



# DIPLOMARBEIT

Titel der Diplomarbeit

„Bioinformatische Methoden zur Auswertung von Daten aus Messungen des Metabolic Profil/ Fingerprints/ Metabolomics“

Verfasserin

Barbara Klug

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.rer.nat.)

Wien, 2011

Studienkennzahl lt. Studienblatt:

A 474

Studienrichtung lt. Studienblatt:

Diplomstudium Ernährungswissenschaften

Betreuerin / Betreuer:

Univ.-Prof. Dr. Jürgen König



<b>1</b>	<b>EINLEITUNG UND FRAGESTELLUNG</b>	<b>1</b>
<b>2</b>	<b>DEFINITION DER RELEVANTEN BEGRIFFE METABOLOMIK, METABONOMIK, TARGETED ANALYSIS, METABOLIC PROFILING UND METABOLIC FINGERPRINTING</b>	<b>3</b>
<b>3</b>	<b>ÜBERBLICK ÜBER DIE BENÖTIGTEN ARBEITSSCHRITTE BEI DER METABOLITENERFASSUNG IM BEREICH ERNÄHRUNG</b>	<b>5</b>
<b>4</b>	<b>AUSWIRKUNGEN VON EXOGENEN UND ENDOGENEN FAKTOREN AUF DEN STOFFWECHSEL</b>	<b>7</b>
<b>5</b>	<b>DIE NOTWENDIGKEIT DER STANDARDISIERUNG DER METHODEN UND BEGRIFFE BEI HUMANSTUDIEN IM BEREICH METABOLOMIK</b>	<b>11</b>
<b>6</b>	<b>PROBENMATERIAL, -ENTNAHME UND -AUFBEWAHRUNG ZUR GEWINNUNG VON METABOLITEN</b>	<b>14</b>
<b>7</b>	<b>EXTRAKTIONSVERFAHREN ZUM AUFTRENNEN DER STOFFMENGE UND ZUR GEWINNUNG DER ZU UNTERSUCHENDEN METABOLITE</b>	<b>17</b>
<b>8</b>	<b>CHEMISCH ANALYTISCHE METHODEN ZUR ANALYSE DER METABOLITE</b>	<b>20</b>
<b>8.1</b>	<b>Beschreibung des Einsatzes von Massenspektrometrie (MS) sowie deren Vor- und Nachteile in der Metabolomik</b>	<b>21</b>
8.1.1	Kopplung von Gaschromatographie mit Massenspektrometrie (GC/MS)	22
8.1.2	Kopplung von Flüssigkeitschromatographie beziehungsweise Hochleistungsflüssigkeitschromatographie mit Massenspektrometrie (LC/MS beziehungsweise HPLC/MS)	23
8.1.3	Kopplung von Kapillarelektrophorese mit Massenspektrometrie (CE/MS)	24
<b>8.2</b>	<b>Beschreibung des Einsatzes von Kernspinresonanzspektroskopie (NMR), sowie deren Vor- und Nachteile in der Metabolomik</b>	<b>25</b>

