Identification of novel high mannose *N*-glycan isomers undescribed by current multicellular eukaryotic biosynthetic pathways

Chia Yen Liew[1,2,3], Hong-Sheng Luo[1,4], Ting-Yi Yang[1,4], An-Ti Hung[1,5], Bryan John Abel Magoling[1,6,7], Charles Pin-Kuang Lai[1,7,8], Chi-Kung Ni*[1,3,5]

[1] Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei 10617, Taiwan

[2] International Graduate Program of Molecular Science and Technology, National Taiwan University (NTU-MST), Taipei 10617, Taiwan

[3] Molecular Science and Technology (MST), Taiwan International Graduate Program (TIGP), Academia Sinica, Taipei 10617, Taiwan

[4] Department of Chemistry, National Taiwan Normal University, Taipei 11677, Taiwan

[5] Department of Chemistry, National Tsing Hua University, Hsinchu, 30013, Taiwan

[6] Institute of Biochemical Sciences, College of Life Science, National Taiwan University, Taipei, 10617 Taiwan

[7] Chemical Biology and Molecular Biophysics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, 11529 Taiwan

[8] Genome and Systems Biology Degree Program, National Taiwan University and Academia Sinica, Taipei, 10617 Taiwan

*Corresponding author, e-mail addresses: ckni@po.iams.sinica.edu.tw

Abstract

N-linked glycosylation is one of the most important post-translational modifications of proteins. Current knowledge of multicellular eukaryote *N*-glycan biosynthesis suggests high mannose *N*-glycans are generated in the endoplasmic reticulum and Golgi apparatus through conserved biosynthetic pathways. As a part of post-translational modifications, lipid dolichol-phosphate linked oligosaccharide $Glc_3Man_9GlcNAc_2$ is transferred to proteins, and glucoses and mannose are sequentially removed by various ER- and Golgi-localized glucosidases and α-1,2-mannosidases. According to reported biosynthetic pathways, four $Man_7GlcNAc_2$ isomers, three $Man_6GlcNAc_2$ isomers, and one $Man_5GlcNAc_2$ isomer are generated during this process. In this study, we applied our latest mass spectrometry method, logically derived sequence tandem mass spectrometry $(LODES/MS^n)$ to re-examine high mannose *N*-glycans extracted from various multicellular eukaryotes. $LODES/MS^n$ identified many high mannose *N*-glycan isomers previously unreported in plantae, animalia, cancer cells, and fungi.

Importantly, their synthesis is yet described by known biosynthetic pathways, thereby suggesting additional and unidentified pathways for these *N*-glycans isomers in multicellular eukaryotic cells.

The asparagine linked (*N*-linked) glycosylation is one of the most important post-translational modifications of proteins in eukaryotes. *N*-linked glycans play important roles in the stabilization of protein structures and the regulations of protein functions[1,2]. The biosynthetic process of *N*-glycans has been studied in details for eukaryotes[3-9], and can be divided into three distinct stages. The first stage, which is the same for all eukaryotes, is the pre-assembly of lipid dolichol-phosphate linked oligosaccharide, $Glc_3Man_9GlcNAc_2$, followed by the transfer of the oligosaccharide moiety to proteins. The second stage in multicellular eukaryotes involves trimming of $Glc_3Man_9GlcNAc_2$ to $Man_5GlcNAc_2$ by various glycosidases located in endoplasmic reticulum (ER) and Golgi apparatus[3-9]. In the final stage, $Man_5GlcNAc_2$ is converted to hybrid and complex *N*-glycans.

The current knowledge of the second stage biosynthetic pathways of multicellular eukaryotes is illustrated in Fig. 1. The second stage begins with removal of the terminal $Glc\alpha1-2$ and $Glc\alpha1-3$ residues by α-1,2-glucosidase I and α-1,3-glucosidase II, respectively, to generate $Man_9GlcNAc_2$, followed by the generation of two isomers of $Man_8GlcNAc_2$, 8E1 and 8E2, by α-1,2-mannosidases in ER. The other enzyme, endo-α-mannosidase, acts on $Glc_1Man_9GlcNAc_2$ to form the third $Man_8GlcNAc_2$ isomer, 8G1. After the generation of $Man_8GlcNAc_2$, various Golgi-localized α-1,2-mannosidases act together to remove all the mannoses connecting by α-1$\rightarrow$2 glycosidic bond and convert $Man_8GlcNAc_2$ to a single isomer of $Man_5GlcNAc_2$ (5E1). According to these biosynthetic pathways, there are three possible isomers of $Man_8GlcNAc_2$ (8E1, 8E2, 8G1), four possible isomers of $Man_7GlcNAc_2$ (7D3, 7E1, 7E2, 7G1), three possible isomers of $Man_6GlcNAc_2$ (6D3, 6F1, 6F2), and only one $Man_5GlcNAc_2$ isomer (5E1). The relative abundances of these isomers vary with species.
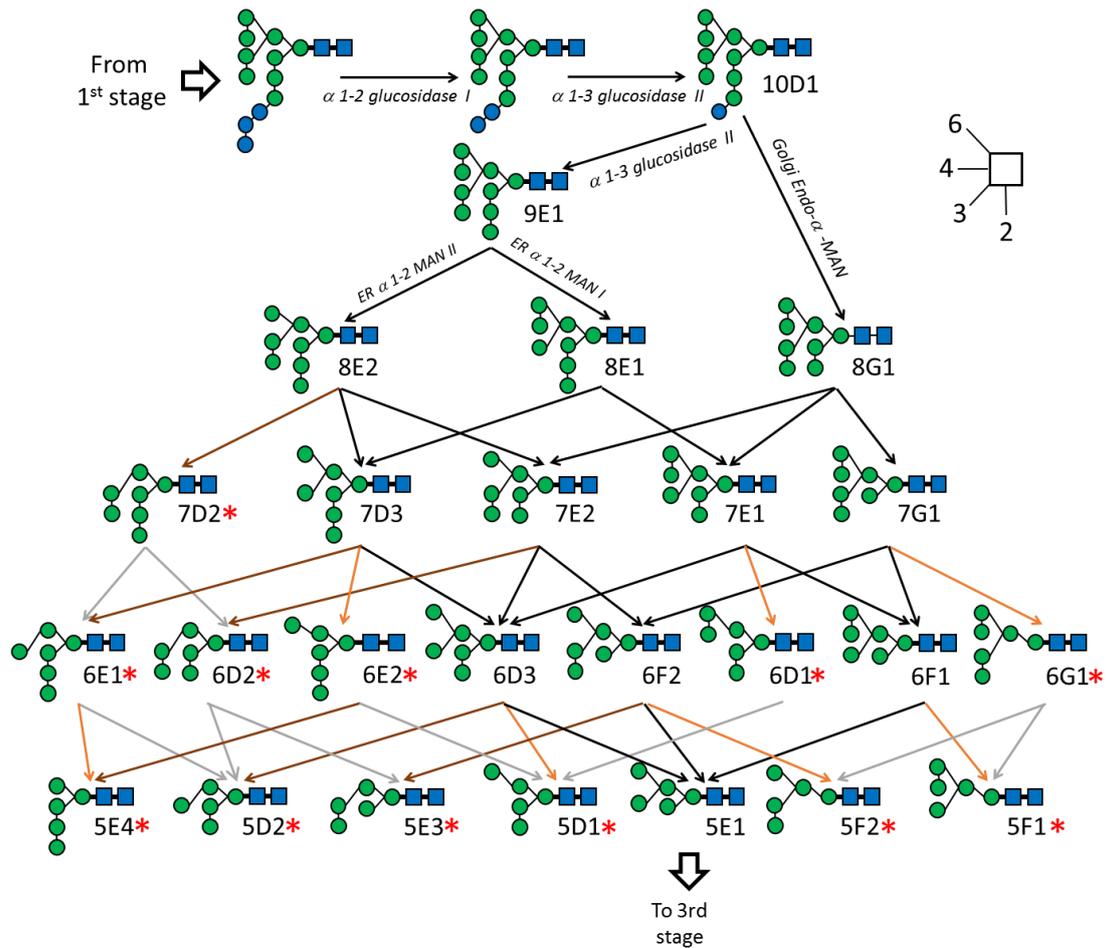
3

Fig. 1. Second stage biosynthetic pathways of *N*-glycans in multicellular eukaryotes. *N*-glycans not labelled by red stars are predicted by the current biosynthetic pathways. *N*-glycans labelled by red stars are not predicted by the current biosynthetic pathways, but they were found in this study. Black arrows represent the current biosynthetic pathways. Brown (α-1,6-mannosidases), orange (α-1,3-mannosidases), and gray (α-1,2-mannosidases) arrows represent the new biosynthetic pathways to generate the *N*-glycans labelled by red stars. The linkage positions are represented by orientation, as illustrated by the insert at upper right corner. Think and thin lines represent beta and alpha anomeric configuration of the glycosidic bond.

