

# Congruent Strategies for Carbohydrate Sequencing. 3. OSCAR: An Algorithm for Assigning Oligosaccharide Topology from MS<sup>n</sup> Data

Anthony J. Lapadula,<sup>†,‡</sup> Philip J. Hatcher,<sup>‡</sup> Andy J. Hanneman,<sup>†</sup> David J. Ashline,<sup>†</sup> Hailong Zhang,<sup>†</sup> and Vernon N. Reinhold<sup>\*,†</sup>

Departments of Chemistry and Computer Science, University of New Hampshire Durham, New Hampshire 03824

This is the third in a sequence of reports devoted to the development of congruent strategies for carbohydrate sequencing. Two previous reports outlined the strategies for observing structural detail from MS<sup>n</sup> data and introduced tools that compile, search, and compare fragment spectra in a bottom-up approach to oligosaccharide sequencing. In this third report, we introduce the operational details of an algorithm that we define as the Oligosaccharide Subtree Constraint Algorithm (OSCAR). This algorithm assimilates analyst-selected MS<sup>n</sup> ion fragmentation pathways into oligosaccharide topology (branching and linkage) using what may be considered a top-down sequencing strategy. Guided by a series of logical constraints, this de novo algorithm provides molecular topology without presumed biosynthetic constraints or external comparisons. In this introductory study, OSCAR is applied to a series of permethylated oligomers and isomeric glycans, and topologies are assigned in a few hundredths of a second.

The two preceding reports in this series introduced methodologies for determining the structural details of carbohydrates from MS<sup>n</sup> data<sup>1</sup> and described tools that extract further details through spectral comparisons against known oligosaccharide fragments.<sup>2</sup> Glycobiology in general has been a fertile area for the development of bioinformatics tools,<sup>3</sup> and a variety of computer programs are now available to support carbohydrate analysis using mass spectrometry data. The web-based tool GlycoMod<sup>4</sup> accepts a glycan MS mass and returns a list of possible compositions, using literature-derived constraints to limit its output. In contrast, StrOligo<sup>5,6</sup> examines oligosaccharide MS<sup>2</sup> spectra to propose a set of candidate structures restricted by biosynthetic constraints.

The candidates are fragmented in silico, and the resulting simulated spectra are ranked against the experimental MS<sup>2</sup> spectrum. Similarly, GlycosidIQ<sup>7</sup> compares an experimental MS<sup>2</sup> spectrum against simulated spectra generated from the contents of GlycoSuiteDB,<sup>8</sup> a curated database of known structures, to produce a ranked list of candidate glycans.

In the catalog library method,<sup>9,10</sup> a catalog contains the characteristic fragmentation patterns of substructures isolated from a library of known oligosaccharides. Total structure assignment is accomplished by matching observed fragmentation patterns with the catalog motif entries.

The web-based Saccharide Topology Analysis Tool (STAT) accepts a native glycan mass as input, infers the glycan's possible compositions over a wide range of monosaccharide residues, and generates a candidate set of all possible branching topologies.<sup>11,12</sup> STAT then accepts a list of ion masses extracted from the MS<sup>n</sup> data tree, where each ion must form a connected substructure within the candidate topologies. Finally, a ranked list of branching topologies is generated. STAT supports native glycans, requires manual intervention to resolve ambiguous ion compositions, and restricts the branch points for N-linked glycans to the core nonreducing termini. STAT executes nearly instantaneously for octasaccharides and in under 1 min for nonasaccharides, with execution time increasing exponentially for higher glycans.

In this report, we describe an algorithm for the assignment of glycan topology called OSCAR, the Oligosaccharide Subtree Constraint Algorithm. OSCAR couples changes in ion composition with continuity relationships for each step of an MS<sup>n</sup> disassembly pathway. Given structurally informative fragmentation pathways, this stepwise-relational process guides the output to a single structure from multiple possibilities. In practice, the topologies proposed are comparable to what experts can deduce without invoking biological insights or requiring data obtained from other analytical methodologies. The algorithm is de novo because it

\* Corresponding author. E-mail: vnr@unh.edu. Tel.: (603) 862-2527.

<sup>†</sup> Department of Chemistry.

<sup>‡</sup> Department of Computer Science.

- (1) Ashline, D.; Singh, S.; Hanneman, A.; Reinhold, V. *Anal. Chem.* **2005**, *77*, 6250–6262. ac050724z
- (2) Zhang, H.; Singh, S.; Reinhold, V. *Anal. Chem.* **2005**, *77*, 6263–6270. ac050725r
- (3) Marchal, I.; Golfier, G.; Dugas, O.; Majed, M. *Biochimie* **2003**, *85*, 75–81.
- (4) Cooper, C. A.; Gasteiger, E.; Packer, N. H. *Proteomics* **2001**, *1*, 340–349.
- (5) Ethier, M.; Saba, J. A.; Ens, W.; Standing, K. G.; Perreault, H. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1743–1754.
- (6) Ethier, M.; Saba, J. A.; Spearman, M.; Krokhn, O.; Butler, M.; Ens, W.; Standing, K. G.; Perreault, H. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2713–2720.

- (7) Joshi, H. J.; Harrison, M. J.; Schulz, B. L.; Cooper, C. A.; Packer, N. H.; Karlsson, N. G. *Proteomics* **2004**, *4*, 1650–1664.
- (8) Cooper, C. A.; Joshi, H. J.; Harrison, M. J.; Wilkins, M. R.; Packer, N. H. *Nucleic Acids Res.* **2003**, *31*, 511–513.
- (9) Tseng, K.; Hedrick, J. L.; Lebrilla, C. B. *Anal. Chem.* **1999**, *71*, 3747–3754.
- (10) Tseng, K.; Xie, Y.; Seeley, J.; Hedrick, J. L.; Lebrilla, C. B. *Glycoconjugate J.* **2001**, *18*, 309–320.
- (11) Gaucher, S. P.; Morrow, J.; Leary, J. A. *Anal. Chem.* **2000**, *72*, 2331–2336.
- (12) Leavell, M. D.; Leary, J. A.; Yamasaki, R. J. *Am. Soc. Mass Spectrom.* **2002**, *13*, 571–576.

neither presumes biosynthetic constraints nor compares unknowns against previously characterized oligosaccharides.

## METHODS/EXPERIMENTAL SECTION

**Data Collection.** *Caenorhabditis elegans* glycan samples were prepared as described.<sup>1,13</sup> The bovine brain gangliosides GM1a and GM1b were purchased from Sigma (G-2375). MS<sup>n</sup> spectra were collected by ESI-LTQ (ThermoFinnigan, San Jose, CA) and MALDI-QIT-TOF (Kratos/Shimadzu, Kyoto, Japan).

