# ENTWICKLUNG NEUER STRATEGIEN ZUR ANALYTIK VON GLYCOPROTEINEN

## Dissertation

zur Erlangung des Doktorgrades Fachbereich Chemie Fakultät für Mathematik, Informatik und Naturwissenschaften Universität Hamburg

vorgelegt von

Meike Fellenberg aus Hamburg

Hamburg 2011

We are drowning in information and starving for knowledge.

(R. D. Rogers)

Die vorliegende Arbeit wurde in der Zeit von September 2008 bis November 2011 am Institut für Organische Chemie, Fachbereich Chemie, Universität Hamburg, Geschäftsführender Direktor Prof. Dr. Bernd Meyer, angefertigt.

Herrn Prof. Dr. Bernd Meyer danke ich für die interessante Themenstellung und für die stets motivierende und freundliche Unterstützung bei der Durchführung dieser Arbeit.

Herrn Prof. Dr. Dr. h. c. mult. Wittko Francke danke ich für die Übernahme der Zweitbegutachtung dieser Arbeit.

1. Gutachter: Prof. Dr. Bernd Meyer

2. Gutachter: Prof. Dr. Dr. h. c. mult. Wittko Francke

Tag der Disputation: 15. Dezember 2011

## **Gliederung**

A	BKÜR	ZUNG	SVERZEICHNIS	III		
1	E	INLEI	ՐՍNG	1		
	1.1	N-C	Ilycane – Struktur und Funktion	1		
	1.2	Ana	llytik von Oligosacchariden	2		
	1.2	2.1	Chromatographische Methoden zur Trennung von Glycanen	3		
	1.2	2.2	Massenspektrometrie zur Analytik von Glycanen	4		
	1.2	2.3	NMR-Spektroskopie – Das Konzept der structural reporter groups	4		
	1.2	2.4	Datenbanken - CCSD, sugabase, glycosciences.de und GlycomeDB	5		
	1.3	Kor	nzepte zur Kombination von analytischen Methoden	7		
2	Μ	[ETHC	DEN	9		
	2.1	Mat	hematische Beschreibungen der Korrelation zweier Variablen	9		
	2.2	Opt	imierte Wasserunterdrückung bei der NMR-Spektrenakquisition	10		
	2.3	Stru	kturaufklärung mit selektiven TOCSY-Experimenten	13		
	2.4	Ter	minologie in der Massenspektrometrie	13		
3	A	UFGA	BENSTELLUNG	14		
4	E	RGEBI	NISSE UND DISKUSSION	15		
	4.1	Ent	wicklung einer Methode zur automatischen Spektrenidentifikation	15		
	4.1	1.1	Berechnung von Spektren anhand von Datenbankeinträgen	15		
	4.1	1.2	Vergleich von berechneten und experimentellen Spektren	19		
	4.1	1.3	utomatische Identifikation von Spektren			
	4.1	1.4	Fehlerhafte Einträge in sugabase und daraus resultierende Probleme	23		
	4.2	Ide	ntifikation von Komponenten in komplexen Gemischen	25		
	4.3 3D-Kreuzkorrelation (3DCC) von LC, MS und NMR – Entwicklung einer neuen Methode zur Charakterisierung von Glycanen		Kreuzkorrelation (3DCC) von LC, MS und NMR – Entwicklung einer en Methode zur Charakterisierung von Glycanen	30		
	4.3	3.1	Schematische Darstellung der 3D-Kreuzkorrelation	31		
	4.3	3.2	Vergleich der Korrelation nach Kendall, Spearman und Pearson	35		
	4.3	3.3	3DCC am Beispiel eines simulierten Chromatographielaufes	39		
	4.3.3.1		1 Dekonvolution von NMR-Spektren	39		
		4.3.3.	2 Automatische Zuordnung der dekonvulierten Spektren	45		
	4.4	Ana	lytik der Glycane des bovinen Fibrinogens	47		
	4.4	4.1	3DCC zur Charakterisierung der neutralen Glycane vom bovinen Fibrinogen	47		
		4.4.1.	1 Methoden zur Optimierung der Spektrenextraktion	55		
		4.4.1.	2 S/N-Verbesserung durch Spektrendekonvolution	65		

	4.4.2 3DCC zur Charakterisierung von massenisobaren Strukturen						
4.4.3 Automatische Spektrenerkennung der extrahierten NMR-Daten							
4.4.4 Untersuchungen zum Abspaltungsprozess vom bovinen Fibrinogen							
	4.4.5	Identifikation bisher unbekannter Glycanstrukturen im bovinen Fibrinogen	73				
	4.4.5	1 Aufklärung des Dodecasaccharids <b>4</b>	74				
	4.4.5	2 Aufklärung der Undecaisomere <b>5</b> und <b>6</b>	76				
	4.4.5	Aufklärung der Nonaisomere 1,7,8 und 9	77				
	4.4.5	.4 Aufklärung der triantennären Struktur <b>10</b>	79				
5	ZUSAM	MENFASSUNG	82				
6	SUMMA	ARY	84				
7	EXPER	IMENTELLER TEIL	86				
	7.1 Ver	rwendete Geräte, Software und Chemikalien	86				
	7.2 Pro	benpräparation	87				
	7.2.1	Präparation des simulierten Chromatographielaufes	87				
	7.2.2	Präparation der Gemische G1 und G3	87				
7.2.3 Gemische <b>G2</b> und <b>G4</b>							
	7.3 Au	fnahme und Auswertung der LC-MS-Daten	88				
	7.4 Au	fnahme und Prozessierung der NMR-Daten	91				
	7.5 Dat	tenbearbeitung mit <i>matlab</i>	93				
	7.5.1	Berechnung von Spektren anhand von Datenbankeinträgen	93				
	7.5.2	Automatische Identifikation von Spektren	96				
	7.5.3	Berechnung des 3DCC-Beispiels	97				
	7.5.4	Berechnung der Beispiele zum Vergleich verschiedener Korrelationsmethoden	97				
	7.5.5	Import und Bearbeitung von MS- und NMR-Daten	98				
	7.5.6	Korrelation, Extraktion und Optimierung	99				
	7.5.7	Export von Grafiken	100				
	7.6 Cha	arakterisierung der untersuchten Glycanstrukturen	101				
	7.7 To:	xikologie	128				
	7.8 KN	IR-Substanzen	128				
8	LITERA	ATUR	129				
D	ANKSAGUN	īG	134				
С	URRICULU	M VITAE	136				
E	RKLÄRUNG		138				

## Abkürzungsverzeichnis

1D	eindimensional
3DCC	3D-Kreuzkorrelation (3D cross correlation)
Asn	Asparagin
BDSDB	Bacterial Carbohydrate Structure Database
BPC	base peak chromatogram
CASPER	Computer Assisted SPectrum Evaluation of Regular polysaccharides
CCSD	Complex Carbohydrate Structure Database
CFG	Consortium for Functional Glycomics
COSY	correlation spectroscopy
dHex	Deoxyhexose
DNA	Desoxyribonukleinsäure (deoxyribnucleic acid)
EDC	extracted delta chromatogram
EIC	extracted ion chromatogram
EM	Multiplikation des FIDs mit einer Exponentialfunktion
ER	Endoplasmatisches Retikulum
es	excitation sculpting
ESI	electrospray ionization
Exp	experimentell erhaltenes Spektrum
Expno	Experimentnummer (experimental number)
gb	gaussian broadening factor
GM	Multiplikation des FIDs mit Gauß'scher Fensterfunktion
Hex	Hexose
HexNAc	<i>N</i> -Acetylhexosamin
HMBC	heteronuclear multiple-bond correlation
HPLC	high performance liquid chromatography
HSQC	heteronuclear single-quantum correlation
HWHM	half width half maximum
KEGG	Kyoto Encyclopedia of Genes and Genomes
lb	line broadening factor
LINUCS	LInear Notation for Unique description of Carbohydrate Sequences
MALDI	matrix-assisted laser desorption/ionization
MS	Massenspektrometrie
MW	Molgewicht

n.d.	not determined
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
oLH	ovine luteinizing hormone
PDB	Protein Data Bank
Procno	Prozessierungsnummer (processing number)
S/N	Signal-zu-Rausch-Verhältnis (signal-to-noise ratio)
sEIC	sub-extracted ion chromatogram
Ser	Serin
SHY	statistical heterospectroscopy
STOCSY	statistical total correlation spectroscopy
TDC	total delta chromatogram
Thr	Threonin
TIC	total ion current
TOCSY	total correlation spectroscopy
TOF	time-of-flight
$V_{obs}$	beobachtbares Volumen (observable volume)
$\alpha_L$	Parameter für die Berechnung der Lorentz-Verteilung (HWHM)
$\mathbf{M}_{(m,t)}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs)
$\mathbf{M}_{(m,t)}$ $\mathbf{N}_{(t,\delta)}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren
$\mathbf{M}_{(m,t)}$ $\mathbf{N}_{(t,\delta)}$ $ N\rangle$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $\mathbf{N}_{(t,\delta)}$ , entsprechen einzelnen EDCs
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen Spektren
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $\mathbf{N}_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $\mathbf{N}_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $\mathbf{M}_{(m,t)}$ , entsprechen einzelnen EICs
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $\mathbf{N}_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $\mathbf{N}_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $\mathbf{M}_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$ $C_{m,\delta}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $\mathbf{N}_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $\mathbf{N}_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $\mathbf{M}_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix Element der Matrix $\mathbf{C}_{(m,\delta)}$ , Korrelationskoeffizient
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$ $c_{m,\delta}$ $(c_{m,\delta})^{4}/(c_{m,\delta})_{s}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $M_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix Element der Matrix $C_{(m,\delta)}$ , Korrelationskoeffizient mit 4. Potenz bzw. sigmoider Funktion gewichteter Korrelationskoeffizient
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$ $c_{m,\delta}$ $(c_{m,\delta})^4/(c_{m,\delta})_s$ $E_{(m,\delta)}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $M_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix Element der Matrix $C_{(m,\delta)}$ , Korrelationskoeffizient mit 4. Potenz bzw. sigmoider Funktion gewichteter Korrelationskoeffizient Matrix, beinhaltet extrahierte NMR-Spektren assoziiert zu einzelnen EICs
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$ $c_{m,\delta}$ $(c_{m,\delta})^4/(c_{m,\delta})_s$ $E_{(m,\delta)}$ $e_{m,\delta}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $M_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix Element der Matrix $C_{(m,\delta)}$ , Korrelationskoeffizient mit 4. Potenz bzw. sigmoider Funktion gewichteter Korrelationskoeffizient Matrix, beinhaltet extrahierte NMR-Spektren assoziiert zu einzelnen EICs Element der Matrix $E_{(m,\delta)}$
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$ $c_{m,\delta}$ $(c_{m,\delta})^4/(c_{m,\delta})_s$ $E_{(m,\delta)}$ $e_{m,\delta}$ $E4_{(m,\delta)}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $M_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix Element der Matrix $C_{(m,\delta)}$ , Korrelationskoeffizient mit 4. Potenz bzw. sigmoider Funktion gewichteter Korrelationskoeffizient Matrix, beinhaltet extrahierte NMR-Spektren assoziiert zu einzelnen EICs Element der Matrix $E_{(m,\delta)}$
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$ $c_{m,\delta}$ $(c_{m,\delta})^4/(c_{m,\delta})_s$ $E_{(m,\delta)}$ $E_{(m,\delta)}$ $E4_{(m,\delta)}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $M_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix Element der Matrix $C_{(m,\delta)}$ , Korrelationskoeffizient mit 4. Potenz bzw. sigmoider Funktion gewichteter Korrelationskoeffizient Matrix, beinhaltet extrahierte NMR-Spektren assoziiert zu einzelnen EICs Element der Matrix $E_{(m,\delta)}$ Matrix, beinhaltet gewichtet-extrahierte NMR-Spektren
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$ $c_{m,\delta}$ $(c_{m,\delta})^4/(c_{m,\delta})_s$ $E_{(m,\delta)}$ $E_{(m,\delta)}$ $E_{(m,\delta)}$ $E_{(m,\delta)}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $\mathbf{N}_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $\mathbf{N}_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $\mathbf{M}_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix Element der Matrix $\mathbf{C}_{(m,\delta)}$ , Korrelationskoeffizient mit 4. Potenz bzw. sigmoider Funktion gewichteter Korrelationskoeffizient Matrix, beinhaltet extrahierte NMR-Spektren assoziiert zu einzelnen EICs Element der Matrix $\mathbf{E}_{(m,\delta)}$ Matrix, beinhaltet gewichtet-extrahierte NMR-Spektren Matrix, beinhaltet durch Korrelation nach Kendall extrahierte NMR-Spektren
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$ $c_{m,\delta}$ $(c_{m,\delta})^4/(c_{m,\delta})_s$ $E_{(m,\delta)}$ $E4_{(m,\delta)}$ $EK_{(m,\delta)}$ $ES_{(m,\delta)}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $M_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix Element der Matrix $C_{(m,\delta)}$ , Korrelationskoeffizient mit 4. Potenz bzw. sigmoider Funktion gewichteter Korrelationskoeffizient Matrix, beinhaltet extrahierte NMR-Spektren assoziiert zu einzelnen EICs Element der Matrix $E_{(m,\delta)}$ Matrix, beinhaltet gewichtet-extrahierte NMR-Spektren Matrix, beinhaltet durch Korrelation nach Kendall extrahierte NMR-Spektren Matrix, beinhaltet durch Korrelation nach Spearman extrahierte NMR-Spektren Vektor, entspricht dem Summenspektrum über alle NMR-Daten
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$ $c_{m,\delta}$ $(c_{m,\delta})^4/(c_{m,\delta})_s$ $E_{(m,\delta)}$ $E4_{(m,\delta)}$ $EK_{(m,\delta)}$ $ES_{(m,\delta)}$ $S_{(\delta)}$	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $N_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix Element der Matrix $C_{(m,\delta)}$ , Korrelationskoeffizient mit 4. Potenz bzw. sigmoider Funktion gewichteter Korrelationskoeffizient Matrix, beinhaltet extrahierte NMR-Spektren assoziiert zu einzelnen EICs Element der Matrix $E_{(m,\delta)}$ Matrix, beinhaltet gewichtet-extrahierte NMR-Spektren Matrix, beinhaltet durch Korrelation nach Kendall extrahierte NMR-Spektren Matrix, beinhaltet durch Korrelation nach Spearman extrahierte NMR-Spektren Vektor, entspricht dem Summenspektrum über alle NMR-Daten Element des Vekors $S_{(\delta)}$
$M_{(m,t)}$ $N_{(t,\delta)}$ $ N\rangle$ $\langle N $ $\langle M $ $C_{(m,\delta)}$ $c_{m,\delta}$ $(c_{m,\delta})^4/(c_{m,\delta})_s$ $E_{(m,\delta)}$ $E4_{(m,\delta)}$ $ES_{(m,\delta)}$ $S_{(\delta)}$ $S_{\delta}$ K	Matrix, beinhaltet MS-Daten entlang des zeitlichen Verlaufs (EICs) Matrix, beinhaltet zeitlich aufeinanderfolgend aufgenommene NMR-Spektren Spaltenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen EDCs Zeilenvektoren der Matrix $N_{(t,\delta)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $M_{(m,t)}$ , entsprechen einzelnen Spektren Zeilenvektoren der Matrix $M_{(m,t)}$ , entsprechen einzelnen EICs Korrelationsmatrix Element der Matrix $C_{(m,\delta)}$ , Korrelationskoeffizient mit 4. Potenz bzw. sigmoider Funktion gewichteter Korrelationskoeffizient Matrix, beinhaltet extrahierte NMR-Spektren assoziiert zu einzelnen EICs Element der Matrix $E_{(m,\delta)}$ Matrix, beinhaltet gewichtet-extrahierte NMR-Spektren Matrix, beinhaltet durch Korrelation nach Kendall extrahierte NMR-Spektren Matrix, beinhaltet durch Korrelation nach Spearman extrahierte NMR-Spektren Vektor, entspricht dem Summenspektrum über alle NMR-Daten Element des Vekors $S_{(\delta)}$

### Abbildung 1: Symbolerklärung Monosaccharid-Bausteine



Oligosaccharid		Bezeichnung in dieser Arbeit
1	$ \begin{array}{c} & & & \\ & & & & \\ & & & \\ & & & & \\ & & & \\ & & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ $	Nona
2	$\begin{array}{c} & & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & \\ & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ &$	Deca
3	$\begin{array}{c} & & & & & \\ & & & & \\ & & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & \\ & & & \\ &$	fucosyliertes Nona
4	$\begin{array}{c} & & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\$	Dodeca
5	$\begin{array}{c} & & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\$	Undeca
6	$\begin{array}{c} & & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & \\ & & & & \\ & & & \\ & & & & \\ & & & \\ & & & \\ & & & \\ & & & & \\ & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & &$	Undecaisomer
7	$ \begin{array}{c} & & & \\ & $	Nonaisomer / 4,3-Isomer
8	$ \begin{array}{c} & & & \\ & $	Nonaisomer / 3,4-Isomer
9	$\begin{array}{c} & & & \\ & & & & \\ & & & \\ & & & & \\ & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\$	Nonaisomer / 3,3-Isomer
10	$\begin{array}{c} & & & & \\ & & & & \\ & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & &$	Triantennäre Struktur

### Abbildung 2: Strukturen der untersuchten Oligosaccharide (Strukturen 1-10)

#### Abbildung 3: Bibliothek der Strukturen I-XIII



### 1 Einleitung

### 1.1 <u>N-Glycane – Struktur und Funktion</u>

Das Glycom bezeichnet in Analogie zum Proteom und dem Genom alle Glycane eines Organismus.<sup>1,2</sup> Dazu gehören Glycolipide, Glycosaminoglycane und Glycoproteine. Die Glycosylierung von Proteinen stellt ihre häufigste und komplexeste Modifikation dar.<sup>3</sup> Die statistische Auswertung bekannter Proteinstrukturen resultiert in der Schätzung, dass etwa 50 % aller Proteine glycosyliert vorliegen.<sup>4</sup> Das humane Proteom ist nach aktuellen Erkenntnissen sogar zu mehr als 70 % glycosyliert.<sup>5,6</sup>

Die Verknüpfung der Glycane kann dabei hauptsächlich in N-Typ und O-Typ Glycosylierungen unterteilt werden. N-Glycane sind im Rahmen der Konsensussequenz Asn-X-Ser/Thr an die Seitenkette des Asparagins gebunden und weisen eine gemeinsame Grundstruktur auf, den core bestehend aus Man<sub>3</sub>GlcNAc<sub>2</sub>. X kann dabei jede beliebige Aminosäure außer Prolin sein. Die an diesen core verknüpften Monosaccharidbausteine entscheiden über die Klassifizierung als hochmannosid-, komplex- oder hybrid-Typ. Der Aufbau von N-Glycanen erfolgt zunächst im endoplasmatischen Retikulum (ER) durch die Übertragung des uniformen Oligosaccharidbausteins Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub> vom membrangebunden Dolichylpyrophosphat zum Protein. Anschließend wird dieser durch enzymatische Ab- und Aufbaureaktionen cotranslational im ER und posttranslational im Golgi-Apparat modifiziert.<sup>7</sup> Schon während der Translation des Proteins können die Zuckerstrukturen entscheidenden Einfluss auf die Faltung haben.<sup>8,9</sup> Den Glycanen kommen darüber hinaus stabilisierende, schützende und vermittelnde Funktionen zu.<sup>10-13</sup> Das Glycosylierungsmuster kann dabei essentiell für z.B. Erkennungsprozesse von Zellen, Antikörpern oder Pathogenen sein, aufgrund derer Erkrankungen abgewehrt werden oder entstehen können.<sup>6</sup>

Eins der unzähligen Beispiele hierfür ist der Infektionsbeginn durch Influenza A Viren, der durch die Wechselwirkung des Virusproteins Hämagglutinin mit terminalen Sialinsäureresten der Wirtszellen zustande kommt. Dabei ist die Bindung der Hämagglutininspezies abhängig vom Virustyp entweder hochspezifisch für die 2,6- oder 2,3-Verknüpfung der terminalen Sialinsäurereste und damit für die von der Infektion betroffene Spezies (z.B. Mensch (H3N2 und H1N1) oder Vogel (H5N1)).<sup>14,15</sup>

Abweichende Glycosylierungsmuster werden häufig im Zusammenhang mit Krebs erkannt. Die Zusammensetzung der Glycanstrukturen codiert gewissermaßen den physiologischen Zustand einer Zelle.<sup>16</sup> Diese Veränderung kann als Biomarker für diagnostische Zwecke oder als *target* für Medikamente und Vakzine genutzt werden.<sup>3,17</sup> Beispielsweise wurde bei Brustund Darmkrebs ein verstärktes Auftreten von GlcNAc in  $\beta$ 1,6-Verknüpfung am Man- $\alpha$ 1,6-Man-Arm beobachtet.<sup>18,19</sup>

Krebs wird häufig mit chimären, rekombinant erzeugten humanen Antikörpern adressiert und behandelt. Dabei kann die Rolle der Glycane entscheidend für die Wirksamkeit und/oder die Verträglichkeit des Medikaments sein. Retuximab, ein zur Behandlung von non-Hodgkin-Lymphom-Erkrankungen eingesetztes Medikament, ist in seiner Wirksamkeit durch einen hohen Fucosylierungsgrad reduziert.<sup>20,21</sup> Patienten mit hypersensitiven Reaktionen auf Cetuximab, einem gegen Darm- und Hautkrebs zugelassenen Medikament, zeigten eine erhöhte Antikörper-Produktion gegen das  $\alpha$ -Gal-Epitop (Gal- $\alpha$ 1,3-Gal- $\beta$ 1,4-GlcNAc) welches im Glycosylierungsmuster von Cetuximab auftritt.<sup>22,23</sup> Eben dieses Epitop ist auch dafür verantwortlich, dass im Menschen eine Abstoßungsreaktion von porcinen Organen stattfindet (Xenotransplantation).<sup>24</sup>

Auch bei Erythropoietin (EPO), einem Glycoprotein im Einsatz zur Therapie von Anämie, ist das Glycosylierungsmuster von entscheidender Bedeutung. Der Grad an Sialylierung entscheidet hier über die Aktivität des Proteins und die Lebensdauer im Blutstrom.<sup>25</sup> EPO ist mit einem weltweiten Umsatz von über 10 Milliarden US\$ das kommerziell erfolgreichste Biopharmazeutikum.<sup>6</sup>

Für alle hier vorgestellten Beispiele ist eine effiziente Analytik essentiell, um die Funktion von Glycanen besser verstehen zu können, um die Früherkennung von Krebs und mögliche Behandlungsstrategien auszubauen und um die Qualitätskontrolle von Biopharmazeutika<sup>26</sup> zu gewährleisten.

### 1.2 Analytik von Oligosacchariden

Die Analytik von Oligosacchariden ist aus mehreren Gründen eine komplexe und zeitaufwändige Aufgabe. Zum einen gibt es für Saccharide keine Amplifikationsmethode, wie sie beispielsweise bei DNA eingesetzt wird, sodass Glycane in physiologischen Mengen untersucht werden müssen.<sup>27</sup> Zum anderen ergibt sich aus den vielfältigen Verknüpfungsmöglichkeiten und den daraus resultierenden Verzweigungen eine ungleich höhere Diversität an verschiedenen Glycanstrukturen im Vergleich zu linear aufgebauten DNA- und Proteinsequenzen. Beispielsweise errechnet sich die Anzahl möglicher Oligomere bei der Verknüpfung von sechs Bausteinen zu 4096 für Nucleotide, 64 Millionen für Peptide

für und 192 Milliarden für Kohlenhydrate (berechnet 10 verschiedene Monosaccharidbausteine, die in  $\alpha$ - und  $\beta$ -Konfiguration auftreten können).<sup>28</sup> Auch wenn durch den Biosyntheseweg eine gewisse Limitierung aller möglichen Strukturen angenommen werden kann, so sind in der bisher umfangreichsten Sammlung von Zucker-Daten - der bis 1998 geführten Complex Carbohydrate Structure Database (CCSD) – über 20000 verschiedene Strukturen von Glycanen und Glycoconjugaten zu finden.<sup>29-31</sup> Innerhalb eines einzelnen Proteins kann Mikroheterogenität beobachtet werden, die dazu führt, das Proteine durch verschiedene Glycosylierungen mehrere Glycoformen ausbilden, die in ihrer Aktivität variieren können.<sup>32</sup> Beispielsweise wurden an der einzigen Glysoylierungsstelle der  $\alpha$ -Untereinheit des oLH (*ovine luteinizing hormone*) mehr als 20 verschiedene Glycanstrukturen gefunden.<sup>33</sup>

Experimentelle Methoden zur Analytik von Glycanen sind enzymatische Ansätze wie Lectin-Bindungsstudien, HPLC (high performance liquid chromatography) in Verbindung mit Exoglycosidase-Verdaustudien und HPLC in Verbindung mit Massenspektrometrie (MS), sowie der Einsatz von NMR-Spektroskopie.<sup>3</sup> Aufgrund der Komplexität von Glycanen ist eine einzelne Methode allein meist nicht ausreichend für eine zuverlässige und vollständige Charakterisierung.<sup>32,34</sup> Für die meisten Verfahren wird zunächst die Abspaltung der Glycane vom Glycoprotein oder von Glycopeptiden vorgenommen. Diese kann enzymatisch beispielsweise Glycosidasen wie PNGase F (Peptide- $N^4$ -(N-Acetyl- $\beta$ durch glucoaminyl)asparagine amidase) für N-Glycane oder chemisch durch Hydrazinolyse erreicht werden.<sup>35,36</sup> Weitere mögliche Schritte sind dann je nach den Erfordernissen der verwendeten Methode das *labelling* der Strukturen, die Reduktion des reduzierenden Endes zum Alditol oder die Behandlung mit Exoglycosidasen wie der Neuraminidase zur Abspaltung von definierten, terminal auftretenden Zuckerresten.<sup>37</sup>

### 1.2.1 Chromatographische Methoden zur Trennung von Glycanen

Da von natürlichen Proben aufgrund der zuvor beschriebenen Mikroheterogenität Glycangemische erhalten werden, ist der Einsatz von modernen Trennmethoden meist unumgänglich. Der Einsatz von HPLC für die Trennung von Saccharidgemischen wird dabei seit den 1980er Jahren weiterentwickelt.<sup>3</sup> Typischerweise werden die Saccharidstrukturen für die Trennung reduziert und/oder mit Modifikationen am anomeren Zentrum versehen, um damit die Trennproblematik durch das Auftreten von  $\alpha$ -/ $\beta$ -Gemischen zu umgehen und beispielsweise durch Fluoreszenzmarker einen Sensitivitätsgewinn bei der Detektion zu erreichen.<sup>38</sup> Dadurch sind einerseits zusätzliche Schritte nötig, andererseits erschwert es die

3

Auswertung von NMR-Daten (s. a. Abschnitt 1.2.3). Häufig für die Trennung von Glycanen eingesetzt werden HILIC- (*hydrophilic-interaction chromatography*) und PGC-Säulen (*porous graphitized carbon*). Die Entwicklung von PGC-Säulen ermöglicht die Trennung von underivatisierten Glycanstrukturen und ergibt charakteristische Retentionszeiten, die die Charakterisierung vereinfachen können.<sup>39,40</sup> Einfluss auf die noch nicht vollständig verstandenen Trennmechanismen haben hydrophobe und ionische Wechselwirkungen.<sup>41-43</sup> Je komplexer aber die Gemische werden, umso schwieriger gestaltet sich gerade bei strukturell ähnlichen Isomeren die Trennung.<sup>41</sup> LC ist eine Methode, die sehr gut für die Kopplung zu weiteren analytischen Verfahren wie MS aber auch NMR-Spektroskopie geeignet ist.<sup>3</sup>

### 1.2.2 Massenspektrometrie zur Analytik von Glycanen

Die Massenspektrometrie ist eines der am häufigsten eingesetzten Verfahren zur Analytik von Glycanen und Glycopeptiden. Die hohe Sensitivität, die das Messen von Substanzmengen im Bereich von Picomol bis Femtomol erlaubt, und die einfache Kombinierbarkeit mit LC-Methoden sind große Vorteile dieses Verfahrens.<sup>32</sup> Allerdings sind die Unterscheidung von isobaren Strukturen, die Identifikation von Verknüpfungen und das Auftreten von Verzweigungen meist schwierig und nur mit großem Aufwand lösbar. Häufig werden aus dem Abgleich mit Datenbanken und auf der Grundlage des Biosynthesewegs von Glycanen Strukturvorschläge erstellt. Genaue Informationen über anomere Konfiguration und erhältlich.40 Verknüpfungen sind dadurch aber nicht Abhilfe bezüglich der Verknüpfungsinformationen können Methylierungsanalysen und die Aufnahme von MS<sup>n</sup>-Daten bringen. Durch die Notwendigkeit von quantitativer chemischer Umsetzung bei Methylierungsreaktionen ist eine Automatisierung schwierig. MS<sup>n</sup>-Daten wiederum benötigen die vollständige Trennung der Analyten und geben nicht immer zuverlässige Informationen.<sup>44</sup> Beispiel hierfür ist die Umlagerung von Fucoseresten die zu fehlerhaften Interpretationen führen kann.<sup>45,46</sup> Die Entwicklung von Programmen zur automatischen Zuordnung von Fragmentspektren kann eine Vereinfachung der Interpretation und die Eingrenzung möglicher Strukturen mit sich bringen. Eine Erhöhung der Informationsdichte ergibt sich durch die Kombination mit den aus LC erhältlichen Retentionszeiten.<sup>47</sup>

### 1.2.3 <u>NMR-Spektroskopie – Das Konzept der structural reporter groups</u>

Durch NMR-spektroskopische Daten ist die zuverlässige und vollständige Aufklärung von Glycanen mit der Information über Verknüpfungspositionen, anomere Konfigurationen und Unterscheidung der Monosaccharidbausteine möglich. Dafür stehen diverse zweidimensionale Methoden wie TOCSY, COSY, NOESY, HSQC und HMBC zur Verfügung, die allerdings aufgrund der relativ geringen Sensitivität der NMR-Spektroskopie gegenüber anderen Verfahren wie MS viel Substanz erfordern, sowie Expertise in der Spektreninterpretation benötigen. Meist wird eine vollständige Trennung der Substanzen benötigt und automatische Methoden zur Zuordnung haben sich hier noch nicht durchgesetzt. Doch auch in 1D-<sup>1</sup>H-NMR-Spektren sind umfassende Informationen vorhanden, die durch Anwendung des structural reporter group-Konzepts nach Vliegenthart et al. genutzt werden können.<sup>48-50</sup> Als structural reporter groups werden Protonen-Signale bezeichnet, die separiert vom bulk-Bereich der meisten Skelettprotonen (etwa 3.4 - 4.0 ppm) bei genau definierten chemischen Verschiebungen auftreten. Dazu gehören anomere Protonen, H-2 von Mannoseresten, H-3 von Neuraminsäurederivaten, H-3, H-4 und H-5 Protonen von Galactoseresten, die H-5 und CH<sub>3</sub>-Protonen von Fucoseeinheiten, sowie N-Acetylsignale von Hexosaminen und N-Acetyl- bzw. N-Glycolylsignale von Neuraminsäureresten. Die Größe der Kopplungskonstanten der H-1-Protonen, die in der Feinstruktur der Signale erkennbar ist, erlaubt schnellen Zugriff auf die anomere Konfiguration. Die genaue chemische Verschiebung ist abhängig von Art und Position der verknüpften Zuckerreste und ergibt unter Standardbedingungen (D<sub>2</sub>O, 300 K, Aceton-CH<sub>3</sub> als Standard bei 2.225 ppm) vergleichbare Ergebnisse. Durch den Aufbau der Datenbank sugabase dokumentierten van Kuik et al. die <sup>1</sup>H-NMR-Daten von über 1400 verschiedenen Saccharidstrukturen, die sich anhand ihrer structural reporter group-Signale unterscheiden lassen.<sup>51,52</sup> So können reine Spektren von Strukturen durch den Abgleich mit dieser Datenbank der eindeutigen Charakterisierung dienen. Im Rahmen der dieser Arbeit vorausgehenden Studien konnte gezeigt werden, dass durch Anwendung des structural reporter group-Konzepts die Charakterisierung eines komplex-Typ-Glycans in Substanzmengen von nur 15 pmol möglich ist (s. a. Abschnitt 2.2).<sup>53</sup> Relevant hierfür war die Entwicklung der Cryotechnologie beim Bau von NMR-Probenköpfen, durch die bis zu vierfach bessere Sensitivität und damit einhergehend die 16 fache Reduktion der Messzeit erreicht wurde.<sup>54</sup> Dadurch treten die Vorteile der NMR-Spektroskopie in Form einer zerstörungsfreien und zuverlässigen Analytik gegenüber der geringeren Sensitivität in den Vordergrund, weil die Spektrenakquisition von physiologischen Mengen in der Größenordnung von nur noch einigen Nanogramm möglich ist.

### 1.2.4 Datenbanken – CCSD, sugabase, glycosciences.de und GlycomeDB

Die erste Zusammenstellung von glycoassoziierten Daten erfolgte 1986 durch die Gründung der *Complex Carbohydrate Structure Database* (CCSD), häufig auch nach dem damit

verbundenen Datenverarbeitungsprogramm *CarbBank* genannt.<sup>30,31</sup> *CarbBank* beinhaltete Struktur- und Literaturinformation von Sacchariden. Aufbauend auf *CarbBank* wurde Anfang der 1990er die Datenbank *sugabase* entwickelt, in der NMR-Daten entsprechend des *structural reporter group*-Konzepts unter Verwendung der gleichen Terminologie hinterlegt wurden. Aufgrund mangelnder Finanzierung wurden sowohl *CarbBank* als auch *sugabase* Mitte bzw. Ende der 1990er nicht mehr fortgeführt<sup>3,29,55</sup>, beide Datenbanken waren aber bis Oktober 2011 *online* verfügbar.<sup>56,57</sup> *Sweet-DB*, später zu *glycosciences.de* umbenannt, stellte schließlich eine Fortsetzung von *sugabase* dar, in der viele *CarbBank*-Einträge, sowie ein Großteil der *sugabase*-Einträge übernommen wurden.<sup>29,55</sup> Allerdings wurde eine neue Identifikationsnomenklatur eingeführt, die eine Möglichkeit zur linearen Repräsentation von Saccharidstrukturen bietet, und den Namen LinucsID (Linear Notation for Unique Description of Carbohydrate Sequences, entwickelt für automatische, computerbasierte Prozessierung) erhielt.<sup>58</sup> Eine Suche anhand der CCSD-Nummern ist in *glycosciences.de* nicht möglich.



CCSD-Schreibweise (CCSD in sugabase: 42857, LINUCS-ID: 1124):



### Abbildung 4: Struktur eines komplex-Typ Oligosaccharids und die entsprechende Darstellung in CFGund CCSD -Schreibweise

*Glycosciences.de* ist zusätzlich mit weiteren NMR-Dateneinträgen, sowie der Möglichkeit zur Berechnung der chemischen Verschiebungen von Kohlenhydratprotonen anhand von empirischen Einträgen ausgestattet (*chemical shift estimation* durch *CASPER*).<sup>59</sup> Die durch

CASPER berechneten Spektren sind in Form von Stabdiagrammen dargestellt, die als alleinige Information die chemische Verschiebung beinhalten. Neben der Möglichkeit, die Qualität der Dateneinträge bezüglich Vollständigkeit und plausibler Zuordnung zu überprüfen, können auch die Daten unbekannter Strukturen errechnet werden.<sup>3</sup> Vermutlich führte diese Überprüfung zu dem Ausschluss einiger *sugabase*-Einträge wie beispielsweise eines Dodecasaccharids (B42857, LinucsID: 23927), bei welchem einige wesentliche *structural reporter group*-Signale nicht dokumentiert sind.

Die Datenbanksuche zur Interpretation eines gemessenen NMR-Spektrums erfolgt sowohl bei *sugabase* als auch bei *glycosciences.de* immer anhand der reinen chemischen Verschiebung, also ohne die Verwendung von Intensitäts- und Kopplungsinformation. Die Struktursuche erfolgt in beiden Datenbanken anhand der CCSD-Schreibweise (Abbildung 4), die eine intuitive Suche erschwert.

Auch für die Auswertung von MS-Daten kann auf Datenbanken zurückgegriffen werden. Ein Werkzeug dazu stellt die GlycoWorkbench dar, welche für eine Datenbanksuche und die Annotation von MS<sup>n</sup>-Daten entwickelt wurde, aber keine *scoring*-Funktion zur Bewertung der Suchergebnisse beinhaltet.<sup>60</sup> Sie nutzt beispielsweise GlycomeDB, eine Meta-Datenbank, die Daten aller frei verfügbaren Datenbanken integriert hat (*CFG (Consortium for Functional Glycomics), KEGG, glycosciences.de, BDSDB, CarbBank, EuroCarbDB, Glycobase,* sowie Kohlenhydrat-assoziierte Daten der *PDB*). GlycomeDB kann nach den verschiedenen Datenbank IDs der integrierten Datenbanken durchsucht werden. Im August 2010 waren 35873 unterschiedliche Saccharidsequenzen in GlycomeDB aufgeführt, von denen 11822 vollständig charakterisiert sind.<sup>61-63</sup>

### 1.3 Konzepte zur Kombination von analytischen Methoden

MS und NMR-Spektroskopie werden meist als komplementäre Methoden betrachtet, deren Informationsgewinn separat interpretiert und dann zusammengeführt wird. Es gibt Methoden, die eine Kombination analytischer Verfahren vorsehen. Beispielsweise wurde in Studien zur Untersuchung von Metaboliten im menschlichen Urin eine Kopplung von HPLC-NMR-MS genutzt. Aus der Auftragung von chemischer Verschiebung gegen die Zeit konnten so anhand der aus MS erhältlichen Retentionszeit einzelne Projektionen für die Auswertung entnommen werden.<sup>64,65</sup> Relevant ist hierbei die vollständige chromatographische Trennung der Substanzen, die durch Überlappung im NMR-Spektrum die Charakterisierung verhindern.<sup>66</sup> Ein weiterer Ansatz zur Integration von NMR-Spektroskopie und MS sieht das *statistical heterospectroscopy* (SHY) Modell vor, welches ebenfalls in der Detektion von Metaboliten

Einsatz findet. Bei dieser Methode zur Koanalyse von spektroskopischen Daten über eine Vielzahl an Proben wurde das Auftreten von NMR-Signalen in einzelnen Proben mit dem Auftreten von MS-Signalen in den gleichen Proben korreliert, um daraus eine Zusammengehörigkeit ableiten zu können.<sup>67</sup> Für den Einsatz bei Metabonomstudien wurde außerdem die *statistical total correlation spectroscopy* (STOCSY) entwickelt, die auf LC-NMR basiert. Bei dieser Methode werden im NMR-Spektrum *driver*-Signale definiert, die zu einzelnen Verbindungen gehören. Alle weiteren NMR-Signale werden innerhalb von aufeinander folgenden NMR-Spektren mit den einzelnen *driver*-Signalen korreliert. Dadurch werden ähnlich wie bei TOCSY-Experimenten die chemischen Verschiebungen einer Komponente (eines Spinsystems) identifiziert.<sup>68,69</sup>

Eine weitere Möglichkeit zur optimierten Datennutzung stellt die Kombination von NMR-Spektroskopie mit Computer-basierten Methoden dar. Durch künstliche neuronale Netzwerke konnten beispielsweise Spektren von Komponenten erkannt werden, die sich lediglich in einem von 20 Saccharidresten unterschieden. Selbst bei geringfügigen Abweichungen der chemischen Verschiebungen, wie sie durch Messung an unterschiedlichen Geräten auftreten können, konnten Spektren mit einem Signal-zu-Rausch-Verhältnis von 1.25 und weniger erkannt werden.<sup>70,71</sup>

### 2 Methoden

### 2.1 Mathematische Beschreibungen der Korrelation zweier Variablen

Die Korrelation zweier Variablen x und y ist eine Möglichkeit zur Überprüfung, ob eine Verknüpfung beider Variablen besteht. Es gibt verschiedene Modelle, nach der die Ähnlichkeit der beiden Variablen bewertet werden kann. Im Rahmen dieser Arbeit wird die Berechnung des Korrelationskoeffizienten nach Pearson, bzw. der Rangkorrelationskoeffizienten nach Spearman oder Kendall verwendet.<sup>72-74</sup>

#### Der Korrelationskoeffizient nach Pearson

Der Korrelationskoeffizient nach Pearson (auch Bravais-Pearson) errechnet sich über die Kovarianz, normiert über die Standardabweichung zweier Variablen (Gleichung 1). Er gilt für die lineare Beziehung zweier Variablen. Die Kovarianz ist ein Maß für die Ähnlichkeit von x und y (Gleichung 2). Wird die Variable y im gleichen Maß größer wie x, so zeigen sie eine lineare Abhängigkeit und damit eine positive Kovarianz. Entsprechend bedeutet negative Kovarianz, dass die Variable y im gleichen Maße kleiner wird wie x größer wird, sie zeigen umgekehrt proportionale Verläufe. Ist kein Zusammenhang erkennbar, wird die Kovarianz null.

$$c = \frac{Cov(\mathbf{x}, \mathbf{y})}{\sigma_{\mathbf{x}} \cdot \sigma_{\mathbf{y}}} \qquad \text{Gleichung 1}$$

$$Cov(x,y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$
 Gleichung 2

c – Korrelationskoeffizient, x<sub>i</sub>, y<sub>i</sub> – Werte von x,y,  $\overline{x}$ ,  $\overline{y}$  – Mittelwerte von x,y, Cov(x,y) - Kovarianz von x,y,  $\sigma_{x,y}$  – Standardabweichung von x,y

#### **Rangkorrelationskoeffizient nach Spearman**

Rang- oder ordinalskalierte Merkmale sind Daten, die zwar durch Einstufung in Rangpositionen eine Vergleichbarkeit aufweisen, deren Abstände aber keine quantitative Aussage zulassen. Ein Beispiel ist die Korrelation von Schwefeldioxidgehalt und der Einstufung der geschmacklichen Qualität von Wein (die nach Spearman einen Koeffizienten von 0.93 ergibt).<sup>72</sup> Während die Schwefeldioxidkonzentrationen unterschiedlicher Weine als absolute Werte auf einer metrischen Skala eingestuft werden können, so ist die Kategorisierung des Geschmacks ein ordinalskaliertes Merkmal. Die einzige zulässige Aussage ist, dass ein Wein mit der Rangposition 1 besser bewertet wurde, als ein Wein mit der Rangposition 2 und dieser als ein Wein mit Rangposition 3. Es kann keine Aussage getroffen werden ob der Wein mit Rangposition 2 geschmacklich eher dem mit Rang 1 oder dem mit Rang 3 ähnelt.<sup>72</sup> Spearman erweiterte die Korrelation nach Pearson durch die Verwendung dieser Rangpositionen anstelle der absoluten Werte bzw. Daten. So lassen sich ordinalskalierte Daten in kardinale Merkmale (also Zahlenwerte) umwandeln und vergleichen. Die Rangkorrelation ist auch für metrische Daten, also solche, für die ein quantifizierbarer Zusammenhang besteht, anwendbar und ist hier toleranter gegenüber einzelnen Werten mit großer Abweichung vom linearen Verlauf. Allerdings geht der metrische Informationsinhalt dabei verloren.

$$c_s = \frac{\sum_{i=1}^{n} (rg(x_i) - \bar{rg}_x) (rg(y_i) - \bar{rg}_y)}{\sigma_{rg(x)} \cdot \sigma_{rg(y)}}$$
Gleichung 3

c<sub>s</sub> – Korrelationskoeffizient nach Spearman, rg(x),rg(y) – Rang von x,y

### **Rangkorrelation nach Kendall**

Ebenso wie die Methode nach Spearman wird die Korrelation durch Zuordnung von Rängen erreicht. Anschließend wird die Korrelation darüber ermittelt, ob die Entwicklung der Ränge in die gleiche Richtung erfolgt, d.h. ob der Rang der beiden Variablen x und y vom ersten Datenpunkt zum zweiten Datenpunkt bei beiden anwächst bzw. abnimmt usw. Im Gegensatz zur Korrelation nach Spearman werden jeweils nur die relativen Rangänderungen zweier Wertepaare betrachtet. Durch den Vergleich der übereinstimmenden Rangentwicklungen mit den unterschiedlich verlaufenden Rangpositionen über die Gesamtanzahl an verglichenen Daten ergibt sich hier der Rangkorrelationskoeffizient. Im Ergebnis zeigt die Verwendung des Rangkorrelationskoeffizienten nach Kendall große Ähnlichkeiten zum Rangkorrelationskoeffizienten nach Spearman.<sup>73,75</sup>

### 2.2 Optimierte Wasserunterdrückung bei der NMR-Spektrenakquisition

Da es für Glycane keine Amplifikationsmethode wie für DNA-Stränge gibt, müssen diese in physiologischen Mengen gemessen werden. Meist wird die NMR-Spektroskopie als unsensitive Methode betrachtet, die nur für Substanzmengen von einigen nanomol einsetzbar ist.<sup>76,77</sup> Dabei liefert die NMR-Spektroskopie wie in Abschnitt 1.2.2 vorgestellt eindeutige und zuverlässige Ergebnisse und ist somit eine äußerst wertvolle Methode zur Charakterisierung von Strukturen. In dieser Arbeit vorausgehenden Untersuchungen wurde gezeigt, dass Substanzmengen von 15 pmol für eine Charakterisierung anhand des *structural reporter* 

*group*-Konzepts reichen. Dafür müssen die experimentellen Parameter an die geringen Substanzmengen angepasst werden.

