Article

Toward automated identification of glycan branching patterns using multistage mass spectrometry with intelligent precursor selection

Shiwei Sun, Chuncui Huang, Yaojun Wang, Yaming Liu, Jingwei Zhang, Jinyu Zhou, Feng Gao, Fei Yang, Runsheng Chen, Barbara Mulloy, Wengang Chai, Yan Li, and Dongbo Bu

Anal. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.analchem.8b03967 • Publication Date (Web): 16 Nov 2018

Toward automated identification of glycan branching patterns using multistage mass spectrometry with intelligent precursor selection

Shiwei Sun[1,3†], Chuncui Huang[2†], Yaojun Wang[1,3†], Yaming Liu[2,3], Jingwei Zhang[1,3], Jinyu Zhou[2,3], Feng Gao[1,3], Fei Yang[1,3], Runsheng Chen[2,3], Barbara Mulloy[4], Wengang Chai[4*], Yan Li[2,3*], and Dongbo Bu[1,3*]

[1]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, 6 Kexueyuan South Road, Beijing 100080, China.

[2]Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China.

[3]University of Chinese Academy of Sciences, 19 Yuquan Road, Beijing 100049, China.

[4]Glycosciences Laboratory, Faculty of Medicine, Imperial College London, London W12 0NN, United Kingdom.

†These authors contributed equally to this work.

*Correspondence should be addressed to W.C. (w.chai@imperial.ac.uk), Y.L. (yanli@ibp.ac.cn) or D.B. (dbu@ict.ac.cn)

ABSTRACT

Glycans play important roles in a variety of biological processes. Their activities are closely related to the fine details of their structures. Unlike the simple linear chains of proteins, branching is a unique feature of glycan structures, making their identification extremely challenging. Multistage mass spectrometry (MS^n) has become the primary method for glycan structural identification. The major difficulty for MS^n is the selection of fragment ions as precursors for the next stage of scanning. Widely-used strategies are either manual selection by experienced experts, which requires considerable expertise and time, or simply selecting the most intense peaks by which the product-ion spectrum generated may not be structurally informative and therefore fail to make the assignment. We here report a glycan 'intelligent precursor selection' strategy (GIPS) to guide MS^n experiments. Our approach consists of two key elements, an empirical model to calculate candidate glycan's '*probability*' and a statistical model to calculate fragment ion's '*distinguishing power*' in order to select the structurally-most informative peak as the precursor for next-stage scanning. Using 15 glycan standards, including 3 pairs with isomeric sequences, and 8 variously fucosylated oligosaccharides on linear or branched hexasaccharide backbones isolated from a human milk oligosaccharide fraction by HPLC, we demonstrate its successful application to branching pattern analysis with improved efficiency and sensitivity, and also the potential for automated operation.

INTRODUCTION

Glycans play important roles in a wide range of biological processes such as protein conformation, molecular recognition, cell proliferation and differentiation[1,2]. These roles are closely related to the fine details of their structures. Unlike the linear chain of proteins, branching is a unique feature of glycan structures and increased branching of *N*-glycans has been shown to be correlated with invasiveness and metastatic potential of tumors[3-5]. Different linkages and multiple isobaric residues (e.g. HexNAc: GalNAc and GlcNAc; and Hex: Man, Gal, and Glc) are also unique to glycans and these make glycan structural identification extremely challenging.

To fully appreciate biological roles of glycans, high-throughput analysis of glycan activities and structures is highly desirable. Major progress has been made during the last decade in revealing glycan activities and glycan microarray has played an important role[6]. However, a high-throughput method for glycan structural identification remains a challenge. Current approaches rely on a combination of analytical techniques such as mass spectrometry (MS), methylation analysis, glycosidase digestion, and immunochemical detection using mAbs and lectins[7]. Although NMR can be used to determine complete structure, the amount of glycan typically required for analysis precludes its use in most cases.

Recent development in MS has opened up new possibilities and it has been the major method for both profiling of glycan mixtures isolated from cells or tissues and elucidation of the complex glycan sequence[8-12]. Tandem MS (MS^2) with collision-induced dissociation (CID) has been used successfully in high-sensitivity glycan sequence analysis[13,14]. Glycosidic bond cleavage with B-/Y- and C-/Z-ions[15] affords sequence and branching pattern information while cross-ring fragmentation with different types of A-/X-ions may occur, although frequently at low abundances, to indicate linkage positions[16].

A variety of software tools[17,18] has also been proposed to assist interpretation of the complex data obtained from MS^2 using different strategies, for example database search[19] and the so-called *de novo* method using sophisticated algorithms[20]. However, MS^2 is not always

sufficient and multiple-stage mass spectrometry (MS^n, n>2) is frequently required to provide detailed structural information of glycans[21-25]. In the past, MS^n can only be carried out with ion-trap and FT-MS types of instrument mainly by electrospray ionization (ESI). It has now become more accessible due to the development in instrumentation and the popularity of orbitrap and quadrupole ion-trap/time-of-flight (QIT-TOF) instruments for both ESI- and MALDI-MS, respectively.

The major difficulty for successful application of MS^n and thereafter automated high-throughput structural identification of glycans is precursor-ion selection. Unlike MS^2, sequential MS^n requires careful selection of fragment ions as precursors for further stages. Currently, precursor-ion selection is either manually by experienced experts[21, 26,27] or simply picking-up the most intense peaks (MIPS)[22]. Manual selection requires considerable expertise and time, whereas product-ion spectrum generated by MIPS may not be structurally informative and therefore fail to make the assignment. The concept of intelligent precursor selection to guide MS^n experiments has been suggested[22,28] but a systematic approach for achieving it has never been proposed and remains a challenge.

In this study, we report such an approach for glycan identification using a glycan intelligent precursor-ion selection (GIPS) strategy. We implemented this approach by computer program, in which we established an empirical model (the *P-model*) to calculate '*probability*' of each candidate glycan and developed a statistical model (the *DP-model*) to select the structurally-most informative peaks as precursors. For the latter, a new concept of '*distinguishing power*' (DP) was proposed, calculated and assigned to each fragment ion. We demonstrated the GIPS approach using Man-5D1 as an example, and applied to assignment of 12 human milk-related oligosaccharides as representatives including 3 pairs of isomeric glycans, and compared with the MIPS method.

For reference glycan database, although there has been a considerable increase in number since 2000, e.g. the KEGG GLYCAN[29] GlycomeDB[30] and EUROCarbDB[31], we selected the most widely-used and well-documented CarbBank[32] (also known as CCSD[33]), developed by the

Complex Carbohydrate Research Center, University of Georgia (Athens) and consisted of 7,837 glycan structures, for our use in this initial proof-of-concept work.

EXPERIMENTAL SECTION

Materials

Glycan standards were purchased from Dextra Laboratories (Reading, England) and Ludger (Abingdon, England). Sodium chloride, acetic acid, sodium hydroxide and standard peptides were purchased from Sigma-Aldrich. Acetonitrile, ethanol and water were obtained from Avantor Performance Materials (Center Valley, PA). C18-Sep-Pak cartridge was obtained from Waters (Milford, MA). All the chemicals were of analytical grade or better, and used as received without further purification.

Human milk oligosaccharides were isolated from a neutral oligosaccharide fraction containing variously fucosylated hexasaccharides by HPLC on an amide column (Waters XBridge, 3.5 μm, 4.6 × 250 mm). A gradient of ACN/H_2O (Sol A: 80:20 and Sol B: 20:80; 22-37%B in 40 min) was used for elution and UV 196 nm for detection.

Permethylation of glycans

Permethylation and purification were performed as previously reported. Briefly, methyl iodide was added to standard oligosaccharides or *N*-glycans in the presence of NaOH slurry, and the sample was then agitated on an automatic shaker at room temperature for 20 min. Chloroform was added and mixed thoroughly with the sample. The mixture was then allowed to settle into two layers. The upper aqueous layer was removed, and the chloroform layer was dried down under a gentle stream of nitrogen. Finally, the mixture was purified on a C18 Sep-Pak cartridge.

MALDI-MS

Permethylated oligosaccharide standards and *N*-glycans were analyzed on an Axima MALDI Resonance mass spectrometer with a QIT-TOF configuration (Shimadzu). A nitrogen laser was used to irradiate samples at 337 nm, with an average of 200 shots accumulated. Permethylated glycan standards dissolved in methanol were applied to a μfocus MALDI plate target (900 μm, 384 circles, Hudson Surface Technology, NJ). A matrix solution (0.5 μL) of 2,5-dihydroxybenzoic acid (20 mg/mL) in a mixture of methanol/water (1:1) containing 0.1% trifluoroacetic acid and 1 mM NaCl was added to the plate and mixed with samples. The mixture was air dried at room temperature before analysis. A standard mass resolution of FWHM 250 and collision energy between 100-400 mV was used for MS^n.

