

### 3.5. Auswertung von Daten aus der Massenspektrometrie

Die Funktionen der drei Hauptanwendungen zur Auswertung von MS-Spektren sind in der folgenden Übersicht dargestellt.

**Tabelle 15: Funktionen innerhalb der Anwendungen für die Massenkalkulation**

	<b>GlycoProtMass</b>	<b>GlycoWorkbench</b>	<b>Peptid Tools</b>
<b>Glykane</b>			
m/z MS	+	+	-
m/z MS <sup>n</sup>	-	+	-
Strukturdarstellung	über Importieren	+	-
Bindungseigenschaften	-	+	-
Export/Import von Daten	+	+ / -	-
<b>Proteine</b>			
m/z MS	+	-	+
m/z MS <sup>n</sup>	-	-	+
<b>Glykoproteine</b>			
m/z MS	+	-	-
Notizen, Suchkriterien	+	-	-
Erzeugung von Listen, Layouts	+	-	-
<b>Daten-Archivierung</b>	+	eingeschränkt	-

#### 3.5.1. GlycoProtMass – Berechnung m/z und Erstellung von Datenbanken

Während der Einarbeitungsphase für die Daten-Analyse der Massenspektrometrie traten zwei Hauptprobleme in den Vordergrund.

1. die Berechnung der Massen von Glykanen, Peptiden, Glykopeptiden und
2. die Archivierung von Daten für spätere Analysen und Vergleiche

Da das manuelle Berechnen von Massen sehr zeitaufwendig ist und in einem Spektrum viele verschiedene Signale möglich sind, sollten bereits erbrachte Ergebnisse in einer systematischen Form abgespeichert werden. Dabei war eine Datensicherung anzustreben, in der nach verschiedenen Kriterien gefiltert werden kann, z.B. Glykoprotein-spezifisch und/oder Projekt-bezogen. Für die Effektivität der gesamten Analytik ist eine Datenbank vorteilhaft, in der jeder Mitarbeiter nachgewiesene Strukturen ablegt. Somit wären Berechnungen nicht mehrfach nötig und Glykan-Profile von bereits untersuchten Zellen/Glykoproteinen würden bereits vorliegen.

Für die Erstellung einer angepassten Anwendung wurde die Software FileMaker 8 Pro bzw. 6 Pro (FileMaker Inc.; Santa Clara, USA) verwendet (Datei-Erweiterung ist *fp7* bzw. *fp5*). Dieses universelle Programm zur Abspeicherung von Daten wie z. B. Adressen, oder CD-Sammlungen

besitzt keinen themenbezogenen Hintergrund, sondern wurde den Anforderungen angepasst. Grundlage dieser Datenbank sind Datensätze. Jeder Datensatz wird durch beliebig viele Felder mit Informationen definiert. Dabei können die Formate Text, Zahl, Formeln u.a. verwendet werden. Somit sind Eingaben von Informationen, aber auch deren Kombination zur Weiterverarbeitung möglich. Eine wichtige Eigenschaft von Datenbanken ist die erweiterte Funktion zur Suche nach Daten, da alle Felder nach verschiedensten Kriterien abgefragt werden können.

Zur Veranschaulichung soll die folgende Tabelle dienen.

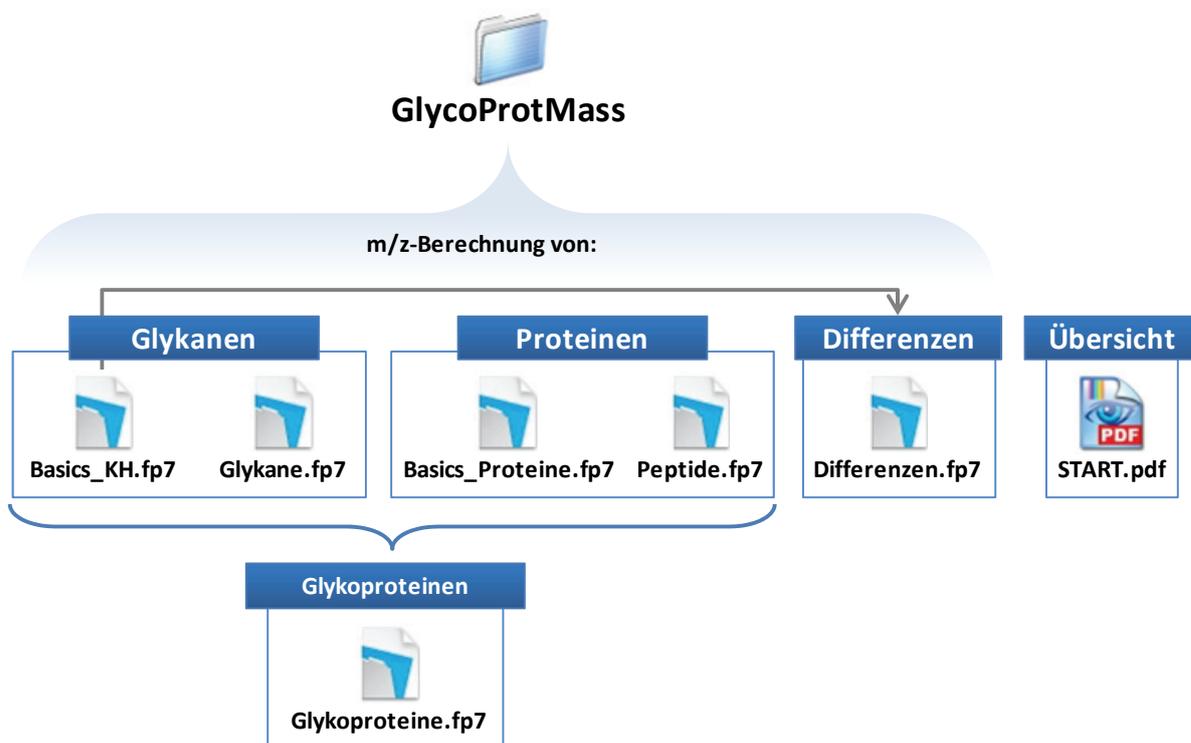
**Tabelle 16: Beispiele für den Aufbau der Datenbank**

Programm-Detail	entspricht in Anwendung	Format	Beispiel
Datensatz	Glykan	Text	Biantennär mit einer N-Acetyl-Neuraminsäure
Feld	Bezeichnung		Bi+1 Neu5Ac
	Information zur Quelle	Text	a1sGP
	Projekt	Text	0639G00
	Nutzer	Text	DaG
	Anzahl an Monosacchariden	Zahl	4 GlcNAc, 5 Hex, 1 Neu5Ac
	Polarität MS	Text	pos / neg
	Ion	Text	Na <sup>+</sup> , K <sup>+</sup> , NH <sub>4</sub> <sup>+</sup> , H <sup>+</sup> , H <sup>-</sup>
	Berechnung m/z	Formel	m/z (Glykan)
Berechnung m/z nach Markierung mit 2AB	Formel	m/z (Glykan) + m/z (2AB)	

