

Entwicklung und statistische Validierung eines Biomarkers für  
Leberkrebs basierend auf den posttranslationalen Modifikationen  
von Fibrinogen und Entwicklung eines Bioinformatischen Tools  
zur Integration von LC-MS und NMR

## **Dissertation**

zur Erlangung des Doktorgrades

Fachbereich Chemie

Fakultät für Mathematik, Informatik und Naturwissenschaften

Universität Hamburg

vorgelegt von

**Tim Nagel**

aus Hamburg

Hamburg 2017

Die vorliegende Arbeit wurde in der Zeit von April 2013 bis September 2017 am Institut für Organische Chemie, Fachbereich Chemie, Fakultät für Mathematik, Informatik und Naturwissenschaften, Universität Hamburg, Geschäftsführender Direktor Prof. Dr. Christian Stark, angefertigt.

Ich danke Herrn Prof. Dr. Bernd Meyer für die interessante Themenstellung sowie die stets freundliche und wertvolle Unterstützung bei dieser Arbeit.

Herrn Prof. Dr. Ansgar Lohse, Herrn Dr. Henning Wege und Frau Dr. Ines Gil-Ibanez danke ich für die Zusammenarbeit und Bereitstellung der Blutplasmaproben.

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

1. Gutachter: Prof. Dr. Bernd Meyer

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

Tag der Disputation: 9. März 2018



# Inhaltsverzeichnis

<b>Abkürzungsverzeichnis.....</b>	<b>V</b>
<b>1 Einleitung .....</b>	<b>1</b>
1.1 Struktur und Funktion von Glycanen.....	1
1.2 Analytik von Glycanen .....	4
1.3 Leberkrebs.....	6
1.4 Fibrinogen.....	7
<b>2 Aufgabenstellung.....</b>	<b>9</b>
<b>3 Methoden.....</b>	<b>11</b>
3.1 Massenspektrometrische Analyse intakter Proteine.....	11
3.2 Quantifizierung der Glycoformen und Proteinspezies von A $\alpha$ -, B $\beta$ - und $\gamma$ -Untereinheit.....	13
3.3 Statistische Methoden.....	18
3.3.1 ANOVA .....	18
3.3.2 Hauptkomponentenanalyse .....	18
3.3.3 MANOVA.....	19
3.4 Dreidimensionale Kreuzkorrelation (3DCC).....	19
3.4.1 3DCC <sub>C</sub> .....	21
3.4.2 3DCC <sub>L</sub> .....	22
3.4.3 3DCC <sub>S</sub> .....	23
<b>4 Ergebnisse und Diskussion .....</b>	<b>25</b>
4.1 Entwicklung eines Biomarkers für Leberkrebs basierend auf den posttranslationalen Modifikationen von Fibrinogen .....	25
4.1.1 Analyse der intakten Fibrinogenuntereinheiten gesunder Spender.....	25
4.1.1.1 Die A $\alpha$ -Untereinheit.....	25

4.1.1.2	Die B $\beta$ - und $\gamma$ -Untereinheiten .....	27
4.1.2	Chromatographische Trennung der Glycoformen bzw. Proteinspezies der Fibrinogenuntereinheiten.....	31
4.1.3	Untersuchung der intakten Fibrinogenuntereinheiten von HCC- und Zirrhosepatienten.....	33
4.1.3.1	Die A $\alpha$ -Untereinheit.....	33
4.1.3.2	Die B $\beta$ - und $\gamma$ -Untereinheiten .....	34
4.1.4	Untersuchung des Fibrinogens mittels enzymatischen Verdaus.....	38
4.1.4.1	Analyse der SNPs anhand der tryptischen Peptide .....	38
4.1.4.2	Identifizierung des Fibrinogen anhand der tryptischen Peptide.....	43
4.1.4.3	Analyse der glycosylierten und phosphorylierten tryptischen Peptide von humanem Fibrinogen.....	46
4.1.4.4	Untersuchung der N-Glycosylierung von humanem Fibrinogen durch Verdau mit Pronase .....	51
4.1.5	Statistische Analyse der quantitativen Daten.....	54
4.1.5.1	Hauptkomponentenanalyse.....	57
4.1.5.2	MANOVA.....	61
4.1.5.3	Einfluss verschiedener Parameter auf die Lage der Datenpunkte im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA.....	66
4.1.6	Validierung.....	70
4.1.7	Diskussion und Fazit zur Entwicklung eines Biomarkers für Leberkrebs.....	76
4.2	Entwicklung einer graphischen Benutzeroberfläche für die dreidimensionale Kreuzkorrelation.....	81
4.2.1	Funktion und Anwendung des 3DCC-GUI.....	84
4.2.1.1	Das Hauptmenü .....	84
4.2.1.2	Bearbeiten der NMR-Spektren .....	85
4.2.1.3	3D <i>Plot</i> der NMR-Daten .....	86
4.2.1.4	Reduzierung der MS-Daten.....	87
4.2.1.5	Erstellen von sub-EICs .....	89
4.2.1.6	Berechnung der Korrelationsspektren.....	90
4.2.1.7	Betrachtung der Korrelationsspektren.....	91
4.2.1.8	Export der Korrelationsspektren .....	92
4.2.1.9	Speichern und Laden von Daten .....	92

4.2.2	Analyse der <i>N</i> -Glycane aus humanem Plasma mittels 3DCC .....	93
4.2.3	Fazit zur Entwicklung einer graphischen Benutzeroberfläche für die dreidimensionale Kreuzkorrelation.....	106
<b>5</b>	<b>Zusammenfassung.....</b>	<b>107</b>
<b>6</b>	<b>Summary.....</b>	<b>109</b>
<b>7</b>	<b>Experimenteller Teil .....</b>	<b>111</b>
7.1	Materialien.....	111
7.2	Software und Datenbanken .....	112
7.3	Untersuchung der intakten Untereinheiten von humanem Fibrinogen.....	113
7.4	Untersuchung der tryptischen Peptide von humanem Fibrinogen.....	114
7.4.1	MASCOT MS/MS Ions Search .....	114
7.5	Untersuchung der Glycopeptide von humanem Fibrinogen durch Verdau mit Pronase .....	115
7.6	Probeninformationen.....	116
7.7	3DCC.....	122
7.8	Sicherheitshinweise.....	127
7.8.1	Arbeiten mit biologischem Probenmaterial.....	128
<b>8</b>	<b>Literaturverzeichnis.....</b>	<b>129</b>
<b>9</b>	<b>Anhang.....</b>	<b>143</b>
9.1	Übersicht der identifizierten tryptischen Peptide von Fibrinogen und Albumin .....	143
9.2	Übersicht der beim Verdau mit Pronase gefundenen Glycopeptide.....	156
9.3	Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der Fibrinogen A $\alpha$ -, B $\beta$ - und $\gamma$ -Untereinheit .....	157
9.4	Boxplots und Ergebnisse der Varianzanalysen.....	181
9.5	Erweiterte Diskussion der statistischen Datenauswertung.....	192

9.5.1	Entwicklung der Methode zur Klassifizierung unbekannter Proben.....	192
9.5.2	Einordnung des Validierungsdatensatzes in das Koordinatensystem der Hauptkomponentenanalyse .....	203
9.5.3	Verfeinerung der Methode durch die Proben der Validierungskohorte.....	204
<b>10</b>	<b>Danksagungen .....</b>	<b>206</b>
<b>11</b>	<b>Eidesstattliche Erklärung .....</b>	<b>207</b>

## Abkürzungsverzeichnis

3DCC	<i>three dimensional cross correlation</i>
AIH	Autoimmunhepatitis
ANOVA	<i>analysis of variance</i>
APF	<i>alpha-fetoprotein</i>
AUC	<i>area-under-the-curve</i>
cfg	<i>Consortium for Functional Glycomics</i>
CI	<i>confidence interval</i>
COPD	<i>chronic obstructive pulmonary disease</i>
CV	<i>canonical variable</i>
df	<i>degrees of freedom</i>
DM	Diabetes mellitus
EDC	<i>extracted delta-Chromatogramm</i>
EIC	<i>extracted ion-Chromatogramm</i>
ER	Endoplasmatisches Reticulum
emPAI	<i>exponentially modified protein abundance index</i>
ESI	<i>electrospray ionization</i>
GUI	<i>graphical user interface</i>
GUIDE	<i>graphical user interface development environment</i>
HBV	Hepatitis-B-Virus
HCC	<i>hepatocellular carcinoma</i>
HCV	Hepatitis-C-Virus
HIV	Humanes Immundefizienz-Virus

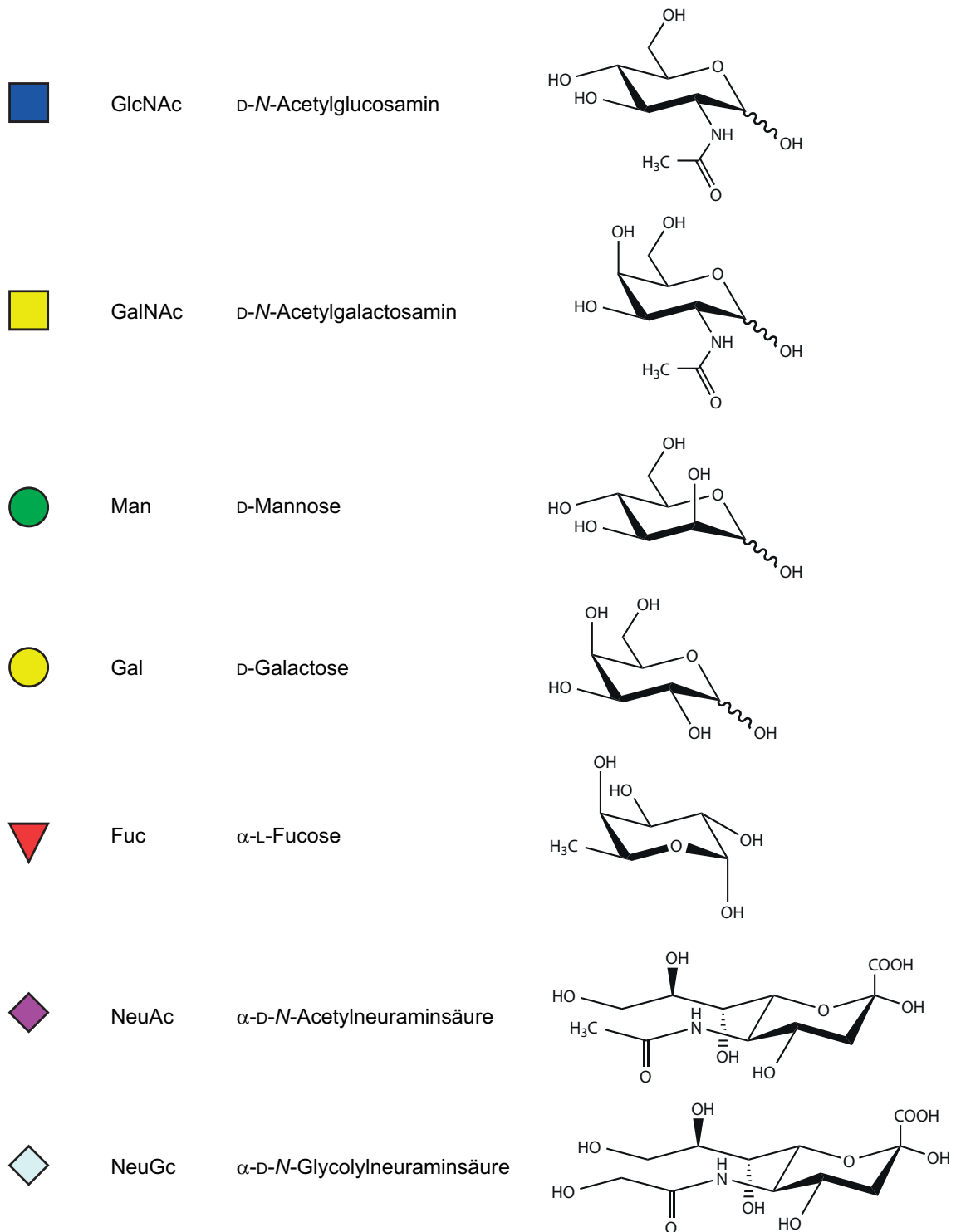


HPLC	<i>high pressure liquid chromatography</i>
HT	arterieller Hypertonus
isCID	<i>in source collision induced dissociation</i>
JCAMP	<i>Joint committee on Atomic and Molecular Physical Data</i>
KHK	koronare Herzerkrankung
LC	<i>liquid chromatography</i>
LTX	Lebertransplantation
MANOVA	<i>multivariate analysis of variance</i>
MS	Massenspektrometrie
NASH	<i>non-alcoholic steatohepatitis</i>
NMR	<i>nuclear magnetic resonance</i>
NRH	Noduläre regenerative Hyperplasie
PBC	primär biliäre Cholangitis
PCA	<i>principal component analysis</i>
PGC	<i>porous graphitized carbon</i>
PNGaseF	<i>Peptide N-glycosidase F</i>
PSC	Primär sklerosierende Cholangitis
RFA	Radiofrequenzablation
ROC	<i>receiver operating characteristic</i>
RP	<i>reversed phase</i>
RPM	<i>revolutions per minute</i>
SNP	<i>single nucleotide polymorphism</i>
SPE	<i>solid phase extraction</i>
SVD	<i>singular value decomposition</i>

TACE	<i>transcatheter arterial chemoembolization</i>
TDC	<i>total delta-Chromatogramm</i>
TIC	<i>total ion current</i>
TKI	Tyrosinkinaseinhibitoren
TOF	<i>time of flight</i>
ZIR	Leberzirrhose

**Tabelle 1: Die 20 proteinogenen Aminosäuren und ihre Abkürzungen im Drei- bzw. Einbuchstabencode.**

<b>Aminosäure</b>	<b>Dreibuchstabencode</b>	<b>Einbuchstabencode</b>
Alanin	Ala	A
Arginin	Arg	R
Asparagin	Asn	N
Asparaginsäure	Asp	D
Cystein	Cys	C
Glutamin	Gln	Q
Glutaminsäure	Glu	E
Glycin	Gly	G
Histidin	His	H
Isoleucin	Ile	I
Leucin	Leu	L
Lysin	Lys	K
Methionin	Met	M
Phenylalanin	Phe	F
Prolin	Pro	P
Pyroglutaminsäure	Pyr	Z
Serin	Ser	S
Threonin	Thr	T
Tryptophan	Trp	W
Tyrosin	Tyr	Y
Valin	Val	V

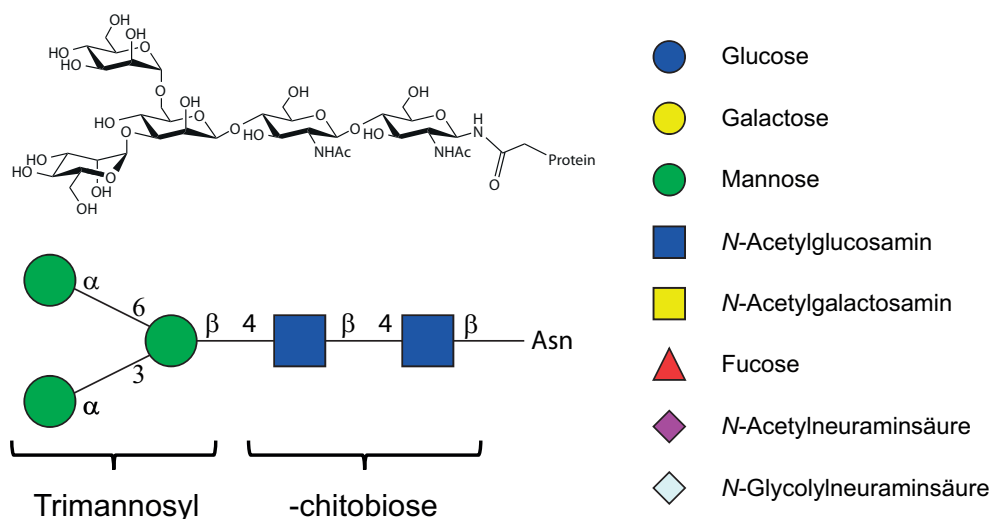


**Abbildung 1: Die in dieser Arbeit relevanten Monosaccharideinheiten und ihre symbolische Schreibweise in der cfg-Notation.**

# 1 Einleitung

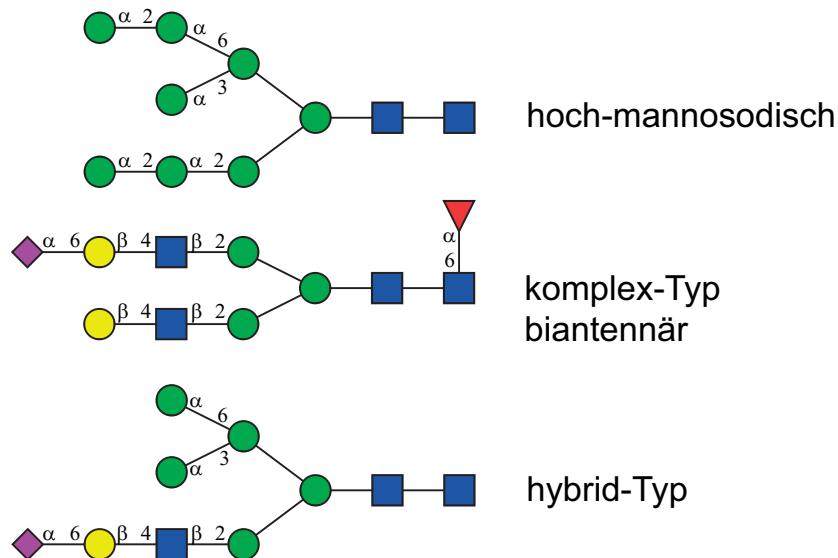
## 1.1 Struktur und Funktion von Glycanen

Die Glycosylierung gehört zu den wichtigsten posttranslationalen Modifikationen von Proteinen und besitzt viele biologische Funktionen.<sup>1,2</sup> Glycosylierungen können die Löslichkeit von Proteinen beeinflussen oder diese vor Proteasen schützen.<sup>1</sup> Vor allem sind Glycokonjugate an Bindungsereignissen zwischen Proteinen und Zellen beteiligt.<sup>3</sup> Oligosaccharide können, ausgehend von wenigen Kohlenhydratbausteinen, sehr viele verschiedene Strukturen bilden und sind aus diesem Grund gut für die Vermittlung von hoch-spezifischen Wechselwirkungen in biologischen Systemen geeignet.<sup>4</sup> Beispielsweise binden Selectine in den Endothelzellen der Blutgefäße an Kohlenhydratstrukturen auf den Leukozyten und ermöglichen so die Leukodiapedese.<sup>5,6</sup> Die Glycosylierung ist extrem wichtig für den Organismus, so führen die meisten genetischen Defekte der Glycosylierung zu schwerwiegenden Erkrankungen.<sup>7</sup> Kohlenhydratstrukturen bilden aber auch Angriffspunkte für Viren, so bindet das Hämagglutinin des Influenzavirus an Sialinsäuren des bronchialen Epithels und ermöglicht die Ansteckung.<sup>6</sup> Durch Anfügen weiterer Saccharidbausteine können die Zielstrukturen maskiert werden, sodass das Virus nicht mehr binden kann. Im Laufe der Evolution haben sich auf diese Weise sehr lange und komplexe Oligosaccharidstrukturen gebildet.<sup>6</sup>



**Abbildung 2: Die Core-Struktur von N-Glycanen.** Alle N-Glycane besitzen einen Core, bestehend aus drei Mannosen und zwei N-Acetylglucosaminen, der N-glycosidisch an eine Asparaginseitenkette des Proteins gebunden ist. Das Asparagin muss Teil der Konsensussequenz Asn-X-Ser/Thr sein, wobei X jede beliebige Aminosäure außer Prolin sein darf. Zur vereinfachten Darstellung der Oligosaccharide dient die cfg-Notation, bei der jede Monosaccharideinheit als Kreis, Quadrat oder Dreieck in einer bestimmten Farbe dargestellt wird.

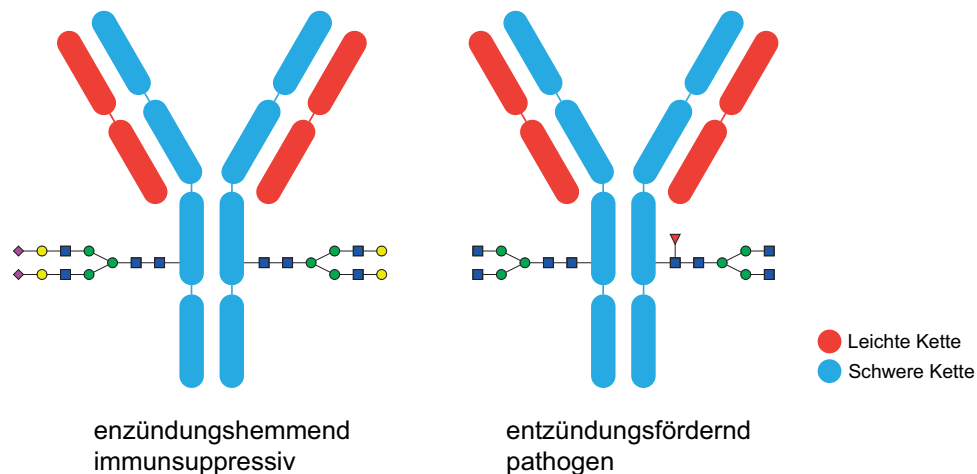
Es werden zwei Arten von Glycosylierungen unterschieden, die *N*-Glycosylierung und die *O*-Glycosylierung. Die Biosynthese der *N*-Glycosylierung beginnt im Lumen des Endoplasmatischen Reticulums (ER), hier wird das Vorläuferglycan GlcNAc<sub>2</sub>Man<sub>9</sub>Glc<sub>3</sub> kotranslational an eine Asparaginseitenkette des Proteins gebunden. Dabei muss das Asparagin Teil der Konsensussequenz Asn-X-Ser/Thr sein, wobei X jede beliebige Aminosäure außer Prolin sein darf. Die drei terminalen Glucosen dienen der korrekten Faltung des Proteins und werden noch im ER abgespalten. Im Golgi-Apparat werden anschließend posttranslational weitere Mannosen abgespalten und andere Monosaccharidbausteine schrittweise angefügt.<sup>8,9</sup> Der sogenannte Trimannosylchitobiose-*Core*, bestehend aus drei Mannosen und zwei *N*-Acetylglucosaminen, bleibt dabei stets erhalten und ist bei allen *N*-Glycanen identisch (Abbildung 2). Es wird aber zwischen hochmannosidischen, komplex-Typ und hybrid-Typ *N*-Glycanen unterschieden; Beispiele für die einzelnen Typen sind in Abbildung 3 dargestellt.



**Abbildung 3: Beispiele für hoch-mannosidische, komplex-Typ und hybrid-Typ *N*-Glycane.** Bei den hoch-mannosidischen *N*-Glycanen sind weitere  $\alpha$ -Mannosen an den *Core* gebunden, welche nicht vom Vorläufer Glycan (GlcNAc<sub>2</sub>Man<sub>9</sub>Glc<sub>3</sub>) abgespalten wurden. Die komplex-Typ *N*-Glycane zeichnen sich durch die sogenannten Antennen aus, welche an den *Core* gebunden sind. Die Antennen beginnen mit einem *N*-Acetylglucosamin und können mit einer Galactose und *N*-Acetylneuraminsäuren fortgesetzt werden. Außerdem ist hier das am proximalen *N*-Acetylglucosamin fucosylierte komplex-Typ *N*-Glycan gezeigt. Der Hybrid-Typ stellt eine Mischung aus hoch-mannosidischem und Komplex-Typ dar.

*O*-Glycane sind an die Hydroxylgruppen von Serin- oder Threoninseitenketten des Proteins gebunden, wobei mehrere Konsensussequenzen bekannt sind. Bei *O*-Glycanen wird ebenfalls zwischen einem *Core* und Antennen unterschieden, es existieren allerdings acht verschiedene *Core*-Strukturen. die Antennen der *O*-Glycane weisen ähnliche Strukturen wie bei den *N*-Glycanen auf.<sup>8</sup>

An einer Glycosylierungsstelle eines Proteins kommen in der Regel verschiedene Glycanstrukturen vor, dies wird als Mikroheterogenität bezeichnet. Die Glycosylierung kann dabei großen Einfluss auf die Funktion des Proteins haben. Ein gutes Beispiel hierfür sind die Immunoglobuline (Antikörper), welche Pathogene spezifisch binden können und somit wichtig für unser Immunsystem sind. Die Wirkungsweise von Immunoglobulinen ist abhängig von ihrer Glycosylierung. Antikörper mit galactosylierten, sialylierten Glycanstrukturen wirken immunsuppressiv und entzündungshemmend, während Antikörper mit kurzen Glycanstrukturen entzündungsfördernd wirken.<sup>3,10</sup> (Abbildung 4) Darüber hinaus wird die Aktivität des Antikörpers durch die *Core*-Fucosylierung des Glycans reguliert, nicht-fucosylierte Antikörper rufen eine 100-fach stärkere Immunantwort hervor als fucosylierte Antikörper.<sup>3,11</sup>

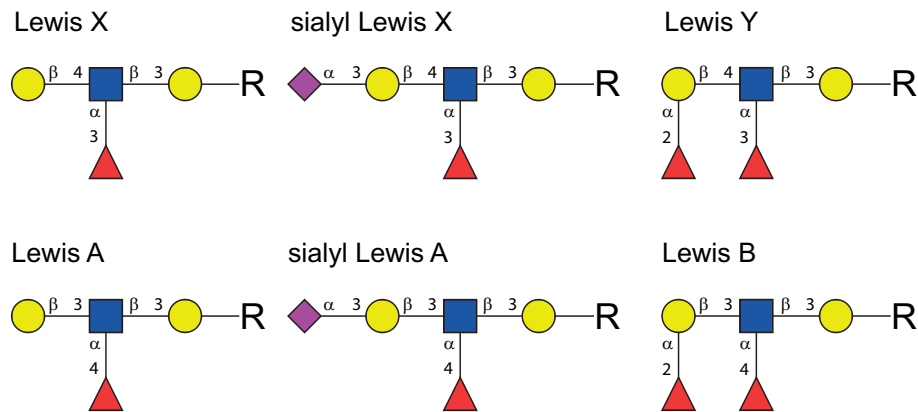


**Abbildung 4:** Es sind zwei Immunoglobuline mit ihren Glycosylierungsstrukturen dargestellt. Antikörper mit galactosylierten, sialylierten Kohlenhydratstrukturen wirken entzündungshemmend und immunsuppressiv, während Antikörper mit kurzen Kohlenhydratstrukturen entzündungsfördernd wirken. Außerdem wird die Aktivität von Antikörpern über die *Core*-Fucosylierung reguliert, nicht-fucosylierte Antikörper rufen eine 100-fach stärkere Immunantwort hervor als fucosylierte Antikörper.

In einem gesunden Organismus stehen Zellproliferation und Zelltod im Gleichgewicht. Bei einer Krebserkrankung ist dieses Gleichgewicht gestört. Signale, die zum Zelltod führen, werden nicht erkannt und Zellen können sich unkontrolliert vermehren.<sup>12</sup> Die fehlerhafte Zellkommunikation ist mit Veränderungen der Glycosylierung verbunden. In Krebszellen werden beispielsweise größere Mengen verzweigter, sialylierter oder fucosylierter *N*-Glycane gebildet.<sup>13-17</sup> Diese Strukturen scheinen die Metastasierung von Tumoren zu fördern. In stark metastasierenden Krebszelllinien wurden große Mengen  $\beta$ 1-6-verzweigter *N*-Glycane gefunden, während die Metastasierung bei *GnT-V knockout* Mäusen unterdrückt ist.<sup>18-20</sup> (*GnT-V* katalysiert die Übertragung des  $\beta$ 1-6 verzweigenden *N*-Acetylglucosamins auf den *Core*.) Sialylierung und Fucosylierung sind zwar essentielle Strukturen von humanen

*N*-Glycanen, Krebszellen bilden aber vermehrt sowohl *Core*-fucosylierte als auch Antennen-fucosylierte und sialylierte Strukturen. Vor allem die in Abbildung 5 gezeigten Lewis-Antigene werden mit der Metastasierung von Krebszellen in Verbindung gebracht, da sie an Selectine in den Endothelzellen der Blutgefäße binden.<sup>20-23</sup>

Die Untersuchung von Glycanstrukturen aus Glycoproteinen ist daher ein vielversprechender Ansatz zur Erkennung von Krebserkrankungen.<sup>24,25</sup> Viele Studien beschäftigen sich mit der Entwicklung von Biomarkern, beispielsweise für Lungenkrebs<sup>26-29</sup>, Eierstockkrebs<sup>30,31</sup> und Darmkrebs<sup>32,33</sup>, basierend auf der Analyse von Glycanstrukturen.



**Abbildung 5: Strukturen der Lewis-Antigene, welche bei humanen *N*- und *O*-Glycanen vorkommen. Die Lewis-Antigene binden an Selectine in den Endothelzellen der Blutgefäße und ermöglichen so die Leukodiapedese. Auf der Oberfläche von Tumorzellen fördern die Lewis-Antigene die Metastasierung.**

## 1.2 Analytik von Glycanen

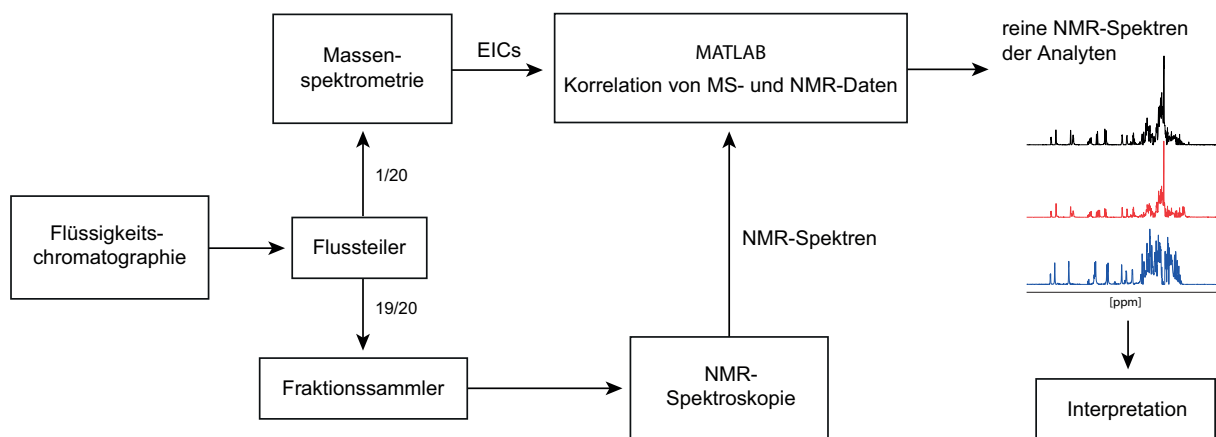
Die wichtigsten Methoden zur Untersuchung von Glycoproteinen sind die Flüssigkeitschromatographie (LC), die Massenspektrometrie (MS) sowie die NMR-Spektroskopie. Für die Analyse mittels LC-MS wird das Glycoprotein meist mit verschiedenen Enzymen behandelt, sodass Glycopeptide oder freie Glycane entstehen.<sup>34</sup> Diese werden chromatographisch getrennt und mittels Massenspektrometrie analysiert.<sup>35-37</sup> Häufig werden die freien Glycane chemisch modifiziert um die chromatographische Trennung oder die Detektion durch ein Fluoreszenzlabel zu verbessern.<sup>34</sup> Der gesamte Prozess des sogenannten *bottom-up* Verfahrens ist zeitaufwendig, kostenintensiv und benötigt gut ausgebildetes Personal. Für die klinische Anwendung ist das *bottom-up* Verfahren daher ungeeignet.

Durch den technischen Fortschritt im Bereich der Massenspektrometrie ist es inzwischen möglich Proteine von bis zu 150 kDa intakt zu ionisieren und mit ausreichender Auflösung zu detektieren, um verschiedene Glycoformen zu unterscheiden.<sup>38-41</sup> Aufgrund des schnellen

und einfachen Arbeitsablaufes ist dieses Verfahren gut für die klinische Anwendung geeignet und könnte zur Früherkennung von Krebserkrankungen eingesetzt werden.

Bei der massenspektrometrischen Untersuchung eines Glycans oder Glycoproteins wird die exakte Masse erhalten, aus der lediglich die Glycankomposition bestimmt werden kann. Durch tandem-MS Experimente können zwar gewisse Rückschlüsse auf die Sequenz der Monosaccharide getroffen werden<sup>42,43</sup>, die genaue Verknüpfung kann mittels Massenspektrometrie aber nicht ermittelt werden. Hierfür ist die NMR-Spektroskopie eine geeignete Methode. Im <sup>1</sup>H-NMR-Spektrum eines typischen *N*-Glycans gibt es neben dem sogenannten Bulk-Bereich zwischen 3.2 und 3.9 ppm, indem sich die Signale vieler Protonen überlagern, mehrere Bereiche in denen isolierte Signale zu finden sind. Dabei handelt es sich vor allem um die Signale der anomeren Protonen und der Methylgruppen von *N*-Acetylgruppen und Desoxyzuckern.<sup>44-46</sup> Diese werden als *structural reporter group*-Signale bezeichnet, da anhand ihrer chemischen Verschiebung die Struktur des Glycans ermittelt werden kann. Zu diesem Zweck wurden die chemischen Verschiebungen tausender Oligosaccharide in Datenbanken gesammelt.<sup>47,48</sup>

Für die NMR-spektroskopische Charakterisierung werden ca. 15 pmol des reinen Oligosaccharids benötigt.<sup>49</sup> Leider liefert die chromatographische Reinigung freier *N*-Glycane häufig kein reines Produkt, da die Elutionsbereiche der chemisch ähnlichen Glycane stark überlappen. Eine Lösung für dieses Problem ist die von Behnken und Fellenberg entwickelte dreidimensionale Kreuzkorrelation von NMR- und MS-Daten.<sup>50-53</sup>



**Abbildung 6: Arbeitsschema der dreidimensionalen Kreuzkorrelation (3DCC). Das Eluat der Flüssigkeitschromatographie wird geteilt und zum Teil massenspektrometrisch detektiert, wodurch die *extracted ion*-Chromatogramme erhalten werden. Der größere Teil des Eluats wird fraktioniert aufgefangen und NMR-spektroskopisch untersucht. Durch mathematische Korrelation der NMR- und MS-Daten werden die reinen Spektren der Probenbestandteile erhalten.**



Das Arbeitsschema der dreidimensionalen Kreuzkorrelation (3DCC) ist in Abbildung 6 dargestellt. Das Eluat der Flüssigkeitschromatographie wird geteilt. Der kleinere Teil des Eluats wird massenspektrometrisch untersucht, wodurch für jeden Analyten ein *extracted ion*-Chromatogramm erhalten wird. Der größere Teil des Eluats wird fraktioniert aufgefangen und NMR-spektroskopisch untersucht. Durch mathematische Korrelation der NMR- und der MS-Daten können reine Spektren der einzelnen Bestandteile erhalten werden. In Abschnitt 3.4 wird das Verfahren in seinen Einzelheiten beschrieben.

### 1.3 Leberkrebs

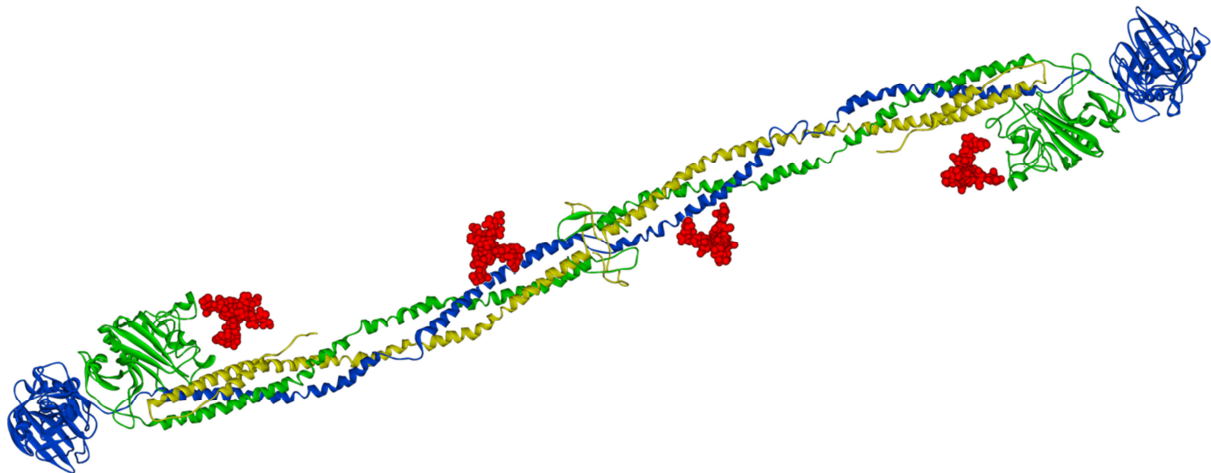
Leberkrebs ist die fünfthäufigste Krebsart weltweit und mit etwa 746.000 Todesfällen im Jahr 2012 die weltweit zweithäufigste Todesursache unter allen Krebsarten.<sup>54</sup> In Deutschland erkranken ungefähr 8.600 Personen jährlich an Leberkrebs mit annähernd gleicher Anzahl an Todesfällen.<sup>55</sup> Die 5-Jahres-Überlebensrate von Leberkrebs beträgt nur ca. 15%, damit gehört Leberkrebs zu den Krebsarten mit der schlechtesten Prognose. Das hepatozelluläre Karzinom (HCC) stellt die häufigste Art von Lebertumoren dar.

Der wichtigste Risikofaktor für Leberkrebs ist Leberzirrhose, welche hauptsächlich durch Infektion mit Hepatitisviren (HBV oder HCV), Alkoholmissbrauch oder nichtalkoholischer Steatohepatitis (NASH) verursacht wird.<sup>56,57</sup> Die niedrige Überlebensrate von Leberkrebs ist vor allem durch die späte Diagnose bedingt. Es konnte gezeigt werden, dass eine halbjährige Untersuchung auf HCC die Sterblichkeitsrate um 37% reduziert.<sup>58</sup>

Neben der Sonographie werden zur Erkennung von Leberkrebs verschiedene Biomarker eingesetzt.<sup>59,60</sup> Der wichtigste Biomarker für HCC ist Alpha-Fetoprotein (AFP), ein humanes Glycoprotein, welches normalerweise während der Embryogenese exprimiert wird. Die Messung der AFP-Konzentration im Blut ergibt eine Sensitivität von 65% und eine Spezifität von 89%.<sup>60</sup> Durch Bestimmung der fucosylierten AFP-Glycoform L3 können Sensitivität und Spezifität zwar auf 97% bzw. 92% erhöht werden<sup>60-62</sup>, trotzdem werden dringend neue Biomarker für HCC benötigt. Vor allem die Analyse fucosylierter Plasmaproteine hat sich als vielversprechender Ansatz zur Erkennung von HCC erwiesen.<sup>63-69</sup>

## 1.4 Fibrinogen

Fibrinogen ist ein 360 kDa schweres Glycoprotein, welches mit einer Konzentration von etwa 2.5 mg/mL im Blut vorhanden ist.<sup>70</sup> Fibrinogen wird hauptsächlich von der Leber produziert und stellt daher ein ideales Ziel für die Erkennung von Leberkrebs dar. Bei der Blutgerinnung entsteht aus Fibrinogen durch Polymerisation Fibrin, welches ein stabiles Gerinnsel bildet.<sup>70,71</sup> Außerdem übernimmt Fibrinogen wichtige Funktionen bei der Hämostase, Angiogenese und bei Entzündungsreaktionen.<sup>70</sup>



**Abbildung 7: Röntgenkristallstruktur von humanem Fibrinogen nach Kollman *et al.*<sup>72</sup> Die  $A\alpha$ -Untereinheiten sind in Gelb, die  $B\beta$ -Untereinheiten in Grün und die  $\gamma$ -Untereinheiten in Blau dargestellt. Das stäbchenförmige Fibrinogenmolekül setzt sich aus zwei  $A\alpha$ -, zwei  $B\beta$ - und zwei  $\gamma$ -Untereinheiten zusammen. Die zentrale Domäne und die beiden äußeren globulären Domänen sind durch die sogenannten *coiled-coil*-Regionen verbunden. Die äußere C-terminale Region der  $A\alpha$ -Untereinheit (etwa 2/3 der Sequenz) ist in der Kristallstruktur nicht vorhanden. Die N-Glycane an Asparagin-364 der  $B\beta$ -Untereinheit und Asparagin-52 der  $\gamma$ -Untereinheit sind in Rot dargestellt. Die N-Glycane an der  $\gamma$ -Untereinheit sind in der Kristallstruktur nicht aufgelöst und wurden nachträglich eingefügt.**

Fibrinogen setzt sich aus drei verschiedenen Untereinheiten zusammen, welche als  $A\alpha$ -,  $B\beta$ - und  $\gamma$ -Untereinheit bezeichnet werden. Ein Fibrinogenmolekül besteht aus zwei  $A\alpha$ -, zwei  $B\beta$ - und zwei  $\gamma$ -Untereinheiten, welche durch insgesamt 29 Disulfidbrücken miteinander verbunden sind. Die Röntgenkristallstruktur von humanem Fibrinogen ist in Abbildung 7 gezeigt.<sup>72</sup> Das stäbchenförmige Molekül besitzt eine zentrale und zwei terminale globuläre Domänen, welche durch alpha-helicale sogenannte *coiled-coil*-Regionen verbunden sind.<sup>70,71</sup> Die C-terminale Region der  $A\alpha$ -Untereinheit ist in der Kristallstruktur nicht vorhanden. An den N-terminalen Enden der  $A\alpha$ - und  $B\beta$ -Untereinheit in der zentralen Domäne befinden sich die Fibrinopeptide A und B. Die Abspaltung der Fibrinopeptide durch Thrombin leitet die Polymerisation ein. Dabei werden in der zentralen Domäne

Bindungsstellen freigelegt, die mit den äußeren Domänen interagieren. Dies führt zur Bildung von doppelsträngigen Protofibrillen. Die C-terminale Region der A $\alpha$ -Untereinheit kann nach Abspaltung des Fibrinopeptid B intermolekulare Wechselwirkungen zur zentralen Domäne eingehen, die zur Quervernetzung der Fibrillen führen.<sup>70,71</sup>

Die B $\beta$ - und  $\gamma$ -Untereinheiten von humanem Fibrinogen sind an Asparagin-364 bzw. Asparagin-52 *N*-glycosyliert. Die Oligosaccharide sind in Abbildung 7 rot dargestellt. Es handelt sich um biantennäre komplex-Typ *N*-Glycane mit einer oder zwei *N*-Acetylneuraminsäuren.<sup>73-75</sup> Es wurde gezeigt, dass die Glycosylierung schwerwiegende Auswirkungen auf das Polymerisationsverhalten und die Struktur des Fibrins hat.<sup>76,77</sup> Die A $\alpha$ -Untereinheit von humanem Fibrinogen ist nicht *N*-glycosyliert, kann aber Phosphorylierungen und eine *O*-Glycosylierung tragen.<sup>78-85</sup>

## 2 Aufgabenstellung

Das Ziel dieser Arbeit war die Entwicklung und statistische Validierung eines Biomarkers für Leberkrebs. Zu diesem Zweck wurde eine LC-MS Methode zur Analyse der posttranslationalen Modifikationen von humanem Fibrinogen entwickelt. Da Fibrinogen hauptsächlich in der Leber hergestellt wird, stellt es ein optimales Ziel zur Erkennung von Lebererkrankungen dar. Fibrinogen wird aus Blutplasma isoliert und die intakten Untereinheiten werden massenspektrometrisch analysiert. Anhand der erhaltenen Massenspektren können die verschiedenen Glycoformen und Proteinspezies der Fibrinogenuntereinheiten quantifiziert werden.

Es sollten Blutplasmaproben von gesunden Spendern, HCC- und Leberzirrhosepatienten untersucht werden, um quantitative Veränderungen in den posttranslationalen Modifikationen der Fibrinogenuntereinheiten festzustellen. Die erhaltenen Daten sollten mit statistischen Methoden ausgewertet werden. Die Betrachtung von jeweils nur einer Glycoform oder Proteinspezies liefert nur geringe Sensitivität und Spezifität und ist daher nicht als Biomarker geeignet. Aus diesem Grund sollte eine Methode entwickelt werden, die eine Kombination aus allen quantifizierten Glycoformen und Proteinspezies der Fibrinogenuntereinheiten zur Unterscheidung der Proben von gesunden Spendern, HCC- und Leberzirrhosepatienten verwendet. Hierzu sollten multivariate statistische Verfahren, wie die Hauptkomponentenanalyse und die multivariate Varianzanalyse (MANOVA) verwendet werden.

Die entwickelte Methode zur Unterscheidung von gesunden Spendern, HCC- und Leberzirrhosepatienten sollte validiert werden. Hierzu sollte ein zweiter Datensatz von Proben mit unbekanntem Status untersucht werden, welche mit der zuvor entwickelten Methode einer der Gruppen zugeordnet werden sollten. Durch den Vergleich mit den Diagnosen der Proben aus der Validierungskohorte sollte die Leistungsfähigkeit der Methode ermittelt werden.

Die Identität und Reinheit des isolierten Fibrinogens sollte durch Analyse der tryptischen Peptide untersucht werden. Die beim Fibrinogen durch SNPs auftretenden Aminosäureaustausche sollten auf Peptidebene bestätigt werden. Außerdem sollten die posttranslationalen Modifikationen, welche am intakten Protein gefunden wurden, durch einen Pronaseverdau untersucht werden.

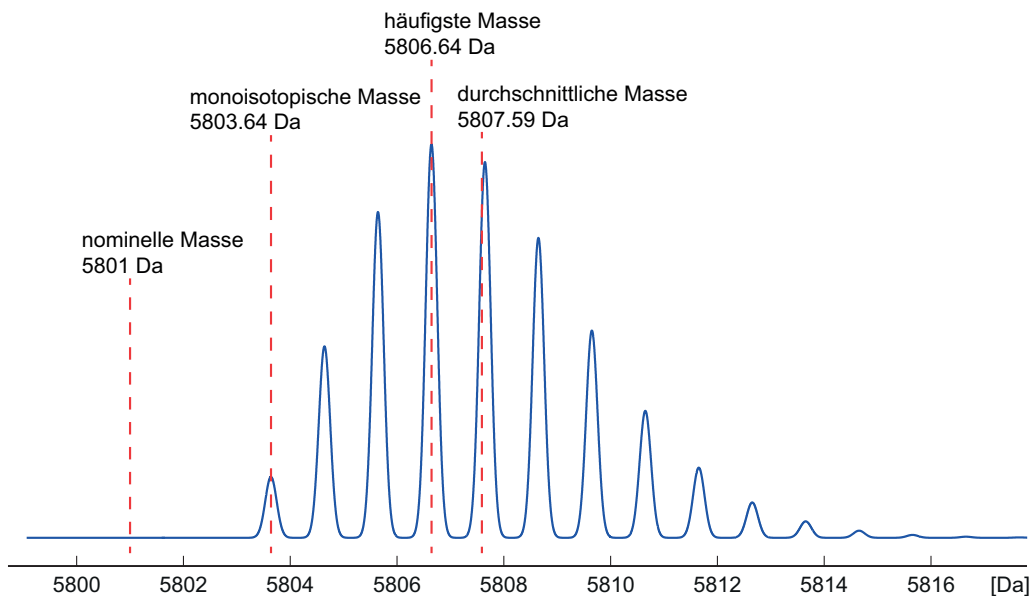
In einem zweiten Projekt sollte eine graphische Benutzeroberfläche (GUI) für die dreidimensionale Kreuzkorrelation entwickelt werden. Die dreidimensionale Kreuz-

korrelation dient der Berechnung von reinen NMR-Spektren aus einem chromatographisch nicht vollständig getrennten Substanzgemisch. Dabei werden NMR- und LC-MS-Daten mathematisch korreliert. Um die Methode möglichst vielen potentiellen Anwendern nutzbar zu machen, sollte eine einfach zu benutzende, graphische Benutzeroberfläche entwickelt werden. Hierzu sollte die Software MATLAB verwendet werden, welche eine Funktion zu Programmierung von graphischen Benutzeroberflächen besitzt. Bei der Entwicklung des GUIs wurde ein Testdatensatz benötigt, dieser sollte auch als Anleitung zur Benutzung des GUIs für externe Anwender dienen. Hierzu sollte die Gesamtheit aller *N*-Glycane einer Blutplasmaprobe isoliert werden und mittels 3DCC untersucht werden.

## 3 Methoden

### 3.1 Massenspektrometrische Analyse intakter Proteine

In dieser Arbeit wurden die Untereinheiten von humanem Fibrinogen als intakte Spezies massenspektrometrisch analysiert. Organische Moleküle besitzen aufgrund der natürlichen Anwesenheit der schwereren Isotope ( $^2\text{H}$ ,  $^{13}\text{C}$  etc.) eine Isotopenverteilung.

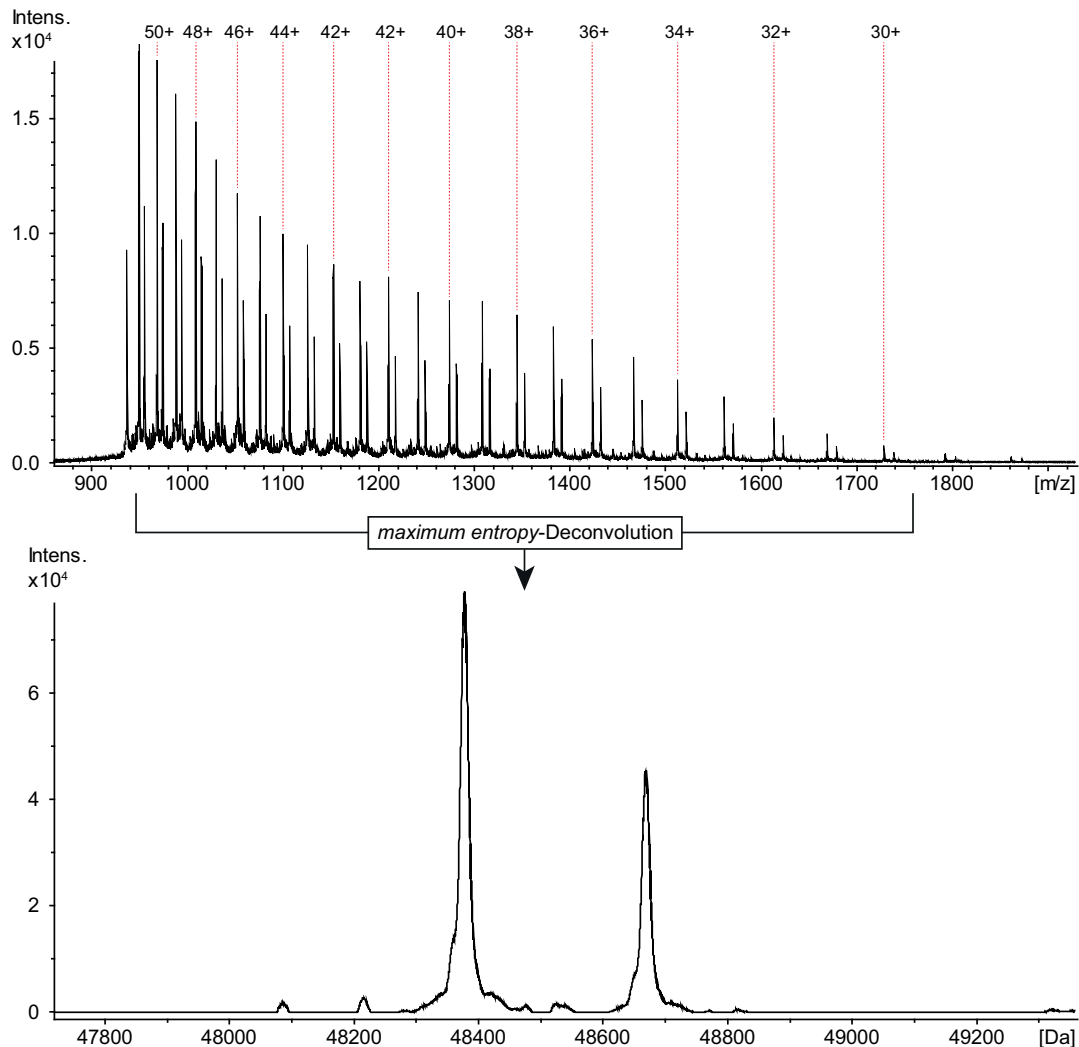


**Abbildung 8: Isotopenverteilung von humanem Insulin ( $\text{C}_{257}\text{H}_{383}\text{N}_{65}\text{O}_{77}\text{S}_6$ ). Die nominelle Masse (5801 Da), die monoisotopische Masse (5803.64 Da), die häufigste Masse (5806.64 Da) und die durchschnittliche Masse (5807.59 Da) sind markiert.**

Abbildung 8 zeigt die theoretische Isotopenverteilung von humanem Insulin. Die nominelle Masse, die monoisotopische Masse, die häufigste Masse und die durchschnittliche Masse sind markiert. Bei organischen Molekülen, die nur die Elemente CHNOSP enthalten, stellt die monoisotopische Masse stets das Signal geringster Masse dar. Die häufigste Masse berechnet sich aus der statistisch wahrscheinlichsten Isotopenzusammensetzung und stellt das intensivste Isotopensignal im Spektrum dar. Je größer das untersuchte Molekül, umso weiter ist die häufigste Masse von der monoisotopischen Masse entfernt. Bei intakten Proteinen ist die monoisotopische Masse meist so wenig intensiv, dass sie nicht mehr beobachtet werden kann.

Die Elektrosprayionisierung (ESI) intakter Proteine liefert viele hochgeladene Spezies, welche als *charge envelope*-Massenspektrum beobachtet werden (Abbildung 9, oben). Je höher der Ladungszustand des Proteins, umso näher liegen die Isotopensignale zusammen.

Bei größeren Proteinen, wie den Fibrinogenuntereinheiten, ist die Auflösung des Massenspektrometers meist nicht ausreichend, um die Isotopensignale aufzulösen. In diesem Fall ist das Isotopenmuster nicht zu erkennen.



**Abbildung 9: Maximum entropy-Deconvolution.** Oben ist das *charge-envelope*-Massenspektrum der  $\gamma$ -Untereinheit von humanem Fibrinogen gezeigt. Die *maximum entropy-Deconvolution* des Bereiches von ca. 950 - 1750 Da liefert das virtuelle Spektrum der ungeladenen  $\gamma$ -Untereinheit (unten).

Im Massenspektrum der  $\gamma$ -Untereinheit von humanem Fibrinogen (Abbildung 9, oben) sind die Ladungszustände 28+ bis 51+ zu erkennen. Je höher die Ladungszustände, umso kleiner sind die Abstände zwischen diesen. Die Auswertung eines solchen *charge envelope*-Spektrums ist kompliziert, da die Signalmuster verschiedener Ladungszustände teilweise überlappen. Außerdem wird die Intensität auf alle Ladungszustände aufgeteilt, wodurch das Signal-zu-Rausch-Verhältnis sinkt. Die *maximum entropy-Deconvolution* berechnet aus dem *charge envelope*-Spektrum das virtuelle Spektrum der ungeladenen Spezies (Abbildung 9, unten). Im deconvolvierten Massenspektrum der  $\gamma$ -Untereinheit ist nur noch

eine Signalgruppe bei der realen Masse des Proteins vorhanden. Die Auswertung des deconvolierten Spektrums ist daher deutlich einfacher.

### 3.2 Quantifizierung der Glycoformen und Proteinspezies von $\text{A}\alpha$ -, $\text{B}\beta$ - und $\gamma$ -Untereinheit

Die verschiedenen Glycoformen und Proteinspezies, welche für die  $\text{A}\alpha$ -,  $\text{B}\beta$ - und  $\gamma$ -Untereinheit von humanem Fibrinogen beobachtet wurden, sollten quantifiziert werden. Hierfür wurden die berechneten Isotopenmuster aller Glycoformen und Proteinspezies an das experimentelle, deconvolierte Massenspektrum angepasst. Die theoretischen Isotopenmuster aller Glycoformen und Proteinspezies wurden in MATLAB mit dem Befehl *isotopicdist* berechnet und durch *least squares*-Optimierung mit dem Befehl *lsqnonneg* an das experimentelle, deconvolierte Spektrum angepasst. Dabei werden Koeffizienten erhalten, die die jeweiligen Anteile der Glycoformen oder Proteinspezies repräsentieren. Mit den berechneten Koeffizienten wurde das theoretische Massenspektrum berechnet, welches die geringste Abweichung zum experimentellen Spektrum aufweist. Da jedes experimentelle Spektrum mit einer individuellen Massenungenauigkeit behaftet ist, welche z. B. durch fehlerhafte Kalibrierung des Massenspektrometers zustande kommt, musste die x-Achse jedes Spektrums korrigiert werden. Die jeweilige Verschiebung wurde ermittelt, indem die Anpassung für Verschiebungen im Abstand von 0.01 Da im Bereich  $\pm 5.00$  Da durchgeführt wurde. Die Anpassung mit der geringsten Abweichung zwischen experimentellem und berechnetem Spektrum wurde für die Quantifizierung verwendet.

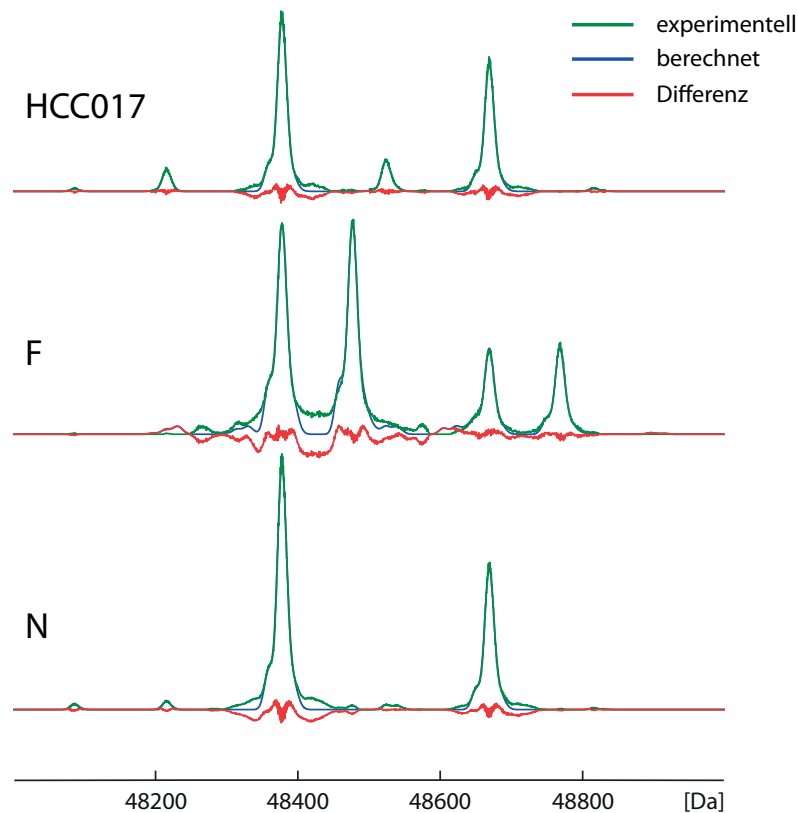
Für die Quantifizierung der Glycoformen der  $\text{B}\beta$ - und  $\gamma$ -Untereinheit wurden die theoretischen Isotopenmuster folgender sechs Glycoformen verwendet:

**Tabelle 2: Glycoformen der  $\text{B}\beta$ - und  $\gamma$ -Untereinheit, welche durch Anpassung der theoretischen Isotopenmuster an das experimentelle, deconvolierte Massenspektrum, mittels *least squares*-Optimierung quantifiziert wurden. H = Hexose, N = N-Acetylglucosamin, S = N-Acetylneuraminsäure, F = Fucose.**

Bezeichnung	Glycankomposition
G2	H <sub>5</sub> N <sub>4</sub>
G1S1	H <sub>4</sub> N <sub>4</sub> S <sub>1</sub>
S1	H <sub>5</sub> N <sub>4</sub> S <sub>1</sub>
S1F	H <sub>5</sub> N <sub>4</sub> S <sub>1</sub> F <sub>1</sub>
S2	H <sub>5</sub> N <sub>4</sub> S <sub>2</sub>
S2F	H <sub>5</sub> N <sub>4</sub> S <sub>2</sub> F <sub>1</sub>



In Abbildung 10 sind das experimentelle, deconvolierte Massenspektrum (grün), das berechnete Massenspektrum (blau) und die Differenz zwischen experimentellem und berechnetem Spektrum (rot) der  $\gamma$ -Untereinheit für drei verschiedene Proben (N, F und HCC017, vgl. Abschnitt 7.7, Tabelle 15) gezeigt. Um eine gute Anpassung zu erreichen wurden neben den Signalen des nativen Proteins zusätzlich die Signale der oxidierten (+16 Da) Spezies und der Verlust eines Wassermoleküls (-18 Da) berücksichtigt, sodass insgesamt 18 Isotopenmuster für die Anpassung der Spektren der  $\gamma$ -Untereinheit verwendet wurden. Zur Berechnung der Anteile der sechs Glycoformen wurden nur die Koeffizienten des nativen Proteins verwendet.



**Abbildung 10: Anpassung der theoretischen Isotopenmuster aller Glycoformen der  $\gamma$ -Untereinheit an die experimentellen, deconvolierten Massenspektren der  $\gamma$ -Untereinheit aus den Proben N, F und HCC017. Grün: experimentelles, deconvoliertes Massenspektrum, blau: berechnetes Massenspektrum, rot: Differenz zwischen experimentellem und berechnetem Massenspektrum.**

Bei Proben, welche heterozygot in Bezug auf den SNP rs6063 sind, wurden die Isotopenmuster der nativen und der veränderten Proteinsequenz (Gly165  $\rightarrow$  Arg) addiert. Die resultierenden Isotopenmuster, welche die Signale beider Spezies im Verhältnis 1:1 enthalten, wurden dann für die Anpassung verwendet. Die Anpassung wurde zunächst mit den Isotopenmustern der vollständig reduzierten  $\gamma$ -Untereinheit durchgeführt. Dabei zeigte sich, dass die experimentellen und berechneten Spektren um etwa 4 Da verschoben waren.

Es ist davon auszugehen, dass die beiden internen Disulfidbrücken der  $\gamma$ -Untereinheit unter den gewählten Bedingungen nicht reduziert werden. Daher wurden die Isotopenmuster der  $\gamma$ -Untereinheit mit zwei Disulfidbrücken verwendet. Die in Abbildung 10 gezeigten Ergebnisse der Anpassung für die Proben N, F und HCC017 zeigen nur geringe Differenzen zwischen experimentellem und berechnetem Spektrum.

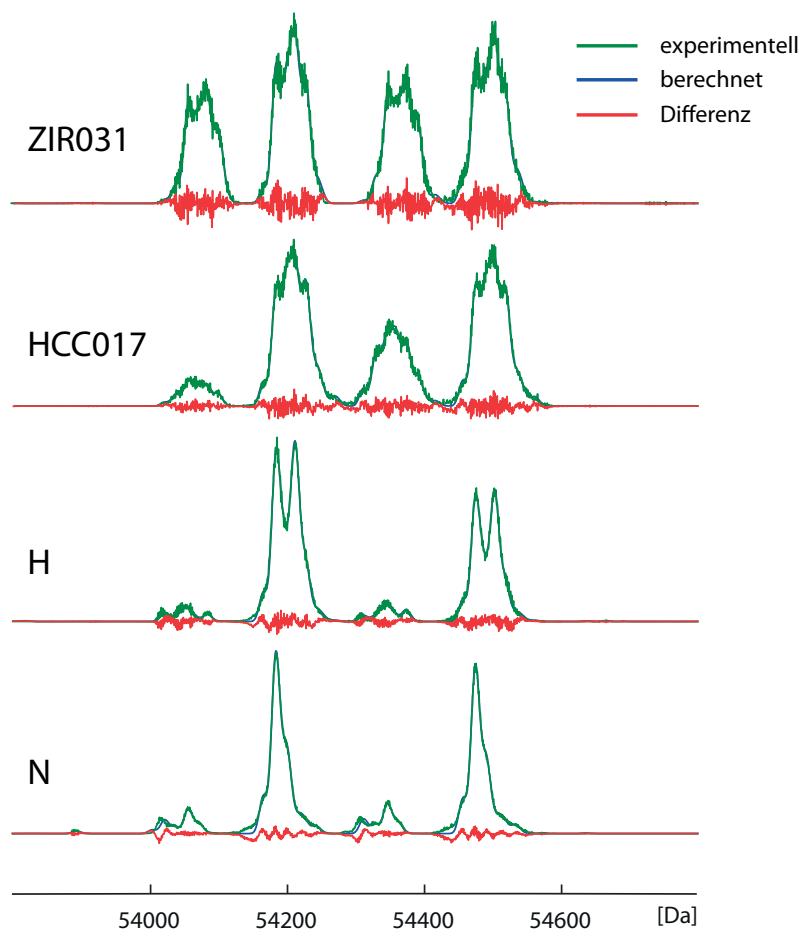
Abbildung 11 zeigt das experimentelle, deconvolierte Massenspektrum der B $\beta$ -Untereinheit (grün), das berechnete Massenspektrum (blau) und die Differenz zwischen experimentellem und berechnetem Spektrum (rot) für die Proben N, H, HCC017 und ZIRO31 (vgl. Tabelle 15). Zur Quantifizierung der Glycoformen der B $\beta$ -Untereinheit wurden die Signale folgender Proteinspezies berücksichtigt:

**Tabelle 3: Proteinspezies der B $\beta$ -Untereinheit, welche bei der Anpassung der theoretischen Isotopenmuster an das experimentelle, deconvolierte Massenspektrum verwendet wurden.**

Proteinspezies	Änderung der Proteinsummenformel
Nativ	
Nativ oxidiert (Met)	+O
Trunkiert (-Gln)	-C <sub>5</sub> H <sub>8</sub> N <sub>2</sub> O <sub>2</sub>
Trunkiert (-Gln), oxidiert	-C <sub>5</sub> H <sub>8</sub> N <sub>2</sub> O
Wasserverlust	-H <sub>2</sub> O
Carbamidiert	+CHNO
Carbamidiert, oxidiert	+CHNO <sub>2</sub>

Neben den Signalen des nativen Proteins, des oxidierten Proteins sowie dem Verlust eines Wassermoleküls, zeigten sich in vielen Spektren intensive Signale bei einer um 128 Da niedrigeren Masse. Es ist davon auszugehen, dass diese Signale durch Abspaltung der C-terminalen Glutamineinheit verursacht werden. Die Signale der um eine Glutamineinheit verkürzten B $\beta$ -Untereinheit und die Signale der carbamidierten B $\beta$ -Untereinheit wurden daher ebenfalls berücksichtigt (Tabelle 3). Insgesamt wurden bei der Anpassung der Spektren der B $\beta$ -Untereinheit 42 Isotopenmuster verwendet. Zur Berechnung der prozentualen Anteile der Glycoformen, dienten nur die Koeffizienten der nativen Proteinspezies. Bei Proben, welche heterozygot in Bezug auf den SNP rs4220 sind, wurden die Isotopenmuster der nativen und veränderten Proteinsequenzen (Arg448  $\rightarrow$  Lys) addiert, sodass die Isotopenmuster, welche beide Signale im Verhältnis 1:1 enthalten, für die Anpassung verwendet wurden. Bei Proben die homozygot in Bezug auf rs4220 sind, wurden die Isotopenmuster der veränderten Proteinsequenz verwendet. In Abbildung 11 sind die

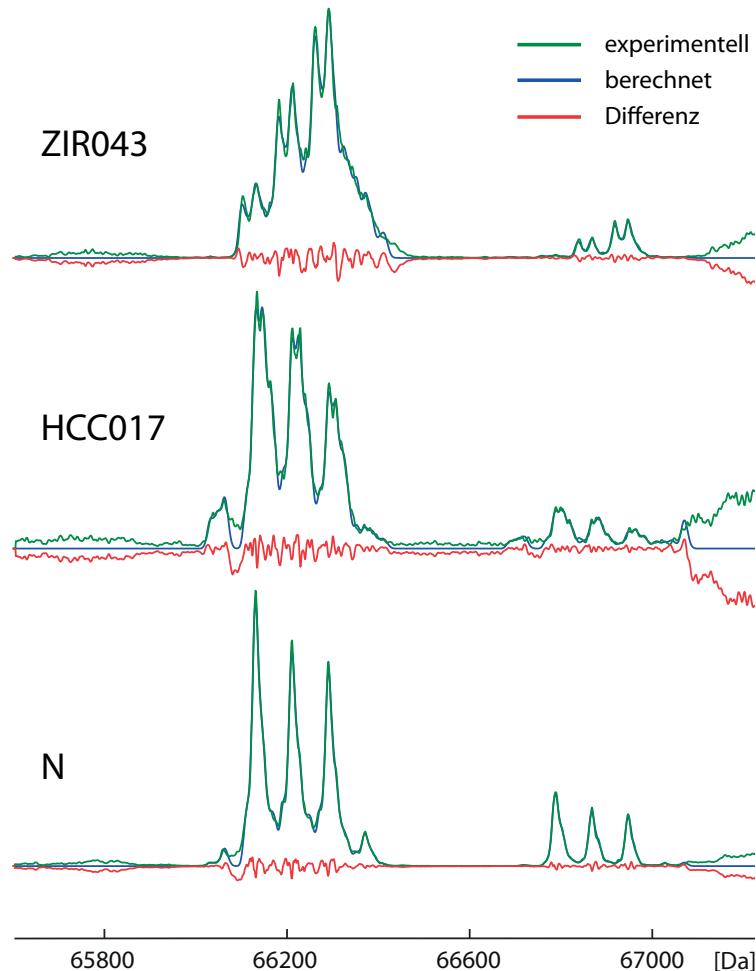
Ergebnisse der Anpassung für die Proben N, H, HCC017 und ZIRO31 gezeigt. Es ist zu erkennen, dass bei allen Proben nur eine geringe Differenz zwischen experimentellem und berechnetem Spektrum auftritt, sogar die breiten Signale in den Spektren der Proben HCC017 und ZIRO31, welche durch das gleichzeitige Auftreten von Oxidation und des SNPs rs4220 verursacht werden, konnten gut angepasst werden.



**Abbildung 11: Anpassung der theoretischen Isotopenmuster aller Glycoformen der B $\beta$ -Untereinheit an die experimentellen, deconvolvierten Spektren der B $\beta$ -Untereinheit aus den Proben N, H, HCC017 und ZIRO31. Grün: experimentelles, deconvolviertes Massenspektrum, blau: berechnetes Massenspektrum, rot: Differenz zwischen experimentellem und berechnetem Massenspektrum.**

Bei der A $\alpha$ -Untereinheit von humanem Fibrinogen sollten die vier Phosphorylierungszustände sowie die O-glycosylierten Zustände quantifiziert werden. Da in vielen Proben bei der A $\alpha$ -Untereinheit ein hohes Maß an Oxidation beobachtet wurde, wurden bei der Anpassung die Signale der einfach und zweifach oxidierten Spezies berücksichtigt. Außerdem wurden sowohl der Verlust eines Wassermoleküls sowie Carbamidierung berücksichtigt. Bei vielen Proben trat ein großes Signal der um die C-terminale Valineinheit verkürzten A $\alpha$ -Untereinheit auf. Die daraus resultierende Massendifferenz von -99 Da führt zu einer starken Überlagerung der Signale von nativer und verkürzter Spezies im jeweils non-,

mono- und diphosphorylierten Zustand. Da die Anpassung unter Berücksichtigung der Isotopenmuster der verkürzten Spezies aller Phosphorylierungszustände zu absurden Ergebnissen führte, wurden nur die Isotopenmuster der verkürzten und nicht-phosphorylierten  $A\alpha$ -Untereinheit verwendet. Insgesamt wurden die Isotopenmuster von 46 Proteinspezies für die Anpassung der Spektren der  $A\alpha$ -Untereinheit verwendet.



**Abbildung 12: Anpassung der theoretischen Isotopenmuster aller Proteinspezies der  $A\alpha$ -Untereinheit an die experimentellen, deconvolierten Massenspektren der  $A\alpha$ -Untereinheit aus den Proben N, HCC017 und ZIR043. Grün: experimentelles, deconvoliertes Massenspektrum, blau: berechnetes Massenspektrum, rot: Differenz zwischen experimentellem und berechnetem Massenspektrum.**

Die prozentualen Anteile der einzelnen Proteinspezies wurden nur ausgehend von den Koeffizienten des nativen Proteins berechnet. In Abbildung 12 sind das experimentelle, deconvolierte Massenspektrum der  $A\alpha$ -Untereinheit (grün), das berechnete Massenspektrum (blau), sowie die Differenz zwischen experimentellem und berechnetem Massenspektrum (rot) für die Proben N, HCC017 und ZIR043 (vgl. Tabelle 15) gezeigt. Es ist zu erkennen, dass nur geringe Abweichungen zwischen experimentellem und berechnetem Spektrum auftreten.

### 3.3 Statistische Methoden

#### 3.3.1 ANOVA

Die Varianzanalyse (ANOVA: *Analysis of Variance*) dient der Untersuchung eines Datensatzes, bei welchem die unabhängigen Variablen in mehr als zwei Gruppen eingeteilt sind. Im hier vorliegenden Fall gibt die unabhängige Variable den Gesundheitszustand des Spenders wieder und ist in drei Gruppen eingeteilt: Gesund, HCC und Zirrhose. Die abhängige Variable ist der Anteil einer bestimmten Glycoform oder Proteinspezies von humanem Fibrinogen. Ziel der ANOVA ist es herauszufinden, ob sich eine Gruppe signifikant von den anderen unterscheidet. Hierzu stellt die ANOVA die Varianz innerhalb der Gruppen und die Varianz zwischen den Gruppen ins Verhältnis und prüft mittels F-Verteilung, ob das Ergebnis signifikant ist. Durch *Post-Hoc*-Tests kann ermittelt werden, zwischen welchen Gruppen Unterschiede bestehen.<sup>86-89</sup>

Die Varianzanalyse wurde in dieser Arbeit verwendet, um Unterschiede in den Anteilen der einzelnen Glycoformen bzw. Proteinspezies zwischen gesunden Spendern, HCC- und Zirrhosepatienten zu erkennen. Da allerdings insgesamt 32 Glycoformen und Proteinspezies analysiert wurden, welche darüber hinaus zum Teil miteinander korrelieren, müssen multivariate statistische Methoden angewendet werden.

#### 3.3.2 Hauptkomponentenanalyse

Die Hauptkomponentenanalyse (PCA: *principal component analysis*) ist eines der wichtigsten Verfahren der multivariaten Statistik, sie dient der Vereinfachung eines komplexen multivariaten Datensatzes. Dabei wird durch eine Hauptachsentransformation ein neues Koordinatensystem berechnet, in welchem die Daten die größtmögliche Varianz aufweisen. Die größte Varianz liegt auf der ersten Hauptkomponente, die zweitgrößte Varianz auf der zweiten Hauptkomponente usw. Dabei sind die Hauptkomponenten orthogonal zueinander. Die Koordinaten der einzelnen Datenpunkte werden durch Linearkombination aus den normierten ursprünglichen Koordinaten berechnet.<sup>87,90</sup>

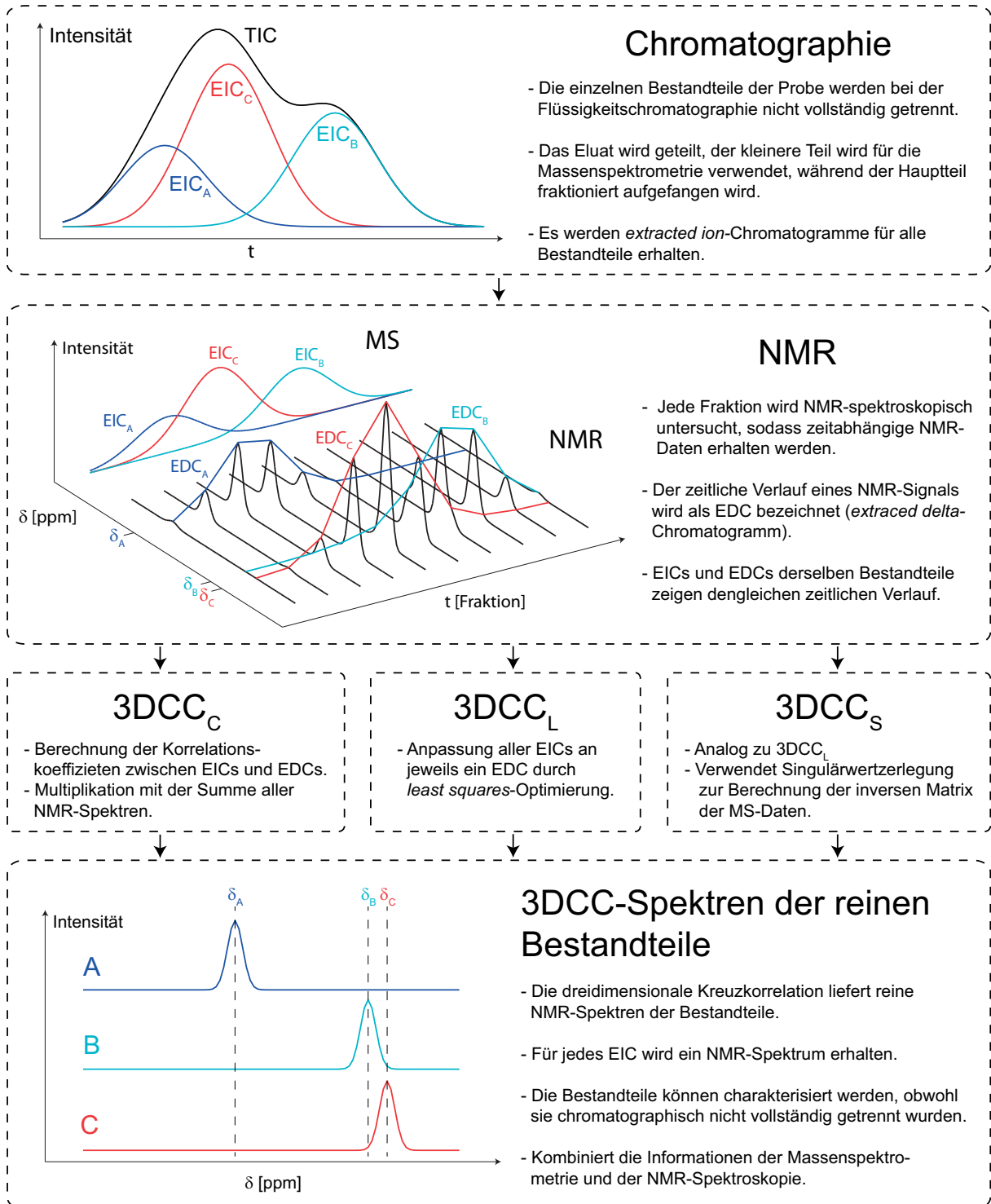
In dieser Arbeit wurde die Hauptkomponentenanalyse verwendet, um einen Datensatz mit mehreren quantifizierten Glycoformen und Proteinspezies zu analysieren. Das Ziel war, eine Trennung zwischen den Proben von gesunden Spendern, HCC- und Zirrhosepatienten zu beobachten. Da die Hauptkomponentenanalyse allerdings keine Information darüber verwendet, welche Probe zu welcher Gruppe gehört, lieferte sie nicht das optimale Ergebnis.

### 3.3.3 MANOVA

Die multivariate Varianzanalyse dient der Untersuchung eines Datensatzes mit mehreren abhängigen Variablen und einer unabhängigen Variablen, welche mehr als zwei Stufen (Gruppen) besitzt. Die Teststatistik der MANOVA ist eine Matrix, daher gibt es verschiedene Möglichkeiten um auf Signifikanz zu prüfen (Pillai-Spur, Wilks-Lambda, Hotelling-Spur, Roys *Largest Root*). Bei der MANOVA wird, wie bei der PCA, eine Hauptachsentransformation durchgeführt. Auf der ersten Hauptachse, welche als erste kanonische Variable bezeichnet wird, wird die beste Trennung zwischen den Gruppen beobachtet. Die kanonischen Variablen werden ebenfalls durch Linearkombination aus den normierten ursprünglichen Variablen erhalten. Dabei ist die erste kanonische Variable die Linearkombination, welche in einer univariaten ANOVA das signifikanteste Ergebnis liefert.<sup>86-90</sup>

### 3.4 Dreidimensionale Kreuzkorrelation (3DCC)

Die dreidimensionale Kreuzkorrelation (3DCC) dient der Berechnung von reinen NMR-Spektren aus einem chromatographisch nicht vollständig getrennten Substanzgemisch.<sup>50-53</sup> Das Prinzip der dreidimensionalen Kreuzkorrelation ist in Abbildung 13 dargestellt. Das Eluat der Flüssigkeitschromatographie wird geteilt. Der kleinere Teil wird für die Massenspektrometrie verwendet, während der Hauptteil fraktioniert aufgefangen wird. Dabei werden *extracted ion*-Chromatogramme (EICs) für alle Bestandteile erstellt. Anschließend werden alle Fraktionen NMR-spektroskopisch analysiert, wobei zeitabhängige NMR-Daten erhalten werden. Der zeitliche Verlauf eines NMR-Signals wird als *extracted delta*-Chromatogramm (EDC) bezeichnet. Die MS-Daten und die NMR-Daten werden nun über die gemeinsame Zeitdomäne mathematisch korreliert, hierfür stehen drei Methoden zur Verfügung, welche als 3DCC<sub>C</sub>, 3DCC<sub>L</sub> und 3DCC<sub>S</sub> bezeichnet werden.



**Abbildung 13: Graphische Darstellung des Prinzips der dreidimensionalen Kreuzkorrelation. Die einzelnen Schritte sind in der Abbildung erläutert.**

### 3.4.1 3DCC<sub>c</sub>

Die dreidimensionale Kreuzkorrelation basiert auf der Tatsache, dass das EIC und das EDC eines Analyten den gleichen zeitlichen Verlauf zeigen. Bei der Methode 3DCC<sub>c</sub> werden die Korrelationskoeffizienten zwischen EICs und EDCs gebildet. Dabei werden die EICs als Spaltenvektoren in der Matrix  $\mathbf{MS}_{(t,m)}$  und die EDCs als Spaltenvektoren in der Matrix  $\mathbf{NMR}_{(t,\delta)}$  gesammelt. (Die Matrix  $\mathbf{NMR}_{(t,\delta)}$  enthält demnach die NMR-Spektren als Zeilenvektoren.) Die EICs und die EDCs müssen dabei die gleiche Anzahl an Datenpunkten besitzen, die Matrizen  $\mathbf{MS}_{(t,m)}$  und  $\mathbf{NMR}_{(t,\delta)}$  haben daher die gleiche Zeitdomäne  $t$ . In Anlehnung an die Dirac-Notation werden die Spaltenvektoren von  $\mathbf{MS}_{(t,m)}$  als  $|MS\rangle$  und die Spaltenvektoren von  $\mathbf{NMR}_{(t,\delta)}$  als  $|NMR\rangle$  bezeichnet. Der Korrelationskoeffizient nach Pearson ist somit definiert als:

$$c_{m,\delta} = \frac{\text{Cov}(|MS\rangle, |NMR\rangle)}{\sigma_{|MS\rangle} \cdot \sigma_{|NMR\rangle}} \quad \text{Gleichung 1}$$

Dabei ist  $\text{Cov}(|MS\rangle, |NMR\rangle)$  die Kovarianz zwischen einem EIC und einem EDC und  $\sigma_{|MS\rangle} \cdot \sigma_{|NMR\rangle}$  stellt das Produkt der Standardabweichungen von  $|MS\rangle$  und  $|NMR\rangle$  dar. Für 3DCC<sub>c</sub> wird jeder Zeilenvektor  $|MS\rangle$  mit jedem Spaltenvektor  $|NMR\rangle$  korreliert, sodass eine Matrix  $\mathbf{C}_{(m,\delta)}$  erhalten wird, die die Korrelationskoeffizienten zwischen jedem EIC und jedem EDC enthält. Der Korrelationskoeffizient nimmt bei hoher Übereinstimmung zwischen einem EIC und einem EDC einen Wert nahe Eins an, tritt keine Übereinstimmung zwischen EIC und EDC auf, beträgt der Wert des Korrelationskoeffizienten Null. Korrelationskoeffizienten mit negativem Wert werden gleich Null gesetzt, da eine anti-Korrelation nicht zu erwarten ist.

