Supplemental Information

Adapting data-independent acquisition for mass spectrometry-based protein site-specific N-glycosylation analysis

Kuan-Ting Pan[1]; Chen-Chun Chen[2]; Henning Urlaub[1,3,*]; Kay-Hooi Khoo[4,*]

[1]Bioanalytical Mass Spectrometry Group, Max Plank Institute for Biophysical Chemistry, Goettingen, Germany.

[2]Genomics Research Center, Academia Sinica, Taipei, Taiwan.

[3]Department of Clinical Chemistry, University Medical Center Goettingen, Goettingen, Germany.

[4]Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan.

* Corresponding authors

Contact:

Henning Urlaub: henning.urlaub@mpibpc.mpg.de

Kay-Hooi Khoo: kkhoo@gate.sinica.edu.tw

Contents

Supplemental Methods

Sample preparation for MS analysis

Filter-aided sample preparation (FASP)[1] was applied to denatured, reduced, alkylated and digested human serum IgM. Briefly, an aliquot (100 μg) of IgM was mixed with and denatured in 25 mM ammonium bicarbonate containing 8 M urea in an Amicon Ultra (Millipore). After a few centrifugation steps for complete buffer exchange, IgM was reduced and alkylated simultaneously by incubating with 5 mM TCEP and 10 mM iodoacetamide for 1 hour at room temperature in the dark. IgM was then digested sequentially by trypsin and Glu-C both in protein-to-enzyme ratio of 50:1 at 37°C for overnight and 8 hours, respectively. The resulting (glyco)peptides were cleaned up with C18 spin column (Hoefer) before subjected to mass spectrometric analysis. To mimic complex samples, yeast lysate was mixed with varying concentration of IgM digest as mentioned in the text.

Mass spectrometric settings for data-dependent acquisition

A Q Exactive HF (Thermo Fisher Scientific) coupled with Dionex UltiMate 3000 LC system was used in this study. Peptides were first trapped by loading buffer (2% acetonitrile, 0.02% TFA in water) at 10 μl/min on a home-made precolumn (ReproSil-Pur 120 C18-AQ, 5 μm, Dr. Maisch GmbH; 100 μm x 5 cm) and then separated on an analytical column (ReproSil-Pur 120 C18-AQ, 1.9 μm, Dr. Maisch GmbH, 75 μm x 30 cm, self-packed) with a 40-min linear gradient ramping from 5% to 40% Buffer B (80% ACN, 0.1%FA; Buffer A: 0.1% FA in water) at a constant flow rate of 300 nl/min.

For data-dependent acquisition, one full MS scan across the 350-1600 m/z range was acquired at a resolution setting of 120,000 FWHM (full width, half maximum) to select up to 20 most abundant peptide precursors of charge states 2 to 6 above a $2x10^4$ intensity threshold, at an isolation width of 1.4 Th. Precursors were fragmented by Higher Collision Energy Dissociation (HCD) with nitrogen at a normalized collision energy setting of 28%, and their product ion spectra were recorded with a start mass of 110 m/z at resolution of 30,000 FWHM. Automatic gain control (AGC) target value and maximum ion injection times for MS and MS/MS were 1×10^6 in 50 ms, and 1×10^5 in 64 ms, respectively. Selected precursor m/z values were then excluded for the following 30 s. For data-independent acquisition, one acquisition cycle consisted of one MS1 scan followed by 50 DIA swaths at a fixed isolation width of 16 Th to cover the mass range of m/z 800-1,600. Acquisition resolution was set to 30,000 for both MS1 and MS2 scans. AGC target value and maximum ion injection times for MS and MS/MS were 1×10^6 in 32 ms, and 2×10^5 in 32 ms.

Data processing for data-dependent acquisition

The MS raw files acquired from all of the LC-MS/MS DDA experiments were searched by Byonic (version 2.3.5)[2] against human Ig mu chain C region (P01871) as well as

immunoglobulin J chain (P01591). The search parameters were set as following: peptide tolerance = 10 ppm; fragment tolerance = 20 ppm; missed cleavages = −2. Allowed flexible modifications included protein N-terminal acetylation, methionine oxidation, deamidation on asparagine and glutamine. Carbamidomethyl cysteine was defined as a fixed modification. Built-in mammalian N-glycan database was used for glycopeptide search. Spectrum-level FDR of 0.1% was set as auto cut.

