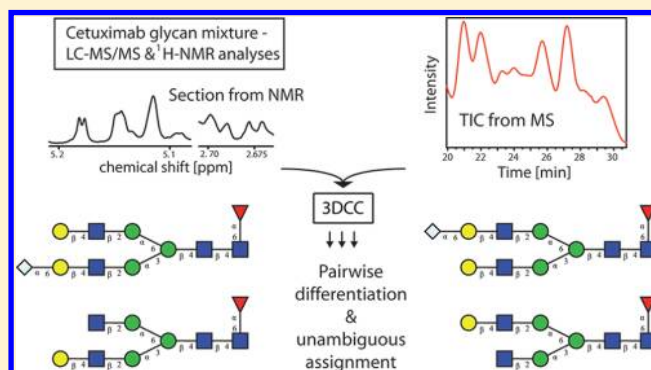


Unambiguous Characterization of *N*-Glycans of Monoclonal Antibody Cetuximab by Integration of LC-MS/MS and ^1H NMR Spectroscopy

Alena Wiegandt and Bernd Meyer*

Organic Chemistry, Department of Chemistry, University of Hamburg, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany

ABSTRACT: Monoclonal antibodies are most rapidly emerging as therapeutic drugs for the treatment of cancer and of various other diseases such as autoimmunity or inflammation. Recently, it was found that nonhuman glycosylation of recombinant antibodies can cause tremendous problems for some patients. Therefore, unambiguous assignment of the glycosylation pattern of therapeutic antibodies is of high importance for assessment of human compatibility. Here we present results from a broad and detailed *N*-glycan analysis of the therapeutic antibody cetuximab by LC-MS/MS analyses tightly integrated with ^1H NMR to obtain unambiguous structures. Thirty-seven *N*-glycan compositions were identified by LC-MS(/MS). Subsequently, ten abundant structures were structurally characterized by applying the recently introduced method called three-dimensional cross correlation (3DCC). It was possible to extract NMR spectra of pure *N*-glycans that were heavily overlapping in a chromatographic separation by mathematically dissecting the NMR spectra obtained from chromatographic fractions. Even mass isobaric structures that differ only in the branching position of one monosaccharide unit were distinguished and characterized. We also developed an improvement of the 3DCC method by introducing singular value decomposition (SVD) for processing of the data. The smallest amount of the *N*-glycan characterized by 3DCC was approximately 400 pmol (836 ng). Among the ten unambiguously identified glycans, six *N*-glycans, representing 24% of all detected glycans, possess the immunogenic α -1,3-Gal epitope and/or *N*-glycolylneuraminic acid. These results illustrate the importance of integrated use of LC-MS(/MS) and ^1H NMR for the glycome analysis of biopharmaceuticals in research, development, and quality control.



With more than 30 clinically approved monoclonal antibodies (mAbs) and mAb derivatives currently used in therapy within the US and the EU and 30 additional mAbs being investigated in late clinical trials, IgGs represent the fastest growing class of therapeutics.^{1–3} The antibodies are developed as treatment for various diseases such as autoimmunity, cancer, inflammation, infection, and cardiovascular diseases.⁴ In 2009, mAbs used in the medical sector generated a revenue of 36 billion dollars in the US.⁵ Because several therapeutic mAbs will be off patent in 2016, biosimilar products are already in production.⁶ These generic versions have to be verified in terms of structural similarity to ensure the biosimilar's quality and safety. Glycosylation of proteins has long been known to significantly affect various properties, e.g., protein solubility, folding, and stability as well as regulatory functions.⁷ Glycans mediate the cellular organization of complex organisms by regulating cell–cell, cell–matrix, or cell–molecule interactions and contacts between different organisms, respectively.⁸ Furthermore, specific alteration in glycan expression is associated with pathogenic processes such as inflammation or formation of tumors.⁹ Hence, the influence of glycosylation on potency and safety of biopharmaceuticals is not surprising. Also, in monoclonal antibodies an enhanced antibody-dependent

cellular cytotoxicity (ADCC) was observed for nonfucosylated glycan core structures as well as bisecting *N*-acetylglucosamines.¹⁰ Another very important aspect is the immunogenicity of several nonhuman glycan motifs. α -1,3-Gal-epitopes and the glycan moiety *N*-glycolylneuraminic acid can both cause anaphylaxis. In fact, recent studies showed hypersensitivity reactions in over 30% of patients that received cetuximab.¹¹

Therefore, analysis of the glycome is important for diagnostic biomarkers and also for development and quality assurance of biopharmaceuticals, such as cetuximab, as well as for the approval of biosimilar mAb versions. Structures of glycans located on antibodies vary greatly with the expression system. The immense heterogeneity of glycan structures results in very low concentrations of individual compounds. This analysis of glycans with low abundance was possible due to immense progress in analytical instrumentation.^{12,13} The proportion of patients with a hypersensitive reaction against cetuximab illustrates the urgent need for reliable glycosylation analysis. Cetuximab is used for the treatment of head, neck, and colorectal cancer.

Received: December 13, 2013

Accepted: April 11, 2014

It blocks the EGFR activation and is often used in combination with chemotherapy, especially in case of colorectal cancer.^{14,15} Studies revealed that many different *N*-glycans are linked to cetuximab at four glycosylation sites: one at each ⁸⁸Asn in the Fab domain and one at each ²⁹⁹Asn in the Fc domain.¹⁶ Recent studies revealed an incorrect primary structure of cetuximab in the databases and therefore demonstrate the necessity of extensive characterization of amino acid sequences as well as of glycoforms to ensure the safety of first generation and bio-similar versions.⁶