In Abhängigkeit von den vorgegebenen Feldern der Datenbank kann für jede Struktur eine Vielzahl an Informationen gespeichert und gleichzeitig auf deren Grundlage Massen und Formeln ausgegeben werden. Jedes Feld kann als Grundlage für eine Sortierung oder Suche von Datensätzen dienen. D.h., am Beispiel der Tabelle 16 kann nach Strukturen die „Bi“ enthalten, die 1 Neu5Ac aufweisen, aus dem AGP stammen oder einer vorgegebenen Masse entsprechen, gesucht werden. Für den Aufbau der eigentlichen Anwendungen waren verschiedene Aufgaben zu erfüllen. Im ersten Schritt mussten alle Atom- und Molekülmassen für spätere Kalkulationen in die Datenbank eingegeben werden. Dann wurde die Form der Datenbank festgelegt, benötigte Felder für die abzuspeichernden Daten und Berechnungen von Massen definiert. Im dritten Schritt waren verschiedene Layouts (Format der Ausgabe) zu entwickeln, die entweder für die Dateneingabe, eine Datensatz-Übersicht oder den Ausdruck optimiert sind. Da sich ständig neue Anforderungen, Modifikationen und Erweiterungen ergaben (verschiedene Messtechniken, Derivatisierungen, usw.), wurden während des gesamten Zeitraumes der Promotionsarbeit

Änderungen vorgenommen und die Richtigkeit der Kalkulationen mit anderen Anwendungen (s. 2.12.2.) überprüft.

Ich habe mich zu Beginn dazu entschlossen, generell eine Datei mit den Massen der Grundbausteine zu erstellen und diese in der eigentlichen Datenbank-Datei über Verknüpfungen zu verwenden (Dateinamen sind im folgenden Text kursiv dargestellt).



**Abbildung 72: Dateien der aktuellen Endversion 3.0 von GlycoProtMass**

Für die Massenkalkulation der verschiedenen Strukturen sind bestimmte Dateien erforderlich, die Differenzenermittlung benötigt die *Basics\_KH*-Daten

In der Datei *START.pdf* sind die Dateien in einer Übersicht mit ihren Verknüpfungen und weiteren Informationen zur Verwendung aufgelistet.

### 3.5.1.1. GlycoProtMass 1.0

In der Version 1.0 (FileMaker 6 Pro) der Anwendung war das Ziel die Ausgabe von Glykanmassen. Für die Glykane wurden die Werte der Monosaccharide (Layout „Basics“), Ionen (Layout „Ionen“) und Modifikationen (Layout „Derivate\_KH“) in der Datei *Basics\_KH* und die Datenbank für die Strukturen in der Datei *Glykane* abgelegt. Die weiteren aufgeführten Bestandteile von GlycoProtMass (Abbildung 72) wurden erst später hinzugefügt. Eine erhöhte

Übersichtlichkeit über die Daten-Felder wird durch verschiedene Layouts ermöglicht, welche über das entsprechende Menü angewählt werden können.



The screenshot shows the 'Basics\_KH' software window. On the left, there is a sidebar with a 'Layout' dropdown menu, which is highlighted by a red arrow. The main area displays a table titled 'Massen der Kohlenhydrate' (Masses of Carbohydrates). The table has columns for 'frei', 'im Oligosaccharid', 'permethyliert', 'Anzahl ~CH2', 'Anzahl ~C2H2O1', and 'Formel'. The rows list different carbohydrate units: Hex, Desoxyhex, Pent, HexN, and HexNAc, each with its mass and corresponding chemical formula.

	frei		im Oligosaccharid	permethyliert	Anzahl		Formel
					~CH2	~C2H2O1	
Hex	180,0634	m_Hex	162,0528	204,0998	3	3	C6H12O6
Desoxyhex	164,0685	m_desHex	146,0579	174,0892	2	2	C6H12O5
Pent	150,0528	m_Pent	132,0423	160,0736	2	2	C5H10O5
HexN	179,0794	m_HexN	161,0688	217,1314	4	5	C6H13NO5
HexNAc	221,0899	m_HexNAc	203,0794	245,1263	3	2	C8H15NO6

**Abbildung 73: Ansicht des Layout „Basics“ für die Glykan-Berechnung in der Datei *Basics\_KH***  
Mit dem Pfeil ist das Auswahlmennü für die verschiedenen Layouts gekennzeichnet.

Die möglichen Modifikationen an Glykanen in vivo, als auch durch Derivatisierungen sind im Layout „Derivate\_KH“ definiert. Das Layout „Ionen“ enthält die bei der Analyse von Spektren zu berücksichtigenden Ionen.

In der Datei *Glykane* können für jedes Oligosaccharid Informationen in die vordefinierten Felder eingetragen und durch Verwendung der Daten aus *Basics\_KH* die gewünschten Massen berechnet werden. Aufgrund der Etablierung der ESI-IonTrap-MS wurden auch Mehrfachladungen bestimmt, da im Gegensatz zur MALDI-TOF-MS nur zu einem geringen Anteil einfach geladene Ionen auftreten. Die kalkulierten Werte sind monoisotopisch und mit drei Stellen nach dem Komma angegeben. Wird die Masse im Eingabefenster direkt angewählt, erscheint die exakte Zahl.

The screenshot shows the 'Glykane' software interface. The main window is titled 'Glykane' and contains several input fields and sections:

- Blättern:** Navigation icons.
- Layout:** 'Eingabe' selected.
- Datensatz:** '1' selected.
- Summe:** '1'.
- Unsortiert:** Status indicator.
- Pol:** Radio buttons for 'pos' (selected) and 'neg'.
- Thema:** 'N-Glykane'.
- Projekt:** '0602000'.
- Beschreibung:** Empty text box.
- Source:** Empty text box.
- Darstellung:** Empty text box.
- Name\_KH:** 'Bi+1NeuAc'.
- redEnd:** '1'.
- HexNAc:** '4'.
- Hex:** '5'.
- desHex:** Empty.
- Pent:** Empty.
- NeuAc:** '1'.
- NeuGc:** Empty.
- SO3:** Empty.
- HPO3:** Empty.
- CH3:** Empty.
- COCH3:** Empty.
- dPent:** Empty.
- Hep:** Empty.
- HexA:** Empty.
- KDH:** Empty.
- KDC:** Empty.
- HexH:** Empty.
- Mur:** Empty.
- Derivat:** 'keine'.
- Ion:** 'Na+'.
- Ion\_Zahl:** '1'.
- KH-Struktur:** 'HN4H5NA1'.
- Formel\_KH:** 'C73H121N5O54'.
- nH:** 0.
- nNa:** 1.
- nHH4:** 0.
- nK:** 0.
- Masse Basic:** 1931,688.
- Masse:** 1954,677 (highlighted in yellow).
- Masse red:** 1956,693 (highlighted in yellow).
- Masse DMB:** 2090,741 (highlighted in blue).
- Summary Table:**

	2x	3x	4x	5x	6x	7x	8x	
Masse	977,339	651,559	488,669	390,935	325,780	279,240	244,335	
Masse 2AB	1.037,373	691,582	518,687	414,949	345,791	296,392	259,343	2074,746
Masse frei	977,339	651,559	488,669	390,935	325,780	279,240	244,335	1954,677

Abbildung 74: Ansicht des Layouts "Eingabe" der Datei *Glykane*, der Datenbank für Glykan-Strukturen

Die Felder zur Dateneingabe (Abbildung 74) werden in der folgenden Tabelle erläutert.

