MARCH 2021

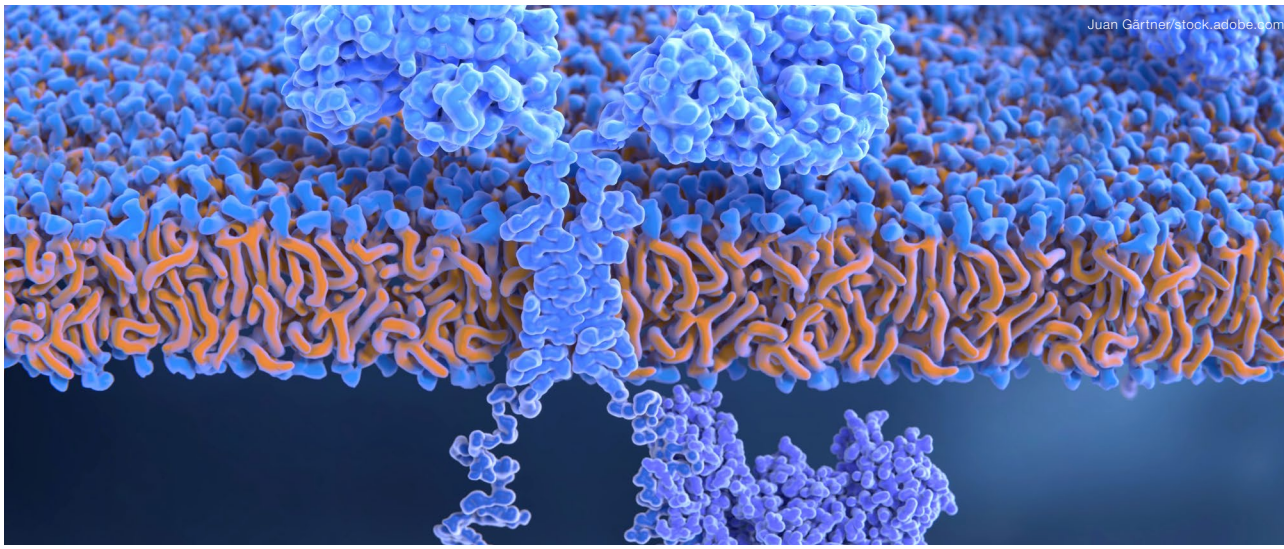
LC GC's CHROMacademy

powered by crawford scientific

Novel Solutions for the Separation, Identification, and Quantitation of Biomolecules

- Glycosylation and Sialic Acid Analysis of Biotherapeutic Glycoproteins
- Two-Dimensional Liquid Chromatography for Biomolecule Separations
- Identification, Mapping and Quantitation of SARS-CoV-2 Spike Glycopeptides

This custom ebook is sponsored by Agilent and presented in partnership with *CHROMAcademy*.



Glycosylation and Sialic Acid Analysis of Biotherapeutic Glycoproteins

John Yan

Approaches to identifying different sialylated N-glycan species and quantitating the sialic acid content of biotherapetuic glycoproteins using plate-based and LC/FLD/MS approaches.

Introduction

Sialic acid serves a critical role in mediating the effectiveness of recombinant therapeutic glycoproteins. The presence or absence of sialic acid on the non-reducing terminal of glycans can affect the pharmacokinetics of the biotherapeutic protein, as well as potential immunogenicity. Consequently, sialylation must be monitored and controlled during the manufacturing process. This article presents an approach to identifying different sialylated N-glycan species and quantitating the sialic acid content of biotherapeutic glycoproteins using plate-based and liquid chromatography (LC) with fluorescence (FL) and mass spectrometry (MS) detection approaches. Also discussed are the benefits of fluorescent dyes for N-glycan analysis, improved sialic acid quantitation, and analysis workflows, and chromatographic separation options that utilize both fluorescence and mass spectrometry (MS) detection.

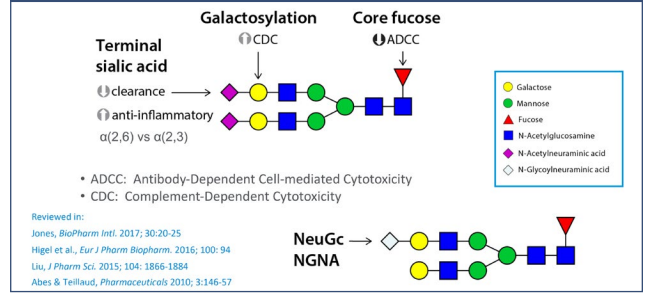
Glycosylated Biotherapeutics

Although glycomics has not enjoyed the same attention as genomics and proteomics, a fifth of all proteins in the SwissProt database are known to be glycosylated. Glycosylation can impact protein folding, stability, and function and impart diversity based on differences in glycosylation site, glycan composition, glycan length, and structure. More than 60% of marketed biotherapeutics are glycosylated. Glycosylation can be a critical quality attribute and/or monitored as a product quality attribute. The two main types of glycosylation are the addition of O-linked glycans and N-linked glycans, but N-linked glycosylation, where glycans are attached to the amide nitrogen of asparagine residues within a defined consensus sequon, is the glycosylation type most seen in biotherapeutics.

In 2010, Epogen was the only biologic of the top 10 selling pharmaceuticals that also happens to be glycosylated. Fast forward to 2018 and seven of the top 10 selling pharmaceutical products were glycosylated biologics, the biggest seller being Humira, which highlights the general shift in the pharmaceutical industry towards biopharmaceuticals.

Biopharmaceuticals are significantly more complex than small molecules from an analytical characterization perspective. For example, aspirin, a small-molecule drug, has a molecular weight of approximately 180 Da. In contrast, biopharmaceutical drugs based on immunoglobulin G (IgG) have a molecular weight of about 150,000 Da and have a

FIGURE 1: Effect of N-Glycan structure on biotherapeutics.

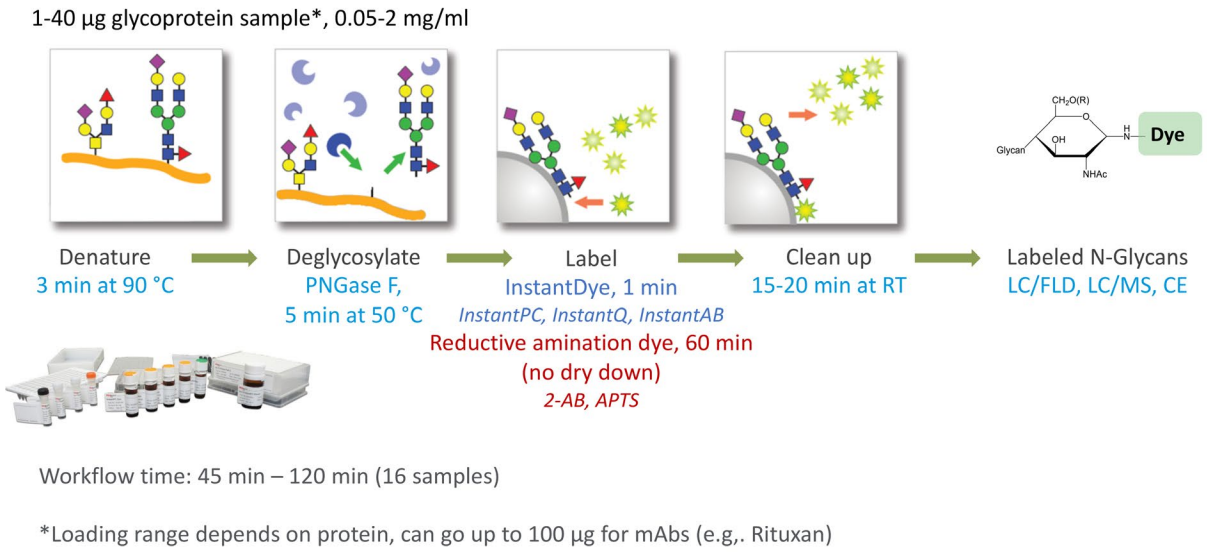


variety of post-translational modifications (including glycosylation) that can impact how the drug works. Glycans are made of multiple monosaccharide units such fucose, mannose, galactose, N-acetylglucosamine and various sialic acids, as illustrated in FIGURE 1. The presence of core fucose reduces antibody-dependent cellular cytotoxicity (ADCC), which is important for cancer therapeutics. For this reason, companies developing biotherapeutics may try to reduce the amount of core fucose to maximize ADCC activity. It has also been shown in the literature that the presence of galactosylation can increase complement dependent cytotoxicity (CDC) activity while the presence of a terminal sialic acid can lower the rate of clearance thereby increasing the amount of time the therapeutic stays within the body.

N-Glycan Analysis

Many different techniques can be used for N-glycan analysis. At the intact protein level, LC separations, in combination with high-resolution accurate mass spectrometry, can be used to provide information on the type of glycans that are present. Another option is to a use a protease to digest the glycoprotein into peptides that can be separated by LC

FIGURE 2: Gly-X N-Glycan Sample Prep Technology.



and analyzed by tandem MS using database searching to obtain site-specific glycosylation information. Another approach involves analyzing the enzymatic release of N-glycans from the glycoprotein using enzymes such as PNGase F. The released glycans can be analyzed as is or are commonly labeled with a fluorescent tag to enable LC or capillary electrophoresis (CE) separation techniques with fluorescence detection where the data obtained are generally expressed as relative % areas of the different glycan species.

Glycans can also be monitored at the individual monosaccharide level following acid hydrolysis either directly or after labeling with a fluorescent tag.

N-Glycan Sample Preparation

Traditional N-glycan sample preparation can be a very time-consuming process that can take one to multiple days with overnight deglycosylation and labeling steps. Several years ago, an all-in-one solution workflow was introduced that reduced the sample preparation time to 5 hr using traditional reductive-amination fluorophores such as 2-aminobenzamide (2-AB) and aminopyrene trisulfonate. Further improvements using a rapid glycosylamine reactive dye were made, which reduced the preparation time to 3–4 hr. More recently, the AdvanceBio Gly-X sample preparation platform was introduced. This new platform reduces the preparation time to 2 hr for reductive amination (2-AB, APTS), and less than 1 hr using instant glycosylamine reactive dyes (InstantPC).

A schematic of the AdvanceBio Gly-X N-glycan sample preparation platform is shown in FIGURE 2. The glycoprotein is first denatured with a proprietary reagent for 3 min at 90 °C, opening the protein for deglycosylation by PNGase F, which takes place over the course of 5 min at 50 °C. Next, released glycans are labeled with

a fluorophore such as the glycosylamine reactive dye InstantPC in approximately 1 min, or a reductive amination dye such as 2-AB in 60 min. After labeling, the labeled glycans are cleaned up to remove excess dye and are then ready for analysis.