Different schemes exist in literature for the analysis of *N*-glycans linked to proteins. Most of the time, LC-MS(/MS) techniques are employed. Historically, the bottom-up approaches starting from the glycoprotein followed by enzymatic cleavage either into glycopeptides or glycans followed by analysis of the compositions dominate the field.^{17,18} More recently, top-down approaches that directly analyze the intact glycoprotein by (LC-)MS/MS techniques have gained ground.¹⁹ In addition, a large amount of data is present for the analysis of isolated glycans by NMR spectroscopy.²⁰ Structurally similar analytes often overlap in chromatographic separations. We developed a new method called 3DCC (three-dimensional cross correlation) to overcome this problem.^{21,22} Another development that has benefited glycan analysis is the development of the porous graphitized carbon (PGC) phase for chromatographic separation, e.g., even mass isobaric glycans that differ only in one branching position can be separated.²³ Combination of PGC separation of *N*-glycans with high precision MS analysis as well as characteristic fragmentation patterns (B- and Y-ions) has also tremendously benefited structural elucidation.^{24–28} In recent years it was also shown that NMR is highly sensitive. Fifteen picomoles of material is sufficient for ¹H NMR spectra so that the technique gains importance for glycome analysis.¹³ Furthermore, NMR is a nondestructive method and produces information on the linkage pattern, chirality, and anomericity. By means of the structural reporter group concept, it is possible to assign anomeric protons, H-2 protons of mannose, H-3, H-4, and H-5 protons from galactose, H-3 protons from neuraminic acids, H-5 and methyl protons from fucose, *N*-glycolylneuraminic acids as well as *N*-acetate signals from *N*-acetylneuraminic acids and *N*-acetylhexosamines from one-dimensional ¹H NMR spectra.²⁹

The use of MS and NMR analyses is a common strategy to obtain reliable data for structure interpretation. 3DCC is a procedure that no longer treats these data as independent but integrates them with LC information to allow a consistent characterization with standard laboratory equipment and data analysis software, developed in preceding work in this laboratory.^{21,22} 3DCC allows the assignment of glycans that are strongly overlapping in LC. By mathematical correlation of NMR signals with information from the LC-run (retention time and accurate *m/z* values), it is possible to extract NMR spectra of pure compounds. Our previous work applied the 3DCC procedure to analyze well-known *N*-glycan structures of the model system bovine fibrinogen, of which many simultaneously eluting *N*-glycans with differing monosaccharide compositions could be structurally characterized by the integrated use of mass spectrometry and NMR spectroscopy.

Here we show the tight integration of NMR and 3DCC with LC-MS information for an unambiguous *N*-glycan characterization of the monoclonal antibody cetuximab. The immunogenic α -1,3-Gal-epitope and the immunogenic glycan motif *N*-glycolylneuraminic acid were both unambiguously identified

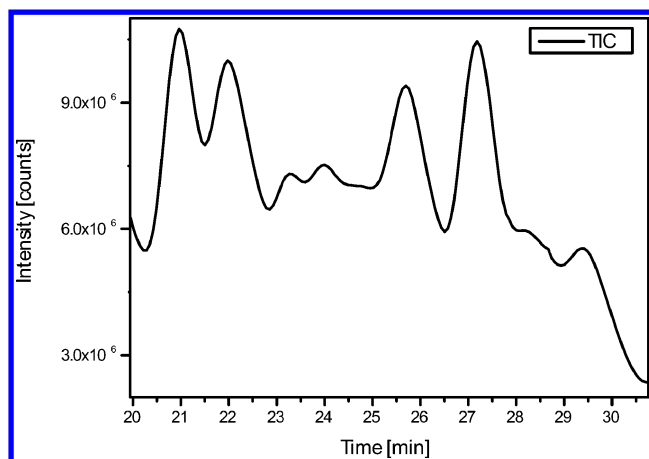


Figure 1. LC-MS chromatogram of underivatized *N*-glycans enzymatically cleaved from cetuximab separated on a PGC column. Ammonia was used as coeluent to avoid separation of α - and β -anomers. Detail of the total ion chromatogram (minute 20 to 30) is shown. Thirty-seven *N*-glycans were identified by LC-MS(/MS)-NMR.

with LC-MS(/MS)-NMR integration realized by the 3DCC method. The 3DCC procedure was improved in terms of robustness and effectivity by implementing SVD (singular value decomposition) as an optimized scheme for the least-squares optimization. The quality of the new method in comparison to the previously published version is being demonstrated for the glycans of cetuximab to provide new insights into biological relevant glycan structures. Ten *N*-glycans strongly overlapping in chromatographic separation and even mass isobaric structures with equal monosaccharide composition were unambiguously identified.

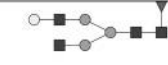


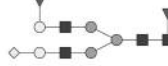
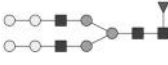
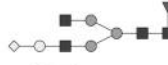
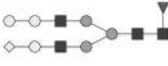






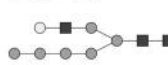




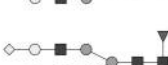

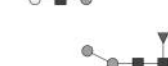
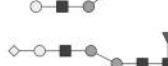
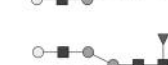

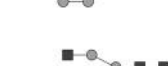
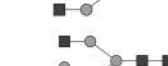
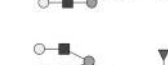
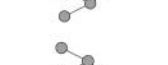


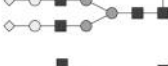
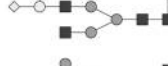


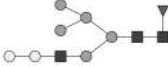


MATERIALS AND METHODS

Reagents. Cetuximab (generic name: Erbitux) infusion solution (5 mg/mL) from MerckSerono was purchased. For enzymatic digestion, Trypsin Gold (Mass spectrometry grade) was purchased from Promega and recombinant *N*-Glycosidase F from Roche Diagnostics. 1,4-D/L-Dithiothreitol (DTT) and iodoacetamide were purchased from AppliChem. If not differently denoted all other reagents were products from Sigma-Aldrich.