The product-ion spectrum acquired at each stage was introduced into the GIPS program as a mzXML file (Shimadzu) for peak identification, using a signal-to-noise ratio 3:1 as the filtering parameter. Based on the primary spectra, candidate glycans were extracted from CarbBank and stored in the GIPS system. The probability and DP were calculated by GIPS and the results were manually fed back to the mass spectrometer data system to instruct further product-ion scanning using the fragment ion with the highest DP to generate next-stage spectrum until the probability of a candidate glycan exceeded the pre-defined threshold 0.70. A mass tolerance was set at 0.5.

For MS^n using the MIPS approach, the most intense peak was selected as the precursor for the next-stage MS scanning whereas for MS^n using the GIPS approach, the peak with the highest DP was selected as the precursor as directed by the computer program.

GIPS approach for glycan identification

The basic operation of the GIPS approach are as follows.

- *Initialization:* From the primary mass spectrum (denoted as S_1) of the glycan of interest, its molecular mass can be determined and all glycans with the identical mass are extracted from CarbBank and listed as candidate glycans (denoted as $G_1, G_2, \ldots, G_n$). Initially, all candidate glycans are assigned with equal probability:

$$P(G_1 \mid S_1) = P(G_2 \mid S_1) = \ldots P(G_n \mid S_1) = \frac{1}{n}$$

As none of the candidate glycans has probability exceeding the pre-defined threshold 0.70, the GIPS program instructs the instrument to carry out MS^2 using the only apparent peak in the primary spectrum as the precursor. When the MS^2 spectrum (denoted as S_2) is generated, the probabilities of the candidate glycans are re-calculated using Formula (2) shown below in the Results and Discussion section.

- *Iteration:* For each fragment ion in the acquired experimental mass spectra, its distinguishing power (DP) is calculated using Formula (4), and the ion with the highest DP is selected as precursor to generate a next-stage mass spectrum. Based on this newly-generated spectrum, the candidates' probabilities are re-calculated using Formula (2).

- *Stopping criteria:* The identification process terminates when a certain candidate glycan has probability exceeding the pre-defined threshold 0.70. In principle, a threshold of 0.50 is sufficient for deducing the actual glycan even if there are two candidate glycans. However, the threshold is set as 0.70 to allow for robust and reliable identification. Other values (e.g. 0.60 and 0.80) were also tested but 0.70 was selected as GIPS is insensitive to these settings (Table S-1).

The calculation of candidate's probability as well as the fragment ion's DP are also described and illustrated in detail in Results and Supplemental Results sections.

Structural determination by ESI-MS/MS and ^{1}H-NMR

Sequence determination of the eight HPLC fractions of human milk oligosaccharides was carried out as described[34,35] on a Waters SYNAPT G2 mass spectrometer (Manchester, UK). Cone voltage was at 80 eV for MS/MS of $[M–H]^–$, 20 eV for $[M–2H]^{2–}$ and 150 eV for $D_{2\beta\text{-}3}$. Argon

was used as the collision gas and 30, 15 and 18 V were used as the collision energies for fragmentation of $[M–H]^–$, $[M–2H]^{2–}$ and $D_{2\beta\text{-}3}$, respectively.

NMR experiment was performed on a Bruker Avance III HD 700 MHz instrument equipped with a 5 mm $^1H/^{13}C/^{15}N$ triple-resonance PFG cryoprobe, using the pulse sequences provided at a temperature of 30 ºC. The TOCSY spectrum was acquired with 2K points in F2 and 512 in F1 and was processed without zero filling; the mixing time was 80 ms. Oligosaccharide fractions HPLC-3 and -5 were dissolved in D_2O (99.9%), exchanged by lyophilization, redissolved in 0.6 mL or 0.3 ml D_2O, and placed in a normal or low volume 5 mm NMR tube, respectively, for data acquisition.

RESULTS AND DISCUSSION

The GIPS approach assigns the structure of a glycan of interest (*actual glycan*) by comparing its MS^n spectra with many possible glycans with identical masses listed in a well-established glycan database, the CarbBank in the present study. When the molecular mass of the glycan of interest is identified from the primary mass spectrum, all glycans with identical molecular masses listed in the database are extracted and listed as possible glycans (*candidate glycans*). Glycan identification is then achieved by assigning the actual glycan from the listed candidate glycans based on MS^n spectra. Although GIPS is suitable for other types of MS, in the present study positive-ion MALDI-MS of permethylated glycans was carried out with the sodiated molecular ion as the precursor for product-ion scanning.

The Concept of GIPS

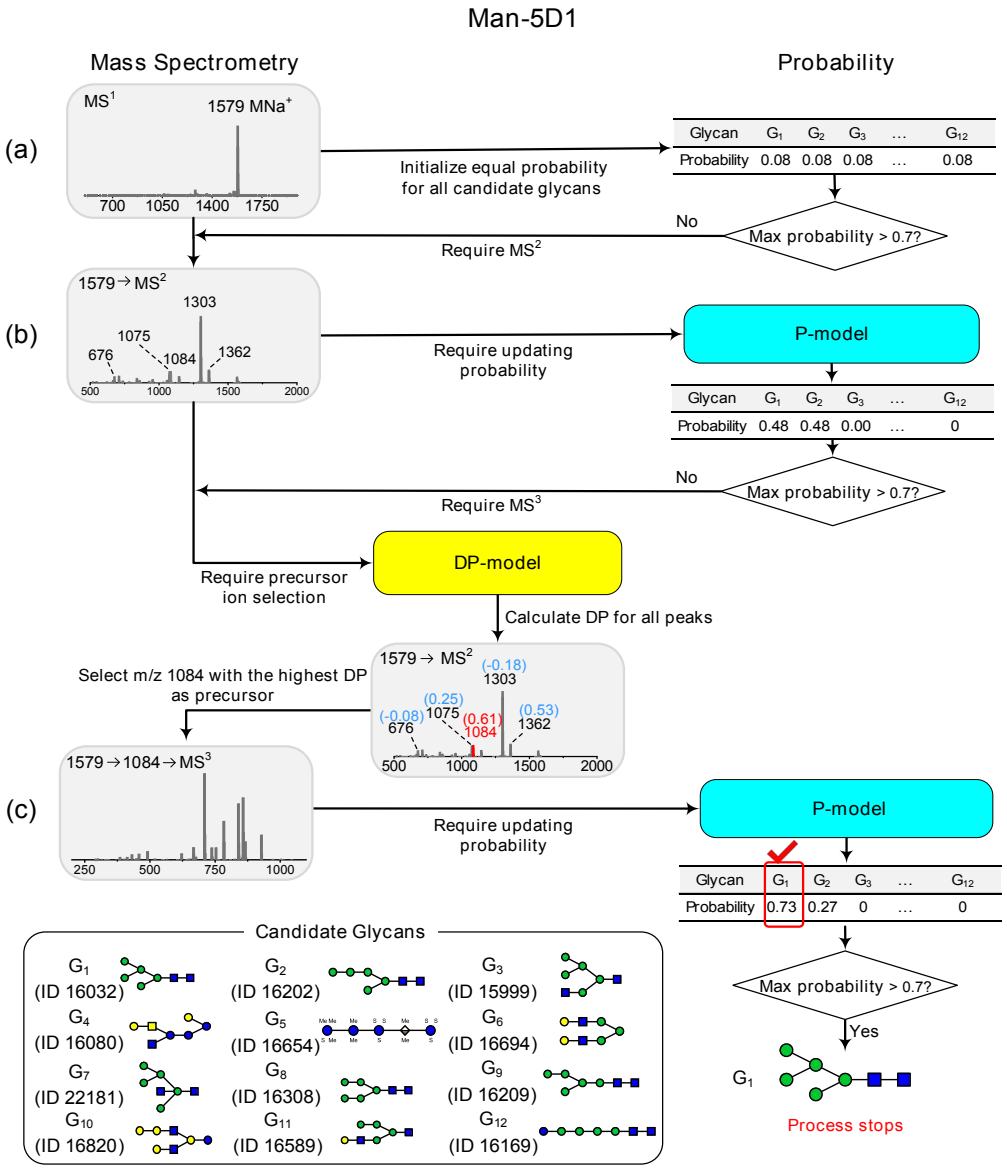


Figure 1. Identification process of Man-5D1 using GIPS.