**8.3 Beschreibung des Einsatzes von Schwingungsspektroskopie, sowie deren Vor- und Nachteile in der Metabolomik 26**

**8.4 Weitere Analysemethoden, welche im Bereich Metabolomik eingesetzt werden 27**

## **9 COMPUTERGESTÜTZTE VERFAHREN ZUR ROH-DATEN-ERFASSUNG DER UMFANGREICHEN METABOLIT-DATEN 28**

**9.1 Datenverarbeitung von MS-Daten 28**

9.1.1 Rohdaten Vorverarbeitung und Filterung 29

9.1.2 Mustererkennung 30

9.1.3 Ausrichtung 31

9.1.4 Normalisierung 32

**9.2 Datenverarbeitung von NMR-Daten 33**

## **10 VERFAHREN FÜR DIE STATISTISCHE DATENANALYSE DER METABOLITDATEN 34**

**10.1 Supervised Methods– Überwachte Methoden 35**

10.1.1 Pfadanalyse 35

10.1.2 Varianzanalyse 36

10.1.3 Künstliche neuronale Netzwerke 36

10.1.4 Evolutionäre Algorithmen 37

10.1.5 Regelinduzierte Algorithmen 38

**10.2 Unsupervised Methods– Unüberwachte Methoden 38**

10.2.1 Hauptkomponentenanalyse 38

10.2.2 Clusterverfahren 39

## **11 SOFTWAREPROGRAMME ZUR DATENBEARBEITUNG IM BEREICH METABOLOMIK 41**

**11.1 Beispiele für kommerzielle Software zur Datenbearbeitung im Bereich Metabolomik 42**

11.1.1 BlueFuse for Biomarkers 42

11.1.2 MarkerLynx 42

11.1.3 MarkerView 43

11.1.4 SIEVE 43

<b>11.2</b>	<b>Beispiele für frei zugängliche Software zur Datenbearbeitung im Bereich Metabolomik</b>	<b>44</b>
11.2.1	Automated Mass Spectral Deconvolution and Identification System (AMDIS)	44
11.2.2	metaP-Server	45
11.2.3	MetAlign	45
11.2.4	MZmine und MZmine 2	46
11.2.5	XCMS und XCMS <sup>2</sup>	47
11.2.6	COMparision of SPectrAI Retention Information (COMSPARI)	48
11.2.7	MetaboAnalyst	48
<b>12</b>	<b>ANSÄTZE ZUR DATENSPEICHERUNG FÜR DEN AUFBAU VON METABOLIT-DATENBANKEN</b>	<b>50</b>
<b>13</b>	<b>AUFLISTUNG BEREITS VORHANDENER DATENBANKEN IM BEREICH METABOLOMIK</b>	<b>52</b>
13.1	Kyoto Encyclopedia of Genes and Genomes (KEGG)	52
13.2	Human Metabolome Database (HMDB)	53
13.3	The Small Molecule Pathway Database (SMPDB)	55
13.4	WikiPathways	56
13.5	HumanCyc	57
13.6	Reactome	57
<b>14</b>	<b>ZUKÜNFTIGE ANWENDUNGSBEREICHE FÜR METABOLITDATEN</b>	<b>59</b>
14.1	Verknüpfung der Datenbanken mit anderen "Omiks-Daten"	59
14.2	Identifizierung von Biomarkern	60
14.3	Individuelle Ernährung und individuelle Medizin	60
<b>15</b>	<b>SCHLUSSBETRACHTUNG</b>	<b>63</b>
<b>16</b>	<b>ZUSAMMENFASSUNG</b>	<b>67</b>

<b>17</b>	<b>ABSTRACT</b>	<b>68</b>
<b>18</b>	<b>ABBILDUNGSVERZEICHNIS</b>	<b>69</b>
<b>19</b>	<b>LITERATURVERZEICHNIS</b>	<b>70</b>
<b>20</b>	<b>LEBENS LAUF</b>	<b>82</b>

## 1 Einleitung und Fragestellung

Viele Menschen beschäftigt die Frage nach der richtigen Ernährung. Zahlreiche Studien belegen, dass ungeeignete Lebensmittelauswahl zu einem metabolischen Ungleichgewicht führen kann, wodurch das Risiko von ernährungsabhängigen Krankheiten wie Arteriosklerose, Adipositas, Diabetes Typ 2, Bluthochdruck, Nahrungsmittelallergien und –unverträglichkeiten, Magen-Darm-Erkrankungen und entzündliche Erkrankungen erhöht wird.

Die bisherige Forschung hat gezeigt, dass auf Grund genetischer und phänotypischer Unterschiede jedoch nicht einfach eine allgemein gültige, optimale Ernährungsweise für jeden einzelnen festzulegen ist. Eine Ernährung, welche für den einen optimal ist, kann bei jemand anderem Krankheiten prädisponieren. Aus diesem Grund ist es für die Wissenschaft wichtig, optimale Ernährungsprofile auf der Grundlage individueller Stoffwechselprofile zu erstellen, welche die Möglichkeiten für jeden einzelnen eröffnen eine bessere Lebensmittelauswahl zu treffen und somit die Lebensqualität zu erhöhen.

Ernährungswissenschaftler versuchen die Grundlagen der Beziehung zwischen Ernährung und Gesundheit zu verstehen, indem sie die Wechselwirkungen zwischen Ernährung und Stoffwechselwegen untersuchen. Um den Schritt in Richtung personalisierte Ernährung (beziehungsweise personalisierte Medizin) zu gehen, sind Erkenntnisse aus den Bereichen Genomik (Gesamtheit aller Gene), Transkriptomik (Gesamtheit aktivierter Gene zu einer Zeit unter bestimmten Bedingungen), Proteomik (Gesamtheit der Proteine zu einer Zeit unter bestimmten Bedingungen), Metabolomik (Gesamtheit der Stoffwechselprodukte zu einer Zeit unter bestimmten Bedingungen) und Bioinformatik nötig [German et al., 2004].

