70, 1:80, 1: 90, 1:100 (w/w), or less. The enzyme digestion can be performed at room temperature or at 37°C. The time of incubation can be, *e.g.*, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more hours. For multi-enzyme digestion, the protein can be digested by two or more enzymes at the same time. Alternatively, the protein can be digested sequentially by two or more enzymes.

Exemplary enzymes include, but not limited to endoproteinases and endoglycosidases. For example, one or more of the pepsin, trypsin, Lys-C, and PNGase F can be selected.

Pepsin hydrolyzes only peptide bonds. It does not hydrolyze non-peptide amide or ester linkages. Pepsin exhibits preferential cleavage for hydrophobic, preferably aromatic, residues in P1 and P1' positions. Increased susceptibility to hydrolysis occurs if there is a sulfur-containing amino acid close to the peptide bond, which has an aromatic amino acid. Pepsin will also preferentially cleave at the carboxyl side of phenylalanine and leucine and to a lesser extent at the carboxyl side of glutamic acid residues. Pepsin will not cleave at valine, alanine, or glycine linkages. Amidation of the C-terminal carboxyl group prevents hydrolysis by pepsin.

Trypsin specifically hydrolyzes peptide bonds at the carboxyl side of lysine and arginine residues.

Lys-C is a serine protease that specifically hydrolyzes amide, ester, and peptide bonds at the carboxylic side of Lys.

Asp-N is an endoproteinase which selectively cleaves peptide bonds N-terminal to Asp, Glu, and Cys residues.

PNGaseF cleaves an entire glycan from a glycoprotein provided the glycosylated asparagine moiety is substituted on its amino and carboxyl terminus with a polypeptide chain.

Liquid Chromatography–Mass Spectrometry

Liquid chromatography–mass spectrometry (LC-MS, or alternatively HPLC-MS) is an analytical chemistry technique that combines the physical separation capabilities of liquid chromatography (or HPLC) with the mass analysis capabilities of mass spectrometry. LC-MS is a technique that has high sensitivity and selectivity and that can be used for various applications. For example, LC-MS can be used for detection and potential identification of chemicals in the presence of other chemicals (in a complex mixture).

The scale of chromatography in LC-MS is generally smaller than that in traditional HPLC, *e.g.*, with respect to the internal diameter of the column and the flow

rate. For example, 1 mm columns with an internal diameter equal to or less than 1 mm (*e.g.*, 300 μm or 75 μm) can be used for LC-MS work, as opposed to 4.6 mm for traditional HPLC. At the low end of these column diameters, the flow rates approach 100 nL/min and can be used with nanospray sources. When standard bore (4.6 mm) columns are used the flow is often split ~10:1. This can be beneficial by allowing the use of other techniques in tandem such as MS and UV. However splitting the flow to UV will decrease the sensitivity of spectrophotometric detectors. The mass spectrometry on the other hand will give improved sensitivity at flow rates of 200 $\mu\text{L}/\text{min}$ or less.

Various different mass analyzers can be used in LC-MS. Exemplary mass analyzers include, but not limited to, single quadrupole, triple quadrupole, ion trap, TOF (time of flight), and quadrupole-time of flight (Q-TOF).

The quadrupole mass analyzer consists of 4 circular rods, set parallel to each other. In a quadrupole mass spectrometer (QMS), the quadrupole is the component of the instrument responsible for filtering sample ions, based on their mass-to-charge ratio (m/z). Ions are separated in a quadrupole based on the stability of their trajectories in the oscillating electric fields that are applied to the rods.

A triple quadrupole mass spectrometer is a tandem mass spectrometer consisting of two quadrupole mass spectrometers in series, with a (non mass-resolving) radio frequency (RF) only quadrupole between them to act as a collision cell for collision-induced dissociation. The first (Q_1) and third (Q_3) quadrupoles serve as mass filters, whereas the middle (q_2) quadrupole serves as a collision cell. This collision cell is an RF only quadrupole (non-mass filtering) using an inert gas such as Ar, He, or N_2 gas to provide collision-induced dissociation of a selected precursor ion that is selected in Q_1 . Subsequent fragments are passed through to Q_3 where they may be filtered or scanned. This configuration is often abbreviated QqQ, here $Q_1q_2Q_3$.

An ion trap is a combination of electric or magnetic fields that captures ions in a region of a vacuum system or tube. Ion traps can be used in mass spectrometry while the ion's quantum state is manipulated.

Time-of-flight mass spectrometry (TOFMS) is a method of mass spectrometry in which an ion's mass-to-charge ratio is determined via a time measurement. Ions are accelerated by an electric field of known strength. This acceleration results in an ion

having the same kinetic energy as any other ion that has the same charge. The velocity of the ion depends on the mass-to-charge ratio. The time that it subsequently takes for the particle to reach a detector at a known distance is measured. This time will depend on the mass-to-charge ratio of the particle (heavier particles reach lower speeds). From this time and the known experimental parameters one can find the mass-to-charge ratio of the ion.

A quadrupole-time of flight (Q-TOF) is a triple quadrupole with the last quadrupole section replaced by a TOF analyzer.

The interface can be an electrospray ion source or variant such as a nanospray source. Atmospheric pressure chemical ionization interface can also be used. Various deposition and drying techniques have also been used such as using moving belts, *e.g.*, off-line MALDI deposition. In addition, Direct-EI LC-MS interface, couples a nano HPLC system and an electron ionization equipped mass spectrometer, can also be used.

Collision induced dissociation

In mass spectrometry, collision-induced dissociation (CID), also known as collisionally activated dissociation (CAD), is a mechanism by which to fragment molecular ions in the gas phase. The molecular ions are usually accelerated by some electrical potential to high kinetic energy and then allowed to collide with neutral molecules (*e.g.*, helium, nitrogen or argon). In the collision, some of the kinetic energy is converted into internal energy which results in bond breakage and the fragmentation of the molecular ion into smaller fragments. These fragment ions can then be analyzed by a mass spectrometer.

CID and the fragment ions produced by CID are used for several purposes. Partial or complete structural determination can be achieved. In some cases, identity can be established based on previous knowledge without determining structure. Another use is in simply achieving more sensitive and specific detection. By looking for a unique fragment ion you can detect a given molecule in the presence of other molecules of the same nominal molecular weight, essentially reducing the background and increasing the limit of detection.

For example, in a triple quadrupole mass spectrometer, the first quadrupole termed "Q1" can act as a mass filter and transmits a selected ion and accelerates it

towards "Q2" which is termed a collision cell. The pressure in Q2 is higher and the ions collide with neutral gas in the collision cell and fragments by CID. The fragments are then accelerated out of the collision cell and enter Q3 which scans through the mass range, analyzing the resulting fragments (as they hit a detector). This produces a mass spectrum of the CID fragments from which structural information or identity can be gained. Many other experiments using CID on a triple quadrupole exist such as precursor ion scans that determine where a specific fragment came from rather than what fragments are produced by a given molecule.

SORI-CID (sustained off-resonance irradiation collision-induced dissociation) is a CID technique used in Fourier Transform Ion Cyclotron Resonance mass spectrometry which involves accelerating the ions in cyclotron motion (in a circle inside of an ion trap) and then increasing the pressure resulting collisions that produce CID fragments.

CID can be performed on charge-reduced species, e.g., isolated from ETD fragment ions. This type of CID is termed CRCID.

Electron transfer dissociation

Electron-transfer dissociation (ETD) is a method of fragmenting ions in a mass spectrometer. ETD induces fragmentation of cations (e.g. peptides or proteins) by transferring electrons to them. ETD does not use free electrons but employs radical anions (e.g. anthracene or azobenzene) for this purpose. ETD cleaves randomly along the peptide backbone (so called c and z ions) while side chains and modifications such as phosphorylation are left intact. The technique is suitable for higher charge state ions ($z > 2$). ETD is also suitable for the fragmentation of longer peptides or even entire proteins. ETD is also effective for peptides with modifications such as phosphorylation.

Reference Values and Standards

A reference value, for example, can be a value determined from a reference sample (e.g., a commercially available sample or a sample from previous production). For example, a reference value can be a value for the presence of a subject entity in a sample, e.g., a reference sample. The reference value can be numerical or non-numerical, e.g., it can be a qualitative value, e.g., yes or no, or present or not present at a preselected

level of detection, or graphic or pictorial. The reference value can also be values for the presence of more than one subject entity in a sample. For example, the reference value can be a map of structures present in a protein, *e.g.*, ASA, *e.g.*, rhASA, when analyzed by LC-MS, *e.g.*, an LC-MS method described herein. The reference value can also be a release standard (a release standard is a standard which should be met to allow commercial sale of a product) or production standard, *e.g.*, a standard which is imposed, *e.g.*, by a party, *e.g.*, the FDA, on a protein, *e.g.*, ASA, *e.g.*, rhASA.

The reference standard can be derived from any of a number of sources. The reference standard can be one which was set or provided by (either solely or in conjunction with another party, *e.g.*, a regulatory agency, *e.g.*, the FDA) the manufacturer of the drug or practitioner of a process to make the drug. The reference standard can be one which was set or provided by (either solely or in conjunction with another party, *e.g.*, a regulatory agency, *e.g.*, the FDA) a party other than the party manufacturing a drug and practicing a method disclosed herein, *e.g.*, another party which manufactures the drug or practices a process to make the drug. The reference standard can be one which was set or provided by (either solely or in conjunction with another party) a regulatory agency, *e.g.*, the FDA, to the manufacturer of the drug or practitioner of the process to make the drug, or to another party licensed to market the drug. For example, the reference standard can be a production, release, or product standard required by the FDA. In some embodiments, a reference standard is a standard required of a pioneer drug (*e.g.*, a drug marketed under an approved NDA) or a generic drug (*e.g.*, a drug marketed or submitted for approval under an ANDA).

The reference standard can be one which was set or provided by Shire, its fully owned subsidiaries, its successors and assigns or agents, either solely or in conjunction with another party, *e.g.*, a regulatory agency, *e.g.*, the FDA, for production or release of a protein, *e.g.*, ASA, *e.g.*, rhASA.

The reference value can be a statistical function, *e.g.*, an average, of a number of values. The reference value can be a function of another value, *e.g.*, of the presence or distribution of a second entity present in the sample, *e.g.*, an internal standard.

Evaluation against a reference value can be used to determine if a particular subject entity is present in an ASA sample, *e.g.*, an rhASA sample.

EXAMPLES

The following examples describe the characterization of unpaired cysteines, disulfide linkages (*e.g.*, nested pairs of cysteines and cystine knot), and posttranslational modification of a cysteine to formylglycine for rhASA preparations. The statuses of these cysteines are critical structure attributes for rhASA function and stability that requires precise examination. In this study, a novel approach, which was comprised of multi-enzyme digestion strategies (*e.g.*, one or more of Lys-C, trypsin, Asp-N, pepsin, and PNGase F) and multi-fragmentation methods in mass spectrometry (*e.g.*, using electron transfer dissociation (ETD), collision induced dissociation (CID), and/or CID with MS3 (after ETD)), was used to determine the status and linkage of each cysteine in rhASA. In addition to generating desired lengths of enzymatic peptides for effective fragmentation, the digestion pH was optimized to minimize the disulfide scrambling. The disulfide linkages, including the cystine knot and a pair of nested cysteines, unpaired cysteines, and the posttranslational modification of a cysteine to formylglycine, were all determined. The primary structure of rhASA is shown in **FIG. 1**. In the disulfide assignment, the disulfide linkages observed were Cys138 with Cys154, Cys143 with Cys150, Cys282 with Cys396, Cys470 with Cys482, Cys471 with Cys484, and Cys475 with Cys481. For the unpaired cysteines, Cys20 and Cys276 were free cysteines, and Cys51 was largely converted to formylglycine (> 70%). These data indicated that the structure of rhASA was preserved throughout the manufacturing process as the disulfide linkages observed in rhASA were the same as those predicted from the crystal structure. Thus, a successful and robust method has been developed which can be routinely used to determine the difficult-to-resolve disulfide linkages that are present in proteins (*e.g.*, rhASA and other recombinant proteins), ensuring drug function and stability.

Example 1: Determination of Cystine knot and Disulfide Linkages in Recombinant Human Arylsulfatase A

Materials and Methods

Samples: rhASA, GMP lot JPT11001, manufactured by Shire Human Genetic Therapies (Lexington, MA, USA) was provided at 39.1 mg/mL. The sample was aliquoted (10 μ L or 391 μ g per vial) and stored at -80°C before analysis.

Reagents: Sequencing-grade trypsin was purchased from Promega (Madison, WI). Achromobacter protease I (Lys-C) was obtained from Roche (Nutley, NJ). Asp-N, pepsin, PNGase F, ammonium bicarbonate (NH_4HCO_3), and formic acid (FA) were purchased from Sigma-Aldrich (St. Louis, MO). LC-MS grade water was from J.T. Baker (Phillipsburg, NJ), and HPLC grade acetonitrile from ThermoFisher Scientific (Fairlawn, NJ). Amicon centrifugal filter units (10kDa MWCO) were obtained from Millipore (Bedford, MA).

Enzymatic Digestion: The protein solution (10 μ L of 4.9 mg/mL) was buffer exchanged with 100 mM ammonium bicarbonate (pH 8) or 50 mM Tris-HCl buffer (pH 6.8) using a 10 kDa molecular weight cutoff filter and concentrated to 1 mg/mL (49 μ L). In addition to pH 8, a slightly less than alkaline pH (pH 6.8) was used to examine the effect of pH on the formation of alternative disulfide linkages during the digestion procedure. If a difference was observed, pepsin digestion at pH 2 was used to eliminate the scrambling that can occur at high pH conditions. For a pepsin digestion, the protein solution was buffer exchanged with 10 mM HCl (pH 2). Pepsin (1:10, w/w) was added to the protein solution and incubated at 37°C for 30 min. The reaction was quenched by adjusting the pH to 6 with sodium hydroxide. For a Lys-C plus trypsin digestion, the protein solution (pH 6.8 or 8) was incubated with endoproteinase Lys-C (1:50 w/w) and trypsin (1:50 w/w) for 8 hr at room temperature, and then added a second time (1:50 w/w for each enzyme) and allowed to incubate for an additional 12 hr at room temperature. For a Lys-C plus trypsin, Asp-N, and PNGase F digestion, the protein solution (pH 6.8 or 8) was treated with endoproteinase Lys-C (1:50 w/w), trypsin (1:50 w/w), Asp-N (1:50 w/w), and PNGase F (10 units/mg) for 8 hr at room temperature and then added a second time (the same ratio for each enzyme) and allowed to incubate for an additional 12 hr at room temperature. In all cases except pepsin digestion, digestion was terminated by the

addition of 1% formic acid. An aliquot of 2 µg of the enzyme digest was analyzed per LC-MS run.