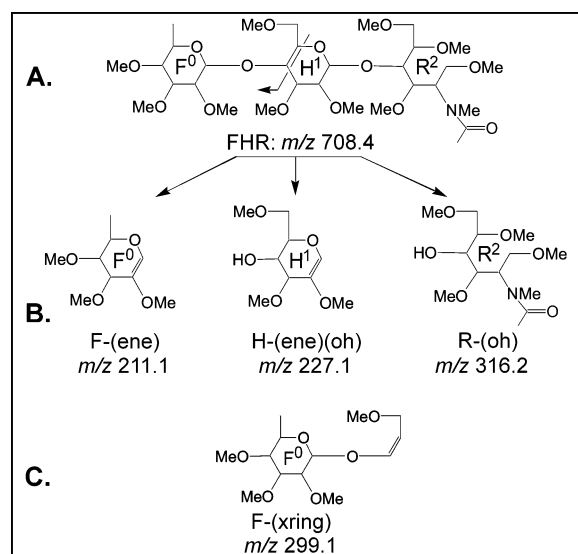
**Mining Details from Mass Spectral Data.** Fragments generated during MS<sup>n</sup> disassembly of permethylated glycans retain telltale features of their oligomer positions (branching and linkage) in the form of mass shifts.<sup>1</sup> The most abundant ions are typically produced by glycosidic bond rupture. At lower abundance, a second set of ions are a consequence of rupture across a monomer ring, thereby retaining details of interresidue linkage. Since these latter fragments vary with linkage position, the mass shifts characterize linkage type. Thus, the mass of a fragment ion is a summation of cleavage events but does not directly specify the fragment's origin or internal structure; OSCAR computes glycan topology by reassembling these analyst-selected fragments.

**Fragment Composition Database and Notation.** With this understanding, we have created a database of oligosaccharide fragment compositions for the common monomers Hex, HexNAc, dHex, and Neu5Ac. When probed with a sodiated ion mass, this in silico composition database returns a list of possible fragment compositions. Because stereoisomers cannot be differentiated by mass, the database labels Hex as H, HexNAc as N, dHex (usually fucose) as F, and Neu5Ac as S. An optionally reduced HexNAc at the glycan's reducing end is labeled R. The database can easily be modified to accommodate additional monomers (e.g., pentoses, uronic acids and *N*-glycolylneuraminic acid), different metal ion adducts (e.g., lithium), and attached phosphate and sulfate groups.

Because conventional nomenclature<sup>14</sup> can be somewhat unwieldy when applied to the complex fragments generated by multiple steps of MS<sup>n</sup> disassembly, we introduce here a simplified composition notation: cleavages that result in an open hydroxyl are abbreviated "(oh)" and those that result in a double-bond are abbreviated "(ene)." Scheme 1A shows a hypothetical reduced and permethylated trisaccharide with composition FHR; Scheme 1B shows three fragments that might arise during MS<sup>n</sup> disassembly, with the compositions F-(ene), H-(ene)(oh), and R-(oh), respectively. A hexose with three open hydroxyl cleavages would be denoted H-(oh)<sub>3</sub>. Scheme 1C shows one possible cross-ring fragment. For clarity when attention must be paid to particular residues, we label each monomer with a superscript (F<sup>0</sup>, H<sup>1</sup>, and R<sup>2</sup>). Labels begin with zero, as is common in computer science.

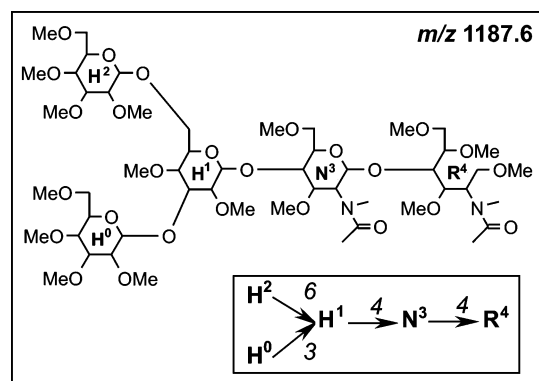
The composition database currently contains 12 378 entries for unfragmented glycans, representing all compositions containing up to 12 H residues, 12 N, 5 F, 5 S, and 1 R. Total glycan size is limited to 20 residues. The database also contains 4 542 720 entries for glycan fragments containing from one to five cleavages, with combinations of cleavages applied to each unfragmented composition. Sixty different cross-ring cleavages are used here, as are (ene) and (oh) cleavages. The database size is ~90 MB and is built in less than 1 min on a 2.66-GHz Pentium 4 laptop PC with 512 MB of memory.

## Scheme 1<sup>a</sup>



<sup>a</sup> (A) Hypothetical trisaccharide with the composition FHR. (B) Three fragments that might arise during MS<sup>n</sup> disassembly. (C) One possible cross-ring fragment. OSCAR computes glycan topology by reassembling fragments such as these.