Tabelle 17: Felder der Dateneingabe in der Datenbank *Glykane*

Daten-	Feld	Bemerkung
Eingabe	Thema, Projekt, Beschreibung usw.	Proben-Definition
	Pol	Polarität am Massenspektrometer
	Name KH	Bezeichnung des Glykans
	N/O	Definition der Bindung des Glykans
	HexNAc, Hex, desHex, Pent usw.	Anzahl der Monosaccharide im Glykan
	SO3, HPO3, CH3, COCH3	Anzahl an Phosphorylierungen, Sulfatierungen, Methylierungen, Acetylierungen
	Derivat	Derivatisierungen von Glykanen
	Ion, Ion_Zahl	Typ und Anzahl des Ions
Ausgabe	s/as	mit / ohne Sialylierung
	KH-Struktur	Kurzformel der Monosaccharid-Zusammensetzung
	Formel-KH	Summenformel aus KH-Struktur
	Glykan bound	m/z des Glykans nach Bindung an Peptid
	Masse Basic	m/z des Glykans
	Masse	m/z des Glykans mit Ionen, ohne Derivatisierungen
	Masse DMB	m/z des Glykans nach DMB-Derivatisierung
	Masse 2AB	m/z des Glykans nach 2AB-Derivatisierung
	Masse frei	m/z des Glykans mit Ionen und Derivatisierungen
1x bis 8x	m/z mit Mehrfachladung unter Einberechnung der jeweiligen Ionen	

### 3.5.1.2. GlycoProtMass 2.0

Da während der Promotionsarbeit Peptid- und- Glykopeptid-Proben charakterisiert wurden, musste die Anwendung in ihren Funktionen erweitert werden. In der Version 2.0 wurde eine Datei (*Basics\_Proteine*) analog zu den *Basics\_Glykane* mit verschiedenen Layouts zu Derivatisierungen und Aminosäuren erstellt. GlycoProtMass wurde gleichzeitig in die neuere FileMaker 8 Pro-Version konvertiert, welche nun standardmäßig Verwendung fand.

Die Glykane-Datenbank wurde mit neuen Feldern für die Peptid-Berechnung (Biemann & James, 1990) erweitert, so dass mit der Verknüpfung auf *Basics-Proteine* die Massen von Glykanen, Peptiden und Glykopeptiden kalkulierbar waren (Abbildung 75).

The screenshot displays the 'Glyk\_Prot' application window. The 'Eingabe' layout is active, showing a form for entering glycan and protein data. Key fields include 'Kurzname' (Ei+1NeuAc mit Peptid), 'redEnd' (1), 'HexNAc' (4), 'Hex' (5), 'Pent' (1), 'NeuAc' (1), 'NeuGc', 'SO3', 'HPO3', 'CH3', 'COCH3', 'n Glykane', 'dPent', 'Hep', 'HexA', 'KDN', 'DC', 'HexN', 'Mur', 'Ion' (Na+), 'Ion\_Zahl' (1), 'Masse' (1954,677), 'Masse red' (1956,693), 'Masse 2AB' (2074,746), 'Masse DMB' (2090,741), 'Masse frei' (1954,677), 'Masse' (1954,677), 'Glykan bound' (1913,677), 'ProteinMass' (18,011), 'ProteinMass Seq' (1,032,539), 'Sequenz' (MADGKRDIK), 'Peptid' (1,032,539), '[Peptid] Mod' (1,055,528), and '[Peptid]' (1,055,528). A table of glycopeptide masses is also visible, ranging from 1,484,603 to 371,151.

Abbildung 75: Ansicht des Layouts "Eingabe" der Datenbank Version 2.0

Die in der Version 1.0 und 2.0 erstellten Formeln erwiesen sich bei der ständigen Anwendung von GlycoProtMass teilweise als zu unflexibel, da die Massen direkt mit den m/z-Werten für Monosaccharide, Derivate, Ionen berechnet wurden. Für einen geringeren Rechenaufwand der Anwendung, welcher bei einer höheren Anzahl an Datensätzen wichtig ist, und ein einfacheres nachträgliches Hinzufügen von Modifikationen müssen die Formeln mit Variablen und FileMaker-Funktionen wie

*Falls(Bedingung1; Ergebnis1 {; {Bedingung2; Ergebnis2;...} Standardergebnis})  
oder  
Wenn(Bedingung; ErgebnisWennWahr; ErgebnisWennFalsch)*

aufgebaut werden. Dies wird am folgenden Beispiel erläutert. Für alle wichtigen Modifikationen des Glykans soll die Masse berechnet werden. In den Versionen 1.0 und 2.0 wurde für jede ein eigenes Feld mit der Berechnungsformel erstellt. Für ein 2AB-markiertes Glykan wäre die Formel in vereinfachter Form folgende:

$$m/z (\text{Glykan}) = \text{Summe } m/z (\text{Monosaccharide}) - m/z \text{ H}_2\text{O} + \text{Ion} + m/z \text{ 2AB}$$

Für 10 Modifikationen sind somit auch 10 Felder mit der entsprechenden Formel zu erstellen. Im Falle einer neuen Modifikation müsste nachträglich ein weiteres hinzugefügt werden. Die effektivere Lösung ist:

$$m/z (\text{Glykan}) = \text{Summe } m/z (\text{Monosaccharide}) - m/z \text{ H}_2\text{O} + \text{Ion} + m/z \text{ X}$$

Der Wert X ist nicht konstant, sondern richtet sich nach der Datenauswahl des Anwenders. Nach Auswahl der Modifikation 2AB im entsprechenden Menü in der Datei *Glykane*, wird X mit der Masse von 2AB gleichgesetzt und die Masse des 2AB-markierten Glykans ist sichtbar. D.h. es existiert nur eine Formel für alle Derivatisierungen, wodurch die Berechnungen innerhalb der Datenbank übersichtlicher bleiben und die Rechenleistung geringer ist (bei 10 parallelen Formeln wird jede ständig angewendet). Für eine neue Modifikation muss diese nur in der Datei *Basics\_KH* abgespeichert werden und ist dann im Auswahlmenü von *Glykane* sichtbar. Eine Veränderung oder Neuerstellung von Feldern und Formeln ist nicht notwendig.

Ein weiterer Punkt für Verbesserungen war die Vereinfachung der Massenberechnungen durch die Verkürzung von Formeln. Zu Beginn der Arbeiten mit FileMaker wurden die Kalkulationen ständig erweitert und die älteren Formeln modifiziert. Hierfür folgt ein Beispiel.