LC-MS: An Ultimate 3000 nano-LC pump (Dionex, Mountain View, CA) and a self-packed C18 column (Magic C18, 200 Å pore and 5 µm particle size, 75 µm i.d. × 15 cm) (Michrom Bioresources, Auburn, CA) was coupled online to an LTQ-Orbitrap-ETD XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a nanospray ion source (New Objective, Woburn, MA). Mobile phase A was 0.1% formic acid in water, and mobile B was 0.1% formic acid in acetonitrile. The peptides were eluted at 200 nL/min using a linear gradient from 2% to 60% B in 90 min, followed by 60% to 80% B for 10 min. The LTQ-Orbitrap-ETD XL mass spectrometer was operated in the data-dependent mode to switch automatically between MS (scan 1 in the Orbitrap), CID-MS2 (scan 2 in the LTQ), and ETD-MS2 (scan 3 in the LTQ). Briefly, after a survey MS spectrum from m/z 300 to 2000, subsequent CID-MS2, and ETD-MS2 steps were performed on the same precursor ion with a ± 2.5 m/z isolation width (Yu *et al.*, *FEBS Lett.* 2007, 581: 5561-5565). CID-MS2 and ETD-MS2 spectra were repeated by targeting specific ions, in order to gain additional linkage information not obtained in the initial run. These targeted approaches, using the Orbitrap in scans 2 and 3 if needed, were repeated (e.g. targeting multiple charges of a precursor ion or the same disulfide-linked peptide but with different enzymatic cleavage sites or missed-cleavages) until the linkage information was elucidated. If necessary, the ions of interest obtained with ETD-MS2 were targeted for CID-MS3.

Disulfide assignment: The expected disulfide-linked tryptic or multi-enzyme digested peptide masses with different charges were first calculated, and then matched to the observed masses in the LC-MS chromatogram. The matched masses (with < 5 ppm mass accuracy for highly abundant ions and <20 ppm for low abundant ions) were further verified by analysis of the corresponding CID-MS2 and ETD-MS2 fragmentation spectra, as well as the CID-MS3 fragmentation spectra, as needed.

Digestion strategy

The digestion strategy for analyzing the disulfide linkages in a protein with a complicated disulfide structure (*e.g.*, rhASA) is, at least in part, based on the following considerations.

Generally, proteases which can cut proteins to peptide sizes containing only a single disulfide are desired for identification of a single disulfide linkage because there is usually only one possibility for connection. However, intertwined disulfides or a cysteine-rich region in a protein (*e.g.*, rhASA) may prevent enzyme digestion to the desired peptide size. Peptide sizes are preferred to be 1 to 5 kDa since recovery and electrospray ionization efficiency can be a problem for larger peptides while smaller peptides less than 1 kDa may not retain well on a typical reversed-phase column. In some cases, disulfide assignment will require further adjustment of peptide sizes to generate peptide lengths with sufficiently high-charge states that result in a low mass-to-charge ratio (*i.e.*, $m/z < 900$) for effective ETD fragmentation. Thus, selection of proper enzymes or multiple enzymes needs to be considered to achieve ideal peptide length. Also, for disulfide-linked peptides containing N-linked glycosylation, an additional PNGase F treatment step should be considered to reduce the complexity of the mass spectra. For peptides containing free cysteines, the digestion pH for the selected enzymes needs also to be optimized to maintain sufficient enzyme activity while avoiding scrambling.

In this study, after surveying several enzyme combinations (Lys-C, trypsin, Asp-N, pepsin, and PNGase F), several protocols were developed for the full cysteine status of rhASA. **Table 1** lists the various digestion protocols including the fragmentation methods for the specific assignments. A detailed description of these steps is included in the Example. **Table 1** also lists the identified unpaired cysteines and disulfide linkages of rhASA. The linkage sites, peptide sequences, digest enzymes, and MS/MS fragmentation methods are summarized in the table. In **Table 1**, “fgly” means formylglycine, in which the side chain of cysteine, $-\text{CH}_2\text{SH}$, is converted to $-\text{CHO}$.

Table 1. Unpaired cysteine and disulfide linkages in rhASA

#	Site	Status	Enzyme	Corresponding peptide sequence	I.D. (method)	Theoretical mass	Observed mass
1	C20	Free Cys	Pepsin	GCGYGHPSSTTPNL (19-31)	CID	667.2957 (2+)	667.2965 (2+)
2	C51	Cys as fgly	Pepsin	YVPVSLC(fgly)TPSRAAL (45-58)	CID	729.9012 (2+)	729.9020 (2+)
3	C51	Free Cys	Pepsin	YVPVSLCTPSRAAL (45-58)	CID	738.8976 (2+)	738.9059 (2+)
4	C276	Free Cys	Pepsin	RMSRGGCSGL (270-279)	CID	512.2448 (2+)	512.2453 (2+)
5A	C282-C396	1 disulfide linkage	Pepsin	LRCGKGTTYEGGVRE (280-294) FTQGS ~~AHSDTTADPACHASSL~~ (381-402)	CID/ETD	1271.9117 (3+)	1271.9153 (3+)
5B	C282-C396	1 disulfide linkage	Lys-C + Trypsin	CGK (282-284) AHFFTQGS ~~AHSDTTADPACHASS~~ SLTAHEPPLLYDL ~~SK~~ (378-415)	CID/ETD	863.8022 (5+)	863.8050 (5+)
6	C138-C154 C143-C150	2 disulfide linkages	Lys-C+ Trypsin+ Asp-N+ PNGase F	DQGP ~~CQ~~ (134-139) DLT ~~CFPP~~ ATPCDGGC (140-154)	CID/ETD	1069.8997 (2+)	1069.9006 (2+)
7	C470-C482 C471-C484 C475-C481	3 disulfide linkages	Pepsin	PALQICCHPGCTPRPACCHCPDP HA (465-489)	CID/ETD/ CIDMS3	874.3564 (3+)	874.3596 (3+)

Unpaired cysteines: Cys20, Cys51, Cys276

When Lys-C or Lys-C plus trypsin digestion was used to assign the unpaired cysteines, disulfide scrambling, which formed various disulfides mainly among the free cysteines, was observed using a standard digestion buffer at pH 8 (~ 40%), and to a lesser extent at pH 6.8 (~ 5%). Scrambled disulfides were not observed with pepsin digestion at pH 2. Although pepsin digestion could be non-specific, major cleavages were found after leucine residues (the cleavage at the C-terminal side), followed by aromatic amino acids, proline, and glutamic acid residues. The major pepsin fragment containing the unpaired Cys20 was identified. The peptide with the corresponding mass and charge is shown in **FIG. 2A**. The monoisotopic mass accurately matched the theoretical peptide mass with an unmodified free cysteine, m/z 667.2965 (observed) and m/z 667.2957 (theoretical) for the 2+ charge state. The site of the free cysteine was determined by CID-MS2 of the precursor ion, as shown in **FIG. 2B**.

The remaining unpaired cysteines were identified in a similar manner, for example, as shown in **FIGS. 3A-3B** (Cys51 converted to formylglycine), **FIGS. 4A-4B** (Cys51 as a free cysteine), and **FIGS. 5A-5B** (Cys276 as a free cysteine). **Table 1** (#1, #2, #3, and #4) summarizes the assignments for all the unpaired cysteines. At Cys51, it contains more than 70% of formylglycin form.

Single disulfide: Cys282-Cys396

For the peptide with a single disulfide (Cys282 with Cys396, #5A and #5B in **Table 1**), the linkage assignment was straightforward. Although alkali pH (*i.e.*, pH8) should not cause the disulfide-linked cysteines to scramble, the other free cysteines in the protein could potentially cross react with the disulfide at alkali pH. Indeed, a minute amount of cross-reacted disulfides was observed using a digestion buffer at pH 8. No cross-reacted disulfides could be observed when trypsin at pH 6.8 or pepsin (pH 2) was used for the digestion. The assignment of the disulfide-linked peptide is illustrated by the pepsin digestion protocol (**FIGS. 6A-6C**). In **FIG. 6A**, the observed accurate mass matched the theoretical peptide mass with one disulfide (a loss of 2 H from the backbone sequence). The corresponding CID-MS2 spectrum, b and y ions in **FIG. 6B**, verified the correct sequence. For the corresponding ETD-MS2 spectrum, the disulfide bond was

preferentially dissociated, which resulted in two dissociated peptides designated as P1 and P2 (**FIG. 6C**), that confirms that the two peptides are linked together.

Nested disulfides: Cys138-Cys154 and Cys143-Cys150

As shown in **FIG. 1**, the cysteines for the nested disulfides are located in Cys138-Cys154, and Cys143-Cys150. Since there are four cysteines, other potential linkages could be either as two separate disulfides (Cys138-Cys143 and Cys150-Cys154) as well as two crossed disulfides (Cys138-Cys150 and Cys143-Cys154) (**FIG. 7**). Furthermore, the complexity is increased by two N-linked glycosylation sites, one within, and the other next to the two disulfides (N underlined in **FIG. 1**). To reduce the complexity, the two N-linked glycans were removed with PNGase F, converting Asn (N) to Asp (D). This conversion provided a target for Asp-N digestion. Thus, in addition to Lys-C plus trypsin (to reduce the protein size), the addition of PNGase F and Asp-N enzymes effectively cut the disulfide-linked peptide to a suitable size for mass spectrometric analysis (see **FIG. 8**). These nested disulfide bonds form a ring, which significantly reduces CID fragmentation efficiency for the amino acids inside the ring, therefore complicating the assignment for disulfide linkages inside the ring. Although ETD is effective to break the disulfides, the peptide length obtained by trypsin or pepsin alone is too large for effective fragmentation ($m/z > 1000$). The digestion protocol required 4 enzymes to obtain the proper size for effective fragmentation by mass spectrometry (see **FIGS. 8-9**). The assignment of the disulfide-linked peptide based on the mass spectra is shown in **FIGS. 10A-10C**. In **FIG. 10A**, the observed accurate mass matched the theoretical peptide mass with two disulfides (a loss of 4 H from the backbone sequence). Since the ring structure formed by nested disulfides was broken by the additional Asp-N digestion, the disulfide linkages could be conclusively assigned as long as cleavages can be observed in the backbone between the CDGGC amino acid residues. As shown in **FIG. 10B**, the y1, y3, b11, and b12 fragments in the CID-MS2 spectrum provide strong evidence for the linkages Cys138 with Cys154, and Cys143 with Cys150. In addition to the CID-MS2 spectrum, the corresponding ETD-MS2 spectrum (**FIG. 10C**) confirms that the two linked peptides (P1 and P2) are connected.

It should be noted that although Asp-N should cleave aspartic acid in the protein backbone, the aspartic acid residue adjacent to a cysteine (a disulfide) inside the ring was not cleaved (see **FIG. 8**). For digestion at pH 8, a significantly scrambled disulfide was observed at a different LC retention time (**FIG. 11**), as the structure resembled to scramble 1 in **FIG. 7**. At pH 6.8, the scrambled disulfide was reduced to a trace amount and could not be observed at pH 2 with pepsin. Although the pepsin-digested disulfide could not be effectively fragmented by CID, the fragmentation did indirectly confirm the nested disulfide linkage (see **FIG. 12**). ETD was also tested to fragment the pepsin-digested disulfide but was not successful, due to minimal fragmentation and mainly charge-reduced species in the ETD spectrum. Although CID-MS3 and even MS4 have been attempted to fragment the charge-reduced species, the fragmentation efficiency was still poor for the peptide of this size. As described above, the use of an additional enzyme (*i.e.*, Asp-N) to obtain a proper size and configuration of the disulfides was critical.

Cystine knot: Cys470-Cys482, Cys471-Cys484, and Cys475-Cys481

The cystine knot could not be broken by all the enzymes or the combination of the enzymes. In addition, CID fragmentation could not produce backbone cleavages within the cystine knot. Thus, ETD was examined. For the amino acid sequence in this region, pepsin digestion was selected in order to obtain the proper peptide length with less acidic residues for effective fragmentation by ETD (*i.e.*, eliminated additional glutamic and aspartic acid residues as compared to the corresponding tryptic fragment). The corresponding mass and charge of the pepsin-digested peptide is shown in **FIG. 13A**. The monoisotopic mass matched the expected peptide mass with three disulfides (a loss of 6 H from the backbone sequence). Limited sequence information was obtained by CID-MS2 (**FIG. 13B**). Nevertheless, and significantly, ETD-MS2 dissociated the disulfides, which allowed cleavage of the peptide backbone, as shown in **FIGS. 13C-13D**. The fragmentation of this disulfide-linked peptide but for two different charge states is shown in **FIG. 13C** (m/z 656.30, 4+) and **FIG. 13D** (m/z 525.20, 5+). The fragmentation data from the two different charge states demonstrates consistency with respect to cleavage sites and verifies that the linkage assignments are correct. Since the

peptide was linked through three intertwined disulfides, a partial reduction of a particular disulfide (with mass shift by only one Dalton), the high resolution-accurate mass instrument (Orbitrap) with ETD provided even more convincing evidence for the disulfide bond assignments. As seen in ETD spectra of **FIGS. 13C-13D**, z_7 and c_{18} along with the internal cleavages from the dissociated disulfide confirm the connection between Cys471 and Cys484. In addition, one of the charge-reduced species (m/z 1312.6, $[M+4H]^{2+}$) in the ETD spectrum was further fragmented (CID-MS3 using the Orbitrap) as shown in **FIG. 13E**. The MS3 spectra contain additional disulfide and backbone cleavages, such as y_{17} and b_8 , confirming the connection between Cys470 and Cys482. The fragmentation pattern and assignments were also observed with the same CID-MS3 spectra generated in the LTQ ion trap (**FIG. 13F**), which makes this method applicable to makes the method applicable even with low resolution MS instruments. After assigning the two disulfide linkages, the non-dissociated (the third) disulfide was left with the only possible remaining connection, which was a linkage between Cys475 and Cys481. In summary, the combination of ETD-MS2 and CID-MS3 mass spectral analysis confirms the linkage sites as Cys470 with Cys482, Cys471 with Cys484, and Cys475 with Cys481. Thus, the combination of ETD-MS2 and CID-MS3 mass spectral analysis confirms the linkage sites as Cys470 with Cys482, Cys471 with Cys484, and Cys475 with Cys481. The theoretical and observed fragment ions are listed in **Table 2**.