Three rounds of mass spectrometric scanning (a: MS^1; b: MS^2, and c, MS^3) were required to assign Man-5D1 from the 12 candidate glycans *G_1 to G_{12}* (structures together with CarbBank ID numbers listed in the box at the bottom). Left column: mass spectrometry process; right column: calculated probabilities; and middle column: required calculation process. The fragment ion shown in red was selected as the precursor for the next round of product-ion scanning. The calculated DPs are in brackets above the *m/z* values in the spectrum.

Using the high mannose *N*-glycan Man-5D1 as an example (Figure 1 and Table S-2), we describe the main concept and steps of the GIPS program for glycan identification, and how the process is accomplished in an interactive manner with the mass spectrometer instrument rather

than using pre-programmed parameters.

MS^1 of Man-5D1 gave a MNa^+ at *m/z* 1579 (Figure 1a), indicating a glycan with a molecular mass of 1556 *Da*. A total of 12 candidate glycans can be extracted from the database and listed as G_1–G_{12} (Figure 1, text box at the bottom). The probability of each candidate glycan was initially calculated as equal at 0.08 ($= \frac{e^1}{\sum_{i=1}^{12} e^1}$). As this is below the termination threshold 0.70, the GIPS program instructed the instrument to carry out MS^2 using the only ion MNa^+ at *m/z* 1579 as the precursor (Figure 1a).

When the MS^2 spectrum was obtained, the probability of the 12 candidate glycans were re-calculated and updated using the *P-model* (see below for detailed calculation of probability). The underlying reason for recalculation is that the fragment ions produced in MS^2 could favor certain candidates and discriminate against others, and thus the probabilities are no longer equal. In the present case, the updated probabilities for both G_1 and G_2 increased from 0.08 to 0.48, while others decreased to nearly 0.00 (Figure 1b, right column). As none of the candidate probabilities exceeded 0.70, further round of product-ion scanning MS^3 is required.

Unlike MS^2, which uses the only molecular ion MNa^+ in the MS^1 spectrum as the precursor, multiple fragment ions were present in the spectrum of MS^2 (e.g. *m/z* 1303, 1084, and 1075, Figure 1b, left column) and could be used as the precursor for MS^3. At this stage, to aid the selection of the best ion as the precursor for the next round of scanning, we assigned a *distinguishing power* (DP) value, to each ion peak using the *DP-model*. DP is a calculated measure of an ion's ability to produce a product-ion spectrum containing distinctive structural information to be used in its differentiation from other isomers. The higher the value, the more distinctive-structural information it can produce. The detailed calculation of DP is described in the section below.

In the present case, among the major ions in the spectrum of MS^2 (Figure 1b), *m/z* 1084 had the highest DP (0.61) compared with other ions, e.g. *m/z* 1362 (DP 0.53) and *m/z* 1075 (DP 0.25) (Figure 1b, middle column shown in brackets), and was therefore selected as the precursor for MS^3.

Based on the newly generated spectrum of MS^3 (Figure 1c), together with that of MS^2, the probability of each candidate glycan was re-calculated. Among the updated probabilities, G_1 has the highest (0.73) and G_2 has a moderate (0.27), while all others are close to 0 (Figure 1c, right column). Thus, the identification process stopped with G_1 reported as the actual glycan, as its probability exceeded the termination threshold 0.70. In the case that MS^3 is insufficient for glycan identification, a further round of product-ion scanning will continue in a similar way until a glycan structure can be found with a probability exceeding 0.70.

Using this approach, we successfully identified different types of human milk-related oligosaccharides, including B-Tetra-T1, LNT, MFLNH-III, LNDFH-II, SLe^a, LST-a, MFiLNO and TFiLNO (Figures S-1a to S-14a, respectively).

Calculations of Candidate's *Probability*

Calculation of probability has taken the two factors into consideration: *i)* how to integrate the information from all experimental MS^n spectra obtained; and *ii)* how to transform the information into an appropriate form such that a universal termination threshold can be set for reliable identification.

The basic idea to establish such a model (the *P-model*) is to consider each peak in an experimental spectrum as a clue to the actual glycan; thus, the candidate that has the most supporting clues is likely to be the actual glycan. These clues are then integrated into a score using an empirical scoring function and normalized into [0, 1] using the *softmax* function, and therefore enabling setting of a universal termination threshold.

To integrate the clues to each candidate G_i provided by experimental mass spectra S_1, S_2, …, S_m, an empirical scoring function $f(G_i, S_1, S_2, \ldots, S_m)$ is used:

$$f(G_i, S_1, \cdots, S_m) = \sum_{k=1}^{m} \sum_{l=1}^{|S_k|} I\left(G_i, S_k^{(l)}\right) \quad \ldots\ldots(1)$$

Here, $S_k^{(l)}$ represents the l-th peak of S_k, $I\left(G_i, S_k^{(l)}\right) = 1$ if $S_k^{(l)}$ matches certain fragment of G_i

and $I(G_i,\ S_k^{(l)}) = 0$ otherwise. The score function $f(G_i, S_1,S_2,...,S_m)$ has an advantage in its additivity property, i.e. when a new spectrum S_{m+1} is obtained, $f(G_i,S_1,\cdots,S_{m+1})$ can be calculated as $f(G_i,S_1,\cdots,S_{m+1}) = f(G_i,S_1,\cdots,S_m) + \sum_{l=1}^{|S_{m+1}|} I(G_i, S_{m+1}^{(l)})$, which simplifies the calculation process.

The score $f(G_i, S_1,S_2,...,S_m)$ is then normalized into candidate probability $P(\ G_i \mid S_1,S_2,\cdots,S_m)$ using the *softmax* function, i.e.,

$$P(\ G_i \mid S_1,S_2,...,S_m) = \frac{e^{f(G_i, S_1,S_2,...,S_m)}}{\sum_{j=1}^{n} e^{f(G_j, S_1,S_2,...,S_m)}} \quad (2)$$

The candidates' probabilities are summed up to 1, thus enabling setting a universal termination threshold that applies for various samples.

In the case of Man-5D1 described above, the probability of the candidate G_1 increased from 0.08 to 0.48 after the MS^2 spectrum (as S_2 in this case) was obtained. The probability $P(\ G_1 \mid S_1,S_2) = 0.48$ was obtained by the following calculation. As shown in Table S-3, the peaks in S_1, S_2 that can be explained by certain fragments of G_1 were counted as $f(G_1, S_1,S_2)$=19, and thus $e^{f(G_1, S_1,S_2)} = 178482300.96$. Similarly, $e^{f(G_i, S_1,S_2)}$ was calculated for all candidate glycans, giving a sum of 371658136.39. Finally, the probabilities of G_1 was calculated as $\frac{178482300.96}{371658136.39} = 0.4802$.

Calculations of Fragment Ion's *Distinguishing Power*

The concept of a fragment ion's DP is the key element of the GIPS strategy. DP is used to guide the acquisition of the product-ion spectra in the MS^n experiment by selecting the ion with the highest DP as the precursor. The product-ion spectrum thus obtained could be most distinctive and structurally informative, leading to unambiguous assignment. For a fragment ion *I* with its experimental production-ion spectrum S', its DP is defined as the amount of information of the actual glycan that S' can provide, which is quantitatively measured by the entropy decrease of

candidates' probabilities. Suppose that we have already acquired experimental spectra S_1,S_2,... S_m, the addition of a new mass spectrum S' increases the ability to assign the actual glycan and the increase of such ability is calculated as below.

According to the existing experimental spectra S_1, S_2, ..., S_m, the candidate's probabilities were calculated using Formula (2), and the ability of these spectra to distinguish candidate glycans was calculated using entropy of the probability distribution:

$$H(S_1,S_2,...,S_m) = -\sum_{i=1}^{n} P(G_i \mid S_1,S_2,...,S_m)\log P(G_i \mid S_1,S_2,...,S_m).$$

Considering adding the experimental product-ion spectrum S' of a fragment ion, the probabilities of candidate glycans were recalculated and updated as

$$P(G_i, S_1,S_2,...,S_m,S') = \frac{e^{f(G_i, S_1,S_2,...,S_m,S')}}{\sum_{j=1}^{n} e^{f(G_j, S_1,S_2,...,S_m,S')}}.$$

The entropy of the updated probability distribution was recalculated accordingly as

$$H(S_1,S_2,...,S_m,S') = -\sum_{i=1}^{n} P(G_i \mid S_1,S_2,...,S_m,S')\log P(G_i \mid S_1,S_2,...,S_m,S').$$

Finally, DP of the fragment ion is calculated as the entropy decrease

$$DP(S' \mid S_1,S_2,...,S_m) = H(S_1,S_2,...,S_m) - H(S_1,S_2,...,S_m,S') \quad \ldots\ldots (\mathbf{3})$$

which measures the extra information that S' has provided on the basis of the existing experimental spectra S_1, S_2, ..., S_m.