Von einem Glykan sollen die Massen für die Summe der Monosaccharide (M1), Summe der Monosaccharide + Ionen (M2) und Summe der Monosaccharide + Ionen + Modifikationen (M3) berechnet werden. Für das Vermeiden von Fehlern wurde immer die Ausgangs-Formel kopiert, für die folgende eingefügt und mit den Erweiterungen ergänzt. Die Kalkulation der drei Werte erfolgte somit parallel und unter Verwendung aller Strukturkomponenten. Da FileMaker die Ergebnisse der Formeln nicht abspeichert, sondern bei Aufruf der Datensätze immer neu berechnet, ist bei einer Vielzahl von Formeln eine bestimmte Rechenzeit notwendig. Diese steigt mit zunehmender Anzahl an Datensätzen. Diesem kann mit seriellem Arbeiten von Kalkulationen entgegengewirkt werden. D.h. die bereits berechnete Masse  $M1 + \text{Masse Ionen} = M2$  und  $M2 + \text{Masse Modifikation} = M3$ .

Ein dritter Punkt für Verbesserungen von GlycoProtMass war die Glykopeptid-Berechnung. In der Version 2.0 mussten für einen weiteren Datensatz sowohl die Glykan-Struktur, als auch die

Aminosäure-Sequenz neu eingegeben werden. Wenn einer dieser Sachverhalte ständig konstant bleibt, ist die Dateneingabe sehr zeitaufwändig. Hierfür waren weitere Änderungen vorzunehmen, die aber eine Aufteilung der Felder auf mehrere FileMaker-Dateien zur Folge hatte.

### 3.5.1.3. GlycoProtMass 3.0

In der Version 3.0 kam es zu einer Vielzahl an Optimierungen und Neuerungen. Die Formeln wurden neu geschrieben und die Richtigkeit musste ständig überprüft werden. Außerdem stand eine Anwender-freundlichere Berechnung von Glykopeptiden im Vordergrund.

Der Aufbau der Anwendung GlycoProtMass 3.0 wurde bereits in der Abbildung 72 dargestellt. Die verschiedenen Dateien sind intern in ihren Formeln verknüpft und dürfen deshalb nicht umbenannt werden. Die notwendigen Dateien für die jeweiligen Massenbestimmungen können ebenfalls der Abbildung entnommen werden. Für die Berechnung von Glykanen und Peptiden existieren jeweils zwei Dateien, eine für die Massen der Komponenten (*Basics\_KH*, *Basics\_Proteine*) und eine für die zu sichernden Strukturen (*Glykane* bzw. *Peptide*). Für Glykopeptide ist die Datenbank-Datei ebenfalls vorhanden (*Glykoproteine*), jedoch werden die Massen für den Glykan- bzw. Peptid-Anteil direkt aus den jeweiligen Datenbanken (*Glykane* bzw. *Peptide*) übernommen und nicht neu eingegeben. Das Eingabefenster der Datenbank *Glykane* ist in der Abbildung 76 dargestellt.

Es sind nur die Änderungen für die Eingabe ersichtlich, Modifikationen in den Formeln werden hier nicht detailliert angegeben. Zusätzlich Felder im Vergleich zur Tabelle 17 (Version 1.0) sind der folgenden Auflistung zu entnehmen.

**Tabelle 18: Felder der Dateneingabe in der Datenbank *Glykane***

Daten-	Feld	Bemerkung
Eingabe	User, usw.	Proben-Definition
	Mod 1, Mod 2, Mod 3 und Anzahl	3 auswählbare Glykan-Modifikationen
	Vorgabe	Vorgabe einer Massendifferenz (auch als Alternative zu Mod 1-3)
	X Ladung	Auswahl einer beliebigen Ladung
	SO <sub>3</sub> , HPO <sub>3</sub> , CH <sub>3</sub> , COCH <sub>3</sub>	Anzahl an Phosphorylierungen, Sulfatierungen, Methylierungen, Acetylierungen
	Derivat	Derivatisierungen von Glykanen
	Ion, Ion_Zahl	Typ und Anzahl des Ions
Ausgabe	X Masse	m/z zu X Ladung
	Mod Summe	m/z der Modifikation(en)
	Glykan bound	m/z Glykan nach Bindung an Peptid

**New** **Copy** **Accept**

Pol  pos  neg

User  Zelllinie  Protein

Thema  Projekt

DaG

Beschreibung

**Hermentin**

Referenz

Name\_KH

redEnd  HexNAc  Hex  desHex  Pent  NeuAc  NeuGc  SO3  HPO3  CH3  COCH3

dPent  Hep  HexA  KDN  KDO  HexN  Mur

Mod 1

Mod 2

Mod 3

Ion  Ion\_Zahl

nH  nNa  nNH4  nK

KH-Struktur

KH Formel

**Glykan bound ohne Ionen** **1.913,677**

X Ladung	10	X Masse	195,468	red	DMB	2090,741	1931,688	Masse Basic
2x	977,339	651,559	488,669	1956,693	8x	244,335	1.954,677	Masse
3x	1.037,373	691,582	518,687	279,240	7x	259,343	2.074,746	Masse 2AB
4x	977,339	651,559	488,669	325,780	6x	279,240	244,335	Masse frei
5x				390,935	5x	345,791		
6x				414,949	4x	488,669		
7x				518,687	3x	651,559		
8x				691,582	2x	977,339		

**Darstellung**

Abbildung 76: Ansicht des Eingabe-Layouts einer Glykan-Datenbank mit GlycoProfMass

Die Berechnung von permethylierten Oligosacchariden ist über ein eigenes Eingabe-Layout vorzunehmen. Da diese Derivatisierung erst zum Ende dieser Arbeit in GlycoProtMass berücksichtigt wurde, ist dies noch nicht vollständig geschehen. Die zusätzlichen Methyl-Gruppen werden nicht bei Modifikationen der Auswahlmenüs „Mod1,2,3“ angewendet. Aus diesem Grund fehlen im Permethylierungs-Layout die entsprechenden Felder. Derivatisierungen die zu einer Massenänderung für das gesamte Glykan führen, können einfach hinzugefügt werden. Im Falle von spezifischen Änderungen für jedes Monosaccharid sind neue Formeln notwendig, da jede der Standard-Berechnungen von den Massen der underivatisierten Kohlenhydrat-Monomere ausgeht.

Da sowohl für Glykane, als auch für Proteine eine eigene Datenbank existieren sollte, musste für Letztere die Vorlage neu erstellt werden. Der Aufbau des Eingabe-Layouts (s. Abbildung 77) ist ähnlich dem der Glykane. Die folgende Tabelle dient der Übersicht der vorhandenen Felder.