Neben einer optimierten Probenpräparation ist die Verwendung einer geeigneten Wasserunterdrückung essentiell, da Deuteriumoxid (D<sub>2</sub>O) aufgrund der Ähnlichkeit zu physiologischen Bedingungen das Standardlösungsmittel für Glycane ist. Neben der Möglichkeit zur Unterdrückung des Lösungsmittelsignals durch *watergate*<sup>78</sup>- und *presaturation*<sup>79</sup>-Pulssequenzen wurde die *excitation sculpting*<sup>80</sup>-Pulssequenz nach Shaka *et al.* als die geeignetste Methode eingestuft.<sup>53</sup> Das *excitation sculpting* basiert auf einer *double pulsed field gradient spin-echo*-Sequenz, bei der durch einen selektiven Puls die Solvenzresonanz invertiert wird, sodass diese als einzige eine Dephasierung durch die flankierenden Gradienten erfährt und so unterdrückt wird.<sup>81</sup> Unter Verwendung der Standardparameter wurde beobachtet, dass Signale in der Nähe zur Einstrahlfrequenz des selektiven Pulses (50-200 Hz) drastisch reduziert werden (Abbildung 6, oben). Außerdem lässt sich bei Spektren mit geringem Signal-zu-Rausch-Verhältnis eine intensive Störung der Basislinie beobachten (Abbildung 5).



Abbildung 5: Auswirkung der Standardparameter der *excitation sculpting*-Pulssequenz. Oben: Referenzspektrum einer 5 mM ( $0.4 \mu mol/V_{obs}$ ) Sucrose-Probe in D<sub>2</sub>O. Unten: Spektrum von 13 pmol/ $V_{obs}$ der Sucrose. Beide Spektren wurden mit den Standardparametern der *excitation sculpting*-Pulssequenz akquiriert. Während im oberen Spektrum eine gute Wasserunterdrückung erreicht wird, zeigt sich bei der geringkonzentrierten Probe eine intensive Störung der Basislinie. Trotzdem sind die Signale der Sucrose neben einigen geringen Verunreinigungen deutlich erkennbar.



Abbildung 6: Auswirkung der optimierten Wasserunterdrückung (excitation sculpting) am Beispiel eines komplex-Typ Decasaccharids (Struktur 2). A und B: Spektren einer 1 mM Probe des Saccharids (entspricht 80 nmol/V<sub>obs</sub>) in H<sub>2</sub>O/D<sub>2</sub>O 9:1. Spektrum A wurde mit den von Shaka *et al.*<sup>80</sup> publizierten Parametern aufgenommen, während Spektrum B mit optimierten Parametern akquiriert wurde. C: Aufnahme von 25 pmol/V<sub>obs</sub> des Decasaccharids in  $D_2O$ , akquiriert mit optimierter Wasserunterdrückung. Durch die Optimierung können für die Charakterisierung wichtige structural reporter group-Signale benachbart zum H<sub>2</sub>O- bzw. HDO-Signal beobachtet werden, die unter Verwendung der Standardparameter unterdrückt bzw. in ihrer Intensität reduziert werden. Diese sind selbst bei dem Spektrum von nur 25 pmol/Vobs besser erkennbar als bei Spektrum A von der um Faktor 3000 größeren Substanzmenge. Außerdem ist durch die Optimierung eine ungestörte Basislinie zu beobachten (s. a. Abbildung 5).<sup>53</sup>

Durch die Verlängerung des selektiven Rechteckpulses von 2 ms auf 8 ms wird die Anregungsbreite auf etwa 30 Hz reduziert, die ausreicht, um das Lösungsmittelsignal zu unterdrücken. Dabei wird die Detektion der Signale in Abständen von 40-200 Hz zur Einstrahlfrequenz entscheidend verbessert und Artefakte in der Form der gestörten Basislinie werden verhindert. Bei der Messung einer Probe von 25 pmol/ $V_{obs}$  ( $V_{obs}$  – in der NMR-Spule beobachtbares Volumen) eines komplex-Typ Decasaccharids konnte so eine Reduktion des HDO-Signals um Faktor 500000 erreicht werden, die für die Auswertbarkeit der NMR-Spektren essentiell ist. Abbildung 6 zeigt, dass durch diese Optimierung Signale im Spektrum der Probe von nur 25 pmol/ $V_{obs}$  des Decasaccharids sichtbar sind, die nicht einmal im Spektrum der 3000fach konzentrieren Probe unter Verwendung der Standardparameter auftreten.<sup>53</sup>

Begleiterscheinung der Verlängerung des selektiven Pulses ist das verstärkte Auftreten eines Dispersionsanteils im Signal durch anti-Phasenmagnetisierung,<sup>82</sup> der in Abhängigkeit von der Größe der Kopplungskonstanten zu einer Verzerrung der Signale führen kann. Daraus kann eine ungenaue Bestimmung der Kopplungskonstanten resultieren. Dieser Effekt muss gegenüber dem Informationsgewinn durch detektierbare Signale in der Nähe des Lösungsmittels und der Vermeidung des Basislinienartefakts abgewogen werden.

### 2.3 Strukturaufklärung mit selektiven TOCSY-Experimenten

Die Aufnahme von selektiven TOCSY-Experimenten erlaubt die gezielte Detektion einzelner Spinsysteme, wie sie jede Monosaccharideinheit eines Glycans darstellt. Dazu wird eine einzelne Resonanz durch einen selektiven Puls angeregt, sodass der TOCSY-Transfer während des in der Pulssequenz integrierten *spin-lock*-Feldes nur auf Spins des gleichen Spinsystems erfolgt. Gerade bei Kohlenhydraten ergibt sich so die Möglichkeit, auch in Überlappungsbereichen die Signale einzelner Saccharidreste zu identifizieren. Ein Vorteil der 1D-Sequenz ist die gezielte Informationsextraktion, die dann entsprechend mit besserer Auflösung und gegebenenfalls einer Reduktion an benötigter Messzeit, wie sie für ein 2D-Experiment erforderlich wäre, einhergeht.<sup>81,83,84</sup>

### 2.4 Terminologie in der Massenspektrometrie

Bei der Aufnahme von LC-MS-Daten werden *total ion currents* (TIC) erhalten, die dem Chromatographieprofil entsprechen, da sie die Detektion des gesamten Ionenstroms zu jedem Zeitpunkt beinhalten. Ein TIC ist somit die Summe aller detektierten Ionen über die Zeit. Eine bereinigte Darstellung des TICs ist das *base peak chromatogram* (BPC), welches zu jedem Zeitpunkt jeweils nur das intensivste Signal darstellt. Die Spur eines Ions (m/z) über die Zeit, also entlang der aufeinanderfolgenden MS-Spektren, wird *extracted ion chromatogram* (EIC, selten auch: *extracted ion current*) genannt.<sup>85</sup>

### 3 Aufgabenstellung

Den Oligosaccharidstrukturen von Glycoproteinen kommen vielfältige Aufgaben zu. Sie können schützende, stabilisierende, aber auch vermittelnde Aufgaben erfüllen und sind an zellulären Ereignissen wie der Zell-Zell-Erkennung und Apoptose beteiligt.<sup>86</sup> Sie sind somit involviert in viele physiologische und pathologische Prozesse wie beispielsweise der Entstehung und Ausbreitung von Krebs. Bei vielen Arten von Krebs werden veränderte Glycosylierungsmuster beobachtet, sodass die Glycane als Biomarker für Diagnose und Behandlung fungieren können. Für das Verständnis der Funktion von Glycanen, für die Erkennung von krankhaften Veränderungen und für die Qualitätssicherung von Biophamarzeutika ist eine konsequente und effiziente Analytikstrategie notwendig. Die Charakterisierung von Glycanen ist eine komplizierte und häufig zeitaufwendige Aufgabe, da Oligosaccharide durch die Vielzahl an möglichen Verknüpfungen und dem Auftreten von Verzweigungen eine hohe Komplexität aufweisen und meist nur limitierte Substanzmengen zur Verfügung stehen. Die Mikroheterogenität von Glycoproteinen führt dazu, dass Glycane meist in komplexen Gemischen erhalten werden, deren Trennung durch ähnliche und häufig massenisobare Strukturen aufwendig ist. Zurzeit bietet keine der vielfach eingesetzten Methoden, wie NMR-Spektroskopie, MS, HPLC oder enzymatische Verfahren alleine eine zuverlässige und schnelle Charakterisierung von Glycangemischen.

Ziel dieser Arbeit ist somit die Entwicklung neuer Strategien zur Analytik von Glycoproteinen. Dabei stehen zwei Aspekte im Vordergrund: zum einen soll durch die Berechnung von NMR-Spektren anhand von Datenbank-Einträgen eine automatisierte Identifikation von <sup>1</sup>H-NMR-Spektren von Glycanstrukturen entwickelt und überprüft werden. Zum anderen soll durch die Integration von MS und NMR-Spektroskopie über den zeitlichen Zusammenhang durch LC eine Nutzung aller im Chromatogramm vorhandener Information ermöglicht werden, auch wenn eine unvollständige Trennung der Substanzen vorliegt. Dazu soll die Nutzbarkeit von Kreuzkorrelation von LC, MS- und NMR-Daten zur Extraktion von reinen NMR-Spektren überprüft und daraus eine neue Methode zur 3D-Kreuzkorrelation entwickelt werden. Die Methodik soll dabei zunächst anhand eines simulierten Systems getestet werden, um sie dann auf ein vom bovinen Fibrinogen erhaltenes Gemisch als reales Beispiel auszuweiten. Dabei soll eine möglicht soll überprüft werden, ob beide Strategien der automatischen Interpretation und der Spektrenextraktion zusammengefügt werden können, um eine automatische Erkennung der extrahierten Daten zu erhalten.

### 4 Ergebnisse und Diskussion

### 4.1 Entwicklung einer Methode zur automatischen Spektrenidentifikation

Durch Anwendung des *structural reporter group*-Konzepts können 1D-<sup>1</sup>H-NMR-Spektren von Oligosacchariden zur Charakterisierung der Strukturen verwendet werden. Eine automatische Zuordnung von Spektren ist insofern problematisch, da die Datenbanksuche nicht für die Identifikation mehrerer Komponenten ausgelegt ist und auf den reinen chemischen Verschiebungen basiert. Kopplungen und Intensitäten werden nicht berücksichtigt. Im Folgenden wird gezeigt, dass der Abgleich von berechneten Spektren mit experimentell erhaltenen Spektren eine automatische Zuordnung erlaubt. Die Berechnung der Spektren basiert dabei auf den Einträgen in *sugabase*.

### 4.1.1 Berechnung von Spektren anhand von Datenbankeinträgen

Die NMR-Datenbank, die im späteren Verlauf den Namen *sugabase* erhielt, wurde 1992 in der Literatur vorgestellt.<sup>51,52</sup> Sie entstand in Anlehnung an die wenige Jahre zuvor ins Leben gerufene *Complex Carbohydrate Structure Database* (CCSD), die den Zugriff auf Publikationen zu komplexen Zuckerstrukturen vereinfachen sollte.<sup>30,31</sup> *Sugabase* wurde entsprechend der Nomenklatur und der hinterlegten bibliographischen Daten der CCSD aufgebaut und mit vereinheitlichten NMR-Datensätzen erweitert. Zunächst war *sugabase* noch als kostenpflichtige Version auf Diskette erhältlich, inzwischen ist *sugabase online* frei verfügbar.<sup>52</sup> Essentiell für diese Datenbank ist das Konzept der *structural reporter groups*, welches auf dem 27th International Congress of Pure and Applied Chemistry in Helsinki, 1979 vorgestellt und anschließend weiter ausgebaut wurde.<sup>48-50</sup> Signale, die separiert vom *bulk*-Bereich der meisten Protonensignale auftreten und charakteristisch für bestimmte Verknüpfungen und Zuckerreste sind, wurden nach der Publikation für *sugabase* in standardisierter tabellarischer Form erfasst. Bis zum Jahr 1997 wurden 1758 Datensätze hinterlegt, von denen insgesamt etwa 1400 <sup>1</sup>H-NMR-Spektren entnommen sind. Seitdem erfolgten keine neuen Einträge, die letzte Aktualisierung wurde auf das Jahr 2000 datiert.

### Übersicht über die in dieser Arbeit verwendeten Daten aus sugabase

- In *sugabase* wurden innerhalb der Jahre 1989-1998 Daten aus Publikationen, die zwischen 1981 und 1997 veröffentlicht wurden, eingetragen.
- Insgesamt wurden 1758 Datensätze in *sugabase* notiert, davon beinhalten 1411
   Einträge Signale von <sup>1</sup>H-NMR-Daten und 557 Einträge mit <sup>13</sup>C-Daten. 210 Datensätze davon beinhalten sowohl <sup>1</sup>H- als auch <sup>13</sup>C-Signale.
- Für diese Arbeit wurden 1403 Einträge importiert. Diese beinhalten 156 verschieden notierte Monosaccharideinheiten.
- Davon wurden f
  ür die Spektrenberechnung Eintr
  äge ber
  ücksichtigt, die nichtreduzierte Derivate von Glucose, Mannose, Galactose, Fucose und Neuramins
  äure beinhalten (insg. 17 Derivate)
- Es wurden 523 Spektren berechnet. Dies entspricht 37% der gesamten Datenbank-Einträge.
- Diese 523 Spektren bestehen aus insgesamt 12595 Signalen.

Ende 2010 / Anfang 2011 war die Webseite der Datenbank sugabase vorübergehend online nicht verfügbar. Nachdem sie anschließend bis Mitte Oktober 2011 gemäß der Version von 2000 abgerufen werden konnte, ist sie inzwischen erneut nicht mehr zugänglich. Um nicht auf die permanente Verfügbarkeit von sugabase angewiesen zu sein und außerdem ein automatisches Verwenden aller Daten zu ermöglichen, wurden die Daten offline gesichert. Diese offline-Daten konnten für das Importieren und Berechnen von Spektren verwendet werden. Dazu wurden zunächst die textbasierten Eintragungen mithilfe von awk, einer Linuxgestützten Programmiersprache zur Interpretation von Textdateien, eingelesen. Durch awk ist es möglich, jedes Leerzeichen innerhalb einer Zeile einzeln auszulesen und bestimmten Spalten zuzuordnen. Anschließend wurden zwischen verschiedenen Einträgen Semikola als Trennzeichen eingefügt und alle Daten als Textdatei gespeichert (B. Meyer, unveröffentlichte Ergebnisse). Im Folgenden wurden die Dateneinträge mit matlab importiert und in eine automatisch interpretierbare Formatierung umgeschrieben (s. Abbildung 7). Im Rahmen dieser Arbeit wurden für die weiteren Untersuchungen nur Spektren von Strukturen interpretiert, die sich aus folgenden fünf Zuckerresten und ihren Derivaten ergeben: Glucose, Galactose, Mannose, Fucose, N-Acetyl- bzw. N-Glycolylneuraminsäure. Außerdem wurden zum Alditol reduzierte Derivate nicht verwendet. Diese Auswahl wurde getroffen, da der Abgleich mit experimentellen Daten, die anhand von Strukturen der klassischen N-Typ-Glycosylierung gewonnen wurden, im Vordergrund stand. Die Einträge, auf die die vorgenannten Vorgaben zutreffen wurden selektiert, dies entspricht 523 Einträgen und damit etwa 37 % aller <sup>1</sup>H-NMR-Datensätze aus *sugabase*.

su	ę	<i>abase</i> -Eintrag	(Ausschnitt CCSD-Nr. 42857)	)
				_

Residue	Linkage	Proton PPM	J	Hz	Note
D-GlcNAc		H-1a 5.180			
		H-1b 4.694			
		NAc 2.039			
a-L-Fucp	6	H-1(a) 4.897			
-		H-1(b) 4.889			
		H-5(a) 4.096			
		H-5(b) 4.133			
		CH3(a) 1.210			
		CH3(b) 1.222			
	Į				
tierung nach I	nterpretation mit <i>awk</i>				
tierung nach In Record: ;1330	nterpretation mit <i>awk</i>				
tierung nach In Record: ;1330 D-GlcNAc	nterpretation mit <i>awk</i>	;H-1a ;	5.180 ;	;;;	
tierung nach In Record: ;1330 D-GlcNAc D-GlcNAc	nterpretation mit <i>awk</i> ; CCSD: ; 42857 ; ; ; ; ;	;H-la ; ;H-lb ;	5.180 ; 4.694 ;	;;;;	
tierung nach In Record: ;1330 D-GlcNAc D-GlcNAc D-GlcNAc	nterpretation mit <i>awk</i>	;H-la ; ;H-lb ; ;NAc ;	5.180 ; 4.694 ; 2.039 ;	;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;	
tierung nach In Record: ;1330 D-GlcNAc D-GlcNAc D-GlcNAc a-L-Fucp	nterpretation mit <i>awk</i>	;H-la ; ;H-lb ; ;NAc ; ;H-l(a) ;	5.180 ; 4.694 ; 2.039 ; 4.897 ;	;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;	
tierung nach In Record: ;1330 D-GlcNAc D-GlcNAc D-GlcNAc a-L-Fucp a-L-Fucp	nterpretation mit <i>awk</i> ; CCSD: ; 42857 ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;	;H-la ; ;H-lb ; ;NAc ; ;H-l(a) ; ;H-l(b) ;	5.180 ; 4.694 ; 2.039 ; 4.897 ; 4.889 ;	;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;	
tierung nach In Record: ;1330 D-GlcNAc D-GlcNAc D-GlcNAc a-L-Fucp a-L-Fucp a-L-Fucp	<pre>terpretation mit awk ccsd: ; 42857 ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;</pre>	;H-la ; ;H-lb ; ;NAc ; ;H-l(a) ; ;H-1(b) ; ;H-5(a) ;	5.180 ; 4.694 ; 2.039 ; 4.897 ; 4.889 ; 4.096 ;	;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;	
tierung nach In Record: ;1330 D-GlcNAc D-GlcNAc a-L-Fucp a-L-Fucp a-L-Fucp a-L-Fucp a-L-Fucp	<pre>terpretation mit awk ccsp: ; 42857 ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;</pre>	;H-la ; ;H-lb ; ;NAc ; ;H-l(a) ; ;H-1(b) ; ;H-5(a) ; ;H-5(b) ;	5.180 ; 4.694 ; 2.039 ; 4.897 ; 4.889 ; 4.096 ; 4.133 ;	;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;	
tierung nach In Record: ;1330 D-GlcNAc D-GlcNAc D-GlcNAc a-L-Fucp a-L-Fucp a-L-Fucp a-L-Fucp a-L-Fucp a-L-Fucp	<pre>terpretation mit awk ccsd: ; 42857 ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;</pre>	;H-1a ; ;H-1b ; ;NAc ; ;H-1(a) ; ;H-1(b) ; ;H-5(a) ; ;H-5(b) ; ;CH3(a) ;	5.180 ; 4.694 ; 2.039 ; 4.897 ; 4.889 ; 4.096 ; 4.133 ; 1.210 ;	;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;	

Formatierung nach Bearbeitung mit matlab

'CCSD'	'Residue'	'Linkage'	'Proton'	'ppm'	'J1 (Hz)'	'J	2 (Hz)'	'J	3 (Hz)"
'42857'	'D-GlcNAc'	1 - 1	'H-la'	'5.180'	[ 3.2000]	[	0]	[	0]
'42857'	'D-GlcNAc'	· _ ·	'H-1b'	'4.694'	[ 8]	[	0]	[	0]
'42857'	'D-GlcNAc'	'_'	'NAc'	'2.039'	[ 0]	[	0]	[	0]
'42857'	'a-L-Fucp'	'6'	'H-1(a)'	'4.897'	[ 3.7000]	[	0]	[	0]
'42857'	'a-L-Fucp'	'6'	'H-1(b)'	'4.889'	[ 3.7000]	[	0]	[	0]
'42857'	'a-L-Fucp'	'6'	'H-5(a)'	'4.096'	[ 6.7000]	[	6.7000]	[	6.7000]
'42857'	'a-L-Fucp'	'6'	'H-5(b)'	'4.133'	[ 6.7000]	[	6.7000]	[	6.7000]
'42857'	'a-L-Fucp'	'6'	'CH3(a)'	'1.210'	[ 6.7000]	[	0]	[	0]
'42857'	'a-L-Fucp'	'6'	'CH3(b)'	'1.222'	[ 6.7000]	[	0]	[	0]

Abbildung 7: Ausschnitt des *sugabase*-Eintrags der CCSD-Nr. 42857 (Decasaccharid 2) und fortlaufende Prozessierung. Oben: Ausschnitt des Originaleintrags in *sugabase*. Mitte: Nach Interpretation mit *awk* werden die einzelnen Inhalte durch Verwendung des Semikola getrennt, jede Spalte enthält hier genau ein Zeichen. Unten: Texteintrag nach Bearbeitung mit *matlab*, jede Spalte enthält einen Informationseintrag (CCSD, *Residue, Linkage* usw.). Zu diesem Zeitpunkt wurden die Kopplungskonstanten bereits zugeordnet. J1, J2 und J3 beziehen sich auf drei unterschiedliche Kopplungskonstanten, die den Wert 0 erhalten, wenn sie nicht benötigt werden. Die aus der Einbeziehung der Kopplungskonstanten Übersichtlichkeit nicht dargestellt.

Anschließend wurden in Abhängigkeit von Zuckerrest und anomerer Konfiguration Kopplungskonstanten zugeordnet, die dann für die Berechnung der Signale mit den



entsprechenden Multiplizitäten verwendet wurden. Die Berechnung ist in Abbildung 8 genauer beschrieben.

Abbildung 8: Berechnung der Multiplizitäten. Die chemische Verschiebung, die anhand des *sugabase*-Eintrags vorgegeben ist, entspricht dem oben genannten  $\delta$ . Jedem Signal werden drei Kopplungskonstanten J1, J2 und J3 zugeordnet. Jedem Maximum des Multipletts kann durch Anwendung der Matrix *M* auf die drei Kopplungskonstanten über einen Schleifenbefehl ein neues  $\delta$  zugeordnet werden. Das Erste ist beispielhaft durch Pfeile dargestellt und resultiert daraus, dass jede der drei Kopplungskonstanten mit +1/2 eingeht. Daraus ergibt sich automatisch die korrekte Intensitätsverteilung. Ist beispielsweise J1=J2=J3≠0, so ergibt sich ein Quartett mit der zu erwartenden Intensitätsverteilung von 1:3:3:1. Entspricht ein Signal einem Singulett, so wurden diesem die Kopplungskonstanten J1=J2=J3=0 zugeordnet, alle Linien fallen also beim ursprünglichen  $\delta$  zusammen welches gegenüber dem Quartett dann eine relative Intensität von 8 aufweist, sodass sich bei gleicher Protonenzahl ein identisches Integral ergibt. Aus einem Spaltenvektor der Länge n, der die ursprünglichen chemischen Verschiebungen aller Signale enthält, wird so eine 8 x n-Matrix berechnet, deren Spaltenvektoren jeweils einer Linie und deren Zeilen jeweils dem Signal einer Protonenspezies entsprechen.

Weiterhin wurden die Intensitäten der Protonen, deren chemische Verschiebung abhängig von der Konfiguration am reduzierenden Ende ist, entsprechend einberechnet. Dazu wurde ein Verhältnis der  $\alpha$ - zu  $\beta$ -Konfiguration von 0.6 zu 0.4, entsprechend dem Integralverhältnis im experimentellen Spektrum des Decasaccharids (Struktur 1), angenommen. Die Bevorzugung

der  $\alpha$ -Konfiguration von Oligosacchariden mit *N*-Acetylglucosamin am reduzierenden Ende entspricht den Literaturangaben.<sup>50,87,88</sup>

Ein Abgleich der berechneten Spektren mit experimentellen Daten vereinfacht nicht nur die manuelle Interpretation, sondern ermöglicht auch eine automatische Identifikation durch den Vergleich von Spektren, ohne dass vorab eine Interpretation nötig ist. Dazu ist eine hohe Genauigkeit der berechneten Spektren erforderlich. Die Qualität der erzeugten Spektren ist abhängig von:

- i) korrekt zugeordneten Signalen in der Originalliteratur
- ii) korrekter Übertragung in sugabase
- iii) korrekter Referenzierung unter Verwendung von Standardbedingungen (Temperatur, pH-Wert, Lösungsmittel, interner/externer Standard)
- iv) korrekter Verwendung der vorgegebenen Formatierung
- v) Qualität des Skripts zum Einlesen und Konvertieren
- vi) den verwendeten Kopplungskonstanten.

Während die Parameter i)-iv) intrinsisch durch *sugabase* vorgegeben sind, stellen die Qualität der Parameter v)-vi) einen Teil dieser Arbeit da. In Abschnitt 4.1.4 wird auf Fehler innerhalb der Datenbank und daraus resultierende Probleme eingegangen. Weitere zu beachtende Faktoren sind die für die Berechnung verwendete Linienbreite, sowie die Berücksichtigung der für die Messung der experimentellen Daten verwendeten Basisfrequenz des NMR-Spektrometers. Als Identifikationsmerkmal wurde die CCSD-Nummer gewählt. Diese wird für das Generieren von Namen für Variablen verwendet, denen dann in einem automatischen Schritt das zugehörige Spektrum zugeordnet wird. *Sugabase* oder GlycomeDB können direkt nach dieser CCSD-Nummer durchsucht werden. Die Kopplungskonstanten wurden anhand experimenteller und in der Literatur genannter Daten ausgewählt.<sup>89</sup> Dabei kann nur ein Standardwert verwendet werden, von dem die tatsächliche Kopplungskonstante in Abhängigkeit von Verknüpfung und Substituenten abweichen kann.

### 4.1.2 Vergleich von berechneten und experimentellen Spektren

Beispielhaft ist in Abbildung 9 das berechnete Spektrum für das Decasaccharid (1) mit der CCSD-Nummer 42857 gezeigt. Spur **a** zeigt die Auftragung der chemischen Verschiebungen, die im Datenbank-Eintrag aufgelistet sind. Die Intensitätsinformation ist zunächst nur indirekt über die Benennung der Protonen gegeben. Spur **b** zeigt das berechnete Spektrum im Vergleich zu einem akquirierten Spektrum des Decasaccharids (Abbildung 9**c**). Ein Maß für

die Ähnlichkeit zweier Variablen, in diesem Fall der Spektren **a** und **b**, ist der Pearson'sche Korrelationskoeffizient (s. Abschnitt 2.1). Aus diesem ergibt sich eine 98% ige Übereinstimmung von gerechnetem gegenüber gemessenem Spektrum des Decasaccharids im Bereich der *structural reporter group*-Signale.



Abbildung 9: Berechnung eines <sup>1</sup>H-NMR-Spektrums am Beispiel des Decasaccharids (1). <u>a</u>: Auftragung der chemischen Verschiebung entsprechend des unveränderten *sugabase*-Eintrages ohne Intensitätsinformation, die Region der Methylprotonen ist auf 1/3 skaliert. <u>b</u>: berechnetes Spektrum des Decasaccharids (berechnet mit einem Faktor  $\alpha_L$  für die Linienbreite von 1.5). <u>c</u>: experimentelles Spektrum des Decasaccharids, das HDO-Signal wurde für die Darstellung entfernt. Für die Berechnung von b wurden die entsprechenden Kopplungskonstanten zugeordnet und einberechnet. Die verminderten Intensitäten der Protonen, deren chemische Verschiebung abhängig von der Konfiguration des reduzierenden Endes ist, wurden ebenso wie die Verteilung der Intensität durch Aufspaltung zu definierten Multiplizitäten berücksichtigt. Im Bereich der *structural reporter group*-Signale weisen Spektren b und c eine Ähnlichkeit von 98 %, betrachtet über die Korrelation beider Spektren, auf.

Das im Rahmen dieser Arbeit entwickelte Skript kann zur Berechnung von Spektren unter Vorgabe von Basisfrequenz (z.B. 500 MHz oder 700 MHz), Anzahl an Datenpunkten, minimaler und maximaler chemischer Verschiebung, verwendet werden. Grundsätzlich ist das Skript dazu ausgelegt, alle importierten Datensätze innerhalb eines Durchgangs in Spektren umzurechnen. Die verwendeten Kopplungskonstanten werden anhand einer externen Tabelle zugeordnet. Die Auswahl der Spektren, die erzeugt werden sollen, kann durch Vorgabe der Standardreste und Erweiterung der Tabelle angepasst werden. Zudem wurde das Skript mit einer *user*-Abfrage versehen, die es ermöglicht, einzelne ausgewählte Spektren anhand der CCSD-Nummer zu erzeugen. Dieses ist insbesondere dann sinnvoll, wenn ein konkreter Strukturvorschlag besteht, der manuell abgeglichen werden soll. Die Berechnung der 523
Spektren mit 32k Datenpunkten dauerte insgesamt etwa 130 min (Prozessor: Intel Core i5, 2.66 GHz, 8 GB RAM). Dementsprechend wird ein einzelnes Spektrum im Schnitt innerhalb von ca. 15 s erzeugt. Durch die Auswahl einzelner Einträge kann also schneller auf ausgesuchte Daten zugegriffen werden. Um auch bei der Berechnung der Spektren den internen Speicher nicht zu überlasten, werden die Datensätze durch Verwendung von Schleifen in Datenpaketen von bis zu 140 Spektren unterteilt, in denen sie gesichert und gelöscht werden, bevor die Bearbeitung des nächsten Pakets erfolgt. Aus diesen Paketen können die Daten wiederum durch eine Schleifenabfrage für die weitere Bearbeitung in *matlab* eingelesen werden.

## 4.1.3 Automatische Identifikation von Spektren

Eine automatische Interpretation von Spektren bietet den Vorteil, nicht auf eine manuelle Auswertung und die manuellen Suche innerhalb der Datenbank angewiesen zu sein. Ungeachtet dessen ist eine kritische Begutachtung nicht nur während der Erprobungsphase notwendig, um die automatische Zuordnung zu bestätigen oder zu verwerfen. Für diese Begutachtung ist der Vergleich von Spektren gegenüber der Verwendung von Signallisten deutlich vereinfacht, da für die Erstellung der Signallisten die Interpretation der Spektren erforderlich ist. Zum einen ist es möglich, Unterschiede von experimentellen und ermittelten Spektren durch Differenzspektren zu visualisieren. Zum anderen entfallen als mögliche Fehlerquellen beispielsweise fehlerhafte Zuordnungen von Signalen bei Überlagerungen und die Vernachlässigung der Intensität von Signalen. Die Intensität ist in der *sugabase*-Peakliste nicht direkt verschlüsselt. Die Suche erfolgt anhand der zuvor erzeugten und in Datenpakete unterteilten Spektren. Diese bilden eine Matrix **K**, aus der die Spektren ausgewählt werden sollen, die die höchste Übereinstimmung mit einem zu untersuchenden experimentellem Spektrum U aufweisen. Für jedes Spektrum der Matrix **K** wird ein Faktor *x* bestimmt, mit dem es zum Spektrum U beiträgt.

$$K \cdot x = U$$
 Gleichung 4  
min<sub>x</sub>  $\|U - K \cdot x\|^2$  Gleichung 5

#### x – Faktor, K – zu überprüfende Datensätze, U – zu identifizierendes Spektrum

Dieser Faktor kann mittels der mathematischen Methode der kleinsten Fehlerquadrate zur Lösung linearer Systeme bestimmt werden (s. Gleichung 5). Dazu wird ein in *matlab* implementierter Algorithmus eingesetzt. Alle Spektren werden dazu normiert, sodass bei perfekter Übereinstimmung eines einzigen Spektrums der Faktor x eins betragen sollte.

die Identifikation anhand der relevanten structural reporter group-Signale Um durchzuführen, wird das Spektrum U entsprechend beschnitten. Es wird auf den Bereich von  $\delta$ =1.18 bis 6.00 ppm eingegrenzt. Weiterhin werden die Intensitäten in den Bereichen des HDO-Signals (4.72-4.76 ppm), der bulk-Protonen (3.4-3.9 ppm) und prominenter Verunreinigungen (1.28-1.36: Lactat, 1.88-1.93: Acetat-Spezies) gleich null gesetzt. Analog erfolgt die Bearbeitung der berechneten Spektren k, deren Signale davon im Wesentlichen nicht betroffen sind. Für jedes berechnete Spektrum wird der zuvor beschriebene Faktor bestimmt, mit dem es in das zu untersuchende Spektrum U eingeht. Anhand eines frei definierbaren *cut offs* werden Spektren aussortiert, deren Beteiligungsfaktor am Untersuchungsspektrum U gleich dem cut off oder unterhalb dieses Wertes liegt. Die Anzahl an Spektren innerhalb K reduziert sich also, sodass sich für die verbleibenden Spektren ein neuer Faktor ergeben kann. Durch Programmierung einer Schleife wird der Faktor für die verbleibenden Spektren solange neu bestimmt, bis die Anzahl der Spektren in K konstant ist. Anschließend erfolgt die Erweiterung um das nächste Datenpaket. Das Ergebnis wird ausgegeben, wenn alle 523 Spektren durchsucht wurden und sich K nur noch aus Spektren, deren Beteiligungsfaktoren am Untersuchungsspektrum U oberhalb des cut offs liegen, zusammensetzt.

Die Anwendung wurde anhand eines Decasaccharids vom komplexen Typ überprüft. Dazu wurde ein experimentelles Spektrum des Decasaccharids mit 523 berechneten Spektren verglichen, die in ihrer Linienbreite, der Basisfrequenz, sowie der Anzahl an Datenpunkten dem experimentellen Spektrum angepasst wurden (700 MHz, Faktor für Linienbreite  $\alpha_L$  1.5, 32k Datenpunkte für spektrale Breite von 10 ppm).

Das Spektrum des Decasaccharids wird bei einem *cut off* für x von 0.2 eindeutig der korrekten CCSD-Nummer zugeordnet. Diese kann in eine Strukturinformation umgewandelt werden. Wird der *cut off* bei 0 gesetzt, wird die korrekte Zuordnung mit einer Präferenz von 21:1 gegenüber dem nächsthöheren Hit getroffen. Über die CCSD Nummer kann dann zum einen der eingelesene und modifizierte Datensatz der gefundenen Spektren ausgelesen, zum anderen über *sugabase* der originale Datensatz aufgerufen werden.

Auch hier ist die Qualität der hinterlegten Daten essentiell. Das Nonasaccharid **1** beispielsweise wird nur dann eindeutig gefunden, wenn die chemischen Verschiebungen des gesamten *sugabase*-Eintrags eine Verschiebung von 0.003 ppm erfahren. Diese Verschiebung ist die mittlere Abweichung aller im Rahmen dieser Arbeit gemessenen chemischen Verschiebungen gegenüber den *sugabase*-Angaben. Die Messungen wurden mit

verschiedenen Proben des Nonasaccharids **1** wiederholt, um hier eine zuverlässige Aussage über die chemischen Verschiebungen treffen zu können (s. a. Abschnitt 4.1.4).

Hit(s):	

t(S):							
'No.'	'CCSD' 'x,noi	cm (co: 0.2)'					
[501]	'42857' [	1]					
'CCSD'	'Residue'	'Linkage'	'Proton'	'ppm'	'J1 (Hz)'	'J2 (Hz)'	'J3 (Hz)'
'42857'	'D-GlcNAc'	'-'	'H-1a'	'5.180'	[ 3.2000]	[ 0]	[ 0]
'42857'	'D-GlcNAc'	'_'	'H-1b'	'4.694'	[ 8]	[ 0]	[ 0]
'42857'	'D-GlcNAc'	'_'	'NAc'	'2.039'	[ 0]	[ 0]	[ 0]
'42857'	'a-L-Fucp'	'6'	'H-1(a)'	'4.897'	[ 3.7000]	[ 0]	[ 0]
'42857'	'a-L-Fucp'	'6'	'H-1(b)'	'4.889'	[ 3.7000]	[ 0]	[ 0]
'42857'	'a-L-Fucp'	'6'	'H-5(a)'	'4.096'	[ 6.7000]	[ 6.7000]	[ 6.7000]
'42857'	'a-L-Fucp'	'6'	'H-5(b)'	'4.133'	[ 6.7000]	[ 6.7000]	[ 6.7000]
'42857'	'a-L-Fucp'	'6'	'CH3(a)'	'1.210'	[ 6.7000]	[ 0]	[ 0]
'42857'	'a-L-Fucp'	'6'	'CH3(b)'	'1.222'	[ 6.7000]	[ 0]	[ 0]
'42857'	'b-D-GlcpNAc'	'4'	'H-1(a)'	'4.664'	[ 8]	[ 0]	[ 0]
'42857'	'b-D-GlcpNAc'	'4'	'H-1(b)'	'4.668'	[ 8]	[ 0]	[ 0]
'42857'	'b-D-GlcpNAc'	'4'	'NAc(a)'	'2.096'	[ 0]	[ 0]	[ 0]
'42857'	'b-D-GlcpNAc'	'4'	'NAc(b)'	'2.094'	[ 0]	[ 0]	[ 0]
'42857'	'b-D-Manp'	'4,4'	'H-2'	'4.250'	[ 3.3000]	[ 0.9000]	[ 0]
'42857'	'a-D-Manp'	'6,4,4'	'H-1'	'4.925'	[ 1.3000]	[ 0]	[ 0]
'42857'	'a-D-Manp'	'6,4,4'	'H-2'	'4.108'	[ 3.3000]	[ 1.3000]	[ 0]
'42857'	'b-D-GlcpNAc'	'2,6,4,4'	'H-1'	'4.582'	[ 8]	[ 0]	[ 0]
'42857'	'b-D-GlcpNAc'	'2,6,4,4'	'NAc'	'2.047'	[ 0]	[ 0]	[ 0]
'42857'	'b-D-Galp'	'4,2,6,4,4'	'H-1'	'4.473'	[ 7.8000]	[ 0]	[ 0]
'42857'	'a-D-Manp'	'3,4,4'	'H-1'	'5.119'	[ 1.3000]	[ 0]	[ 0]
'42857'	'a-D-Manp'	'3,4,4'	'H-2'	'4.189'	[ 3.3000]	[ 1.3000]	[ 0]
'42857'	'b-D-GlcpNAc'	'2,3,4,4'	'H-1'	'4.582'	[ 8]	[ 0]	[ 0]
'42857'	'b-D-GlcpNAc'	'2,3,4,4'	'NAc'	'2.051'	[ 0]	[ 0]	[ 0]
<b>'</b> 42857 <b>'</b>	'b-D-Galp'	'4,2,3,4,4'	'H-1'	'4.467'	[ 7.8000]	[ 0]	[ 0]

>>

Abbildung 10: *matlab command window* – Ausgabe des Suchergebnisses. Zunächst wird eine Tabelle ausgegeben, die die interne Nummer (No.), die CCSD-Nummer, sowie den Faktor x,norm enthält. Faktor x wird normiert als prozentualer Anteil an allen Hits für die gilt x > cut off (co). In der Kopfzeile wird zusätzlich benannt, mit welchem *cut off* die Suche durchgeführt wurde. In einer zweiten Tabelle werden die Daten der gefundenen CCSD-Nummer(n) aufgeführt. Die CCSD-Nummer kann dazu verwendet werden, den Originaleintrag bei *sugabase* zu überprüfen und ist damit strukturassoziiert.

Die automatische Erkennung von Spektren, die hier noch unter grundlegenden Aspekten getestet wurde, ist vielversprechend, sodass eine ausgeweitete Untersuchung und Verfeinerung der Suchalgorithmen beispielsweise durch Integration von MS Daten in die Suchfunktion sinnvoll erscheint.

## 4.1.4 Fehlerhafte Einträge in sugabase und daraus resultierende Probleme

Einige Einträge in *sugabase* sind fehlerhaft oder wegen der Aufnahmeparameter der zugrunde liegenden Spektren ungenau. Die hier vorgestellten problembehafteten Einträge lassen sich meist auf die Originalliteratur zurückführen und sind hier rein dokumentarisch aufgeführt. Ungenauigkeiten der Einträge können die automatische Spektreninterpretation beeinträchtigen. Beispielsweise wird das Spektrum des Nonasaccharids erst erkannt, wenn der *sugabase*-Eintrag um 0.003 ppm tieffeldverschoben wird.

CCSD	Bezeichnung	Beobachtung	Überprüfung	Literatur
6518	Nona 1	$\Delta \delta$ (ppm) = (-0.002) - (-0.003) ppm	Messung verschiedener Proben des Nona <b>1</b> , Vergleich zum Decasaccharid (42857)	Bendiak <i>et al</i> . <sup>90</sup>
42857	Deca 2	Vertauscht: H-1( $\alpha$ ) und H-1( $\beta$ ) von Fuc	Intensitätsverhältnis α:β ~ 0.6:0.4 deckt sich z.B. bei Fuc H-5 und Fuc H-6	Bergwerff et al. <sup>91</sup>
42859 und 42858	Dodeca <b>4</b> und Undeca <b>5</b>	keine Unterscheidung: H-1 von Gal-6/Gal-6', H-1α von Gal-7/Gal-7' fehlend (Dodeca): H-1, H-5 von Fuc	700 MHz Spektren: verbesserte Auflösung, eindeutige Unterscheidbarkeit (s. Abbildung 38)	Bergwerff <i>et</i> <i>al.</i> <sup>91</sup>
3910	-	Falscher Syntax in sugabase: H-2a statt H-2(a) von Man	H-2a: Konfiguration des Restes H-2(a): abhängig von Konfiguration des reduzierenden Endes	Van Pelt <i>et</i> <i>al.</i> <sup>92</sup>

Tabelle 1: Fehlerdokumentation von sugabase-Einträgen.