The rational of using the entropy can be clearly illustrated through examining the following two extreme cases of candidates' probabilities: the most even case where all candidate glycans have equal probabilities, and the most uneven case where a specific candidate has a probability of 1 while the probabilities of all others are 0. Intuitively, the more uneven the candidate's probabilities are, the more information the spectra obtained can provide for use in the differentiation from the candidate glycans. A larger entropy decrease gives a higher DP value, indicating a greater contribution of the spectrum S' to the differentiation of candidate glycans.

When the experimental product-ion spectrum of fragment ion *I* is unknown, all possible product-ion spectra of *I* are simulated by computer program and used as substitute of its experimental product-ion spectrum (Figure 2). Based on each possible product-ion spectrum $\tilde{S}_k$ ($k = 1,2,\ldots, t$), the entropy decrease can be similarly calculated, giving the average entropy decrease as

$$DP(I \mid S_1,S_2,\ldots,S_m) = \frac{1}{t}\sum_{k=1}^{t} DP(\tilde{S}_k \mid S_1,S_2,\ldots,S_m) \quad \ldots\ldots (4)$$

The average entropy decrease is used as the DP of fragment ion *I*. Further details of DP calculation is described in Supplementary Methods.

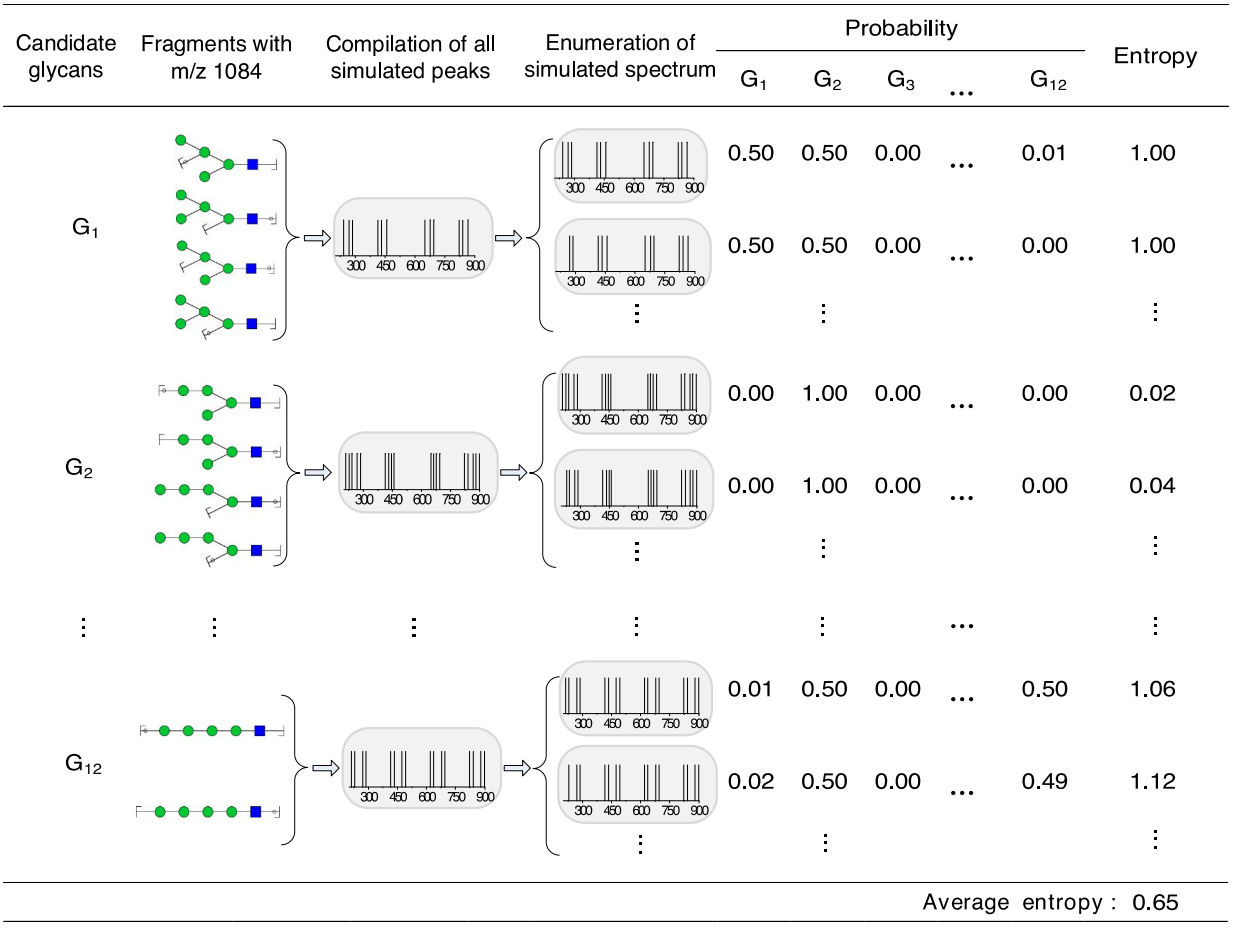


Figure 2. Calculation of average entropy of probability distribution using the ion *m/z* 1084 of Man-5D1 as the example**.**

The average entropy was calculated by three steps: **(1)** Construction of theoretical product-ion spectrum of *m/z* 1084 by computer simulation. From each candidate glycan, we identified the fragments ions with *m/z* 1084 (Column 2). Next, for each fragment, we simulated its fragmentation process during mass spectrometry experiment, and compiled all of its possible product peaks (Column 3). Since each peak might or might not appear in the final product-ion mass spectrum, we enumerated all combinations and obtained multiple possible simulated spectra

(Column 4); **(2)** Based on each possible simulated spectrum, we calculated probabilities of the candidate glycans (Column 5) together with entropy of these probabilities (Column 6); **(3)** Finally, the average entropy of probability distribution given experimental and simulated spectra was calculated as 0.65.

In the case of the assignment of Man-5D1, *m/z* 1084 in the MS^2 spectrum had the highest DP (0.61). The value was obtained by the following calculation. When considering the MS^1 and MS^2 spectra, S_1 and S_2, the probabilities of candidate glycans were calculated with an entropy $H(S_1,S_2) = 1.26$ (further details in Table S-3). The product-ion spectrum *S* of the ion *m/z* 1084 was simulated and candidate probabilities were re-calculated accordingly. As shown in Figure 2, the average entropy of probability was then calculated as $\overline{H(S_1,S_2,S)} = 0.65$. Thus, the entropy decrease $H(S_1,S_2) - \overline{H(S_1,S_2,S)} = 1.26 - 0.65 = 0.61$, and was assigned as the DP of *m/z* 1084.