Separation of InstantPC Labeled N-Glycans

One commonly used method for separation and analysis of fluorescent labeled glycans is hydrophilic interaction LC (HILIC) with fluorescence detection. In-line MS detection is also an option to aid in the identification of different glycan species. FIGURE 3 highlights a 60-min HILIC gradient method used for the separation and analysis of InstantPC labeled glycans from rituximab, an IgG therapeutic.

MS detection is also used. FIGURE 4 presents the total ion chromatogram response for labeled N-glycans of Enbrel. The MS total ion chromatogram for the InstantPC labeled sample can be seen to almost mirror the fluorescent response, which can help to identify low abundant glycan species. In addition, there are a wide variety of labeled N-glycan standards available that can help with identification efforts when MS may not be available. These include individual glycan standards, such as common biantennary N-glycans likely to be present on IgG molecules expressed in Chinese hamster ovary cells, as well as libraries of prelabeled glycans such as human IgG and both alpha-(2,6) and alpha-(2,3)-sialylated libraries.

2-AB Workflows

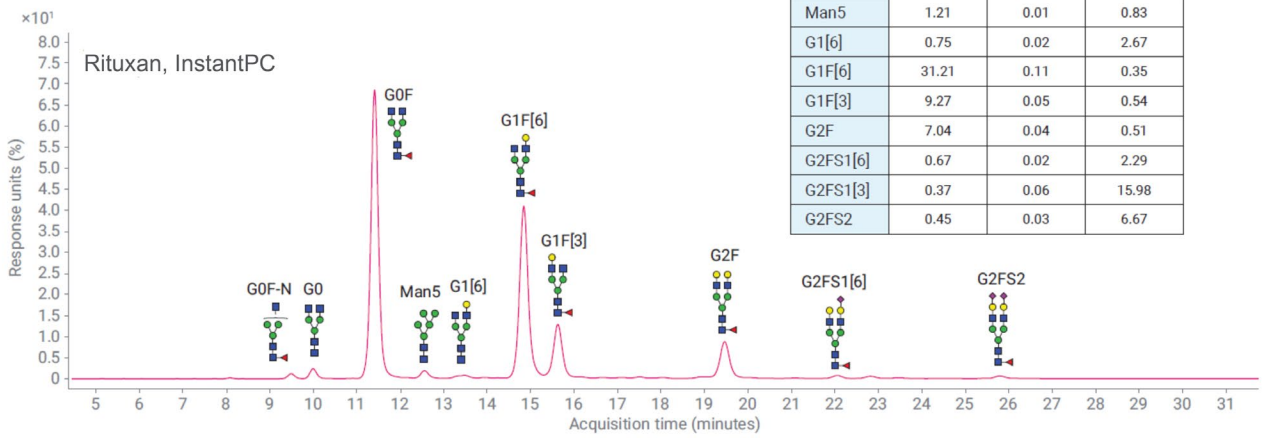
2-AB is a traditionally used dye for released glycan analysis. Traditional workflows with 2-AB can be time consuming, but when used in an updated and efficient workflow, 2-AB labeling can be a good fit for those that want the benefit of a rapid workflow but need to be able to relate new data to historical data. An express workflow using the AdvanceBio Gly-X technology is illustrated in FIGURE 2. Familiar in-solution denaturation and deglycosylation steps are used with on-matrix reductive-amination in a 96-well cleanup plate, which avoids the need for an additional dry down step. After labeling, the glycans are cleaned up to remove excess dye and then are ready for analysis. The preparation time is about 2 hr versus older methods that can take much longer. FIGURE 5 shows an example of the separation and analysis of glycans from rituximab; shown are mainly biantennary neutral 2-AB labeled glycans G0F, G1F, and G2F with low % CVs for all of the major glycan species.

Sialic Acid Quantitation

Sialic acid serves a critical role in mediating the effectiveness of therapeutic glycoproteins. The presence or absence of sialic acid at the non-reducing terminal of N- or O-glycans can impact the pharmacokinetics as well as the immunogenicity of the protein. It is therefore essential that the sialic acid content of protein therapeutics be maintained and controlled. The AdvanceBio Total Sialic Acid Quantitation kit offers a sensitive and high-throughput approach to total sialic acid quantitation. The 96-well plate-based assays are based on enzyme reactions that begin by converting sialic acid released by sialidase-A to hydrogen peroxide, which then reacts one-to-one with a fluorescent

FIGURE 3: Rituximab InstantPC Glycans.

	Average Rel % Area	Standard Deviation	%CV
G0F-N	0.75	0.01	1.55
G0	1.47	0.02	1.18
G0F	46.82	0.07	0.15
Man5	1.21	0.01	0.83
G1[6]	0.75	0.02	2.67
G1F[6]	31.21	0.11	0.35
G1F[3]	9.27	0.05	0.54
G2F	7.04	0.04	0.51
G2FS1[6]	0.67	0.02	2.29
G2FS1[3]	0.37	0.06	15.98
G2FS2	0.45	0.03	6.67



UHPLC-HILIC, 60 minute method

AdvanceBio Glycan Mapping column

2.1 x 150 mm, 1.8 µm

FLD: 285/345 nm

1 µl injection (equivalent to glycans from 0.4 µg protein)

N = 4

FIGURE 4: MS total ion chromatogram for InstantPC N-Glycans from Enbrel.

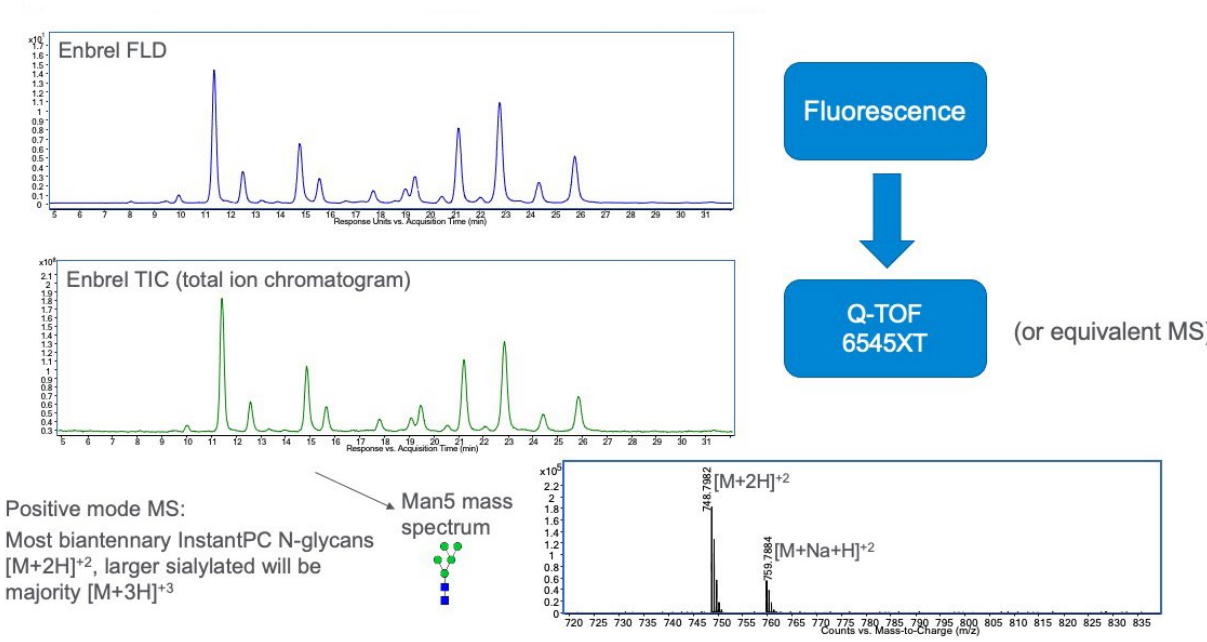
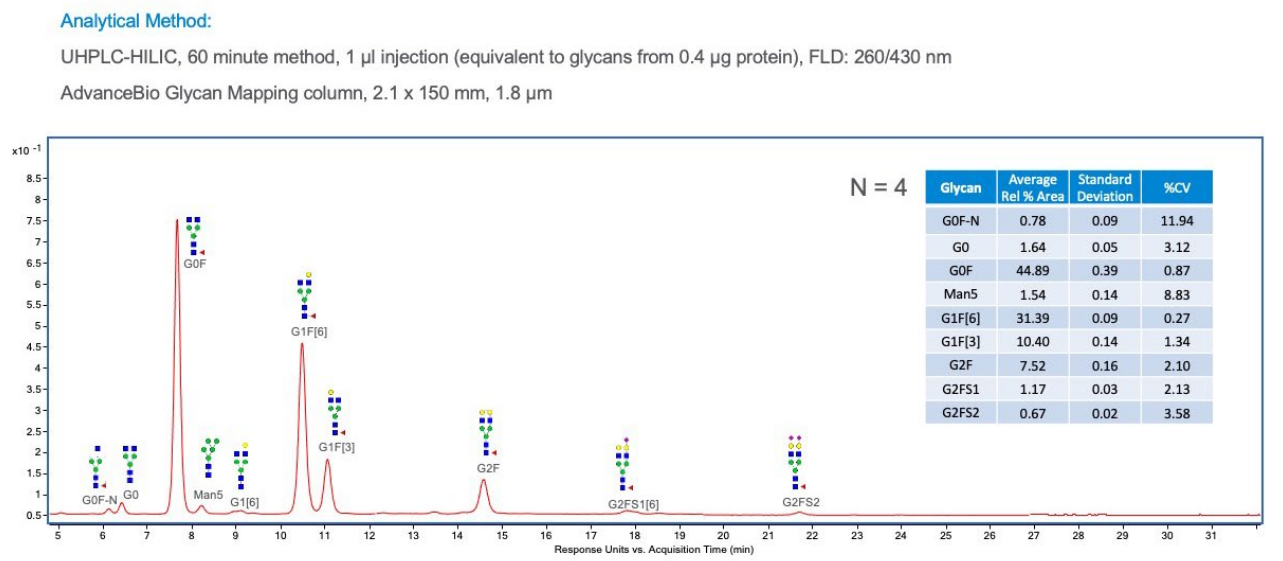


FIGURE 5: Rituximab Gly-X 2-AB Express N-Glycans.