## Scheme 2. Topology of the Reduced, Permethylated Man<sub>3</sub>GlcNAc<sub>2</sub> Glycan ( $m/z$ 1187.6)<sup>a</sup>



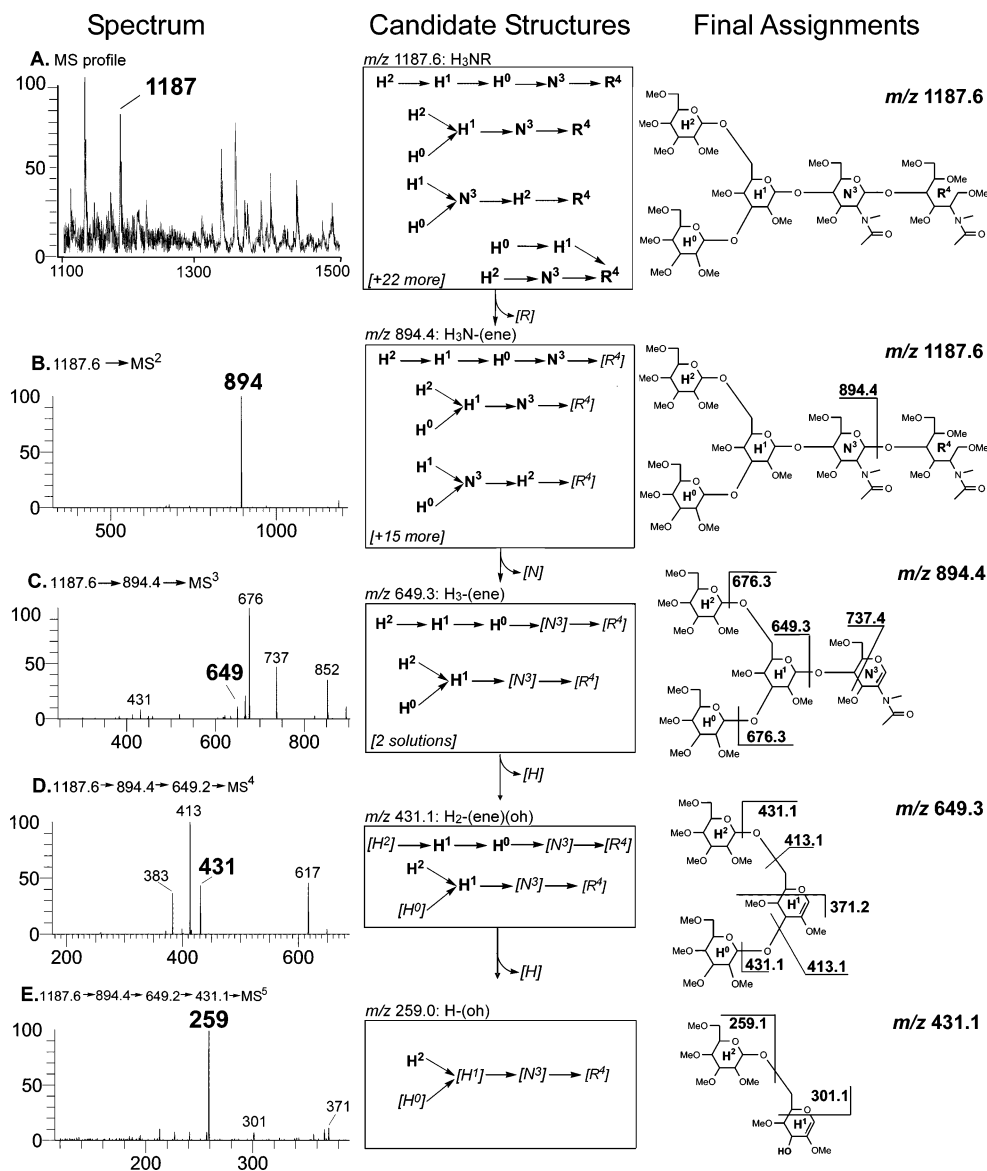
<sup>a</sup> Inset is a tree representation of the glycan where italicized numbers indicate linkage position. The absence of notation for linkage position means undetermined.

**Algorithm Overview.** The input data for the algorithm is a series of fragmentation pathways obtained from MS<sup>n</sup> analysis of permethylated oligosaccharides. Aided by the composition database, the algorithm fits pathway ions to compositions that satisfy logical structural constraints and provides as output a set of one or more structures that satisfy all the selected pathways. Given structurally informative pathways, the output will converge to a single topology. The algorithm does not search a database of reported structures and in this sense is de novo. OSCAR is designed to be a high-performance algorithm; all execution times reported are from the same laptop used to generate the composition database.

**Algorithm Details.** Given a glycan MS mass, the algorithm retrieves the matching unfragmented ion composition from the database and constructs an optimized representation of all possible glycan topologies fitting that composition. These are the candidate structures. For each ion in the provided MS<sup>n</sup> fragmentation

(13) Hanneman, A.; Reinhold, V. *Glycobiology* **2003**, *13*, 899–900.

(14) Domon, B.; Costello, C. E. *Glycoconjugate J.* **1988**, *5*, 397.



**Figure 1.** MS<sup>n</sup> disassembly of the N-linked glycan in Scheme 2 along the pathway  $m/z$  1187.6 → 894.4 → 649.2 → 431.1 → 259.0. The left-hand column shows the spectrum for each step, the middle column shows the isomeric candidate structures reported by OSCAR at each step, and the right-hand column contains the final assignments of selected ions. Boldface type substructures match the product ion composition introduced at each step; lost residues are shown italicized in brackets.

### Chart 1. Input Listing<sup>a</sup>

```
-RootIsR      ; Assume that R is the glycan's reducing end residue
AddPathway 1187.6_894.4_649.2_431.1_259.0
Summarize
```

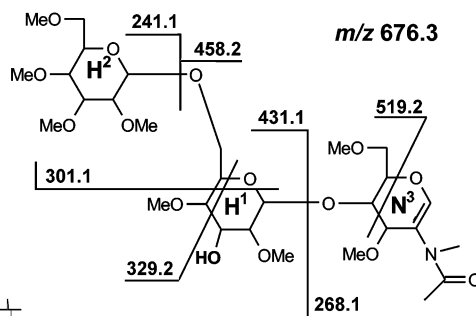
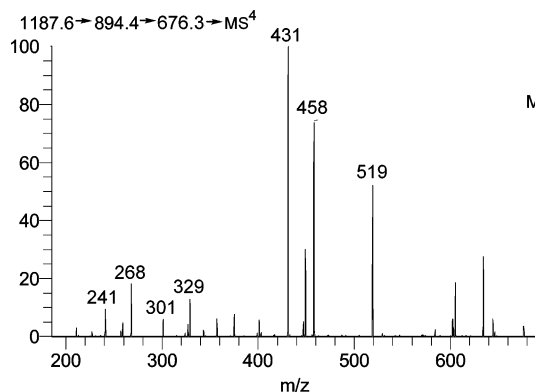
<sup>a</sup> Input to compute the branching of the glycan in Figure 1 in 0.01 s. The option “-RootIsR” instructs the algorithm that the reducing end residue of the glycan is R, and the semicolon character introduces a human-readable comment that is ignored by the program. The command “AddPathway” adds the given MS<sup>n</sup> fragmentation pathway to the set of constraints considered by the algorithm. The command “Summarize” instructs the program to output the set of glycans that are consistent with the constraints imposed by all selected pathways.

pathways, a series of structural inference rules are applied that constrain the candidate glycan set using the branching structure of glycans to eliminate “impossible” structures and retain only those that are consistent with all input pathways. For example, if a selected ion reveals a terminal hexose residue, then all candidate structures having that hexose in a nonterminal position are eliminated. Similarly, a fragment bearing three cleavages eliminates all candidates lacking branching. The algorithm iteratively applies over 50 inference rules, thereby restricting the candidate set to contain only those structures consistent with all the selected

fragmentation pathways. A full description of the data structures and inference rules used by OSCAR is available.<sup>15</sup>

The algorithm evaluates many possible solutions in parallel and discards those that generate no valid structures. With this generic de novo approach, opportunities arise to propose topologies not previously reported. Moreover, the algorithm’s design allows simple changes to cover new monomers or the possibility

(15) Lapadula, A. J. Technical report, Department of Computer Science, University of New Hampshire, Durham, NH, 03824, <http://www.cs.unh.edu/~pjh/bioinformatics/glycomics/GlySpy-and-OSCAR.pdf>, 2004.