Die hier vorliegende Arbeit beschäftigt sich mit einige Fragestellungen und Probleme im Hinblick auf Erforschung des menschlichen Metaboloms. Im Detail werden die folgenden Punkte diskutiert:

- Die Problematik der teilweise überschneidenden Begriffsdefinition, sowie der Definition von Standards im Bereich der Metabolomik.

- Die Wahl des zu untersuchenden Probenmaterials für ernährungswissenschaftlichen Studien, sowie die Erfassung verschiedener endogener und exogener Faktoren, welche Einfluss auf das Metabolitenprofil haben.
- Die Extraktionsverfahren und die Wahl der geeigneten Technologie zur chemischen Analyse, mit deren Hilfe die gewonnenen Metabolite ausgewertet werden.
- Die Verarbeitung und statistische Analyse der Daten, sowie die Aufzählung vorhandener Softwareprogramme, welche die Bearbeitung der enormen Menge an Metabolitendaten erleichtern.
- Die notwendigen Kriterien, welche Metabolit-Datenbanken erfüllen sollen und die Aufzählung bereits zur Verfügung stehender Datenbanken.
- Der Ausblick auf mögliche Anwendungsgebiete für die gewonnenen Erkenntnisse aus den Metabolitdaten in der Ernährungswissenschaft.



## **2 Definition der relevanten Begriffe Metabolomik, Metabonomik, targeted analysis, metabolic profiling und metabolic fingerprinting**

Die Metabolomik befasst sich mit der Identifizierung und quantitativen Analyse aller oder zumindest einem großen Teil der zellulären Metabolite [DETTMER et al., 2007]. Der Begriff Metabolomik gilt als Ergänzung zu den Begriffen Genomik, Transkriptomik und Proteomik [WHITFIELD et al., 2004].

Als Metabolom wird die Gesamtheit der endogenen und exogenen Metabolite in einem Organismus bezeichnet. Metabolite sind kleine Moleküle mit niedrigem Molekulargewicht (ca. <1000 Dalton), welche bei Stoffwechselreaktionen beteiligt sind [DETTMER et al., 2007]. Zu den häufigsten Metaboliten zählen Aminosäuren, Lipide, Vitamine, kleine Peptide und Kohlenhydrate.

Da Metabolite die Endprodukte von Genexpression und physiologischen Regulationsvorgängen darstellen, können sie herangezogen werden um vorliegende biochemische Zustände zu beschreiben. Im Bereich der Nutrigenomik werden vor allem nahrungs- und ernährungsabhängige metabolische Veränderungen und deren Auswirkungen auf den Stoffwechsel untersucht [GARCIA-CANAS et al., 2010].

Es gibt einige Fachbegriffe im Bereich der Metabolomik, welche häufig kongruierend verwendet werden, da genaue Definitionen noch nicht festgelegt wurden. So kann zum Beispiel der Begriff Metabolomik synonym zur Bezeichnung metabolic fingerprinting verwendet werden [WHITFIELD et al., 2004].

Es gibt drei wichtige Ansätze im Bereich Metabolomik:

- Targeted analysis, also eine zielgerichtete Analyse, bezieht sich auf eine quantitative Messung ausgewählter Metabolite, also spezifischer Biomarker oder Reaktionsprodukten [GARCIA-CANAS et al., 2010]. Gemessen werden die Konzentrationen von einer begrenzten Anzahl bekannter Metabolite. Dafür muss die Struktur der Metabolite bekannt sein, eine Analysetechnik zur Verfügung stehen und die Verbindungen für die Kalibrierung gereinigt werden. Der Nachteil ist demnach, dass mit dieser Methode keine neuen Biomarker identifiziert werden können [SHULAEV, 2006].