Durch falsche  $\alpha$ -/ $\beta$ -Zuordnung oder nicht korrekt verwendeten *sugabase* Syntax können die Intensitätsverhältnisse oder die Kopplungszuordnung beeinflusst werden. Gerade die beiden letzteren Probleme sind in der in *sugabase* direkt implementierten Suche allerdings irrelevant, weil dort lediglich die chemische Verschiebung abgefragt wird. Somit ergeben sich erst bei der automatischen Interpretation Probleme. Da bei einer Datenbank mit mehr als 1500 Einträgen Ungenauigkeiten zu erwarten sind, müsste hier überprüft werden, ob einige Fehler bei der Umsetzung der Daten in Spektren korrigierbar sind und ob die Suchabfragen bei der automatischen Spektreninterpretation mit entsprechender Flexibilität ausgestattet werden, um inkorrekte Referenzierungen zu kompensieren. Aufgrund der Heterogenität von Glycoproteinen werden bei Abspaltungsprozessen der Glycane komplexe Gemische erhalten. Die Aufklärung der darin enthaltenden Komponenten ist dann aufgrund von Überlagerungen erschwert oder unmöglich. Eine Erweiterung zu der bisher beschriebenen Identifikation von Spektren von reinen Komponenten stellt also die Charakterisierung von Gemischen dar. Am Beispiel einer Bibliothek von 13 Glycanen und vier zunächst unbekannten Gemischen A-D, die sich aus einer beliebigen Anzahl an Komponenten dieser Bibliothek zusammensetzten, wird im Folgenden diskutiert, inwiefern eine direkte Identifikation von Substanzen im Gemisch möglich ist. Dabei steht die Interpretation der NMR-Daten im Vordergrund. Alle Substanzen und Gemische wurden von der Firma TheraProteins bezogen. Die Bibliothek setzt sich aus zwei hochmannosidischen Strukturen und 11 sialylierten komplex-Typ-Glycanen zusammen, die sich in Anzahl ihrer Reste und Antennen, sowie im Vorkommen von repeats (\beta1,4-Gal-\beta1,3-GlcNAc) und Fucbzw. α1,3-Gal-Resten unterscheiden (s. Abkürzungsverzeichnis S. VII, Abbildung 3, Strukturen I-XIII). Alle Strukturen liegen in absoluten Mengen von 230-450 pmol als Referenzsubstanzen vor. Die Gemische setzten sich aus Substanzmengen von 200-1000 pmol zusammen. In vorangegangen Arbeiten konnte gezeigt werden, dass diese geringen Substanzmengen unter Berücksichtigung optimierter Probenpräparation und Spektrenakquisition NMR-spektroskopisch gemessen werden können (s. a. Abschnitt 2.2).<sup>53</sup> Alle Spektren weisen entsprechend des structural reporter group-Konzepts Charakteristika auf, die sie eindeutig unterscheidbar machen. Beispielsweise zeigt Struktur X, welche als einzige Struktur der Bibliothek einen zusätzlichen α1,3-Gal-Rest aufweist, im NMR-Spektrum ein charakteristisches Signal bei  $\delta$ =5.146 ppm (Gal-7' H-1) auf, welches ihr Alleinstellungsmerkmal ist (vgl. Abbildung 11). Zudem kann über die chemische Verschiebung der anomeren Protonen der Gal-6/6' die Verknüpfung des α1,3-Gal an der Man-6-Antenne identifiziert werden.

Ausgehend von der An- bzw. Abwesenheit einzelner Signale kann eine eindeutige Aussage dazu getroffen werden, ob eine Komponente im Gemisch vorliegt, ohne dass eine Trennung erforderlich ist. Dieses ist allerdings nicht mehr möglich, wenn die Bibliothek auf mehrere Hundert Strukturen ausgeweitet wird, weil dann nur noch die Kombination aus vielen charakteristischen *structural reporter group*-Signalen, die in Spektren der reinen Verbindungen auftreten, zu einer eindeutigen Zuordnung führt.



Abbildung 11: Vergleich der Spektren von Gemisch C und Komponente X. Während die Charakterisierung der Komponente X im Spektrum von Gemisch C nicht eindeutig vorgenommen werden kann, da beispielsweise das anomere Signal der Gal-6' überlagert ist, ist diese Komponente eindeutig als Bestandteil von Gemisch C identifizierbar. Beispiel hierfür ist das H-1 der Gal-7', welche eine chemische Verschiebung bei  $\delta$ =5.146 ppm aufweist.

Durch manuelle Interpretation der NMR-Daten und der kombinierten Analyse von MS-Daten der Gemische konnten die in Tabelle 2 notierten Zusammensetzungen ermittelt werden. Beispielhaft ist in Abbildung 12 zudem der Vergleich des gemessenen Spektrums von Gemisch **A** mit dem Summenspektrum der einzelnen Komponenten dargestellt. Dabei ist erkennbar, dass aufgrund des starken Rauschens die Signale im Erscheinungsbild variieren, obwohl sie mit gleicher Prozessierung bearbeitet wurden. Das Differenzspektrum zeigt, dass trotzdem eine relativ gute Übereinstimmung der Daten gegeben ist, da die Abweichungen hauptsächlich im *bulk*-Bereich auftreten. MS von Gemischen ermöglicht als zusätzliche Informationsquelle einen schnellen Zugriff auf die Anzahl massenunterschiedlicher Strukturen im Gemisch. NMR-Daten erlaubten außerdem eine Quantifizierung der Substanzmengen der einzelnen Komponenten, sowie eine Verifikation der gegebenen 2,3-Verknüpfung der terminalen Neuraminsäurereste anhand der chemischen Verschiebungen der NeuAc H-3a/e Protonen nachgewiesen werden. Außerdem konnten die Verknüpfungen der Neuraminsäure am 3-Arm und der  $\alpha$ 1,3-Gal am 6-Arm bestimmt werden. Probe VI konnte zudem als Gemisch zweier triantennärer Strukturen mit dem Auftreten des *repeats* an unterschiedlichen Antennen nachgewiesen werden.



Abbildung 12: Gemisch A – Vergleich des experimentellen Spektrums mit dem Summenspektrum der Einzelkomponenten (I, III, V, IX). Unten ist das Differenzspektrum abgebildet, welches im Bereich der *structural reporter groups* gute Übereinstimmung der beiden Spektren (oben) zeigt. Abweichungen sind im Bereich des HDO-Signals, bei 4.1 ppm (Lactat als Verunreinigung) und im *bulk* Bereich erkennbar.

Allerdings konnten nicht von allen 13 Substanzen der Bibliothek die Isomerenunterscheidung vorgenommen werden, was neben starken Überlagerungen in den Spektren auch daran lag, dass nicht alle potentiellen Strukturen in *sugabase* verfügbar waren. In diesem Fall sind die möglichen Verknüpfungen durch eine geschweifte Klammer angedeutet. Um eine Zuordnung von neuen Strukturen anhand der 1D-<sup>1</sup>H-NMR-Spektren machen zu können, müssen die Daten ähnlicher Strukturen zur Verfügung stehen. Dann können durch den Vergleich von identischen und abweichenden Signalen Zuordnungen zuverlässig getroffen werden. Stehen solche Daten nicht zur Verfügung, müssen die Zuordnungen anhand von 2D-Spektren durchgeführt werden, für die allerdings im Gegensatz zu 1D-Spektren mehr Substanz und/oder Messzeit nötig ist.

2: Identifikation der Komponenten in vier Gemischen (A-D) anhand einer Biblioth						
rukturen (s. Abbildung 3, S. VII)						
Gemisch	Komponenten	Anteil [%]	Substanzmenge im			
		manuelle	Gemisch (manuell)			
		Bestimmung	(pmol)			
А	I	21 %	420			
	III	16 %	310			
	V	46 %	910			

Tabelle 2: Ide Bibliothek von 13 Glycanstruktu

		manuelle	Gemisch (manuell)
		Bestimmung	(pmol)
А	Ι	21 %	420
	ш	16 %	310
	V	46 %	910
	IX	17 %	340
В	I	31 %	340
	VI	19 %	210
	VIII	49 %	530
С	П	27 %	590
	VI	29 %	650
	Х	23 %	510
	XI	21 %	450
D	Ι	15 %	420
	IV	26 %	730
	VII	22 %	620
	VIII	26 %	730
	XII	10 %	270

Die manuelle Bestimmung der Anteile der Komponenten im Gemisch ist insofern fehlerbehaftet, da beispielsweise gerade im bulk Bereich Unterschiede nur schwer wahrnehmbar sind. Eine zuverlässige automatische Erkennung und Quantifizierung der Spektren im Gemisch würde hier eine Erleichterung und eine Verbesserung der Genauigkeit mit sich bringen.

Die in Abschnitt 4.1.3 vorgestellte Methode der kleinsten Fehlerquadrate zur Identifikation von Komponenten wurde hier angewendet, um zu überprüfen, ob auch mehrere Komponenten, aus denen sich ein Summenspektrum mit überlappenden Signalen ergibt, identifiziert werden können.

Wird das Summenspektrum aus den Einzelspektren der in Tabelle 2 aufgelisteten Komponenten mit den ermittelten prozentualen Anteilen eines Gemisches zusammengesetzt, so werden die Komponenten eindeutig erkannt. Außerdem kann auch der prozentuale Anteil anhand des Faktors x abgelesen werden. Weiterhin wurde eine einzelne Komponente, die zu mindestens 5 % vertreten ist, aus dem Summenspektren von 4 verschiedenen Substanzen in variierender Zusammensetzung zu 100 % eindeutig erkannt, solange das Summenspektrum aus den zuvor gemessenen Einzelspektren zusammengesetzt ist.

Wird der gleiche Algorithmus angewendet, um die Komponenten im gemessenen Summenspektrum zu ermitteln, so erfolgt keine eindeutige und zuverlässige Zuordnung. Bildung von Differenzspektren zeigt hier, dass eine gute Übereinstimmung über den Gesamtbereich erreicht wird, aber einzelne essentielle Signale vernachlässigt werden. Vermutlich resultiert diese Problematik aus geringfügigen experimentellen Unterschieden des gemessenen Summenspektrums im Gegensatz zu den Einzelspektren, die umso stärker hervortreten, je geringer das Signal-Rausch-Verhältnis (S/N) ist. Aufgrund der geringen Substanzmengen weisen alle Spektren ein geringes S/N auf. Dieses liegt in der Größenordnung von 4-9 für das Man-4 H-1 (komplex-Typ-Glycane, Prozessierung: Multiplikation des FIDs mit Exponentialfunktion (EM) mit einem *line broadening factor* (*lb*) von 1 vor der Fouriertransformation). Beispielsweise stimmt nicht jedes Signal genau überein, weil es teilweise durch das Rauschen moduliert ist.

Denkbar ist eine Optimierung durch Erweiterung des Skriptes durch Zyklen, in denen einzelne Bereiche des Spektrums nach charakteristischen Signalen der Komponenten untersucht werden und daraus Restriktionen für die Menge der zu untersuchenden Komponenten abzuleiten. Die Anwendung von neuronalen Netzwerken könnte die Interpretation von Spektren mit geringem S/N sowie den Fehler durch rauschmodulierte Signale verbessern. Dazu beschrieben Radomski *et al.* und Meyer *et al.*, dass neuronale Netzwerke dazu trainiert werden können, Spektren zu erkennen, die genau die zuvor beschriebenen Probleme des geringen S/N und der experimentellen Ungenauigkeit aufweisen.<sup>70,93</sup> In einem Datensatz von mehr als 100 NMR-Spektren wurden 90 % der Daten bei einem S/N von nur 1.25 erkannt. Außerdem konnten neuronale Netzwerke imperfekte Spektren erkennen, nachdem sie mit Daten trainiert wurden, die durch das Einbringen von geringfügigen Abweichungen den intrinsisch experimentellen Fehler beinhalten. Dieser resultiert einfach daraus, dass die Spektren zweier Proben einer Substanz niemals perfekt identisch sein werden.

Grundsätzlich ist es also möglich, auch aus NMR-Spektren von Gemischen Informationen über die einzelnen Bestandteile zu erhalten, wie es an der Bibliothek der 13 Substanzen gezeigt wurde.

# 4.3 <u>3D-Kreuzkorrelation (3DCC) von LC, MS und NMR – Entwicklung einer</u> neuen Methode zur Charakterisierung von Glycanen

Die in diesem Abschnitt 4.3, sowie in Abschnitt 4.4 diskutierten Ergebnisse sind in Zusammenarbeit mit Henning N. Behnken entstanden. Einige Ergebnisse der Abschnitte 4.3.3, 4.4.4 und 4.4.5.1 stammen außerdem aus den von mir /uns betreuten Bachelorarbeiten von Tim Nagel und Raffael Jirmann.<sup>94,95</sup>

In den vorherigen Abschnitten wurde gezeigt, dass eine automatische Identifikation von Glycanen anhand von NMR-Daten dann zuverlässig möglich ist, wenn Spektren reiner Substanzen vorliegen. Auch aus Spektren von Substanzgemischen können noch einzelne Informationen gewonnen werden. Die Auswertung kann aber auch durch zu starke Überlagerungen verhindert werden. Gerade bei strukturell ähnlichen Substanzen wie Glycanen kann die chromatographische Trennung von komplexen Gemischen jedoch schwierig sein.

Das Konzept der 3D-Kreuzkorrelation (*3D cross correlation* – 3DCC) ist eine neu entwickelte Methode zur Verknüpfung von LC-, MS- und NMR-Daten, die eine schnelle und zuverlässige Charakterisierung von Glycanen ermöglicht. 3DCC ist dabei nicht auf Glycane als Substanzklasse limitiert, sondern ist für alle Problemstellungen mit unvollständiger Trennung einsetzbar. Die Entwicklung und Anwendung von 3DCC wird in dieser Arbeit am Beispiel von Oligosacchariden diskutiert, die durch ihre komplexen Strukturen eine analytische Herausforderung darstellen. Die Kombination von MS und NMR-Spektroskopie ist eine konsequente Analytikstrategie, die die Vorteile beider Methoden vereint.

MS- und NMR-Daten werden anhand einer gemeinsamen Zeitabhängigkeit korreliert, die aus der LC entsteht. Massenspektrometrie und NMR-Spektroskopie sind zwei in der Analytik von Glycanen meist unabhängig voneinander eingesetzte Methoden zur Charakterisierung der Saccharidstrukturen. MS- oder MS/MS Daten bieten schnellen Zugriff auf die Zusammensetzung der Strukturen und Angaben über die Koexistenz mehrerer Strukturen. Eine Unterscheidung isomerer Monosaccharidbausteine, sowie der anomeren Konfiguration ist jedoch im Regelfall nicht zugänglich und auch die Verknüpfungsinformation ist nicht immer unzweifelhaft erhältlich. Im Gegensatz dazu bieten NMR-Spektren anhand der charakteristischen chemischen Verschiebung und der Kopplungskonstante schnellen Zugriff auf diese Informationen. Sobald eine Probe aber mehrere Komponenten beinhaltet, ist hier die Auswertung erschwert oder nicht mehr durchführbar. Dann müssen die Komponenten durch chromatographische Methoden getrennt werden. Gerade bei sehr ähnlichen Strukturen erfordert die chromatographische Trennung Expertise und durch die Notwendigkeit, ideale Bedingungen zu finden, Zeit und Substanz. Eine neue Methode, die also die Vorteile von Massenspektrometrie mit denen der NMR-Spektroskopie vereint und dabei keine vollständige chromatographische Trennung erfordert, kann die Analytik von Glycanstrukturen wesentlich vereinfachen.

#### 4.3.1 Schematische Darstellung der 3D-Kreuzkorrelation

Das Prinzip der 3D-Keuzkorrelation ist in Abbildung 13 anhand eines berechneten Beispiels dargestellt und kann in fünf Abschnitte unterteilt werden: A: Durchführung einer Chromatographie zur Trennung oder Antrennung der Substanzen (1. Dimension), B: Aufnahme von MS- und NMR-Daten in gleicher Zeitabhängigkeit (2. und 3. Dimension), C: mathematische Korrelation der MS- und NMR-Daten, D: Extraktion der NMR-Signale, die im Rahmen dieser Arbeit auch als Dekonvolution bezeichnet wird, und E: Charakterisierung durch Kombination von m/z mit den zugehörigen extrahierten NMR-Spektren und Retentionszeiten.

Liegt eine komplexe Mischung vor, so kann diese also zunächst durch LC partiell getrennt werden, eine vollständige Separation ist nicht notwendig. Das Chromatographieprofil kann beispielsweise durch Kopplung mit MS anhand des *total ion currents* (TIC) detektiert werden. Durch LC ergibt sich die Zeit als Dimension über die eine Korrelation vorgenommen werden kann. Die Extraktion vom TIC ergibt bei einem bestimmten m/z zeitabhängige Profile jedes einzelnen Ions (*extracted ion chromatograms*, EIC). Die EICs bilden die Zeilenvektoren einer Matrix  $\mathbf{M}_{(m,t)}$ , die in Anlehnung an die Dirac-Notation durch (M| (*bra*) beschrieben werden können. Entsprechend können in gleicher Zeitabhängigkeit entweder durch Fraktionierung oder durch *on-line*-LC-NMR <sup>1</sup>H-NMR-Spektren akquiriert werden. Von jedem NMR-Datenpunkt kann ähnlich zum EIC ein zeitabhängiges Profil über alle aufgenommenen Spektren erstellt werden. In Analogie zum EIC wurde dieses Profil als *extracted delta chromatogram* (EDC) benannt. Die EDCs bilden die Spaltenvektoren |*N*⟩ (*ket*) der Matrix  $\mathbf{N}_{(t,\delta)}$ .



Darstellung der Methode anhand eines gerechneten Beispiels. <u>A:</u> Verlauf eines LC-Laufs ohne erkennbare Trennung nach Detektion durch MS (TIC). <u>B:</u> Extraktion von EICs aus dem TIC ergibt die gezeigten Intensitätsverläufe der Komponenten A-C. Analog ergeben konsekutiv aufgenommene NMR-Spektren zeitabhängige Profile jedes einzelnen Datenpunkts (EDCs). Jedes hervorgehobene EDC (cyan: EDC<sub>A</sub>, blau: EDC<sub>B</sub>, rot: EDC<sub>C</sub>) korreliert mit genau einem EIC (gleiche Farbkodierung). <u>C:</u> Kreuzkorrelation der MS-Daten mit den NMR-Daten ergibt Matrix C<sub>(m,δ)</sub>, die die Korrelationskoeffizienten jedes EDCs mit jedem EIC beinhaltet. <u>D:</u> Durch Multiplikation der Korrelationsmatrix C<sub>(m,δ)</sub> mit dem Summenspektrum S<sub>(δ)</sub> über alle Fraktionen lassen sich reine NMR-Spektren, assoziiert mit einem definierten *m/z*, extrahieren (E<sub>(m,δ)</sub>). <u>E:</u> Aus Retentionszeit, *m/z* und NMR-Signalen ergibt sich die eindeutige Charakterisierung der Substanzen A, B und C.

Jeder Zeilenvektor  $\langle M |$  wird mit jedem Spaltenvektor  $|N \rangle$  mathematisch korreliert, es resultiert daraus die Korrelationsmatrix  $C_{(m,\delta)}$  (Gleichung 6). Bei hoher Übereinstimmung eines EDC-Verlaufs mit einem EIC-Verlauf (z.B. Abbildung 13, EDCA und EICA) ergibt sich ein Korrelationskoeffizient ~1, sind die Vektoren jedoch stochastisch unabhängig (z.B. EDC<sub>B</sub> und EDC<sub>C</sub> im Vergleich zu EIC<sub>A</sub>), so ergibt sich ein Koeffizient von ~0. Schematisch wird dieses in Abbildung 13B dargestellt, hier sind die EICs von drei Verbindungen A, B und C gezeigt, die jeweils mit den EDCs eines NMR-Signals bei  $\delta_A$ ,  $\delta_B$  bzw.  $\delta_C$  eine hohe Korrelation aufweisen, während sie mit den anderen beiden (nahezu) nicht korrelieren. Da eine anti-Korrelation nicht zu erwarten ist, werden Korrelationskoeffizienten mit negativem Wert gleich Null gesetzt. Ein zu einer definierten Molekülionenspezies gehörendes NMR-Spektrum kann extrahiert werden, indem das Summenspektrum  $S_{(\delta)}$  mit dem entsprechenden Vektor der Korrelationsmatrix  $C_{(m,\delta)}$  paarweise multipliziert wird, woraus sich eine Matrix  $\mathbf{E}_{(m,\delta)}$  ergibt. Beispielsweise errechnet sich das in Abbildung 13D gezeigte extrahierte NMR-Spektrum der Verbindung A durch Multiplikation des Summenspektrums mit dem ersten Zeilenvektor von  $C_{(m,\delta)}$ .  $C_{(m,\delta)}$  beinhaltet die Korrelationskoeffizienten jedes EDCs mit EIC<sub>A</sub>, die nur in der Region von  $\delta_A$  von 0 verschieden sind (s. a. Abbildung 14).

Nach Pearson errechnet sich der Korrelationskoeffizient zweier Vektoren über die Kovarianz *Cov* normiert über die Standardabweichungen  $\sigma$  beider Vektoren (s. a. Abschnitt 2.1). Die Elemente  $c_{m,\delta}$  der Matrix  $\mathbf{C}_{(m,\delta)}$  lassen sich also folgendermaßen berechnen:

$$c_{m,\delta} = \frac{Cov(\langle M|, |N\rangle)}{\sigma_{\langle M|} \cdot \sigma_{|N\rangle}} \qquad \text{Gleichung 6}$$

Der Vektor  $S_{(\delta)}$ , der das Summenspektrum über alle NMR-Daten beinhaltet, ergibt sich aus dem Summieren der Zeilenvektoren  $\langle N |$  von  $N_{(t,\delta)}$ :

$$S_{(\delta)} = \sum_{t} \langle N |$$
 Gleichung 7

Um die extrahierten NMR-Spektren zu erhalten, die jeweils einer definierten Molekülionenspezies zuordenbar sind  $(\mathbf{E}_{(m,\delta)})$ , erfolgt die paarweise Multiplikation der Elemente von  $\mathbf{C}_{(m,\delta)}$  mit denen von  $\mathbf{S}_{(\delta)}$  entlang der delta-Dimension.

$$e_{m,\delta} = c_{m,\delta} \cdot s_{\delta}$$
 Gleichung 8

Das Einführen einer gewichteten Extraktion (s. Abschnitt 4.4.1.1), bei der der Korrelationskoeffizient mit der 4. Potenz in die Berechnung der extrahierten NMR-Spektren eingeht resultiert in der Matrix  $E4_{(m,\delta)}$ :

$$e4_{m,\delta} = (c_{m,\delta})^4 \cdot s_\delta$$
 Gleichung 9

Im oben gezeigten Beispiel (Abbildung 13) wurden drei chromatographische Peaks, die eine starke Überlappung aufweisen anhand einer Gauß-Verteilung berechnet. Im TIC, welches hier der Summe der drei Peaks entspricht, ist keine Trennung beobachtbar. Dieses Beispiel soll verdeutlichen, dass trotz der starken Überlappung im Chromatogramm eine Extraktion von NMR-Spektren möglich ist. Die Intensität der drei NMR-Signale in den jeweiligen Spektren wurde anhand der Chromatographiepeaks berechnet.



Abbildung 14: Darstellung der Spektrenextraktion. A) Summenspektrum, B) Verlauf der Korrelationskoeffizienten, die sich aus der Korrelation jedes EDCs mit  $EIC_A$ ,  $EIC_B$  oder  $EIC_C$  berechnen lassen. C) Multiplikation jedes Datenpunktes des Summenspektrums mit jedem Datenpunkt der einzelnen Korrelationskoeffizienten ergibt die extrahierten Spektren der Verbindungen A,B und C.

Anschließend wurde die Korrelationsmatrix wie zuvor beschrieben bestimmt und mit dem Summenspektrum multipliziert (Abbildung 14). Dadurch konnten die drei gezeigten NMR-Spektren, die jeweils nur genau ein Signal beinhalten, extrahiert werden.

Die Methode ist nicht darauf beschränkt, den Pearson'schen Korrelationskoeffizienten zu verwenden, sondern kann analog mit den Rangkorrelationskoeffizienten nach Kendall  $(\mathbf{EK}_{(m,\delta)})$  oder Spearman verwendet werden  $(\mathbf{ES}_{(m,\delta)})$ .

Diese Methode ist prinzipiell für jede Substanzklasse einsetzbar und mit beliebigen weiteren analytischen Verfahren wie MS/MS und/oder Infrarot-Spektroskopie kombinierbar.

Durch 3DCC lassen sich extrahierte NMR-Spektren erhalten, die eindeutig einem definierten m/z zuordenbar sind, obwohl in jedem einzelnen aufgenommenen Spektrum alle Signale auftauchen können. Daraus lassen sich die Informationen aus der Massenspektrometrie mit denen aus der NMR-Spektroskopie verknüpfen, sodass eine eindeutige Charakterisierung vorgenommen werden kann. Die Überprüfung der Methode an der Klasse der Oligosaccharide anhand eines simulierten Chromatographielaufes (Abschnitt 4.3.3) und die Anwendung bei komplexen Gemischen (Abschnitt 4.4) werden im Folgenden diskutiert.

## 4.3.2 Vergleich der Korrelation nach Kendall, Spearman und Pearson

Bei der Korrelation der Intensitätsverläufe von NMR- und MS-Signalen müssen zwei Aspekte berücksichtigt werden. Zum einen ist die Extraktion der korrekten NMR-Signale einer Verbindung B von der Überlappung des Chromatographiepeaks von B mit dem von Verbindung A abhängig. Dieses wird in Abbildung 15 thematisiert.

Zum anderen ist gerade bei Kohlenhydraten relevant, wie gut NMR-Signale, die zu beiden Komponenten zuordenbar sind und daher im Folgenden als globale Signale bezeichnet werden, mit den EICs der jeweiligen Verbindungen korrelieren und dementsprechend korrekt extrahierbar sind (s. Abbildung 16).

Zwar liegt der Berechnung des Pearson'schen Korrelationskoeffizienten die metrische Skala und damit die umfassendere Verwendung der zur Verfügung stehenden Informationen zu Grunde, trotzdem kann im experimentellen Beispiel auch die Anwendung der Rangkorrelation nach Spearman oder Kendall eine geeignete Möglichkeit für die Datenextraktion darstellen. Deshalb wurden die Abhängigkeiten der drei Korrelationsberechnungen von Peakabstand und Intensitätsverhältnis am berechneten Beispiel überprüft. Abbildung 15 zeigt den Vergleich der Verläufe der Korrelationskoeffizienten, berechnet nach Kendall, Spearman und Pearson, in Abhängigkeit vom Abstand des Peaks B zu Peak A.



Abbildung 15: Vergleich der Korrelationskoeffizienten nach Kendall, Spearman und Pearson anhand eines berechneten Beispiels. Peak A in Ausschnitt A ist in Rot dargestellt, die Verschiebung von Peak B relativ zu Peak A ist in Grau angedeutet, schwarz markiert sind vier Beispiele, für die oben die Summensignale von A und B dargestellt sind. Ausschnitte B-D zeigen die Verläufe der Koeffizienten von B zum Summensignal (A+B) in schwarz und zu Peak A in Rot. Dabei zeigt der Rangkoeffizient nach Kendall einen nahezu linearen Verlauf, während nach Spearman die Peaks A und B auch bei waschsenden Abständen stark korrelieren und sich nach Pearson ein schmalerer Verlauf ergibt. Durch die Verwendung des Pearson'schen Korrelationskoeffizienten lassen sich Spektren von Chromatographiepeaks mit größerer Überlappung extrahieren.

Der Pearson'sche Korrelationskoeffizient fällt am schnellsten ab, sodass er am besten geeignet ist, NMR-Signale der Verbindung B auch bei großer Überlappung mit Peak A im Chromatographielauf eindeutig der richtigen Verbindung zuzuordnen. Bei der Korrelation eines NMR-Signals, welches für Verbindungen A und B bei der gleichen chemischen Verschiebung auftritt und demzufolge dem Summenpeak von A und B entspricht, nähern sich die Korrelationskoeffizienten bei großen Abständen Grenzwerten von 0.6 (Kendall, Pearson) bzw. 0.8 (Spearman) an. Damit liegen sie oberhalb des eigentlichen Anteils von Peak B am Summenpeak von 0.5, woraus sich die Problematik ergibt, dass nur innerhalb eines extrahierten NMR-Spektrums eine zuverlässige Integration erhältlich ist, während der Vergleich zu den gemessenen Spektren oder zu weiteren extrahierten Daten Abweichungen ergeben muss.

In Abbildung 16 ist aufgetragen, wie sich die Intensität eines globalen, extrahierten NMR-Signals in Abhängigkeit vom Abstand der Peaks B und A im Chromatogramm und in Abhängigkeit vom Intensitätsverhältnis der Komponenten B und A verhält. Ein NMR-Signal, welches alleinig Peak B zuordenbar ist, weist den Idealverlauf auf, bei dem die Intensität des Signals exakt das Verhältnis von Komponente B zu A widerspiegelt. Ein globales Signal, welches zu beiden Komponenten A und B gehört, wird durch die Extraktion unter Verwendung des Pearson'schen Korrelationskoeffizienten unterrepräsentiert und zwar umso stärker, je größer der Abstand der Peaks A und B und je geringer der Anteil von Komponente B im Gemisch ist. Die Intensitätsverläufe, die sich aus Extraktion unter Verwendung des Rangkorrelationskoeffizienten nach Spearman bzw. Kendall ergeben, zeigen deutlich geringere Abweichungen.

Daraus ergibt sich, dass der Pearson'sche Korrelationskoeffizient gegenüber der Rangkorrelation umso effizienter ist, je größer die Überlappung im Chromatogramm ist, weil sich dann auch der Effekt der Unterrepräsentation geringkonzentrierter Komponenten verringert. Je größer der Abstand der Komponenten, umso eher können globale Signale der geringkonzentrierten Verbindung durch Korrelation nach Kendall oder Spearman extrahiert werden. Ist eine vollständige Trennung der Komponenten erreicht, ist die Korrelation allerdings nicht mehr notwendig.



Abbildung 16: Vergleich der Intensitätsabnahme von extrahierten globalen Signalen berechnet mit den Korrelationskoeffizienten nach Kendall, Spearman und Pearson. Peak A ist in Abschnitt A rot dargestellt. Der Abstand von Peak B zu Peak A wird durch Graustufen dargestellt (schwarz:  $\Delta$ (B-A) geht gegen 0, hellgrau: Der Abstand  $\Delta$ (B-A) geht gegen unendlich). Für jeden Abstand wurde berechnet, wie gut die Korrelation den realen Anteil am Summensignal widerspiegelt, wenn Peak B vom Verhältnis A zu B von 1:1 gegen 1:0 abnimmt. In Abschnitt B stellt die rote Linie den idealen Verlauf (Int<sub>soll</sub> B) dar, die in Graustufen gezeigten Verläufe geben wieder, wie stark der Intensitätsverlauf, der sich aus der Korrelation ergibt, in Abhängigkeit vom Abstand ( $\Delta$ (B-A)) abweicht. Dabei wird sichtbar, dass nach Pearson'schem Korrelationsberechnung die Abweichung am stärksten auftritt.

Im Folgenden wurde für die Experimente die Korrelation nach Pearson durchgeführt. Ein Vergleich zu den Ergebnissen der Kendall'schen oder Spearman'schen Korrelation ist in Einzelfällen diskutiert und ist dann entsprechend benannt.

### 4.3.3 <u>3DCC am Beispiel eines simulierten Chromatographielaufes</u>

#### 4.3.3.1 Dekonvolution von NMR-Spektren

Zur Überprüfung des in Abschnitt 4.3.1 vorgestellten Konzepts der 3D-Kreuzkorrelation wurde ein Chromatographielauf simuliert, um die Korrelierbarkeit von MS- und NMR-Daten unabhängig von LC als zusätzlichen experimentellen Parameter nachweisen zu können. Dazu wurden mathematisch Verläufe zweier Substanzen berechnet (Abbildung 17). Ausgehend von den berechneten Verläufen wurden 15 Fraktionen ausgewählt und diese dann mit den berechneten Mengen jeder Komponente präpariert. Als Substanzen für dieses Modellsystem wurden zwei Oligosaccharide (1 und 2) vom komplexen Typ ausgewählt, die sich lediglich durch eine Fucose-Einheit unterscheiden. (s. Abbildung 17). Diese sind im Folgenden mit Nona (1) und Deca (2) benannt (s. Abbildung 17).



Abbildung 17: Planung des Experiments. Die EICs wurden anhand einer Gauß'schen Normalverteilung so berechnet und gegeneinander verschoben, dass im TIC keine Trennung von Nona und Deca erkennbar ist. Die Kreuze markieren die 15 Fraktionen, die präpariert wurden um so den Chromatographielauf zu simulieren. Von diesen Fraktionen wurden jeweils MS- und NMR-Daten aufgenommen.



Abbildung 18: Darstellung der Kovarianz am Beispiel des Fuc H-6 mit den EICs von Nona 1 und Deca 2. Projektion A zeigt das EDC welches einen dem EIC vom Deca ähnlichen Verlauf aufweist und mit diesem einen Korrelationskoeffizienten von 0.9 bildet, während es mit EIC<sub>Nona</sub> mit einem Koeffizienten von -0.1 korreliert. Die Auftragung des EDCs gegen EIC<sub>Deca</sub> zeigt deutlich den linearen Zusammenhang, der durch eine Regressionsgerade charakterisierbar ist (D). Die Darstellung der Kovarianz des EDCs mit EIC<sub>Nona</sub> zeigt, dass die Werte zufällig verteilt sind, es ist keine Abhängigkeit beobachtbar.

Von den 15 Fraktionen wurden jeweils ESI-TOF-MS- und  $1D^{-1}H$ -NMR-Spektren akquiriert. Die EICs jeder einzelnen Fraktion für **1** bzw. **2** wurden gegen die Fraktionen aufgetragen, um daraus einen rekonstruierten EIC-Verlauf über alle Fraktionen zu erhalten ( $[M+2Na]^{2+}$ -Addukte, Nona **1**: m/z=843.2859, Deca **2**: m/z=916.3147). Daraus ergab sich die Matrix **M**<sub>(m,t)</sub>, mit zwei Zeilenvektoren (M|, die den Verläufen der m/z für **1** und **2** entsprechen. Jeder der 32k Datenpunkte der NMR-Spektren ( $\cong |N\rangle$ ) wurde in gleicher zeitproportionaler Abhängigkeit mit den EIC-Verläufen von Nona- bzw. Decasaccharid verglichen. Die Kovarianz des EDC bei  $\delta=1.214$  ppm, welches dem Fuc H-6 des Deca **2** entspricht, ist in Abbildung 18 dargestellt. Es ist deutlich erkennbar, dass sich nur für die Auftragung des EDCs gegen das EIC<sub>Deca</sub> eine lineare Abhängigkeit ergibt. Niedrige bzw. hohe Werte des EDCs gehen also mit niedrigen bzw. hohen Werten des EIC<sub>Deca</sub> einher. Die Auftragung gegen das EIC<sub>Nona</sub> hingegen zeigt eine zufällige Verteilung, die sich in dem Koeffizienten von -0.1 widerspiegelt. Dieser errechnet sich aus der normierten Kovarianz (s. a. Abschnitt 2.1).



Abbildung 19: Vergleich der EIC-Verläufe (gemessen) zu den EDCs bei  $\delta$ =2.081 und 2.095 ppm (jeweils GlcNAc-2, *N*-Acetylsignal). Der vergleichbare Anstieg und Abfall in den Signalintensitäten ist deutlich erkennbar (gestrichelte Linie: Deca, durchgezogene Linie: Nona). Die Korrelationskoeffizienten der gezeigten EDCs mit den EICs ist in der nebenstehenden Tabelle aufgeführt.

In Abbildung 19 ist beispielhaft das *N*-Acetylsignal des GlcNAc-2 von Nona bzw. Deca gezeigt. Hier zeigt das Signal bei  $\delta$ =2.08 ppm ein dem EIC-Verlauf des Nona entsprechenden Verlauf. Es korreliert mit EIC<sub>Nona</sub> mit einem Korrelationskoeffizienten von 0.91, während sich mit EIC<sub>Deca</sub> ein Koeffizient von 0.08 ergibt. Somit kann dieses Signal eindeutig dem EIC des Nonasaccharids zugeordnet und für eine Charakterisierung verwendet werden. Analog verhält sich der Verlauf des Signals bei  $\delta$ =2.10 ppm mit dem EIC des Decasaccharids.

Die Korrelationskoeffizienten können für die in Abschnitt 4.3.1 beschriebene Extraktion der Spektren mit den für das Nona- bzw. Decasaccharid spezifischen Signalen verwendet werden. Daraus ergeben sich die in Abbildung 20 gezeigten dekonvulierten Spektren.



Abbildung 20: Durch 3DCC dekonvulierte NMR-Spektren von Nona (1) und Deca (2). Aus dem in Abbildung 17 gezeigten simulierten Chromatographielauf ergibt sich das Summenspektrum über alle Fraktionen (oben). Es ist deutlich erkennbar, dass es Signale von beiden Strukturen beinhaltet (beispielsweise GlcNAc-1, H-1 $\alpha$  bei  $\delta$ =5.19 ppm und GlcNAc-2, *N*-Acetylsignal bei  $\delta$ =2.08-2.10 ppm). Durch 3DCC können diese Signale eindeutig dem Nona (mitte) bzw. Deca (unten) zugeordnet werden, wodurch die Charakterisierung möglich wird. Beispielsweise sind die Fucose Signale H-1 und H-6 ( $\delta$ =4.9 und 1.2 ppm) im Spektrum des unfucosylierten Nona 1 nicht erkennbar.

Die obere Spur von Abbildung 20 zeigt jeweils das Summenspektrum ( $S_{(\delta)}$ ). Dieses zeigt beispielsweise im Bereich von  $\delta$ =5.18-5.20 zwei einander überlappende Dubletts auf, die jeweils dem anomeren Proton des Nona- bzw. Decasaccharids entsprechen. Durch die beschriebene Dekonvolution der Spektren können die Dubletts eindeutig dem jeweiligen

Saccharid zugeordnet werden, da sie nur in einem der beiden extrahierten Spektren auftreten. Die structural reporter group-Signale der Fucose sind selektiv im dekonvulierten Spektrum des Decasaccharids, welches alleiniger Träger der Fucose ist, zu erkennen. Gleichzeitig sind sie im Spektrum des Nonasaccharids vollständig unterdrückt. Auch die in Abbildung 19 aufgeführten N-Acetylsignale des GlcNAc-2 können eindeutig einem der beiden Saccharide zugeordnet werden. Sogar minimale Verschiebungen des H-1 der Man-4' werden in den dekonvulierten Spektren aufgelöst. Signale, die für beide Saccharide identisch sind, wie beispielsweise das des H-1 von Man-4 und der H-2 Protonen der Mannosereste, sind in beiden Spektren in korrekter Intensität aufzufinden. Sie weisen im Summenspektrum etwa doppelte Intensität durch den Beitrag durch beide Saccharide auf und korrelieren mit einem Korrelationskoeffizienten von etwa 0.5 bzw. 0.6 (s. a. Abschnitt 4.3.2). Die durch Integration erhältliche Information über die Anzahl zu einem Signal beitragender Protonen konnte auch nach Extraktion der Spektren erhalten werden. Dazu wurden die Flächen der Signale der extrahierten Spektren mittels numerischer Integration in den Integrationsgrenzen der Vergleichsspektren bestimmt und auf das H-1 der Man-4 normiert. Tabelle 3 listet die erhaltenen Integrale im Vergleich zu den theoretisch zu erwartenden Werten auf.

<b>Tabelle 3: Integration der</b>	dekonvulierten Spektren,	die Integrale sind auf	das H-1 von Man-4 normiert.
0	L /	0	

Proton(an)	Nona			Deca		
rroion(en)	$\mathbf{E}_{(m,\delta)}$	theor.	$\Delta$ in %	$\mathbf{E}_{(m,\delta)}$	theor.	$\Delta$ in %
GlcNAc H-1α	0.31	0.60	-48%	0.39	0.60	-36%
Man-4 H-1*	1.0	1.0	-	1.0	1.0	-
Man-4' H-1	0.87	1.0	-13%	1.0	1.0	0%
Fuc H-1	0.0	0.02	2%	0.90	1.0	-10%
GlcNAc5/5' H-1 + GlcNAc-2 H-1(Nona)	2.6	3.0	-14%	1.7	2.0	-16%
Gal-6/6' H-1	1.6	2.0	-19%	1.5	2.0	-24%
Man-3 H-2	1.0	1.0	2%	1.2	1.0	2%
Man-4 H-2	1.1	1.0	5%	1.2	1.0	19%
Man-4' H-2 + Fuc H-5 (Deca)	1.1	1.0	10%	1.9	2.0	-5%
N-Acetylregion 1	12	12	4%	2.8	3.0	-7%
N-Acetylregion 2	-	-	-	9.5	9.0	6%
Fuc H-6	0.10	0	6%	3.0	3.0	-1%
<i>bulk</i> (4.05-3.40 ppm)	45	51	-11%	58	54	7%

 $\Delta$  – Abweichung der Integrale (berechnet - theor.) in %

\* – Integral wurde für Normierung verwendet



Abbildung 21: Grafische Darstellung der prozentualen Abweichung der Integrale vom theoretisch zu erwartenden Wert. Die Asterisk-Markierung bezeichnet das Integral des Man-4 H-1, welches für die Normierung verwendet wurde. Die Abweichungen betragen im Schnitt 10 %. Größere Abweichungen resultieren aus den experimentellen Bedingungen. Beispielsweise sind die H-1α Protonen des GlcNAc-1 auch in experimentellen Spektren meist reduziert. Die Abweichung der Integrale der H-1 von Gal-6/6' und GlcNAc-5/5' erfahren eine leichte Reduktion durch den Einfluss der Wasserunterdrückung. Anhand der in den Korrelationsspektren erhaltenen Integrale lassen sich eindeutige Rückschlüsse auf die Anzahl der zu einem Signal beitragenden Protonen ziehen.

Die bestimmten Integrale bestätigen nochmals die erfolgreiche Extraktion der beiden NMR-Spektren. Beispielsweise werden für das H-1 und die Methylgruppe (H-6) des Fucoserests beim Spektrum des Nonasaccharids Integrale von 0 und bei dem des Decasaccharids Integrale von 1 bzw. 3 erhalten. Auch das H-5-Proton der Fucose, welches in Abhängigkeit von anomerer Konfiguration des GlcNAc-1 unterschiedliche chemische Verschiebungen aufweist und dessen quartett-ähnliche Signale (dq) sich in den Flanken des Man-4' H-2 befinden, tragen nur beim Decasaccharid zum Integral bei. Alle Abweichungen von Integralen vom theoretischen Wert resultieren aus den experimentellen Bedingungen. Beispiel hierfür sind die H-1 Protonen der Galactosereste, deren Integral auch in experimentellen Spektren vermutlich aufgrund der Lösungsmittelunterdrückung leicht reduziert ist.

Durch die Anwendung von 3DCC auf diesen simulierten Chromatographielauf konnten also für beide Saccharide NMR-Spektren extrahiert werden, die zu einem definierten m/z gehören. Aus den extrahierten NMR-Daten ergeben sich die chemischen Verschiebungen wichtiger *structural reporter groups*, sowie das Integral der Signale. Die Kombination von MS und den extrahierten NMR-Daten ermöglicht so eine eindeutige Charakterisierung der Strukturen mit vollständiger Verknüpfungs- und Konfigurationsinformation.