**Figure 2.** Spectrum and fragmentation scheme for the ion  $m/z$  676.3 from the pathway  $m/z$  1187.6  $\rightarrow$  894.4  $\rightarrow$  676.3. Ions  $m/z$  301.1, 329.2, and 519.2 are cross-ring fragments that indicate linkage positions.

### Chart 2. Input Listing<sup>a</sup>

```
-RootIsR      ; Assume that R is the glycan's reducing-end residue

AddPathway 1187.6_894.4_649.2_431.1_259.0      ; Same pathway to define branching

; The following pathways use cross-ring fragments to assign partial linkage
AddPathway 1187.6_894.4_737.3                  ; H¹ (1-4/6)N³
AddPathway 1187.6_894.4_676.3_329.1           ; H² (1-4/6)H¹
AddPathway 1187.6_894.4_649.2_431.1_371.1     ; H⁰ (1-2/3/4)H¹

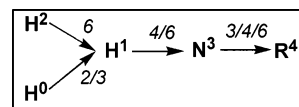
Summarize
```

<sup>a</sup> Input to compute the branching and partial interresidue linkage of the glycan in Figure 1 in 0.01 s. Multiple fragmentation pathways are used to impose additional structural constraints, resulting in improved structural characterization.

of introducing different cross-ring ion fragments to enhance linkage specificity. Although the program applies many inference rules to eliminate candidate structures, here we introduce two of the most powerful: the Connected Substructure constraint and the Precursor/Product constraint.

**Connected Substructure Constraint.** The composition of each input ion must match a connected substructure embedded within each candidate structure. If a candidate does not contain the substructure, it is removed from the candidate set. For example, given the N-linked glycan core structure (Scheme 2), the composition  $H_3\text{(oh)}$  is satisfied by the connected substructure  $H^0/H^1/H^2$ . However, the composition  $HR\text{(oh)}$  cannot be satisfied by this glycan because the intermediate residue  $N^3$  prevents extraction of the connected disaccharide fragment  $HR$ . Other compositions satisfied by this glycan include  $H\text{(oh)}$  (satisfied by  $H^0$  or  $H^2$ ),  $HN\text{(oh)}_3$  (satisfied by  $H^1/N^3$ ), and  $R\text{(oh)}$  (satisfied by  $R^4$ ); in contrast, the compositions  $H\text{(oh)}_2$ ,  $HN\text{(oh)}$ , and  $R\text{(oh)}_2$  cannot be satisfied by this glycan.

**Precursor/Product Constraint.** Tracking changes from precursor to product ions along fragmentation pathways is a unique aspect of the algorithm and is perhaps its strongest attribute. OSCAR utilizes the precursor/product relationship in an important way: product substructures are assigned within precursor structures. In Scheme 2, assuming OSCAR has assigned the composition  $HN\text{(oh)}_3$  to residues  $H^1/N^3$  and further fragmentation produces an ion with the composition  $H\text{(oh)}_3$ , that product ion must be assigned to residue  $H^1$ ; residues  $H^0$  and  $H^2$  are not considered because they are not part of the  $H^1/N^3$  precursor structure. This relationship illustrates how  $MS^n$  disassembly provides evidence regarding substructure location, with each product assignment made within the boundaries of its precursor.



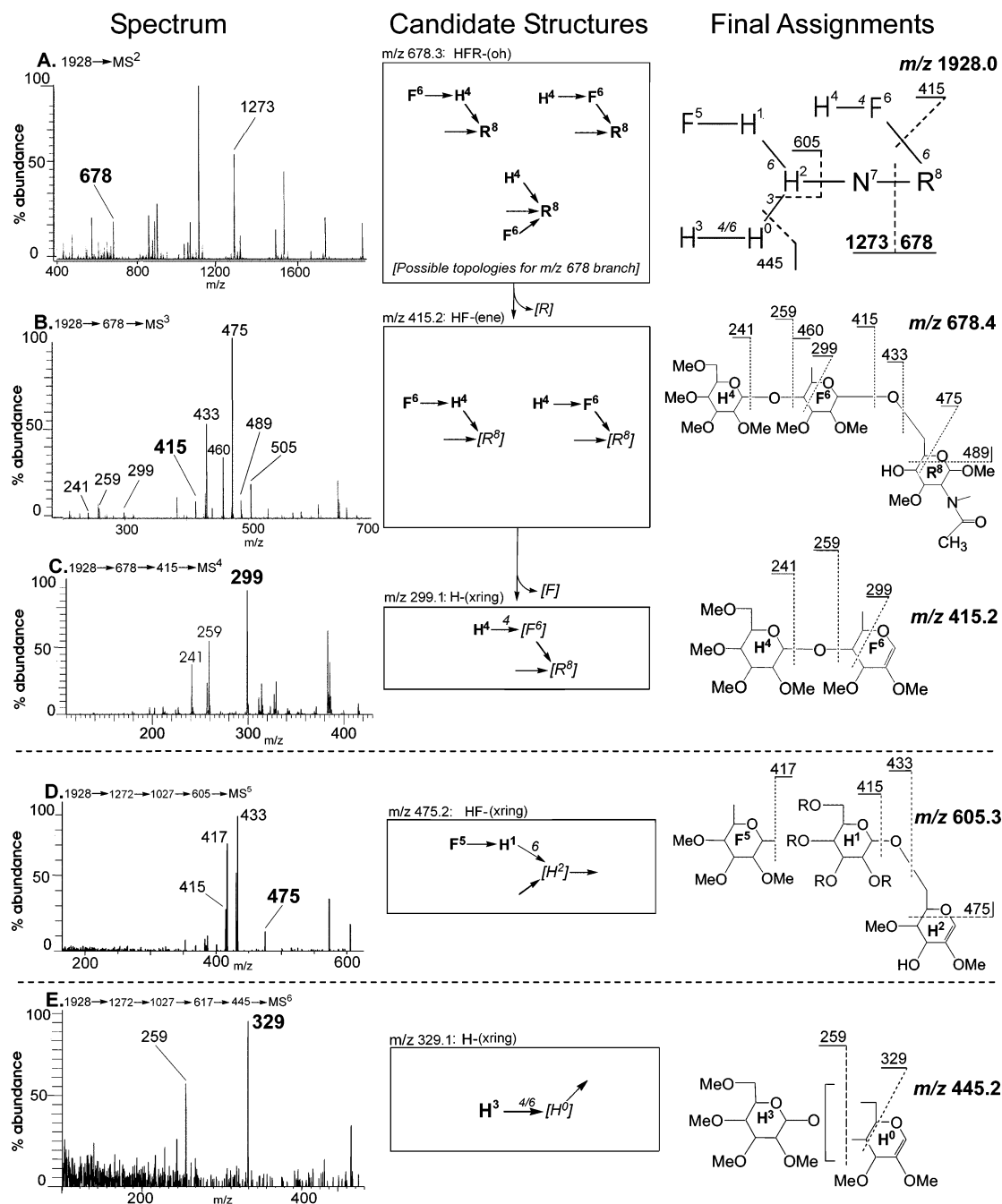
**Figure 3.** OSCAR's computed branching and linkage for the glycan in Figure 1 given the input shown in Chart 2. This analysis completes in 0.01 s.