Sample Preparation. *N*-Glycans were enzymatically cleaved from erbitux with trypsin and *N*-glycosidase F. Briefly, 580 μ L of Erbitux infusion solution ($c = 5$ mg/mL) was purified from excipients by centrifugal filters (Amicon Ultra, Millipore), freeze-dried, and dissolved in 6 M urea (Grüssing). Prior to tryptic digestion, the glycoprotein was reduced by addition of DTT for 10 min at 60 °C and alkylated with iodoacetamide for 30 min at room temperature in the dark. The solution was diluted with 100 mM NaHCO₃, pH 8.3 (Grüssing), and trypsin was added in a protein:enzyme ratio of 75:1. After 16 h of incubation at 37 °C, trypsin was inactivated for 10 min at 99 °C. Glycans were enzymatically cleaved from erbitux by incubation with 0.58 U of *N*-Glycosidase F for 48 h at 37 °C. Subsequently, the digest was freeze-dried, dissolved in bidistilled H₂O and subjected to solid-phase extraction to separate the glycans from the tryptic peptides (SPE cartridges, Agilent Technologies).

PGC-LC-ESI-q/TOF. The resultant glycan mixture was subjected to PGC-LC-MS/MS. The analysis was performed on an ESI-q/TOF (maXis, Bruker) coupled to an Ultimate 3000

Table 1. Compositions of the *N*-Glycans of Cetuximab Characterized by LC-MS Based on High Mass Accuracy and Several of Them by CID-ESI Fragment Spectra (cf., Figure 2)

Glycan composition ^a	Identified by		Quant.	Glycan composition ^a	Identified by		Quant.
	MS [Da]	Acc. [ppm]			MS [Da]	Acc. [ppm]	
	1624.604*	3.7	31.4		2475.888*	1.9	0.4
	1462.552*	3.7	19.6		2239.802	1.5	0.4
	2110.762*	2.7	13.9		1931.692	1.9	0.3
	2255.800*	2.3	7.7		2313.838	0.6	0.3
	1786.656*	2.7	7.5		2620.931*	1.5	0.3
	1234.436	2.2	2.9		2637.945*	1.6	0.3
	1259.465	1.0	2.0		1761.621	1.6	0.3
	1745.626*	1.7	1.9		2296.8198	0.3	0.2
	2094.763*	1.3	1.5		2458.868	0.6	0.1
	2093.746*	2.2	1.5		1827.678*	0.2	0.3
	1421.519	1.1	0.4		1890.758	2.2	0.2
	1583.574*	1.3	1.0		1665.625	0.6	0.2
	1478.542*	1.4	1.0		1437.514	1.8	0.2
	2151.782	1.3	0.8		1396.486	0.7	0.2
	2400.832*	1.9	0.6		2077.751	4.5	0.1
	1989.734*	0.8	0.6		2052.718	1.9	0.1
	1907.680*	1.5	0.6		2928.016*	1.9	0.1
	2782.984*	1.4	0.5		1316.486*	2.3	<0.1
	1599.566*	1.4	0.5				

^aSquare: *N*-acetylhexosamine; triangle: fucose; tilted square: *N*-glycolylneuraminic acid, dark grey circle: mannose; light grey circle: galactose.
 *Analyzed by CID-ESI fragment spectra. The two structures highlighted in boldface each contain two isomers (cf., Table 2) identified by 3DCC (cf., discussion in the next subsection).

UPLC (Dionex). MS data were acquired in positive-ion-mode with an m/z range of 100 to 2975 and a spectra rate of 1.0 Hz. The capillary voltage was set to 5000 V, nebulizer pressure to 3 bar, dry gas to 7 L/min, and drying temperature to 190 °C. The three most intense ions (if an intensity threshold of 1000 counts was minimally reached) were automatically chosen for fragmentation. CID fragmentation occurred in the collision cell with nitrogen as collision gas and a collision energy of 40 to 55 eV depending on charge state and m/z range. Singly charged ions were excluded, and ions that were fragmented four times were excluded for 1 min. The separation on PGC was carried out with a HyperCarb column (4.6 × 150 mm, 3 μ m particle size, Thermo Fisher) with ultrapure water containing 65 mM ammonium formate and 10 mM ammonia as solvent A and acetonitrile containing 0.1% FA as solvent B using the following gradient: 0 min, 5% B; 2 min, 8% B; 53 min, 35% B; 57 min, 90% B; 61 min, 90% B; 63 min, 5% B; 65 min, 5% B. Ammonia was added to avoid separation of α - and β -anomers. The column temperature was set to 40 °C and the flow rate to 800 μ L/min. The liquid flow was split with a postcolumn splitter, 5% were directly injected into the mass spectrometer, and 95% of the LC flow were further fractionated in well plates in 15 s (~190 μ L) fractions. Calibration was performed by a prerun calibration (phosphazine mix, Agilent) and a lock mass calibration.

¹H NMR. Glycan-containing fractions from the deep well plate were freeze-dried and redissolved in 180 μ L of D₂O each. NMR spectra of all fractions were acquired on an Avance 700 MHz NMR spectrometer equipped with a cryoprobe. All spectra were recorded with 4 k scans utilizing an excitation sculpting technique for solvent suppression, a spectral width of 10 ppm, an acquisition time of 2.34 s per scan, and a relaxation delay of 1 s.

3DCC. 3DCC spectra were generated as described previously.²² Briefly, NMR data were imported into MATLAB utilizing the script `rbnmr` by Nils Nyberg (SLU, 2001-05-02, `nn@farma.ku.dk`). Signals of impurities and residual signals of suppressed solvent were set to zero. Exported EIC data points (MS data) were reduced using Origin 8.6 to the number of collected fractions (= number of NMR spectra) and imported into MATLAB from Excel. For quantification, the integrals of the EICs were normalized to the maximum intensity of the NMR spectra. The extracted 3DCC spectra were calculated by least-squares optimization and by Pearson correlation coefficients as described previously. Furthermore, a new variant calculates 3DCC spectra by singular value decomposition using the MATLAB SVD routine. A detailed description for all variants can be found in Results and Discussion.

RESULTS AND DISCUSSION

The *N*-glycans of cetuximab were analyzed and quantified using different analytical methods. The glycans were prepared and analyzed by LC-MS(/MS) and integration of LC-MS with NMR spectroscopy.