An important technique for structure determination of *N*-glycans is nuclear magnetic resonance (NMR) spectroscopy[10, 11]. However, NMR requires milligrams of samples. Mass spectrometry, on the other hand, has high sensitivity and capable of

detecting oligosaccharides of low abundance[12-15]. However, conventional mass spectrometry cannot distinguish stereo-isomers easily (for example, mannose and galactose, α- and β-anomers, etc.), additional techniques are needed for differentiation of precise structure of N-glycans. These additional methods include methods based on modification of the sample oligosaccharides (e.g., methylation), enzyme digestion[18,19], comparison to the existing mass spectrum database of oligosaccharides or N-glycan standards[20-26,] and the deduction from the currently understood N-glycan biosynthetic pathways[16,17]. These additional methods are confined by the current knowledge of N-glycan biosynthesis, the availability of enzymes, and incomplete database of oligosaccharides and N-glycan standards. For example, the current multicellular eukaryote biosynthetic pathway suggests there is only one isomer of $Man_5GlcNAc_2$ (5E1). Therefore, if ion of $Man_5GlcNAc_2$ is found in the mass spectra of multicellular eukaryote samples, it is typically assumed to be the isomer 5E1 without further structural identification[16,17].

We recently developed a mass spectrometry method, namely logically derived sequence (LODES) tandem mass spectrometry (MS^n), for oligosaccharide structural determination[27-34]. $LODES/MS^n$ does not depends on the mass spectrum database of oligosaccharides or N-glycan standards, thus $LODES/MS^n$ is particularly useful for the structural determination of the newly discovered N-glycans. In this study, we applied $LODES/MS^n$ to determine the structures of high mannose N-glycans of various species. Surprisingly, many N-glycan isomers newly found in plantae, animalia, cancer cells, and fungi cannot be explained by the current biosynthetic pathways. In some cases, the abundances of these unexpected N-glycans are larger than that of known N-glycans, indicating the existence of other biosynthetic pathways than what we now understood.

Results and Discussions

A. Databases of $Man_nGlcNAc_2$ (n=5, 6, 7) isomers

To facilitate the rapid *N*-glycan structural determination of various biological samples, we first constructed the *N*-glycan database which includes high performance liquid chromatography (HPLC) retention time and multistage collision-induced dissociation (MS^n CID) mass spectra. The sources of these *N*-glycans include (1) commercial products, (2) isolation and purification from biological samples, and (3) generation from large *N*-glycans by enzymic degradation. Various *N*-glycan databases have been reported previously[35-39]. The important differences between our new database and the existing databases are: (1) Our database consists of all possible high mannose *N*-glycan isomers by removing arbitrary numbers and positions of mannoses from $Man_9GlcNAc_2$ (9D1). These isomers are not limited to the *N*-glycans according to the current biosynthetic pathways; (2) We used intact *N*-glycans for structural analysis. No permethylation, reduction, or labeling at the reducing end of *N*-glycans is required. Therefore, the potential interference by the products generated by the side reactions during permethylation, reduction, or labeling is completely avoided.

In the construction of *N*-glycan database, the structures of commercial *N*-glycans were carefully examined using LODES/MS^n. The *N*-glycans isolated from biological samples were separated by two-dimensional HPLC chromatography and structurally determined by LODES/MS^n. The structures of parts of *N*-glycans isolated from biological samples were confirmed using enzyme digestion.

Here we used $Man_5GlcNAc_2$ *N*-glycans extracted from black beans as examples to demonstrate the processes of structural determination. After *N*-glycans were released from black bean glycoproteins by ammonia catalyzed reaction[40], they were

purified by using solid phase extraction and size exclusion. Next, the *N*-glycans were separated according to their sizes by HPLC using TSKgel Amide-80 column, the chromatogram is shown in Fig. 2(a). The separation of *N*-glycans by different sizes avoids the interference in structural determination of small *N*-glycans by large *N*-glycans due to the ESI in-source decay[41]. Fig. 2(a) shows that the retention time 19-23 min mainly consists of $Man_6GlcNAc_2$ (*m/z* 1419) and a small amount of $Man_5GlcNAc_2$ (*m/z* 1257). The $Man_5GlcNAc_2$ in this retention time region likely was produced by the ESI in-source decay of $Man_6GlcNAc_2$, thus this part of $Man_5GlcNAc_2$ were not analyzed in the structural determination of $Man_5GlcNAc_2$. The $Man_5GlcNAc_2$ collected from retention time t=15-18 min was sent into the second HPLC using PGC column for isomer separation. The chromatogram is illustrated in Fig. 2(b). Since we used intact *N*-glycans, there are two anomers for each isomer, corresponding to the α and β anomers of the GlcNAc at the reducing end of *N*-glycans. PGC column is known for the separation of anomers. To identify which two peaks belongs to the same isomer, the eluents from PGC column were collected every 30 seconds such that different peaks were collected in different tubes. These fractions were stored at room temperature for 6 hours before they were concentrated and re-injected into the same PGC column individually. If two peaks in Fig. 2(b) belong to the same isomer, the re-injection of the eluents into the same PGC column would show two peaks again in chromatogram, and the relative intensities and the retention times of these two peaks must remain the same as that in Fig. 2(b). This is because the α and β anomers of the same isomer change to each other through mutarotation, which typically takes about 30 min to 2 hours in aqueous solution at room temperature to reach equilibrium. The chromatograms of the reinjection are illustrated in Fig. 2(c). Four pairs of peaks, highlighted by blue, green, red, and orange bars, representing four isomers were identified from these chromatograms. There is

only one isomer each in tubes 17, 32, and 40-52, and these isomers were sent into MS separately for structural analysis. Tube 26 consists of two isomers. The eluents in tube 26 was sent into PGC column again for isomer separation.
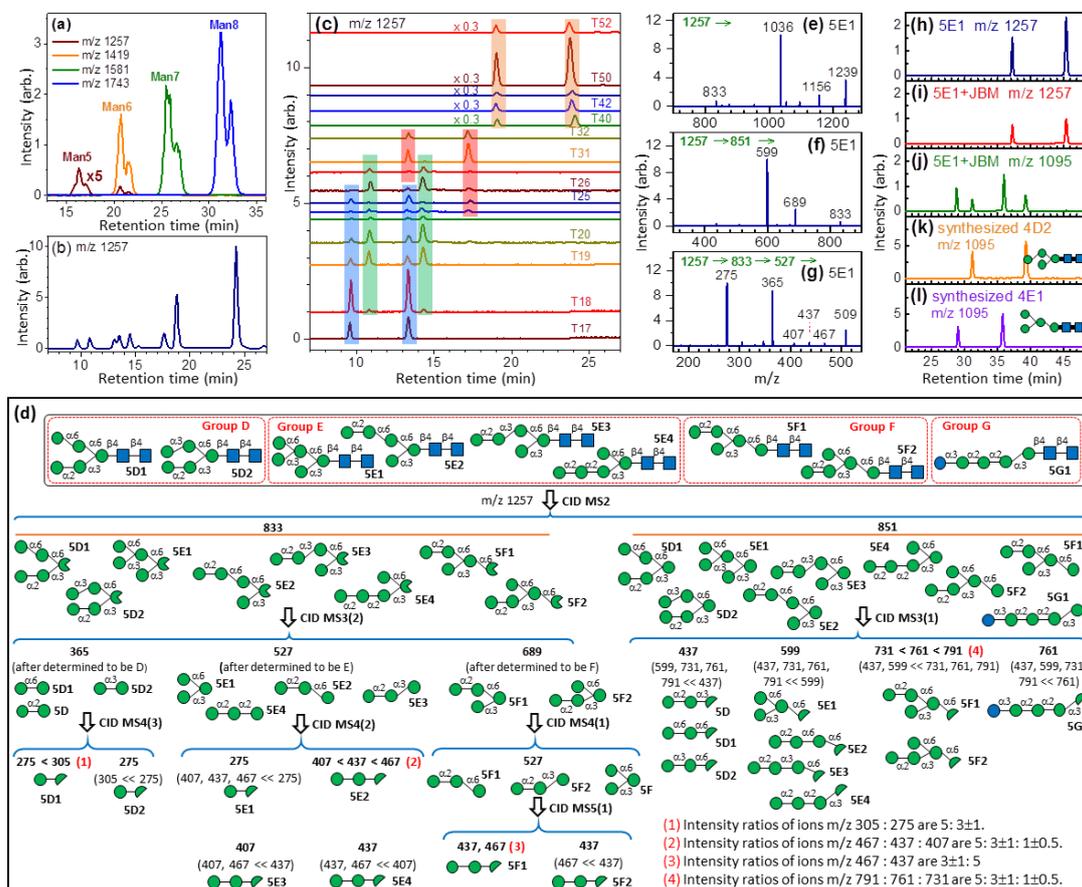