### RESULTS AND DISCUSSION

To illustrate detailed operation of the algorithm, we present two N-linked glycans of increasing complexity, followed by an analysis of the isomeric carbohydrate moieties obtained from the bovine brain gangliosides GM1a and GM1b. In the first N-linked glycan (Scheme 2), each of the five monosaccharide residues is labeled with its monomer class, H, N, or R, and an identifying superscript from 0 to 4. OSCAR uses a classical computer science-based tree notation<sup>15</sup> to represent the glycan, where the interresidue linkages are represented in italics (Scheme 2, inset).

**Defining Branching.** The N-linked core structure  $\text{Man}_3\text{GlcNAc}_2$  is used to demonstrate the principles of  $MS^n$  connectivity and the assistance provided by OSCAR for rapidly assigning topology. The ion  $m/z$  1187.6, fitting the composition  $H_3NR$ , was detected in a glycan profile of *C. elegans* and disassembled along the pathway  $m/z$  1187.6  $\rightarrow$  894.4  $\rightarrow$  649.2  $\rightarrow$  431.1  $\rightarrow$  259.0. The spectrum for each  $MS^n$  step appears in the left-hand column of Figure 1; final assignments for many ions in the spectra are shown in the right-hand column; the center column tracks the shrinking candidate set as the algorithm processes the pathway. This simple example demonstrates analysis of a single fragmentation pathway; the more complicated structures presented later will require multiple pathways.

To begin, 26 isobaric candidate structures match the initial composition (Figure 1A, center column). OSCAR starts with no



**Figure 4.** Sequential  $MS^n$  disassembly of the glycan  $H_5F_2NR$  ( $m/z$  1928.0). Given only the ion  $m/z$  1928.0 as input, OSCAR produces 1006 candidate N-linked structures. Steps A–C show sequential disassembly of the reducing end  $m/z$  678 branch to determine branching and complete linkage of the  $H^4/F^6/R^8$  branch. Pathway D positions the terminal residue  $F^5$ . Pathway E reveals partial linkage of the terminal  $H^3$  residue.

structural prejudice or insight, and this large number of isomeric structures for a pentasaccharide is expected. In subsequent panels, the moieties lost are retained in the candidate structures for clarity, but are shown in brackets.

The first disassembly step yields the major product ion  $m/z$  894.4 ( $MS^2$ , Figure 1B), which the composition database identifies as  $H_3N$ -(ene). Applying the connected substructure constraint, OSCAR eliminates all candidates that fail to contain a connected substructure with the composition  $H_3N$ -(ene). Figure 1B (center) shows that 18 structures meet this constraint, with 8 having been eliminated.

In the subsequent step ( $MS^3$ , Figure 1C), the spectrum provides multiple product ions. The selected ion,  $m/z$  649.2, has

the composition  $H_3$ -(ene). Again applying the connected substructure constraint, OSCAR retains the only two structures, one linear and one branched, that embed a substructure with this composition (Figure 1C, center). Thus, in two disassembly steps, the algorithm has identified 2 possible structures and eliminated 24 incorrect isomers.

The next disassembly step ( $MS^4$  Figure 1D) suggests a branched structure because no fully methylated terminal disaccharide ion is observed ( $m/z$  445 or 463). However, the algorithm can supply structural insight only with observed ions and draws no conclusions from missing fragments. The next selected ion,  $m/z$  431.1, has the composition  $H_2$ -(ene)(oh). Applying the connected substructure constraint again, OSCAR determines that

### Chart 3. Input Listing<sup>a</sup>

```
-unreducedr      ; This glycan sample has an unreduced R
-nlinked         ; Specify that result must be embed N-linked core

; Examine non-reducing end terminals H3 and F5 via 1027 branch
AddPathway 1928.0_1272.6_1027.5_617.2_445.2_329.1
AddPathway 1928.0_1272.6_1027.5_605.3_475.1

; Examine reducing end via 678 branch: H4(1-4)F6(1-6)R8
AddPathway 1928.0_678.4_415.2_259.1 ; Branching
AddPathway 1928.0_678.4_415.2_299.1 ; H4(1-4)F6
AddPathway 1928.0_678.4_489.2      ; F6(1-4/6)R8
AddPathway 1928.0_678.4_475.2      ; F6(1-6)R8

Summarize
```

<sup>a</sup> Input to compute the branching and interresidue linkage of the glycan in Figure 4 in 0.03 s. The “-nlinked” option restricts candidate structures to those that embed the Man<sub>3</sub>GlcNAc<sub>2</sub> core.

both structures in Figure 1C satisfy this constraint, so ion *m/z* 431.1 fails to eliminate either candidate.

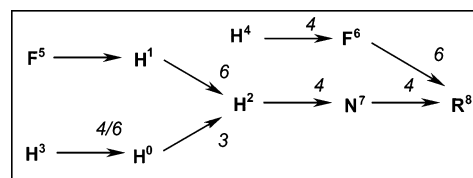
Disassembly of *m/z* 431.1 (MS,<sup>5</sup> Figure 1E) provides the major ion *m/z* 259.0, H-(oh), a terminal hexose. This time, applying the precursor/product constraint, OSCAR eliminates the linear candidate because neither H<sup>0</sup> nor H<sup>1</sup> is a terminal residue; the linear candidate's sole terminal hexose was lost in the previous step. The branched candidate is retained, and its residue H<sup>2</sup> is assigned to ion *m/z* 259.0. In this example, branching is assigned from the single pathway *m/z* 1187.6 → 894.4 → 649.2 → 431.1 → 259.0.

The exact input given to OSCAR for this analysis is presented in Chart 1; the analysis is completed in 0.01 s. Providing corroborating evidence for the proposed structure, OSCAR also reports each ion's deduced composition and location (Figure 1, right-hand column).