Die Extraktion der 3DCC<sub>c</sub>-Korrelationsspektren erfolgt durch paarweise Multiplikation der entsprechenden Korrelationskoeffizienten mit dem NMR-Summenspektrum  $\mathbf{S}_{(\delta)}$ .

$$S_{(\delta)} = \sum_t \langle NMR | \quad \text{Gleichung 2}$$

Das NMR-Summenspektrum ergibt sich durch Summieren der Zeilenvektoren  $\langle NMR |$  von  $\mathbf{NMR}_{(t,\delta)}$ . Durch Multiplikation der Elemente von  $\mathbf{C}_{(m,\delta)}$  mit den Elementen von  $\mathbf{S}_{(\delta)}$  ergibt sich die Matrix  $\mathbf{E}_{(m,\delta)}$ , welche die Korrelationsspektren als Zeilenvektoren enthält.

$$e_{m,\delta} = c_{m,\delta} \cdot S_{\delta} \quad \text{Gleichung 3}$$



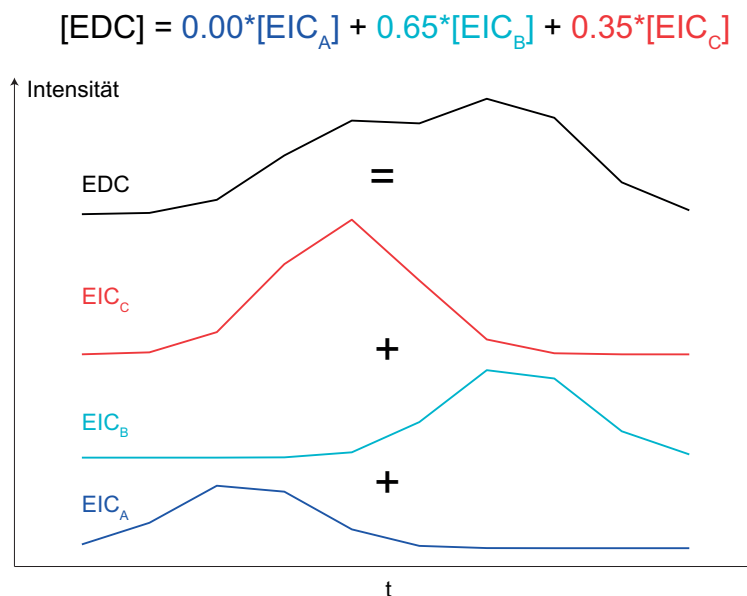
Jedes Korrelationsspektrum gehört zu einem EIC und ist somit einem  $m/z$ -Verhältnis zuzuordnen.<sup>50-53</sup>

### 3.4.2 3DCC<sub>L</sub>

Die Methode 3DCC<sub>L</sub> basiert auf der Tatsache, dass jedes EDC als Linearkombination aller EICs darstellbar ist. Dies ist in Abbildung 14 am Beispiel eines EDCs dargestellt, welches Anteile der Signale B und C enthält. Diejenigen Koeffizienten, für die die Abweichung zwischen experimentellem und berechnetem EDC am geringsten ist, werden bei 3DCC<sub>L</sub> durch *least squares*-Optimierung ermittelt.

$$F = \min_x \left( |NMR\rangle - \sum_1^m x_i |MS\rangle_i \right)^2 \quad \text{Gleichung 4}$$

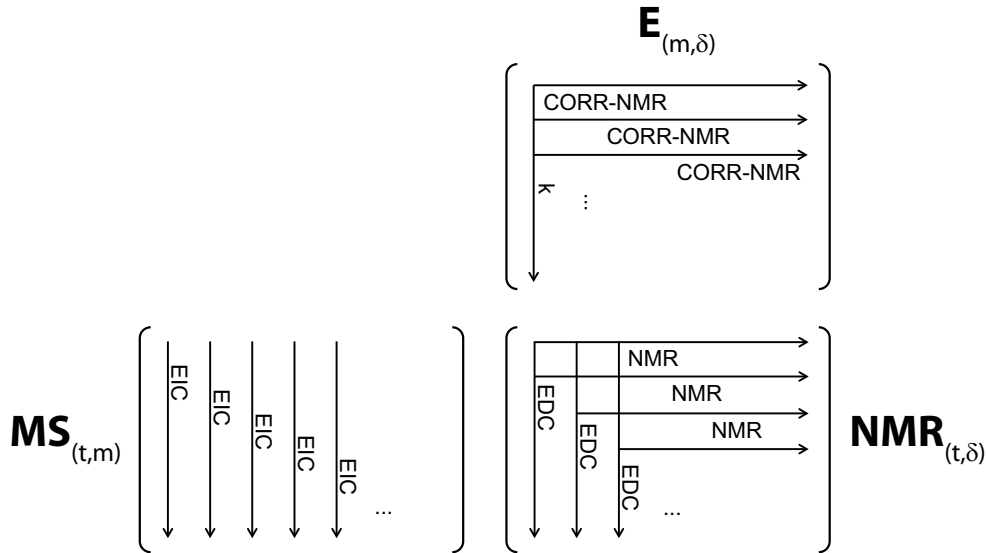
Für jedes EIC wird durch Auftragung der Koeffizienten gegen die chemische Verschiebung ein Korrelationsspektrum erhalten.



**Abbildung 14:** *Least squares*-Anpassung der EICs an ein EDC. Das in schwarz gezeigte EDC zeigt die Signale der Bestandteile B und C. Die *least squares* Anpassung ergibt für EIC<sub>B</sub> einen Koeffizienten von 0.65 und für EIC<sub>C</sub> einen Koeffizienten von 0.35.

### 3.4.3 3DCC<sub>s</sub>

Die Methode 3DCC<sub>s</sub> basiert auf dem gleichen Prinzip wie 3DCC<sub>L</sub>. Der Unterschied besteht darin, dass die Koeffizienten nicht durch *least squares*-Optimierung berechnet werden, sondern durch Multiplikation der inversen MS-Daten mit den NMR-Daten.



**Abbildung 15:** Darstellung der NMR-Daten (einzelne Spektren als Zeilen, EDC als Spalten), MS-Daten (EICs als Spalten) und der Korrelationsspektren (als Zeilen) als Matrizen. Durch Multiplikation der MS-Daten mit den Korrelationsspektren werden die MS-Daten erhalten.

In Abbildung 15 sind die NMR-Daten, MS-Daten und die Korrelationsspektren als Matrizen dargestellt. Durch Multiplikation der MS-Daten (EICs als Spalten) und der Korrelationsspektren (als Zeilen) können die NMR-Daten (Spektren als Zeilen, EDCs als Spalten) erhalten werden. In Dirac-Notation kann dies mit folgender Gleichung dargestellt werden:

$$\langle MS|E \rangle = nmr_{t,\delta} \quad \text{Gleichung 5}$$

Dabei ist  $\langle MS|$  eine Zeilenvektor von  $\mathbf{MS}_{(t,m)}$ , welcher die Datenpunkte aller EICs zu einem bestimmten Zeitpunkt enthält, und  $|E\rangle$  ist ein Spaltenvektor von  $\mathbf{E}_{(m,\delta)}$ , welcher die entsprechenden Koeffizienten enthält. Daraus folgt, dass die Korrelationsspektren durch Multiplikation der NMR-Daten mit der inversen Matrix der MS-Daten erhalten werden können.

$$\mathbf{MS}_{(t,m)} \cdot \mathbf{E}_{(m,\delta)} = \mathbf{NMR}_{(t,\delta)} \quad \text{Gleichung 6}$$

$$\mathbf{MS}_{(t,m)}^{-1} \cdot \mathbf{NMR}_{(t,\delta)} = \mathbf{E}_{(m,\delta)} \quad \text{Gleichung 7}$$

Da bei der dreidimensionalen Kreuzkorrelation die EDCs bzw. die NMR-Spektren gegeben, die Korrelationsspektren hingegen gesucht sind, muss die inverse Matrix der MS-Daten ermittelt werden. Die inverse Matrix der MS-Daten kann durch Singulärwertzerlegung (SVD: *singular value decomposition*) erhalten werden.

Jede Matrix  $M$  besitzt mindestens eine Singulärwertzerlegung der Form:

$$M = U\Sigma V^* \quad \text{Gleichung 8}$$

Dabei ist  $M$  eine beliebige  $m \times n$  Matrix mit dem Rang  $r$ ,  $U$  ist eine unitäre  $m \times m$  Matrix,  $V^*$  ist die Adjungierte einer unitären  $n \times n$  Matrix und  $\Sigma$  ist eine reelle  $m \times n$  Matrix folgender Gestalt:

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \dots & 0 \\ 0 & 0 & \sigma_r & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} \quad \text{Gleichung 9}$$

Die Diagonaleinträge  $\sigma_i$  der Matrix  $\Sigma$  werden als Singulärwerte von  $M$  bezeichnet. Für die inverse  $M^{-1}$  der Matrix  $M$  gilt:

$$M^{-1} = V\Sigma^{-1}U^* \quad \text{Gleichung 10}$$

Die Elemente der inversen  $\Sigma^{-1}$  der Matrix  $\Sigma$  können wie folgt berechnet werden:

$$(\Sigma^{-1})_{ij} = \begin{cases} \frac{1}{\sigma_i}, & \text{falls } i = j \leq r \\ 0 & \text{sonst} \end{cases} \quad \text{Gleichung 11}$$

Auf diese Weise kann die inverse Matrix der MS-Daten erhalten werden, welche durch Multiplikation mit den NMR-Daten die Korrelationsspektren ergibt.

## 4 Ergebnisse und Diskussion

### 4.1 Entwicklung eines Biomarkers für Leberkrebs basierend auf den posttranslationalen Modifikationen von Fibrinogen

Im Rahmen dieser Arbeit wurde ein Biomarker für Leberkrebs entwickelt. Hierfür wurde eine Methode zur Untersuchung der posttranslationalen Modifikationen von humanem Fibrinogen entwickelt.<sup>91</sup> Fibrinogen wird hauptsächlich in der Leber hergestellt und stellt daher ein optimales Ziel zur Erkennung von Leberkrankheiten dar. Fibrinogen wurde durch Ethanolfällung aus humanem Blutplasma isoliert. Nach Denaturierung und Reduktion wurden die drei Fibrinogenuntereinheiten ( $A\alpha$ ,  $B\beta$  und  $\gamma$ ) flüssigkeitschromatographisch getrennt und als intakte Spezies massenspektrometrisch analysiert. Mittels *maximum entropy*-Deconvolution wurden die Massenspektren der virtuell neutralen Fibrinogenuntereinheiten erhalten.

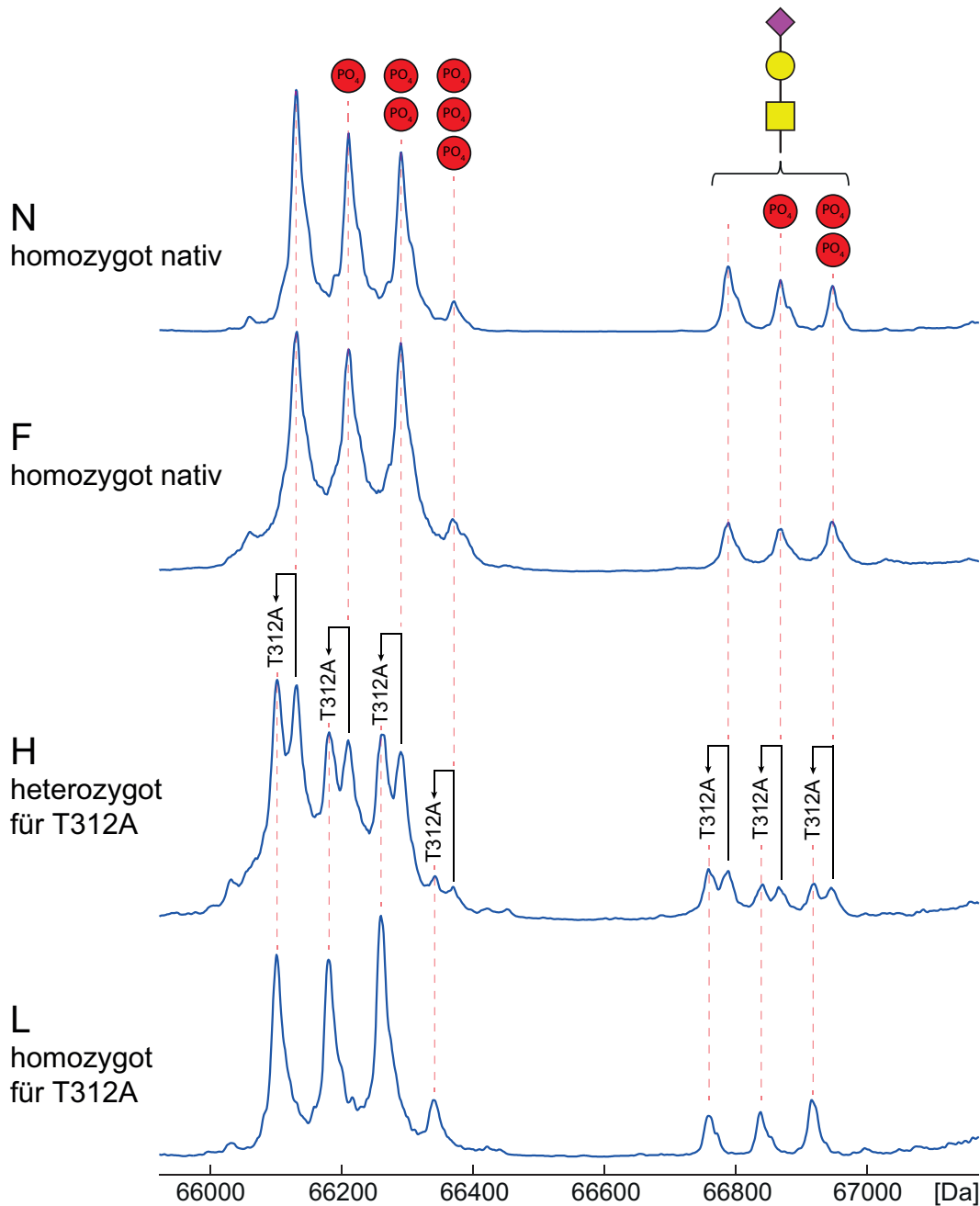
#### 4.1.1 Analyse der intakten Fibrinogenuntereinheiten gesunder Spender

Es wurden zunächst 20 Proben von gesunden Spendern untersucht, welche mit den Buchstaben A-T bezeichnet werden (Tabelle 15). Die Proben wurden vom Zentralinstitut für Transfusionsmedizin GmbH, Hamburg erhalten. Die deconvolierten Massenspektren von jeweils vier Proben (N, F, H und L) werden im Folgenden diskutiert.

##### 4.1.1.1 Die $A\alpha$ -Untereinheit

Die deconvolierten Massenspektren der  $A\alpha$ -Untereinheit aus den Proben N, F, H, und L sind in Abbildung 16 gezeigt. Auf der linken Seite der Spektren, im Bereich 66100 – 66300 Da, sind drei intensive Signale im Abstand von jeweils 80 Da zu erkennen. Es handelt sich hierbei um die Signale der unmodifizierten, der monophosphorylierten und der diphosphorylierten  $A\alpha$ -Untereinheit. Außerdem ist bis zu 7% der triphosphorylierten  $A\alpha$ -Untereinheit zu erkennen. Die Phosphorylierung der  $A\alpha$ -Untereinheit von humanem Fibrinogen ist bekannt.<sup>78-84</sup> Diese Spektren zeigen darüber hinaus, dass die un-, mono- und diphosphorylierten Spezies der  $A\alpha$ -Untereinheit ungefähr im Verhältnis 1:1:1 vorkommen. Die drei weniger intensiven Signale im Bereich 66700 – 67000 Da stammen von einer O-glycosylierten Spezies der  $A\alpha$ -Untereinheit, welche ebenfalls un-, mono- und

diphosphoryliert vorliegen kann. Bei dem *O*-Glycan handelt es sich um das in Abbildung 16 gezeigte sialyl-T-Antigen.<sup>85,92,93</sup>



**Abbildung 16: Deconvolierte Massenspektren der A $\alpha$ -Untereinheit von Fibrinogen aus den Proben N, F, H und L. Die drei intensiven Signale im Bereich 66100 – 66300 Da zeigen die un-, mono- und diphosphorylierte Spezies der A $\alpha$ -Untereinheit. Die drei weniger intensiven Signale im Bereich 66700 – 67000 Da stammen von einer *O*-glycosylierten Spezies der A $\alpha$ -Untereinheit, welche ebenfalls in un-, mono- oder diphosphorylierter Form vorliegen kann. Im Spektrum von Probe H sind alle Signale dupliziert. Dies ist auf den Aminosäureaustausch Thr<sub>312</sub>Ala zurückzuführen, welcher durch den SNP rs6050 verursacht wird. Spender H ist demnach heterozygot in Bezug auf rs6050. Spender L ist homozygot für rs6050, da in diesem Spektrum nur die Signale der veränderten Proteinsequenz vorhanden sind.**

Im Spektrum von Spender H sind alle Signale dupliziert. Der Grund hierfür ist der Aminosäureaustausch Thr<sub>312</sub>Ala, welcher durch den *single nucleotide polymorphism* (SNP) rs6050 auf dem Gen der A $\alpha$ -Untereinheit verursacht wird.<sup>94</sup> Die A $\alpha$ -Untereinheit mit dem Aminosäureaustausch Thr<sub>312</sub>Ala hat eine um 30 Da geringere Masse. Da im Spektrum der A $\alpha$ -Untereinheit von Spender H sowohl die Signale der nativen, als auch die Signale der veränderten Proteinsequenz vorliegen, ist Spender H heterozygot in Bezug auf rs6050. Im Spektrum der A $\alpha$ -Untereinheit von Spender L sind nur die Signale der veränderten Proteinsequenz vorhanden, Spender L ist demnach homozygot in Bezug auf rs6050.

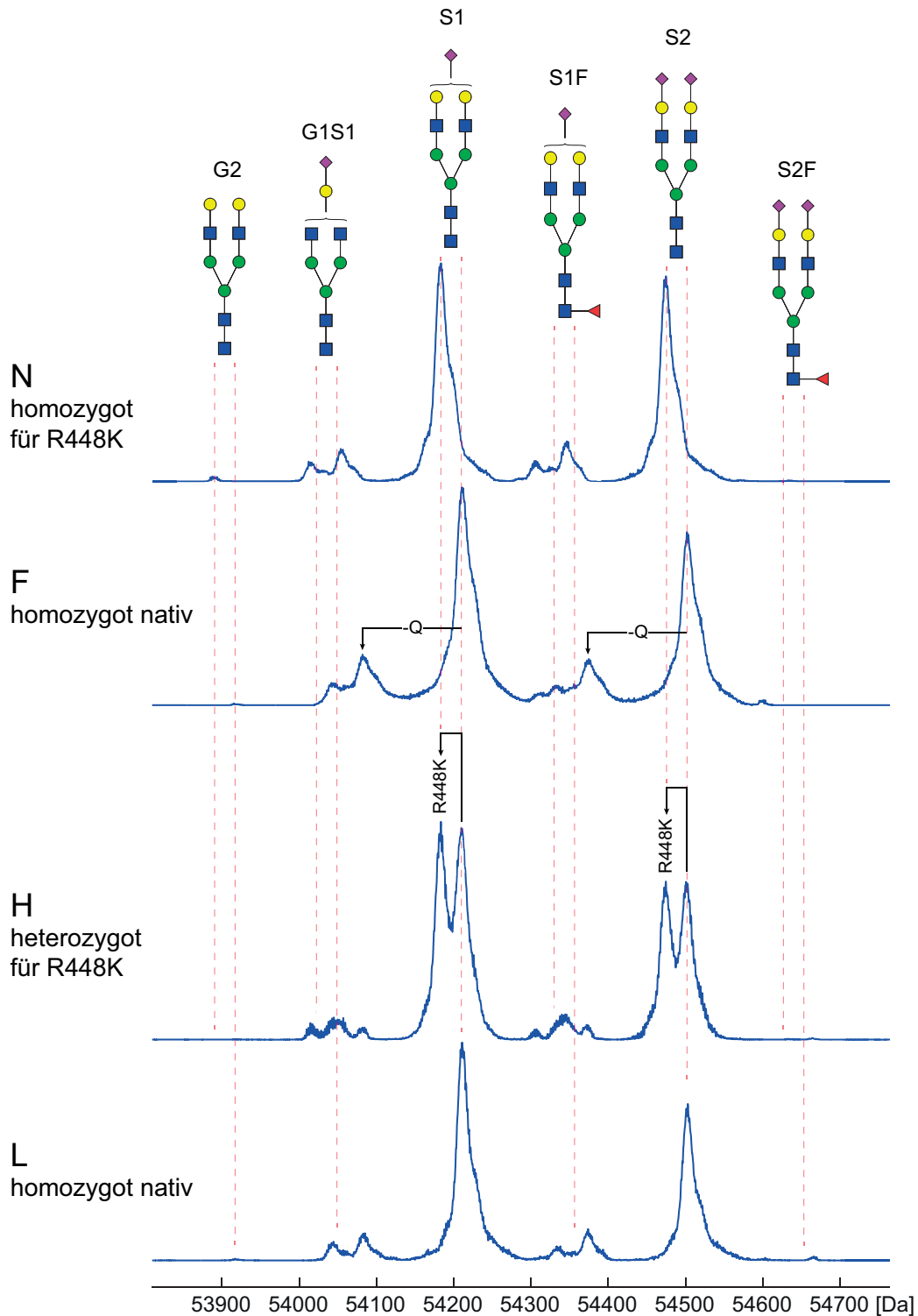
Die Phosphorylierung der A $\alpha$ -Untereinheit ist medizinisch relevant, es wurde beispielsweise gezeigt, dass neugebildetes Fibrinogen von Patienten, die nach einem Herzinfarkt mit Streptokinase behandelt wurden, bis zu dreimal mehr Phosphat enthielt, als vor der Behandlung.<sup>95</sup> Außerdem wurde im Blut von Patientinnen mit Eierstockkrebs ein erhöhter Gehalt der phosphorylierten A $\alpha$ -Untereinheit nachgewiesen.<sup>96</sup>

Der SNP rs6050 wurde bei den gesunden Spendern mit einer Allelfrequenz von 25% beobachtet. Dieser Wert steht in guter Übereinstimmung mit den Literaturdaten, welche für europäische Bevölkerungsgruppen Werte von ca. 24% angeben.<sup>94</sup> Die durch den SNP rs6050 verursachte Fibrinogenvariante A $\alpha$ -Thr<sub>312</sub>Ala bildet festere Gerinnsel als natives Fibrinogen.<sup>97</sup> Aus diesem Grund wird der SNP rs6050 mit einem höheren Risiko für Schlaganfälle<sup>98,99</sup>, Hirnblutungen<sup>100</sup> und venöse Thromboembolien<sup>101,102</sup> assoziiert.

Die O-Glycosylierung der A $\alpha$ -Untereinheit von humanem Fibrinogen wurde 1995/96 von L'Hôte *et al.* durch Lectinbindung des sialyl-T-Antigens entdeckt.<sup>92,93</sup> Seit 2012 haben mehrere Gruppen das sialyl-T-Antigen auf der A $\alpha$ -Untereinheit auch massenspektrometrisch nachgewiesen.<sup>85,103,104</sup> Die entsprechenden O-Glycopeptide wurden im Urin von Patienten mit Harnwegsinfektion gefunden und als Biomarker dafür vorgeschlagen.<sup>105</sup> Die in dieser Arbeit gewonnenen Daten zeigen, dass die A $\alpha$ -Untereinheit von humanem Fibrinogen bei gesunden Spendern zu 13-21% O-glycosyliert vorliegt.

### 4.1.1.2 Die B $\beta$ - und $\gamma$ -Untereinheiten

In Abbildung 17 sind die deconvolierten Massenspektren der B $\beta$ -Untereinheit aus den Proben N, F, H und L gezeigt. Die Strukturen von sechs Glycoformen sind über den Spektren dargestellt. Die beiden dominanten Signale stammen von den Glycoformen S1 und S2, welche ein mono- oder disialyliertes biantennäres N-Glycan vom komplex-Typ tragen. Die Signale verkürzter und fucosylierter Glycoformen (G2, G1S1, S1F und S2F) sind in Abbildung 17 markiert, in den Spektren gesunder Spender sind diese Signale aber nur schwach zu erkennen



**Abbildung 17: Deconvolierte Massenspektren der B $\beta$ -Untereinheit von Fibrinogen aus den Proben N, F, H und L. Die Strukturen von sechs Glycoformen sind über den Spektren dargestellt. Die beiden dominanten Signale stammen von den Glycoformen S1 und S2. Die Glycoformen G2, G1S1, S1F und S2F sind in den Proben gesunder Spender kaum vorhanden. Außerdem sind Signale der um eine Glutamineinheit verkürzten B $\beta$ -Untereinheit zu erkennen, welche mit den Signalen der Glycoformen G1S1 und S1F überlagern. Im Spektrum von Spender H sind alle Signale dupliziert, da Spender H heterozygot bezüglich des SNPs rs4220 ist, welcher den Aminosäureaustausch Arg448Lys verursacht. Im Spektrum von Spender N sind nur die Signale der veränderten Proteinisof orm vorhanden, daher ist Spender N homozygot in Bezug auf rs4220.**

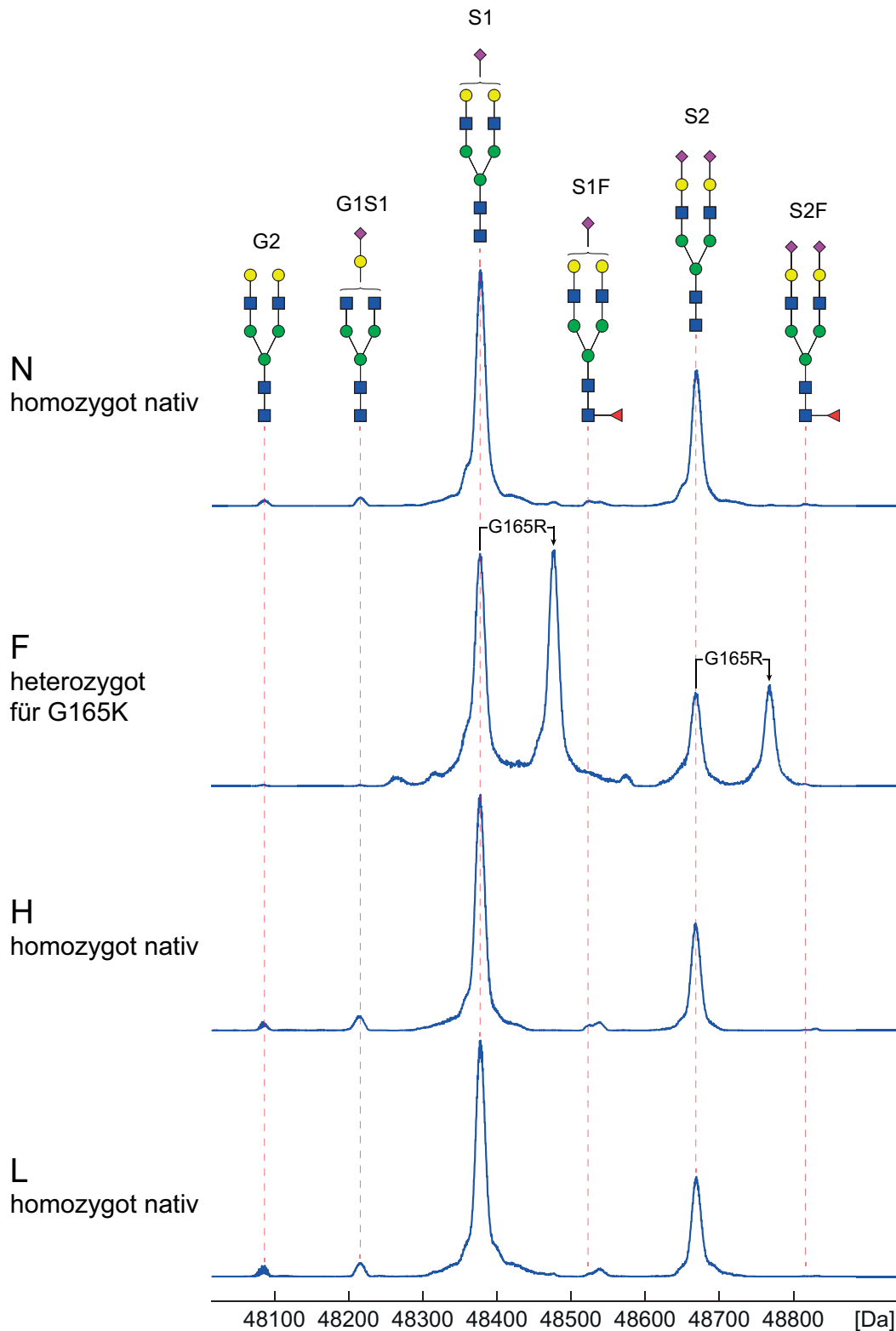
Die Identifizierung der Glycoformen G1S1 und S1F ist zusätzlich erschwert, da ihre Signale von den Signalen einer trunkierten Form der B $\beta$ -Untereinheit überlagert werden, welche um die C-terminale Glutamineinheit verkürzt ist. Im Spektrum der B $\beta$ -Untereinheit von Spender H sind zusätzlich die Signale der Proteinisofom mit dem Aminosäureaustausch Arg448Lys zu erkennen, welche durch den SNP rs4220 verursacht wird,<sup>106</sup> und zu einer um 28 Da geringeren Masse führt. Spender H ist demnach heterozygot in Bezug auf rs4220. Im Spektrum der B $\beta$ -Untereinheit von Spender N sind nur die Signale des veränderten Proteins vorhanden, Spender N ist daher homozygot in Bezug auf rs4220.

Unter den gesunden Spendern trat rs4220 mit einer Allelfrequenz von 18% auf. Dieser Wert stimmt gut mit den Literaturdaten überein, welche Werte von ca. 20% angeben.<sup>106</sup> Es wurde gezeigt, dass rs4220 in Verbindung mit erhöhten Fibrinogenkonzentrationen im Blut steht und damit zum Risiko für kardiovaskuläre Krankheiten beiträgt.<sup>107-109</sup> Außerdem wurde rs4220 in einer Studie mit verringerter Leukozytenanzahl verbunden.<sup>110</sup>

Die deconvolierten Massenspektren der  $\gamma$ -Untereinheit aus den Proben N, F, H und L sind in Abbildung 18 gezeigt. Bei den über den Spektren dargestellten Glycanen handelt es sich um die gleichen Strukturen, die auch auf der B $\beta$ -Untereinheit zu finden sind. Die beiden dominanten Signale stammen ebenfalls von den Glycoformen S1 und S2. Die Signale der Glycoformen G2 und G1S1 sind wenig intensiv, aber in den Spektren der  $\gamma$ -Untereinheit gut zu erkennen. Die fucosylierten Glycoformen S1F und S2F sind auch auf der  $\gamma$ -Untereinheit gesunder Spender nicht vorhanden. Das Spektrum der  $\gamma$ -Untereinheit von Spender F zeigt zusätzliche Signale bei einer 99 Da höheren Masse, welche durch den Aminosäureaustausch Gly165Arg verursacht werden. Diese Mutation wird durch den SNP rs6063, auf dem für die  $\gamma$ -Untereinheit codierenden Gen, verursacht.<sup>111</sup> Spender F ist demnach heterozygot für rs6063. Insgesamt wurde der SNP rs6063 nur bei einem von 20 gesunden Spendern beobachtet. In der Literatur kommt rs6063 mit einer Allelfrequenz von 0.3% vor<sup>112</sup>, und wurde mit erhöhter Erythrozytenverteilungsbreite in Verbindung gebracht.<sup>110</sup>

Aus Abbildung 17 und Abbildung 18 geht hervor, dass die B $\beta$ - und  $\gamma$ -Untereinheiten von humanem Fibrinogen fast identisch glycosyliert sind. Bei gesunden Spendern liegt die B $\beta$ -Untereinheit durchschnittlich zu 39% in der Glycoform S2 vor, während die  $\gamma$ -Untereinheit nur zu 29% disialyliert vorliegt. Im Mittel liegt humanes Fibrinogen also zu 34% disialyliert vor. Dieser Wert stimmt gut mit den Ergebnissen von Adamczyk *et al.* überein, welche die Glycoform S2 zu 33% fanden.<sup>75</sup> Der Sialyierungsgrad des Fibrinogens beeinflusst die Struktur des gebildeten Fibrins: ein hoher Sialyierungsgrad führt zu dünnen, quervernetzten Fibrinfasern, während desialyliertes Fibrinogen dicke, unvernetzte Fibrinfasern bildet.<sup>77,113</sup>





**Abbildung 18: Deconvolierte Massenspektren der  $\gamma$ -Untereinheit von Fibrinogen aus den Proben N, F, H und L. Die Strukturen der sechs Glycoformen sind über den Spektren dargestellt. Die beiden dominanten Signale stammen von den Glycoformen S1 und S2. Die Signale der Glycoformen G2 und G1S1 sind mit sehr geringer Intensität zu erkennen. Die fucosylierten Glycoformen S1F und S2F sind in den Spektren gesunder Spender nicht vorhanden. Im Spektrum der  $\gamma$ -Untereinheit von Spender F sind zusätzliche die Signale der Proteinisoform mit dem Aminosäureaustausch Gly165Arg vorhanden. Spender F ist daher heterozygot in Bezug auf den SNP rs6063, welcher diesen Aminosäureaustausch verursacht.**

#### 4.1.2 Chromatographische Trennung der Glycoformen bzw. Proteinspezies der Fibrinogenuntereinheiten

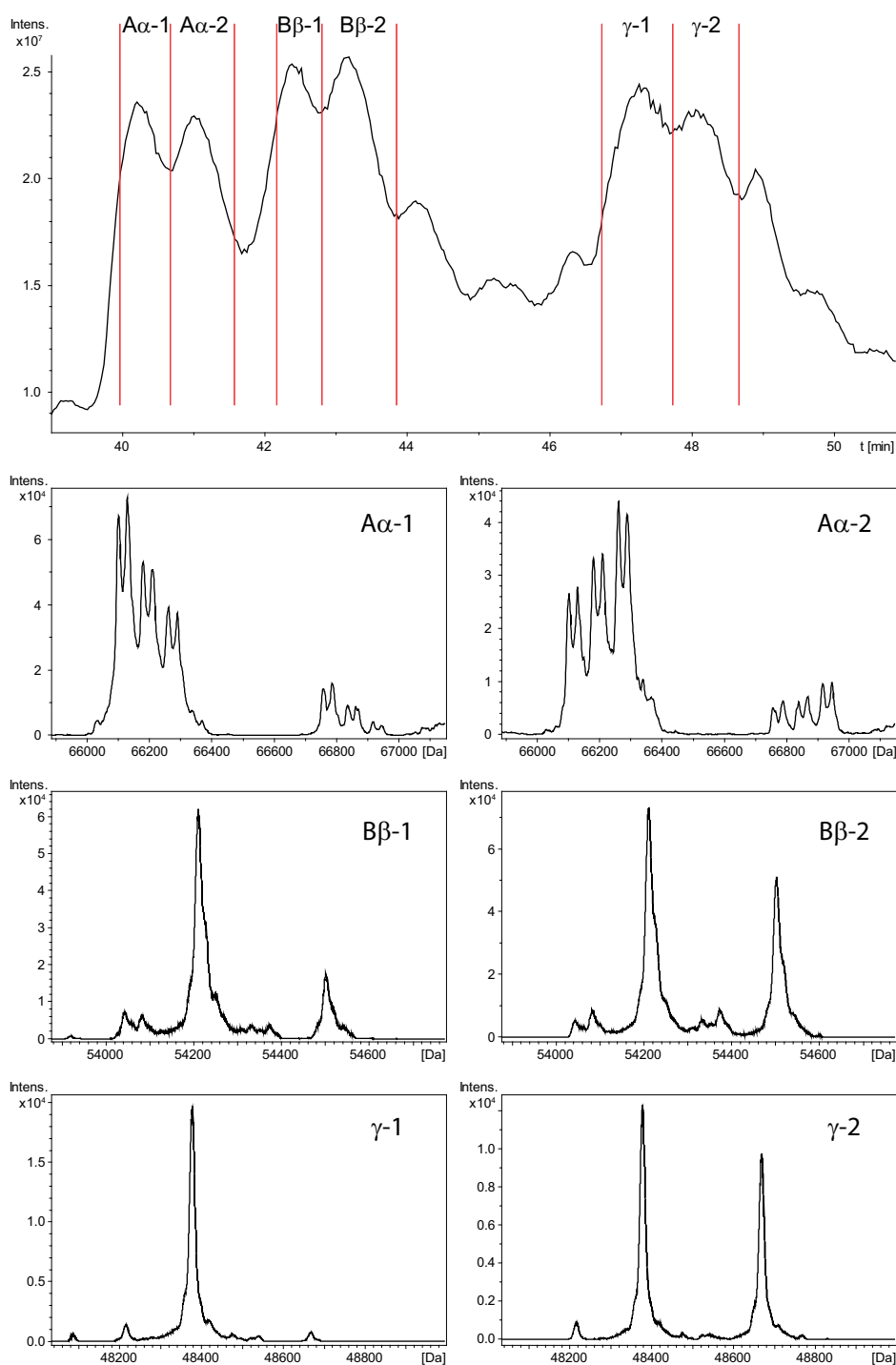
Die chromatographische Trennung der Fibrinogenuntereinheiten wurde auf einer, für intakte Proteine geeigneten, C8-reversed phase-Säule durchgeführt. Dabei werden zum Teil auch die Glycoformen oder Proteinspezies der  $A\alpha$ -,  $B\beta$ - und  $\gamma$ -Untereinheit chromatographisch voneinander getrennt. Dies ist gut an dem in Abbildung 19 gezeigten Ausschnitt des *total ion current* von Probe Q zu erkennen. Die drei Untereinheiten von humanem Fibrinogen eluieren in den Bereichen 40.0-41.6 min ( $A\alpha$ ), 42.2-43.9 min ( $B\beta$ ) und 46.7-48.7 min ( $\gamma$ ). Für jede der drei Untereinheiten sind deutlich zwei Maxima zu erkennen, welche durch partielle Trennung der Glycoformen bzw. Proteinspezies entstehen. In Abbildung 19 sind jeweils zwei Bereiche für eine Untereinheit markiert, in welchen die Massenspektren gemittelt wurden. Die dazugehörigen deconvolvierten Spektren sind ebenfalls in Abbildung 19 gezeigt.

Die beiden Spektren der  $A\alpha$ -Untereinheit von Probe Q in den Bereichen  $A\alpha$ -1 (40.0-40.7 min) und  $A\alpha$ -2 (40.7-41.6 min) zeigen einen deutlichen Unterschied. Im Spektrum des Bereichs  $A\alpha$ -1 nimmt die Intensität der Signale der unglycosylierten  $A\alpha$ -Untereinheit von links nach rechts ab. Der Anteil der unphosphorylierten  $A\alpha$ -Untereinheit ist am größten, der Anteil der diphosphorylierten  $A\alpha$ -Untereinheit am geringsten. Im Spektrum des Bereiches  $A\alpha$ -2 ist die Verteilung umgekehrt, die Intensität der Signale der unglycosylierten Spezies nimmt von links nach rechts zu. Der Anteil der unphosphorylierten  $A\alpha$ -Untereinheit ist niedriger als der Anteil der diphosphorylierten  $A\alpha$ -Untereinheit. Diese Beobachtung zeigt, dass die höher phosphorylierten Spezies der  $A\alpha$ -Untereinheit später eluieren als die niedrig phosphorylierten Spezies.

Die Massenspektren der  $B\beta$ - und  $\gamma$ -Untereinheit zeigen ebenfalls große Unterschiede. Das Spektrum der  $B\beta$ -Untereinheit im Bereich  $B\beta$ -1 (42.2-42.8 min) zeigt einen deutlich geringeren Anteil der Glycoform S2 als das Spektrum im Bereich  $B\beta$ -2 (42.8-43.9 min). Bei der  $\gamma$ -Untereinheit ist der Effekt noch größer. Im Spektrum des Bereichs  $\gamma$ -1 (46.7-47.7 min) ist fast kein Signal der Glycoform S2 zu erkennen. Im Spektrum des Bereichs  $\gamma$ -2 (47.8-48.7 min) beträgt der Anteil der Glycoform S2 etwa 44%. Diese Beobachtungen zeigen, dass die disialylierten Glycoformen der  $B\beta$ - und  $\gamma$ -Untereinheit höhere Retentionszeiten besitzen als die monosialylierten Glycoformen.

Die partielle Trennung der Phosphorylierungszustände der  $A\alpha$ -Untereinheit und der Glycoformen der  $B\beta$ - und  $\gamma$ -Untereinheit muss beachtet werden, um eine verlässliche und

reproduzierbare Quantifizierung der verschiedenen Spezies zu erreichen. Dieses Problem wird in Abschnitt 4.1.5.3 untersucht.



**Abbildung 19:** Ausschnitt aus dem *total ion current* der chromatographischen Trennung der Fibrinogenuntereinheiten aus Probe Q (Aα: 40.0-41.6 min; Bβ: 42.2-43.9 min; γ: 46.6-48.7 min). Für jede Untereinheit sind zwei Maxima zu erkennen, welche durch partielle Trennung der Glycoformen bzw. Proteinspezies zustande kommen. Die deconvolvierten Massenspektren wurden erhalten, indem die Massenspektren in den markierten Bereichen gemittelt wurden. Es ist zu erkennen, dass die höher phosphorylierten Spezies der Aα-Untereinheit später eluieren als die niedrig phosphorylierten Spezies. Außerdem eluieren die disialylierten Glycoformen der Bβ- und γ-Untereinheit später als die monosialylierten Glycoformen.

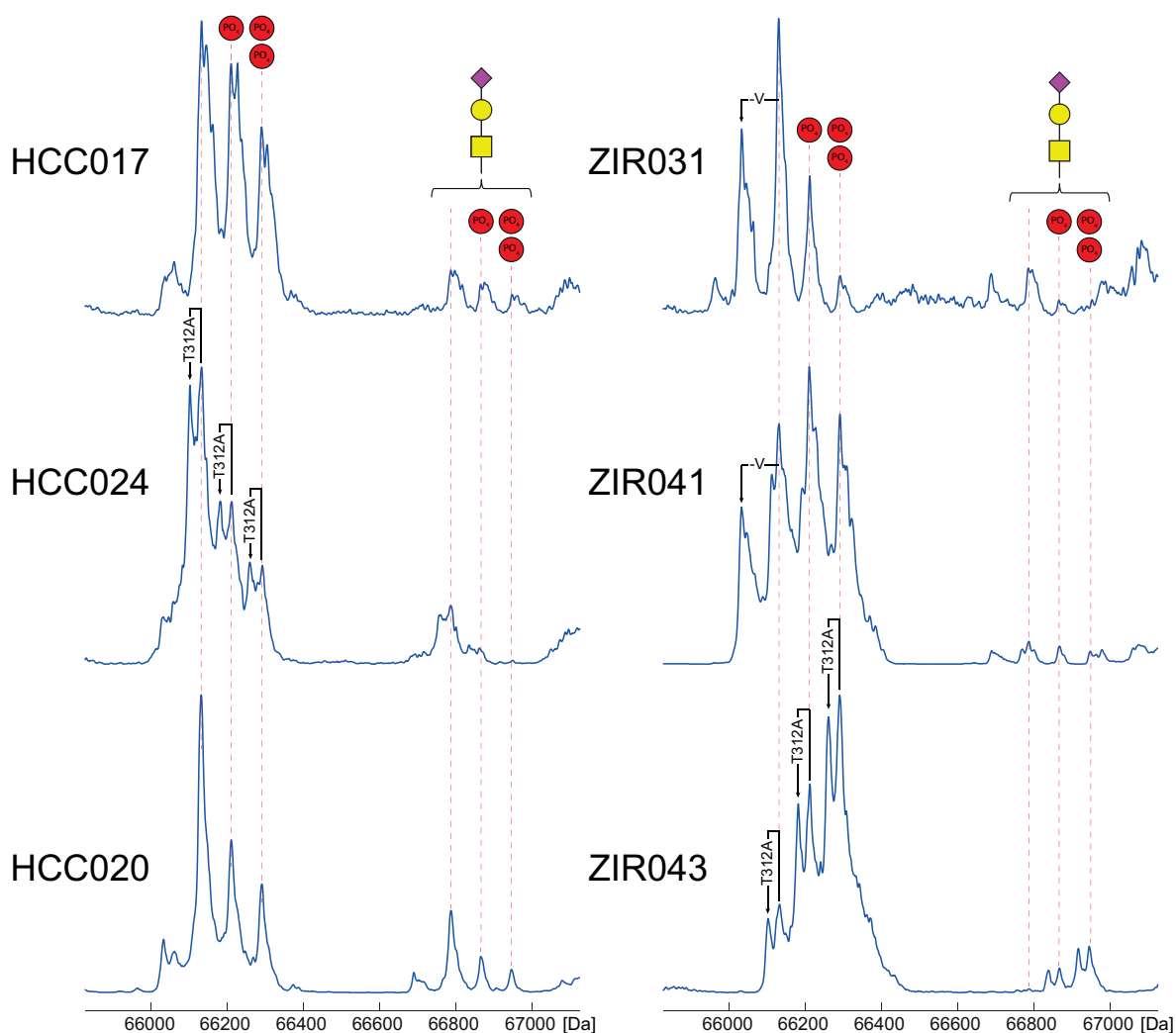
### 4.1.3 Untersuchung der intakten Fibrinogenuntereinheiten von HCC- und Zirrhosepatienten

Es wurden 19 Proben von HCC-Patienten sowie 17 Proben von Leberzirrhosepatienten untersucht (Tabelle 15). Der Großteil der HCC-Patienten hatte auch eine Leberzirrhose, welche in den meisten Fällen die Ursache für den Krebs darstellt. Aus diesem Grund ist die Unterscheidung zwischen Leberzirrhosepatienten ohne HCC und denen, die bereits HCC entwickelt haben, von entscheidender Bedeutung. Die deconvolvierten Massenspektren der Proben von drei HCC-Patienten und drei Leberzirrhosepatienten sind im Folgenden gezeigt.

#### 4.1.3.1 Die $A\alpha$ -Untereinheit

Abbildung 20 zeigt die deconvolvierten Massenspektren der  $A\alpha$ -Untereinheit aus den Proben verschiedener HCC- und Zirrhosepatienten. Die Patienten HCC024 und ZIRO43 sind heterozygot in Bezug auf rs6050. Im Vergleich zu den in Abbildung 16 gezeigten Spektren gesunder Spender sind einige Unterschiede zu erkennen. Der Phosphorylierungsgrad der  $A\alpha$ -Untereinheit ist bei den Patienten HCC024, HCC020 und ZIRO31 deutlich geringer als bei gesunden Spendern. Im Gegensatz dazu zeigt das Spektrum der  $A\alpha$ -Untereinheit von Probe ZIRO43 einen extrem hohen Phosphorylierungsgrad.

In den Spektren der Proben ZIRO31 und ZIRO41 sind zusätzlich die Signale der um eine Valineinheit verkürzten  $A\alpha$ -Untereinheit zu erkennen. Das Signal, welches von der unphosphorylierten Spezies der  $A\alpha$ -Untereinheit ausgeht, ist in Abbildung 20 markiert, die von der mono- und diphosphorylierten Spezies ausgehenden Signale überlagern allerdings mit den Signalen der unverkürzten  $A\alpha$ -Untereinheit. Da diese Signale nur in wenigen Proben auftraten, ist davon auszugehen, dass die trunkeierte Form der  $A\alpha$ -Untereinheit nicht durch Fragmentierung in der Ionenquelle des Massenspektrometers, sondern durch proteolytischen Verdau entsteht. Ob die proteolytische Aktivität eine Folge der Leberzirrhose ist, konnte allerdings nicht geklärt werden. Außerdem sind im Spektrum der  $A\alpha$ -Untereinheit von Probe HCC017 deutlich die Signale der oxidierten  $A\alpha$ -Untereinheit zu erkennen, welche jeweils um 16 Da verschoben sind.

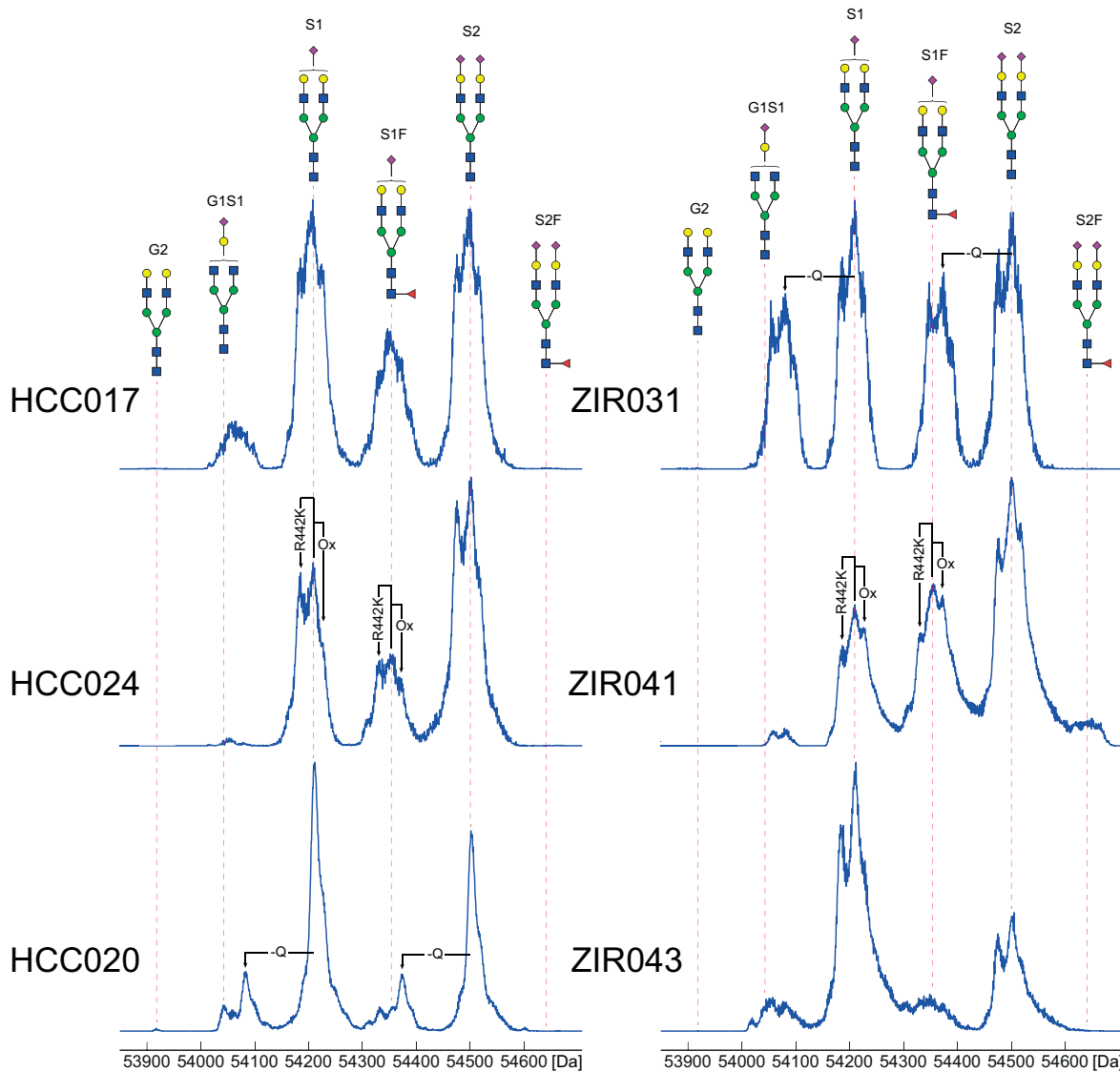


**Abbildung 20: Deconvolvierte Massenspektren der  $A\alpha$ -Untereinheit von humanem Fibrinogen aus den Proben HCC017, HCC024, HCC020, ZIR031, ZIR041 und ZIR043. Die Patienten HCC024 und ZIR043 sind heterozygot in Bezug auf rs6050. Die Signale der mono- und diphosphorylierten Spezies der  $A\alpha$ -Untereinheit sind in den Spektren von ZIR031, HCC024 und HCC020 deutlich kleiner als in den Spektren gesunder Spender (Abbildung 16). Im Gegensatz dazu zeigt das Spektrum von ZIR043 einer sehr hohen Phosphorylierungsgrad. In den Spektren von ZIR031 und ZIR041 sind zusätzlich Signale einer trunkierten, um die C-terminale Valineinheit verkürzten,  $A\alpha$ -Untereinheit zu erkennen. Im Spektrum von HCC017 sind außerdem die Signale der oxidierten  $A\alpha$ -Untereinheit vorhanden.**

#### 4.1.3.2 Die $B\beta$ - und $\gamma$ -Untereinheiten

Die deconvolvierten Massenspektren der  $B\beta$ -Untereinheit aus verschiedenen Proben von HCC und Zirrhosepatienten sind in Abbildung 21 gezeigt. Die auffällig breiten Signale in allen Spektren, außer dem von Probe HCC020, werden durch zwei Gegebenheiten verursacht. Zum einen sind die Spender dieser Proben heterozygot in Bezug auf den SNP rs4220 und besitzen demnach zur Hälfte Fibrinogen mit dem Aminosäureaustausch Arg448Lys auf der  $B\beta$ -Untereinheit. Zum anderen zeigen fast alle HCC- und Zirrhosepatienten einen auffällig

hohen Grad an Oxidation. Die Oxidation der Thioetherfunktion in Methioninseitenketten zum Sulfoxid führt zu einer um 16 Da höheren Masse, welche sich als Schulter in der rechten Flanke der Peaks zeigt. In den Spektren der B $\beta$ -Untereinheit aus den Proben HCC017, HCC024, ZIRO31, ZIRO41 und ZIRO43 überlagern die Signale von nativer und veränderter Proteinsequenz mit den Signalen der oxidierten Spezies und führen so zu den auffällig breiten Signalen.



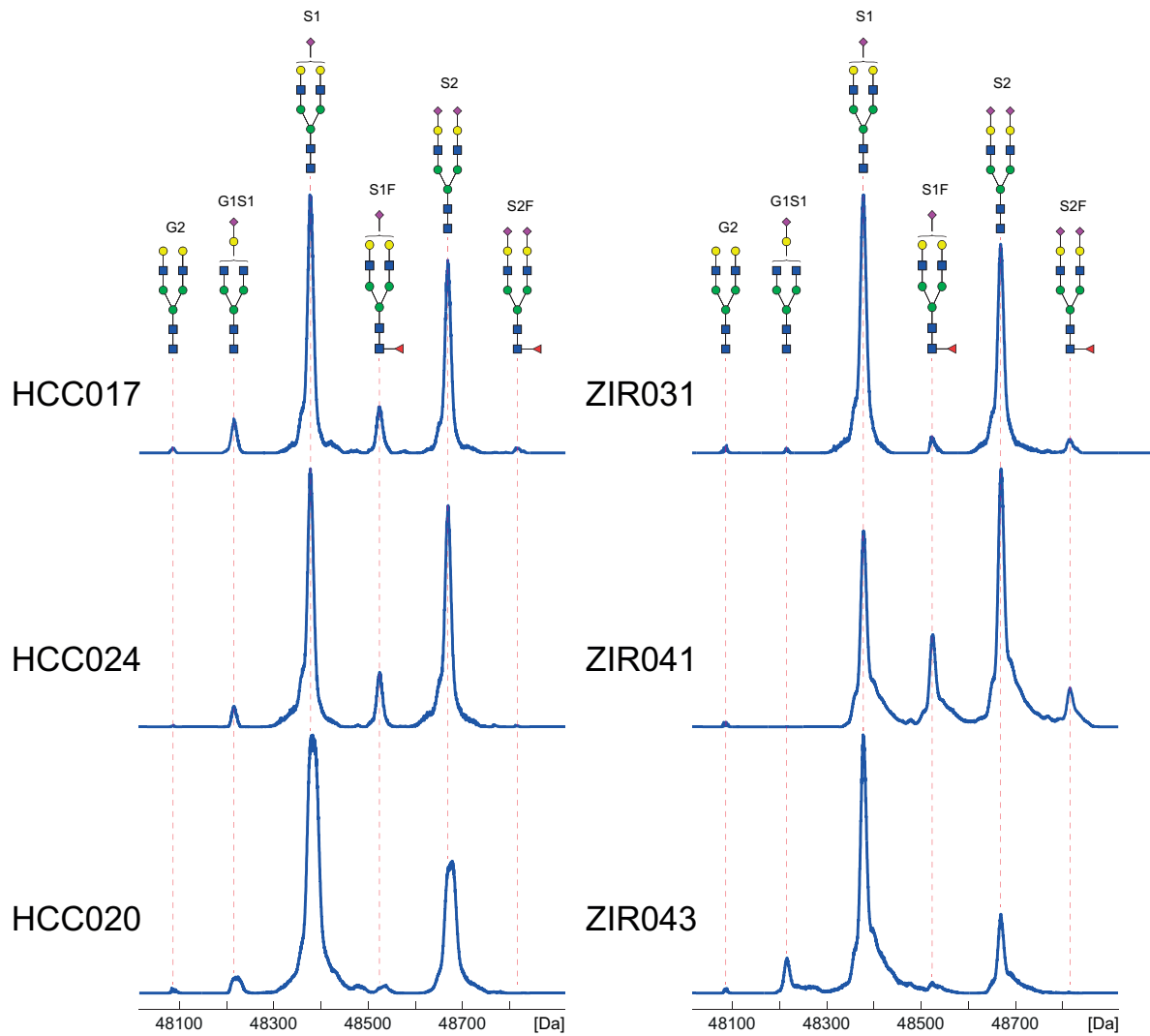
**Abbildung 21:** Deconvolvierte Massenspektren der B $\beta$ -Untereinheit von humanem Fibrinogen aus den Proben HCC017, HCC024, HCC020, ZIRO31, ZIRO41 und ZIRO43. Die breiten Signale in den Spektren entstehen durch die Überlagerung der Signale von nativer und veränderter Proteinsequenz mit den Signalen der oxidierten Spezies. Die um eine Glutamineinheit verkürzte Form der B $\beta$ -Untereinheit ist in den Spektren von ZIRO31 und HCC020 markiert, im Spektrum von ZIRO31 sind diese Signale sehr intensiv und dürfen nicht mit den Signalen der Glycoformen G1S1 und S1F verwechselt werden. Die fucosylierte Glycoform S1F ist in den Spektren der Proben HCC017, HCC024 und ZIRO41 gut erkennbar, die disialylierte, fucosylierte Glycoform S2F ist nur im Spektrum von Probe ZIRO41 als breite Schulter erkennbar. Mit Ausnahme der Probe ZIRO43 ist der Anteil der disialylierten Glycoform S2 höher als in den Proben gesunder Spender (vgl. Abbildung 17).

Im Spektrum der B $\beta$ -Untereinheit von Probe ZIRO31 sind die Signale der um eine Glutamineinheit verkürzten Proteinsequenz (-Q) sehr intensiv. Diese dürfen nicht mit den Signalen der Glycoformen G1S1 und S1F verwechselt werden, mit denen sie sich überschneiden. Da die Signale der Glycoformen G1S1 und S1F im Spektrum der  $\gamma$ -Untereinheit von Probe ZIRO31 (vgl. Abbildung 22) nur schwach zu erkennen sind, ist davon auszugehen, dass es sich hauptsächlich um die trunkierte Proteinsequenz handelt.

Die deconvolvierten Massenspektren der  $\gamma$ -Untereinheit aus den Proben von HCC- und Zirrhosepatienten sind in Abbildung 22 gezeigt, der Vergleich mit den Spektren der B $\beta$ -Untereinheit zeigt, dass die *N*-Glycosylierung der B $\beta$ - und  $\gamma$ -Untereinheit weitestgehend übereinstimmt. In den Spektren der B $\beta$ - und  $\gamma$ -Untereinheit von HCC017, HCC024 und ZIRO41 sind die Signale der monosialylierten, fucosylierten Glycoform S1F gut zu erkennen, allerdings ist die disialylierte, fucosylierte Glycoform S2F nur in der Probe ZIRO41 deutlich vorhanden. In den Spektren der Proben HCC020 und ZIRO43 sind die fucosylierten Glycoformen nicht vorhanden.

Der Vergleich zwischen gesunden und erkrankten Spender zeigt außerdem, dass der Sialylierungsgrad in den Proben von HCC und Zirrhosepatienten erhöht ist. Bei gesunden Spendern lag der Anteil der disialylierten Glycoform S2 bei 22-35%. In den Spektren der B $\beta$ - und  $\gamma$ -Untereinheit aus den Proben HCC017, HCC024 und ZIRO31 liegen die Glycoformen S1 und S2 ungefähr zu gleichen Teilen vor. In der Probe von Patient ZIRO41 ist der Anteil der disialylierten Glycoform S2 sogar deutlich höher, während die Probe von Patient ZIRO43 einen sehr geringen Sialylierungsgrad aufweist.

Das Spektrum der  $\gamma$ -Untereinheit aus der Probe von Patient HCC020 zeigt ungewöhnlich breite Signale. Die genauere Betrachtung zeigt, dass die Signale nur zu einer Seite um ca. 11 Da verbreitert sind, was die Vermutung nahe legt, dass der Patient heterozygot in Bezug auf einen SNP ist. Hierfür infrage kommen eine Reihe von SNPs (rs766346280, rs747416023, rs748671641, rs138511699, rs751212238) auf dem für die  $\gamma$ -Untereinheit codierenden Gen, welche alle den Aminosäureaustausch Threonin  $\rightarrow$  Isoleucin, mit einer Massenzunahme von 12 Da, verursachen.<sup>114</sup>



**Abbildung 22: Deconvolvierte Massenspektren der  $\gamma$ -Untereinheit von humanem Fibrinogen aus den Proben HCC017, HCC024, HCC020, ZIR031, ZIR041 und ZIR043. Die fucosylierte Glycoform S1F ist in den Spektren der Proben HCC017, HCC024 und ZIR041 gut erkennbar. Die disialylierte, fucosylierte Glycoform S2F ist nur im Spektrum der Probe ZIR041 vorhanden. In den Spektren der Proben HCC020 und ZIR043 sind die Signale der fucosylierten Glycoformen S1F und S2F nicht vorhanden. Im Vergleich zu den Spektren gesunder Spender (Abbildung 18) ist der Anteil der disialylierten Glycoform S2 höher. Nur die Probe ZIR043 zeigt einen niedrigen Sialyierungsgrad. Im Spektrum von Probe HCC020 ist eine einseitige Signalverbreiterung um ca. 11 Da zu erkennen, welche vermutlich durch einen heterozygoten SNP verursacht wird.**



#### 4.1.4 Untersuchung des Fibrinogens mittels enzymatischen Verdaus

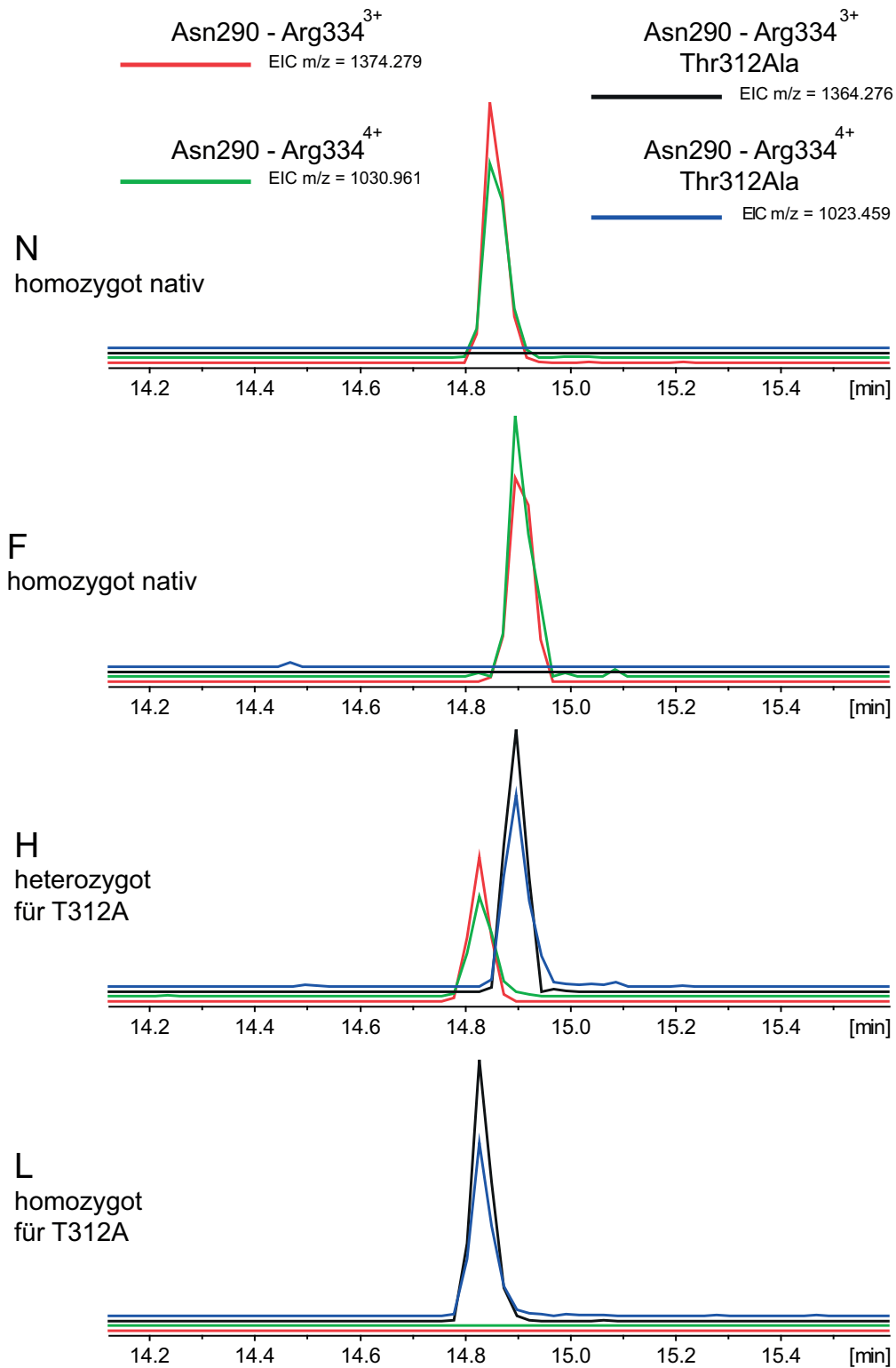
Bei der Untersuchung der intakten Untereinheiten von humanem Fibrinogen wurde die Phosphorylierung und O-Glycosylierung der A $\alpha$ -Untereinheit sowie die N-Glycosylierung der B $\beta$ - und  $\gamma$ -Untereinheit gezeigt. Die posttranslationalen Modifikationen wurden auch auf Peptidebene untersucht. Außerdem wurden auf allen Untereinheiten durch SNPs verursachte Aminosäureaustausche gefunden, welche anhand der tryptischen Peptide nachgewiesen wurden. Darüber hinaus wurde die Identität und Reinheit des isolierten Fibrinogens durch die Analyse der tryptischen Peptide bestätigt.

##### 4.1.4.1 Analyse der SNPs anhand der tryptischen Peptide

Die SNPs, die zum Austausch einer Aminosäure führen, wurden auch anhand der tryptischen Peptide identifiziert. Hierzu wurde Fibrinogen der Spender N, F, H und L mit Trypsin verdaut und mittels LC-MS untersucht. In den Daten wurde nach den Signalen der entsprechenden Peptide mit und ohne den Aminosäureaustausch gesucht.

Der SNP rs6050 führt zum Aminosäureaustausch Thr<sub>312</sub>Ala auf der A $\alpha$ -Untereinheit von humanem Fibrinogen. Threonin-312 befindet sich auf dem 45 Aminosäuren langen tryptischen Peptid Asn<sup>290</sup>-Arg<sup>334</sup>, welches dominant im drei- und vierfach positiv geladenen Zustand auftritt. In Abbildung 23 sind die *extracted ion*-Chromatogramme (EICs) des nativen Peptides (rot und grün) und des veränderten Peptides (schwarz und blau) dargestellt. In den Chromatogrammen der Proben N und F sind nur die Signale des nativen Peptides zu erkennen. Die Chromatogramme von Probe H zeigen sowohl die Signale des nativen als auch des veränderten Peptides, da Spender H heterozygot in Bezug auf rs6050 ist. In den Chromatogrammen von Probe L sind nur die Signale des veränderten Peptides vorhanden, da Spender L homozygot für rs6050 ist.

Der SNP rs4220 verursacht den Aminosäureaustausch Arg<sub>448</sub>Lys auf der B $\beta$ -Untereinheit von humanem Fibrinogen. Da Trypsin das Peptidrückgrat sowohl hinter Arginin als auch hinter Lysin schneidet, bleibt die Schnittstelle vorhanden. In Abbildung 24 sind die EICs des nativen tryptischen Peptides <sup>442</sup>GSWYSMR<sup>448</sup> in grün (einfach positiv geladener Zustand) und in rot (zweifach positiv geladener Zustand) dargestellt. Die EICs des Peptides <sup>442</sup>GSWYSMK<sup>448</sup>, welches den Aminosäureaustausch Arg<sub>448</sub>Lys trägt, sind in blau (einfach positiv geladen) und schwarz (zweifach positiv geladen) dargestellt. In den Chromatogrammen der Proben F und L sind nur die Signale des nativen Peptides vorhanden, da die Spender F und L den SNP



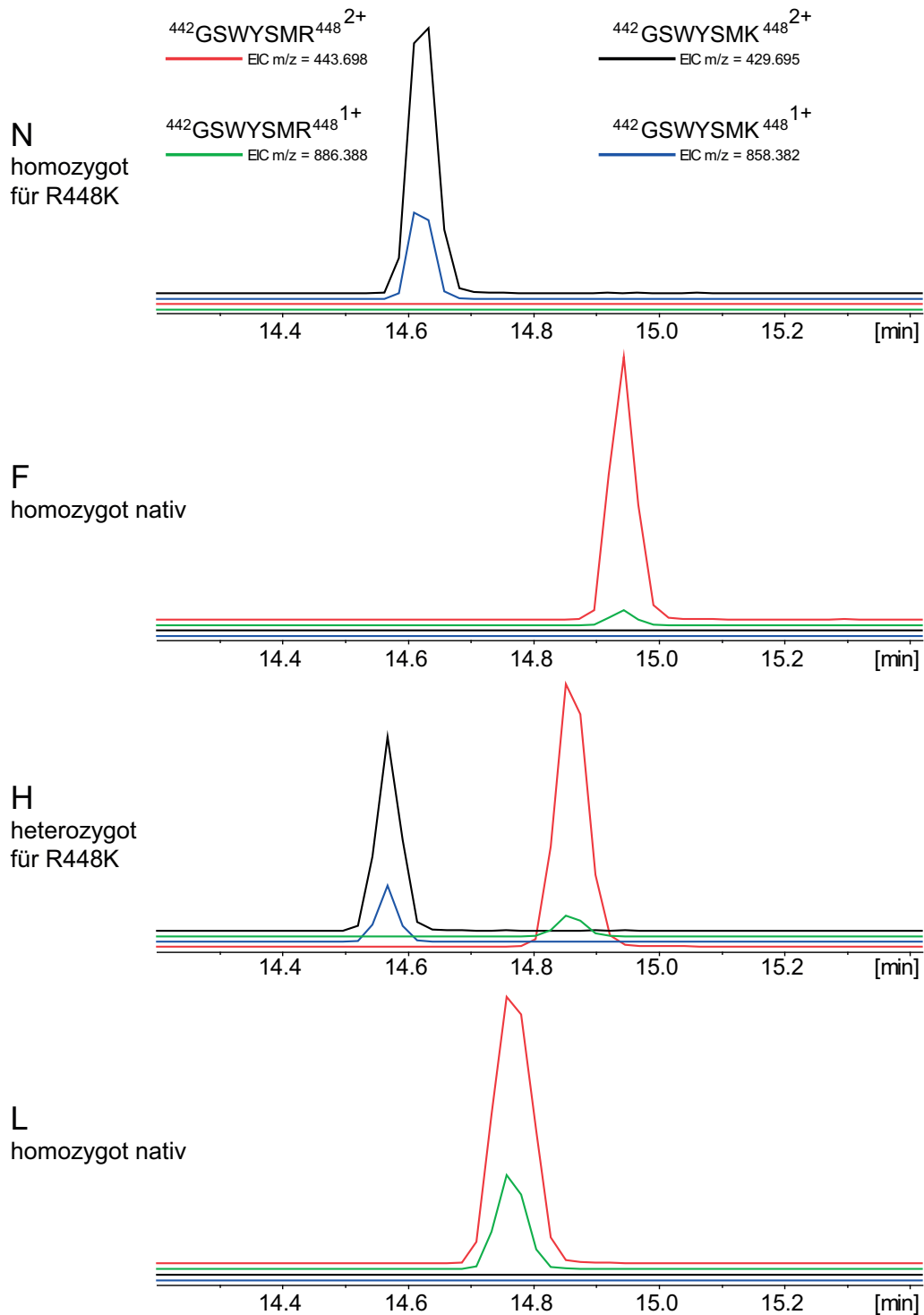
**Abbildung 23:** *Extracted ion*-Chromatogramme des tryptischen Peptids Asn290-Arg334 der  $\alpha$ -Untereinheit von humanem Fibrinogen. In Rot und Grün sind die EICs des dreifach und vierfach positiv geladenen Zustands des nativen Peptids dargestellt. In Schwarz und Blau sind die EICs des dreifach und vierfach positiv geladenen Zustands des Peptids mit dem Aminosäureaustausch Thr312Ala gezeigt. In den Chromatogrammen der Proben N und F sind nur die Signale des nativen Peptids (rot und grün) zu erkennen. In den Chromatogrammen von Probe H, sind sowohl die Signale des nativen als auch des veränderten Peptids vorhanden, da Spender H heterozygot in Bezug auf rs6050 ist. In den Chromatogrammen von Probe L sind nur die Signale des veränderten Peptids vorhanden, da Spender L homozygot für rs6050 ist.

nicht tragen. In den Chromatogrammen von Probe H sind die Signale beider Peptide zu erkennen, da Spender H heterozygot in Bezug auf rs4220 ist. Spender N ist homozygot für rs4220, aus diesem Grund sind in den Chromatogrammen von Probe N nur die Signale des veränderten Peptids vorhanden.

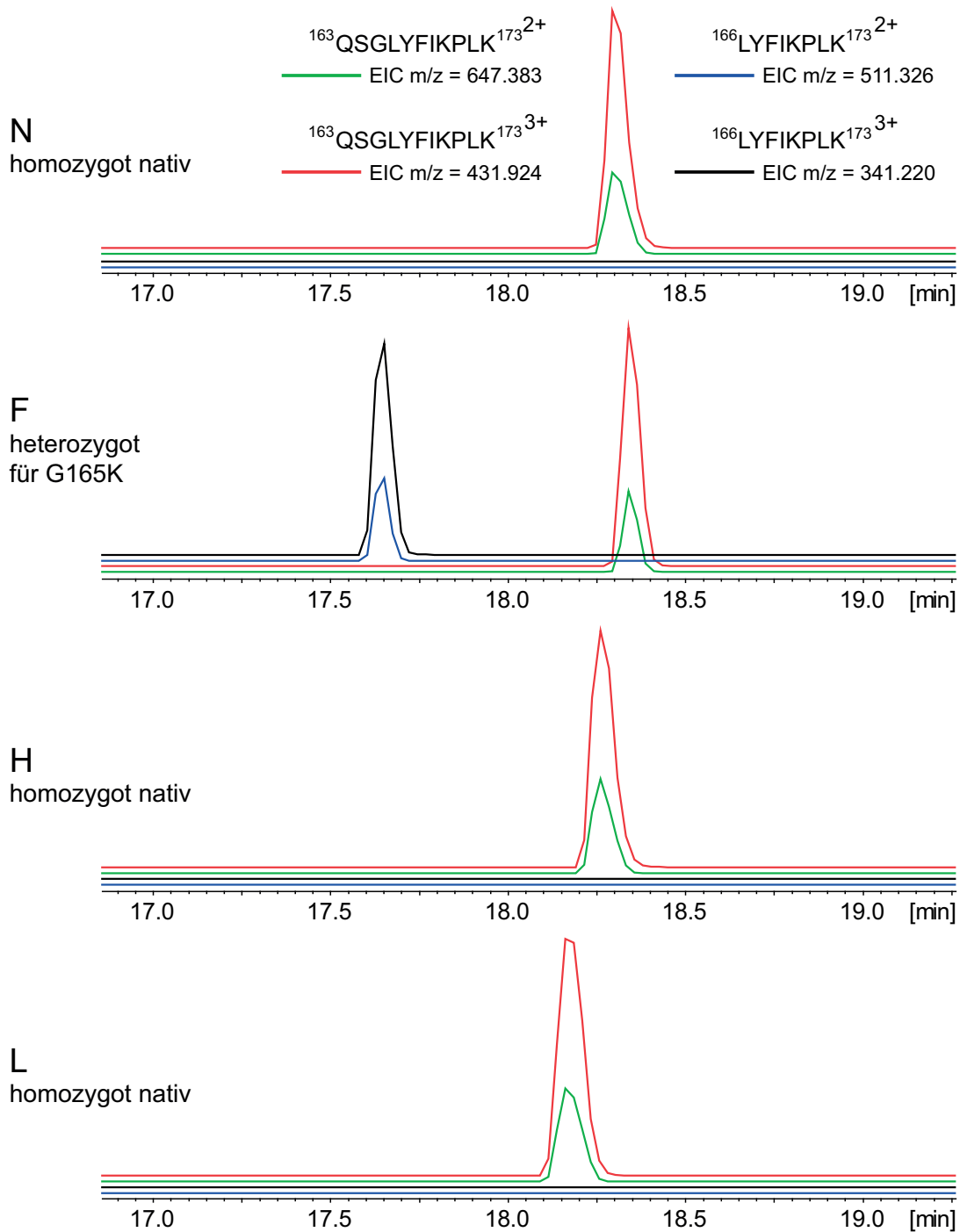
Der SNP rs6063 führt zum Aminosäureaustausch Gly165Arg auf der  $\gamma$ -Untereinheit von humanem Fibrinogen. Dabei wird eine neue Schnittstelle für Trypsin eingeführt. Die EICs des nativen tryptischen Peptids  $^{163}\text{QSGLYFIKPLK}^{173}$  sind in Abbildung 25 dargestellt (grün: zweifach positiv geladener Zustand, rot: dreifach positiv geladener Zustand). Außerdem sind die EICs des um drei Aminosäuren verkürzten Peptids  $^{166}\text{LYFIKPLK}^{173}$  dargestellt (blau: zweifach positiv geladener Zustand, schwarz: dreifach positiv geladener Zustand). In den Proben N, H und L ist nur das native Peptid vorhanden, da keiner der Spender den SNP rs6063 trägt. Im Chromatogramm der Probe F sind die Signale des verkürzten Peptids neben den Signalen des nativen Peptids zu erkennen, da Spender F heterozygot in Bezug auf rs6063 ist.

Leider waren bei diesen Experimenten die Retentionszeiten der Peptide nicht reproduzierbar, dies hat aber keine Auswirkungen auf die Interpretation, da die Peptide eindeutig anhand ihrer Masse-zu-Ladungs-Verhältnisse identifiziert werden können. Die in den Spektren der intakten Fibrinogenuntereinheiten gefundenen Aminosäureaustausche konnten somit auf Peptidebene bestätigt werden.

## Ergebnisse und Diskussion



**Abbildung 24:** *Extracted ion*-Chromatogramme des tryptischen Peptids  $^{442}\text{GSWYSMK}^{448}$  der nativen  $\beta$ -Untereinheit von humanem Fibrinogen (grün: einfach positiv geladener Zustand, rot: zweifach positiv geladener Zustand) und des tryptischen Peptids  $^{442}\text{GSWYSMR}^{448}$ , welches den Aminosäureaustausch Arg448Lys trägt (blau: einfach positiv geladener Zustand, schwarz: zweifach positiv geladener Zustand). Die Chromatogramme der Proben F und L zeigen nur das native Peptid. In den Chromatogrammen von Probe H sind sowohl die Signale des nativen als auch des veränderten Peptids zu erkennen, da Spender H heterozygot bezüglich rs4220 ist. In den Chromatogrammen von Probe N ist nur das veränderte Peptid vorhanden, da Spender N homozygot für rs4220 ist.



**Abbildung 25:** *Extracted ion*-Chromatogramme des tryptischen Peptids  $^{163}\text{QSGLYFIKPLK}^{173}$  der nativen  $\gamma$ -Untereinheit von humanem Fibrinogen (grün: zweifach positiv geladener Zustand, rot: dreifach positiv geladener Zustand) und des tryptischen Peptids  $^{166}\text{LYFIKPLK}^{173}$ , welches durch den Aminosäureaustausch G165R entsteht, da eine zusätzliche Trypsinschnittstelle eingefügt wird (blau: zweifach positiv geladener Zustand, schwarz: dreifach positiv geladener Zustand). Die Chromatogramme der Proben N, H, und L zeigen nur die Signale des Peptids  $^{163}\text{QSGLYFIKPLK}^{173}$ . Im den Chromatogrammen von Probe F sind auch die Signale des Peptids  $^{166}\text{LYFIKPLK}^{173}$  zu erkennen, da Spender F heterozygot in Bezug auf rs6063 ist.

#### 4.1.4.2 Identifizierung des Fibrinogen anhand der tryptischen Peptide

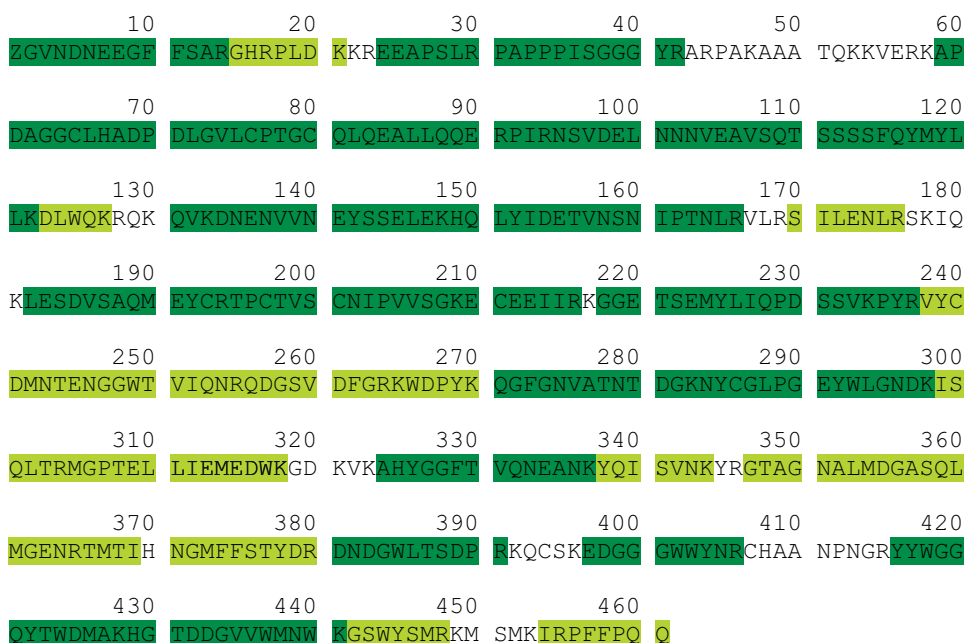
Die Identität und Reinheit des isolierten Fibrinogens wurde durch einen tryptischen Verdau analysiert. Hierfür wurde Fibrinogen aus den Proben N, O, Q und T mit Trypsin verdaut und mittels LC-MS/MS untersucht. Die erhaltenen Daten wurden manuell nach den Signalen der tryptischen Peptide von Fibrinogen durchsucht. Falls vorhanden, wurden die MS/MS-Spektren extrahiert und nach den für die Aminosäuresequenz spezifischen Massendifferenzen durchsucht. Außerdem wurde eine MASCOT-Suche durchgeführt. Die Ergebnisse sind in Tabelle 4 zusammengefasst.