- Metabolic profiling befasst sich mit der Untersuchung spezifischer Substanzklassen oder Metabolite, welche an einem bestimmten Stoffwechselweg beteiligt sind. Es ist eines der grundlegenden Konzepte für die Untersuchung von metabolischen Profilen einer Zelle zur genaueren Beschreibung des Phänotyps [GARCIA-CANAS et al., 2010]. Körperflüssigkeiten können mit Hilfe von metabolic profiling analysiert werden, und bei der Diagnose des Gesundheitszustandes von Probanden helfen. Die Schwierigkeiten, welche bei dieser Methode auftreten, sind die hohe Anzahl an Metaboliten, welche sich pro Probe ergeben und deren Auswertung mit Hilfe von bioinformatischen Methoden [SHULAEV, 2006]. Laut Dettmer et al. ist bei metabolic profiling die targeted analysis inkludiert [DETTMER et al., 2007].
- Metabolic fingerprinting beschäftigt sich mit dem Vergleich von Metabolitmustern, welche sich auf Grund unterschiedlicher zellulärer Bedingungen, welche zum Beispiel durch Umwelteinflüsse, Krankheiten, genetische Störungen, etc. beeinflusst werden, verändern [GARCIA-CANAS et al., 2010]. Diese Methode wird vor allem bei der Identifizierung von Biomarkern und in der Diagnostik, zur Erkennung spezifischer Muster von metabolischen Krankheiten, verwendet [SHULAEV, 2006].

Es findet sich in der Literatur auch Unterscheidung, welche nur zwischen target beziehungsweise metabolitspezifischen Studien und globaler Metabolomik differenzieren [ISSAQ et al., 2009].

### **3 Überblick über die benötigten Arbeitsschritte bei der Metabolitenerfassung im Bereich Ernährung**

Bei Studien, welche im Bereich Metabolomik im Rahmen der Ernährungswissenschaft durchgeführt werden, ergeben sich eine Reihe von Arbeitsschritten, auf die in der folgenden Arbeit detailliert eingegangen wird.

Je nach Studiendesign werden von den Probanden unterschiedliche Proben (Urin, Blut, Speichel,...) genommen und die darin enthaltenen Metabolite extrahiert.

Zur Auftrennung werden diverse Trennmethoden, wie zum Beispiel Gaschromatographie, Flüssigkeitschromatographie oder Kapillarelektrophorese, verwendet. Schwierigkeiten ergeben sich hinsichtlich der Tatsache, dass die zu untersuchenden Metabolite sehr unterschiedliche chemische und physikalische Eigenschaften besitzen.

Bei der anschließenden Analyse der Metabolite haben sich in der Praxis zwei Verfahren durchgesetzt, nämlich die Massenspektrometrie und die Kernspinresonanzspektroskopie.

Die gewonnenen Daten werden, auf Grund ihrer enormen Anzahl, mit Hilfe von Computerprogrammen, welche meist Methoden der Datenverarbeitung und der statistischen Analyse (Hauptkomponentenanalyse, Pfadanalyse, etc.) kombinieren, analysiert.

Die gewonnenen Informationen können anschließend zum Beispiel in diverse Datenbanken (KEGG, HMDB, SMPDB, etc.) eingetragen oder mit vorhanden Daten aus diesen Datenbanken verglichen werden.

Die nachfolgende Abbildung zeigt die schematische Darstellung der notwendigen Arbeitsschritte zur Erfassung der Metabolite im Bereich der Nutrigenomik [GARCIA-CANAS et al., 2010].

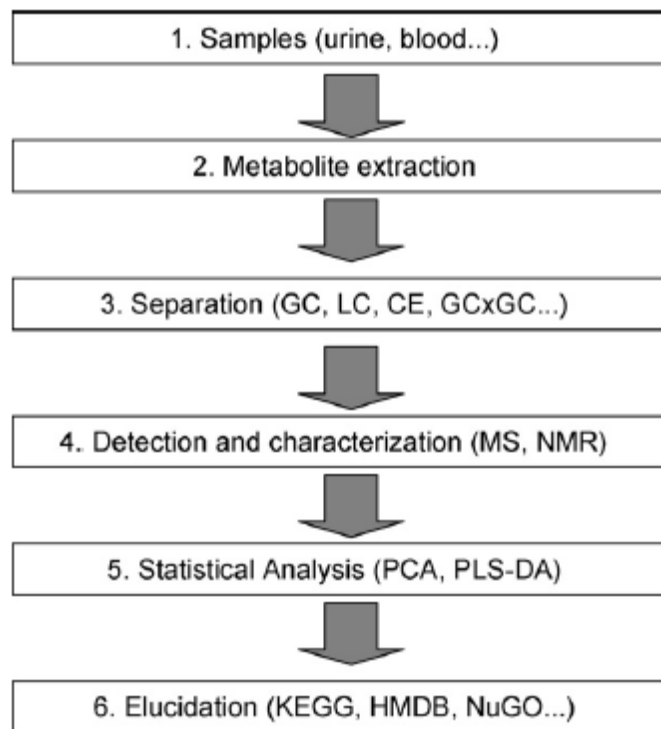


Abbildung 1: Arbeitsschritte zur Metabolitenerfassung in der Nutrigenomik  
[GARCIA-CANAS et al., 2010]