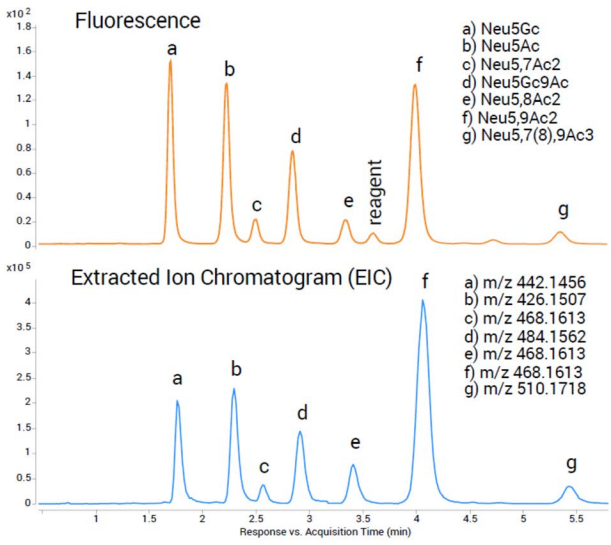
Glycan	Average Rel % Area	Standard Deviation	%CV
G0F-N	0.78	0.09	11.94
G0	1.64	0.05	3.12
G0F	44.89	0.39	0.87
Man5	1.54	0.14	8.83
G1F[6]	31.39	0.09	0.27
G1F[3]	10.40	0.14	1.34
G2F	7.52	0.16	2.10
G2FS1	1.17	0.03	2.13
G2FS2	0.67	0.02	3.58

dye generating intense fluorescence or absorbance signal. Different glycoproteins can have varying amounts of glycosylation, which also means varying amounts of sialylation. Consequently, it is important to start with an amount of protein that will give a good signal. For example, MabThera has low levels of sialic acid, so a larger amount of protein is necessary, whereas fetuin or Enbrel have higher sialic acid contents and thus requires less protein.

N-acetylneuraminic acid (NANA) and N-glucoylneuraminic acid (NGNA) are types of sialic acids that are similar in structure and differ only by the addition of hydroxyl group (NGNA). This small difference is enough to make NGNA potentially immunogenic to humans. To distinguish between these two types of sialic acid, one commonly used method is to release the sialic acid from the glycan labeling it with the dye: 1,2-diamino-4,5-methylenedioxybenzene-2HCl (DMB), a process utilized by the AdvanceBio Sialic Acid Profiling and Quantitation kit, which also includes quantitative NANA and NGNA standards. Labeling is followed by LC separation and detection by fluorescence and/or MS. An example of the analysis of a sialic acid reference panel, a mixture of different types of sialic acids labeled with DMB, is shown in FIGURE 6 using LC followed by both fluorescent and MS detection. The first two peaks on the chromatogram are NGNA and NANA, respectively, and

"The presence or absence of sialic acid at the non-reducing terminal of N- or O-glycans can impact the pharmacokinetics as well as the immunogenicity of the protein."

FIGURE 6: DMB labeled Sialic Acid Reference Panel (SARP).



Reversed-phase UHPLC conditions

Parameter	Value
Instrument	Agilent 1290 Infinity II LC System
Column	Agilent InfinityLab Poroshell 120 EC-C18, 2.1 x 75 mm, 2.7 µm (p/n 697775-902).
Column Temp	30 °C
Mobile Phase	A) Methanol:acetonitrile:water (4:8:88) B) Acetonitrile
Gradient Program	Time (min) / %A / %B / Flow rate (mL/min) 0.00 / 100 / 0 / 0.4 — Isocratic elution 6.00 / 100 / 0 / 0.4 6.25 / 20 / 80 / 0.4 — Wash 7.30 / 20 / 80 / 0.4 7.50 / 100 / 0 / 0.4 — Re-equilibration 10.00 / 100 / 0 / 0.4
Injection Volume	10 µL (Equivalent to 0.25 pmol of Enbrel, 10 pmol of Rituxan)
Detection	Agilent 1260 Infinity II FLD λ_{Ex} 373 nm, λ_{Em} 448 nm

Pressure > 400 bar

MS parameters

6545XT AdvanceBio LC/ Q-TOF	
Source	Dual AJS ESI
Gas Temperature	350 °C
Drying Gas Flow	11 L/min
Nebulizer	15 psi
Sheath Gas Temperature	400 °C
Sheath Gas Flow	12 L/min
Vcap	1400 V
Nozzle Voltage	1800 V
Fragmentor	120 V
Skimmer	65 V
Oct 1 VF Vpp	600 V
Mass Range (MS)	*m/z* 400-1000
Mass Range (MS/MS)	*m/z* 100-550
Acquisition Mode	High resolution (4 GHz)

Extracted ion chromatogram of DMB labeled sialic acid species, $[M+H]^{1+}$

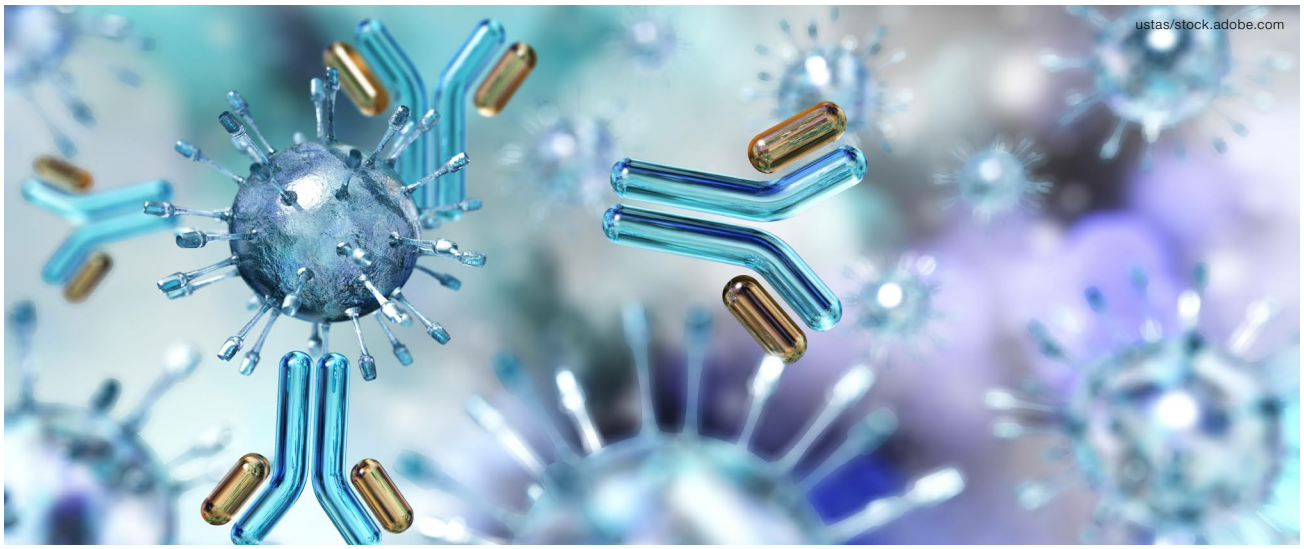
the remaining peaks in the chromatogram are different versions of sialic acid that are observed in nature.

Conclusion

Many different tools are available to researchers to prepare, separate, detect, and identify N-glycans that are present on a typical biotherapeutic glycoprotein. Modern, workflows such as the AdvanceBio Gly-X sample preparation kit, utilize techniques to rapidly release, label, and separate glycans for quantitation with fluoresce detection. In addition to these new and rapid sample preparation workflows, the high ionization efficiency of new dyes such as InstantPC, allows for robust characterization of released glycans by MS. These new workflows are also available with traditional glycan labeling dyes such as 2-AB, which offers a convenient and rapid sample preparation option for those who wish to relate new and historical data. In addition to released glycan analysis, Agilent offers products for sialic acid quantitation and profiling, which rounds out a robust tool kit offering for the analysis of glycosylated biotherapeutics.

John Yan

John Yan, Ph.D., is an Applications Chemist at Agilent.



Introduction to Two-Dimensional Liquid Chromatography for Biomolecule Separations

Dwight R. Stoll

2D-liquid chromatography fundamentals and contemporary applications in bioanalysis.

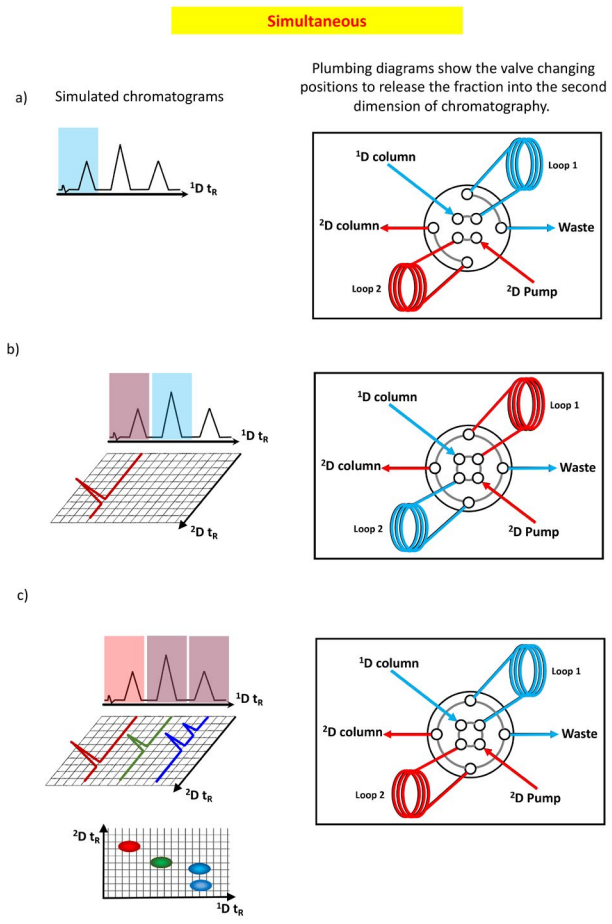
Introduction

While conventional one-dimensional liquid chromatography (1D-LC) remains a valuable tool in bioanalysis, adding another dimension of chromatographic separation may be advantageous. Analytical separations with two columns in tandem, or two-dimensional liquid chromatography (2D-LC), can provide tangible gains in analyte resolution without requiring onerous extensions in separation time. Tailoring the separation parameters to specific mixture conditions and targeted analytes will help to maximize the analytical yield of 2D-LC. This summary will serve as an introduction to the concept of 2D liquid chromatographic separations, cover some key modes and operating principles for 2D-LC, and also provide pertinent examples of how 2D-LC yields tangible benefits for solving bioanalytical challenges.

2D-LC Separation Basics

Very often in biomolecular analysis, even seemingly simple samples may be complex enough to have unresolved peaks in

FIGURE 1: How two-dimensional liquid chromatography works.