**Tabelle 19: Felder der Dateneingabe in der Datenbank *Peptide***

Daten-	Feld	Bemerkung
Eingabe	Pol	Polarität am Massenspektrometer
	Thema, Projekt, usw.	Proben-Definition
	Name	Protein-Bezeichnung
	Ion und n	Typ und Anzahl der Ionen
	Modifikation und n	Typ und Anzahl der Modifikationen
	X Ladung	Auswahl einer beliebigen Ladung
	A, R, N, D, usw.	direkte Eingabe der Anzahl an Aminosäuren
	Sequenz	direkte Eingabe der Peptid-Sequenz aus Datenbanken u.a.
	SO <sub>3</sub> , HPO <sub>3</sub>	Anzahl an Sulfatierungen bzw. Phosphorylierungen
	Protein-Vorgabe	Angabe einer zusätzlichen Massendifferenz
	X Ladung	Auswahl einer beliebigen Ladung
	aa	Auswahl einer Aminosäure zur Anzahlbestimmung in der Sequenz
	Glykosylierungen N und O	Anzahl der Glykosylierungsstellen in der Proteinsequenz
Ausgabe	Formel P	Summenformel Protein
	Modifikation Gesamt m/z	m/z Summe der Modifikationen
	AA / AS	Anzahl der Aminosäuren
	ProteinMass (Seq)	m/z Protein monoisotopisch
	ProteinMass (Seq) av	m/z Protein average
	Protein Basic	m/z Protein (nur Aminosäuren)
	Protein SPM	Protein Basic + Modifikationen
	Protein	Protein SPM + Ionen
	X Masse	m/z zu X Ladung
n aa	Anzahl der Aminosäure aa in Sequenz	



Die Datenbankdatei für die Erstellung von Glykoproteinen ist im Folgenden dargestellt.

**Tabelle 20:**Felder der Dateneingabe in der Datenbank *Glykoproteine*

Daten-	Feld	Bemerkung
Eingabe	Pol	Polarität am Massenspektrometer
	mit / ohne Glykan	Kalkulation mit/ohne Glykan
	Thema, Projekt, usw.	Proben-Definition
	Name	Glykoprotein-Bezeichnung
	Ion und n	Typ und Anzahl der Ionen
	Glykan	Auswahlmenü mit den Namen der Oligosaccharide in der Datenbankdatei <i>Glykane</i>
	n	Anzahl des gewählten Glykans
	Typ	N- oder O-Glykan
	Peptid	Auswahlmenü mit den Namen der Proteine in Datenbankdatei <i>Peptide</i>
	Mod1, Mod2, Mod3	Auswahlmenü für Modifikationen
	Protein-Vorgabe	Angabe einer zusätzlichen Massendifferenz
	X Ladung	Auswahl einer beliebigen Ladung
	aa	Auswahl einer Aminosäure zur Anzahlbestimmung in der Sequenz
Ausgabe	s / as	sialo / asialo Glykan
	m/z geb	
	m/z KH	
	Summe	m/z geb aller Glykane am Protein
	Struktur-Kurzformel	
	Summenformel	
	Glykosylierungen N und O	Anzahl der angegebenen Glykosylierungsstellen in der Sequenz in der Datenbankdatei <i>Peptide</i>
	ProteinMass (Seq)	m/z Protein monoisotopisch
	ProteinMass (Seq) av	m/z Protein average
	AS	Anzahl der Aminosäuren
	n aa	Anzahl der Aminosäure aa in Sequenz
	Formel P	Summenformel Protein
	Formel GP	Summenformel Glykoprotein
	X Masse	m/z zu X Ladung
	Modifikation Gesamt m/z	m/z Summe der Modifikationen
	Protein Basic	m/z Protein (nur Aminosäuren)
	Protein Mod	Protein Basic + Modifikationen
Protein	Protein Mod + Ionen	
GlycoProtein	m/z Glykoprotein bzw. Protein	

Die Felder „Glykane“ und „Peptide“ sind Verknüpfungen zu der jeweiligen Datenbank-Datei, d.h. beim Anwählen dieser wird diese automatisch geöffnet. Somit können zu wählende Strukturen schneller überprüft bzw. neu erstellt werden.

**New** **Copy**

**Accept**

**Pol**  pos  neg  mit  ohne

**Glykan**

Thema

Projekt

User

Trypsin-Verdau AGP

Beschreibung

Glykoprotein

Hermentin

Referenz

Ion	n	Glykan	m/z geb	m/z KH	n	Typ
1. Na	1	1. Bi+1NeuAc	1.913,677	1.931,688	1	N
2.						
3.						
4.						
5.						
6.						
7.						
8.						
<b>Glykane</b>			<b>Summe</b>			
						1.913,677

**Peptide**

Glykosylierungen

Peptid

AS

Mod1 Phosphorylierung

Mod2

Mod3

**Protein**

Formel P

Formel GP

2X

Protein Basic

Protein Mod

Protein Vorgabe

**Glycoprotein**

Abbildung 78: Ansicht des Eingabe-Layouts einer Glykoprotein-Datenbank mit GlycoProtMass

Die Kalkulation von Glykanen und Peptiden wird durch die eigenständigen Dateien ermöglicht, deren Funktionen wie Modifikationen und Ionen nachträglich erweiterbar sind. Durch das Verknüpfen der Datenbanken können Glykoproteine mit relativ geringem Zeitaufwand berechnet werden. Das Kombinieren von Proteinen mit verschiedenen Glykanen ist, vorhandene Struktureinträge in *Glykane* und *Peptide* vorausgesetzt, über eine einfache Auswahl aus den Einblendmenüs „Glykan“ und „Peptid“ möglich. D.h. nach Fertigstellung eines Glykans oder Peptids in der jeweiligen Datenbank stehen diese in den Auswahlmenüs von *Glykoproteine* zur Verfügung (rote Umrahmung mit Pfeil in Abbildung 78).

#### 3.5.1.4. Anwendungen der FileMaker-Datei *Differenzen.fp7*

Bei der Erstellung von GlycoProtMass und Auswertungen von Spektren waren weitere Berechnungen notwendig, die in der Datei *Differenzen* vorliegen. Hier sind drei verschiedene Layouts vorhanden, die jeweils eine eigene Datenbank innerhalb der Datei aufbauen.

1. Glykane Berechnungen von Massendifferenzen aufgrund von Mono- oder Oligosacchariden
2. Ionen Massenänderungen in Abhängigkeit vom Kation
3. Masse Kalkulation der Masse organischer Moleküle

In Spektren können regelmäßige Abstände oder auch zusätzliche Peaks mit definierten Werten erkennbar sein. Gründe hierfür sind z.B. unvollständige Derivatisierungen und Wiederholungen bzw. Modifikationen innerhalb einer bestimmten Struktur. Für die einfachere Zuordnung dieser Eigenschaften können im Layout „Glykane“ Derivate und Monosaccharide ohne/mit Angaben zur Ionisierung kombiniert werden. D.h. nach Abbildung 79 entspricht eine Massendifferenz von ca. 169,0 u einer Desoxy-Hexose mit einem Na-Ion.