Fig. 2. (a) Chromatograms of the *N*-glycans extracted from black beans, separated by HPLC using TSKgel Amide-80 column (150 mm × 2.0 mm). The *m/z* values represent the sodium ion adducts of $Man_5GlcNAc_2$ (*m/z* 1257), $Man_6GlcNAc_2$ (*m/z* 1419), $Man_7GlcNAc_2$ (*m/z* 1581), $Man_8GlcNAc_2$ (*m/z* 1743). (b) Chromatograms of the $Man_5GlcNAc_2$ eluents collected from t=15 to 18 min in chromatogram (a), separated by HPLC using Hypercarb PGC column (2.1 mm × 100 mm). (c) Chromatograms of the eluents collected from chromatogram (b) every 30 s, separated by HPLC using Hypercarb PGC column (2.1 mm × 100 mm). (d) Logically derived sequence (LODES) for sodium ion adducts of $Man_5GlcNAc_2$ and $GlcMan_4GlcNAc_2$ isomers. Green circles and blue squares represent mannose and *N*-acetylglucosamine, respectively. Half circles and three quarter circles represent

cross-ring dissociation and dehydration, respectively. (e)-(g) CID spectra of sodium ion adduct of Man$_5$GlcNAc$_2$ isomer 5E1. The CID sequences are shown on the top of spectra. (h) Chromatogram of Man$_5$GlcNAc$_2$ isomer 5E1 before adding enzyme α-mannosidase of *Canavalia ensiformis*. (i) Chromatogram of Man$_5$GlcNAc$_2$ isomer 5E1 3 minutes after adding enzyme α-mannosidase of *Canavalia ensiformis*. (j) Chromatogram of Man$_4$GlcNAc$_2$ isomers 3 minutes after adding enzyme α-mannosidase of *Canavalia ensiformis* to Man$_5$GlcNAc$_2$ isomer 5E1. (k) Chromatogram of synthesized Man$_4$GlcNAc$_2$ isomer 4D2. (l) Chromatogram of synthesized Man$_4$GlcNAc$_2$ isomer 4E1.

The LODES for structural determination of Hex$_5$GlcNAc$_2$ isomers, including all possible Man$_5$GlcNAc$_2$ and GlcMan$_4$GlcNAc$_2$ isomers, is illustrated in Fig. 2(d). These isomers include the removal of arbitrary numbers and positions of glucose and mannose from canonical Glc$_3$Man$_9$GlcNAc$_2$, the *N*-glycan transferred from lipid dolichol-phosphate to proteins before any removal of glucose and mannose by enzyme in ER and Golgi. The CID sequences in LODES were derived from carbohydrate dissociation mechanisms[42-45]. LODES enables us to determine the *N*-glycan structures without the requirements of oligosaccharide mass spectrometry library or *N*-glycan standards. The CID spectra of the isomer in tube 50 are illustrated in Fig. 2(e)-2(g). The MS^2 CID spectrum in Fig 2(e) represents the dissociation of precursor ion, sodium ion adduct of Hex$_5$GlcNAc$_2$ (*m/z* 1257). Fragment ions *m/z* 1239 (dehydration product) and *m/z* 1156 (cross-ring dissociation product) suggest a HexNAc is located at the reducing end and it is connected to the other sugars by 1→4 or 1→6 linkage, according to the dissociation mechanism of HexNAc[45]. Fragment ion *m/z* 1036 represents the elimination of HexNAc; fragment ion *m/z* 833 represents the elimination of two HexNAc, indicating five hexoses of Hex$_5$GlcNAc$_2$ are connected without any HexNAc between them. This MS^2 spectrum is only useful for the differentiation of *N*-glycans from impurities, it is not useful for isomer structural

9

identification. The MS^3 spectrum [Fig. 2(f)] of CID sequence, 1257→851→fragments [MS^2→MS^3(1) →fragments in Fig. 2(d)], classifies isomers into four groups, namely D, E, F, and G, according to the dissociation mechanisms of hexose[27-31, 41-43]. Fragment ion m/z 599 in Fig. 2(f) suggests that there is a branch consisting of 1→3 and 1→6 linkages at the reducing end in the structure of the oligosaccharide consisting of five hexoses, and there is one hexose in one linkage (1→3 or 1→6 linkage) and three hexoses in the other linkage. Thus, the spectrum in Fig. 2(f) indicates the isomer in tube 50 belongs to group E. The MS^4 spectrum [Fig. 2(g)] of CID sequence, 1257→833→527→fragments [MS^2→MS^3(2)→MS^4(2) →fragments in Fig. 2(d)] differentiate the isomers in group E. Fragment ion m/z 275 in Fig. 2(g) indicates the three hexoses in one linkage (1→3 or 1→6 linkage) is a branched trisaccharide with 1→3 and 1→6 linkages. This structure is only consistent to the structure of isomer 5E1.

The structure of the isomer in tube 50 was crosschecked using enzyme digestion. The chromatogram of ion m/z 1257 (sodium ion adduct of $Hex_5GlcNAc_2$) before and 3 minutes after enzyme (α-mannosidase of *Canavalia ensiformis*) digestion are illustrated in Fig. 2(h) and 2(i), respectively. The decrease in intensity indicates the degradation of $Hex_5GlcNAc_2$ by enzyme digestion. The chromatogram of ion m/z 1095 (sodium ion adduct of $Hex_4GlcNAc_2$) produced from the degradation of $Hex_5GlcNAc_2$ by enzyme is illustrated in Fig. 2(j). The retention time of the four peaks in Fig. 2(j) agree with that of synthesized $Man_4GlcNAc_2$ isomers 4D2 and 4E1 (structures of 4D2 and 4E1 were crosschecked using LODES/MS^n). The CID MS^2, MS^3, and MS^4 spectra of the $Hex_4GlcNAc_2$ produced from the isomer in tube 50 by enzyme digestion are also the same as that of the synthesized $Man_4GlcNAc_2$ isomers 4D2 and 4E1, as presented in Supplementary Information.

In our database, some isomers were generated by the enzyme degradation of large synthesized N-glycans. There may be more than one isomer produced by enzyme degradation. Fortunately, we had the retention time and CID spectra of all the potential isomers produced by enzyme degradation except the one we wanted to generate by enzymatic degradation, therefore we were able to assign these isomers unambiguously. For example, both isomers 5F1 and 5F2 could be generated by the enzyme degradation of $Man_6GlcNAc_2$ isomers 6G1. Because we have the retention time of isomer 5F1, we assigned the enzyme generated isomer which the retention time is different from that of isomer 5F1 to isomer 5F2. Details of the extraction and purification of N-glycans from biological samples, structural determination of commercial N-glycans and N-glycans extracted from biological sample, and enzyme digestion process of N-glycans are described in Methods and Supplementary Information.