**Tabelle 4: Ergebnisse der MASCOT-Suchen der Proben N, O, Q und T. Der Protein score, die Anzahl der identifizierten Sequenzen (Seq.) und der emPAI (*Exponentially Modified Protein Abundance Index*) sind für alle Proteine, die in mindestens einer Probe mit einem Score von über 100 identifiziert wurden, angegeben. Die drei Fibrinogenuntereinheiten und Albumin wurden in allen Proben mit sehr hohen scores identifiziert. In den Proben Q und T wurde außerdem Fibronectin mit sehr hohem score identifiziert. Bei allen anderen identifizierten Proteinen handelt es sich um abundante Plasmaproteine wie Immunglobuline, Serotransferrin,  $\alpha$ 1-Antitrypsin oder Komplementfaktoren.**

Protein	Probe N			Probe O			Probe Q			Probe T		
	Score	Seq.	emPAI	Score	Seq.	emPAI	Score	Seq.	emPAI	Score	Seq.	emPAI
ALBU_HUMAN	1948	23	2.44	2316	24	2.62	2339	27	3.31	2368	26	3.04
FIBA_HUMAN	1776	19	1.14	2062	21	1.32	2224	22	1.43	2203	19	1.14
FIBB_HUMAN	1530	15	1.75	2071	19	2.60	2012	19	2.64	2024	18	2.37
FIBG_HUMAN	1564	13	1.59	1712	14	1.79	1807	16	2.26	1760	16	2.23
FINC_HUMAN	463	8	0.12	585	9	0.14	2599	21	0.36	2261	21	0.36
IGHG1_HUMAN	782	7	1.07	924	7	1.06	1008	7	1.08	665	6	0.86
IGHG2_HUMAN	301	4	0.52	230	4	0.52	209	4	0.52	174	3	0.37
IGKC_HUMAN	622	5	3.78	815	5	3.77	900	5	3.82	827	5	3.78
LAC2_HUMAN	575	4	2.62	768	4	2.61	798	4	2.65	527	4	2.62
IGLL5_HUMAN	554	4	0.90	739	4	0.90	778	4	0.91	517	4	0.90
TRFE_HUMAN	182	1	0.05	187	1	0.05	106	1	0.05	269	4	0.21
APOA1_HUMAN	107	1	0.13	244	2	0.28	137	1	0.13	371	3	0.45
CO3_HUMAN	156	3	0.06	468	7	0.15	774	9	0.20	163	3	0.06
A1AT_HUMAN	144	2	0.18	246	2	0.18	238	2	0.18	237	2	0.18
A2MG_HUMAN	117	2	0.05	183	2	0.05	283	6	0.15	305	6	0.15
IGHA1_HUMAN	-	-	-	-	-	-	125	1	0.10	179	2	0.22
APOA2_HUMAN	113	1	0.38	120	1	0.38	90	1	0.39	153	2	0.91
IGHM_HUMAN				271	4	0.36	247	3	0.26	96	2	0.16
C1QC_HUMAN	121	2	0.34	131	2	0.34	121	2	0.34	-	-	-
C1R_HUMAN	110	1	0.05	60	1	0.05	103	2	0.10	-	-	-
HV305_HUMAN	-	-	-	151	1	0.32	-	-	-	-	-	-
IGHG3_HUMAN	-	-	-	410	5	0.57	431	5	0.57	-	-	-

In Tabelle 4 sind der Protein score, die Anzahl der identifizierten Sequenzen und der empPAI (*Exponentially Modified Protein Abundance Index*) für alle Proteine, die in mindestens einer Probe mit einem score von über 100 identifiziert wurden, angegeben. Die drei Fibrinogenuntereinheiten sind in allen Proben mit einem sehr hohen score identifiziert worden und befinden sich unter den ersten fünf Treffern. In allen Proben wurde Albumin mit einem noch höheren score identifiziert. Dies ist nicht verwunderlich, da Albumin mit sehr hoher Konzentration im Blutplasma vorkommt (35 – 50 mg/mL) und Fibrinogen bindende Eigenschaften besitzt.<sup>70</sup>

Die manuelle Suche nach den tryptischen Peptiden der Fibrinogenuntereinheiten und Albumin ergab Sequenzabdeckungen von 83 – 88%. Die Sequenzabdeckung der  $\beta$ -Untereinheit von humanem Fibrinogen aus Probe N ist in Abbildung 26 dargestellt, die grün markierten Bereiche wurden massenspektrometrisch nachgewiesen, die dunkelgrün markierten Bereiche wurden anhand von MS/MS-Spektren eindeutig identifiziert. In Tabelle 4 sind die Sequenzabdeckungen der Fibrinogenuntereinheiten sowie Albumin aus Probe N zusammengefasst. Im Anhang ist die Liste aller identifizierten Peptide zu finden (Abschnitt 9.1).



**Abbildung 26: Aminosäuresequenz der  $\beta$ -Untereinheit von humanem Fibrinogen. Die grün markierten Bereiche der Sequenz wurden massenspektrometrisch nachgewiesen. Die Sequenzabdeckung beträgt 88%. Die dunkelgrün markierten Bereiche wurden anhand von MS/MS-Spektren eindeutig identifiziert.**

**Tabelle 5: Sequenzabdeckungen der Fibrinogenuntereinheiten und Albumin aus Probe N sowie Fibronectin aus Probe Q.**

<b>Protein</b>	<b><i>Sequence coverage</i></b>	<b><i>MS/MS coverage</i></b>
A $\alpha$ -Untereinheit	83%	38%
B $\beta$ -Untereinheit	88%	60%
$\gamma$ -Untereinheit	86%	63%
Albumin	86%	61%
Fibronectin (Probe Q)	70%	29%

Da Fibronectin in den Proben Q und T mit sehr hohem *score* identifiziert wurde (Tabelle 4), wurde in den Daten von Probe Q manuell nach den tryptischen Peptiden von Fibronectin gesucht. Tatsächlich wurde Fibronectin mit einer Sequenzabdeckung von 70% identifiziert. Fibronectin bindet spezifisch an Fibrinogen und ist bei der Isolierung daher nur schwierig von Fibrinogen zu trennen.<sup>70,115</sup> Bei den anderen in der MASCOT-Suche identifizierten Proteinen handelt es sich um hoch abundante Plasmaproteine wie Immunglobuline, Serotransferrin,  $\alpha$ 1-Antitrypsin und Komplementfaktoren, welche aber deutlich niedrigere *scores* aufweisen.

Die Identität des isolierten Fibrinogens konnte durch die LC-MS/MS-Analyse der tryptischen Verdäue bestätigt werden. Es zeigte sich aber, dass das isolierte Fibrinogen Verunreinigungen aufweist. Die abundantesten Verunreinigungen sind Albumin und Fibronectin, welche in sehr hohen Konzentrationen im Blutplasma vorkommen und an Fibrinogen binden. Bei der Isolierung des Fibrinogens im kleinen Maßstab können pH-Wert, Salzkonzentration und Temperatur nicht immer exakt eingehalten werden. Außerdem ist bei der Zugabe kleiner Volumina Ethanol mit einem großen Fehler zu rechnen. Aus diesen Gründen ist es nicht verwunderlich, dass neben Fibrinogen auch einige andere Proteine isoliert werden. Aufgrund der Fibrinogen-bindenden Eigenschaften von Albumin und Fibronectin kann aber vermutlich nicht verhindert werden, dass diese, selbst bei perfekter Einhaltung der Bedingungen, mitgefällt werden. Die Verunreinigungen stören bei der hier entwickelten Methode nicht, da die Proteine im intakten Zustand analysiert werden. Außerdem werden die Verunreinigungen durch die Flüssigkeitschromatographie von den Fibrinogenuntereinheiten getrennt.



#### 4.1.4.3 Analyse der glycosylierten und phosphorylierten tryptischen Peptide von humanem Fibrinogen

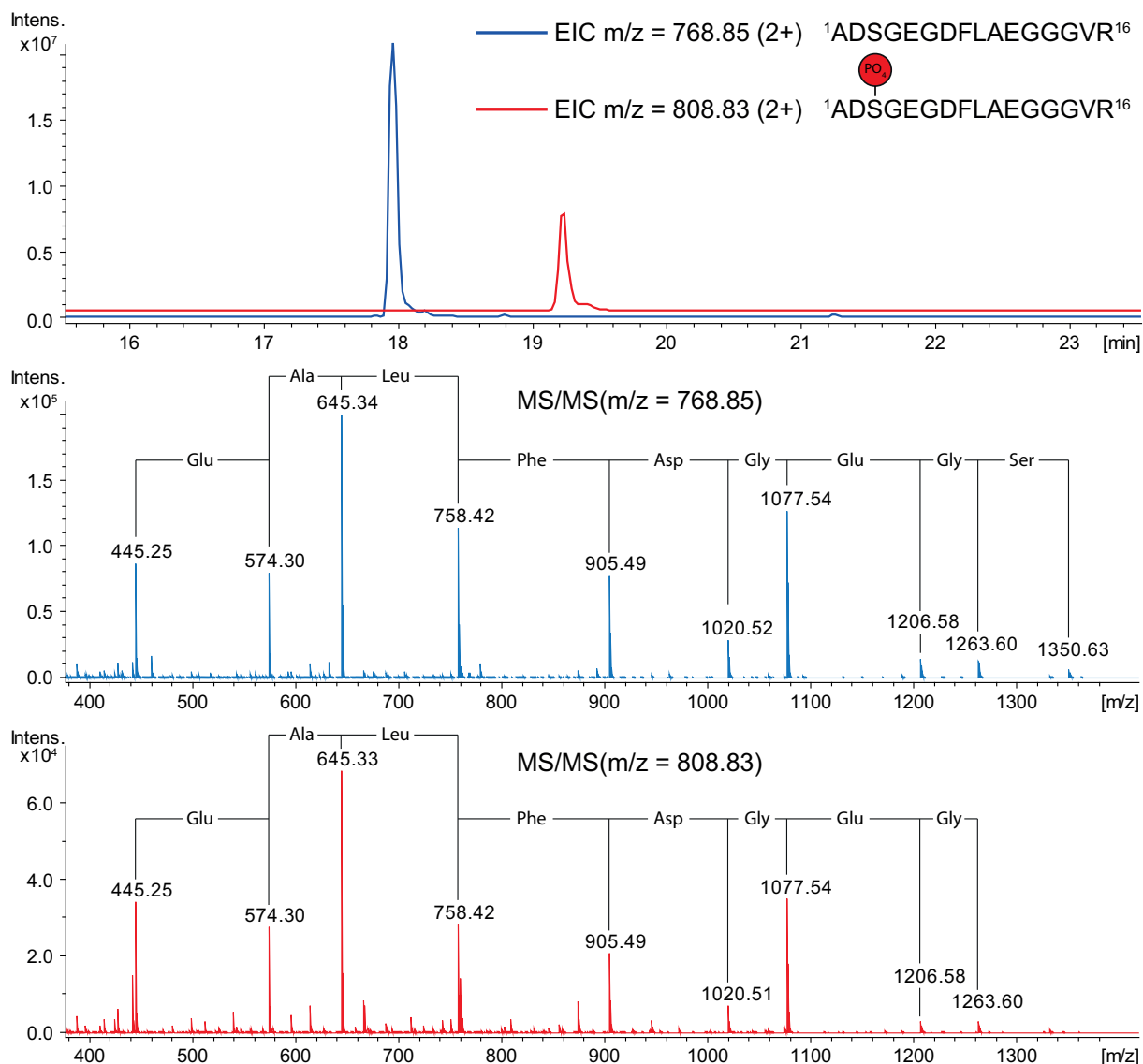
In Abschnitt 4.1.1 wurde die Phosphorylierung und O-Glycosylierung der A $\alpha$ -Untereinheit sowie die N-Glycosylierung der B $\beta$ - und  $\gamma$ -Untereinheit diskutiert. Es wurde deshalb gezielt nach phosphorylierten und glycosylierten tryptischen Peptiden gesucht.

Alle literaturbekannten Phosphorylierungsstellen der A $\alpha$ -Untereinheit sind in Tabelle 6 aufgeführt. Es konnten insgesamt fünf phosphorylierte Peptide identifiziert werden. Die beiden Phosphorylierungsstellen Serin-3 und Serin-345 konnten mit hoher Intensität detektiert werden, während die Phosphorylierungsstellen Thr-393, Serin-272/Serin-275 sowie Ser-482/Thr-486 nur mit geringer Intensität nachgewiesen werden konnten. (Die beiden letzten Phosphorylierungsstellen liegen jeweils auf den gleichen tryptischen Peptiden und können daher nicht differenziert werden.) Dieses Ergebnis steht in Übereinstimmung mit der Literatur, in der Serin-3 und Serin-345 als die dominanten Phosphorylierungsstellen angegeben werden.<sup>79-81</sup>

In Abbildung 27 sind die EICs und die MS/MS-Spektren des Peptids <sup>1</sup>ADSGEGDFLAEGGV<sup>16</sup> im unphosphorylierten und phosphorylierten Zustand gezeigt. Beide Peptide konnten eindeutig anhand der  $\gamma$ -Fragmentationenserie identifiziert werden. Beim unphosphorylierten Peptid kann die  $\gamma$ -Fragmentationenserie bis zum Serin-3 verfolgt werden, beim phosphorylierten Peptid bricht die  $\gamma$ -Fragmentationenserie nach Glycin-4 ab, da Serin-3 phosphoryliert ist.

**Tabelle 6: Übersicht der literaturbekannten Phosphorylierungsstellen der A $\alpha$ -Untereinheit von humanem Fibrinogen. Neben der phosphorylierten Aminosäure und der Literaturstelle sind die Peptidsequenz und das m/z-Verhältnis (Ladungszustand) angegeben.**

Phosphorylierungsstelle		Peptid	m/z	
Ser-3	78,79,81	<sup>1</sup> ADS(PO <sub>4</sub> )GEGDFLAEGGGVR <sup>16</sup>	808.83 (2+)	✓✓
Ser-26	84	<sup>24</sup> HQS(PO <sub>4</sub> )ACK <sup>29</sup>	810.30 (1+)	×
Ser-31	83	<sup>30</sup> DS(PO <sub>4</sub> )DWPFCSEDEDWNYK <sup>44</sup>	681.91 (3+)	×
Ser-37	84	<sup>30</sup> DSDWPFCS(PO <sub>4</sub> )DEDWNYK <sup>44</sup>		
Ser-262	83	<sup>253</sup> GGSTSYGTGS(PO <sub>4</sub> )ETESPR <sup>268</sup>	826.83 (2+)	×
Ser-272	83	<sup>269</sup> NPSS(PO <sub>4</sub> )AGSWNSGSSGPGSTGNR <sup>289</sup>	681.94 (3+)	✓
Ser-275	83	<sup>269</sup> NPSSAGS(PO <sub>4</sub> )WNSGSSGPGSTGNR <sup>289</sup>		
Ser-345	80-82,84	<sup>335</sup> PGSTGTWNPGS(PO <sub>4</sub> )SER <sup>348</sup>	756.81 (4+)	✓✓
Thr-393	82	<sup>349</sup> GSAGHWTSESSVSGSTGQWHSESGSFRPDSP	1211.73	✓
		GSGNARPNNPDWGT(PO <sub>4</sub> )FEEVSGNVSPGTR <sup>406</sup>	(5+)	
Ser-432	83	<sup>430</sup> VTS(PO <sub>4</sub> )GSTTTTR <sup>439</sup>	545.74 (2+)	×
Ser-482	83	<sup>462</sup> EVVTS(PO <sub>4</sub> )GSDCPEAMDGLGTLG(PO <sub>4</sub> )GIGTLDGFR <sup>491</sup>	1070.13	✓
Thr-486	83	<sup>462</sup> EVVTS(PO <sub>4</sub> )GSDCPEAMDGLGTLG(PO <sub>4</sub> )LDGFR <sup>491</sup>	(3+)	
Ser-505	82,84	<sup>494</sup> HPDEAAFFDTAS(PO <sub>4</sub> )TGK <sup>508</sup>	837.35 (2+)	×
Ser-530	82	<sup>529</sup> GS(PO <sub>4</sub> )ESGIFTNTK <sup>539</sup>	610.76 (2+)	×
Ser-541	84	<sup>540</sup> ES(PO <sub>4</sub> )SSHHPGIAEFPSR <sup>554</sup>	859.37 (2+)	×
Ser-590	82,84	<sup>584</sup> MADEAGS(PO <sub>4</sub> )EADHEGTHSTK <sup>601</sup>	651.58 (3+)	×



**Abbildung 27: Extracted ion-Chromatogramme und MS/MS-Spektren des zweifach geladenen Peptids  ${}^1\text{ADSGEGDFLAEGGGVR}^{16}$  im unphosphorylierten und phosphorylierten Zustand. Beide Peptide konnten anhand der y-Fragmentationsserie eindeutig identifiziert werden. Beim unphosphorylierten Peptid kann die y-Fragmentationsserie bis zum Serin-3 verfolgt werden. Beim phosphorylierten Peptid bricht die y-Fragmentationsserie nach Glycin-4 ab, da Serin-3 phosphoryliert ist.**

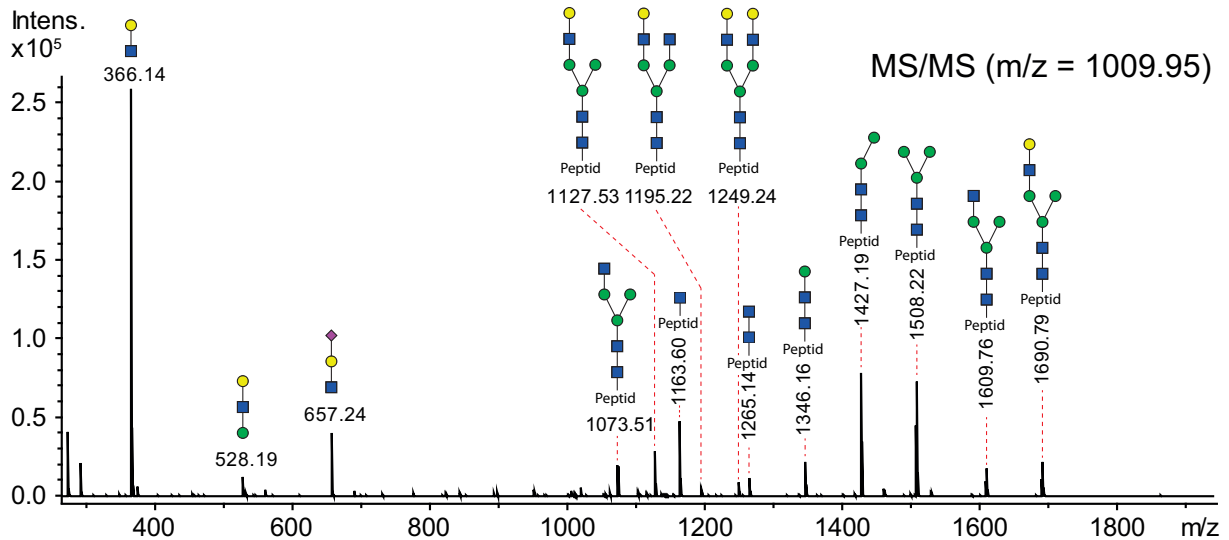
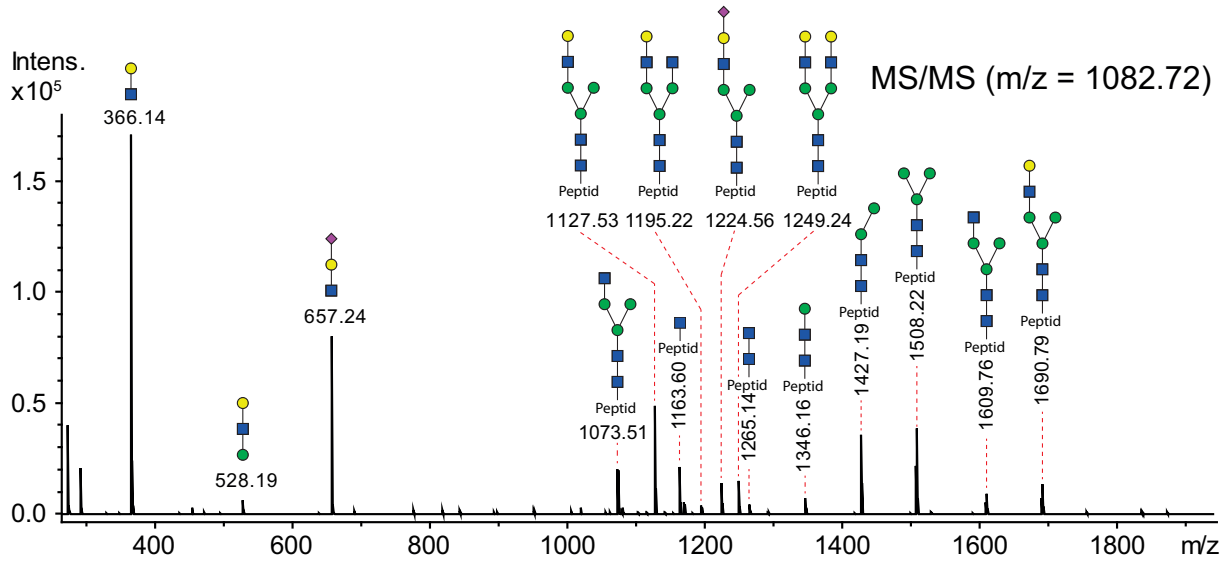
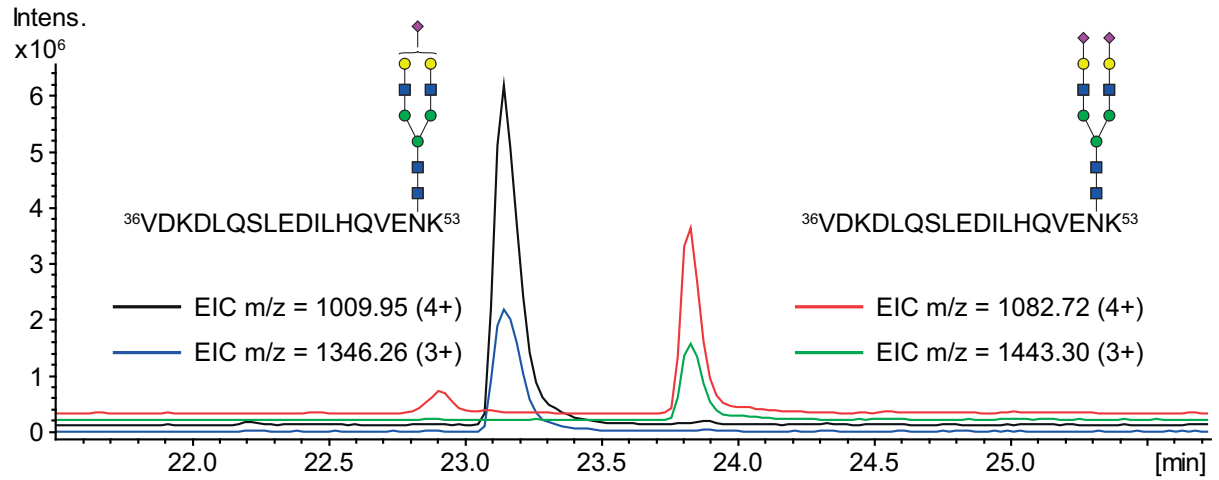
In der Veröffentlichung von Zauner *et al.* wurden verschiedene O-glycosylierte Peptide im C-terminalen Bereich der A $\alpha$ -Untereinheit von humanem Fibrinogen identifiziert.<sup>85</sup> In Tabelle 7 sind die potentiellen O-glycosylierten Peptide der A $\alpha$ -Untereinheit zusammengefasst, vier davon konnten in Probe N beobachtet werden. Auf eine Quantifizierung wurde verzichtet, da glycosylierte Peptide deutlich schlechter ionisierbar sind als unglycosylierte Peptide.

**Tabelle 7: Potentielle O-glycosylierte Peptide der C-terminalen Region der A $\alpha$ -Untereinheit von humanem Fibrinogen nach Zauner *et al.*<sup>85</sup> Vier O-glycosylierte Peptide konnten in Probe N nachgewiesen werden.**

Peptidsequenz	m/z	
<sup>494</sup> HPDEAAFFDT <b>ASTGK</b> <sup>508</sup>	750.65 (3+)	✓
<sup>529</sup> G <b>SESGIFTNTK</b> <sup>539</sup>	898.90 (2+)	×
<sup>540</sup> E <b>SSSHHPGIAEFPSR</b> <sup>554</sup>	765.34 (3+)	✓
<sup>529</sup> G <b>SESGIFTNTKESSSHHPGIAEFPSR</b> <sup>554</sup>	854.64 (4+)	×
<sup>563</sup> Q <b>FTSSTSYNR</b> <sup>572</sup>	923.89 (2+)	×
<sup>573</sup> GD <b>STFESK</b> <sup>580</sup>	763.81 (2+)	×
<sup>563</sup> Q <b>FTSSTSYNRGDSTFESK</b> <sup>580</sup>	900.05 (3+)	✓
<sup>529</sup> G <b>SESGIFTNTKESSSHHPGIAEFPSRGK</b> <sup>556</sup>	900.92 (4+)	×
<sup>557</sup> <b>SSSYSKQFTSSTSYNRGDSTFESK</b> <sup>580</sup>	835.11 (4+)	×
<sup>557</sup> <b>SSSYSKQFTSSTSYNR</b> <sup>572</sup>	829.36 (3+)	✓
<sup>557</sup> <b>SSSYSK</b> <sup>562</sup>	657.77 (2+)	×

Die *extracted ion*-Chromatogramme und MS/MS-Spektren des N-glycosylierten Peptids <sup>36</sup>VDKDLQSLEDILHQVENK<sup>53</sup> der  $\gamma$ -Untereinheit von humanem Fibrinogen in der mono- und disialylierten Glycoform sind in Abbildung 28 gezeigt. Anhand der Peakflächen der EICs wurde der Anteil der disialylierten Glycoform zu 34.5% bestimmt. Die MS/MS-Spektren beider Glycoformen zeigen die Fragmentierung des komplexen N-Glycans. Die Spektren sind sich sehr ähnlich, einzig das Signal des Fragmentions bei m/z = 1224.56 ist im Spektrum des disialylierten Glycopeptids vorhanden, im Spektrum des monosialylierten jedoch nicht. Außerdem ist die Intensität des sialylierten Antennenfragmentes m/z = 657.24 im Spektrum des disialylierten Glycopeptids deutlich höher. Das Glycopeptid der B $\beta$ -Untereinheit <sup>347</sup>GTAGNALMDGASQLMGENR<sup>365</sup>, welches an Asparagin-364 N-glycosyliert ist, wurde nur mit schwacher Intensität nachgewiesen.

## Ergebnisse und Diskussion



**Abbildung 28:** Oben: *Extracted ion*-Chromatogramme des *N*-glycosylierten Peptids  $^{36}\text{VDKDLQSLEDILHQVENK}^{53}$  der  $\gamma$ -Untereinheit von humanem Fibrinogen in der mono- und disialylierten Glycoform (Probe N). Mitte: MS/MS-Spektrum der disialylierten Glycoform des obigen Peptids. Unten: MS/MS-Spektrum der monosialylierten Glycoform des obigen Peptids. Die Fragmentierung des komplexen *N*-Glycans ist gut erkennbar.

#### 4.1.4.4 Untersuchung der N-Glycosylierung von humanem Fibrinogen durch Verdau mit Pronase

Fibrinogen aus den Proben R, T, HCC017, HCC024 und ZIRO31 (Tabelle 15) wurde mit Pronase verdaut. Pronase ist eine Mischung verschiedener Proteasen, welche das Protein in Peptide von ca. 1-4 Aminosäuren schneidet. Auf diese Weise entstehen Glycopeptide mit einem kleinen Peptidanteil, welche sich gut zur Untersuchung der Glycanstruktur eignen. Die Analyse des Glycopeptidgemisches erfolgte mittels LC-MS/MS mit einer *porous graphitized carbon*-Säule. Leider arbeitet die Pronase nicht homogen, sodass Glycopeptide mit unterschiedlichem Peptidanteil entstehen. Außerdem enthält das isolierte Fibrinogen, wie in Abschnitt 4.1.4.2 gezeigt, größere Mengen anderer Glycoproteine, wie Fibronectin, Immunoglobuline und Serotransferrin. Daher konnte eine Vielzahl verschiedener Glycopeptide anhand ihrer MS/MS-Spektren identifiziert werden (Anhang, Tabelle 22).

Abbildung 29 zeigt die EICs der Glycopeptide GEN(H<sub>5</sub>N<sub>4</sub>S<sub>1</sub>)R und GEN(H<sub>5</sub>N<sub>4</sub>S<sub>1</sub>F)R, welche eindeutig der B $\beta$ -Untereinheit von humanem Fibrinogen zugeordnet werden können. Während das unfucosylierte Glycopeptid in allen Proben beobachtet werden kann, ist das fucosylierte Glycopeptid nur in den Proben HCC017 und HCC024 vorhanden. Dies bestätigt die Ergebnisse aus Abschnitt 4.1.3.2, in welchem die intakte B $\beta$ -Untereinheit untersucht wurde. In Abbildung 21 wurde gezeigt, dass die monosialylierte, fucosylierte Glycoform der B $\beta$ -Untereinheit in den Proben HCC017 und HCC024 vorhanden ist. Im Spektrum von Probe ZIRO31 waren intensive Signale der trunkierten Spezies zu erkennen, es war aber unklar, ob auch die fucosylierte Glycoform vorhanden ist. Die durch den Pronaseverdau erhaltenen Daten zeigen eindeutig, dass die fucosylierte Glycoform in Probe ZIRO31 nicht vorhanden ist. Die beträchtlichen Unterschiede der Retentionszeiten in den verschiedenen Proben sind auf eine Fehlfunktion der Säule zurückzuführen, beeinflussen das Ergebnis allerdings nicht, da alle Glycopeptide eindeutig anhand ihres m/z-Verhältnisses und MS/MS-Spektren identifiziert wurden. Abbildung 30 zeigt die MS/MS-Spektren der Glycopeptide GEN(H<sub>5</sub>N<sub>4</sub>S<sub>1</sub>)R und GEN(H<sub>5</sub>N<sub>4</sub>S<sub>1</sub>F)R. Das Vorhandensein der fucosylierten Glycoform konnte durch das MS/MS-Spektrum eindeutig nachgewiesen werden, die exakte Position der Fucose kann aber aufgrund der *Fucose migration*<sup>116</sup> nicht bestimmt werden. Die hohe Intensität des Fragmentions m/z = 512.19 (N<sub>1</sub>H<sub>1</sub>F<sub>1</sub>) ist aber ein Hinweis auf die Fucosylierung an der Antenne (Lewis-Strukturen). Die Fucose wird trotzdem stets als *Core-Fucose* dargestellt.

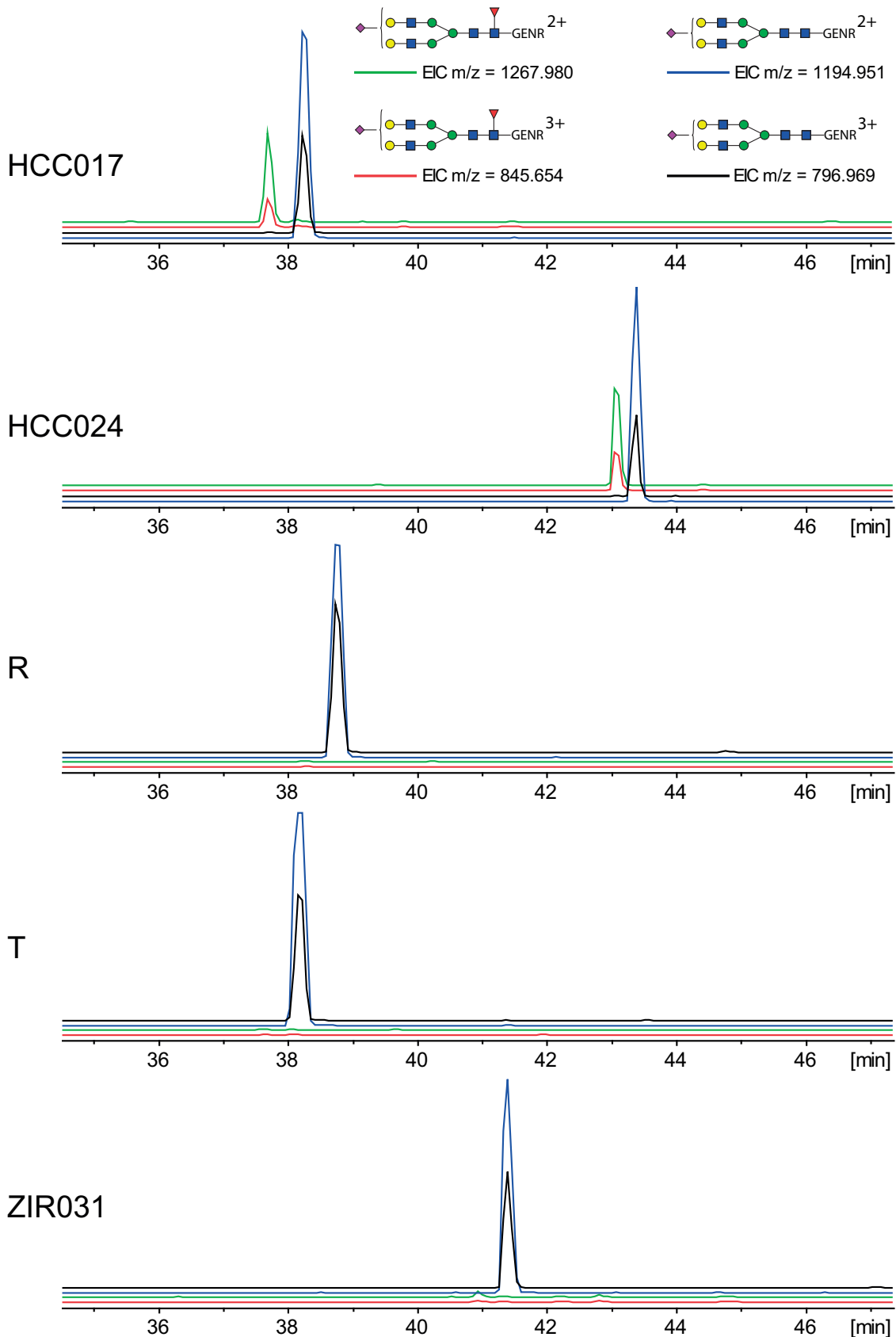
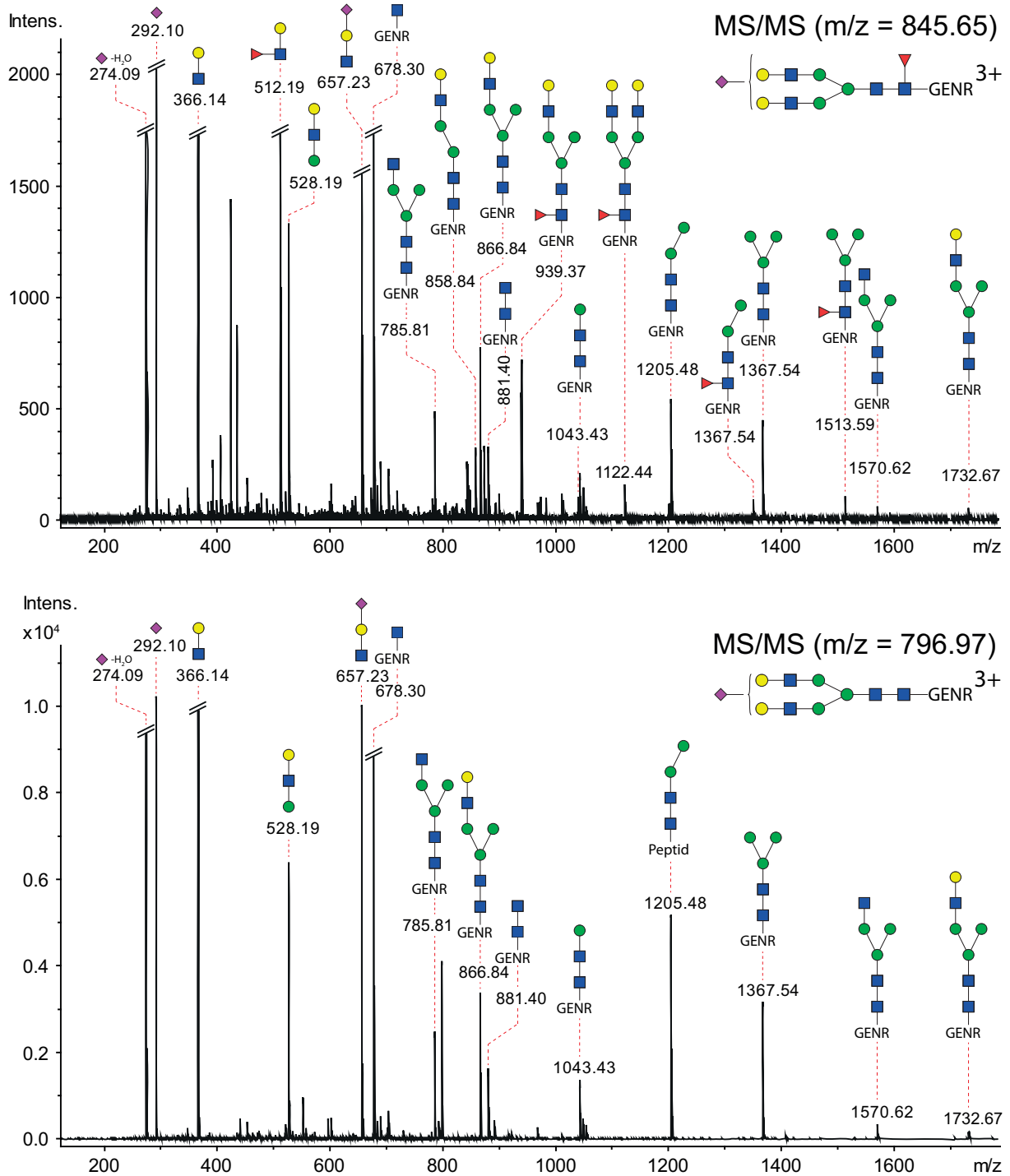


Abbildung 29: *Extracted ion*-Chromatogramme der Glycopeptide GEN(H<sub>5</sub>N<sub>4</sub>S<sub>1</sub>)R und GEN(H<sub>5</sub>N<sub>4</sub>S<sub>1</sub>F)R aus der B $\beta$ -Untereinheit von humanem Fibrinogen. Die Signale der nicht-fucosylierten Glycoformen (blau und schwarz) sind in allen Proben vorhanden. Die Signale der fucosylierten Glycoform (rot und grün) sind nur in den Proben HCC017 und HCC024 zu erkennen.

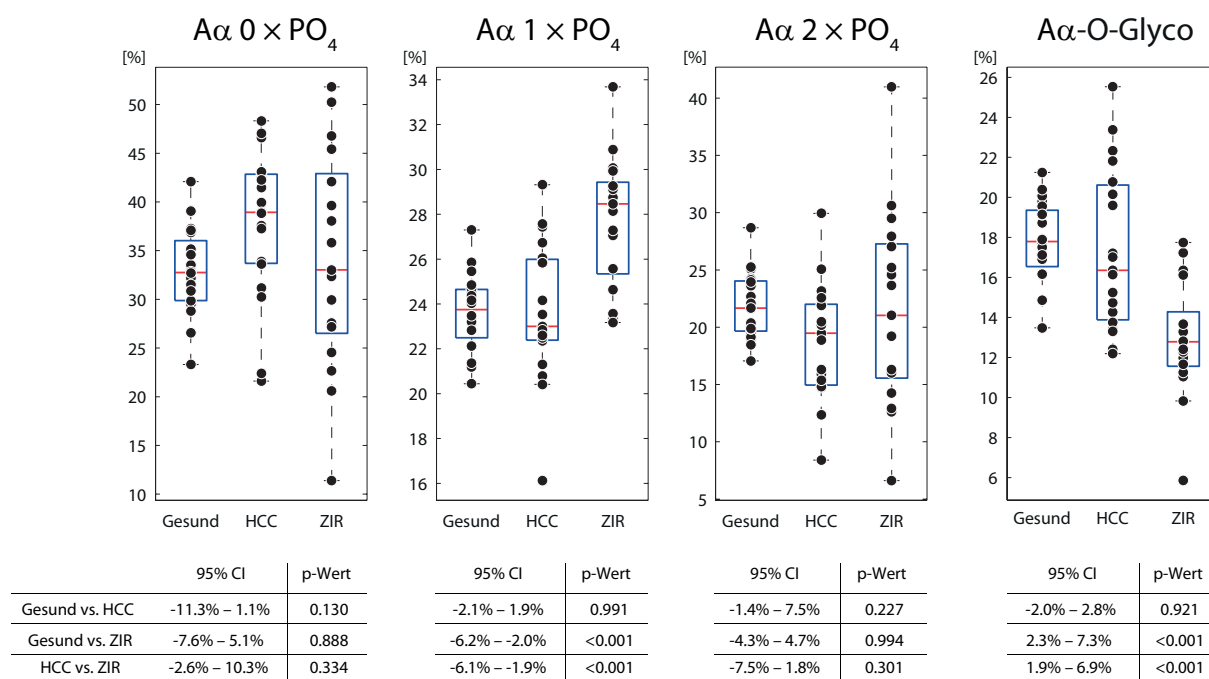


**Abbildung 30: MS/MS-Spektren der Glycopeptide GEN(H<sub>5</sub>N<sub>4</sub>S<sub>1</sub>)R (unten) und GEN(H<sub>5</sub>N<sub>4</sub>S<sub>1</sub>F)R (oben). Die Fragmentierung der komplexen N-Glycanstruktur ist in beiden Spektren gut erkennbar. Die Fucosylierung konnte anhand verschiedener Fragmentationen eindeutig nachgewiesen werden, die genaue Position der Fucose kann mittels MS/MS allerdings nicht bestimmt werden. Die Fucose wird trotzdem stets als Core-Fucose dargestellt.**



#### 4.1.5 Statistische Analyse der quantitativen Daten

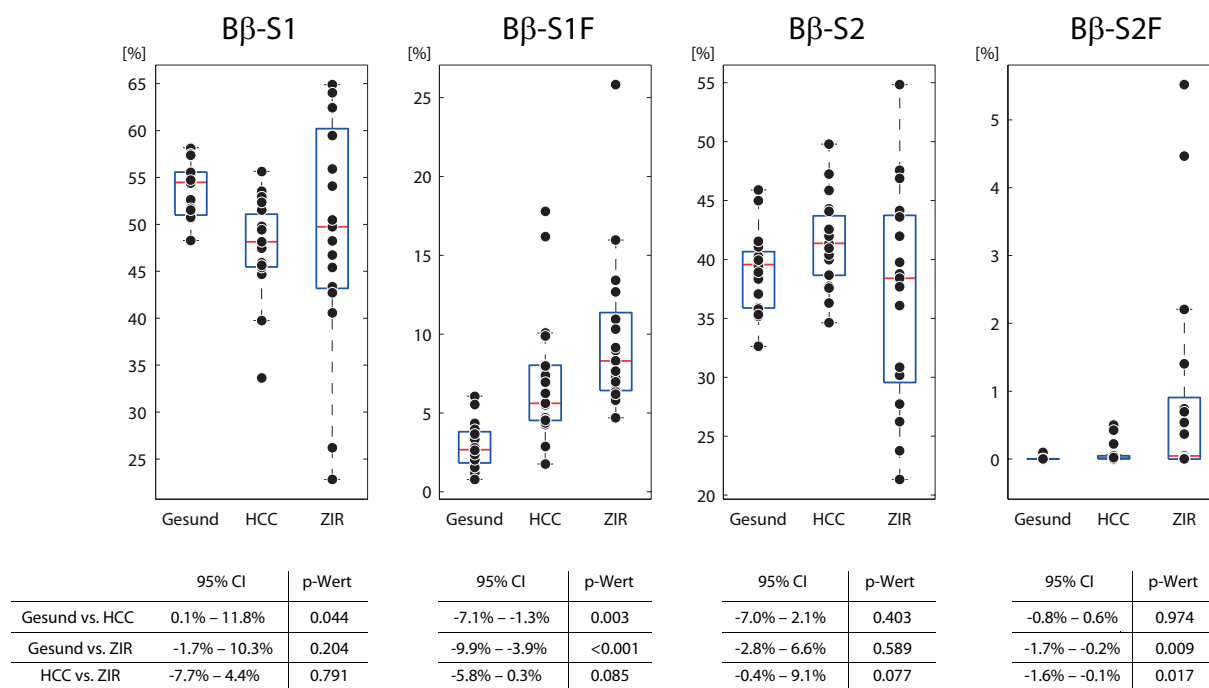
Die Unterschiede in der Glycosylierung und Phosphorylierung von humanem Fibrinogen zwischen gesunden Spendern, HCC- und Zirrhosepatienten wurden statistisch analysiert. Hierzu wurden die Anteile aller Glycoformen und Proteinspezies der Fibrinogen-untereinheiten durch Anpassung der theoretischen Isotopenmuster an die experimentellen, deconvolvierten Spektren bestimmt. (Das Verfahren ist in Abschnitt 3.2 genau erläutert und die vollständigen Ergebnisse sind im Anhang, Abschnitt 9.3, zu finden.) Um festzustellen, ob signifikante Unterschiede zwischen gesunden Spendern, HCC- und Zirrhosepatienten vorliegen, wurden einfaktorische Varianzanalysen für jede Glycoform bzw. Proteinspezies durchgeführt. (Die vollständigen Ergebnisse der Varianzanalysen sind im Anhang, Abschnitt 9.4, zu finden.)



**Abbildung 31: Boxplot-Diagramme der Anteile an unphosphorylierter, monophosphorylierter, diphosphorylierter und O-glycosylierter A $\alpha$ -Untereinheit von humanem Fibrinogen für gesunde Spender, HCC- und Zirrhosepatienten. Der Median ist als roter Strich, der Interquartilbereich als blaue Box dargestellt. Unter den Diagrammen sind die 95%-Konfidenzintervalle der Differenzen der Mittelwerte sowie die dazugehörigen p-Werte angegeben. Der Anteil der unphosphorylierten A $\alpha$ -Untereinheit ist bei HCC-Patienten im Vergleich zu gesunden Spendern erhöht (p-Wert = 0.130), während der Anteil der diphosphorylierten A $\alpha$ -Untereinheit verringert ist (p-Wert = 0.227). Der Anteil der monophosphorylierten A $\alpha$ -Untereinheit ist bei Zirrhosepatienten signifikant erhöht (p-Werte < 0.001). Der Anteil O-glycosylierter A $\alpha$ -Untereinheit ist bei Zirrhosepatienten im Vergleich zu gesunden Spendern und HCC-Patienten signifikant verringert (p-Werte < 0.001).**

In Abbildung 31-33 sind die Boxplot-Diagramme der wichtigsten Glycoformen bzw. Proteinspezies der A $\alpha$ -, B $\beta$ - und  $\gamma$ -Untereinheit von humanem Fibrinogen gezeigt. Die

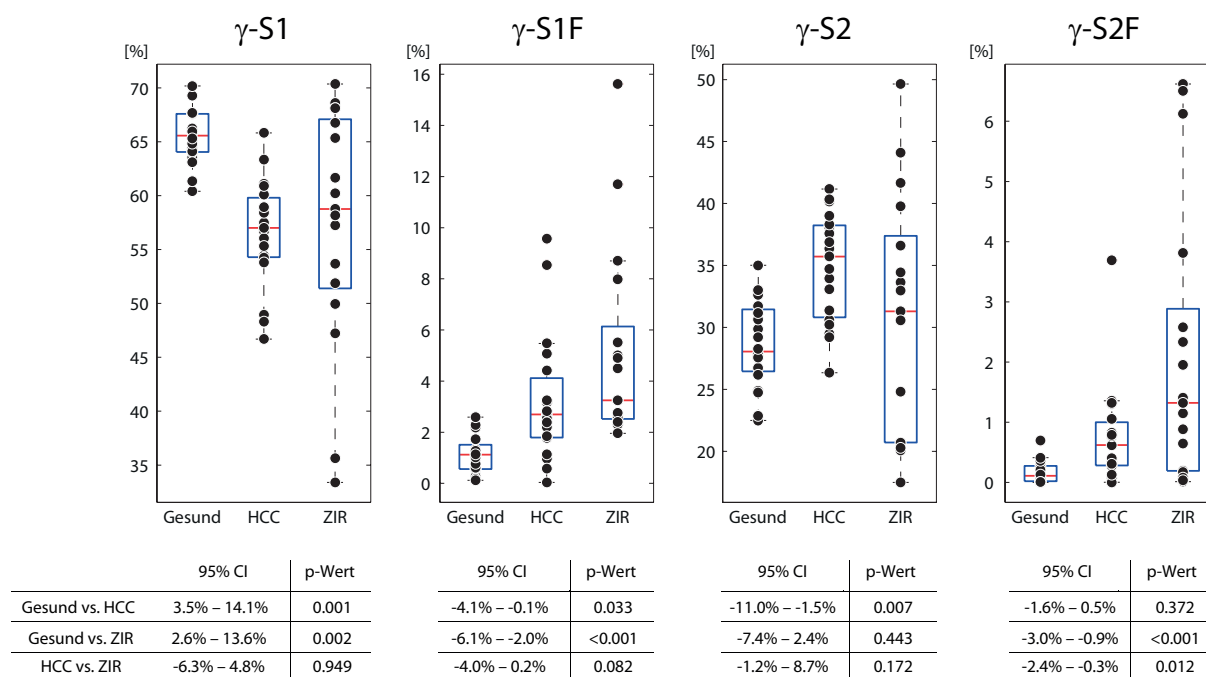
95%-Konfidenzintervalle der Differenzen der Mittelwerte sowie die dazugehörigen p-Werte sind unterhalb der Boxplots angegeben. Es sind einige Unterschiede zwischen gesunden Spendern, HCC- und Zirrhosepatienten zu erkennen. So ist der Anteil der unphosphorylierten A $\alpha$ -Untereinheit bei HCC-Patienten im Vergleich zu gesunden Spendern leicht erhöht (p-Wert = 0.130), während der Anteil der diphosphorylierten A $\alpha$ -Untereinheit leicht verringert ist (p-Wert = 0.227). Die Unterschiede sind zwar nicht signifikant, trotzdem deuten die Ergebnisse drauf hin, dass die A $\alpha$ -Untereinheit bei HCC-Patienten weniger phosphoryliert ist. Der Anteil der monophosphorylierten A $\alpha$ -Untereinheit ist bei Zirrhosepatienten interessanterweise signifikant erhöht (p-Wert < 0.001). Außerdem ist der Anteil der O-glycosylierten A $\alpha$ -Untereinheit bei Zirrhosepatienten im Vergleich zu gesunden Spendern und HCC-Patienten signifikant verringert (p-Werte < 0.001).



**Abbildung 32: Boxplot-Diagramme der Anteile der Glycoformen S1, S1F, S2 und S2F der B $\beta$ -Untereinheit von humanem Fibrinogen für gesunde Spender, HCC- und Zirrhosepatienten. Der Median ist als roter Strich, der Interquartilbereich als blaue Box dargestellt. Die 95%-Konfidenzintervalle der Differenzen der Mittelwerte sowie die dazugehörigen p-Werte sind unterhalb der Boxplots angegeben. Der Anteil der Glycoform B $\beta$ -S1 ist bei HCC-Patienten im Vergleich zu gesunden Spendern verringert (p-Wert = 0.044). Außerdem ist zu erkennen, dass der Anteil der fucosylierten Glycoformen B $\beta$ -S1F und B $\beta$ -S2F bei HCC- und Zirrhosepatienten erhöht ist.**

Die Boxplot-Diagramme der Glycoformen der B $\beta$ - und  $\gamma$ -Untereinheit zeigen, dass der Sialylierungsgrad bei HCC-Patienten erhöht ist. Der Anteil der Glycoform  $\gamma$ -S1 ist bei HCC-Patienten im Vergleich zu gesunden Spender verringert (p-Wert = 0.001), während der

Anteil der Glycoform  $\gamma$ -S2 erhöht ist (p-Wert = 0.007). Bei Zirrhosepatienten weisen die Daten eine große Streuung auf, sodass keine Aussagen getroffen werden können.



**Abbildung 33: Boxplot-Diagramme der Anteile der Glycoformen S1, S1F, S2 und S2F der  $\gamma$ -Untereinheit von humanem Fibrinogen für gesunde Spender, HCC- und Zirrhosepatienten. Der Median ist als roter Strich, der Interquartilbereich als blaue Box dargestellt. Die 95%-Konfidenzintervalle der Differenzen der Mittelwerte sowie die dazugehörigen p-Werte sind unterhalb der Boxplots angegeben. Es ist zu erkennen, dass der Anteil der Glycoform  $\gamma$ -S1 bei HCC-Patienten im Vergleich zu gesunden Spendern verringert ist (p-Wert = 0.001), während der Anteil der Glycoform  $\gamma$ -S2 erhöht ist (p-Wert = 0.007). Außerdem sind die Anteile der fucosylierten Glycoformen  $\gamma$ -S1F und  $\gamma$ -S2F bei HCC- und Zirrhosepatienten im Vergleich zu gesunden Spendern erhöht.**

Außerdem ist ein klarer Trend bei der Fucosylierung der  $\beta$ - und  $\gamma$ -Untereinheit von humanem Fibrinogen zu erkennen. Die Anteile der Glycoformen S1F und S2F sind bei HCC- und Zirrhosepatienten im Vergleich zu gesunden Spendern deutlich erhöht. Außerdem zeigt sich, dass Zirrhosepatienten einen noch höheren Grad an Fucosylierung besitzen als HCC-Patienten.

Die Boxplot-Diagramme und Ergebnisse der univariaten Varianzanalysen aller quantifizierten Glycoformen und Proteinspezies sind im Anhang zu finden (Abschnitt 9.4). Diese zeigen unter anderem, dass die Anteile der oxidierten Fibrinogenuntereinheiten sowie der trunkierten Spezies der  $\alpha$ - und  $\beta$ -Untereinheit bei HCC- und Zirrhosepatienten im Vergleich zu gesunden Spendern erhöht sind. Eine Erklärung hierfür könnte sein, dass die fortschreitende Leberschädigung oxidative und proteolytische Prozesse fördert.

Die statistische Analyse der quantitativen Daten zeigte, dass es Unterschiede in den posttranslationalen Modifikationen von humanem Fibrinogen zwischen gesunden Spendern, HCC- und Zirrhosepatienten gibt. Dabei ist zu beachten, dass die meisten der HCC-Patienten auch eine Leberzirrhose aufweisen, sodass teilweise unklar bleibt, ob die Unterschiede durch den Krebs oder die Zirrhose verursacht werden. Die Unterscheidung zwischen krebsfreien Zirrhosepatienten und denen, die bereits ein hepatozelluläres Karzinom entwickelt haben, ist aber von großem Interesse.

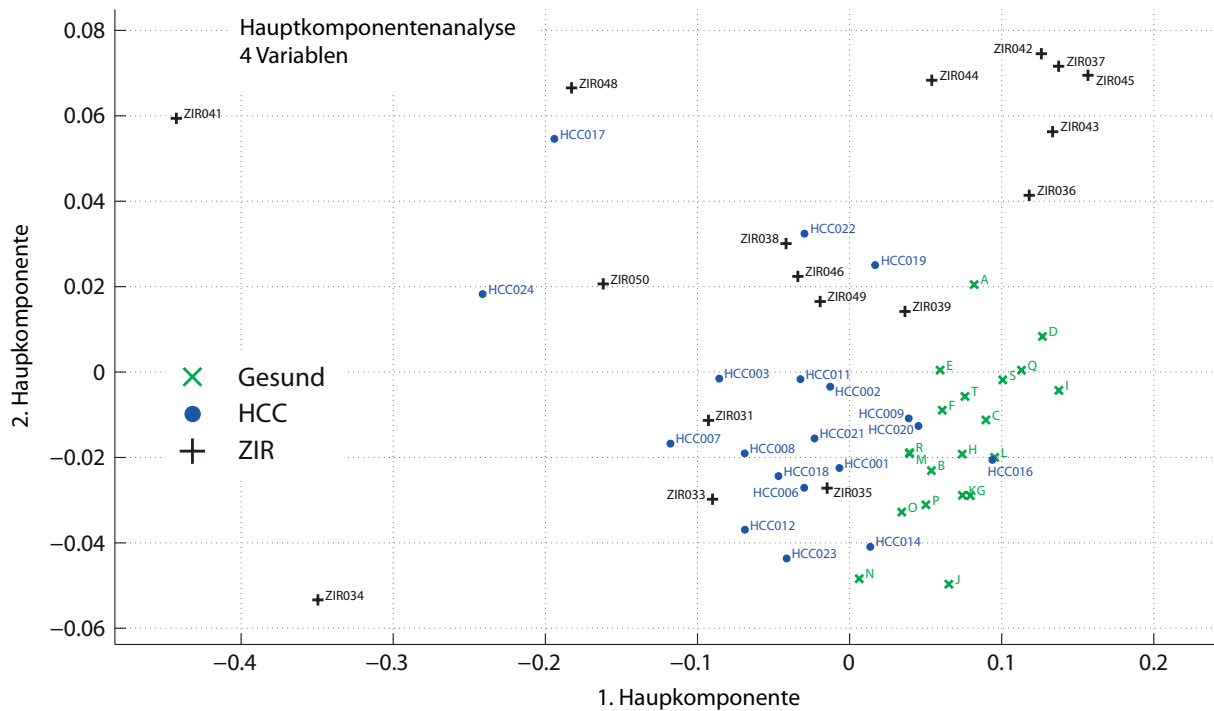
Die Zuordnung einer unbekannt Probe zu einer Gruppe ist allerdings nicht durch die Betrachtung von jeweils nur einer Glycoform oder Proteinspezies möglich, sondern muss durch Kombination mehrerer Glycoformen oder Proteinspezies geschehen. Hierfür stehen verschiedene multivariate statistische Methoden zur Verfügung.

### 4.1.5.1 Hauptkomponentenanalyse

Die Hauptkomponentenanalyse dient der Vereinfachung eines multivariaten Datensatzes. Dabei wird eine Hauptachsentransformation durchgeführt, welche die größte Varianz der Daten auf der ersten Hauptkomponente abbildet.

Eine entscheidende Frage in dieser Arbeit war, welche der quantifizierten Anteile der Glycoformen und Proteinspezies der Fibrinogenuntereinheiten als abhängige Variablen der Hauptkomponentenanalyse und MANOVA (Abschnitt 4.1.5.2) verwendet werden sollten. Eine vollständige Diskussion dieser Frage ist im Anhang (Abschnitt 9.5.1) zu finden. Im Folgenden werden stets die vier Glycoformen  $B\beta/\gamma$ -S1 und  $B\beta/\gamma$ -S1F als abhängige Variablen verwendet, welche sich als bestes Motiv für die multivariaten Methoden erweisen haben.

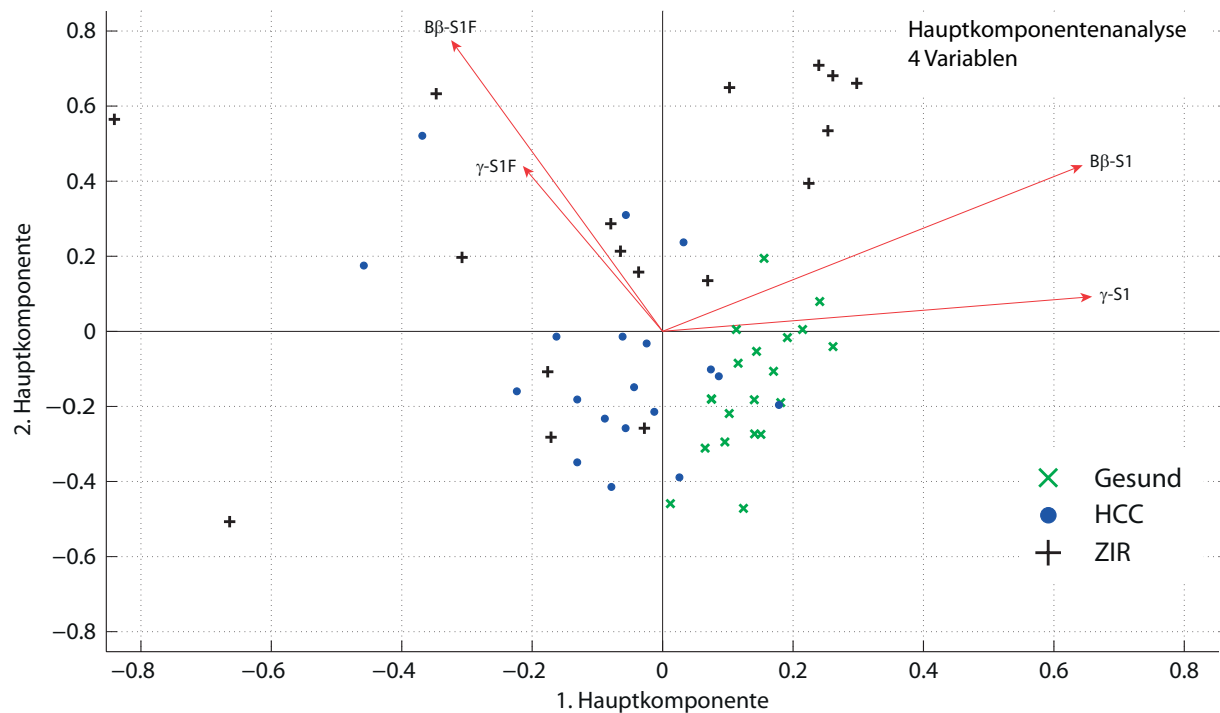
Das Ergebnis der Hauptkomponentenanalyse mit vier Variablen ist in Abbildung 34 dargestellt, die erste und zweite Hauptkomponente sind gegeneinander aufgetragen. Die Datenpunkte der gesunden Spender (grüne Kreuze) liegen in einem kleinen Bereich im unteren rechten Teil des Diagrammes, während sich die meisten Datenpunkte von HCC- und Zirrhosepatienten außerhalb dieses Bereiches befinden. Eine deutliche Trennung zwischen HCC- und Zirrhosepatienten ist nicht zu erkennen. Die Datenpunkte der HCC-Patienten (blaue Punkte) liegen fast alle in einem Bereich links von den Datenpunkten der gesunden Spender, während die Datenpunkte der Zirrhosepatienten über einem größeren Bereich verteilt liegen.



**Abbildung 34:** Auftragung der ersten und zweiten Hauptkomponente aus der Hauptkomponentenanalyse mit vier Variablen für alle 56 Proben. Die Datenpunkte gesunder Spender sind als grüne Kreuze dargestellt, die Datenpunkte von HCC-Patienten als blaue Punkte und die Datenpunkte von Zirrhosepatienten als schwarze Kreuze. Es ist deutlich zu erkennen, dass alle Datenpunkte gesunder Spender in einem kleinen Bereich im unteren linken Teil des Diagramms liegen. Die meisten Datenpunkte von HCC- und Zirrhosepatienten liegen außerhalb dieses Bereiches. Zwischen HCC- und Zirrhosepatienten ist keine klare Trennung zu erkennen. Die Datenpunkte der HCC-Patienten liegen fast alle in einem Bereich links von den Datenpunkten gesunder Spender, während die Datenpunkte der Zirrhosepatienten über den gesamten Bereich des Diagramms verteilt liegen.

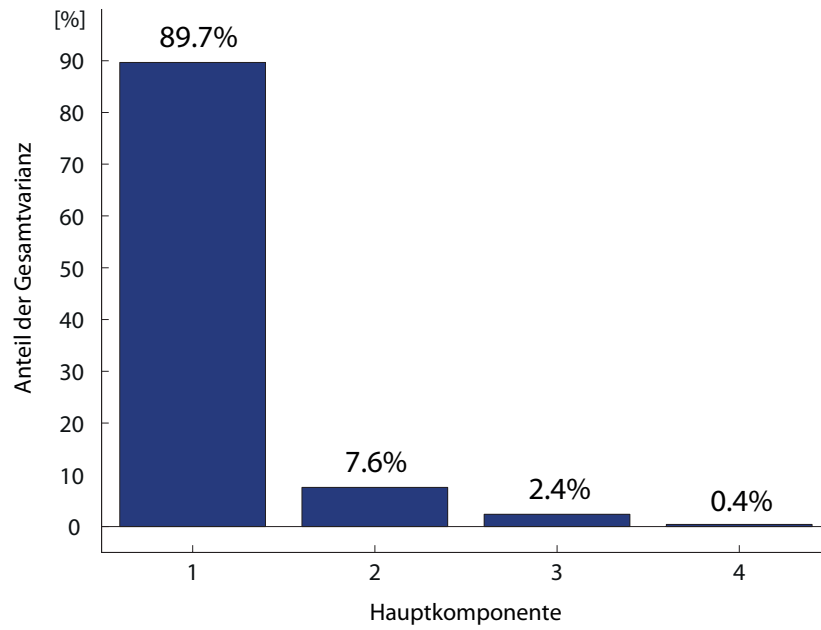
In Abbildung 35 sind neben den Datenpunkten aller Proben auch die Vektoren gezeigt, welche die Koeffizienten der einzelnen Variablen in der Hauptkomponentenanalyse darstellen. Die Richtung der einzelnen Vektoren gibt Auskunft über den Zusammenhang der jeweiligen Glycoform bzw. Proteinspezies mit HCC oder Zirrhose.

Die Vektoren der Glycoformen  $\beta$ -S1 und  $\gamma$ -S1 zeigen nach rechts in Richtung der Datenpunkte der gesunden Spender. Dies steht in Übereinstimmung mit dem Ergebnis, dass der Anteil der monosialylierten Glycoformen bei HCC-Patienten niedriger als bei gesunden Spendern ist. Die Vektoren der fucosylierten Glycoformen zeigen nach oben links, zu den Datenpunkte von HCC- und Zirrhosepatienten, da der Anteil der fucosylierten Glycoformen bei HCC- und Zirrhosepatienten erhöht ist.



**Abbildung 35: Darstellung der PCA-Koeffizienten aller vier Variablen als Vektoren. Jeder in Rot dargestellte Vektor zeigt die Koeffizienten der ersten und zweiten Hauptkomponente der Hauptkomponentenanalyse für eine Variable. Es ist zu beachten, dass sich die Achsenbeschriftung nicht auf die Datenpunkte, welche als grüne Kreuze (gesunde Spender), blaue Punkte (HCC-Patienten) und schwarze Kreuze (Zirrhosepatienten) dargestellt sind, sondern auf die Vektoren bezieht. Die Datenpunkte sind so skaliert, dass sie in den dargestellten Bereich passen.**

Die orthogonalen Hauptkomponenten stellen absteigend eine immer geringere Varianz dar. In Abbildung 36 sind die Anteile jeder Hauptkomponente der Hauptkomponentenanalyse mit vier Variablen an der Gesamtvarianz des Datensatzes angegeben. Demnach decken die ersten beiden Hauptkomponenten 97% der Gesamtvarianz ab. Die Hauptkomponentenanalyse verwendet keinerlei Information über die zu trennenden Gruppen des Datensatzes (in diesem Fall gesunde Spender, HCC- und Zirrhosepatienten). Eine Trennung zwischen den einzelnen Gruppen kann daher auch in den höheren Hauptkomponenten vorliegen. Bei der vorliegenden Problemstellung sind multivariate statistische Methoden, welche die Gruppen berücksichtigen, daher vorzuziehen.



**Abbildung 36: Anteile aller Hauptkomponenten der Hauptkomponentenanalyse mit vier Variablen an der Gesamtvarianz. Die ersten beiden Hauptkomponenten beinhalten 97% der Gesamtvarianz.**

#### 4.1.5.2 MANOVA

Die multivariate Varianzanalyse (MANOVA: *multivariate analysis of variance*) dient der Untersuchung eines multivariaten Datensatzes mit zwei oder mehr Gruppen. Die Ergebnisse der MANOVA sind denen der Hauptkomponentenanalyse ähnlich. Neben der in Tabelle 8 gezeigten MANOVA-Tabelle generiert die MANOVA eine Reihe von kanonischen Variablen (CVs), welche mit den Hauptkomponenten der PCA verglichen werden können. Die kanonischen Variablen werden, wie die Hauptkomponenten, durch Linearkombination aus den abhängigen Variablen gebildet. Die Koeffizienten sind dabei so gewählt, dass die erste kanonische Variable die beste Trennung zwischen den Gruppen aufweist (die zweite kanonische Variable die zweitbeste Trennung zwischen den Gruppen usw.).

In Tabelle 8 ist die Standard MANOVA-Tabelle der MANOVA mit vier Variablen gezeigt, die verschiedenen Tests auf Unterschiede zwischen den Gruppen fallen alle signifikant aus (p-Werte < 0.001).

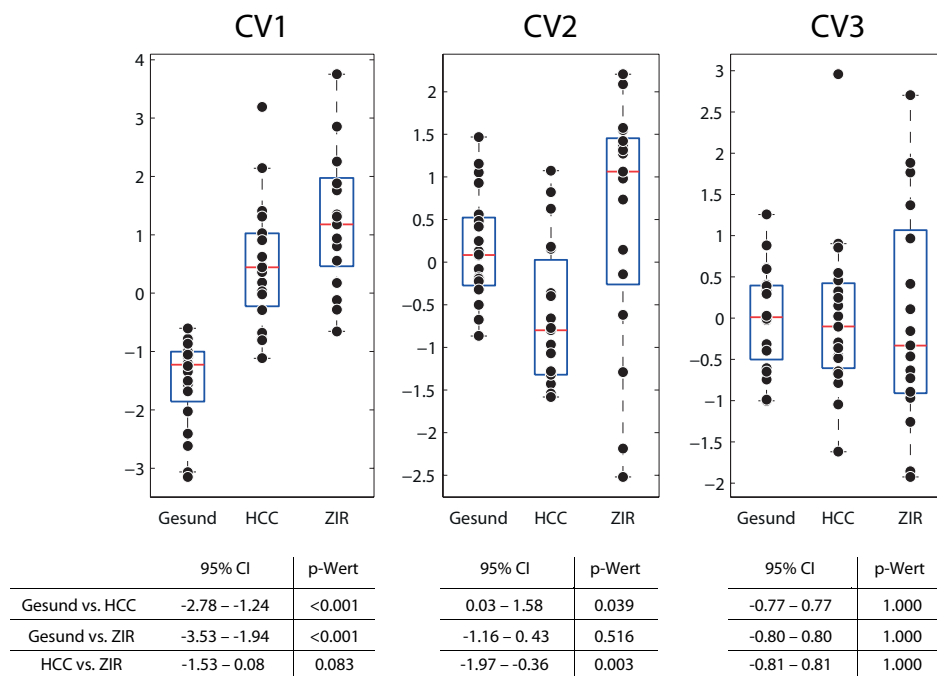
**Tabelle 8: Ergebnisse der MANOVA mit vier Variablen. Die Pillai-Spur Wilks-Lambda, die Hotelling-Spur und Roys größte charakteristische Wurzel sind verschiedene Arten die Signifikanz der Trennung zwischen den Gruppen zu bestimmen. In diesem Fall sind alle Werte klar signifikant (p-Wert < 0.001).**

	Wert	F	Hypothesen df	Fehler df	p-Wert	Effektgröße
Pillai-Spur	0.787	8.278	8	102	<0.001	0.397
Wilks-Lambda	0.329	9.278	8	100	<0.001	0.426
Hotelling-Spur	1.681	10.297	8	98	<0.001	0.457
Größte charakteristische Wurzel nach Roy	1.434	18.283	4	51	<0.001	0.589

Die Werte der ersten drei kanonischen Variablen der MANOVA mit vier Variablen sind in Abbildung 37 als Boxplot-Diagramme für gesunde Spender, HCC- und Zirrhosepatienten gezeigt. Die kanonischen Variablen wurden univariaten Varianzanalysen unterzogen, die 95% Konfidenzintervalle der Differenzen der Mittelwerte sowie die dazugehörigen p-Werte der Folgetests sind unter den Boxplots angegeben. In Abbildung 37 ist zu erkennen, dass HCC- und Zirrhosepatienten in der ersten kanonischen Variablen signifikant von gesunden Spendern getrennt sind (p-Werte < 0.001). In der zweiten kanonischen Variablen tritt eine Trennung zwischen gesunden Spendern und HCC-Patienten (p-Wert = 0.039) bzw. zwischen HCC- und Zirrhosepatienten (p-Wert = 0.003) auf. In der dritten kanonischen Variablen zeigen sich keine Unterschiede zwischen den Gruppen (p-Werte = 1.000). Die Varianz

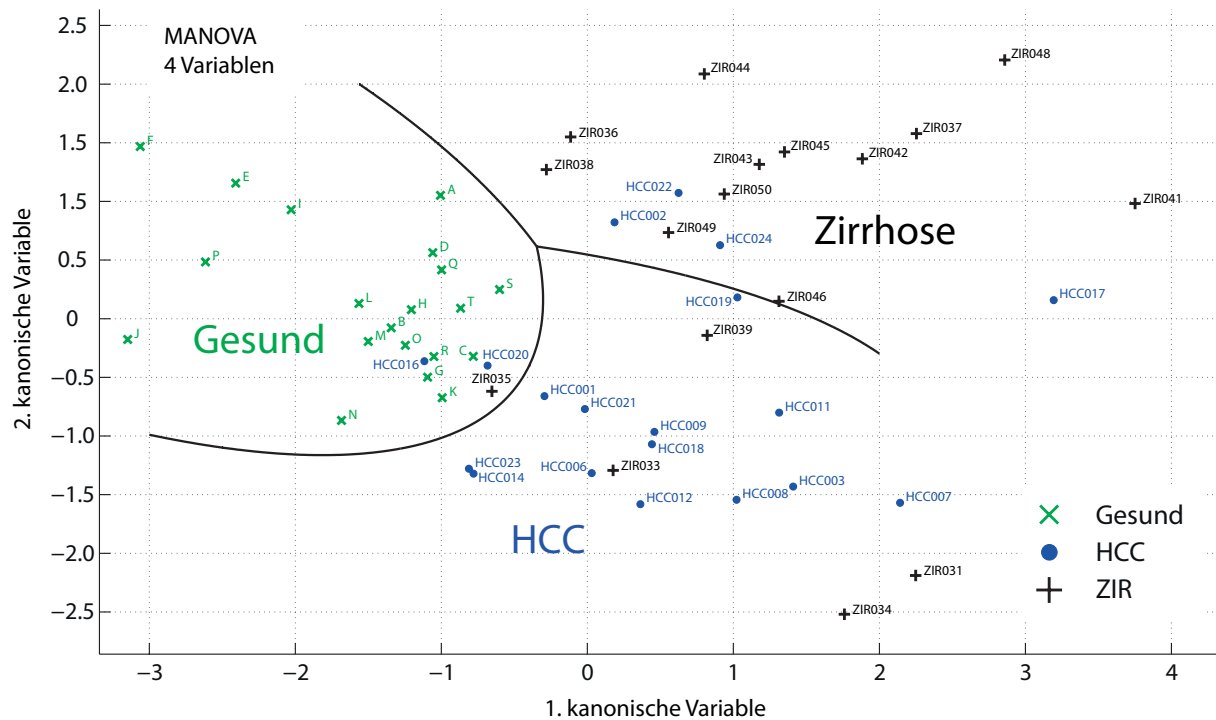


zwischen den Gruppen wird demnach vollständig durch die ersten beiden kanonischen Variablen abgebildet, die dritte und vierte kanonische Variable brauchen daher nicht betrachtet werden.



**Abbildung 37: Boxplot-Diagramme der ersten drei kanonischen Variablen (CV1-3) der MANOVA mit vier Variablen für gesunde Spender, HCC- und Zirrhosepatienten. Der Median ist als rote Linie, der Interquartilbereich als blaue Box dargestellt. Die 95% Konfidenzintervalle der Differenzen der Mittelwerte, sowie die dazugehörigen p-Werte sind unterhalb der Boxplots angegeben. In der ersten kanonischen Variable weisen die HCC- und Zirrhosepatienten einen signifikanten Unterschied zu gesunden Spendern auf (p-Werte < 0.001). In der zweiten kanonischen Variable tritt eine Trennung zwischen HCC-Patienten und gesunden Spendern (p-Wert = 0.039) sowie zwischen HCC- und Zirrhosepatienten (p-Wert = 0.003) auf. In der dritten kanonischen Variable sind keine Unterschiede zwischen den Gruppen zu erkennen (p-Werte = 1.000).**

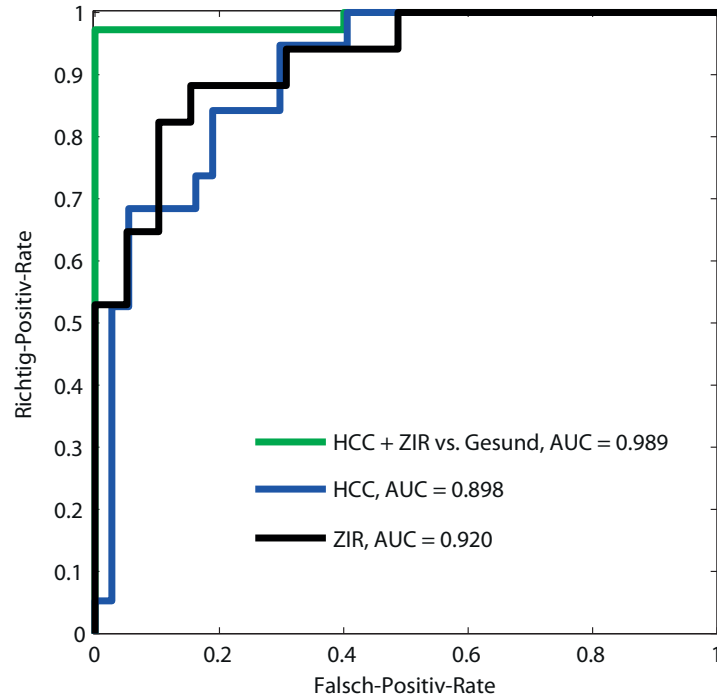
In Abbildung 38 sind die ersten beiden kanonischen Variablen aller Proben gegeneinander aufgetragen. Es ist eine ausgezeichnete Trennung zwischen gesunden Spendern und HCC- bzw. Zirrhosepatienten zu erkennen. Die eingezeichneten Grenzlinien wurden mittels Diskriminanzanalyse berechnet. Hierzu wurden die ersten beiden kanonischen Variablen aller 56 Proben für die Diskriminanzanalyse verwendet. Alle gesunden Spender sind korrekt klassifiziert. Zwei Datenpunkte von HCC-Patienten (HCC016 und HCC020) sowie ein Datenpunkte eines Zirrhosepatienten (ZIR035) liegen im Bereich der gesunden Spender und sind somit falsch negativ klassifiziert. Die Trennung zwischen HCC- und Zirrhosepatienten ist weniger deutlich. Vier Datenpunkte von HCC-Patienten liegen im Bereich der Zirrhosepatienten und vier Datenpunkte von Zirrhosepatienten liegen im Bereich der HCC-Patienten.



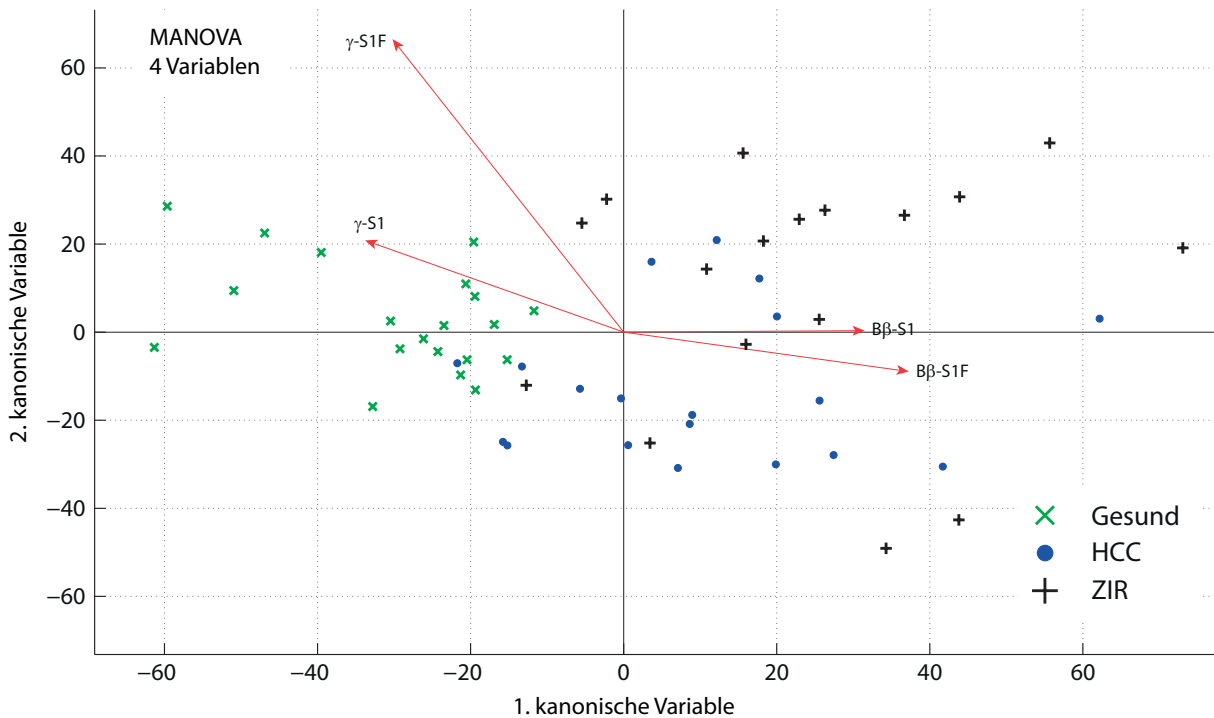
**Abbildung 38: Ergebnisse der MANOVA mit vier Variablen. Die ersten beiden kanonischen Variablen jeder Probe sind gegeneinander aufgetragen. Grüne Kreuze: gesunde Spender, blaue Punkte: HCC-Patienten, schwarze Kreuze: Zirrhosepatienten. Es ist eine ausgezeichnete Trennung zwischen den Datenpunkten von gesunden Spendern, HCC-Patienten und Zirrhosepatienten zu erkennen.**

Anhand der *a posteriori*-Wahrscheinlichkeiten der Diskriminanzanalyse können ROC-Kurven (*receiver operating characteristic*) berechnet werden. Die in Abbildung 39 gezeigte ROC-Kurve für die Klassifizierung der Proben von HCC- und Zirrhosepatienten gegen gesunde Spender (grün) weist einen fast perfekten *area-under-the-curve*-Wert von 0.989 auf. Die *area-under-the-curve*-Werte für die Klassifizierung von HCC- und Zirrhosepatienten betragen 0.898 bzw. 0.920.

In Abbildung 40 sind die Koeffizienten der MANOVA mit vier Variablen als Vektoren dargestellt. Der Vektor der Variable  $\gamma$ -S1 zeigt in Richtung der Datenpunkte von gesunden Spendern, da der Anteil der Glycoform  $\gamma$ -S1 bei diesen höher war als bei HCC-Patienten. Der Vektor der Variable  $\beta$ -S1F zeigt erwartungsgemäß in Richtung der HCC- und Zirrhosepatienten. Die Vektoren der Variablen  $\gamma$ -S1F und  $\beta$ -S1 erscheinen allerdings unlogisch, da der Vektor der Variablen  $\gamma$ -S1F in Richtung der Datenpunkte gesunder Spender bzw. der Vektor der Variable  $\beta$ -S1 in Richtung der Datenpunkte von HCC- und Zirrhosepatienten zeigt. Nichtsdestotrotz liefern diese Koeffizienten die beste Trennung zwischen gesunden Spendern, HCC- und Zirrhosepatienten und stellen somit das beste Modell zur Klassifizierung einer unbekannter Proben dar.