***N*-Glycans from Cetuximab Analyzed by LC-MS/MS.** Cetuximab was treated with 6 M urea and subsequently with DTT and iodoacetamide to produce a denatured protein. Enzymatic cleavage of the glycans from the protein was achieved by *N*-glycosidase F. The oligosaccharides were separated by PGC-HPLC. Within a 10 min period, 37 underivatized *N*-glycans eluted from the column (Figure 1, Table 1). The amount of the *N*-glycan with the lowest abundance was approximately 38 pmol (50 ng). The compositions of the sugars were identified by ESI-q/TOF mass spectrometry yielding a precise mass with

97% showing a deviation of less than 4 ppm. Among the most abundant structures, critical motifs such as α -1,3-Gal-epitopes and *N*-glycolylneuraminic acids were detected.

Figure 2 shows the fragment spectrum of glycan 3 (cf., Table 2). Fragment spectra were recorded and analyzed for half of them

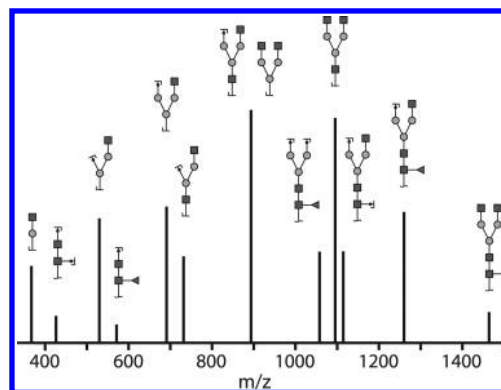


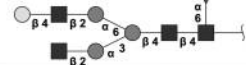


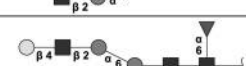
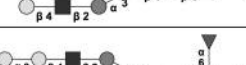
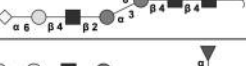
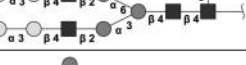



Figure 2. CID MS/MS spectrum of *N*-glycan 3 with the composition Hex3HexNac4dHex1. For clarity, only B- and Y-fragment ions are shown in the annotation. Several ions may also be generated by C-/Z-fragmentation, but B- and Y-ions occur most frequently in fragmentation of *N*-glycans in positive mode mass spectra. Fragment spectra give information about the monosaccharide sequence. However, details about linkages are hard or even impossible to obtain by MS. For explanation of the symbols, see footnote to Table 1.

(Table 1). Fragmentation of glycans generates glycosidic and cross-ring fragments. Glycosidic fragments allow the determination of monosaccharide sequences, while cross-ring fragments hint linkage positions.

***N*-Glycans from Cetuximab Analyzed by Integration of LC-MS and ¹H NMR Spectroscopy.** MS alone is not sufficient to elucidate the full structure of glycans; hexoses, e.g., mannose, galactose, and glucose, cannot be differentiated by MS because of their identical molecular formula. However, these hexoses exhibit altered biological activities and therefore have to be distinguished clearly. Because mass isobaric structures that differ only in linkage positions and branching are difficult or even impossible to differentiate by MS/MS, integration of MS and ¹H NMR was used for unambiguous structure characterization. The 3DCC procedure utilizes an LC flow fractionation into well plates (15 s/fraction) in parallel to the MS acquisition. Each fraction measured by 1D-¹H NMR contains multiple *N*-glycans and, as a result, signals of various analytes. 3DCC was applied to mathematically separate NMR spectra using information from the LC-MS run, i.e., the EICs containing retention time and accurate m/z values.²² Ten *N*-glycans were unambiguously identified by the 3DCC procedure (Figure 3, Table 2). *N*-Glycans with an abundance as low as 0.1% (relative quantification by integration of EIC signals for all detected 37 structures) were detected. Compounds 1 and 2 as well as 8 and 10 are pairwise mass isobaric structures that could only be characterized by 3DCC methods. These structures cannot be distinguished by LC-MS solely which demonstrates the analytical power of combining LC-MS and NMR in the 3DCC procedure.

EDCs (extracted delta chromatograms) are defined as the intensity at each data point in the NMR spectra as a function of the fraction number, i.e., retention time; that is, a plot of the intensity at each chemical shift as a function of retention time. 3DCC mathematically correlates LC-MS information with

Table 2. Results Obtained by Applying the 3DCC_L, 3DCC_C, and 3DCC_S Methods To Elucidate the Structure of the Prominent Glycans^a

Glycan structure ^b	Identified by			Quantification		Compound Number
	MS [Da]	Accuracy [ppm]	NMR (L,C,S) ^c	MS [%] ^d	NMR [%/nmol] ^e	
	1624.604	3.9	L,C,S	20.4	13.9/5.2	1
	1624.604	3.9	L,C,S	11.4	14.6/5.5	2
	1462.552	4.9	L,C,S	19.6	22.9/8.6	3
	1786.656	3.1	L,S	7.5	3.0/1.1	4
	2255.800	2.8	L,C,S	7.7	5.8/2.2	5
	2110.762	2.8	L,C,S	13.9	15.9/5.9	6
	1234.436	1.9	L,C,S	2.9	7.9/3.0	7
	2093.746	2.5	L,S	1.4	4.8/1.8	8
	2400.832	0.4	L,S	0.6	10.0/3.7	9
	2093.746	2.5	L,S	0.1	1.2/0.4	10

^aAbout 67% of the structures are covered by the 3DCC technique. The EICs obtained from the separation of the LC-MS run were used for the generation of the ¹H NMR spectra (cf., Figures 1 and 3). Compounds 1 and 2 as well as 8 and 10 are pairwise mass isobaric structures that could only be characterized by 3DCC methods. ^bFor explanation of the symbols, see footnote to Table 1. ^cL: identification by applying 3DCC_L (least-squares optimization), C: identification by applying 3DCC_C (Pearson coefficients), S: identification by applying 3DCC_S (singular value decomposition) cf., text. ^dMS quantification by integration of EIC signals for all 37 detected N-glycans by MS. ^eRelative NMR quantification by integration of H-1 (b-D-GlcNAc) signal. Absolute quantification by calibration relative to an external sucrose standard.