**Defining Linkage.** More pathways may subsequently be added to assign interresidue linkages. OSCAR tracks all allowable linkage positions for each residue, adding structural constraints to reduce the set of possible linkages when A-type cross-ring fragment ions are added. Because some linkages are resolved by detecting multiple cross-ring fragments, the absence of one or more fragments will cause several possibilities to be included in the linkage solution. Figure 1 lists many of the applicable cross-ring fragments, and more are given in Figure 2, which shows the ion *m/z* 676.3 isolated using the pathway *m/z* 1187.6 → 894.4 → 676.3. In this example, the input has been extended with additional A-type cross-ring fragments to yield the input shown in Chart 2, from which OSCAR computes many linkage details of the glycan in 0.01 s (Figure 3).

**Analyzing a Larger N-Linked Glycan.** Adding the “-nlinked” program option restricts the initial candidate structures to only those with an embedded Man<sub>3</sub>GlcNAc<sub>2</sub> core. This option is utilized in the analysis of an unusual *N*-glycan from *C. elegans* (H<sub>5</sub>F<sub>2</sub>NR, *m/z* 1928, Figure 4). OSCAR resolves the correct topology of this glycan given the six pathways shown in Chart 3; three of these pathways are detailed in Figure 4.

Given the ion *m/z* 1928.0, the candidate set initially contains 1006 topologies. To begin, the analyst inspects the MS<sup>2</sup> spectrum and employs simple composition database queries to identify a pair of complementary reducing- and nonreducing end fragments from among a number of observed structural isomers. The reducing end ion *m/z* 678.3, composition HFR-(oh), is disassembled along the pathway *m/z* 1928.0 → 678.3 → 415.2 → 259.1 to reveal the linear structure of the HexFuc branch (residues H<sup>4</sup>/



**Figure 5.** Topology proposed by OSCAR in 0.03 s for the glycan in Figure 4 given the input shown in Chart 3.

F<sup>6</sup>/R<sup>8</sup>); additional ions determine all interresidue linkages (Figure 4A–C). With the *m/z* 678.3 branch complete, the complementary nonreducing end ion *m/z* 1272.6 is analyzed. Ion *m/z* 605.3, isolated along the pathway *m/z* 1928.0 → 1272.6 → 1027.5 → 605.3, is fragmented to resolve placement of the terminal residue F<sup>5</sup> (Figure 4D). Similarly, the product spectrum of ion *m/z* 445.2, isolated along *m/z* 1928.0 → 1272.6 → 1027.5 → 617.3 → 445.2, contains the cross-ring cleavage ion *m/z* 329.1, indicating the terminal residue H<sup>3</sup> is either (1–4)- or (1–6)-linked to H<sup>0</sup> (Figure 4E). A number of other fragmentation pathways, excluded for brevity, corroborate this topology. A complete structural analysis is also available;<sup>13</sup> the unusual reducing end motif Gal-β(1–4)-Fuc-α(1–6)-GlcNAc was reported for *N*-glycans from keyhole limpet haemocyanin.<sup>16</sup>

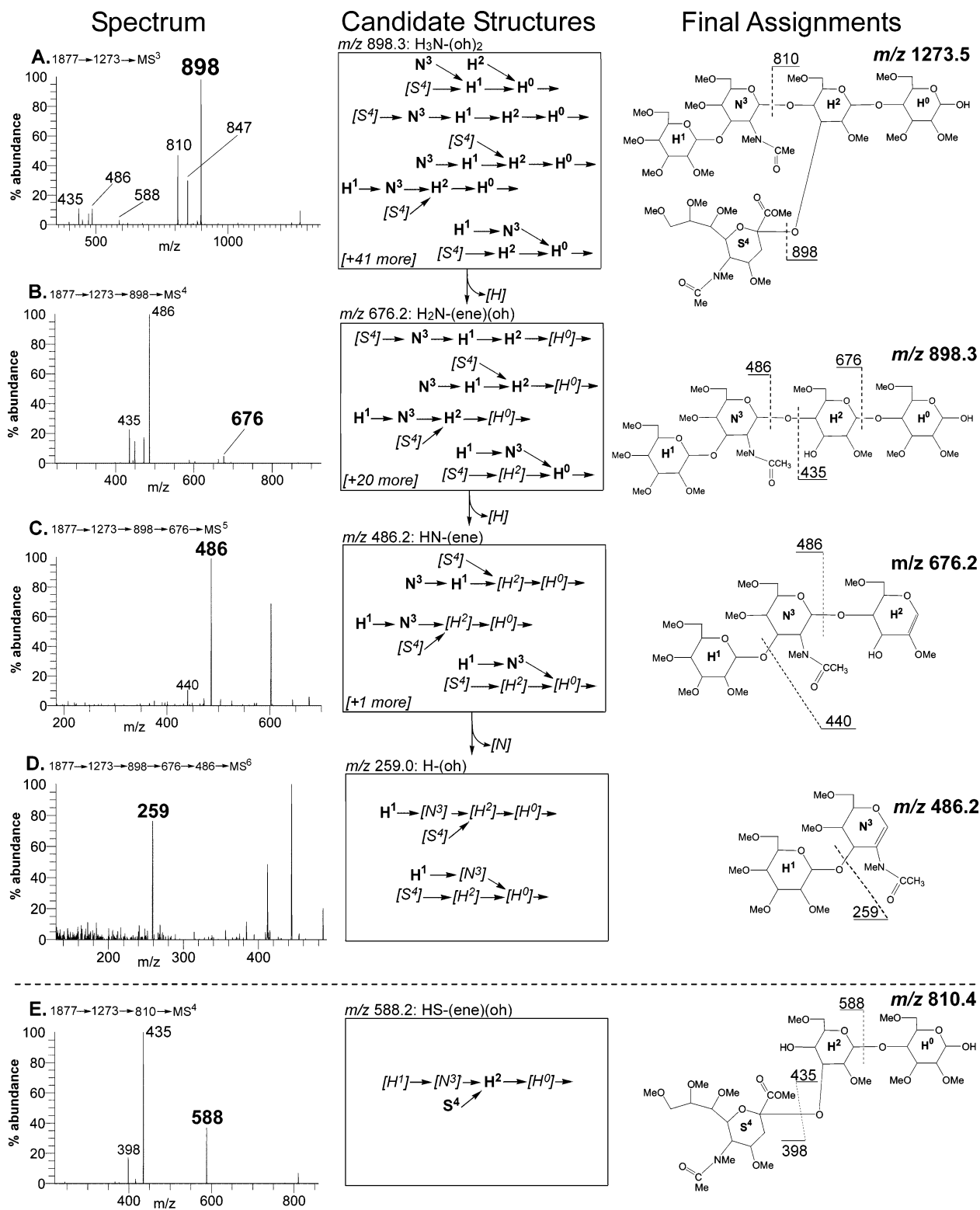
It is possible to assign this novel glycan because OSCAR is a de novo algorithm. Tools that rely on comparisons to previously characterized glycans might have difficulty with this structure. Moreover, the utility of the precursor/product relationship is again demonstrated here: ion *m/z* 433.2 appears in two spectra of Figure 4 (B and D), but because the context of the ions differs, OSCAR assigns one ion to the HexFuc disaccharide H<sup>4</sup>/F<sup>6</sup> and the other to the FucHex disaccharide F<sup>5</sup>/H<sup>1</sup>. The program produces the topology shown in Figure 5 in 0.03 s, indicating that OSCAR's performance appears to scale well to larger glycans.