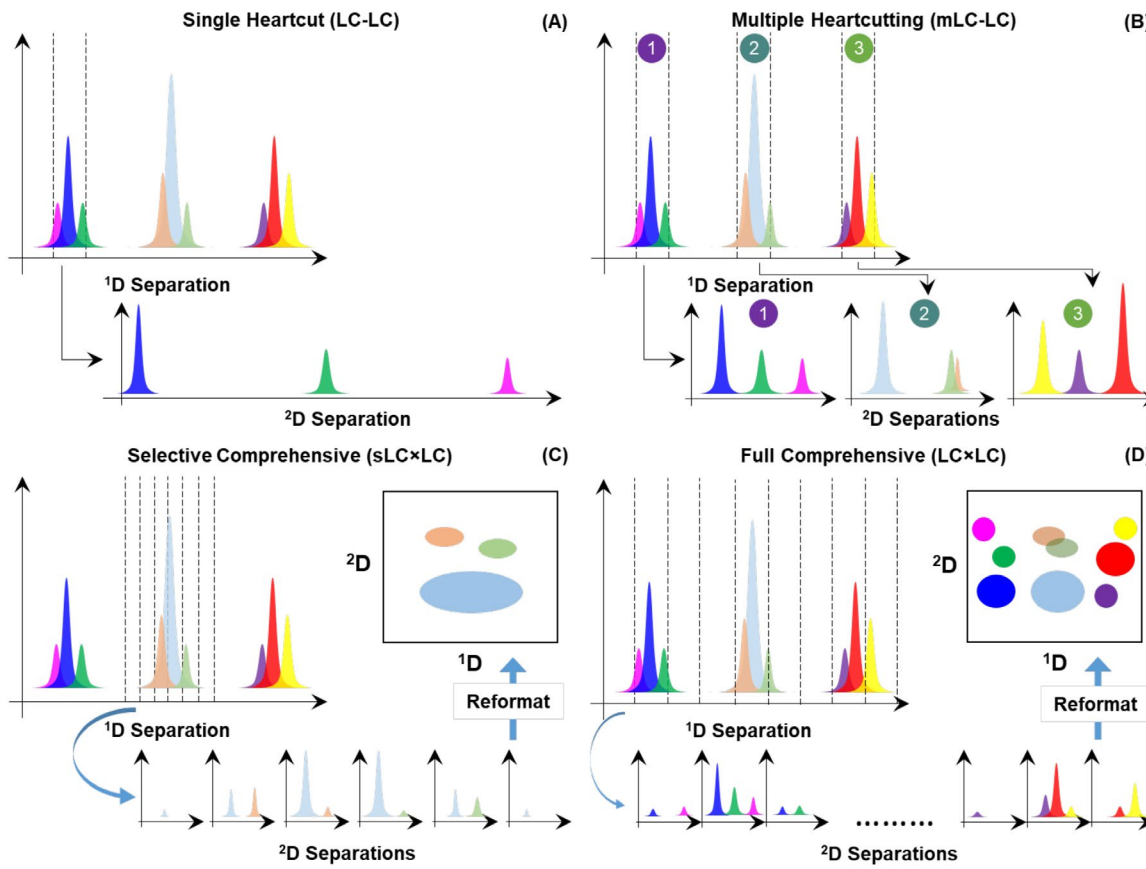


chromatographic separations. As a result, there may be times when a single stage of liquid chromatography is simply not adequate to separate species in a mixture, particularly for analytes that may be closely related in structure as with isomeric mixtures, or for unknown species requiring analyte identification. While FIGURE 1A offers an idealized example, the simulated chromatograms shown can clarify why using two LC dimensions is valuable. Under circumstances where separations according to no single property alone, such as size or charge, will yield sufficient analyte resolution, separating the analytes by both properties can yield fully resolved species.

As a way of quantifying the enhancement of separations observed with 2D-LC, the commonly used metric of peak capacity can provide a numerical explanation of how a second LC dimension is analytically more valuable than simply lengthening the column for 1D-LC. For each dimension of chromatography, the analytical measurements are essentially grouped into bins according to the effectiveness of the separation and the instrument response duty cycle. This creates a linear set of bins with instrument response over time. With the addition of a second dimension, the peak capacity is multiplied to generate a 2D set of measurement bins. With judicious selection of complementary columns and separation conditions, it is possible to make use of this separation space to differentiate previously unresolved analytes. Furthermore, the second dimension can be added as necessary for a given analysis—2D-LC operating modes permit the targeted application of the second dimension (2D) separation stage.

As with 1D-LC, the instrumentation needed for 2D-LC involves the use of a switching valve but with some added components for timing sample injections. Appropriate timing of the valve switching and loading of the two sample loops permit the separation process to occur on both columns simultaneously (FIGURE 1B). A portion of the effluent from the 1D column is diverted onto a sample loop of known volume, and once filled, the solution within the sample loop is injected on to the 2D column. As a result, segments of the separation from first dimension

FIGURE 2: Modes of 2D separation (6).



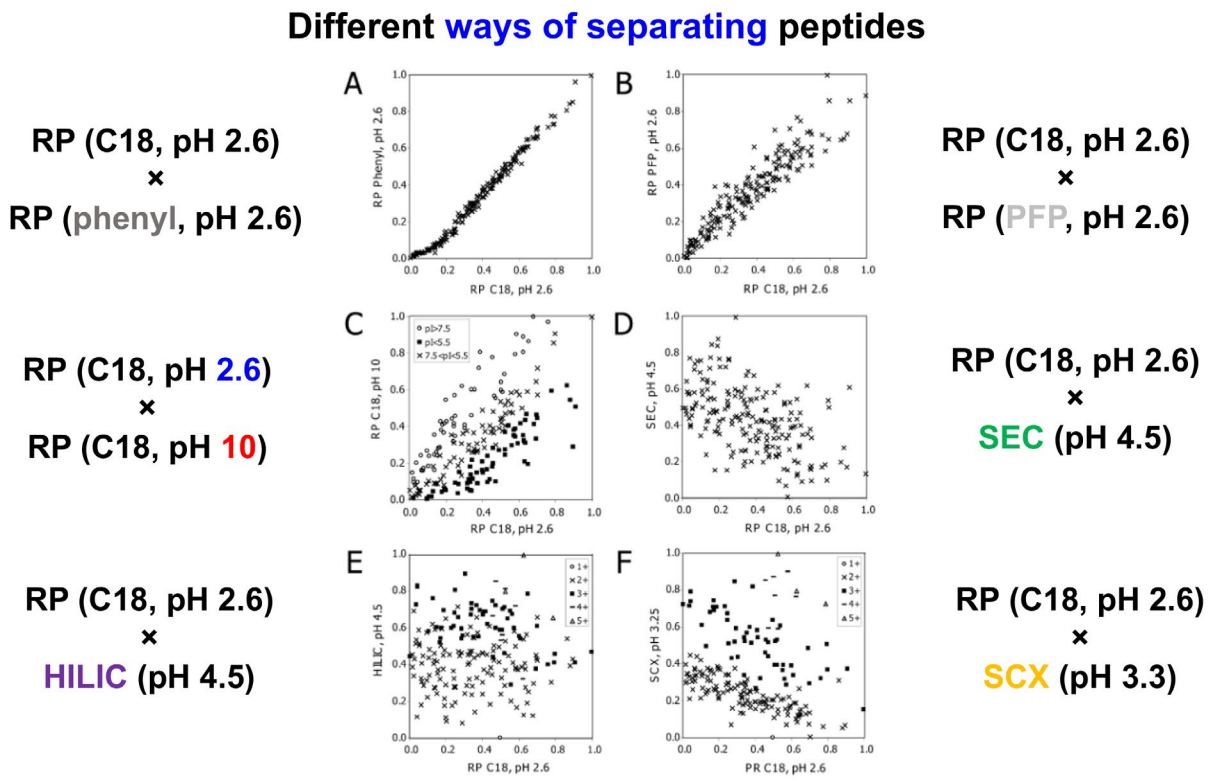
(1D) are effectively pooled before being injected on to the second column. The chromatographic data from 2D-LC reveals how analyte from each of the 1D peaks are then further separated according to the 2D conditions. Any subsequent separation of components that can occur in 2D beyond that in 1D will be revealed as additional peaks in the 2D chromatogram.

Targeting 2D-LC Operational Modes by Application

Within the confines of 2D-LC, there are two primary modes of operation as well as two modes that are effectively hybrids of the primary operating modes; all four modes are depicted in FIGURE 2.

The first of the primary operating modes is known as single heartcut and abbreviated as LC-LC (FIGURE 2A). The approach in single heartcut 2D-LC involves taking effluent from a single time segment from the 1D chromatogram, such as a single peak or cluster of peaks, and subjecting that one portion of effluent to separation on the 2D column. LC-LC is most effective for providing additional characterization of a single region of interest identified from 1D.

FIGURE 3: Optimize usage of the 2D separation space (10).



For the second primary operating mode in 2D-LC, all of the eluent from 1D is directed to 2D in fractions for separation in what is known as full comprehensive 2D-LC and abbreviated as LCxLC (Figure 2d). LCxLC provides the most complete dataset regarding the components in a target mixture. This mode is most appropriate for highly complex samples that have the potential for multiple unresolved and overlapping species, or for the complete assessment of an unknown mixture.

In addition to the two primary operating modes for 2D-LC, the two hybrid modes available include multiple heartcutting (mLC-LC) and selective comprehensive (sLCxLC). Multiple heartcutting applies similar operational principles as single heartcut where a specific section of eluent from 1D is injected for separation on 2D, but with more than one heartcut sample taken per chromatographic run (FIGURE 2B). The mLC-LC technique is useful for circumstances necessitating enhanced characterization of more than one region of the chromatogram, but not to the far extent granted by LCxLC. For selective comprehensive, the LCxLC concept of sampling multiple time segments over a specific time window is applied but not for the entire duration of the analysis (FIGURE 2C). The sLCxLC operational mode is ideal for thorough evaluation of a specific target species, such as for deep protein characterization, where only a portion of

the chromatographic run requires greater analytical scrutiny.

"As with traditional 1D-LC, performing 2D-LC merits careful tuning of the separation conditions according to the sensitivity of the detector in use."