All identified peptide-to-spectrum matches (PSMs) were manually validated. PSMs were accepted if the following criteria were fulfilled:

1) At least two or more glycan oxonium ions including m/z 204.09 (HexNAc), m/z 274.09 (NeuAc-H_2O), m/z 292.08 (NeuAc), m/z 366.14 (HexNAcHex), m/z 512.20, (dHexHexNAcHex), m/z 528.19 ($HexNAc_2Hex$) and m/z 657.23 (HexNAcHexNeuAc) should be detected on the PSM.
2) Charge state and detected m/z of the precursor and its corresponding Y1 ion (peptide + HexNAc) should match the assigned glycopeptide.
3) At least 5 data points should be acquired on the MS1 XIC of each identified glycopeptide.
4) At least 3 b/y ions of the peptide backbone should be observed on the PSM.

Notable PSMs with Byonic-assigned glycan composition apparently in conflict with the detected glycan oxonium ions were not discarded (Table s1). Other than these additional filtering criteria, no further manual attempt was made to ensure true positives and accurate glycan composition assignment. Instead, the same search and acceptance criteria were standardized and used throughout for comparative purposes.

As retention time constraint of glycopeptides in built spectra library is critical to uphold reliable detection in following targeted analysis[3], we additionally required our final target glycopeptides to be eluted within 3 min of the median retention time of all identified glycopeptides sharing a common peptide core (Supplemental Table s1 and s2). By applying the same constraint to our targeted DIA analysis, we ensured a tight correlation between the retention times of target glycopeptides monitored in DIA analysis with that determined based on DDA evidence (Figure s4).

Supplemental Figures

Figure s1. Overview of the intensities of generated peptide fragment ions and glycan Y ions from various glycopeptides under different settings of Normalized collisional energy (NCE). For each glycosylation site, three most abundant glycoforms (as listed) were selected for generating the heat maps. Intensities of peptide fragments (left panel) or glycan Y ions (Y0/Y1/Y1f/Y2, right panel) were summed and then normalized against the median across all NCE settings. Color scale from green to red indicates low to high normalized intensities.

Figure s2. Differences in chromatographic retention time of IgM/IgJ glycopeptides bearing various glycans on the same peptide backbone. For each glycosylation site, only one unique peptide sequence was considered here, as shown on the figure. Each squire represents an unique glycoform on the indicated glycosylation sites.

Figure s3. Site-specific glycosylation on human serum IgM/IgJ. Intensities of MS1 precursor XICs of targeted glycopeptides were extracted and then normalized to the most abundant glycoform on each glycosylation site separately. Different glycoforms were categorized as complex, hybrid or high-mannose types based on the glycan compositions. The normalized relative intensities of glycoforms were color-coded in red scale for better visualization. N46, N209 and N272 on human serum IgM and N71 on IgJ carried mostly bi-antennary complex type N-glycans, whereas N279 and N439 on IgM bore mainly high-mannose type. Among them, N46 appeared to bear the most heterogeneous glycosylation pattern with dHex1Hex5HexNAc4NeuAc1 as the most abundant glycoform followed by dHex1Hex5HexNAc5NeuAc1 and dHex1Hex4HexNAc3NeuAc1. The determined site-specific glycosylation of IgM is consistent with a previous study[4,5].

Figure s4. Comparison of the observed retention times for all target glycopeptides in DDA and DIA analysis. Each blue square represents a targeted glycopeptide. The formula for the trend line and the value of R^2 are shown on the figure.

Supplemental Tables

Table s1. List of all peptide-to-spectrum matches reported by Byonic at 1% of false discovery rate.

Table s2. List of glycopeptides identified by DDA and Byonic search (A) and the further selected DIA target list representing unique site-specific glycoforms to be assayed.

Table s3. List of targeted glycopeptides identified by DIA at different amount of injected sample in the presence of spiked-in yeast lysates.

Reference:

(1) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. *Nat. Methods* **2009**, *6* (5), 359–362.

(2) Bern, M.; Kil, Y. J.; Becker, C. *Curr. Protoc. Bioinformatics* **2012**, *Chapter 13*, Unit13.20.

(3) Dong, Q.; Yan, X.; Liang, Y.; Stein, S. E. *J. Proteome Res.* **2016**, *15* (5), 1472–1486.

(4) Loos, A.; Gruber, C.; Altmann, F.; Mehofer, U.; Hensel, F.; Grandits, M.; Oostenbrink, C.; Stadlmayr, G.; Furtmüller, P. G.; Steinkellner, H. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (17), 6263–6268.

(5) Moh, E. S. X.; Lin, C.-H.; Thaysen-Andersen, M.; Packer, N. H. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (7), 1143–1155.