Table 2. The theoretical and observed fragment ions in FIGS. 4C-4E

Assignment for fragment ions in FIG. 13C			
	Theoretical mass (Mono)	Observed m/z	
c3	299.2084	299.2082 (1+)	
c4	427.2669	427.2668(1+)	
c18	932.3952	932.3967 (2+)	
C19+S	1016.9106	1016.9135 (2+)	
c21	1100.9556	1100.9581 (2+)	
c23	1206.9955	1206.4959 (2+)	
c24	1275.5249	1275.5307 (2+)	
z6-S	590.2575	590.2575 (1+)	
z7	759.2885	759.2897 (1+)	
z20	1041.8639	1041.8681 (2+)	
z21	1098.4060	1098.4093 (2+)	
z22	1162.4352	1162.4383 (2+)	
z23	1218.9773	1218.9815 (2+)	
z24	1275.5249	1275.5307 (2+)	
Assignment for fragment ions in FIG. 13D			
	Theoretical mass (Mono)	Observed m/z	
c3	299.2084	299.2082(1+)	
c4	427.2669	427.2665(1+)	
c5	540.351	540.3509 (1+)	
c18	932.8991	932.3996(2+)*	
c19	1001.4285	1001.4224 (2+)	
C19+S	1017.4145	1017.4154(2+)	
c21	1100.9556	1100.9608 (2+)	
c23	1206.9954	1206.9976 (2+)	
c24	1275.5249	1275.5314 (2+)	
z7	758.2807	758.3070 (1+)	
z20	1041.8639	1042.3701 (2+)*	
z21	1098.406	1098.9122 (2+)*	
z22	1162.4352	1162.9413 (2+)*	
z23	1218.9772	1219.4836 (2+)*	
z24	1275.5249	1276.0336 (2+)*	
*observed the second isotope			
Assignment for fragment ions in FIG. 13E			
	Theoretical mass (Mono)	Observed m/z	
b8	864.38613	864.3884 (1+)	
b18	1847.76375	1847.7614 (1+)	
b19	1984.82266	1984.8352 (1+)	
b20	1044.41986	1044.4200 (2+)	
b22	1150.45971	1150.4601 (2+)	
b24	1267.51555	1267.5189 (2+)	
y6	639.25612	639.2559 (1+)	
y17	1759.69264	1759.6974 (1+)	

To summarize, in this study, in-depth LC-MS protocols have been developed to assign the status of all 15 cysteine residues in rhASA, including the disulfide linkages from the nested disulfide and cystine knot. An effective multi-enzyme digestion approach was particularly useful for determining intertwined disulfides in peptides with N-linked glycosylation. Although both cystine knot and nested disulfides are difficult to resolve, strategies with a combination of different enzymes and MS fragmentation methods successfully determined the assignments. The successful assignment of the disulfide linkages in the cystine knot demonstrated the power of the approach, which should be generally useful for other cystine knots. This methodology can be used to monitor the disulfide linkages of manufactured proteins, such as rhASA.

The references, patents and patent applications cited herein are incorporated by reference. Modifications and variations of these methods and products thereof will be obvious to those skilled in the art from the foregoing detailed description and are intended to be encompassed within the scope of the appended claims.

What is claimed:

1. A method of analyzing a recombinant human arylsulfatase A (rhASA) preparation, the method comprising:
 - subjecting an rhASA preparation to enzyme digestion;
 - subjecting the digested preparation to liquid chromatography-mass spectrometry (LC-MS);
 - determining if the preparation has one or more structures depicted in Table 1, thereby analyzing the preparation.
2. The method of claim 1, furthering comprising providing an rhASA preparation.
3. The method of claim 1, wherein the structure is selected from the group consisting of an unpaired cysteine residue, a disulfide linkage, and a cystine knot.
4. The method of claim 3, wherein the unpaired cysteine residue is C20, C51, or C276 in rhASA.
5. The method of claim 3, wherein the disulfide linkage is between C282 and C396, between C138 and C154, between C143 and C150, between C470 and C482, between C471 and C484, or between C475 and C481, in rhASA.
6. The method of claim 3, wherein the cystine knot comprises the disulfide linkages between C470 and C482, between C471 and C484, and between C475 and C481, in rhASA.
7. The method of claim 1, wherein the enzyme digestion comprises two or more enzymes.

8. The method of claim 1, wherein the enzyme is selected from the group consisting of pepsin, trypsin, Lys-C, and PNGase.

9. The method of claim 1, wherein the LC-MS comprises collision induced dissociation (CID) or electron transfer dissociation (ETD).

10. A method of processing an rhASA preparation, the method comprising:
providing an LC-MC determination of whether one or more structures depicted in Table 1 are present in an rhASA preparation, and

if the one or more structures are present in the rhASA preparation, processing the preparation, wherein processing comprises one or more of selecting, accepting, processing into drug product, shipping, formulating, labeling, packaging, or selling the preparation.

11. A method of analyzing a process of making an rhASA preparation, the method comprising:

providing an rhASA preparation;

analyzing the rhASA preparation using the method of claim 1; and

if the one or more structures are present in the rhASA preparation, maintaining the process based on the analysis.

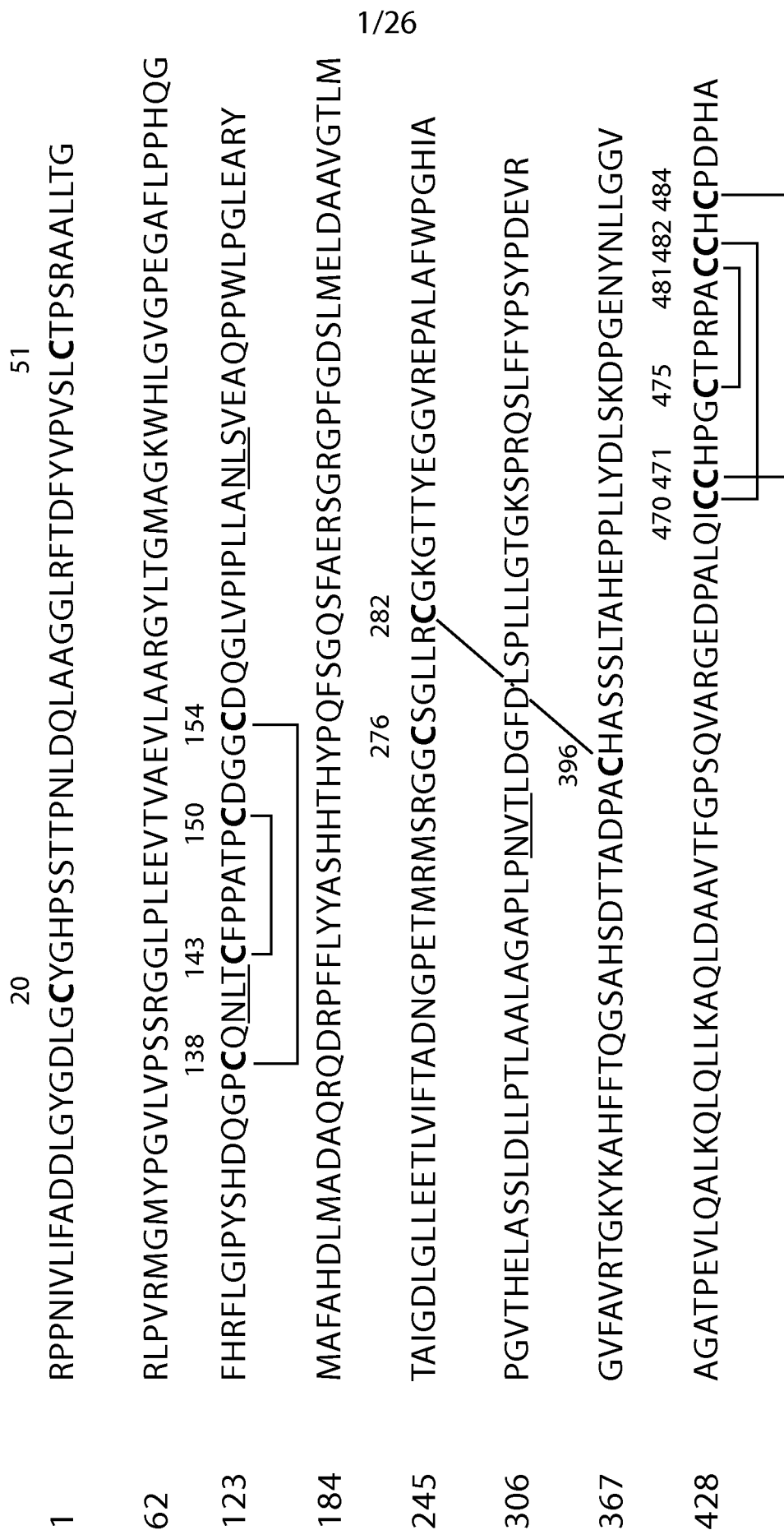


Fig. 1

2/26

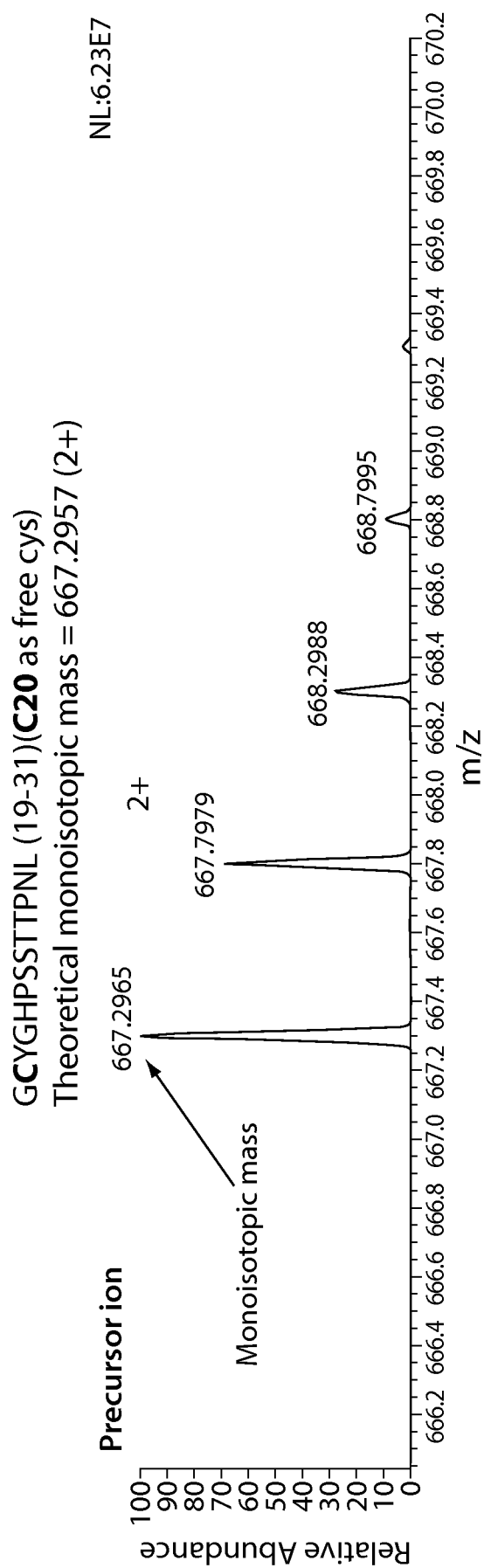


Fig. 2A

3/26

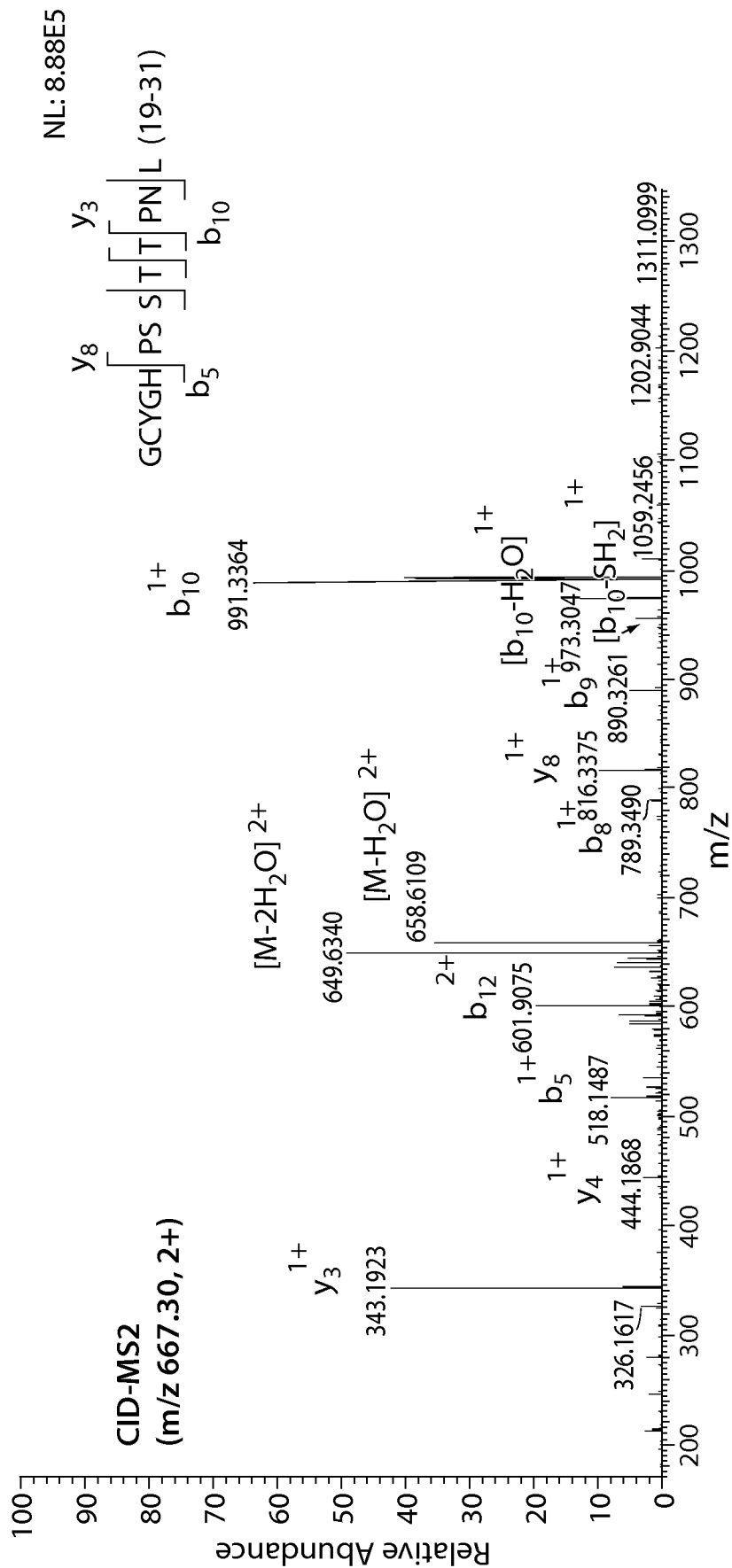


Fig. 2B

4/26

YVPVSLC(fgly)TPSRAAL (45-58)(**C51** as fgly)
Theoretical monoisotopic mass = 729.9012 (2+)