Identification of Isomeric Glycan Pairs

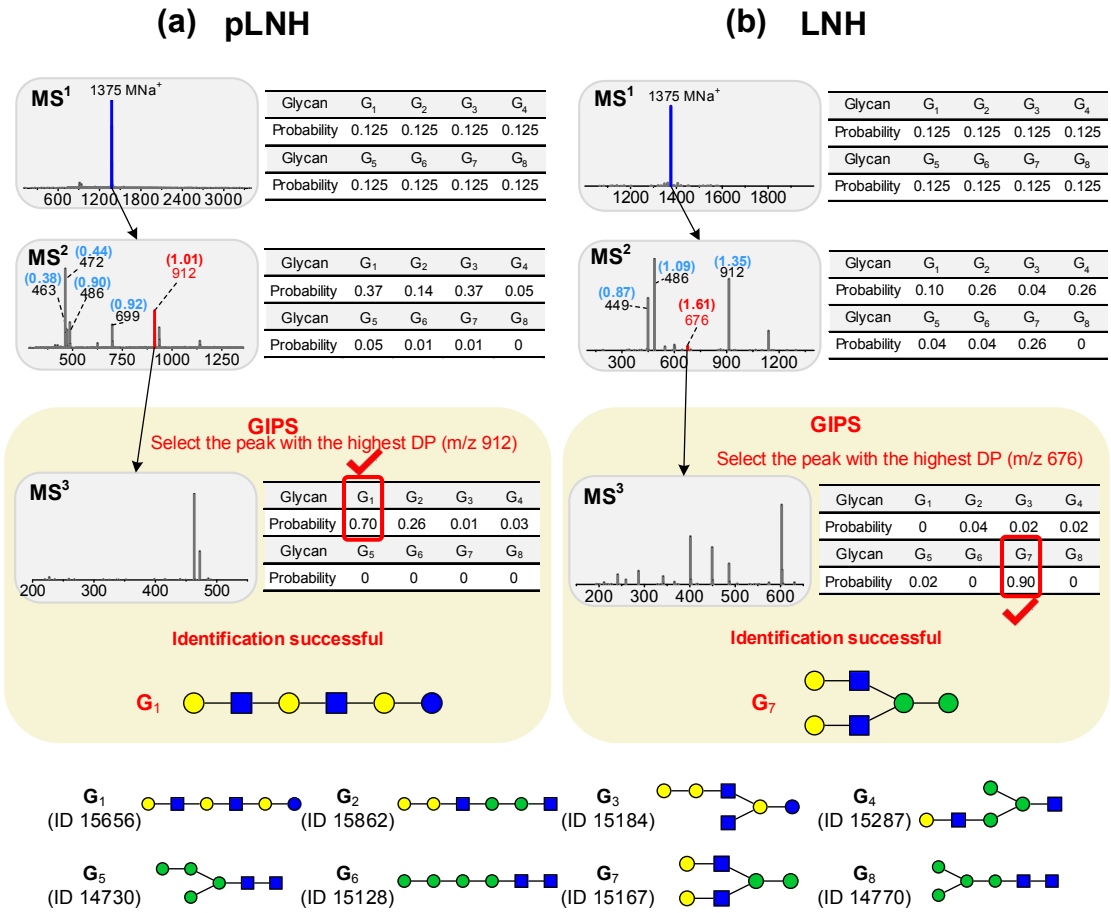


Figure 3. Identification process of the isomeric pair pLNH and LNH using GIPS.

MS^3 was required for identification of pLNH (a), while two rounds of MS^3 was required for LNH (b) from the 8 candidate glycans G_1 to G_8 (structures together with CarbBank ID numbers listed in the box at the bottom). Mass spectrometry process is shown in the left column and the calculated probabilities are in the right column. The fragment ions in red were selected as the precursors for the next round of product-ion scanning. The calculated DPs are in brackets above the m/z values in the spectrum. Note that for LNH, only one round of MS^3 scanning was shown here for simplicity of graphics. The detailed identification process was provided in Figure S-15.

LNH and pLNH (Table S-2) are different in sequence, one is linear and the other is branched, but identical in composition. Here, we assess the ability of the GIPS approach to differentiate these two isomers.

Based on the primary mass spectra with MNa^+ at m/z 1375, a total of 8 candidates were extracted (G_1 to G_8, Figure 3), and each was assigned with an equal probability 0.125 in both cases of pLNH and LNH. The MS^2 spectrum of pLNH led to the recalculated probabilities: the probabilities of G_1 and G_3 increased to 0.37 while others decreased to either 0.14, 0.05 or 0 (Figure 3a). Among the fragment ions in the spectrum of MS^2, m/z 912 had the highest DP

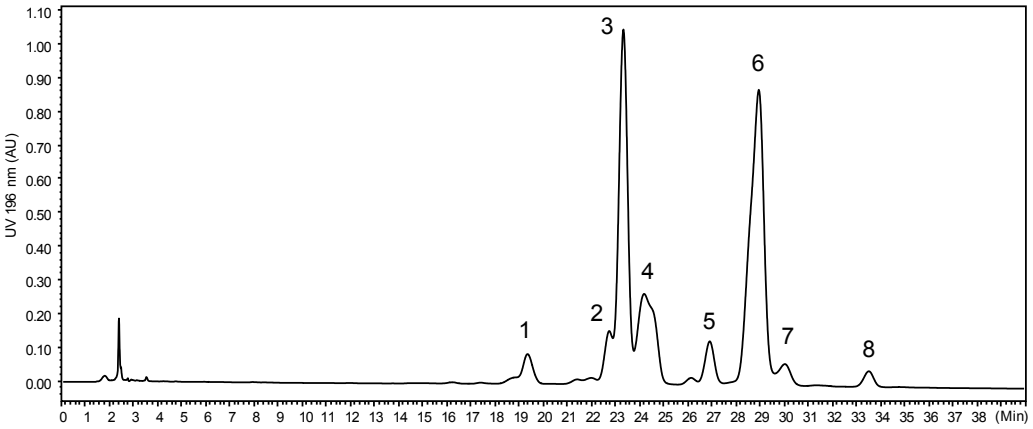
(1.01), and based on which MS^3 was carried out. When the MS^3 spectrum was obtained, further probability calculation indicated a significantly high one for G_1 (0.70), and therefore pLNH (G_1; ID 15656) was correctly reported as the actual glycan.

The identity of the LNH was similarly assigned. Although LNH gave a MS^2 spectrum (Figure 3b) different from that of pLNH, making it readily differentiated from pLNH, more round of scanning was required to assign its structure from the rest 7 candidates (G_2–G_8). Unlike all other glycans used in this study, two rounds of MS^3 scanning were carried out (see Supplementary Results and Figure S-15 for details). The MS^3 spectrum obtained from m/z 676 (with a DP of 1.61) in the MS^2 spectrum as the precursor ion gave the correct assignment with G_7 as the actual structure (further details in Figure S-15 and Supplementary Results). Thus after 3 and 4 rounds of scanning, the identities of pLNH and LNH were successfully determined (Figure 2 and Figure S-15).

Other isomeric pairs, LNFP-I and LNFP-II, and B-Hexa-T2 and Globo-H-Hexa were similarly assigned using this approach (Figure S-3 to S-6).

Analysis of HPLC Fractions of Neutral Human Milk Oligosaccharides

The method was applied to identification of oligosaccharides isolated by normal-phase HPLC from a neutral human milk oligosaccharide fraction which contains variously fucosylated oligosaccharides with hexasaccharide-backbones (Figure 4). Eight subfractions were obtained for



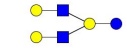
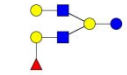
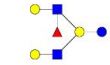
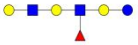
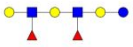
HPLC Fraction	MNa^+ (m/z)	Assignment	Candidate Numbers	Probability MS^2	Probability MS^3	Probability MS^4	Assigned glycan
HP-1	1375		8	0.06	**0.82***		LNH
HP-2	1549		9	0.45	**0.84**		MFLNH-I
HP-3	1549		9	0.13	0.50	**0.98***	MFLNH-III
HP-4	1549		9	**0.96**			MFpLNH-IV
HP-5	1723		5	0.48	**0.84**		DFLNH-a
HP-6	1723		5	0.44	**0.86**		DFLNH-b
HP-7	1723		5	**0.99**			DFpLNH-II
HP-8	1898		2	**0.98**			TFLNH

Figure 4. Separation and structural analysis of a neutral human milk oligosaccharide fraction.

HPLC was carried out on a amide column and eight subfractions were obtained. GIPS was performed to assign the branching patterns and the results were corroborated by ESI-CID-MS/MS and NMR (Supplementary Results and Figure S-16)

analysis by GIPS. Although there are main and minor components and some may not be of high purity due to incomplete resolution (e.g. HPLC-2 and -4), GIPS assigned all eight subfractions. These are of different branching patterns including two linear and six branched sequences with

various fucose residues at different positions (Figure 4). The branching patterns of all fractions were corroborated by the established ESI-MS/MS[34,35] (Table S-4) and two selected ones (HPLC-3 and -5) with high purity were also assigned by ^{1}H-NMR (Supplementary Results and Figure S-16).