#### **4 Auswirkungen von exogenen und endogenen Faktoren auf den Stoffwechsel**

Die Erstellung von metabolischen Profilen ist einerseits vom ernährungsbedingten und biorhythmischen Status des jeweiligen Organismus abhängig. Andererseits auch von der Durchführung der anschließenden Analysetechnik und der Identifizierung der relevanten Metaboliten, worauf später noch detaillierter eingegangen wird.

Schwierigkeiten bestehen auf Grund der hohen biologischen Variation im menschlichen Stoffwechsel und den teilweise nur schwer oder gar nicht kontrollierbaren Bedingungen im Bereich Umwelt, Ernährung, Tagesschwankungen, Geschlecht, Krankheiten, Lebensstil, Zeitpunkt der Probensammlung und Einnahme von Arzneimittel. Um biologische Schwankungen möglichst gering zu halten ist zum Beispiel die Einführung einer Kontrollgruppe oder das Fasten der Probanden vor der Probenentnahme empfehlenswert. So können auch Schwankungen, welche sich auf Grund unterschiedlicher Metabolitprofile von Gesunden und Kranken ergeben, abgeleitet werden [DUNNE et al., 2005].

Auf Grund der erwähnten Gründe ist es notwendig, bei Experimenten ausführlich über die jeweiligen Umweltbedingungen, als auch über den Allgemeinzustand der Probanden zu berichten [CASTLE et al., 2006].

Hinsichtlich der biologischen Variationen innerhalb der Bevölkerung gibt es noch einige Unklarheiten, sei es den Einfluss von Genetik betreffend oder die Unterschiede in der Körperzusammensetzung, welche Auswirkungen auf den Stoffwechsel haben können [BRENNAN, 2008].

Ein wichtiger Punkt ist demnach ebenfalls die möglichst genaue Erfassung der Nahrungsmittel, sowie der darin enthaltenen Stoffe, welche möglicherweise Einfluss auf das metabolische Profil haben. Des Weiteren müssen die endogenen Faktoren und exogene Faktoren, welche den Stoffwechsel beeinflussen, berücksichtigt werden. Zu den endogenen Faktoren zählen zum Beispiel Ernährung, Arzneimittelkonsum, physikalische Aktivität, Dickdarmflora oder Stress. Die exogenen Faktoren beziehen sich zum Beispiel auf Körperzusammensetzung, Alter, Gesundheitsstatus, Stoffwechselrate im Ruhezustand oder Genotyp [GIBNEY et al., 2005].

Die nachfolgende Abbildung zeigt zusammenfassend die Einflüsse der Omiks-Kaskade (Genomik, Transkriptomik, Proteomik und Metabolomik) – oder anders ausgedrückt der DNA, RNA, Proteine, Metabolite – sowie die Einflüsse von Medizin, Umwelt, Lebensstil und der Ernährung auf den Phänotyp. Der Phänotyp, das Erscheinungsbild eines Organismus, kann direkt mit Krankheit oder Gesundheit in Verbindung gebracht werden [ZEISEL et al., 2005].