NL: 7.73E8

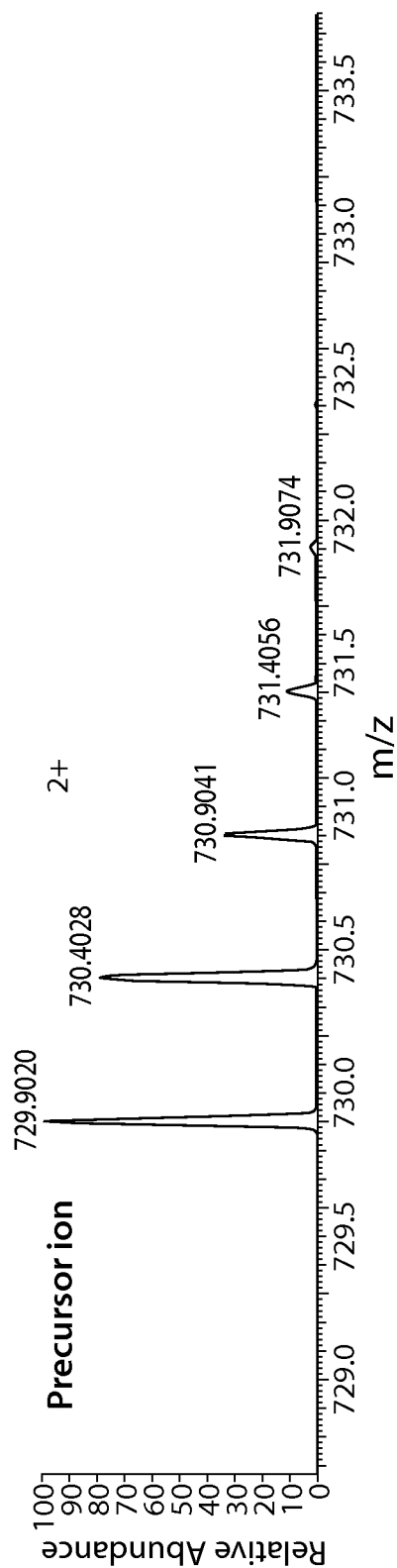


Fig. 3A

5/26

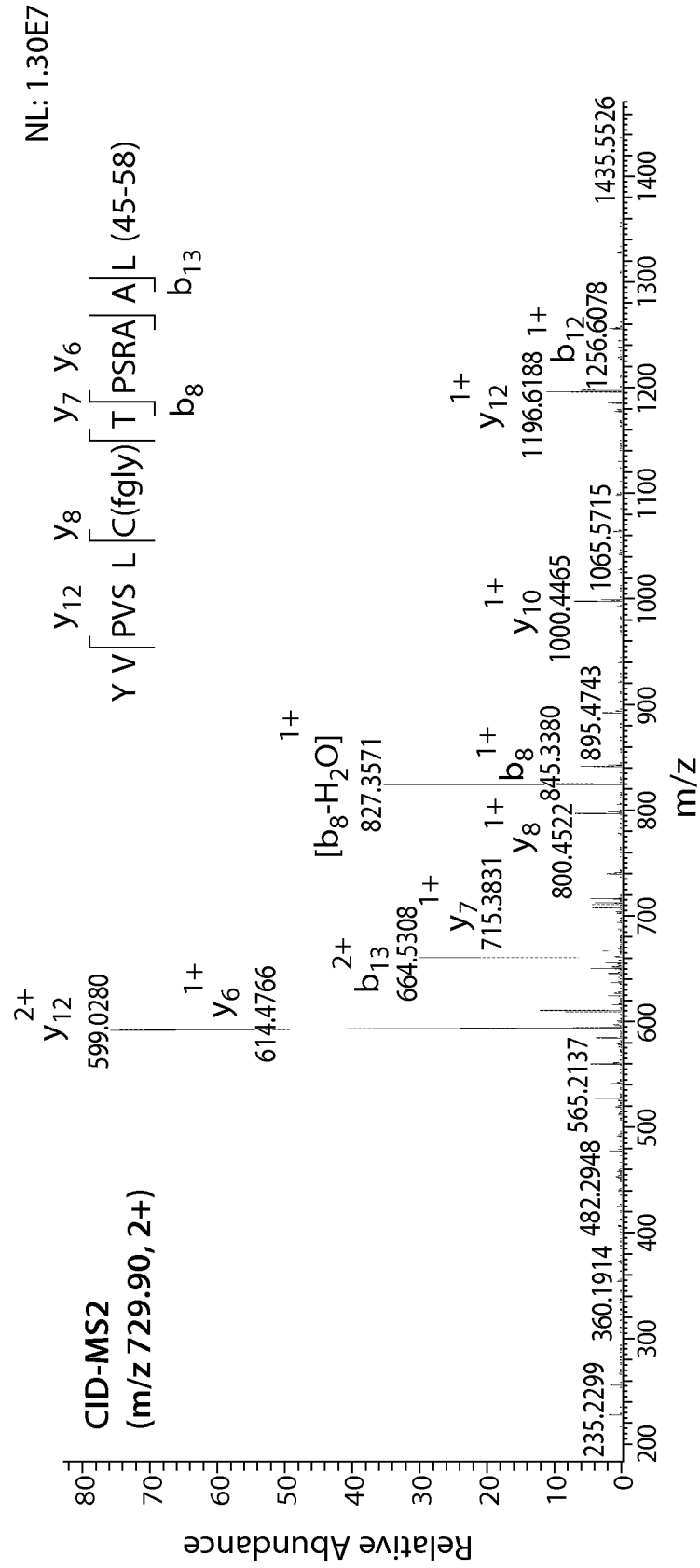


Fig. 3B

6/26

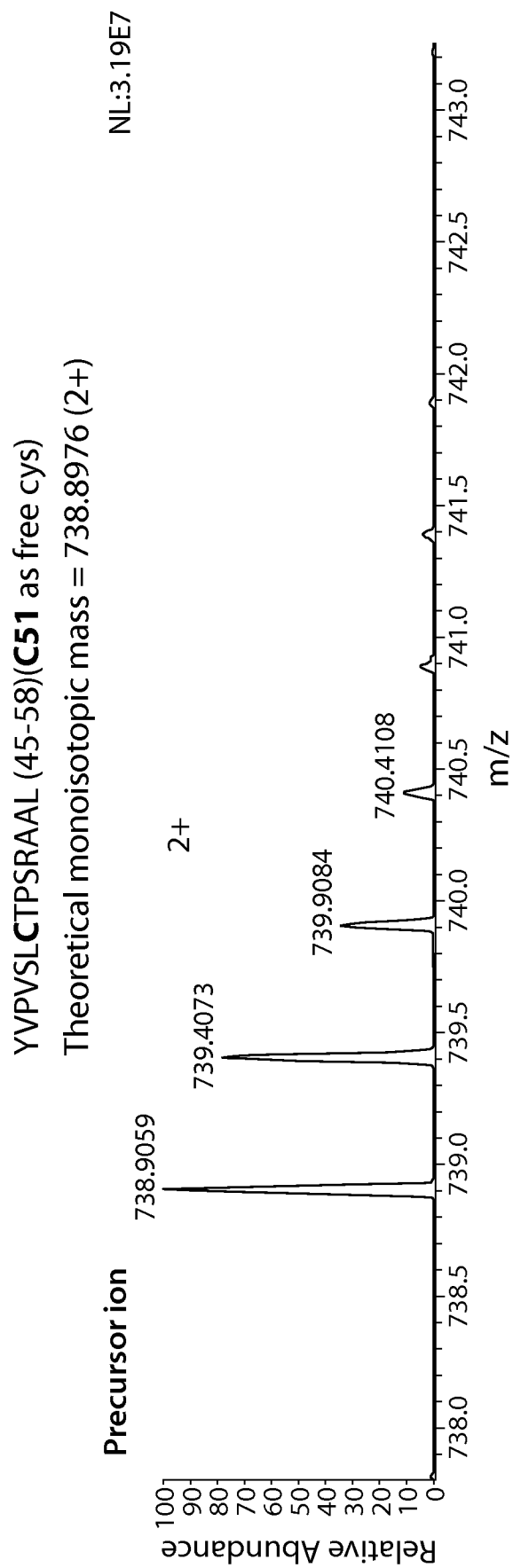


Fig. 4A

7/26

NL: 4.67E5

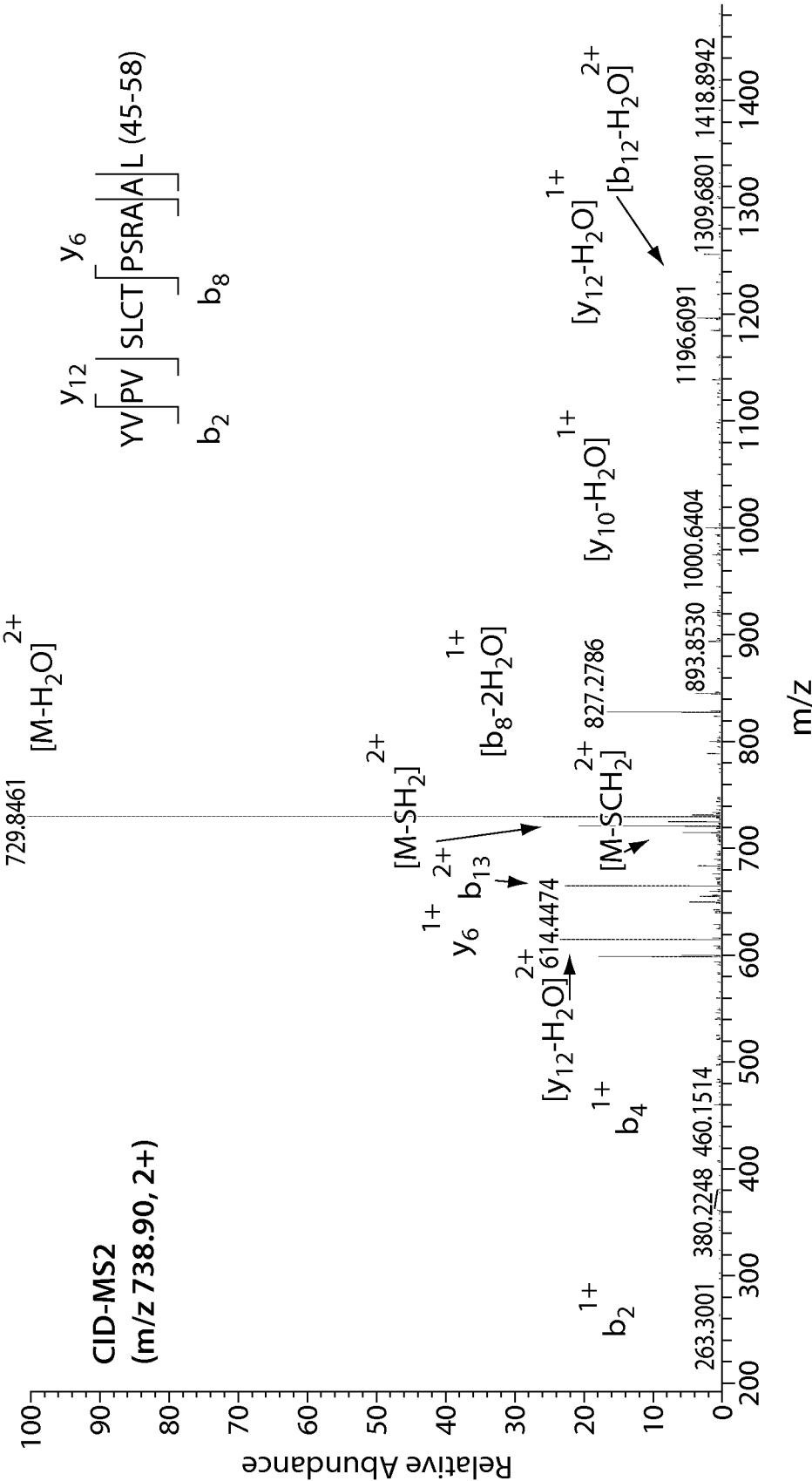


Fig. 4B

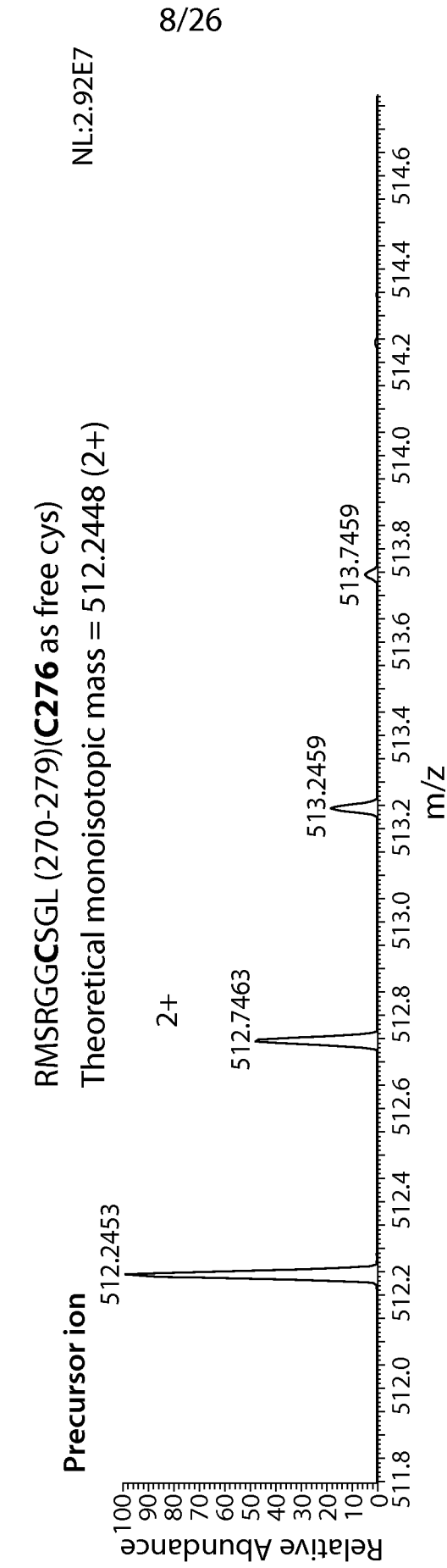