**Abbildung 39: Receiver operating characteristic-Kurve für die Klassifizierung der Proben von gesunden Spendern (grün), HCC-Patienten (blau) und Zirrhosepatienten (schwarz) basierend auf den *a posteriori* Wahrscheinlichkeiten der Diskriminanzanalyse, welche die ersten beiden kanonischen Variablen der MANOVA mit vier Variablen verwendet. Die *area-under-the-curve*-Werte betragen 0.989 (Gesund), 0.898 (HCC) und 0.920 (Zirrhose).**



**Abbildung 40: Darstellung der Koeffizienten aller vier Variablen der MANOVA als Vektoren. Die Koeffizienten der ersten beiden kanonischen Variablen jeder Variable sind als Vektoren dargestellt. Zusätzlich sind die Datenpunkte aller Proben gezeigt. Dabei ist zu beachten, dass sich die Achsenbeschriftung nicht auf die Datenpunkte der Proben, sondern auf die Koeffizienten bezieht.**

**Tabelle 9: Koeffizienten der MANOVA mit vier Variablen, welche zur Berechnung der ersten und zweiten kanonischen Variablen verwendet werden. Außerdem sind die Mittelwerte der vier Proteinspezies aus den 56 verwendeten Proben angegeben. Zur Berechnung der ersten und zweiten kanonischen Variablen werden die jeweiligen Mittelwerte von den Anteilen der Proteinspezies abgezogen. Die normierten Werte werden mit den Koeffizienten multipliziert und summiert, um die kanonische Variable zu ergeben (Gleichung 12).**

Abhängige Variable	Koeffizienten der ersten kanonischen Variablen [ <i>COEFF1</i> ]	Koeffizienten der zweiten kanonischen Variablen [ <i>COEFF1</i> ]	Mittelwert der Proteinspezies aus den 56 Proben [ <i>MEAN</i> ]
B $\beta$ -S1	31.51	0.34	0.504
B $\beta$ -S1F	37.25	-8.93	0.064
$\gamma$ -S1	-33.80	20.85	0.600
$\gamma$ -S1F	-30.21	66.53	0.031

Die kanonischen Variablen einer unbekannt Probe werden nach Gleichung 12 berechnet. Die hierzu verwendeten Koeffizienten (vgl. Abbildung 40) sind in Tabelle 9 gezeigt. Außerdem werden die Mittelwerte der Anteile jeder Variable aus den 56 Proben benötigt.

$$CV_x = \sum_{i=1}^4 COEFFx_i \times (OBSERVED_i - MEAN_i) \quad \text{Gleichung 12}$$

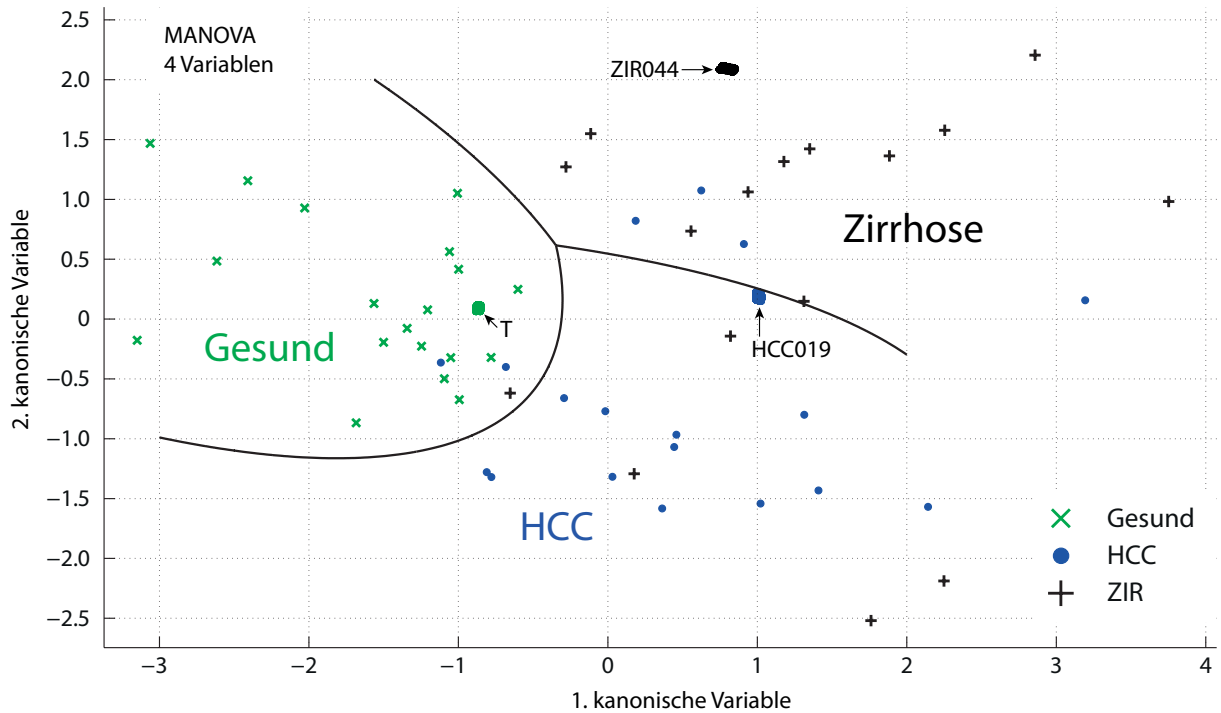
Die jeweiligen Mittelwerte (*MEAN*) werden von den beobachteten Werten (*OBSERVED*) einer Probe abgezogen. Die normierten Werte werden dann mit den Koeffizienten (*COEFFx*) multipliziert und summiert.

#### 4.1.5.3 Einfluss verschiedener Parameter auf die Lage der Datenpunkte im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA

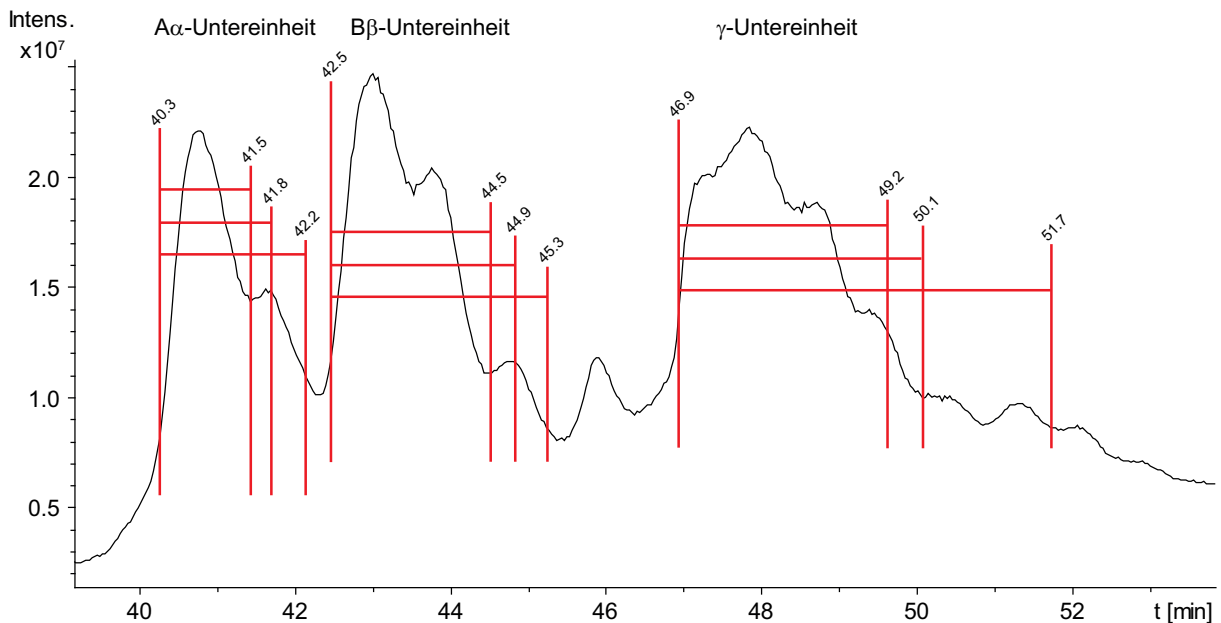
Die Lage eines Datenpunktes im Plot der ersten und zweiten kanonischen Variablen der MANOVA wird, wie jeder Messwert, einer gewissen Streuung unterliegen. Um das Ausmaß dieser Streuung abzuschätzen, wurde der Einfluss verschiedener Parameter untersucht. Wie bereits in Abschnitt 3.2 beschrieben, wird bei der Quantifizierung der Glycoformen und Proteinspezies der Fibrinogenuntereinheiten ein *shift* angewendet, um experimentelle Ungenauigkeiten der Massenspektrometrie auszugleichen. Dabei wird zu den x-Werten (in Dalton) des experimentellen Massenspektrums ein konstanter Wert, welcher als *shift* bezeichnet wird, addiert. Der *shift* wurde bestimmt, indem die Anpassung mit Werten im Bereich  $\pm 5.00$  Da und einem Abstand von  $0.01$  Da durchgeführt wurde. Diejenige Anpassung, welche die geringste Abweichung zwischen experimentellem und berechnetem Spektrum lieferte, wurde für die Quantifizierung herangezogen.

Es wurde untersucht, welche Auswirkung der *shift* auf die Lage eines Datenpunktes im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA hat. Hierzu wurden die drei Proben T, HCC019 und ZIRO44 ausgewählt. Die Spektren der B $\beta$ - und  $\gamma$ -Untereinheit der drei Proben wurden mit verschiedenen Werten für den *shift* quantifiziert. Dabei wurden Werte von  $\pm 0.5$  Da und  $\pm 1.0$  Da ausgehend vom optimalen *shift* verwendet, sodass für jede Untereinheit fünf verschiedene Quantifizierungen erhalten wurden. Durch Kombination der jeweils fünf Quantifizierungen der B $\beta$ - und  $\gamma$ -Untereinheit entstehen für jede Probe 25 Datensätze, welche in das Diagramm der ersten und zweiten kanonischen Variablen der MANOVA eingefügt werden können. Abbildung 41 zeigt die durch Variation des *shift* erhaltenen 25 Datenpunkte der Proben T, HCC019 und ZIRO44 im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA mit vier Variablen. Es ist nur eine sehr kleine Streuung, von kleiner als  $\pm 0.1$  in den ersten beiden kanonischen Variablen, zu erkennen.

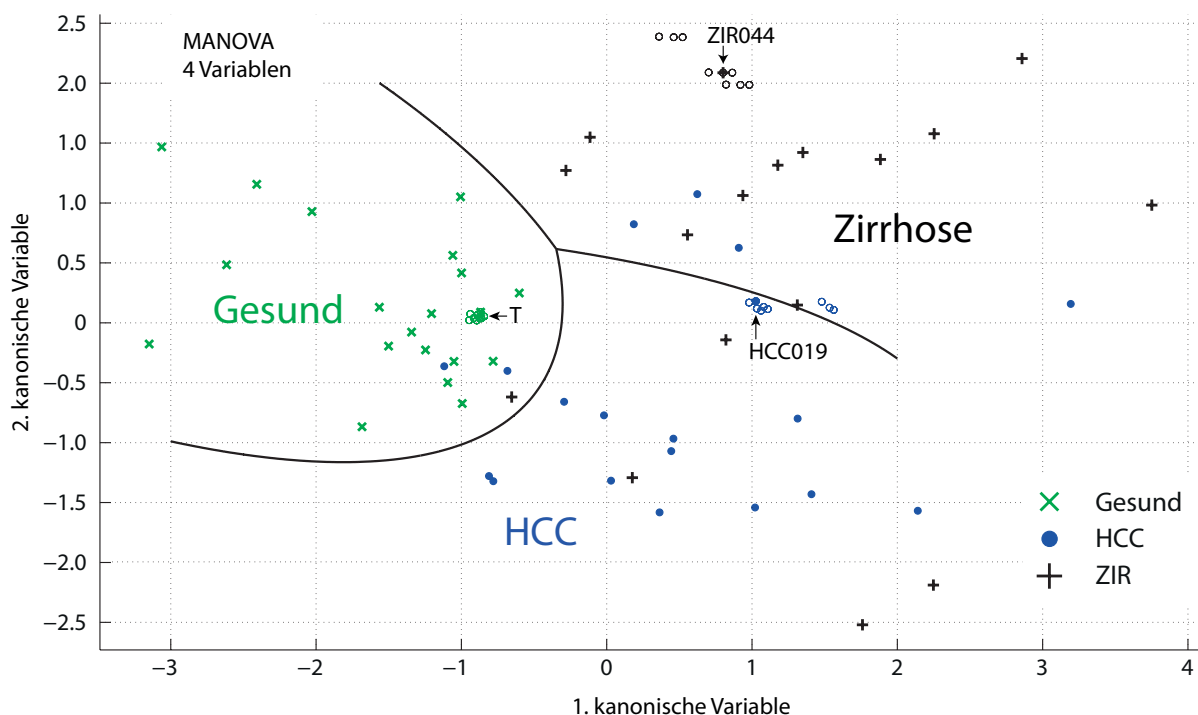
Ein weiterer Faktor der sich auf die Streuung der Datenpunkte im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA auswirkt, ist der chromatographische Bereich über welchen die Massenspektren vor der Deconvolution gemittelt werden. Dieser wurde manuell für jede Probe gewählt und unterliegt daher einer gewissen Unsicherheit. In Abschnitt 4.1.2 wurde gezeigt, dass die verschiedenen Glycoformen und Proteinspezies der Fibrinogenuntereinheiten leicht unterschiedliche Retentionszeiten aufweisen. Aus diesem Grund wirkt sich der Bereich in welchem die Massenspektren gemittelt werden auf das Ergebnis der Quantifizierung aus.



**Abbildung 41:** Streuung der durch Variation des *shift* generierten Datenpunkte der Proben T, HCC019 und ZIR044 im Diagramm der ersten und zweiten kanonischen Variablen aus der MANOVA mit vier Variablen. Grüne Kreise: T, blaue Kreise: HCC019, schwarze Kreise: ZIR044. Die beobachtete Streuung der jeweils 25 Datenpunkte fällt sehr gering aus.



**Abbildung 42:** Bereiche des Chromatogramms von Probe T, in denen die Massenspektren gemittelt wurden, um die Streuung in Diagramm der ersten und zweiten kanonischen Variable der MANOVA zu untersuchen. Der Start des Elutionsbereiches ist gut zu erkennen, daher wurde nur der Endpunkt variiert.



**Abbildung 43:** Die durch Variation der in den Chromatogrammen gemittelten Bereiche erhaltenen jeweils neun Datenpunkte der Proben T, HCC019 und ZIRO44 im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA mit vier Variablen.

Es wurde untersucht, wie sich die Wahl des gemittelten Bereiches im Chromatogramm auf die Lage der Datenpunkte im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA auswirkt. Hierzu wurden für die Proben T, HCC019 und ZIRO44 jeweils drei verschiedene Bereiche der  $A\alpha$ -,  $B\beta$ - und  $\gamma$ -Untereinheit gemittelt, deconvolviert und quantifiziert. In Abbildung 42 sind diese Bereiche am Beispiel des Chromatogramms von Probe T gezeigt. (Die  $A\alpha$ -Untereinheit wird in der MANOVA mit vier Variablen nicht benötigt. Der Vollständigkeit halber sind die Bereiche der  $A\alpha$ -Untereinheit trotzdem angegeben.) Der Start des zu wählenden Bereiches ist in der Regel gut zu erkennen, daher wurde nur das Ende des gewählten Bereiches variiert. Aus den drei Quantifizierungen für die  $B\beta$ - und  $\gamma$ -Untereinheit ergeben sich für jede Probe neun verschiedene Kombinationen.

In Abbildung 43 sind diese neun Datenpunkte der Proben T, HCC019 und ZIRO44 im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA mit vier Variablen dargestellt. Die Streuung der Datenpunkte von Probe T ist sehr klein (ca.  $\pm 0.1$ ). Die Streuung der Datenpunkte der Proben HCC019 und ZIRO44 ist etwas größer (ca.  $\pm 0.5$ ). Aufgrund des Unterschiedes der Retentionszeiten von mono- und disialylierter  $B\beta/\gamma$ -Untereinheit wirkt sich der Bereich über welchen die Spektren gemittelt werden stark auf die Quantifizierung aus. (vgl. Abschnitt 4.1.2) Ist der für die Quantifizierung gemittelte Bereich der  $B\beta$ - oder  $\gamma$ -Untereinheit zu klein, fällt der Anteil der Glycoform S1 zu hoch aus, da die Glycoform S1

früher als die Glycoform S2 eluiert. Aus diesem Grund ist es bei der Extraktion der Massenspektren wichtig den Bereich, über welchen gemittelt wird, groß genug zu wählen, damit alle Glycoformen vollständig erfasst werden.

Diese Ergebnisse zeigen, dass der *shift* und der chromatographische Bereich, über welchen die Massenspektren vor der Deconvolution und Quantifizierung gemittelt werden, einen gewissen Einfluss auf die Lage der Datenpunkte im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA haben. Durch die Optimierung des *shift* kann dieser sehr genau bestimmt werden. Der Einfluss auf die Position der Datenpunkte im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA mit vier Variablen fällt daher gering aus.

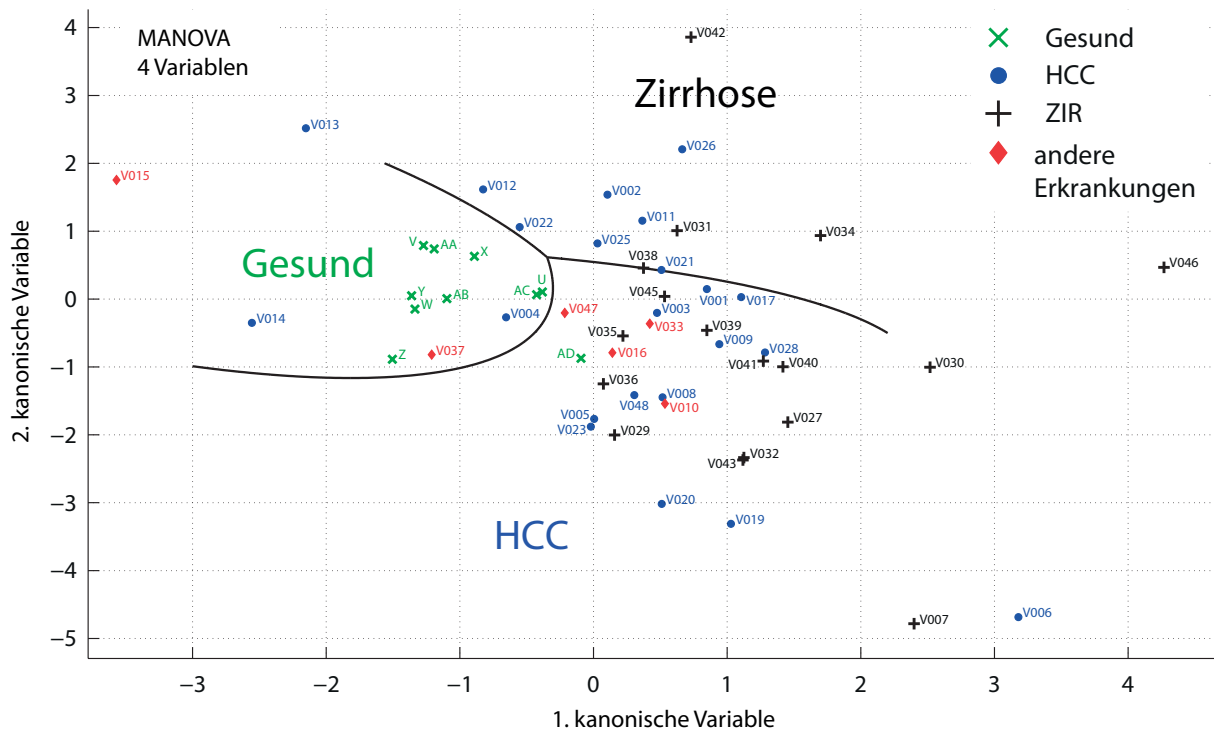
Um eine zuverlässige und reproduzierbare Einordnung unbekannter Proben zu gewährleisten, muss der gemittelte Bereich der Massenspektren so groß gewählt werden, dass alle Glycoformen vollständig erfasst werden. Ist der chromatographische Bereich, über welchen die Massenspektren gemittelt wurden, groß genug, haben die Datenpunkte im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA nur einen kleinen Fehlerbereich. Werden diese Aspekte bei der Analyse unbekannter Proben berücksichtigt, ist eine zuverlässige und reproduzierbare Einordnung möglich.



### 4.1.6 Validierung

Für die Validierung der entwickelten Methode zur Detektion von Leberkrebs und Leberzirrhose wurde eine zweite Kohorte von Proben untersucht. Die Validierungskohorte bestand aus insgesamt 55 Proben. Zehn Proben stammen von gesunden Spendern (vom Zentralinstitut für Transfusionsmedizin GmbH) und 39 Proben stammen von HCC- oder Zirrhosepatienten, deren Krankheitszustand unbekannt war. Weitere sechs Proben stammen von Patienten mit anderen Erkrankungen.

Die Proben der Validierungskohorte wurden exakt wie die anderen Proben analysiert und nach der in Abschnitt 4.1.5.2 beschriebenen Methode in das existierende Koordinatensystem der MANOVA mit vier Variablen eingeordnet. Die Ergebnisse wurden danach mit den tatsächlichen Gesundheitszuständen der Patienten (Tabelle 16) verglichen.



**Abbildung 44: Einordnung der Proben des Validierungsdatensatzes in das Koordinatensystem der MANOVA mit vier Variablen. Es sind die ersten beiden kanonischen Variablen jeder Probe gegeneinander aufgetragen. Grüne Kreuze: gesunde Spender, blaue Punkte: HCC-Patienten, schwarze Kreuze: Zirrhosepatienten, rote Rauten: Patienten mit anderen Erkrankungen (Tabelle 16). Es ist eine gute Trennung zwischen gesunden Spendern und HCC- oder Zirrhosepatienten zu erkennen. Die Trennung zwischen HCC- und Zirrhosepatienten konnte nicht bestätigt werden.**

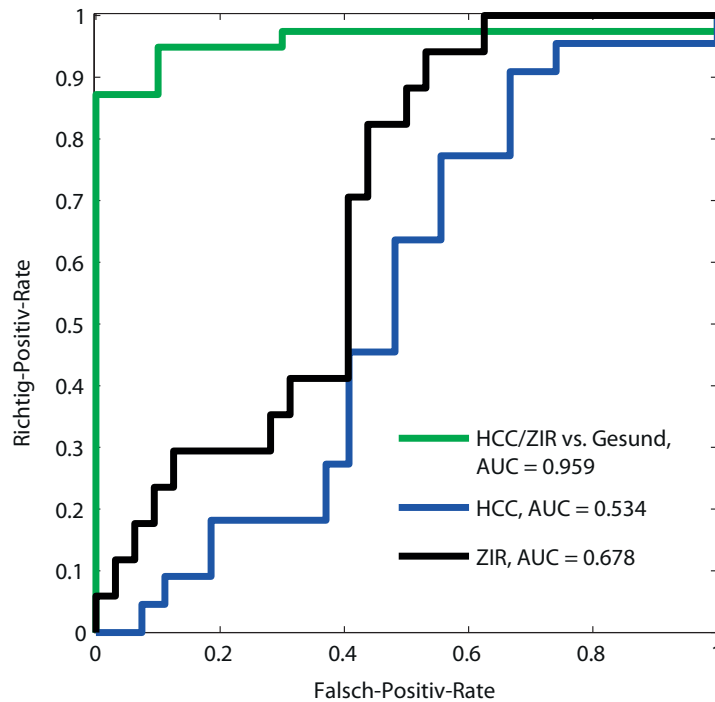
Abbildung 44 zeigt die Ergebnisse der Validierungskohorte bei Einordnung in das Koordinatensystem der MANOVA mit vier Variablen. Die Trennung zwischen gesunden Spendern und HCC- oder Zirrhosepatienten ist sehr gut. Neun der zehn Datenpunkte von

gesunden Spendern liegen im korrekten Bereich. Einzig der Datenpunkt von Probe AD liegt im Bereich der HCC-Patienten. Nur zwei Datenpunkte von Proben von HCC- oder Zirrhosepatienten liegen im Bereich der gesunden Spender (V004 und V014). Demnach konnte die Unterscheidung zwischen HCC/Zirrhosepatienten und gesunden Spendern mit einer Sensitivität von 94.9% und einer Spezifität von 90.0% erreicht werden. Die Trennung zwischen HCC- und Zirrhosepatienten wird durch den Validierungsdatensatz nicht bestätigt. Nur zwölf von 22 HCC-Patienten bzw. fünf von 17 Zirrhosepatienten wurden richtig klassifiziert. Im Anhang (Abschnitt 9.5.1) werden MANOVA-Modelle mit acht und zwölf abhängigen Variablen diskutiert. Die Einordnung der Validierungskohorte in die Koordinatensysteme dieser Modelle (Daten nicht gezeigt) brachte keine besseren Ergebnisse. Die Verwendung von mehr als vier Glycoformen oder Proteinspezies als abhängige Variablen bringt demnach keinen Vorteil.

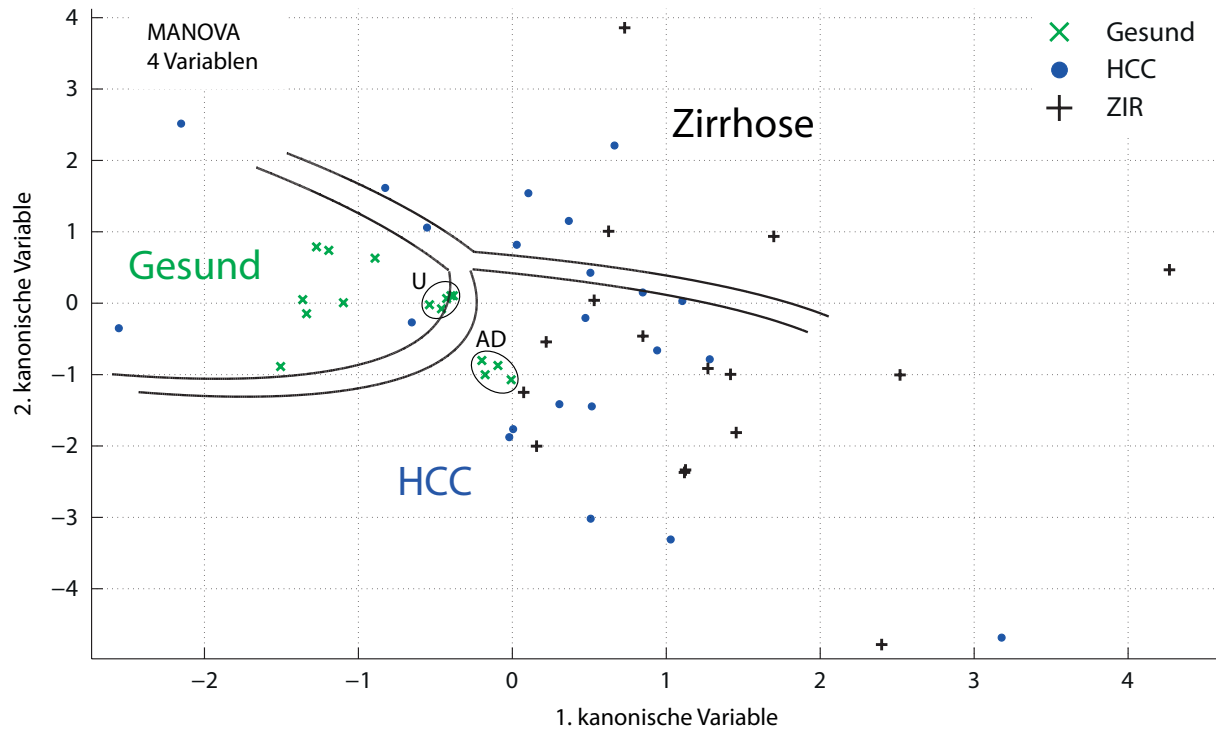
Sechs Proben des Validierungsdatensatzes stammen von Patienten, die weder ein hepatozelluläres Karzinom noch eine Leberzirrhose hatten. Da diese Patienten allerdings andere Erkrankungen aufweisen, sind die entsprechenden Datenpunkte in Abbildung 44 rot dargestellt. Zwei dieser Patienten (V015 und V016) wiesen gutartige Lebertumore auf. Der Datenpunkt der Probe V016 liegt im Bereich der HCC-Patienten, während der Datenpunkt von Proben V015 weit links im Bereich der gesunden Spender liegt. Da der gutartige Lebertumor von Patient V015 sehr klein ist (<1 cm) und kein Transformationsrisiko besteht, kann diese Probe als richtig klassifiziert angesehen werden. Der Patient V010 weist einen semimaligen Lebertumor (Angiomyolipom) auf und wird daher wie ein HCC-Patient behandelt. Der Datenpunkt von Probe V010 liegt im Bereich der HCC-Patienten. Diese Ergebnisse deuten darauf hin, dass auch gutartige Lebertumore das Glycosylierungsmuster von Fibrinogen beeinflussen und mit der hier entwickelten Methode erkannt werden. Der Datenpunkt der Probe V037 liegt im Bereich gesunder Spender. Dieses Ergebnis ist korrekt, da der Patient keine Lebererkrankungen, sondern koronare Herzkrankheiten aufweist. Die Datenpunkte der Proben V033 (chronische Pankreatitis) und V047 (Hashimoto Thyreoiditis) liegen im Bereich der HCC-Patienten. Es ist möglich, dass sich diese Krankheiten, da sie entzündliche Prozesse darstellen, auf die Glycosylierung von Fibrinogen, welches ein Akut-Phase-Protein ist, auswirken und somit zu einem positiven Ergebnis führen.

Die *receiver-operating-characteristic*-Kurven für die Einordnung der Validierungskohorte mit der MANOVA mit vier Variablen ist in Abbildung 45 gezeigt. Die ROC-Kurven basieren auf den *a posteriori*-Wahrscheinlichkeiten, welche mittels Diskriminanzanalyse berechnet wurden. Hierzu wurden nur diejenigen 49 Proben verwendet, welche eindeutig einer Gruppe zugeordnet werden konnten. Die sechs Proben mit anderen Erkrankungen wurden nicht berücksichtigt. Es ist zu erkennen, dass die Klassifizierung von HCC- und Zirrhosepatienten

nicht funktioniert. Die *area-under-the-curve*-Werte betragen 0.534 (HCC) bzw. 0.678 (Zirrhose). Dies liegt vor allem an der schlechten Differenzierung von HCC- und Zirrhosepatienten. Die ROC-Kurve für die Unterscheidung von HCC- und Zirrhosepatienten von gesunden Spendern ist mit einem AUC-Wert von 0.959 sehr gut.



**Abbildung 45: Receiver operating characteristic-Kurven für die Klassifizierung der Proben des Validierungsdatensatzes, basierend auf der Einordnung in das Koordinatensystem der MANOVA vier Variablen. Die ROC-Kurve basiert auf den *a posteriori*-Wahrscheinlichkeiten, welche mittels Diskriminanzanalyse berechnet wurden. Es wurden nur diejenigen 49 Proben verwendet, welche eindeutig einer Gruppe zugeordnet werden konnten.**



**Abbildung 46: Ergebnisse der Vierfachbestimmung der Proben U und AD im Diagramm der ersten beiden kanonischen Variablen der MANOVA mit vier Variablen. Die Streuung der vier Datenpunkte beider Proben beträgt ca.  $\pm 0.2$  in den ersten beiden kanonischen Variablen. Anhand dieser Werte wurde ein Unsicherheitsbereich um die Grenzlinien gebildet.**

In Abschnitt 4.1.5.3 wurde bereits der Einfluss des *shift* und des gemittelten Bereiches der Massenspektren auf die Lage der Datenpunkte im Diagramm der ersten und zweiten kanonischen Variablen untersucht. Um die verbleibende Streuung eines Datenpunktes abzuschätzen, wurden zwei Proben des Validierungsdatensatzes (U und AD) jeweils viermal untersucht. Die Ergebnisse sind in Abbildung 46 dargestellt. Es ist zu erkennen, dass die Streuung der jeweils vier Datenpunkte der Proben U und AD ca.  $\pm 0.2$  in den ersten beiden kanonischen Variablen beträgt. Anhand dieser Werte lässt sich ein Unsicherheitsbereich um die Grenzlinien abschätzen, welcher in Abbildung 46 dargestellt ist.

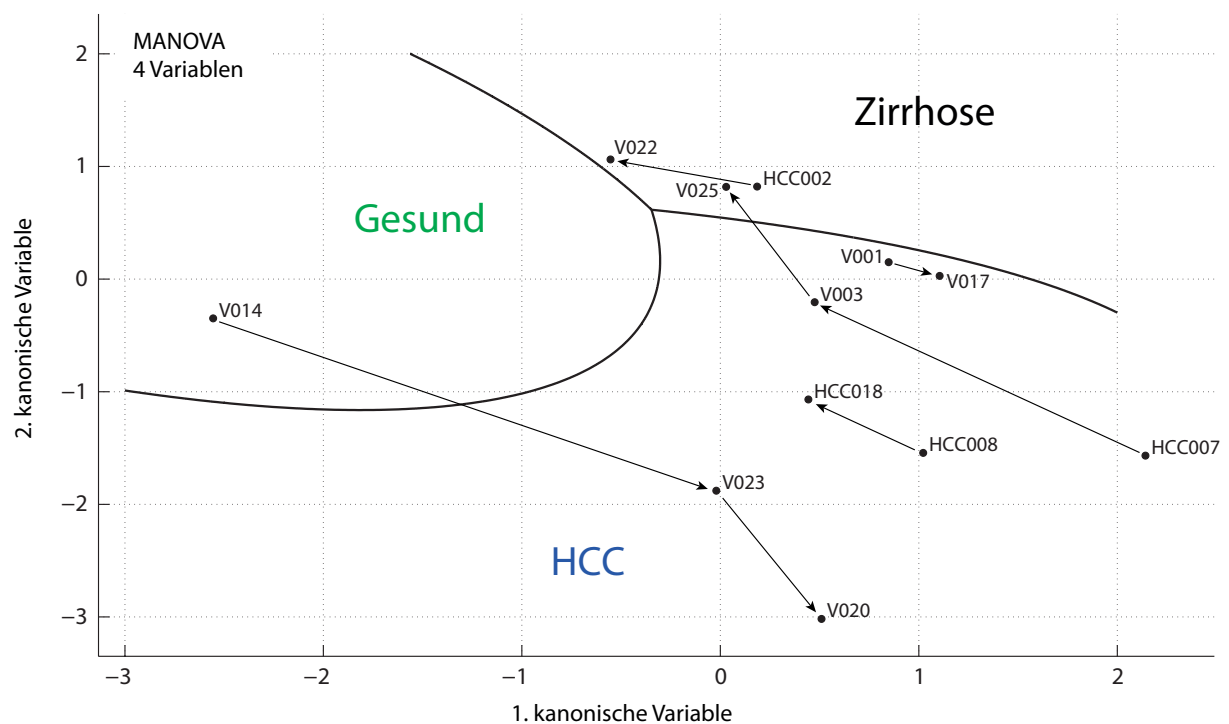
Im Laufe dieser Studie wurden von einigen Patienten mehrere Proben zu unterschiedlichen Zeitpunkten genommen (vgl. Tabelle 10). Die Lage dieser Datenpunkte im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA mit vier Variablen ist in Abbildung 47 dargestellt. Die zeitliche Änderung der Datenpunkte eines Patienten ist mit Pfeilen gekennzeichnet.

Der Datenpunkt der Probe V014 liegt zunächst klar im Bereich gesunder Spender und ist somit falsch negativ klassifiziert. Der Datenpunkt der Probe V023, welche ca. zwei Monate später genommen wurde, liegt bereits deutlich im Bereich der HCC-Patienten und ist somit richtig klassifiziert. Weitere sieben Monate später liegt der Datenpunkt der Probe V020 noch

weiter unten rechts im Bereich der HCC-Patienten. Diese Veränderung ist vor allem auf die Glycoform  $\gamma$ -S1 zurückzuführen, deren Wert von Probe V014 über V023 zu V020 stetig kleiner wird.

Der Datenpunkt von Probe HCC007 liegt rechts im Bereich der HCC-Patienten. Die Probe V003, welche ca. 18 Monate später genommen wurde liegt deutlich weiter links. Die Probe V025, welche weitere vier Monate später genommen wurde, liegt noch weiter oben links, bereits im Bereich der Zirrhosepatienten. Eine mögliche Ursache für die Veränderung dieses Patienten ist die Therapie. Zum Zeitpunkt der ersten Probe (HCC007) hatte der Patient noch keine Therapie erhalten. Danach wurde der Patient mit Tyrosinkinaseinhibitoren behandelt (vgl. Tabelle 10). Die Verschiebung der anderen Datenpunkte von Proben, welche von denselben Patienten zu unterschiedlichen Zeitpunkten genommen wurden (HCC008/HCC018, V001/V017, HCC002/V022), ist deutlich kleiner. Die Klassifizierung dieser Proben ändert sich dabei nicht.

Die genaue Ursache für die Veränderung dieser Patienten konnte allerdings nicht endgültig geklärt werden. Sowohl das Fortschreiten der Krebserkrankung als auch die Therapie können Einfluss auf die posttranslationalen Modifikationen von Fibrinogen haben und somit die Lage des Datenpunktes beeinflussen. Es kann aber nicht ausgeschlossen werden, dass sich die Proben aufgrund des Lebensstils des Patienten verändern.



**Abbildung 47: Datenpunkte von Patienten, welche im Laufe dieser Studie mehrere Proben beigetragen haben im Diagramm der ersten beiden kanonischen Variablen der MANOVA mit vier Variablen. Die zeitliche Veränderung der Lage der Datenpunkte ist mit Pfeilen gekennzeichnet.**

**Tabelle 10: Liste der Proben von Patienten, welche im Laufe dieser Studie mehrere Proben beigetragen haben.**

<b>Patient</b>	<b>Probe</b>	<b>Blutentnahmedatum</b>	<b>Änderungen</b>
HCC008/HCC018	HCC008	05.01.2015	-
	HCC018	05.10.2015	-
HCC007/V003/V025	HCC007	05.01.2015	keine Therapie
	V003	27.06.2016	TKI
	V025	24.10.2016	TKI
V001/V017	V001	27.06.2016	< Mailand
	V017	08.05.2017	> Mailand
V014/V023/V020	V014	01.08.2016	-
	V023	10.10.2016	-
	V020	15.05.2017	-
HCC002/V022	HCC002	01.12.2014	> Mailand, keine Therapie
	V022	10.10.2016	Metastasen, Resektion

#### 4.1.7 Diskussion und Fazit zur Entwicklung eines Biomarkers für Leberkrebs

Die Ergebnisse der Untersuchung der Validierungskohorte zeigen, dass die Unterscheidung zwischen Proben gesunder und kranker Spender möglich ist, die Differenzierung zwischen HCC- und Zirrhosepatienten ist allerdings problematisch. Die Methode zur Einordnung unbekannter Proben wurde daher auf Grundlage aller 105 Proben des Trainings- und Validierungsdatensatzes verfeinert (Anhang, Abschnitt 9.5.3). Die univariaten Varianzanalysen unter Verwendung aller 104 Proben (Daten nicht gezeigt) bestätigten die folgenden Ergebnisse:

- Der Fucosylierungsgrad der  $\beta$ - und  $\gamma$ -Untereinheit ist bei HCC-Patienten im Vergleich zu gesunden Spendern erhöht.
- Bei Zirrhosepatienten ist der Anteil fucosylierter Glycoformen sogar noch höher als bei HCC-Patienten.
- Der Sialylierungsgrad der  $\gamma$ -Untereinheit ist bei HCC- und Zirrhosepatienten im Vergleich zu gesunden Spendern erhöht.
- Bei HCC-Patienten ist die  $A\alpha$ -Untereinheit weniger phosphoryliert als bei gesunden Spendern.
- Der Anteil der monophosphorylierten Spezies der  $A\alpha$ -Untereinheit ist bei Zirrhosepatienten erhöht.
- Der Anteil an *O*-glycosylierter  $A\alpha$ -Untereinheit ist bei Zirrhosepatienten verringert.
- Die Anteile der oxidierten und trunkierten Spezies der  $A\alpha$ - und  $\beta$ -Untereinheit sind bei HCC- und Zirrhosepatienten erhöht.

Da der Großteil der HCC-Patienten auch eine Leberzirrhose aufwies, kann nicht eindeutig geklärt werden, ob die Unterschiede durch die Zirrhose oder den Krebs verursacht werden. Es wurden nur acht Proben von HCC-Patienten ohne Leberzirrhose untersucht (siehe Tabellen 15 & 16). Die Untersuchung dieser Proben lässt vermuten, dass die Unterschiede zum größeren Teil, aber nicht ausschließlich, durch die Leberzirrhose verursacht werden. Außerdem muss der Einfluss des Krebsstadiums, der Genese, des Alters und Geschlechts sowie der Begleiterkrankungen berücksichtigt werden. Statistische Untersuchungen bezüglich dieser Parameter blieben allerdings größtenteils ohne eindeutiges Ergebnis. Dies liegt zum Teil an der kleinen Probenmenge und an der ungleichen Verteilung der Proben, zum anderen an der großen Varianz der Daten. Um diese Fragen eindeutig zu klären, müssen deutlich mehr Proben untersucht werden, dabei ist auf eine systematische Probennahme zu achten. Bei den in dieser Studie verwendeten Proben von gesunden Spendern handelt es sich um abgelaufene oder nicht verwendbare Blutspenden vom Zentralinstitut für Transfusionsmedizin GmbH. Außer Alter und Geschlecht der Spender sind deshalb keinerlei

Informationen zu den Proben verfügbar. Zur Weiterentwicklung und Etablierung dieser Methode ist ein Datensatz von Proben gesunder Spender, die unter medizinischer Kontrolle genommen wurden, essentiell.

Einige Proben stammen von denselben Patienten, wurden aber zu unterschiedlichen Zeitpunkten genommen. Die zeitliche Verschiebung der Datenpunkte dieser Proben im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA ist in Abbildung 47 gezeigt. Die Frage, wie sich das Fortschreiten der Krankheit oder die Therapie auf das Ergebnis der hier entwickelten Methode auswirkt, konnte anhand dieser Proben nicht endgültig geklärt werden. Außerdem wurden sechs Proben von Patienten mit anderen Erkrankungen untersucht, welche kein hepatozelluläres Karzinom und keine Zirrhose aufwiesen (Abbildung 44). Die Lage der Datenpunkte dieser Proben zeigte, dass gutartige Lebertumore und entzündliche Prozesse in der hier entwickelten Methode zu einem positiven Ergebnis führen können. Es sind daher weitere Studien nötig, um den Einfluss von Begleiterkrankungen oder der Therapie auf das Glycosylierungsmuster von Fibrinogen zu untersuchen.

Die hier entwickelte Methode basiert auf den prozentualen Anteilen der einzelnen Glycoformen und Proteinspezies von humanem Fibrinogen. Die Quantifizierung der Glycoformen und Proteinspezies der drei Fibrinogenuntereinheiten ist deshalb von entscheidender Bedeutung. Hierzu wurden die theoretischen Isotopenmuster aller einzelnen Glycoformen und Proteinspezies berechnet und an die experimentellen Massenspektren angepasst. Um eine gute Anpassung zu erreichen, mussten dabei die Signale vieler Proteinspezies beachtet werden (z.B. Oxidation, Wasserverlust, Carbamidierung, Trunkierung). Die Quantifizierungsmethode wird dadurch sehr komplex. Außerdem erschwert die Überlagerung verschiedener Signale die Quantifizierung bestimmter Glycoformen erheblich. Beispielsweise überlappen die Signale der trunkierten B $\beta$ -Untereinheit mit den Glycoformen G1S1 und G1F. Schließlich stellt sich die Frage, ob die Anteile der einzelnen Glycoformen nur durch jeweiligen Koeffizienten der unmodifizierten Spezies oder durch alle zu dieser Glycoform beitragenden Spezies berechnet werden sollten. Die hier verwendeten Daten wurden nur ausgehend von den Signalen der unmodifizierten Spezies berechnet. Um eine zuverlässige Einordnung zu erreichen, müssen alle Proben mit guter Anpassung quantifiziert werden. Um dies zu erreichen wurde für jedes Spektrum ein *shift* angewendet, welcher kleine Fehler in der Massengenauigkeit korrigiert. Außerdem müssen Spektren von hetero- oder homozygoten Trägern eines der gefunden SNPs mit den jeweiligen Isotopenmustern der veränderten Proteinsequenzen angepasst werden.



Die drei *single nucleotide polymorphisms* rs6050 ( $\text{A}\alpha\text{-Thr312Ala}$ ), rs4220 ( $\text{B}\beta\text{-Arg448Lys}$ ) und rs6063 ( $\gamma\text{-Gly165Arg}$ ), welche in den Spektren der intakten Fibrinogenuntereinheiten beobachtet wurden, konnten durch Verdau mit Trypsin bestätigt werden. Die Allelfrequenzen im Datensatz mit allen 105 Proben stimmen sehr gut mit den Literaturdaten überein. Unterschiede in den Allelfrequenzen zwischen gesunden Spendern, HCC- und Zirrhosepatienten konnten mit Chi-Quadrat-Tests nicht identifiziert werden.

Die in dieser Arbeit verwendeten multivariaten statistischen Methoden wurden stets mit folgenden vier Variablen durchgeführt:  $\text{B}\beta\text{-S1}$ ,  $\text{B}\beta\text{-S1F}$ ,  $\gamma\text{-S1}$  und  $\gamma\text{-S1F}$ . Im Anhang (Abschnitt 9.5.1) ist erläutert, aus welchen Gründen diese vier Variablen ausgewählt wurden. Die Anwendung der Varianzanalysen (ANOVA bzw. MANOVA) unterliegt einigen Voraussetzungen:

- Die Beobachtungen (Proben) müssen unabhängig voneinander sein.
- Die abhängigen Variablen müssen innerhalb jeder Gruppe normalverteilt sein.
- Die Fehlervarianzen der unabhängigen Variablen müssen in allen Gruppen ungefähr gleich sein (Homogenität der Varianz-Kovarianz-Matrizen).
- Die abhängigen Variablen dürfen keine Kollinearität zeigen.
- Der Stichprobenumfang muss ausreichend groß sein.

Obwohl ANOVA und MANOVA robust gegenüber Verletzungen der obigen Voraussetzungen sind, sollten diese trotzdem diskutiert werden. Die in Abschnitt 4.1.5 gezeigten Boxplot-Diagramme geben einen groben Eindruck über die Verteilung der Daten. Zweifel an der Normalverteilung der Daten besteht nur für die fucosylierten Glycoformen der  $\text{B}\beta\text{-}$  und  $\gamma\text{-}$ Untereinheit. Genauere Analysen durch QQ-Plots und Shapiro-Wilk-Tests bestätigen, dass die Daten für die Variablen  $\text{B}\beta/\gamma\text{-S1F}$  und  $\text{B}\beta/\gamma\text{-S2F}$  nicht normalverteilt sind. Durch Kruskal-Wallis-Tests, welche keine Normalverteilung der Daten voraussetzen, konnten die Unterschiede in der Fucosylierung der  $\text{B}\beta\text{-}$  und  $\gamma\text{-}$ Untereinheit von gesunden Spendern, HCC- und Zirrhosepatienten dennoch nachgewiesen werden. Bei der MANOVA haben die fucosylierten Glycoformen einen großen Einfluss auf die Trennung zwischen gesunden Spendern, HCC- und Zirrhosepatienten. Es erscheint daher sinnvoll die Variablen  $\text{B}\beta/\gamma\text{-S1F}$  trotzdem zu verwenden.

Die Gleichheit der Fehlervarianzen kann durch Box' M- und Levene-Tests überprüft werden. Diese Tests zeigen, dass die Fehlervarianzen in den drei Gruppen nicht gleichgroß sind. Dieses Ergebnis ist auch in den Boxplot-Diagrammen zu erkennen. Die Zirrhosepatienten zeigen häufig eine deutlich größere Streuung als die gesunden Spender und die

HCC-Patienten. Die Ergebnisse der MANOVA (Tabelle 8) zeigen allerdings, dass diese robust gegenüber Verletzungen dieser Voraussetzung ist.

Bei der MANOVA muss darauf geachtet werden, dass die Anzahl der abhängigen Variablen im Verhältnis zur Anzahl der Beobachtungen nicht zu hoch ist, da es sonst zu einer Überanpassung (engl.: *Overfitting*) kommen kann. Aus diesem Grund wurde die Anzahl der als Variablen verwendeten Glycoformen auf vier festgesetzt (Anhang, Abschnitt 9.5.1). Um die Anzahl der benötigten Proben für die MANOVA zu bestimmen, wurde eine *Power*-Analyse durchgeführt (G\*Power 3.1.9.2). Bei einer mittleren Effektgröße von 0.4, einem Signifikanzniveau von 5%, drei Gruppen, vier abhängigen Variablen und einer Teststärke von 99% beträgt die benötigte Anzahl an Proben 45. Der untersuchte Trainingsdatensatz besteht aus 56 Proben. Mit einem 14-fachen Überschuss von Beobachtungen im Verhältnis zu abhängigen Variablen kann davon ausgegangen werden, dass keine Überanpassung vorliegt.

Im Endeffekt dient die MANOVA allerdings nur der Berechnung der Koeffizienten, welche zur Einordnung unbekannter Proben verwendet werden. Es ist daher eine Validierung mit unbekanntem Proben nötig (Abschnitt 4.1.6). Diese ergab, dass die Unterscheidung zwischen den Proben gesunder Spender und kranker Patienten möglich ist. Eine Differenzierung zwischen HCC- und Zirrhosepatienten war nicht zu erreichen. Es zeigte sich allerdings, dass das MANOVA-Modell zur Einordnung unbekannter Proben weiterentwickelt werden muss, indem mehr als 56 Proben verwendet werden. Hierfür sind weitere Studien mit deutlich mehr Proben erforderlich.

Die hier entwickelte Methode ist ein vielversprechender Ansatz zur Detektion von Krebserkrankungen. Es handelt sich um die erste Studie, in der multivariate Datenanalyse in Verbindung mit der massenspektrometrischen Analyse intakter Proteine genutzt wird, um Proben von Krebspatienten zu identifizieren. Viele andere Studien beschäftigen sich mit der Analyse von Glycoproteinen von HCC- und Zirrhosepatienten. So wurde gezeigt, dass Haptoglobin bei HCC-Patienten stärker fucosyliert ist.<sup>117-120</sup> Comunale *et al.* untersuchten unter anderem Hemopexin und  $\alpha$ -1-Antitrypsin und fanden heraus, dass die Fucosylierung bei Zirrhosepatienten in  $\alpha$ 1-3 Position an der Antenne vorliegt, während bei HCC-Patienten eine *core*-Fucosylierung vorliegt.<sup>64,121,122</sup> Callewaert *et al.* entwickelten verschiedene Tests zur Diagnose von HCC<sup>67</sup>, Zirrhose<sup>123</sup> und Fibrose<sup>124</sup>, basierend auf der Analyse von *N*-Glycanen aus Serum mittels Kapillarelektrophorese.

In den meisten dieser Studien werden die Glycane vom Protein abgespalten oder dieses wird proteolytisch verdaut. Dieser Prozess ist zeit-/kostenintensiv und benötigt gut ausgebildetes Personal. Aus diesem Grund sind die meisten dieser Methoden schlecht für die medizinische Routinediagnostik geeignet. Die in dieser Arbeit vorgestellte Methode basiert auf der

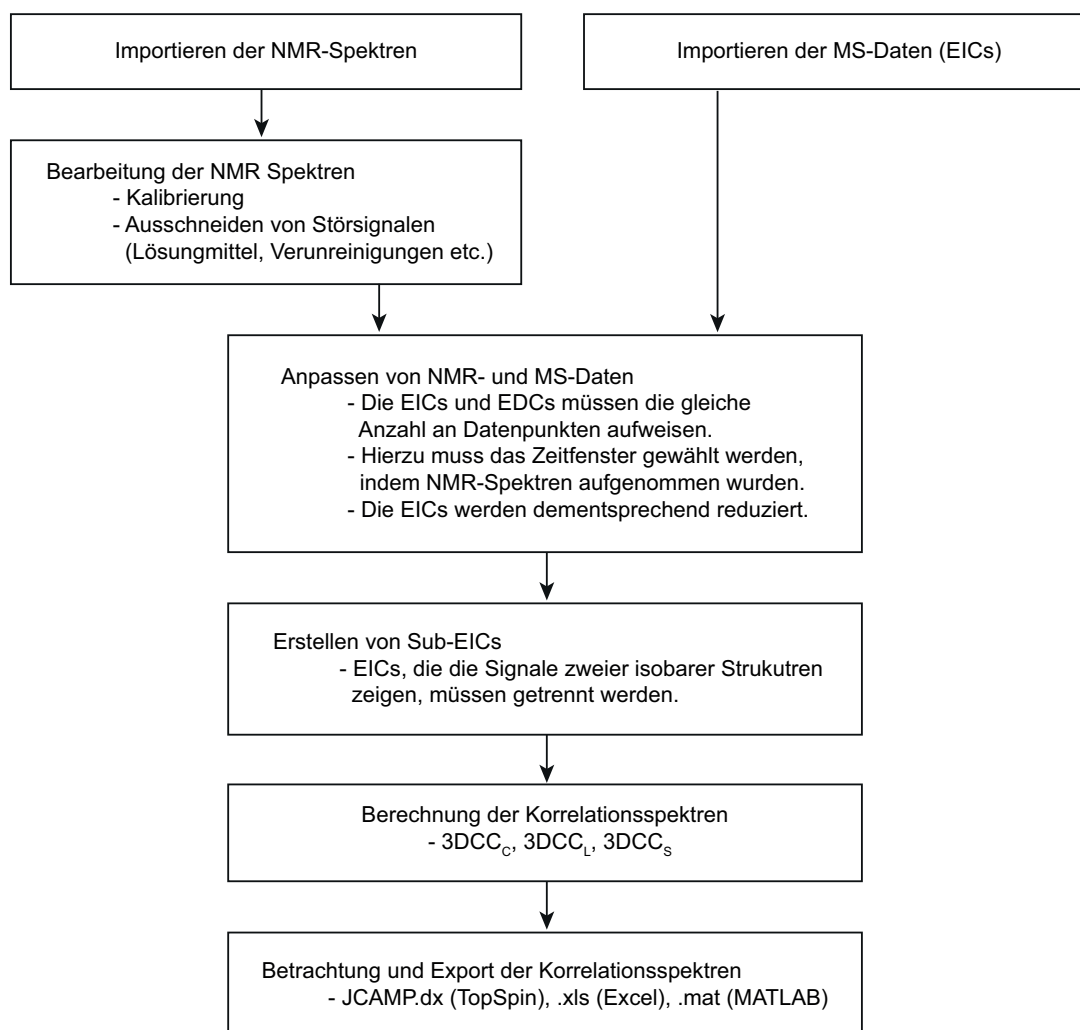
massenspektrometrischen Analyse der intakten Untereinheiten von humanem Fibrinogen. Fibrinogen kommt in hoher Konzentration im Blutplasma vor und lässt sich einfach durch Ethanol-fällung isolieren. Nach der Reduzierung wird die Probe direkt mittels LC-MS analysiert, um Spektren der intakten Untereinheiten zu erhalten. Dieser Prozess benötigt etwa drei Stunden und ist einfach durchzuführen, sodass er gut in der Routinediagnostik eingesetzt werden kann.

Obwohl die entwickelte Methode noch statistische Validierung benötigt, konnte gezeigt werden, dass das Verfahren prinzipiell zur Detektion von Krebs geeignet ist. Die Vorgehensweise kann auch auf andere Plasmaproteine und Krebsarten erweitert werden. Aus diesen Gründen könnte diese Arbeit die Früherkennung von Krebs revolutionieren.

## 4.2 Entwicklung einer graphischen Benutzeroberfläche für die dreidimensionale Kreuzkorrelation

In Abschnitt 1.2 wurde bereits erläutert, dass zur Untersuchung der exakten Stereochemie von Glycanen eine Kombination von *bottom-up* Massenspektrometrie und NMR-Spektroskopie erforderlich ist. Die von Behnken und Fellenberg entwickelte Methode der dreidimensionalen Kreuzkorrelation (3DCC) wurde mehrfach erfolgreich zur Untersuchung von *N*-Glycanen aus Glycoproteinen eingesetzt.<sup>50,51,125</sup> Das Prinzip der dreidimensionalen Kreuzkorrelation ist in Abschnitt 3.4 erklärt. Die Verarbeitung der Daten wurde mit dem Programm MATLAB (The MathWorks, Inc.) durchgeführt. Da die Vorbereitung der Daten dabei von großer Bedeutung ist, werden erweiterte Kenntnisse in der MATLAB-Programmierung benötigt, um die dreidimensionale Kreuzkorrelation anzuwenden. Für die meisten potentiellen Anwender ist 3DCC deshalb nicht geeignet. Um dies zu ändern wurde eine graphische Benutzeroberfläche entwickelt, die den Anwender durch alle wichtigen Schritte der dreidimensionalen Kreuzkorrelation führt. Hierzu wurde das in MATLAB integrierte *Graphical User Interface Development Environment* (GUIDE) verwendet.

In Abbildung 48 ist der *Workflow* der Datenverarbeitung für die dreidimensionale Kreuzkorrelation in MATLAB dargestellt. Für den Import der NMR-Spektren wird das *script* `rbnmr` von Nils Nyberg verwendet. Nach dem Import der NMR-Daten müssen diese bearbeitet werden, um Störsignale zu entfernen und eine Kalibrierung durchzuführen. Die Kalibrierung von NMR-Spektren ist von entscheidender Bedeutung, um reproduzierbare chemische Verschiebungen zu gewährleisten. Bei der dreidimensionalen Kreuzkorrelation ist es außerdem wichtig, dass alle NMR-Spektren exakt dieselbe x-Achse besitzen, da sonst fehlerhafte EDCs resultieren. Die Kalibrierung in TopSpin generiert für jedes Spektrum eine eigene x-Achse, deshalb muss vor der Korrelation eine Kalibrierung in MATLAB durchgeführt werden. Die Signale von Verunreinigungen oder dem Lösungsmittel sollten vor der Korrelation ausgeschnitten werden, da sie zu Artefakten in den Korrelationsspektren führen.



**Abbildung 48: Workflow der Datenverarbeitung für die dreidimensionale Kreuzkorrelation in MATLAB.** Nach dem Import der NMR-Daten müssen diese bearbeitet werden. Dabei wird eine Kalibrierung vorgenommen, damit alle NMR-Spektren dieselbe x-Achse besitzen. Außerdem können Bereiche aus den NMR-Spektren ausgeschnitten oder auf Null gesetzt werden. Da die EDCs und EICs für die Korrelation die gleiche Anzahl an Datenpunkten haben müssen, werden die EICs reduziert. Danach können aus EICs, welche die Signale mehrerer isobarer Strukturen ausweisen, mehrere sub-EICs gebildet werden. Es folgt die eigentliche Korrelation der NMR- und MS-Daten durch  $3DCC_C$ ,  $3DCC_L$  oder  $3DCC_S$ . Die erhaltenen Korrelationsspektren können exportiert werden.

Die EICs haben in der Regel eine deutlich höhere Auflösung als die EDCs, da pro Fraktion mehr als ein MS-Spektrum aufgezeichnet wurde. Für die Korrelation müssen EICs und EDCs allerdings die gleiche Anzahl an Datenpunkten besitzen. Aus diesem Grund müssen die MS-Daten vor der Korrelation reduziert werden. Außerdem werden meistens nur in einem bestimmten Bereich des Chromatogramms NMR-Spektren aufgenommen. Zur Reduzierung der MS-Daten werden die EICs in den Intervallen der Fraktionen gemittelt, hierzu muss der Benutzer den Zeitpunkt wählen zu dem die erste Fraktion beginnt. Dieser Zeitpunkt sollte dem Benutzer ungefähr bekannt sein, aufgrund technischer Ungenauigkeiten muss der exakte Wert aber empirisch bestimmt werden. Hierzu kann die Abweichung zwischen dem

*total ion current* (TIC) und dem *total delta*-Chromatogramm (TDC) minimiert werden. Funktioniert dies nicht, muss der Anwender ein EIC und ein EDC von denen er weiß, dass sie zusammengehören, manuell vergleichen. Die Reduzierung der MS-Daten und die damit verbundene zeitliche Abstimmung der MS- und NMR-Daten ist von entscheidender Bedeutung für die dreidimensionale Kreuzkorrelation, da kleine Abweichungen große Auswirkungen auf das Ergebnis haben können. Im Fall isobarer Strukturen sind in einem EIC die Signale mehrerer Bestandteile vorhanden. Um für alle Bestandteile Korrelationsspektren zu erhalten, müssen aus diesem EICs mehrere sogenannte Sub-EICs erstellt werden.

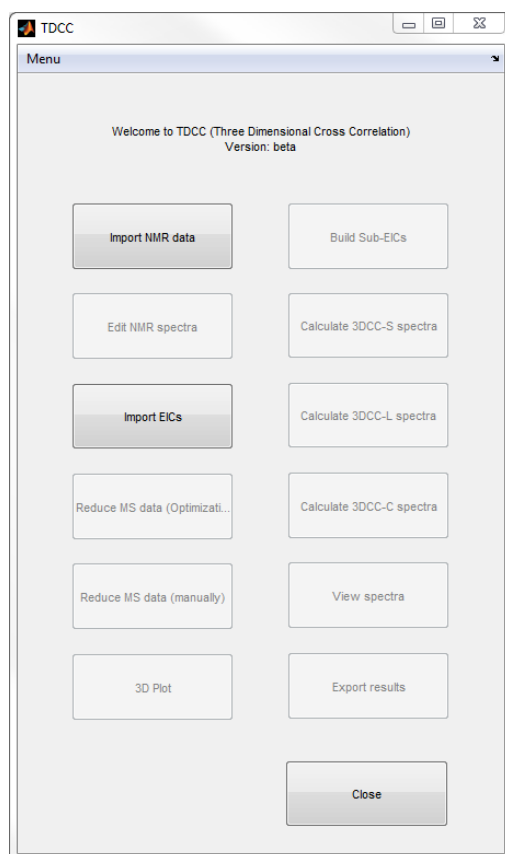
Die Korrelation von MS- und NMR-Daten mittels  $3DCC_C$  verwendet die Korrelationskoeffizienten zwischen den EICs und EDCs. Vor der Multiplikation mit dem Summen-NMR-Spektrum können die Korrelationskoeffizienten mit verschiedenen Funktionen gewichtet werden.<sup>52</sup> Dabei muss der Anwender einige Parameter der Gewichtungsfunktion angeben. Für die Korrelation mittels  $3DCC_L$  oder  $3DCC_S$  sind keine weiteren Eingaben nötig. Bei der Methode  $3DCC_L$  werden durch die Methode der kleinsten Fehlerquadrate (*least squares optimization*) die Koeffizienten gebildet, welche gegen die chemische Verschiebung aufgetragen das Korrelationsspektrum ergeben.  $3DCC_S$  verwendet die Singulärwertzerlegung (*singular value decomposition*), um die inverse Matrix der MS-Daten zu bilden, welche dann mit der Matrix der NMR-Daten multipliziert wird, um die Korrelationsspektren zu ergeben. Die erhaltenen Korrelationsspektren können in MATLAB betrachtet und ausgewertet werden. Außerdem ist es möglich die Korrelationsspektren als Excel-Datei oder als JCAMP.dx-Datei, welche in TopSpin geöffnet werden kann, zu exportieren.

## 4.2.1 Funktion und Anwendung des 3DCC-GUI

Die wesentlichen Aspekte des hier entwickelten *User Interface* werden im Folgenden erläutert.

### 4.2.1.1 Das Hauptmenü

Durch den Befehl *TDCC* wird das Programm gestartet, es öffnet sich zunächst das in Abbildung 49 gezeigte Hauptmenü. Zunächst sind nur die Schaltflächen „*Import NMR data*“ und „*Import EICs*“ verfügbar. Das Programm ist modular aufgebaut, die weiteren Funktionen werden schrittweise freigeschaltet. Mit den Schaltflächen „*Import NMR data*“ und „*Import EICs*“ können die NMR-Spektren und die EICs importiert werden. Dabei müssen alle NMR-Spektren, die verwendet werden sollen, im Standard Bruker Format in aufsteigender Reihenfolge in einem Ordner vorhanden sein. Die EICs müssen als .xy- oder .txt-Datei vorliegen.

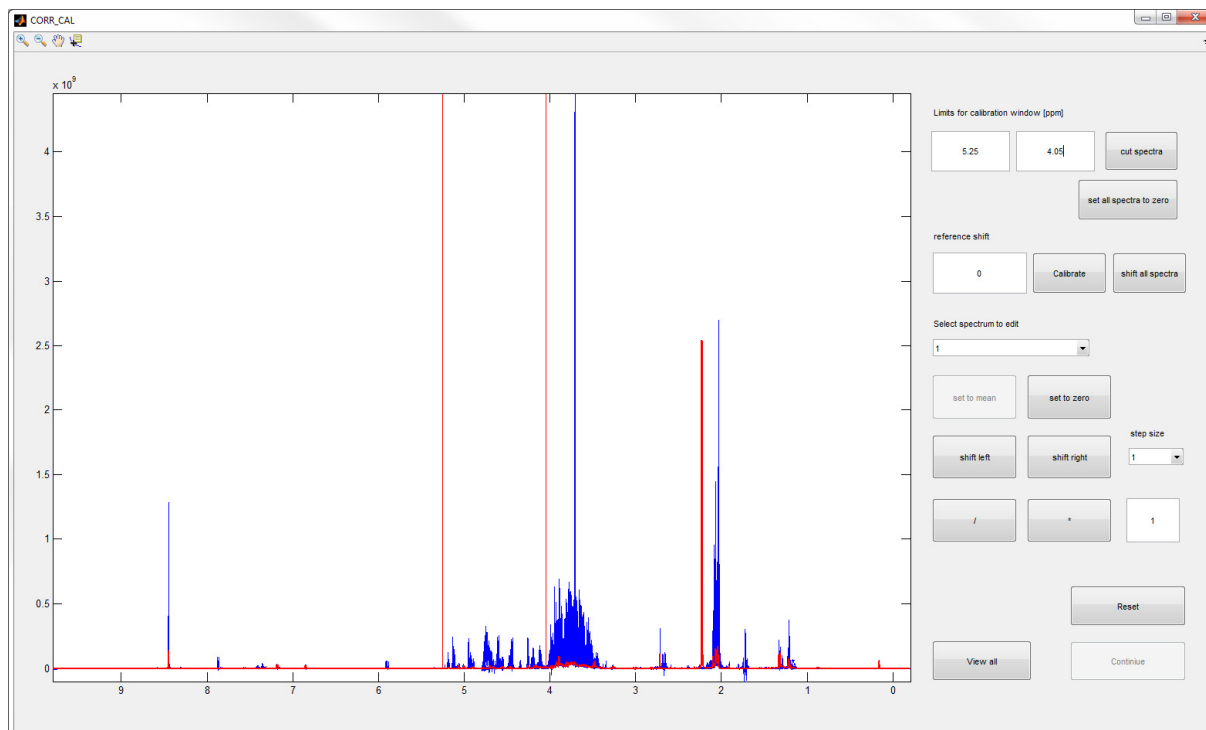


**Abbildung 49: TDCC (Three dimensional cross correlation) Hauptmenü. Das Programm ist modular aufgebaut. Zunächst sind nur die Funktionen „*Import NMR data*“ und „*Import EICs*“ verfügbar. Die anderen Funktionen werden schrittweise freigeschaltet.**

#### 4.2.1.2 Bearbeiten der NMR-Spektren

Nach dem Import der NMR-Spektren, wird die Schaltfläche „*Edit NMR spectra*“ freigeschaltet, ein Klick darauf öffnet das in Abbildung 50 gezeigte Fenster. Die Bearbeitung der NMR-Spektren ist von entscheidender Bedeutung für die dreidimensionale Kreuzkorrelation. Zum einen müssen die NMR-Spektren kalibriert werden, damit sie eine gemeinsame x-Achse besitzen, zum anderen sollten Störsignale aus den Spektren entfernt werden.

Das Programm bietet zwei Möglichkeiten zur Kalibrierung der NMR-Spektren. Ist in allen Spektren das Signal einer Referenzsubstanz vorhanden, kann dieses zur Kalibrierung verwendet werden. (Dabei kann es sich auch um ein Signal einer ubiquitären Verunreinigung handeln.) Hierzu wird der Bereich ausgewählt, in dem das Signal liegt. Das Programm sucht den höchsten Datenpunkt in diesem Bereich und setzt ihn auf die angegebene chemische Verschiebung. Ist kein Referenzsignal vorhanden, kann der Anwender einen Punkt im Spektrum auswählen, welcher auf die angegebene chemische Verschiebung gesetzt wird. Dabei werden die Spektren nur um maximal einen Datenpunkt gegeneinander verschoben. Liefern beide Möglichkeiten kein zufriedenstellendes Ergebnis, kann jedes Spektrum einzeln, datenpunktweise verschoben werden.



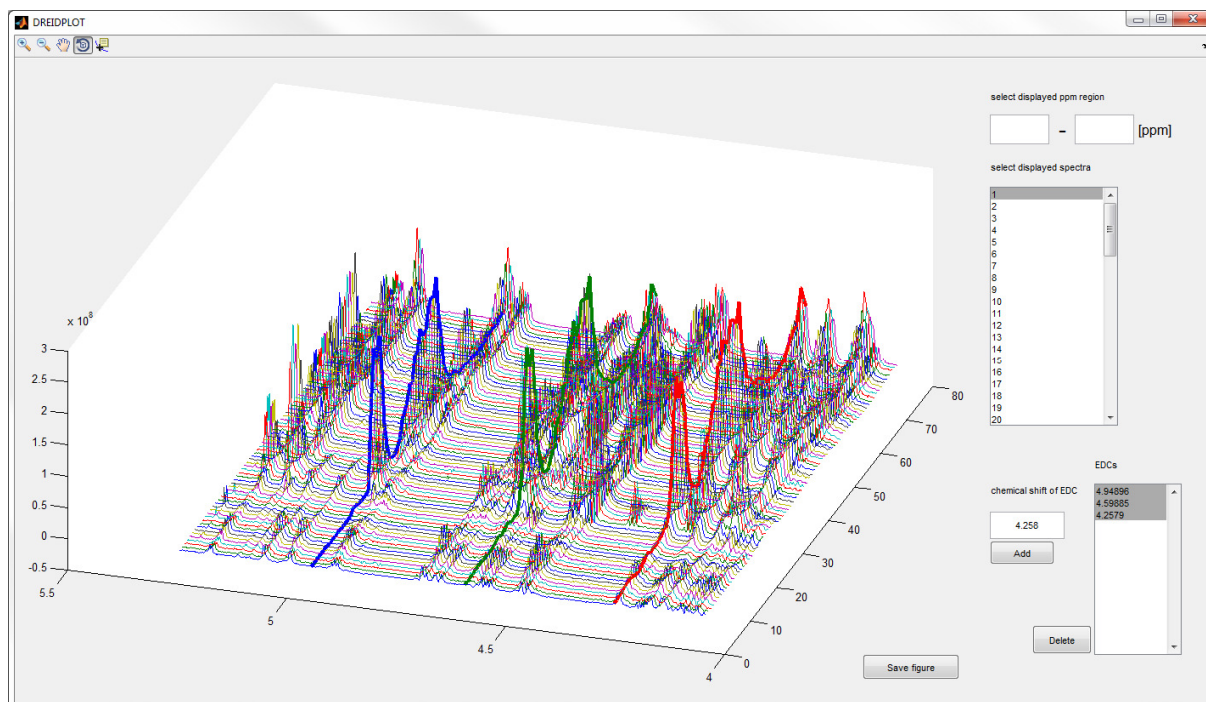
**Abbildung 50: Fenster zur Bearbeitung der NMR-Spektren. In diesem Fenster können die NMR-Spektren kalibriert werden. Dies ist für die dreidimensionale Kreuzkorrelation von entscheidender Bedeutung, da alle NMR-Spektren dieselbe x-Achse besitzen müssen. Außerdem bietet dieses Fenster die Möglichkeit Bereiche aus den NMR-Spektren auszuschneiden oder auf Null zu setzen.**



In den meisten Fällen ist es nicht nötig die Korrelation über den gesamten spektralen Bereich durchzuführen. Zu diesem Zweck bietet das Programm die Möglichkeit einen Bereich aus den Spektren auszuschneiden. Außerdem ist es nötig Störsignale, z.B. vom Lösungsmittel oder von Verunreinigungen, zu entfernen. Hierzu wird der betreffende Bereich ausgewählt und entweder in allen oder in einem einzelnen Spektrum auf Null gesetzt. Wird ein Bereich, in dem wichtige Signale liegen, in einem Spektrum auf Null gesetzt, kommt es im entsprechenden EDC zu einem Einschnitt, welcher die Korrelation stört. Aus diesem Grund gibt es die Möglichkeit die Datenpunkte in einem Bereich eines Spektrums auf den Mittelwert des vorherigen und folgenden Spektrums zu setzen. Außerdem bietet das Programm die Möglichkeit die Intensität einzelner Spektren zu erhöhen oder zu verringern, falls einzelne Spektren mit einer anderen Anzahl von *Scans* gemessen wurden.

#### 4.2.1.3 3D Plot der NMR-Daten

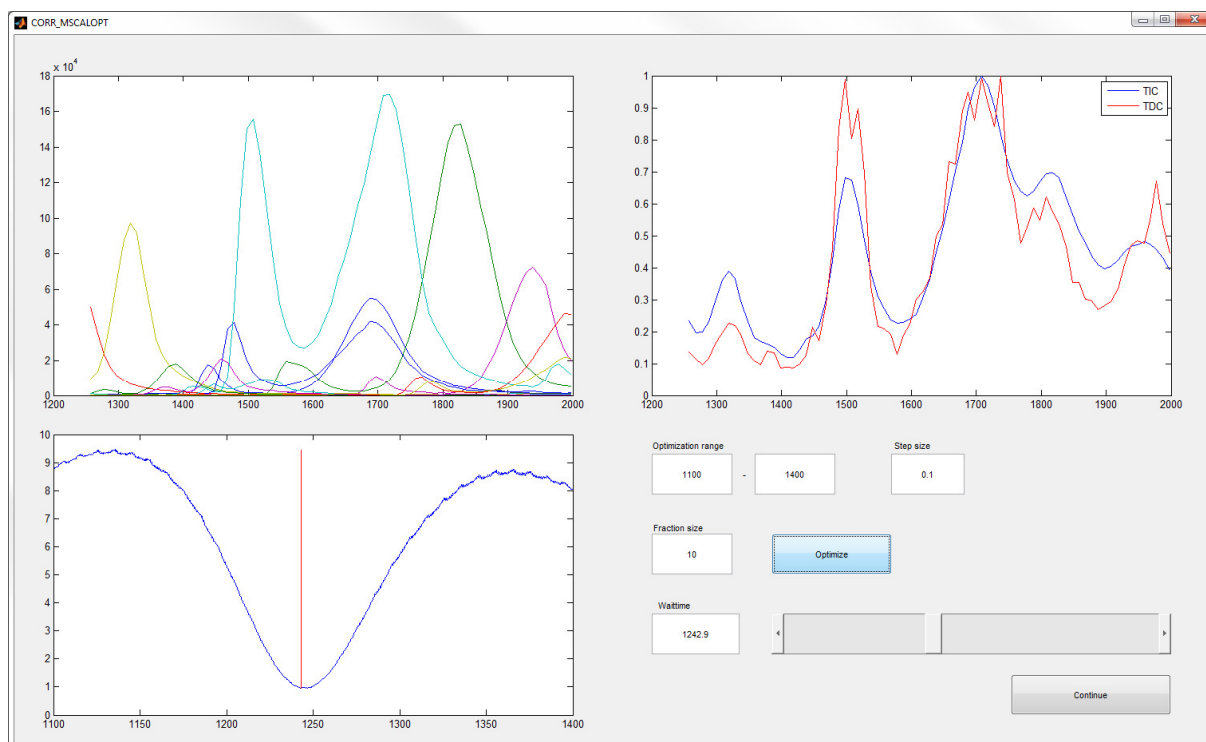
Zur Veranschaulichung der NMR-Daten wurde die Funktion „3D Plot“ implementiert. Diese dient lediglich der Veranschaulichung der Daten und hat keinen Einfluss auf die Korrelation. Das Fenster der Funktion „3D Plot“ ist in Abbildung 51 gezeigt. Die ausgewählten NMR-Spektren sind im linken Teil des Fensters dreidimensional dargestellt. Der Bereich der chemischen Verschiebung und die NMR-Spektren, die dargestellt werden sollen, können im rechten Teil des Fensters gewählt werden. Außerdem kann der Benutzer bestimmte EDCs anzeigen lassen.



**Abbildung 51: Fenster für die dreidimensionale Darstellung der NMR-Spektren. Der Benutzer hat außerdem die Möglichkeit EDCs anzeigen zu lassen.**

#### 4.2.1.4 Reduzierung der MS-Daten

Für die Korrelation von MS- und NMR-Daten müssen die MS-Daten (EICs) genauso viele Datenpunkte besitzen, wie NMR-Spektren vorhanden sind. Die EICs müssen daher in den Bereichen der einzelnen Fraktionen gemittelt werden. Hierfür muss der Zeitpunkt gewählt werden zu dem die erste Fraktion beginnt, dieser Zeitpunkt wird als „*waittime*“ bezeichnet.

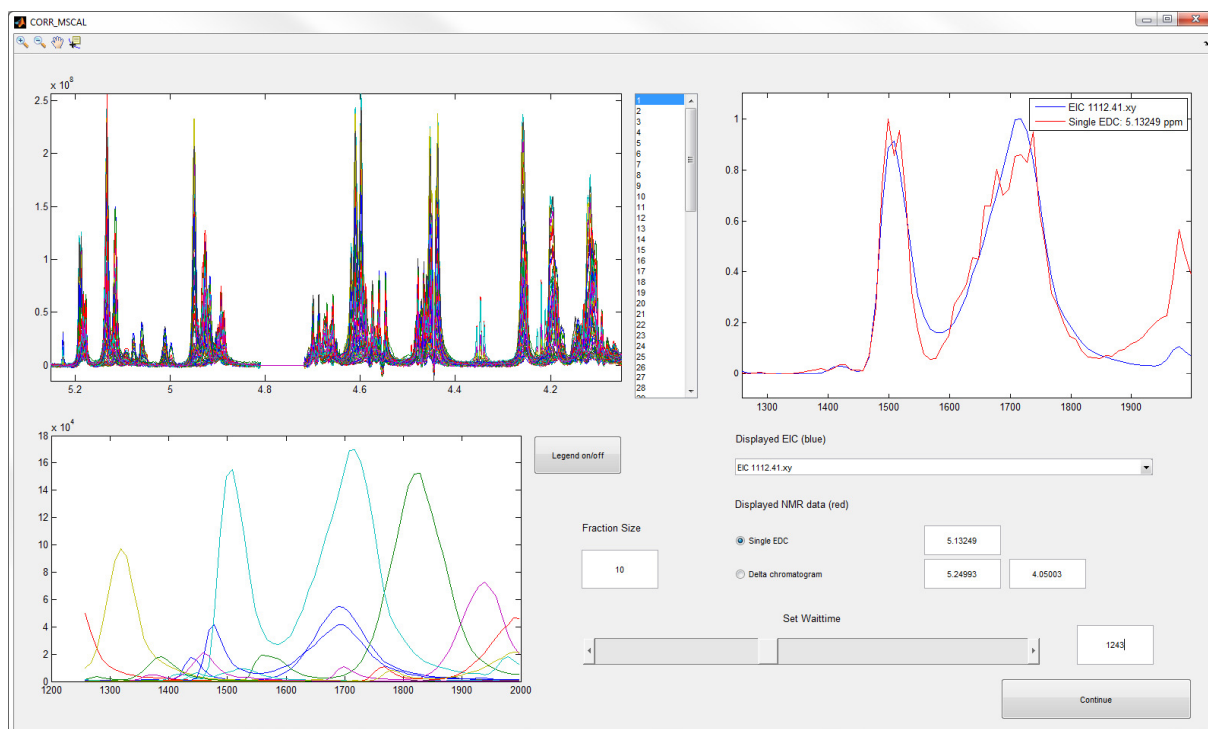


**Abbildung 52: Fenster zur Reduzierung der MS-Daten durch Optimierung.** Das Programm berechnet für alle Werte der *waittime* im angegebenen Bereich die Abweichung zwischen TDC und TIC. Die Kurve unten links stellt die Abweichung als Funktion der *waittime* dar. Die *waittime* mit der geringsten Abweichung ist mit dem roten Strich markiert. Das TIC und das TDC dieser *waittime* sind oben rechts dargestellt. Die neu berechneten EICs sind oben links gezeigt. Die *waittime* kann durch den Schieberegler verändert werden.

Das Programm bietet zwei Möglichkeiten die *waittime* zu bestimmen. Die erste Möglichkeit besteht darin die Abweichung zwischen *total ion*-Chromatogramm (TIC) und *total delta*-Chromatogramm (TDC) zu minimieren. Die Schaltfläche „Reduce MS data (Optimization)“ öffnet das in Abbildung 52 gezeigte Fenster. Hier kann ein Bereich und ein Intervall angegeben werden, in dem die *waittime* optimiert werden soll. Das Programm berechnet für jede *waittime* in diesem Bereich die reduzierten EICs sowie die Quadratsumme der Abweichung zwischen TIC und TDC und stellt diese im unteren linken Diagramm dar. Diejenige *waittime* mit der geringsten Abweichung wird automatisch ausgewählt und die entsprechenden EICs werden im oberen linken Diagramm dargestellt. Oben rechts sind das entsprechende TIC sowie das entsprechende TDC gezeigt. Durch den Schieberegler kann die

gewünschte *waittime* manuell verstellt werden, die Auswirkungen auf die EICs sowie das TIC sind sofort sichtbar. Es ist drauf zu achten, dass die Fraktionsgröße korrekt angegeben wird, da sonst falsche EICs berechnet werden. Die im unteren linken Diagramm von Abbildung 52 dargestellt Kurve der Abweichungen zwischen TIC und TDC gegen die *waittime* hat die Form einer Parabel. Um die optimale *waittime* zu finden kann die Kurve an eine quadratische Funktion angepasst werden, deren Minimum als *waittime* definiert wird.

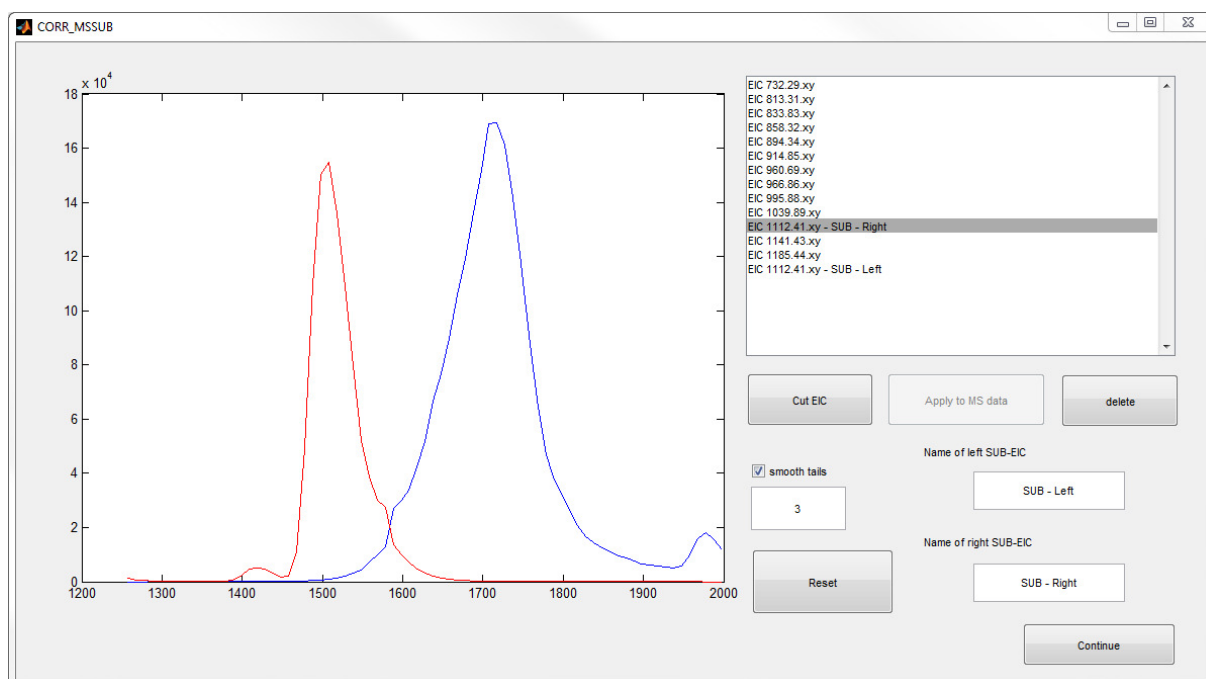
Die zweite Möglichkeit besteht darin die *waittime* vollständig manuell zu ermitteln. Hierfür muss der Anwender einem NMR-Signal ein EIC zuordnen können. Die Schaltfläche „Reduce MS data (manually)“ öffnet das in Abbildung 53 gezeigte Fenster. Im oberen rechten Diagramm dieses Fensters kann sich der Anwender ein EIC und ein EDC anzeigen lassen. Stammen das EIC und das EDC von derselben Substanz, sollten sie den gleichen Verlauf zeigen. Der Anwender kann nun die *waittime* manuell variieren und das EIC mit dem EDC vergleichen. Auf diese Weise kann die korrekte *waittime* ermittelt werden. Die chemische Verschiebung des angezeigten EDCs ist in den NMR-Spektren oben links markiert. Darüber hinaus kann ein Bereich der NMR-Spektren ausgewählt werden, der summiert und als *delta*-Chromatogramm dargestellt wird. Die reduzierten EICs sind unten links gezeigt.