Obtaining Optimal Resolution with Column Orthogonality

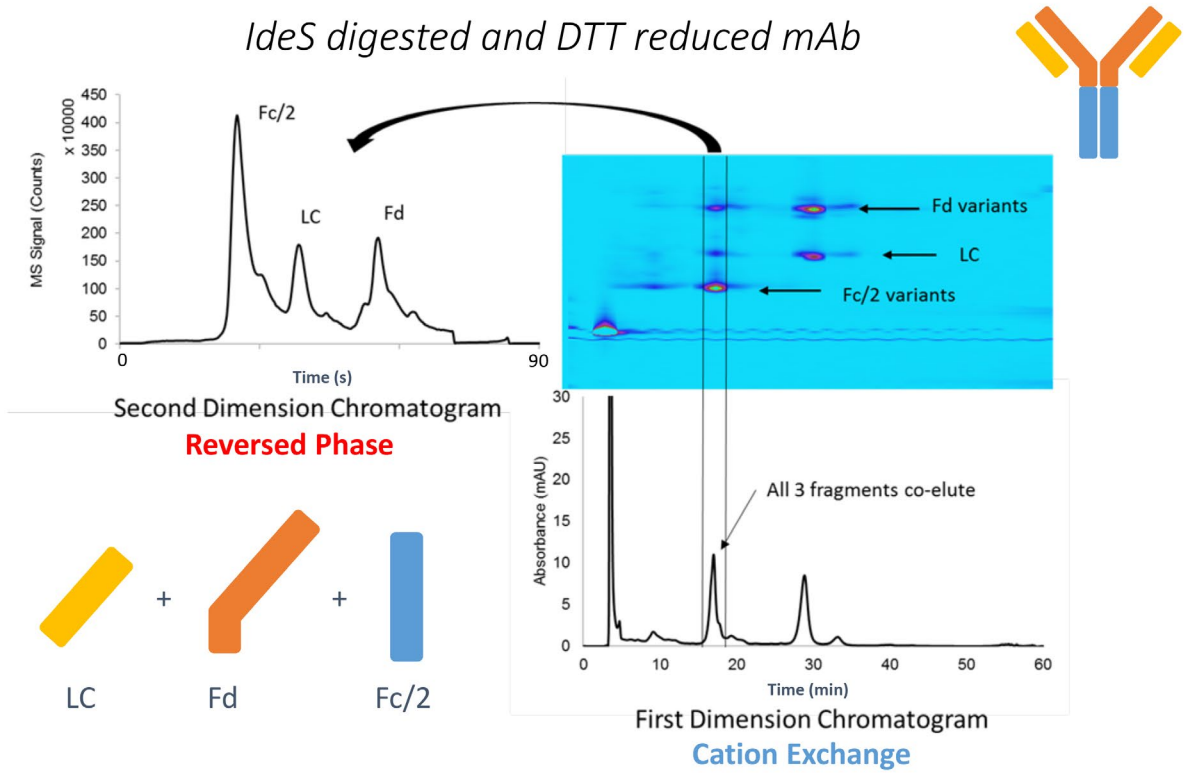
The multiplicative enhancement in peak capacity that may result from 2D-LC hinges largely on the use of appropriately complementary, or orthogonal, stationary phases on the 1D and 2D columns. Attempting 2D-LC separations with two columns of essentially the same stationary phase and mobile phase conditions will not grant any productive separation improvement as a result of adding the second dimension. Instead, there should be some form of differentiation between the dimensions to provide separation enhancement over the 2D space. Per the example in Figure 3a, the combination of C18 and phenyl stationary phases with the same mobile phase pH yields essentially no spread of analyte peaks to regions of the separation space outside of a roughly 1:1 correlation line for analyte elution between columns. However, by simply replacing the 2D column with a fluorinated phenyl stationary phase (FIGURE 3B), additional distinction between analytes in each separation stage was created. Even if the same column type is used for both 1D and 2D (FIGURE 3C), altering the mobile phase pH between dimensions creates a large enough difference in separation conditions for noticeably improved peak resolution across the analytical space. Likewise, resolution enhancement can be obtained through even more distinctly different columns between dimensions as shown in FIGURE 3D-F.

Optimizing 2D Sampling Speed to Avoid Undersampling

A key takeaway for 2D-LC separation timings is to avoid undersampling the effluent from the 1D column for 2D separation, which is an issue known as either the undersampling or re-mixing problem. By taking too few samples across a 1D peak, the net effect created is essentially an averaging, or re-mixing, of any species that were separated in 1D and thereby reduce the value of the first separation stage. According to Murphy-Schure-Foley theory (1), there should be no fewer than four samples taken from 1D across the region of a chromatographic peak in order for the impact of undersampling to be minimized. Furthermore, it is necessary to optimize the sampling speed so as to balance the speeds in both 1D and 2D to yield an optimal net peak capacity for the system overall.

As with traditional 1D-LC, performing 2D-LC merits careful tuning of the separation conditions according to the sensitivity of the detector in use. For each stage of chromatography (1D and 2D), the analyte concentrations in eluent will be lower than upon initial injection on to each column.

FIGURE 4: Subunit-level 2D-LC-MS of mAbs by LCxLC (7).



Consequently, there is a minimum injection volume and sampling time to ensure that enough analyte will be present.

Overcoming Mobile Phase Mismatch Between Dimensions

Beyond accounting for undersampling, another important aspect of 2D-LC to consider lies in preventing negative effects from a mobile phase mismatch between columns. This issue arises from the mobile phase eluting with analyte from 1D effectively becoming the injection solvent for 2D, and thereby requiring mobile phase compatibility. This can be of minimal impact for some combination of columns and mobile phases, but of substantial importance in others. As an example, the mobile phase composition of an ion exchange separation typically has an aqueous content and therefore is generally compatible with the primarily aqueous mobile phase at the beginning of a reversed phase separation. In contrast, the highly organic mobile phase used in normal phase chromatography would create challenges coupling a normal phase 1D with a reversed phase (RP) 2D and may result in fronting or other loss of analyte separation.

However, mobile phase incompatibility can be overcome with active solvent modulation (ASM) technology. Developed in concert with Agilent, 2D-LC systems with ASM act to adjust the solvent composition of the sample that is injected on to the 2D column for improved compatibility. By introducing a parallel sample loop that dilutes the collected

effluent from the first dimension before the sample reaches the 2D column, the strength of the 2D injection solvent is reduced. ASM enhances 2D resolution despite sample dilution because the lowered solvent strength prevents the analyte breakthrough that would result otherwise due to the inappropriately high analyte solubility in the mobile phase at the start of the separation.

Leveraging 2D-LC to Obtain 2-for-1 Analyses

One category where 2D-LC would be an ideal approach is where two or more separate experiments would be required to achieve the results obtained from a single 2D-LC analysis. This can be exemplified when characterizing an unknown peak in a chromatogram under mass spectrometry (MS)-incompatible conditions, which was examined by Luo et al (2). Rather than devising an additional separation under MS-compatible separation conditions, Luo et al. added a second dimension that was MS-compatible on to the first stage of LC, permitting straightforward identification of unknown peaks. As observed by Karongo et al., sLCxLC applied to the characterization of impurities in therapeutic peptides revealed the presence of three impurities that were coeluting with the peptide of interest and would have been missed with 1D-LC (3). Through a combination of an achiral RP column in 1D and a chiral WAX column in 2D, Woiwode et al. demonstrated that sLCxLC permits the enantioselective separation of all amino acids in a single analysis (4). For characterizing plant-derived extracts such as those from tea or wine, LCxLC had a vital role in enabling the identification of phenolic species by Muller et al. (5).

Methods for Deep Protein Characterization

Another type of analysis where 2D-LC would yield particularly helpful insights is in the deep characterization of proteins. When combined with mass spectrometry on the back end, two dimensional separations can greatly streamline the sample identification process. Thorough assessment of monoclonal antibody structure and composition can be accomplished with the appropriate sample treatments and 2D-LC followed by MS detection. Antibody subunits, light and heavy chains, attached glycans, and component peptides have been demonstrated as identifiable through this type of multidimensional approach (6). Furthermore, results reported by Sorensen et al. substantiate the need for two stages of LC separation due to the coelution of digested antibody fragments if the first LC dimension were used alone (FIGURE 4) (7).

Development of Automated 2D-LC Analyses

One of the challenges in bioanalysis that 2D-LC can help overcome is the issue of time-consuming sample preparation and manual selection of peaks for further analysis. Rather than performing each sample treatment such as trypsin digestion or dithiothreitol (DTT) reduction offline, these steps can be integrated with automated 2D-LC. A specific example involving antibody separation by Gstöttner et al. included a workflow where samples were first separated on an ion exchange column and

FIGURE 5: Contemporary applications—automated characterization systems for mAbs (9).

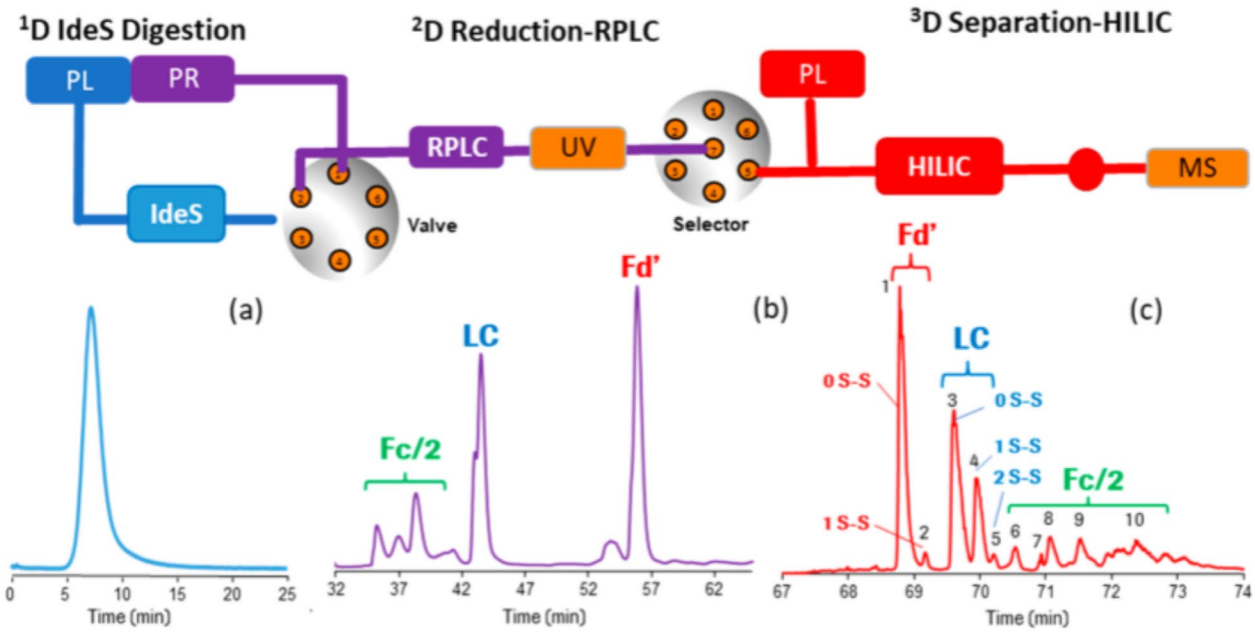


Figure 2. Schematic representation of the online 3D-LC workflow for automated middle-up analysis of mAbs using an online IdeS digestion with enzyme-immobilized column (^{1}D), followed by on-column reduction by RP-HPLC (^{2}D) and HILIC (^{3}D)/MS analysis of the proteolytic fragments Fd', LC, and Fc/2. (a) IdeS digestion of mAb-A (offline, detection UV at 280 nm), (b) RPLC (^{2}D), and (c) HILIC (^{3}D) chromatograms showing the three subunits (Fd', LC, and Fc/2). See Table S2 for retention times and mass assignments.

collected online in fractions, which could then be sent through a reducing reactor column followed by the 2D-RP separation and MS detection (8). Similarly, Camperi et al. devised a procedure for characterizing the glycosylation variants of monoclonal antibodies wherein samples were subject to IdeS digestion, reduction and separation on a RP column, and hydrophilic interaction LC (HILIC) separation before detection by mass spectrometry (FIGURE 5) (9). The automated method by Camperi et al. was found to have no significant differences in comparison with offline sample treatment in proof-of-concept experiments.