Uniqueness of the GIPS Approach

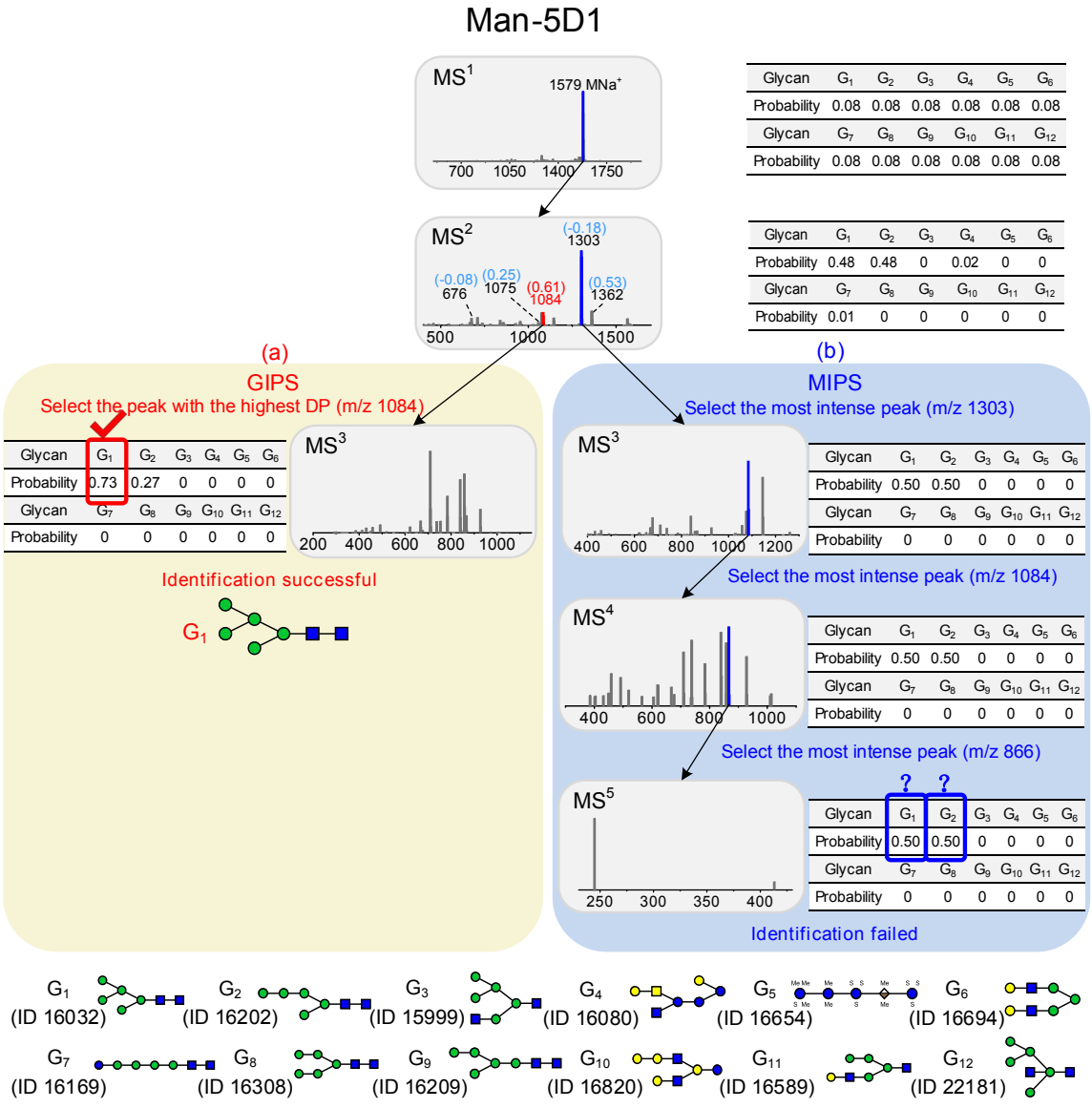


Figure 5. Comparison of GIPS and MIPS approaches using Man-5D1 as the example.

After MS^2, GIPS approach selected *m/z* 1084 with the highest DP (0.65) as the precursor for MS^3 (a) whereas MIPS selected the most intense peak *m/z* 1303 as the precursor MS^3 (b). After MS^3 the calculated probability of the

candidate G_1 became 0.75 by GIPS (a), exceeded the termination threshold (0.70) and Man-5D1 was reported as the actual glycan. However, none of the candidate glycans has a sufficiently high probability after MS^3 by MIPS (b). Further product-ion scanning was required using the most intense peak as the precursor. Even after MS^5 the two candidates, G_1 and G_2, were still indistinguishable by MIPS, leading to failed identification.

We compared our GIPS with the widely-used MIPS strategy. For Man5-D1, the MIPS selected the most intense peak m/z 1303 in the MS^2 spectrum as the precursor to carry out MS^3 (Figure 5b). Although the most intense peak may help to generate a product-ion spectrum of good quality, the spectrum may not be structurally informative. After MS^3, the probabilities of the two candidates G_1 and G_2 only reached 0.50 (Figure 5b), indicating less information in the MS^3 spectrum. However, at the MS^3 stage with the GIPS approach, the probabilities of these two candidates became 0.73 and 0.27, respectively (Figure 5a). GIPS used 3 rounds of scanning to assign Man-5D1 whereas MIPS required further scanning. Even after MS^5, MIPS still cannot distinguish the two candidates as the probabilities remained at 0.5 for both G_1 and G_2.

Using 15 glycans as representatives (Table 1) we compared the results of the two approaches (Figure 5, and Figures S-1 to S-14). The GIPS approach successfully assigned these sequences within a maximum of MS^4. However, the MIPS strategy failed identification of four samples (pLNH, LNH, LNT and Man-5D1). For pLNH, LNH, and LNT, the MIPS strategy could not identify the actual structure after MS^3, and the generated MS^4 spectrum contained only noisy peak. Furthermore, the MIPS strategy could not make a distinction between the isomeric glycan pair pLNH/LNH (Figures S-1 and S-2).

We also examined the reproducibility of the results and possible factors which may affect the GIPS results. The complete identification procedure, including the acquisition of MS^n spectra guided by GIPS, for 9 samples was repeated at least twice (4 or 5 times for selected samples). The results were reproducible (Table S-5). For the present work with oligosaccharide chain length no more than 11 monosaccharide residues, a mass tolerance was set at 0.5 although a value between 0.5-1.0 does not affect the results (Table S-6), and a medium/standard (FWHM 250) mass resolution was used for optimum sensitivity although other resolution settings gave similar results (Table S-7). The collision energy was between 100-400 mV and within this range

the relative intensities of fragment ions varied, but the identification results remained largely the same (Table S-8).

Table 1. Comparison of performance of GIPS and MIPS approaches on 1 glycans.

Glycans	MNa^+ (m/z)	Number of Candidates	Method	Probability of actual glycans			
				MS^2	MS^3	MS^4	MS^5
Man-5D1	1579	12	GIPS	0.49	**0.75**		
			MIPS	0.49	0.50	0.50	0.50 **failed**
LNFP-I	1100	6	GIPS	**0.79**			
			MIPS	**0.79**			
LNFP-II	1100	6	GIPS	0.59	**0.99**		
			MIPS	0.59	**0.98**		
B-Hexa-T2	1304	2	GIPS	**0.99**			
			MIPS	**0.99**			
Globo-H-Hexa	1304	2	GIPS	**0.99**			
			MIPS	**0.99**			
pLNH	1375	8	GIPS	0.64	**0.96**		
			MIPS	0.64	0.66	**failed**	
LNH	1375	8	GIPS	0.27	**0.90**		
			MIPS	0.27	0.42	**failed**	
B-Tetra-T1	896	7	GIPS	0.49	**0.99**		
			MIPS	0.49	**0.99**		
LNT	926	11	GIPS	0.40	**0.84**		
			MIPS	0.40	0.46	**failed**	
MFLNH-III	1550	9	GIPS	0.32	0.50	**0.73**	
			MIPS	0.32	**0.78**		
LNDFH-II	1274	4	GIPS	0.41	**0.78**		
			MIPS	0.41	**0.84**		
SLe[a]	1053	2	GIPS	0.50	0.50	**0.88**	
			MIPS	0.50	0.50	**0.88**	
LST-a	1288	6	GIPS	0.45	**0.86**		
			MIPS	0.45	**0.86**		
MFiLNO	1999	4	GIPS	**0.99**			
			MIPS	**0.99**			
TFiLNO	2347	2	GIPS	**0.98**			
			MIPS	**0.98**			

GIPS successfully identified all 15 glycan standards within a maximum of 4 stages of MS^n experiments. MIPS generally required more stages of MS^n than GIPS and failed for 4 glycan structures (pLNH, LNH, LNT and Man-5D1). In addition, GIPS was able to distinguish all 3 isomeric pairs, whereas MIPS failed with 1 pair (pLNH/LNH). Probability exceeding the pre-defined threshold 0.70 was considered as successful and shown in underlined bold.

CONCLUSIONS

The results presented here for glycan identification by MS^n using the GIPS approach have highlighted the special feature of intelligent precursor selection. The ability to identify glycan branching structures has been clearly demonstrated with the 15 glycan standards and 8 oligosaccharides isolated from a human milk sample within a maximum of four stages of MS^n experiments. Intelligent selection of precursors by experienced specialists is possible but the GIPS method proposed here represents a major step forward towards high-throughput and automated glycomics, and also for use in the non-expert laboratories. Improved sensitivity and reduced time required for analysis are additional advantages, due mainly to fewer rounds of product-ion scanning.