After structural determination of all collected $Man_5GlcNAc_2$ isomers, a database consisting of the chromatograms and CID spectra of these isomers was constructed for rapid isomeric identification [Fig. 3]. The chromatograms of $Man_7GlcNAc_2$ and $Man_6GlcNAc_2$ in the database are illustrated in Fig. 4, while the CID spectra in the $Man_7GlcNAc_2$ and $Man_6GlcNAc_2$ database are presented in Supplementary Information. Notably, these N-glycans were not reduced at the reducing end. They were in intact form, and there are two peaks in the chromatogram for each isomer.
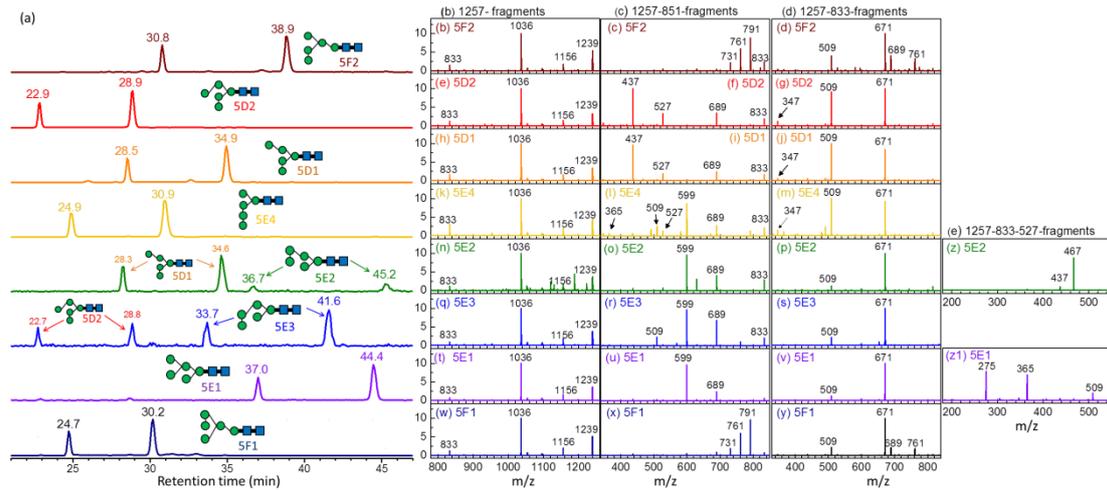
Figure 3. Database of Man$_5$GlcNAc$_2$ isomers. (a) HPLC chromatograms of ion *m/z* 1257 (sodium ion adduct), (b) CID spectra through the CID sequence 1257→fragments, (c) CID spectra through the CID sequence 1257→851→fragments, (d) CID spectra through the CID sequence 1257→833→fragments, (e) CID spectra through the CID sequence 1257→833→527→fragments. Isomers 5E1 and 5E2 have close retention time and similar MS^2 and MS^3 spectra. MS^4 spectra are useful for differentiation of these two isomers when the HPLC separation is not very good. These *N*-glycans were not reduced at the reducing end. They were in intact form, and there are two peaks in the chromatogram for each isomer. All Y axes represent intensity in arbitrary units.
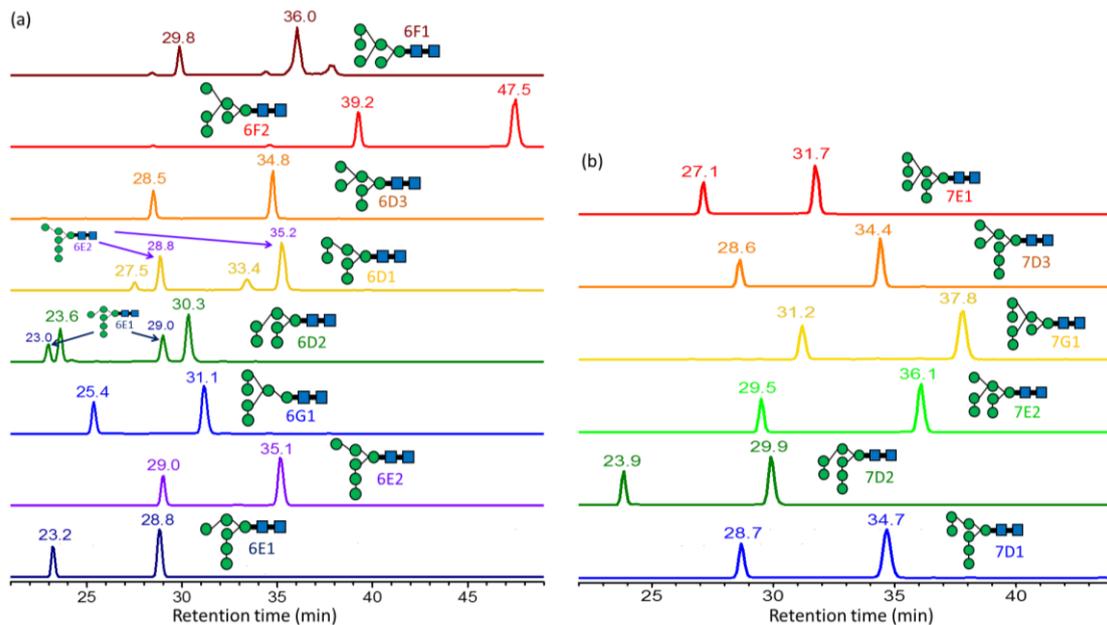

Figure 4. HPLC chromatograms of (a) Man$_6$GlcNAc$_2$ isomers (*m/z* 1419, sodium ion adduct), (b)

12

Man$_7$GlcNAc$_2$ isomers (*m/z* 1581, sodium ion adduct). They are part of the database. The other part of the database, CID spectra, are illustrated in Supplementary Information.

B. High mannose *N*-glycans of various biological samples

The structural assignments of *N*-glycans extracted from biological samples were made by comparing to the database on the similarities of the following three properties: (1) HPLC retention time in chromatogram of the selected *m/z* value, (2) one MS^2 and two MS^3 CID mass spectra at the corresponding retention time, (3) the relative intensity of two peaks (resulting from α and β anomeric configuration of the GlcNAc at the reducing end) in chromatogram of the selected isomer.

Fig. 5(a) and (b) shows the HPLC chromatograms and structural assignments of Man$_7$GlcNAc$_2$ isomers extracted from bovine lactoferrin. The isomers with large abundances belong to the *N*-glycan isomers, 7D3, 7E1, 7E2, and 7G1 predicted by biosynthesis. Surprisingly, isomer 7D2, one of the *N*-glycans not predicted by the current biosynthetic pathways (denoted as unusual *N*-glycans) was found, although the abundance is low. Here we used two methods, enzyme PNGase F [Fig. 5(a)] and ammonia catalyzed reaction [Fig. 5(b)], to release *N*-glycans from bovine lactoferrin. The unusual *N*-glycan isomer 7D2 were found using both methods, although the relative abundances of isomers were different between these two release methods.
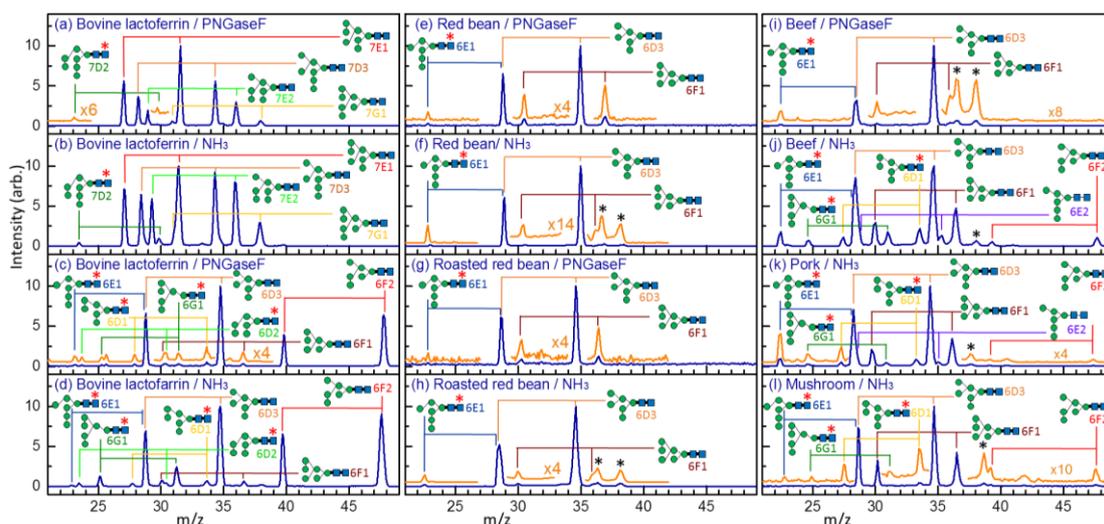
Fig. 5. Chromatograms and structural assignments of (a)-(b) $Man_7GlcNAc_2$ (c)-(l) $Man_6GlcNAc_2$ *N*-glycan isomers extracted from various samples using enzyme PNGase F or ammonia catalyzed reaction to release *N*-glycans from proteins. *N*-glycans labelled by red stars represent the *N*-glycans not predicted by current biosynthesis but found in this study. Peaks labelled by black stars represent impurities.