Conclusion

Multidimensional separations are fast becoming known for providing ample analytical information beyond 1D-LC while requiring little additional effort. However, for this technology to yield optimal results, traditional principles underlying chemical analysis should still be taken into account. As with other analytical instruments, method development for 2D-LC requires that the operating mode used (LC-LC, mLC-LC, sLCxLC, or LCxLC) be targeted to the sample type and what kind of information is needed from the analysis. After accounting for operating mode, it is vital to use separation conditions that are sufficiently orthogonal. Maximal resolution across the 2D space can result not only from complementary stationary phases, but also from mobile phase conditions such as pH depending on analyte class, so considering molecular properties

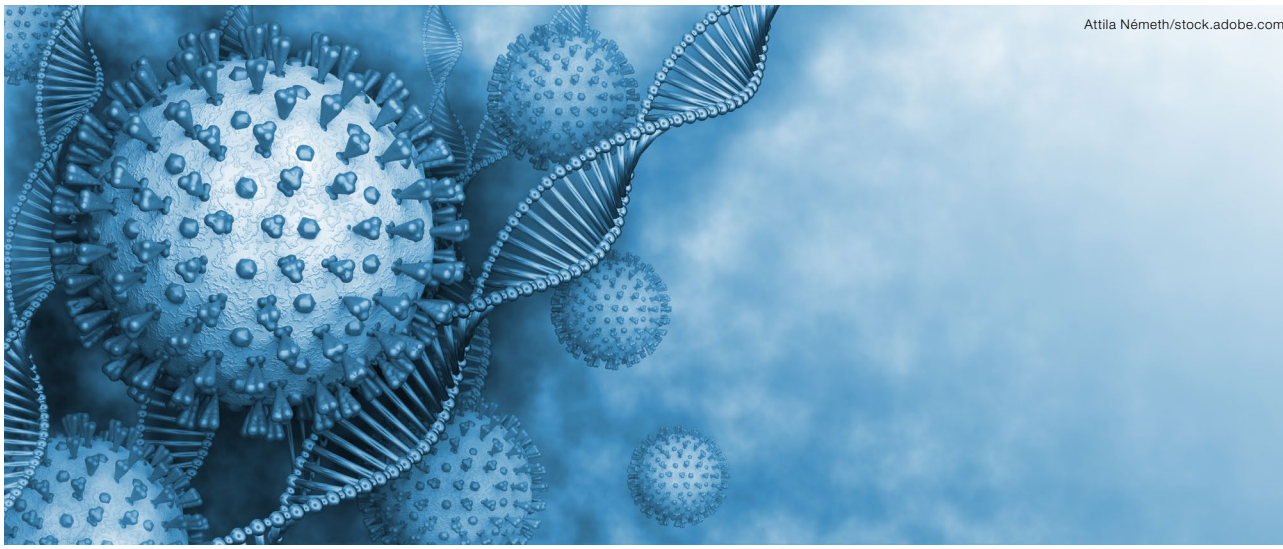
is critical. Additionally, care should be taken to not undersample the 1D stage as this has the potential to void separation gains made in the first dimension. However, this needs to be balanced with the sensitivity of the detector; the fractions separated in 2D should not be so low in volume, and therefore low in analyte, as to go undetected. Furthermore, the compatibility of mobile phases between dimensions should be examined to avoid mismatch when possible. When mobile phase mismatch is unavoidable, appropriate workaround methods and ASM technology should be used.

Dwight R. Stoll, PhD

Professor
Gustavus Adolphus College

References

1. R. E. Murphy, et al., *Anal. Chem.* **70**(8), 1585-1594 (1998). DOI: 10.1021/ac971184b
2. H. Luo, et al., *J. Pharm. Biomed. Anal.* **137**, 139-145 (2017). DOI: 10.1016/j.jpba.2016.11.012
3. R. Karongo, et al,. *J. Chrom. A*, **1627**, 461430 (2020). DOI: 10.1016/j.chroma.2020.461430
4. U. Woiwode, et al., *J. Chrom. A*, **1562**, 69–77 (2018). DOI: 10.1016/j.chroma.2018.05.062
5. M. Muller, et al., *Chromatographia*, **82**, 181–196 (2019). DOI: 10.1007/s10337-018-3662-6
6. D. R. Stoll, et al.,in *Advances in Chromatography*, (CRC Press: Boca Raton, FL, 2018).
7. M. Sorensen, et al., *mAbs*, **8**(7), 1224–1234 (2016). DOI: 10.1080/19420862.2016.1203497
8. C. Gstöttner et al., *Anal. Chem.* **90**(3), 2119–2125 (2017). DOI: 10.1021/acs.analchem.7b04372
9. J. Camperi et al., *Anal. Chem.* **92**(6), 4357–4363 (2020). DOI: 10.1021/acs.analchem.9b05193
10. M. Gilar, et al., *Anal. Chem.* **77**(19), 6426–6434 (2005). DOI: 10.1021/ac050923i



Identification, Mapping and Relative Quantitation of SARS-CoV-2 Spike Glycopeptides by Mass-Retention Time Fingerprinting

Rod Chalk

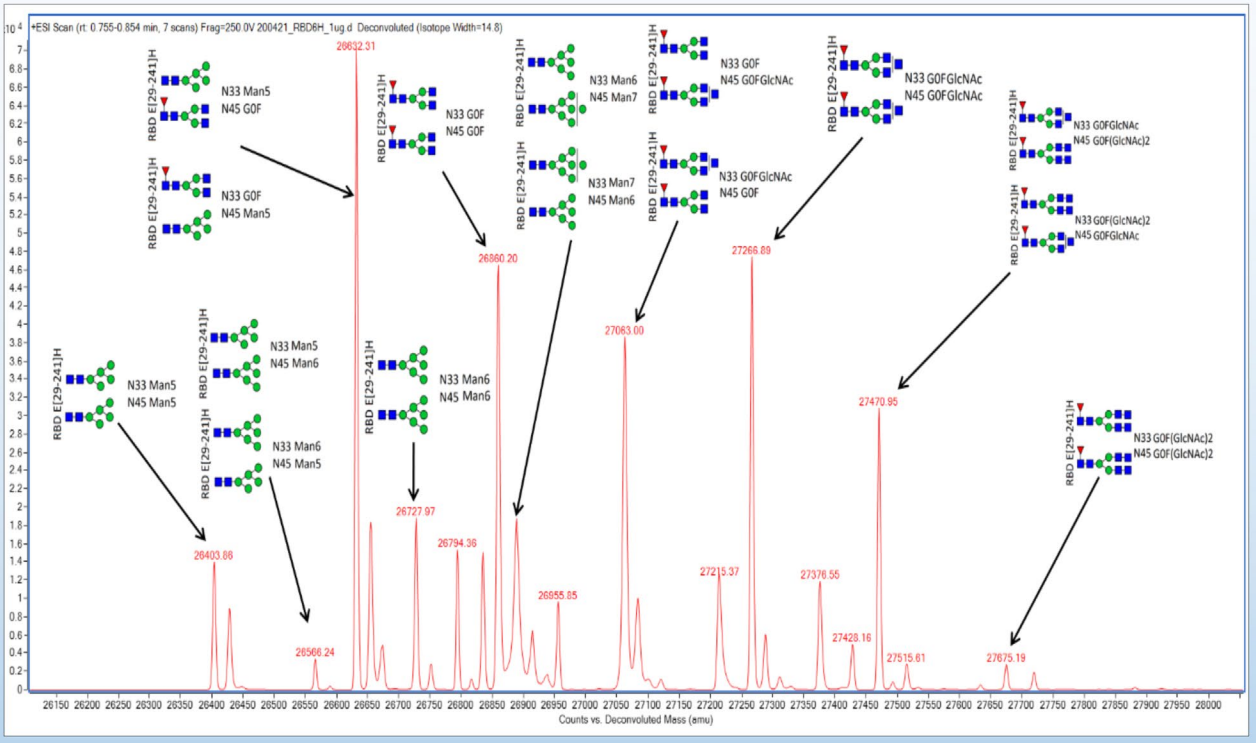
A master database for fingerprinting and mapping glycopeptides on the spike protein greatly simplifies this otherwise complex and time-consuming process.

Introduction

The complex glycosylation present in the Spike protein of the SARS-CoV-2 virus is believed to be involved in immune-evasion, and may make it more pathogenic. Many laboratories are expressing Spike protein for research purposes, and need reliable analytical workflows to fully characterize it. This is important because the glycosylation pattern is determined by the host cell expression system used and also by the growth conditions. These conditions are difficult to precisely control and can lead to batch-to-batch variations in the Spike protein being produced. The complexity of this glycosylation presents numerous technical obstacles to the full characterization of the protein. Intact mass analysis by reversed-phase liquid chromatograph mass spectrometry (LC-MS) is probably not feasible because the large number of possible glycans at each of the 22 glycosylation sites results in exceptionally high polydispersity and consequent dilution of signal. Even if it were possible, it would not yield the desired positional information for the glycans. Released glycan analysis is possible, but this also does not give positional information.

IDENTIFICATION, MAPPING AND RELATIVE QUANTITATION OF SARS-COV-2 SPIKE GLYCOPEPTIDES BY MASS-RETENTION TIME FINGERPRINTING

FIGURE 1: Spike receptor binding domain of SARS-CoV2 has two glycosylation sites (N343, N331) Several glycans can be associated with those sites.



This article discusses a novel approach that involves, as a first step, generation of a glycopeptide mass and retention time database using LC-Q-TOF MSMS. This database, which is in essence a collection of accurate mass-retention time fingerprints for each glycopeptide, can then be incorporated into a simple, powerful, rapid and robust characterization workflow.