### 4.3.3.2 Automatische Zuordnung der dekonvulierten Spektren

Durch die in Abschnitt 4.3.3.1 gezeigte Dekonvolution von NMR-Spektren kann in Kombination mit den MS-Ergebnissen eine eindeutige Zuordnung vorgenommen werden. Neben der erfolgreichen manuellen Interpretation können auch dekonvulierte Spektren der automatischen Identifikation unterzogen werden (s. Abschnitt 4.1.3). So werden die dekonvulierten Spektren von Nona- und Decasaccharid unter gleichen Bedingungen identifiziert wie die reinen Spektren. Damit ergibt sich aus 1.) chromatographischer Trennung mit starker Überlappung der Komponenten im Chromatogramm 2.) Akquisition von NMR-und MS-Daten 3.) 3D-Kreuzkorrelation und 4.) automatischer Spektreninterpretation die eindeutige Strukturzuordnung.

Abbildung 22 zeigt das extrahierte Spektrum des Decasaccharids 2 (s. Abschnitt 4.3.3.1) im Vergleich mit dem durch automatische Zuordnung gemäß Abschnitt 4.1.3 ermittelten Spektrum. Außerdem ist das Differenzspektrum (extrahiertes Spektrum - berechnetes Spektrum) abgebildet. Unterschiede werden im Differenzspektrum durch negative oder positive Abweichungen von der Basislinie kenntlich gemacht. Diese resultieren einerseits aus unterschiedlicher Linienbreite (z. B. Man-4, H-1), andererseits aus Signalmodulation durch die Wasserunterdrückung in den experimentellen Daten. Beispielsweise sind die H-1 Protonen des GlcNAc-2 ( $\alpha$ , $\beta$ ) aufgrund der Nähe zum HDO-Signal nahezu vollständig unterdrückt, während die Signale der anomeren Protonen von GlcNAc-5/5' und Gal-6/6' einen Dispersionsanteil tragen, der an dem "Einschnitt" erkennbar ist. Unterschiede bei den H-1 der Fucose-Einheit resultiert aus der in sugabase fehlerhaften Zuordnung von  $\alpha$ - und  $\beta$ -Signalen (s. a. Tabelle 1, Abschnitt 4.1.4). Unterschiede von den experimentell erhaltenen Daten gegenüber den berechneten Spektren sind zu erwarten, da beispielsweise von Durchschnittswerten für Kopplungskonstanten und Linienbreiten ausgegangen wird und die berechneten Daten perfekte Linienformen aufweisen und keine Verunreinigungen beinhalten. Durch optimierte Anpassung der berechneten Daten an die experimentellen Spektren könnte eine weitere Verbesserung erreicht werden. Denkbar wäre beispielsweise, die Auswirkungen der Wasserunterdrückung, die abhängig von der Kopplungskonstante und der gewählten Pulslänge in der Unterdrückungssequenz (excitation sculpting) sind, in die Berechnung der Signale einzubeziehen. Gezeigt ist jedoch, dass die automatische Spektrenerkennung tolerant gegenüber geringfügigen Unterschieden ist, solange die experimentellen Daten ein hohes S/N aufweisen. Durch die Verwendung von 3DCC geht generell eine S/N-Verbesserung einher, da für die Berechnung des extrahieren Spektrums das Summenspektrum verwendet wird (s. a. Abschnitt 4.4.1.2).



Abbildung 22: Automatische Spektrenidentifikation, a: zu untersuchendes Spektrum U, entspricht hier dem durch Korrelation mit dem EIC der Masse m/z=916.3147 extrahierten Decasaccharid-Spektrum; b: durch automatische Spektreninterpretation gefundenes, berechnetes Spektrum, entspricht hier der CCSD-No. 42857 (Decasaccharid 2), also der korrekten Zuordnung; c: Differenzspektrum (a-b). Verwendete Kurzschreibweise: G (Gal), GN (GlcNAc), M (Man), F (Fuc). Signale die in c negativ sind, entsprechen also Bereichen, die im berechneten Spektrum b zu stark vertreten sind. Signale die in c positiv sind, entsprechen Signalen, die im berechneten Spektrum b nicht oder vermindert auftreten. Es sind einige Abweichungen erkennbar, was zeigt, dass eine gewisse Toleranz gegenüber Linienbreite (M4, H-1) und z.B. durch Wasserunterdrückung modulierte (z.B. G6/G6' H-1) oder unterdrückte Signale (z.B. GN2 H-1) gegeben ist.

#### **Eindeutige Zuordnung von Korrelationsspektren**

Das Korrelationsspektrum des Decasaccharids wird beim Durchsuchen von 523 möglichen berechneten Spektren bei einem *cut off* von 0.2 eindeutig der richtigen CCSD-Nr. 42857 zugeordnet. Das Korrelationsspektrum des Nona wird beim gleichen *cut off* ebenfalls eindeutig der richtigen CCSD-Nr. zugeordnet, wenn, wie zuvor beschrieben, alle Einträge um 0.0026 ppm zu tieferem Feld verschoben werden (s. a. 4.1.4).

## 4.4 <u>Analytik der Glycane des bovinen Fibrinogens</u>

Bovines Fibrinogen ist ein Glycoprotein, welches an der Blutgerinnungskaskade beteiligt ist. Es weist eine dimere Struktur auf, deren monomere Einheiten aus jeweils drei unterschiedlichen Ketten (A $\alpha$ , B $\beta$  und  $\gamma$ ) aufgebaut sind.<sup>96,97</sup> Die Glycane sind an zwei Glycosylierungsstellen innerhalb der  $\beta$ - und  $\gamma$ -Kette verknüpft und bilden einen Anteil am Glycoprotein von 3.9 Gew.-%.<sup>47,98</sup> Aufgrund der guten Verfügbarkeit eignet es sich als Modellsystem. Im Folgenden wird dargelegt, dass durch die Kombination von MS- und NMR-Daten Strukturen nachgewiesen werden konnten, die für das bovine Fibrinogen bisher nicht publiziert sind. Außerdem konnten NMR-Daten für massenisobare Strukturen erhalten werden, die bisher nicht literaturbekannt sind. Im folgenden Abschnitt wird zudem gezeigt, dass aus einem Chromatographielauf mit starker Überlappung der einzelnen Komponenten durch die Anwendung von 3DCC Strukturinformationen erhalten werden können, die eine eindeutige Charakterisierung erlauben.

## 4.4.1 <u>3DCC zur Charakterisierung der neutralen Glycane vom bovinen Fibrinogen</u>

Die Glycane des bovinen Fibrinogens wurden nach modifiziertem literaturbekannten Protokoll abgespalten und gereinigt.<sup>99,100</sup> Dazu wurde das Protein zunächst denaturiert, anschließend durch einen tryptischen Verdau in Peptide und Glycopeptide gespalten und die Glycane nach zwischenzeitlicher Dialyse durch Verdau mit PNGase F abgespalten. Die Umsetzung mit Sialidase lieferte neutrale Glycanstrukturen, die über eine anschließende gemischte Ionentauschersäule gereinigt wurden. Dabei verlief die Umsetzung mit der Sialidase nicht quantitativ, was in einer Anreicherung von in der nativen Form desialylierten Strukturen resultierte (s. a. Abschnitt 4.4.4). Von diesem Gemisch **G1** wurde eine modifizierte chromatographische Trennung nach Muddiman *et al.* an PGC durchgeführt.<sup>101</sup> Jede der fünf Hauptkomponenten weist unter den verwendeten Bedingungen im Chromatogramm zwei Peaks auf, je einen für  $\beta$ - und  $\alpha$ -Anomer. Daraus resultiert eine starke Überlappung der Peaks und die Substanzen sind nur sehr unvollständig getrennt (s. Abbildung 23), da entweder  $\alpha$ -oder  $\beta$ -Peak oder aber beide eine starke Überlappung mit Peaks können durch 3DCC die zur Strukturaufklärung notwendigen Informationen erhalten werden.



Abbildung 23: 3DCC am Beispiel eines Gemisches komplexer biantennärer Glycanstrukturen aus dem bovinen Fibrinogen mit sämtlichen Informationen die im Chromatogramm enthalten sind. A: Chromatogramm, detektiert als TIC, mit EICs. B: Tabelle mit Retentionszeiten und Auswertung der MS-Daten. C: extrahierte NMR-Daten ( $E_{(m,\delta)}$  bzw. mit Asterisk\*:  $E4_{(m,\delta)}$ ), die zusammen mit den extrahierten m/z den rechts dargestellten Zuckerstrukturen zugeordnet werden können.

Durch Anwendung von 3DCC können nach Akquisition von MS- und NMR-Daten in der zeitlichen Abhängigkeit des Chromatographielaufes die Spektren a-e) extrahiert werden. Die Kombination der extrahierten NMR-Spektren mit den definierten m/z resultiert in der Zuordnung der als Piktogramme gezeigten Strukturen.

# **Extraktion der MS-Daten**

Anhand des TIC können die in Abbildung 23 gezeigten EICs extrahiert werden. Diese sind zusammengefasst in der Matrix  $\mathbf{M}_{(m,t)}$ . Zu erkennen ist, dass jedes EIC zwei Maxima aufweist, die aus der Trennung der  $\alpha$ - und  $\beta$ -Anomere resultieren. Die Peaks zeigen starke Überlappungen. Aus der Integration der EICs errechnen sich die Anteile der Komponenten zu a=40 %, b=20 %, c=18 %, d=14 % und e=9 %.

Tabelle 4: Korrelation der EICs untereinander, grau hinterlegt sind die paarweise maximalenKorrelationskoeffizienten.

	EIC <sub>a</sub>	EIC <sub>b</sub>	EIC <sub>c</sub>	EIC <sub>d</sub>	EIC <sub>e</sub>
	(Deca <b>2</b> )	(fuc. Nona <b>3</b> )	(Nona 1)	(Dodeca 4)	(Undeca 5)
EIC <sub>a</sub> (Deca 2)	1.00	0.27	0.20	-0.11	0.00
$EIC_b$ (fuc. Nona 3)	0.27	1.00	0.11	-0.27	0.10
$EIC_c$ (Nona 1)	0.20	0.11	1.00	-0.28	-0.33
$EIC_d$ (Dodeca 4)	-0.11	-0.27	-0.28	1.00	0.04
$EIC_e$ (Undeca 5)	0.00	0.10	-0.33	0.04	1.00

Eine Abschätzung des minimalen Korrelationskoeffizienten um ein NMR-Signal (EDC) eindeutig einem EIC zuweisen zu können, ergibt sich aus der Korrelation der EICs untereinander (s. Tabelle 4). Beispielsweise korreliert das EIC<sub>b</sub> (fuc. Nona 3) mit EIC<sub>a</sub> (Deca 2) mit einem Koeffizienten von 0.27. Demzufolge können nur EDCs die mit dem EIC<sub>b</sub> des fucosylierten Nonasaccharids 3 einen Korrelationskoeffizienten von >0.27 aufweisen eindeutig EIC<sub>b</sub> zugeordnet werden.

## Komponente a: Decasaccharid (2)

Das Decasaccharid (2) ist die Hauptkomponente des Gemisches (Spektrum, s. Abbildung 23a). Durch 3DCC kann das Subspektrum **a** extrahiert werden, das mit dem experimentellen Spektrum des Decasaccharids im Bereich der *structural reporter groups* vollständig übereinstimmt.

#### Komponente b: fucosyliertes Nonasaccharid (3)

Die Auswertung des MS-Spektrums ergibt die Zusammensetzung Hex4HexNAc4dHex (Hex: HexNAc: *N*-Acetylhexosamin, dHex: Deoxyhexose), woraus sich Hexose, als Strukturvorschlag das um ein Galactoserest reduzierte Decasaccharid ergibt. Die Struktur und Position der Galactose kann durch Anwendung des structural reporter group-Konzepts aus einem reinen NMR-Spektrum schnell bestätigt bzw. identifiziert werden, während diese Information anhand eines Mischungsspektrums oder allein durch MS oder MS/MS nicht erhältlich ist. Die 1,4-Verknüpfung des Galactoserestes am 3-Arm ist sichtbar an der Verschiebung des H-1 des Galactoserestes, sowie der Verschiebung des anomeren Protons des Man-4'. Das dekonvulierte Spektrum zeigt gegenüber dem Summenspektrum die Verstärkung der Intensität des Gal-6 H-1 und des Man-4' H-1 (s. Abbildung 24, jeweils durch Pfeile hervorgehobenen Signale), welche die Position am 3-Arm im Vergleich zu den Signalen einer möglichen Verknüpfung am 6-Arm bestätigen. Jedoch ist eine eindeutige Aussage so erschwert. Das Einführen einer gewichteten Dekonvolution ermöglicht die eindeutige Zuordnung. Dazu geht der Korrelationskoeffizient in vierfacher Potenz in die Multiplikation mit dem Summenspektrum ein. Daraus resultiert die stärkere Gewichtung hoher Korrelationskoeffizienten gegenüber niedrigen Korrelationskoeffizienten im Vergleich zur linearen Gewichtung (s. a. Abbildung 28). Abbildung 24 zeigt einen Ausschnitt des dekonvulierten, sowie des gewichtet-dekonvulierten Spektrums im Vergleich mit einem experimentell erhaltenen Spektrum des isolierten Oligosaccharids 3 sowie dem Summenspektrum aller NMR-Daten ( $S_{(\delta)}$ ).

Die Zuordnung des Isomers erfolgt über das Dublett des Galactoserestes bei  $\delta$ =4.467 ppm, sowie das anomere Proton der Man-4' bei  $\delta$ =4.916 ppm, welches durch die fehlende Galactose gegenüber dem Signal bei  $\delta$ =4.925 ppm eines galactosetragenden 6-Arms verschoben ist. Auch das GlcNAc-5' H-1 Signal bei  $\delta$ =4.554 ppm bestätigt die in Abbildung 24 gezeigte Struktur, sodass diese in Kombination mit MS zweifelsfrei charakterisiert werden kann. Weiterhin ist auch eine Auflösungsverbesserung im gewichtetdekonvulierten Spektrum erkennbar, die beispielsweise die Auswertung des Fuc H-1 erleichtert. Beim Durchsuchen von *sugabase* und *glycosciences.de* wird eindeutig die korrekte Struktur gefunden, wenn die Signalliste anhand der Peakmaxima von Man-4' H-1, Fuc H-1 ( $\alpha$ , $\beta$ ), GlcNAc-5 und GlcNAc-5' H-1, und des anomeren Protons von Gal-6 im gewichtetextrahierten Spektrum erzeugt wird.



5'

4'

1'

(3)

Abbildung 24: Fucosyliertes Nonasaccharid (3): Vergleich von extrahierten Spektren mit dem Summenspektrum  $(S_{(\delta)})$  und einem experimentellen Spektrum (Exp) der gereinigten Komponente. Die chemischen Verschiebungen des H-1 der Man-4<sup>4</sup> und Gal-6 sind essentiell für die Unterscheidung der Position der Galactosylierung. Durch die Gewichtung wird neben der Verbesserung des S/N eine zusätzliche Auflösungsverbesserung durch die bessere Diskriminierungseigenschaft der mit der 4. Potenz gewichteten Korrelationskoeffizienten erreicht (E4<sub>(m,\delta)</sub>).

Die im einfach-dekonvulierten NMR-Spektrum auftretenden Signale bei  $\delta$ =4.473 ppm und  $\delta$ =4.925 ppm, die nicht zu der gezeigten Struktur gehören, resultieren vermutlich aus einer partiellen Korrelation mit dem im Chromatogramm überlappenden Decasaccharid. Die

Korrelation von  $EIC_b$  (fuc. Nona 3) und  $EIC_a$  ergibt einen Koeffizienten von 0.27, sodass dieser einen Grenzwert für die Unterscheidbarkeit darstellt.

Das Isomer, welches den Galactoserest am 6-Arm trägt, würde zwar ebenso die zuvor genannten Signale aufweisen, in diesem Falle müsste aber die Intensität der beiden anomeren Protonen von GlcNAc-5 bzw. 5' bei  $\delta$ =4.580 ppm und  $\delta$ =4.554 ppm identisch sein. Da die Intensität des Dubletts bei  $\delta$ =4.580 ppm im einfach-dekonvulierten Spektrum höher ist, muss sich diese aus den zwei GlcNAc-5/5' H-1 des Deca und dem GlcNAc-5 H-1 des fucosylierten Nona **3** zusammensetzen. Im Gegensatz dazu entspricht das Signal bei  $\delta$ =4.554 ppm nur einem Signal, dem H-1 des GlcNAc-5' von **3**. Das bestätigt sich auch darin, dass laut MS-Daten einer vergleichbaren Probe das Verhältnis des fucosylierten Nonasaccharids (**3**) zu einer isomeren Struktur von 14:1 ergibt und somit die Intensität deutlich geringer sein müsste.

## Komponente c: Nonasaccharid (1)

Anhand der MS-Daten kann Komponente c als Hex5HexNAc4 nachgewiesen werden. Die *structural reporter group*-Signale im dekonvulierten NMR-Spektrum bestätigen die Abwesenheit der Fucose und führen zu der Zusammensetzung eines biantennären Glycans vom komplexen Typ. Wichtige Merkmale des gewichtet-extrahierten Spektrums sind die Position des GlcNAc-1 H-1 $\alpha$  bei  $\delta$ =5.188 ppm, die Unterdrückung der Fucose-Signale (H-1 bei  $\delta$ =4.89 ppm und H-6 bei  $\delta$ =1.21 ppm), die anomeren Protonen zweier Galactosereste und die Position der anomeren Protonen von Man-4 und -4' bei  $\delta$ =5.119 ppm und  $\delta$ =4.927 ppm. Allerdings erfahren einige Signale, wie beispielsweise die H-2 Protonen der Mannosereste, sowie das *N*-Acetylsignal von GlcNAc-5 eine Reduktion gegenüber dem einfach extrahierten Spektrum oder einem experimentellem Spektrum. Dieses tritt insbesondere dann auf, wenn NMR-Signale global auftreten, also identisch für mehrere verschiedene Strukturen sind. Dieses Problem wird in Abschnitt 4.4.1.1 diskutiert. Anhand der eindeutigen *structural reporter group*-Signale in Kombination mit der aus dem Massenspektrum erhältlichen Information ergibt sich zwingend die oben gezeigte Struktur (s. Abbildung 23).

#### Komponente d: Dodeca 4

Komponente **d** hat nach MS die Zusammensetzung Hex7HexNAc4dHex. Dem Korrelationsspektrum sowie dem gewichtet-korrelierten Spektrum können folgende NMR-Signale sicher entnommen werden: i) die H-1 $\alpha$  zweier 1,3-verknüpfter Galactosereste bei  $\delta$ =5.147 und 5.145 ppm, ii) die H-1 $\beta$  Protonen bei  $\delta$ =4.54 ppm, charakteristisch für die Galactosereste 6 und 6', an deren 3-Position eine weitere Verknüpfung zu finden ist, sowie

iii) ein intensives Signal im Bereich von 4.20 - 4.18 ppm, bei dem neben dem Man-4 H-2 die H-5 Protonen von Gal-7/7<sup>+</sup>, sowie die H-4 von Gal-6/6<sup>+</sup> zu finden sind, und iv) ein Signal bei 4.02 ppm, welches charakteristisch für das H-4 Proton von α1,3-verknüpften Galactoseresten ist (s. a. Abschnitt 4.4.5). Weitere Informationen werden beispielsweise durch die nur im einfach korrelierten Spektrum sichtbaren Signale der GlcNAc-5/5<sup>+</sup> H-1 Protonen erhalten. Fehlende Signale resultieren wiederum daraus, dass Signale global auftreten. Es werden also bei geringer konzentrierten Komponenten im Wesentlichen die charakteristischen Signale, wie die zuvor beschriebenen der zusätzlichen Galactosereste, hervorgehoben. Auch durch die Berechnung von Kendalls Rangkoeffizienten konnte hier keine Verbesserung beobachtet werden.

Alle fehlenden Signale müssen aber zumindest im Summenspektrum auffindbar sein. Deshalb kann beispielsweise mit großer Sicherheit von einer 1,6-verknüpften Fucose-Einheit ausgegangen werden, da die für eine 1,3-verknüpfte Fucose charakteristischen Signale auch im Summenspektrum nicht auftreten. Auch das H-1 $\alpha$  des proximalen GlcNAc-1 muss sich mit der chemischen Verschiebung der Hauptkomponente decken, da es bei individueller Verschiebung im Korrelationsspektrum auftreten würde. Anhand der wichtigen Signale kann in Kombination mit der Information aus den MS-Daten die in Abbildung 23 gezeigte Struktur erhalten werden.

## Komponente e: Undeca 5

Bei Komponente e handelt es sich nach MS um ein Oligosaccharid der Komposition Hex6HexNAc4dHex. Die Zuordnung erfolgt anhand ähnlicher Argumente wie Komponente d. Wiederum werden die Signale hervorgehoben, die sich von den Signalen der weiteren Strukturen unterscheiden. Die chemische Verschiebung des H-1 von  $\alpha$ 1,3verknüpften Gal-7' bei  $\delta$ =5.147 ppm, tieffeldverschoben gegenüber dem einer Gal-7, ist charakteristisch für die Verknüpfung mit der Gal-6'. Außerdem erscheint zumindest partiell das Dublett des H-1 der terminalen Gal-6 bei  $\delta$ =4.47 ppm. So kann eindeutig gesagt werden, an welcher der beiden Antennen der zusätzliche Galactoserest verknüpft ist. Weiterhin ist im einfach-dekonvulierten Spektrum neben dem H-1α von GlcNAc-1, sowie den H-1 von Man-4' und Fuc-1', auch das Auftreten der weiteren Gal-Signale (H-4, H-5) zu erkennen. Ein Vergleich der Extraktion durch Verwendung der Korrelation nach Pearson, Spearman und Kendall ist in Abbildung 25 aufgeführt. Beispielsweise ist das H-1 der Man-4 bei  $\delta$ =5.12 ppm nach Kendall und Spearman deutlich besser extrahiert, gleichzeitig ist im Bereich von  $\delta$ =4.47 ppm nur anhand der Pearson-Korrelation die Unterscheidung der H-1 von Gal-6 und Gal-6' und damit eine Unterscheidung von isomeren Strukturen möglich. Detailliertere Angaben zur Zuordnung finden sich in Abschnitt 4.4.5.



Abbildung 25: Extraktion des Undecasaccharids nach Pearson, Spearman und Kendall. Die Korrelationsspektren, die nach Pearson  $(E_{(m,\delta)})$ , Spearman  $(ES_{(m,\delta)})$  und Kendall  $(ES_{(m,\delta)})$  erhalten wurden sind im Vergleich zum Summenspektrum  $(S_{(\delta)})$  und einem experimentellen Spektrum des gereinigten Undecas (Exp) gezeigt. Beispielsweise wird das Man-4 H-1 Signal deutlich besser durch die Korrelation nach Spearman und Kendall erhalten, jedoch ist dafür die Unterscheidung des Gal-6 H-1 bei  $\delta$ =4.467 ppm gegenüber dem Gal-6' H-1, welches im Undeca von  $\delta$ =4.473 zu  $\delta$ =4.544 ppm verschoben ist, durch Pearson besser gegeben. Zusammen mit der chemischen Verschiebung des H-1 $\alpha$  des 1,3-Gal-Restes bei  $\delta$ =5.147 ppm lässt sich so das Undecaisomer identifizieren.

# Informationsgewinn durch 3DCC

Das hier gezeigte Chromatogramm stellt ein Extrembeispiel bezüglich suboptimaler Trennung der Komponenten dar. Selbst bei diesem Lauf, bei dem die Komponenten sehr starke Überlappungen aufwiesen, konnten die zur Strukturzuordnung nötigen Informationen erhalten werden. Entsprechend universell ist 3DCC also bei unvollständig getrennten Substanzgemischen einsetzbar.

# 4.4.1.1 Methoden zur Optimierung der Spektrenextraktion

In den oben genannten Beispielen (Abschnitt 4.4.1) wird ersichtlich, dass in Abhängigkeit der Substanzmenge und des Überlappungsgrades der EIC-Daten einzelne Signale nicht perfekt korreliert werden, d. h. entweder nicht vollständig unterdrückt oder zu stark unterdrückt sind. Aus diesem Grund wurden unterschiedliche Methoden auf Grundlage der Pearson'schen Korrelation darauf hin überprüft, ob eine Verbesserung der extrahierten Daten erreicht werden kann (Tabelle 5).

Methode	Beschreibung der Methode
RT	NMR-Spektren von Fraktionen für die gilt: $I_{EIC} < max(I_{EIC})/100$ werden
(Retentionszeit)-	gleich Null gesetzt.
Filter	
cut off	Korrelationskoeffizienten $c_{m,\delta} < cut \ off$ werden = 0 gesetzt. ( <i>cut off</i> wird entsprechend Tabelle 4 gewählt)
Signifikanztest	Es gilt: wenn $p_{EIC,EDC}$ <0.05, dann ist die Korrelation ( $c_{EIC,EDC}$ ) signifikant, Konfidenzlevel 5 %
Gewichtung	Gewichtung mit 4. Potenz:
	$c_{m,\delta} \rightarrow c_{m,\delta}^4$
	Gewichtung durch sigmoide Funktion:
	$(c_{m,\delta})_s \rightarrow \frac{1}{1+e^{-k\cdot(c_{m,\delta}-x)}}$

Tabelle 5: Methoden zur	Optimierung der	r Spektrenextraktion.
-------------------------	-----------------	-----------------------

Abbildung 26 zeigt das Ergebnis der verschiedenen Methoden am Beispiel der Korrelationsdaten des fucosylierten Nonasaccharids (3). Eingerahmt ist das Spektrum, welches durch gewichtete Extraktion ( $\mathbf{E4}_{(m,\delta)}$ ) erhalten und als die optimale der verglichenen Methoden bewertet wurde.



Abbildung 26: Vergleich verschiedener Optimierungsvarianten am Beispiel des fucosylierten Nonasaccharids. Die Pfeile 1-4 markieren unterschiedliche Signale, die hier detaillierter betrachtet werden. Mit Rahmen versehen ist das gewichtet-extrahierte Spektrum, welches für die Charakterisierung ausgewählt wurde (s. a. Abbildung 24). Pfeil 1 markiert das H-1 der Man-4<sup>c</sup>, Pfeil 2 den Bereich der H-1-Protonen von terminalen 1,4-verknüpften Galactoseresten. Beide Signalgruppen sind nur im Spektrum B eindeutig zuordenbar. Pfeil 3 markiert die verbesserte Auflösung des *N*-Acetylsignals des GlcNAc-2 ( $\alpha$  und  $\beta$ ), Pfeil 4 das in den Spektren A, C-F zu viel auftretende *N*-Acetylsignal eines galactosylierten 6-Arm GlcNAc (5<sup>c</sup>). Insgesamt wurde die Methode der mit vierter Potenz des Korrelationskoeffizienten gewichteten Extraktion, aus der Spektrum B erhalten werden kann, als die optimale der vorgestellten Methoden im Vergleich zur einfachen Extraktion bewertet.
Wichtig für die Unterscheidung, welcher Arm galactosyliert vorliegt, sind die chemischen Verschiebungen der Signale von Man-4' H-1 und Gal-6 H-1 (markiert in Abbildung 26 mit den Pfeilen 1 und 2). Die Spektren A und C-F weisen hier doppelte Signalsätze auf, die vermutlich aus der Überlappung mit dem Decasaccharid (2) resultieren (Man-4' H-1:  $\delta$ =4.916 ppm und  $\delta$ =4.925 ppm, Gal H-1:  $\delta$ =4.467 ppm und  $\delta$ =4.473 ppm). Dadurch ist eine eindeutige Interpretation nur anhand des gewichtet-extrahierten Spektrums möglich.

Am Beispiel der EDCs der Signale bei  $\delta$ =4.926 ppm und  $\delta$ =4.916 ppm werden im Folgenden die Resultate der verschiedenen Methoden erläutert (Abbildung 27).

# **Einfache Extraktion ohne Gewichtung**

Die beiden Signale bei  $\delta$ =4.926 (A) und 4.916 ppm (B) weisen im Summenspektrum ein Intensitätsverhältnis von 1:0.6 auf. EDC<sub>B</sub> korreliert mit EIC<sub>fuc. Nona</sub> mit 0.90, während EDC<sub>A</sub> entsprechend mit 0.36 korreliert. Daraus ergibt sich für das einfach extrahierte Spektrum (ohne Filter, Abbildung 26A) ein Intensitätsverhältnis von A zu B von 0.7:1, beide Signale sind also vorhanden.

### Extraktion nach Gewichtung des Korrelationskoeffizienten mit der 4. Potenz

Durch die gewichtete Extraktion ergibt sich dann ein Verhältnis von 0.04:1, wodurch eine eindeutige Interpretation möglich ist. Erkennbar ist, dass die NMR-Signalverläufe von größerem Rauschen betroffen sind als die EICs, was sich aus der intrinsischen Gerätesensitivität ergibt. Vermutlich ergeben sich daraus Variationen des Korrelationskoeffizienten.

### **Signifikanztest**

Der p-Filter ist die Einbeziehung eines Signifikanztests, der die Wahrscheinlichkeit einer zufälligen Korrelation ermittelt und damit ein Maß für die Zuverlässigkeit des bestimmten Korrelationskoeffizienten darstellt. Für p<0.05 ist die Wahrscheinlichkeit einer zufälligen Korrelation <5%. Generell gilt, dass die Bestimmung des Korrelationskoeffizienten umso genauer ist, desto mehr Datenpunkte für die Korrelation zur Verfügung stehen. Dadurch, dass die Korrelation von  $c_{EIC_{fuc. Nona}, EIC_A}$ =0.36 signifikant ist (für p<0.05), kann durch die Anwendung des Signifikanztests hier keine Optimierung bezüglich der Unterscheidung der Signale erwirkt werden.



Abbildung 27: Vergleich von EICs und EDCs bei  $\delta$ =4.926 ppm und  $\delta$ =4.916 ppm. Im Summenspektrum sind in cyan die nach *sugabase* vorgegebenen chemischen Verschiebungen der Verbindungen 1-5 markiert. Das NMR-Signal bei  $\delta$ =4.916 ppm (EDC<sub>B</sub>, rot) ist essentiell, um die Struktur des fucosylierten Nonasaccharids (3) eindeutig charakterisieren zu können. Es korreliert mit dem EIC von 3 mit einem Koeffizienten von 0.90. EDC<sub>A</sub> entspricht dem Verlauf des NMR-Signals bei  $\delta$ =4.926 ppm. Diese chemische Verschiebung weist das H-1 der Man-4<sup>c</sup> auf, wenn die 6-Arm Antenne mindestens eine Galactose trägt, wie es für die Verbindungen 1,2,4 und 5 der Fall ist. EDC<sub>A</sub> ( $\delta$ =4.926 ppm) korreliert mit dem EIC von 3 mit einem Korrelationskoeffizienten von 0.36. Dadurch erscheint dieses Signal im Korrelationsspektrum des fucosylierten Nonasaccharids, obwohl es keinem Proton zuordenbar ist. Durch die Gewichtung wird das Verhältnis von Signal A:B von 0.7:1 (im einfach extrahierten Spektrum) zu 0.04:1 (im gewichtet extrahierten Spektrum) verschoben.

### **Definition des** *cut off* **Filters**

Tabelle 4 zeigt, dass die EICs von Deca (2) und fucosyliertem Nona (3) eine Korrelation von 0.27 aufweisen, weshalb dieser Wert als *cut off* für die Extraktion der Spektren D und F gewählt wurde. Analog können die Werte für die Extraktion der anderen vier Komponenten gewählt werden. Der *cut off* Filter wurde hier so definiert, dass die Korrelationskoeffizienten gleich Null gesetzt werden, wenn sie unterhalb des definierten *cut offs* liegen. Daraus ergibt sich allerding ein unstetiger Verlauf des Gewichtungsfaktors beim *cut off.* Werte knapp oberhalb des *cut offs* sind also überbewertet gegenüber Werten, die unterhalb liegen (s. Abbildung 28).



Abbildung 28: Entwicklung des Korrelationskoeffizienten bei verschiedener Gewichtung. Ähnlich wie beim hier gewählten *cut off* von 0.27 werden auch bei der Gewichtung des Koeffizienten mit der 4. Potenz Werte unterhalb von 0.3 nahezu vollständig unterdrückt.

Dadurch, dass  $c_{EIC_{fuc. Nona}, EIC_A}$  oberhalb des *cut offs* liegt, wird hier keine verbesserte Unterscheidung der Signale erreicht. Prinzipiell ließe sich der *cut off* sicherlich manuell anpassen, sodass eine Verbesserung auftreten kann. Zielsetzung ist jedoch, ein automatisch einsetzbares Verfahren zu entwickeln, das auch bei unbekannten Verbindungen die korrekte Extraktion erlaubt, ohne dass abgeschätzt werden muss, welcher *cut off* für welche Komponente im Gemisch geeignet sein könnte.

# **Anwendung eines Retentionszeit-Filters**

Weist das EIC einer Komponente in einer Fraktion weniger als 1% der maximalen Intensität auf, so sind die NMR-Signale dieser Komponente im Spektrum der entsprechenden Fraktion vernachlässigbar gering. Deshalb wurden beim Retentionszeit-Filter (RT-Filter) diese Spektren parallel zu den MS-Daten entweder gleich Null gesetzt, wie es für das Spektrum in Abbildung 26 (E) vorgenommen wurde oder können alternativ aus der Korrelation komplett entfernt werden. Ein Problem bei der Anwendung des RT-Filters ist, dass die Anzahl an für die Korrelation relevanten Datenpunkten verringert wird, sodass im Endeffekt die Wahrscheinlichkeit einer zufälligen Korrelation erhöht werden kann. Entsprechend kann die zufällige Korrelation eines Datenpunktes durch eine bessere zeitliche Auflösung, also die Aufnahme von mehr Datenpunkten, verringert werden. Auch die Einbeziehung des Signifikanztests ist denkbar. Allerdings wurden hier ebenfalls keine optimalen Ergebnisse beobachtet, da essentielle Signale nicht erhalten wurden (Spektrum nicht gezeigt). Häufig ähnelten die Korrelationsspektren, welche durch Anwendung des RT-Filters gewonnen wurden, dem Summenspektrum der verbleibenden Fraktionen.

### Kombination des Retentionszeitfilters mit einem cut off

Die Kombination mit einem *cut off* kann sinnvoll sein, unterliegt aber den gleichen Problemen, wie zuvor beschrieben. Der RT-Filter ist dann gut einsetzbar, wenn Substanzen im Chromatogramm Peaks ohne starke Überlappung aufweisen, wie es für den  $\alpha$ -Peak des Dodeca der Fall ist (Abbildung 27). Beispielsweise zeigt EDC<sub>A</sub> in den letzten Fraktionen (27-31) ein Profil, welches eindeutig dem  $\alpha$ -Peak im EIC des Dodeca entspricht. Trotzdem beträgt der Korrelationskoeffizient beider Verläufe nur 0.02, weshalb das Signal im extrahierten Spektrum des Dodecasaccharids nicht auftritt. Das liegt daran, dass das anomere Signal der Man-4<sup>+</sup> bei  $\delta$ =4.926 ppm für 4 der 5 Verbindungen bei der gleichen chemischen Verschiebung und somit global auftritt und deshalb als Profil eher der Summe der EICs entspricht. So ist die Korrelation zu den höher konzentrierten Substanzen wie dem Deca (2) stärker ausgeprägt. Hier kann durch den Retentionszeitfilter die Korrelation verbessert werden, da der globale Charakter des Signals verringert wird. Allerdings ist der  $\alpha$ -Peak des Dodeca von keiner Überlappung betroffen, sodass hier auch die Verwendung des Summenspektrums über die Fraktionen 27-31 ein interpretierbares Spektrum ergibt, wobei die Signalintensitäten der Spektren im Bereich des  $\beta$ -Peaks verloren gehen.

### Einfluss der gewichteten Extraktion auf die Korrelation globaler Signale

Die gewichtete Extraktion wurde hier als am besten geeignete der verwendeten Methoden vorgestellt, ist aber ebenfalls von der Problematik global auftretender Signale betroffen. Das wird am Beispiel des *N*-Acetylsignals in den Spektren des Nonasaccharids deutlich (Abbildung 29).



Abbildung 29: Auswirkung der Gewichtung am Beispiel der *N*-Acetylregion des fucosylierten Nonasaccharids (3) und des Nona (1).  $E_{(m,\delta)}$ -  $E4_{(m,\delta)}$ : bezeichnen die Potenz mit der der Korrelationskoeffizient  $c_{(m,\delta)}$  in die Berechnung der Spektren eingeht. Gezeigt ist die Region der *N*-Acetylprotonen der GlcNAc-Reste (GN) für das fucosylierte Nona 3 und das Nona 1. Die Pfeile a-c, sowie die gestrichelte Linie markieren Besonderheiten in den Spektren. Am Beispiel des fucosylierten Nona 3 ist die Optimierung des Spektrums durch stärkere Gewichtung erkennbar. Pfeil a markiert das GlcNAc-2 (GN-2) des fucosylierten Nona 3, welches durch das Auftreten von  $\alpha$ - und  $\beta$ -Konfiguration des benachbarten GlcNAc-1 zwei Signale im Spektrum aufweist. Hier ist eine deutliche Unterscheidung des zum  $\alpha$ -Anomer gehörenden Signals im Gegensatz zu dem des  $\beta$ -Anomers zu erkennen, was einer Auflösungsverbesserung entspricht. Dieses ist auch im reinen Spektrum erst bei entsprechend harter Prozessierung erkennbar. Pfeil b weist auf ein Signal hin, welches nicht zum fucosyliertem Nona gehört und durch die Gewichtung mit 4. Potenz weitgehend eliminiert wird. Gleichzeitig wird im Fall des Nonasaccharids das mit Pfeil c markierte *N*-Acetylsignal des GlcNAc-5 des Nona 1 unterdrückt und das Signal des GlcNAc-1 (gestrichelte Linie) in der Intensität vom  $\alpha$ - zum  $\beta$ -Maximum verschoben. Dieses resultiert aus dem globalen Charakter der Signale (s. a. Abbildung 30).

GN-5

2.04

GN-1

Hierbei wurde der Korrelationskoeffizient mit den Potenzen 1-4 in die Berechnung der extrahierten Spektren einbezogen. Während die *N*-Acetylregion des fucosylierten Nonasaccharids durch die stärkere Gewichtung eine Auflösungsverbesserung erfährt (Abbildung 29, GN-2, Pfeil a) und das mit Pfeil b markierte Signal besser unterdrückt wird, ist in der gleichen Region des Nona (1) eine Modulation der Signale erkennbar. Zwar wird das Signal bei  $\delta$ =2.096 ppm ebenfalls unterdrückt, gleichzeitig erfährt aber auch das zum Spektrum gehörende *N*-Acetylsignal des GlcNAc-5 eine drastische Intensitätsreduktion (Pfeil c). Außerdem lässt sich beim *N*-Acetylsignal eine Hochfeld-Verschiebung des Maximums beobachten, die beim gleichen Signal des fucosylierten Nonasaccharids nicht auftritt (jeweils gestrichelte Linie). Das Signal liegt bei beiden reinen Komponenten bei  $\delta$ =2.039 ( $\alpha$ ) und 2.037 ppm ( $\beta$ ). Im Falle des Nonasaccharids wird die Intensität zum quantitativ geringer vorkommenden  $\beta$ -Signal bei  $\delta$ =2.037 ppm verschoben. Die NMR-Spektren der *N*-Acetylregion sind in Abbildung 30 zusammen mit den EDCs bei  $\delta$ =2.039 ppm und  $\delta$ =2.037 ppm aufgeführt. Außerdem sind die EICs zum Vergleich abgebildet.

Es ist erkennbar, dass beide EDCs ein dem EIC des Nona (1) folgenden Verlauf aufweisen und dann den Anstieg entsprechend der weiteren Komponenten im Gemisch durchlaufen. Dabei verläuft das EDC<sub>2.037</sub> in den Fraktionen 27-31 flach und weist somit im Gegensatz zum EDC<sub>2.039</sub> keine Zugehörigkeit zum Dodeca auf. Daraus resultiert ein Korrelationskoeffizient von 0.76 bzw. 0.52 der beiden EDCs (2.037 bzw. 2.039 ppm) zum EIC des Nona. Demzufolge wird das *N*-Acetylsignal des GlcNAc-1 $\beta$  bei stärkerer Gewichtung stärker hervorgehoben woraus eine Umkehr der Intensitäten von  $\alpha$  zu  $\beta$  erfolgt.

Eine Verbesserung könnte die Verwendung der Gewichtung durch eine sigmoide Funktion mit sich bringen (Abbildung 31). Bei dieser Funktion erfahren alle hohen Korrelationskoeffizienten eine starke und alle niedrigen in Abhängigkeit von dem gewählten Wendepunkt eine schwache Gewichtung. Dadurch konnte die Problematik der Verschiebung des Maximums von GlcNAc-1 reduziert werden. Es bedarf aber noch weiterer Untersuchungen zur Optimierung. Grundsätzlich ist das Problem der Zuordnung globaler Signale zu niedrig konzentrierten Komponenten eines Gemisches damit aber noch nicht gelöst.



Abbildung 30: Ausschnitt des *N*-Acetylbereichs von GlcNAc-1 mit den EDCs der Signale bei  $\delta$ =2.039 und 2.037 ppm im Vergleich zu den EICs. Beide EDCs zeigen anfänglich einen parallelen Verlauf zu EDC<sub>c</sub> (rot, Nona 1), entsprechen im weiteren Profil aber eher der Summe der EICs und werden deshalb als globale Signale bezeichnet. Erkennbar ist, dass EDC<sub>2.037</sub> (cyan) einen geringeren Anteil des Dodeca beinhaltet (Fr. 27-31) und deshalb einen höheren Korrelationskoeffizienten in der Korrelation zum EIC<sub>c</sub> aufweist als das EDC<sub>2.039</sub> (blau). Deshalb verschiebt sich bei der Gewichtung mit 4. Potenz im Korrelationsspektrum des Nonas (1) das Maximum des Signals vom  $\alpha$ - (2.039 ppm) zum  $\beta$ -Signal (2.037 ppm).

# Grundlage für weitere Optimierungen

Jedes Signal, welches einerseits charakteristisch für eine bestimmte Verbindung ist und andererseits nur bei dieser Verbindung auftritt wird zweifelsfrei auch in der einfachen Extraktion dieser Verbindung zugeordnet. Ein Beispiel dafür ist das Man-4<sup>+</sup> H-1 bei  $\delta$ =4.916 ppm des fucosylierten Nonasaccharids (**3**). Durch die Anwendung von Filtern kann die Spektreninterpretation dann vereinfacht werden, wenn Korrelationskoeffizienten durch die starke Überlappung im Chromatogramm höher als erwartet sind. Dann hat sich die gewichtete Extraktion als vorteilhaft erwiesen. Neben der besseren Unterscheidbarkeit der Signale wird beispielsweise im Vergleich zum *cut off* auch die natürliche Signalform beibehalten. Probleme resultieren aus Signalen, die bei mehreren Komponenten die gleiche chemische Verschiebung aufweisen und deshalb globaler Natur sind. Diese globalen Signale werden umso stärker durch die Korrelation unterdrückt, je geringer der Anteil der Komponente im Gemisch ist.