Massen-Differenzen											06.04.2008															
(Derivat)	Komponente						Ionen / Zahl						Ladung													
	1	2	3	4	5	6							1x	2x	3x	4x	5x	6x								
2AB	<input type="text"/>	Na	<input type="text"/>	143,1	71,5	47,7	35,8	28,6	23,8																	
Pent	<input type="text"/>	Na	<input type="text"/>	155,0	77,5	51,7	38,8	31,0	25,8																	
DMB	<input type="text"/>	Na	<input type="text"/>	159,1	79,5	53,0	39,8	31,8	26,5																	
dHex	<input type="text"/>	Na	<input type="text"/>	169,0	84,5	56,3	42,3	33,8	28,2																	

Abbildung 79: Eingabe-Layout „Glykane“ in *Differenzen.fp7*

Die Unterscheidung von Kationen ist primär bei Messungen mit der ESI-IonTrap-Methode notwendig. Hier können die vier Haupt-Ionen (s. Abbildung 80) mit Mehrfachladungen auftreten, während mit dem MALDI-TOF-Gerät fast nur Na-Addukte (1-fache Ladung) sichtbar sind.

Massen-Differenzen																															
Ion	n	1x					2x					3x					4x					5x					6x				
		m/z	H	Na	K	NH4	m/z	H	Na	K	NH4	m/z	H	Na	K	NH4	m/z	H	Na	K	NH4	m/z	H	Na	K	NH4	m/z	H	Na	K	NH4
H	1	1,0		22,0	38,0	17,0	0,5		11,0	19,0	8,5	0,3		7,3	12,7	5,7	0,3		5,5	9,5	4,3	0,2		4,4	7,6	3,4	0,2		3,7	6,3	2,8
Na	1	23,0	22,0		16,0	5,0	11,5	11,0		8,0	2,5	7,7	7,3		5,3	1,7	5,7	5,5		4,0	1,2	4,6	4,4		3,2	1,0	3,8	3,7		2,7	0,8
K	1	39,0	38,0	16,0		20,9	19,5	19,0	8,0		10,5	13,0	12,7	5,3		7,0	9,7	9,5	4,0		5,2	7,8	7,6	3,2		4,2	6,5	6,3	2,7		3,5
NH4	1	18,0	17,0	5,0	20,9		9,0	8,5	2,5	10,5		6,0	5,7	1,7	7,0		4,5	4,3	1,2	5,2		3,6	3,4	1,0	4,2		3,0	2,8	0,8		3,5

Abbildung 80: Massendifferenzen zwischen Kationen im Layout „Ionen“ in *Differenzen.fp7*

Für das Auswerten von Glykan-Profilen, die im positiven Modus gemessen werden, sind diese Werte essentiell.

Das Layout „Masse“ (s. Abbildung 81) gibt die Möglichkeit das Molekulargewicht von Kohlenhydraten oder anderen organische Verbindungen zu bestimmen.

Name	H	C	N	O	P	S	Masse
Glukose frei	12	6		6			180,0634
Phosphorylierung	1			3	1		79,9663
Methylierung	3	1					15,0235
Hexose	10	6		5			162,0528
Desoxyhexose	10	6		4			146,0579
Pentose	8	5		4			132,0423
Hexosamine	11	6	1	4			161,0688
N-Acetylhexosamine	13	8	1	5			203,0794
NeuAc	17	11	1	8			291,0954
NeuGc	17	11	1	9			307,0903
Perac	2	2		1			42,0106
Permeth	2	1					14,0157

Abbildung 81: Eingabe-Layout „Masse“ in *Differenzen.fp7*

### 3.5.1.5. Kalkulation von Fragmentierungen

Die Identifizierung von Glykanen mittels Massenspektrometrie im MS-Modus ist keine eindeutige Methode. Da sich verschiedene Hexosen nicht in ihrer molekularen Masse, sondern in der Anordnung der Hydroxyl-Gruppen unterscheiden bzw. das Wissen über die beteiligten Bausteine eines Oligosaccharides keine Aussage über die einzelnen Verknüpfungen zulässt, kann eine

Fragmentierung der Ziel-Struktur vorgenommen werden. Dieses ist nicht nur bei Kohlenhydraten, sondern im Prinzip bei allen Molekülen möglich, die im Spektrometer detektierbar sind.

Diese Analyse der durch die Zerstörung eines Mutter-Ions entstandenen Fragmente kann genauere Informationen über die Gesamtstruktur liefern. Die Fragmentierungsmethode ist Geräte- und Molekül-abhängig, d.h. einige organische Verbindungen lassen sich nur mit bestimmten MS-Geräten detektieren bzw. gezielt spalten. Für die Untersuchungen an Glykanen standen in dieser Arbeit zwei MALDI-TOF-Geräte von Bruker (Biflex 99 und Ultraflex III) und ein ESI-IonTrap-Gerät von Agilent zur Verfügung. Dabei sind am Biflex keine Fragmentierungen möglich.

Im Vergleich zu Peptiden (Johnson *et al.*, 1987; Falick *et al.*, 1993) ist die theoretische Kalkulation von Glykan-Fragmenten komplexer, da nicht nur lineare, sondern auch verzweigte Strukturen möglich sind. Um die Erstellung einer eigenen Anwendung zu testen, wurde zuerst mit der Spaltung von Peptiden gearbeitet. Hierfür diente wieder FileMaker als Software und für die Theoriehintergründe Quellen im Internet (z.B. <http://www.matrixscience.com>), (Roepstorff & Fohlman, 1984; Papayannopoulos, 1995; Johnson *et al.*, 1998).

Nach dem Neuerstellen und Austesten von Formeln und Layouts war eine Ausgabe von Fragmenten möglich, jedoch nicht in dem erwünschten Umfang. Die in der Software vorhandenen Funktionen sind nicht für Auswertungen dieser Art vorgesehen, so dass das momentane Ergebnis nur über umständliche Kombinationen erreicht wurde. In Abbildung 82 ist das Layout „AS“ abgebildet. Die in der Datenbank *Peptide.fp7* eingegebene Peptid-Sequenz wird hier angezeigt, mit der Möglichkeit für jede Aminosäure eine zusätzliche Modifikation (Glykan, Phosphorylierung, usw.) anzugeben. Unter deren Berücksichtigung sind bei Auswahl des Layouts „Frag“ die Aminosäuren der Fragmente, deren Einzel-Massen (mx1-3, ma1-3) und die Fragment-Größen (y1-3, a1-3) zu finden.