**Analyzing Isobars.** The algorithm also assists in assigning mixtures of structural isomers, as demonstrated by application to the bovine brain gangliosides GM1a and GM1b.<sup>17,18</sup> The ganglioside mixture (*m/z* 1877.2) is fragmented in MS<sup>2</sup> to release the ceramide moiety, yielding the isomeric glycans (*m/z* 1273.5). This ion has composition H<sub>3</sub>NS-(oh); the (oh) cleavage records where the glycan was fragmented from the ceramide. (OSCAR

(16) Wuhler, M.; Robijn, M. L.; Koeleman, C. A.; Balog, C. I.; Geyer, R.; Deelder, A. M.; Hokke, C. H. *Biochem J.* **2004**, *378*, 625–32.

(17) Svennerholm, L. *J. Neurochem.* **1963**, *10*, 613–623.

(18) Chester, M. A. *Pure Appl. Chem.* **1997**, *12*, 2475–2487.

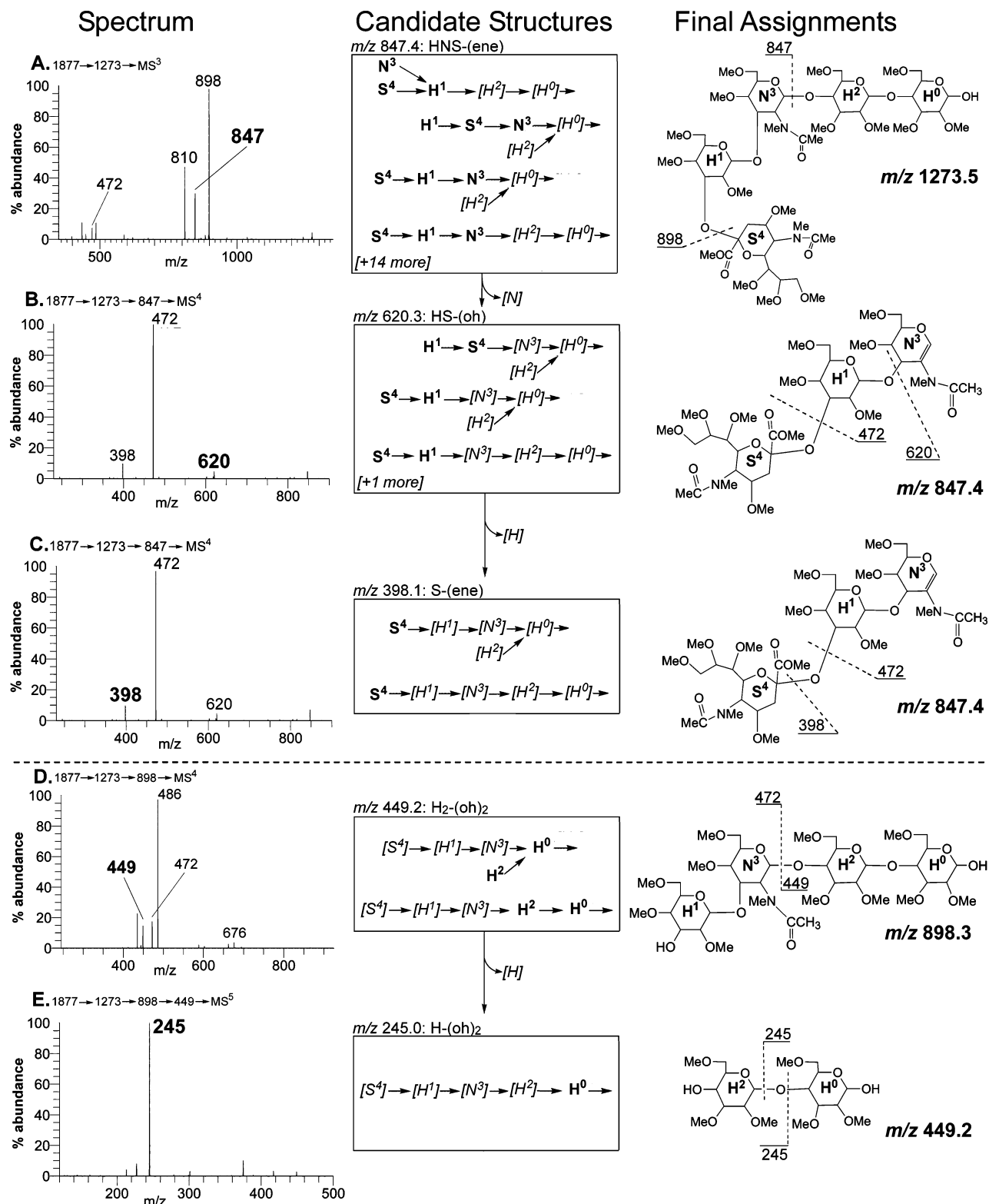


**Figure 6.** Sequential MS<sup>n</sup> disassembly of the carbohydrate moiety fragmented from ganglioside GM1a. Steps A–D show disassembly along the pathway  $m/z$  1273.5 → 898.3 → 676.2 → 486.2 → 259.0. Step E shows how the additional pathway  $m/z$  1273.5 → 810.3 → 588.2 resolves the glycan's reducing end. Analysis completes in 0.02 s.

accepts glycans of this type when the “Unmethylated-ReducingEnd” option is given.) Given the ion  $m/z$  1273.5, the initial candidate set contains 118 structures. MS<sup>3</sup> of the ion  $m/z$  1273.5 (Figure 6A, left column) produces the major ion  $m/z$  898.3, composition H<sub>3</sub>N-(oh)<sub>2</sub>, indicating loss of a terminal neuraminyl

moiety, reducing the candidate set to 46 structures (Figure 6A, center).