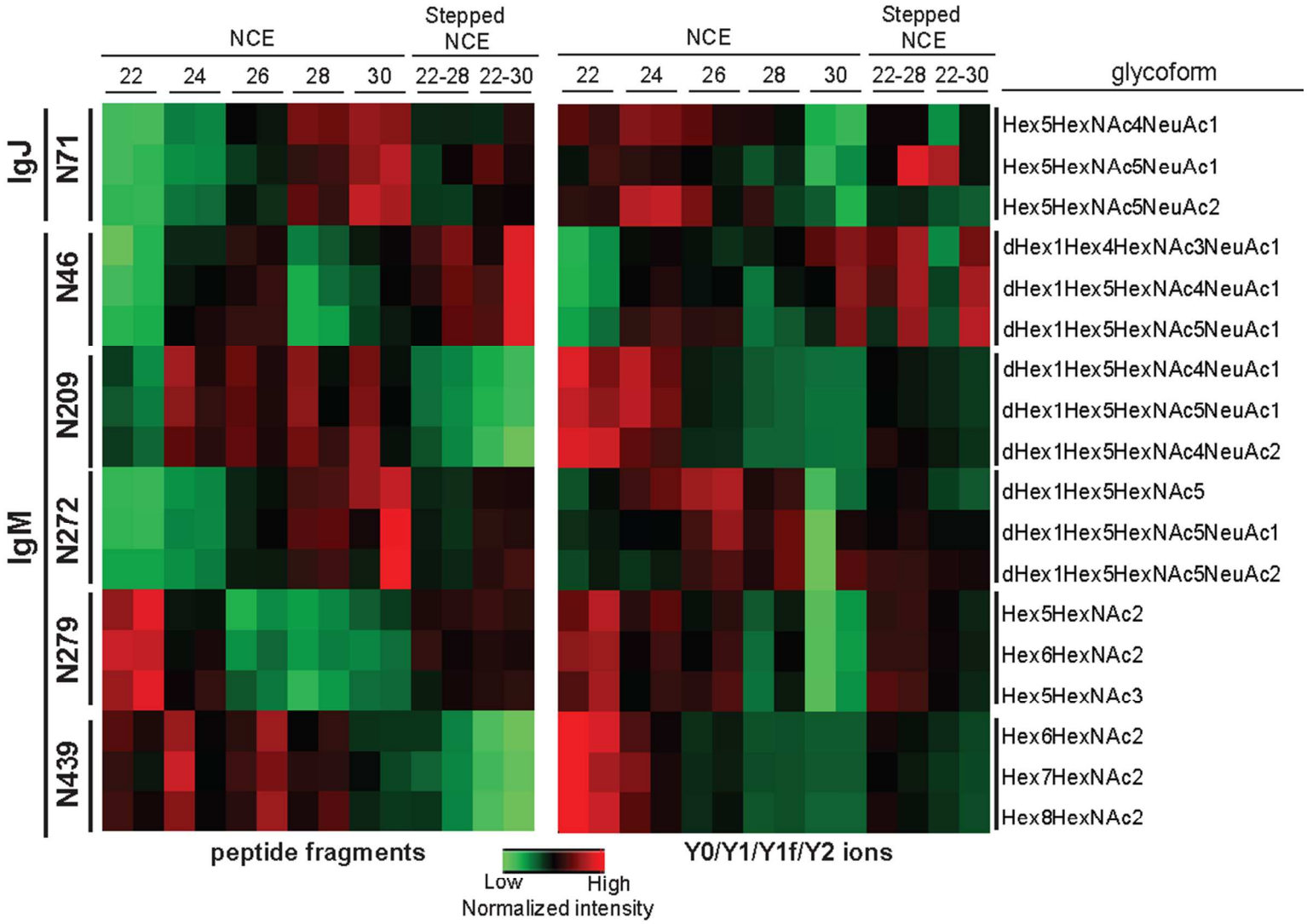