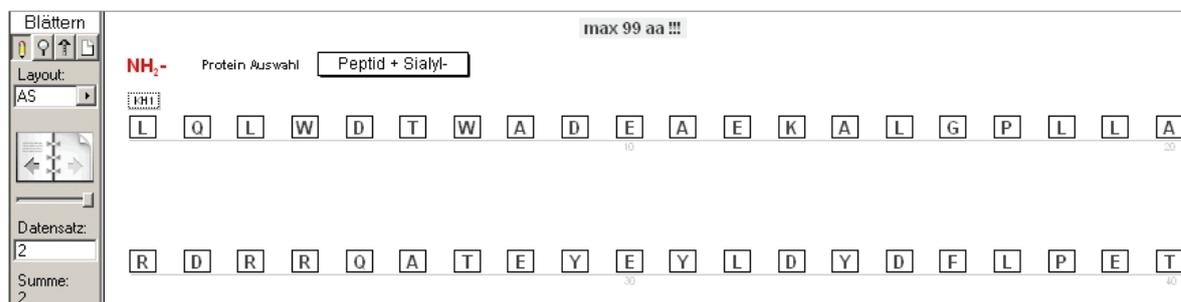


Abbildung 82: Ausschnitt aus dem Eingabe-Layout „AS“ in *Peptide.fp7*

Fragment	Mass	Amino Acid	Fragment	Mass	Amino Acid
mx1	128,0950	K	a1	86,0970	L
mx2	115,0269	D	a2	214,1556	Q
mx3	115,0269	D	a3	327,2396	L

Abbildung 83: Ausschnitt aus dem Eingabe-Layout „Frag“ in *Peptide.fp7*

Die Arbeiten zu den Kalkulationen von Peptid-Fragmenten wurden nicht weiter geführt. Für dieses Ziel und ein Übernehmen dieser Anwendung auf die Glykan-Charakterisierung wäre eine eigene Promotionsarbeit im Bereich der Bioinformatik denkbar.

Ein wichtiges Hilfsmittel für die Fragmentierungs-Analyse von Glykanen war ab dem Jahr 2007 in Form der GlycoWorkbench (s. 2.12.2.2) vorhanden.

### 3.5.1.6. Berechnung von Massen mit GlycoProtMass

Mit der Erstellung von GlycoProtMass ist sowohl die Kalkulation der Massen von Glykanen, als auch deren Modifikation möglich. Es können dabei natürlich nur die Basis-Moleküle verwendet werden, die in der Datenbank hinterlegt sind (s. Tabelle 21). In der GlycoProtMass-Version 3.0 ist ein nachträgliches Hinzufügen von Derivaten ohne, eine Änderung oder Hinzufügen eines Monosaccharids/einer Aminosäure hingegen nur mit Änderungen der Berechnungsformeln möglich. Modifikationen können von jedem Nutzer über einen Eintrag in der jeweiligen *Basics*-Datei hinzugefügt werden.

Tabelle 21: vorhandene Basisdaten in GlycoProtMass zur Massenberechnung

Monosaccharide	Aminosäuren	Modifikationen	Ionen
Desoxy-Pentosen	Alanin	2-Aminobenzamid	H <sup>+</sup>
Desoxy-Hexosen	Asparaginsäure	1,2-Diamono-4,5-methylenedioxybenzenbenzoat	Na <sup>+</sup>
Pentosen	Arginin	Aminoethylthioharnstoff	K <sup>+</sup>
Hexosen	Asparagin	Ethylierung	NH <sub>4</sub> <sup>+</sup>
Hexosamine	Cystein	Sulfatierung	
N-Acetyl-Hexosamine	Glutamin	Phosphorylierung	
N-Acetyl-Neuraminsäure	Glutaminsäure	Methylierung	
N-Acyl-Neuraminsäuren der Analoga-Experimente	Glycin	Acetylierung	
N-Glycolyl-Neuraminsäure	Histidin	Amidierung	
KDN	Isoleucin		
KDO	Leucin		
Hexuronic Acid	Lysin		
Heptosen	Methionin		
Mureinsäure	Phenylalanin		
	Prolin		
	Selenocystein		

	Serin		
	Threonin		
	Tryptophan		
	Tyrosin		
	Valin		

Die Kalkulation der Glykane erfolgt nur mit monoisotopischen Werten, für Proteine kann auch der Durchschnittswert der natürlichen Isotopenverteilung (*average*) verwendet werden. Aufgrund der Anwendungserstellung während der Etablierungszeit der ESI-IonTrap-MS sind die Layouts zusätzlich auf die Ausgabe von Mehrfachladungen ausgerichtet. Für Analysen von MALDI-TOF-Spektren wären diese nicht notwendig

Die optimale Funktionsfähigkeit der Anwendung benötigt das Erstellen von Formeln mit Hilfe der in FileMaker vorhandenen Funktionen und die Verknüpfung von Feldern. In der Tabelle 22 ist ein Beispiel für die Berechnung eines einfachen Glykans aufgezeigt. Die in der *Glykane*-Datei eingetragenen Werte für die Struktur (Monosaccharide, Ionen, Derivate, ...) werden mit den vorhandenen Werten der einzelnen Komponenten der *Basics\_KH*-Datei kombiniert und nach den Vorgaben des Anwenders (Layouts) ausgegeben.