Successive fragmentations yield the pathway  $m/z$  1273.5 → 898.3 → 676.2 → 486.2 → 259.0 (Figure 6A–D). The MS<sup>4</sup> ion  $m/z$  676.2 clarifies that the neuraminyl residue is not linked to the



**Figure 7.** Sequential MS<sup>n</sup> disassembly of the carbohydrate moiety fragmented from ganglioside GM1b. The major ion *m/z* 847.4 from Figure 6A was unexplained by GM1a, prompting the search for isomer GM1b. Steps B and C show how ions *m/z* 620.3 and 398.1 reveal the linear structure of the nonreducing end S<sup>4</sup>/H<sup>1</sup>/N<sup>3</sup>. Steps D and E resolve the reducing end H<sup>0</sup>/H<sup>2</sup>. Taken together, these pathways reveal the linear structure of GM1b. Analysis completes in 0.02 s.

reducing end hexose. The MS<sup>5</sup> ion *m/z* 486.2 denotes an unbranched nonreducing end fragment with the composition HN-(ene), locating the neuraminyl branch point to the internal hexose H<sup>2</sup>. A terminal hexose is established by the MS<sup>6</sup> ion *m/z* 259.0.

At each step along this pathway, the candidate set is restricted until only two topologies remain (Figure 6D). To resolve a single structure, analysis must focus on the glycan's reducing end. The MS<sup>3</sup> ion *m/z* 810.3 (Figure 6A), the reducing end complement of

#### Chart 4. Input Listing<sup>a</sup>

```
-UnmethylatedReducingEnd  
AddPathway 1273.5_898.3_676.2_486.2_259.0  
AddPathway 1273.5_810.4_588.2  
Summarize
```

<sup>a</sup> Input to compute the branching structure of GM1a's oligosaccharide (Figure 6) in 0.02 s.

#### Chart 5. Input Listing<sup>a</sup>

```
-UnmethylatedReducingEnd  
AddPathway 1273.5_847.4_620.3  
AddPathway 1273.5_847.4_398.1  
AddPathway 1273.5_898.3_449.2_245.0  
Summarize
```

<sup>a</sup> Input to compute the branching structure of GM1b's oligosaccharide (Figure 7) in 0.02 s.

ion  $m/z$  486.2, yields the pathway  $m/z$  1273.5  $\rightarrow$  810.3  $\rightarrow$  588.2. This pathway locates the HN-(ene) fragment as attached to H<sup>2</sup>, yielding the topology of GM1a (Figure 6E). Given the input shown in Chart 4, OSCAR produces this result in 0.02 s.

However, the major ion  $m/z$  847.4 (Figures 6A and 7A) is inconsistent with the GM1a oligosaccharide, possibly indicating the presence of a structural isomer. The composition database identifies the ion  $m/z$  847.3 as HNS-(ene); however, the only possible match in GM1a (H<sup>2</sup>/N<sup>3</sup>/S<sup>4</sup>) cannot be extracted from the glycan by a single cleavage. Ion  $m/z$  847.4 is fragmented in MS<sup>4</sup> to yield the ions  $m/z$  620.3 and 398.1 (Figure 7B and C), allowing OSCAR to infer the linear nature of the HNS-(ene) fragment. Now only 2 candidate structures of the initial 118 remain, and the analysis focuses on residues H<sup>0</sup> and H<sup>2</sup>. The ion  $m/z$  449.2, isolated as  $m/z$  1273.5  $\rightarrow$  898.3  $\rightarrow$  449.2, has the composition H<sub>2</sub>(oh)<sub>2</sub>, linking the two hexoses (Figure 7D). Fragmenting  $m/z$  449.2 yields the major ion  $m/z$  245.0, with the composition H-(oh)<sub>2</sub>; the double cleavage eliminates the branched candidate because residue H<sup>2</sup> requires one cleavage to be removed while H<sup>0</sup> requires three. The remaining candidate indicates the linear topology corresponding to GM1b (Figure 7E). Corroborating pathways are excluded for clarity. Given the input shown in Chart 5, OSCAR produces the topology of GM1b in 0.02 s.

**Future Work.** OSCAR is a proven computational assistant for glycan topology analysis, and further improvements are planned, including enhanced support for resolving structural isomers. One planned improvement is for automated identification of ions that are inconsistent with the candidate structure(s). As with the ion  $m/z$  847.4 in the GM1 ganglioside example, inconsistent ions may indicate the presence of isomers. If a set of proposed isomeric glycans accounts for all the major ions, the analyst can have confidence that no obvious isobars remain undetected. This

capability may eventually reduce or eliminate the need for tedious separations of isomeric mixtures, with the added benefit of overall sensitivity enhancement.

As discussed in the previous report,<sup>2</sup> UNH-CSB is constructing an MS<sup>n</sup> oligosaccharide spectral library that associates known glycan substructures with experimentally observed spectra. Using mass/intensity comparisons, this library shows promise for allowing assignment of linkage, anomericity, and monosaccharide identification of the substructural motifs obtained from unknown glycans during MS<sup>n</sup> disassembly. This library may be integrated with OSCAR, with the goal of producing glycan sequences for automated, high-throughput glycomics.

## CONCLUSIONS

Disassembly of permethylated oligosaccharides by MS<sup>n</sup> generates a series of ion compositions that help position each fragment in the total glycan topology, allowing topologies to be reconstructed from precursor/product relationships and other logical constraints. With this premise, an algorithm, OSCAR, has been implemented that proposes structures from MS<sup>n</sup> data without imposing biosynthetic constraints or comparing against previously reported glycans. Thus, the program may be considered *de novo*. To accomplish this, the program uses a series of structural inference rules pertaining to permethylated glycans to narrow the set of candidate topologies while processing one or more ion fragmentation pathways, utilizing the important precursor/product constraint to maximize the structural information extracted from each ion. The algorithm shows potential for resolving structural isomers and may reduce the need for chromatographic separations. The algorithm processes N-linked and O-linked glycans and may be extended to cover a wider range of oligosaccharides. Branching and interresidue linkages are a component of the proposed topologies, and how comprehensively this degree of detail may extend is dependent on further ongoing studies. On the samples examined to date, the algorithm has demonstrated excellent execution time performance. GlySpy, a Windows-based command-line tool that implements the algorithm, is available at <http://glycome.unh.edu/tools/GlySpy>.

## ACKNOWLEDGMENT

Funding provided by the National Institutes of Health NCRB BRIN P20 RR16459 (V.N.R.) and NIGMS R01 GM54045 (V.N.R.) is gratefully acknowledged. We thank Joe Gieser, Suddham Singh, and Kevin Bullock of the UNH-CSB for their help.

Received for review April 27, 2005. Accepted July 29, 2005.

AC050726J