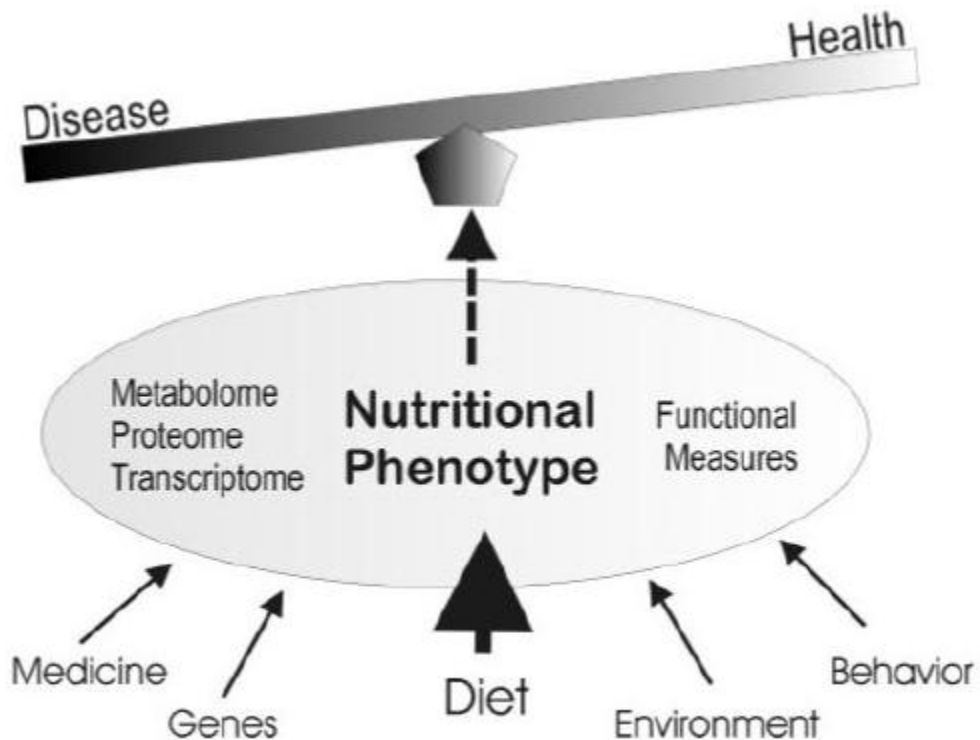


Abbildung 2: Phänotyp beeinflussende Faktoren [ZEISEL et al., 2005]

Die Auswirkungen von genetischen und umweltbedingten Faktoren auf das Metabolom lassen sich, laut Oresic, am besten an Hand von Zwillingsstudien zeigen.

Generell zeigt sich die Bedeutsamkeit metabolische Abweichungen, welche durch Veränderungen der Nahrungsaufnahme hervorgerufen werden, genau zu untersuchen. Studien, wie zum Beispiel die unten angeführte von Walsh et al., welche eine Standard-Diät am Tag vor der Probenentnahme vorschreiben zeigten, dass Variationen bei derselben Person im Hinblick auf die enthaltenen Metabolite im Urin reduziert werden können, während Schwankungen sowohl

bei der Plasma- als auch bei der Speichelprobe gegeben waren [ORESIC, 2009].

Walsh et al. erforschten die Auswirkungen standardisierter Ernährung auf das Stoffwechselprofil in Urin, Plasma und Speichel. Sie zeigten, dass Urin Änderungen in der Ernährung widerspiegelt, Plasma und Speichel hingegen nicht. Des Weiteren traten tagesabhängige Schwankungen im Metabolitprofil von Urin und Speichel auf. Untersucht wurden 30 gesunde Probanden, Männer und Frauen, deren Proben an 4 separaten Tagen gesammelt wurden. An den ersten beiden Tagen dokumentierten die Probanden ihren normalen Tagesablauf inklusive Ernährung und durchgeführte körperliche Aktivität. Am dritten Tag musste möglichst genau der Tagesablauf vom zweiten Tag nachvollzogen werden, um intraindividuale Schwankungen festzustellen. Für den vierten Tag bekamen die Probanden eine Standard-Diät vorgeschrieben, um etwaige interindividuale Schwankungen einzugrenzen. Die Proben wurden jeweils am nächsten Morgen abgenommen, Urin- und Speichelproben jeweils auch am Abend davor, um tagesabhängige Schwankungen zu erfassen. Die Ermittlung des Metabolitprofils erfolgte mittels NMR und anschließender multivariater Datenanalyse [WALSH et al., 2006].

Stella et al. untersuchten kurzfristige Auswirkungen von vegetarischer Ernährung und Ernährung mit geringem beziehungsweise hohem Fleischanteil auf das metabolische Profil. 12 männliche Probanden konsumierten jede der beschriebenen Ernährungsformen jeweils 15 Tage lang. Dazwischen wurde jeweils eine 7 tägigen Pause eingelegt. Die gesammelten Urinproben wurden mittels  $^1\text{H}$  NMR Spektroskopie und multivariater statistischer Analysemethoden analysiert. Mit Hilfe von Programmen zur Mustererkennung konnten charakteristische metabolisch bedingte Unterschiede zwischen den drei Ernährungsweisen aufgezeigt werden [STELLA et al., 2006].