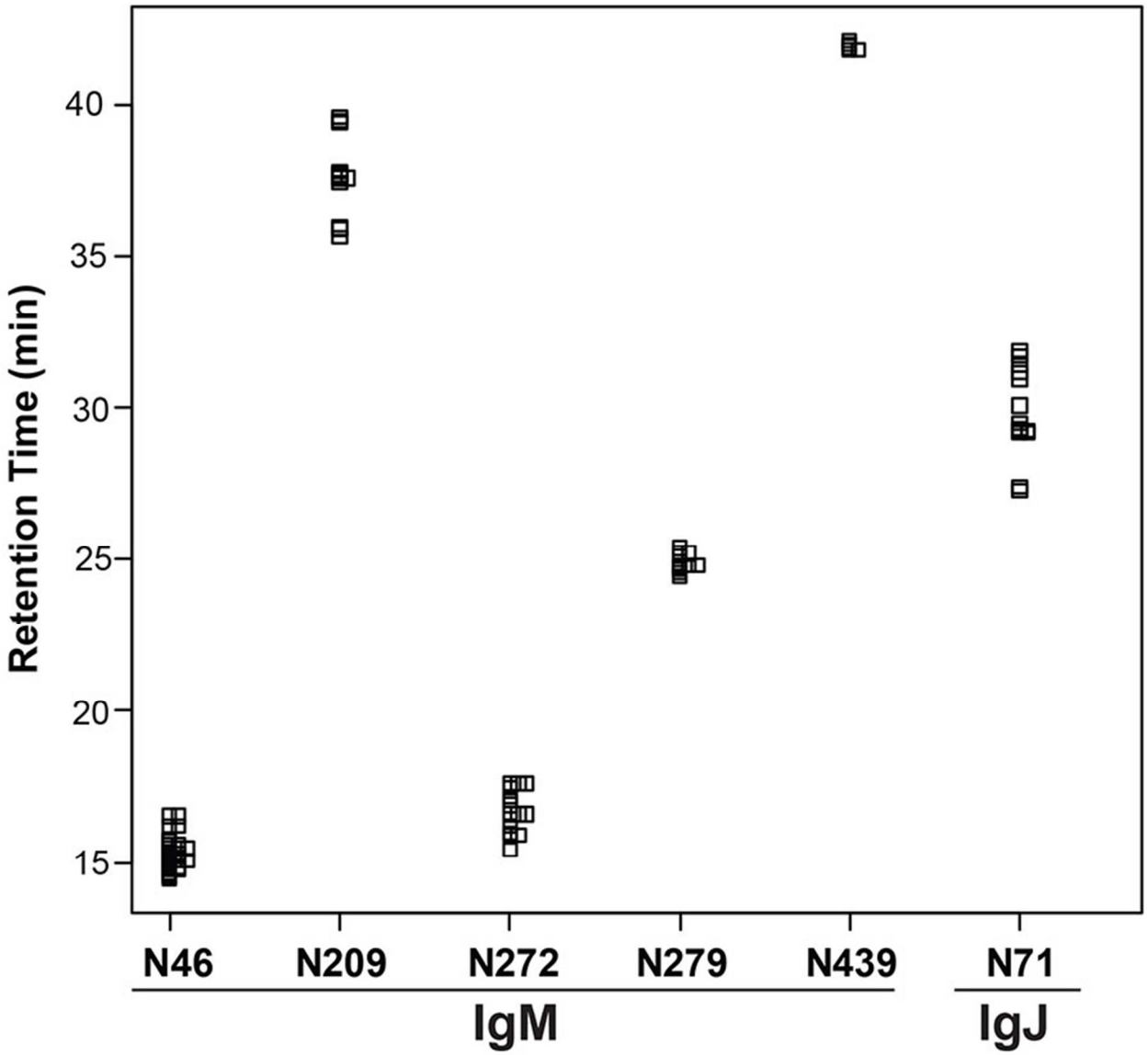