NMR data to extract NMR spectra of pure compounds from mixture spectra. All NMR signals belonging to one analyte rise and fall at the same time the analyte generates a signal in the MS spectrum (this is referred to as EIC) (cf., Figure 3). At a given chemical shift in the NMR often more than one analyte exhibit signals. This results in an EDC that is composed of several signals rising and decaying differently with the retention time. Thus, if an EDC contains signals from several analytes, the EICs of all structures that cause this experimental EDC are summed up in a weighted proportion (according to their abundance) to generate a calculated EDC (Figure 4). The factors that have to be multiplied with each EIC for the calculation of the theoretical EDC are proportional to the abundance of the corresponding analyte in the mixture.

The correlation of increasing and decreasing NMR signals of one analyte with the rising and falling of the corresponding *m/z* values detected by MS, led to ten extracted NMR spectra of pure compounds (Table 2, Figure 3). By means of 3DCC, it is easily possible to identify the presence of the immunogenic α -1,3-gal epitope utilizing highly specific signals at chemical

shifts of 5.145 (H1), 4.019 (H4), or 4.192 (H5) ppm and of the immunogenic N-glycolylneuraminic acid moieties at chemical shifts of 4.118 (NGc), 2.684 (H3e), and 1.739 (H3a) ppm, respectively. Figure 5 shows four 1D-¹H NMR spectra of these critical glycosylation patterns. This represents the first unambiguous identification by MS/MS and NMR of these structures. At the top of Figure 5 the sum of all NMR spectra is shown. The spectrum expectedly contains all characteristic chemical shifts. Only by means of the mathematical separation into spectra of pure analytes, the unambiguous characterization of each structure was possible. Accordingly, spectra of analytes that do not carry a specific motif, e.g. an N-glycolylneuraminic acid, do not show signals at the chemical shifts of 2.684 ppm (cf., traces for compounds 1, 3, and 6) characteristic for these residues.

Description and Comparison of 3DCC_L, 3DCC_C, and 3DCC_S. Originally, 3DCC was realized in two mathematical ways: 3DCC_L that uses least-squares optimization and 3DCC_C that applies correlation coefficients (Pearson).²² Here 3DCC was extended with a new variant to solve the set of equations by using singular value decomposition (SVD). The new variant is

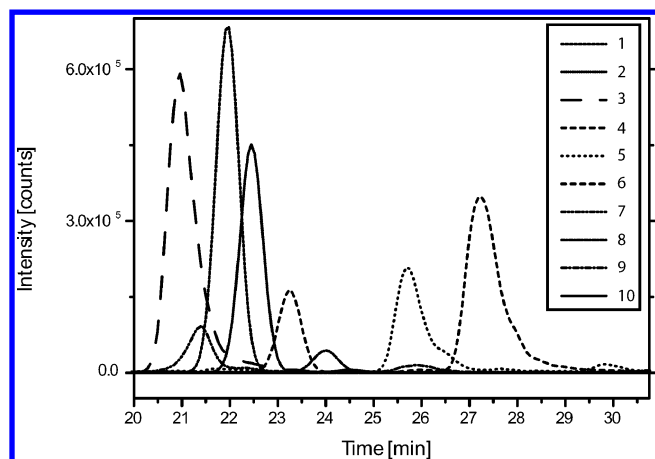


Figure 3. LC-MS chromatogram of underivatized *N*-glycans enzymatically cleaved from cetuximab separated on a PGC column. Ammonia was used as coeluent to avoid separation of α - and β -anomers. Detail of ten extracted ion chromatograms of *N*-glycans that yielded unambiguous structure characterization by ^1H NMR analysis. Within the time depicted, 32 further *N*-glycans coelute that were identified by MS/(MS) spectra.

called 3DCC_S. SVD is considered to be the best method for dealing with sets of equations or matrices that are singular or very close to singular.³⁰

Implementation of experimental data is identical for all variants. NMR data are imported to Matlab by using the “rnbmr” script of Nils Nyberg. As a result, two objects are generated: a vector holding the chemical shift for each data point and a two-dimensional matrix (*matrix_NMR*) containing the signal intensities of all acquired NMR spectra as a function of chemical shifts and of fraction number. EICs were generated (Data-Analysis, Bruker), and the number of data points was reduced to the number of LC fractions, which is equal to the number of acquired NMR spectra. In case an EIC for a given mass has more than one peak, the EIC is separated into as many EICs as peaks occur. This is achieved by a line shape fitting procedure. A two-dimensional matrix containing the intensities for all EIC data points as a function of LC retention time and the number of analytes that can be extracted from the EICs is imported to Matlab (*matrix_MS*). EDCs are vectors (i.e., intensities at a given NMR data point as a function of the retention time, which is proportional to the fraction number) and are contained as columns in *matrix_NMR*. The following subsections describe briefly the two recently introduced solutions to the cross-correlation problem of correlating the EICs with the EDCs by least-squares minimization (3DCC_L) and by correlation coefficients (3DCC_C) followed by a detailed section on the new variant 3DCC_S utilizing SVD.

3DCC_L.²² Theoretical EDCs (the number of EDCs is equal to the number of NMR data points) are calculated to yield the fit to each experimental EDC using a linear combination of all EICs. EICs are summed up with individual weights to generate the theoretical EDC. This variant uses least-squares optimization (*lsqlin* command of Matlab) to fit the EDCs. Each EIC receives a coefficient that is required for the calculation of each theoretical EDC. The coefficients for one particular EIC are then plotted for all NMR data points, resulting in the 3DCC_L NMR spectra for pure components for each EIC, i.e., for each analyte.