Fig. 5A

9/26

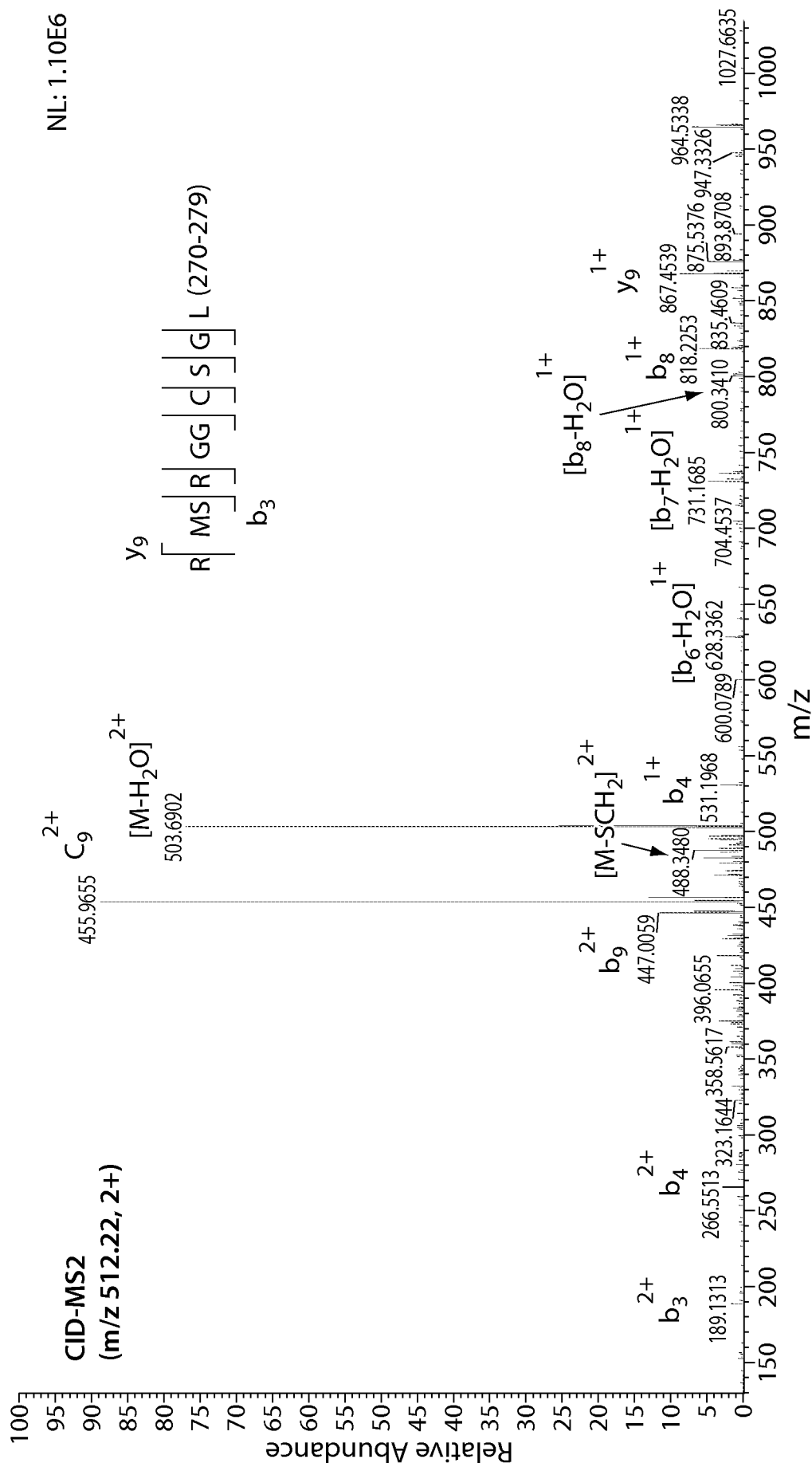


Fig. 5B

10/26

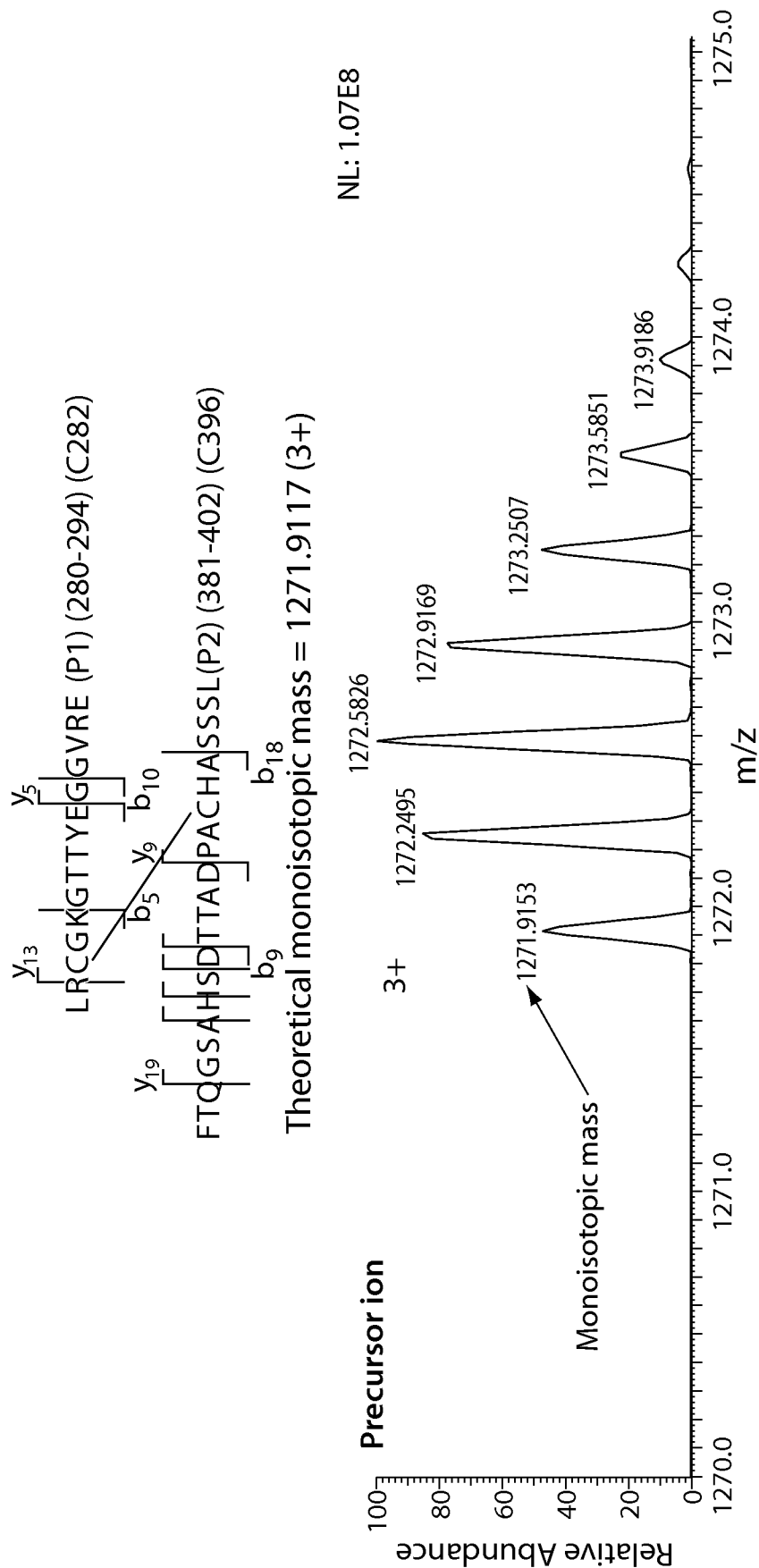


Fig. 6A

11/26

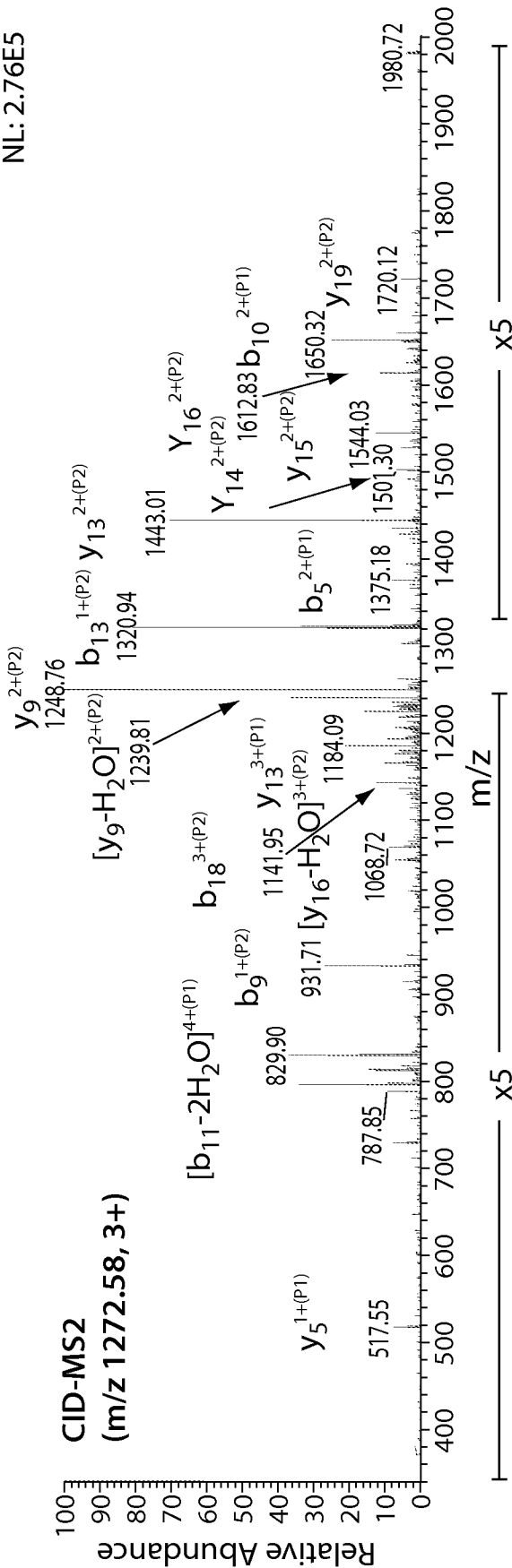


Fig. 6B

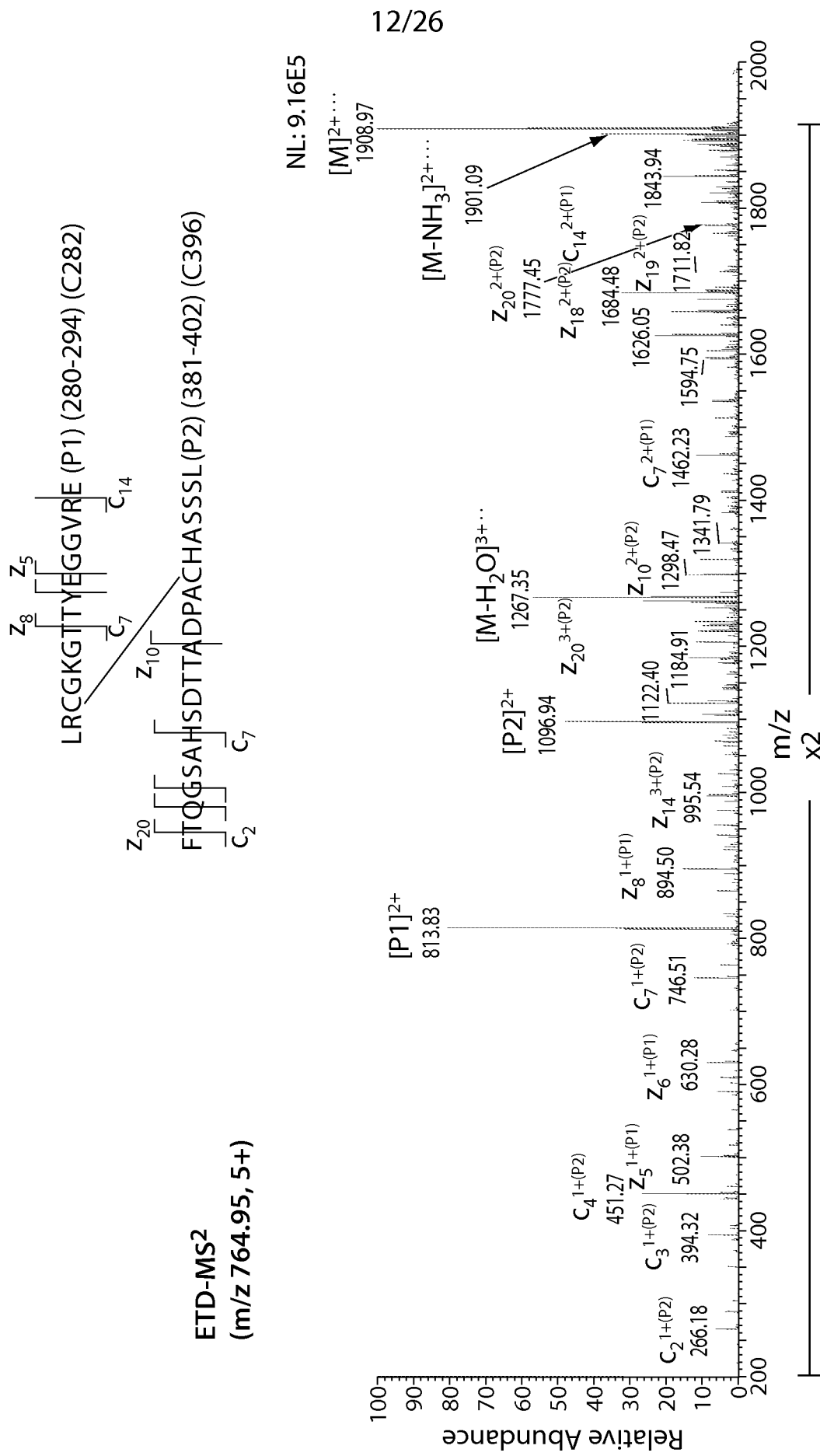
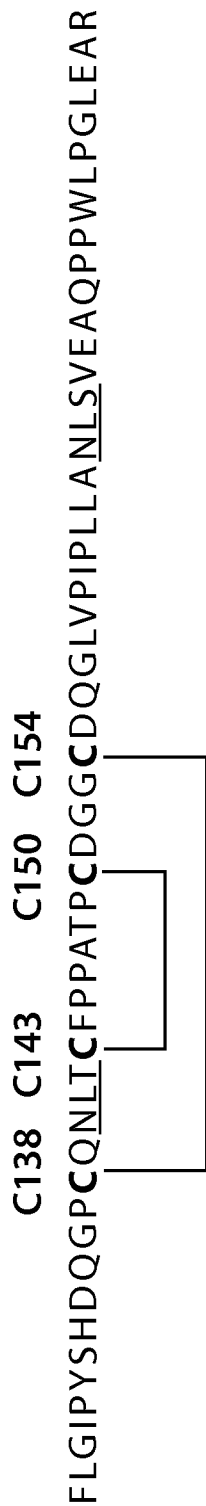