**Abbildung 53:** Fenster zu Reduzierung der MS-Daten durch manuelles Einstellen der *waittime*. Der Benutzer hat die Möglichkeit ein TDC (oder einen Bereich der NMR-Spektren) auszuwählen, welches im oberen rechten Diagramm zusammen mit einem EIC dargestellt wird. Mit dem Schieberegler kann die *waittime* angepasst werden, sodass EIC und TDC übereinstimmen. Auf diese Weise kann die *waittime* manuell bestimmt werden.

#### 4.2.1.5 Erstellen von sub-EICs

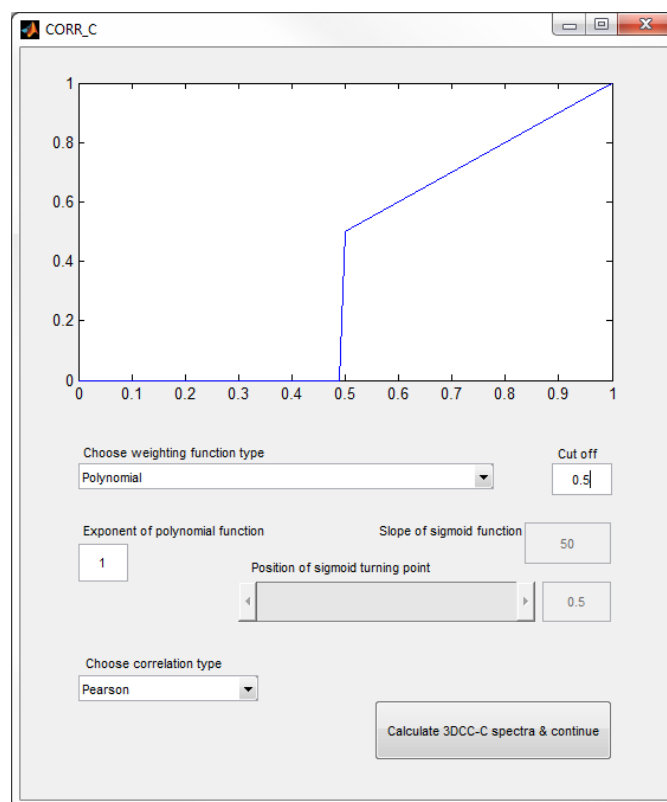
Sind in einem EIC die Signale mehrerer isobarer Bestandteile vorhanden, müssen aus diesem EIC mehrere sogenannte sub-EICs erstellt werden. Hierzu dient die Funktion „Build Sub-EICs“, das entsprechende Fenster ist in Abbildung 54 gezeigt. Der Benutzer kann ein EIC auswählen und an einer bestimmten Stelle schneiden, sodass zwei EICs entstehen. Dabei werden die Datenpunkte auf jeweils einer Seite der Schnittstelle auf Null gesetzt. Ist die Funktion „smooth tails“ aktiviert, werden auf diesen Seiten Flanken erzeugt.



**Abbildung 54:** Fenster für die Erstellung von sub-EICs. In diesem Fenster kann der Benutzer ein EIC zerschneiden. Die Datenpunkte auf jeweils einer Seite der Schnittstelle werden dabei auf Null gesetzt und zwei sub-EICs werden gebildet. Mit der Funktion „smooth tails“ werden an der Schnittstelle Flanken gebildet.

#### 4.2.1.6 Berechnung der Korrelationsspektren

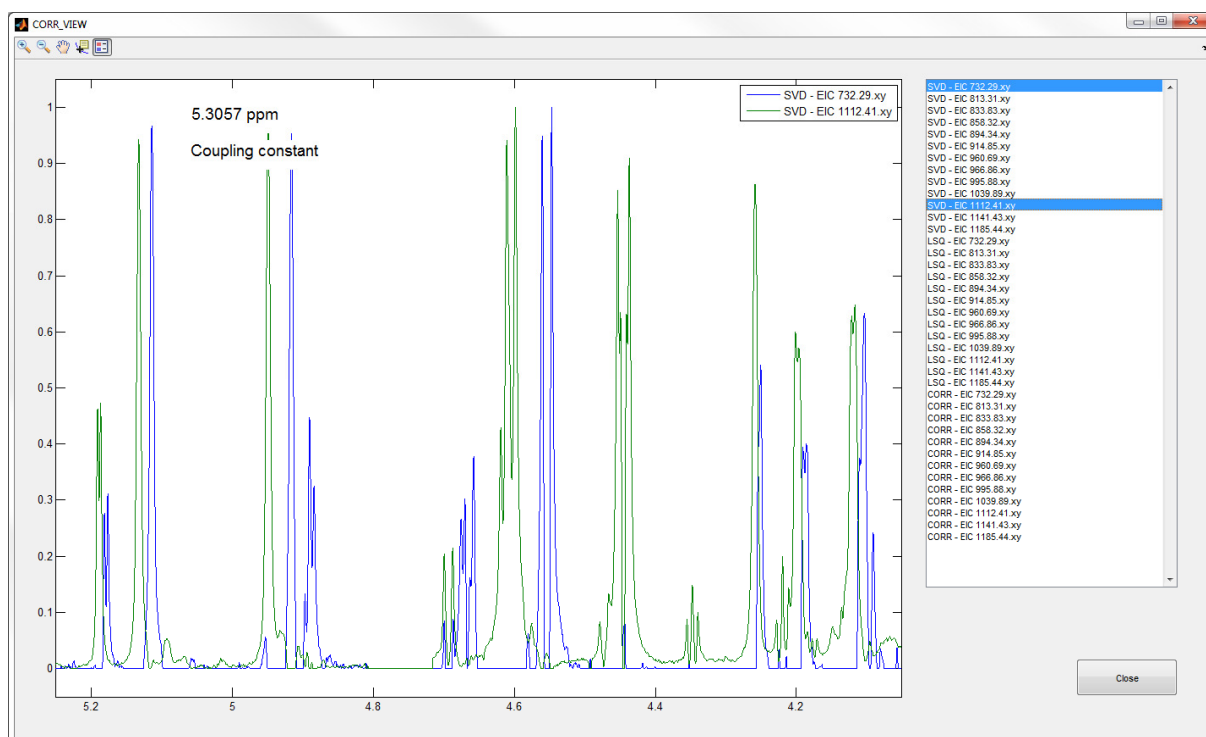
Sind alle Voraussetzungen erfüllt, werden die Schaltflächen zur Berechnung der Korrelationsspektren aktiv. Zur Berechnung von  $3DCC_S$ - und  $3DCC_L$ -Spektren sind keine weiteren Eingaben erforderlich. Bei der Berechnung von  $3DCC_C$ -Spektren kann eine Gewichtungsfunktion angewendet werden. Ein Klick auf die Schaltfläche „Calculate  $3DCC_C$  spectra“ öffnet das in Abbildung 55 gezeigte Fenster. Hier kann der Benutzer die Parameter der Gewichtungsfunktion angeben, es stehen Polynomial- und Sigmoidalfunktion zur Verfügung. Außerdem kann der Benutzer einen *cut-off*-Wert definieren. Korrelationskoeffizienten unterhalb dieses Wertes werden auf Null gesetzt.



**Abbildung 55:** Fenster zur Berechnung von  $3DCC_C$ -Korrelationsspektren. Für die Berechnung von  $3DCC_C$ -Korrelationsspektren muss eine Gewichtungsfunktion gewählt werden. Hierfür stehen Polynomial- und Sigmoidalfunktion zur Verfügung. Die Parameter können beliebig angepasst werden. Außerdem kann ein *cut-off*-Wert bestimmt werden unterhalb dessen alle Korrelationskoeffizienten auf Null gesetzt werden.

#### 4.2.1.7 Betrachtung der Korrelationsspektren

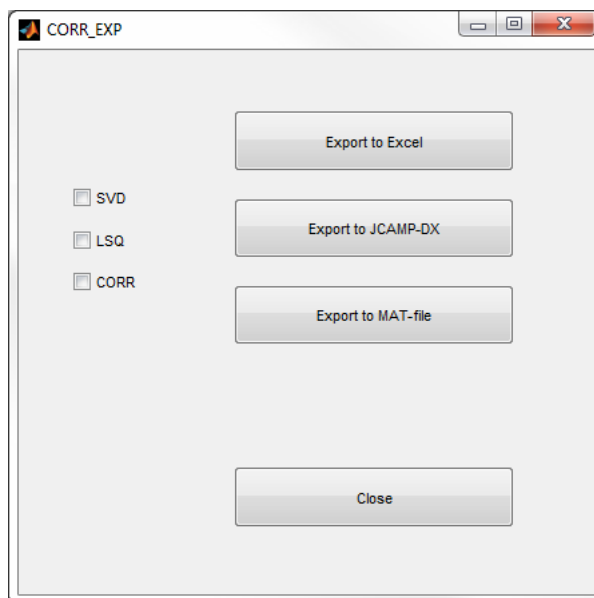
Die „View spectra“-Funktion bietet die Möglichkeit die Korrelationsspektren direkt zu betrachten. In Abbildung 56 ist das Fenster zur Betrachtung der Korrelationsspektren gezeigt. Rechts sind alle berechneten Korrelationsspektren aufgelistet, die ausgewählten Spektren werden links dargestellt. Die chemische Verschiebung an der Position des Mauszeigers wird im Spektrum angezeigt. Außerdem hat der Benutzer die Möglichkeit Abstände im Spektrum (z. B. Kopplungskonstanten) zu messen.



**Abbildung 56: Fenster für die Betrachtung der Korrelationsspektren. In diesem Fenster können alle berechneten Korrelationsspektren ausgewählt und betrachtet werden. Die chemische Verschiebung an der Position des Mauszeigers wird im Diagramm angezeigt. Außerdem kann ein Abstand (z. B. eine Kopplungskonstante) im Spektrum bestimmt werden.**

#### 4.2.1.8 Export der Korrelationsspektren

Für die Bearbeitung oder Auswertung der Korrelationsspektren können diese exportiert werden. Ein Klick auf die Schaltfläche „*Export results*“ öffnet das in Abbildung 57 gezeigte Fenster. Hier kann der Benutzer die Korrelationsspektren als Excel-Datei, JCAMP-DX-Datei oder als MAT-Datei exportieren. Die JCAMP-DX-Datei kann dann in TopSpin geöffnet werden.



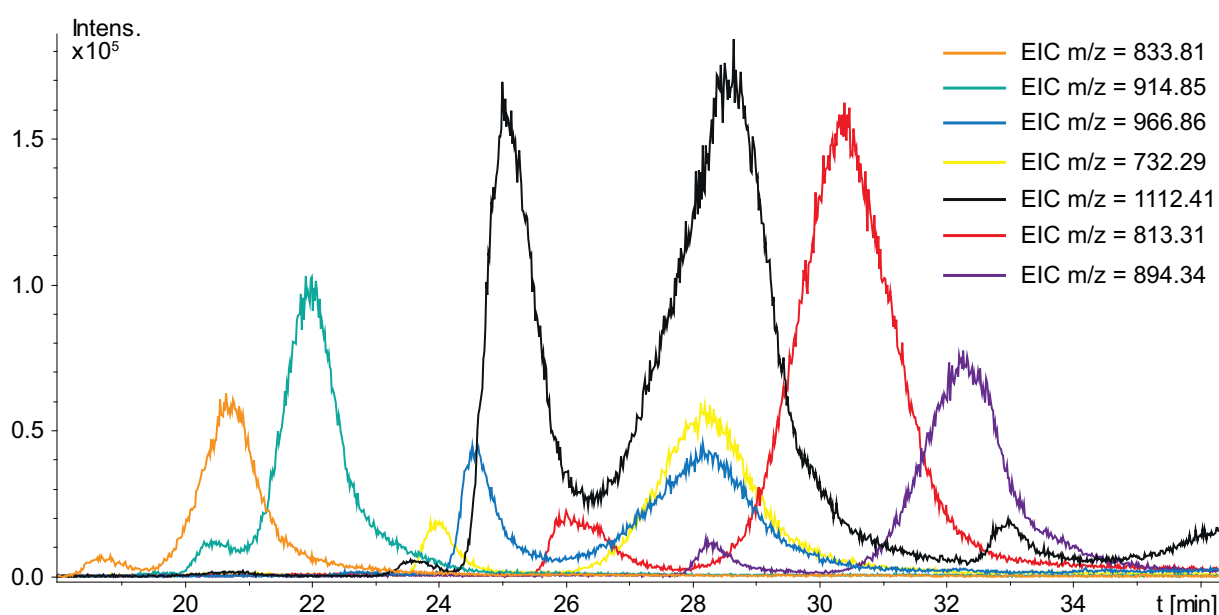
**Abbildung 57:** Fenster für den Export der Korrelationsspektren. Die berechneten Korrelationsspektren können als Excel-Datei, JCAMP-DX-Dateien oder als MAT-Datei exportiert werden. Die JCAMP-DX-Dateien können in TopSpin geöffnet werden.

#### 4.2.1.9 Speichern und Laden von Daten

Über die Schaltfläche „*Menu*“ im TDCC Hauptmenü und die Funktionen „*Save project as*“ und „*Load project*“ könne die Daten jederzeit gespeichert und geladen werden. So kann der Benutzer die Daten nach dem Import oder nach der Bearbeitung der NMR-Spektren speichern und muss diese Schritte nicht wiederholen, wenn beispielsweise eine andere *waittime* ausprobiert werden soll.

#### 4.2.2 Analyse der *N*-Glycane aus humanem Plasma mittels 3DCC

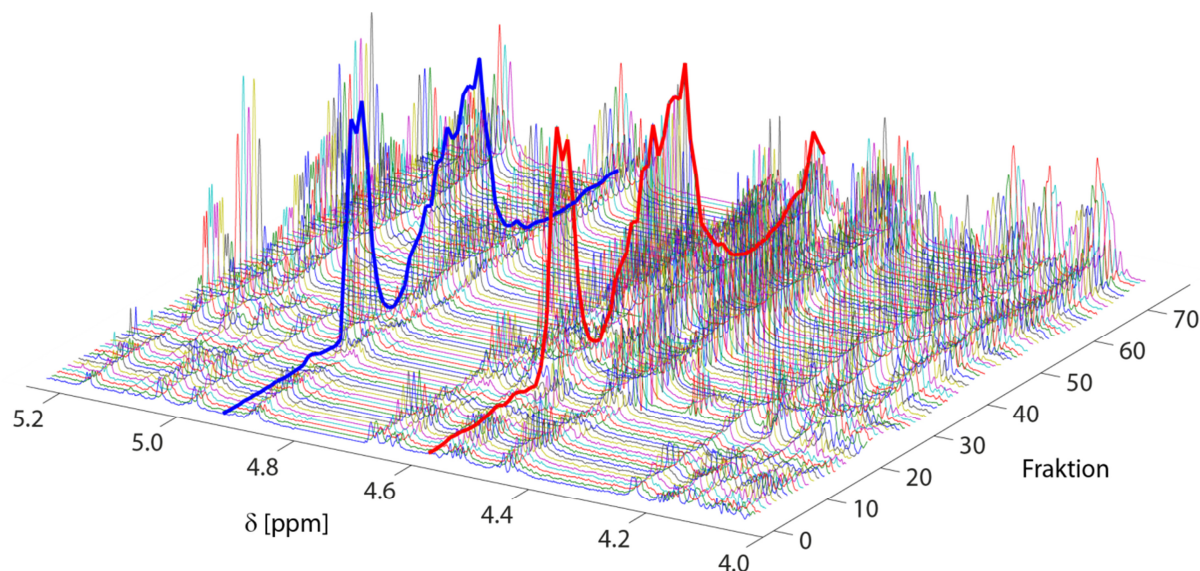
Bei der Entwicklung und zum Testen des *User Interface* wurde der im Folgenden diskutierte Datensatz verwendet. Hierfür wurden die *N*-Glycane aus humanem Plasma untersucht. Nach der Abspaltung und Reinigung aller *N*-Glycane aus humanem Plasma wurden diese mittels HPLC auf einer *porous graphitized carbon*-Säule getrennt (siehe Abschnitt 7.8). Die sieben *extracted ion*-Chromatogramme, für welche Korrelationspektren erhalten werden konnten, sind in Abbildung 58 gezeigt.



**Abbildung 58:** *Extracted ion*-Chromatogramme der Trennung von *N*-Glycanen aus humanem Plasma auf einer *porous graphitized carbon*-Säule. Die Signale zeigen starke Überlappung. Die dreidimensionale Kreuzkorrelation kann in diesem Fall reine Spektren der einzelnen Bestandteile berechnen. Es sind nur diejenigen EICs gezeigt, für die gute Korrelationspektren erhalten wurden.

Um die Trennung von  $\alpha$ - und  $\beta$ -Anomeren bei der Chromatographie zu vermeiden, wurde dem Laufmittel 50 mM Ammoniak zugesetzt. Dies führt allerdings auch zu einer Verbreiterung der Signale. Die Signale in Abbildung 58 zeigen daher deutliche Überlagerung. In einem solchen Fall ist die dreidimensionale Kreuzkorrelation eine geeignete Methode um reine NMR-Spektren der einzelnen Bestandteile zu erhalten. Zu diesem Zweck wurde der Fluss der Flüssigkeitschromatographie geteilt und das Eluat wurde in Fraktionen zu je 10 Sekunden aufgefangen. 75 Fraktionen wurden NMR-spektroskopisch untersucht, Ausschnitte aus den NMR-Spektren sind in Abbildung 59 dreidimensional dargestellt.

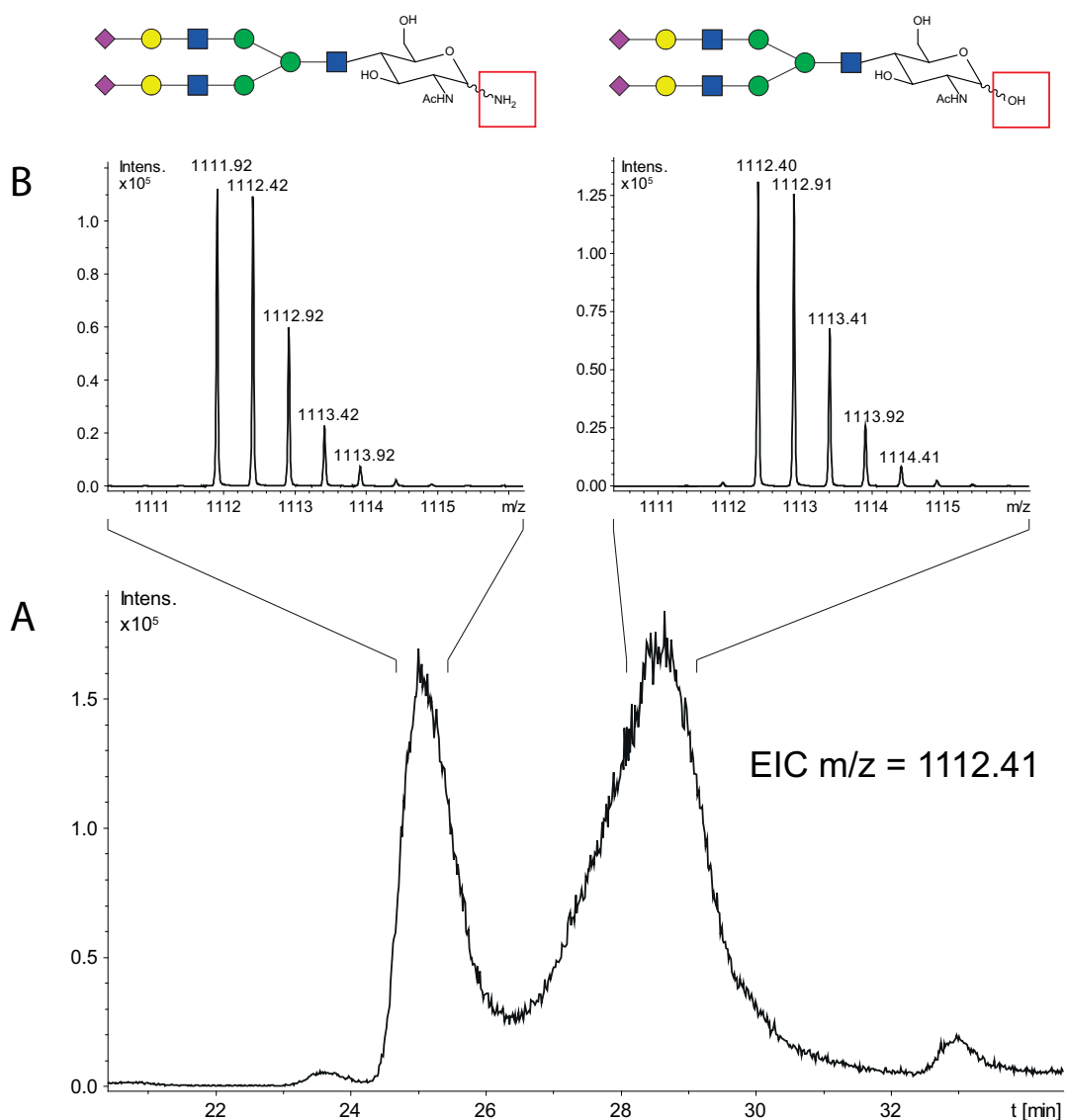




**Abbildung 59: Dreidimensionale Darstellung aller 75 NMR-Spektren im Bereich 4.05 – 5.25 ppm. Die EDCs mit den chemischen Verschiebungen 4.949 ppm und 4.598 ppm sind in Blau und Rot dargestellt.**

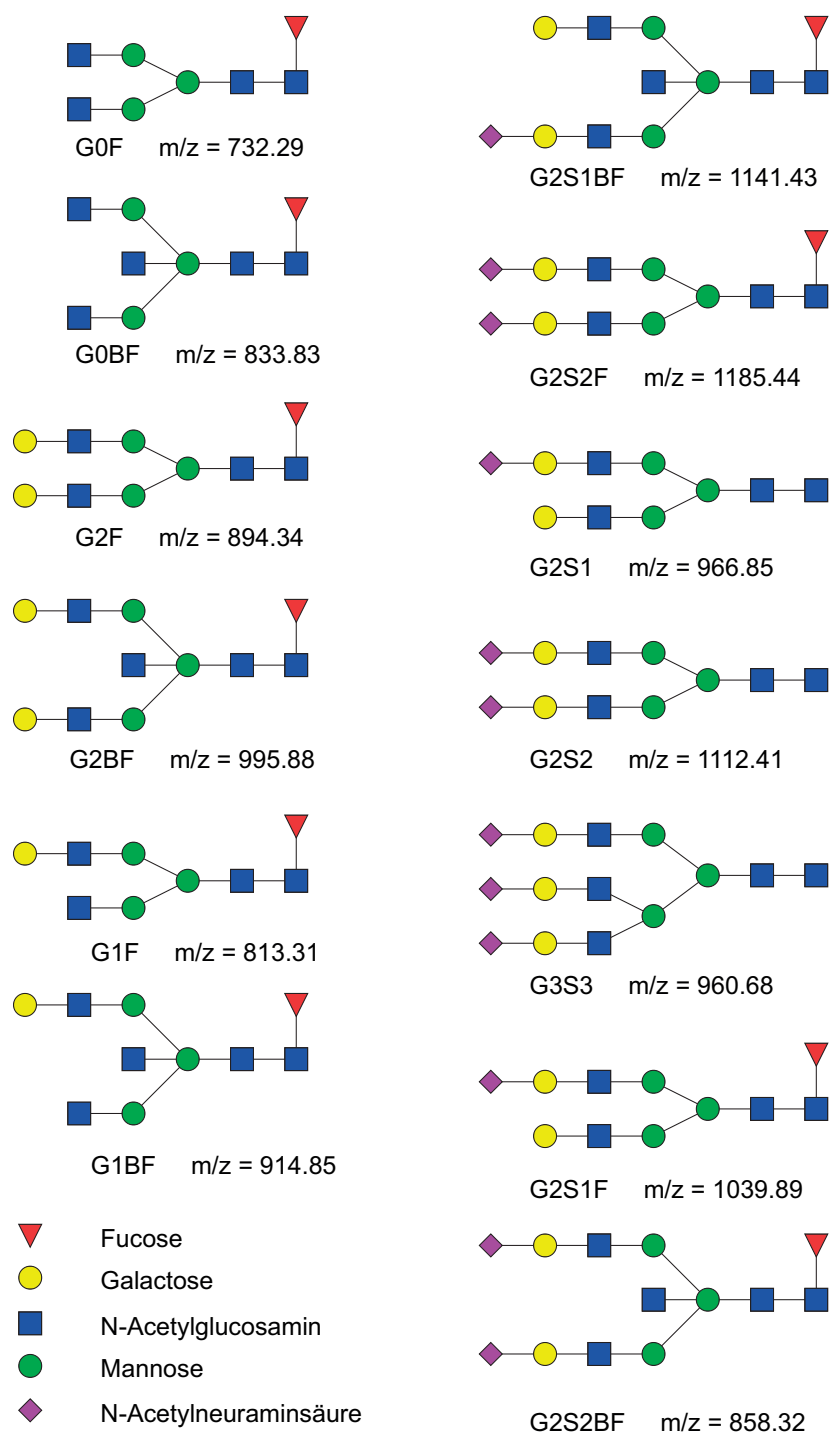
In Abbildung 60 ist das EIC des abundantesten Glycans ( $\text{Hex}_5\text{HexNAC}_4\text{NeuAc}_2$ ) mit dem  $m/z$ -Verhältnis 1112.41 gezeigt. Das EIC weist zwei Signale bei etwa 25 und 29 Minuten auf. Die Massenspektren zeigen allerdings, dass nur das rechte Signal bei 29 Minuten eine monoisotopische Masse von 1112.41 besitzt. Das linke Signal bei 25 Minuten hat eine monoisotopische Masse von  $m/z = 1111.92$ , sodass das erste Isotopensignal des zweifach geladenen Zustandes auf  $m/z = 1112.42$  fällt. Bei dem linken Signal handelt es sich um die 1-Amino Form des Glycans, welche von der PNGaseF freigesetzt wird und sich im leicht alkalischen nur langsam in die 1 Dalton schwerere 1-Hydroxy Form umwandelt. In diesem LC-Lauf hatte sich die 1-Amino Form noch nicht vollständig in die 1-Hydroxy Form umgewandelt, sodass für alle Glycane zwei Signale zu beobachten sind. Da vor der Messung der NMR-Spektren genügend Zeit vergangen war, zeigen diese nur die 1-Hydroxy Form.

Für die dreidimensionale Kreuzkorrelation stellt die Tatsache, dass alle EICs zwei Signale aufweisen kein Problem dar. Im Gegenteil, die zusätzlichen Signale sorgen für eine bessere Differenzierung zwischen den einzelnen Bestandteilen. Beispielsweise wären die drei Glycane mit den  $m/z$ -Verhältnissen 1112.41 (schwarz,  $\text{Hex}_5\text{HexNAC}_4\text{NeuAc}_2$ ), 732.29 (gelb,  $\text{Hex}_3\text{HexNAC}_4\text{Fuc}_1$ ) und 966.86 (blau,  $\text{Hex}_5\text{HexNAC}_4\text{NeuAc}_1$ ) anhand des rechten Signals der EICs nicht zu trennen gewesen. Das linke Signal der 1-Amino Form zeigt aber deutliche Unterschiede in der Retentionszeit, sodass für alle drei Glycane gute Korrelationspektren berechnet werden konnten.



**Abbildung 60:** Das EICs des Glycans Hex<sub>5</sub>HexNAc<sub>4</sub>NeuAc<sub>2</sub> mit einem m/z-Verhältnis von 1112.41 zeigt zwei Signale bei etwa 25 und 29 Minuten (A). Die Massenspektren (B) zeigen, dass das linke Signal eine monoisotopische Masse von m/z = 1111.92 besitzt, sodass das erste Isotopensignal des zweifach geladenen Zustand auf 1112.42 fällt und von dem EIC erfasst wird. Das linke Signal stammt von der um 1 Da leichteren 1-Amino Form des N-Glycans, welche von der PNGaseF freigesetzt wird und sich nur langsam in die 1-Hydroxy Form umwandelt.

Für die dreidimensionale Kreuzkorrelation wurden die EICs der 13 in Abbildung 61 gezeigten N-Glycane verwendet (In Abbildung 58 sind nur die EICs der Glycane gezeigt, für die gute Korrelationsspektren erhalten werden konnten). Die besten Ergebnisse lieferte die Methode 3DCCs, für sieben Glycane wurden gute Korrelationsspektren erhalten, diese sind in den Abbildungen 62 – 67 gezeigt. (Die Spektren der Glycane GoF und G2F sind zweimal gezeigt, um die Unterschiede zwischen den Spektren zu verdeutlichen.) Die jeweiligen Strukturen sowie die Vergleichsdaten für die chemischen Verschiebungen der *structural reporter group*-Signale sind rechts neben den Spektren gezeigt.<sup>126-130</sup>



**Abbildung 61: Strukturen der 13 N-Glycane deren EICs für die dreidimensionale Kreuzkorrelation verwendet wurden. Für die sieben N-Glycane, deren EICs in Abbildung 58 gezeigt sind, wurden gute Korrelationspektren erhalten.**

In Abbildung 62 sind Ausschnitte im Bereich 4.05 – 5.25 ppm der  $^3\text{DCC}_s$ -Korrelationspektren der Glycane GoF ( $m/z = 732.29$ ), GoBF ( $m/z = 833.83$ ) und G1BF ( $m/z = 914.85$ ) gezeigt. Die Signale von GlcNAc-1 H-1 (5.180 ppm), GlcNAc-2 H-1 (4.658/4.667) und Fuc H-1 (4.898/4.907) sind in allen drei Spektren vorhanden. Das H-1 Signal der Mannose-3 liegt in den *Bisec*-Strukturen bei 4.687 (G1BF) bzw. 4.697 ppm

(GoBF). Ist kein *Bisec*-GlcNAc vorhanden, befindet sich das Signal bei 4.773 ppm und ist nicht zu erkennen, da es im Bereich der Wasserunterdrückung liegt. Die H-1 Signale der Mannosen 4 und 4' liegen in GoF bei 5.114 bzw. 4.916 ppm. Das *Bisec*-GlcNAc hat großen Einfluss auf die chemische Verschiebung dieser Signale. Das Signal vom Man-4 H-1 liegt in GoBF und G1BF bei ca. 5.058 ppm, das Signal von Man-4' H-1 liegt in GoBF bei 4.998 ppm und in G1BF aufgrund der zusätzlichen Galactose bei 5.011 ppm.

Für das Glycan GoF weisen die H-1 Signale von GlcNAc-5 und 5' dieselbe chemische Verschiebung von 4.554 ppm auf. Im Spektrum von GoBF sind zwei leicht verschobene Signale bei 4.546 und 4.555 ppm zu erkennen, da das *Bisec*-GlcNAc einen kleinen Einfluss auf die chemische Verschiebung des H-1 von GlcNAc-5' hat. Im Spektrum von G1BF liegt das H-1 Signal von GlcNAc-5 wie erwartet bei 4.556 ppm, das H-1 Signal von GlcNAc-5' liegt durch den Einfluss der zusätzlichen Galactose bei 4.579 ppm. Das H-1 Signal des *Bisec*-GlcNAc liegt in den Spektren von GoBF und G1BF bei 4.464 ppm und ist im Spektrum von GoF wie erwartet nicht vorhanden. Im Spektrum von G1BF ist zusätzlich das Signal vom Gal-6' H-1 bei 4.478 ppm vorhanden, welches sich mit dem H-1 Signal des *Bisec*-GlcNAc überlagert.

Die H-2 Signale der *core*-Mannosen befinden sich am Rand des *bulk*-Bereiches. Im Korrelationsspektrum von GoF sind die H-2 Signale von Man-3 (4.250 ppm) Man-4 (4.188 ppm) und Man-4' (4.104 ppm) gut zu erkennen. In den Spektren von GoBF und G1BF sind in diesem Bereich viele Störsignale vorhanden. Das H-2 Signal von Man-4 befindet sich in den Spektren der *Bisec*-Strukturen bei etwa 4.250 ppm, während das H-2 Signal der Man-3 bei ca. 4.175 ppm zu finden ist. Das H-2 Signal von Man-4' liegt in Spektrum von G1BF bei 4.142 ppm und im Spektrum von GoBF bei 4.149 ppm.

In Abbildung 63 sind Ausschnitte im Bereich 1.95 – 2.2 ppm der Korrelationsspektren der Glycane GoF ( $m/z = 732.29$ ) GoBF ( $m/z = 833.83$ ) und G1BF ( $m/z = 914.85$ ) gezeigt. In diesem Bereich befinden sich die Signale der *N*-Acetylgruppen von *N*-Acetylglucosaminen und *N*-Acetylneuraminsäuren. Im Korrelationsspektrum des Glycans GoF sind drei Signale zu erkennen: Das NAc-Signal des proximalen GlcNAc bei 2.039 ppm, das NAc-Signal der beiden terminalen *N*-Acetylglucosamine 5 und 5' bei 2.052 ppm sowie das NAc-Signal von GlcNAc-2 bei 2.095 ppm. Im Korrelationsspektrum von GoBF ist zusätzlich das NAc-Signal des *Bisec*-GlcNAc bei 2.065 ppm vorhanden. Außerdem sind die NAc-Signale der terminalen *N*-Acetylglucosamine durch den Einfluss des *Bisec*-GlcNAc etwas verschoben (2.050 ppm bzw. 2.058 ppm). Das mit einem Sternchen markierte Signal kann nicht erklärt werden. Bei dem Glycan G1BF ist das NAc-Signal von GlcNAc-5' durch den Einfluss der Galactose zu 2.039 verschoben und überlagert mit dem NAc-Signal des proximalen GlcNAc.

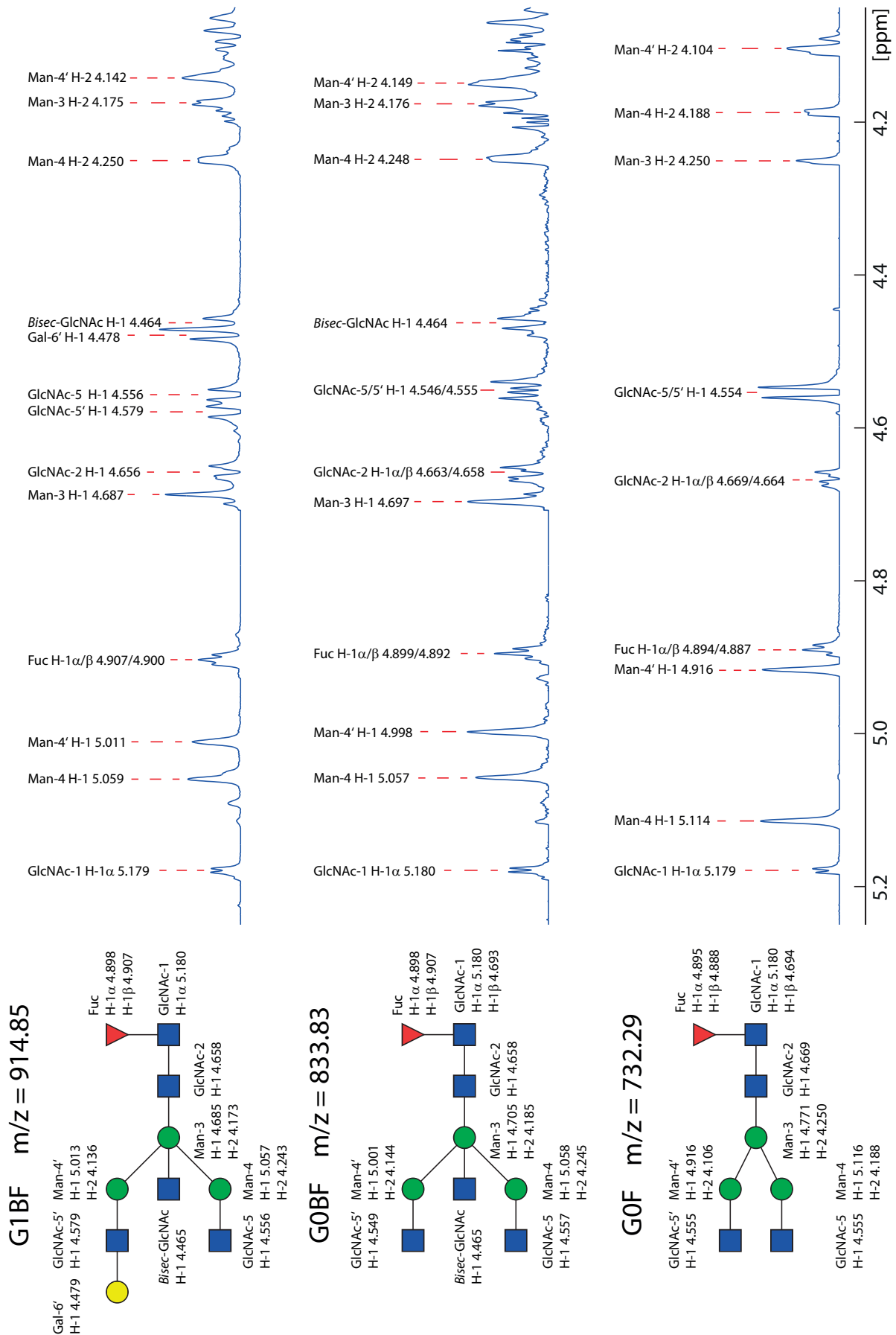
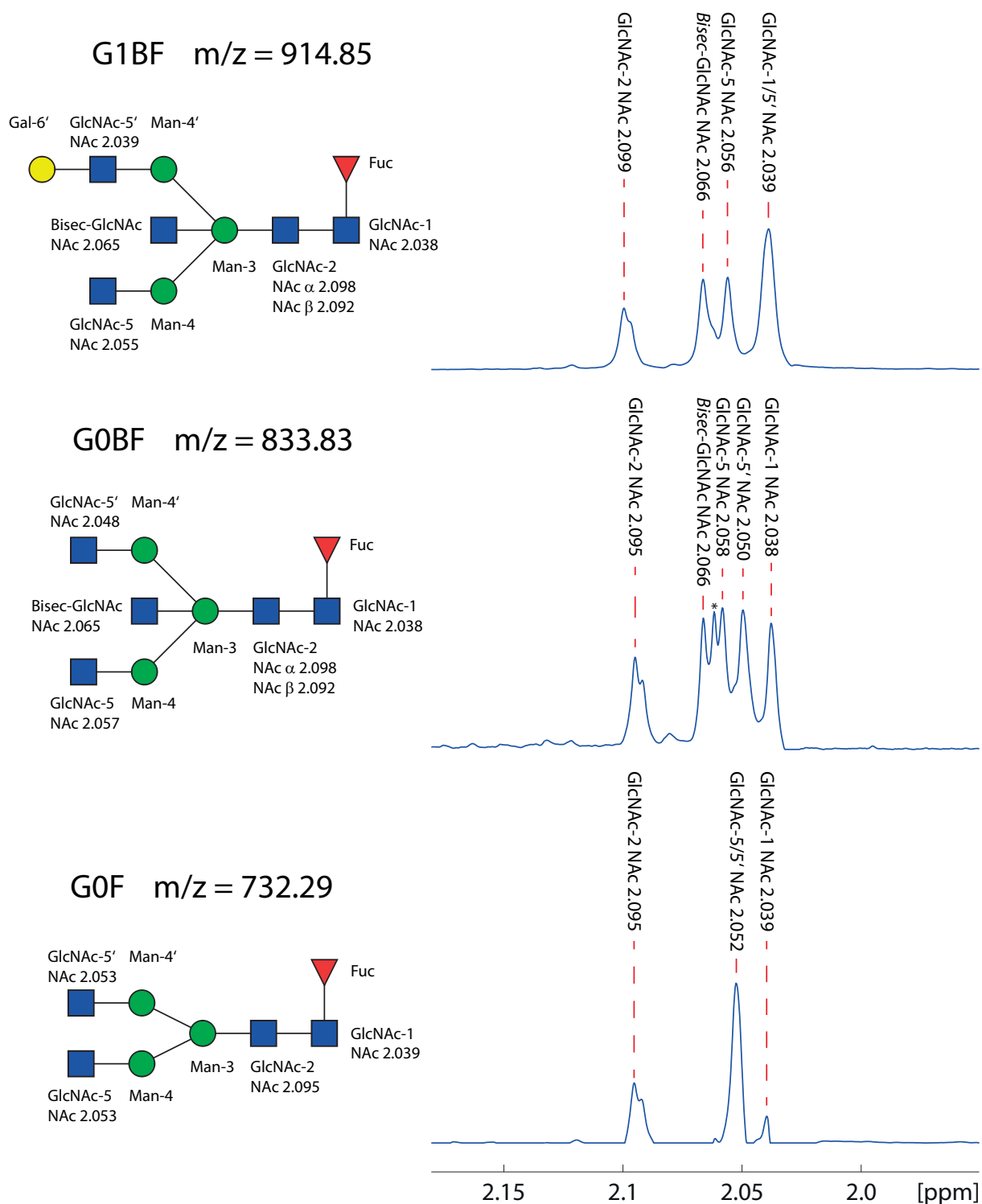


Abbildung 62: 3DCCs Korrelationsspektren der Glycane GoF ( $m/z = 732.29$ ), GoBF ( $m/z = 833.83$ ) und G1BF ( $m/z = 914.85$ ) im Bereich 4.05 – 5.25 ppm.



**Abbildung 63:** 3DCCs Korrelationspektren der Glycane GoF ( $m/z = 732.29$ ), GoBF ( $m/z = 833.83$ ) und G1BF ( $m/z = 914.85$ ) im Bereich 1.95 – 2.20 ppm.

Die Ausschnitte im Bereich 4.05 – 5.25 ppm der 3DCCs-Korrelationspektren der Glycane GoF ( $m/z = 732.29$ ) G1F ( $m/z = 813.31$ ) und G2F ( $m/z = 894.34$ ) sind in Abbildung 64 gezeigt. Das Korrelationspektrum von GoF, welches bereits erläutert wurde, ist hier zum Vergleich nochmal dargestellt. Die Signale von GlcNAc-1, GlcNAc-2, Man-4 und der Fucose

sind in den drei Spektren weitestgehend identisch. Die H-1 Signale der terminalen *N*-Acetylglucosamine-5 und 5' im Glycan GoF befinden sich bei 4.554 ppm. Im Spektrum des Glycans G1F ist nur noch ein Signal bei 4.554 ppm vorhanden, da das Signal des zweiten GlcNAc durch den Effekt der Galactose zu 4.581 ppm verschoben ist. Außerdem ist das H-1 Signal der Mannose-4' bei 4.925 ppm zu erkennen. Die Verschiebung im Vergleich zum Glycan GoF (Man-4' H-1 4.916 ppm) deutet darauf hin, dass die zusätzliche Galactose dominant am  $\alpha$ 1-6 verknüpften Arm lokalisiert ist. Die H-1 Signale der terminalen Galactosen befinden sich bei 4.466 bzw. 4.471 ppm. Im Korrelationsspektrum von G1F ist nur das Signal von Gal-6' bei 4.473 ppm zu erkennen, während im Korrelationsspektrum von G2F zwei überlappende Signale bei 4.467 und 4.473 ppm zu erkennen sind.

Der Bereich der *N*-Acetylsignale aus den Korrelationsspektren der Glycane GoF ( $m/z = 732.29$ ) G1F ( $m/z = 813.31$ ) und G2F ( $m/z = 894.34$ ) ist in Abbildung 65 gezeigt. Die NAc-Signale von GlcNAc-1 und GlcNAc-2 befinden sich in allen Spektren bei 2.039 bzw. 2.097 ppm. Im Korrelationsspektrum von GoF liegen die NAc-Signale der beiden terminalen *N*-Acetylglucosamine bei 2.052 ppm. Der Effekt der zusätzlichen Galactosen auf die chemische Verschiebung dieser Signale ist zwar gering, trotzdem sind in den Spektren von G1F und G2F jeweils zwei Signale zu erkennen.

Die Ausschnitte im Bereich 4.05 – 5.25 ppm aus den Korrelationsspektren der Glycane G2F ( $m/z = 894.34$ ), G2S1 ( $m/z = 966.86$ ) und G2S2 ( $m/z = 1112.41$ ) sind in Abbildung 66 gezeigt. In den Korrelationsspektren der nicht fucosylierten Glycane G2S1 und G2S2 ist das H-1 Signal der Fucose, welches sich im Spektrum von G2F bei 4.897/4.890 ppm befindet, nicht vorhanden. Außerdem hat die Fucose einen Einfluss auf die chemische Verschiebung der H-1 Signale von GlcNAc-1 und GlcNAc-2. In der  $\alpha$ -Form des fucosylierten Glycans befindet sich das H-1 Signal von GlcNAc-1 bei 5.180 ppm. Im nicht fucosylierten Glycan befindet sich dieses Signal bei 5.189 ppm. Erstaunlicherweise hat das Vorhandensein der Fucose in der  $\beta$ -Form kaum Einfluss auf die chemische Verschiebung des H-1 von GlcNAc-1. Die H-1 Signale von GlcNAc-2 befindenden sich im Korrelationsspektrum von G2F bei 4.668/4.664 ppm. In den Spektren der nicht fucosylierten Strukturen G2S1 und G2S2 befinden sich diese Signale bei ca. 4.606/4.614 ppm und sind durch Überlagerung mit den H-1 Signalen der *N*-Acetylglucosamine-5 und 5' nur schwer zu erkennen.

Im Korrelationsspektrum von G2F liegen die H-1 Signale von Man-4 und Man-4' bei 5.118 bzw. 4.926 ppm. Im Spektrum von G2S1 befinden sich diese Signale bei 5.136 bzw. 4.932 ppm. Die deutliche Verschiebung des H-1 Signals von Man-4 deutet darauf hin, dass sich die *N*-Acetylneuraminsäure dominant am  $\alpha$ 1-3 verknüpften Arm befindet. Ist wie bei G2S2 auch am  $\alpha$ 1-6 verknüpften Arm eine *N*-Acetylneuraminsäure vorhanden, liegt das H-1

Signal von Man-4' bei 4.949 ppm. Die *N*-Acetylneuraminsäuren haben außerdem großen Einfluss auf die chemische Verschiebung der *N*-Acetylglucosamine-5 und 5' sowie der Galactosen-6 und 6'. Im Korrelationsspektrum von G2F befinden sich die H-1 Signale von GlcNAc-5 und 5' bei 4.581 ppm. Im Spektrum des disialylierten Glycans G2S2 sind diese Signale bei 4.604 ppm zu finden, während im Spektrum des monosialylierten Glycans G2S1 sowohl bei 4.582 ppm als auch bei 4.607 ppm ein Signal zu erkennen ist. Die H-1 Signale der Galactosen-6 und 6' befinden sich im Korrelationsspektrum von G2F bei 4.467 bzw. 4.473 ppm. Im Spektrum von S2S2 sind die beiden Signale aufgrund der zusätzlichen *N*-Acetylneuraminsäuren zu 4.444 bzw. 4.448 ppm verschoben. Das Spektrum der monosialylierten Struktur G2S1 weist zwei Signale bei 4.473 ppm (Gal-6' H-1) und 4.445 ppm (Gal-6 H-1) auf.

Der Bereich der *N*-Acetylsignale aus den Korrelationsspektren der Glycane G2F ( $m/z = 894.34$ ), G2S1 ( $m/z = 966.86$ ) und G2S2 ( $m/z = 1112.41$ ) ist in Abbildung 67 gezeigt. Das NAc-Signal von GlcNAc-2 ist in den Spektren der fucosylierten Strukturen bei 2.097 ppm zu finden. In den Spektren der nicht fucosylierten Strukturen G2S1 und G2S2 befindet sich das NAc-Signal von GlcNAc-2 bei 2.081 bzw. 2.085 ppm. Die chemische Verschiebung des NAc-Signals von GlcNAc-1 ist weitestgehend unabhängig von der *core*-Fucose und liegt in allen Spektren bei etwa 2.039 ppm. Die NAc-Signale von GlcNAc-5 und 5' befinden sich im Spektrum der nicht sialylierten Struktur G2F bei 2.048 bzw. 2.051 ppm. Im Spektrum des disialylierten Glycans G2S2 sind diese Signale bei 2.070 bzw. 2.066 ppm zu finden. Im Spektrum des monosialylierten Glycans G2S1 liegt das NAc-Signal von Gal-6 bei 2.070 ppm, da diese sialyliert ist, während das NAc-Signal der unsialylierten Gal-6' bei 2.048 zu finden ist. Das *N*-Acetylsignal der *N*-Acetylneuraminsäuren ist in den Korrelationsspektren der Glycane G2S1 und G2S2 bei 2.030 ppm zu erkennen, im Spektrum von G2S2 ist das Signal erwartungsgemäß deutlich größer als im Spektrum von G2S1, da zwei *N*-Acetylneuraminsäuren vorhanden sind.



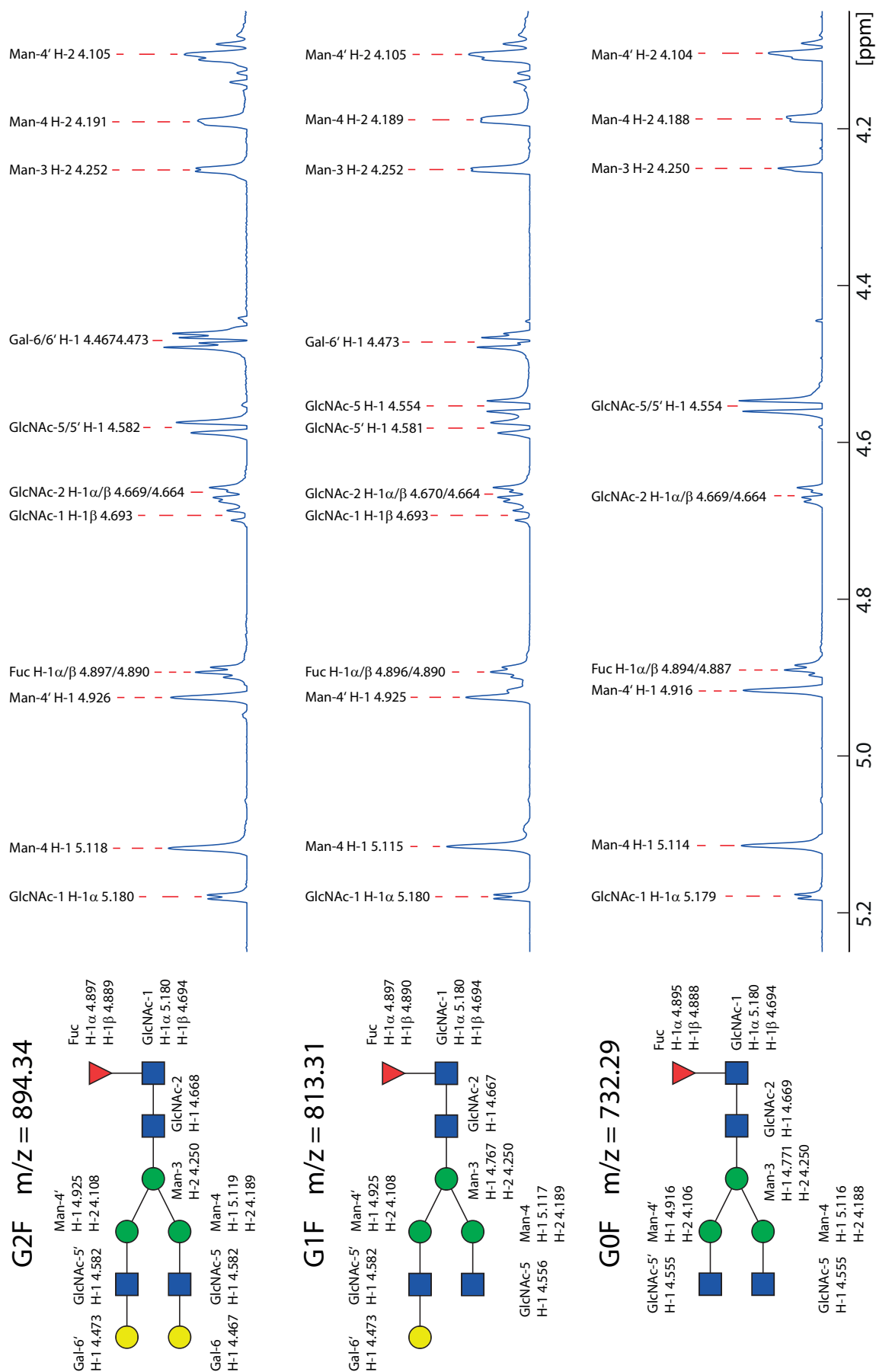
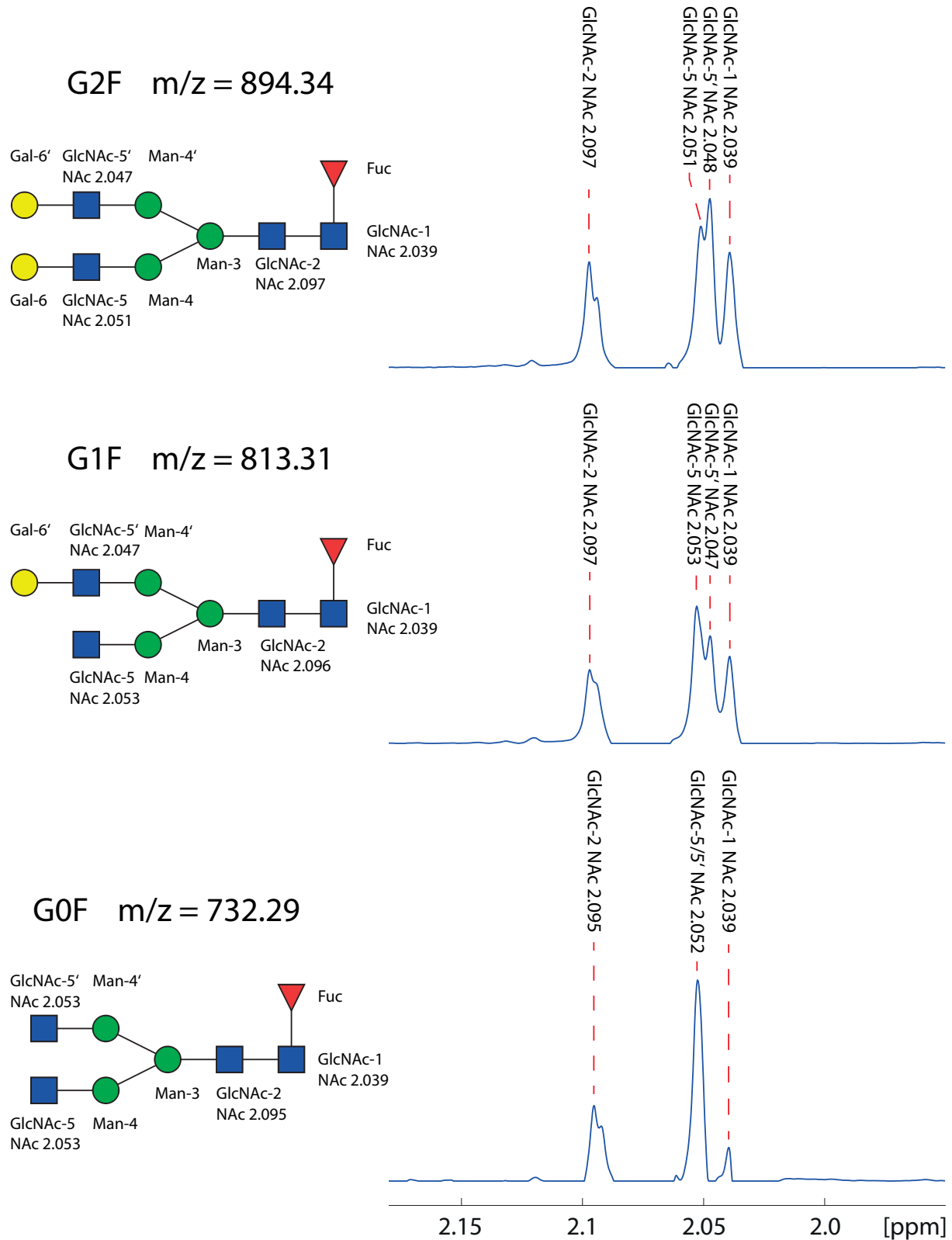


Abbildung 64: 3DCCs Korrelationspektren der Glycane GoF ( $m/z = 732.29$ ), G1F ( $m/z = 813.31$ ) und G2F ( $m/z = 894.34$ ) im Bereich 4.05 – 5.25 ppm.



**Abbildung 65:** 3DCCs Korrelationspektren der Glycane G0F ( $m/z = 732.29$ ), G1F ( $m/z = 813.21$ ) und G12F ( $m/z = 894.34$ ) im Bereich 1.95 – 2.20 ppm.

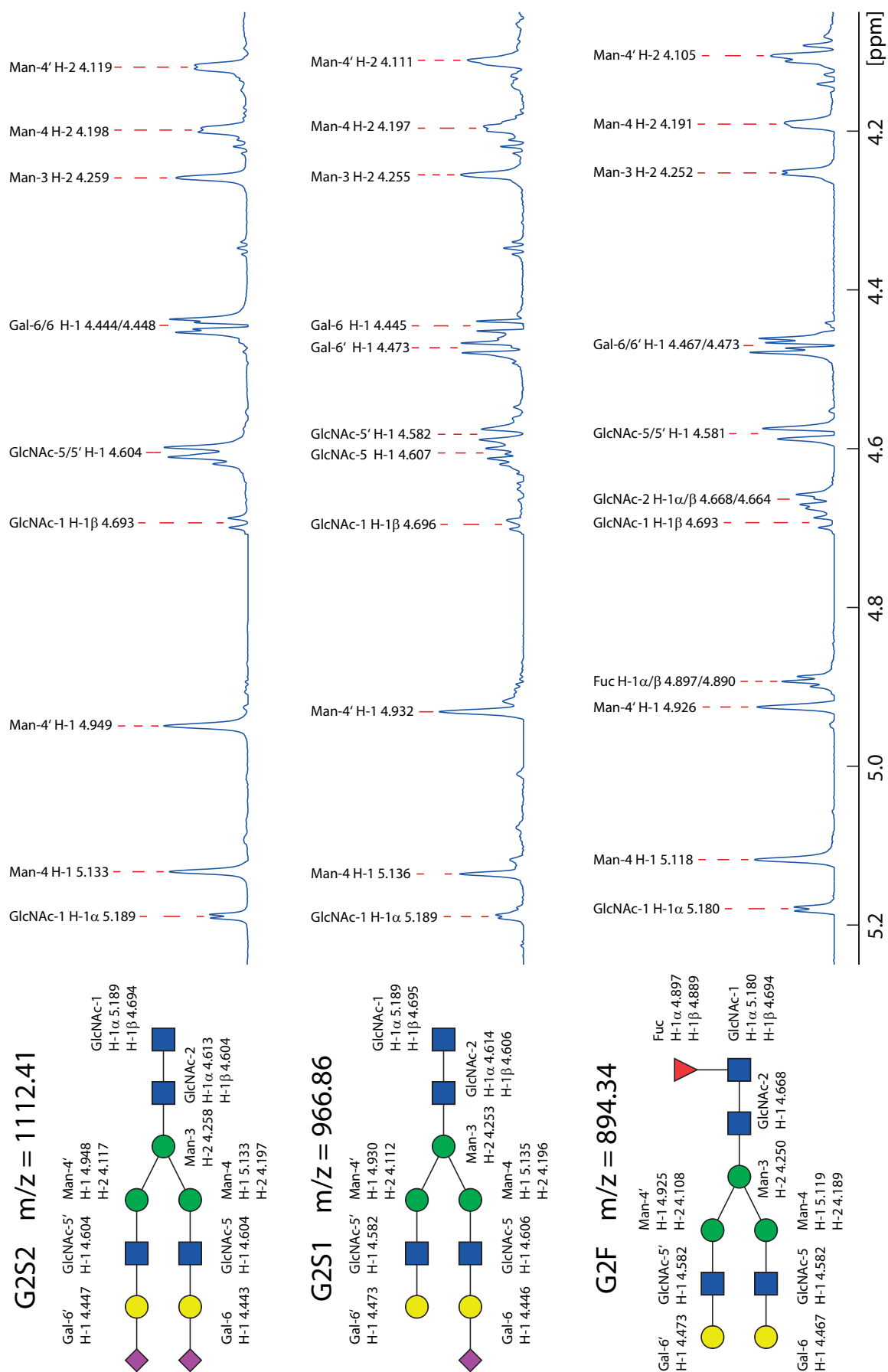
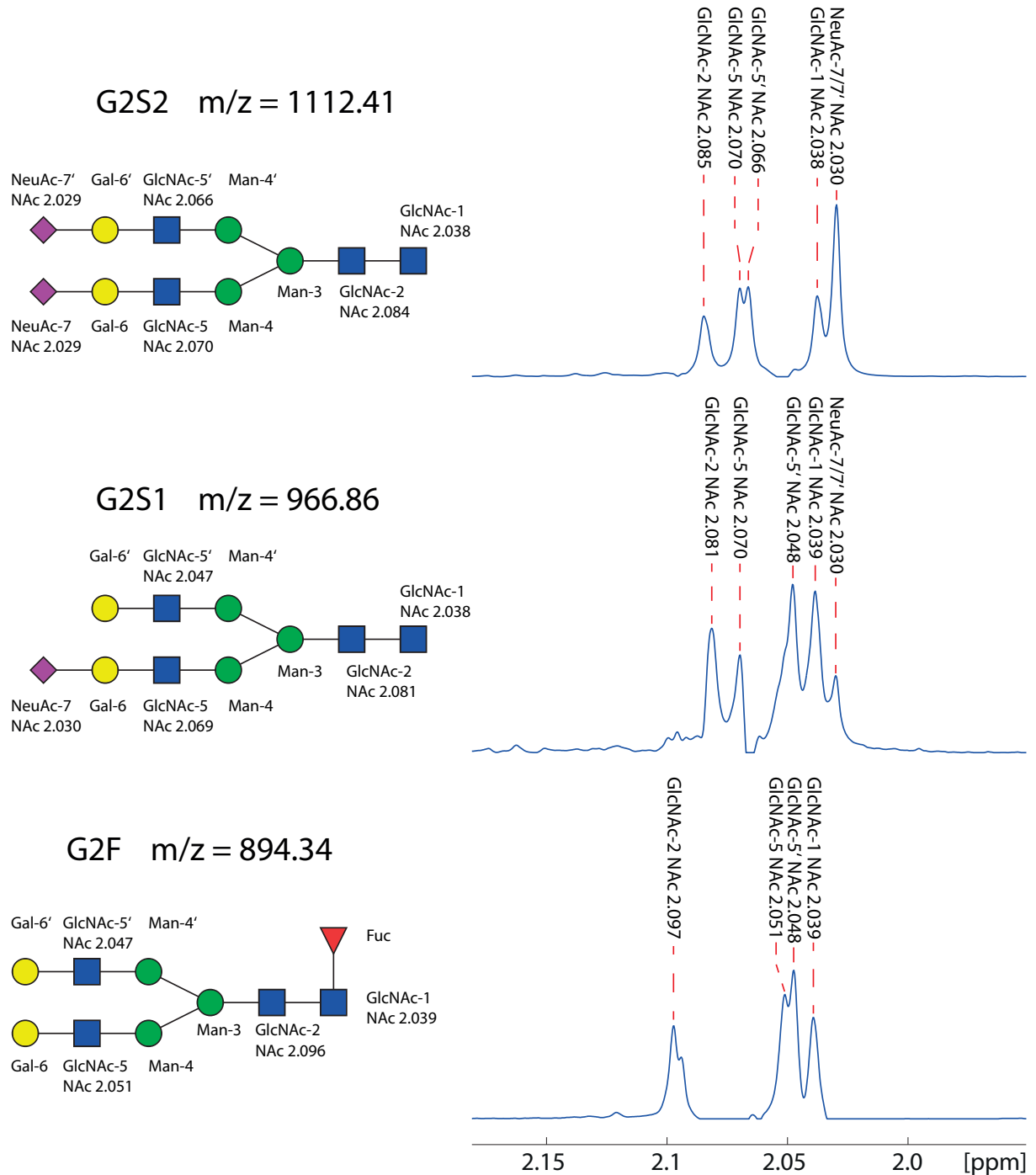


Abbildung 66: 3DCCs Korrelationspektren der Glycane G2F ( $m/z = 894.34$ ), G2S1 ( $m/z = 966.86$ ) und G2S2 ( $m/z = 1112.41$ ) im Bereich 4.05 – 5.25 ppm.



**Abbildung 67:** 3DCCs Korrelationspektren der Glycane G2F ( $m/z = 894.34$ ), G1S1 ( $m/z = 966.86$ ) und G2S2 ( $m/z = 1112.41$ ) im Bereich 1.95 – 2.20 ppm.

### **4.2.3 Fazit zur Entwicklung einer graphischen Benutzeroberfläche für die dreidimensionale Kreuzkorrelation**

Die im Vorherigen diskutierten Spektren zeigen ausschließlich die Signale des jeweils erwarteten Glycans. Signale anderer Glycane wurden durch die Korrelation vollständig entfernt, sodass die reinen Spektren der jeweiligen Glycane erhalten wurden. Dies gilt auch für die Signale der CH<sub>3</sub>-Protonen der Fucosen und H-3-Protonen der *N*-Acetylneuraminsäuren, welche nicht gezeigt wurden. Dieses Ergebnis verdeutlicht die Leistungsfähigkeit der dreidimensionalen Kreuzkorrelation.

Die Möglichkeit reine NMR-Spektren zu erhalten, ohne die Analyten chromatographisch vollständig zu trennen, kann eine erhebliche Zeitersparnis bedeuten. Insbesondere für die Untersuchung von Oligosacchariden aus Glycoproteinen, welche sich strukturell sehr ähnlich sind und chromatographisch schwer trennen lassen, wurde die dreidimensionale Kreuzkorrelation erfolgreich eingesetzt.<sup>51,125</sup>

Die dreidimensionale Kreuzkorrelation kann aber auch in anderen Bereichen von großem Nutzen sein. Die hier entwickelte graphische Benutzeroberfläche vereinfacht die Anwendung der dreidimensionalen Kreuzkorrelation erheblich und wird die Methode vielen potentiellen Nutzern zugänglich machen.

## 5 Zusammenfassung

In dieser Arbeit wurde eine Methode zur Erkennung von Leberkrebs entwickelt. Leberkrebs wird häufig spät entdeckt und hat daher eine der niedrigsten 5-Jahres-Überlebensraten. Der Hauptrisikofaktor für Leberkrebs ist Leberzirrhose, welche meist durch Alkoholmissbrauch, NASH (*non-alcoholic steatohepatitis*) oder Infektion mit Hepatitisviren ausgelöst wird. Aus diesem Grund sind einfache und schnelle Detektionsmethoden für Leberkrebs von größtem medizinischem Interesse.

Fibrinogen wird hauptsächlich in der Leber hergestellt und stellt daher ein optimales Ziel zur Erkennung von Leberkrankheiten dar. Bei der hier entwickelten Methode wird Fibrinogen aus humanem Blutplasma mittels Ethanol-fällung isoliert. Die drei Untereinheiten ( $A\alpha$ ,  $B\beta$  und  $\gamma$ ) werden reduziert, flüssigkeitschromatographisch voneinander getrennt und massenspektrometrisch analysiert. Die deconvolvierten Massenspektren zeigen die Phosphorylierung und *O*-Glycosylierung der  $A\alpha$ -Untereinheit sowie die *N*-Glycosylierung der  $B\beta$ - und  $\gamma$ -Untereinheit. Außerdem wurden auf jeder Untereinheit *single nucleotide polymorphisms* beobachtet, welche zum Austausch einer Aminosäure führen. Durch Verdau mit Trypsin konnte die Identität des Fibrinogen und das Auftreten der SNPs bestätigt werden. Es zeigte sich aber auch, dass das isolierte Fibrinogen Verunreinigungen wie z. B. Albumin und Fibronectin enthält, welche allerdings bei der Flüssigkeitschromatographie abgetrennt werden und die Analyse der intakten Fibrinogenunterheiten nicht stören.

Es ist bekannt, dass die posttranslationalen Modifikationen von Proteinen in Krebszellen verändert sind. So ist zum Beispiel die Fucosylierung von Glycanstrukturen in Krebszellen erhöht. Deshalb wurden Proben von HCC-Patienten, Zirrhosepatienten und gesunden Spendern untersucht, um Unterschiede im Glycosylierungs- und Phosphorylierungsmuster von humanem Fibrinogen festzustellen. Die verschiedenen Glycoformen und Proteinspezies wurden durch Anpassung der theoretischen Isotopenmuster an das experimentelle, deconvolvierte Spektrum quantifiziert. Die Ergebnisse wurden statistisch analysiert. Hierzu wurden univariate Varianzanalysen (ANOVA) durchgeführt. Es konnte gezeigt werden, dass der Anteil sialylierter und fucosylierter Glycoformen der  $B\beta$ - und  $\gamma$ -Untereinheit bei HCC-Patienten im Vergleich zu gesunden Spendern erhöht ist. Der Anteil fucosylierter Glycoformen ist bei Zirrhosepatienten sogar noch höher als bei HCC-Patienten. Außerdem ist der Anteil phosphorylierter Spezies bei HCC-Patienten verringert. Der Anteil einer einzigen Glycoform oder Proteinspezies eignet sich allerdings nicht als Biomarker. Es wurde daher eine Methode entwickelt, die eine Kombination verschiedener Proteinspezies zur

Unterscheidung von gesunden Spendern, HCC- und Zirrhosepatienten verwendet. Hierzu wurden multivariate statistische Methoden wie die Hauptkomponentenanalyse (PCA) und die multivariate Varianzanalyse (MANOVA) verwendet. Hauptkomponentenanalyse und MANOVA wurden mit den vier wichtigsten Glycoformen und Proteinspezies durchgeführt. Die Hauptkomponentenanalyse verwendet keine Informationen über die zu trennenden Gruppen und liefert daher nicht das optimale Ergebnis. Es zeigte sich aber bereits eine gute Trennung zwischen den Proben von gesunden Spendern und HCC-/Zirrhosepatienten. Mit der MANOVA, welche die Information über die zu trennenden Gruppen verwendet, konnte eine ausgezeichnete Trennung zwischen den Datenpunkten von gesunden Spendern, HCC- und Zirrhosepatienten erreicht werden. Die *Receiver-operating-characteristic*-Kurven ergaben ausgezeichnete *area-under-the-curve*-Werte von 0.989 (HCC/Zirrhose vs. gesund), 0.898 (HCC) und 0.920 (Zirrhose). Zur Validierung der entwickelten Methode wurde ein zweiter Datensatz von Proben mit unbekanntem Status analysiert (*blinded data set*). Die Proben der HCC- und Zirrhosepatienten konnten dabei sehr gut von denen gesunder Spender unterschieden werden. Eine Unterscheidung zwischen Proben von HCC- und Zirrhosepatienten konnte allerdings nicht beobachtet werden.

Es konnte gezeigt werden, dass die hier entwickelte Methode zur Detektion von Leberkrebs und Leberzirrhose geeignet ist. Es handelt sich um die erste Studie, welche multivariate Datenanalyse in Verbindung mit massenspektrometrischer Analyse intakter Proteine verwendet, um Proben von Krebspatienten zu identifizieren. Aufgrund der einfachen und schnellen Vorgehensweise, hat diese Methode das Potential die Krebsfrüherkennung zu revolutionieren.

In einem zweiten Projekt wurde eine graphische Benutzeroberfläche für die dreidimensionale Kreuzkorrelation (3DCC) entwickelt. Die dreidimensionale Kreuzkorrelation dient der Berechnung reiner NMR-Spektren aus einem chromatographisch nicht vollständig getrennten Substanzgemisch. Dabei werden NMR-Spektren und *extracted ion*-Chromatogramme mathematisch korreliert. Hierfür müssen die Rohdaten vorbereitet werden. Dies geschieht beispielsweise in MATLAB und bedarf fortgeschrittener Kenntnisse im Umgang damit. Um die dreidimensionale Kreuzkorrelation möglichst vielen Benutzern zugänglich zu machen, wurde ein Programm mit graphischer Benutzeroberfläche entwickelt, welches den Nutzer durch alle erforderlichen Schritte führt. Zum Zweck eines Testdatensatzes wurden die *N*-Glycane aus humanem Plasma mittels 3DCC analysiert. Dabei konnten saubere NMR-Spektren von sieben *N*-Glycanen erhalten werden. Anhand dieses Testdatensatzes wurde ein Tutorial für die entwickelte graphische Benutzeroberfläche verfasst. Diese wird die dreidimensionale Kreuzkorrelation für viele potentielle Benutzer anwendbar machen und zu einer Verbreitung der Methode beitragen.

## 6 Summary

A method for detection of liver cancer has been developed. Liver cancer is often diagnosed late and hence its 5-year-survival-rate is very low. The main risk factor for liver cancer is cirrhosis, primarily caused by alcohol abuse, NASH (*non-alcoholic steatohepatitis*) or viral hepatitis (HBV or HCV). Therefore, fast and easy methods for detection of liver cancer are of great medical interest.

Fibrinogen is dominantly synthesized in the liver and thus an optimal target for detection of liver diseases. The method developed here uses ethanol precipitation to purify fibrinogen from human plasma. The three fibrinogen subunits ( $A\alpha$ ,  $B\beta$  and  $\gamma$ ) are reduced, separated by liquid chromatography and analyzed by mass spectrometry. The deconvoluted mass spectra display phosphorylation and *O*-glycosylation of the  $A\alpha$ -subunit and *N*-Glycosylation of  $B\beta$ - and  $\gamma$ -subunits. Furthermore, single nucleotide polymorphisms that lead to exchanges of an amino acid were found at all subunits. The identity of the fibrinogen and the presence of the SNPs were confirmed by tryptic digests. But it was also shown that the isolated fibrinogen contains impurities like albumin and fibronectin. However, these impurities are separated by liquid chromatography and do not interfere in the analysis of the intact fibrinogen subunits.

It is known that the posttranslational modifications of proteins are changed during cancer. For example fucosylation of glycan structures is increased in cancer cells. Therefore, samples of HCC patients, cirrhosis patients and healthy controls were analyzed to detect differences in the glycosylation and phosphorylation pattern of human fibrinogen. The different glycoforms and protein species were quantified by fitting the theoretical isotopic patterns to the experimental deconvoluted spectra. The results were analyzed statistically. Univariate analyses of variance (ANOVA) show that the amount of sialylated and fucosylated glycoforms is increased in HCC patients compared to healthy donors. The amount of fucosylated glycoforms is even higher in cirrhosis patients. Furthermore, the amount of phosphorylated species is decreased in HCC patients. However, a single glycoform or protein species cannot be used as a biomarker. Therefore, a method that uses a combination of different protein species for differentiation of healthy donors, HCC- and cirrhosis patients was developed. For this purpose, multivariate statistical methods like principal component analysis (PCA) and multivariate analysis of variance (MANOVA) were used. Principal component analysis and MANOVA were performed with the four most important protein species. Since principal component analysis does not use information about the groups that are to be separated, it does not give the optimal results. It shows, however, a good separation between samples of



healthy donors and HCC/cirrhosis patients. With MANOVA that uses the information about the groups, an excellent separation between healthy donors, HCC- and cirrhosis patients was achieved. The area-under-the-curve-values of the receiver-operating-characteristic-curves were 0.989 (healthy vs. HCC/cirrhosis), 0.898 (HCC) and 0.920 (cirrhosis). For validation a second dataset containing samples of unknown status (blinded data set) was analyzed. Samples of HCC- and cirrhosis patients could be distinguished from samples of healthy donors, but differentiation between HCC- and cirrhosis samples is not yet possible.

It was shown that the method developed here is applicable for detection of liver cancer and liver cirrhosis. It is the first study that uses multivariate data analysis in conjunction with mass spectrometry of intact proteins to identify samples of cancer patients. Due to its fast and simple approach, this method could have great impact on cancer screening.

In a second project a graphical user interface for three dimensional cross correlation (3DCC) was developed. Three dimensional cross correlation calculates pure NMR-spectra from a mixture of analytes that is not completely resolved by chromatography. Therefore, NMR-spectra and extracted ion chromatograms are correlated mathematically. Data preparation is usually performed with MATLAB. For this, advanced knowledge in MATLAB programming is required. To make 3DCC applicable for more potential users, a program with a graphical user interface was developed that leads the user through all important steps of three dimensional cross correlation. As an example dataset, the *N*-glycans from human plasma were analyzed by 3DCC. Pure NMR-spectra of seven *N*-glycans were obtained. On the basis of this example dataset, a tutorial for the graphical user interface for 3DCC was written. This will make 3DCC applicable for more potential users and help to increase the methods publicity.

## 7 Experimenteller Teil

### 7.1 Materialien

**Tabelle 11: Verwendete Chemikalien und Biomaterialien**

<b>Verwendete Chemikalie / Biomaterialien</b>	<b>Firma</b>
Acetonitril HPLC grade $\geq$ 99.9%	Sigma Aldrich Chemie GmbH
Ameisensäure 98%	VWR International GmbH
2-Amino-2-(hydroxymethyl)-propan-1,3-diol	Sigma Aldrich Chemie GmbH
Ammoniumformiat	Fluka Chemie AG
Ammoniumhydrogencarbonat	Grüssing GmbH
Blutplasmakonserven	Zentralinstitut für Transfusionsmedizin GmbH
Calciumchlorid	Honeywell International Inc.
Deuteriumoxid	Deutero GmbH
1,4-D/L-Dithiothreitol	AppliChem GmbH
Ethanol	k. A.
Harnstoff	Merck KGaA
Iodacetamid	AppliChem GmbH
Natriumdodecylsulfat	Merck Schuchardt OHG
Nonident P-40	Sigma Aldrich Chemie GmbH
N-Glycosidase F	F. Hoffman – La Roche AG
Pronase E	Sigma Aldrich Chemie GmbH
Tuning Mix ES (ESI) (Pos)	Agilent Technologies GmbH
Trypsin Gold Mass Spectrometry Grade	Promega Corporation

**Tabelle 12: Verwendete Geräte**

<b>Verwendete Geräte</b>	<b>Firma</b>
5 mm TXI-Cryoprobenkopf	Bruker BioSpin GmbH
Aeris WIDEPOR XB-C8	Phenomenex Inc.
Analysenwaage ABS 220-4	Kern & Sohn GmbH
Analysenwaage AE240	Mettler-Toledo GmbH
Centrifuge 5415 D	Eppendorf AG
Centrifuge 5417 R	Eppendorf AG

## Experimenteller Teil

Centrifuge 5804 R	Eppendorf AG
ESI-q/TOF Massenspektrometer (maXis)	Bruker Daltonik GmbH
Hypercarb 3 µm, PGC, 250 Å, 150 × 2.1 mm	Thermo Fisher Scientific Inc.
Kinetex 1.7µm, C18, 100Å, 150 × 2.1 mm	Phemenex Inc.
Lyophilisator, Alpha 1-2	CHRIST® Gefriertrocknungsanlagen GmbH
NMR (700 MHz, Avance I)	Bruker Corporation
Reinstwasseranlage Ultra Clear UV 18.2 MΩ	SGwater GmbH
Thermomixer comfort	Eppendorf AG
Sicherheitswerkbank HERAsafe	Kendro Laboratory Products GmbH
Vortex Mixer, REAX top	Heidolph Instruments GmbH & Co. KG

## 7.2 Software und Datenbanken

**Tabelle 13: Verwendete Software**

<b>Verwendete Software</b>	<b>Firma/Literatur</b>
MATLAB R2012a	The MathWorks, Inc.
SPSS (Version 22)	IBM Corp.
DataAnalysis 4.2	Bruker Daltonik GmbH
Microsoft Excel 2010	Microsoft Corp.
Glycoworkbench 2.1 <sup>131</sup>	Ceroni, A., Maass, K., Geyer, H., Dell, A., Haslam, S.M. <i>J Proteome Res</i> , 7, 1650-1659 (2008)
TopSpin 3.2	Bruker BioSpin GmbH
Mascot 2.6	Matrix Science Inc.
G*Power 3.1.9.2 <sup>132,133</sup>	Faul, F., Erdfelder, E., Lang, A.-G. <i>Behavior Research Methods</i> , 39, 175-191, (2007) Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G. <i>Behavior Research Methods</i> , 41, 1149-1160, (2009)

**Tabelle 14: Verwendete Datenbanken**

<b>Verwendete Datenbank</b>	<b>Organisation/Literatur</b>
Sugabase	Vankuik, J.A., Hard, K. & Vliegthart, J.F.G. <i>Carbohydr Res</i> 235, 53-68 (1992)
GlycomeDB	Deutsches Krebsforschungszentrum Heidelberg
UniProt	The UniProt Consortium
NCBI SNP database	National Center for Biotechnology Information

### 7.3 Untersuchung der intakten Untereinheiten von humanem Fibrinogen

Ein Volumenanteil Blutplasma (ca. 400  $\mu\text{L}$ ) wurde mit einem Volumenanteil Ethanol (16%, v/v) versetzt, sodass eine Ethanolkonzentration von 8% erreicht wurde. Die Lösung wurde 15 min bei 5000 RPM und  $-5\text{ }^{\circ}\text{C}$  zentrifugiert. Das Pellet wurde mit Ethanol (8%, v/v) gewaschen und in 200  $\mu\text{L}$  einer Lösung aus Dithiothreitol (100 mM) und Harnstoff (6 M) gelöst. Die Lösung wurde 30 min bei  $37\text{ }^{\circ}\text{C}$  unter Schütteln inkubiert und mit 1300  $\mu\text{L}$  Ameisensäurelösung (0.1%) verdünnt.

Es wurden ca. 30  $\mu\text{L}$  (ca. 64 pmol Fibrinogen) der Lösung für die HPLC-MS verwendet. Die Trennung wurde auf einer für intakte Proteine geeigneten C8-reversed phase Säule (Aeris WIDEPOR XB-C8) bei einer Flussrate von 250  $\mu\text{L}/\text{min}$  und einer Temperatur von  $60\text{ }^{\circ}\text{C}$  durchgeführt. Als Lösungsmittel A wurde Wasser (0.1% Ameisensäure) verwendet, Lösungsmittel B war Acetonitril (0.1% Ameisensäure). Die Trennung erfolgte mit folgendem Gradienten: 0 min: 5% B, 5 min 5% B, 82 min: 60% B, 84 min: 90% B, 87 min: 90% B, 89 min 5% B, 90 min: 5% B.

Die massenspektrometrische Detektion erfolgte unter folgenden Parametern: *End Plate Offset*: 500 V; *Capillary Voltage*: 4500 V; *Nebulizer Pressure*: 3.0 bar; *Dry Gas Flow Rate*: 9.0 L/min; *Dry Gas Temperature*:  $190\text{ }^{\circ}\text{C}$ ; *Mass Range*: 500-4000 m/z; *Ion Polarity*: Positiv; *isCID*: 40 eV; *Quadrupole Ion Energy*: 5 eV; *Quadrupole Low Mass*: 1200 m/z; *Collision Cell Energy*: 5.0 eV; *Collision Cell RF*: 1200 Vpp; *Collision Cell Transfer Time*: 200  $\mu\text{s}$ ; *Collision Cell Pre Pulse Storage*: 20.0  $\mu\text{s}$ ; *Transfer Funnel 1 RF*: 400.0 Vpp; *Spectra Rate*: 0.5 Hz; *Multipole RF*: 300 Vpp.