3DCC_C.²² The second previously described variant uses cross correlation (Pearson coefficients) of every EIC with every

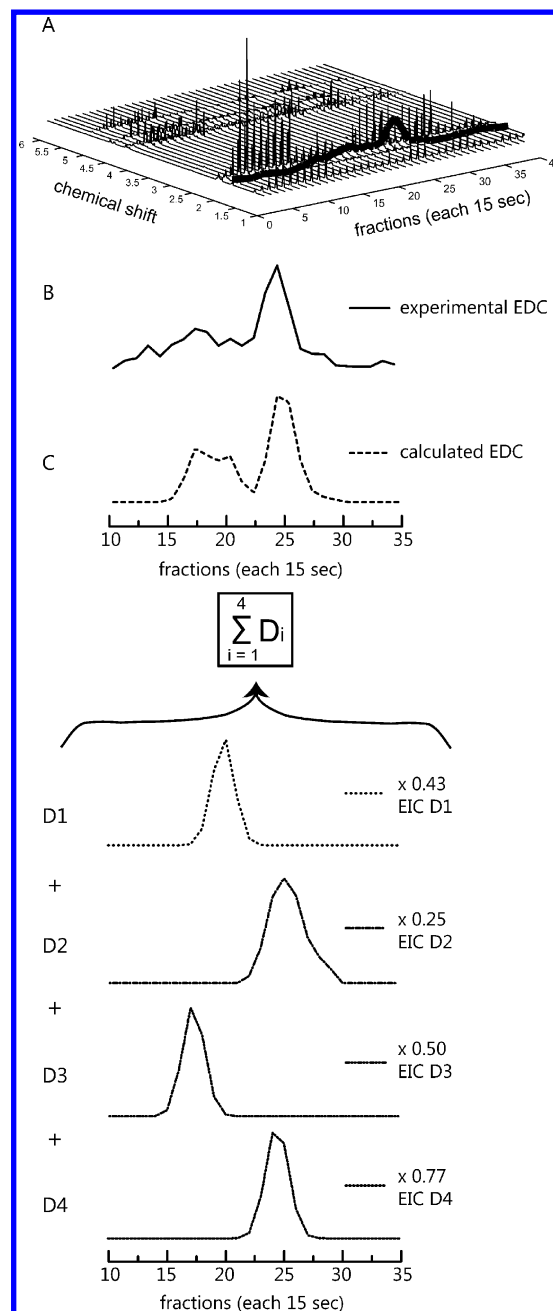


Figure 4. Exemplified 3DCC results by applying singular value decomposition using LC-MS and ^1H NMR information. Forty LC-fractions were analyzed by ^1H -NMR (A). An EDC represents a distinct chemical shift over all fractions (A+B). The 3DCC method calculates a theoretical EDC (C) by multiplying weighted EICs (D1–D4) that fits an arbitrary experimental EDC (B). The factor for the weighting of an EIC is proportional to the portion of the glycan that generates the considered EDC signal.

EDC. Therefore, the correlation coefficients are built pairwise (*corr* command in Matlab). The 3DCC NMR spectra are then calculated by pairwise multiplying these correlation coefficients with the experimental NMR sum spectrum at a given chemical shift and *m/z* value. Negative intensity values were set to zero. This procedure leads to 3DCC NMR spectra that contain unique signals. In combination with MS spectra the characterization is achieved.

3DCC_S. This new variant uses *matrix_MS* to generate individual EDCs with singular value decomposition (SVD) using