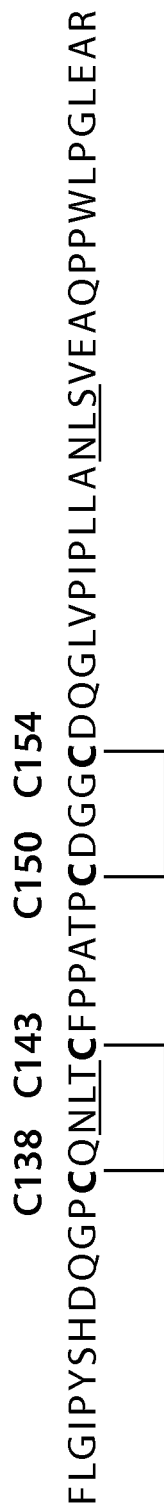
Fig. 6C

Nested Disulfides in a Tryptic Peptide

A (correct)



B (scramble 1)



C (scramble 2)

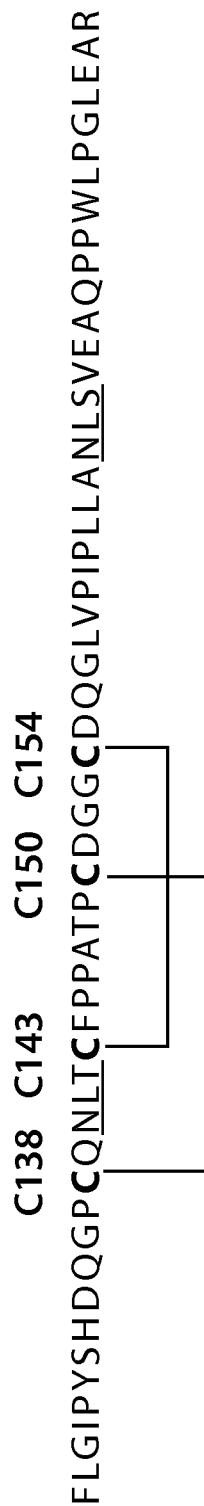
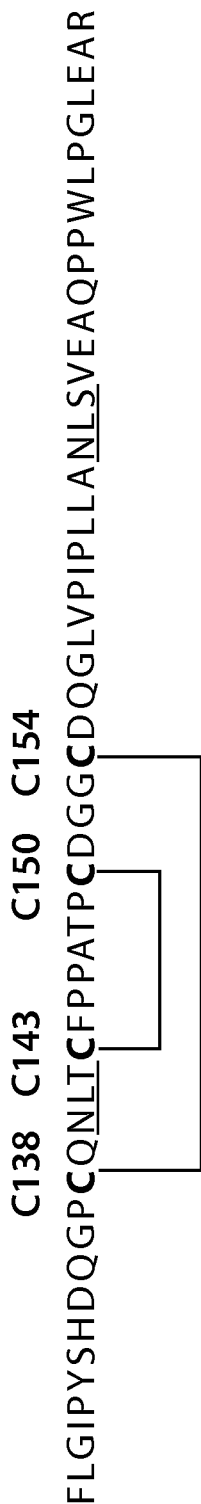


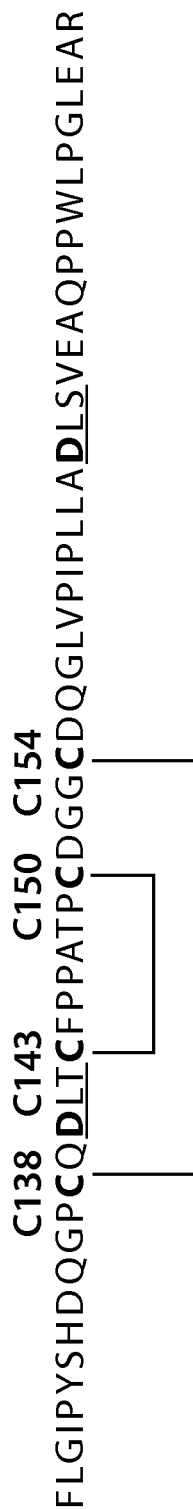
Fig. 7

Digestion Strategy for Nested Disulfides

A (Lys-C + trypsin)



B (Lys-C + trypsin + PNGase F)



C (Lys-C + trypsin + PNGase F + Asp-N)

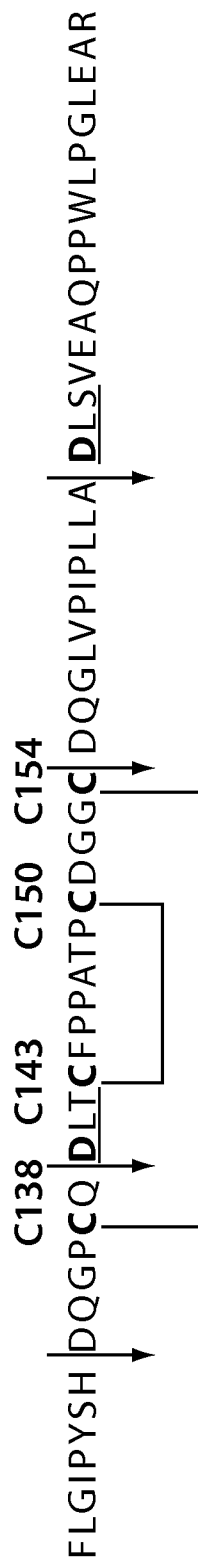


Fig. 8

Digestion Strategy for Nested Disulfides

Lys-C + trypsin + PNGase F + Asp-N

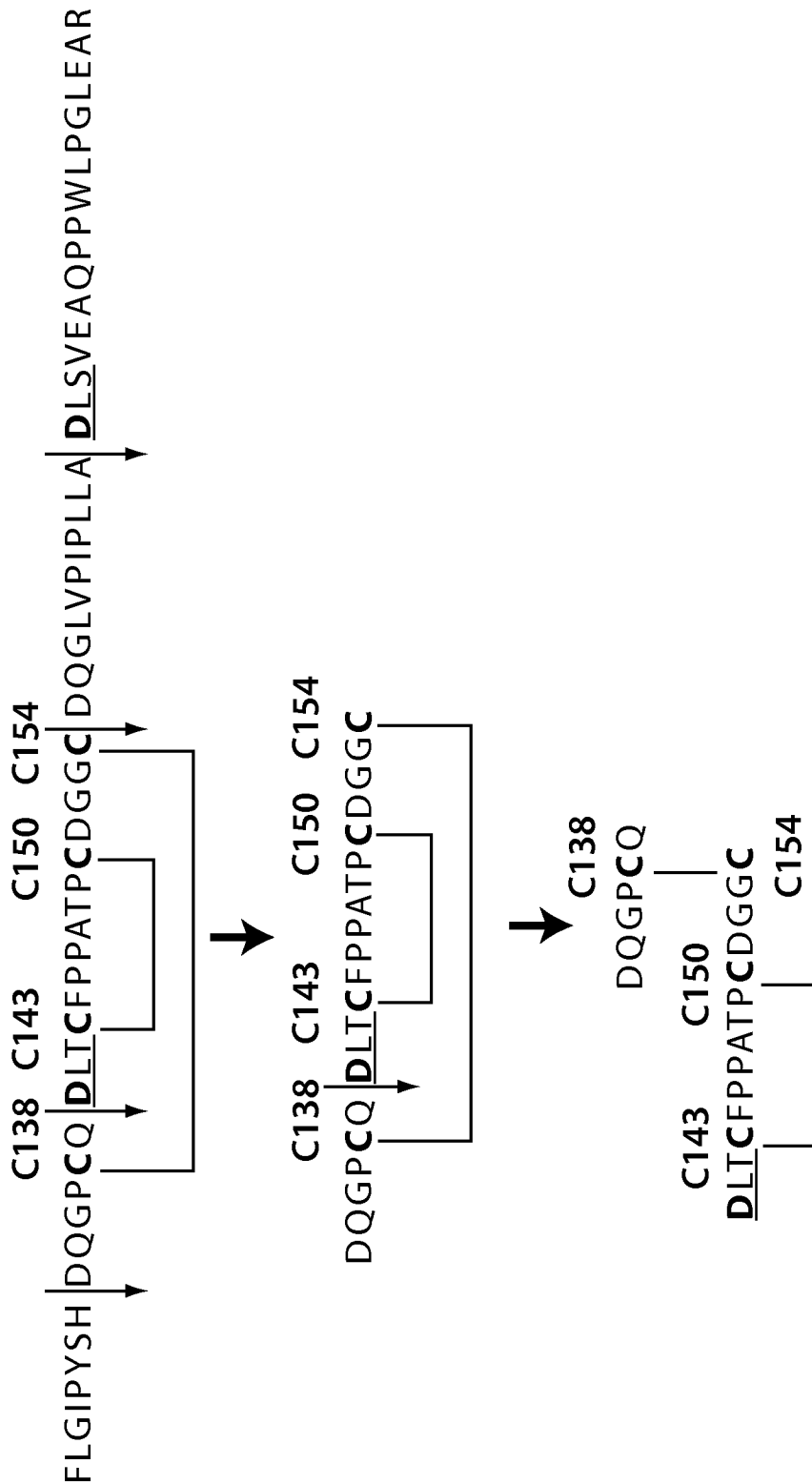
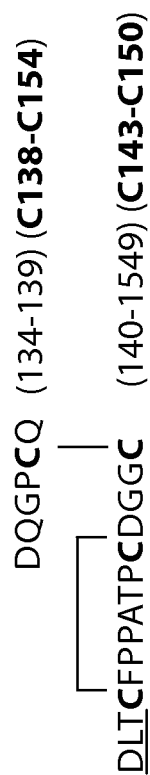


Fig. 9

16/26

NL: 1.71E6



Theoretical monoisotopic mass = 1069.8997 (2+)