Fig. 5(c)-(l) show the HPLC chromatograms of $Man_6GlcNAc_2$ isomers extracted from bovine lactoferrin, red bean, beef, pork, and mushroom. Both *N*-glycan release methods were used for bovine lactoferrin, red bean, and beef. The relative abundances of isomers are slightly different between these two N-glycan release methods. Only ammonia catalyzed reaction was used in pork and mushroom. The unusual $Man_6GlcNAc_2$ isomers were found in all the samples we studied. None of these unusual *N*-glycans have the largest abundance, but they are not negligible.

The HPLC chromatograms of $Man_5GlcNAc_2$ extracted from lactoferrin, beef, red bean, human cell line M10, breast cancer cell line MDA-MB-231, pork, mushroom, and the structural assignments are illustrated in Fig. 6. The *N*-glycans not predicted by the current biosynthetic pathways were found in all the samples we studied. Isomer 5E1 which is the only isomer predicted by current biosynthesis has the largest

14

abundance in bovine lactoferrin, beef, human cell line, breast cancer cell line, pork, and mushroom, but the abundances of unusual *N*-glycans found in beef and pork are not small. Interestingly, the abundances of unusual *N*-glycans found in red bean are as large as that of isomer 5E1, some of them even larger than that of isomer 5E1.
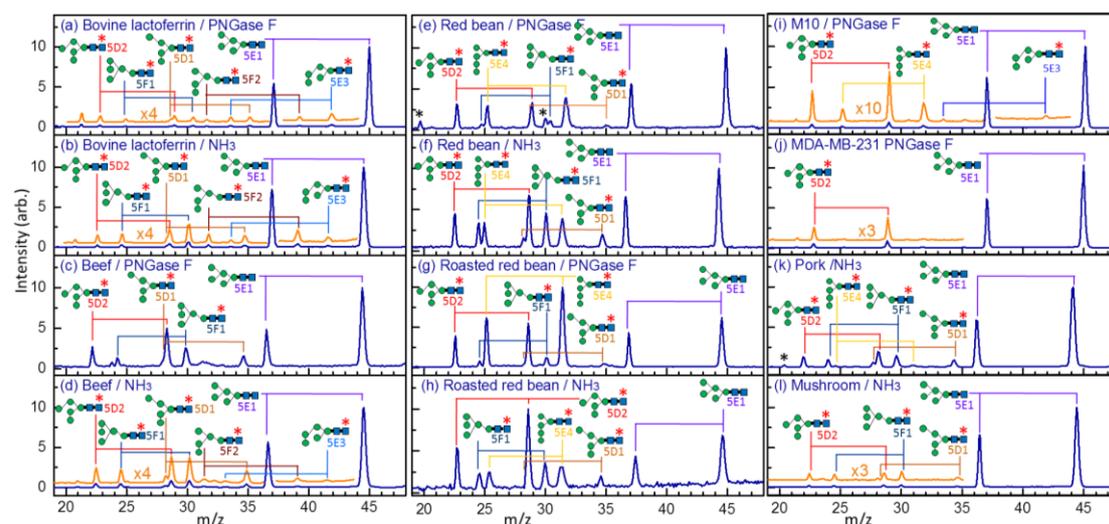

Fig. 6. Chromatograms and the structural assignments of $Man_5GlcNAc_2$ *N*-glycan isomers extracted from various biological samples. *N*-glycans labelled by red stars represent the *N*-glycans not predicted by current biosynthesis but found in this study.

The current biosynthesis suggests all the mannoses connecting by Manα-1→2 glycosidic bond were removed from $Man_9GlcNAc_2$ by α-1,2-mannosidases before the cleavage of Manα-1→3 or Manα-1→6 glycosidic bond by other enzymes. These unusual *N*-glycans indicate that there are enzymes removing Manα-1→3 and Manα-1→6 from $Man_nGlcNAc_2$ (n=8, 7, 6) before the complete removal of Manα-1→2. These processes are not included in the current biosynthesis. One possibility is that these enzymes are not located in ER and Golgi, and these unexpected *N*-glycans are produced during cell rupture in the *N*-glycan extraction process such that glycoproteins in ER or Golgi encounter the enzymes not located in ER and Golgi. To exclude this possibility, we compared the red bean powder samples

prepared by two different methods. In one method, red beans were grounded into powder by grinder directly. In the other method, red beans were baked in an oven at 90°C for 20 minutes and then immersed in liquid nitrogen for 20 minutes, followed by reiteration of the same process (baking in oven and immersing in liquid nitrogen) twice before grounding into powder. The process inactivates the potential enzymes that could contact glycoproteins during cell rupture. Both enzyme PNGase F and ammonia catalyzed reaction were used to release N-glycans from red bean powder prepared in both methods. In ammonia catalyzed reaction, red bean powder was added into 25% ammonia solution. The high pH value (>11) of ammonia solution ensures the potential enzymes which is active in low pH value, such as the enzymes in lysosome, do not involve in the generation of the unusual N-glycans. Fig. 5(e)-(h) and Fig. 6(e)-(h) show the chromatograms of the N-glycans released by PNGase F and ammonia-catalyzed reaction from red beans and roasted red beans, respectively. The relative abundances of isomers between two sample preparation methods and two N-glycans release methods are similar. The results suggest the unusual isomers found in this study were not produced by the artifact during the sample preparation.

In the first stage of biosynthesis, $Man_5GlcNAc_2$ isomer 5E4, $Man_6GlcNAc_2$ isomer 6E1, and $Man_7GlcNAc_2$ isomer 7D2 are generated and connected to the lipid dolichol-phosphatelocated on the cytoplasmic face of the ER membrane. The release of these oligosaccharides from dolichol typically requires acidic solution[38]. The observation of isomers 7D2, 6E1, and 5E4 are not likely due to the release from dolichol during the sample preparation. This is because (1) the pH values of ammonia solution and PNGase F buffer solution are 11 and 7.5, respectively, (2) isomers 6E1 and 5E4 are not the only unusual N-glycan isomers of $Man_6GlcNAc_2$ and $Man_5GlcNAc_2$ we found in various samples.

16

The N-glycans beyond the current multicellular eukaryote biosynthetic pathways have been reported in previous studies. $Man_5GlcNAc_2$ isomer 5D2 has been found in human recombinant gene expressed in CHO cells[46,47], royal jelly from bee[48], crayfish[49], flagellated protozoa[50]; isomer 5D1 has been found in horseshoe crab's proteins[51]; isomers 5D2 and 5E3 found in bovine ribonuclease B[25, 52]. Most of these unusual N-glycans were found in glycosylation mutants or in low abundance, or the abundances of these unusual N-glycans were not reported. Consequently, $Man_5GlcNAc_2$ and $Man_6GlcNAc_2$ found in the mass spectra of multicellular eukaryote samples are usually assumed to be isomer 5E1 and isomers 6D3, 6F1, and 6F2, respectively, and no further structural identification was made in most studies.

The high abundances of previously unidentified high mannose N-glycans in this study suggest additional biosynthetic pathways are involved in the generation of high mannose N-glycans. The enzymes involved in the additional biosynthetic pathways act on the terminal Manα1-3 and Manα1-6 before the complete removal of terminal Man-α-(1→2) from N-glycans. One possible enzymes are the enzymes in the third stage of N-glycan synthesis. In the third stage of current biosynthesis, the enzymes only act on terminal Man-α-(1→3) or Man-α-(1→6) after the complete removal of Man-α-(1→2) from N-glycans. These enzymes have low reactivity on terminal Man-α-(1→3) or Man-α-(1→6) of the N-glycans which have terminal Man-α-(1→2). The reactivity of these enzymes that act on terminal Man-α-(1→3) or Man-α-(1→6) before the complete removal of Man-α-(1→2) maybe high in some biological systems such as beef, pork, and red beans, resulting in high abundances of unusual high mannose N-glycans. Other possible enzymes are the enzymes not located in ER and Golgi. Glycoproteins released from the ER or Golgi may encounter other enzymes which were not located in ER and Golgi and have a high reactivity to act on the

terminal Man-α-(1→3) or Man-α-(1→6) of high mannose *N*-glycans before the complete removal of Man-α-(1→2). Yet, no such enzymes have been reported. In summary, our current findings on new high mannose *N*-glycan isomers warrant future investigation to identify enzymes and pathways responsible for their synthesis in multicellular eukaryotic cells.