## 7.4 Untersuchung der tryptischen Peptide von humanem Fibrinogen

Jeweils 500  $\mu\text{L}$  Blutplasma wurde mit 500  $\mu\text{L}$  Ethanol (16%, v/v) versetzt und 15 min bei 5000 RPM und  $-5\text{ }^\circ\text{C}$  zentrifugiert. Das Pellet wurde mit Ethanol (8%, v/v) gewaschen und in 100  $\mu\text{L}$  Dithiothreitollösung (10 mM) gelöst. Es wurde 20 min bei  $60\text{ }^\circ\text{C}$  unter schütteln inkubiert. Nach Zugabe von 20  $\mu\text{L}$  Iodacetamid (0.3 M) wurde weitere 30 min bei Raumtemperatur unter Lichtausschluss inkubiert. Die Lösung wurde mit 450  $\mu\text{L}$  Ammoniumhydrogencarbonatlösung (50 mM) verdünnt und es wurde 1  $\mu\text{L}$  Trypsin (1  $\mu\text{g}/\mu\text{L}$ ) hinzugegeben. Nach 24 Stunden Inkubation bei  $37\text{ }^\circ\text{C}$  wurde nochmals 1  $\mu\text{L}$  Trypsin (1  $\mu\text{g}/\mu\text{L}$ ) zugegeben und weitere 24 Stunden bei  $37\text{ }^\circ\text{C}$  inkubiert. Die Proben wurden filtriert und jeweils 100  $\mu\text{L}$  wurden für die HPLC-MS Analyse verwendet.

Die Trennung wurde auf einer C18-*reversed phase* Säule (Kinetex 1.7 $\mu$ , C18, 100 $\text{\AA}$ , 150  $\times$  2.1 mm) mit einer Flussrate von 250  $\mu\text{L}/\text{min}$  und einer Temperatur von  $55\text{ }^\circ\text{C}$  durchgeführt. Als Lösungsmittel A wurde Wasser (0.1% Ameisensäure) verwendet, Lösungsmittel B war Acetonitril (0.1% Ameisensäure). Die Trennung erfolgte mit folgendem Gradienten. 0 min: 2% B, 4 min: 2% B, 42 min: 60% B, 44 min: 90% B, 50 min: 90% B, 55 min: 2% B, 60 min: 2% B.

Die massenspektrometrische Detektion erfolgte unter folgenden Parametern: *End Plate Offset*: 500 V; *Capillary Voltage*: 4500 V; *Nebulizer Pressure*: 4.0 bar; *Dry Gas Flow Rate*: 9.0 L/min; *Dry Gas Temperature*:  $200\text{ }^\circ\text{C}$ ; *Mass Range*: 150-2500 m/z; *Ion Polarity*: Positiv; *isCID*: 0 eV; *Quadrupole Ion Energy*: 4 eV; *Quadrupole Low Mass*: 322 m/z; *Collision Cell Energy*: 8.0 eV; *Collision Cell RF*: 1200 Vpp; *Collision Cell Transfer Time*: 120  $\mu\text{s}$ ; *Collision Cell Pre Pulse Storage*: 10.0  $\mu\text{s}$ ; *Transfer Funnel 1 RF*: 400.0 Vpp; *Spectra Rate*: 1 Hz; *Multipole RF*: 400 Vpp; *Absolute Threshold (MS/MS)*: 500 cts per 1000 sum.; *Exclude after*: 3 Spectra; *Release after*: 1.00 min; *Exclude Singly*: yes

### 7.4.1 MASCOT MS/MS Ions Search

Die MS/MS-Spektren der Proben N, O, Q und T wurden mit folgenden Parametern als *mascot generic file* aus Compass DataAnalysis exportiert: *Intensity threshold*: 1000; *Maximum number of compounds*: 1200; *Fragments filtered by*: Amino Acids. Die MASCOT MS/MS Ions Search wurde mit folgenden Parametern durchgeführt: *Database*: SwissProt; *Taxonomy*: *homo sapiens*; *Enzyme*: Trypsin; *Allow up to 1 missed cleavages*; *Fixed modifications*: Carbamidomethyl (C); *Peptide tol.*: 0.2 Da; *MS/MS tol.*: 0.2 Da; *Peptide charge*: 2+, 3+ and 4+, *monoisotopic*.

## 7.5 Untersuchung der Glycopeptide von humanem Fibrinogen durch Verdau mit Pronase

Jeweils 500 µL Blutplasma der Proben T, R, HCC017, HCC024 und ZIRO31 wurden mit 500 µL Ethanol (16%, v/v) versetzt und 15 min bei 5000 RPM und -5 °C zentrifugiert. Das Pellet wurde mit Ethanol (8%, v/v) gewaschen und in 90 µL TRIS-Puffer (100 mM) mit Calciumchlorid (50 mM) aufgenommen. Es wurden 10 µL Pronaselösung (1 mg/mL) zugegeben und bei 37 °C inkubiert. Nach 24 Stunden wurden weitere 10 µL Pronaselösung (1 mg/mL) zugegeben und weitere 24 Stunden bei 37 °C inkubiert. Die Proben wurden filtriert und 100 µL wurden für die HPLC-MS Analyse verwendet

Die Trennung erfolgte auf einer *porous graphitized carbon* Säule (Hypercarb) mit einer Flussrate von 150 µL/min, bei einer Temperatur von 40 °C. Als Lösungsmittel A wurde Wasser (0.1% Ameisensäure) verwendet, Lösungsmittel B war Acetonitril (0.1% Ameisensäure). Die Trennung erfolgte mit folgendem Gradienten: 0 min: 0% B, 5 min: 0% B, 68 min: 40% B, 70 min: 90% B, 72 min: 90% B, 75 min: 0% B, 80 min: 0% B.

Die massenspektrometrische Detektion erfolgte unter folgenden Parametern: *End Plate Offset: 500 V; Capillary Voltage: 5000 V; Nebulizer Pressure: 3.0 bar; Dry Gas Flow Rate: 6.0 L/min; Dry Gas Temperature: 180 °C; Mass Range: 100-2800 m/z; Ion Polarity: Positiv; isCID: 0 eV; Quadrupole Ion Energy: 4 eV; Quadrupole Low Mass: 322 m/z; Collision Cell Energy: 8.0 eV; Collision Cell RF: 1200 Vpp; Collision Cell Transfer Time: 80 µs; Collision Cell Pre Pulse Storage: 10.0 µs; Transfer Funnel 1 RF: 400.0 Vpp; Spectra Rate: 0.5 Hz; Multipole RF: 400 Vpp; Absolute Threshold (MS/MS): 160 cts per 1000 sum.; Exclude after: 5 Spectra; Release after: 1.00 min; Exclude: m/z = 100-700; Exclude Singly: yes*

## 7.6 Blutplasmaproben

Die Blutplasmaproben von gesunden Spendern (A-AD) wurden vom Zentralinstitut für Transfusionsmedizin GmbH erhalten.

Blutplasmaproben von HCC- und Leberzirrhosepatienten wurden in Kooperation mit Prof. Dr. Ansgar Lohse, Dr. Henning Wege und Dr. Inez Gil-Ibanez vom Universitätsklinikum Hamburg Eppendorf erhalten. Vor der Blutentnahme wurde die Einverständniserklärung der Patienten eingeholt. Der Untersuchung von Blutplasmaproben zur Entwicklung neuer Biomarker für Leberkrebs wurde von der lokalen Ethikkommission zugestimmt (PV3578).

## 7.7 Probeninformationen

Tabelle 15: Informationen zu den Proben des Trainingsdatensatzes.

Probenname	Geschlecht	Alter	HCC	Stadium	Zirrhose	Genese	Therapie	Begleiterkrankungen
A	M	62	Nein	-	Nein	-	-	-
B	M	60	Nein	-	Nein	-	-	-
C	M	34	Nein	-	Nein	-	-	-
D	M	53	Nein	-	Nein	-	-	-
E	M	26	Nein	-	Nein	-	-	-
F	M	19	Nein	-	Nein	-	-	-
G	M	58	Nein	-	Nein	-	-	-
H	M	60	Nein	-	Nein	-	-	-
I	W	29	Nein	-	Nein	-	-	-
J	W	21	Nein	-	Nein	-	-	-
K	W	48	Nein	-	Nein	-	-	-
L	W	51	Nein	-	Nein	-	-	-
M	M	45	Nein	-	Nein	-	-	-
N	M	21	Nein	-	Nein	-	-	-
O	M	54	Nein	-	Nein	-	-	-
P	W	28	Nein	-	Nein	-	-	-
Q	M	24	Nein	-	Nein	-	-	-
R	M	55	Nein	-	Nein	-	-	-
S	M	46	Nein	-	Nein	-	-	-
T	M	47	Nein	-	Nein	-	-	-

Tabelle 15 Fortsetzung: Informationen zu den Proben des Trainingsdatensatzes.

Probenname	Geschlecht	Alter	HCC	Stadium	Zirrhose	Genese	Therapie	Begleiterkrankungen
HCC001	W	65	Ja	Metastasen	Ja	NASH	TACE	Diabetes mellitus
HCC002	W	80	Ja	> Mailand	Ja	Unklar	-	Nierenzellkarzinom
HCC003	M	72	Ja	> Mailand	Ja	Alkohol	TACE	Diabetes mellitus
HCC006	M	62	Ja	> Mailand	Ja	NASH	TACE	Diabetes mellitus
HCC007	M	58	Ja	> Mailand	Ja	Alkohol	-	-
HCC008	M	73	Ja	> Mailand	Ja	HCV	TACE, Ablation	Plattenepithelkarzinom
HCC009	M	63	Ja	Metastasen	Nein	HCV	-	-
HCC011	M	77	Ja	Metastasen	Ja	Alkohol	-	Diabetes mellitus
HCC012	M	70	Ja	< Mailand	Ja	HCV	-	Diabetes mellitus
HCC014	M	36	Ja	< Mailand	Nein	HBV	-	-
HCC016	W	38	Ja	> Mailand	Nein	Unklar	Resektion	-
HCC017	M	59	Ja	Metastasen	Ja	Unklar	-	-
HCC018*	M	73	Ja	> Mailand	Ja	HCV	TACE, Ablation	Plattenepithelkarzinom
HCC019	W	72	Ja	> Mailand	Ja	Alkohol	TACE	-
HCC020	M	71	Ja	> Mailand	Ja	Alkohol	-	-
HCC021	M	78	Ja	> Mailand	Ja	HCV	TACE	Rektumkarzinom
HCC022	M	52	Ja	Metastasen	Ja	HCV	TACE, Sorafenib	-
HCC023	M	63	Ja	< Mailand	Ja	HBV	TACE, Ablation	Chronisches Nierenversagen
HCC024	M	57	Ja	< Mailand	Ja	HCV	TACE	Multiples Myelom

\*HCC018 ist die Wiederholung von HCC008.



Tabelle 15 Fortsetzung: Informationen zu den Proben des Trainingsdatensatzes.

Probenname	Geschlecht	Alter	HCC	Stadium	Zirrhose	Genese	Therapie	Begleiterkrankungen
ZIRO31	W	54	Nein	-	Ja	Alkohol	-	Hypothyreose, HT
ZIRO33	W	56	Nein	-	Ja	Unklar	-	arterieller Hypertonus
ZIRO34	W	66	Nein	-	Ja	Alkohol	-	arterieller Hypertonus
ZIRO35	W	55	Nein	-	Ja	NRH	-	Hypothyreose, HT
ZIRO36	M	42	Nein	-	Ja	Morbus Wilson	-	Diabetes mellitus
ZIRO37	W	39	Nein	-	Ja	Vaskulär	-	Polyzythämie Vera
ZIRO38	M	54	Nein	-	Ja	Alkohol	-	HT, Diabetes mellitus
ZIRO39	M	55	Nein	-	Ja	HCV	-	Osteopenie
ZIRO41	W	59	Nein	-	Ja	Alkohol	-	Hypothyreose, DM, COPD
ZIRO42	W	44	Nein	-	Ja	Alkohol	-	Paroxymales Vorhofflimmern
ZIRO43	W	62	Nein	-	Ja	NASH	-	KHK, Diabetes mellitus
ZIRO44	M	44	Nein	-	Ja	Alkohol	-	-
ZIRO45	W	61	Nein	-	Ja	Alkohol	-	Paroxymales Vorhofflimmern, HT
ZIRO46	M	49	Nein	-	Ja	PBC, NASH	-	Diabetes mellitus, Adipositas
ZIRO48	M	57	Nein	-	Ja	Alkohol, HCV	-	Hypothyreose
ZIRO49	M	39	Nein	-	Ja	Alkohol	-	-
ZIRO50	W	56	Nein	-	Ja	Alkohol	-	Nikotinabusus, COPD

NASH = *non-alcoholic steatohepatitis*; HBV = Hepatitis B virus; HCV = Hepatitis C virus; TACE = *trans arterial chemoembolization*; HT = arterieller Hypertonus; DM = Diabetes mellitus; COPD = Chronisch obstruktive Lungenerkrankung; KHK = koronare Herzerkrankung, PBC = primär biliäre Cholangitis; TKI = Tyrosinkinaseinhibitoren; LTX = Lebertransplantation; RFA = Radiofrequenzablation; NRH = Noduläre regenerative Hyperplasie.

**Tabelle 16: Informationen zu den Proben der Validierungskohorte.**

Probenname	Geschlecht	Alter	HCC	Stadium	Zirrhose	Genese	Therapie	Begleiterkrankungen
U	W	27	Nein	-	Nein	-	-	-
V	M	51	Nein	-	Nein	-	-	-
W	M	22	Nein	-	Nein	-	-	-
X	M	32	Nein	-	Nein	-	-	-
Y	M	39	Nein	-	Nein	-	-	-
Z	M	38	Nein	-	Nein	-	-	-
AA	M	27	Nein	-	Nein	-	-	-
AB	M	46	Nein	-	Nein	-	-	-
AC	M	66	Nein	-	Nein	-	-	-
AD	M	46	Nein	-	Nein	-	-	-
V001	M	76	Ja	< Mailand	Ja	Alkohol	Resektion	Diabetes mellitus
V002	M	56	Ja	Metastasen	Ja	HCV	TKI	-
V003*	M	58	Ja	> Mailand	Ja	Alkohol	TKI	-
V004	M	36	Ja	> Mailand	Nein	HBV	TACE	-
V005	M	58	Ja	> Mailand	Ja	HCV	TKI	-
V006	M	62	Ja	> Mailand	Ja	HCV	TACE	COPD
V007	M	58	Nein	-	Ja	Alkohol	Resektion	Dysplastische Leberherde
V008	M	51	Ja	> Mailand	Ja	HBV	TACE	-
V009	M	83	Ja	> Mailand	Ja	NASH	TACE	Diabetes mellitus

\*V003 ist die Wiederholung von HCC007; V017: Wdh. V007; V020: Wdh. V014; V022: Wdh. HCC002; V023: Wdh. V020 bzw. V014.

**Tabelle 16 Fortsetzung: Informationen zu den Proben der Validierungskohorte.**

Probenname	Geschlecht	Alter	HCC	Stadium	Zirrhose	Genese	Therapie	Begleiterkrankungen
V010	M	50	Nein	-	Nein	HBV	-	Leberangiomyolipom
V011	M	79	Ja	> Mailand	Ja	Alkohol	TACE	Prostatakarzinom
V012	W	70	Ja	Metastasen	Nein	NASH	TKI	-
V013	M	65	Ja	Metastasen	Nein	HBV	Radiation	Diabetes mellitus
V014	M	58	Ja	< Mailand	Ja	HCV	TACE, RFA	Nasopharynxkarzinom
V015	W	44	Nein	-	Nein	-	-	Leberadenom
V016	W	46	Nein	-	Nein	-	-	Leberadenom
V017*	M	76	Ja	> Mailand	Ja	Alkohol	Resektion	Diabetes mellitus
V019	M	53	Ja	Metastasen	Nein	HCV	LTX	-
V020*	M	58	Ja	< Mailand	Ja	HCV	TACE, RFA	Nasopharynxkarzinom
V021	M	47	Ja	Metastasen	Nein	Adenom	TKI	-
V022*	M	80	Ja	Metastasen	Ja	Unklar	Resektion	Nierenzellkarzinom
V023*	M	58	Ja	< Mailand	Ja	HCV	TACE, RFA	Nasopharynxkarzinom
V025*	M	58	Ja	> Mailand	Ja	Alkohol	TKI	-
V026	M	63	Ja	Metastasen	Ja	Unklar	TACE, TKI	Metabolisches Syndrom
V027	M	55	Nein	-	Ja	HCV	Resektion	-
V028	W	76	Ja	< Mailand	Ja	PBC	TACE	Niereninsuffizienz
V029	M	45	Nein	-	Ja	HCV	-	Diabetes mellitus
V030	W	45	Nein	-	Ja	Alkohol	-	-

\*V025 ist die Wiederholung von HCC007 bzw. V003.

Tabelle 16 Fortsetzung: Informationen zu den Proben der Validierungskohorte.

Probenname	Geschlecht	Alter	HCC	Stadium	Zirrhose	Genese	Therapie	Begleiterkrankungen
Vo31	M	55	Nein	-	Ja	Alkohol	-	-
Vo32	M	35	Nein	-	Ja	Alkohol	-	-
Vo33	W	84	Nein	-	Nein	-	-	Chronische Pankreatitis
Vo34	W	55	Nein	-	Ja	Alkohol	-	Diabetes mellitus
Vo35	W	44	Nein	-	Ja	PSC, LTX	-	Colitis ulcerosa
Vo36	M	50	Nein	-	Ja	PSC	-	Diabetes mellitus
Vo37	M	75	Nein	-	Nein	-	-	KHK
Vo38	M	60	Nein	-	Nein	NASH, LTX	-	Diabetes mellitus
Vo39	M	46	Nein	-	Ja	Alkohol	-	-
Vo40	W	53	Nein	-	Ja	PSC	-	Colitis ulcerosa
Vo41	M	69	Nein	-	Ja	Alkohol	-	Niereninsuffizienz
Vo42	M	57	Nein	-	Ja	Alkohol, LTX	-	Niereninsuffizienz
Vo43	M	39	Nein	-	Ja	PSC, AIH	-	Colitis ulcerosa
Vo45	W	63	Nein	-	Ja	AIH	-	Ösophaguskarzinom
Vo46	M	47	Nein	-	Ja	Alkohol	-	Asthma bronchiale
Vo47	M	35	Nein	-	Nein	-	-	Hashimoto-Thyreoiditis
Vo48	M	70	Ja	< Maitland	Ja	HCV	Resektion	Diabetes mellitus

NASH = *non-alcoholic steatohepatitis*; HBV = Hepatitis B virus; HCV = Hepatitis C virus; TACE = *trans arterial chemoembolization*; HT = arterieller Hypertonus; DM = Diabetes mellitus; COPD = Chronisch obstruktive Lungenerkrankung; KHK = koronare Herzerkrankung, PBC = primär biliäre Cholangitis; TKI = Tyrosinkinaseinhibitoren; LTX = Lebertransplantation; RFA = Radiofrequenzablation; NRH = Noduläre regenerative Hyperplasie; PSC = Primär sklerosierende Cholangitis; AIH = Autoimmunhepatitis.

## 7.8 3DCC

Die *N*-Glycane aus humanem Blutplasma wurden wie folgt isoliert. 2 mL Blutplasma (Probe D) wurden mit 4 mL Natriumdodecylsulfatlösung (2%) versetzt und für 30 min bei 60 °C inkubiert. Es wurden 2 mL Nonoxinol 40 (4%) und 1 µL PNGaseF (10 u) zugegeben und 24 Stunden bei 37 °C inkubiert. Es wurden erneut 1 µL PNGaseF (10 u) zugegeben und weitere 24 Stunden bei 37 °C inkubiert.

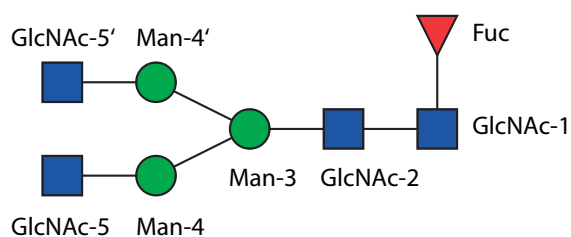
Die Lösung wurde mit 42 mL Ethanol (96%) versetzt und 10 min bei Raumtemperatur/5000 RPM zentrifugiert. Der Überstand wurde unter vermindertem Druck vom Ethanol befreit und mittels C18-RP-SPE gereinigt. Hierzu wurde der Rückstand auf die SPE-Kartusche aufgebracht und mit etwa 10 mL Wasser eluiert. Das Eluat der C18-RP-SPE wurde gefriergetrocknet und mittels PGC-SPE gereinigt. Hierzu wurde der Rückstand in 1 mL Wasser (0.1% Ameisensäure) aufgenommen und auf die PGC-SPE-Kartusche aufgebracht. Der Durchfluss wurde insgesamt zehnmal erneut auf die PGC-SPE-Kartusche aufgebracht. Anschließend wurde dreimal mit jeweils 2 mL Wasser (0.1% Ameisensäure) gewaschen und zweimal mit Wasser/Acetonitril (1:1, 0.1% Ameisensäure) eluiert. Das Eluat wurde unter vermindertem Druck vom Lösungsmittel befreit. Der Rückstand wurde in 50 µL Wasser (0.1% Ameisensäure) aufgenommen und vollständig für die LC-MS-Analyse verwendet.

Die flüssigkeitschromatographische Trennung erfolgte auf einer *porous graphitized carbon*-Säule (Hypercarb, 4.6 × 150 mm) mit einer Flussrate von 0.8 mL/min, bei einer Temperatur von 40 °C. Als Lösungsmittel A wurde Wasser (65 mM Ammoniumformiat, 50 mM Ammoniak) verwendet, Lösungsmittel B war Acetonitril (0.1% Ameisensäure). Die Trennung erfolgte mit folgendem Gradienten: 0 min: 0% B, 2 min: 8% B, 53 min: 25% B, 55 min: 90% B, 60 min: 90% B, 62 min: 0% B, 65 min: 0% B. Der Fluss wurde nach der Säule geteilt, etwa 50 µL/min wurden zur massenspektrometrischen Detektion verwendet, während etwa 750 µL/min in Fraktionen von jeweils 10 s gesammelt wurden.

Die massenspektrometrische Detektion erfolgte unter folgenden Parametern: *End Plate Offset*: 500 V; *Capillary Voltage*: 5000 V; *Nebulizer Pressure*: 3.0 bar; *Dry Gas Flow Rate*: 7.0 L/min; *Dry Gas Temperature*: 200 °C; *Mass Range*: 150-3000 m/z; *Ion Polarity*: Positiv; *isCID*: 0 eV; *Quadrupole Ion Energy*: 4 eV; *Quadrupole Low Mass*: 322 m/z; *Collision Cell Energy*: 10.0 eV; *Collision Cell RF*: 1200 Vpp; *Collision Cell Transfer Time*: 120 µs; *Collision Cell Pre Pulse Storage*: 2.0 µs; *Transfer Funnel 1 RF*: 400.0 Vpp; *Spectra Rate*: 1 Hz; *Multipole RF*: 400 Vpp.

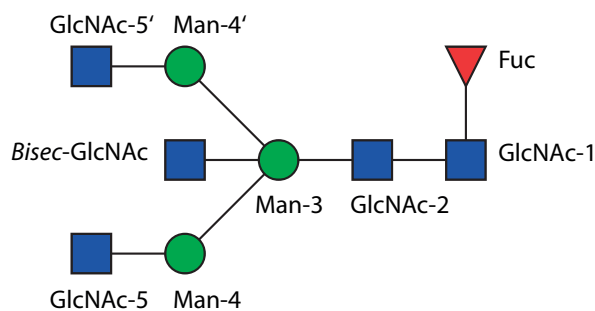
Die Fraktionen wurden unter vermindertem Druck vom Lösungsmittel befreit und in jeweils 180  $\mu\text{L}$  Deuteriumoxid aufgenommen. Die Aufnahme von NMR-Spektren erfolgte an einem 700 MHz Spektrometer mit TXI-Kryoprobenkopf. Es wurde die Pulssequenz *zgesgp* mit einem  $180^\circ$  *shaped*-Puls von 8 ms zur Wasserunterdrückung mittels *excitation sculpting* verwendet.

### GOF $m/z = 732.29$

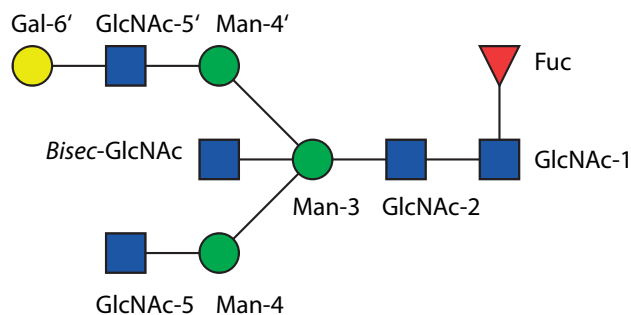


$^1\text{H-NMR}$  (700 MHz,  $3\text{DCC}_6\text{S}$ , 300 K)  
 $\delta$  [ppm] = 5.179 (d,  $^3J = 3.5$  Hz, 0.3 H, GlcNAc-1  $\alpha$  H-1), 5.114 (s, 1 H, Man-4 H-1), 4.916 (s, 1 H, Man-4' H-1), 4.984 (d,  $^3J = 4.6$  Hz, 0.3 H, Fuc  $\alpha$  H-1), 4.888 (d,  $^3J = 4.3$  Hz, 0.7 H, Fuc  $\beta$  H-1), 4.669 (d,  $^3J = 9.2$  Hz, 0.3 H, GlcNAc-2  $\alpha$  H-1), 4.664 (d,  $^3J = 8.3$  Hz, 0.7 H, GlcNAc-2 H-1), 4.554 (d,  $^3J = 9.4$  Hz, 2 H, GlcNAc-5/5' H-1), 4.250 (s, 1 H, Man-3 H-2), 4.188 (d,  $^3J = 3.1$  Hz, 1 H, Man-4 H-2), 4.104 (s, 1 H, Man-4' H-2), 2.095 (s, 2 H, GlcNAc-2  $\beta$  NAc), 2.092 (s, 1 H, GlcNAc-2  $\alpha$  NAc), 2.052 (s, 6 H, GlcNAc-5/5' NAc), 2.039 (s, 3 H, GlcNAc-1 NAc), 1.219 (d,  $^3J = 7.3$  Hz, 2 H, Fuc  $\beta$   $\text{CH}_3$ ), 1.209 (d,  $^3J = 7.3$  Hz, 1 H, Fuc  $\alpha$   $\text{CH}_3$ ).

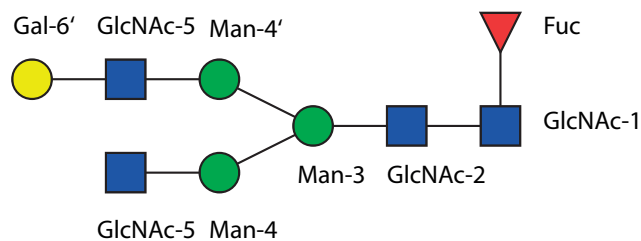
### GOBF $m/z = 833.83$



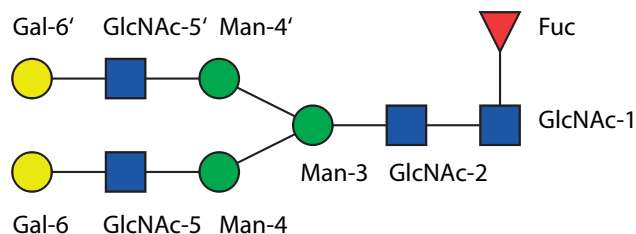
$^1\text{H-NMR}$  (700 MHz,  $3\text{DCC}_6\text{S}$ , 300 K)  
 $\delta$  [ppm] = 5.179 (d,  $^3J = 3.6$  Hz, 0.3 H, GlcNAc-1  $\alpha$  H-1), 5.057 (s, 1 H, Man-4 H-1), 4.998 (s, 1 H, Man-4' H-1), 4.899 (d,  $^3J = 4.5$  Hz, 0.3 H, Fuc  $\alpha$  H-1), 4.892 (d,  $^3J = 4.5$  Hz, 0.7 H, Fuc  $\beta$  H-1), 4.697 (s, 1 H, Man-3 H-1), 4.663 (d,  $^3J = 8.5$  Hz, 0.3 H, GlcNAc-2  $\alpha$  H-1), 4.658 (d,  $^3J = 9.0$  Hz, 0.7 H, GlcNAc-2 H-1), 4.555 (d,  $^3J = 9.4$  Hz, 1 H, GlcNAc-5 H-1), 4.546 (d,  $^3J = 9.6$  Hz, 1 H, GlcNAc-5' H-1), 4.464 (d,  $^3J = 9.2$  Hz, 1 H, Bisec-GlcNAc H-1), 4.248 (m, 1 H, Man-4 H-2), 2.095 (s, 2 H, GlcNAc-2  $\beta$  NAc), 2.092 (s, 1 H, GlcNAc-2  $\alpha$  NAc), 2.066 (s, 3 H, Bisec-GlcNAc NAc), 2.058 (s, 3 H, GlcNAc-5 NAc), 2.050 (s, 3 H, GlcNAc-5' NAc), 2.038 (s, 3 H, GlcNAc-1 NAc), 1.217 (d,  $^3J = 7.3$  Hz, 2 H, Fuc  $\beta$   $\text{CH}_3$ ), 1.207 (d,  $^3J = 7.1$  Hz, 1 H, Fuc  $\alpha$   $\text{CH}_3$ ).

**G1BF**  $m/z = 914.85$ 


$^1\text{H-NMR}$  (700 MHz,  $^3\text{DCCS}$ , 300 K)  
 $\delta$  [ppm] = 5.179 (d,  $^3J = 3.5$  Hz, 0.3 H, GlcNAc-1  $\alpha$  H-1), 5.059 (s, 1 H, Man-4 H-1), 5.011 (s, 1 H, Man-4' H-1), 4.907 (d,  $^3J = 4.3$  Hz, 0.3 H, Fuc  $\alpha$  H-1), 4.900 (d,  $^3J = 4.5$  Hz, 0.7 H, Fuc  $\beta$  H-1), 4.687 (s, 1 H, Man-3 H-1), 4.660 (d,  $^3J = 9.0$  Hz, 0.3 H, GlcNAc-2  $\alpha$  H-1), 4.657 (d,  $^3J = 9.2$  Hz, 0.7 H, GlcNAc-2 H-1), 4.579 (d,  $^3J = 9.8$  Hz, 1 H, GlcNAc-5' H-1), 4.556 (d,  $^3J = 9.4$  Hz, 1 H, GlcNAc-5 H-1), 4.478 (d,  $^3J = 8.8$  Hz, 1 H, Gal-6' H-1), 4.464 (d,  $^3J = 9.8$  Hz, 1 H, Bisec-GlcNAc H-1), 4.250 (m, 1 H, Man-4 H-2), 4.175 (d,  $^3J = 3.1$  Hz, 1 H, Man-3 H-2), 4.142 (s, 1 H, Man-4' H-2), 2.099 (s, 2 H, GlcNAc-2  $\beta$  NAc), 2.097 (s, 1 H, GlcNAc-2  $\alpha$  NAc), 2.066 (s, 3 H, Bisec-GlcNAc NAc), 2.056 (s, 3 H, GlcNAc-5 NAc), 2.039 (s, 3 H, GlcNAc-5' NAc), 2.039 (s, 3 H, GlcNAc-1 NAc), 1.218 (d,  $^3J = 7.5$  Hz, 2 H, Fuc  $\beta$   $\text{CH}_3$ ), 1.208 (d,  $^3J = 7.3$  Hz, 1 H, Fuc  $\alpha$   $\text{CH}_3$ ).

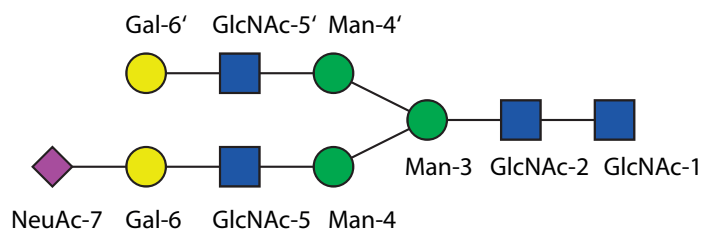
**G1F**  $m/z = 813.31$ 


$^1\text{H-NMR}$  (700 MHz,  $^3\text{DCCS}$ , 300 K)  
 $\delta$  [ppm] = 5.180 (d,  $^3J = 3.5$  Hz, 0.3 H, GlcNAc-1  $\alpha$  H-1), 5.115 (s, 1 H, Man-4 H-1), 4.926 (s, 1 H, Man-4' H-1), 4.896 (d,  $^3J = 4.3$  Hz, 0.3 H, Fuc  $\alpha$  H-1), 4.890 (d,  $^3J = 4.1$  Hz, 0.7 H, Fuc  $\beta$  H-1), 4.693 (d,  $^3J = 8.5$  Hz, 0.7 H, GlcNAc-1  $\beta$  H-1), 4.670 (d,  $^3J = 8.8$  Hz, 0.3 H, GlcNAc-2  $\alpha$  H-1), 4.664 (d,  $^3J = 8.8$  Hz, 0.7 H, GlcNAc-2  $\beta$  H-1), 4.582 (d,  $^3J = 9.4$  Hz, 1 H, GlcNAc-5' H-1), 4.554 (d,  $^3J = 9.4$  Hz, 1 H, GlcNAc-5 H-1), 4.473 (d,  $^3J = 8.8$  Hz, 1 H, Gal-6' H-1), 4.252 (d,  $^3J = 1.9$  Hz, 1 H, Man-3 H-2), 4.189 (m, 1 H, Man-4 H-2), 4.105 (s, 1 H, Man-4' H-2), 2.097 (s, 2 H, GlcNAc-2  $\beta$  NAc), 2.095 (s, 1 H, GlcNAc-2  $\alpha$  NAc), 2.053 (s, 3 H, GlcNAc-5 NAc), 2.047 (s, 3 H, GlcNAc-5' NAc), 2.039 (s, 3 H, GlcNAc-1 NAc), 1.220 (d,  $^3J = 7.3$  Hz, 2 H, Fuc  $\beta$   $\text{CH}_3$ ), 1.209 (d,  $^3J = 7.3$  Hz, 1 H, Fuc  $\alpha$   $\text{CH}_3$ ).

**G2F**  $m/z = 894.34$ 


$^1\text{H-NMR}$  (700 MHz,  $3\text{DCCs}$ , 300 K)  
 $\delta$  [ppm] = 5.180 (d,  $^3J = 3.4$  Hz, 0.3 H, GlcNAc-1  $\alpha$  H-1), 5.118 (s, 1 H, Man-4 H-1), 4.926 (s, 1 H, Man-4' H-1), 4.897 (d,  $^3J = 4.5$  Hz, 0.3 H, Fuc  $\alpha$  H-1), 4.890 (d,  $^3J = 4.3$  Hz, 0.7 H, Fuc  $\beta$  H-1), 4.693 (d,  $^3J = 8.5$  Hz, 0.7 H, GlcNAc-1  $\beta$  H-1), 4.669

(d,  $^3J = 8.8$  Hz, 0.3 H, GlcNAc-2  $\alpha$  H-1), 4.665 (d,  $^3J = 8.8$  Hz, 0.7 H, GlcNAc-2  $\beta$  H-1), 4.582 (d,  $^3J = 9.2$  Hz, 2 H, GlcNAc-5/5' H-1), 4.473 (d,  $^3J = 9.0$  Hz, 1 H, Gal-6' H-1), 4.467 (d,  $^3J = 8.5$  Hz, 1 H, Gal-6 H-1), 4.252 (d,  $^3J = 2.6$  Hz, 1 H, Man-3 H-2), 4.191 (m, 1 H, Man-4 H-2), 4.105 (s, 1 H, Man-4' H-2), 2.097 (s, 2 H, GlcNAc-2  $\beta$  NAc), 2.094 (s, 1 H, GlcNAc-2  $\alpha$  NAc), 2.051 (s, 3 H, GlcNAc-5 NAc), 2.048 (s, 3 H, GlcNAc-5' NAc), 2.039 (s, 3 H, GlcNAc-1 NAc), 1.220 (d,  $^3J = 7.1$  Hz, 2 H, Fuc  $\beta$   $\text{CH}_3$ ), 1.210 (d,  $^3J = 7.1$  Hz, 1 H, Fuc  $\alpha$   $\text{CH}_3$ ).

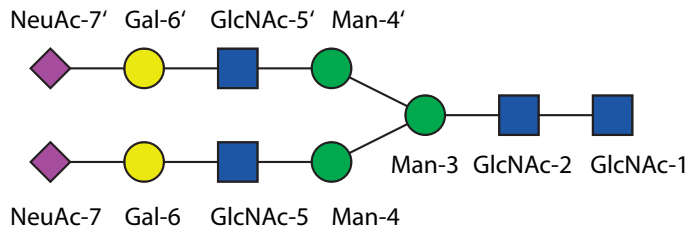
**G2S1**  $m/z = 966.86$ 


$^1\text{H-NMR}$  (700 MHz,  $3\text{DCCs}$ , 300 K)  
 $\delta$  [ppm] = 5.189 (d,  $^3J = 2.9$  Hz, 0.3 H, GlcNAc-1  $\alpha$  H-1), 5.136 (s, 1 H, Man-4 H-1), 4.932 (s, 1 H, Man-4' H-1), 4.696 (d,  $^3J = 9.0$  Hz, 0.7 H, GlcNAc-1  $\beta$  H-1), 4.607 (d,  $^3J = 8.8$  Hz, 1 H, GlcNAc-5 H-1),

4.582 (d,  $^3J = 9.4$  Hz, 1 H, GlcNAc-5' H-1), 4.473 (d,  $^3J = 8.8$  Hz, 1 H, Gal-6' H-1), 4.445 (d,  $^3J = 8.8$  Hz, 1 H, Gal-6 H-1), 4.255 (s, 1 H, Man-3 H-2), 4.197 (d,  $^3J = 3.4$  Hz, 1 H, Man-4 H-2), 4.111 (s, 1 H, Man-4' H-2), 2.667 (dd,  $^3J_{\text{H-3}\ddot{\text{a}}\text{q}, \text{H-4}} = 5.5$  Hz,  $^3J_{\text{H-3}\ddot{\text{a}}\text{q}, \text{H-3ax}} = 13.9$  Hz, 1 H, NeuAc-7 H-3 $\ddot{\text{a}}\text{q}$ ), 2.081 (s, 3 H, GlcNAc-2 NAc), 2.070 (s, 3 H, GlcNAc-5 NAc), 2.048 (s, 3 H, GlcNAc-5' NAc), 2.039 (s, 3 H, GlcNAc-1 NAc), 2.030 (s, 3 H, NeuAc-7 NAc), 1.725 (dd,  $^3J_{\text{H-3ax}, \text{H-4}} = 14.0$  Hz,  $^3J_{\text{H-3ax}, \text{H-3}\ddot{\text{a}}\text{q}} = 13.9$  Hz, 1 H, NeuAc-7 H-3ax).



G2S2  $m/z = 1112.41$



$^1\text{H-NMR}$  (700 MHz,  $^3\text{DCC}_8$ , 300 K)

$\delta$  [ppm] = 5.189 (d,  $^3J = 3.1$  Hz, 0.3 H,

GlcNAc-1  $\alpha$  H-1), 5.133 (s, 1 H, Man-4

H-1), 4.949 (s, 1 H, Man-4' H-1), 4.693

(d,  $^3J = 8.2$  Hz, 0.7 H, GlcNAc-1  $\beta$

H-1), 4.614 (d,  $^3J = 8.5$  Hz, 0.3 H,

GlcNAc-2  $\alpha$  H-1), 4.604 (d,

$^3J = 9.4$  Hz, 2.7 H, GlcNAc-5/5' H-1, GlcNAc-2  $\beta$  H-1), 4.448 (d,  $^3J = 8.8$  Hz, 1 H, Gal-6' H-1),

4.444 (d,  $^3J = 8.8$  Hz, 1 H, Gal-6 H-1), 4.259 (s, 1 H, Man-3 H-2), 4.199 (d,  $^3J = 3.4$  Hz, 1 H,

Man-4 H-2), 4.119 (d,  $^3J = 3.4$  Hz, 1 H, Man-4' H-2), 2.673 (dd,  $^3J_{\text{H-3äq, H-4}} = 4.6$  Hz,  $^3J_{\text{H-3äq, H-3ax}} = 13.6$  Hz, 1 H, NeuAc-7 H-3äq),








2.667 (dd,  $^3J_{\text{H-3äq, H-4}} = 4.4$  Hz,  $^3J_{\text{H-3äq, H-3ax}} = 13.6$  Hz, 1 H, NeuAc-7' H-3äq), 2.085 (s, 3 H, GlcNAc-2 NAc), 2.070 (s, 3 H, GlcNAc-5 NAc), 2.066 (s,





3 H, GlcNAc-5' NAc), 2.038 (s, 3 H, GlcNAc-1 NAc), 2.030 (s, 6 H, NeuAc-7/7' NAc), 1.722

(dd,  $^3J_{\text{H-3ax, H-4}} = 13.8$  Hz,  $^3J_{\text{H-3ax, H-3äq}} = 13.8$  Hz, 1 H, NeuAc-7/7' H-3ax).

## 7.9 Sicherheitshinweise

**Tabelle 17: Gefahrenpiktogramme, H- und P-Sätze der in dieser Arbeit verwendeten Chemikalien.**

Substanz	Gefahrenpiktogramm	H/P-Sätze
Acetonitril		H225, H332, H302, H312, H319, P210, P240, P302+P352, P305+P351+P338, P403+P233
Ameisensäure 98%		H226, H302, H314, H331, EUH071, P210, P280, P303+P361+P353, P304+P340+P310, P305+P351+P338, P403+P233
2-Amino-2-(hydroxy- methyl)- propan-1,3-diol	Kein gefährlicher Stoff nach GHS	
Ammoniumformiat		H315, H319, H335, P261, P305+P351+P338
Ammoniumhydrogen- carbonat		H302, P301+P312, P330
Calciumchlorid		H319, P305+P351+P338
Deuteriumoxid	Kein gefährlicher Stoff nach GHS	
1,4-D/L-Dithiothreitol		H302, H315, H319, H335, P261, P305+P351+P338
Ethanol		H225, H319, P210, P240, P305+P351+P338, P403+P233
Harnstoff	Kein gefährlicher Stoff nach GHS	
Iodacetamid		H301, H317, H334, H413, P261, P280, P301+P310, P342+P311
Natriumdodecylsulfat		H228, H302+H332, H315, H318, H335, H412, P210, P261, P280, P301+P312+P330, P305+P315+P338+P310, P370+P378

Substanz	Gefahrenpiktogramm	H/P-Sätze
Nonident P-40		H302, H318, H411, P273, P280, P305+P451,P338
N-Glycosidase F	Kein gefährlicher Stoff nach GHS	
Pronase E		H315, H319, H334, H335, P261, P305+P351+P338, P342+P311
Tuning Mix ES (ESI) (Pos)		H225, H302+H332, H319, P210, P305+P351+P338
Trypsin Gold Mass Spectrometry Grade		H315, H319, H334, H335, P264, P280, P284, P302+P352, P305+P351+P338, P312

### 7.9.1 Arbeiten mit biologischem Probenmaterial

Nicht gezielte Tätigkeiten mit Blutplasmaproben, welche negativ auf HIV, HBV und HCV getestet wurden, wurden der Risikogruppe 1 zugeordnet und unter den Bedingungen der Schutzstufe 1 durchgeführt (BioStoffV, TRBA100).

Nicht gezielte Tätigkeiten mit Blutplasmaproben, welche nicht auf HIV, HBV und HCV getestet wurden bzw. positiv getestet sind, wurden der Risikogruppe 2 zugeordnet und unter den Bedingungen der Schutzstufe 2 durchgeführt (BioStoffV, TRBA100).

## 8 Literaturverzeichnis

1. Varki, A., Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **1993**, 3, (2), 97-130.
2. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V., Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Bio* **2012**, 13, (7), 448-462.
3. Lauc, G.; Pezer, M.; Rudan, I.; Campbell, H., Mechanisms of disease: The human N-glycome. *Biochim Biophys Acta* **2016**, 1860, (8), 1574-82.
4. Cummings, R. D., The repertoire of glycan determinants in the human glycome. *Mol Biosyst* **2009**, 5, (10), 1087-1104.
5. Ley, K., The role of selectins in inflammation and disease. *Trends Mol Med* **2003**, 9, (6), 263-8.
6. Varki, A.; Gagneux, P., Multifarious roles of sialic acids in immunity. *Ann N Y Acad Sci* **2012**, 1253, 16-36.
7. Hennet, T., Diseases of glycosylation beyond classical congenital disorders of glycosylation. *Biochim Biophys Acta* **2012**, 1820, (9), 1306-17.
8. Dell, A., Structures of glycoprotein glycans. *Aust J Chem* **2002**, 55, (1-2), 27-37.
9. Dennis, J. W.; Granovsky, M.; Warren, C. E., Protein glycosylation in development and disease. *Bioessays* **1999**, 21, (5), 412-421.
10. Collin, M.; Ehlers, M., The carbohydrate switch between pathogenic and immunosuppressive antigen-specific antibodies. *Exp Dermatol* **2013**, 22, (8), 511-514.
11. Iida, S.; Misaka, H.; Inoue, M.; Shibata, M.; Nakano, R.; Yamane-Ohnuki, N.; Wakitani, M.; Yano, K.; Shitara, K.; Satoh, M., Nonfucosylated therapeutic IgG1 antibody can evade the inhibitory effect of serum immunoglobulin g on antibody-dependent cellular cytotoxicity through its high binding to Fc gamma RIIIa. *Clin Cancer Res* **2006**, 12, (9), 2879-2887.
12. Hanahan, D.; Weinberg, R. A., The hallmarks of cancer. *Cell* **2000**, 100, (1), 57-70.

13. Zhao, Y. Y.; Takahashi, M.; Gu, J. G.; Miyoshi, E.; Matsumoto, A.; Kitazume, S.; Taniguchi, N., Functional roles of N-glycans in cell signaling and cell adhesion in cancer. *Cancer Sci* **2008**, 99, (7), 1304-1310.
14. Pinho, S. S.; Reis, C. A., Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **2015**, 15, (9), 540-555.
15. Stowell, S. R.; Ju, T. Z.; Cummings, R. D., Protein Glycosylation in Cancer. *Annu Rev Pathol-Mech* **2015**, 10, 473-510.
16. Kobata, A.; Amano, J., Altered glycosylation of proteins produced by malignant cells, and application for the diagnosis and immunotherapy of tumours. *Immunol Cell Biol* **2005**, 83, (4), 429-439.
17. Christiansen, M. N.; Chik, J.; Lee, L.; Anugraham, M.; Abrahams, J. L.; Packer, N. H., Cell surface protein glycosylation in cancer. *Proteomics* **2014**, 14, (4-5), 525-46.
18. Gu, J. G.; Isaji, T.; Xu, Q. S.; Kariya, Y.; Gu, W.; Fukuda, T.; Du, Y. G., Potential roles of N-glycosylation in cell adhesion. *Glycoconjugate J* **2012**, 29, (8-9), 599-607.
19. Asada, M.; Furukawa, K.; Segawa, K.; Endo, T.; Kobata, A., Increased expression of highly branched N-glycans at cell surface is correlated with the malignant phenotypes of mouse tumor cells. *Cancer Res* **1997**, 57, (6), 1073-80.
20. Hauselmann, I.; Borsig, L., Altered tumor-cell glycosylation promotes metastasis. *Front Oncol* **2014**, 4, 28.
21. Varki, A., Selectin Ligands. *P Natl Acad Sci USA* **1994**, 91, (16), 7390-7397.
22. Vestweber, D.; Blanks, J. E., Mechanisms that regulate the function of the selectins and their ligands. *Physiol Rev* **1999**, 79, (1), 181-213.
23. Peters, T., A matter of order: how E-selectin makes sweet contacts. *Chembiochem* **2012**, 13, (16), 2325-6.
24. Reis, C. A.; Osorio, H.; Silva, L.; Gomes, C.; David, L., Alterations in glycosylation as biomarkers for cancer detection. *J Clin Pathol* **2010**, 63, (4), 322-329.
25. Ruhaak, L. R.; Miyamoto, S.; Lebrilla, C. B., Developments in the Identification of Glycan Biomarkers for the Detection of Cancer. *Mol Cell Proteomics* **2013**, 12, (4), 846-855.

26. Liang, Y.; Ma, T.; Thakur, A.; Yu, H.; Gao, L.; Shi, P.; Li, X.; Ren, H.; Jia, L.; Zhang, S.; Li, Z.; Chen, M., Differentially expressed glycosylated patterns of alpha-1-antitrypsin as serum biomarkers for the diagnosis of lung cancer. *Glycobiology* **2015**, *25*, (3), 331-40.
27. Arnold, J. N.; Saldova, R.; Galligan, M. C.; Murphy, T. B.; Mimura-Kimura, Y.; Telford, J. E.; Godwin, A. K.; Rudd, P. M., Novel glycan biomarkers for the detection of lung cancer. *J Proteome Res* **2011**, *10*, (4), 1755-64.
28. Ruhaak, L. R.; Taylor, S. L.; Stroble, C.; Nguyen, U. T.; Parker, E. A.; Song, T.; Lebrilla, C. B.; Rom, W. N.; Pass, H.; Kim, K.; Kelly, K.; Miyamoto, S., Differential N-Glycosylation Patterns in Lung Adenocarcinoma Tissue. *J Proteome Res* **2015**, *14*, (11), 4538-49.
29. Ruhaak, L. R.; Stroble, C.; Dai, J. L.; Barnett, M.; Taguchi, A.; Goodman, G. E.; Miyamoto, S.; Gandara, D.; Feng, Z. D.; Lebrilla, C. B.; Hanash, S., Serum Glycans as Risk Markers for Non-Small Cell Lung Cancer. *Cancer Prev Res* **2016**, *9*, (4), 317-323.
30. Ruhaak, L. R.; Kim, K.; Stroble, C.; Taylor, S. L.; Hong, Q.; Miyamoto, S.; Lebrilla, C. B.; Leiserowitz, G., Protein-Specific Differential Glycosylation of Immunoglobulins in Serum of Ovarian Cancer Patients. *J Proteome Res* **2016**, *15*, (3), 1002-10.
31. Alley, W. R., Jr.; Vasseur, J. A.; Goetz, J. A.; Svoboda, M.; Mann, B. F.; Matei, D. E.; Menning, N.; Hussein, A.; Mechref, Y.; Novotny, M. V., N-linked glycan structures and their expressions change in the blood sera of ovarian cancer patients. *J Proteome Res* **2012**, *11*, (4), 2282-300.
32. Snyder, C. M.; Alley, W. R., Jr.; Campos, M. I.; Svoboda, M.; Goetz, J. A.; Vasseur, J. A.; Jacobson, S. C.; Novotny, M. V., Complementary Glycomic Analyses of Sera Derived from Colorectal Cancer Patients by MALDI-TOF-MS and Microchip Electrophoresis. *Anal Chem* **2016**, *88*, (19), 9597-9605.
33. Sethi, M. K.; Hancock, W. S.; Fanayan, S., Identifying N-Glycan Biomarkers in Colorectal Cancer by Mass Spectrometry. *Acc Chem Res* **2016**, *49*, (10), 2099-2106.
34. Jensen, P. H.; Karlsson, N. G.; Kolarich, D.; Packer, N. H., Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc* **2012**, *7*, (7), 1299-310.

35. Wuhrer, M.; Deelder, A. M.; Hokke, C. H., Protein glycosylation analysis by liquid chromatography-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **2005**, 825, (2), 124-33.
36. Ruhaak, L. R.; Deelder, A. M.; Wuhrer, M., Oligosaccharide analysis by graphitized carbon liquid chromatography-mass spectrometry. *Anal Bioanal Chem* **2009**, 394, (1), 163-74.
37. Zauner, G.; Deelder, A. M.; Wuhrer, M., Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. *Electrophoresis* **2011**, 32, (24), 3456-66.
38. Kellie, J. F.; Tran, J. C.; Lee, J. E.; Ahlf, D. R.; Thomas, H. M.; Ntai, I.; Catherman, A. D.; Durbin, K. R.; Zamdborg, L.; Vellaichamy, A.; Thomas, P. M.; Kelleher, N. L., The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. *Mol Biosyst* **2010**, 6, (9), 1532-9.
39. Zhou, H.; Ning, Z.; Starr, A. E.; Abu-Farha, M.; Figeys, D., Advancements in top-down proteomics. *Anal Chem* **2012**, 84, (2), 720-34.
40. Capriotti, A. L.; Cavaliere, C.; Foglia, P.; Samperi, R.; Lagana, A., Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics. *J Chromatogr A* **2011**, 1218, (49), 8760-76.
41. Lanucara, F.; Eyers, C. E., Top-down mass spectrometry for the analysis of combinatorial post-translational modifications. *Mass Spectrom Rev* **2013**, 32, (1), 27-42.
42. Domon, B.; Costello, C. E., A Systematic Nomenclature for Carbohydrate Fragmentations in Fab-MS MS Spectra of Glycoconjugates. *Glycoconjugate J* **1988**, 5, (4), 397-409.
43. 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, (1-2), 115-28.
44. Vliegthart, J. F. G.; Dorland, L.; Vanhalbeek, H., High-Resolution, H-1-Nuclear Magnetic-Resonance Spectroscopy as a Tool in the Structural-Analysis of Carbohydrates Related to Glycoproteins. *Adv Carbohydr Chem Bi* **1983**, 41, 209-374.

45. Vliegthart, J. F. G., Introduction to NMR spectroscopy of carbohydrates. *Acs Sym Ser* **2006**, 930, 1-19.
46. Vliegthart, J. F. G.; Vanhalbeek, H.; Dorland, L., The Applicability of 500-Mhz High-Resolution H-1-Nmr Spectroscopy for the Structure Determination of Carbohydrates Derived from Glycoproteins. *Pure Appl Chem* **1981**, 53, (1), 45-77.
47. Vankuik, J. A.; Hard, K.; Vliegthart, J. F. G., A H-1-Nmr Database Computer-Program for the Analysis of the Primary Structure of Complex Carbohydrates. *Carbohydr Res* **1992**, 235, 53-68.
48. Vankuik, J. A.; Vliegthart, J. F. G., Databases of Complex Carbohydrates. *Trends Biotechnol* **1992**, 10, (6), 182-185.
49. Fellenberg, M.; Coksezen, A.; Meyer, B., Characterization of picomole amounts of oligosaccharides from glycoproteins by 1H NMR spectroscopy. *Angew Chem Int Ed Engl* **2010**, 49, (14), 2630-3.
50. Behnken, H. N.; Fellenberg, M.; Koetzler, M. P.; Jirmann, R.; Nagel, T.; Meyer, B., Resolving the problem of chromatographic overlap by 3D cross correlation (3DCC) processing of LC, MS and NMR data for characterization of complex glycan mixtures. *Anal Bioanal Chem* **2012**, 404, (5), 1427-37.
51. Fellenberg, M.; Behnken, H. N.; Nagel, T.; Wiegandt, A.; Baerenfaenger, M.; Meyer, B., Glycan analysis: scope and limitations of different techniques--a case for integrated use of LC-MS(/MS) and NMR techniques. *Anal Bioanal Chem* **2013**, 405, (23), 7291-305.
52. Fellenberg, M. Entwicklung neuer Strategien zur Analytik von Glycoproteinen, Dissertation. Universität Hamburg, Hamburg, **2011**.
53. Behnken, H. N. New Methods for the Characterization of N-Type Glycans of Glycoproteins: Integration of LC-MS, MS/MS and NMR Spectroscopy, Dissertation. University of Hamburg, Hamburg, **2013**.
54. Ferlay, J.; Soerjomataram, I.; Ervik, M.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D. M.; Forman, D.; Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; **2013**. Available from: <http://globocan.iarc.fr>, accessed on 31/10/2016.



55. Krebs in Deutschland 2011/2012. 10. Ausgabe. Robert-Koch-Institut (Hrsg) und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (Hrsg). Berlin, **2015**.
56. Schutte, K.; Schulz, C.; Malfertheiner, P., Hepatocellular Carcinoma: Current Concepts in Diagnosis, Staging and Treatment. *Gastrointest Tumors* **2014**, 1, (2), 84-92.
57. Flores, A.; Marrero, J. A., Emerging trends in hepatocellular carcinoma: focus on diagnosis and therapeutics. *Clin Med Insights Oncol* **2014**, 8, 71-6.
58. Zhang, B. H.; Yang, B. H.; Tang, Z. Y., Randomized controlled trial of screening for hepatocellular carcinoma. *J Cancer Res Clin Oncol* **2004**, 130, (7), 417-22.
59. Stefaniuk, P.; Cianciara, J.; Wiercinska-Drapalo, A., Present and future possibilities for early diagnosis of hepatocellular carcinoma. *World J Gastroenterol* **2010**, 16, (4), 418-24.
60. Zhou, L.; Liu, J.; Luo, F., Serum tumor markers for detection of hepatocellular carcinoma. *World J Gastroenterol* **2006**, 12, (8), 1175-81.
61. Sato, Y.; Nakata, K.; Kato, Y.; Shima, M.; Ishii, N.; Koji, T.; Taketa, K.; Endo, Y.; Nagataki, S., Early recognition of hepatocellular carcinoma based on altered profiles of alpha-fetoprotein. *N Engl J Med* **1993**, 328, (25), 1802-6.
62. Aoyagi, Y.; Isokawa, O.; Suda, T.; Watanabe, M.; Suzuki, Y.; Asakura, H., The fucosylation index of alpha-fetoprotein as a possible prognostic indicator for patients with hepatocellular carcinoma. *Cancer* **1998**, 83, (10), 2076-82.
63. Naitoh, A.; Aoyagi, Y.; Asakura, H., Highly enhanced fucosylation of serum glycoproteins in patients with hepatocellular carcinoma. *J Gastroen Hepatol* **1999**, 14, (5), 436-445.
64. Comunale, M. A.; Wang, M. J.; Hafner, J.; Krakover, J.; Rodemich, L.; Kopenhaver, B.; Long, R. E.; Junaidi, O.; Di Bisceglie, A. M.; Block, T. M.; Mehta, A. S., Identification and Development of Fucosylated Glycoproteins as Biomarkers of Primary Hepatocellular Carcinoma. *Journal of Proteome Research* **2009**, 8, (2), 595-602.
65. Goldman, R.; Resson, H. W.; Varghese, R. S.; Goldman, L.; Bascug, G.; Loffredo, C. A.; Abdel-Hamid, M.; Gouda, I.; Ezzat, S.; Kyselova, Z.; Mechref, Y.; Novotny, M. V.,

- Detection of Hepatocellular Carcinoma Using Glycomic Analysis. *Clin Cancer Res* **2009**, 15, (5), 1808-1813.
66. Wang, M. J.; Long, R. E.; Comunale, M. A.; Junaidi, O.; Marrero, J.; Di Bisceglie, A. M.; Block, T. M.; Mehta, A. S., Novel Fucosylated Biomarkers for the Early Detection of Hepatocellular Carcinoma. *Cancer Epidem Biomar* **2009**, 18, (6), 1914-1921.
67. Liu, X. E.; Desmyter, L.; Gao, C. F.; Laroy, W.; Dewaele, S.; Vanhooren, V.; Wang, L.; Zhuang, H.; Callewaert, N.; Libert, C.; Contreras, R.; Chen, C., N-glycomic changes in hepatocellular carcinoma patients with liver cirrhosis induced by hepatitis B virus. *Hepatology* **2007**, 46, (5), 1426-1435.
68. Yi, C. H.; Weng, H. L.; Zhou, F. G.; Fang, M.; Ji, J.; Cheng, C.; Wang, H.; Liebe, R.; Dooley, S.; Gao, C. F., Elevated core-fucosylated IgG is a new marker for hepatitis B virus-related hepatocellular carcinoma. *Oncoimmunology* **2015**, 4, (12), e1011503.
69. Tanabe, K.; Kitagawa, K.; Kojima, N.; Iijima, S., Multifucosylated Alpha-1-acid Glycoprotein as a Novel Marker for Hepatocellular Carcinoma. *J Proteome Res* **2016**, 15, (9), 2935-44.
70. Weisel, J. W., Fibrinogen and fibrin. *Adv Protein Chem* **2005**, 70, 247-299.
71. Mosesson, M. W., Fibrinogen and fibrin structure and functions. *J Thromb Haemost* **2005**, 3, (8), 1894-904.
72. Kollman, J. M.; Pandi, L.; Sawaya, M. R.; Riley, M.; Doolittle, R. F., Crystal structure of human fibrinogen. *Biochemistry* **2009**, 48, (18), 3877-86.
73. Townsend, R. R.; Hilliker, E.; Li, Y. T.; Laine, R. A.; Bell, W. R.; Lee, Y. C., Carbohydrate structure of human fibrinogen. Use of 300-MHz <sup>1</sup>H-NMR to characterize glycosidase-treated glycopeptides. *J Biol Chem* **1982**, 257, (16), 9704-10.
74. Townsend, R. R.; Heller, D. N.; Fenselau, C. C.; Lee, Y. C., Determination of the sialylation pattern of human fibrinogen glycopeptides with fast atom bombardment. *Biochemistry* **1984**, 23, (26), 6389-92.
75. Adamczyk, B.; Struwe, W. B.; Ercan, A.; Nigrovic, P. A.; Rudd, P. M., Characterization of fibrinogen glycosylation and its importance for serum/plasma N-glycome analysis. *J Proteome Res* **2013**, 12, (1), 444-54.

76. Langer, B. G.; Weisel, J. W.; Dinauer, P. A.; Nagaswami, C.; Bell, W. R., Deglycosylation of fibrinogen accelerates polymerization and increases lateral aggregation of fibrin fibers. *J Biol Chem* **1988**, 263, (29), 15056-63.
77. Dang, C. V.; Shin, C. K.; Bell, W. R.; Nagaswami, C.; Weisel, J. W., Fibrinogen sialic acid residues are low affinity calcium-binding sites that influence fibrin assembly. *J Biol Chem* **1989**, 264, (25), 15104-8.
78. Beranova-Giorgianni, S.; Zhao, Y.; Desiderio, D. M.; Giorgianni, F., Phosphoproteomic analysis of the human pituitary. *Pituitary* **2006**, 9, (2), 109-20.
79. Blombaeck, B.; Blombaeck, M.; Edman, P.; Hessel, B., Amino-acid sequence and the occurrence of phosphorus in human fibrinopeptides. *Nature* **1962**, 193, 833-4.
80. Seydewitz, H. H.; Kaiser, C.; Rothweiler, H.; Witt, I., The location of a second in vivo phosphorylation site in the A alpha-chain of human fibrinogen. *Thromb Res* **1984**, 33, (5), 487-98.
81. Wind, M.; Gosenca, D.; Kubler, D.; Lehmann, W. D., Stable isotope phospho-profiling of fibrinogen and fetuin subunits by element mass spectrometry coupled to capillary liquid chromatography. *Anal Biochem* **2003**, 317, (1), 26-33.
82. Zahedi, R. P.; Lewandrowski, U.; Wiesner, J.; Wortelkamp, S.; Moebius, J.; Schutz, C.; Walter, U.; Gambaryan, S.; Sickmann, A., Phosphoproteome of resting human platelets. *J Proteome Res* **2008**, 7, (2), 526-34.
83. Bian, Y.; Song, C.; Cheng, K.; Dong, M.; Wang, F.; Huang, J.; Sun, D.; Wang, L.; Ye, M.; Zou, H., An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *J Proteomics* **2014**, 96, 253-62.
84. Tagliabracci, V. S.; Wiley, S. E.; Guo, X.; Kinch, L. N.; Durrant, E.; Wen, J.; Xiao, J.; Cui, J.; Nguyen, K. B.; Engel, J. L.; Coon, J. J.; Grishin, N.; Pinna, L. A.; Pagliarini, D. J.; Dixon, J. E., A Single Kinase Generates the Majority of the Secreted Phosphoproteome. *Cell* **2015**, 161, (7), 1619-32.
85. Zauner, G.; Hoffmann, M.; Rapp, E.; Koeleman, C. A.; Dragan, I.; Deelder, A. M.; Wuhrer, M.; Hensbergen, P. J., Glycoproteomic analysis of human fibrinogen reveals novel regions of O-glycosylation. *J Proteome Res* **2012**, 11, (12), 5804-14.
86. Wentura, D.; Pospeschill, M., *Multivariate Datenanalyse - Eine kompakte Einführung*. Springer: Wiesbaden, **2015**.

87. Handl, A., *Multivariate Analysemethoden - Theorie und Praxis multivariater Verfahren unter besonderer Berücksichtigung von S-PLUS (2. Auflage)*. Springer: Heidelberg Dordrecht London New York, **2010**.
88. Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R., *Multivariate Analysemethoden - Eine anwendungsorientierte Einführung (14. Auflage)*. Springer Gabler: Berlin Heidelberg, **2016**.
89. Zaiontz, C., Real Statistics Using Excel. [www.real-statistics.com](http://www.real-statistics.com). **2015**.
90. Härdle, W. K.; Simar, L., *Applied Multivariate Statistical Analysis (4th Edition)*. Springer: Heidelberg New York Dordrecht London, **2015**.
91. Nagel, T.; Meyer, B., Simultaneous characterization of sequence polymorphisms, glycosylation and phosphorylation of fibrinogen in a direct analysis by LC-MS. *Biochim Biophys Acta* **2014**, 1844, (12), 2284-9.
92. Lhote, C. L.; Berger, S.; Bourgerie, S.; Duvaliflah, Y.; Julien, R.; Karamanos, Y., Use of Porcine Fibrinogen as a Model Glycoprotein to Study the Binding-Specificity of the 3 Variants of K88 Lectin. *Infect Immun* **1995**, 63, (5), 1927-1932.
93. LHote, C.; Berger, S.; Karamanos, Y., O-glycosylation of fibrinogen from different mammalian species as revealed by the binding of Escherichia coli biotinylated lectins. *Thromb Haemostasis* **1996**, 76, (5), 710-714.
94. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. dbSNP accession: rs6050, (dbSNP Build ID: 52/147). Available from: <http://www.ncbi.nlm.nih.gov/SNP/>.
95. Haglund, A. C.; Ronquist, G.; Frithz, G.; Ek, P., Alteration of the fibrinogen molecule and its phosphorylation state in myocardial infarction patients undergoing thrombolytic treatment. *Thromb Res* **2000**, 98, (2), 147-56.
96. Ogata, Y.; Heppelmann, C. J.; Charlesworth, M. C.; Madden, B. J.; Miller, M. N.; Kalli, K. R.; Cilby, W. A.; Bergen, H. R., 3rd; Saggese, D. A.; Muddiman, D. C., Elevated levels of phosphorylated fibrinogen-alpha-isoforms and differential expression of other post-translationally modified proteins in the plasma of ovarian cancer patients. *J Proteome Res* **2006**, 5, (12), 3318-25.

97. Standeven, K. F.; Grant, P. J.; Carter, A. M.; Scheiner, T.; Weisel, J. W.; Ariens, R. A., Functional analysis of the fibrinogen Aalpha Thr312Ala polymorphism: effects on fibrin structure and function. *Circulation* **2003**, 107, (18), 2326-30.
98. Carter, A. M.; Catto, A. J.; Grant, P. J., Association of the alpha-fibrinogen Thr312Ala polymorphism with poststroke mortality in subjects with atrial fibrillation. *Circulation* **1999**, 99, (18), 2423-2426.
99. Zhang, J. J.; Yu, L. H.; Yin, Y. Y.; Lu, Q.; Lei, L.; Xiao, J.; Guo, J.; Zhao, J. J.; Wang, Y. A.; He, G. Q.; Xu, Y. M.; He, L., Association Between Two Functional Fibrinogen-Related Polymorphisms and Ischemic Stroke: A Case-Control Study. *Genet Test Mol Bioma* **2013**, 17, (11), 789-793.
100. Zeng, Y.; Zhang, L.; Hu, Z. P.; Yang, Q. D.; Ma, M. M.; Liu, B. Q.; Xia, J.; Xu, H. W.; Liu, Y. H.; Du, X. P., Fibrinogen polymorphisms associated with sporadic cerebral hemorrhage in a Chinese population. *J Clin Neurosci* **2012**, 19, (5), 753-756.
101. Carter, A. M.; Catto, A. J.; Kohler, H. P.; Ariens, R. A.; Stickland, M. H.; Grant, P. J., alpha-fibrinogen Thr312Ala polymorphism and venous thromboembolism. *Blood* **2000**, 96, (3), 1177-9.
102. Ko, Y. L.; Hsu, L. A.; Hsu, T. S.; Tsai, C. T.; Teng, M. S.; Wu, S.; Chang, C. J.; Lee, Y. S., Functional polymorphisms of FGA, encoding alpha fibrinogen, are associated with susceptibility to venous thromboembolism in a Taiwanese population. *Hum Genet* **2006**, 119, (1-2), 84-91.
103. Stavenhagen, K.; Plomp, R.; Wuhrer, M., Site-Specific Protein N- and O-Glycosylation Analysis by a C18-Porous Graphitized Carbon-Liquid Chromatography-Electrospray Ionization Mass Spectrometry Approach Using Pronase Treated Glycopeptides. *Anal Chem* **2015**, 87, (23), 11691-9.
104. Hoffmann, M.; Marx, K.; Reichl, U.; Wuhrer, M.; Rapp, E., Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins. *Mol Cell Proteomics* **2016**, 15, (2), 624-41.
105. Pacchiarotta, T.; Hensbergen, P. J.; Wuhrer, M.; van Nieuwkoop, C.; Nevedomskaya, E.; Derks, R. J.; Schoenmaker, B.; Koeleman, C. A.; van Dissel, J.; Deelder, A. M.; Mayboroda, O. A., Fibrinogen alpha chain O-glycopeptides as possible markers of urinary tract infection. *J Proteomics* **2012**, 75, (3), 1067-73.

106. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. dbSNP accession: rs4220, (dbSNP Build ID: 36/147). Available from: <http://www.ncbi.nlm.nih.gov/SNP/>.
107. Kathiresan, S.; Yang, Q.; Larson, M. G.; Camargo, A. L.; Tofler, G. H.; Hirschhorn, J. N.; Gabriel, S. B.; O'Donnell, C. J., Common genetic variation in five thrombosis genes and relations to plasma hemostatic protein level and cardiovascular disease risk. *Arterioscler Thromb Vasc Biol* **2006**, 26, (6), 1405-12.
108. Ong, K. L.; Tso, A. W.; Cherny, S. S.; Sham, P. C.; Lam, K. S.; Jiang, C. Q.; Thomas, G. N.; Lam, T. H.; Cheung, B. M., A genetic variant in the gene encoding fibrinogen beta chain predicted development of hypertension in Chinese men. *Thromb Haemost* **2010**, 103, (4), 728-35.
109. Klovaite, J.; Nordestgaard, B. G.; Tybjaerg-Hansen, A.; Benn, M., Elevated fibrinogen levels are associated with risk of pulmonary embolism, but not with deep venous thrombosis. *Am J Respir Crit Care Med* **2013**, 187, (3), 286-93.
110. Jeff, J. M.; Brown-Gentry, K.; Crawford, D. C., Replication and characterisation of genetic variants in the fibrinogen gene cluster with plasma fibrinogen levels and haematological traits in the Third National Health and Nutrition Examination Survey. *Thromb Haemost* **2012**, 107, (3), 458-67.
111. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. dbSNP accession: rs6063, (dbSNP Build ID: 52/147). Available from: <http://www.ncbi.nlm.nih.gov/SNP/>.
112. Lek, M.; Karczewski, K.; Minikel, E.; Samocha, K.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.; Ware, J.; Hill, A.; Cummings, B.; Tukiainen, T.; Birnbaum, D.; Kosmicki, J.; Duncan, L.; Estrada, K.; Zhao, F.; Zou, J.; Pierce-Hoffman, E.; Cooper, D.; DePristo, M.; Do, R.; Flannick, J.; Fromer, M.; Gauthier, L.; Goldstein, J.; Gupta, N.; Howrigan, D.; Kiezun, A.; Kurki, M.; Levy Moonshine, A.; Natarajan, P.; Orozco, L.; Peloso, G.; Poplin, R.; Rivas, M.; Ruano-Rubio, V.; Ruderfer, D.; Shakir, K.; Stenson, P.; Stevens, C.; Thomas, B.; Tiao, G.; Tusie-Luna, M.; Weisburd, B.; Won, H.-H.; Yu, D.; Altshuler, D.; Ardissino, D.; Boehnke, M.; Danesh, J.; Roberto, E.; Florez, J.; Gabriel, S.; Getz, G.; Hultman, C.; Kathiresan, S.; Laakso, M.; McCarroll, S.; McCarthy, M.; McGovern, D.; McPherson, R.; Neale, B.; Palotie, A.; Purcell, S.; Saleheen, D.; Scharf, J.; Sklar, P.;

- Patrick, S.; Tuomilehto, J.; Watkins, H.; Wilson, J.; Daly, M.; MacArthur, D., Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv* **2015**.
113. Martinez, J.; Keane, P. M.; Gilman, P. B.; Palascak, J. E., The abnormal carbohydrate composition of the dysfibrinogenemia associated with liver disease. *Ann N Y Acad Sci* **1983**, 408, 388-96.
114. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. (dbSNP Build ID: 144/147). Available from: <http://www.ncbi.nlm.nih.gov/SNP/>.
115. Makogonenko, E.; Tsurupa, G.; Ingham, K.; Medved, L., Interaction of fibrin(ogen) with fibronectin: further characterization and localization of the fibronectin-binding site. *Biochemistry* **2002**, 41, (25), 7907-13.
116. Wuhrer, M.; Koeleman, C. A.; Hokke, C. H.; Deelder, A. M., Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments. *Rapid Commun Mass Spectrom* **2006**, 20, (11), 1747-54.
117. Ang, I. L.; Poon, T. C. W.; Lai, P. B. S.; Chan, A. T. C.; Ngai, S. M.; Hui, A. Y.; Johnson, P. J.; Sung, J. J. Y., Study of serum haptoglobin and its glycoforms in the diagnosis of hepatocellular carcinoma: A glycoproteomic approach. *Journal of Proteome Research* **2006**, 5, (10), 2691-2700.
118. Pompach, P.; Brnakova, Z.; Sanda, M.; Wu, J.; Edwards, N.; Goldman, R., Site-specific Glycoforms of Haptoglobin in Liver Cirrhosis and Hepatocellular Carcinoma. *Mol Cell Proteomics* **2013**, 12, (5), 1281-1293.
119. Lee, H. J.; Cha, H. J.; Lim, J. S.; Lee, S. H.; Song, S. Y.; Kim, H.; Hancock, W. S.; Yoo, J. S.; Paik, Y. K., Abundance-ratio-based semiquantitative analysis of site-specific N-linked glycopeptides present in the plasma of hepatocellular carcinoma patients. *J Proteome Res* **2014**, 13, (5), 2328-38.
120. Zhu, J.; Lin, Z.; Wu, J.; Yin, H.; Dai, J.; Feng, Z.; Marrero, J.; Lubman, D. M., Analysis of serum haptoglobin fucosylation in hepatocellular carcinoma and liver cirrhosis of different etiologies. *J Proteome Res* **2014**, 13, (6), 2986-97.
121. Comunale, M. A.; Lowman, M.; Long, R. E.; Krakover, J.; Philip, R.; Seeholzer, S.; Evans, A. A.; Hann, H. W. L.; Block, T. M.; Mehta, A. S., Proteomic analysis of serum

- associated fucosylated glycoproteins in the development of primary hepatocellular carcinoma. *Journal of Proteome Research* **2006**, 5, (2), 308-315.
122. Comunale, M. A.; Rodemich-Betesh, L.; Hafner, J.; Wang, M. J.; Norton, P.; Di Bisceglie, A. M.; Block, T.; Mehta, A., Linkage Specific Fucosylation of Alpha-1-Antitrypsin in Liver Cirrhosis and Cancer Patients: Implications for a Biomarker of Hepatocellular Carcinoma. *Plos One* **2010**, 5, (8).
123. Callewaert, N.; Van Vlierberghe, H.; Van Hecke, A.; Laroy, W.; Delanghe, J.; Contreras, R., Noninvasive diagnosis of liver cirrhosis using DNA sequencer-based total serum protein glycomics. *Nat Med* **2004**, 10, (4), 429-34.
124. Vanderschaeghe, D.; Laroy, W.; Sablon, E.; Halfon, P.; Van Hecke, A.; Delanghe, J.; Callewaert, N., GlycoFibroTest is a highly performant liver fibrosis biomarker derived from DNA sequencer-based serum protein glycomics. *Mol Cell Proteomics* **2009**, 8, (5), 986-94.
125. Wiegandt, A.; Meyer, B., Unambiguous characterization of N-glycans of monoclonal antibody cetuximab by integration of LC-MS/MS and (1)H NMR spectroscopy. *Anal Chem* **2014**, 86, (10), 4807-14.
126. Bergwerff, A. A.; Stroop, C. J.; Murray, B.; Holtorf, A. P.; Pluschke, G.; Van Oostrum, J.; Kamerling, J. P.; Vliegthart, J. F., Variation in N-linked carbohydrate chains in different batches of two chimeric monoclonal IgG1 antibodies produced by different murine SP2/o transfectoma cell subclones. *Glycoconj J* **1995**, 12, (3), 318-30.
127. Damm, J. B.; Voshol, H.; Hard, K.; Kamerling, J. P.; Vliegthart, J. F., Analysis of N-acetyl-4-O-acetylneuraminic-acid-containing N-linked carbohydrate chains released by peptide-N4-(N-acetyl-beta-glucosaminy)asparagine amidase F. Application to the structure determination of the carbohydrate chains of equine fibrinogen. *Eur J Biochem* **1989**, 180, (1), 101-10.
128. 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, (1), 96-105.
129. Hard, K.; Damm, J. B.; Spruijt, M. P.; Bergwerff, A. A.; Kamerling, J. P.; Van Dedem, G. W.; Vliegthart, J. F., The carbohydrate chains of the beta subunit of human chorionic gonadotropin produced by the choriocarcinoma cell line BeWo. Novel O-



- linked and novel bisecting-GlcNAc-containing N-linked carbohydrates. *Eur J Biochem* **1992**, 205, (2), 785-98.
130. Grey, A. A.; Narasimhan, S.; Brisson, J. R.; Schachter, H.; Carver, J. P., Structure of the glycopeptides of a human gamma 1-immunoglobulin G (Tem) myeloma protein as determined by 360-megahertz nuclear magnetic resonance spectroscopy. *Can J Biochem* **1982**, 60, (12), 1123-31.
131. 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, (4), 1650-9.
132. Faul, F.; Erdfelder, E.; Lang, A. G.; Buchner, A., G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* **2007**, 39, (2), 175-91.
133. Faul, F.; Erdfelder, E.; Buchner, A.; Lang, A. G., Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods* **2009**, 41, (4), 1149-60.