The Challenge Posed by Spike Protein

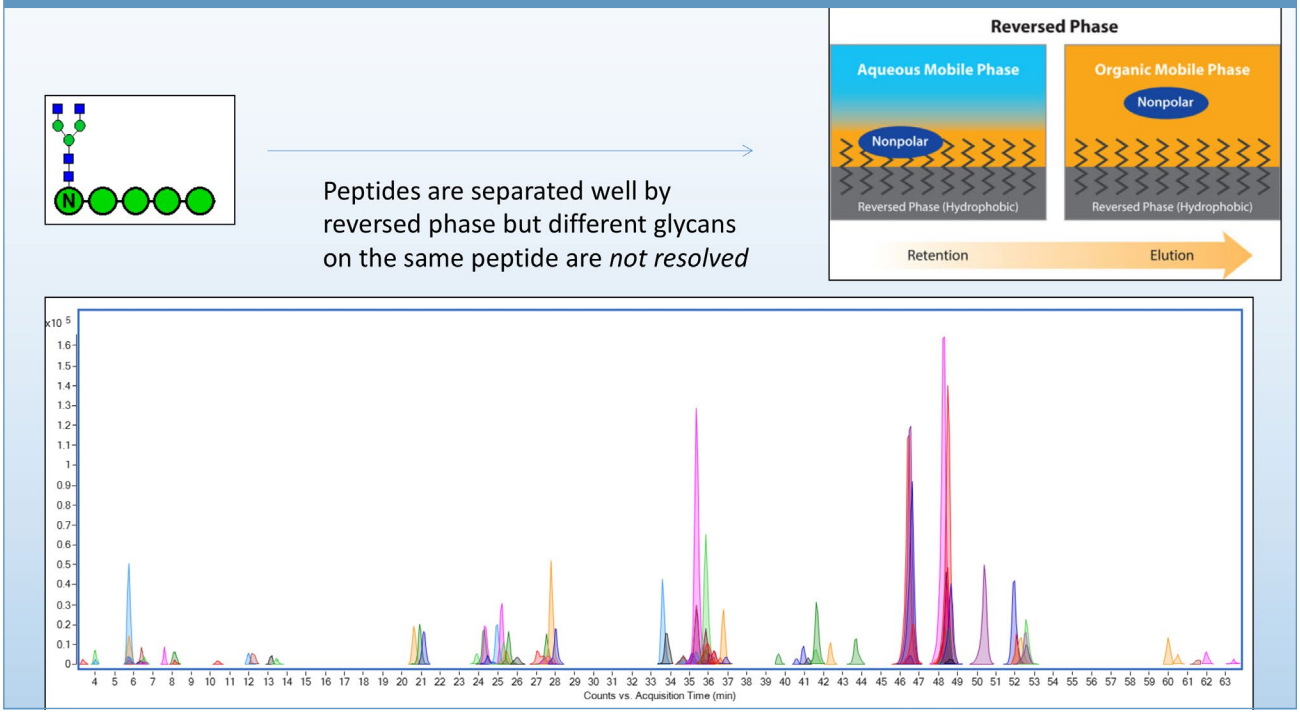
Full-length Spike protein exists as a trimer. The receptor binding domain (RBD) portion of the Spike monomer has just two glycosylation sites (N343 and N331), but each bearing several possible glycans. In spite of the low number of glycosylation sites for RBD, its glycosylation profile is complex and also displays sample to sample variation as can be seen by its intact mass spectrum (FIGURE 1). Because there are two sites, in some cases, there are two possible structures for a given mass, hence one cannot determine which of these structures is correct, precluding full identification.

The situation is more complicated for the full-length Spike protein. Here there are 22 glycosylation sites, and even assuming only six glycans per site, the total number of possible glycoforms would be around 130 billion, rendering intact mass analysis for the Spike protein unfeasible. Furthermore, such an analysis would not provide positional information.

An alternative would be glycopeptide analysis. In this case it is necessary to generate one glycosylation site per peptide,

IDENTIFICATION, MAPPING AND RELATIVE QUANTITATION OF SARS-COV-2 SPIKE GLYCOPEPTIDES BY MASS-RETENTION TIME FINGERPRINTING

FIGURE 2: Glycopeptide separation by reversed phase.



and for each site to be definitively mapped. Trypsin alone cannot be used because the two glycosylation sites within the RBD lack tryptic cleavage sites between them. In addition, because the precursor masses are unknown, one cannot use precursor database matching, therefore MS/MS is needed. There are, however, some issues that must be considered: the glycopeptide molecule is a hybrid consisting of the peptide backbone and a glycan structure with different optimal fragmentation conditions for each. Optimal fragmentation for the glycan leaves the peptide unfragmented; whereas optimal peptide fragmentation means the glycan information will be lost.

Improved Separation and Detection

As the use of multiple enzymes would present additional complications, researchers from the Centre for Medicines Discovery at Oxford University decided to go with a single enzyme, elastase, which cuts more frequently than trypsin and generates shorter peptides. More importantly, elastase generates a single peptide per glycosylation site, which is essential for unambiguous mapping. Since glycans are predominantly polar, they separate well using hydrophobic interaction liquid chromatography (HILIC), while the peptides, which are comparatively non-polar, separate best by reverse-phase (RP)-HPLC. The shorter peptides increase the total ion chromatogram (TIC) signal in the MS and are more readily resolved by RP-HPLC. The separation shows excellent resolution for peptides, but fails to separate individual glycans which coelute (FIGURE 2) appearing as multiple overlaid peaks.

IDENTIFICATION, MAPPING AND RELATIVE QUANTITATION OF SARS-COV-2 SPIKE GLYCOPEPTIDES BY MASS-RETENTION TIME FINGERPRINTING

FIGURE 3: Glycopeptide discovery workflow.

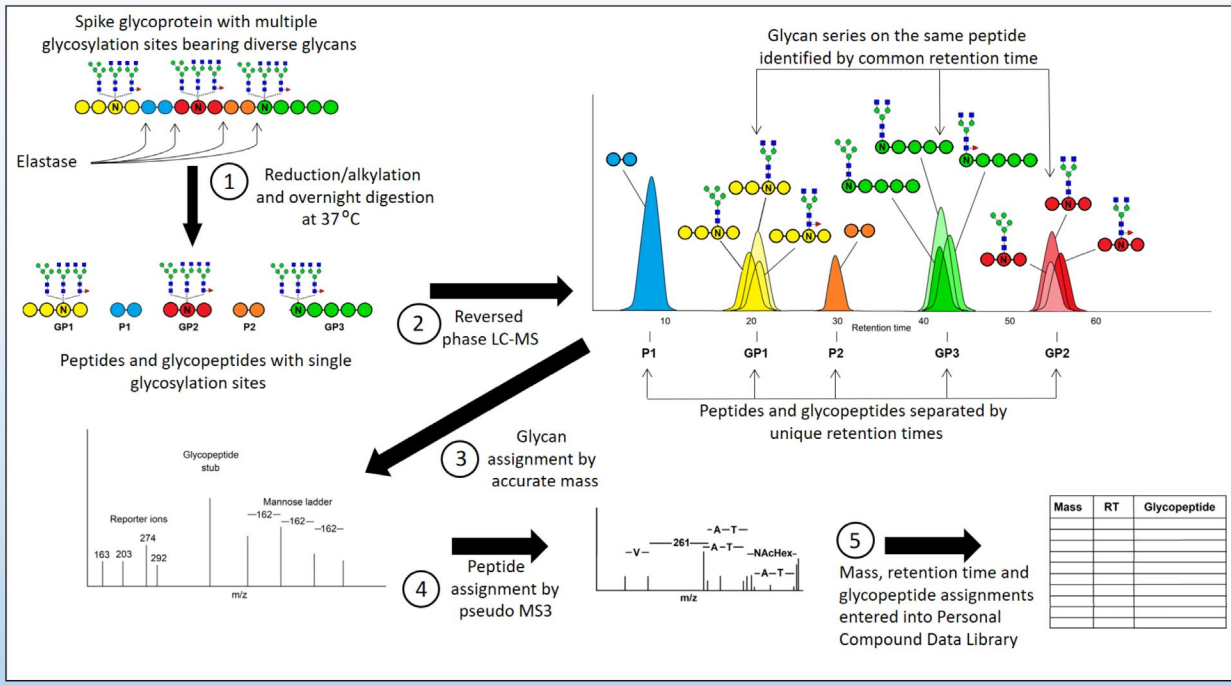
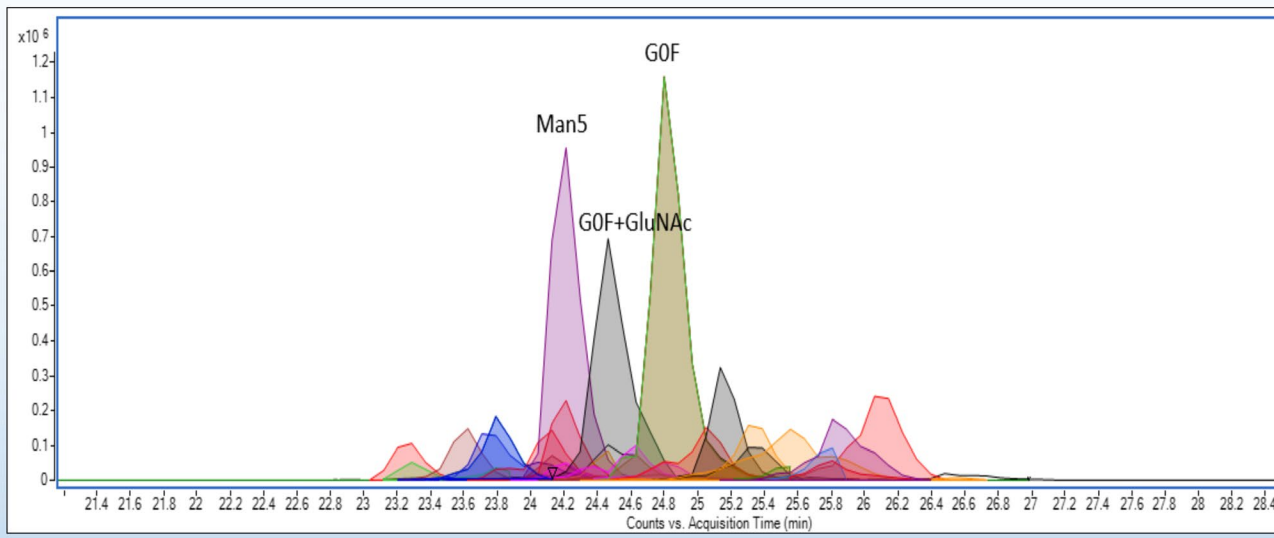
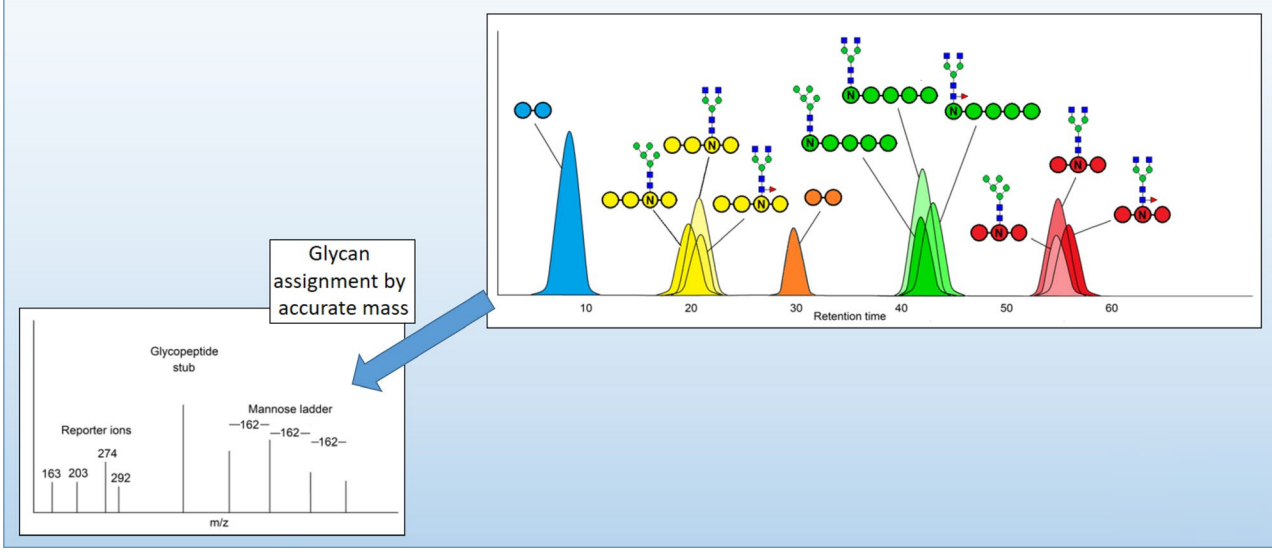


FIGURE 4: 27 GEVFNAT-Man5 glycopeptides all elute within ±2-minute retention time window



IDENTIFICATION, MAPPING AND RELATIVE QUANTITATION OF SARS-COV-2 SPIKE GLYCOPEPTIDES BY MASS-RETENTION TIME FINGERPRINTING

FIGURE 5: A mannose ladder is the structural feature of a mannose-containing glycan.