Methods

(a) Sources of materials

Red beans, black beans, beef, pork, mushroom were purchased from local market. Bovine lactoferrin and bovine whey protein were purchased from AOR Inc. (Clifton, NJ, USA) and Raw GrassFed Whey (Fairfield, IA, USA), respectively. Enzymes PNGase F, α-mannosidase of *Canavalia ensiformis* (Jack bean), and α1-6 mannosidase were purchased from New England Biolabs Inc. (Ipswich, MA, USA). Both cell lines were gifted by Dr. Ruey-Hwa Chen, Institute of Biological Chemistry, Academia Sinica, Taiwan.

Bovine lactoferrin and bovine whey proteins were used directly. Beef, pork, and mushroom were lyophilized before ground into powder by grinder. Black beans were ground into powder by grinder directly. Red beans were separately into two groups. One of them was ground into powder by grinder directly. The other was baked in oven at 90°C for 20 minutes and then immersed in liquid nitrogen for 20 minutes, followed by repeating the same process (baking in oven and immersing in liquid nitrogen) twice before grounding into powder.

Isomers 5E1 and 5F1 were extracted from black bean; isomers 5D1 and 5D2 were purchased from Omicron Biochemicals, Inc. (South Bend, IN, USA); isomer 5E4 was generated from the degradation of $Man_6GlcNAc_2$ isomer 6E2 by enzyme α1-6

mannosidase; isomer 5F2 was produced from the degradation of $Man_6GlcNAc_2$ isomers 6G1 by α-mannosidase of *Canavalia ensiformis*; isomers 5E2 and 5E3 were generated from the degradation of $Man_7GlcNAc_2$ isomers 7D1 and 7D2, respectively, by α-mannosidase of *Canavalia ensiformis*. Isomers 6E1, 6E2, and 6G1 were purchased from Omicron Biochemicals, Inc.; isomers 6D3 and 6F2 were extracted from bovine whey proteins; isomers 6F1 was extracted from black bean; isomers 6D1 and 6D2 were generated from the degradation of $Man_7GlcNAc_2$ isomers 7D1 and 7D2, respectively, by α-mannosidase of *Canavalia ensiformis*. Isomers 7D1 and 7D2 were purchased from Omicron Biochemicals, Inc.; and 7E1, 7E2, 7D3, and 7G1 were extracted from bovine lactoferrin.

(b) Extraction of membrane proteins from human cell lines

The human mammary epithelial M10 cells and human triple-negative breast cancer MDA-MB-231 cells were cultured in DMEM (high glucose (4500 mg L^{-1}), Hyclone, South Logan, Utah, USA) supplemented with 10% (v/v) fetal bovine serum (Hyclone), 100 U/mL penicillin, and 100 mg/L streptomycin (Hyclone). All culturing media were supplemented with plasmocin (2.5 μg mL^{-1}; Invivogen, San Diego, California, USA) to mitigate mycoplasma contamination. The cells were maintained at 37 °C in a humidified incubator supplied with 5% CO_2. Cell membrane proteins were extracted using the differential ultracentrifugation protocol adopted from Li et al[53]. In brief, M10 and MDA-MB-231 cells were collected, washed with 0.22 μm-filtered PBS, pelleted, and each suspended in 2 mL of homogenization buffer (0.25 M sucrose, 20 mM HEPES–KOH (pH 7.4), and 1% protease inhibitor cocktail). Around 1×10^7 cells of each cell line were collected. The cell suspensions were kept at 4 °C prior to the succeeding steps of the extraction process. The cells were lysed using a sonicator system with standard probe (Q700; QSonica, Newton, Connecticuts, USA). The cell

lysates were centrifuged at 2,000 x *g* for 10 min at 4 ℃ to pellet the nuclei and mitochondria. The supernatants were transferred into new ultracentrifuge tubes and were centrifuged at 200,000 x *g* for 45 min at 4 ℃ to pellet plasma membrane proteins. The resulting pellets were washed with 0.2 M sodium carbonate solution to remove cell membrane–associated proteins and were centrifuged at 200,000 x *g* for 45 min at 4 ℃. The pellets were finally washed with double-distilled water and centrifuged at 200,000 x *g* for 45 min at 4 ℃. The cell membrane pellets were stored at −20 °C prior to downstream analyses. The concentration of the membrane protein was determined by BCA protein concentration assay.

(c) Extraction of *N*-glycans from various biological samples

The *N*-glycans were released from proteins using two methods. One is the ammonia-catalyzed reaction described in a previous study[40]. In brief, sample powder was dissolved in 25% ammonia aqueous solution and then underwent 16-h reaction at 60 ℃. After the reaction, the ammonia in the solution was removed using a rotary evaporator, and proteins were removed through ethanol (60%) precipitation. The samples were further purified using a C18 cartridge (Sep-Pak C18, Waters, Milford, MA, USA) to remove residual proteins and size exclusion chromatography (TOYOPEARL HW-40F, Tosoh Bioscience GmbH, Griesheim, Germany) to remove other impurities.

For the N-glycans released from both cell lines using PNGase F (New England Biolabs, Ipswich, MA, USA) digestion, 500 μL of denaturing buffer containing 0.5% SDS and 40 mM DTT were used to re-suspend the pellet and heated at 100 ℃ for 10 min followed by cooling at 0 ℃ for 10 min. To release the N-glycans from the proteins, 50 μL of PNGase F solution (250,000 units), 100 μL of GlycoBuffer 2, 100 μL of 10% NP-40, and 300 μL deionized water were added and incubated at 37 °C

overnight. The pH value of the buffer solution is 7.5. The released N-glycans were purified by 2 volumes of ethanol precipitation in −20 °C overnight. After centrifugation, the supernatant was completely dried down using dry centrifugal concentrator to remove the ethanol prior to solid phase extraction (SPE) using C18 and graphitized carbon (Extract-Clean SPE Carbo, Grace, Columbia, MD, USA) cartridges.

(d) Separation of *N*-glycan isomers

The fractions from size exclusion were sent into two-dimensional high-performance liquid chromatography (HPLC) for *N*-glycan isomer separation. They were first separated by HPLC with TSKgel Amide-80 column (150 mm × 2.0 mm, particle size of 5 µm, Tosoh Bioscience GmbH, Griesheim, Germany), followed by the separation by HPLC with Hypercarb PGC column. The amide-80 column separated *N*-glycans into different sizes, as illustrated in Fig. 2(a), to eliminate the interference by ESI in-source decay[41]. The fractions collected from the eluents of amide-80 column were sent into the second HPLC using a PGC Hypercarb column (2.1 mm × 100 mm, particle size of 3 µm, Thermo Fisher Scientific, Waltham, MA, USA) for *N*-glycan isomer separation. The eluents from PGC column were collected every 30 s using fraction collector (FC204, Gilson, Middleton, WI, USA). The eluents were analyzed by HPLC using the same PGC column again to ensure at most only one isomer is collected in each fraction collection.

The HPLC conditions for the TSKgel Amide-80 column were as follows: The flow rate was 0.2 mL/min, the gradient was changed linearly from A (H_2O) = 35% and B (acetonitrile) = 65% at t = 0 to A = 45% and B = 55% at t = 50 min. The HPLC conditions for the Hypercarb column (10 cm) were as follows: The flow rate was 0.2 mL/min; the gradient was changed linearly from A = 100%, B=0% at t = 0 to A = 82%,

B = 18% at t = 30 min.

(e) MS^n Mass Spectrometry

For *N*-glycan structural identification, samples purified by multidimensional HPLC were prepared in a NaCl (5×10^{-5}M) water/methanol (50:50 vol/vol) solution. A nanoelectrospray ionization instrument with a Nanospray Flex housing (Thermo Fisher Scientific) coupled to collision induced dissociation in a linear ion trap mass spectrometer (LTQ, Thermo Fisher Scientific Inc., Waltham, MA, USA) was used for the multistage tandem mass spectrometry (MS^n) measurement. Typically, 2 µL of each sample was loaded into a ESI emitter. The ESI source voltage was 1.5 kV, capillary voltage was 130 V, heated capillary temperature was 120 ℃, and tube-lens voltage was 230 V. Helium gas was used as a buffer gas for the ion trap as well as a collision gas in CID. The activation Q value was 0.25, activation time was 30 ms, normalised collision energy was 30-40%. The number of ions was regulated by injection time (10-20 ms) or automatic gain control (1×10^5 for full scan, and 1×10^4 for MS^n). The precursor ion isolation width was set to 1u. The sequences of CID in MS^n were guided by LODES. Details of LODES for *N*-glycan structural identification were described in our previous study[32].