Precursor ion

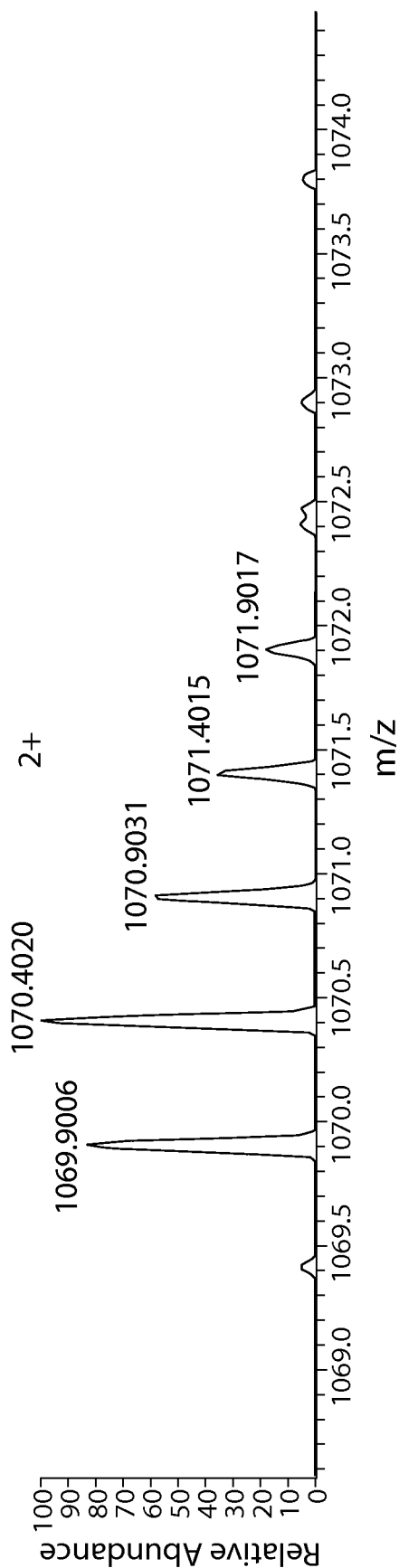


Fig. 10A

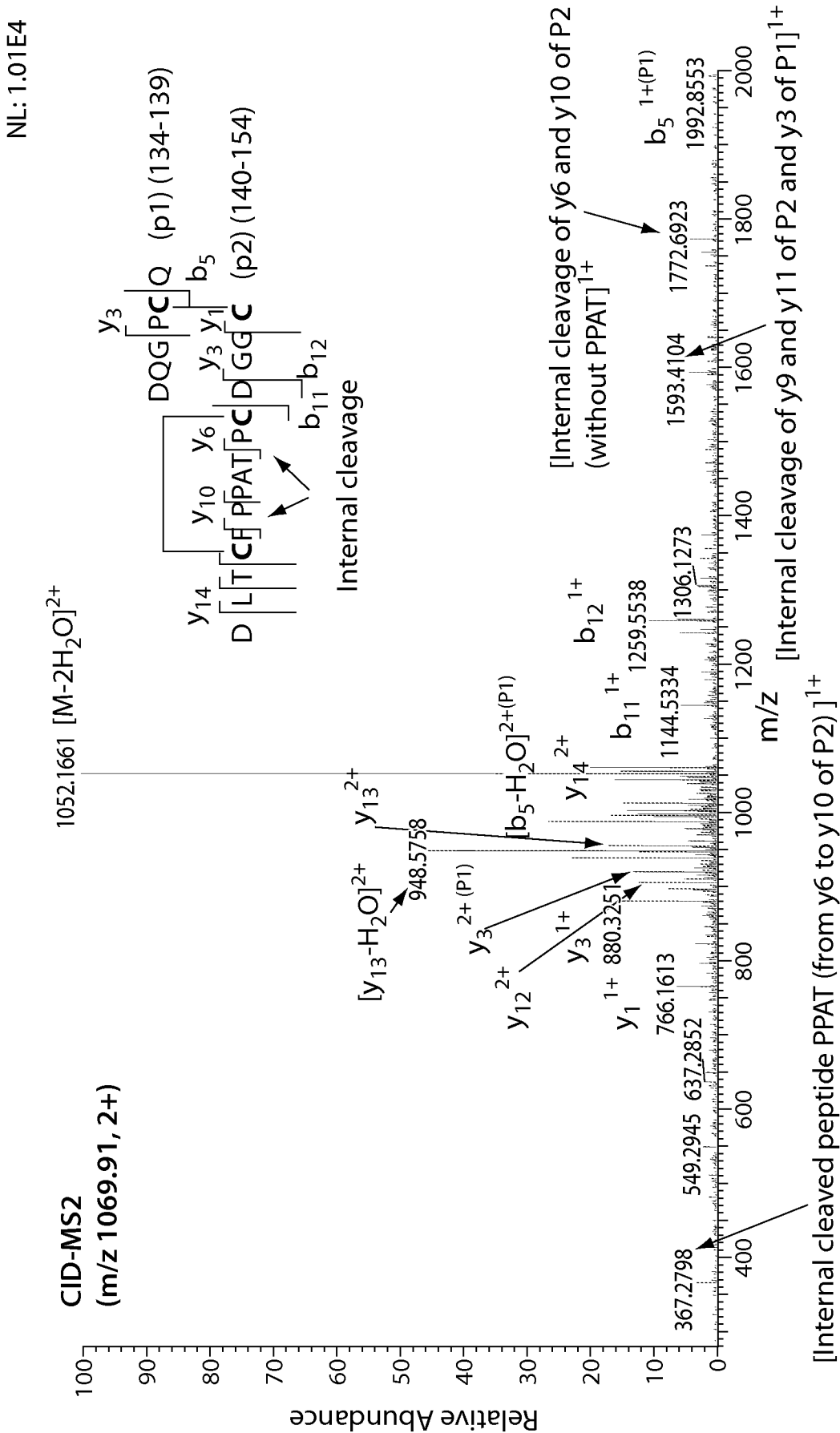


Fig. 10B

18/26

NL: 7.67E3

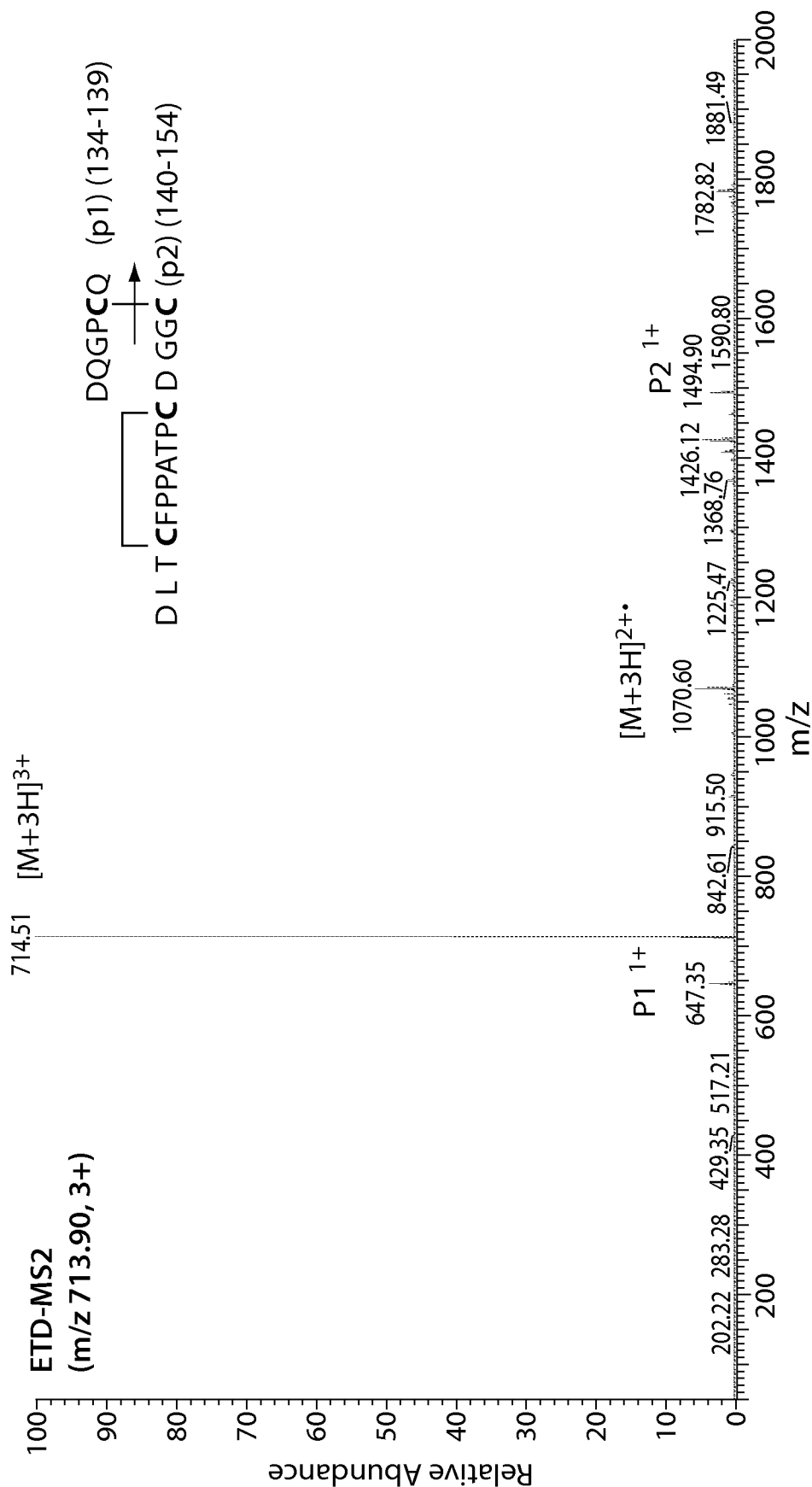


Fig. 10C

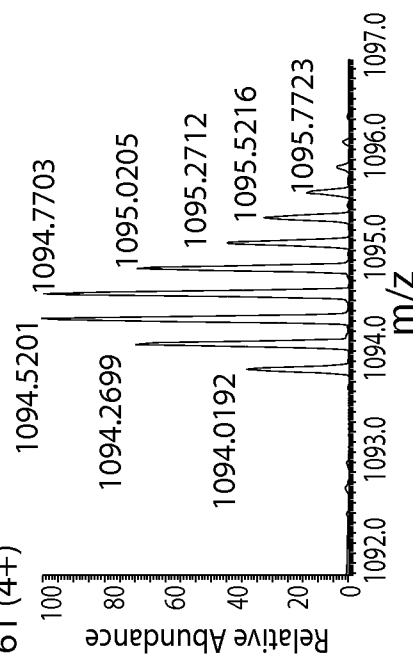
Nested Disulfide (pepsin digestion)

HRFLGIPYSHDQGPCQNLTICFPATPCDGGCDQGLVPIPLL (124-164) (**C138-C154, C143-C150**)

Precursor ion, 4+

Theoretical monoisotopic mass = 1094.0161 (4+)

Enzyme digestion: pepsin at pH 2 (since the N-linked site is not fully glycosylated, a non-glycosylated counterpart was selected for interpretation)



NL:2.17E5

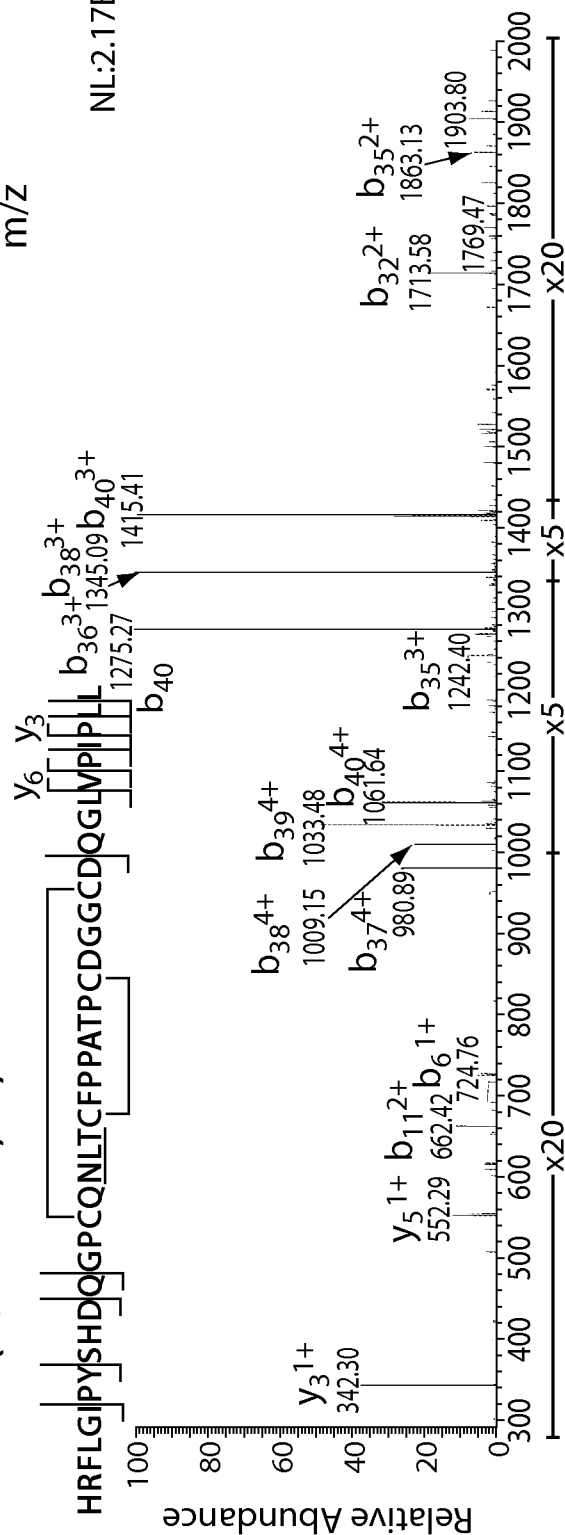


Fig. 12

21/26

Precursor ion



Theoretical monoisotopic mass = 874.3564 (3+)