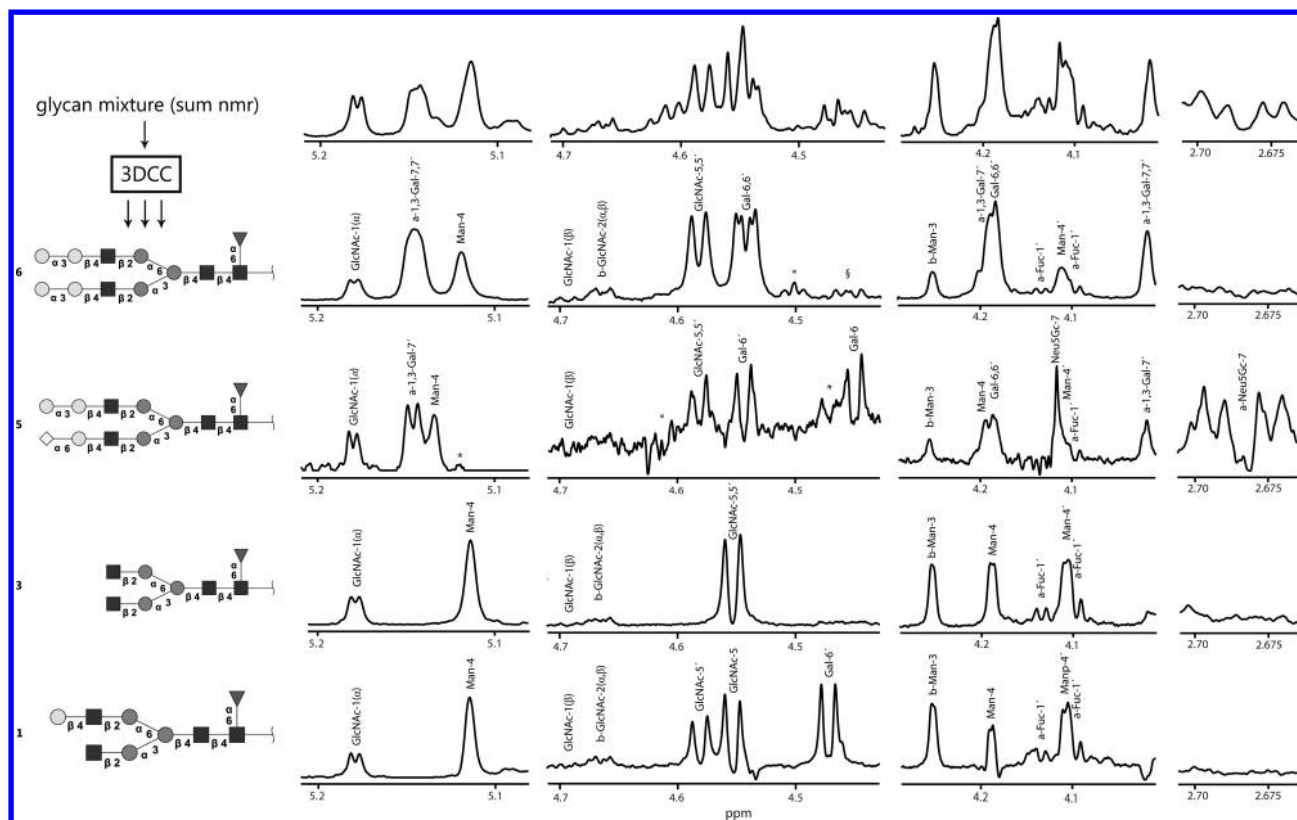


Figure 5. Extracted NMR spectra by applying 3DCC_S. The top spectrum shows the sum of the NMR spectra of all fractions of the glycan mixture. 3DCC mathematically correlates LC-MS information with rising and falling NMR signals to extract NMR spectra of pure compounds. In total, ten structures were unambiguously identified; four of these are shown. Critical moieties were identified by characteristic chemical shifts, e.g. N-glycolylneuraminic acids at 2.684 (H3e) and 4.119 (NGc) or α-1,3-Gal epitopes at 5.147 (H1). For explanation of the symbols see legend to Table 1. *Minor signals of simultaneously eluting structure in the LC run with the composition Hex6HexNac4dHex1 (very low abundant glycan structure). §Minor signal of a glycan (Hex5HexNac4dHex1NeuGc2) that is overlapping in the elution profile. †Impurity probably resulting from a short peptide that is simultaneously eluting.

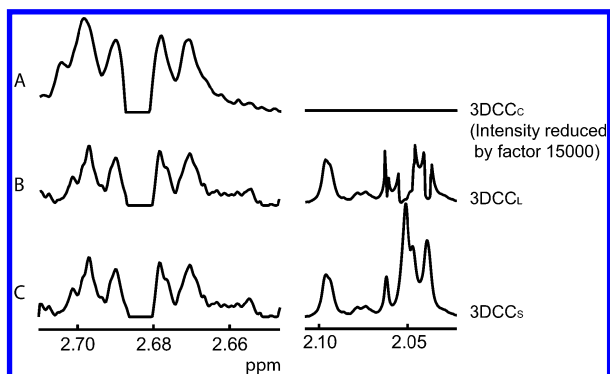


Figure 6. Comparison of spectra obtained by the 3DCC_S, 3DCC_L, and 3DCC_C methods, respectively. 3DCC_L and 3DCC_S generate complete spectra containing all signals of the compounds. In comparison, 3DCC_C results in unique signals (see NAc signals at about 2.0 ppm, 6A: right-hand side) but shows significantly higher signal-to-noise (S/N) ratios (see N-glycolylneuraminic acid signals at about 2.68 ppm, 6A left-hand side). The depicted detail of the 3DCC_S spectrum (6C, right-hand side) shows more continuous curves than 3DCC_L (cf., 6B, right-hand side).

the Matlab routine (*svd* command). SVD is known to be the method of choice to solve most linear squares problems and can be seen to enhance effectiveness and accuracy of 3DCC_L.³⁰ The system of equations

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b} \quad (1)$$

has to be solved for every EDC, so that \mathbf{A} is represented by *matrix_MS*, \mathbf{b} is represented by one experimental EDC (individual columns in *matrix_NMR*), and \mathbf{x} denotes a vector of the coefficients that is necessary to fit the weighted sum of EICs to the experimental EDC (compare 3DCC_L) and has the same size as the number of fractions. Therefore, the SVD routine decomposes matrix \mathbf{A} to prepare the inverse matrix \mathbf{A}^{-1} , because

$$\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{b} \quad (2)$$

Summarized, any $j \times k$ matrix \mathbf{A} can be expressed as the product of three unique matrices: a $j \times j$ matrix \mathbf{U}_A , a $k \times k$ matrix $(\mathbf{V}_A)^+$ (+: conjugate transpose), and a $j \times k$ diagonal matrix \mathbf{D}_A :

$$\mathbf{A} = \mathbf{U}_A \cdot \mathbf{D}_A \cdot (\mathbf{V}_A)^+ \quad (3)$$

Because \mathbf{U}_A and \mathbf{V}_A are unitary matrices $(\mathbf{U}_A)^+ = (\mathbf{U}_A)^{-1}$ and $(\mathbf{V}_A)^+ = (\mathbf{V}_A)^{-1}$, it is

$$\mathbf{A}^{-1} = ((\mathbf{V}_A)^+)^{-1} \cdot (\mathbf{D}_A)^{-1} \cdot (\mathbf{U}_A)^{-1} \quad (4)$$

$$= \mathbf{V}_A \cdot (\mathbf{D}_A)^{-1} \cdot (\mathbf{U}_A)^+ \quad (5)$$

With eq 2, coefficient vector \mathbf{x} can be calculated for every EDC, e.g., \mathbf{b} :

$$\mathbf{x} = \mathbf{V}_A \cdot (\mathbf{D}_A)^{-1} \cdot (\mathbf{U}_A)^+ \cdot \mathbf{b} \quad (6)$$

3DCC_S spectra are then calculated in the same way as for 3DCC_L spectra.

3DCC_L and 3DCC_S generate complete extracted spectra of the compounds. In comparison, 3DCC_C results in unique signals (Figure 6) but shows significantly higher signal-to-noise (S/N) ratios. Therefore, 3DCC_C should be used in case of low S/N ratios. However, quantification is possible only with 3DCC_L and 3DCC_S. 3DCC_S spectra show a more continuous progress of curves than 3DCC_L spectra (cf., Figure 6). In addition, effectivity in terms of computing time is enhanced drastically. In the case of 3DCC_S, less than 1 s is required for the whole calculation on a modern workstation PC compared to at least 20 min for the 3DCC_L variant.