After structural identification, HPLC chromatogram of each *N*-glycan isomer was measured using PGC Hypercarb column (2.1 mm × 150 mm, particle size of 3 µm, Thermo Fisher Scientific, Waltham, MA, USA) coupled to collision induced dissociation in a linear ion trap mass spectrometer (LTQ, Thermo Fisher Scientific). The HPLC conditions for the PGC Hypercarb column (15cm) were as follows: The flow rate was 0.15 mL/min, the gradient was changed linearly from A = 100% and B = 0% at t = 0, A = 93% and B = 7% at t = 5 to A = 80% and B = 20% at t = 60 min. The retention time of the chromatogram was recorded and collected as the database.

Meanwhile, the MS^2 and MS^3 mass spectra of each *N*-glycan isomer was measured during the chromatography and collected in the database.

(f) Enzyme digestion

The reaction conditions for α-mannosidase of *Canavalia ensiformis* and α1-6 mannosidase experiments were adapted from the protocols provided from the vendor. 1 µL of high mannose *N*-glycan (1 mM) from Omicron Biochemicals was added into the reaction buffer. The reaction buffer of α-mannosidase of *Canavalia ensiformis* containing 15 µL of DI water, 2 µL of Glycobuffer 4 (10X), and 2 µL of Zinc (10X), whereas the reaction buffer for the α1-6 mannosidase containing 15 µL of DI water, 2 µL of Glycobuffer 1 (10X), and 2 µL of BSA (100 µg/ml). The reaction mixture was incubated in 37℃, 800 rpm shaking for 5 minutes before 0.1 µL of α-mannosidase of *Canavalia ensiformis* or α1-6 mannosidase was added. Aliquots of 10 µL at different time points (5 minutes and 15 minutes for α-mannosidase of *Canavalia ensiformis* and α1-6 mannosidase, respectively) were taken, and the reaction was immediately quenched using a 0.6µL ZipTip C4 (Merck Ltd., Taipei, Taiwan). The aliquots were aspirated slowly for at least 15 times to bind the enzyme onto the pre-conditioned ZipTip C4 before eluted into a clean 300 µL HPLC vail. Additional 10 µL of DI water was added to wash the ZipTip C4 and the elution was combined into the same HPLC vail for HPLC-MS analysis.

Data availability

The data supporting the findings of this study are available within the article and its Supplementary Information.

Code availability

No custom code or mathematical algorithm that is deemed central to the conclusions

was used in this study.

References

1. Shental-Bechor, D. & Levy, Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 8256–8261 (2008).

2. Live, D. H., Kumar, R. A., Beebe, X. & Danishefsky, S. J. Conformational influences of glycosylation of a peptide: a possible model for the effect of glycosylation on the rate of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 12759–12761 (1996).

3. Schachter, H. Processing of protein-bound *N*-glycans. In *Comprehensive glycoscience from chemistry to systems biology*. Edited by Kamerling, J. P. Elsevier Science Ltd, Oxford (2007).

4. Stanley, P., Taniguchi, N., & Aebi, M. *N*-glycans. In *Essentials of Glycobiology*, 3rd edition, edited by Varki A. et al. Cold Spring Harbor Laboratory Press, New York (2017).

5. Parodi, A., Cummings, R. D., & Aebi, M. Glycans in glycoprotein quality control. In *Essentials of Glycobiology*, 3rd edition, edited by Varki A. et al. Cold Spring Harbor Laboratory Press, New York (2017).

6. Kornfeld, R., & Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* **54,** 631–664 (1985).

7. Schenk, B., Fernandez, F., & Waechter, C. J. The ins(ide) and out(side) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum. *Glycobiology* **11,** 61R–70R (2001).

8. Aebi, M. N-linked protein glycosylation in the ER. *Biochim Biophys Acta* **1833,** 2430–2437 (2013).

9. Shrimal, S., Cherepanova, N.A., & Gilmore, R. Cotranslational and posttranslocational N-glycosylation of proteins in the endoplasmic reticulum. *Semin Cell Dev Biol* **41,** 71–78 (2015).

10. Lundborg, M. & Widmalm, G. Structural analysis of glycans by NMR chemical shift prediction. *Anal. Chem.* **83**, 1514–1517 (2011).

11. Wang, N., Seko, A., Takeda, Y. & Ito, Y. Preparation of asparagine-linked monoglucosylated high-mannosetype oligosaccharide from egg yolk. *Carbohydr. Res.* **411**, 37−41 (2015).

12. Dell, A. & Morris, H. R. Glycoprotein structure determination by mass spectrometry. *Science* **291**, 2351–2356 (2001).

13. Ashline, D., Singh, S., Hanneman, A. & Reinhold, V. Congruent strategies for carbohydrate sequencing. 1. Mining structural details by MSn. *Anal. Chem.* **77**, 6250–6262 (2005).

14. Kailemia, M. J., Ruhaak, L. R., Lebrilla, C. B. & Amster, I. J. Oligosaccharide analysis by mass spectrometry: a review of recent developments. *Anal. Chem.* **86**, 196–212 (2014).

15. Zaia, J. Mass Spectrometry and Glycomics. *OMICS* **14**(4), 401–418 (2010).

16. Nwosu, C.C. et al. Comparison of the human and bovine milk N-Glycome via high-performance microfluidic chip liquid chromatography and tandem mass spectrometry. *J. Proteome Res.* **11**, 2912−2924 (2012).

17. Dong X., Zhou, S. & Mechref, Y. LC‑MS/MS analysis of permethylated free oligosaccharides and N‑glycans derived from human, bovine, and goat milk samples. *Electrophoresis* **37**, 1532–1548 (2016).

18. Houel, S. et al. N- and O-glycosylation analysis of etanercept using liquid chromatography and quadrupole time-of-flight mass spectrometry equipped with electron-transfer dissociation functionality. A*nal. Chem.* **86**, 1, 576–584 (2014).

19. Song, T., Aldredge, D. & Lebrilla, C. B. A method for in-depth structural annotation of human serum glycans that yields biological variations. *Anal. Chem.* **87**, 7754−7762 (2015).

20. Harvey, D. J. & Abrahams, J.L. Fragmentation and ion mobility properties of negative ions from N-linked carbohydrates: part 7. Reduced glycans. *Rapid Commun. Mass Spectrom.* **30**, 627−634 (2016).

21. Ashwood, C.; Lin, C. H.; Thaysen-Andersen, M. & Packer, N. H. Discrimination of isomers of released N- and O-glycans using diagnostic product ions in negative ion PGC-LC-ESI-MS/MS. *J. Am. Soc. Mass Spectrom.* **29**, 1194−1209. (2018).

22. Morimoto, K. et al. *Bioinformatics* **31**, 2217−2219 (2015).

23. Klein, J. & Zaia, J. *J. Proteome Res.* **18**, 3532−3537 (2019).

24. Rojas-Macias, et al. *Nat. Commun.* **10**, 1−10 (2019).

25. Prien, J. M., Ashline, D. J., Lapadula, A. J., Zhang, H. & Reinhold, V. N. The high mannose glycans from bovine ribonuclease B isomer characterization by ion trap MS. *J. Am. Soc. Mass Spectrom.* **20**, 539−556 (2009).

26. Wei, J. et al. Toward automatic and comprehensive glycan characterization by online PGC-LC-EED MS/MS. *Anal. Chem.* **92**, 782−791 (2020).

27. Hsu, H. C., Liew, C. Y., Huang, S. P., Tsai, S. T. & Ni, C. K. Simple approach for De novo structural identification of mannose Trisaccharides. *J. Am. Soc. Mass Spectrom.* **29**, 470−480 (2018).

28. Hsu, H. C., Liew, C. Y., Huang, S. P., Tsai, S. T. & Ni, C. K. Simple method for De novo structural determination of underivatised glucose oligosaccharides. *Sci. Rep.* **8**, 5562 (2018).

29. Tsai, S. T. et al. Automatically full glycan structural determination with logically derived sequence tandem mass spectrometry. *ChemBioChem* **20**, 2351−2359 (2019).

30. Hsu, H. C., Huang, S. P., Liew, C. Y., Tsai, S. T. & Ni, C. K. De novo structural determination of mannose oligosaccharides by using a logically derived sequence for tandem mass spectrometry. *Anal. Bioanal. Chem.* **411**, 3241–3255 (2019).

