Electronic Supplementary Material (ESI) for Analyst.

SUPPORTING INFORMATION

A New Approach for Identifying Positional Isomers of Glycans Cleaved from Monoclonal Antibodies

Irina Dyukova,[a] Ahmed Ben Faleh,[a] Stephan Warnke,[a] Natalia Yalovenko,[a] Vasyl Yatsyna,[a,b] Priyanka Bansal,[a] and Thomas R. Rizzo*[a]

[a]Laboratoire de Chimie Physique Moléculaire, École Polytechnique Fédérale de Lausanne, EPFL SB ISIC LCPM, Station 6, CH-1015 Lausanne, Switzerland.

[b]University of Gothenburg, Department of Physics, 412 96 Gothenburg, Sweden

Table of Contents

I. Conversion efficiency of the chemoenzymatic reaction.

The synthesis of G1(α1-6)F (Fig. 2) was performed in a total volume of 50 µL containing 0.27 mM of an acceptor G-NGA2 N-linked glycan (Dextra Laboratories, UK), 0.54 mM of GDP-Fuc (guanosine 5'-diphospho-β-L-fucose sodium salt) (Sigma-Aldrich), a MES (2-(N-morpholino)ethanesulfonic acid) buffer solution (100 mM, pH 7.0), and 0.12 mg/mL of recombinant human α1,6-fucosyltransferase FUT8 (Creative BioMart, USA). The reaction was incubated overnight at 37 °C. To determine the optimal incubation time of the reaction, samples were collected at 1, 2, 3, 4, 6, 8, 16 and 48 h. The reactions were then quenched by adding 30 µL of ice-cold acetonitrile to the mixture, and subject to LC analysis and purification. We monitored the efficiency of the enzymatic reaction with a UPLC (ACQUITY™ H-Class Plus, Waters, UK) coupled to Micromass Q-TOF Premier (Waters, UK). We used a Waters XBridge BEH amide column (130 Å, 3.5 μm, 4.6 × 150 mm) at 0.5 mL/min at 60 °C. A gradient elution (%A: 36-44% within 20 min) was performed with Solvent A (100 mM ammonium formate, pH 4.5) and Solvent B (Acetonitrile). The conversion efficiency (see Table S1) was calculated as % = Product peak area/ (Product peak area + Substrate peak area) × 100.

Table S1. Conversion efficiency at different incubation times of the reaction

	G1(α1-6)F
Time	**Conversion (%)**
1h	17.1
2h	18.2
3h	19.6
4h	21.5
6h	26.8
8h	30.3
16h	43.7
48h	57.7

Our assumption is that we start with a 50/50 mixture of the positional isomers of G-NGA2. Since we are producing only one of the positional isomers, the conversion efficiency should be less than 50%. According to the results in Table S1, the best incubation time of the reaction was 16h. It can be seen from Table S1 that a conversion rate of more than 50% was observed after 48 h (i.e., 57%), possibly due to the reaction of G(α1-3)-NGA2 and formation of the G1(α1-3)F isomer. For this reason, we chose 16h as the optimal incubation time of the reaction assuming that G1(α1-6)F is exclusively formed.

II. IR spectral comparison using a machine-learning approach.

The IR spectra were compared using an algorithm that combines principal component analysis (PCA) and automatic cluster detection to identify and assign the positional isomers of G1F. The algorithm makes use of the *'scikit-learn'* python software library using both the 'PCA' fit function from the decomposition library and the 'kmeans' *fit* and *predict* functions from the cluster library (Stow et al. 2017). As a first step, PCA is used to reduce the dimensionality of the considered IR spectra to a few principal components that sufficiently describe the original data set (in this case 3 principal components). Following this, the different reference spectra are automatically identified and classified in the so-called 'clusters' in principle-component space. Finally, the *predict* function automatically assigns the analyte spectra to the corresponding references. This approach allows for a rapid and accurate identification and assignment of the IR spectra to the different isomers of the considered molecule. Moreover, no user interpretation is required, making the analysis fully automated and thus putting machine learning at the service of glycan analysis.

III. Integrated peak areas of each signal in the ATDs of G1F.

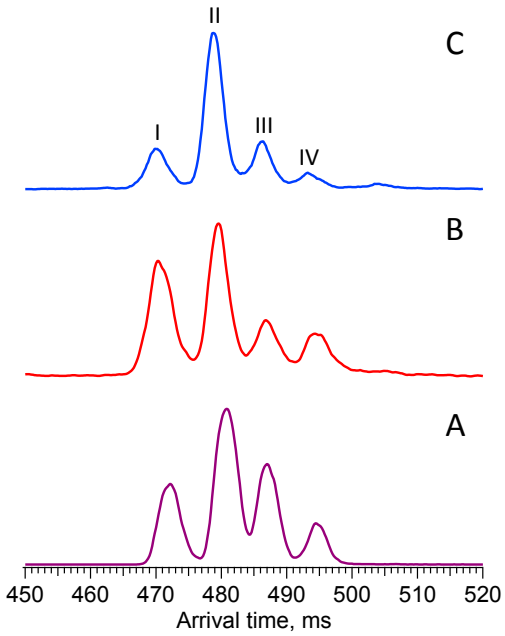

Figure S1. Integrated peak areas after gaussian peak fitting of each signal in the ATDs of G1F standard (A), G1F from HEK-293 cell line (B) and G1F from CHO cell line (C).

G1F	Integrated peak areas, %				G1(α1-3)F/G1(α1-6)F
	G1(α1-3)F		G1(α1-6)F		
	Peak I	Peak II	Peak III	Peak IV	
Standard	22	43	26	9	65/35
HEK-293	33	37	16	14	70/30
CHO	17	58	18	7	75/25

Sto w SM, Causon TJ, Zheng XY, Kurulugama RT, Mairinger T, May JC, Rennie EE, Baker ES, Smith RD, McLean JA, *et al.* 2017. An Interlaboratory Evaluation of Drift

Tube Ion Mobility-Mass Spectrometry Collision Cross Section Measurements. *Anal. Chem.*, 89:9048-9055.