Abbildung 31: Sigmoide Gewichtung des Korrelationskoeffizienten  $(c_{m,\delta})_s$  als alternative Möglichkeit zur mathematischen Gewichtung des Korrelationskoeffizienten. Dabei erfahren die Korrelationskoeffizienten in der Nähe des Wendepunktes (hier: x=0.5, k=20) eine Umkehrung der Gewichtung, sodass hohe Korrelationskoeffizienten möglichst ähnlich hoch und niedrige Korrelationskoeffizienten ähnlich gering gewichtet werden.

Dieser Umstand kann auch die Integrierbarkeit herabsetzen, wenn stark gewichtete Signale gegenüber globalen Signalen ein ungleich intensiveres Integral aufweisen. Auch die Korrelation nach Kendall oder Spearman hat in diesem Fall keine Verbesserung der Extraktion gebracht, wie es am Beispiel des Undeca **5** in Abbildung 25 diskutiert wurde. Für die Interpretation kann helfen, dass wichtige Signale, wie beispielsweise das H-1 der zusätzlichen α1,3-verknüpften Galactose(n) (Dodeca und Undeca) eindeutig zugeordnet sind und alle fehlenden Signale zumindest im Summenspektrum vertreten sein müssen. Dadurch kann beispielsweise 1,6-verknüpfter von 1,3-verknüpfter Fucose unterschieden werden. Wichtig ist hier der Einbezug der MS-Daten um eine eindeutige Charakterisierung durchführen zu können. Außerdem konnte in Abschnitt 4.2 gezeigt werden, dass auch anhand von Spektren von Substanzgemischen Informationen über die Zusammensetzung erhalten werden können. Demzufolge könnten auch Verbindungen, die aufgrund zu starker Überlappung im Chromatogramm keine vollständige Separierung in zwei unabhängige NMR-Spektren ermöglichen, identifiziert werden.

3DCC konnte hier als Methode am praktischen Beispiel erprobt werden. Durch 3DCC war die Charakterisierung der fünf Hauptkomponenten möglich, indem Informationen aus MS und NMR-Spektroskopie verknüpft wurden und zusammen eine signifikante Verbesserung des Informationsgehalts erfolgte. Damit konnte die Grundlage für weitere Optimierungen gelegt werden.

### 4.4.1.2 S/N-Verbesserung durch Spektrendekonvolution

Mit der Verwendung von 3DCC geht meist eine Signal-zu-Rausch-Verbesserung einher, da die einzelnen Spektren nur für die Korrelation benötigt werden, die eigentliche Extraktion aber anhand der Summe aller aufgenommenen NMR-Daten des Chromatographielaufes erfolgt. Damit geht jedes gemessene Signal in die Berechnung des Korrelationsspektrums ein. Durch die Einführung von Gewichtungsfaktoren können Signale gegenüber dem Rauschen weiter intensiviert werden. Abbildung 32 zeigt den Vergleich von einfach- und gewichtetextrahierten NMR-Spektren des Nona 1 im Vergleich zum NMR-Spektrum der Fraktion 5, bei der der  $\beta$ -Peak des Nona im Chromatogramm maximal ist.



Abbildung 32: Verbesserung des S/N am Beispiel des Nona 1. Gezeigt sind das einfach- und das gewichtetextrahierte Spektrum ( $E_{(m,\delta)}$  bzw.  $E4_{(m,\delta)}$ , es gilt für die Intensität I: I(I<0) wurde gleich 0 gesetzt) im Vergleich zum Spektrum der Fraktion 5, dem Maximum des  $\beta$ -Peaks des Nona im Chromatogramm. Alle Spektren sind mit einem gleitenden Durchschnitt geglättet. Das S/N wird hier durch die Extraktion und die Reduktion des Rauschens um einen Faktor von 140 für die einfache und einen Faktor von 500 für die gewichtete Extraktion verbessert (s. a. Tabelle 6).

Die Verbesserung im S/N ist deutlich erkennbar. Eine Quantifizierung kann durch Berechnung des S/N nach Gleichung 10 vorgenommen werden, ist aber durch die Schwierigkeit, vergleichbare Bedingungen zu finden, in ihren absoluten Werten vorsichtig zu

interpretieren. Die nach Gleichung 10 erhaltenen S/N-Verhältnisse liegen im Schnitt 5-7 % unterhalb der mit TopSpin (Bruker BioSpin) erhaltenen Werte, da bei der Berechnung in TopSpin zuvor eine Magnitudenkalkulation anhand von Real- und Imaginärteil des Spektrums durchgeführt wird. Da für alle Berechnungen immer ausschließlich der Realteil der Spektren verwendet wurde, ist die Vergleichbarkeit des S/N unter Verwendung von Gleichung 10 gegeben.

$$S/N = \frac{\max(I_{NMR})}{2 \cdot \sigma_{Noise}}$$
 Gleichung 10

S/N – Signal-zu-Rausch-Verhältnis, max $(I_{NMR})$  – Intensitätsmaximum des NMR-Signals,  $\sigma_{Noise}$  – Standardabweichung des Rauschbereichs

Tabelle 6 zeigt den Vergleich der erhaltenen NMR-Daten. Für die Komponenten Nona 1, Deca 2 und fucosyliertes Nona 3 wurde das Man-4, H-1 als Signal für die Berechnung des S/N gewählt. Für die beiden  $\alpha$ 1,3-galactosylierten Oligosaccharide (4 und 5) wurde die Intensität des charakteristischen H-1 $\alpha$  der zusätzlichen Galactose zur S/N-Berechnung verwendet.

 Tabelle 6: S/N-Verbesserung der extrahierten Spektren gegenüber einzelnen Fraktionen (F), berechnet

 mit Gleichung 10.

	Signal	S/N einzelner Fraktionen		S/N einzelner Fraktionen $S/N$ von $E_{(m,\delta)}^*$	
Nona 1		Fr. 5	3.9	570	2070
Deca 2	Man-4, H-1	Fr. 16	12	730	9480
Fuc. Nona 3		Fr. 13	6.9	1170	24600
Dodeca 4	Gal-7/7',	Fr. 29	4.6	250	4190
Undeca 5	H-1α	Fr. 21	1.7	95	140

Rauschen: 8.5-9.5 ppm, Prozessierung mit EM, *lb* 1, Glättung (gleitender Durchschnitt), \*es gilt für Korrelationskoeffizienten c und Intensitäten I: c(c<0) = 0, I(I<0)=0

Für die Bewertung der S/N muss berücksichtigt werden, dass durch den Ausschluss negativer Korrelation alle Korrelationskoeffizienten kleiner Null gleich Null gesetzt werden (Spalte 4, *S/N von*  $E_{(m,\delta)}$ ). Außerdem wurden in den meisten Beispielen Intensitäten kleiner Null ebenfalls gleich Null gesetzt, sowie eine Glättung (gleitender Mittelwert, n=5) durchgeführt, wodurch sich eine Verbesserung des S/N ergibt. Die Gegenüberstellung zeigt, dass charakteristische Signale durch die einfache Extraktion und die mathematische Prozessierung um Faktoren von 50-150 verbessert wurden, während durch die Gewichtung des Korrelationskoeffizienten mit 4. Potenz im Schnitt eine Verbesserung in der Größenordnung von  $10^2$ - $10^3$  erreicht wurde. Neben der Dekonvolution von Signalen kann diese Verbesserung des S/N durch mathematische Operationen essentiell für die Auswertbarkeit der Spektren sein.

### 4.4.2 <u>3DCC zur Charakterisierung von massenisobaren Strukturen</u>

3DCC kann auch bei massenisobaren Strukturen eingesetzt werden. Bedingung für die Verwendung von 3DCC sind lediglich geringfügig unterschiedliche Retentionszeiten. Ein weiteres Gemisch G2 vom bovinen Fibrinogen (s. a. Abbildung 36) wurde diesmal unter Zusatz von Ammoniak an PGC chromatographiert. Dadurch wurde die Mutarotation so beschleunigt, dass keine Trennung in  $\alpha$ - und  $\beta$ -Peaks der Glycanstrukturen auftritt. Daraus resultiert ein Sensitivitätsgewinn, da die Substanz nicht mehr über zwei Peaks verteilt ist.

Es wurde ein Bereich des Chromatogramms untersucht, in dem vier massenisobare Strukturen mit der anhand von MS bestimmbaren Zusammensetzung Hex5HexNAc4 auftreten (Nonaisomere 1,7,8 und 9, Charakterisierung s. Abschnitt 4.4.5.3). Durch Fraktionierung resultiert ein Auflösungsverlust des EIC-Verlaufs über die Fraktionen gegenüber dem direkt aus LC-MS erhaltenen EIC. Eine Unterscheidung der einzelnen Isomere im EIC kann anhand der unterschiedlichen Retentionszeiten vorgenommen werden. Dazu wurde das nach Fraktionierung erhaltene rekonstruierte EIC auf Minima untersucht, die den Übergang von einem Chromatographiepeak zum nächsten markieren und damit die Abgrenzung der sogenannten sub-EICs (sEICs) mit gleichem m/z darstellen (Abbildung 33A). Dabei wurden fünf Zeitprofile erhalten, so dass mindestens ein Minimum zufällig auftrat. Aus der Korrelation der NMR-Daten mit sEIC1 resultierte Nona 1 ist und im weiteren Verlauf als 4,4-Isomer bezeichnet wird.

Die weiteren Isomere werden im Folgenden entsprechend der Verknüpfung der terminalen Galactosereste als 4,3-, 3,4- und 3,3-Isomer bezeichnet. Die Spektren dieser Isomere wurden unvollständig erhalten. Eine Verfeinerung der sEICs und somit Identifikation des artifiziellen Minimums kann prinzipiell durch den Vergleich von EDC-Verläufen vorgenommen werden.



Abbildung 33: Korrelation von massenisobaren Komponenten aus Gemisch G2 vom bovinen Fibrinogen. Chromatographische Trennung unter Zusatz von Ammoniak führt dazu, dass  $\alpha$ - und  $\beta$ -Peaks bei gleicher Retentionszeit auftreten. A) EIC bei m/z=821.3038, getrennt in subEICs über die Minima. B) ausgewählte EDCs der chemischen Verschiebungen, die in den NMR-Spektren in C) durch Striche markiert sind. Die vertikalen Linien deuten eine mögliche Neuzuordnung der sEICs anhand der EDCs an. Es ist erkennbar, dass ein Minimum innerhalb von sEIC<sub>2</sub> zufällig war. Jedes gezeigte EDC weist zwei Maxima auf (markiert mit jeweils roten \*), weil es von zwei Isomeren erzeugt wird. Die gezeigten NMR-Spektren entsprechen jeweils den Fraktionen der markierten Maxima.

Beispielsweise ist in Abbildung 33 erkennbar, dass  $sEIC_2$  sich aus zwei zunächst gefundenen Verläufen zusammensetzt. Auch  $sEIC_3$  und  $sEIC_4$  ließen sich durch die EDC-Verläufe verfeinern. Allerdings muss dabei berücksichtigt werden, dass beispielsweise EDCs von globalen Signalen ebenfalls zu mehreren Komponenten gehören können. Gerade weil bei massenisobaren Strukturen häufig sehr ähnliche Strukturmerkmale und demzufolge auch sehr ähnliche chemische Verschiebungen auftreten, ergibt sich allerdings auch bei dieser Fragestellung wieder der Effekt der globalen Signale (s. a. Abschnitt 4.4.1.1).

Jedes in Abbildung 33 gezeigte EDC gehört zu mindestens zwei isomeren Strukturen. Dieses ist in Abbildung 33B und C deutlich erkennbar. Durch die Korrelation wird das Signal dem intensivsten sEIC zugewiesen. Obwohl also beispielsweise im EDC<sub>B</sub> deutlich zwei Maxima erkennbar sind, korreliert es mit einem Korrelationskoeffizienten 0.97 nahezu ausschließlich mit sEIC<sub>1</sub>. Die Komponente, die im Verhältnis am geringsten vertreten ist (Verhältnisse der vier Komponenten sind etwa 17:6:3:1, bestimmt über das Verhältnis der sEICs des LC-MS-Laufs), weist hingegen im Spektrum nahezu keine Signale auf, da jedes Signal dominanter in einem der Spektren der anderen Isomere auftritt.

Eine Verbesserung bringt an dieser Stelle die Verwendung der Korrelationsberechnung nach Kendall oder Spearman, wodurch neben dem Hauptisomer noch zwei der drei Nebenisomere (Korrelation mit sEIC<sub>2</sub>, sEIC<sub>3</sub> und sEIC<sub>4</sub>) identifizierbar sind (s. Abbildung 34). Im Gegensatz zur Dekonvolution nach Pearson'scher Korrelation können hier für das 4,3- und das 3,4-Isomer auch globale Signale extrahiert werden. Für das 3,3-Isomer sind die extrahierten Signale auch hier nicht ausreichend, um die Unterscheidung vornehmen zu können. Die Aufklärung der Strukturen wird in Abschnitt 4.4.5 (Nonaisomere **1**, **7**, **8**, **9**) diskutiert.

An diesem Beispiel wird nochmals deutlich, dass alle benötigten Informationen in den Spektren vorhanden sind, die Extraktion dieser Informationen aber noch nicht optimal funktioniert. Einerseits könnten Optimierungsvarianten auf Basis der in Abschnitt 4.4.1.1 gezeigten Methoden eine Verbesserung mit sich bringen.

Andererseits könnte eine alternative Herangehensweise an die Extraktion, beispielsweise durch Anwendung von Clusteranalysen, das Verwenden sämtlicher vorhandener Informationen ermöglichen. Das hier gezeigte Beispiel ist tatsächlich gut genug getrennt, um aus einzelnen ausgewählten Fraktionen die Zuordnung der NMR-Signale vornehmen zu können (s. Abbildung 33C und Abschnitt 4.4.5 –Aufklärung der Nonaisomere 1,7,8 und 9).



Abbildung 34: Vergleich der Korrelation der drei Nebenisomere nach Kendall, Spearman und Pearson. Die Extraktion der charakteristischen Signale des 4,3- und des 3,4-Isomers ist bei Kendall und Spearman gegeben, wie der Vergleich zu den experimentellen Spektren zeigt. Dadurch ist die Charakterisierung möglich. Die Signale des 3,3-Isomer als Komponente mit dem geringsten Anteil am Gemisch werden unvollständig erhalten, sodass keine eindeutige Zuordnung anhand der Extraktionsspektren möglich ist.

Es eignet sich insofern als Modellsystem für weitere Extraktionsexperimente, weil durch den Abgleich mit den zugeordneten Spektren gezielt die Effizienz der Methode überprüft werden kann.

Die Korrelation nach Pearson ist also dann vorteilhaft, wenn Substanzen im Chromatogramm starker Überlappung ausgesetzt sind, während die Korrelation nach Kendall oder Spearman vorteilhaft für die Extraktion globaler Signale einer niedrig konzentrierten Substanz sein kann.

# 4.4.3 Automatische Spektrenerkennung der extrahierten NMR-Daten

Die automatische Erkennung des extrahierten Spektrums von Deca 2, Hauptkomponente im Gemisch G1, ist unter den gleichen Parametern möglich wie zuvor beschrieben (Abschnitte 4.1.3 und 4.3.3.2). Allerdings kann die Erkennung von zwei Faktoren kompromittiert werden. Zum einen können Substanzen nicht erkannt werden, wenn die Strukturen nicht oder nur unvollständig in *sugabase* eingetragen sind. Zum anderen benötigt das aktuell verwendete Skript vollständige Signalsätze im Bereich der *structural reporter groups*. Aus den beiden vorgenannten Gründen ist die Spektrenerkennung bezüglich der meisten Komponenten der Gemische G1 und G2 nicht erfolgreich. Eine Verfeinerung der Suche sollte eine Verbesserung der Identifikation ermöglichen. Grundsätzlich zeigt aber das positive Beispiel des Decasaccharids (2), dass Korrelationsspektren für die automatische Erkennung genauso nutzbar sind, wie direkt von reinen Komponenten erhaltene NMR-Daten.

### 4.4.4 <u>Untersuchungen zum Abspaltungsprozess vom bovinen Fibrinogen</u>

Das in Abschnitt 4.4.1 verwendete Glycan-Gemisch **G1** wurde aus dem bovinen Fibrinogen durch mehrere enzymatische Schritte gewonnen. Ein Vergleich der Spektren nach den einzelnen Schritten des Abspaltungsprozesses zeigt Informationen über die sialylierten Spezies (in Abbildung 35 von einem vergleichbaren Gemisch **G3**).<sup>94</sup>

Das bovine Fibrinogen wurde dazu zunächst durch einen tryptischen Verdau in Peptide und Glycopeptide aufgespalten. Das nach anschließender Dialyse aufgenommene Spektrum A zeigt, dass die Antenne am 3-Arm quantitativ sialyliert ist (Man-4 bei  $\delta$ =5.134 ppm), während der 6-Arm etwa zu 50 % sialyliert vorliegt und demzufolge zwei Signale der Man-4<sup>+</sup> bei  $\delta$ =4.926 ppm und  $\delta$ =4.943 ppm auftreten. Durch den PNGaseF-Verdau erfolgt die Abspaltung der Glycane von den Peptiden, wodurch keine Änderung im Sialylierungsmuster erhalten wird (Abbildung 35B). Der anschließende Neuraminidase-Verdau zur Abspaltung der terminalen Sialinsäurereste verlief im gezeigten Beispiel etwa zu 50%, was daran erkennbar ist, dass das Verhältnis des Man-4 H-1 bei  $\delta$ =5.120 ppm etwa 1:1 ist, während die Signale der H-1 von Man-4<sup>+</sup> etwa 1:3 vorliegen. Erste Ergebnisse deuten darauf hin, dass die als sialylierte Spezies verbleibenden Glycanstrukturen verknüpft an eine definierte Peptidsequenz vorliegen und in Form dieses Konjugats nicht von der PNGaseF und der Neuraminiadase erkannt werden (*Henning N. Behnken, unveröffentlichte Ergebnisse*). Durch die anschließende Reinigung mittels gemischter Ionentauschersäule wurden in Spektrum D

nur noch Strukturen erhalten, die desialyliert vorliegen. Erkennbar ist, dass das Signal bei  $\delta$ =5.15 ppm (markiert mit einem Pfeil) gegenüber den Spektren A-C deutlich angereichert ist.



Abbildung 35: Spektren von verschiedenen Schritten des Abspaltungsprozess einer Probe vom bovinen Fibrinogen. Markiert sind die anomeren Signale der Man-4 und Man-4' an denen sich der Sialylierungsgrad ablesen lässt. Durch den nur partiell verlaufenen Neuraminidase-Verdau zur Abspaltung der terminalen Sialinsäuren ergibt sich eine Anreicherung der nicht-sialylierten Glycane wie z.B. Dodeca 4 und Undeca 5 zu denen das mit Pfeil markierte Signal bei  $\delta$ =5.15 ppm gehört. Das Glycangemisch ist gegenüber Spektrum C von Peptiden abgetrennt worden.

Das in Abschnitt 4.4.1 verwendete Gemisch **G1** wurde nach gleichem Verfahren erhalten, allerdings ist der Neuraminidase-Verdau nur zu etwa 27 % erfolgreich verlaufen, weshalb die Anreicherung von anfänglich nicht-sialylierten Strukturen noch dominanter ist. Die Überrepräsentation der  $\alpha$ 1,3-galactosylierten Spezies (Dodeca 4 und Undeca 5) vereinfachte deren Identifikation und Charakterisierung.



# 4.4.5 Identifikation bisher unbekannter Glycanstrukturen im bovinen Fibrinogen



LC-MS: Gradient 2

t (min)

Abbildung 36: Gezeigt sind die *base peak chromatograms* (BPCs) zweier Chromatographieläufe der vergleichbaren Gemische 2 und 4. Lauf 1 zeigt die Trennung bei Verwendung eines gegenüber Lauf 2 steileren Gradienten im Bereich von t>65 min. Lauf 2 wurde zur Charakterisierung der Undecaisomere (5 und 6), sowie der triantennären Struktur 10 verwendet.

In weiteren Untersuchungen von kommerziell erhältlichem bovinen Fibrinogen (Gemisch G2) konnten neue Strukturen nachgewiesen werden. Zum einen handelt es sich dabei um biantennäre Strukturen, die um  $\alpha$ 1,3-Galactosereste (Dodeca 4 und Undeca 5, sowie dessen Isomer 6) erweitert sind. Auf das Auftreten dieser Strukturen im bovinen Fibrinogen deuteten bisher nur biologische Tests durch Bindung von *anti*-Gal Antikörpern hin.<sup>23</sup> Die genauen Strukturen konnten durch NMR-Spektroskopie und MS nachgewiesen werden. Zum anderen konnten 3 Isomere des Nonasaccharids 1, die von Altmann *et al.*<sup>47</sup> durch massenspektrometrische Untersuchungen und Vergleiche mit enzymatisch aufgebauten Strukturen gezeigt wurden, erstmals NMR-spektroskopisch belegt werden. Außerdem wurde eine triantennäre Struktur (10) nachgewiesen, die bisher nicht literaturbekannt ist. Neue Erkenntnisse ergeben sich daraus, dass die Strukturen entweder unbekannt für das bovine Fibrinogen sind (Dodeca 4, Undeca 5 und Undecaisomer 6), es für sie keine Einträge in NMR-Datenbanken gibt (Undecaisomer 6, Nonaisomere 7,8 und 9) oder keine Daten literaturbekannt sind (Triantennäre Struktur 10).

Abbildung 36 zeigt Ausschnitte aus zwei chromatographischen Läufen, bei denen durch Wahl des Gradienten einerseits die biantennären Strukturen mit niedrigeren Retentionszeiten, andererseits die  $\alpha$ 1,3-galactosylierten biantennären Strukturen, sowie eine triantennäre Struktur aufgeklärt werden konnten.

#### 4.4.5.1 Aufklärung des Dodecasaccharids 4

In Untersuchungen der Glycangemische (**G1-G4**) von zwei von unterschiedlichen Herstellern erworbenen Chargen des bovinen Fibrinogens konnte die Struktur des Dodecasaccharids (**4**) nachgewiesen und um ein *structural reporter group*-Signal erweitert werden. Durch die Verwendung höherer Feldstärke (700 MHz) konnten zudem die anomeren Signale der terminalen  $\alpha$ 1,3-Gal (7,7<sup>c</sup>), sowie der benachbarten  $\beta$ 1,4-Gal (6,6<sup>c</sup>)-Reste separiert werden, die bei der bisherigen Dokumentation und dem Datenbankeintrag in *sugabase* aufgrund geringerer Auflösung als ein gemeinsames Signal angegeben sind.<sup>91</sup> Dieses wird insbesondere dann relevant, wenn bei mono- $\alpha$ 1,3-Galactosylierung der Undecaisomere **5** und **6** die Antennen unterschieden werden sollen (s. a. Abbildung 38). Abbildung 37 zeigt die Anwendung von selektiven TOCSY-Experimenten zur Aufklärung der Struktur des Dodecasaccharids. Die Spektren wurden von einer Probe eines Glycangemisches vom bovinen Fibrinogen aufgenommen, sodass in den Spektren Signale unterschiedlicher Strukturen auftreten können. Trotzdem ist das  $\alpha$ 1,3-Gal Epitop deutlich erkennbar. Die mit Pfeilen markierten Regionen wurden selektiv angeregt, sodass nur Signale beobachtet werden, die im gleichen Spinsystem auftreten wie das angeregte Signal und so den TOCSY Transfer erfahren. Spektrum E zeigt als Vergleich das <sup>1</sup>H-Spektrum der Probe.



Abbildung 37: Selektive TOCSY-Spektren einer Probe eines Oligosaccharidgemisches. Untersuchung der  $\alpha$ 1,3-verknüpften Galactosereste. E: Vergleichsspektrum der Probe. A-D: Selektive Anregung bei der mit rotem Pfeil markierten Resonanz. Spektren A und B zeigen, dass die Signale bei  $\delta$ =5.15 und 4.02 ppm einem gemeinsamen Spinsystem zuzuordnen sind (H-1 und H-4 der Gal-7/7<sup>4</sup>). Spektren C und D zeigen, dass die H-1 Protonen bei  $\delta$ =4.54 ppm und das H-4 bei  $\delta$ =4.18 ppm den Gal-6/6<sup>4</sup> als einem gemeinsamen Spinsystem zuzuordnen sind.

Bei Spektrum A wurde im Bereich der H-1 Protonen des  $\alpha$ 1,3-Gal bei  $\delta$ =5.15 ppm angeregt, wovon das H-1 $\alpha$  des GlcNAc-1, sowie das H-1 der Man-4 ebenfalls betroffen sind. Durch den TOCSY-Transfer sind ebenfalls das H-2 der Man-4 und als neues *structural reporter group*- Signal das H-4 der  $\alpha$ -Gal zu erkennen, welches anhand des Kopplungsmusters zuordenbar ist. Spektrum B bestätigt bei Anregung bei  $\delta$ =4.02 ppm den TOCSY Transfer zurück zum H-1 $\alpha$  der Gal-7/7<sup>+</sup>, außerdem wird ein weiteres GlcNAc-1 Signal angeregt, weshalb ebenfalls das GlcNAc-1 H-1 $\alpha$  erkennbar ist. Anregung bei  $\delta$ =4.19 ppm (Abbildung 37C) resultiert in der Detektion der beiden anomeren Dubletts von Gal-6 und -6<sup>+</sup>, welche im Gemisch nicht eindeutig erkennbar waren und die um 0.1 ppm gegenüber den H-1 der Gal-6/6<sup>+</sup> des Nonabzw. Decasaccharids (1 und 2) verschoben sind, welche keine zusätzlichen  $\alpha$ 1,3-Galactosereste tragen. Daraus ergibt sich die chemische Verschiebung der H-4 Protonen von Gal-6 und -6<sup>+</sup> bei der angeregten Resonanz. Spektrum D belegt diesen Transfer von den H-1 zu den H-4. Der TOCSY Transfer ausgehend von den H-5 Protonen von Gal-7 und -7<sup>+</sup> ebenfalls bei  $\delta$ =4.19 ppm wird aufgrund der kleinen Kopplungskonstante zum H-4 unterbrochen. Das Multiplett bei  $\delta$ =4.19 ppm sowie das Integral bestätigen aber die chemische Verschiebung für die H-5 Protonen, die im *sugabase*-Eintrag notiert sind.

## 4.4.5.2 Aufklärung der Undecaisomere 5 und 6



Beide Undecaisomere können durch Vergleich zum Dodeca 4 eindeutig zugeordnet werden.

Abbildung 38: Vergleich der *structural reporter group*-Signale der  $\alpha$ 1,3-galactosylierten Strukturen 4-6 (Dodeca, Undeca 5, Undecaisomer 6). Markiert mit gestrichelter Linie sind Signale, die charakteristisch für die am 6-Arm verknüpfte  $\alpha$ 1,3-Gal sind, während mit durchgezogener Linie die 3-Arm Signale hervorgehoben sind. \* - Signal stammt von einer triantennären Struktur

Dafür ist die Auflösungsverbesserung durch Messung an einem 700 MHz Spektrometer im Vergleich zum 500 MHz Spektrometer notwendig. Beim Undeca **5** ist genau ein Datensatz an anomeren Protonen von Gal-7<sup>•</sup> und Gal-6<sup>•</sup> in gleicher Verschiebung wie im Dodeca (**4**) vorhanden ( $\delta$ =5.147 ppm und  $\delta$ =4.544 ppm), während das anomere Proton der ungalactosylierten Gal-6 im hohen Feld bei  $\delta$ =4.467 ppm, entsprechend dem H-1 der Gal-6 im Decasaccharid (**2**), zu erkennen ist. Analog dazu ist der zweite Datensatz, der also zu den H-1 der Gal-7 und Gal-6 des Undecaisomers **6** gehört, zu erkennen, während das anomere Proton der Gal-6<sup>•</sup> hochfeldverschoben bei  $\delta$ =4.473 ppm auftritt. Dieses ist flankiert von einem weiteren Signal (Abbildung 38, Asterisk-Markierung) welches sich aber aus einer partiellen Überlagerung mit einer triantennären Struktur ergibt.

### 4.4.5.3 Aufklärung der Nonaisomere 1,7,8 und 9

Durch die Kombination von MS- und NMR-Daten können auch Strukturen der in Abbildung 39 gezeigten Nonaisomere (Strukturen 1, 7, 8 und 9) eindeutig charakterisiert werden. Das untere NMR-Spektrum zeigt das bekannte Nonasaccharid, bei dem die beiden terminalen Galactosereste 1-4 verknüpft sind (4,4-Isomer). Die weiteren drei Spektren wurden von massenisobaren Strukturen mit leicht veränderten Retentionszeiten aufgenommen (s. Abbildung 36 und Abschnitt 4.4.2). Struktur 7 weist im Spektrum bei  $\delta$ =4.445 ppm ein gegenüber 1 um 0.020 ppm verschobenes Dublett auf, welches dem H-1 der Gal-6 entspricht (Abbildung 39, Pfeil a). Gleichzeitig erscheint das Signal des H-1 der Gal-6', die sich am 6-Arm befindet, an gleicher Stelle wie das des 4,4-Isomers bei  $\delta$ =4.473 ppm. Auch das H-1 des GlcNAc-5 erfährt eine Verschiebung, diese ist charakteristisch für die Verknüpfung des nachfolgenden Restes. Allerdings kann hier keine Unterscheidung von GlcNAc-5 und GlcNAc-5' vorgenommen werden, da beide bei gleicher Verknüpfung nahezu identische chemische Verschiebungen aufweisen. Außerdem verschiebt sich das Signal der *N*-Acetylgruppe von GlcNAc-5 von  $\delta$ =2.051 ppm zu  $\delta$ = 2.044 ppm. Daraus ergibt sich die zwingende Interpretation, dass Struktur 7 im Gegensatz zu 1 eine 1,3-verknüpfte Galactose am 3-Arm aufweist also dem 4,3-Isomer entspricht. Analog verschiebt sich bei Struktur 8 das Dublett des H-1 von Gal-6' (zu δ=4.453 ppm), sowie die GlcNAc-5'-Signale, während das Signal des anomeren Protons von Gal-6 bei  $\delta$ =4.467 ppm verbleibt (Abbildung 39, Pfeil b). Außerdem reagiert das H-1 der Man-4' empfindlicher auf die veränderte Verknüpfung des terminalen Restes, es ist eine Hochfeld-Verschiebung von 0.010 ppm beobachtbar. Deshalb kann Struktur 8 als 3,4-Isomer charakterisiert werden, bei dem die Gal-6' am 6-Arm 1-3 verknüpft ist, während die Gal-6 am 3-Arm eine 1-4 Verknüpfung aufweist.



Abbildung 39: Charakterisierung der Nonaisomere 1, 7-9. Spektrum a zeigt das in *sugabase* charakterisierte 4,4-Isomer. Die neue Zuordnung der isomeren Strukturen 7-9, die sich in der Verknüpfung der terminalen Galactosereste unterscheiden, ergibt sich aus der Verschiebung einzelner Signale. Pfeil a markiert die Verschiebung des Gal-6 H-1 zu  $\delta$ =4.445 ppm, woraus sich die 1,3-Verknüpfung des Gal-6 erschließen lässt. Analog zeigt Pfeil b die Verschiebung des Gal-6' im Spektrum c. Beide Signale sind im Spektrum d des 3,3-Isomers hochfeldverschoben. Auch die Verschiebung der *N*-Acetylsignale, sowie des Man-4' H-1 (vertikale Linien) zeigen signifikante Veränderungen der Spektren.

Struktur 9 vereint die vorher beschriebenen charakteristischen Veränderungen von 7 und 8 und kann mit beiden verschobenen Gal H-1 Signalen, beiden verschobenen GlcNAc-5/5'-Signalen sowie dem verschobenen Man-4' H-1 eindeutig als 3,3-Isomer identifiziert werden. Außerdem kann für das literaturbekannte 4,4-Isomer (1) die Unterscheidung der *N*-Acetylsignale von GlcNAc-5 (2.051 ppm) und GlcNAc-5' (2.047 ppm) verifiziert werden, welche im *sugabase* Eintrag und in der Originalliteratur nicht eindeutig zugewiesen sind.<sup>90</sup>

#### 4.4.5.4 Aufklärung der triantennären Struktur 10

Die mono- $\alpha$ 1,3-galactosylierte triantennäre Struktur **10** mit einer Retentionszeit von 73.1 min und einem *m/z* von 1166.44 (M+H+NH<sub>4</sub>)<sup>2+</sup> kann durch 3DCC eindeutig charakterisiert werden (s. Abbildung 41). Sie ist bis jetzt weder in *sugabase* dokumentiert, noch ist sie als Teil der Glycanstrukturen des bovinen Fibrinogen bekannt. In *glycosciences.de* und GlycomeDB ist sie zwar verzeichnet, aber es sind ebenfalls keinerlei Daten hinterlegt. Durch die Informationen aus den MS-Daten und den Vergleich mit ähnlichen Strukturen kann eine eindeutige Zuordnung vorgenommen werden. Zunächst ist aufgrund der starken Überlappung mit dem Dodeca im Chromatogramm die Anwendung von 3DCC nötig.



Abbildung 40: Vergleich TIC und TDC. Da MS über die Zeit eine bessere Auflösung als die fraktionierten NMR-Daten hat, werden die Daten des TIC auf die gleiche Anzahl an Datenpunkten wie im TDC reduziert, daraus ergibt sich das  $TIC_R$ . Durch den Abgleich von TDC und TIC kann eine Überprüfung der korrekten Zeitberechnung der Fraktionen erfolgen.

Werden die MS-Daten in besserer zeitlicher Auflösung aufgenommen als durch die Fraktionierung erreichbar ist, so muss die Anzahl der Datenpunkte der MS-Daten reduziert werden, um eine Korrelation mit den NMR-Daten zu ermöglichen. Ein Abgleich des reduzierten TICs (TIC<sub>R</sub>) mit dem *total delta chromatogramm* (TDC), welches der Summe aller NMR-Signale jeder Fraktion, aufgetragen über die Zeit entspricht, kann zur Überprüfung der korrekten zeitlichen Zuordnung benutzt werden und um gegebenenfalls zeitliche Verschiebungen zwischen MS- und NMR-Daten korrigieren zu können (Abbildung 40).



Abbildung 41: Charakterisierung der mono- $\alpha$ 1,3-galactosylierten triantennären Struktur 10. A: EICs von Dodeca (4) und der triantennären Struktur 10 als Hauptkomponenten im Bereich von 72-76 min des in Abbildung 36 gezeigten LC-Laufs von G4, sowie drei beispielhafte EDCs mit Korrelationskoeffizienten. B: mittels 3DCC extrahiertes NMR-Spektrum von m/z = 1166.44 (Triant. 10). EDC<sub>a</sub> entspricht dem für triantennäre Strukturen charakteristischen *N*-Acetylsignal bei  $\delta$ =2.08 ppm des GlcNAc-T5, der Verlauf korreliert stark mit EIC<sub>10</sub>, nicht aber mit dem des Dodeca. EDC<sub>b</sub> entspricht dem H-1 des  $\alpha$ 1,3-Gal bei  $\delta$ =5.15 und gehört somit zu beiden Strukturen und korreliert mit beiden EICs. EDC<sub>c</sub> entspricht dem Man-3 H-2 Signal bei  $\delta$ =4.25 ppm, welches ausschließlich zum biantennären Dodeca (4) gehört. Aufgrund des partiell ähnlichen Verlaufs zum EIC<sub>10</sub> erscheint es als Restsignal in Spektrum B. Die Verzweigung am 3-Arm ergibt sich aus der Position des H-1 der Man-4' bei  $\delta$ =4.925 ppm, sowie der Verschiebung des H-2 von Man-3 von 4.25 ppm zu höherem Feld. Die Position des  $\alpha$ 1,3-Gal-Restes am 6-Arm ergibt sich aus den chemischen Verschiebungen der H-1 von Gal-6/T6 bei  $\delta$ =4.463 und  $\delta$ =4.467, sowie der galactosylierten Gal-6' bei  $\delta$ =4.542 ppm. \*Asterisk-Markierung: H-2 von Man-4 oder Man-3.

Außerdem lässt sich durch den Vergleich von TDC und TIC<sub>R</sub> schnell Aufschluss darüber erhalten, ob Verbindungen im MS unterschiedliches Ionisierungsverhalten aufweisen und deshalb die Intensität im MS nicht dem Anteil einer Komponente am Gemisch entspricht. Abbildung 40 zeigt, dass eine gute Übereinstimmung von TIC<sub>R</sub> und TDC besteht, die sich durch einen Korrelationskoeffizienten von 0.95 bestätigt.

Abbildung 41 zeigt die Charakterisierung der triantennären Struktur **10**. Die Identifikation der Verzweigungsposition der dritten Antenne am 3-Arm der Struktur kann anhand der chemischen Verschiebungen von Man-4<sup>•</sup> H-1, sowie anhand der Hochfeldverschiebung des H-2 von Man-3 erfolgen. Das H-1 der Man-4<sup>•</sup> mit einer Verschiebung von  $\delta$ =4.924 ppm schließt die Verzweigung am 6-Arm aus, da es ansonsten hochfeldverschoben zu  $\delta$ <4.90 ppm aufträte. Das Korrelationsspektrum zeigt weiterhin, dass das Signal der H-2 der Man-3 von  $\delta$ =4.250 ppm hochfeldverschoben wird. Diese Hochfeldverschiebung des Man-3 H-2 ist ebenfalls charakteristisch für die benannte Verzweigung am 3-Arm. Das Signal bei  $\delta$ =4.21 ppm, welches eindeutig korreliert, entspricht entweder dem H-2 von Man-3 oder Man-4. Die Überlagerung mit den H-5 von Gal-7<sup>•</sup> und H-4 von Gal-6<sup>•</sup> verhindert die Identifikation des Kopplungsmusters und des zweiten H-2 Signals (s. Abbildung 41, Asterisk-Markierung).

Das *N*-Acetylsignal des GlcNAc-T5 bei  $\delta$ =2.078 ppm ist charakteristisch für den zusätzlichen Arm. T bezeichnet hier den bei der Triantenne zusätzlich auftretenden Arm, der an der Man-4 1,4-verknüpft auftritt. Die Identifikation als 4-Verknüpfung basiert auf Daten von ähnlichen Strukturen, allerdings existieren keine Daten von Strukturen mit 1,6-verknüpftem Arm am 3-Mannosearm. Die Position des  $\alpha$ 1,3-Gal-Restes ergibt sich aus den chemischen Verschiebungen der anomeren Protonen der Gal-6 und Gal-T6 bei  $\delta$ =4.463 ppm und  $\delta$ =4.468 ppm, sowie dem durch die  $\alpha$ 1,3-Galactosylierung tieffeldverschobenen Signal der Gal-6' bei  $\delta$ =4.542 ppm. Dieses träte ungalactosyliert bei  $\delta$ =4.473 ppm auf. Somit kann durch die Information aus MS und den Vergleich zu Strukturen wie dem Deca (**2**) und dem Dodeca (**4**), sowie in *sugabase* auftretenden ähnlichen tri- und tetraantennären Strukturen eine eindeutige Charakterisierung vorgenommen werden.

# 5 Zusammenfassung

Als häufigste und komplexeste Modifikation von Proteinen haben Glycane Einfluss auf viele zelluläre Prozesse und sind beispielsweise an der Entstehung von Krebs beteiligt.<sup>19,102,103</sup> Die systematische Charakterisierung des Glycoms ist aufgrund limitierter Substanzmengen, der vielfältigen Verknüpfungsmöglichkeiten und dem verzweigten Aufbau aufwendig und schwierig.<sup>104</sup> Dabei ist eine effiziente und präzise Analytikstrategie notwendig, um die Funktion der Glycane besser verstehen zu können.

Deshalb wurden im Rahmen dieser Arbeit Methoden zur einfachen und schnellen Charakterisierung von Glycanen entwickelt. Zum einen wurde die Möglichkeit einer automatischen Spektrenerkennung durch Abgleich mit berechneten NMR-Spektren überprüft. Zum anderen wurde das Konzept der dreidimensionalen Kreuzkorrelation (3DCC) entwickelt, welches durch die gemeinsame Interpretation von MS- und NMR-Daten entlang einer zeitlichen Domäne der LC selbst bei unvollständig getrennten Substanzen eine Spektrenextraktion erlaubt.

Für eine automatische Spektrenerkennung wurden im Rahmen dieser Arbeit die tabellarisch gelisteten chemischen Verschiebungen von 523 Einträgen von der Datenbank *sugabase* zur Berechnung von NMR-Spektren verwendet. Dazu wurden Kopplungsmuster und relative Intensitäten einbezogen, um die Spektren zu berechnen und vergleichbar machen. Anhand dieses Datensatzes konnte das experimentelle Spektrum eines komplex-Typ Decasaccharids eindeutig der korrekten Struktur zugeordnet werden. Somit war eine automatisierte Erkennung der Daten möglich, ohne auf die manuelle Interpretation der Spektren angewiesen zu sein. Weiterhin konnte gezeigt werden, dass auch NMR-Spektren von Gemischen unter bestimmten Bedingungen eine Zuordnung der einzelnen Komponenten erlauben. Ein relevanter Faktor für eine zuverlässige Zuordnung stellt hier das Signal-zu-Rausch-Verhältnis der Spektren dar.

3DCC konnte als geeignete Methode zur Charakterisierung von Glycanen in komplexen Gemischen vorgestellt werden. Dabei wurden MS- und NMR-Daten über die gemeinsame Zeitdomäne, die auch aus einer unvollständigen chromatographischen Trennung erhältlich ist, mathematisch korreliert und damit zur Extraktion von NMR-Daten der reinen Komponenten genutzt. Nachdem das Prinzip zunächst mathematisch belegt werden konnte, zeigte sich am Beispiel eines simulierten Chromatographielaufes von zwei biantennären komplex-Typ Glycanen die erfolgreiche Spektrendekonvolution. Anhand dieser extrahierten Spektren konnte eine eindeutige Zuordnung der einzelnen Saccharidstrukturen vorgenommen werden, die sich lediglich durch eine Fucoseeinheit unterschieden. Die Ausweitung auf ein komplexes Gemisch vom bovinen Fibrinogen erlaubte anschließend die Charakterisierung der fünf Hauptkomponenten, die im Chromatogramm starke Überlappungen ihrer Peaks zeigten. Diese Komponenten waren biantennäre Strukturen, die sich durch Fucose-, sowie terminale Galactoseeinheiten unterschieden. Für sehr starke Überlappungen wurden verschiedene Optimierungsmethoden überprüft, um eine verbesserte Extraktion zu erreichen. Durch die Gewichtung des Korrelationskoeffizienten mit der 4. Potenz konnte die Unterscheidung von Signalen verschärft werden, sodass beispielsweise bei einer biantennären Komponente des Gemisches die Identifikation der Verknüpfung des einzigen Galactoserestes am 3-Arm erreicht werden konnte. Außerdem wurde gezeigt, dass durch 3DCC eine Verbesserung des S/N sowie der Auflösung im Spektrum erreichbar ist. Weiterhin wurde diskutiert, dass Signale, die zu mehreren Komponenten gehören und die gleichen chemischen Verschiebungen aufweisen und deshalb als globale Signale bezeichnet werden, problematisch für die Spektrenextraktion sein können. Hierfür wurden neben der Korrelation nach Pearson die Rangkorrelation nach Spearman und Kendall überprüft, die eine Verbesserung der Extraktion von Signalen von Komponenten mit geringem Anteil am Gemisch ergeben können. Schließlich konnte gezeigt werden, dass bei Extraktion der Hauptkomponente des analysierten Gemisches, sowie bei den dekonvulierten Spektren des simulierten Chromatographielaufes die automatische Spektrenidentifikation durch Abgleich mit den anhand von sugabase-Einträgen berechneten Spektren erfolgreich durchgeführt werden konnte.