Kochhar et al. untersuchten Urin- und Plasmaproben von 150 männlichen und weiblichen Probanden unter Verwendung von  $^1\text{H}$  NMR und multivariater statisti-

scher Analysemethode. Ziel der Studie war es Zusammenhänge zwischen Geschlecht, Alter, Body Mass Index und dem metabolischen Profil aufzuzeigen [KOCHHAR et al., 2006].

Gibney et al. geben zu bedenken, dass der Anteil der an sekundären Pflanzeninhaltsstoffen in der Nahrung, welche potentielle Auswirkungen auf metabolische Effekte haben, weitaus größer ist, als der Anteil der Hauptnährstoffe. Des Weiteren weisen die Autoren darauf hin, dass auch chemische Veränderungen in der Lebensmittelmatrix – hervorgerufen durch Kochen oder Verdauung – beachtet werden müssen [GIBNEY et al., 2005].

Es gibt mittlerweile, wie bei Kussmann et al. angeführt, eine Vielzahl von Beispielen verschiedener Studien, welche sich mit den Auswirkungen unterschiedlicher Ernährung beziehungsweise einzelner Nahrungsmittel (wie zum Beispiel Tee (vergleiche Law et al., Van Dorsten et al.), Kamille (vergleiche Wang et al.) oder Soja-Isoflavone (vergleiche Solanky et al., 2003 beziehungsweise Solanky et al., 2005) auf den Stoffwechsel befassen. Die Autoren merken auch an, dass sich bei vielen Studien zeigt, dass ernährungsbedingte Vorlieben und Gewohnheiten auch von kulturellen, sozio-ökonomischen, psychologischen und biologischen Faktoren beeinflusst werden [KUSSMANN et al., 2008].



## **5 Die Notwendigkeit der Standardisierung der Methoden und Begriffe bei Humanstudien im Bereich Metabolomik**

Standardisierte Begriffe und bioinformatische Standards sollen den Austausch, das Management, die Interpretation und den Vergleich der Informationen im Bereich der Metabolomik erleichtern [CASTLE et al., 2006]. Vor allem im Hinblick darauf, dass dieser Bereich in Bezug auf die Ernährungswissenschaft erst mehr oder weniger am Beginn der Erforschung steht, ist es ratsam sich möglichst bald auf klare Standards einigen.

Betreffend das Studiendesign müssen einheitliche Richtlinien hinsichtlich der Probenentnahme und Probenaufbereitung, sowie der Standardisierung von Flüssigkeiten, Zeiten, Mengen, und Verarbeitungshilfsmitteln erarbeitet werden. Die Nutzung von Datenbanken zum Vergleich von Probanden und der Identifikation von Metaboliten macht nur dann Sinn, wenn gewisse Kriterien für die Datenerfassung festgelegt werden. Mehrere Ansätze in diese Richtung wurden bereits unternommen [GIBNEY et al., 2005].

Die Arbeitsgruppe Standard Metabolic Reporting Structure (SMRS), eine Sammlung von Interessensvertretern aus Wissenschaft, Industrie und Regierung, stellt Empfehlungen für die Standardisierung von Durchführung und Berichterstattung für Studien im Bereich Metabolomik und Metabolomik auf. Ziel ist vor allem, die Kommunikation zwischen den einzelnen Tätigkeitsbereichen zu erleichtern [LINDON et al., 2005].

SMRS beteiligte sich zum Beispiel bei einem Workshop, der 2005 im National Institutes of Health (NIH) abgehalten wurde. Dieser beschäftigte sich vor allem mit dem Thema Standardisierung in Bezug auf die Bereiche Methoden, Technologien und Datenbearbeitung. Dabei wurden gewisse minimale Standards festgelegt um die Berichterstattung in Bezug auf Metabolitdaten und den zugrunde liegenden Studiendesigns zu strukturieren.

Da Metabolite sowohl vom biologischen Zustand des Organismus, als auch von der Analysetechnik beeinflusst werden können, muss eine genaue Beschreibung des jeweiligen Experiments vorliegen. Dies erleichtert außerdem die anschließende Datenanalyse am Computer. Das bedeutet aber auch die Entwick-

lung von Ontologien in Bezug auf die Experimente und der dabei verwendeten Instrumente. Wichtig ist eine Vereinheitlichung im Bereich Technologie und Terminologie – vor allem bezüglich der Vernetzung mit anderen Informationen, sei es zum Beispiel aus dem Bereich der Genomik oder Proteomik.