## 9 Anhang

### 9.1 Übersicht der identifizierten tryptischen Peptide von Fibrinogen und Albumin

**Tabelle 18: Übersicht der bei der LC-MS Analyse des tryptischen Verdaus von Fibrinogen aus Probe N gefundenen Peptide der  $\alpha_2$ -Untereinheit. Es sind die Peptidsequenz sowie das Masse-zu-Ladungs-Verhältnis und der Ladungszustand des jeweils intensivsten Signals angegeben. Weist das Peptid eine ausgelassene Trypsinschnittstelle auf, ist dies unter Anmerkungen angegeben.**

Peptidsequenz	m/z	Ladungszustand	Anmerkungen
<sup>1</sup> ADSGEGFLAEGGGV <sup>16</sup>	768.85	2+	MS/MS
<sup>30</sup> DSDWPFCSDEDEWNYK <sup>44</sup>	982.34	2+	MS/MS
<sup>53</sup> GLIDEVNIQDFTNR <sup>65</sup>	760.87	2+	MS/MS
<sup>71</sup> NSLFEYQK <sup>78</sup>	514.76	2+	MS/MS
<sup>82</sup> DSHSLTTNIMEILR <sup>95</sup>	543.95	3+	
<sup>95</sup> GDFSSANNR <sup>104</sup>	484.22	2+	
<sup>111</sup> VSEDLR <sup>116</sup>	718.37	1+	
<sup>119</sup> IEVLK <sup>123</sup>	601.39	1+	
<sup>126</sup> VIEKVQHILQLLQK <sup>138</sup>	525.99	3+	missed cleavage, MS/MS
<sup>142</sup> AQLVDMIK <sup>148</sup>	402.72	2+	
<sup>149</sup> RLEVDIDIK <sup>157</sup>	550.82	2+	missed cleavage, MS/MS
<sup>168</sup> ALAR <sup>171</sup>	430.28	1+	
<sup>172</sup> EVDLKDYEQQK <sup>183</sup>	755.35	2+	missed cleavage, MS/MS
<sup>184</sup> QLEQVIK <sup>191</sup>	464.78	2+	
<sup>195</sup> DLLPSR <sup>197</sup>	700.40	1+	
<sup>200</sup> QHLPLIK <sup>206</sup>	424.77	2+	
<sup>207</sup> MKPVDPDLVPGNFK <sup>219</sup>	481.27	3+	
<sup>220</sup> SQLQK <sup>224</sup>	603.35	1+	

**Tabelle 18 Fortsetzung: Übersicht der bei der LC-MS Analyse des tryptischen Veraus von Fibrinogen aus Probe N gefundenen Peptide der  $\alpha_2$ -Untereinheit. Es sind die Peptidsequenz sowie das Masse-zu-Ladungs-Verhältnis und der Ladungszustand des jeweils intensivsten Signals angegeben. Weist das Peptid eine ausgelassene Trypsinschnittstelle auf ist dies unter Anmerkungen angegeben.**

Peptidequenz	m/z	Ladungszustand	Anmerkungen
<sup>224</sup> VPPEWK <sup>230</sup>	378.21	2+	
<sup>231</sup> ALDMPQMR <sup>239</sup>	531.76	2+	
<sup>240</sup> MELERPGGNEITR <sup>252</sup>	501.25	3+	
<sup>253</sup> GGSTSYGTGETESPR <sup>268</sup>	786.84	2+	
<sup>269</sup> NPSSAGSWNSGSSGPGSTGNR <sup>289</sup>	982.43	2+	
<sup>290</sup> NPGSSGTGGTATWKPGSSGPGSTGWSNSGSSGTGNQNPSPR <sup>334</sup>	1374.28	3+	
<sup>335</sup> PGSTGTWNPSSER <sup>348</sup>	716.83	2+	MS/MS
<sup>349</sup> GSAGHWTESSVSGSTGQWHSESGFRPDSPGSGNARNPNPDWGTFFEVSGNVSPGTR <sup>406</sup>	996.61	6+	
<sup>430</sup> VTSGSTTTTR <sup>439</sup>	505.76	2+	
<sup>414</sup> LVTSK <sup>418</sup>	547.35	1+	
<sup>449</sup> TVIGPDGHK <sup>457</sup>	923.50	1+	
<sup>462</sup> EVTSEGDSDCPEAMDLGTLGGIGLDGFR <sup>491</sup>	1043.47	3+	MS/MS
<sup>492</sup> HRHIPDEAAFFDTASTGK <sup>508</sup>	629.63	3+	missed cleavage, MS/MS
<sup>509</sup> TPGFFSPMLGEFVSETESR <sup>528</sup>	1133.03	2+	MS/MS
<sup>529</sup> GSESGIFTNTK <sup>539</sup>	570.78	2+	MS/MS
<sup>540</sup> ESSSHHPGIAEPPSR <sup>554</sup>	546.59	3+	MS/MS
<sup>563</sup> QFTSSTSYNRGDSSTFESK <sup>580</sup>	681.31	3+	missed cleavage, MS/MS
<sup>584</sup> MADEAGSEADHEGTHSTK <sup>601</sup>	624.93	3+	
<sup>607</sup> SRPV <sup>610</sup>	458.27	1+	

ALPHA UNTEREINHEIT

10            20            30            40            50            60  
 ADSGEGDFLA EGGGVRGPRV VERHQSAKLI SDWPFCSDEL WNYKCPSGCR MKGLIDEVNG  
 70            80            90            100           110           120  
 DFTNRINKLK NSLFEYQKNN KDSHSLTTNI MEILRGDFSS ANNRDNTYNR VSEDLRSRLE  
 130           140           150           160           170           180  
 VLKRKVIIEKV OHIQLLQKNV RAQLVDMKRL EVDIDIKIRS CRGSCSRALA REVDLKDYEI  
 190           200           210           220           230           240  
 QOKQLEQVIA KDLLPSRDRQ HLPLIKMKPV PDLVPGNFKS QLQKVPPEWK ALTDMPQMRM  
 250           260           270           280           290           300  
 ELERPGGNEI TRGGSTSYGT SSETESPRNE SSAGSWNSGS SGPGSTGNFN PGSSGTGTA  
 310           320           330           340           350           360  
 TWKPGSSGPG STGSWNSGSS GTGSTGNQNP GSPRPGSTGT WNPSSSERGS AGHWTSSESV  
 370           380           390           400           410           420  
 SGSTQWHSE SGSFRPDSPG SGNARPNNDP WGTFFEEVSGN VSPGTRREYH TEKLVTSKGD  
 430           440           450           460           470           480  
 KELRTGKEKV TSGSTTTTR SCSKTVTKIV IGPDGHKVET KFWVTSSEDS DCPEAMDLEI  
 490           500           510           520           530           540  
 LSGIGTLDGF RHRHPDEAAF EDTASTGKTE PGFFSPMLGE FVSETESRGS ESGIFNTKE  
 550           560           570           580           590           600  
 SSSHHPGIAE FPSRKGSSSY SKQFTSSTSY NRGDSTFESK SYKMADEAGS EADHEGTHST  
 610  
 KRGHAKSRPV

**Abbildung 68: Aminosäuresequenz der  $\alpha$ -Untereinheit von humanem Fibrinogen. Die grün markierten Bereiche der Sequenz wurden massenspektrometrisch nachgewiesen. Die Sequenzabdeckung beträgt 83%. Die dunkelgrün markierten Bereiche wurden anhand von MS/MS-Spektren eindeutig identifiziert.**

**Tabelle 19: Übersicht der bei der LC-MS Analyse des tryptischen Veraus von Fibrinogen aus Probe N gefundenen Peptide der B $\beta$ -Untereinheit. Es sind die Peptidsequenz sowie das masse-zu-Ladungs-Verhältnis und der Ladungszustand des jeweils intensivsten Signals angegeben. Weist das Peptid eine ausgelassene Trypsinschnittstelle auf ist dies unter Anmerkungen angegeben.**

Peptidequenz	m/z	Ladungszustand	Anmerkungen
<sup>17</sup> GVNDNEEGFFSAR <sup>14</sup>	776.84	2+	MS/MS
<sup>15</sup> GHRPLDK <sup>21</sup>	411.73	2+	
<sup>24</sup> EEAPSLRPAPPPIGGGYR <sup>42</sup>	651.01	3+	
<sup>59</sup> APDAGGCLHADPDLGVLCTGCQLQEALLQQRPIR <sup>94</sup>	989.74	4+	MS/MS
<sup>95</sup> NSVDELNNNEAVSQTSSSFQMYLLK <sup>122</sup>	1056.50	3+	MS/MS
<sup>1</sup> DLWQK <sup>127</sup>	689.36	1+	
<sup>131</sup> QVKDNENVNEYSSELEK <sup>148</sup>	708.68	3+	missed cleavage, MS/MS
<sup>149</sup> HQLYIDETVNSIPTNLR <sup>166</sup>	709.70	3+	MS/MS
<sup>170</sup> SILENLR <sup>176</sup>	422.75	1+	
<sup>182</sup> LESDVSAQMEYCR <sup>194</sup>	794.34	2+	
<sup>195</sup> TPCTVSCNIPVVSQKECEEIIR <sup>216</sup>	850.08	3+	missed cleavage, MS/MS
<sup>218</sup> GGETSEMYLIQPDSSVKPYR <sup>237</sup>	753.03	3+	MS/MS
<sup>238</sup> VYCDMMTENGWTVIQNR <sup>255</sup>	1078.98	2+	
<sup>256</sup> QDGSVDFGR <sup>264</sup>	980.44	1+	
<sup>265</sup> KWDPYK <sup>270</sup>	418.72	2+	missed cleavage
<sup>271</sup> QGFGNVATNTDGG <sup>283</sup>	654.81	2+	MS/MS
<sup>284</sup> NYCGLPGEYWLGNDK <sup>298</sup>	893.40	2+	MS/MS
<sup>299</sup> ISQLTR <sup>304</sup>	359.22	1+	

**Tabelle 19 Fortsetzung: Übersicht der bei der LC-MS Analyse des tryptischen Veraus von Fibrinogen aus Probe N gefundenen Peptide der B $\beta$ -Untereinheit. Es sind die Peptidsequenz sowie das masse-zu-Ladungs-Verhältnis und der Ladungszustand des jeweils intensivsten Signals angegeben. Weist das Peptid eine ausgelassene Trypsinschnittstelle auf ist dies unter Anmerkungen angegeben.**

Peptidequenz	m/z	Ladungszustand	Anmerkungen
<sup>305</sup> MGPTLLIEMEDWK <sup>318</sup>	846.41	2+	
<sup>324</sup> AHYGGFTVQNEANK <sup>337</sup>	512.58	3+	
<sup>338</sup> YQJSVNIK <sup>344</sup>	426.24	2+	
<sup>347</sup> GTAGNALMDGASQLMGEN(H5N4S1)R <sup>365</sup>	1269.52	3+	
<sup>347</sup> GTAGNALMDGASQLMGEN(H5N4S2)R <sup>365</sup>	1025.16	4+	
<sup>366</sup> TMTHNGMFFSTYDR <sup>380</sup>	910.91	2+	
<sup>381</sup> DNDGWLTS DPR <sup>391</sup>	638.28	2+	MS/MS
<sup>397</sup> EDGGGWYNYR <sup>406</sup>	620.26	2+	MS/MS
<sup>416</sup> YYWGGQYTWDMAK <sup>428</sup>	834.86	2+	MS/MS
<sup>429</sup> HGTDDGVVWMMNWK <sup>441</sup>	515.57	3+	MS/MS
<sup>442</sup> GSWYSMR <sup>448</sup>	443.70	2+	
<sup>454</sup> IRPFFPQQ <sup>461</sup>	516.79	2+	

## BETA UNTEREINHEIT

```

      10      20      30      40      50      60
ZGVNDNEEGF FSARGHRPLD KKREAEAPSLR PAPPPIISGGG VPARPAKAAA TQKKVERKAE
      70      80      90     100     110     120
DAGGCLHADP DLGVLCPPTGC QLQEALLQEE KPIRNSVDEL VNNVEAVSQT SSSSFQMYL
      130     140     150     160     170     180
LKDLWQKRQK QVKDNENVVN EYSSELEKHC LYIDETVNSN IPTNLEVLRS ILENLRSKIQ
      190     200     210     220     230     240
KLES DVSAQM EYCRTPCTVS CNIPVVSQKE CEEIIRKSGE TSEMYLIQPL SSVKPYRVYC
      250     260     270     280     290     300
DMNTENGGWT VIQNRQDGSV DFGRKWDPYK QGFGNVATNT DGKNYCGLPG EYWLGNDRIS
      310     320     330     340     350     360
QLTRMGPTL  LIEMEDWKGD KVKAHYGGFT VQNEANKYQI SVNKYRGTAG NALMDGASQL
      370     380     390     400     410     420
MGENRTMTIH NGMFFSTYDR QNDGWLTSDP EKQCSKEDGG GWWYNRCHAA NPNGRQYWGQ
      430     440     450     460
QYTWDMAKHG TDDGVVVMNW KGSWYSMRKM SMKIRPFFPQ Q

```

**Abbildung 69: Aminosäuresequenz der  $\beta$ -Untereinheit von humanem Fibrinogen. Die grün markierten Bereiche der Sequenz wurden massenspektrometrisch nachgewiesen. Die Sequenzabdeckung beträgt 88%. Die dunkelgrün markierten Bereiche wurden anhand von MS/MS-Spektren eindeutig identifiziert.**

**Tabelle 20: Übersicht der bei der LC-MS Analyse des tryptischen Veraus von Fibrinogen aus Probe N gefundenen Peptide der  $\gamma$ -Untereinheit. Es sind die Peptidsequenz sowie das masse-zu-Ladungs-Verhältnis und der Ladungszustand des jeweils intensivsten Signals angegeben. Weist das Peptid eine ausgelassene Trypsinschnittstelle auf ist dies unter Anmerkungen angegeben.**

Peptidequenz	m/z	Ladungszustand	Anmerkungen
<sup>1</sup> YVATR <sup>5</sup>	609.34	1+	
<sup>6</sup> DNCCILDER <sup>14</sup>	597.75	2+	MS/MS
<sup>15</sup> FGSYPTTCGIADFLSTYQK <sup>35</sup>	1209.04	2+	MS/MS
<sup>39</sup> VDKDLOSLEDILHQVEN(H5N4S1)K <sup>53</sup>	1009.95	4+	missed cleavage, MS/MS
<sup>39</sup> VDKDLOSLEDILHQVEN(H5N4S2)K <sup>53</sup>	1082.72	4+	missed cleavage, MS/MS
<sup>59</sup> QLIK <sup>62</sup>	501.34	1+	
<sup>63</sup> AIQLTYPDESSKPNMIDAATLK <sup>85</sup>	840.76	3+	MS/MS
<sup>89</sup> MLEEIMK <sup>95</sup>	447.23	2+	
<sup>96</sup> YEASILTHDSSIR <sup>108</sup>	746.38	2+	MS/MS
<sup>109</sup> YLQEIYNSNNQK <sup>120</sup>	757.37	2+	MS/MS
<sup>121</sup> IVNLK <sup>125</sup>	586.39	1+	
<sup>128</sup> VAQLEAQCQEPCKDTVQIHDTGK <sup>151</sup>	692.84	4+	Missed cleavage, MS/MS
<sup>152</sup> DCQDIANK <sup>159</sup>	482.21	2+	
<sup>163</sup> QSGLYFIKPLK <sup>173</sup>	431.92	3+	MS/MS
<sup>174</sup> ANQFLVYCEIDGSGNGWTVFQK <sup>196</sup>	887.75	3+	MS/MS
<sup>197</sup> RLDGSVDFK <sup>205</sup>	440.73	2+	missed cleavage, MS/MS
<sup>207</sup> NWIOYK <sup>212</sup>	426.23	2+	
<sup>213</sup> EGFGLSPTGTTEFWLGNK <sup>232</sup>	736.35	3+	MS/MS



**Tabelle 20 Fortsetzung: Übersicht der bei der LC-MS Analyse des tryptischen Veraus von Fibrinogen aus Probe N gefundenen Peptide der  $\gamma$ -Untereinheit. Es sind die Peptidsequenz sowie das masse-zu-Ladungs-Verhältnis und der Ladungszustand des jeweils intensivsten Signals angegeben. Weist das Peptid eine ausgelassene Trypsinschnittstelle auf ist dies unter Anmerkungen angegeben.**

Peptidequenz	m/z	Ladungszustand	Anmerkungen
<sup>233</sup> IHLISTQSAIPYALR <sup>247</sup>	561.66	3+	MS/MS
<sup>248</sup> VELEDWNGR <sup>256</sup>	559.27	2+	MS/MS
<sup>257</sup> TSTADYAMFK <sup>266</sup>	567.76	2+	MS/MS
<sup>267</sup> VGPEADKYR <sup>275</sup>	517.77	2+	missed cleavage, MS/MS
<sup>276</sup> LTYAYFAGGDAGDAFDGDFGDDPSDK <sup>302</sup>	1417.59	2+	
<sup>335</sup> CHAGHLNGVYQGGTYSK <sup>356</sup>	503.73	4+	
<sup>357</sup> ASTPNGYDNGIHWATWK <sup>373</sup>	947.46	2+	MS/MS
<sup>376</sup> WYSMK <sup>380</sup>	714.33	1+	
<sup>386</sup> IIPFNR <sup>391</sup>	380.23	2+	
<sup>392</sup> LTIGEGQQHHLGGAK <sup>406</sup>	515.94	3+	MS/MS
<sup>407</sup> QAGDV <sup>411</sup>	489.23	1+	

## GAMMA UNTEREINHEIT

10	20	30	40	50	60
YVATRDNCCI	LDERFGSYCP	ITCGIADFLS	TYQTKVDDKI	QSLEDILHQV	ENK <sup>T</sup> SEVK <sup>Q</sup> L
70	80	90	100	110	120
IKAIQLTYNF	DESSKFNMD	ATLK <sup>S</sup> SRK <sup>M</sup> L	EEIMK <sup>Y</sup> EAST	ITHDSSIRYL	DELYNSNNQK
130	140	150	160	170	180
IVNLKEK <sup>V</sup> AO	LEAQCOEPCK	DTVQIHDTG	KDCQDIANK <sup>G</sup>	AK <sup>S</sup> SGLYFIK	PLKANQQFLV
190	200	210	220	230	240
YCEIDGSGNG	WTVFQKRLDG	SVDFK <sup>K</sup> NW <sup>I</sup> Q	YKEGFGHLSP	TGTTEFWLGN	EKIHLISTQS
250	260	270	280	290	300
AIPYALRVEL	EDWNGRTSTA	DYAMFKVGP	ADKYRLTYAY	FAGGDAGDAF	DGFDFGDDPS
310	320	330	340	350	360
DKFFTS <sup>H</sup> NGM	QFSTWDNDND	KFEGNCAEQD	GSGWWMNK <sup>C</sup> H	AGHLNGVYYQ	GGTYSK <sup>A</sup> STE
370	380	390	400	410	
NGYDNGI <sup>I</sup> WA	TWK <sup>T</sup> R <sup>W</sup> YSMK	KTTMK <sup>I</sup> IPFN	RLTIGEGQOH	HLGGAK <sup>Q</sup> AGD	V

**Abbildung 70: Aminosäuresequenz der  $\gamma$ -Untereinheit von humanem Fibrinogen. Die grün markierten Bereiche der Sequenz wurden massenspektrometrisch nachgewiesen. Die Sequenzabdeckung beträgt 86%. Die dunkelgrün markierten Bereiche wurden anhand von MS/MS-Spektren eindeutig identifiziert.**

**Tabelle 21: Übersicht der bei der LC-MS Analyse des tryptischen Veraus von Fibrinogen aus Probe N gefundenen Peptide von Albumin. Es sind die Peptidsequenz sowie das masse-zu-Ladungs-Verhältnis und der Ladungszustand des jeweils intensivsten Signals angegeben. Weist das Peptid eine ausgelassene Trypsinschnittstelle auf ist dies unter Anmerkungen angegeben.**

Peptidequenz	m/z	Ladungszustand	Anmerkungen
<sup>11</sup> FKDLGEENFK <sup>20</sup>	409.54	3+	missed cleavage
<sup>21</sup> ALVLIFAFYLQQCFEDHVK <sup>41</sup>	830.77	3+	MS/MS
<sup>42</sup> LVNEVTEFAK <sup>51</sup>	575.31	2+	MS/MS
<sup>52</sup> TCVADESAENCDK <sup>64</sup>	749.79	2+	MS/MS
<sup>65</sup> SLHTLFGDKLCTVATLR <sup>81</sup>	483.77	4+	missed cleavage, MS/MS
<sup>82</sup> ETYGEMADCCAK <sup>93</sup>	717.77	2+	
<sup>94</sup> QEPERNEFLQHK <sup>106</sup>	527.27	3+	missed cleavage, MS/MS
<sup>107</sup> DDNPNLPR <sup>114</sup>	470.73	2+	
<sup>115</sup> LVRPEVDVMCTAFHDNEETFLK <sup>136</sup>	663.32	4+	
<sup>138</sup> YLYEIARI <sup>144</sup>	464.25	2+	
<sup>145</sup> RHPYFYAPELFFFAK <sup>159</sup>	633.67	3+	missed cleavage, MS/MS
<sup>163</sup> AAFTECCQAADK <sup>174</sup>	686.29	2+	MS/MS
<sup>175</sup> AAALLPK <sup>181</sup>	386.72	2+	
<sup>182</sup> LDELRDEGK <sup>190</sup>	358.85	3+	missed cleavage, MS/MS
<sup>200</sup> CASLQK <sup>205</sup>	353.68	2+	
<sup>206</sup> FGER <sup>209</sup>	508.25	1+	
<sup>213</sup> AWAVAR <sup>218</sup>	337.19	2+	
<sup>226</sup> AEFAEVSK <sup>233</sup>	440.73	2+	

**Tabelle 21 Fortsetzung: Übersicht der bei der LC-MS Analyse des tryptischen Veraus von Fibrinogen aus Probe N gefundenen Peptide von Albumin. Es sind die Peptidsequenz sowie das masse-zu-Ladungs-Verhältnis und der Ladungszustand des jeweils intensivsten Signals angegeben. Weist das Peptid eine ausgelassene Trypsinschnittstelle auf ist dies unter Anmerkungen angegeben.**

Peptidsequenz	m/z	Ladungszustand	Anmerkungen
<sup>234</sup> LVTDLTK <sup>240</sup>	395.24	2+	
<sup>241</sup> VHTECCHGDLLECADDR <sup>257</sup>	696.29	3+	
<sup>263</sup> YICENQDSISSK <sup>274</sup>	722.33	2+	MS/MS
<sup>275</sup> LKECEKPLLEK <sup>286</sup>	516.27	3+	missed cleavage
<sup>287</sup> SHCIAEVENDEMPADLPDLAADFVESK <sup>313</sup>	992.12	3+	MS/MS
<sup>318</sup> NYAEAKDVLGMFLYEYAR <sup>336</sup>	767.37	3+	missed cleavage, MS/MS
<sup>337</sup> RHPDYSVLLLR <sup>348</sup>	489.95	3+	missed cleavage, MS/MS
<sup>352</sup> TYETLEK <sup>359</sup>	492.75	2+	
<sup>360</sup> CCAAADPHECYAK <sup>372</sup>	518.21	3+	MS/MS
<sup>373</sup> VDEFKPLVEEPQNLIK <sup>389</sup>	682.37	3+	MS/MS
<sup>390</sup> QNCLEFEQLGEYKFNALLVR <sup>410</sup>	867.11	3+	missed cleavage, MS/MS
<sup>414</sup> KVPQVSTPTLVEVSR <sup>428</sup>	756.43	2+	missed cleavage, MS/MS
<sup>429</sup> NLIGK <sup>432</sup>	431.26	1+	
<sup>433</sup> VGSK <sup>436</sup>	390.24	1+	
<sup>445</sup> RMPCAEDYLSVVLNQLCVLHEK <sup>466</sup>	669.34	4+	missed cleavage, MS/MS
<sup>467</sup> TPVSDR <sup>472</sup>	674.35	2+	
<sup>476</sup> CCTESLVNR <sup>484</sup>	569.75	2+	MS/MS
<sup>485</sup> RPCFSALEVDETYVPK <sup>500</sup>	637.65	3+	MS/MS

**Tabelle 21 Fortsetzung: Übersicht der bei der LC-MS Analyse des tryptischen Veraus von Fibrinogen aus Probe N gefundenen Peptide von Ablumin. Es sind die Peptidsequenz sowie das masse-zu-Ladungs-Verhältnis und der Ladungszustand des jeweils intensivsten Signals angegeben. Weist das Peptid eine ausgelassene Trypsinschnittstelle auf ist dies unter Anmerkungen angegeben.**

Peptidequenz	m/z	Ladungszustand	Anmerkungen
<sup>501</sup> EFNAETTFHADICTLSEKER <sup>521</sup>	637.05	4+	missed cleavage, MS/MS
<sup>525</sup> KQTALVELVK <sup>534</sup>	564.85	2+	missed cleavage, MS/MS
<sup>546</sup> AVMDDFAAFVEK <sup>557</sup>	671.82	2+	
<sup>561</sup> ADDKETCFAEEGK <sup>573</sup>	500.55	3+	missed cleavage
<sup>575</sup> LVAASQAALGL <sup>585</sup>	507.30	2+	

ALBUMIN

10	20	30	40	50	60
DAHKSEVAHR	FKDLGEENFK	ALVLIATAFAQY	LOQCPFEDHV	KLVNEVTEFA	KTCVADESAA
70	80	90	100	110	120
NCDKSLHTLF	GDKLCTVATL	RETYGEMADC	CAKQEPERNE	CFLQHKDDNF	NLPRLVRPEV
130	140	150	160	170	180
DVMCTAFHDN	EETFLKLYLY	EIARRHPYFY	APELLFFAKR	YKAAFTTECCG	AADKAACLLP
190	200	210	220	230	240
KLDELREDEGK	ASSAKQRLKC	ASLQKFGERA	FKAWAVARLS	QRFPKAEFAE	VSKLVTDLTK
250	260	270	280	290	300
VHTECCHGDL	LECADDRADL	AKYICENQDS	LSSKLEKCE	KPLLEKSHGI	AEVENDEMPA
310	320	330	340	350	360
DLPSLAADFV	ESKDVCKNYA	EAKDVFLGME	IYEYARRHPD	YSVWLLLELA	KTYETTLEK
370	380	390	400	410	420
CAAADPHECY	AKVFDEFKPL	VEEPQNLTK	NCELFEQLGE	YKFNALLVR	YTKRVPOVST
430	440	450	460	470	480
PTLVEVSRNL	GKVGSKCCKH	PEAKFMPCAE	SYLSWVLNQL	SVLHEKTPVS	DRVTKECTES
490	500	510	520	530	540
LVNRRPCFSA	LEVDETYVPK	EFNAETFTFH	ADICTLSEKE	HQIKKOTALM	ELVHKPKAT
550	560	570	580		
KEQLKAVMDD	FAAFVEKCKK	ADDKETCFAE	EGKKLVAASQ	AALGL	

**Abbildung 71: Aminosäuresequenz humanem Albumin. Die grün markierten Bereiche der Sequenz wurden massenspektrometrisch nachgewiesen. Die Sequenzabdeckung beträgt 86%. Die dunkelgrün markierten Bereiche wurden anhand von MS/MS-Spektren eindeutig identifiziert.**

## 9.2 Übersicht der beim Verdau mit Pronase gefundenen Glycopeptide

**Tabelle 22: Übersicht der bei der LC-MS-Analyse des mit Pronase verdauten Fibrinogens aus den Proben R, T, HCCo17, HCCo24 und ZIRO31 gefundenen Glycopeptide.**

Peptid	Glycan	m/z	Ladungszustand	Protein
NK	H5N4S1	1087.92	2+	Fibrinogen $\gamma$
NST	H4N4F1	964.364	2+	IgG
NST	H5N4F1	1045.39	2+	IgG
HN	H5N4S1	1092.40	2+	Fibronectin
ENK/NRT	H5N4S1	1152.44	2+	Fibrinogen $\gamma$ /B $\beta$
NK	H5N4S2	1233.47	2+	Fibrinogen $\gamma$
GENR	H5N4S1F1	1267.99	2+	Fibrinogen B $\beta$
GENR	H5N4S1	1194.96	2+	Fibrinogen B $\beta$
NST	H5N4S1F1	1190.94	2+	IgG
N	H5N4S1	1023.87	2+	-
ENK/NRT	H5N4S2	1297.98	2+	Fibrinogen $\gamma$ /B $\beta$
NH	H5N4S2	825.63	3+	Fibronectin
GENR	H5N4S2	1340.50	2+	Fibrinogen B $\beta$
NVT	H5N4S2	846.65	3+	Serotransferrin
NET	H5N4S2	856.65	3+	Fibronectin
ENR	H5N5S2F1	991.38	3+	Fibrinogen B $\beta$
EN	H5N4S2	822.96	3+	Fibrinogen $\gamma$ /B $\beta$
NDT	H5N4S2	851.97	3+	Fibronectin
ENR	H5N4S2F1	923.68	3+	Fibrinogen B $\beta$
GEN	H5N4S2	841.97	3+	Fibrinogen B $\beta$
GEN/SVN	H5N5S2F	958.36	3+	Fibrinogen B $\beta$ /Fibronectin
NVT	H6N5S2	968.36	3+	Serotransferrin
N	H5N4S2	1169.42	2+	-

### 9.3 Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der $\gamma$ -Untereinheit aus den Proben des Trainingsdatensatzes.

Probe	Zustand	$\gamma$ -Untereinheit									
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Wasser- verlust		
A	Gesund	0.4%	2.1%	67.6%	2.2%	27.5%	0.2%	17.1%	18.7%		
B	Gesund	0.6%	4.6%	63.6%	1.3%	29.9%	0.1%	11.1%	14.9%		
C	Gesund	0.7%	6.4%	65.8%	0.2%	26.7%	0.0%	8.9%	13.7%		
D	Gesund	1.0%	6.7%	69.3%	0.5%	22.5%	0.0%	9.0%	14.1%		
E	Gesund	0.9%	4.2%	67.6%	2.3%	24.9%	0.2%	18.4%	14.1%		
F	Gesund	0.0%	1.7%	67.6%	2.6%	27.8%	0.3%	12.2%	17.8%		
G	Gesund	0.7%	3.9%	64.2%	0.3%	30.6%	0.4%	8.9%	10.3%		
H	Gesund	1.4%	3.8%	64.4%	1.2%	29.2%	0.1%	7.3%	10.4%		
I	Gesund	1.0%	5.3%	70.2%	0.6%	22.9%	0.0%	8.8%	12.2%		
J	Gesund	0.3%	0.4%	66.0%	0.2%	32.9%	0.3%	9.3%	16.5%		
K	Gesund	1.8%	2.3%	64.0%	0.1%	31.7%	0.0%	6.8%	10.7%		
L	Gesund	1.7%	3.6%	66.3%	0.6%	27.8%	0.1%	10.5%	9.6%		
M	Gesund	1.3%	2.0%	64.1%	1.2%	31.2%	0.2%	7.8%	12.2%		
N	Gesund	1.2%	1.9%	60.4%	1.1%	35.0%	0.4%	7.8%	12.1%		
O	Gesund	2.2%	1.5%	61.3%	1.7%	32.6%	0.7%	7.3%	11.5%		



**Tabelle 23 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\gamma$ -Untereinheit aus den Proben des Trainingsdatensatzes.**

Probe	Zustand	$\gamma$ -Untereinheit									
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Wasser- verlust		
P	Gesund	2.5%	2.9%	64.8%	1.7%	27.6%	0.4%	8.1%	10.6%		
Q	Gesund	0.9%	4.5%	67.7%	0.8%	26.2%	0.0%	12.5%	12.8%		
R	Gesund	0.6%	1.9%	63.1%	1.3%	33.0%	0.1%	14.2%	12.5%		
S	Gesund	0.9%	7.4%	66.0%	1.0%	24.7%	0.0%	13.2%	12.3%		
T	Gesund	0.7%	4.6%	65.3%	1.1%	28.3%	0.0%	12.9%	12.9%		
HCC001	HCC	0.7%	3.6%	58.8%	2.2%	34.0%	0.8%	11.0%	14.5%		
HCC002	HCC	0.5%	3.4%	56.7%	5.1%	30.6%	3.7%	12.2%	13.4%		
HCC003	HCC	0.4%	2.5%	54.4%	3.2%	38.1%	1.4%	11.3%	14.2%		
HCC006	HCC	1.4%	1.5%	57.5%	1.8%	37.6%	0.3%	8.7%	16.3%		
HCC007	HCC	0.4%	4.7%	49.0%	4.4%	40.2%	1.3%	10.9%	14.0%		
HCC008	HCC	0.6%	3.8%	54.2%	2.7%	38.3%	0.4%	11.1%	14.5%		
HCC009	HCC	3.8%	4.5%	61.1%	1.0%	29.5%	0.1%	11.2%	14.3%		
HCC011	HCC	1.9%	4.5%	56.0%	3.2%	33.1%	1.3%	7.4%	18.1%		
HCC012	HCC	0.6%	1.3%	53.8%	2.5%	41.2%	0.6%	8.8%	13.3%		
HCC014	HCC	0.6%	2.0%	60.1%	0.6%	36.3%	0.3%	9.8%	13.2%		

**Tabelle 23 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\gamma$ -Untereinheit aus den Proben des Trainingsdatensatzes.**

Probe	Zustand	$\gamma$ -Untereinheit									
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Wasser- verlust		
HCC016	HCC	1.3%	6.5%	65.8%	0.0%	26.3%	0.0%	5.8%	12.2%		
HCC017	HCC	0.7%	5.9%	48.3%	8.5%	35.7%	0.8%	5.6%	10.9%		
HCC018	HCC	0.6%	3.6%	55.3%	2.8%	36.9%	0.8%	11.8%	11.8%		
HCC019	HCC	0.7%	5.6%	60.9%	3.2%	29.2%	0.4%	11.6%	12.5%		
HCC020	HCC	0.6%	3.6%	63.3%	1.1%	31.4%	0.0%	39.8%	8.6%		
HCC021	HCC	0.5%	3.7%	58.4%	2.4%	34.7%	0.3%	9.6%	12.7%		
HCC022	HCC	0.9%	3.7%	58.9%	5.5%	30.2%	0.8%	7.0%	13.1%		
HCC023	HCC	0.1%	0.9%	57.0%	1.8%	39.0%	1.1%	9.2%	15.8%		
HCC024	HCC	0.1%	3.2%	46.7%	9.6%	40.3%	0.1%	6.5%	14.9%		
ZIR031	ZIR	0.7%	0.5%	51.9%	2.7%	41.6%	2.6%	8.6%	13.6%		
ZIR033	ZIR	0.4%	1.6%	53.7%	3.2%	39.8%	1.4%	7.4%	13.5%		
ZIR034	ZIR	0.1%	0.0%	35.6%	8.0%	49.6%	6.6%	17.6%	11.5%		
ZIR035	ZIR	0.9%	3.5%	58.8%	2.3%	33.6%	0.9%	15.0%	11.7%		
ZIR036	ZIR	1.0%	6.9%	68.6%	2.6%	20.7%	0.2%	17.8%	10.8%		
ZIR037	ZIR	0.8%	11.7%	66.8%	3.2%	17.5%	0.0%	17.2%	12.6%		

**Tabelle 23 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\gamma$ -Untereinheit aus den Proben des Trainingsdatensatzes.**

Probe	Zustand	$\gamma$ -Untereinheit									
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Wasser- verlust		
ZIR038	ZIR	0.8%	0.9%	60.2%	5.5%	30.6%	2.0%	17.3%	11.4%		
ZIR039	ZIR	0.5%	4.1%	61.7%	2.3%	31.3%	0.2%	12.4%	13.0%		
ZIR041	ZIR	0.4%	0.0%	33.4%	15.6%	44.1%	6.5%	18.6%	12.0%		
ZIR042	ZIR	0.6%	7.2%	68.1%	2.8%	20.7%	0.6%	17.8%	10.9%		
ZIR043	ZIR	0.6%	8.7%	68.1%	2.4%	20.1%	0.1%	18.0%	11.1%		
ZIR044	ZIR	0.5%	3.2%	65.4%	5.0%	24.8%	1.1%	17.7%	11.7%		
ZIR045	ZIR	0.8%	6.6%	70.4%	2.0%	20.3%	0.0%	18.2%	11.5%		
ZIR046	ZIR	0.9%	3.0%	57.3%	4.5%	33.0%	1.3%	12.3%	11.9%		
ZIR048	ZIR	0.4%	0.1%	47.2%	11.7%	34.4%	6.1%	13.5%	12.9%		
ZIR049	ZIR	1.1%	2.9%	58.2%	4.9%	30.6%	2.3%	13.2%	12.3%		
ZIR050	ZIR	0.9%	0.0%	49.9%	8.7%	36.6%	3.8%	11.0%	10.4%		

**Tabelle 24: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\alpha$ -Untereinheit aus den Proben des Trainingsdatensatzes.**

Probe	Zustand	$\alpha$ -Untereinheit													
		unglycosyliert				O-glycosyliert				Gesamt O-glycosyliert	Einfach oxidiert	Zweifach oxidiert	Wasserverlust	Carbamido	Trunkiert
		0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>	0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>						
A	Gesund	35.2%	23.1%	19.5%	4.8%	9.2%	4.9%	2.9%	0.4%	17.5%	20.0%	10.6%	20.1%	8.7%	3.4%
B	Gesund	32.2%	25.9%	23.6%	3.5%	5.7%	5.2%	4.0%	0.0%	14.9%	18.8%	6.8%	19.4%	7.9%	3.4%
C	Gesund	29.9%	27.3%	25.1%	4.2%	4.9%	4.2%	3.8%	0.6%	13.5%	24.2%	10.3%	16.2%	8.5%	2.6%
D	Gesund	33.1%	25.5%	21.7%	3.7%	6.9%	4.8%	4.1%	0.3%	16.2%	17.5%	7.5%	17.9%	7.3%	4.4%
E	Gesund	37.3%	21.2%	20.4%	4.3%	8.1%	3.7%	4.2%	0.9%	16.9%	23.5%	9.6%	17.0%	7.7%	4.8%
F	Gesund	26.6%	24.6%	25.3%	5.8%	5.6%	4.9%	5.7%	1.5%	17.7%	22.0%	10.0%	18.8%	9.9%	4.3%
G	Gesund	33.5%	24.6%	19.0%	3.2%	8.3%	5.9%	4.9%	0.7%	19.8%	21.0%	8.0%	17.5%	7.5%	5.9%
H	Gesund	32.8%	21.4%	24.3%	4.1%	7.1%	4.8%	5.0%	0.6%	17.5%	16.9%	4.4%	20.5%	9.1%	6.1%
I	Gesund	34.6%	23.2%	19.1%	4.4%	8.1%	4.9%	4.5%	1.2%	18.7%	19.9%	9.9%	18.8%	10.9%	3.6%
J	Gesund	42.1%	20.4%	17.1%	2.5%	10.9%	4.3%	2.8%	0.0%	17.9%	9.4%	4.6%	21.6%	11.4%	7.5%
K	Gesund	29.5%	22.1%	22.7%	5.6%	7.5%	5.4%	6.0%	1.2%	20.1%	17.8%	7.4%	17.5%	7.9%	4.4%
L	Gesund	23.3%	23.2%	28.7%	6.9%	4.8%	5.1%	7.0%	1.0%	17.9%	20.7%	8.4%	11.7%	7.6%	2.1%
M	Gesund	28.8%	23.5%	24.2%	4.0%	7.8%	5.7%	5.6%	0.5%	19.6%	15.2%	4.9%	13.8%	6.3%	3.8%
N	Gesund	29.9%	24.0%	22.1%	3.7%	8.3%	6.2%	5.6%	0.3%	20.4%	21.4%	6.7%	12.4%	6.1%	1.6%
O	Gesund	31.5%	24.7%	21.7%	3.0%	7.9%	5.6%	5.0%	0.5%	19.1%	12.8%	4.6%	14.0%	5.8%	4.4%

**Tabelle 24 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\alpha\alpha$ -Untereinheit der Proben des Trainingsdatensatzes.**

Probe	Zustand	$\alpha\alpha$ -Untereinheit													
		unglycosyliert				O-glycosyliert				Gesamt O-glycosyliert	Einfach oxidiert	Zweifach oxidiert	Wasserverlust	Carbamido	Trunkiert
		0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>	0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>						
P	Gesund	32.7%	22.1%	19.9%	4.1%	9.6%	6.3%	5.1%	0.3%	21.2%	19.5%	6.5%	13.0%	5.4%	1.9%
Q	Gesund	30.9%	24.4%	23.9%	3.7%	7.0%	5.6%	4.4%	0.0%	17.1%	19.0%	5.4%	14.4%	7.0%	1.8%
R	Gesund	36.9%	22.8%	18.5%	2.6%	9.5%	5.5%	4.1%	0.1%	19.2%	17.9%	7.7%	13.7%	7.3%	3.5%
S	Gesund	39.1%	24.8%	20.4%	2.2%	7.2%	4.4%	1.9%	0.0%	13.5%	23.0%	7.3%	14.7%	7.1%	3.4%
T	Gesund	37.1%	24.1%	19.9%	2.7%	7.9%	5.6%	2.7%	0.0%	16.2%	21.0%	7.4%	14.1%	7.2%	3.6%
HCC001	HCC	39.9%	22.4%	14.8%	2.1%	10.3%	5.4%	4.0%	1.1%	20.8%	19.2%	7.7%	17.9%	8.4%	11.4%
HCC002	HCC	21.6%	26.0%	25.1%	3.9%	5.8%	7.7%	7.9%	2.0%	23.4%	23.6%	12.1%	17.2%	9.3%	9.0%
HCC003	HCC	37.6%	22.5%	20.5%	3.1%	6.1%	4.4%	5.8%	0.0%	16.4%	24.2%	13.1%	19.1%	10.7%	11.2%
HCC006	HCC	42.6%	23.0%	15.9%	1.3%	9.2%	4.7%	3.1%	0.3%	17.2%	16.0%	7.0%	18.5%	6.8%	9.4%
HCC007	HCC	33.9%	27.4%	22.0%	2.3%	4.7%	3.6%	4.1%	1.9%	14.3%	23.8%	11.2%	22.7%	9.0%	32.2%
HCC008	HCC	42.9%	20.8%	12.3%	1.6%	13.4%	5.3%	2.7%	0.9%	22.3%	25.6%	14.0%	16.3%	10.1%	8.2%
HCC009	HCC	41.5%	23.5%	19.5%	0.7%	7.3%	2.9%	4.3%	0.3%	14.7%	27.8%	15.8%	15.7%	9.7%	12.3%
HCC011	HCC	48.3%	23.0%	14.7%	1.5%	9.2%	2.0%	1.2%	0.0%	12.4%	26.1%	16.3%	14.2%	10.3%	14.4%
HCC012	HCC	38.9%	25.8%	20.2%	1.6%	7.4%	4.1%	1.9%	0.0%	13.4%	12.1%	0.4%	14.8%	5.1%	3.1%
HCC014	HCC	31.1%	25.8%	23.2%	3.7%	6.8%	5.1%	4.2%	0.1%	16.1%	19.8%	8.1%	12.5%	9.4%	1.2%

**Tabelle 24 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\alpha\alpha$ -Untereinheit der Proben des Trainingsdatensatzes.**

Probe	Zustand	$\alpha\alpha$ -Untereinheit													
		unglycosyliert				O-glycosyliert				Gesamt O-glycosyliert	Einfach oxidiert	Zweifach oxidiert	Wasserverlust	Carbamido	Trunkiert
		0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>	0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>						
HCC016	HCC	43.1%	20.4%	14.8%	2.1%	11.9%	4.6%	2.7%	0.4%	19.6%	25.7%	9.4%	12.9%	6.1%	5.2%
HCC017	HCC	33.6%	29.3%	21.9%	2.9%	5.3%	3.6%	2.4%	0.9%	12.2%	30.0%	17.0%	11.3%	9.7%	6.1%
HCC018	HCC	46.6%	16.1%	8.4%	3.3%	16.9%	5.2%	1.9%	1.5%	25.5%	27.6%	15.2%	12.7%	13.3%	6.2%
HCC019	HCC	37.2%	26.7%	20.5%	1.8%	7.1%	4.2%	2.3%	0.1%	13.8%	27.1%	12.1%	13.5%	8.1%	6.7%
HCC020	HCC	42.3%	21.3%	15.4%	0.9%	11.6%	5.2%	3.2%	0.2%	20.1%	20.7%	8.7%	13.7%	6.8%	10.1%
HCC021	HCC	38.8%	24.2%	18.9%	1.1%	8.1%	5.4%	3.3%	0.1%	17.0%	33.5%	15.4%	6.0%	5.2%	4.2%
HCC022	HCC	22.4%	27.6%	29.9%	4.9%	4.7%	0.8%	6.1%	3.7%	15.2%	25.4%	11.6%	18.2%	13.4%	24.0%
HCC023	HCC	30.3%	22.8%	22.6%	2.5%	10.4%	5.7%	5.7%	0.1%	21.8%	16.4%	5.3%	18.5%	5.4%	9.9%
HCC024	HCC	47.0%	22.6%	16.3%	0.8%	9.6%	3.1%	0.6%	0.0%	13.3%	22.6%	8.8%	16.3%	8.7%	8.1%
ZIRO31	ZIR	51.8%	23.3%	6.6%	1.9%	8.4%	2.5%	1.8%	3.7%	16.4%	24.3%	10.0%	12.6%	10.5%	25.0%
ZIRO33	ZIR	46.8%	25.6%	14.2%	0.6%	7.3%	3.8%	1.4%	0.2%	12.8%	31.9%	14.7%	8.3%	6.2%	7.8%
ZIRO34	ZIR	39.7%	28.5%	19.2%	0.6%	8.0%	2.9%	0.7%	0.5%	12.0%	25.9%	16.4%	15.3%	11.7%	14.5%
ZIRO35	ZIR	42.1%	23.6%	16.0%	1.1%	10.0%	4.7%	2.4%	0.1%	17.2%	25.8%	12.5%	14.0%	8.7%	6.4%
ZIRO36	ZIR	20.6%	27.1%	30.6%	8.4%	2.3%	5.9%	5.2%	0.0%	13.3%	28.6%	16.4%	13.5%	10.3%	0.8%
ZIRO37	ZIR	22.7%	30.1%	27.1%	6.9%	3.7%	5.1%	4.5%	0.0%	13.3%	26.1%	15.4%	15.2%	12.7%	1.5%

**Tabelle 24 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\alpha\alpha$ -Untereinheit der Proben des Trainingsdatensatzes.**

Probe	Zustand	$\alpha\alpha$ -Untereinheit													
		unglycosyliert				O-glycosyliert				Gesamt O-glycosyliert	Einfach oxidiert	Zweifach oxidiert	Wasserverlust	Carbamido	Trunkiert
		0 × PO <sub>4</sub>	1 × PO <sub>4</sub>	2 × PO <sub>4</sub>	3 × PO <sub>4</sub>	0 × PO <sub>4</sub>	1 × PO <sub>4</sub>	2 × PO <sub>4</sub>	3 × PO <sub>4</sub>						
ZIR038	ZIR	38.0%	30.9%	16.3%	1.1%	7.0%	4.2%	1.6%	0.8%	13.7%	27.2%	17.7%	14.6%	13.1%	4.0%
ZIR039	ZIR	32.4%	28.8%	23.7%	3.0%	5.1%	4.6%	2.5%	0.1%	12.3%	27.7%	14.9%	13.7%	9.4%	5.8%
ZIR041	ZIR	27.6%	33.7%	27.9%	5.0%	2.4%	2.0%	1.2%	0.3%	5.9%	23.6%	13.5%	21.4%	11.3%	12.6%
ZIR042	ZIR	29.9%	29.9%	24.6%	5.7%	4.6%	3.2%	1.8%	0.2%	9.8%	26.8%	17.6%	14.1%	13.0%	1.2%
ZIR043	ZIR	11.4%	27.3%	41.0%	9.3%	0.5%	3.5%	7.0%	0.0%	11.0%	20.7%	10.4%	13.6%	10.3%	0.2%
ZIR044	ZIR	24.5%	29.1%	29.5%	5.6%	3.4%	3.9%	3.9%	0.0%	11.2%	25.0%	13.0%	14.3%	10.6%	2.2%
ZIR045	ZIR	27.2%	29.3%	25.2%	6.7%	4.7%	4.5%	2.5%	0.0%	11.7%	26.5%	14.0%	12.9%	9.7%	0.7%
ZIR046	ZIR	50.3%	23.2%	12.6%	1.1%	8.3%	3.7%	0.9%	0.0%	12.8%	24.4%	12.2%	13.2%	9.3%	9.5%
ZIR048	ZIR	45.4%	24.6%	12.9%	0.9%	11.0%	3.9%	1.2%	0.0%	16.1%	22.3%	11.0%	17.6%	10.9%	8.5%
ZIR049	ZIR	35.8%	28.1%	21.0%	2.6%	6.1%	4.1%	2.1%	0.2%	12.4%	26.4%	13.4%	13.3%	9.8%	2.1%
ZIR050	ZIR	33.0%	28.5%	19.2%	1.6%	8.4%	7.4%	1.9%	0.0%	17.7%	20.5%	4.7%	15.3%	8.1%	3.7%

**Tabelle 25: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der B $\beta$ -Untereinheit aus den Proben des Trainingsdatensatzes.**

Probe	Zustand	B $\beta$ -Untereinheit										
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Trunkiert	Wasser- verlust	Carbamido	
A	Gesund	0.0%	3.2%	54.9%	6.1%	35.8%	0.0%	20.8%	21.0%	16.3%	18.5%	
B	Gesund	0.2%	4.7%	52.6%	2.7%	39.9%	0.0%	23.2%	15.6%	15.1%	6.0%	
C	Gesund	0.6%	5.0%	55.6%	2.9%	35.9%	0.0%	31.1%	20.8%	11.9%	6.5%	
D	Gesund	0.1%	5.8%	58.2%	3.4%	32.6%	0.0%	23.0%	12.5%	15.6%	9.3%	
E	Gesund	0.1%	4.8%	51.2%	5.6%	38.3%	0.1%	30.8%	23.7%	12.5%	7.8%	
F	Gesund	0.2%	4.9%	50.8%	4.4%	39.7%	0.0%	30.8%	20.6%	13.4%	7.9%	
G	Gesund	0.1%	3.6%	54.9%	1.2%	40.2%	0.0%	24.0%	18.8%	9.5%	6.4%	
H	Gesund	0.0%	3.0%	54.6%	2.1%	40.3%	0.0%	22.1%	6.7%	11.9%	1.5%	
I	Gesund	0.1%	4.3%	58.1%	1.6%	35.8%	0.0%	24.0%	17.3%	11.5%	4.7%	
J	Gesund	0.1%	2.5%	50.6%	0.8%	45.9%	0.1%	18.1%	9.9%	16.8%	3.4%	
K	Gesund	0.4%	4.2%	54.4%	1.6%	39.5%	0.0%	18.2%	15.6%	11.2%	0.9%	
L	Gesund	0.2%	3.8%	55.5%	1.5%	38.9%	0.0%	20.8%	20.1%	10.8%	4.9%	
M	Gesund	0.1%	3.6%	50.7%	4.4%	41.3%	0.0%	19.0%	19.0%	12.8%	6.1%	
N	Gesund	0.6%	3.6%	48.3%	2.4%	45.0%	0.0%	25.5%	17.3%	12.0%	4.7%	
O	Gesund	1.5%	3.7%	51.7%	2.0%	41.1%	0.0%	14.9%	13.2%	13.7%	3.6%	



**Tabelle 25 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der B $\beta$ -Untereinheit aus den Proben des Trainingsdatensatzes.**

Probe	Zustand	B $\beta$ -Untereinheit									
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Trunkiert	Wasser- verlust	Carbamido
P	Gesund	1.5%	3.8%	50.8%	2.4%	41.5%	0.0%	26.1%	17.4%	11.3%	4.3%
Q	Gesund	0.1%	4.4%	57.5%	2.8%	35.2%	0.0%	24.8%	19.8%	12.0%	8.8%
R	Gesund	0.6%	3.9%	51.5%	4.0%	39.9%	0.0%	19.9%	26.0%	11.6%	8.4%
S	Gesund	0.1%	4.6%	57.4%	2.6%	35.3%	0.0%	28.0%	21.1%	10.8%	7.7%
T	Gesund	0.0%	4.6%	54.7%	3.7%	37.1%	0.0%	22.6%	18.2%	12.6%	11.7%
HCC001	HCC	0.1%	4.4%	49.4%	4.7%	41.4%	0.1%	24.8%	17.5%	14.3%	5.2%
HCC002	HCC	0.0%	3.9%	51.5%	4.6%	40.0%	0.0%	32.9%	20.8%	11.6%	5.9%
HCC003	HCC	0.0%	4.9%	44.6%	10.1%	40.4%	0.0%	31.0%	28.3%	10.9%	12.2%
HCC006	HCC	0.6%	4.1%	47.5%	5.6%	42.2%	0.0%	21.8%	17.2%	15.6%	3.9%
HCC007	HCC	0.0%	4.7%	44.7%	8.0%	42.5%	0.0%	34.2%	19.1%	11.8%	5.3%
HCC008	HCC	0.0%	4.7%	45.9%	7.4%	42.0%	0.0%	33.3%	26.3%	10.8%	8.9%
HCC009	HCC	0.0%	5.9%	53.5%	4.3%	36.3%	0.0%	39.9%	27.9%	8.8%	7.4%
HCC011	HCC	0.0%	4.6%	49.7%	7.0%	38.7%	0.0%	37.6%	28.1%	9.2%	8.7%
HCC012	HCC	0.1%	2.9%	45.4%	5.5%	45.8%	0.2%	18.4%	17.4%	13.0%	6.2%
HCC014	HCC	0.0%	3.0%	49.8%	2.9%	44.3%	0.0%	23.9%	21.3%	11.0%	12.0%

**Tabelle 25 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der B $\beta$ -Untereinheit aus den Proben des Trainingsdatensatzes.**

Probe	Zustand	B $\beta$ -Untereinheit										
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Trunkiert	Wasser- verlust	Carbamido	
HCC016	HCC	0.2%	4.6%	55.7%	1.8%	37.7%	0.0%	32.1%	14.3%	11.1%	4.2%	
HCC017	HCC	0.0%	1.4%	39.8%	17.8%	41.0%	0.0%	32.7%	29.1%	9.0%	5.8%	
HCC018	HCC	0.1%	2.7%	47.5%	5.6%	44.1%	0.0%	37.1%	26.3%	8.4%	7.7%	
HCC019	HCC	0.0%	4.5%	52.9%	8.0%	34.6%	0.0%	34.3%	27.6%	9.0%	9.1%	
HCC020	HCC	0.2%	4.4%	52.3%	4.4%	38.7%	0.0%	25.3%	23.7%	11.1%	7.4%	
HCC021	HCC	0.0%	3.0%	48.1%	6.2%	42.5%	0.0%	35.3%	24.9%	9.1%	5.1%	
HCC022	HCC	0.0%	2.6%	49.4%	9.9%	37.6%	0.5%	36.8%	16.3%	10.5%	1.9%	
HCC023	HCC	0.1%	2.1%	45.7%	4.5%	47.2%	0.4%	26.8%	17.4%	13.8%	3.9%	
HCC024	HCC	0.0%	0.3%	33.6%	16.2%	49.8%	0.0%	24.3%	17.8%	14.0%	6.9%	
ZIRO31	ZIR	0.1%	1.3%	45.4%	9.0%	44.2%	0.0%	37.9%	43.9%	4.7%	3.6%	
ZIRO33	ZIR	0.0%	1.3%	43.3%	7.2%	47.6%	0.5%	30.5%	17.3%	12.7%	2.3%	
ZIRO34	ZIR	0.0%	0.0%	26.2%	13.4%	54.8%	5.5%	34.8%	27.8%	7.7%	19.1%	
ZIRO35	ZIR	0.0%	3.5%	48.3%	4.7%	43.6%	0.0%	31.2%	20.2%	16.3%	12.3%	
ZIRO36	ZIR	0.0%	4.5%	59.5%	5.8%	30.2%	0.0%	41.3%	27.4%	7.6%	15.6%	
ZIRO37	ZIR	0.0%	7.3%	64.9%	6.5%	21.3%	0.0%	36.2%	27.1%	9.2%	15.1%	

**Tabelle 25 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der B $\beta$ -Untereinheit aus den Proben des Trainingsdatensatzes.**

Probe	Zustand	B $\beta$ -Untereinheit										
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Trunkiert	Wasser- verlust	Carbamido	
ZIR038	ZIR	0.0%	2.8%	46.7%	10.9%	38.8%	0.7%	41.3%	28.6%	7.3%	13.8%	
ZIR039	ZIR	0.0%	3.5%	54.1%	6.4%	36.1%	0.0%	38.7%	25.3%	8.7%	11.2%	
ZIR041	ZIR	0.0%	0.0%	22.8%	25.8%	46.9%	4.5%	25.6%	28.8%	10.5%	21.5%	
ZIR042	ZIR	0.0%	5.4%	62.5%	8.3%	23.8%	0.0%	47.2%	39.5%	5.1%	20.3%	
ZIR043	ZIR	0.0%	3.7%	62.5%	6.2%	27.7%	0.0%	22.8%	24.9%	10.6%	18.5%	
ZIR044	ZIR	0.0%	2.9%	55.9%	10.3%	30.8%	0.0%	29.2%	26.8%	9.6%	21.1%	
ZIR045	ZIR	0.0%	2.8%	64.0%	7.0%	26.2%	0.0%	32.0%	24.2%	9.6%	22.2%	
ZIR046	ZIR	0.0%	2.3%	49.7%	9.2%	38.4%	0.4%	32.5%	22.0%	10.1%	9.1%	
ZIR048	ZIR	0.0%	0.2%	42.7%	16.0%	39.7%	1.4%	30.5%	35.3%	7.7%	12.3%	
ZIR049	ZIR	0.0%	3.5%	50.5%	7.7%	37.7%	0.7%	32.1%	25.4%	10.0%	11.2%	
ZIR050	ZIR	0.0%	2.6%	40.6%	12.7%	42.0%	2.2%	30.2%	20.6%	9.9%	7.7%	

**Tabelle 26: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\gamma$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	$\gamma$ -Untereinheit									
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Wasser- verlust		
U	Gesund	0.8%	5.3%	65.6%	1.3%	27.0%	0.0%	12.6%	11.4%		
V	Gesund	0.4%	3.7%	65.6%	2.2%	27.8%	0.3%	16.9%	12.0%		
W	Gesund	0.4%	4.6%	62.5%	1.6%	30.5%	0.4%	11.0%	11.8%		
X	Gesund	1.2%	7.1%	67.4%	1.1%	23.1%	0.0%	17.1%	9.5%		
Y	Gesund	1.4%	3.7%	64.0%	1.5%	29.2%	0.2%	9.6%	10.4%		
Z	Gesund	0.6%	0.6%	59.8%	1.3%	37.2%	0.6%	11.8%	12.3%		
AA	Gesund	2.0%	4.6%	68.9%	0.9%	23.5%	0.0%	9.8%	10.1%		
AB	Gesund	0.9%	4.5%	64.7%	1.1%	28.8%	0.1%	10.3%	11.3%		
AC	Gesund	0.6%	3.2%	63.8%	1.8%	30.4%	0.4%	11.6%	12.2%		
AD	Gesund	1.0%	6.0%	61.8%	1.1%	30.0%	0.1%	9.5%	10.6%		
V001	HCC	1.3%	3.7%	63.8%	2.1%	28.7%	0.3%	13.5%	13.5%		
V002	HCC	1.2%	3.4%	61.1%	5.3%	27.3%	1.7%	13.2%	13.7%		
V003	HCC	1.7%	3.0%	58.5%	3.4%	32.5%	0.9%	12.5%	13.8%		
V004	HCC	1.2%	2.1%	61.9%	2.0%	32.3%	0.6%	12.6%	13.7%		
V005	HCC	1.3%	0.3%	52.8%	2.6%	40.9%	2.0%	13.2%	13.8%		
V006	HCC	0.1%	5.0%	40.7%	2.7%	50.2%	1.3%	14.1%	11.1%		

**Tabelle 26 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\gamma$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	$\gamma$ -Untereinheit									
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Wasser- verlust		
V007	ZIR	0.8%	0.1%	36.1%	4.2%	57.2%	1.6%	13.9%	14.3%		
V008	HCC	0.5%	1.2%	49.2%	4.4%	43.2%	1.6%	9.9%	10.7%		
V009	HCC	0.7%	4.4%	60.0%	2.4%	32.0%	0.6%	13.9%	13.9%		
V010	-	0.9%	6.5%	57.2%	1.4%	33.5%	0.4%	12.7%	13.7%		
V011	HCC	0.5%	8.5%	66.9%	2.5%	21.4%	0.2%	11.5%	12.0%		
V012	HCC	0.9%	6.8%	68.5%	2.5%	21.1%	0.1%	11.7%	12.6%		
V013	HCC	1.8%	6.2%	74.0%	2.1%	15.8%	0.0%	12.2%	12.3%		
V014	HCC	1.7%	0.7%	59.2%	2.8%	34.5%	1.2%	12.3%	12.4%		
V015	-	2.1%	5.2%	70.9%	1.9%	19.8%	0.1%	12.4%	12.6%		
V016	-	0.6%	2.3%	56.3%	2.9%	36.3%	1.7%	14.7%	11.3%		
V017	HCC	0.7%	5.6%	62.9%	2.0%	28.4%	0.5%	11.7%	10.8%		
V019	HCC	0.3%	0.4%	47.0%	2.6%	47.5%	2.1%	36.9%	9.1%		
V020	HCC	0.3%	2.9%	50.1%	1.5%	43.9%	1.2%	11.8%	12.6%		
V021	HCC	1.0%	3.9%	60.0%	3.9%	29.4%	1.8%	11.8%	12.8%		
V022	HCC	0.4%	5.0%	65.1%	2.6%	25.8%	1.2%	11.2%	11.6%		
V023	HCC	0.2%	2.3%	54.1%	2.2%	39.8%	1.5%	11.5%	10.8%		
V025	HCC	0.5%	6.1%	64.3%	2.8%	25.8%	0.3%	11.2%	11.5%		

**Tabelle 26 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\gamma$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	$\gamma$ -Untereinheit									
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Wasser- verlust		
V026	HCC	0.6%	3.6%	61.6%	6.3%	26.0%	2.0%	11.1%	11.5%		
V027	ZIR	0.4%	2.4%	47.9%	4.7%	43.2%	1.4%	11.2%	10.6%		
V028	HCC	0.6%	2.4%	56.1%	3.3%	36.1%	1.4%	11.9%	11.3%		
V029	ZIR	1.1%	1.1%	54.3%	2.0%	40.4%	1.1%	14.0%	14.1%		
V030	ZIR	0.5%	2.1%	52.9%	4.6%	36.6%	3.3%	14.3%	14.3%		
V031	ZIR	1.1%	0.6%	61.6%	4.6%	30.5%	1.6%	12.2%	14.1%		
V032	ZIR	0.6%	0.0%	41.9%	5.8%	47.5%	4.2%	13.7%	14.7%		
V033	-	1.4%	2.9%	58.0%	3.2%	33.2%	1.4%	7.3%	11.4%		
V034	ZIR	1.4%	7.7%	62.9%	3.9%	23.5%	0.5%	10.2%	11.5%		
V035	ZIR	0.5%	2.4%	57.6%	2.9%	35.1%	1.5%	9.7%	10.9%		
V036	ZIR	1.2%	3.8%	58.7%	1.4%	34.5%	0.5%	8.1%	11.7%		
V037	-	0.9%	2.1%	58.8%	1.9%	35.6%	0.8%	8.8%	12.1%		
V038	-	1.2%	3.5%	62.0%	2.9%	29.3%	1.1%	6.2%	40.3%		
V039	ZIR	0.6%	4.1%	60.6%	2.1%	32.3%	0.3%	13.9%	10.8%		
V040	ZIR	1.3%	1.8%	56.7%	3.0%	35.5%	1.7%	9.3%	12.4%		
V041	ZIR	0.8%	0.0%	47.7%	6.6%	42.3%	2.5%	14.7%	12.0%		

**Tabelle 26 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\gamma$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	$\gamma$ -Untereinheit							
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Wasser- verlust
V042	ZIR	1.5%	2.8%	67.7%	6.4%	20.3%	1.4%	8.7%	10.4%
V043	ZIR	1.4%	1.2%	51.7%	2.1%	42.3%	1.3%	9.9%	10.3%
V045	ZIR	0.6%	6.5%	62.5%	2.0%	28.1%	0.4%	13.2%	10.9%
V046	ZIR	0.3%	7.7%	44.3%	10.6%	32.1%	4.9%	15.0%	10.2%
V047	-	0.7%	5.5%	63.5%	1.4%	28.8%	0.1%	14.9%	10.8%
V048	HCC	0.5%	2.7%	56.7%	1.9%	37.4%	0.7%	12.6%	12.9%

**Tabelle 27: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der A $\alpha$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	A $\alpha$ -Untereinheit													
		unglycosyliert				O-glycosyliert				Gesamt O-glycosyliert	Einfach oxidiert	Zweifach oxidiert	Wasserverlust	Carbamido	Trunkiert
		0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>	0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>						
U	Gesund	38.9%	21.8%	17.1%	2.6%	10.9%	5.3%	3.4%	0.1%	19.7%	20.8%	9.2%	12.4%	7.4%	2.6%
V	Gesund	21.1%	24.6%	28.4%	5.9%	5.5%	6.5%	7.4%	0.5%	20.0%	23.1%	10.5%	12.1%	8.6%	0.6%
W	Gesund	25.6%	23.9%	26.9%	6.2%	6.1%	4.9%	6.1%	0.2%	17.3%	22.1%	8.9%	12.9%	7.5%	0.7%
X	Gesund	28.0%	24.6%	24.0%	5.6%	6.4%	5.6%	5.6%	0.3%	17.8%	26.4%	12.0%	11.0%	8.6%	0.8%
Y	Gesund	38.2%	21.2%	17.8%	2.6%	10.8%	5.2%	3.9%	0.2%	20.2%	22.9%	9.3%	11.5%	7.2%	2.9%
Z	Gesund	29.7%	23.5%	22.0%	4.1%	8.3%	6.3%	5.9%	0.2%	20.6%	14.5%	7.4%	13.8%	6.5%	1.5%
AA	Gesund	32.3%	20.6%	20.5%	3.8%	9.9%	7.5%	5.5%	0.0%	22.8%	24.2%	8.1%	8.6%	5.5%	2.6%
AB	Gesund	24.3%	24.7%	27.7%	5.0%	5.6%	6.1%	6.5%	0.2%	18.4%	12.4%	3.7%	12.5%	6.2%	0.5%
AC	Gesund	34.2%	22.8%	21.4%	4.1%	8.0%	5.5%	4.0%	0.0%	17.5%	12.2%	4.3%	16.7%	7.5%	1.9%
AD	Gesund	31.3%	22.7%	21.2%	3.7%	9.1%	6.3%	5.5%	0.2%	21.1%	19.5%	7.5%	11.3%	6.2%	1.9%
V001	HCC	34.7%	23.0%	21.2%	2.9%	9.1%	5.4%	3.6%	0.1%	18.2%	17.3%	10.3%	13.2%	7.6%	2.2%
V002	HCC	36.7%	25.6%	22.2%	4.5%	5.5%	2.0%	2.1%	1.2%	10.9%	29.8%	16.1%	10.0%	8.1%	6.8%
V003	HCC	46.0%	23.6%	14.7%	1.0%	9.8%	2.6%	1.8%	0.6%	14.7%	23.9%	11.2%	12.4%	9.0%	18.7%
V004	HCC	32.9%	20.5%	23.6%	3.5%	8.8%	5.5%	4.6%	0.7%	19.5%	21.1%	10.5%	10.6%	8.9%	1.1%



**Tabelle 27 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\alpha\alpha$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	$\alpha\alpha$ -Untereinheit													
		unglycosyliert				O-glycosyliert				Gesamt O-glycosyliert	Einfach oxidiert	Zweifach oxidiert	Wasserverlust	Carbamido	Trunkiert
		0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>	0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>						
V005	HCC	24.9%	26.4%	24.4%	2.9%	7.9%	8.3%	4.5%	0.6%	21.4%	14.5%	5.2%	17.5%	8.7%	6.6%
V006	HCC	43.3%	26.3%	15.0%	0.4%	9.6%	5.0%	0.6%	0.0%	15.1%	28.3%	17.9%	13.4%	10.9%	3.8%
V007	ZIR	42.5%	21.8%	14.6%	1.3%	11.4%	5.3%	2.2%	1.0%	19.8%	14.4%	8.7%	12.8%	7.6%	18.5%
V008	HCC	39.0%	24.6%	21.3%	1.0%	7.8%	4.1%	2.2%	0.0%	14.1%	14.1%	5.1%	13.3%	8.3%	5.7%
V009	HCC	36.6%	25.2%	18.2%	3.3%	9.2%	5.4%	1.7%	0.4%	16.7%	22.9%	13.0%	14.3%	9.2%	14.9%
V010	-	31.8%	25.7%	21.4%	3.7%	7.8%	5.7%	3.8%	0.0%	17.4%	20.8%	12.4%	14.4%	10.5%	0.7%
V011	HCC	41.4%	22.2%	18.0%	3.2%	9.1%	4.7%	1.5%	0.0%	15.2%	18.5%	8.8%	13.2%	10.0%	2.2%
V012	HCC	41.6%	24.3%	15.3%	3.2%	9.0%	4.6%	2.0%	0.0%	15.6%	16.7%	11.1%	14.2%	10.5%	1.6%
V013	HCC	44.2%	26.0%	12.5%	3.0%	8.8%	4.4%	0.9%	0.2%	14.3%	23.6%	12.6%	15.1%	11.0%	2.5%
V014	HCC	46.8%	23.8%	9.3%	1.3%	13.5%	4.7%	0.4%	0.1%	18.8%	26.3%	18.5%	14.7%	14.7%	2.2%
V015	-	60.0%	16.6%	5.7%	1.6%	13.3%	3.0%	0.1%	0.0%	16.3%	12.5%	7.9%	18.3%	14.3%	5.5%
V016	-	28.8%	21.4%	21.3%	4.2%	10.0%	7.8%	7.3%	0.0%	25.1%	19.3%	8.9%	13.5%	7.5%	0.9%
V017	HCC	28.0%	25.3%	25.2%	3.9%	6.4%	5.8%	5.4%	0.1%	17.7%	27.8%	14.3%	10.5%	7.9%	1.0%
V019	HCC	29.9%	27.1%	21.8%	3.8%	8.7%	5.5%	2.9%	0.2%	17.4%	29.7%	20.3%	12.0%	10.2%	5.8%

**Tabelle 27 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\alpha$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	$\alpha$ -Untereinheit													
		unglycosyliert				O-glycosyliert				Gesamt O-glycosyliert	Einfach oxidiert	Zweifach oxidiert	Wasserverlust	Carbamido	Trunkiert
		0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>	0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>						
V020	HCC	32.7%	24.6%	22.1%	4.3%	7.6%	4.5%	4.0%	0.0%	16.2%	28.5%	15.7%	12.1%	9.6%	1.2%
V021	HCC	50.7%	22.3%	9.5%	1.7%	11.9%	3.4%	0.3%	0.2%	15.8%	25.1%	15.5%	16.2%	12.0%	6.2%
V022	HCC	26.3%	24.9%	24.2%	4.9%	7.1%	7.2%	5.0%	0.5%	19.8%	20.6%	11.2%	13.2%	9.9%	4.2%
V023	HCC	36.6%	26.4%	16.8%	4.5%	10.1%	4.9%	0.7%	0.0%	15.7%	27.5%	20.9%	13.9%	13.1%	0.8%
V025	HCC	54.9%	20.6%	9.7%	1.2%	11.1%	2.2%	0.1%	0.2%	13.7%	24.0%	13.1%	14.6%	10.2%	13.4%
V026	HCC	52.5%	22.3%	9.3%	1.6%	11.8%	2.0%	0.3%	0.1%	14.2%	25.4%	16.3%	14.2%	12.4%	8.0%
V027	ZIR	51.3%	19.7%	9.1%	1.0%	14.0%	4.9%	0.0%	0.0%	18.8%	29.3%	16.0%	10.8%	7.9%	4.1%
V028	HCC	35.9%	23.9%	19.8%	2.1%	8.7%	5.3%	4.2%	0.2%	18.3%	17.9%	8.0%	12.9%	6.8%	9.6%
V029	ZIR	31.5%	22.3%	25.3%	4.4%	7.3%	4.7%	4.4%	0.0%	16.5%	23.2%	9.8%	12.9%	8.0%	1.0%
V030	ZIR	26.7%	28.9%	27.8%	6.8%	2.4%	4.4%	2.9%	0.0%	9.8%	28.0%	19.3%	12.8%	11.0%	0.8%
V031	ZIR	21.3%	27.6%	32.0%	6.3%	3.5%	4.6%	4.6%	0.0%	12.7%	28.8%	19.9%	12.0%	10.0%	2.5%
V032	ZIR	49.0%	23.5%	15.9%	1.8%	8.7%	0.7%	0.2%	0.3%	9.8%	24.8%	11.8%	14.7%	8.1%	22.0%
V033	-	30.9%	23.8%	21.3%	3.1%	8.5%	6.5%	5.4%	0.3%	20.8%	17.0%	6.6%	13.9%	4.4%	6.3%
V034	ZIR	38.2%	29.5%	20.8%	2.8%	5.0%	2.8%	0.8%	0.0%	8.6%	27.2%	14.2%	14.1%	9.3%	3.2%

**Tabelle 27 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der  $\alpha$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	$\alpha$ -Untereinheit													
		unglycosyliert				O-glycosyliert				Gesamt O-glycosyliert	Einfach oxidiert	Zweifach oxidiert	Wasserverlust	Carbamido	Trunkiert
		0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>	0 x PO <sub>4</sub>	1 x PO <sub>4</sub>	2 x PO <sub>4</sub>	3 x PO <sub>4</sub>						
V035	ZIR	47.0%	22.3%	15.3%	0.6%	10.0%	4.1%	0.8%	0.0%	14.9%	27.1%	11.8%	12.8%	9.5%	3.7%
V036	ZIR	35.8%	22.2%	24.3%	2.8%	7.6%	3.6%	3.7%	0.0%	15.0%	15.7%	3.7%	15.9%	5.3%	7.4%
V037	-	24.4%	23.0%	30.0%	4.9%	5.6%	5.2%	6.7%	0.2%	17.7%	10.4%	6.0%	16.7%	6.4%	1.2%
V038	-	24.1%	26.4%	27.4%	6.9%	4.8%	5.3%	5.0%	0.1%	15.2%	29.7%	18.1%	11.9%	8.4%	0.8%
V039	ZIR	30.9%	27.7%	25.1%	3.2%	5.1%	4.5%	3.5%	0.1%	13.1%	29.3%	14.6%	12.8%	8.0%	5.5%
V040	ZIR	27.3%	24.6%	25.8%	3.3%	6.7%	5.4%	6.4%	0.4%	19.0%	26.1%	11.6%	14.2%	5.9%	18.1%
V041	ZIR	31.3%	28.9%	28.6%	3.6%	3.5%	2.4%	1.7%	0.0%	7.6%	16.1%	7.3%	22.4%	10.2%	15.8%
V042	ZIR	34.8%	27.1%	20.1%	0.3%	8.9%	6.5%	2.4%	0.0%	17.8%	18.2%	4.7%	13.8%	5.6%	6.1%
V043	ZIR	44.5%	21.1%	16.6%	1.3%	12.1%	3.7%	0.7%	0.0%	16.5%	32.2%	15.0%	5.6%	4.2%	11.7%
V045	ZIR	18.4%	21.9%	37.3%	6.7%	3.9%	4.3%	7.6%	0.0%	15.8%	21.8%	8.7%	10.3%	6.9%	0.2%
V046	ZIR	47.2%	33.7%	14.7%	0.6%	2.3%	1.1%	0.3%	0.2%	3.8%	30.5%	21.0%	9.2%	8.8%	17.2%
V047	-	21.6%	24.4%	31.3%	6.0%	4.2%	5.7%	6.8%	0.0%	16.6%	24.1%	9.3%	11.6%	6.2%	0.4%
V048	HCC	27.9%	25.2%	25.8%	4.7%	5.8%	5.7%	4.9%	0.0%	16.3%	13.6%	5.5%	16.1%	7.6%	1.9%

**Tabelle 28: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der B $\beta$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	B $\beta$ -Untereinheit									
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Trunktiert	Wasser- verlust	Carbamido
U	Gesund	0.0%	5.3%	55.0%	5.1%	34.6%	0.0%	24.9%	19.7%	11.4%	14.0%
V	Gesund	0.0%	3.9%	54.2%	4.1%	37.6%	0.1%	30.2%	25.4%	9.4%	15.0%
W	Gesund	0.0%	3.9%	51.2%	3.3%	41.4%	0.2%	28.0%	18.6%	11.7%	8.8%
X	Gesund	1.2%	4.8%	58.2%	2.5%	33.3%	0.0%	28.0%	18.9%	10.3%	7.5%
Y	Gesund	0.8%	4.2%	52.2%	3.5%	39.2%	0.0%	28.0%	22.9%	9.9%	7.7%
Z	Gesund	0.1%	3.3%	47.9%	2.9%	45.8%	0.1%	20.6%	21.1%	12.4%	8.5%
AA	Gesund	0.2%	3.7%	57.5%	3.5%	35.0%	0.0%	27.9%	10.9%	12.1%	10.1%
AB	Gesund	0.6%	4.4%	54.3%	2.8%	38.0%	0.0%	18.4%	20.7%	12.3%	7.7%
AC	Gesund	0.0%	4.8%	53.6%	4.9%	36.7%	0.0%	19.5%	11.0%	14.7%	15.3%
AD	Gesund	0.1%	5.3%	50.9%	5.7%	37.9%	0.0%	22.9%	19.9%	11.5%	10.7%
V001	HCC	1.1%	6.0%	55.8%	6.8%	30.4%	0.0%	29.4%	32.5%	9.6%	12.9%
V002	HCC	0.0%	3.7%	51.8%	8.2%	34.8%	1.4%	36.3%	22.1%	10.2%	11.8%
V003	HCC	0.0%	3.5%	49.2%	7.6%	39.4%	0.4%	35.0%	22.8%	10.8%	11.8%
V004	HCC	0.4%	3.6%	49.9%	5.9%	40.2%	0.0%	32.0%	30.2%	9.3%	12.2%

**Tabelle 28 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der B $\beta$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	B $\beta$ -Untereinheit									
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Trunktiert	Wasser- verlust	Carbamido
V005	HCC	0.1%	2.6%	42.8%	6.0%	48.3%	0.3%	30.3%	28.5%	10.0%	11.2%
V006	HCC	0.0%	1.7%	34.7%	10.4%	52.5%	0.8%	32.4%	20.8%	11.3%	13.7%
V007	ZIR	0.0%	0.9%	27.2%	11.7%	59.8%	0.4%	22.9%	29.9%	11.9%	10.7%
V008	HCC	0.1%	2.8%	40.8%	7.2%	48.6%	0.5%	18.9%	17.9%	12.1%	5.8%
V009	HCC	0.0%	2.2%	50.2%	8.5%	38.3%	0.8%	30.1%	20.4%	12.4%	17.5%
V010	-	0.0%	3.8%	49.3%	4.9%	41.4%	0.6%	29.6%	21.6%	12.5%	11.7%
V011	HCC	0.6%	5.7%	58.9%	6.1%	28.8%	0.0%	32.8%	30.6%	8.4%	14.9%
V012	HCC	0.1%	4.5%	58.9%	4.2%	32.3%	0.1%	25.1%	23.8%	11.2%	13.0%
V013	HCC	0.0%	4.0%	60.3%	4.2%	31.4%	0.1%	36.0%	24.6%	8.7%	13.7%
V014	HCC	0.0%	2.0%	41.6%	6.1%	48.3%	2.1%	41.8%	23.2%	8.6%	15.3%
V015	-	0.1%	3.4%	53.1%	3.5%	39.7%	0.3%	23.6%	25.4%	11.0%	12.2%
V016	-	0.1%	3.7%	48.0%	5.2%	42.8%	0.1%	30.1%	25.9%	10.2%	10.7%
V017	HCC	0.0%	5.4%	57.6%	5.0%	32.0%	0.0%	36.9%	27.2%	8.5%	9.8%
V019	HCC	0.0%	0.7%	35.8%	9.3%	52.0%	2.2%	34.0%	21.8%	10.9%	13.7%

**Tabelle 28 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der B $\beta$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

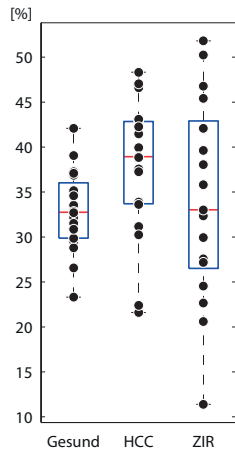
Probe	Zustand	B $\beta$ -Untereinheit										
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Trunktiert	Wasser- verlust	Carbamido	
V020	HCC	0.0%	3.2%	41.1%	5.4%	48.9%	1.5%	35.6%	21.4%	11.4%	9.9%	
V021	HCC	0.0%	0.5%	51.0%	7.9%	39.0%	1.6%	35.6%	18.0%	10.6%	20.6%	
V022	HCC	0.0%	2.8%	57.2%	3.4%	36.0%	0.6%	27.5%	22.9%	10.0%	11.7%	
V023	HCC	0.0%	2.6%	42.8%	6.6%	46.2%	1.8%	43.5%	24.5%	7.6%	14.7%	
V025	HCC	0.0%	4.0%	55.5%	5.9%	34.1%	0.5%	33.6%	21.7%	9.2%	12.5%	
V026	HCC	0.0%	2.8%	53.7%	9.4%	33.0%	1.1%	39.6%	26.5%	7.5%	14.0%	
V027	ZIR	0.0%	2.8%	38.9%	10.4%	47.7%	0.3%	38.0%	28.7%	7.9%	9.6%	
V028	HCC	0.0%	3.8%	48.5%	8.2%	39.5%	0.0%	22.0%	20.6%	11.9%	11.4%	
V029	ZIR	0.0%	1.8%	42.6%	7.4%	46.0%	2.2%	26.7%	13.0%	15.1%	13.6%	
V030	ZIR	0.0%	1.9%	45.3%	12.3%	36.5%	4.0%	34.6%	26.1%	11.7%	15.4%	
V031	ZIR	0.0%	3.8%	50.9%	10.3%	33.6%	1.4%	41.7%	30.3%	9.0%	9.9%	
V032	ZIR	0.0%	1.5%	32.9%	10.0%	52.0%	3.6%	32.1%	21.6%	12.2%	9.5%	
V033	-	0.0%	4.1%	49.3%	6.7%	39.8%	0.0%	23.7%	20.6%	11.4%	8.7%	
V034	ZIR	0.0%	5.5%	55.9%	9.6%	28.9%	0.0%	38.1%	20.7%	9.3%	9.7%	

**Tabelle 28 Fortsetzung: Ergebnisse der Quantifizierung der Glycoformen und Proteinspezies der B $\beta$ -Untereinheit aus den Proben des Validierungsdatensatzes.**

Probe	Zustand	B $\beta$ -Untereinheit										
		G2	G1S1	S1	S1F	S2	S2F	Oxidiert	Trunktiert	Wasser- verlust	Carbamido	
V035	ZIR	0.0%	2.8%	49.1%	5.8%	41.7%	0.5%	32.5%	17.7%	10.1%	7.4%	
V036	ZIR	0.0%	3.6%	49.5%	4.8%	42.0%	0.0%	25.4%	13.4%	12.2%	11.8%	
V037	-	0.0%	3.4%	47.3%	3.7%	45.6%	0.1%	18.5%	3.4%	16.1%	13.0%	
V038	-	0.0%	5.2%	54.9%	5.3%	34.6%	0.0%	41.1%	21.5%	9.0%	9.4%	
V039	ZIR	0.0%	3.9%	53.3%	5.9%	36.8%	0.0%	40.0%	26.8%	7.9%	9.8%	
V040	ZIR	0.0%	3.8%	48.0%	9.1%	39.1%	0.0%	32.5%	27.7%	9.3%	15.4%	
V041	ZIR	0.0%	2.3%	35.8%	13.9%	45.5%	2.6%	28.3%	23.7%	11.0%	15.5%	
V042	ZIR	0.3%	5.3%	64.3%	6.2%	23.7%	0.2%	26.4%	15.3%	11.3%	6.6%	
V043	ZIR	0.1%	3.6%	44.6%	5.9%	45.8%	0.0%	37.9%	26.7%	7.9%	9.0%	
V045	ZIR	0.2%	4.9%	56.8%	3.7%	34.3%	0.0%	24.3%	21.6%	10.9%	11.2%	
V046	ZIR	0.0%	3.8%	38.1%	20.1%	35.1%	2.8%	41.6%	37.7%	5.8%	13.9%	
V047	-	0.0%	4.5%	54.0%	4.6%	36.9%	0.0%	28.0%	17.8%	11.8%	14.8%	
V048	HCC	0.2%	3.7%	47.1%	6.1%	42.8%	0.0%	20.6%	27.6%	12.2%	9.8%	

## 9.4 Boxplots und Ergebnisse der Varianzanalysen

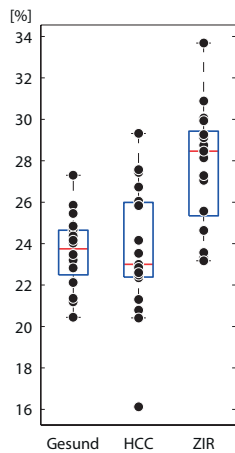
### $A\alpha-0 \times PO_4$



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	266.41	2	133.21	2.076	0.136
Within groups	3401.30	53	64.18		
Total	3667.72	55			

	Mean	SE	95% CI	p-value
Healthy	32.83%	1.79%		
HCC	37.88%	1.84%		
ZIR	34.06%	1.94%		
Healthy vs. HCC			-11.34 - 1.14%	0.130
Healthy vs. ZIR			-7.60 - 5.14%	0.888
HCC vs. ZIR			-2.63 - 10.27%	0.334

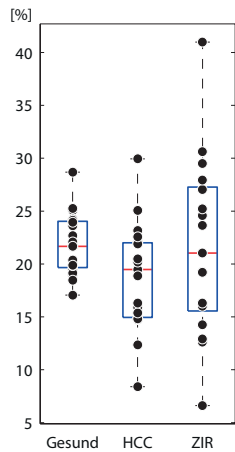
### $A\alpha-1 \times PO_4$



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	192.39	2	96.20	14.01	<0.011
Within groups	361.85	53	6.83		
Total	554.24	55			

	Mean	SE	95% CI	p-value
Healthy	23.65%	0.58%		
HCC	23.75%	0.60%		
ZIR	27.73%	0.63%		
Healthy vs. HCC			-2.12 - 1.91%	0.991
Healthy vs. ZIR			-6.16 - -2.0%	<0.001
HCC vs. ZIR			-6.08 - -1.87%	<0.001

### $A\alpha-2 \times PO_4$

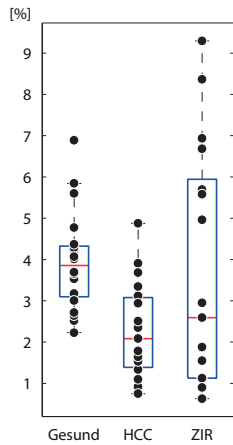


Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	110.17	2	55.09	1.69	0.194
Within groups	1727.52	53	32.60		
Total	1837.69	55			

	Mean	SE	95% CI	p-value
Healthy	21.83%	1.28%		
HCC	18.78%	1.31%		
ZIR	21.63%	1.38%		
Healthy vs. HCC			-1.36 - 7.46%	0.227
Healthy vs. ZIR			-4.34 - 4.74%	0.994
HCC vs. ZIR			-7.45 - 1.75%	0.301



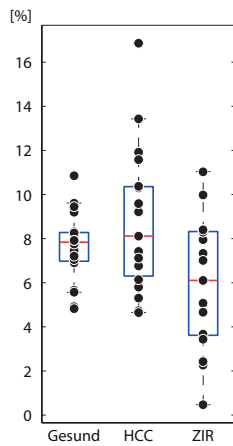
$\text{A}\alpha\text{-3} \times \text{PO}_4$



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	32.32	2	16.16	4.55	0.015
Within groups	188.04	53	3.55		
Total	220.35	55			

	Mean	SE	95% CI	P-value
Healthy	3.95%	0.42%		
HCC	2.23%	0.43%		
ZIR	3.65%	0.46%		
Healthy vs. HCC			0.27 - 3.18%	0.017
Healthy vs. ZIR			-1.20 - 1.80%	0.880
HCC vs. ZIR			-2.94 - 0.09%	0.070

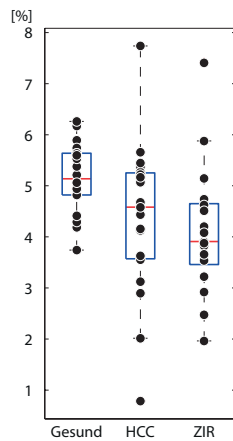
$\text{A}\alpha\text{-0} \times \text{PO}_4\text{-O-glycosyliert}$



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	69.64	2	34.82	5.02	0.010
Within groups	367.58	53	6.94		
Total	437.22	55			

	Mean	SE	95% CI	P-value
Healthy	7.61%	0.59%		
HCC	8.73%	0.60%		
ZIR	5.95%	0.64%		
Healthy vs. HCC			-3.15 - 0.92%	0.389
Healthy vs. ZIR			-0.44 - 3.75%	0.146
HCC vs. ZIR			0.66 - 4.90%	0.007

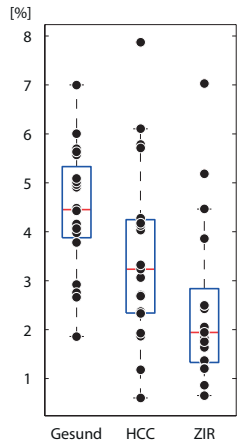
$\text{A}\alpha\text{-1} \times \text{PO}_4\text{-O-glycosyliert}$



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	11.25	2	5.63	3.92	0.026
Within groups	76.05	53	1.44		
Total	87.31	55			