In kommerziell erhältlichen Proben des Fibrinogens konnten außerdem unbekannte Strukturen mit zusätzlichem  $\alpha 1,3$ -verknüpften Galactoseresten nachgewiesen werden, die Teil eines für den Menschen immunogenen Epitops darstellen.<sup>24</sup> Diese Strukturen sind bisher nicht in der Literatur als Teil der Glycane des bovinen Fibrinogens publiziert. Weiterhin konnten erstmalig NMR-Daten von vier Strukturen dokumentiert werden. Bei drei biantennären Glycanen konnte die variierender Verknüpfung der terminalen Galactosereste gezeigt werden, die entweder am 3- oder am 6-Arm oder an beiden 1,3- statt 1,4-Verknüpfungen aufwiesen. Die Struktur einer triantennären Verbindung konnte erst durch Anwendung von 3DCC aufgeklärt werden, wodurch die Verzweigung der 3. Antenne am 3-Arm und die Position des  $\alpha 1,3$ -Galactoserestes am 6-Arm identifiziert wurde.

Die hier erhaltenen Ergebnisse dienen als Grundlage für eine effiziente, konsequente und automatisierbare Analytikstrategie zur Charakterisierung von Glycanstrukturen, die die Vorteile von MS und NMR-Spektroskopie vereint, um zuverlässige und schnelle Strukturinformationen zu erhalten. Die Methode der 3D-Kreuzkorrelation kann weitere analytische Verfahren wie IR und MS/MS einschließen und für jede Substanzklasse, die chromatographischen Trennproblemen unterworfen ist, angewendet werden.

# 6 <u>Summary</u>

Glycan structures are the most common and most complex modifications of proteins and contribute to pathological processes such as cancer.<sup>19,102,103</sup> A systematic characterization of the glycome is tedious and complicated as glycan structures are often of limited availability and exhibit complex structural properties, such as the occurrence of diverse linkage isomers and branching. In order to understand their biological function, an efficient approach to a fast and reliable characterization of glycan structures is necessary.

Therefore, the scope of this study is to develop analytical strategies for an easy and unambiguous identification of glycans. First of all, a method to automatically identify NMR spectra by comparison of experimental and calculated NMR data is described. Secondly, the development of 3DCC - three dimensional cross correlation - is discussed. This method integrates LC, MS and NMR to extract pure NMR spectra even from incompletely separated structures.

For the automatic recognition of spectra, 523 entries of the database *sugabase* were transformed from chemical shift tables to NMR spectra. To every proton species a relative intensity and a correspondent coupling pattern was assigned. The mathematical comparison of these 523 spectra with the experimental spectrum of a complex *N*-type decasaccharide yielded the unequivocal identification of the correct structure. Therefore, the interpretation of the spectra. Furthermore, identification of components in spectra of unseparated mixtures could be demonstrated, although some limitations, such as restrictions due to low signal-to-noise ratio, were documented.

In order to overcome problems of signal overlap in spectra of mixtures, 3DCC was developed to extract pure NMR spectra. This method was proven powerful for obtaining data hidden in chromatographic profiles of poorly separated components. NMR and MS data were mathematically correlated along a common time dimension provided through LC. Correlation data were used for the deconvolution of the sum spectrum into subspectra of the contributing structures. Proof of principle was established in the first place with a mathematical model and was subsequently verified for a constructed chromatographic run of two biantennary oligosaccharides, which differed only by one fucose residue. 3DCC yielded pure NMR spectra for both structures without interfering signals of the other component. Thus, the unambiguous assignment of both glycans was realized. Further expansion of the method to a

mixture of glycans, obtained by enzymatic digestion of bovine fibrinogen, allowed the characterization of the five main components. These five structures were all biantennary and differed in presence of terminal galactose and fucose residues. The components exhibited strong overlap in the chromatographic profile. Therefore, several strategies to optimize the extraction of spectra with severe overlap were discussed. For example, the weighting of the correlation by raising correlation coefficients to the power of four yielded the definite assignment of a biantennary structure with only one galactose residue. The position of this residue was confirmed to be at the 3-man branch. Furthermore, 3DCC can yield enhanced signal-to-noise ratios and enhanced resolution of spectra. Some problems arose from the occurrence of so called global signals, which are identical signals for several of the components in mixtures. Thus, calculation of the Pearson correlation coefficient was compared to the use of rank correlation according to Kendall or Spearman.

Finally, combination of both analytical strategies yielded the automatic assignment of the deconvolved spectra to a structure by comparison to the calculated NMR spectra. This was proven to be successful for both spectra of the simulated chromatographic run as well as for the main component of the mixture discussed above.

Additionally, the studies revealed the occurrence of  $\alpha 1,3$ -galactosylated glycan species in commercially available samples of bovine fibrinogen. This is of particular interest as the motif is part of an epitope which is immunogenic to humans. These structures have so far not been documented in literature. Also, NMR data of four glycan structures were documented for the first time. Three of these structures show varying linkage position of the terminal galactose residues, which were either 1,3- or 1,4-linked. The fourth component is a triantennary structure, which was identified by means of 3DCC. The position of the third antenna at the 3-man branch and the position of the additional  $\alpha 1,3$ -galactose residue at the 6-man branch were assigned from the extracted NMR spectrum.

The results presented are the basis for an efficient and consequent analytical strategy for automatic and reliable characterization of glycan structures. 3DCC combines the advantages of MS and NMR for fast and reliable assignments. However, 3DCC is not limited to glycan structures, but can be used for all classes of molecules that are difficult to separate chromatographically. Also, it could prove to be powerful if combined with other methods such as IR and MS/MS.

# 7 <u>Experimenteller Teil</u>

# 7.1 Verwendete Geräte, Software und Chemikalien

# Geräte und Software

NMR-Spektrometer	DRX500			
1	5 mm TXI-Probenkopf			
	Avance 700			
	Probenköpfe: 5 mm TXI-Cryoprobe, 5 mm TXI-Probenkopf			
	AvanceIII 500			
	5 mm TCI- <i>Cryoprobe</i>			
	(Bruker BioSpin GmbH)			
LC-MS	LC-ESI-Q/TOF maXis (Bruker Daltonics)			
	UHPLC: UltiMate 3000 (Dionex)			
	LC-ESI-TOF MS 6224			
	LC 1200 (Agilent Technologies)			
PC	Optiplex 980, Intel Core i5, 2.66 GHz, 8 GB RAM (Dell)			
Software	matlab, Version 7.11.0.584 (R2010b), 64bit, Mathworks, Inc.			
	<i>TopSpin</i> , Versionen 2.1 (pl 0) und 3.0 (pl 3), Bruker Biospin GmbH			

# **Chemikalien**

Deutero	Deuteriumoxid, 99.9 % / 99.98 %

Die Gemische G1-G4 wurden freundlicherweise zur Verfügung gestellt (s. 7.2.2 und 7.2.3).

Die Proben der Substanzen I-XIII, sowie die Gemische A-D wurden freundlicherweise von der Hexal AG (Dr. Seidl) zur Verfügung gestellt. Bezogen wurden die Substanzen von TheraProteins.

# 7.2 Probenpräparation

# 7.2.1 Präparation des simulierten Chromatographielaufes

1

Es wurden 15 Fraktionen mit unterschiedlichen Anteilen des Nona (1) und des Deca (2) in 200  $\mu$ L D<sub>2</sub>O präpariert. Dazu wurden anhand eines Gaußprofils die Substanzmengen in jeder Fraktion berechnet.<sup>95</sup> Die Akquisition der NMR-Daten erfolge entsprechend Abschnitt 7.4. Für die Aufnahme der MS-Daten wurden die Proben gefriergetrocknet und in 200  $\mu$ L H<sub>2</sub>O gelöst. Davon wurde 1  $\mu$ L für die massenspektrometrischen Untersuchungen direkt injiziert (Abschnitt 7.3).<sup>95</sup>

I

Fraktion	Fraktion C <sub>Nona</sub>		
	(nmol/200 µL)	(nmol/200 µL)	
1	18.4	0	
2	29.9	4.4	
3	42.8	10.1	
4	50	18.4	
5	48.4	24.9	
6	45	29.9	
7	40.4	35.1	
8	37.8	37.8	
9	35.1	40.4	
10	29.9	45	
11	24.9	48.4	
12	18.4	50	
13	8.8	40.4	
14	3.6	27.3	
15	0	18.4	

 Tabelle 7: Präparation der Fraktionen 1-15

\_ . .

# 7.2.2 Präparation der Gemische G1 und G3

Die Gemische **G1** und **G3** wurden im Rahmen der Bachelorarbeit von Tim Nagel durch enzymatische Abspaltung vom bovinen Fibrinogen erhalten. Die Isolierung erfolgte nach modifiziertem Protokoll nach Tamura *et al.*<sup>99,105</sup> **G1** entspricht dem Abspaltungsprodukt vom Fibrinogen, welches von der Firma Sigma erhalten wurde (Fibrinogen vom bovinen Plasma (Type I-S, 65-85 % protein)). **G3** entspricht dem Abspaltungsprodukt vom Fibrinogen der Firma Calbiochem (Fibrinogen, Bovine Plasma). Die Gemische wurden jeweils nach Denaturierung und Alkylierung des Glycoproteins, Trypsin-Verdau, PNGase F-Verdau, Neuraminidase-Verdau und anschließender Abtrennung der Peptide durch

säulenchromatographische Trennung an gemischtem Ionentauscher erhalten (**G1** 0.9 mg, **G3** 0.4 mg).<sup>94</sup>

Gemisch **G1** wurde für die in den Abschnitten 4.4.1 und 4.4.5.1 diskutierte Analytik der Glycane des bovinen Fibrinogens verwendet. Gemisch **G3** ist das Endprodukt des in Abschnitt 4.4.4 vorgestellten Abspaltungsprozesses. Die Daten entstammen ebenfalls der Bachelorarbeit von Tim Nagel.<sup>94</sup>

# 7.2.3 Gemische G2 und G4

Das Gemisch **G2** wurde freundlicherweise von Miriam Kötzler zur Verfügung gestellt. Es wurde für die Aufklärung der Nonasaccharidisomere (Abschnitte 4.4.2 und 4.4.5.3, Abbildung 36 oben) verwendet.

Das Gemisch **G4** wurde freundlicherweise von Henning Behnken zur Verfügung gestellt. Es wurde für die in den Abschnitten 4.4.5.2 und 4.4.5.3 vorgestellte Auswertung der Undecaisomere und der triantennären Struktur 10 verwendet (s. a. Abbildung 36 unten).

Der Abspaltungsprozess wurde jeweils analog zu Abschnitt 7.2.2 durchgeführt.

# 7.3 Aufnahme und Auswertung der LC-MS-Daten

Alle massenspektrometrischen Messungen wurden von Henning N. Behnken konzipiert und durchgeführt (Henning N. Behnken, Dissertation, in Bearbeitung).

# Aufnahme der Gemische A-D (Bibliothek I-XIII)

Die Gemische A-D wurden jeweils in 20  $\mu$ L H<sub>2</sub>O gelöst. Es wurde jeweils 1  $\mu$ L im *positive ion mode* am ESI-TOF (Agilent) gemessen. Die Auswertung erfolgte mit der Software Masshunter (Version B.03.01, Agilent Technologies).

Gemisch	Komponente	[M+2H] <sup>2+</sup> gemessen	[M+2H] <sup>2+</sup> theoretisch
А	Ι	1228.1079 ([M+3H] <sup>3+</sup> )	1228.1079 ([M+3H] <sup>3+</sup> )
	III	1471.5419 ([M+3H] <sup>3+</sup> )	1471.5419 ([M+3H] <sup>3+</sup> )
	V	1513.5324	1513.5415
	IX	1112.3937	1112.3988
В	Ι	1228.0956 ([M+3H] <sup>3+</sup> )	1228.1060 ([M+3H] <sup>3+</sup> )
	VI	1696.1085	1696.1076
	VIII	1185.4223	1185.4277
С	Π	2024.2319	2024.2214
	VI	1696.0999	1696.1076
	Х	1120.9043	1120.9064
	XI	1258.4464	1258.4567

Tabelle 8: Übersicht über die nach MS bestimmten Komponenten der Gemische A-D

Gemisch	Komponente	[M+2H] <sup>2+</sup> gemessen	[M+2H] <sup>2+</sup> theoretisch
D	Ι	1841.6335	1841.6553
	IV	1768.6038	1768.6264
	VII	1440.5030	1440.5126
	VIII	1185.4227	1185.4277
	XII	1337.3849 ([M+H] <sup>+</sup> )	$1337.3890([M+H]^+)$

### Simulierter Chromatographielauf

Die MS-Daten des simulierten Chromatographielaufes wurden durch Messung von 1/200 jeder Fraktion erhalten. Die Substanzen wurden im *positive ion mode* am ESI-TOF (Agilent) gemessen.<sup>95</sup> Die Auswertung erfolgte mit der Software *Masshunter*. Für die Korrelation wurden die Intensitäten der [M+2Na]<sup>2+</sup>-Addukte verwendet. Für die Korrelationsberechnung wurde die Intensitätsinformation von Fraktion 2 als Mittelwert der MS-Daten der Fraktionen 1 und 3 verwendet.

### Gemisch G1

Gemisch **G1** wurde in 20  $\mu$ L H<sub>2</sub>O gelöst. Die Trennung erfolgte an PGC nach einem nach Muddiman *et al.*<sup>101</sup> modifiziertem Protokoll (HyperCarb<sup>TM</sup>, 2.1x150 mm, 3  $\mu$ m Partikelgröße, (Thermo-Fischer), Lösungsmittel A: Wasser, B: Acetonitril, Gradient: 0 min, 2% B, 2 min, 2% B, 4 min, 20% B, 16 min, 98% B, 22min, 2% B, 25 min, 2% B, Flussrate: 200  $\mu$ L/min). Die Substanzen wurden im *positive ion mode* am ESI-TOF (Agilent) gemessen. Beim ersten Lauf wurden 10  $\mu$ L des 1/10 verdünnten Gemisches injiziert und direkt im ESI-TOF gemessen. Beim zweiten Lauf wurden 10  $\mu$ L des Gemisches injiziert und tropfenweise fraktioniert. Jede Fraktion wurde mit 200  $\mu$ L H<sub>2</sub>O verdünnt und es wurde jeweils 1 $\mu$ L mittels Direktinjektion nochmal am MS-ESI-TOF gemessen.<sup>95</sup> Die Datenauswertung erfolgte mit der Software *Masshunter*. Die Fraktionen wurden lyophilisiert und für die NMR-Messungen in 180  $\mu$ L D2O gelöst. Für die Korrelation wurden die über alle Fraktionen rekonstruierten EICs der [M+2H]<sup>2+</sup>-Addukte verwendet (entsprechen im Chromatogramm dem Bereich von 10-13.5 min).

### Gemisch G2

Es wurden 8 mg des Gemisch **G2** für eine erste chromatographische Trennung verwendet. Die erhaltenen Fraktionen wurden in 100  $\mu$ L gelöst. Von jeder Fraktion wurden je 20  $\mu$ L entnommen, vereinigt und lyophilisiert. Das Gemisch wurde in 15  $\mu$ L H<sub>2</sub>O gelöst. Es wurde dann filtriert, mit H<sub>2</sub>O gespült, lyophilisiert und in 12  $\mu$ L H<sub>2</sub>O gelöst. Davon wurden 10  $\mu$ L injiziert. Die LC-MS-Daten wurden im *positive ion mode* am ESI-TOF (Agilent) erhalten. Die Trennung erfolgte an PGC anhand eines nach Altmann *et al.*<sup>47</sup> modifizierten Protokolls (HyperCarb<sup>TM</sup>, 2.1x150 mm, 3  $\mu$ m Partikelgröße, (Thermo-Fischer), Lösungsmittel A: Wasser, B: Acetonitril, jeweils mit 10 mM NH<sub>3</sub>,pH 9.6, Gradient: 0 min, 2% B, 2 min, 9 % B, 65 min, 17% B, 75 min, 40% B, 80 min, 100% B, 85 min 100% B, 90 min 2% B, Flussrate: 200  $\mu$ L/min). Die Fraktionen für die Korrelation wurden tropfenweise fraktioniert. Die Fraktionen wurden im Exsikkator getrocknet und in 20  $\mu$ L H<sub>2</sub>O/Acetonitril (1:1) gelöst. Davon wurde jeweils 1 $\mu$ L für die Messung am ESI-TOF direkt injiziert. Anschließend wurden die Fraktionen erneut im Exsikkator getrocknet und jeweils drei Fraktionen in 160  $\mu$ L D<sub>2</sub>O für die nachfolgenden NMR-Messungen vereinigt. Für die Korrelation wurden die über die Fraktionen rekonstruierten EICs der [M+2H]<sup>2+</sup>-Addukte verwendet.

### Gemisch G4

Es wurden 2.6 mg des Gemisches in 25  $\mu$ L H<sub>2</sub>O gelöst. Davon wurden 20  $\mu$ L für die Trennung injiziert. LC-MS-Daten wurden am ESI-QTOF-MS/MS (Bruker) im *positive ion mode* erhalten. Die Trennung von Gemisch erfolgte nach an PGC nach einem nach Altmann *et al.*<sup>47</sup> modifiziertem Protokoll. (HyperCarb<sup>TM</sup>, 4.6x150 mm, 3  $\mu$ m Partikelgröße, (Thermo-Fischer), Lösungsmittel A: Wasser, B: Acetonitril, jeweils mit 10 mM NH<sub>3</sub>, pH 9.6, Gradient: 0 min, 2% B, 2 min 9% B, 65 min 17% B, 105 min 28% B, 110 min, 90% B, 112 min, 90% B, 115 min, 2% B, Flussrate: 800  $\mu$ L/min). 1/20 des Eluats wurde direkt gemessen, 19/20 des Eluats wurden in Fraktonsgrößen von 15 s gesammelt (Fraction Collector (Advantec) SF-3120). Die Auswertung erfolgte mit *Data Analysis* (Version 4.0, SP 4, (build 281), Bruker Daltonics GmbH). Für die anschließenden NMR-Messungen wurden die Proben lyophilisiert und in 180  $\mu$ L D<sub>2</sub>O aufgenommen. Für die Korrelation wurden die EICs der [M+2H<sup>+</sup>+NH<sub>3</sub>]<sup>+</sup>-Addukte verwendet.

# 7.4 Aufnahme und Prozessierung der NMR-Daten

Alle NMR-Daten wurden entweder mit einem Avance700 oder einem AvanceIII 500 gemessen. Die Substanzen wurden jeweils in Mengen zwischen 150 und 200  $\mu$ L D<sub>2</sub>O in 3 mm NMR-Probenröhrchen gemessen. Zur Unterdrückung der HDO-Resonanz wurde eine *excitation sculpting* Pulssequenz nach Shaka *et al.* verwendet.<sup>80</sup> Der selektive Rechteckpuls wurde mit einer Länge von 8 ms und einem  $\gamma B_1/2\pi$  von 62.5 Hz eingesetzt. Alle Daten wurden mit einer spektralen Breite von 10 ppm, mit 32k Datenpunkten und einem Relaxationsdelay von 1 s aufgenommen. Alle Daten wurden auf internes oder externes Aceton ( $\delta$ =2.225 ppm) referenziert.

# Aufnahme der NMR-Spektren der Bibliothek (I-XIII) und der Gemische A-D

Alle NMR-Daten wurden am AV700 mit cryoprobe aufgenommen.

Struktur	Bezeichnung	Menge (ca. pmol)	Gelöst in µL	NS
		(Herstellerangaben)	$D_2O$	
Ι	4A+F	270	150	16k
II	4A+1R+F	250	180	4k
III	4A+2R+F	230	180	бk
IV	4A	280	180	10k
V	3A(2,6)+F	330	150 / 180	3k / 10k
VI	3A(2,6)+1R+F	290	180	4k
VII	3A(2,6)	350	180	бk
VIII	2A+F	420	170	1k
IX	2A	450	200	1k
Х	2A+1a1,3Gal+F	450	180	1k
XI	2A+Lx+F	400	180	2k
XII	Man5-6P	380	180	2k
XIII	Man6-6P	340	180	3k
Gemisch A	I+III+V+IX	543+226+1320+450	170	10k
Gemisch B	I+VI+VIII	543+295+676	200	10k
Gemisch C	II+VI+X+XI	988+295+446+397	180	10k
Gemisch D	I+IV+VII+VIII+XII	543+565+700+844+380	180	10k

	••					
<b>T</b> I II A	TTI • 14	<b>11 D I</b>	•• 4•	1 1 1	T X7TTT 1 1	<i>a</i> • • • •
	I borciont	t iinar Pranc	nnronorotior	h dar Prohan	I XIII und dor	I omicono A II
1 2016116 7.	UTDET SICTI			ее		ATCHING HE ATT
I GOODIC / I	C NOL DICHT		mpi apai atioi		A THEFT WING GUT	O children c

A – Antenne, F – Fucoseeinheit, R – *repeat*, (2,6)–Verzweigung der 3. Antenne (1,6-verknüpft an Man-4'), Lx –Lewis<sup>x</sup> Struktur, P – Phosphorylierung, NS – *number of scans* 

# Aufnahme der NMR-Spektren für die Korrelationsberechnung

Proben	Μ	lHz	Probenkopf	NS	Prozessierung
	500	700			
Simulierter Chromatographielauf, Fraktionen 1-15		x	TXI Probenkopf	1k	TD: 32k SI: 32k EM, <i>lb</i> 1
Gemisch G1		x	TXI Probenkopf	1k	TD: 32k SI: 32k EM, <i>lb</i> 0.5
Gemisch G2	x		TCI cyroprobe	1k	TD: 32k SI: 32k EM, <i>lb</i> 1
Gemisch G4	x		TCI cryoprobe	2k	TD: 32k SI: 64k GM <i>lb</i> -2 <i>gb</i> 0.1

Tabelle 10: Akquisitions- und Prozessierungsparameter der NMR-Daten für die Korrelationsberechnung

TD – Anzahl aufgenommener Datenpunkte, SI – Anzahl Datenpunkte des prozessierten Spektrums, EM – Multiplikation des FID mit einer Exponentialfunktion, *lb – line broadening factor*, GM – Gaußmultiplikation, *gb – gaussian broadening factor* 

# Aufnahme der NMR-Spektren einzelner Verbindungen

# **Deca** (2)

Es wurde eine 1 mM Probe des Decasaccharids mit 40 *scans* am *Avance700* mit *cryoprobe* gemessen (Pulsprogramm *zg*, Bruker). Für Abbildung 9 verwendete Prozessierung: EM, *lb* 0.3.

# Dodeca (4)

Eine Probe des Dodecasaccharids wurde am *Avance700* mit *cryoprobe* gemessen. Das <sup>1</sup>H-NMR-Spektrum mit Wasserunterdrückung wurde mit den oben beschriebenen Parametern mit 2k *scans* akquiriert. (Pulsprogramm *zgesgp*, Bruker). Ein TOCSY wurde mit einer *mlev*-Pulsabfolge mit folgenden Parametern akquiriert: Scans: 600, Datenpunkte in F2 und F1: 8k, 256; Mischzeit: 100 ms; Relaxationsdelay: 0.1 s. Die Wasserunterdrückung erfolgte mit der *es*-Pulssequenz wie oben beschrieben. Prozessierung für Abbildung 38: GM, *gb* 0.1, *lb* -2.
#### **NMR-Daten einzelner Fraktionen**

Die Proben wurden am *Avance700* gemessen. Die Spektren wurden mit der oben beschriebenen Wasserunterdrückung akquiriert. Alle Proben entstammen weiteren Chromatographieläufen von **G2** oder **G4**.

Substanz	Probenkopf	Scans	Abbildung	Prozessierung für Abbildung
Undere 5	cryoprobe	3k	Abbildung 25	EM, <i>lb</i> 1
Undeca 5	TXI-Probenkopf	2k	Abbildung 38	GM, gb 0.1, lb 2
Undecaisomer 6	TXI-Probenkopf	38k	Abbildung 38	GM, gb 0.1, lb 2
	Cryoprobe	4k	Abbildung 24	EM, lb 0.5,
fucosyliertes Nona 3				average smooth
rucosyneries Nona 3	TXI-Probenkopf	2k	-	-
Nona (1)	TXI-Probenkopf	512	-	-

Tabelle 11: Aufnahme und Prozessierungsparameter der NMR-Daten ausgewählter Substanzen

average smooth - siehe Datenbearbeitung mit matlab: Export von Grafiken

Die Aufklärung der Nonaisomere erfolgte anhand der Spektren von Gemisch **G2**. Prozessierung für Abbildung 39: Bereiche 4.43-4.49 ppm und 4.55-5.2 ppm, EM, *lb* 1; *N*-Acetylregion GM, *lb* -2, *gb* 0.1.

#### Aufnahme der selektiven TOCSY-Experimente

Das Glycangemisch G1 wurde mit der Pulssequenz *selmlgp.2* (Bruker) mit 1k *scans* an einem *DRX500*-Spektrometer gemessen. Es wurden ein selektiver Gauß-Puls und eine *mlev*-Pulsabfolge verwendet. Die in Abbildung 37 gezeigten Spektren A, B und D wurden mit einem selektiven Puls mit der Pulslänge 60 ms ( $\gamma$ B<sub>1</sub> = 20 Hz) und Spektrum C mit 30 ms ( $\gamma$ B<sub>1</sub>=20 Hz) gemessen. Es wurde eine *spin-lock* Zeit von 200 ms verwendet, die Einstrahlpunkte (vor Referenzierung) lagen bei A: 5.098 ppm, B: 3.972 ppm, C: 4.141 ppm und D: 4.498 ppm. Für die Abbildung verwendete Prozessierung: EM, *lb* 1. Die Spektren sind auf externes Aceton referenziert ( $\Delta$ =23.6 Hz).<sup>94</sup>

### 7.5 <u>Datenbearbeitung mit matlab</u>

#### 7.5.1 Berechnung von Spektren anhand von Datenbankeinträgen

Die *sugabase*-Daten wurden als Textdatei von *awk* übergeben. Jeder *sugabase*-Eintrag enthält eine Record-Zeile, die die CCSD-Nummer beinhaltet, und in weiteren Zeilen die Spaltenmerkmale "Residue", "Linkage", "Proton", "Delta", und drei weitere Spalten für Bemerkungen und Kopplungskonstanten. Einige Felder enthalten keine Einträge. Die Textdatei wurde in matlab über den textscan-Befehl eingelesen. Die output-Variable ist eine 1x7 Zelle, die in ihren Spalten die zuvor bezeichneten Merkmale beinhaltet. Diese wurde über einen Schleifenbefehl in einzelne Zeilen zerlegt. Im Folgenden wird nicht mehr jede for- bzw. while-Schleife gesondert aufgeführt. Diese wurden eingesetzt, wenn bestimmte Aktionen für mehrere Einträge durchgeführt werden sollten. Leerzeichen wurden durch strtrim entfernt. Die Gesamtmatrix wurde mit den Spalten "Residue", "Linkage", "Proton" und "Delta" definiert. Durch ein logical indexing wurde der Beginn jeden Eintrags anhand des strings 'Record' identifiziert (strcmp). Bei leerem "Linkage"-Feld in der ersten Zeile wurde diesem ein '-' zugeordnet. Anschließend wurde die CCSD-Nummer aus jeder Record-Zeile in jede zugehörige Eintragszeile geschrieben. Die Record-Zeile wurde anschließend gelöscht. Jedes leere "Linkage"-Feld wurde nun durch Abgleich mit den vorigen Zeilen mit der korrekten Verknüpfungsinformation aufgefüllt. Die nachfolgend implementierte user-Abfrage erlaubt die Wahl der Spektrometerbasisfrequenz und eines linebroadening-Faktors (input). Anschließend wurden alle nicht-reduzierten Derivate von Glucose, Mannose, Galactose, Fucose und Neuraminsäure als Standardreste definiert. Nur wenn jeder Rest eines Eintrags ein Standardrest war, wurde dieser Eintrag weiterprozessiert (ismember). Dann wurden die Kopplungskonstanten anhand einer extern abgelegten Excel-Tabelle zugeordnet, diese enthält als Zeileninformation den Zuckerrest und als Spalteninformation die Protonenspezies (Tabelle 12). Zunächst wurde über regexp jeder Rest der Gesamtmatrix in matlab mit den Resten der Tabelle abgeglichen, sodass diesem ein Zeilenidentifikator zugewiesen wurde. Die Protonenspezies der Gesamtmatrix werden über strcmp/strncmp mit denen der Tabelle abgeglichen, wobei allen H-1 und H-2 Protonen bei offener Konfiguration die des vorstehenden Restes zugewiesen wurde. Jede Protonenspezies der Gesamtmatrix erhielt so einen Spaltenidentifikator. Die Zuordnung der Kopplungskonstanten erfolgte für jede Zeile anhand des zuvor zugewiesenen Spalten- und Zeilenidentifikators. Jeder Zeile wurden bis zu drei Kopplungskonstanten zugeordnet, sodass die Gesamtmatrix um drei Spalten erweitert wurde. Dann wurde eine 3x8-Matrix erzeugt, die die Einträge +1 oder -1 als Faktor für die Kopplungskonstanten für die Bestimmung der Maxima jedes Signals beinhaltete (s. a. Abbildung 8). Es wurde eine nx8 delta-Matrix erzeugt, in der jede Zeile alle zu einem Proton gehörenden Maxima beinhaltet. Jedem Eintrag wurde zunächst die Intensität 1 zugeordnet. Dann wurde zur Berechnung der relativen Intensitäten über regexp die Protonenspalte der Gesamtmatrix auf Zuordnung als  $\alpha$ - oder β-Anomer durchsucht und der entsprechende Eintrag mit 0.6 bzw. 0.4 multipliziert. Methylgruppen wurden mit strcncmp identifiziert und die Intensitäten mit drei

multipliziert. Anschließend wurde eine user-Abfrage implementiert (input), in der zunächst geprüft wird, ob einzelne Spektren ausgelesen werden sollen. Wird diese Abfrage mit 'y' bestätigt, kann anschließend die Anzahl angegeben werden und dann eine Eingabe der CCSD-Nummer(n) der einzelnen Einträge, sowie ein Dateiname, der abschließend durch ein Datum ergänzt wird, erfolgen. Es wurde außerdem die Rückmeldung integriert, ob die entsprechende CCSD-Nummer im (reduzierten) Datensatz vorhanden ist. Zur Berechnung der Spektren wurden nun die CCSD-Nummern ausgelesen (unique) und die Anzahl an Datenpunkten für die Berechnung, sowie die minimale und maximale chemische Verschiebung definiert. Anhand dieser Daten wurde die Berechnung einer Frequenzskala durchgeführt. Weiterhin wurde eine Schleifenabfrage zur Prozessierung von Datenpaketen mit maximal 140 Spektren/Paket geschrieben. In einer weiteren Schleife wurde jeder einzelne CCSD-Eintrag prozessiert. Dazu wurde zunächst die CCSD-Nummer des aktuellen Eintrags vorgegeben und anhand dieser die zugehörigen Daten selektiert. Durch den Einsatz von genvarname wurde eine Variable intCCSD-No. erzeugt. Durch den Befehl cat wurden die 8 Spalten der delta-Matrix aneinandergehängt, sodass ein einzelner Spaltenvektor resultierte. Dieser wurde für die Berechnung der Intensität an jedem einzelnen Datenpunkt der Frequenzskala entsprechend der Linienform einer Lorentzfunktion verwendet (s. a. Gleichung 12). Anschließend wurde das so berechnete Spektrum der erzeugten Variablen zugeordnet, sodass ein eindeutiges Identifikationsmerkmal durch die CCSD-Nummer gegeben ist. Nach Durchlaufen eines Datenpakets wurden die berechneten Spektren zusammen mit einer CCSD-Tabelle als .mat-File gespeichert und anschließend gelöscht, bevor das nächste Datenpaket berechnet wurde. Ein .mat-File ist ein matlab-internes Datenformat zur Speicherung beliebiger Variablen. Die in dieser Arbeit verwendeten Spektren wurden mit einem Faktor für die Linienbreite  $(\alpha_{L}, \text{HWHM})$  von 1.5 berechnet.

	H-1a	H-1b	H-2	H-3	H-3e	H-3a	H-4	H-5	H-6	H-6'
D-Glc	3.2	8	9.6	9.2			9.2	1.6	12	4.8
			3.2	9.6			9.2	9.2	1.6	12
			8					4.8		
D-Gal	3	7.8	9.9	3			1	8.2	12	4.6
			3	9.9			3	1	8.2	12
			7.8					4.6		
D-Man	1.3	0.9	3.3	10			10	1.9	12	5.8
			1.3	3.3			10	10	1.9	12
			0.9					5.8		
D-Neu					12.3	12	4.9			
					4.9	12	12			
L-Fuc	3.7							6.7	6.7	6.7
								6.7		
								6.7		

Tabelle 12: Tabelle der Kopplungskonstanten, die für die Abfrage der Kopplungskonstanten genutzt wird. Alle Werte in Hz.

Für die Zuordnung der Kopplungskonstanten wird jedem eingetragenen Proton ein Spalten- und ein Zeilenindex zugeordnet. Anhand dieser beiden Identifikatoren können jedem Proton die entsprechenden, in dieser Tabelle verzeichneten Kopplungskonstanten zugeordnet werden.

#### 7.5.2 Automatische Identifikation von Spektren

Für die automatische Identifikation von Spektren wurden die *sugabase*-Einträge mit dem gleichen Punkt-Punkt-Abstand wie die experimentellen Spektren in gerechnete Spektren umgewandelt. Alle Daten wurden normiert auf das maximale Signal im jeweiligen Spektrum. Zu identifizierende Spektren wurden für den Abgleich so kalibriert, dass sie eine identische Indexierung, also gleiche minimale und maximale chemische Verschiebung erhielten. Die in .mat-files gespeicherten berechneten Spektren wurden in ihren Datenpaketen ausgelesen, wobei einer Variablen die CCSD-Nummern in alphabetischer Sortierung zugeordnet wurden und einer weiteren die Spektren (*orderfields*, *fieldnames*, *struct2cel1*). Außerdem wurde die ppm-Achse aus diesem Datensatz importiert. Die Daten wurden in einem ppm-Bereich von 1.18-6.0 ppm verglichen, wobei Daten in den Bereichen 1.28-1.36 ppm, 1.88-1.93 ppm 3.4-4.0 ppm und 4.72-4.76 ppm gleich 0 gesetzt wurden. In einer *while*-Schleife wurde nun durch den *lsqlin*-Befehl der Faktor x, mit dem jedes Spektren mit einem Faktor x  $\leq$  *cut off* wurden die Einträge aus dem Datenpaket entfernt. Veränderte sich die Zuordnung der verbleibenden Spektren nicht mehr, so wurde das nächste

Datenpaket hinzugefügt. Anschließend wurde eine Tabelle generiert, die die verbleibenden Spektren in Form ihrer CCSD-Nummer benennt, sowie den normierten Faktor x. In einer grafischen Darstellung wurden das zu untersuchende Spektrum, die Summe der gefundenen Einträge, sowie eine Differenz beider Spektren abgebildet.

#### 7.5.3 Berechnung des 3DCC-Beispiels

Anhand der Gauß'schen Normalverteilung (Gleichung 11) wurden drei EICs mit 20 Datenpunkten für  $1 \le x \le 5$ , mit den Erwartungswerten  $\mu$ =1.8, 3.0 und 4.3, den Intensitäten 0.6, 1.0 und 0.8 und einer Standardabweichung  $\sigma$  von 1.0 berechnet.

$$f(x) = \frac{Int}{\sqrt[2]{2\pi} \cdot \sigma} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{Gleichung 11}$$

## Gaußverteilung: Int – Intensität, σ – Standardabweichung, Maß für Breite und Höhe der Kurve, μ – Mittelwert, bei x=μ Maxmimum der Verteilung

Es wurden anschließend 20 NMR-Spektren mit drei NMR-Signalen in Form der Lorentzverteilung berechnet, von denen fünf graphisch aufgetragen wurden. Jedes NMR-Spektrum bestand aus 200 Datenpunkten, die Halbwertsbreite (HWHM) der NMR-Signale betrug  $\alpha_L$ =1.1. Von jedem EIC wurden die Datenpunkte für die Intensitätsberechnung des zugehörigen NMR-Signals verwendet.

$$f(x) = \frac{1}{\pi} \frac{Int \cdot \alpha_L}{(x - x_0)^2 + \alpha_L^2}$$
 Gleichung 12

Lorentzfunktion: Int – Intensität, *a*<sub>L</sub> – HWHM, x<sub>0</sub> – chemische Verschiebung

Die Berechnung der Korrelationsmatrix nach Pearson und der Spektrendekonvolution erfolgte entsprechend der in Abschnitt 7.5.6 beschriebenen Durchführung.

#### 7.5.4 Berechnung der Beispiele zum Vergleich verschiedener Korrelationsmethoden

Es wurde eine Lorentzfunktion entsprechend Abschnitt 7.5.3 berechnet ( $-1 \le x \le 1$ , Datenpunkte 1000,  $\alpha_L=0.1$ ,  $x_0=0$ ). Anschließend wurde eine zweite identische Funktion in dem Bereich - 0.75<x<0.75 mit 51 Inkrementen verschoben und jeweils der Korrelationskoeffizient zur Ursprungsfunktion und zur Summenfunktion nach Kendall, Spearman und Pearson berechnet (Abschnitt 7.5.6). In einem zweiten Beispiel wurde die zweite Funktion im Bereich von 0<x<0.8 mit 10 Inkrementen verschoben und für jede Verschiebung die Intensität von 1 bis 0.01 mit 10 Inkrementen reduziert. Es wurde jeweils der Korrelationskoeffizient zur Summenfunktion nach Kendall, Spearman und Pearson bestimmt. Dieser wurde jeweils mit

der Intensität der zweiten Funktion multipliziert. Das Ergebnis wurde gegen die reale Intensität der zweiten Funktion aufgetragen.

#### 7.5.5 Import und Bearbeitung von MS- und NMR-Daten

#### MS-Daten

Wurden die MS-Daten mit höherer zeitlicher Auflösung gemessen als fraktioniert wurde, so mussten die Daten entsprechend der Anzahl an Fraktionen reduziert werden. Dazu wurde zunächst eine Wartezeit definiert, die dem Beginn der ersten Fraktion, die für die Korrelation verwendet wurde, entspricht. Anhand der Fraktionsgröße wurde außerdem bestimmt, wie viele MS-Datenpunkte einer Fraktion entsprechen. Davon ausgehend wurde entweder über die Summe der Datenpunkte oder den Mittelwert die reduzierte Datenmenge berechnet (*mean*, *sum*). Die Daten wurden nach Normierung als .mat-file gespeichert.

#### **NMR-Daten**

NMR-Daten wurden nach Prozessierung in TopSpin in *matlab* importiert (*matlab funtion*: *rbmmr*, *Nils Nyberg*, *SLU*, 2001-05-02, *nn@farma.ku.dk*). Anschließend wurden die Daten in *matlab* nachkalibriert. Dazu wurden die Signale bei 2.225 (Aceton), 0.157 (prominente Verunreinigung, kalibriert auf externes Aceton) oder 1.320 (Dublett von Lactat, kalibriert auf externes Aceton) gewählt. In Abhängigkeit vom gewählten Kalibriersignal wurde der entsprechende Spektrenbereich definiert. Innerhalb dieses Bereichs wurde mithilfe des *findpeaks*-Befehls bzw. durch *max* ein *peakpicking* durchgeführt und dadurch der Index des Peakmaximums erhalten. Die Intensitätsdaten wurden daraufhin so beschnitten, dass der Index genau dem der entsprechenden chemischen Verschiebung der delta-Achse entsprach. Die delta-Achse entspricht der x-Achse, gegen die die Intensitäten aufgetragen wurden. Die Daten wurden anschließend in einem .mat-file gespeichert.

Die Berechnung des S/N erfolgte anhand von Gleichung 10. Dazu wurde der Rauschbereich zwischen 8.5 und 9.5 ppm definiert. Der Signalbereich wurde für das Man-4, H-1 von 5.11-5.13 ppm, der des Gal-7/7<sup>4</sup> H-1 $\alpha$  von 5.13-5.16 ppm gewählt. Es wurde jeweils durch *max* die maximale Intensität und durch *std* die Standardabweichung im Rauschbereich bestimmt. Die in Tabelle 6 aufgeführten S/N wurden nach Glättung der Daten durch einen gleitenden Durchschnitt (n=5) erhalten (*smooth*).

Zur Integralberechnung wurde der *trapz*-Befehl verwendet, dieser nutzt eine trapezoide numerische Annäherung. Die Bereiche wurden entweder manuell definiert oder es wurden die

zuvor in TopSpin definierten Integralregionen aus einem vergleichbaren NMR-Spektrum importiert (*importdata*, <Dateipfad>\Name des Spektrums\Expno\pdata\Procno\intrng).

#### 7.5.6 Korrelation, Extraktion und Optimierung

#### Korrelation und Optimierungsvarianten

Die MS- und NMR-Daten wurden entsprechend Abschnitt 7.5.5 bearbeitet. Nach Einlesen der Daten (*load*) wurden diese mit *corrcoef* oder *corr* korreliert. Dafür wurde in einer doppelten Schleifenabfrage je eine Laufvariable für die verschiedenen EICs und eine für die EDCs definiert. Der *corrcoef*-Befehl wurde für die Berechnung der Koeffizienten nach Pearson eingesetzt. Implementiert in dieser Funktion ist auch der Signifikanztest. Für die Anwendung des Signifikanztest wurden alle Korrelationskoeffizienten mit einer Signifikanzt von < 95 % gleich Null gesetzt (p<0.05). Der *corr*-Befehl wurde für den Vergleich der Korrelation nach Kendall, Spearman und Pearson verwendet und benötigt nur für die verschiedenen EDCs eine Laufvariable.

Für den Einsatz des Retentionszeitfilters wurden für die NMR- und die MS-Daten jeder Fraktion, bei der die Intensität im EIC weniger als 1 % der maximalen Intensität des EICs betrug, entweder gleich Null gesetzt oder gelöscht.

Der Wert des *cut offs* wurde anhand der maximalen Korrelation der EICs untereinander definiert. Diese wurde wie oben beschrieben berechnet. Jeder Korrelationskoeffizient der kleiner als dieser *cut off* war, wurde gleich Null gesetzt.

Zur Gewichtung mit der Potenzfunktion wurde für jeden Korrelationskoeffizienten die n.te Potenz bestimmt. Zur Gewichtung mit der sigmoiden Funktion wurden erste Experimente mit folgender Funktion durchgeführt:

$$(c_{m,\delta})_s = \frac{1}{1 + exp^{-k(c_{m,\delta}-x)}}$$
 Gleichung 13

$$(c_{m,\partial})_s$$
 – Korrelationskoeffizient nach Gewichtung mit sigmoider Funktion, k –  
Maß für die Steigung, x – Wendepunkt,  $c_{m,\partial}$ – Korrelationskoeffizient

Die Extraktion der NMR-Spektren erfolgte durch elementweise Multiplikation des einfachen oder optimierten Korrelationskoeffizienten mit dem Summenspektrum. Im Falle des Retentionszeitfilters wurde das Summenspektrum anhand der bereinigten NMR-Spektren gebildet.

#### 7.5.7 Export von Grafiken

Korrelationsspektren und zum Vergleich gezeigte NMR-Spektren, die im Abschnitt 4.4 abgebildet sind, wurden vor dem Export als Grafik anhand des gleitenden Durchschnitts geglättet (*moving average, smooth*). Alle MS-Daten wurden mit einem Savitzky-Golay Filter mit einer Polynomregression 2. Ordnung geglättet. Einzelne EDCs wurden durch *findpeaks* oder *find* gefunden und ohne Glättung exportiert. Zur Darstellung von dreidimensionalen Abbildungen wurden die Dimensionen durch *meshgrid* generiert. Alle Grafiken wurden aus *matlab* im eps-Format exportiert.