**Tabelle 22: Formel-Beispiel in FileMaker zur Berechnung einer Glykan-Masse**

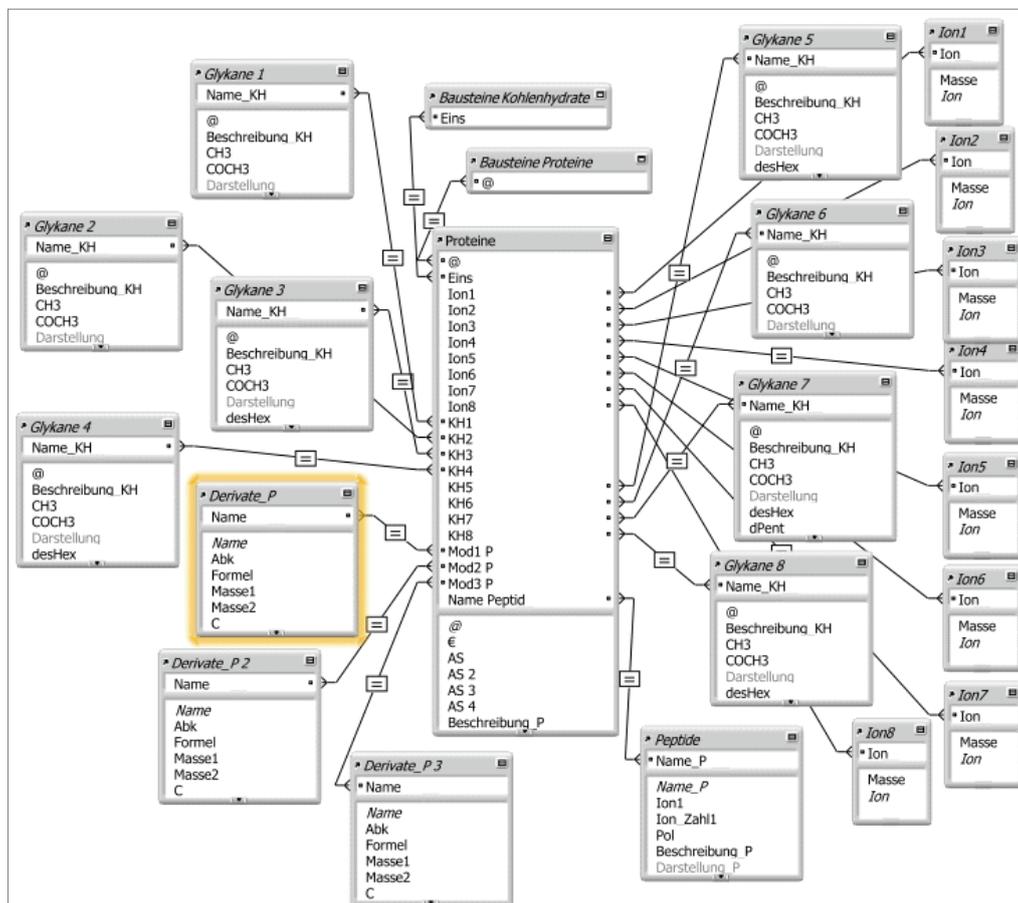
```
= Erweitern(Bausteine Kohlenhydrate::m_redEnd)*redEnd
+Erweitern(Bausteine Kohlenhydrate::m_HexNAc_OS)*HexNAc
+Erweitern(Bausteine Kohlenhydrate::m_Hex_OS)*Hex
+Erweitern(Bausteine Kohlenhydrate::m_desHex_OS)*desHex
+Erweitern(Bausteine Kohlenhydrate::m_Pent_OS)*Pent
+Erweitern(Bausteine Kohlenhydrate::m_NANA_OS)*Neu5Ac
+Bausteine Kohlenhydrate::SO3*SO3_KH
+Bausteine Kohlenhydrate::HPO3*HPO3_KH
+Bausteine Kohlenhydrate::Methyl*CH3
+Bausteine Kohlenhydrate::Acetyl*COCH3
+Bausteine Kohlenhydrate::m_KDN_OS*KDN
+Bausteine Kohlenhydrate::m_KDO_OS*KDO
+Bausteine Kohlenhydrate::m_dPent_OS*dPent
+Bausteine Kohlenhydrate::m_HexA_OS*HexA
+Bausteine Kohlenhydrate::m_Hep_OS*Hep
+Bausteine Kohlenhydrate::m_Mur_OS*Mur
+Bausteine Kohlenhydrate::m_HexN*HexN
+Falls(Derivat="keine";0;)
Derivat="2_AB";Bausteine Kohlenhydrate::m_2_AB_OS;
Derivat="DMB";Bausteine Kohlenhydrate::m_DMB_OS; Derivat="Reduziert";Bausteine
+Falls(Pol="pos"; Falls(Ion="H+";Bausteine Kohlenhydrate::m_H*Ion_Zahl; Ion="Na+";Bausteine
Ion="K+";Bausteine Kohlenhydrate::m_K*Ion_Zahl; Ion="NH4+";Bausteine
-Wenn(Pol="neg"; Ion_Zahl*Bausteine Kohlenhydrate::m_H;0)
```

Mit zunehmender Anzahl an Feldern für Dateneingabe, Formeln, Verknüpfungen innerhalb und zwischen Dateien und der Ausgabe an Informationen (s. Tabelle 23) ist es schwierig, den Einfluss einer Änderung in einem Feld auf andere zu kalkulieren. Für nachfolgende Arbeiten ist es deshalb ratsam zu einem Zeitpunkt stets nur ein Feld zu modifizieren

**Tabelle 23: Anzahl der Felder der Datenbank-Dateien von GlycoProtMass 3.0**

Datei	Anzahl der Felder
<i>Basics_KH</i>	136
<i>Glykane</i>	155
<i>Basics_Proteine</i>	113
<i>Peptide</i>	110
<i>Glykoproteine</i>	267
<i>Differenzen</i>	69
<b>Summe</b>	<b>850</b>

Eine wichtige Regel für das Arbeiten mit GlycoProtMass ist das Verhindern des Umbenennens der Dateien. Für die Funktionsfähigkeit der Anwendung müssen die absoluten Bezüge in den Formeln erhalten bleiben (fehlende oder falsche Verknüpfungen werden durch die Software angezeigt). Die Abbildung 84 zeigt an einem Beispiel den Umfang der Beziehungen der Datei *Glykoproteine* zu internen und externen Datenbank-Feldern.



**Abbildung 84: Übersicht der Verknüpfungen innerhalb der Datei *Glykoproteine***

Die Kalkulation eines Glykans, Peptids oder Glykoproteins mit Hilfe von GlycoProtMass ist die Voraussetzung für die Nutzung der Datenbankfunktion, welche im Folgenden beschrieben wird.

### 3.5.1.7. Erstellung von Datenbanken mit GlycoProtMass

Wird ein Datensatz für die Kalkulation einer Struktur nicht gelöscht, so ist dieser automatisch Bestandteil der Datei und somit der jeweiligen Datenbank. Die Vorteile im Vergleich zu z.B. einer Standard-Excel-Tabelle ist die einfachere Sortierung der Daten nach Kriterien, die durch die vorgegeben Felder zur Informations-Ein(Aus)-gabe definierbar sind. Außerdem können die Werte in verschiedenen Layouts dargestellt werden, d.h. spezifische Ansichten für die Berechnung, Ergebnisvergleiche, Export und Ausdrücke sind möglich. Die Darstellungen der Glykan-Struktur in GlycoProtMass wurde ab 2007 immer mit der GlycoWorkbench vorgenommen, da hier der Zeitaufwand relativ gering ist und sich gleichzeitig die Massenberechnungen für zukünftige Formeländerungen überprüfen lassen. Die Struktur war einfach von GlycoProtMass über die Zwischenablage in das Feld „Darstellung“ der Glykan-Datenbank zu kopieren. Für die Bearbeitung einzelner Datensätze bzw. Strukturen können diese ausgeschlossen und dann als Gruppe verändert oder ausgegeben werden. Um für mehrere Datensätze gleichzeitig ein Datenfeld zu ändern ist die Verwendung von Skripten zu bevorzugen, da bei höheren Anzahlen von Datensätzen der Zeitaufwand bei manueller Veränderung jedes einzelnen zu hoch wäre. Skripte sind abgespeicherte Befehlsabläufe, d.h. bestimmte Vorgehensweisen können archiviert und für späterer Wiederholungen genutzt werden. In den Dateien von GlycoProtMass sind für viele Felder Skripte vorhanden, um z.B. für ausgewählte Datensätze das Kation von  $\text{Na}^+$  auf  $\text{K}^+$  zu ändern, die Polarität zu variieren und Anzahlen an Komponenten zu modifizieren. Die Erstellung neuer Skripte ist ohne größere Aufwände möglich, so dass wie auch bei den Layouts ein Nutzerangepasstes Arbeiten ermöglicht wird. Die vorhandenen Daten können zusätzlich in andere Datenformate, wie z.B. Excel, Text, u.a. exportiert werden, um abgespeicherten Daten auch ohne die Software Filemaker zu nutzen.