31. Ni, C. K.; Hsu, H. C.; Liew, C. Y.; Huang, S. P.; Tsai, S. T.: Modern mass spectrometry techniques for oligosaccharide structure determination: logically derived sequence tandem mass spectrometry for automatic oligosaccharide structural determination. In *Comprehensive glycoscience from chemistry to systems biology*, 2nd edition, edited by Barchi, J., Elsevier Science Ltd, Oxford, (2021).

32. Liew, C. Y.; Yen, C. C.; Chen, J. L.; Tsai, S. T.; Pawar, S.; Wu, C. Y., Ni, C. K.: Structural identification of *N*-glycan isomers using logically derived sequence tandem mass spectrometry. *Commun. Chem.* **4,** 92 (2021).

33. Liew, C. Y., Chan, C. K., Huang, S. P., Cheng, Y. T., Tsai, S. T., Hsu, H. C., Wang, C. C. & Ni, C. K. De novo structural determination of oligosaccharide isomers in glycosphingolipids using logically derived sequence tandem mass spectrometry. *Analyst*, **146**, 7345-7357 (2021).

34. Weng, WC., Liao, HE., Huang, SP. et al. Unusual free oligosaccharides in human bovine and caprine milk. *Sci Rep* **12**, 10790 (2022).

35. Tomiya, N., Lee, Y. C., Yoshida, T., Wada, Y., Awaya, J., Kurono, M. & Takahashi, N. Calculated two-dimensional sugar map of pyridylaminated oligosaccharides: Elucidation of the jack bean a-mannosidase digestion pathway of $Man_9GlcNAc_2$, *Anal. Biochem.* **193**, 90-100 (1991).

36. Pabst, M., Bondili, J. S., Stadlmann, J., Mach, L. & Altmann, F. Mass+retention time=structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin *N*-glycans. *Anal. Chem.* **79**, 5051–5057 (2007).

37. Aldredge, D., An, H. J., Tang, N., Waddell, K. & Lebrilla, C. B. Annotation of a

serum N-glycan library for rapid identification of structures. *J. Proteome Res.*, **11**(3), 1958–1968 (2012).

38. Pabst, M., Grass, J., Toegel, S., Liebminger, E., Strasser, R. & Altmann, F. Isomeric analysis of oligomannosidic N-glycans and their dolichol-linked precursors *Glycobiology* **22**(3), 389–399 (2012).

39. Abrahams, J. L., Campbell, M. P. & Packer, N. H. Building a PGC-LC-MS N-glycan retention library and elution mapping resource, *Glycoconjugate J.* **35**(1), 15–29 (2018).

40. Yang, M. et al. Separation and preparation of N-glycans based on ammonia-catalyzed release method. *Glycoconjugate J.* **37**, 165–174 (2020).

41. Liew, C. Y., Chen, J. L. & Ni, C. K. Electrospray ionization in-source decay of *N*-glycans and the effects on *N*-glycan structural identification. *Rapid Commun. Mass Spectrom.* **36**, e9352 (2022).

42. Tsai, S. T., Chen, J. L. & Ni, C. K. Does low-energy collision-induced dissociation of lithiated and sodiated carbohydrates always occur at anomeric carbon of the reducing end? *Rapid Commun. Mass Spectrom.* **31**, 1835−1844 (2017).

43. Chen, J. L.et al. Collision-induced dissociation of sodiated glucose and identification of anomeric configuration. *Phys. Chem. Chem. Phys.* **19**, 15454−15462 (2017).

44. Huynh, H. T. et al. Collision induced dissociation of sodiated glucose, galactose, and mannose, and the identification of anomeric configurations. *Phys. Chem. Chem. Phys.* **20**, 19614–19624 (2018).

45. Chiu, C.C. et al. Unexpected dissociation mechanism of Sodiated N-Acetylglucosamine and N-Acetylgalactosamine. *J. Phys. Chem. A.* **123**, 3441–3453 (2019).

46. Hwa, K. & Khoo, K. Structural analysis of the asparagine-linked glycans from the procyclic Trypanosoma brucei and its glycosylation mutants resistant to Concanavalin A killing. *Mol. Biochem. Parasitol*. **111**, 173–184 (2000).

47. Ohta, M., Ohnishi, T., Ioannou, Y., Hodgson, M., Matsuura, F. & Desnick, R. Human alpha-N-acetylgalactosaminidase: site occupancy and structure of N-linked oligosaccharides. *Glycobiology* **10**, 251–261 (2000).

48. Kimura, Y., Miyagi, C., Kimura, M., Nitoda, T., Kawai. N. & Sugimoto, H. Structural features of N-glycans linked to royal jelly glycoproteins: structures of high-mannose type, hybrid type, and biantennary type glycans. *Biosci. Biotechnol. Biochem.* **64**, 2109–2120 (2000).

49. Tseneklidou-Stoeter, D., Gerwig, G. J., Kamerling, J. P. & Spindler, K. D. Characterization of N-linked carbohydrate chains of the crayfish, Astacus leptodactylus hemocyanin. *Biol. Chem. Hoppe. Seyler.* **376**, 531–537 (1995).

50. Olafson, R., Thomas, J., Ferguson, M., Dwek, R., Chaudhuri, M., Chang, K., Rademacher, T. Structures of the N-linked oligosaccharides of Gp63, the major surface glycoprotein, from Leishmania mexicana amazonensis. *J. Biol. Chem.* **265**, 12240–12247 (1990).

51. Amatayakul-Chantler, S., Dwek, R., Tennent, G., Pepys, M. & Rademacher, T. Molecular characterization of Limulus polyphemus C-reactive protein. II. Asparagine-linked oligosaccharides. Eur. J. Biochem. **214**, 99–110 (1993).

52. Prien, J. M., Prater, B. D. & Cockrill, S. L. A multi-method approach toward de novo glycan characterization: A Man-5 case study. *Glycobiology* **20**(5), 629–647 (2010).

53. Li, Q., Xie, Y., Wong, M., Barboza, M. & Lebrilla, C. B. Comprehensive structural glycomic characterization of the glycocalyxes of cells and tissues. *Nat. Protoc.* **15**(8) 2668-2704 (2020).

Author contributions

C.Y.L. extracted parts of the *N*-glycans from lactoferrin and bovine why proteins, measured and analyzed the chromatograms and CID spectra of the *N*-glycans extracted from lactoferrin, black bean, bovine why protein, mushroom, human cell line M10, breast cancer cell line MDA-MB-231, construct all the chromatograms and mass spectra in database, and wrote the manuscript and Supplementary Information. HSL extracted and analyzed the *N*-glycans from black bean, red bean, beef, and pork, TYY extracted the *N*-glycans from mushroom, ATH extracted *N*-glycans from bovine whey proteins. BJAM and CPKL grew M10 and MDA-MB-231 cells. C.K.N designed the experiments, extracted part of the *N*-glycan from lactoferrin, conceived the LODES/MS^n method, analyzed CID spectra, and wrote the manuscript and Supplementary Information.

Author Information

Corresponding authors

Chi-Kung Ni − Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, Taiwan, Molecular Science and Technology, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan, and Department of Chemistry, National Tsing Hua University, Hsinchu, Taiwan; orcid.org/0000-0001-6503-8905; Email: ckni@po.iams.sinica.edu.tw

Authors

Chia Yen Liew − Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, Taiwan; International Graduate Program of Molecular Science and Technology, National Taiwan University (NTU-MST), Taipei, Taiwan; Molecular Science and

Technology (MST), Taiwan International Graduate Program (TIGP), Academia Sinica, Taipei, Taiwan; orcid.org/0000-0003-1748-8981.

Hong-Sheng Luo − Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, Taiwan and Department of Chemistry, National Taiwan Normal University, Taipei, Taiwan.

Ting-Yi Yang − Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, Taiwan and Department of Chemistry, National Taiwan Normal University, Taipei, Taiwan.

An-Ti Hung − Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, Taiwan and Department of Chemistry, National Tsing Hua University, Hsinchu, Taiwan.

Bryan John Abel Magoling − Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, Taiwan, Institute of Biochemical Sciences, College of Life Science, National Taiwan University, Taipei, Taiwan, and Chemical Biology and Molecular Biophysics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan

Charles Pin-Kuang Lai − Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, Taiwan, Chemical Biology and Molecular Biophysics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan, and Genome and Systems Biology Degree Program, National Taiwan University and Academia Sinica, Taipei, Taiwan.

Competing Interests

C.Y.L. and C.K.N. are co-inventors of a United States patent (US 10,796,788 B2) that part of the method described in the patent to determine the carbohydrate structure was used in this work. All other authors declare no competing interests.