## 7.6 Charakterisierung der untersuchten Glycanstrukturen

## Struktur 1: Nona

 $C_{62}H_{104}N_4O_{46}$ ; monoisotopisches MW = 1640.59

 $[M+2H]^{2+} = 821.3038$  (theoretisch: 821.3034)



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.189	d	2.9
		H-1b	n.d.	n.d.	n.d.
		NAc(a)	2.039	s	-
		NAc(b)	2.037	s	-
b-D-GlcpNAc	4	H-1(a)	4.613	d	8.5
		H-1(b)	4.604	d	8.1
		NAc(a)	2.082	s	-
		NAc(b)	2.080	s	-
b-D-Manp	4,4	H-2	4.248	dd	2,1: *
					2,3: 3.9
a-D-Manp	6,4,4	H-1	4.927	*	*
		H-2	4.112	dd	2,1:1.6
					2,3: 3.5
b-D-GlcpNAc	2,6,4,4	H- 1	4.582	d	8.5
		NAc	2.047	s	-
b-D-Galp	4,2,6,4,4	H-1	4.474	d	8.2
a-D-Manp	3,4,4	H-1	5.119	*	*
		H-2	4.191	dd	2,1:1.8
					2,3: 3.3
b-D-GlcpNAc	2,3,4,4	H-1	4.582	d	8.5
		NAc	2.051	s	
b-D-Galp	4,2,3,4,4	H-1	4.467	d	8.2

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined

Weitere Signale: 4.00-3.47 (bulk), 4.76 (HDO), 2.72 (s, DMSO), 2.59, 0.157 (s, unbekannt)

### Struktur 2: Deca

 $C_{68}H_{114}N_4O_{50}$ ; monoisotopisches MW=1786.65 g/mol

 $[M+2H]^{2+} = 894.3329$  (theoretisch: 894.3323)



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, internes Aceton, <sup>1</sup>H-Spektrum ohne *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.180	d	3.4
		H-1b	4.693	d	8.1
		NAc(a)	2.039	8	-
		NAc(b)	2.037	8	-
a-L-Fucp	6	H-1(a)	4.889	d	3.9
		H-1(b)	4.896	d	4.0
		H-5(a)	4.097	dq	5,4: 0.9
					5,6: 7.1
		H-5(b)	4.133	dq	5,4: 1.3
					5,6: 6.6
		CH3(a)	1.209	d	6.8
		CH3(b)	1.219	d	6.8
b-D-GlcpNAc	4	H-1(a)	4.668	d	8.1
		H-1(b)	4.664	d	8.1
		NAc(a)	2.096	8	-
		NAc(b)	2.093	8	-
b-D-Manp	4,4	H-2	4.250	dd	2,1:1.0
					2,3: 3.5
a-D-Manp	6,4,4	H-1	4.925	d	1.7
		H-2	4.108	dd	2,1:1.3
					2,3: 3.0
b-D-GlcpNAc	2,6,4,4	H- 1	4.581	d	7.7
		NAc	2.047	8	-
b-D-Galp	4,2,6,4,4	H-1	4.473	d	7.9
a-D-Manp	3,4,4	H-1	5.119	d	1.3
		H-2	4.191	dd	2,1:1.5
					2,3: 3.4
b-D-GlcpNAc	2,3,4,4	H-1	4.581	d	7.7
		NAc	2.051	s	
b-D-Galp	4,2,3,4,4	H-1	4.467	d	7.7

Weitere Signale: 4.02-3.45 (*bulk*), 4.10 + 1.32 (q+d, Lactat), 2.225 (s, Aceton), 1.91 (s, unbekannt)

### Struktur 3: fucosyliertes Nona

C<sub>62</sub>H<sub>104</sub>N<sub>4</sub>O<sub>45</sub>; monoisotopisches MW=1624.60 g/mol

 $[M+2H]^{2+} = 813.3065$  (theoretisch: 813.3059)



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, internes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.180	d	3.9
		H-1b	n.d.	n.d.	n.d.
		NAc	2.038	8	-
a-L-Fucp	6	H-1(a)	4.887	d	4.5
		H-1(b)	4.894	d	4.7
		H-5(a)	4.096	q	6.9
		H-5(b)	n.d.	n.d.	n.d.
		CH3(a)	1.209	d	6.8
		CH3(b)	1.219	d	6.8
b-D-GlcpNAc	4	H-1(a)	4.664	d	7.9
		H-1(b)	4.669	d	7.9
		NAc(a)	2.095	S	-
		NAc(b)	2.091	S	-
b-D-Manp	4,4	H-2	4.250	dd	2,1: *
					2,3: 3.4
a-D-Manp	6,4,4	H-1	4.915	*	*
		H-2	4.107	dd	2,1:1.4
					2,3:4.1
b-D-GlcpNAc	2,6,4,4	H- 1	4.554	d	9.5
-		NAc	2.051	S	-
a-D-Manp	3,4,4	H-1	5.118	*	*
-		H-2	4.190	dd	2,1:1.7
					2,3: 3.8
b-D-GlcpNAc	2,3,4,4	H-1	4.580	d	7.6
-		NAc	2.051	s	-
b-D-Galp	4,2,3,4,4	H-1	4.467	d	8.1

\* – Kopplungskonstante konnte nicht bestimmt werden, n.d. – not determined

Weitere Signale: 4.03-3.32 (*bulk*), 1.32 (d, Lactat), 2.225 (s, Aceton), 2.72 (s, DMSO), 2.59, 0.157 (s, unbekannt)

## Struktur 4: Dodeca

 $C_{80}H_{134}N_4O_{60}$ ; monoisotopisches MW=2110.76 g/mol

 $[M+2H]^{2+} = 1056.3857$  (theoretisch: 1056.3851)



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.180	d	3.6
		H-1b	n.d.	n.d.	n.d.
		NAc	2.039	s	-
a-L-Fucp	6	H-1(a)	4.890	d	4.5
		H-1(b)	4.897	d	4.7
		H-5(a)	4.096	q	7.5
		H-5(b)	4.133	q	7.0
		CH3(a)	1.209	d	7.1
		CH3(b)	1.219	d	7.1
b-D-GlcpNAc	4	H-1(a)	4.663	d	8.2
		H-1(b)	n.d.	n.d.	n.d.
		NAc(a)	2.097	s	-
		NAc(b)	2.093	s	-
b-D-Manp	4,4	H-2	4.249	dd	2,1:1.2
-					2,3: 3.5
a-D-Manp	6,4,4	H-1	4.925	*	*
-		H-2	4.111	dd	2,1:1.6
					2,3: 3.0
b-D-GlcpNAc	2,6,4,4	H- 1	4.583	d	8.5
		NAc	2.047	s	-
b-D-Galp	4,2,6,4,4	H-1	4.544	d	8.3
		H-4	4.183	*	*
a-D-Galp	3,4,2,6,4,4	H-1	5.147	d	4.3#
-		H-4	4.02	*	*
		H-5	4.191	*	*
a-D-Manp	3,4,4	H-1	5.120	*	*
-		H-2	4.191	dd	2,1:1.7
					2,3: 3.5
b-D-GlcpNAc	2,3,4,4	H-1	4.583	d	8.5
-		NAc	2.051	s	-
b-D-Galp	4,2,3,4,4	H-1	4.539	d	8.6
-		H-4	4.183	*	*
a-D-Galp	3,4,2,3,4,4	H-1	5.144	d	4.1#
-		H-4	4.02	*	*
		H-5	4.191	*	*

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined,

Weitere Signale: 4.01-3.45 (*bulk*), 3.35, 2.13, 2.11, (s, unbekannt), 2.29 - 2.23, 2.18 - 2.15, 1.65 - 1.52, 1.29 - 1.25, 0.90 - 0.83 (m, unbekannt), 0.157 (s, unbekannt)

## Struktur 5: Undeca

C<sub>74</sub>H<sub>124</sub>N<sub>4</sub>O<sub>55</sub>; monoisotopisches MW=1948.70 g/mol

 $[M+2H]^{2+} = 975.3593$  (theoretisch: 975.3587)



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.180	d	3.4
		H-1b	n.d.	n.d.	n.d.
		NAc	2.038	S	-
a-L-Fucp	6	H-1(a)	4.890	d	4.5
		H-1(b)	4.896	d	4.6
		H-5(a)	4.096	q	5,6: 7.4
		H-5(b)	4.132	q	5,6: 7.0
		CH3(a)	1.209	d	7.1
		CH3(b)	1.219	d	7.1
b-D-GlcpNAc	4	H-1	n.d.	n.d.	n.d.
		NAc(a)	2.097	s	-
		NAc(b)	2.093	8	-
b-D-Manp	4,4	H-2	4.249	dd	2,1:1.5
					2,3: 3.9
a-D-Manp	6,4,4	H-1	4.925	*	*
		H-2	4.111	dd	2,1:1.5
					2,3: 3.4
b-D-GlcpNAc	2,6,4,4	H- 1	4.581	d	8.8
		NAc	2.047	S	-
b-D-Galp	4,2,6,4,4	H-1	4.544	d	8.4
		H-4	4.21-4.18	m*	*
a-D-Galp	3,4,2,6,4,4	H-1	5.147	d	4.3#
		H-4	4.022	dd	4,3: 3.6
					4,5: 1.1
		H-5	4.21-4.18	m*	*
a-D-Manp	3,4,4	H-1	5.120	d	1.5
		H-2	4.21-4.18	m*	*
b-D-GlcpNAc	2,3,4,4	H-1	4.581	d	8.8
		NAc	2.050	s	
b-D-Galp	4,2,3,4,4	H-1	4.467	d	8.5

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined,

Weitere Signale: 4.01-3.46 (*bulk*), 4.10 + 1.32 (q+d, Lactat), 2.72 (*s*, DMSO), 2.59, 0.157 (je s, unbekannt)

### Struktur 6: Undecaisomer

C<sub>74</sub>H<sub>124</sub>N<sub>4</sub>O<sub>55</sub>; monoisotopisches MW=1948.70 g/mol

 $[M+2H]^{2+} = 975.3593$  (theoretisch: 975.3587)



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.180	d	3.4
		H-1b	n.d.	n.d.	n.d.
		NAc	2.039	s	-
a-L-Fucp	6	H-1(a)	4.890	d	4.1
		H-1(b)	4.896	d	4.3
		H-5(a)	4.097	q	7.4
		H-5(b)	4.133	q	7.5
		CH3(a)	1.209	d	7.1
		CH3(b)	1.219	d	6.9
b-D-GlcpNAc	4	H-1	n.d.	n.d.	n.d.
		NAc(a)	2.096	s	-
		NAc(b)	2.093	s	-
b-D-Manp	4,4	H-2	4.250	dd	2,1:1.2
					2,3: 2.8
a-D-Manp	6,4,4	H-1	4.925	*	*
		H-2	4.11	*	*
b-D-GlcpNAc	2,6,4,4	H- 1	4.582	d	8.8
		NAc	2.046	s	-
b-D-Galp	4,2,6,4,4	H-1	4.473	d	8.4
		H-4	4.23-4.17	m*	*
a-D-Manp	3,4,4	H-1	5.121	*	*
		H-2	4.23-4.17	m*	*
b-D-GlcpNAc	2,3,4,4	H-1	4.582	d	8.8
		NAc	2.051	s	-
b-D-Galp	4,2,3,4,4	H-1	4.539	d	8.3
a-D-Galp	3,4,2,3,4,4	H-1	5.144	d	4.3#
-		H-4	4.019	dd	4,3: 3.4
					4,5: 1.5
		H-5	4.23-4.17	m*	*

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined,

Weitere Signale: 4.01-3.45 (*bulk*), 4.468 (d, von triantennärer Struktur), 2.09-2.07 + 2.054 (Signale von triantennärer Struktur), 4.10 + 1.32 (q+d, Lactat), 2.72 (s, DMSO), 2.91, 2.89, 2.59, 0.89, 0.157 (je s, unbekannt), 1.36-1.26 (m, unbekannt)

## Struktur 7: 4,3-Nonaisomer

C<sub>62</sub>H<sub>104</sub>N<sub>4</sub>O<sub>46</sub>; monoisotopisches MW=1640.59 g/mol

 $[M+2H]^{2+} = 821.3038$  (theoretisch: 821.3034)



<sup>1</sup> H-NMR (500 M	IHz, D <sub>2</sub> O, 300K,	externes A	Aceton, <sup>1</sup> H-S	pektrum mit	excitation	sculpting):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.188	d	2.4
		H-1b	n.d.	n.d.	n.d.
		NAc(a)	2.039	s	-
		NAc(b)	2.037	s	-
b-D-GlcpNAc	4	H-1(a)	4.613	d	8.4
		H-1(b)	4.604	d	9.0
		NAc(a)	2.082	s	-
		NAc(b)	2.080	s	-
b-D-Manp	4,4	H-2	4.249	dd	2,1:1.0
					2,3: 2.3
a-D-Manp	6,4,4	H-1	4.928	*	*
		H-2	4.112	dd	2,1:1.0
					2,3: 2.8
b-D-GlcpNAc	2,6,4,4	H- 1	4.582	d	8.7
		NAc	2.046	s	-
b-D-Galp	4,2,6,4,4	H-1	4.474	d	8.2
a-D-Manp	3,4,4	H-1	5.120	*	*
		H-2	4.199	dd	2,1:1.6
					2,3: 3.7
b-D-GlcpNAc	2,3,4,4	H-1	4.602	d	8.8
		NAc	2.044	s	-
b-D-Galp	3,2,3,4,4	H-1	4.445	d	8.2

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined

Weitere Signale: 4.00-3.42 (*bulk*), 4.10 + 1.32 (q+d, Lactat), 3.35, 0.156 (je s, unbekannt)

#### Struktur 8: 3,4-Nonaisomer

C<sub>62</sub>H<sub>104</sub>N<sub>4</sub>O<sub>46</sub>; monoisotopisches MW=1640.59 g/mol

 $[M+2H]^{2+} = 821.3038$  (theoretisch: 821.3034)



<sup>1</sup>H-NMR (500 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.188	d	3.3
		H-1b	n.d.	n.d.	n.d.
		NAc(a)	2.038	s	-
		NAc(b)	2.036	s	-
b-D-GlcpNAc	4	H-1(a)	4.611	d	9.0
		H-1(b)	n.d.	n.d.	n.d.
		NAc	2.079	s	-
b-D-Manp	4,4	H-2	4.250	dd	2,1:1.3
					2,3: 2.9
a-D-Manp	6,4,4	H-1	4.915	*	*
		H-2	4.118	dd	2,1:1.3
					2,3: 2.9
b-D-GlcpNAc	2,6,4,4	H- 1	4.603	d	8.2
		NAc	2.044	s	-
b-D-Galp	3,2,6,4,4	H-1	4.453	d	8.1
a-D-Manp	3,4,4	H-1	5.119	*	*
		H-2	4.189	dd	2,1:2.1
					2,3: 3.9
b-D-GlcpNAc	2,3,4,4	H-1	4.580	d	8.2
		NAc	2.050	s	
b-D-Galp	4,2,3,4,4	H-1	4.467	d	8.0

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined

Weitere Signale: 4.02-3.44 (*bulk*), 4.10 + 1.32 (*q*+*d*, Lactat), 4.35, 4.31 (je d, unbekannt), 2.13, 2.11 (je s, unbekannt), 2.25-2.15, 1.30-1.25 (m, unbekannt), 0.157 (s, unbekannt)

### Struktur 9: 3,3-Nonaisomer

C<sub>62</sub>H<sub>104</sub>N<sub>4</sub>O<sub>46</sub>; monoisotopisches MW=1640.59 g/mol

 $[M+2H]^{2+} = 821.3038$  (theoretisch: 821.3034)



<sup>1</sup>H-NMR (500 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.189	d	3.0
		H-1b	n.d.	n.d.	n.d.
		NAc	2.038	8	-
b-D-GlcpNAc	4	H-1(a)	n.d.	n.d.	n.d.
		H-1(b)	n.d.	n.d.	n.d.
		NAc	2.080	8	-
b-D-Manp	4,4	H-2	4.251	*	*
a-D-Manp	6,4,4	H-1	4.914	*	*
		H-2	n.d.	n.d.	n.d.
b-D-GlcpNAc	2,6,4,4	H- 1	4.603	d	9.0
		NAc	2.044	s	-
b-D-Galp	3,2,6,4,4	H-1	4.453	d	8.1
a-D-Manp	3,4,4	H-1	5.119	*	*
		H-2	4.19	*	*
b-D-GlcpNAc	2,3,4,4	H-1	4.600	d	8.7
		NAc	2.044	s	
b-D-Galp	3,2,3,4,4	H-1	4.445	d	9.0

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined

Weitere Signale: 3.97-3.44 (*bulk*), 4.10 + 1.32 (*q*+*d*, Lactat), 2.12, 2.11 (je s, unbekannt), 2.10-2.06, 1.30-1.25 (je m, unbekannt), 0.157 (s, unbekannt)

### Struktur 10: Triantennäre Struktur

C<sub>88</sub>H<sub>147</sub>N<sub>5</sub>O<sub>65</sub>; monoisotopisches MW=2313.84 g/mol

 $[M+H]^+ = 2313.8447, [M+2H]^{2+} = 1157.9302, [M+H+NH_4]^{2+} = 1166.4435, [M+2NH_4]^{2+} = 1174.9567$ 



Auswertung anhand des Korrelationsspektrums, basierend auf Datensatzaufnahme:

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.180	d	3.5
		H-1b	n.d.	n.d.	n.d.
		NAc	2.039	s	-
a-L-Fucp	6	H-1(a)	4.891	d	3.6
		H-1(b)	4.898	d	3.6
		H-5(a,b)	n.d.	n.d.	n.d.
		CH3(a)	1.210	d	6.9
		CH3(b)	1.220	d	6.9
b-D-GlcpNAc	4	H-1(a,b)	n.d.	n.d.	n.d.
		NAc(a)	2.097	s	-
		NAc(b)	2.094	s	-
b-D-Manp	4,4	H-2	4.22-4.17	m	*
a-D-Manp	6,4,4	H-1	4.922	*	*
		H-2	n.d.	n.d.	n.d.
b-D-GlcpNAc	2,6,4,4	H- 1	4.582	d	7.9
		NAc	2.048	s	-
b-D-Galp	4,2,6,4,4	H-1	4.545	d	8.4
		H-4	4.22-4.17	m	*
a-D-Galp	3,4,2,6,4,4	H-1	5.148	d	4.1#
		H-4	4.023	dd	4,3: 3.4
					4.5: 0.8
		H-5	4.22-4.17	m	*
a-D-Manp	3,4,4	H-1	5.118	*	*
		H-2	4.22-4.17	m	*
b-D-GlcpNAc	2,3,4,4	H-1	4.567	d	7.5
		NAc	2.048	s	-
b-D-Galp	4,2,3,4,4	H-1	4.463	d	8.5
b-D-GlcpNAc	4,3,4,4	H-1	4.545	d	8.4
*		NAc	2.078	s	-
b-D-Galp	4,4,3,4,4	H-1	4.468	d	8.1

<sup>1</sup> H-NMR (500 MHz, D <sub>2</sub> O, 300K, internes Aceton, <sup>1</sup> H-Spe	ektrum mit excitation sculpting)
--	----------------------------------

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined

Weitere Signale: 4.0-3.48 (*bulk*), 4.26-4.24 (Man-3, H-2, von Dodeca **4**), 2.91, 3.89 (s, unbekannt), 4.10+1.32 (q+d, Lactat), 0.157 (s, unbekannt)

### Strukturen der Bibliothek I-XIII

Die Proben wurden von Hexal zusammen mit einer Strukturinformation zur Verfügung gestellt. Eine Verifikation, sowie falls angegeben die Isomerenunterscheidung erfolgte anhand des *structural reporter group*-Konzepts.

Die Berechnung der Substanzmengen in den Gemischen erfolgte über die manuelle Bestimmung eines Beteiligungsfaktors des jeweiligen Referenzspektrums am Spektrum des jeweiligen Gemischs. Dieser wurde normiert über die Menge an verwendetem Lösungsmittel und die Anzahl an *scans*. Aus den Herstellerangaben über die Substanzmenge der Referenzverbindung wurde dann anhand des Faktors die Menge im Gemisch berechnet.

## Struktur I



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
D-GlcNAc	-	H-1a	5.182	d	4.0	
		H-1b	n.d.	n.d.	n.d.	
		NAc	2.039	s	-	
a-L-Fucp	6	H-1(a)	4.902	d	5.1	
		H-1(b)	4.909	d	4.8	
		H-5(a,b)	4.17-4.07	m	*	
		CH3(a)	1.212	d	7.4	
		CH3(b)	1.222	d	7.0	
b-D-GlcpNAc	4	H-1(a,b)	4.69-4.64	m	*	
		NAc(a)	2.094	s	-	
		NAc(b)	2.091	s	-	
b-D-Manp	4,4	H-2	4.204	dd	2,1:2.0	
					2,3: 3.5	
a-D-Manp	6,4,4	H-1	n.d.	n.d.	n.d.	
		H-2	4.17-4.07	m	*	
b-D-GlcpNAc	6,6,4,4	H- 1				b
		NAc	2.039	s	-	
b-D-Galp	4,6,6,4,4	H-1	4.560	d	9.2	
		H-3	4.17-4.07	m	*	

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
a-D-Neup5Ac	3,4,6,6,4,4	H-3a	1.805	dd	3a,3e: 13.9 <sup>\$</sup>	
					3a,4: 13.9 <sup>\$</sup>	
		H-3e	2.756	dd	3e,3a: 14.4 <sup>\$</sup>	
					3e,4: 5.4 <sup>\$</sup>	
		NAc	2.031	S	-	
b-D-GlcpNAc	2,6,4,4	H-1	4.594	d	9.0	
		NAc	2.039	S	-	
b-D-Galp	4,2,6,4,4	H-1				b
		H-3	4.17-4.07	m	*	
a-D-Neup5Ac	3,4,2,6,4,4	H-3a	1.805	dd	3a,3e: 13.9 <sup>\$</sup>	
					3a,4: 13.9 <sup>\$</sup>	
		H-3e	2.756	dd	3e,3a: 14.4 <sup>\$</sup>	
					3e,4: 5.4 <sup>\$</sup>	
		NAc	2.031	s	-	
a-D-Manp	3,4,4	H-1	5.130	d	1.5	
		H-2	4.220	dd	2,1: *	
					2,3: 3.2	
b-D-GlcpNAc	4,3,4,4	H-1				b
		NAc	2.075	s	-	
b-D-Gal	4,4,3,4,4	H-1				b
		H-3	4.17-4.07	m	*	
a-D-Neup5Ac	3,4,4,3,4,4	H-3a	1.805	dd	3a,3e: 13.9 <sup>\$</sup>	
					3a,4: 13.9 <sup>\$</sup>	
		H-3e	2.756	dd	3e,3a: 14.4 <sup>\$</sup>	
					3e,4: 5.4 <sup>\$</sup>	
		NAc	2.031	S	-	
b-D-GlcpNAc	2,3,4,4	H-1	4.560	d	9.2	
		NAc	2.048	S	-	
b-D-Galp	4,2,3,4,4	H-1				b
		H-3	4.17-4.07	m	*	
a-D-Neup5Ac	3,4,2,3,4,4	H-3a	1.805	dd	3a,3e: 13.9 <sup>\$</sup>	
					3a,4: 13.9 <sup>\$</sup>	
		H-3e	2.756	dd	3e,3a: 14.4 <sup>\$</sup>	
					3e,4: 5.4 <sup>\$</sup>	
		NAc	2.031	s	-	
b) 2 GlcNAc +		H-1	4.546	d	8.5	Zu b
3 Gal		H-1	4.542	d	7.9	

Weitere Signale: 4.07-3.35 (*bulk*), 4.17-4.07, 1.32 (m+d, Lactat), 1.91, 1.39, 1.24, 1.18 (s, unbekannt), 1.31-1.27, 0.89-0.82 (m, unbekannt), 0.157 (s, unbekannt)

## <u>Struktur II</u>



Vermutlich handelt es sich um ein Gemisch aus zwei Isomeren.

<sup>1</sup> H-NMR (700 MHz, D <sub>2</sub> O, 300K, externes Ac	eton, <sup>1</sup> H-Spektrum mit <i>excitation sculpting</i> ):
--	--

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
D-GlcNAc	-	H-1a	5.182	d	3.1	
		H-1b	n.d.	n.d.	n.d.	
		NAc	2.038	s	-	
a-L-Fucp	6	H-1(a)	4.900	d	5.3	
		H-1(b)	4.907	d	5.3	
		H-5(a)	n.d.	n.d.	n.d.	
		H-5(b)	4.15-4.12	m	*	
		CH3(a)	1.210	d	7.9	
		CH3(b)	1.221	d	7.5	
b-D-GlcpNAc	4	H-1(a,b)	4.70-4.64	m	*	
		NAc(a)	2.094/2.090	s/s	-/-	с
		NAc(b)	2.092/2.088	s/s	-/-	с
b-D-Manp	4,4	H-2	4.207	*	*	
					2,3: 4.0	
a-D-Manp	6,4,4	H-1	n.d.	n.d.	n.d.	
		H-2	4.10-4.07	m	*	
a-D-Manp	3,4,4	H-1	5.129	d	2.2	
		H-2	4.219	dd	2,1:2.0	
					2,3: 4.0	
b-D-GlcpNAc	4,3,4,4	NAc	2.075	s	-	
b-D-GlcpNAc	2,3,4,4	NAc	2.048	s	-	
5 b-D-Gal	3 ohne R	H-1	4.61-4.52	m	*	
		H-3	4.13-4.10	m	*	
	1 mit R	H-1	4.41-4.43	m	*	
		H-4	4.162/4.154	d/d	2.9/2.3	с
	1 im R	H-1	4.61-4.52	m	*	
		H-3	4.13-4.10	m	*	
5 b-D-GlcpNAc	4	H-1	4.61-4.52	m	*	
-	(2/6,6,4,4)	NAc	2.038	s	-	
	1 im R	H-1	n.d.	n.d.	n.d.	
		NAc	2.038	s	-	

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
5 a-D-Neup5Ac		H-3a	1.803	dd	3a,3e: 14.6 <sup>\$</sup>	
					3a,4: 14.6 <sup>\$</sup>	
		H-3e	2.756	dd	3e,3a: 14.2 <sup>\$</sup>	
					3e,4: 5.2 <sup>\$</sup>	
		NAc	2.031	s	-	

c - vermutlich zwei Substanzen, R - *repeat,* \* – Kopplungskonstante konnte nicht bestimmt werden, n.d. – *not determined*, \$ – vermutlich ist die Kopplungskonstante aufgrund eines Dispersionsanteils im Signal zu groß

Weitere Signale: 4.02-3.36 (*bulk*), 4.896 (d, 3.6 Hz) + 4.890 (d, 4.7 Hz, vermutl. Fuc H-1), 4.10 +1.32 (q+d, Lactat), 2.29-2.12 (m, unbekannt), 2.225 (s, Aceton), 2.06, 1.91, 1.41,1.24 (s, unbekannt), 1.34-1.25, 0.90-0.82 (m, unbekannt), 0.157 (s, unbekannt)

#### **Struktur III**



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, internes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.181	d	3.8
		H-1b	n.d.	n.d.	n.d.
		NAc	2.037	s	-
a-L-Fucp	6	H-1(a)	4.897	d	4.8
		H-1(b)	4.904	d	4.2
		H-5(a)	n.d.	n.d.	n.d.
		H-5(b)	4.14-4.12	m	*
		CH3(a)	1.209	d	7.5
		CH3(b)	1.220	d	7.1
b-D-GlcpNAc	4	H-1(a,b)	4.71-4.63	m	*
		NAc(a)	2.091	s	-
		NAc(b)	2.088	s	-
b-D-Manp	4,4	H-2	4.24-4.19	m	*
a-D-Manp	6,4,4	H-1	4.854	*	*
		H-2	4.09-4.07	m	*
a-D-Manp	3,4,4	H-1	5.128	*	*
		H-2	4.24-4.19	m	*
b-D-GlcpNAc	4,3,4,4	H-1	4.540	d	8.3
		NAc	2.075	s	-
b-D-GlcpNAc	2,3,4,4	H-1	4.61-4.50	m	*
-		NAc	2.047	s	-

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
6 b-D-Gal	2 ohne R	H-1	4.540	d	8.3
		H-3	4.116	d	3,2: 10.9
					3,4: 3.4
	2 mit R	H-1	4.48-4.44	m	*
		H-4	4.161	d	4,3:4.0
	2 im R	H-1	4.556	d	8.3
		H-3	4.116	d	3,2: 10.9
					3,4: 3.4
4 b-D-GlcpNAc	2 (2/6,6,4,4)	H-1	4.61-4.50	m	*
		NAc	2.037/2.035	s/s	-/-
	2 im R	H-1	4.71-4.63	m	*
		NAc	2.037/2.035	s/s	-/-
5 a-D-Neup5Ac		H-3a	1.800	dd	3a,3e: 14.7 <sup>\$</sup>
					3a,4: 14.7 <sup>\$</sup>
		H-3e	2.757	dd	3e,3a: 13.6 <sup>\$</sup>
					3e,4: 4.5
		NAc	2.030	s	-

Weitere Signale: 4.02-3.37 (*bulk*) 4.84 - 4.82, 4.37-4.32, 2.29-2.12, 1.31-1.25, 0.89-0.83 (m, unbekannt), 2.37, 2.06, 1.91, 1.42, 1.40, 1.39, 1.24, 1.20, 1.18, 1.16 (s, unbekannt), 2.225 (s, Aceton), 1.32 (d, Lactat), 0.98 (d, unbekannt), 0.157 (s, unbekannt)

## **Struktur IV**



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
D-GlcNAc	-	H-1a	5.191	d	3.5	
		H-1b	n.d.	n.d.	n.d.	
		NAc	2.038	s	-	
b-D-GlcpNAc	4	H-1(a,b)	n.d.	n.d.	n.d.	
		NAc	2.080	s	-	
b-D-Manp	4,4	H-2	4.204	dd	2,1:1.5	
					2,3: 3.5	
a-D-Manp	6,4,4	H-1	4.860	*	*	
		H-2	4.09	dd	2,1:0.9	
					2,3:*	

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
b-D-GlcpNAc	6,6,4,4	H- 1	4.548	d	8.5	
		NAc	2.038	8	-	
b-D-Galp	4,6,6,4,4	H-1	4.56	*	*	
		H-3	4.119	*	*	
a-D-Neup5Ac	3,4,6,6,4,4	H-3a	1.804	dd	3a,3e: 13.5 <sup>\$</sup>	
					3a,4: 13.5 <sup>\$</sup>	
		H-3e	2.756	dd	3e,3a: 14.1 <sup>\$</sup>	
					3e,4: 5.6 <sup>\$</sup>	
		NAc	2.031	8	-	
b-D-GlcpNAc	2,6,4,4	H-1	4.594	d	8.5	
		NAc	2.038	8	-	
b-D-Galp	4,2,6,4,4	H-1	4.547	d	8.5	а
		H-3	4.119	*	*	
a-D-Neup5Ac	3,4,2,6,4,4	H-3a	1.804	dd	3a,3e: 13.5 <sup>\$</sup>	
					3a,4: 13.5 <sup>\$</sup>	
		H-3e	2.756	dd	3e,3a: 14.1 <sup>\$</sup>	
					3e,4: 5.6 <sup>\$</sup>	
		NAc	2.031	8	-	
a-D-Manp	3,4,4	H-1	5.130	*	*	
		H-2	4.220	*	*	
		H-3	n.d.	n.d.	n.d.	
b-D-GlcpNAc	4,3,4,4	H-1	4.542	d	9.4	а
		NAc	2.074	s	-	
b-D-Gal	4,4,3,4,4	H-1	4.542	d	9.4	а
		H-3	4.119	*	*	
a-D-Neup5Ac	3,4,4,3,4,4	H-3a	1.804	dd	3a,3e: 13.5 <sup>\$</sup>	
					3a,4: 13.5 <sup>\$</sup>	
		H-3e	2.756	dd	3e,3a: 14.1 <sup>\$</sup>	
					3e,4: 5.6 <sup>\$</sup>	
		NAc	2.031	s	-	
b-D-GlcpNAc	2,3,4,4	H-1	4.564	d	8.8	
		NAc	2.048	s	-	
b-D-Galp	4,2,3,4,4	H-1	4.542	d	9.4	а
		H-3	4.119	*	*	
a-D-Neup5Ac	3,4,2,3,4,4	H-3a	1.804	dd	3a,3e: 13.5 <sup>\$</sup>	
-					3a,4: 13.5 <sup>\$</sup>	
		H-3e	2.756	dd	3e,3a: 14.1 <sup>\$</sup>	
					3e,4: 5.6 <sup>\$</sup>	
		NAc	2.031	S	-	

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined
\$ - vermutlich ist die Kopplungskonstante aufgrund eines Dispersionsanteils im
Signal zu groß, a - sugabase: Assignments may have to be interchanged

Weitere Signale: 4.01-3.49 (bulk), (4.10+1.32 (q+d, Lactat), 4.08, 4.05 (d, unbekannt), 3.44,3.40, 3.38, 2.06, 1.91, 1.24, 1.20 (s, unbekannt), 1.35-1.25, 0.91-0.82 (m, unbekannt), 0.157 (s, unbekannt)

## <u>Struktur V</u>



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, internes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.181	d	3.4
		H-1b	n.d.	n.d.	n.d.
		NAc	2.039	s	-
a-L-Fucp	6	H-1(a)	4.899	d	4.7
-		H-1(b)	4.906	d	4.9
		H-5(a,b)	4.16-4.07	m	*
		CH3(a)	1.211	d	7.7
		CH3(b)	1.221	d	7.2
b-D-GlcpNAc	4	H-1(a,b)	4.68-4.64	m	*
		NAc(a)	2.094	s	-
		NAc(b)	2.091	s	-
b-D-Manp	4,4	H-2	4.252	dd	2,1:1.3
					2,3: 3.4
a-D-Manp	6,4,4	H-1	4.871	*	*
		H-2	4.16-4.07	m	*
b-D-GlcpNAc	6,6,4,4	H-1	4.544	d	9.2
		NAc	2.039	s	-
b-D-Galp	4,6,6,4,4	H-1	4.559	d	9.5
		H-3	4.16-4.07	m	*
a-D-Neu5pAc	3,4,6,6,4,4	H-3a	1.803	dd	3a,3e: 14.2 <sup>\$</sup>
					3a,4: 14.2 <sup>\$</sup>
		H-3e	2.757	dd	3e,3a: 14.0 <sup>\$</sup>
					3e,4: 5.2 <sup>\$</sup>
		NAc	2.031	s	-
b-D-GlcpNAc	2,6,4,4	H-1	4.588	d	7.9
		NAc	2.039	S	-
b-D-Galp	4,2,6,4,4	H-1	4.544	d	9.2
		H-3	4.16-4.07	m	*
a-D-Neup5Ac	3,4,2,6,4,4	H-3a	1.803	dd	3a,3e: 14.2 <sup>\$</sup>
					3a,4: 14.2 <sup>\$</sup>
		H-3e	2.757	dd	3e,3a: 14.0 <sup>\$</sup>
					3e,4: 5.2 <sup>\$</sup>
		NAc	2.031	S	-
a-D-Manp	3,4,4	H-1	5.122	d	1.4
		H-2	4.195	d	2,1:1.7
					2,3: 3.7
b-D-GlcpNAc	2,3,4,4	H-1	4.578	d	9.2
		NAc	2.052	S	-
b-D-Galp	4,2,3,4,4	H-1	4.544	d	9.2
		H-3	4.16-4.07	m	*

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
a-D-Neup5Ac	3,4,2,3,4,4	H-3a	1.803	dd	3a,3e: 14.2 <sup>\$</sup>
					3a,4: 14.2 <sup>\$</sup>
		H-3e	2.757	dd	3e,3a: 14.0 <sup>\$</sup>
					3e,4: 5.2 <sup>\$</sup>
		NAc	2.031	s	-

Weitere Signale: 4.03-3.37 (*bulk*), 4.16-4.07+1.32 (m+d, Lactat), 2.225 (s, Aceton), 2.063 (NAc-Spezies), 1.91, 0.157 (s, unbekannt)

### Struktur VI (a,b)



Mischung aus zwei Strukturen: repeat 3,4,6,6,4,4 (6R) und 3,4,2,6,4,4 (2R) verknüpft

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
D-GlcNAc	-	H-1a	5.182	d	3.2	6R+2R
		H-1b	n.d.	n.d.	n.d.	6R+2R
		NAc	2.038	s	-	6R+2R
a-L-Fucp	6	H-1(a,b)	4.91-4.88	m	*	6R+2R
		H-5(a,b)	4.15-4.07	m	*	6R+2R
		CH3(a)	1.210	d	7.7	6R+2R
		CH3(b)	1.220	d	7.3	6R+2R
b-D-GlcpNAc	4	H-1(a,b)	4.71-4.64	m	*	6R+2R
		NAc(a)	2.092	s	-	6R+2R
		NAc(b)	2.090	s	-	6R+2R
b-D-Manp	4,4	H-2	4.27-4.23	m	*	6R+2R
a-D-Manp	6,4,4	H-1	4.869	*	*	Mitte
						6R+2R
		H-2	4.15-4.07	m	*	6R+2R
b-D-GlcpNAc	6,6,4,4	H-1	4.554	d	8.1	R6+R2
		NAc	2.038	s	-	6R+2R
b-D-Galp	4,6,6,4,4	H-1	4.467/4.560	d/d	7.9/9.3	6R/2R
		H-3	n.d./	n.d./m	n.d./*	6R/2R
			4.15-4.07			
		H-4	4.17-4.15/	m/n.d.	*/n.d.	6R/2R
			n.d.			