NCE
Stepped NCE
NCE
Stepped NCE
22
24
26
28
30
22-28
22-30
22
24
26
28
30
22-28
22-30
glycoform
IgJ
N71
IgM
N46
N209
N272
N279
N439
Hex5HexNAc4NeuAc1
Hex5HexNAc5NeuAc1
Hex5HexNAc5NeuAc2
dHex1Hex4HexNAc3NeuAc1
dHex1Hex5HexNAc4NeuAc1
dHex1Hex5HexNAc5NeuAc1
dHex1Hex5HexNAc4NeuAc1
dHex1Hex5HexNAc5NeuAc1
dHex1Hex5HexNAc4NeuAc2
dHex1Hex5HexNAc5
dHex1Hex5HexNAc5NeuAc1
dHex1Hex5HexNAc5NeuAc2
Hex5HexNAc2
Hex6HexNAc2
Hex5HexNAc3
Hex6HexNAc2
Hex7HexNAc2
Hex8HexNAc2
peptide fragments
Y0/Y1/Y1f/Y2 ions
Low
High
Normalized intensity

Figure s1.



Peptide core sequence:

IgM

N46: YKNN[46]SDISSTR

N209: GLTFQQN[209]ASSMCamVPDQDTAIR

N272: THTN[272]ISE

N279: SHPN[279]ATFSAVGE

N439: STGKPTLYN[439]VSLVMSDTAGTCamY

IgJ

N71: EN[71]ISDPTSPLR

Figure s2.

	Glycoforms	IgM					IgJ
		N46	N209	N272	N279	N439	N71
High mannose	Hex5HexNAc2	0.5%	0.0%	0.0%	100.0%	0.0%	0.0%
	Hex6HexNAc2	0.5%	0.0%	0.0%	40.7%	100.0%	0.0%
	Hex7HexNAc2	0.0%	0.0%	0.0%	21.1%	24.9%	0.0%
	Hex8HexNAc2	0.0%	0.0%	0.0%	18.7%	51.2%	0.0%
	Hex9HexNAc2	0.0%	0.0%	0.0%	7.6%	5.6%	0.0%
Hybrid	dHex1Hex4HexNAc3	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%
	dHex1Hex4HexNAc3NeuAc1	28.3%	0.0%	0.0%	0.0%	0.0%	0.0%
	dHex1Hex5HexNAc3	0.2%	0.0%	0.0%	6.7%	0.0%	0.0%
	dHex1Hex6HexNAc3	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
	dHex1Hex6HexNAc3NeuAc1	3.8%	0.0%	0.0%	0.0%	0.0%	0.0%
	Hex3HexNAc3	0.0%	0.0%	0.0%	6.6%	0.0%	0.0%
	Hex4HexNAc3	0.0%	0.0%	0.0%	9.7%	0.0%	0.0%
	Hex4HexNAc3NeuAc1	1.2%	0.0%	0.0%	0.0%	0.0%	0.0%
	Hex5HexNAc3	0.0%	0.0%	0.0%	17.4%	0.0%	0.0%
	Hex5HexNAc3NeuAc1	1.6%	0.0%	0.0%	0.0%	0.0%	0.0%
	Hex6HexNAc3	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%
	Hex6HexNAc3NeuAc1	2.2%	0.0%	0.0%	0.0%	0.0%	0.0%
	Hex6HexNAc4	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%
	Hex6HexNAc4NeuAc1	1.2%	0.0%	0.0%	0.0%	0.0%	0.0%
Complex	dHex1Hex3HexNAc5	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
	dHex1Hex4HexNAc4	1.2%	0.0%	0.0%	0.0%	0.0%	0.0%
	dHex1Hex4HexNAc4NeuAc1	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
	dHex1Hex4HexNAc5	2.2%	2.4%	19.8%	0.0%	0.0%	0.0%
	dHex1Hex4HexNAc5NeuAc1	3.1%	22.4%	5.7%	0.0%	0.0%	0.0%
	dHex1Hex5HexNAc4	7.4%	3.7%	0.0%	0.0%	0.0%	0.0%
	dHex1Hex5HexNAc4NeuAc1	100.0%	62.6%	13.0%	0.0%	0.0%	6.3%
	dHex1Hex5HexNAc4NeuAc2	6.2%	14.8%	11.0%	0.0%	0.0%	2.6%
	dHex1Hex5HexNAc5	6.3%	5.3%	16.9%	0.0%	0.0%	0.0%
	dHex1Hex5HexNAc5NeuAc1	61.4%	100.0%	100.0%	0.0%	0.0%	0.0%
	dHex1Hex5HexNAc5NeuAc2	1.3%	12.0%	24.5%	0.0%	0.0%	0.0%
	dHex1Hex6HexNAc5	0.1%	0.0%	0.4%	0.0%	0.0%	0.0%
	dHex1Hex6HexNAc5NeuAc1	1.0%	3.6%	2.8%	0.0%	0.0%	0.0%
	dHex2Hex4HexNAc4	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%
	dHex2Hex5HexNAc4NeuAc1	1.2%	0.0%	0.2%	0.0%	0.0%	0.0%
	dHex2Hex5HexNAc5	0.0%	0.0%	0.8%	0.0%	0.0%	0.0%
	dHex2Hex6HexNAc5	0.0%	0.0%	8.6%	0.0%	0.0%	2.7%
	dHex2Hex6HexNAc5NeuAc1	0.0%	0.0%	1.2%	0.0%	0.0%	0.0%
	Hex4HexNAc4	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
	Hex4HexNAc5	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%
	Hex5HexNAc4	0.0%	0.0%	0.0%	0.0%	0.0%	2.1%
	Hex5HexNAc4NeuAc1	8.0%	3.8%	0.0%	0.0%	0.0%	100.0%
	Hex5HexNAc4NeuAc2	0.9%	0.0%	0.0%	0.0%	0.0%	44.4%
	Hex5HexNAc5	1.5%	0.0%	0.0%	0.0%	0.0%	8.5%
	Hex5HexNAc5NeuAc1	8.6%	19.7%	0.0%	0.0%	0.0%	32.7%
	Hex5HexNAc5NeuAc2	0.6%	4.6%	0.5%	0.0%	0.0%	5.0%
	Hex6HexNAc5NeuAc1	0.0%	1.0%	0.0%	0.0%	0.0%	1.3%