Although in the proof-of-concept study we demonstrated the application of GIPS in glycan branching pattern analysis, the concept, the basic idea, the algorithms and the operation can be extended to elucidation of other structural features. In the present work, to assign branching patterns, only glycosidic cleavages were considered and the fragment ions thus produced contain little linkage information. Some specific double glycosidic cleavage has been shown to be useful for sequence analysis[34] and cross-ring fragments of the A/X types for linkage assignment[16]. In future studies, these will be incorporated into the GIPS approach with potential for sequence and partial linkage analysis. At present, only the m/z values are used but possible extension to include peak intensity may assist in the assignment of linkages and anomeric configurations[13,21]. Identification of isomeric structures has been demonstrated. However, assignment of multiple isomeric structures within a mixture is undoubtedly very challenging. Recent work has indicated such possibility using MS^n by manual assignment in expert laboratories[36].

In the current setting of MS^n, permethylated glycan and positive-ion MALDI were employed and these are identical to the commonly applied methods for glycan profiling. The potentially automated MS^n with the GIPS approach for detailed structural analysis will certainly add a new dimension to the currently widely-used profiling method. The GIPS strategy can also

be readily modified to suit other MS methods (e.g. ESI-MS and negative-ion mode) and glycan derivatives by changing some parameters without much revision of the statistical models.

Our work on glycan structural analysis together with recent methodological development in various aspects of glycomics, e.g. novel chemical method for the release of *O*- and *N*-glycans[37] and N-glycosylation site analysis[38], will undoubtedly contribute to a better understanding of glycan structure and function in health and disease.

ASSOCIATED INFORMATION

Supporting Information

Supplementary methods, Supplementary Results, 8 additional tables, and 16 figures to illustrate in detail the calculation process of GIPS approach as supporting information.

Data and material availability

The software GIPS is available through http://glycan.ict.ac.cn/GIPS and the MS^n spectra are available upon request.

AUTHOR INFORMATION

Author Contributions

YL, SS, CH and DB conceived the study. SS, RC and DB designed the GIPS concept and computational model, and YL, CH and WC designed the mass spectrometry methodology. CH established the MALDI-MS^n and glycan structural analysis procedure and performed and analyzed the mass spectral data. YW, JZ, FG and FY implemented the GIPS approach. YML and JZ carried out glycan preparation and participated in MALDI-MS^n data acquisition, and BM performed NMR experiments and analyzed the data. WC provided critical comments and advice on the GIPS concept and mass spectrometry experiments. SS, CH, WC, YL and DB wrote the manuscript. All authors discussed the results and commented on the manuscript.

Corresponding Author

E-mail: w.chai@imperial.ac.uk, yanli@ibp.ac.cn and dbu@ict.ac.cn.

ORCID

Chuncui Huang: 0000-0002-0713-3186

Wengang Chai: 0000-0003-2977-5347.

Yan Li: 0000-0002-4281-573X

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We would like to thank the National High-Tech Research and Development Project (2014AA021101), Scientific Equipment Development Project of Chinese Academy of Sciences (YZ201249), the National Natural Science Foundation of China (31270834, 61272318, 31270909, 31600650, 31671369, and 31770775), and from March of Dimes research centre grant for providing partial financial supports. NMR work was supported by the Francis Crick Institute through provision of access to the MRC Biomedical NMR Centre.

Figure Legends

Figure 1. Identification process of Man-5D1 using GIPS.

Three rounds of mass spectrometric scanning (a: MS^1; b: MS^2, and c, MS^3) were required to assign Man-5D1 from the 12 candidate glycans G_1 *to* G_{12} (structures together with CarbBank ID numbers listed in the box at the bottom). Left column: mass spectrometry process; right column: calculated probabilities; and middle column: required calculation process. The fragment ion shown in red was selected as the precursor for the next round of product-ion scanning. The calculated DPs are in brackets above the *m/z* values in the spectrum.

Figure 2. Calculation of average entropy of probability distribution using the ion *m/z* 1084 of Man-5D1 as the example.

The average entropy was calculated by three steps: **(1)** Construction of theoretical product-ion spectrum of *m/z* 1084 by computer simulation. From each candidate glycan, we identified the fragments ions with *m/z* 1084 (Column 2). Next, for each fragment, we simulated its fragmentation process during mass spectrometry experiment, and compiled all of its possible product peaks (Column 3). Since each peak might or might not appear in the final product-ion mass spectrum, we enumerated all combinations and obtained multiple possible simulated spectra (Column 4); **(2)** Based on each possible simulated spectrum, we calculated probabilities of the candidate glycans (Column 5) together with entropy of these probabilities (Column 6); **(3)** Finally, the average entropy of probability distribution given experimental and simulated spectra was calculated as 0.65.

Figure 3. Identification process of the isomeric pair pLNH and LNH using GIPS.

MS^3 was required for identification of pLNH (a), while two rounds of MS^3 was required for LNH (b) from the 8 candidate glycans G_1 to G_8 (structures together with CarbBank ID numbers listed in the box at the bottom). Mass spectrometry process is shown in the left column and the

calculated probabilities are in the right column. The fragment ions in red were selected as the precursors for the next round of product-ion scanning. The calculated DPs are in brackets above the m/z values in the spectrum. Note that for LNH, only one round of MS^3 scanning was shown here for simplicity of graphics. The detailed identification process was provided in Figure S-15.

Figure 4. Separation and structural analysis of a neutral human milk oligosaccharide fraction. HPLC was carried out on a amide column and eight subfractions were obtained. GIPS was performed to assign the branching patterns and the results were corroborated by ESI-CID-MS/MS and NMR (Supplementary Results and Figure S-16).

Figure 5. Comparison of GIPS and MIPS approaches using Man-5D1 as the example. After MS^2, GIPS approach selected *m/z* 1084 with the highest DP (0.65) as the precursor for MS^3 (a) whereas MIPS selected the most intense peak *m/z* 1303 as the precursor MS^3 (b). After MS^3 the calculated probability of the candidate G_1 became 0.75 by GIPS (a), exceeded the termination threshold (0.70) and Man-5D1 was reported as the actual glycan. However, none of the candidate glycans has a sufficiently high probability after MS^3 by MIPS (b). Further product-ion scanning was required using the most intense peak as the precursor. Even after MS^5 the two candidates, G_1 and G_2, were still indistinguishable by MIPS, leading to failed identification.

REFERENCES

1. Raman, R.; Raguram, S.; Venkataraman, G.; Paulson, J.C.; Sasisekharan, R. Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat. Methods* **2005**, 2, 817– 824.

2. Rudd, P.-M.; Elliott, T.; Cresswell, P.; Wilson, I.-A.; Dwek, R.-A. Glycosylation and the immune system. *Science* **2001**, 291, 2370–2376.

3. Everest-Dass, A.-V.; Briggs, M.T.; Kaur, G.; Oehler, M.K.; Hoffmann, P.; Packer, N.H. N-glycan MALDI imaging mass spectrometry on formalin fixed paraffin embedded tissue enables the delineation of ovarian cancer tissues. *Mol. Cell. Proteomics* **2016**, 15, 3013–3016.

4. McEver, R.-P. Selectin-carbohydrate interactions during inflammation and metastasis. *Glycoconjugate J.* **1997**, 14, 585–591.

5. Taniguchi, N.; Korekane, H. Branched N-glycans and their implications for cell adhesion, signaling and clinical applications for cancer biomarkers and in the rapeutics. *BMB Rep.* **2011**, 44, 772–781.

6. Smith, D.-F.; Cummings, R.-D. Application of microarrays for deciphering the structure and function of the human glycome. *Mol. Cell. Proteomics* **2013**, 12, 902–912.

7. Ruhaak, L.-R.; Lebrilla, C.-B. Advances in analysis of human milk oligosaccharides. *Adv. Nutr.* **2012**, 3, 406S–414S.

8. North, S.-J.; Huang, H.H.; Sundaram, S.; Jang-Lee, J.; Etienne, A.T.; Trollope, A.; Chalabi, S.; Dell, A.; Stanley, P.; Haslam, S.M. Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. *J. Biol. Chem.* **2010**, 285, 5759–5775.

9. Yu, S.-Y.; Wu, S.-W.; Khoo, K.-H. Distinctive characteristics of MALDI-Q/TOF and TOF/TOF tandem mass spectrometry for sequencing of permethylated complex type N-glycans. *Glycoconjugate J.* **2006**, 23, 355–369.

10. Pang, P. C.; Chiu, P. C.; Lee, C. L.; Chang, L. Y.; Panico, M.; Morris, H. R.; Haslam, S. M.; Khoo, K. H.; Clark, G. F.; Yeung, W. S.; Dell, A. Human sperm binding is mediated by the sialyl-Lewisx oligosaccharide on the Zona Pellucida. *Science* **2011,** 333, 1761–1764.

11. Yu, G.; Zhao, X.; Yang, B.; Ren, S.; Guan, H.; Zhang, Y.; Lawson, A.M.; Chai, W. Sequence determination of sulfated carrageenan-derived oligosaccharides by high-sensitivity negative-ion electrospray tandem mass spectrometry. *Anal. Chem.* **2006**, 78, 8499–8505.