## <sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, internes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

		_				
Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
a-D-Neu5pAc	3,4,6,6,4,4	H-3a	1.800	dd	3a,3e: 14.9 <sup>\$</sup>	Nur 2R
					3a,4: 14.9 <sup>\$</sup>	
		H-3e	2.757	dd	3e,3a: 14.5 <sup>\$</sup>	Nur 2R
					3e,4: 6.0 <sup>\$</sup>	
		NAc	2.031	S	-	Nur 2R
b-D-GlcpNAc	2,6,4,4	H-1	n.d.	n.d.	n.d.	6R+2R
		NAc	2.038	S	-	6R+2R
b-D-Galp	4,2,6,4,4	H-1	4.554/4.445	d/d	8.1/8.5	6R/2R
		H-3	4.15-4.07/	m/n.d.	*/n.d.	6R/2R
			n.d.			
		H-4	n.d./	n.d./m	n.d./*	6R/2R
			4.17-4.15			
a-D-Neup5Ac	3,4,2,6,4,4	H-3a	1.800	dd	3a,3e: 14.9 <sup>\$</sup>	Nur 6R
1					3a,4: 14.9 <sup>\$</sup>	
		H-3e	2.757	dd	3e.3a: 14.5 <sup>\$</sup>	Nur 6R
					$3e.4:6.0^{\$}$	
		NAc	2.031	s	-	Nur 6R
a-D-Manp	3.4.4	H-1	5.126/5.122	*/*	*/*	6R/2R
1	, ,	H-2	4.21-4.19	m	*	6R+2R
b-D-GlcpNAc	2.3.4.4	H-1	n.d./4.578	n.d./d	/8.5	6R/2R
	7-7-7	NAc	2.051	S	-	6R+2R
b-D-Galp	4,2,3,4,4	H-1	4.554	d	8.1	6R+2R
I.		H-3	4.15-4.07	m	*	2R+6R
a-D-Neup5Ac	3.4.2.3.4.4	H-3a	1.800	dd	3a.3e: 14.9 <sup>\$</sup>	6R+2R
1					$3a.4: 14.9^{\$}$	
		H-3e	2.757	dd	$3e.3a: 14.5^{\$}$	6R+2R
					$3e.4:6.0^{\$}$	
		NAc	2.031	s	-	6R+2R
Repeat	$6R \cdot ? = 6 2R \cdot ? = 2$	)		-		
h-GlenNAc	347644	- H-1	nd	nd	n d	6R+2R
0-OlepitAe	5,7,1,0,7,7	NAc	2.038	n.u.	n.u.	6R+2R
h D Colp	1212611		2.038	3 d/d	- 9 1/0 /	6D/2D
b-D-Gaip	4,5,4,?,0,4,4		4.330/4.300	d/d	0.1/9.4 *	$\frac{0R}{2R}$
		н-э	4.15-4.07	m	*	ZR+0R
Neup5Ac	3,4,3,4,?,6,4,4	H-3a	1.800	dd	3a,3e: 14.9*	6R+2R
					3a,4: 14.9*	
		H-3e	2.757	dd	3e,3a: 14.5*	6R+2R
					3e,4: 6.0 <sup>\$</sup>	
		NAc	2.031	s	-	6R+2R

Weitere Signale: 4.03-3.37 (*bulk*), 4.15-4.07+1.32 (m+d, Lactat), 2.225 (s, Aceton), 2.062 (NAc-Spezies), 1.91 (s, unbekannt), 1.43-1.24, 1.20-1.16 (m, unbekannt), 0.156 (s, unbekannt)

## Struktur VII



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
D-GlcNAc	-	H-1a	5.190	d	3.4	
		H-1b	n.d.	n.d.	n.d.	
		NAc	2.038	s	-	
b-D-GlcpNAc	4	H-1(a)	4.605	d	9,1	
		H-1(b)	n.d.	n.d.	n.d.	
		NAc(a)	2.080	s	-	
		NAc(b)	2.079	S	-	
b-D-Manp	4,4	H-2	4.250	*	*	
a-D-Manp	6,4,4	H-1	4.874	*	*	
		H-2	n.d.	n.d.	n.d.	
b-D-GlcpNAc	6,6,4,4	H-1				b)
		NAc	2.038	s	-	
b-D-Galp	4,6,6,4,4	H-1	4.559	d	8.1	
a-D-Neu5pAc	3,4,6,6,4,4	H-3a	1.802	dd	3a,3e: 14.4 <sup>\$</sup>	
					3a,4: 14.4 <sup>\$</sup>	
		H-3e	2.757	dd	3e,3a: 13.9 <sup>\$</sup>	
					3e,4: 4.5	
		NAc	2.031	s	-	
b-D-GlcpNAc	2,6,4,4	H-1	4.588	d	9.3	
		NAc	2.038	S	-	
b-D-Galp	4,2,6,4,4	H-1				b)
a-D-Neup5Ac	3,4,2,6,4,4	H-3a	1.802	dd	3a,3e: 14.4 <sup>\$</sup>	
					3a,4: 14.4 <sup>\$</sup>	
		H-3e	2.757	dd	3e,3a: 13.9 <sup>\$</sup>	
					3e,4: 4.5	
		NAc	2.031	S	-	
a-D-Manp	3,4,4	H-1	5.122	d	1.3	
		H-2	4.197	*	*	
b-D-GlcpNAc	2,3,4,4	H-1	4.578	d	7.9	
		NAc	2.052	S	-	
b-D-Galp	4,2,3,4,4	H-1				b)
a-D-Neup5Ac	3,4,2,3,4,4	H-3a	1.802	dd	3a,3e: 14.4 <sup>\$</sup>	
					3a,4: 14.4 <sup>\$</sup>	
		H-3e	2.757	dd	3e,3a: 13.9 <sup>\$</sup>	
					3e,4: 4.5	
		NAc	2.031	s	-	
3 b-D-Galp		H-3	4.119	dd	3,2: 10.4	
					3,4: 3.3	
		H-3	4.114	dd	3,2: 10.6	
					3,4: 3.1	

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
b) 2 b-Galp +		H-1	4.547	d	8.3	Zu b)
GlcpNAc		H-1	4.544	d	8.3	

Weitere Signale: 4.04-3.36 (*bulk*), 4.10-4.08 (m, unbekannt), 2.062 (NAc-Spezies), 0.91-0.83, 1.43-1.13 (m, unbekannt), 0.157 (s, unbekannt)

## **Struktur VIII**



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.181	d	3.2
		H-1b	n.d.	n.d.	n.d.
		NAc	2.038	s	-
a-L-Fucp	6	H-1(a)	4.889	d	4.3
		H-1(b)	4.895	d	4.9
		H-5(a)	4.10	*	*
		H-5(b)	4.134	*	*
		CH3(a)	1.209	d	7.4
		CH3(b)	1.220	d	6.9
b-D-GlcpNAc	4	H-1(a)	4.666	d	8.8
		H-1(b)	4.669	d	9.1
		NAc(a)	2.099	s	-
		NAc(b)	2.095	s	-
b-D-Manp	4,4	H-2	4.257	dd	2,1:1.3
					2,3: 3.1
a-D-Manp	6,4,4	H-1	4.941	*	*
		H-2	4.114	dd	2,1:1.4
					2,3: 3.3
b-D-GlcpNAc	2,6,4,4	H-1	4.604	d	8.3
		NAc	2.066	s	-
b-D-Galp	4,2,6,4,4	H-1	4.446	d	8.8
a-D-Neup5Ac	6,4,2,6,4,4	H-3a	1.719	dd	3a,3e: 13.4 <sup>\$</sup>
					3a,4: 13.4 <sup>\$</sup>
		H-3e	2.673	dd	3e,3a: 13.1 <sup>\$</sup>
					3e,4: 3.8
		NAc	2.030	s	-
a-D-Manp	3,4,4	H-1	5.132	d	1.9
-		H-2	4.197	dd	2,1:1.5
					2,3: 4.0

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
b-D-GlcpNAc	2,3,4,4	H-1	4.604	d	8.3
		NAc	2.069	s	-
b-D-Galp	4,2,3,4,4	H-1	4.442	d	8.8
a-D-Neup5Ac	6,4,2,3,4,4	H-3a	1.719	dd	3a,3e: 13.4 <sup>\$</sup>
					3a,4: 13.4 <sup>\$</sup>
		H-3e	2.666	dd	3e,3a: 13.5 <sup>\$</sup>
					3,4: 3.4
		NAc	2.030	s	-

Weitere Signale: 4.02-3.48 (*bulk*), 4.10 + 1.32 (q+d, Lactat), 1.91, 0.157 (je. s, unbekannt), 1.35-1.25 (m, unbekannt)

### Struktur IX



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.189	d	3.4
		H-1b	n.d.	n.d.	n.d.
		NAc (a)	2.038	S	-
		NAc(b)	2.036	S	-
b-D-GlcpNAc	4	H-1	n.d.	n.d.	n.d.
		NAc(a)	2.085	s	-
		NAc(b)	2.083	S	-
b-D-Manp	4,4	H-2	4.257	dd	2,1:1.4
					2,3: 3.5
a-D-Manp	6,4,4	H-1	4.948	*	*
		H-2	4.119	dd	2,1:1.2
					2,3: 3.1
b-D-GlcpNAc	2,6,4,4	H-1	4.604	d	8.8
		NAc	2.066	S	-
b-D-Galp	4,2,6,4,4	H-1	4.448	d	8.8
a-D-Neup5Ac	6,4,2,6,4,4	H-3a	1.719	dd	3a,3e: 13.7
					3a,4: 12.8
		H-3e	2.673	dd	3e,3a: 13
					3e,4: 3.6
		NAc	2.030	S	-
a-D-Manp	3,4,4	H-1	5.132	*	*
		H-2	4.198	*	*
b-D-GlcpNAc	2,3,4,4	H-1	4.604	d	8.8
		NAc	2.069	S	-

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
b-D-Galp	4,2,3,4,4	H-1	4.443	d	8.5
a-D-Neup5Ac	6,4,2,3,4,4	H-3a	1.719	dd	3a,3e: 13.7 <sup>\$</sup>
					3a,4: 12.8
		H-3e	2.667	dd	3e,3a: 13.8 <sup>\$</sup>
					3e,4: 4.1
		NAc	2.030	s	-

Weitere Signale: 4.00-3.50 (*bulk*), 4.10 + 1.32 (q+d, Lactat), 2.44, 3.38, 0.157 (s, unbekannt)

## Struktur X



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.180	d	3.6
		H-1b	n.d.	n.d.	n.d.
		NAc	2.038	s	-
a-L-Fucp	6	H-1(a)	4.889	d	4.5
		H-1(b)	4.896	d	4.7
		H-5(a)	4.096	*	*
		H-5(b)	4.134	q	7.9
		CH3(a)	1.209	d	7.3
		CH3(b)	1.219	d	7.0
b-D-GlcpNAc	4	H-1	n.d.	n.d.	n.d.
		NAc(a)	2.097	s	-
		NAc(b)	2.094	s	-
b-D-Manp	4,4	H-2	4.255	dd	2,1:*
					2,3: 2.8
a-D-Manp	6,4,4	H-1	4.928	*	*
		H-2	4.11	*	*
b-D-GlcpNAc	2,6,4,4	H-1	4.581	d	9.2
		NAc	2.048	s	-
b-D-Galp	4,2,6,4,4	H-1	4.544	d	9.0
		H-4	4.185	*	*
a-D-Galp	3,4,6,4,4	H-1	5.146	d	4.5#
-		H-4	4.023	dd	4,3: 3.7
					4,3: 1.2
		H-5	4.21-4.18	m*	*
a-D-Manp	3,4,4	H-1	5.134	d	0.8
-		H-2	4.195	*	*

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
b-D-GlcpNAc	2,3,4,4	H-1	4.605	d	9.2
		NAc	2.068	S	-
b-D-Galp	4,2,3,4,4	H-1	4.445	d	8.5
a-D-Neup5Ac	6,4,2,3,4,4	H-3a	1.720	dd	3a,3e: 13.1 <sup>\$</sup>
					3a,4: 13.1 <sup>\$</sup>
		H-3e	2.667	dd	3e,3a: 15.2 <sup>\$</sup>
					3e,4: 6.7 <sup>\$</sup>
		NAc	2.029	S	-

 \* – Kopplungskonstante konnte nicht bestimmt werden, n.d. – not determined
 \$ – vermutlich ist die Kopplungskonstante aufgrund eines Dispersionsanteils im Signal zu groß, <sup>#</sup> – vermutlich Anteil virtueller Kopplung (Spinsystem höherer Ordnung)

Weitere Signale: 4.02-3.42 (*bulk*), 4.10 + 1.32 (q+d, Lactat), 0.157 (s, unbekannt), 1.35-1.25 (m, unbekannt)

### Struktur XI



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
D-GlcNAc	-	H-1a	5.181	d	3.2	
		H-1b	n.d.	n.d.	n.d.	
		NAc	2.039	s	-	с
a-L-Fucp	6	H-1(a,b)	4.94-4.88	m	*	
		H-5(a,b)	4.14-4.06	m	*	
		CH3(a)	1.211	d	7.5	
		CH3(b)	1.226	d	7.2	
b-D-GlcpNAc	4	H-1(a,b)	4.70-4.63	m	*	
		NAc(a)	2.096	s	-	с
		NAc(b)	2.093	S	-	с
b-D-Manp	4,4	H-2	4.251	*	*	
a-D-Manp	6,4,4	H-1	4.94-4.88	m	*	
		H-2	4.14-4.06	m	*	
b-D-GlcpNAc	2,6,4,4	H-1	4.60-4.48	m	*	
		NAc	n.d.	n.d.	-	с
b-D-Galp	4,2,6,4,4	H-1	4.60-4.48	m	*	
a-D-Neup5Ac	3,4,2,6,4,4	H-3a	1.797	m	*	
-		H-3e	2.78-2.74	*	*	
		NAc	2.031	S	-	с

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)	Bemerkung
a-D-Manp	3,4,4	H-1	5.13-5.09	m	*	
		H-2	4.187	dd	2,1: *	
					2,3: 3.4	
b-D-GlcpNAc	2,3,4,4	H-1	4.60-4.48	m	*	
		NAc	n.d.	n.d.	-	с
b-D-Galp	4,2,3,4,4	H-1	4.60-4.48	m	*	
a-D-Neup5Ac	3,4,2,3,4,4	H-3a	1.797	m	*	
		H-3e	2.78-2.74	*	*	
		NAc	2.031	S	-	с
a-L-Fucp	3,2,?,4,4	H-1	5.13-5.09	m	*	
		H-5	4.94-4.88	m	*	
		CH3(a)	1.165	d	6.6	
		CH3(b)	1.171	d	7.3	

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined,
 c - siehe Kommentar

Kommentar zu c:

Die *N*-Acetylregion weist 10 Signale unterschiedlicher Intensität auf. Außerdem wurden viele Signalgruppen nur als Multipletts erhalten. Es ist deshalb zu vermuten, dass die Verbindung nicht rein vorlag.

*N*-Acetylregion - Signal in ppm (relatives Integral):

2.099 (0.87), 2.096 + 2.093 (3), 2.062 (0.5), 2.048 (1.3), 2.044 (2.0), 2.041+2.039+2.038 (7.2), 2.031 (7.8)

Weitere Signale: 4.03-3.44 (*bulk*), 4.14-4.06 + 1.32 (m+d, Lactat), 1.91 (s, unbekannt), 0.156 (s, unbekannt)

## **Struktur XII**



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.191	d	3.3
		H-1b	n.d.	n.d.	n.d.
		NAc (a)	2.037	8	-
		NAc(b)	2.036	8	-
b-D-GlcpNAc	4	H-1(a)	4.597	d	8.5
		H-1(b)	4.590	d	9.1
		NAc(a)	2.063	s	-
		NAc(b)	2.061	8	-
b-D-Manp	4,4	H-1	n.d.	n.d.	n.d.
		H-2	4.210	dd	2,1: *
					2,3: 3.1
4 a-D-Manp		H-1	5.408	d	1.4
			5.110	*	*
			5.030	*	*
			4.903	d	1.1
		H-2	4.142	dd	2,1: 1.6
					2,3: 3.7
			4.103	*	*
		2 Signale	4.08-4.06	m	*
Unbekannt			4.009	dd	3.7
					10.7

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined

Weitere Signale: 3.99-3.51 (*bulk*), 4.10 + 1.32 (q+d, Lactat), 3.44, 3.43, 1.91 je (s, unbekannt), 4.14-4.11, 1.34-1.15 (m, unbekannt), 0.157 (s, unbekannt)

## **Struktur XIII**



<sup>1</sup>H-NMR (700 MHz, D<sub>2</sub>O, 300K, externes Aceton, <sup>1</sup>H-Spektrum mit *excitation sculpting*):

Rest	Verknüpfung	Proton	ppm	Multiplizität	J (Hz)
D-GlcNAc	-	H-1a	5.191	d	3.6
		H-1b	n.d.	n.d.	n.d.
		NAc (a)	2.037	8	-
		NAc(b)	2.036	s	-
b-D-GlcpNAc	4	H-1(a)	4.598	d	9.3
		H-1(b)	4.590	d	9.2
		NAc(a)	2.063	8	-
		NAc(b)	2.062	8	-
b-D-Manp	4,4	H-1	n.d.	n.d.	n.d.
		H-2	4.211	dd	2,1: 1.9
					2.3: 3.8
5 a-D-Manp		H-1	5.424	d	1.7
			5.103	*	*
			5.024	*	*
			4.916	*	*
			4.877	*	*
		H-2	4.152	dd	2,1: 1.9
					2,3: 3.6
			4.071	dd	2,1:1.8
					2,3: 3.8
			4.065	dd	2,1: 1.8
					2,3: 3.5
			4.112	dd	2,1:1.5
					2,3: 3.2
			n.d.	n.d.	n.d.

\* - Kopplungskonstante konnte nicht bestimmt werden, n.d. - not determined

Weitere Signale: 4.03-3.52 (*bulk*), 4.10+1.32 (q+d, Lactat), 3.44, 3.43, 3.79, 3.17, 0.157 (s, unbekannt), 1.36-1.16 0.89- 0.83 (m, unbekannt)

# 7.7 <u>Toxikologie</u>

Deuteriumoxid Kein gefährliches Produkt im Sinne der Richtlinie 67/548/EWG

## 7.8 KMR-Substanzen

Es wurden keine KMR-Substanzen der Kategorien GHS 1A und 1B verwendet.
# 8 Literatur

- 1. Hirabayashi, J.; Arata, Y.; Kasai, K., Glycome project: concept, strategy and preliminary application to Caenorhabditis elegans. *Proteomics* **2001**, *1*, 295.
- 2. von der Lieth, C. W.; Bohne-Lang, A.; Lohmann, K. K.; Frank, M., Bioinformatics for glycomics: status, methods, requirements and perspectives. *Brief Bioinform* **2004**, *5*, 164.
- 3. von der Lieth, C.-W.; Luetteke, T.; Frank, M.; (Editors) *Bioinformatics For Glycobiology And Glycomics: An Introduction*; John Wiley & Sons Ltd.: Chichester (West Sussex), **2009**. S. 12-13, 143ff, 196ff, 295ff.
- 4. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1999**, *1473*, 4.
- 5. Endo, T., Human genetic deficits in glycan formation. *Proc. Jpn. Acad. B* 2004, 80, 128.
- 6. Bedford, C.; Cas, T.; Francois, I.; Harding, S., Glycomics: From glycobiology to diagnostics and therapeutics. *Drug News Perspect.* **2006**, *19*, 163.
- 7. Freeze, H. H.; Aebi, M., Altered glycan structures: the molecular basis of congenital disorders of glycosylation. *Curr Opin Struc Biol* **2005**, *15*, 490.
- 8. Helenius, A.; Aebi, M., Intracellular functions of N-linked glycans. *Science* **2001**, *291*, 2364.
- Petrescu, A. J.; Butters, T. D.; Reinkensmeier, G.; Petrescu, S.; Platt, F. M.; Dwek, R. A.; Wormald, M. R., The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding. *EMBO J* 1997, *16*, 4302.
- 10. Varki, A., Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **1993**, *3*, 97.
- 11. Wyss, D. F.; Wagner, G., The structural role of sugars in glycoproteins. *Curr Opin Biotechnol* **1996**, 7, 409.
- 12. Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A., Glycosylation and the immune system. *Science* **2001**, *291*, 2370.
- 13. Roseman, S., Reflections on glycobiology. *J Biol Chem* **2001**, 276, 41527.
- 14. Chandrasekaran, A.; Srinivasan, A.; Raman, R.; Viswanathan, K.; Raguram, S.; Tumpey, T. M.; Sasisekharan, V.; Sasisekharan, R., Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat Biotechnol* **2008**, *26*, 107.
- 15. Vigerust, D. J., Protein glycosylation in infectious disease pathobiology and treatment. *Cent Eur J Biol* **2011**, *6*, 802.
- 16. Dube, D. H.; Bertozzi, C. R., Glycans in cancer and inflammation. Potential for therapeutics and diagnostics. *Nat. Rev. Drug Discovery* **2005**, *4*, 477.
- 17. Taniguchi, N.; Paulson, J. C., Frontiers in glycomics; bioinformatics and biomarkers in disease. *Proteomics* **2007**, *7*, 1360.
- 18. Fernandes, B.; Sagman, U.; Auger, M.; Demetrio, M.; Dennis, J. W., Beta 1-6 branched oligosaccharides as a marker of tumor progression in human breast and colon neoplasia. *Cancer Res* **1991**, *51*, 718.
- 19. Dennis, J. W.; Granovsky, M.; Warren, C. E., Glycoprotein glycosylation and cancer progression. *Biochim Biophys Acta* **1999**, *1473*, 21.
- 20. Shinkawa, T.; Nakamura, K.; Yamane, N.; Shoji-Hosaka, E.; Kanda, Y.; Sakurada, M.; Uchida, K.; Anazawa, H.; Satoh, M.; Yamasaki, M.; Hanai, N.; Shitara, K., The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. *J Biol Chem* **2003**, *278*, 3466.
- 21. Niwa, R.; Hatanaka, S.; Shoji-Hosaka, E.; Sakurada, M.; Kobayashi, Y.; Uehara, A.; Yokoi, H.; Nakamura, K.; Shitara, K., Enhancement of the antibody-dependent cellular cytotoxicity of low-fucose IgG1 Is independent of FcgammaRIIIa functional polymorphism. *Clin Cancer Res* **2004**, *10*, 6248.

- Platts-Mills, T. A. E.; Chung, C. H.; Mirakhur, B.; Chan, E.; Le, Q.; Berlin, J.; Morse, M.; Murphy, B. A.; Satinover, S. M.; Hosen, J.; Mauro, D.; Slebos, R. J.; Zhou, Q. W.; Gold, D.; Hatley, T.; Hicklin, D. J., Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. *New Engl J Med* 2008, *358*, 1109.
- 23. Thall, A.; Galili, U., Distribution of Gal $\alpha$ 1-3Gal $\beta$ 1-4GlcNAc residues on secreted mammalian glycoproteins (thyroglobulin, fibrinogen, and immunoglobulin G) as measured by a sensitive solid-phase radioimmunoassay. *Biochemistry* **1990**, *29*, 3959.
- 24. Macher, B. A.; Galili, U., The Gal alpha 1,3Gal beta 1,4GlcNAc-R (alpha-Gal) epitope: A carbohydrate of unique evolution and clinical relevance. *Bba-Gen Subjects* **2008**, *1780*, 75.
- 25. Takeuchi, M.; Kobata, A., Structures and functional roles of the sugar chains of human erythropoietins. *Glycobiology* **1991**, *1*, 337.
- 26. Jefferis, R., Glycosylation as a strategy to improve antibody-based therapeutics. *Nat. Rev. Drug Discovery* **2009**, *8*, 226.
- 27. Timmer, M. S.; Stocker, B. L.; Seeberger, P. H., Probing glycomics. Curr Opin Chem Biol 2007, 11, 59.
- 28. Werz, D. B.; Ranzinger, R.; Herget, S.; Adibekian, A.; von der Lieth, C. W.; Seeberger, P. H., Exploring the structural diversity of mammalian carbohydrates ("glycospace") by statistical databank analysis. *ACS Chem Biol* **2007**, *2*, 685.
- 29. Lütteke, T.; Bohne-Lang, A.; Loss, A.; Goetz, T.; Frank, M.; von der Lieth, C. W., GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiology research. *Glycobiology* **2006**, *16*, 71R.
- 30. Doubet, S.; Bock, K.; Smith, D.; Darvill, A.; Albersheim, P., The Complex Carbohydrate Structure Database. *Trends Biochem Sci* **1989**, *14*, 475.
- 31. Doubet, S.; Albersheim, P., Carbbank. *Glycobiology* **1992**, *2*, 505.
- 32. Brooks, S. A., Strategies for analysis of the glycosylation of proteins: current status and future perspectives. *Mol Biotechnol* **2009**, *43*, 76.
- 33. Jiang, H.; Desaire, H.; Butnev, V. Y.; Bousfield, G. R., Glycoprotein profiling by electrospray mass spectrometry. *J Am Soc Mass Spectrom* **2004**, *15*, 750.
- 34. Dwek, R. A., Glycobiology: Toward Understanding the Function of Sugars. *Chem. Rev.* **1996**, *96*, 683.
- 35. Tarentino, A. L.; Plummer, T. H., Jr., Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from Flavobacterium meningosepticum. *Methods Enzymol* **1994**, *230*, 44.
- 36. Patel, T.; Bruce, J.; Merry, A.; Bigge, C.; Wormald, M.; Jaques, A.; Parekh, R., Use of Hydrazine to Release in Intact and Unreduced Form Both N-Linked and O-Linked Oligosaccharides from Glycoproteins. *Biochemistry* **1993**, *32*, 679.
- 37. Morelle, W.; Michalski, J. C., Analysis of protein glycosylation by mass spectrometry. *Nat Protoc* **2007**, *2*, 1585.
- 38. Royle, L.; Radcliffe, C. M.; Dwek, R. A.; Rudd, P. M., Detailed structural analysis of N-glycans released from glycoproteins in SDS-PAGE gel bands using HPLC combined with exoglycosidase array digestions. *Methods Mol Biol* **2006**, *347*, 125.
- 39. Koizumi, K.; Okada, Y.; Fukuda, M., High-Performance Liquid-Chromatography of Monosaccharides and Oligosaccharides on a Graphitized Carbon Column. *Carbohydr. Res.* **1991**, *215*, 67.
- 40. Pabst, M.; Bondili, J. S.; Stadlmann, J.; Mach, L.; Altmann, F., Mass plus retention time = structure: A strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Anal Chem* **2007**, *79*, 5051.
- 41. Ruhaak, L. R.; Deelder, A. M.; Wuhrer, M., Oligosaccharide analysis by graphitized carbon liquid chromatography-mass spectrometry. *Anal Bioanal Chem* **2009**, *394*, 163.
- 42. Fan, J. Q.; Kondo, A.; Kato, I.; Lee, Y. C., High-performance liquid chromatography of glycopeptides and oligosaccharides on graphitized carbon columns. *Anal Biochem* **1994**, *219*, 224.
- 43. Melmer, M.; Stangler, T.; Premstaller, A.; Lindner, W., Solvent effects on the retention of oligosaccharides in porous graphitic carbon liquid chromatography. *J Chromatogr A* **2010**, *1217*, 6092.

- 44. Stephens, E.; Maslen, S. L.; Green, L. G.; Williams, D. H., Fragmentation characteristics of neutral Nlinked glycans using a MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* **2004**, *76*, 2343.
- 45. Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H., Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci* **2007**, *849*, 115.
- 46. Harvey, D. J.; Mattu, T. S.; Wormald, M. R.; Royle, L.; Dwek, R. A.; Rudd, P. M., "Internal residue loss": rearrangements occurring during the fragmentation of carbohydrates derivatized at the reducing terminus. *Anal Chem* **2002**, *74*, 734.
- 47. Altmann, F.; Pabst, M.; Bondili, J. S.; Stadlmann, J.; Mach, L., Mass plus retention time = structure: A strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Anal Chem* **2007**, *79*, 5051.
- 48. Vliegenthart, J. F. G.; van Halbeek, H.; Dorland, L., High-resolution proton NMR spectroscopy in the structure analysis of carbohydrates derived from glycoproteins. *Int. Congr. Pure Appl. Chem., [Proc.]* **1980**, *27th*, 253.
- 49. Vliegenthart, J. F. G.; van Halbeek, H.; Dorland, L., The Applicability of 500-MHz High-Resolution <sup>1</sup>H-NMR Spectroscopy for the Structure Determination of Carbohydrates Derived from Glycoproteins. *Pure Appl. Chem.* **1981**, *53*, 45.
- 50. Vliegenthart, J. F. G.; Dorland, L.; van Halbeek, H., High-Resolution, <sup>1</sup>H-Nuclear Magnetic Resonance Spectroscopy as a Tool in the Structural Analysis of Carbohydrates Related to Glycoproteins. *Adv. Carbohydr. Chem. Biochem.* **1983**, *41*, 209.
- 51. van Kuik, J. A.; Hard, K.; Vliegenthart, J. F., A 1H NMR database computer program for the analysis of the primary structure of complex carbohydrates. *Carbohydr Res* **1992**, *235*, 53.
- 52. van Kuik, J. A.; Vliegenthart, J. F., Databases of complex carbohydrates. *Trends Biotechnol* **1992**, *10*, 182.
- 53. Fellenberg, M.; Coksezen, A.; Meyer, B., Characterization of picomole amounts of oligosaccharides from glycoproteins by 1H NMR spectroscopy. *Angew Chem Int Ed* **2010**, *49*, 2630.
- 54. Kovacs, H.; Moskau, D.; Spraul, M., Cryogenically cooled probes a leap in NMR technology. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 131.
- 55. Loss, A.; Bunsmann, P.; Bohne, A.; Schwarzer, E.; Lang, E.; von der Lieth, C. W., SWEET-DB: an attempt to create annotated data collections for carbohydrates. *Nucleic Acids Res* **2002**, *30*, 405.
- 56. van Kuik, J. A., *SUGABASE*, Last Update: 2000-09-28 Accessed: 2011-10-15,
  - http://www.boc.chem.uu.nl/sugabase/sugabase.html
- 57. van Kuik, J. A., *CarbBank*, Last Update: 2001-11-15 Accessed: 2011-10-15, http://www.boc.chem.uu.nl/sugabase/carbbank.html
- 58. Bohne-Lang, A.; Lang, E.; Forster, T.; von der Lieth, C. W., LINUCS: linear notation for unique description of carbohydrate sequences. *Carbohydr Res* **2001**, *336*, 1.
- 59. Jansson, P. E.; Stenutz, R.; Widmalm, G., Sequence determination of oligosaccharides and regular polysaccharides using NMR spectroscopy and a novel Web-based version of the computer program CASPER. *Carbohydr Res* **2006**, *341*, 1003.
- 60. Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M., GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **2008**, *7*, 1650.
- 61. Ranzinger, R.; Herget, S.; Wetter, T.; von der Lieth, C. W., GlycomeDB integration of open-access carbohydrate structure databases. *BMC Bioinformatics* **2008**, *9*, 384.
- 62. Ranzinger, R.; Frank, M.; von der Lieth, C. W.; Herget, S., Glycome-DB.org: a portal for querying across the digital world of carbohydrate sequences. *Glycobiology* **2009**, *19*, 1563.
- 63. Ranzinger, R.; Herget, S.; von der Lieth, C. W.; Frank, M., GlycomeDB a unified database for carbohydrate structures. *Nucleic Acids Res* 2011, *39*, D373.
- 64. Shockcor, J. P.; Unger, S. E.; Wilson, I. D.; Foxall, P. J.; Nicholson, J. K.; Lindon, J. C., Combined HPLC, NMR spectroscopy, and ion-trap mass spectrometry with application to the detection and characterization of xenobiotic and endogenous metabolites in human urine. *Anal Chem* **1996**, *68*, 4431.

- 65. Spraul, M.; Freund, A. S.; Nast, R. E.; Withers, R. S.; Maas, W. E.; Corcoran, O., Advancing NMR sensitivity for LC-NMR-MS using a cryoflow probe: application to the analysis of acetaminophen metabolites in urine. *Anal Chem* **2003**, *75*, 1536.
- 66. Godejohann, M., Hydrophilic interaction chromatography coupled to nuclear magnetic resonance spectroscopy and mass spectroscopy--a new approach for the separation and identification of extremely polar analytes in bodyfluids. *J Chromatogr A* **2007**, *1156*, 87.
- 67. Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. K., Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabonomic toxicology studies. *Anal Chem* **2006**, *78*, 363.
- 68. Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J., Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Anal Chem* **2005**, *77*, 1282.
- Cloarec, O.; Campbell, A.; Tseng, L. H.; Braumann, U.; Spraul, M.; Scarfe, G.; Weaver, R.; Nicholson, J. K., Virtual chromatographic resolution enhancement in cryoflow LC-NMR experiments via statistical total correlation spectroscopy. *Anal Chem* 2007, *79*, 3304.
- 70. Meyer, B.; Hansen, T.; Nute, D.; Albersheim, P.; Darvill, A.; York, W.; Sellers, J., Identification of the 1H-NMR Spectra of Complex Oligosaccharides with Artificial Neural Networks. *Science* **1991**, *251*, 542.
- 71. Radomski, J. P.; van Halbeek, H.; Meyer, B., Neural-Network-Based Recognition of Oligosaccharide 1H-NMR Spectra. *Nat. Struct. Biol.* **1994**, *1*, 217.
- 72. Miller, J. C.; Miller, J. N. *Statistics and Chemometrics for Analytical Chemistry*; 6th ed.; Pearson Education Limited: Harlow (Essex), **2010**. S. 171-173.
- 73. Ehrenberg, A. S. C. *Statistik oder der Umgang mit Daten*; VCH Verlagsgesellschaft, Weinheim, **1986**. S. 186-191.
- 74. Mittag, H.-J. *Statistik Eine interaktive Einführung*; Springer-Verlag Berlin Heidelberg, **2011**. S. 15, 101-110.
- 75. Kendall, M. G., The Variance of Tau When Both Rankings Contain Ties. *Biometrika* 1947, 34, 297.
- 76. Broberg, A.; Thomsen, K. K.; Duus, J. O., Application of nano-probe NMR for structure determination of low nanomole amounts of arabinoxylan oligosaccharides fractionated by analytical HPAEC-PAD. *Carbohydr. Res.* **2000**, *328*, 375.
- 77. Lacey, M. E.; Subramanian, R.; Olson, D. L.; Webb, A. G.; Sweedler, J. V., High-Resolution NMR Spectroscopy of Sample Volumes from 1 nL to 10 μL. *Chem. Rev.* **1999**, *99*, 3133.
- 78. Liu, M. L.; Mao, X. A.; Ye, C. H.; Huang, H.; Nicholson, J. K.; Lindon, J. C., Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy. *J. Magn. Reson.* **1998**, *132*, 125.
- 79. Campbell, I. D.; Dobson, C. M.; Jeminet, G.; Williams, R. J. P., Pulsed Nmr Methods for Observation and Assignment of Exchangeable Hydrogens Application to Bacitracin. *FEBS Lett* **1974**, *49*, 115.
- 80. Hwang, T. L.; Shaka, A. J., Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson. A* **1995**, *112*, 275.
- 81. Claridge, T. *High-Resolution NMR Techniques in Organic Chemistry*; 2nd ed.; Elsevier Ltd.: Oxford, **1999**.
- 82. Torres, A. M.; Dela Cruz, R.; Price, W. S., Removal of J-coupling peak distortion in PGSE experiments. J. Magn. Reson. 2008, 193, 311.
- 83. Kessler, H.; Oschkinat, H.; Griesinger, C., Transformation of Homonuclear Two-Dimensional Nmr Techniques into One-Dimensional Techniques Using Gaussian Pulses. *J. Magn. Reson.* **1986**, *70*, 106.
- 84. Davis, D. G.; Bax, A., Simplification of 1H NMR Spectra by Selective Excitation of Experimental Subspectra. J. Am. Chem. Soc. **1985**, 107, 7197.
- 85. Watson, J. T.; Sparkman, O. J. Introduction to Mass Spectrometry; 4th ed.; John Wiley & Sons Ltd., 2007. S. 22-25.

- 86. Jimenez-Barbero, J.; Asensio, J. L.; Canada, F. J.; Poveda, A., Free and protein-bound carbohydrate structures. *Curr Opin Struct Biol* **1999**, *9*, 549.
- 87. Fukamizo, T.; Hayashi, K., Separation and Mutarotation of Anomers of Chitooligosaccharides. *J Biochem-Tokyo* **1982**, *91*, 619.
- 88. Blumberg, K.; Liniere, F.; Pustilnik, L.; Bush, C. A., Fractionation of Oligosaccharides Containing N-Acetyl Amino-Sugars by Reverse-Phase High-Pressure Liquid-Chromatography. *Anal. Biochem.* **1982**, *119*, 407.
- 89. Bock, K.; Thøgersen, H. Nuclear Magnetic Resonance Spectroscopy in the Study of Mono- and Oligosaccharides; Ann Rep NMR Spectrosc; Webb, G. A.; Academic Press: 1982; Volume 13, 1.
- 90. Bendiak, B.; Orr, J.; Brockhausen, I.; Vella, G.; Phoebe, C., Separation of Neutral Reducing Oligosaccharides Derived from Glycoproteins by Hplc on a Hydroxylated Polymeric Support. *Anal. Biochem.* **1988**, *175*, 96.
- 91. Bergwerff, A. A.; Stroop, C. J. M.; Murray, B.; Holtorf, A. P.; Pluschke, G.; Vanoostrum, J.; Kamerling, J. P.; Vliegenthart, J. F. G., Variation in N-Linked Carbohydrate Chains in Different Batches of 2 Chimeric Monoclonal Igg1 Antibodies Produced by Different Murine Sp2/O Transfectoma Cell Subclones. *Glycoconjugate J.* **1995**, *12*, 318.
- 92. Van Pelt, J.; Van Kuik, J. A.; Kamerling, J. P.; Vliegenthart, J. F.; Van Diggelen, O. P.; Galjaard, H., Storage of sialic acid-containing carbohydrates in the placenta of a human galactosialidosis fetus. Isolation and structural characterization of 16 sialyloligosaccharides. *Eur J Biochem* **1988**, *177*, 327.
- 93. Radomski, J. P.; van Halbeek, H.; Meyer, B., Neural network-based recognition of oligosaccharide 1H-NMR spectra. *Nat Struct Biol* **1994**, *1*, 217.
- 94. Nagel, T., NMR-spektroskopische Analyse von Kohlenhydratstrukturen in bovinem Fibrinogen unterschiedlicher Herkunft, Bachelorarbeit, Fachbereich Chemie, Universität Hamburg, 2010.
- 95. Jirmann, R., Korrelation NMR-spektroskopischer mit massenspektrometrischen Daten zur Analyse von Oligosacchariden, Bachelorarbeit, Fachbereich Chemie, Universität Hamburg, 2010.
- 96. Henschen, A.; Lottspeich, F.; Kehl, M.; Southan, C., Covalent structure of fibrinogen. *Ann N Y Acad Sci* **1983**, *408*, 28.
- 97. Madrazo, J.; Brown, J. H.; Litvinovich, S.; Dominguez, R.; Yakovlev, S.; Medved, L.; Cohen, C., Crystal structure of the central region of bovine fibrinogen (E5 fragment) at 1.4-A resolution. *Proc Natl Acad Sci U S A* **2001**, *98*, 11967.
- 98. Damm, J. B. L.; Bergwerff, A. A.; Hard, K.; Kamerling, J. P.; Vliegenthart, J. F. G., Sialic-Acid Patterns in N-Linked Carbohydrate Chains - Structural-Analysis of the N-Acetyl/N-Glycolyl-Neuraminic-Acid-Containing N-Linked Carbohydrate Chains of Bovine Fibrinogen. *Recl Trav Chim Pay B* 1989, 108, 351.
- 99. Tamura, T.; Wadhwa, M. S.; Rice, K. G., Reducing-end modification of N-linked oligosaccharides with tyrosine. *Anal Biochem* **1994**, *216*, 335.
- 100. Rice, K. G.; DaSilva, M. L. C., Preparative purification of tyrosinamide N-linked oligosaccharides. *J Chromatogr A* **1996**, *720*, 235.
- 101. Muddiman, D. C.; Bereman, M. S.; Williams, T. I., Development of a nanoLC LTQ Orbitrap Mass Spectrometric Method for Profiling Glycans Derived from Plasma from Healthy, Benign Tumor Control, and Epithelial Ovarian Cancer Patients. *Anal Chem* **2009**, *81*, 1130.
- 102. Varki, A.; Gagneux, P., Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* **1999**, *9*, 747.
- Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E.; (Editors) *Essentials of Glycobiology*; 2nd ed.; Cold Spring Harbor Laboratory Press: NY, 2009.
- 104. Pabst, M.; Altmann, F., Glycan analysis by modern instrumental methods. *Proteomics* 2011, 11, 631.
- 105. Tamura, T.; Wadhwa, M. S.; Chiu, M. H.; Dasilva, M. L. C.; Mcbroom, T.; Rice, K. G., Preparation of Tyrosinamide Oligosaccharides as Iodinatable Glycoconjugates. *Method Enzymol* **1994**, *247*, 43.

# **Danksagung**

Mein Dank gilt

#### Alex

Danke, dass Du immer für mich da bist und mir während dieser Zeit stets den Rücken gestärkt hast.

#### Henning

Danke für unsere Zusammenarbeit. Ich konnte mich immer auf Dich verlassen und darauf, dass Du mit vollem Einsatz und Herzblut dabei bist. Danke für Deine permanente Hilfsbereitschaft, unsere Besprechungen und den Fun, den wir hatten!

#### <u>Miriam</u>

Danke für Deine Unterstützung! Sei es nun die Betreuung der NMR-Geräte, die Hilfe bei meinen Bachelorstudenten, die Korrektur von Praktikumsprotokollen, die Anfertigung von Zuckerproben oder moralischer Beistand, ich konnte immer auf Dich zählen.

#### <u>Anna</u>

Ich danke Dir für unsere gemeinsame Zeit in der Uni (Singen im "tollen" Computerraum) und außerhalb der Uni (på norsk kurset) und die Beantwortung jeglicher Fragen, selbst wenn Du selber gerade im Stress warst.

#### Alex, Henning, Miriam und Anna

Danke für die kritische Begutachtung dieser Arbeit.

#### **Thomas**

Danke für Dein NMR-Wissen, auf das ich immer zurückgreifen durfte und für alles, was Du mir beigebracht hast!

#### Tim und Raffael, meinen Bachelorstudenten

Danke für Eure Begeisterung und Eure Hilfe bei den komplexen Themenstellungen, die ich/wir Euch zugemutet haben. Tim, Dir danke ich für die "Wiederholungstat" im Naturstoffpraktikum und dein Engagement!

#### **Dr. Volker Sinnwell**

Ich möchte mich für jede Unterstützung bei NMR-Fragen und der Chance zur Teilnahme am Bruker-Seminar bedanken.

## <u>Dirk</u>

Für den ersten Kontakt zum Chemiestudium, zur Uni Hamburg und zum AK BM, sowie für unsere gemeinsame NMR-Admin-Zeit möchte ich Dir danken!

#### allen aktuellen und ehemaligen Mitgliedern des AK BM

Alexander, Bea, Dennis, Felix, Ilona, Karolina, Karsten, Katharina, Kathrin, Katrin, Kolja, Martin, Martin, Moritz, Nadja, Patrizia, Robin, Tobias und Wei danke ich für die gemeinsame, schöne Zeit im AK:

Danke für jede Aufmunterung und Hilfe, die ich von Euch erhalten habe und für jeden Spaß, der die Zeit im AK verschönert hat.

#### Bea und Katrin

Euch danke ich für gemeinsam durchlebten Endzeitstress.

#### den Computer-Admins Robin, Alex und Karsten

Danke für Eure Unterstützung bei sämtlichen Rechnerproblemen.

#### dem IT-Service, insbesondere Christian und Sören

Danke für Notfallhilfen was Dateirettungen anbelangt und die Möglichkeit zur reibungslosen Arbeit am PC.

#### allen Lübeckern

Prof. Dr. Thomas Peters danke ich für die Möglichkeit zur Messung am Lübecker NMR-Gerät und für die Unterstützung. Thorsten, Dir danke ich für Deine Hilfe bei allen NMR-Messungen an Eurem 500er und für alle Ratschläge und Anregungen, die ich von Dir erhalten habe. Sarah, Dir danke ich für die Hilfsbereitschaft bei allen Transportproblematiken und allen Lübeckern für die tolle Zeit, die wir bei gemeinsamen Seminaren in Lübeck, Hamburg oder Dänemark hatten.

#### Alina und Klaus

Danke, dass Ihr mir die Benutzung des Plasmaofens ermöglicht habt.

#### meinen Brüdern Michael und Christian

Christian, Dir danke ich für Tipps und Tricks rund um *matlab*. Euch beiden danke ich für alles, was uns verbindet.

### **Yvonne**

Danke für unsere Freundschaft.

#### allen Personen, die mich in meinem Werdegang unterstützt haben

Ich danke allen Assistenten, Professoren und Kommilitonen, sowie Familie und Freunden, die mich ausgebildet, unterstützt und motiviert haben.

# Curriculum Vitae

PERSÖNLICHE DATEN		
Name	Meike Fellenberg	
Geboren am	12. Dezember 1983 in Hamburg	
Berufserfahrung		
10/2008 – 09/2011	Wissenschaftliche Mitarbeiterin Universität Hamburg Fachbereich Chemie, Institut für Organische Chemie Martin-Luther-King-Platz 6 20146 Hamburg	
Lehrtätigkeiten	Lehrbeauftragte im Grundpraktikum für Organische Chemie	
Gerätebetreuung	Betreuung von zwei NMR-Spektrometern (Bruker AVI 700 MHz mit Cryoprobenkopf, Bruker DRX 500 MHz)	
WISSENSCHAFTLICHE AUSBILDUNG		
seit 09/2008	Promotionsstudium in der Forschungsgruppe von Prof. Dr. Bernd Meyer, Institut für Organische Chemie, Fachbereich Chemie, Universität Hamburg	
	Dissertation: Entwicklung neuer Strategien zur Analytik von Glycoproteinen	
03/2010	Fortbildung: HR NMR Spectrometer Service and Maintenance, 08 12. März 2010, Bruker Biospin AG, Fällanden, Schweiz	
01/2008 - 07/2008	Anfertigung der Diplomarbeit in der Arbeitsgruppe von Prof. Dr. Bernd Meyer, Institut für Organische Chemie, Fachbereich Chemie, Universität Hamburg	
	Titel: NMR-spektroskopische Untersuchungen zur Analytik von Glycoproteinen	
	Abschluss: Diplom; Note: Sehr gut	
09/2006 - 11/2006	ERASMUS Forschungsprojekt in der Arbeitsgruppe von Prof. H. Fjellvåg an der Universität Oslo, Norwegen	
	Thema: Synthese und Charakterisierung von metall-organischen Gerüst- verbindungen (MOFs)	
10/2003 - 07/2008	Chemiestudium an der Universität Hamburg	
07/2003	Abitur, Gymnasium Grootmoor, Hamburg,	
	Note: 1.3	

## WISSENSCHAFTLICHE BEITRÄGE UND ERFOLGE

Auszeichnung	Preis für Exzellenz in der Glycoforschung 2010, Verein zur Förderung der Glycowissenschaften e.V., Hamburg, 05. November 2010
	Für: Characterization of Picomole Amounts of Oligosaccharides from Glycoproteins by <sup>1</sup> H NMR Spectroscopy
	Vortrag im Rahmen des Symposiums des Glycovereins, Hamburg, 05. November 2010
Vortrag	Characterization of Picomole Amounts of Oligosaccharides from Glycoproteins by <sup>1</sup> H NMR Spectroscopy, Joint EUROMAR 2010 and 17th ISMAR Conference, Florenz, Italien, 08. Juli 2010
Posterpräsentationen	M. Fellenberg, A. Çoksezen, B. Meyer, <i>Characterization of Picomole Amounts of Oligosaccharides from Glycoproteins by</i> <sup>1</sup> H NMR Spectroscopy, Joint EUROMAR 2010 and 17th ISMAR Conference, Florenz, Italien, Juli 2010
	M. Fellenberg & H. N. Behnken, B. Meyer, Integration of LC, MS and NMR for a Sensitive and Fast Characterization of Oligosaccharides from Glycoproteins, 7th National Carbohydrate Symposium, Banff, Canada, Mai 2011
Publikation	M. Fellenberg, A. Çoksezen, B. Meyer, <i>Characterization of Picomole Amounts of Oligosaccharides from Glycoproteins by</i> <sup>1</sup> H NMR Spectroscopy, Angew. Chem. Int. Ed. <b>2010</b> , 49, 2630 – 2633

# **Erklärung**

Ich erkläre hiermit an Eides statt, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich versichere weiterhin, dass die vorliegende Dissertation weder in gleicher noch in veränderter Form bereits in einem Prüfungsverfahren vorgelegen hat.

Hamburg, den 11.11.2011

Meike Fellenberg