Figure s3.

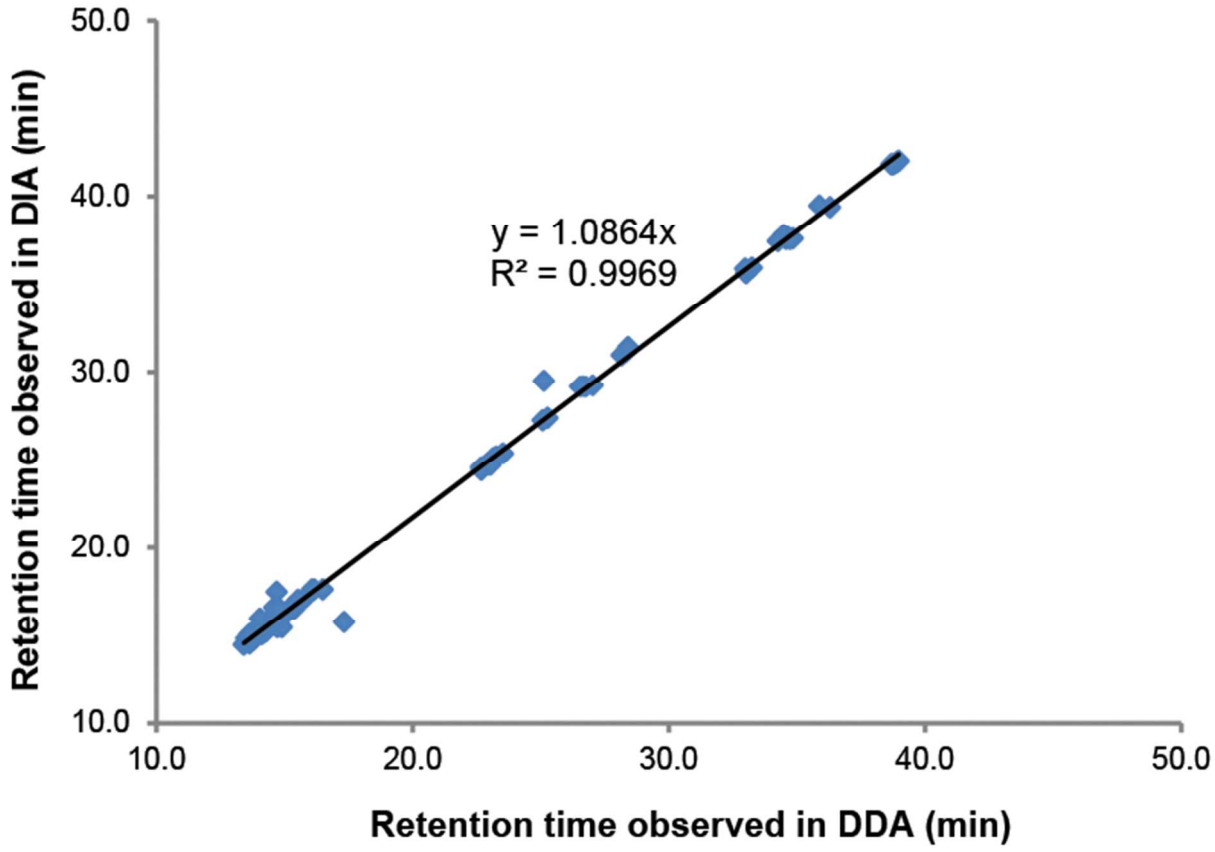

50.0
40.0
30.0
20.0
10.0
Retention time observed in DIA (min)
y = 1.0864x
R² = 0.9969
10.0
20.0
30.0
40.0
50.0
Retention time observed in DDA (min)

Figure s4.