While this could be viewed as a problem, researchers exploited this feature to develop a new method to identify and map these glycopeptides. Using an AdvanceBio Peptide Mapping column from Agilent, the team separated the peptides and observed that the same peptides bearing different glycans elute at the same time. Thus, glycans with the same retention time could be mapped to the same peptide. Since glycans from the same peptide will have the same retention time, and their exact mass is known, the combination of accurate mass and retention time provides a fingerprint that allows complete glycopeptide identification.

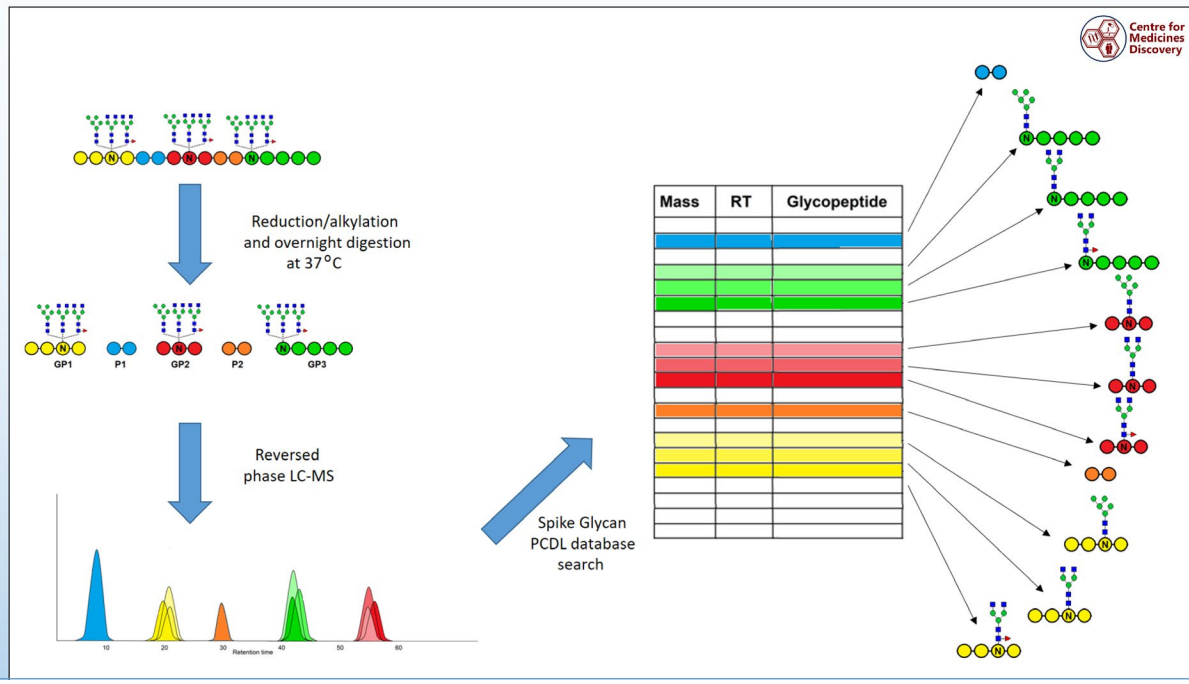
Building the Database

Glycopeptide discovery workflow. The workflow (FIGURE 3) starts with an overnight elastase digestion, producing a mix of peptides and glycopeptides. Next RP-HPLC is performed, where glycans from the same peptide co-elute and can be identified by their retention time. For example, the peptide GEVFNAT- has 27 different glycoforms, and the expanded ion chromatogram in FIGURE 4 shows that all of these elute within a four-minute retention time window. The next step is to identify the glycans using accurate mass. There are many species in the LC-MS/MS run. Glycopeptides are characterized by specific reporter ions representing the free sugars, the peptide stub as well as a mannose ladder (a structural feature of a mannose-containing glycan) as illustrated schematically (FIGURE 5). The deconvoluted mass spectrum reveals the full sequence of the glycan, with the individual sugars, to a mass accuracy within 10 ppm.

The second phase of the workflow involves peptide identification. This is done by pseudo MS3, a procedure that involves using a high fragmentor voltage on the MS instrument to perform in-source fragmentation of the glycopeptide and loss of the glycans. Selecting the peptide stub as the target mass

IDENTIFICATION, MAPPING AND RELATIVE QUANTITATION OF SARS-COV-2 SPIKE GLYCOPEPTIDES BY MASS-RETENTION TIME FINGERPRINTING

FIGURE 6: Mass-retention time fingerprinting workflow



in the quadrupole and performing collision-induced dissociation leads to the generation of the peptide sequence.

The key feature of this workflow involves populating a database combining accurate mass information with the specific retention time for all observed Spike protein glycopeptides. This database then enables the identification, mapping, and relative quantitation of any of these glycopeptides in subsequent analyses. (Chalk, R., Greenland, W. et al., "Identification, Mapping and Relative Quantitation of SARS-CoV-2 Spike Glycopeptides by Mass-Retention Time Fingerprinting", in press.) The database contains information on 140 observed glycopeptides over 13 glycosylation sites, with a further 6 glycopeptides families as yet unassigned to a specific site. In addition, it lists 306 inferred glycopeptides (assigned by presumed retention times and known theoretical masses). It should be noted that glycopeptides derived from sites N17, N234 and N331 contain the isobaric residues leucine or isoleucine and are thus not distinguishable by mass.

Mass-retention time fingerprinting workflow. With the database in hand, the workflow becomes much simpler. All that is required is an overnight elastase digestion on the Spike protein sample, reduction/alkylation, and then reverse phase LC-MS using a column like the AdvanceBio Peptide Mapping column. This takes about 60 minutes. The complete method is transferable to any HPLC system and any accurate mass instrument. Analysis for this work was performed on the Agilent 1290 Infinity II LC System, coupled to an Agilent 6530 Q-TOF mass spectrometer.

There is no need for MS/MS in this case since one can now interrogate the database using the mass retention times from the LC-MS run to assign the glycopeptide structures with a good level of confidence. This can be done automatically with the Agilent MassHunter Qualitative Analysis version B.07 and the Molecular Feature Extraction tool (FIGURE 6). A further advantage of this workflow is that it does not require a specialized knowledge of glycan analysis.

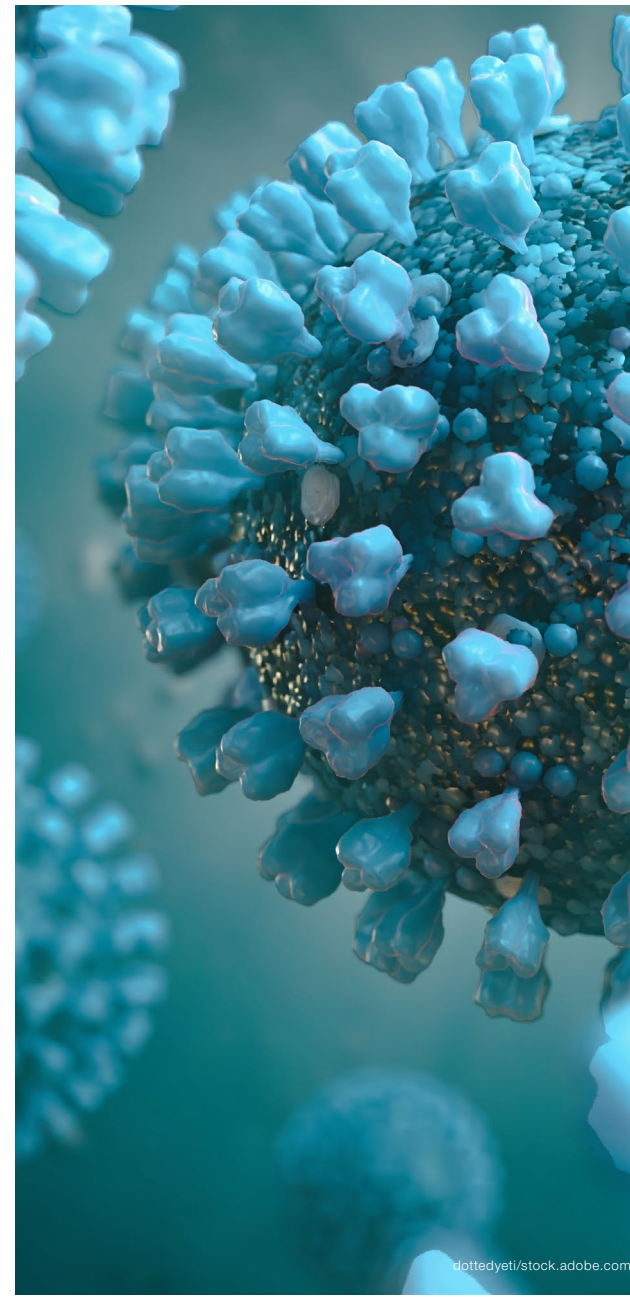
Conclusions

The mass-retention time database, resulting from detailed peptide discovery analysis, greatly simplifies the mapping of SARS-CoV-2 spike glycopeptides, eliminating the need for complex MS/MS analysis. It allows identification of glycopeptide structures by using only their LC retention times and accurate masses. This simple workflow is transferrable to a variety of HPLC-MS instruments and configurations. The detailed method will be freely available (Chalk, R., Greenland, W. et al., "Identification, Mapping and Relative Quantitation of SARS-CoV-2 Spike Glycopeptides by Mass-Retention Time Fingerprinting", in press.) and the database can be obtained online (Chalk, R., Greenland, W. et al., (2020). SARS-COV2 Spike Glycopeptide Mass-Retention Time PCDL database (Version 1) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3958218).

For Research Use Only. Not for use in diagnostic procedures. Agilent products are NOT authorized for COVID-19 testing, diagnosis, treatment, or mitigation. Agilent has not validated a product to detect the novel coronavirus.

Rod Chalk, Ph.D.

Head of Mass Spectrometry,
Centre for Medicines Discovery
Oxford University



RA44260.5453472222