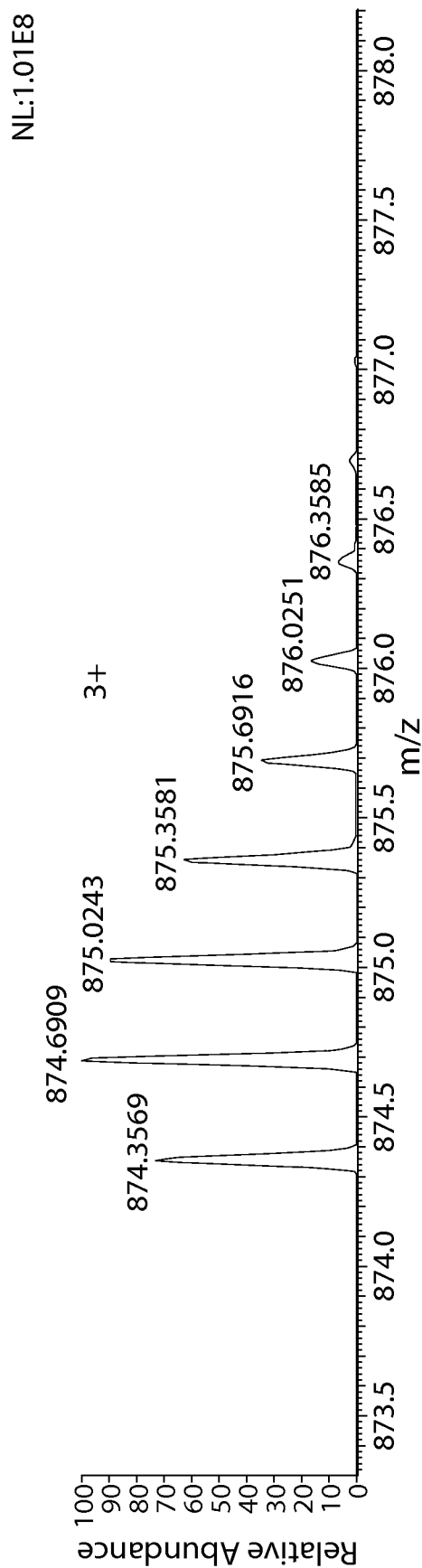


Fig. 13A

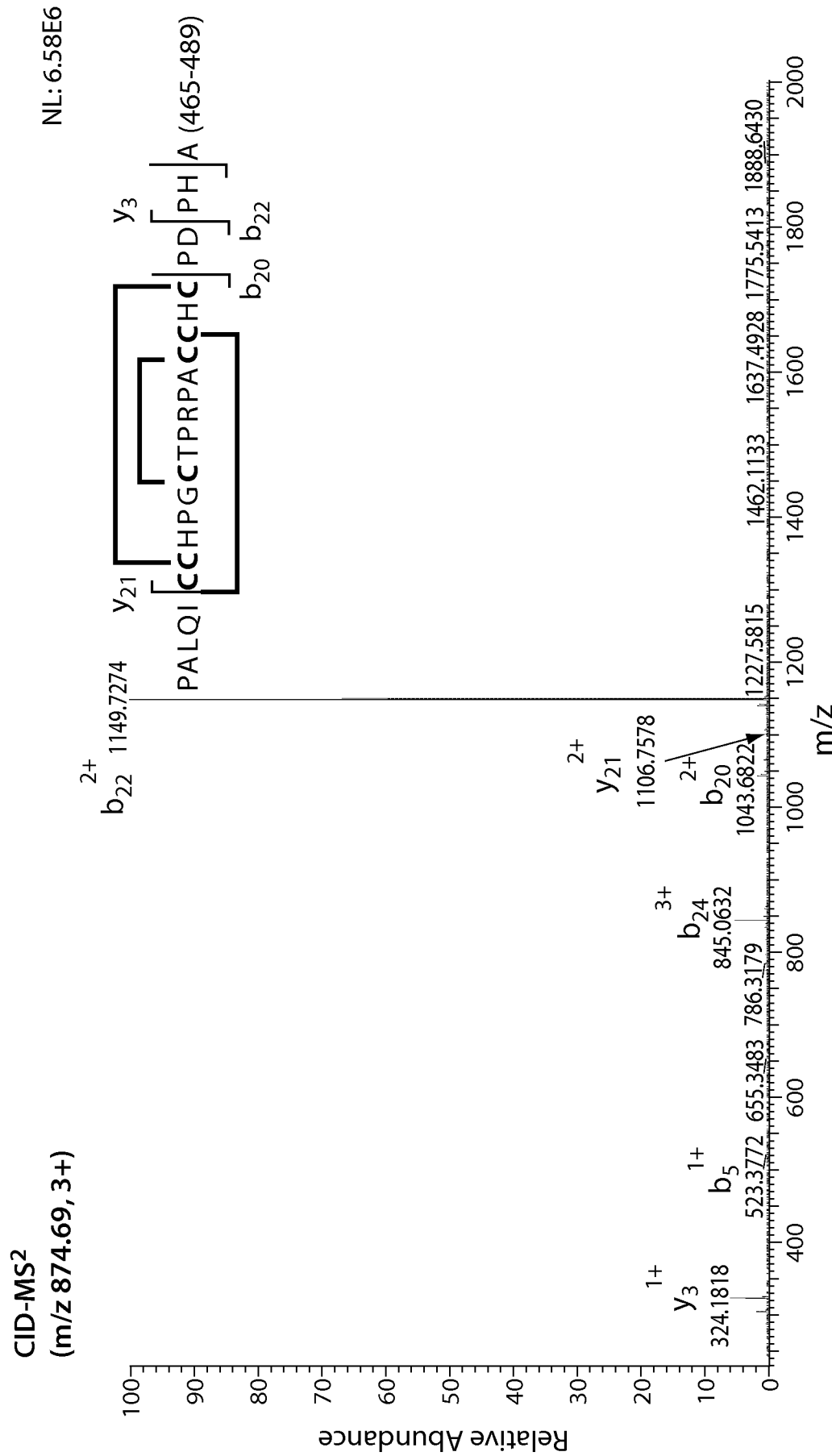
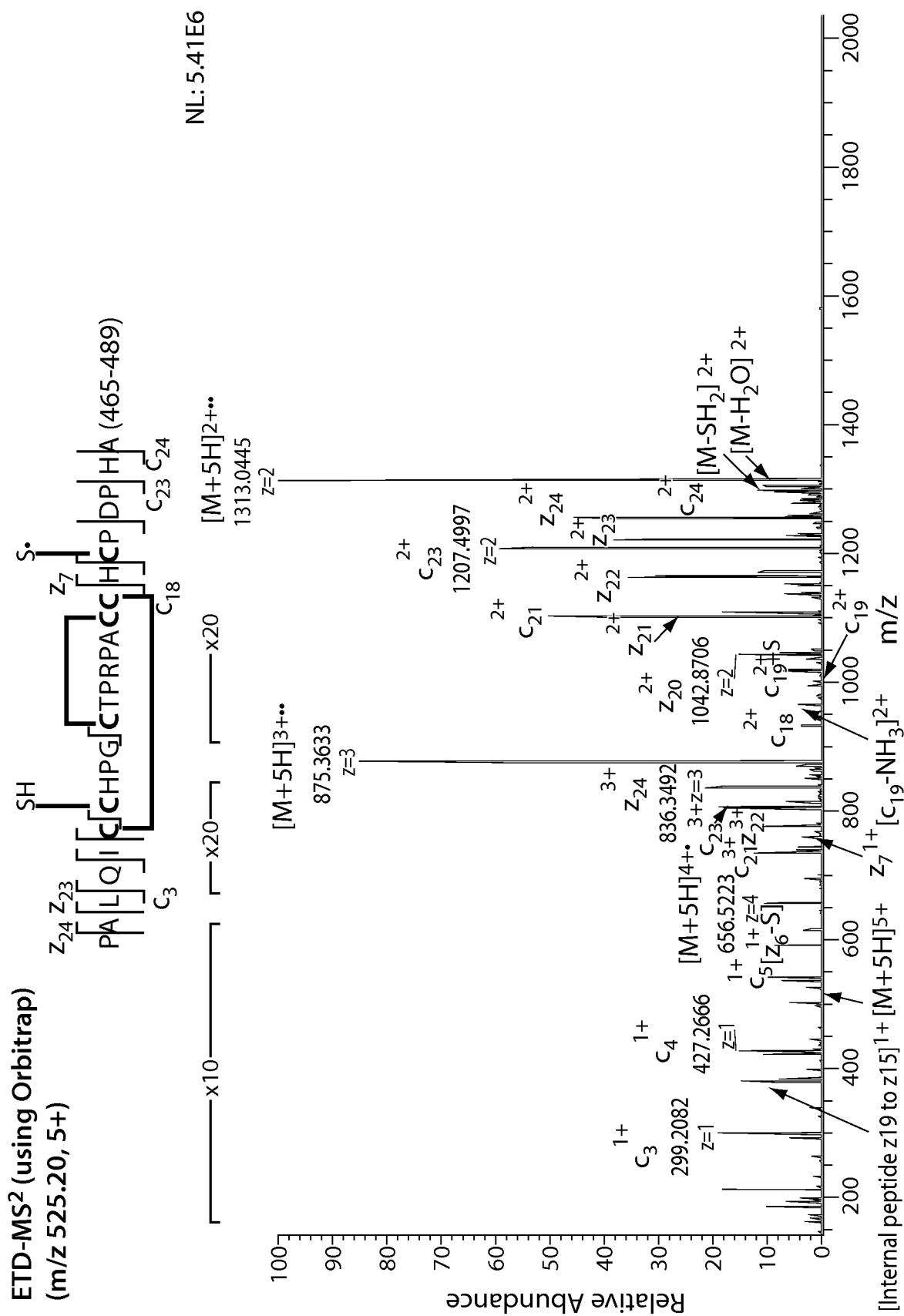
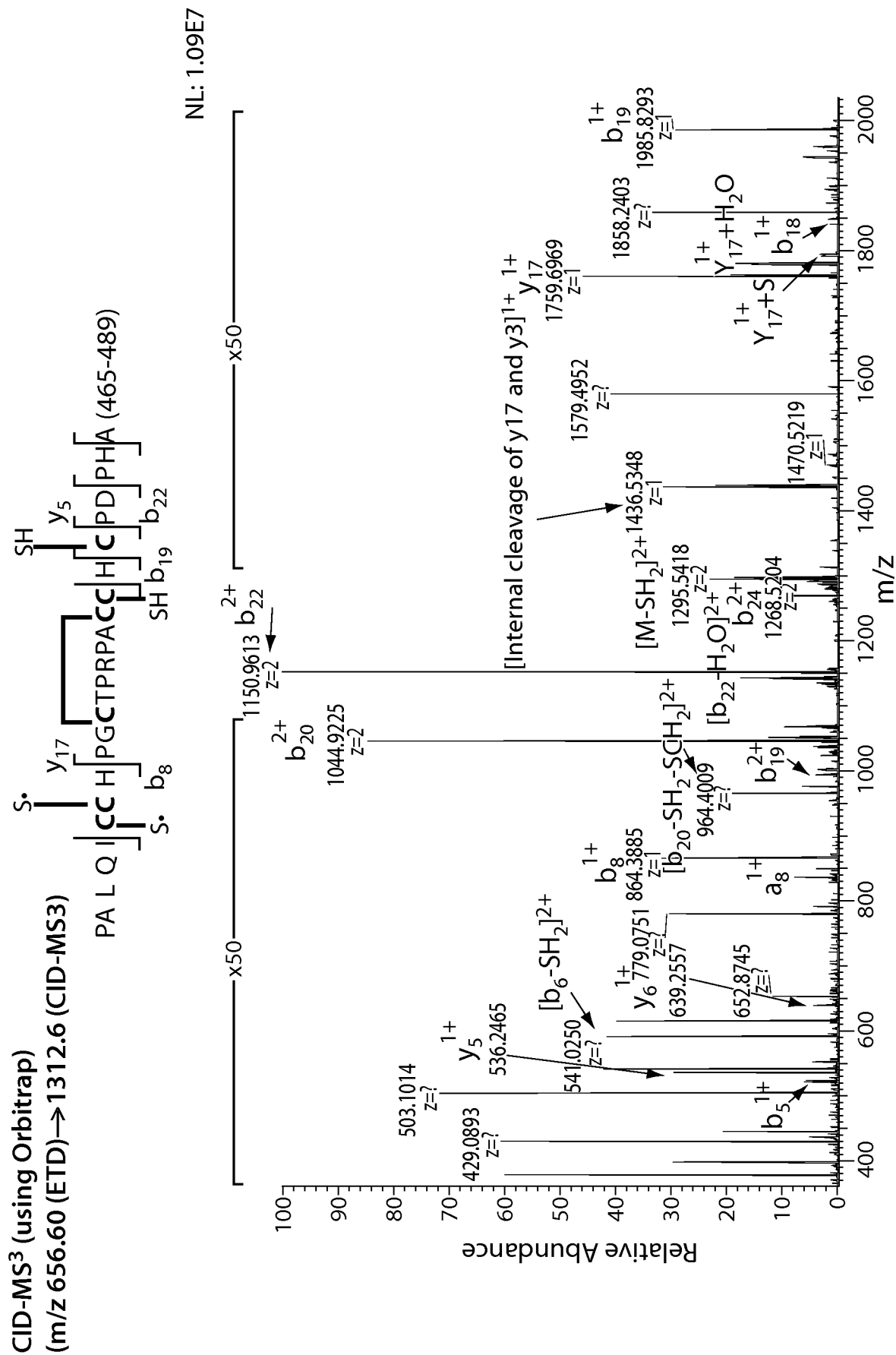


Fig. 13B





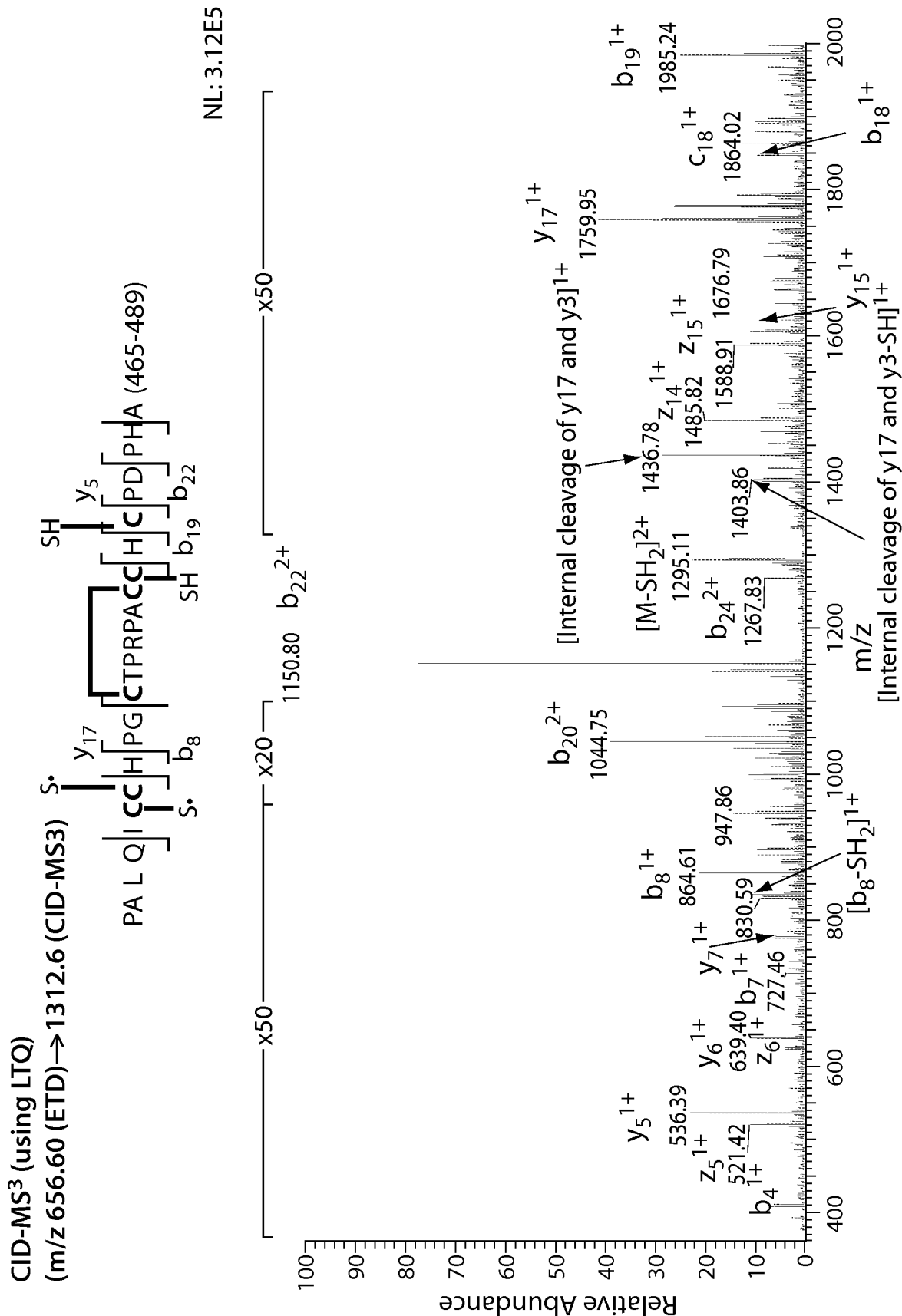


Fig. 13F

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2013/032171

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - A61K 38/46 (2013.01)

USPC - 424/94.6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - A61K 38/46; A61P 25/00; C12N 9/16; C12P 21/00 (2013.01)

USPC - 424/94.6; 435/18, 19, 184, 196

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

CPC - A61K 38/00; C12N 9/14; C12Y 301/06008; G01N 2400/40, 2500/04 (2013.01)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Patbase, Google Patents, Google, Pubmed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SCHRODER et al. "Site-specific analysis of N-linked oligosaccharides of recombinant lysosomal arylsulfatase A produced in different cell lines," Glycobiology, 28 October 2009 (28.10.2009), Vol. 20, No 2, Pgs. 248-259. entire document	1-11
A	WO 2005/113765 A2 (ZANKEL et al) 1 December 2005 (01.12.2005) entire document	1-11
A	WO 2002/098455 A2 (FOGH et al) 12 December 2002 (12.12.2002) entire document	1-11
A	WO 2004/072275 A2 (VON FIGURA et al) 26 August 2004 (26.08.2004) entire document	1-11
A	WO 2009/091994 A2 (VELLARD et al) 23 July 2009 (23.07.2009) entire document	1-11
P, X	NI et al. "Complete Mapping of a Cystine Knot and Nested Disulfides of Recombinant Human Arylsulfatase A by Multi-Enzyme Digestion and LC-MS Analysis Using CID and ETD," Journal of The American Society for Mass Spectrometry, 04 December 2012 (04.12.2012), Vol. 24, Pgs. 125-133. entire document	1-11

☐ Further documents are listed in the continuation of Box C.


* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

26 June 2013

Date of mailing of the international search report

08 JUL 2013

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774