CONCLUSION

The data presented here demonstrate clearly the benefit of combining NMR and LC-MS for an unambiguous identification and characterization of glycan structures. Using this technology on a routine basis could avoid problems that arose in the case of cetuximab. 3DCC yields a reliable determination of critical epitopes and also provides an improved quantification. Using the combination of NMR and LC-MS it is easy to analyze and quantify the immunogenic α -1,3-Gal and the N-glycolylneuraminic acid epitopes. Mass isobaric structures can be distinguished and unambiguously characterized. NMR is certainly less sensitive than LC-MS, but it allows identification of compounds down to about 15 pmol, which is approximately equal to 30 ng of oligosaccharide.¹³ Given the quantities available from therapeutic antibodies, it is easy to perform this type of analysis on a routine basis for quality assurance of the products.

AUTHOR INFORMATION

Corresponding Author

*Tel: +49 40 42838 5913. E-mail: bernd.meyer@chemie.uni-hamburg.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the Deutsche Forschungsgemeinschaft (DFG) for a grant for the 700 MHz NMR spectrometer (BM-1831/1-1). This work was supported by the Helmholtz Association by a stipend for A.W.

REFERENCES

- (1) Reichert, J. *mAbs* **2012**, *4*, 413–415.
- (2) Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N. *Nat. Rev. Immunol.* **2010**, *10*, 345–352.
- (3) Davis, J.; Deng, R.; Boswell, C.; Zhang, Y.; Li, J.; Fielder, P.; Joshi, A.; Kenkare-Mitra, S. *Pharmaceutical Biotechnology*; Springer: New York, 2013; p 143.
- (4) Reichert, J. *mAbs* **2010**, *2*, 84–100.
- (5) Ledford, H. *Nature* **2010**, *468*, 18–19.
- (6) Ayoub, D.; Jabs, W.; Resemann, A.; Evers, W.; Evans, C.; Main, L.; Baessmann, C.; Wagner-Rousset, E.; Suckau, D.; Beck, A. *mAbs* **2013**, *5*, 699–710.
- (7) Lis, H.; Sharon, N. *Eur. J. Biochem./FEBS* **1993**, *218*, 1–27.
- (8) Berger, E.; Buddecke, E.; Kamerling, J.; Kobata, A.; Paulson, J.; Vliegthart, J. *Experientia* **1982**, *38*, 1129–1258.
- (9) Rademacher, T.; Parekh, R.; Dwek, R. *Annu. Rev. Biochem.* **1988**, *57*, 785–838.
- (10) Chan, A.; Carter, P. *Nat. Rev. Immunol.* **2010**, *10*, 301–316.
- (11) Chung, C.; Mirakhur, B.; Chan, E.; Le, Q.-T.; Berlin, J.; Morse, M.; Murphy, B.; Satinover, S.; Hosen, J.; Mauro, D.; Slebos, R.; Zhou, Q.; Gold, D.; Hatley, T.; Hicklin, D.; Platts-Mills, T. *New Engl. J. Med.* **2008**, *358*, 1109–1117.
- (12) Dell, A.; Morris, H. *Science* **2001**, *291*, 2351–2356.
- (13) Fellenberg, M.; Coksezen, A.; Meyer, B. *Angew. Chem., Int. Ed.* **2010**, *49*, 2630–2633.
- (14) Humblet, Y. *Expert Opin. Pharmacother.* **2004**, *5*, 1621–1633.
- (15) Cunningham, D.; Humblet, Y.; Siena, S.; Khayat, D.; Bleiberg, H.; Santoro, A.; Bets, D.; Mueser, M.; Harstrick, A.; Verslype, C.; Chau, I.; Van Cutsem, E. *New Engl. J. Med.* **2004**, *351*, 337–345.
- (16) Qian, J.; Liu, T.; Yang, L.; Daus, A.; Crowley, R.; Zhou, Q. *Anal. Biochem.* **2007**, *364*, 8–18.
- (17) Wührer, M.; Catalina, M.; Deelder, A.; Hokke, C. *J. Chromatogr., B* **2007**, *849*, 115–128.
- (18) Fridriksson, E.; Beavil, A.; Holowka, D.; Gould, H.; Baird, B.; McLafferty, F. *Biochemistry* **2000**, *39*, 3369–3376.
- (19) Han, X.; Jin, M.; Breuker, K.; McLafferty, F. *Science* **2006**, *314*, 109–112.
- (20) Rahul, R.; Raguram, S.; Ganesh, V.; James, C. P.; Ram, S. *Nat. Methods* **2005**, *2*, 817–824.
- (21) Fellenberg, M.; Behnken, H.; Nagel, T.; Wiegandt, A.; Baerenfaenger, M.; Meyer, B. *Anal. Bioanal. Chem.* **2013**, *405*, 7291–7305.
- (22) Behnken, H.; Fellenberg, M.; Koetzler, M.; Jirmann, R.; Nagel, T.; Meyer, B. *Anal. Bioanal. Chem.* **2012**, *404*, 1427–1437.
- (23) Ruhaak, L.; Deelder, A.; Wührer, M. *Anal. Bioanal. Chem.* **2009**, *394*, 163–174.
- (24) König, S.; Leary, J. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1125–1134.
- (25) Viseux, N.; de Hoffmann, E.; Domon, B. *Anal. Chem.* **1998**, *70*, 4951–4959.
- (26) Weiskopf, A.; Vouros, P.; Harvey, D. *Anal. Chem.* **1998**, *70*, 4441–4447.
- (27) Domon, B.; Costello, C. E. *Glycoconjugate J.* **1988**, *5*, 397–409.
- (28) Harvey, D. *Proteomics* **2005**, *5*, 1774–1786.
- (29) Vliegthart, J. F. G.; Dorland, L.; van Halbeek, H. *Adv. Carbohydr. Chem. Biochem.* **1983**, *41*, 209–374.
- (30) Press, W.; Flannery, B.; Teukolsky, S.; Vetterling, W. *Numerical recipes: The art of scientific computing*; Cambridge University Press: New York, 1986; p 818.