12. Both, P. Discrimination of epimeric glycans and glycopeptides using IM-MS and its potential for carbohydrate sequencing. *Nat. Chem.* **2014**, 6, 65–74.

13. Chai, W.; Piskarev, V.E.; Mulloy, B.; Liu, Y.; Evans, P.G.; Osborn, H.M.; Lawson, A.M. Analysis of chain and blood group type and branching pattern sialylated oligosaccharides by negative ion electrospray tandem mass spectrometry. *Anal. Chem.* **2006**, 78, 1581–1592.

14. Gao, C.; Zhang, Y.; Liu, Y.; Feizi, T.; Chai, W. Negative-ion electrospray tandem mass spectrometry and microarray analyses of developmentally regulated antigens based on Type 1 and Type 2 backbone sequences. *Anal. Chem.* **2015**, 87, 11871−11878.

15. Domon, B.; Costello, C. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate J.* **1988,** 5, 397–409.

16. Palma, A. S.; Liu, Y.; Zhang, Y.; McCleary, B.V.; Huang, Q.; Guidolin, A.E.; Torosantucci, A.; Wang, D.; Carvalho, A.L.; Fontes, C.M.; Mulloy, B.; Childs, R.A.; Feizi, T.; Chai, W. Unravelling glucan recognition systems by glycome microarrays using the designer approach and mass spectrometry. *Mol. Cell. Proteomics* **2015**, 14, 974–988.

17. Walsh, I.; Zhao, S.; Campbell, M.; Taron, C.-H.; Rudd, P.-M. Quantitative profiling of glycans and glycopeptides: an informatics' perspective. *Curr. Opin. Struct. Biol.* **2016**, 40, 70–80.

18. Hu, H.; Khatrl, K.; Zaia, J. Algorithms and design strategies towards automated glycoproteomics analysis. *Mass Spectrom. Rev.* **2017**, 36, 475–498.

19. Morimoto, K.; Nishikaze, T.; Yoshizawa, A.C.; Kajihara, S.; Aoshima, K.; Oda, Y.; Tanaka, K. Glycan Analysis Plug-in: a database search tool for N-glycan structures using mass spectrometry. *Bioinformatics*, **2015**, 31, 2217–2219.

20. Maass, K.; Ranzinger, R.; Geyer, H.; von der Lieth, C.-W.; Geyer, R. ''Glyco-peakfinder'' – de novo composition analysis of glycoconjugates. *Proteomics* **2007**, 7, 4435–4444.

21. Ashline, D.-J.; Singh, S.; Hanneman, A.; Reinhold, V. Congruent strategies for carbohydrate sequencing. 1. mining structural details by MS^n. *Anal. Chem.* **2005**, 77, 6250–6262.

22. Ashline, D.J.; Lapadula, A.J.; Liu, Y.H.; Lin, M.; Grace, M.; Pramanik, B.; Reinhold, V.N. Carbohydrate structural isomers analyzed by sequential mass spectrometry. *Anal. Chem.* **2007**, 79, 3830–3842.

23. Reinhold, V.; Zhang, H.; Hanneman, A.; Ashline, D. Toward a platform for comprehensive glycan sequencing. *Mol. Cell. Proteomics* **2013**, 12, 866–873.

24. Zaia, J. Mass spectrometry and the emerging field of glycomics. *Chem. Bio.* **2008**, 15, 881–892.

25. Ashline, D.-J.; Zhang, H.; Reinhold, V.-N. Isomeric complexity of glycosylation documented by MS^n. *Anal. Bioanal. Chem.* **2017**, 409, 439–451.

26. Lapadula, A.J.; Hatcher, P.J.; Hanneman, A.J.; Ashline, D.J.; Zhang, H.; Reinhold, V.N. Congruent strategies for carbohydrate sequencing. 3. OSCAR: An algorithm for assigning oligosaccharide topology from MS^n Data. *Anal. Chem.* **2005**, 77, 6271–6279.

27. Zhang, H.; Singh, S.; Reinhold, V. Congruent strategies for carbohydrate sequencing. 2. FragLib: an MS^n spectral library. *Anal. Chem.* **2005**, 77, 6263–6270.

28. Kameyama, A.; Kikuchi, N.; Nakaya, S.; Ito, H.; Sato, T.; Shikanai, T.; Takahashi, Y.; Takahashi, K.; Narimatsu, H. A strategy for identification of oligosaccharide structures using observational multistage mass spectral library. *Anal. Chem.* **2005**, 77, 4719–4725.

29. Hashimoto, K.; Goto, S.; Kawano, S.; Aoki-Kinoshita, K.F.; Ueda, N.; Hamajima, M.; Kawasaki, T.; and Kanehisa, M. KEGG as a glycome informatics resource. *Glycobiology* **2006**, 16, 63R–70R.

30. Ranzinger, R.; Herget, S.; von der Lieth, C.-W.; Frank, M. GlycomeDB–a unified database for carbohydrate structures. *Nucleic Acids Res.* **2011**, 39, D373–D376.

31. von der Lieth, C. W.; Freire, A. A.; Blank, D.; Campbell, M. P.; Ceroni, A.; Damerell, D. R.; Dell, A.; Dwek, R. A.; Ernst, B.; Fogh, R.; Frank, M.; Geyer, H.; Geyer, R.; Harrison, M. J.; Henrick, K.; Herget, S.; Hull, W. E.; Ionides, J.; Joshi, H. J.; Kamerling, J. P.; Leeflang, B. R.; Lutteke, T.; Lundborg, M.; Maass, K.; Merry, A.; Ranzinger, R.; Rosen, J.; Royle, L.; Rudd, P. M.; Schloissnig, S.; Stenutz, R.; Vranken, W. F.; Widmalm, G.; Haslam, S. M. EUROCarbDB: An open-access platform for glycoinformatics. *Glycobiology* **2011**, 493–502.

32. Doubet, S.; Albersheim, P. Letter to the Glyco-Forum. *Glycobiology* **1992**, 2, 505–507.

33. Doubet, S.; Bock, K.; Smith, D.; Darvill, A.; Albersheim, P. The complex carbohydrate structure database. *Trends Biochem. Sci.* **1989**, 14, 475–477.

34. Chai, W.; Piskarev, V.; Lawson A. M. Negative-ion electrospray mass spectrometry of neutral underivatized oligosaccharides. *Anal. Chem.* **2001**, 73, 651–657.

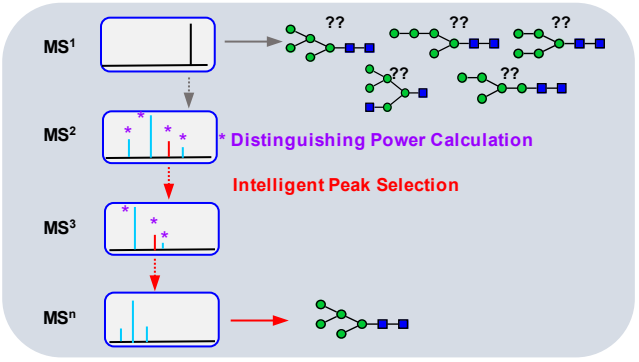
35. Chai, W.; Piskarev, V.; Lawson A. M. J. Branching pattern and sequence analysis of underivatized oligosaccharides by combined MS/MS of singly and doubly charged molecular ions in negative-ion electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2002**, 13, 670–679.

36. Ashline, D.J.; Yu, Y.; Lasanajak, Y.; Song, X.; Hu, L.; Ramani, S.; Prasad, V.; Estes, M.K.; Cummings, R.D.; Smith, D.F.; Reinhold, V.N. Structural characterization by multistage mass spectrometry (MS^n) of human milk glycans recognized by human rotaviruses. *Mol. Cell. Proteomics* **2014**, 13, 2961–2974.

37. Song, X.; Ju, H.; Lasanajak,Y.; Smith, D.F.; Cummings, R. D. Oxidative release of natural glycans for functional glycomics. *Nat. Methods* **2016**, 13, 528–534.

38. Sun, S.; Shah, P.; Eshghi, S. T.; Yang, W.; Trikannad, N.; Yang, S.; Chen, L.; Aiyetan, P.; Hoti, N.; Zhang, Z.; Chan, D. W.; Zhang, H. Comprehensive analysis of protein glycosylation by solid-phase extraction of *N*-linked glycans and glycosite-containing peptides. *Nat. biotech.* **2016,** 34, 84–88.

for TOC only


MS1
??
??
??
??
??
MS2
* Distinguishing Power Calculation
Intelligent Peak Selection
MS3
MSn