	Mean	SE	95% CI	P-value
Healthy	5.16%	0.27%		
HCC	4.36%	0.27%		
ZIR	4.12%	0.29%		
Healthy vs. HCC			-0.13 - 1.72%	0.105
Healthy vs. ZIR			0.09 - 2.00%	0.029
HCC vs. ZIR			-0.72 - 1.21%	0.718

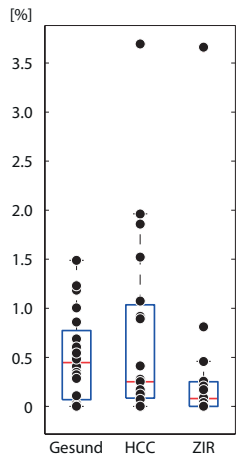
$\alpha$ -2 × PO<sub>4</sub> - O-glycosyliert



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	35.15	2	17.57	6.76	0.002
Within groups	137.70	53	2.60		
Total	172.84	55			

	Mean	SE	95% CI	P-value
Healthy	4.45%	0.36%		
HCC	3.55%	0.37%		
ZIR	2.50%	0.39%		
Healthy vs. HCC			-0.34 - 2.15%	0.194
Healthy vs. ZIR			0.67 - 3.24%	0.002
HCC vs. ZIR			-0.25 - 2.35%	0.136

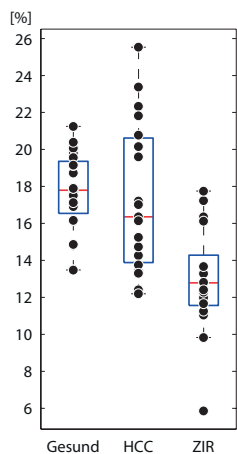
$\alpha$ -3 × PO<sub>4</sub> - O-glycosyliert



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	1.16	2	0.58	0.93	0.402
Within groups	33.10	53	0.62		
Total	34.25	55			

	Mean	SE	95% CI	P-value
Healthy	0.51%	0.18%		
HCC	0.71%	0.18%		
ZIR	0.35%	0.19%		
Healthy vs. HCC			-0.81 - 0.41%	0.708
Healthy vs. ZIR			-0.47 - 0.78%	0.822
HCC vs. ZIR			-0.28 - 0.99%	0.372

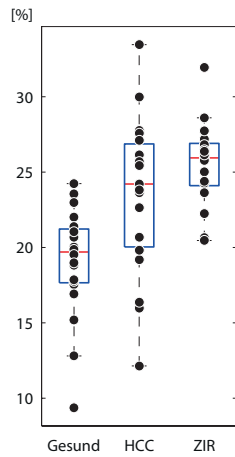
$\alpha$ -O-glycosyliert



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	254.77	2	127.39	13.15	<0.001
Within groups	513.33	53	9.69		
Total	768.10	55			

	Mean	SE	95% CI	P-value
Healthy	17.73%	0.70%		
HCC	17.35%	0.71%		
ZIR	12.92%	0.75%		
Healthy vs. HCC			-2.02 - 2.79%	0.921
Healthy vs. ZIR			2.34 - 7.29%	<0.001
HCC vs. ZIR			1.92 - 6.93	<0.001

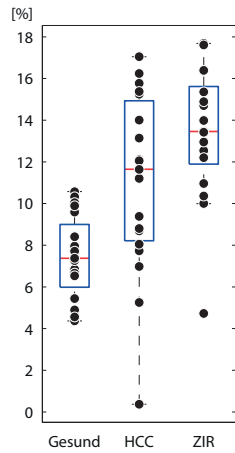
$\text{A}\alpha\text{-1} \times \text{oxidiert}$



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	408.83	2	204.42	12.57	<0.001
Within groups	861.87	53	16.26		
Total	1270.70	55			

	Mean	SE	95% CI	P-value
Healthy	19.08%	0.90%		
HCC	23.54%	0.93%		
ZIR	25.51%	0.98%		
Healthy vs. HCC			-7.58 - -1.35%	0.003
Healthy vs. ZIR			-9.64 - -3.23%	<0.001
HCC vs. ZIR			-5.22 - 1.28%	0.316

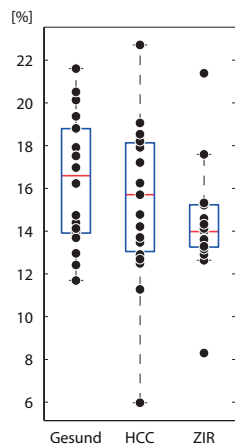
$\text{A}\alpha\text{-2} \times \text{oxidiert}$



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	339.71	2	169.85	15.63	<0.001
Within groups	576.09	53	10.87		
Total	915.80	55			

	Mean	SE	95% CI	P-value
Healthy	7.39%	0.74%		
HCC	11.01%	0.76%		
ZIR	13.39%	0.80%		
Healthy vs. HCC			-6.16 - -1.06%	0.003
Healthy vs. ZIR			-8.62 - -3.37%	<0.001
HCC vs. ZIR			-5.04 - 0.27%	0.086

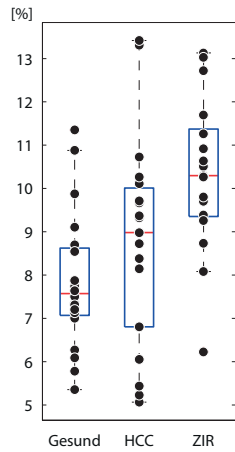
$\text{A}\alpha\text{-Wasserverlust}$



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	39.49	2	19.75	2.03	0.141
Within groups	514.98	53	9.71		
Total	554.47	55			

	Mean	SE	95% CI	P-value
Healthy	16.36%	0.70%		
HCC	15.36%	0.72%		
ZIR	14.29%	0.76%		
Healthy vs. HCC			-1.41 - 3.41%	0.579
Healthy vs. ZIR			-0.41 - 4.55%	0.118
HCC vs. ZIR			-1.44 - 3.58%	0.562

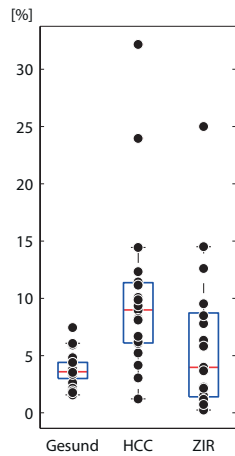
**A $\alpha$ -carbamidiert**



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	58.52	2	29.26	7.63	0.001
Within groups	203.36	53	3.84		
Total	261.88	55			

	Mean	SE	95% CI	P-value
Healthy	7.83%	0.44%		
HCC	8.71%	0.45%		
ZIR	10.33%	0.48%		
Healthy vs. HCC			-2.4 - 0.63%	0.344
Healthy vs. ZIR			-4.06 - -0.95%	0.001
HCC vs. ZIR			-3.20 - -0.04%	0.043

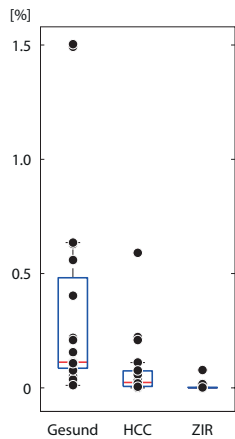
**A $\alpha$ -trunkiert (minus Valin)**



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	394.36	2	197.18	6.31	0.003
Within groups	1656.88	53	31.26		
Total	2051.23	55			

	Mean	SE	95% CI	P-value
Healthy	3.83	1.25		
HCC	10.15	1.28		
ZIR	6.27	1.36		
Healthy vs. HCC			-10.64 - -2.00%	0.002
Healthy vs. ZIR			-6.89 - 2.01%	0.389
HCC vs. ZIR			-0.62 - 8.38%	0.104

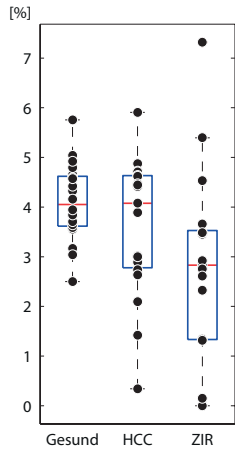
**B $\beta$ -G2**



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	1.11	2	0.56	7.19	0.002
Within groups	4.11	53	0.08		
Total	5.22	55			

	Mean	SE	95% CI	P-value
Healthy	0.33%	0.06%		
HCC	0.08%	0.06%		
ZIR	0.01%	0.07%		
Healthy vs. HCC			0.04 - 0.47%	0.016
Healthy vs. ZIR			0.10 - 0.55%	0.002
HCC vs. ZIR			-0.15 - 0.29%	0.728

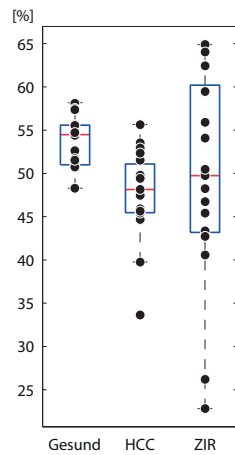
B $\beta$ -G1S1



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	15.82	2	7.91	3.98	0.025
Within groups	105.37	53	1.99		
Total	121.19	55			

	Mean	SE	95% CI	P-value
Healthy	4.10%	0.32%		
HCC	3.63%	0.32%		
ZIR	2.80%	0.34%		
Healthy vs. HCC			-0.61 - 1.57%	0.546
Healthy vs. ZIR			0.18 - 2.43%	0.019
HCC vs. ZIR			-0.31 - 1.96%	0.193

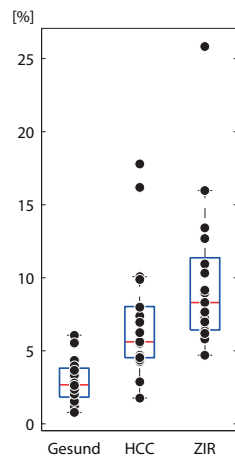
B $\beta$ -S1



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	368.66	2	184.33	3.238	0.047
Within groups	3017.11	53	56.93		
Total	3385.77	55			

	Mean	SE	95% CI	P-value
Healthy	53.69%	1.69%		
HCC	47.74%	1.73%		
ZIR	49.39%	1.83%		
Healthy vs. HCC			0.12 - 11.78%	0.044
Healthy vs. ZIR			-1.70 - 10.31%	0.204
HCC vs. ZIR			-7.72 - 4.42%	0.791

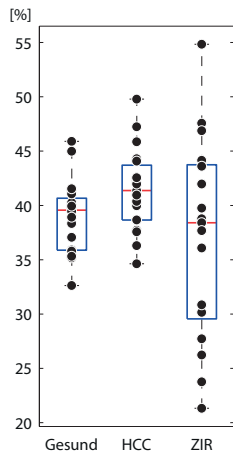
B $\beta$ -S1F



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	452.47	2	226.24	15.78	<0.001
Within groups	759.70	53	14.33		
Total	1212.18	55			

	Mean	SE	95% CI	P-value
Healthy	2.90%	0.85%		
HCC	7.07%	0.87%		
ZIR	9.82%	0.92%		
Healthy vs. HCC			-7.10 - -1.25%	0.003
Healthy vs. ZIR			-9.93 - -3.91%	<0.001
HCC vs. ZIR			-5.80 - 0.30%	0.085

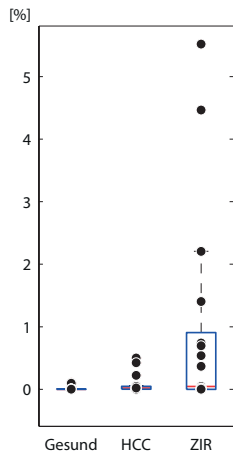
### B $\beta$ -S2



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	172.83	2	86.42	2.49	0.092
Within groups	1838.58	53	34.69		
Total	2011.42	55			

	Mean	SE	95% CI	P-value
Healthy	38.96%	1.32%		
HCC	41.41%	1.35%		
ZIR	37.04%	1.43%		
Healthy vs. HCC			-7.00 - 2.10%	0.403
Healthy vs. ZIR			-2.77 - 6.60%	0.589
HCC vs. ZIR			-0.38 - 9.11%	0.077

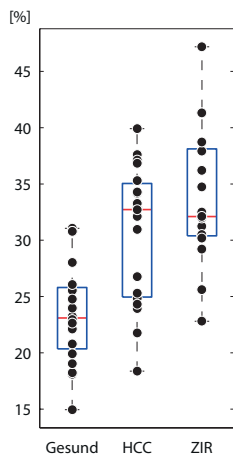
### B $\beta$ -S2F



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	9.62	2	4.81	5.80	0.005
Within groups	43.95	53	0.83		
Total	53.56	55			

	Mean	SE	95% CI	P-value
Healthy	0.01%	0.20%		
HCC	0.08%	0.21%		
ZIR	0.94%	0.22%		
Healthy vs. HCC			-0.77 - 0.64%	0.974
Healthy vs. ZIR			-1.65 - -0.21%	0.009
HCC vs. ZIR			-1.60 - -0.13%	0.017

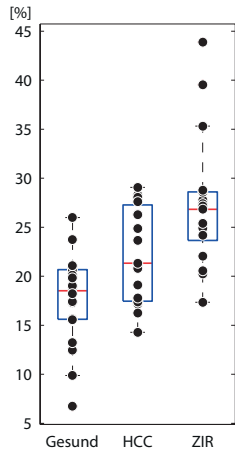
### B $\beta$ -oxidiert



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	1068.07	2	534.04	17.17	<0.001
Within groups	1648.21	53	31.10		
Total	2716.28	55			

	Mean	SE	95% CI	P-value
Healthy	23.39%	1.25%		
HCC	30.66%	1.28%		
ZIR	33.77%	1.35%		
Healthy vs. HCC			-11.58 - -2.96%	<0.001
Healthy vs. ZIR			-14.81 - -5.94%	<0.001
HCC vs. ZIR			-7.59 - 1.38%	0.227

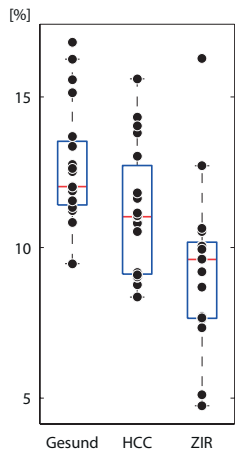
B $\beta$ -trunkiert (minus Glutamin)



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	851.78	2	425.89	14.28	<0.001
Within groups	1581.14	53	29.83		
Total	2432.92	55			

	Mean	SE	95% CI	P-value
Healthy	17.74%	1.22%		
HCC	22.16%	1.25%		
ZIR	27.36%	1.32%		
Healthy vs. HCC			-8.65 - -0.21%	0.038
Healthy vs. ZIR			-9.63 - -5.28%	<0.001
HCC vs. ZIR			-9.60 - -0.80%	0.017

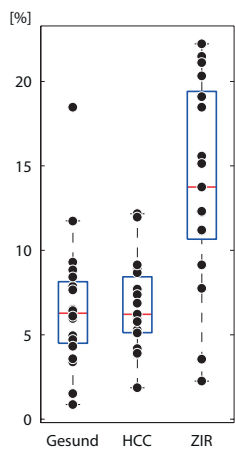
B $\beta$ -Wasserverlust



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	106.30	2	53.15	10.49	<0.001
Within groups	268.51	53	5.07		
Total	374.81	55			

	Mean	SE	95% CI	P-value
Healthy	12.66%	0.50%		
HCC	11.22%	0.52%		
ZIR	9.26%	0.55%		
Healthy vs. HCC			-0.29 - 3.18%	0.121
Healthy vs. ZIR			1.61 - 5.19%	<0.001
HCC vs. ZIR			0.14 - 3.76%	0.032

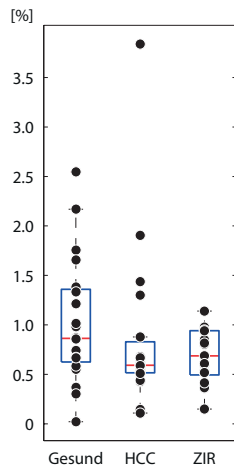
B $\beta$ -carbamidiert



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	621.77	2	310.89	16.39	<0.001
Within groups	1005.10	53	18.96		
Total	1626.87	55			

	Mean	SE	95% CI	P-value
Healthy	6.66%	0.97%		
HCC	6.71%	1.00%		
ZIR	13.93%	1.06%		
Healthy vs. HCC			-3.42 - 3.31%	0.999
Healthy vs. ZIR			-10.74 - -3.81%	<0.001
HCC vs. ZIR			-10.73 - -3.71%	<0.001

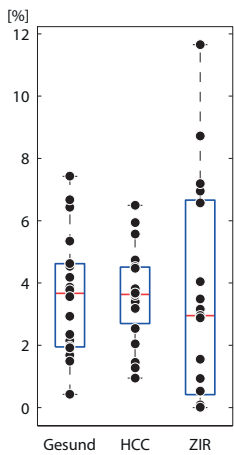
$\gamma$ -G2



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	1.07	2	0.54	1.33	0.274
Within groups	21.43	53	0.40		
Total	22.51	55			

	Mean	SE	95% CI	P-value
Healthy	1.03%	0.14%		
HCC	0.87%	0.15%		
ZIR	0.69%	0.15%		
Healthy vs. HCC			-0.33 - 0.65%	0.713
Healthy vs. ZIR			-0.16 - 0.85%	0.242
HCC vs. ZIR			-0.33 - 0.69%	0.670

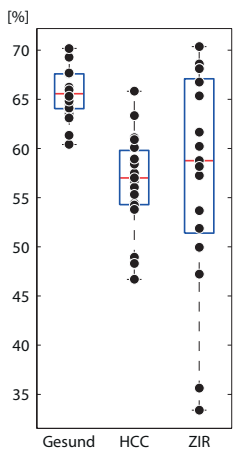
$\gamma$ -G1S1



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	0.01	2	0.01	0.00	0.999
Within groups	307.79	53	5.81		
Total	307.80	55			

	Mean	SE	95% CI	P-value
Healthy	3.59%	0.54%		
HCC	3.61%	0.55%		
ZIR	3.57%	0.58%		
Healthy vs. HCC			-1.88 - 1.85%	1.000
Healthy vs. ZIR			-1.89 - 1.94%	1.000
HCC vs. ZIR			-1.90 - 1.98%	0.999

$\gamma$ -S1

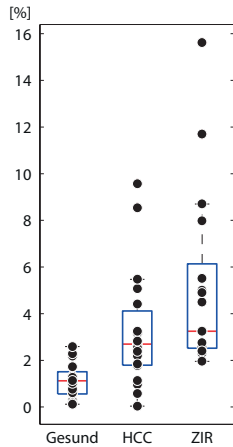


Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	927.33	2	463.67	9.78	<0.001
Within groups	2514.12	53	47.44		
Total	3441.45	55			

	Mean	SE	95% CI	P-value
Healthy	65.46%	1.54%		
HCC	56.65%	1.58%		
ZIR	57.36%	1.67%		
Healthy vs. HCC			3.49 - 14.13%	0.001
Healthy vs. ZIR			2.62 - 13.57%	0.002
HCC vs. ZIR			-6.26 - 4.83%	0.949



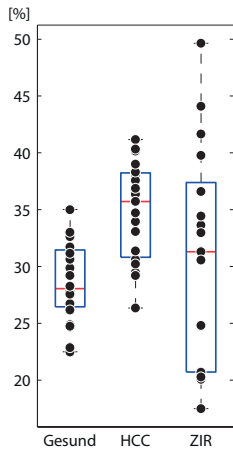
$\gamma$ -S1F



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	150.70	2	75.35	11.24	<0.001
Within groups	355.21	53	6.70		
Total	505.91	55			

	Mean	SE	95% CI	P-value
Healthy	1.10%	0.58%		
HCC	3.24%	0.59%		
ZIR	5.14%	0.63%		
Healthy vs. HCC			-4.14 - -0.14%	0.033
Healthy vs. ZIR			-6.10 - -1.98%	<0.001
HCC vs. ZIR			-3.98 - 0.19%	0.082

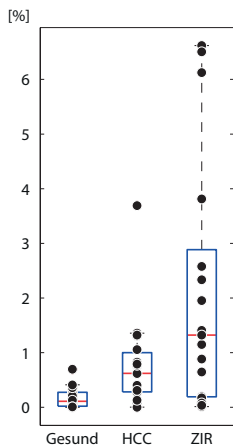
$\gamma$ -S2



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	380.83	2	190.42	5.05	0.010
Within groups	2000.45	53	37.74		
Total	2381.28	55			

	Mean	SE	95% CI	P-value
Healthy	28.65%	1.37%		
HCC	34.87%	1.41%		
ZIR	31.13%	1.49%		
Healthy vs. HCC			-10.97 - -1.48%	0.007
Healthy vs. ZIR			-7.37 - 2.40%	0.443
HCC vs. ZIR			-1.21 - 8.68%	0.172

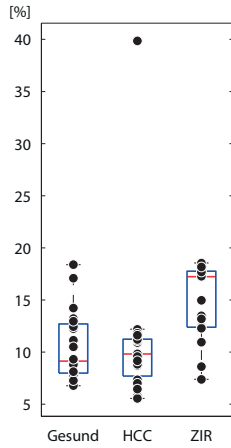
$\gamma$ -S2F



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	35.70	2	17.85	9.64	<0.001
Within groups	98.10	53	1.85		
Total	133.79	55			

	Mean	SE	95% CI	P-value
Healthy	0.17%	0.30%		
HCC	0.76%	0.31%		
ZIR	2.11%	0.33%		
Healthy vs. HCC			-1.64 - 0.46%	0.372
Healthy vs. ZIR			-3.02 - -0.86%	<0.001
HCC vs. ZIR			-2.44 - -0.25%	0.012

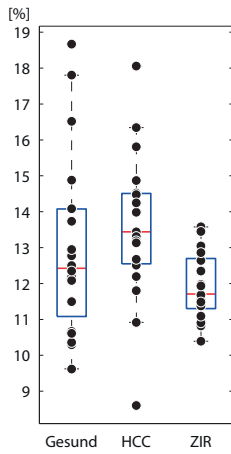
$\gamma$ -oxidiert



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	200.19	2	100.09	3.87	0.027
Within groups	1370.09	53	25.85		
Total	1570.27	55			

	Mean	SE	95% CI	P-value
Healthy	10.61%	1.14%		
HCC	11.01%	1.17%		
ZIR	14.90%	1.23%		
Healthy vs. HCC			-4.33 - 3.53%	0.967
Healthy vs. ZIR			-8.33 - -0.25%	0.035
HCC vs. ZIR			-7.98 - 0.20%	0.066

$\gamma$ -Wasserverlust



Source of variance	Sum of squares	df	Mean squares	F-ratio	P-value
Between groups	25.27	2	12.63	3.25	0.047
Within groups	206.33	53	3.89		
Total	231.59	55			

	Mean	SE	95% CI	P-value
Healthy	12.99%	0.44%		
HCC	13.59%	0.45%		
ZIR	11.93%	0.48%		
Healthy vs. HCC			-2.13 - 0.92%	0.607
Healthy vs. ZIR			-0.51 - 2.63%	0.243
HCC vs. ZIR			0.08 - 3.25%	0.038

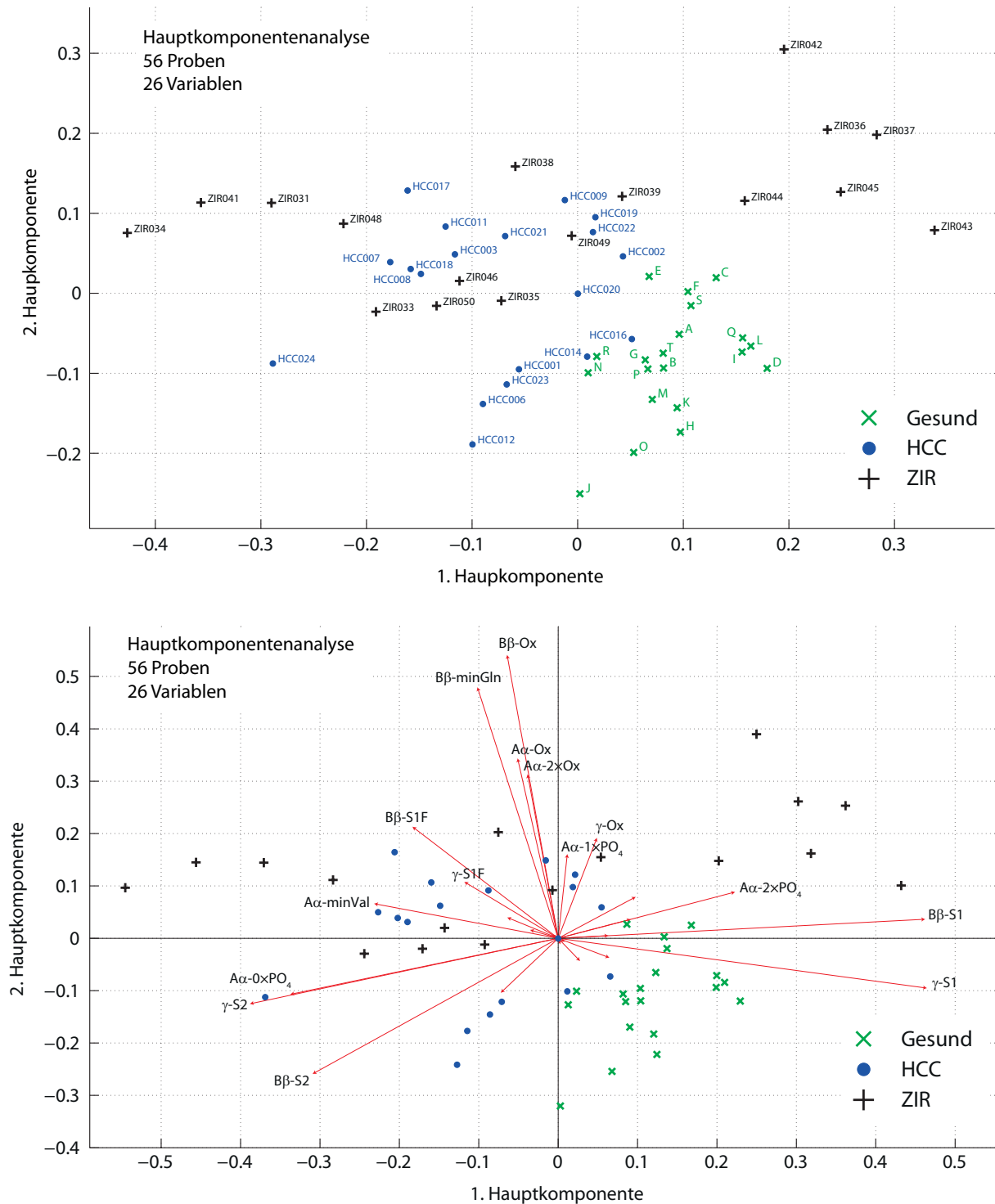
## 9.5 Erweiterte Diskussion der statistischen Datenauswertung

### 9.5.1 Entwicklung der Methode zur Klassifizierung unbekannter Proben

In Abschnitt 4.1.5 wurden stets vier Variablen als Motiv für die Hauptkomponentenanalyse und die MANOVA verwendet. Dabei handelt es sich um folgende vier Glycoformen bzw. Proteinspezies:  $B\beta$ -S1,  $B\beta$ -S1F,  $\gamma$ -S1 und  $\gamma$ -S1F. Die Frage, warum diese vier Variablen für die Hauptkomponentenanalyse und MANOVA ausgewählt wurden, soll im Folgenden diskutiert werden.

Die Hauptkomponentenanalyse wurde zunächst mit 26 Variablen durchgeführt. Dabei wurden alle quantifizierten Glycoformen und Proteinspezies als Variablen verwendet, außer die Anteile der carbamidierten Untereinheiten und die Spezies mit Wasserverlust, da diese ausschließlich bei der Probenvorbereitung bzw. in der Quelle des Massenspektrometers entstehen. Die Ergebnisse der Hauptkomponentenanalyse mit 26 Variablen sind in Abbildung 72 gezeigt. Im Diagramm der ersten und zweiten Hauptkomponente für jede Probe ist zu erkennen, dass die Datenpunkte der gesunden Spender in einem kleinen Bereich im unteren rechten Teil des Diagrammes liegen. Die Datenpunkte von HCC- und Zirrhosepatienten liegen außerhalb dieses Bereiches. Eine klare Trennung zwischen HCC- und Zirrhosepatienten ist nicht zu erkennen, die Datenpunkte der HCC-Patienten liegen allerdings alle im zentralen Bereich des Diagrammes, während die Datenpunkte der Zirrhosepatienten über den gesamten oberen Bereich des Diagrammes verteilt liegen.

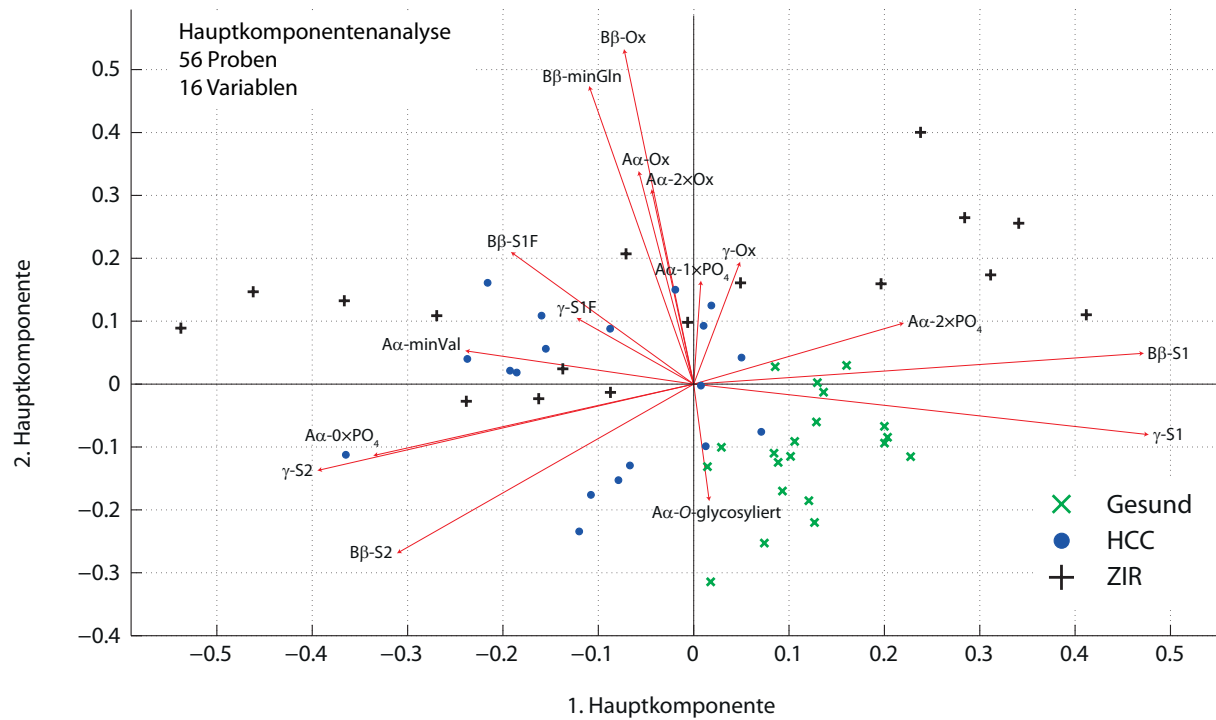
In Abbildung 72 sind außerdem die Koeffizienten der ersten beiden Hauptkomponenten aller 26 Variablen als Vektoren dargestellt. Die Vektoren geben Auskunft darüber, wie sich die einzelnen Variablen auf die Lage eines Datenpunktes im Diagramm der ersten beiden Hauptkomponenten auswirken. So zeigen zum Beispiel die Vektoren der  $B\beta/\gamma$ -S1 nach rechts in Richtung der Datenpunkte von gesunden Spendern, während die Vektoren der Glycoformen  $B\beta/\gamma$ -S2 nach unten links zu den Datenpunkten der HCC- Patienten zeigen. Dies stimmt mit dem Ergebnis überein, dass die Anteile der Glycoformen  $B\beta/\gamma$ -S2 bei HCC-Patienten erhöht sind. Außerdem ist zu erkennen, dass sechs Datenpunkte von Zirrhosepatienten isoliert im oberen rechten Teil des Diagrammes liegen, da diese Proben einen sehr geringen Sialylierungsgrad aufweisen. Die anderen Vektoren stimmen ebenfalls mit den Ergebnissen aus Abschnitt 4.1.5 überein. Die kurzen Vektoren sind in Abbildung 72 nicht beschriftet. Diese Variablen besitzen nur kleine Koeffizienten und haben daher nur wenig Einfluss auf die Lage eines Datenpunktes im Diagramm der ersten beiden Hauptkomponenten. Es wurde deshalb überprüft, ob auf diese Variablen in der Hauptkomponentenanalyse verzichtet werden kann.



**Abbildung 72: Ergebnisse der Hauptkomponentenanalyse mit 26 Variablen. Oben: Auftragung der ersten beiden Hauptkomponenten jeder Probe. Unten: Darstellung der Koeffizienten der ersten beiden Hauptkomponenten jeder Variable als Vektoren.**

Die Hauptkomponentenanalyse wurde mit 16 Variablen wiederholt, dabei wurden die folgenden Variablen weggelassen:  $A\alpha\text{-}3\times\text{PO}_4$ ,  $B\beta\text{-}G2$ ,  $B\beta\text{-}S2F$ ,  $B\beta\text{-}G1S1$ ,  $\gamma\text{-}G2$ ,  $\gamma\text{-}S2F$ ,  $\gamma\text{-}G1S1$ . Die O-glycosylierten Glycoformen der  $A\alpha$ -Untereinheit wurden zu einer Variablen vereint. Die Ergebnisse der Hauptkomponentenanalyse mit 16 Variablen sind in Abbildung 73

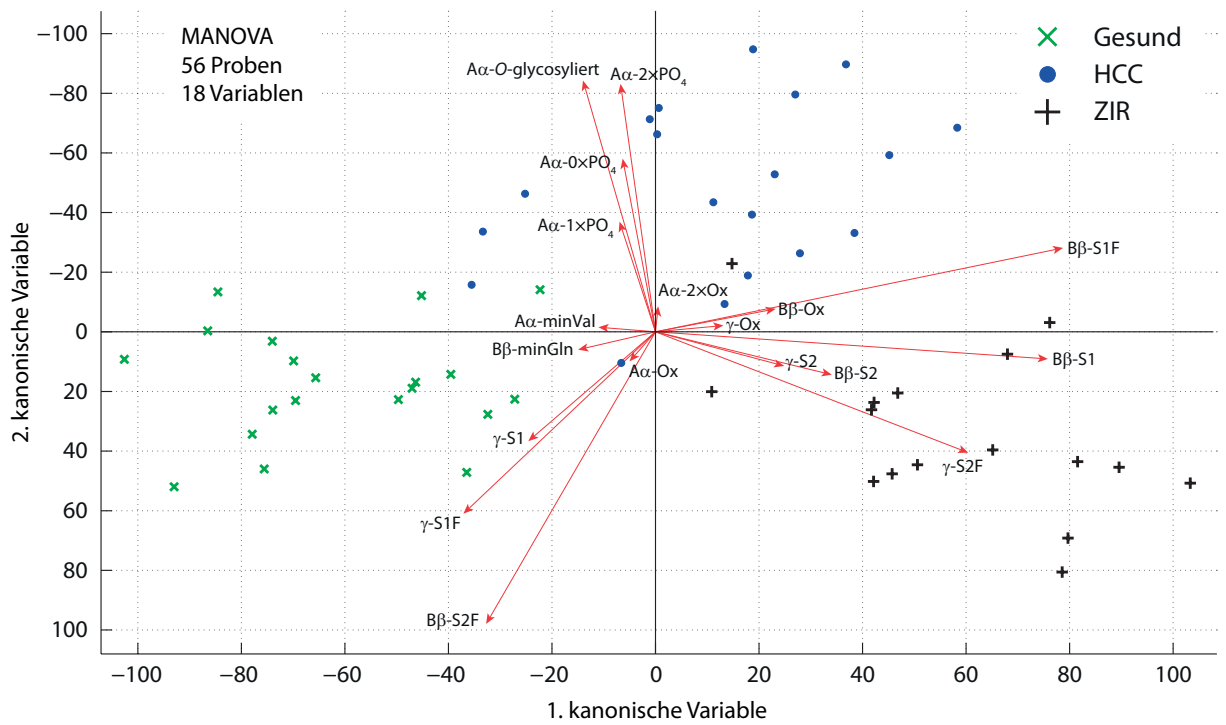
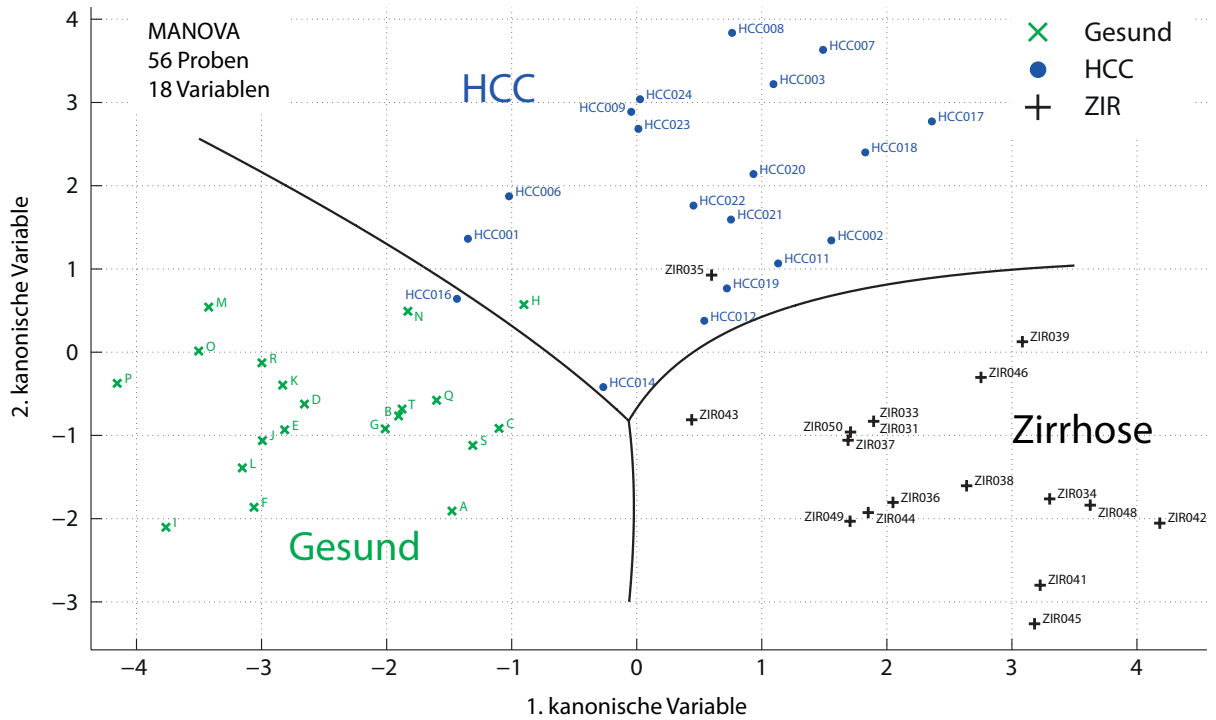
dargestellt. Erwartungsgemäß ist kein großer Unterschied zur Hauptkomponentenanalyse mit 26 Variablen festzustellen.



**Abbildung 73: Darstellung der Koeffizienten der ersten beiden Hauptkomponenten der Hauptkomponentenanalyse mit 16 Variablen als Vektoren.**

Die MANOVA wurde deshalb zunächst mit diesen 16 Variablen durchgeführt. Es zeigte sich allerdings, dass die beiden Variablen  $B\beta$ -S2F und  $\gamma$ -S2F großen Einfluss besitzen. Die Variablen  $B\beta$ -S2F und  $\gamma$ -S2F wurden daher für die MANOVA mit 18 Variablen verwendet, deren Ergebnisse in Tabelle 29 und Abbildung 74 dargestellt sind. Die Trennung zwischen den Datenpunkten von gesunden Spendern, HCC- und Zirrhosepatienten ist ausgezeichnet, nur drei Proben sind falsch klassifiziert. Der Datenpunkt von Probe H liegt knapp im Bereich der HCC-Patienten, der Datenpunkt von Probe HCC016 liegt knapp im Bereich gesunder Spender und der Datenpunkt von ZIR035 liegt im Bereich der HCC-Patienten.

In Abbildung 74 sind außerdem die Koeffizienten der ersten und zweiten kanonischen Variablen aller 18 verwendeten Variablen als Vektoren dargestellt. Die Vektoren erscheinen teilweise unlogisch. Beispielsweise zeigen die Vektoren der Glycoformen  $B\beta$ -S2F und  $\gamma$ -S1F in Richtung der Datenpunkte von gesunden Spendern, obwohl die Anteile dieser Glycoformen bei HCC- und Zirrhosepatienten erhöht sind. Andere Vektoren, wie zum Beispiel  $B\beta$ -S1F und  $\gamma$ -S2F zeigen erwartungsgemäß in Richtung der Datenpunkte von HCC- und Zirrhosepatienten.



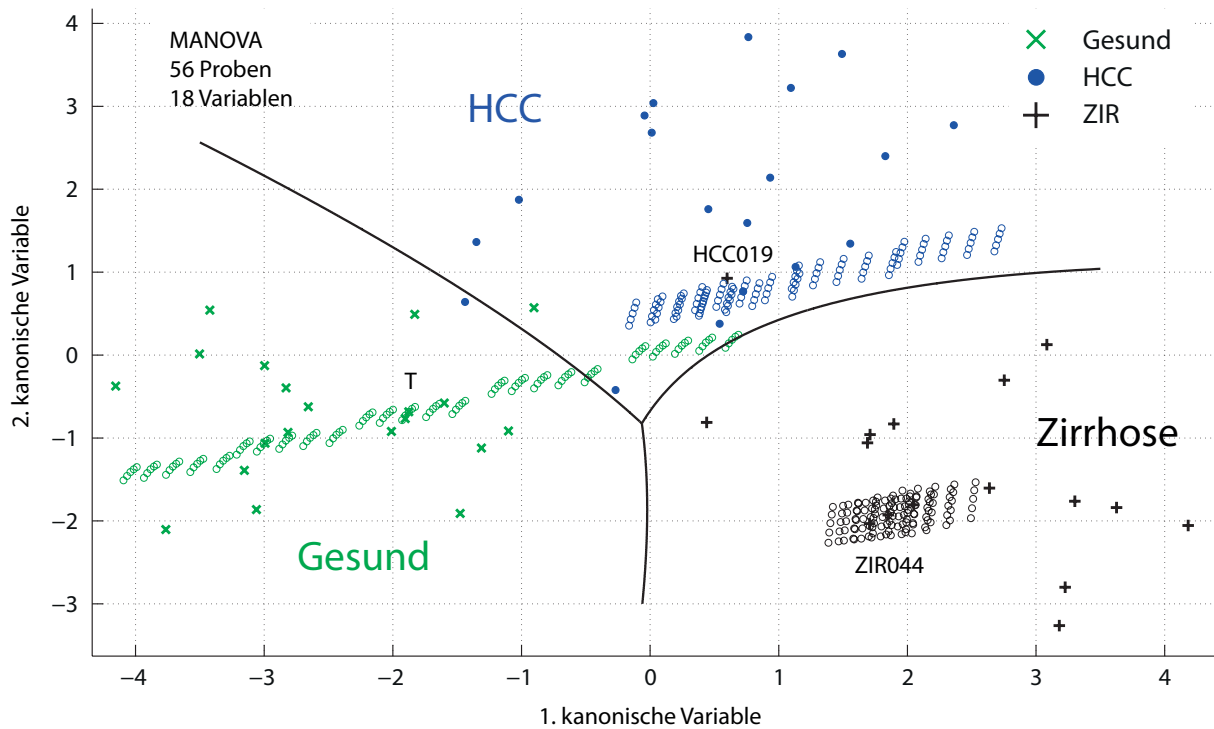
**Abbildung 74: Ergebnisse der MANOVA mit 18 Variablen. Oben: Auftragung der ersten und zweiten kanonischen Variablen jeder der 56 Proben. Unten: Darstellung der Koeffizienten der ersten und zweiten kanonischen Variablen jeder Variable als Vektor.**

**Tabelle 29: Ergebnisse der MANOVA mit 18 Variablen. Die Pillai-Spur Wilks-Lambda, die Hotelling-Spur und Roys größte charakteristische Wurzel sind verschiedene Arten die Signifikanz der Trennung zwischen den Gruppen zu bestimmen. In diesem Fall sind alle Werte klar signifikant (p-Wert < 0.001).**

	Wert	F	Hypothesen df	Fehler df	p-Wert	Effektgröße
Pillai-Spur	1.486	5.939	36	74	<0.001	0.743
Wilks-Lambda	0.062	6.022	36	72	<0.001	0.751
Hotelling-Spur	6.274	6.100	36	70	<0.001	0.758
Größte charakteristische Wurzel nach Roy	4.150	8.530	18	37	<0.001	0.806

In Abschnitt 4.1.5.3 wurde der Einfluss des *shift* und des gemittelten Bereiches des Chromatogrammes auf die Lage der Datenpunkte im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA diskutiert. In Abbildung 75 sind die durch Variation des *shift* erhaltenen 125 Datenpunkte der Proben T, HCC019 und ZIRO44 im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA mit 18 Variablen dargestellt. Es ist eine extrem große Streuung der jeweils 125 Datenpunkte zu erkennen. Die genaue Analyse der Daten zeigt, dass die Streuung hauptsächlich auf die Quantifizierung der  $\beta$ -Untereinheit zurückzuführen ist. Die Quantifizierung der Anteile oxidierter und trunkierter  $\beta$ -Untereinheit ist stark abhängig vom *shift*, so variiert der Anteil an Oxidation bei Probe T zwischen 16.8% und 29.5% ( $\pm 1.0$  Da) und der Anteil trunkierter  $\beta$ -Untereinheit zwischen 14.4% und 21.8% ( $\pm 1.0$  Da). Diese großen Unterschiede führen im Diagramm der ersten und zweiten kanonischen Variablen zu einer großen Streuung der Datenpunkte, da die MANOVA mit 18 Variablen die Anteile oxidierter und trunkierter  $\beta$ -Untereinheit zur Berechnung der kanonischen Variablen verwendet.

Die MANOVA mit 18 Variablen ist aus diesem Grund schlecht zur Einordnung unbekannter Proben geeignet, da bereits ein kleiner Fehler des *shift* große Auswirkung auf die Lage eines Datenpunktes haben kann. Die MANOVA wurde daher ohne die Anteile der oxidierten und trunkierten Spezies als Variablen wiederholt. In Tabelle 30 und Abbildung 76 sind die Ergebnisse der MANOVA mit zwölf Variablen ( $A\alpha\text{-}0\times\text{PO}_4$ ,  $A\alpha\text{-}1\times\text{PO}_4$ ,  $A\alpha\text{-}2\times\text{PO}_4$ ,  $A\alpha\text{-}O\text{-Glycosyliert}$ ,  $\beta\beta\text{-}S1$ ,  $\beta\beta\text{-}S1F$ ,  $\beta\beta\text{-}S2$ ,  $\beta\beta\text{-}S2F$ ,  $\gamma\text{-}S1$ ,  $\gamma\text{-}S1F$ ,  $\gamma\text{-}S2$  und  $\gamma\text{-}S2F$ ) dargestellt. Die Streuung der durch Variation des *shift* erhaltenen 125 Datenpunkte bei der MANOVA mit zwölf Variablen ist sehr klein (Abbildung 77).

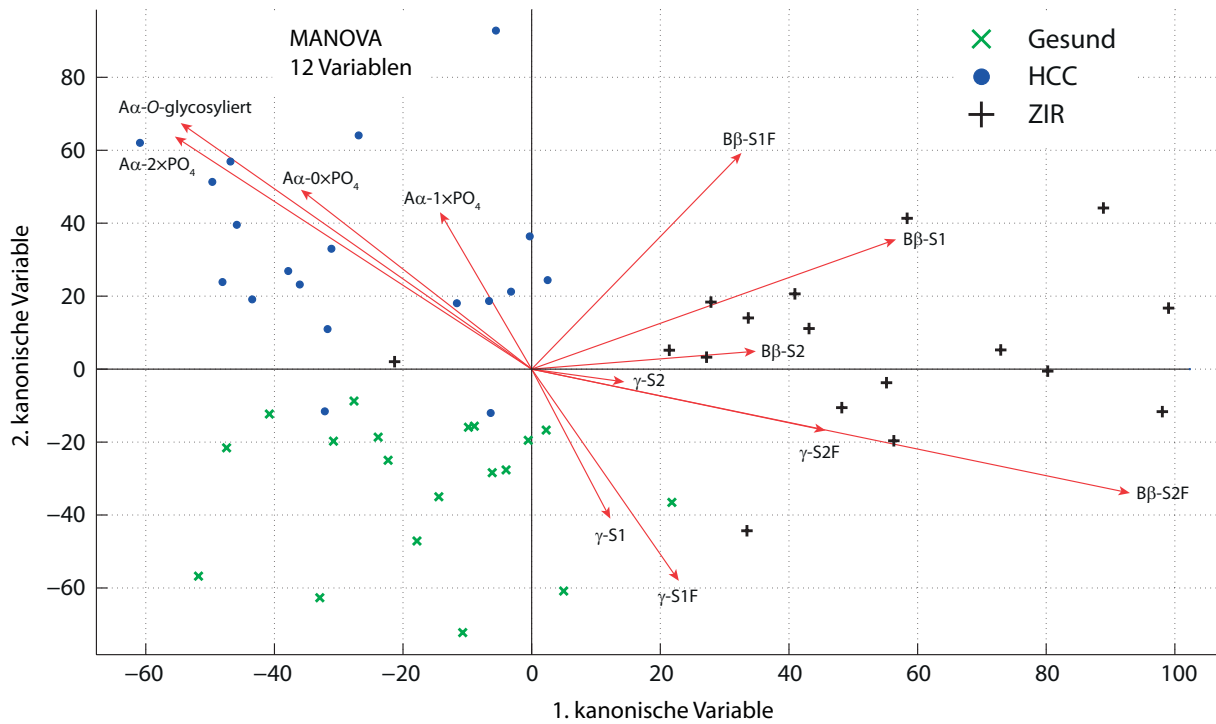
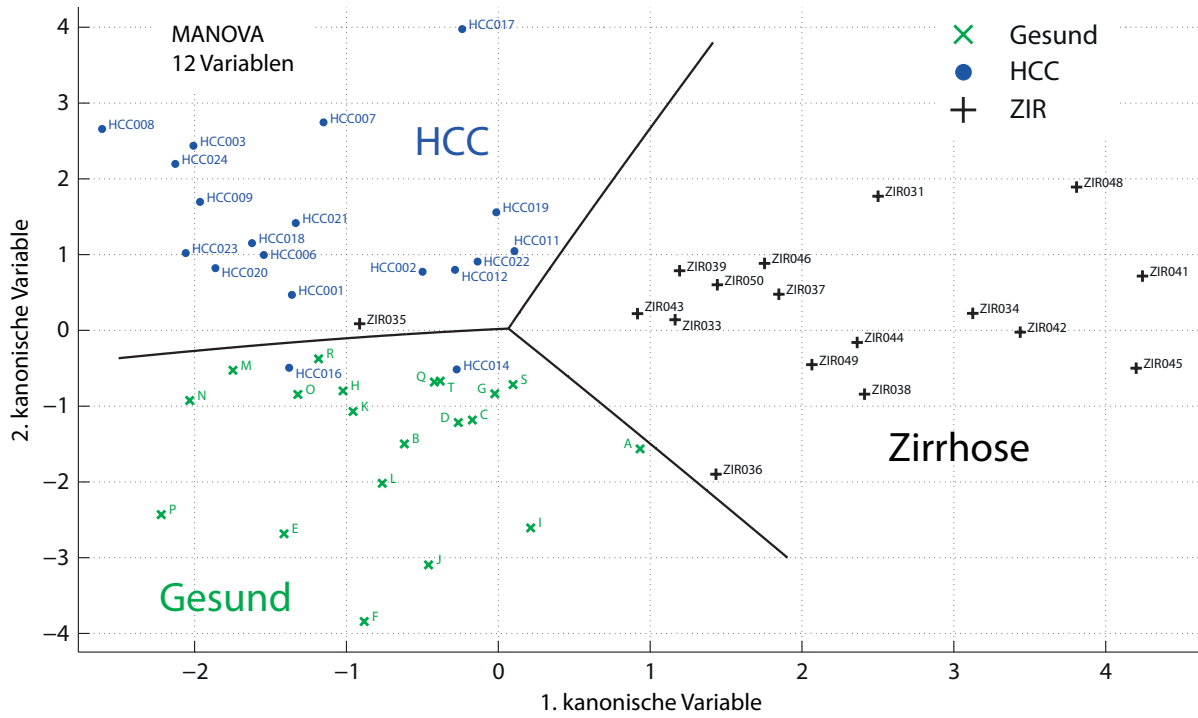


**Abbildung 75:** Die durch Variation des *shift* erhaltenen jeweils 125 Datenpunkte der Proben T, HCC019 und ZIR044 im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA mit 18 Variablen.

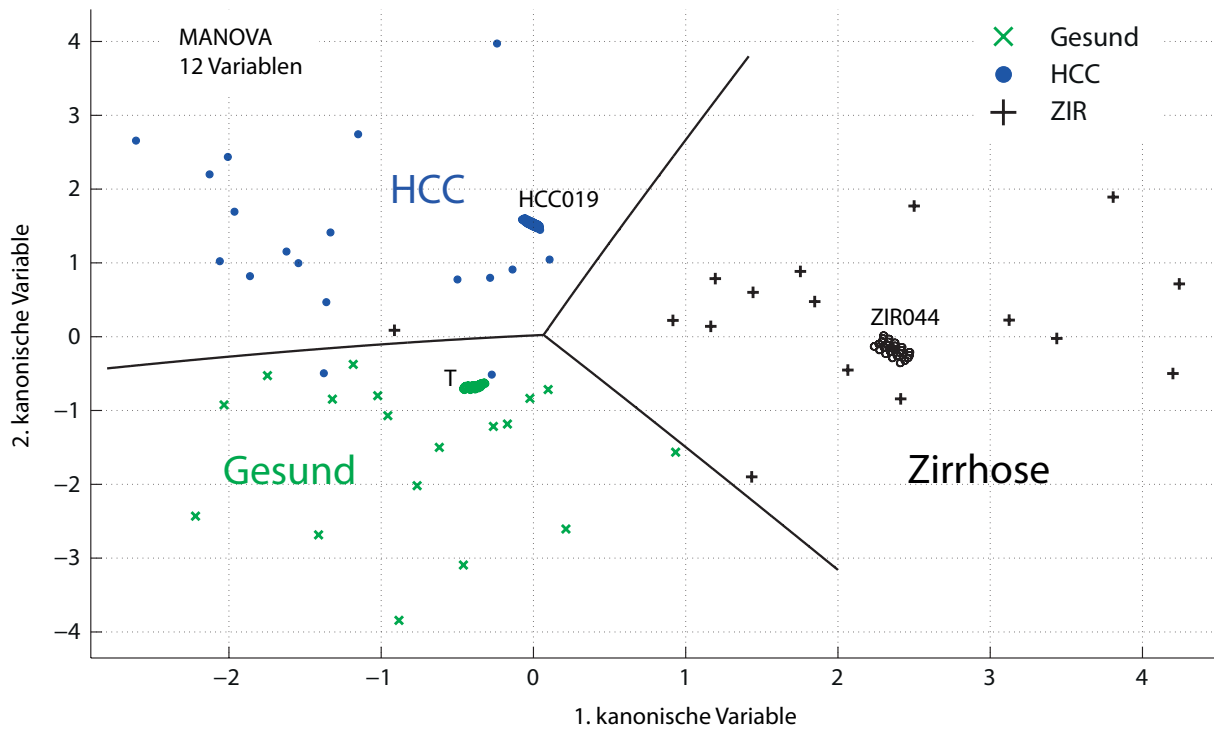
**Tabelle 30:** Ergebnisse der MANOVA mit 12 Variablen. Die Pillai-Spur Wilks-Lambda, die Hotelling-Spur und Roys größte charakteristische Wurzel sind verschiedene Arten die Signifikanz der Trennung zwischen den Gruppen zu bestimmen. In diesem Fall sind alle Werte klar signifikant (p-Wert < 0.001).

	Wert	F	Hypothesen df	Fehler df	p-Wert	Effektgröße
Pillai-Spur	1.289	6.491	24	86	<0.001	0.644
Wilks-Lambda	0.125	6.419	24	84	<0.001	0.647
Hotelling-Spur	3.714	6.344	24	82	<0.001	0.650
Größte charakteristische Wurzel nach Roy	2.217	7.945	12	12	<0.001	0.689





**Abbildung 76: Ergebnisse der MANOVA mit zwölf Variablen. Oben: Auftragung der ersten und zweiten kanonischen Variablen jeder der 56 Proben. Unten: Darstellung der Koeffizienten der ersten und zweiten kanonischen Variablen jeder Variable als Vektor.**



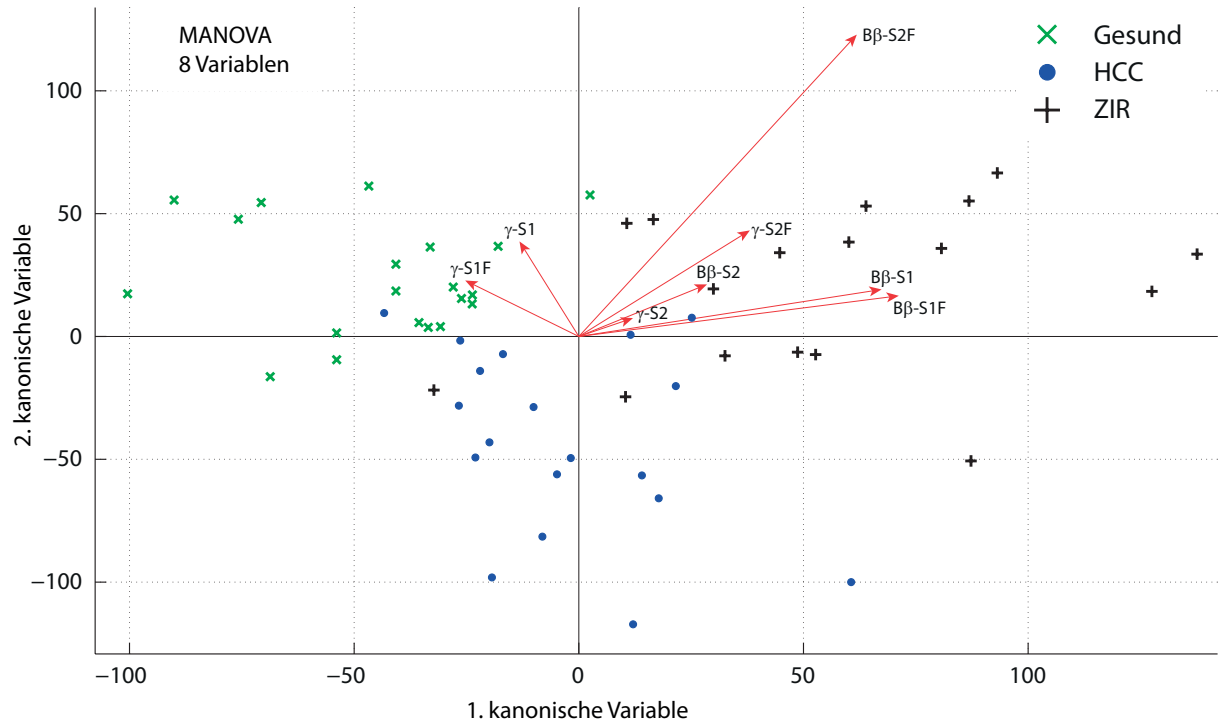
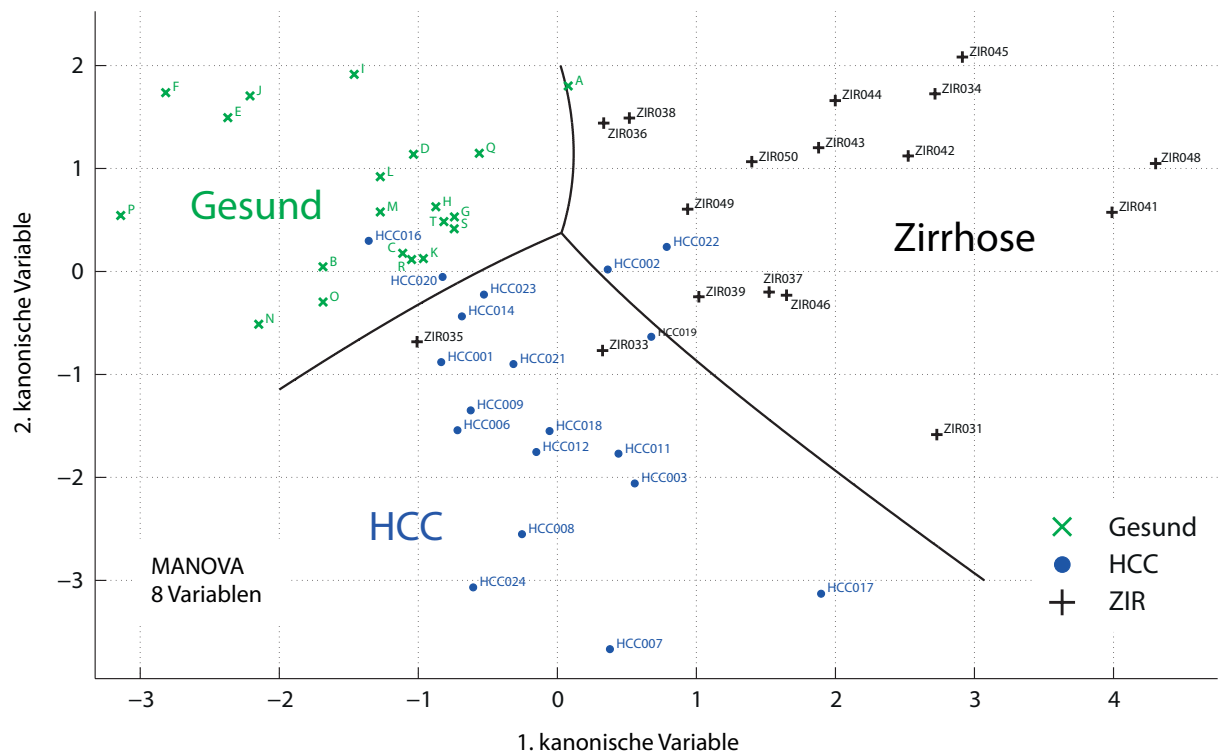
**Abbildung 77:** Die durch Variation des *shift* erhaltenen jeweils 125 Datenpunkte der Proben T, HCC019 und ZIR044 im Diagramm der ersten und zweiten kanonischen Variablen der MANOVA mit zwölf Variablen.

Die Trennung zwischen den Datenpunkten von gesunden Spendern, HCC- und Zirrhosepatienten in der MANOVA mit zwölf Variablen ist sehr gut. Alle Datenpunkte gesunder Spender liegen im korrekten Bereich. Nur drei Datenpunkte von HCC- oder Zirrhosepatienten sind falsch klassifiziert (HCC014, HCC016 und ZIR035).

Die Reduzierung der Anzahl der verwendeten Variablen hat einen weiteren Grund. Da die MANOVA im Gegensatz zur Hauptkomponentenanalyse die zu trennenden Gruppen berücksichtigt, birgt sie die Gefahr von Überanpassung (engl.: *Overfitting*). Überanpassung entsteht, wenn die Anzahl der verwendeten Variablen im Verhältnis zur Anzahl der Beobachtungen zu hoch ist. Daher wurde, um Überanpassung zu vermeiden, die Anzahl der für die MANOVA verwendeten Variablen weiter verringert.

Hierzu wurden zunächst die vier  $\alpha$ -Spezies als Variablen weggelassen und eine MANOVA mit acht Variablen ( $B\beta$ -S1,  $B\beta$ -S1F,  $B\beta$ -S2,  $B\beta$ -S2F,  $\gamma$ -S1,  $\gamma$ -S1F,  $\gamma$ -S2 und  $\gamma$ -S2F) durchgeführt. Die Ergebnisse der MANOVA mit acht Variablen sind in Tabelle 31 und Abbildung 78 gezeigt. Die Trennung zwischen gesunden Spendern, HCC- und Zirrhosepatienten ist immer noch sehr gut, es sind allerdings zwei HCC-Patienten und zwei Zirrhosepatienten falsch klassifiziert. Trotzdem wurde die Anzahl der verwendeten Variablen auf vier verringert. Hierzu erschien es sinnvoll Variablen wegzulassen deren Informationsgehalt redundant ist.

Die Korrelationskoeffizienten zwischen den zwölf wichtigsten Glycoformen und Proteinspezies sind in Tabelle 32 gezeigt.



**Abbildung 78: Ergebnisse der MANOVA mit acht Variablen. Oben: Auftragung der ersten und zweiten kanonischen Variablen jeder der 56 Proben. Unten: Darstellung der Koeffizienten der ersten und zweiten kanonischen Variablen jeder Variable als Vektor.**

**Tabelle 31: Ergebnisse der MANOVA mit acht Variablen. Die Pillai-Spur Wilks-Lambda, die Hotelling-Spur und Roys größte charakteristische Wurzel sind verschiedene Arten die Signifikanz der Trennung zwischen den Gruppen zu bestimmen. In diesem Fall sind alle Werte klar signifikant (p-Wert < 0.001).**

	Wert	F	Hypothesen df	Fehler df	p-Wert	Effektgröße
Pillai-Spur	1.117	7.438	16	94	<0.001	0.559
Wilks-Lambda	0.189	7.464	16	92	<0.001	0.565
Hotelling-Spur	2.661	7.485	16	90	<0.001	0.571
Größte charakteristische Wurzel nach Roy	1.719	10.099	8	47	<0.001	0.632

So sind zum Beispiel die Variablen  $B\beta$ -S1 und  $B\beta$ -S2 sowie  $\gamma$ -S2 und  $\gamma$ -S2 antikorreliert ( $\rho = -0.82$  bzw.  $-0.87$ ), die Variablen  $B\beta$ -S1F und  $B\beta$ -S2F sowie  $\gamma$ -S1F und  $\gamma$ -S2F sind hingegen positiv korreliert ( $\rho = 0.62/0.80$ ). Deshalb wurden wie Variablen  $B\beta$ -S2,  $B\beta$ -S1F,  $\gamma$ -S2 und  $\gamma$ -S2F weggelassen, sodass die MANOVA mit den vier Variablen  $B\beta$ -S1,  $B\beta$ -S1F,  $\gamma$ -S1 und  $\gamma$ -S1F durchgeführt wurde. Die hohen Korrelationskoeffizienten zwischen den Variablen  $B\beta$ -S1 und  $\gamma$ -S1 sowie  $B\beta$ -S1F und  $\gamma$ -S1F ( $r = 0.93$  bzw.  $0.95$ ) mussten akzeptiert werden. Andere Kombinationen von jeweils vier der acht Variablen der  $B\beta$ - und  $\gamma$ -Glycoformen erzielten ähnliche Ergebnisse.

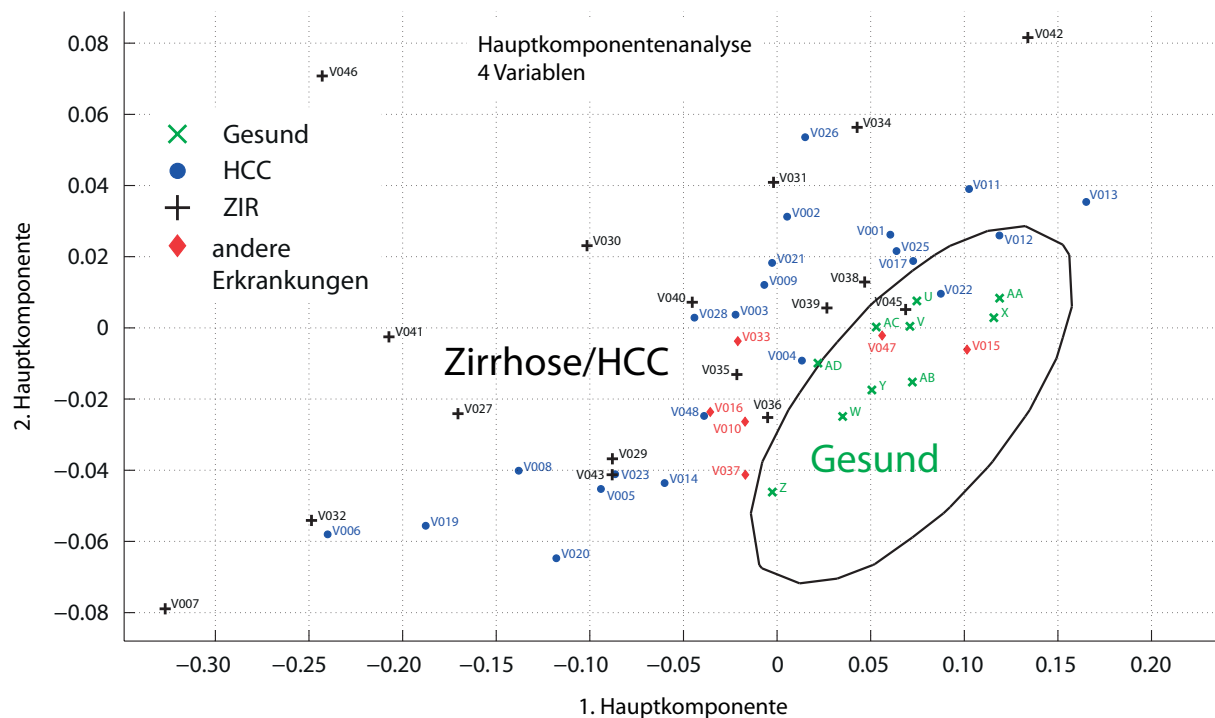
In Abschnitt 4.1.5 wurden alle multivariaten statistischen Methoden stets mit den vier Variablen  $B\beta$ -S1,  $B\beta$ -S1F,  $\gamma$ -S1 und  $\gamma$ -S1F durchgeführt. Das Verhältnis von Beobachtungen zu abhängigen Variablen beträgt somit 14 zu 1. Die Gefahr von Überanpassung der Daten ist somit weitestgehend eliminiert. Weitere Aspekte der statistischen Auswertung werden in Abschnitt 4.2.3 diskutiert.

**Tabelle 32: Matrix der Korrelationskoeffizienten zwischen den wichtigsten zwölf Glycoformen und Proteinspezies der  $\text{A}\alpha$ -,  $\text{B}\beta$ - und  $\gamma$ -Untereinheit von humanem Fibrinogen.**

	$\text{A}\alpha\text{-0}\times\text{PO}_4$	$\text{A}\alpha\text{-1}\times\text{PO}_4$	$\text{A}\alpha\text{-2}\times\text{PO}_4$	$\text{A}\alpha\text{-O-glycosyliert}$	$\text{B}\beta\text{-S1}$	$\text{B}\beta\text{-S1F}$	$\text{B}\beta\text{-S2}$	$\text{B}\beta\text{-S2F}$	$\gamma\text{-S1}$	$\gamma\text{-S1F}$	$\gamma\text{-S2}$	$\gamma\text{-S2F}$
$\text{A}\alpha\text{-0}\times\text{PO}_4$	1.00	-0.48	-0.94	0.13	-0.39	0.11	0.48	0.03	-0.36	0.06	0.51	0.12
$\text{A}\alpha\text{-1}\times\text{PO}_4$	-0.48	1.00	0.53	-0.75	-0.09	0.52	-0.31	0.42	-0.21	0.50	-0.10	0.38
$\text{A}\alpha\text{-2}\times\text{PO}_4$	-0.94	0.53	1.00	-0.37	0.34	-0.03	-0.47	0.01	0.31	-0.01	-0.48	-0.13
$\text{A}\alpha\text{-O-glycosyliert}$	0.13	-0.75	-0.37	1.00	0.12	-0.56	0.27	-0.35	0.22	-0.45	0.07	-0.25
$\text{B}\beta\text{-S1}$	-0.39	-0.09	0.34	0.12	1.00	-0.69	-0.82	-0.70	0.93	-0.72	-0.89	-0.70
$\text{B}\beta\text{-S1F}$	0.11	0.52	-0.03	-0.56	-0.69	1.00	0.19	0.62	-0.76	0.95	0.44	0.69
$\text{B}\beta\text{-S2}$	0.48	-0.31	-0.47	0.27	-0.82	0.19	1.00	0.41	-0.68	0.26	0.90	0.39
$\text{B}\beta\text{-S2F}$	0.03	0.42	0.01	-0.35	-0.70	0.62	0.41	1.00	-0.70	0.68	0.50	0.84
$\gamma\text{-S1}$	-0.36	-0.21	0.31	0.22	0.93	-0.76	-0.68	-0.70	1.00	-0.79	-0.87	-0.78
$\gamma\text{-S1F}$	0.06	0.50	-0.01	-0.45	-0.72	0.95	0.26	0.68	-0.79	1.00	0.44	0.80
$\gamma\text{-S2}$	0.51	-0.10	-0.48	0.07	-0.89	0.44	0.90	0.50	-0.87	0.44	1.00	0.54
$\gamma\text{-S2F}$	0.12	0.38	-0.13	-0.25	-0.70	0.69	0.39	0.84	-0.78	0.80	0.54	1.00

## 9.5.2 Einordnung des Validierungsdatensatzes in das Koordinatensystem der Hauptkomponentenanalyse

In Abbildung 79 sind die Ergebnisse des Validierungsdatensatzes bei Einordnung in das Koordinatensystem der Hauptkomponentenanalyse mit vier Variablen gezeigt. Die Trennung zwischen den Datenpunkten gesunder Spender und HCC- bzw. Zirrhosepatienten ist sehr gut. Neun der zehn Datenpunkte gesunder Spender liegen im korrekten Bereich. Einzig der Datenpunkt von Probe AD liegt knapp außerhalb des Bereiches gesunder Spender. Drei Datenpunkte von Proben von HCC- oder Zirrhosepatienten (V012, V022 und V045) liegen im Bereich der gesunden Spender. Die Datenpunkte der Proben V015 (Leberadenom, < 1 cm) und V047 (Hashimoto Thyreoiditis) liegen ebenfalls im Bereich gesunder Spender und sind somit korrekt klassifiziert. Die Datenpunkte V010 (Leberangiomyolipom), V016 (Leberadenom) V033 (chronische Pankreatitis) und V037 (koronare Herzerkrankung) liegen im Bereich der HCC- bzw. Zirrhosepatienten. Eine Trennung zwischen den Datenpunkten von HCC- und Zirrhosepatienten war nicht zu erkennen.

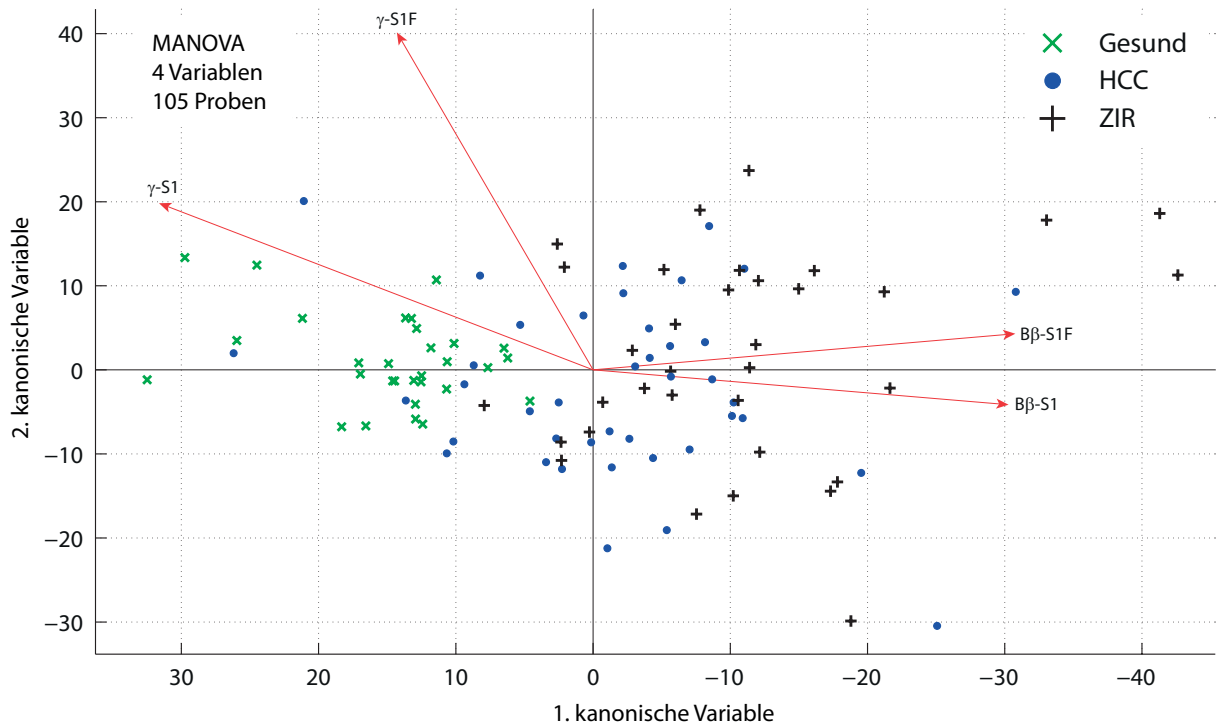
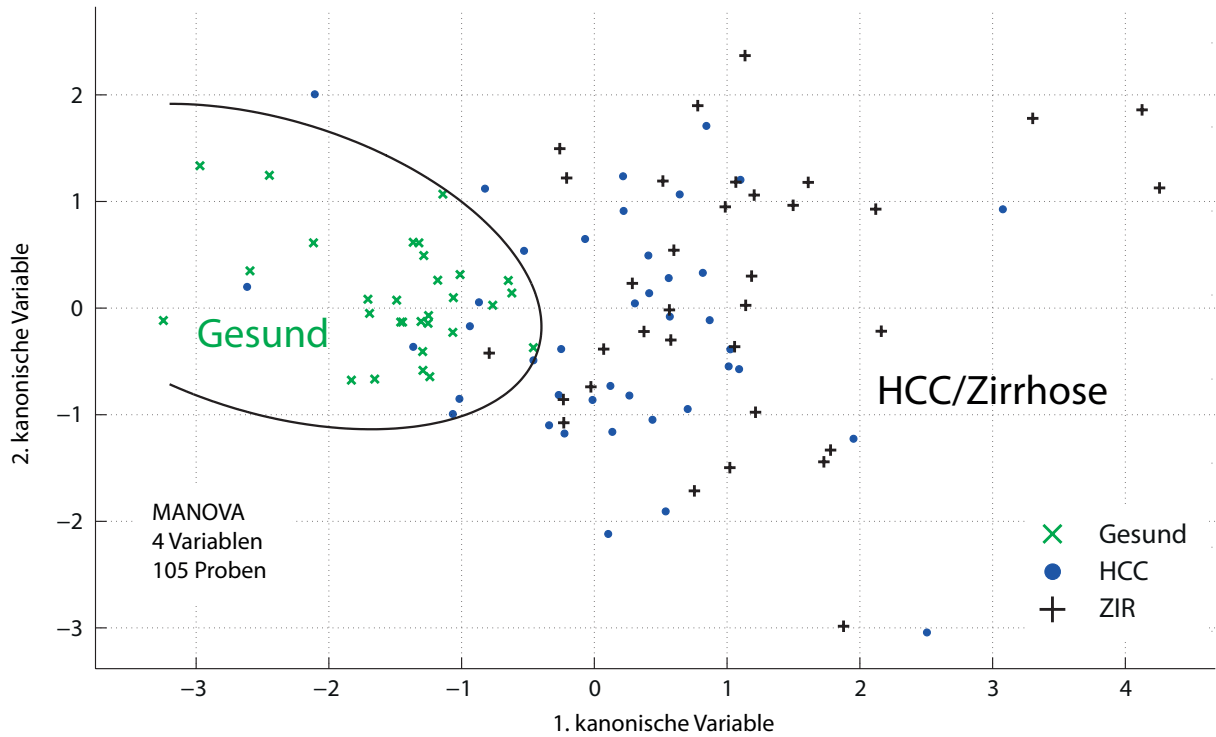


**Abbildung 79: Einordnung der Proben des Validierungsdatensatzes in das Koordinatensystem der Hauptkomponentenanalyse mit vier Variablen. Es sind die ersten beiden Hauptkomponenten gegeneinander aufgetragen. Grüne Kreuze: gesunde Spender, blaue Punkte: HCC-Patienten, schwarze Kreuze: Zirrhosepatienten, rote Rauten: Patienten mit anderen Erkrankungen (Tabelle 16). Es ist eine gute Trennung zwischen den Datenpunkten von Proben gesunder Spender und den Datenpunkten von Proben von HCC- bzw. Zirrhosepatienten zu erkennen.**

### 9.5.3 Verfeinerung der Methode durch die Proben der Validierungskohorte

In Abschnitt 4.1.6 wurde die Einordnung der Proben der Validierungskohorte in das Koordinatensystem der MANOVA mit vier Variablen diskutiert. Die Ergebnisse zeigen, dass die Unterscheidung von HCC- und Zirrhosepatienten nicht möglich ist. Die allgemeine Unterscheidung zwischen Proben gesunder und kranker Spender ist aber sehr erfolgversprechend.

Um die Stabilität des MANOVA-Modells zu überprüfen wurde die MANOVA auf Basis aller 105 Proben des Trainings- und Validierungsdatensatzes durchgeführt. Die Ergebnisse der MANOVA mit 105 Proben und vier Variablen sind in Abbildung 80 dargestellt. Die Trennung der Datenpunkte gesunder Spender zu den Datenpunkten von HCC- und Zirrhosepatienten ist sehr gut. Die Trennung zwischen den Datenpunkten von HCC- und Zirrhosepatienten ist allerdings schlecht. Die Vektoren, welche die Koeffizienten der vier Variablen in der ersten und zweiten kanonischen Variablen darstellen, sind in Abbildung 80 gezeigt. Die Vektoren der MANOVA mit 105 Proben unterscheiden sich nur geringfügig von denen der MANOVA mit 56 Proben (Abbildung 40). Das MANOVA-Modell scheint demnach stabil zu sein, sodass davon ausgegangen werden kann, dass es sich um das optimale Modell zur Einordnung unbekannter Proben handelt.



**Abbildung 80: Ergebnisse der MANOVA mit vier Variablen und allen 105 Proben des Trainings- und Validierungsdatensatzes. Oben: Auftragung der ersten und zweiten kanonischen Variablen jeder der 104 Proben. Unten: Darstellung der Koeffizienten der ersten und zweiten kanonischen Variablen jeder Variable als Vektor.**



## 10 Danksagungen

Ich danke meinen Eltern für ihre Unterstützung, ihr Vertrauen und ihre Liebe, die sie mir stets entgegengebracht haben und auf die Ich mich insbesondere während des Studiums verlassen konnte.

Ich danke allen meinen Freunden für die vielen wundervollen Erlebnisse und die tolle Zeit, die wir miteinander verbracht haben.

Ich danke allen aktuellen und ehemaligen Mitgliedern des AKBM für die schöne Zeit im Arbeitskreis.

Den Analytikern Alena, Melissa, Jan-Mirco und Raffael danke ich für die gute Zusammenarbeit und die vielen fachlichen Diskussionen.

Meike, Henning und Miriam danke ich für die Einführung in den Arbeitskreis und alles was ihr mit beigebracht habt.

Florentine und René danke ich für die fleißige Mitarbeit während der Bachelorarbeit bzw. des Praktikums.

Alena und Hinnerk danke ich für alle KaKa-Pausen, Dartpartien und Kreuzworträtsel.

Melissa, Alena und Jan-Mirco danke ich für das Korrekturlesen dieser Arbeit.

Dr. Maria Riedner und der Abteilung für Massenspektrometrie danke ich für die technische Unterstützung am maXis.

Dr. Thomas Hackl und der Abteilung für NMR-Spektroskopie danke ich für die technische Unterstützung im Bereich NMR.

Dem IT-Service des Fachbereiches Chemie danke ich für die technische Unterstützung in allen IT-Fragen.

## 11 Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Die eingereichte, schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich versichere, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde.

Hamburg, den

A handwritten signature in black ink, appearing to read 'Tim Nagel', written on a light-colored background.

Tim Nagel