Eine genaue Beschreibung der gegebenen Umweltbedingungen, ernährungsbedingter und biorhythmischer Faktoren, als auch des körperlichen Zustandes der Probanden ist nötig um Ergebnisse zu vergleichen beziehungsweise Studien gegebenenfalls zu wiederholen.

Die gewonnenen Metabolite werden mit Hilfe von Datenbanken identifiziert, welche die physikalischen und chemischen Eigenschaften von Metaboliten beinhalten. Die genaue Beschreibung der Metabolite benötigt eine Vielzahl von gebündelten Technologien. Daher ergibt sich die Notwendigkeit alle verwendeten Technologien detailliert zu beschreiben, um zu gewährleisten, dass die Resultate nachvollziehbar und reproduzierbar sind. Unterschiedliche Arbeitsschritte bei der Bearbeitung des Probenmaterials können zu Veränderungen in der Metabolitkonzentration führen.

Identifizierte Metabolite sollen mit Literaturquellen, experimentellen Daten und Datenbanken verlinkt werden. Die Stoffwechseldatenbanken sollen sich aus Referenzdatenbanken, Datengruppen (welche sich auf ein bestimmtes Thema, wie zum Beispiel Organismen, Technologien oder Krankheiten, fokussieren), Indexes (welche Zugang zu vielen anderen Datenbanken gewährleisten) und Labor-Informationen-Management-Systemen (zur Organisation von komplexen Experimenten) zusammensetzen.

Es wurde festgehalten, dass für die weitere Ausarbeitung und Entwicklung von Standards im Bereich Metabolomik eine Zusammenarbeit von möglichst vielen beteiligten Forschungsgruppen wünschenswert ist [CASTLE et al., 2006].

Goodacre et al. versuchen einen möglichen Ansatz hinsichtlich der standardisierten Berichterstattung von statistischen Analysen bei Stoffwechseldaten vorzugeben. Es werden die Themenbereiche Studiendesign, Datenerfassung und Datenspeicherung, Präprozessor, Datenanalyse und Interpretation angesprochen [GOODACRE et al., 2007].

Fiehn et al. beschäftigen sich auch mit Standardisierungsrichtlinien für Publikationen in Journalen. Dabei soll unter anderem darauf geachtet werden, dass bei Artikeln über experimentelle Studien eine detaillierte Angaben über das Studiendesign, sowie die Metadaten vorliegt, so dass die Argumentationen nachvollziehbar sind, um Studien gegebenenfalls zu wiederholen. Metadaten sollen auch in Datenbanken gesammelt werden, um die Wiederverwendung der Daten zu erleichtern und neue Beziehungen zwischen einzelnen Metaboliten aufzuzeigen und daraus neue Erkenntnisse zu gewinnen [FIEHN et al., 2006].

Bino et al. empfehlen MIAMET (Minimum Information About a METabolomics experiment) als Standardisierungsrichtlinie, da MIAMET besonders auf die standardisierte Berichterstattung von Experimenten im Bereich Metabolomik ausgerichtet ist. Es werden die Themen experimentelles Design, Probensammlung und –vorbereitung, Extraktion, Derivatisierung, metabolic profiling Design, sowie Datenverarbeitung diskutiert [BINO et al., 2004].

Sumner et al. konzentrieren sich vor allem auf die Standardisierung im Bereich chemische Analysen, wobei vor allem auf Massenspektrometrie und Kernspinresonanzspektroskopie eingegangen wird. Dabei werden die Themenbereiche Probenvorbereitung, experimentelle Analyse, Qualitätskontrolle, Identifizierung von Metaboliten und die Datenvorverarbeitung angesprochen [SUMNER et al., 2007].

## **6 Probenmaterial, -entnahme und -aufbewahrung zur Gewinnung von Metaboliten**

In Bezug auf die Handhabung von Probenmaterial gibt es einige wichtige Parameter zu beachten, da diese die Qualität der Proben, die Validität der Ergebnisse und die Gesundheit des Analytisten beeinflussen können [ISSAQ et al., 2009].

Laut Gibney et al. sind Blut, Urin und Speichel die drei wichtigsten Körperflüssigkeiten, aus welchen menschliche Metabolite gewonnen werden können [GIBNEY et al., 2005]. Der Vorteil bei Urin- und Speichelprobenentnahme besteht darin, dass es sich um eine nichtinvasive Methode handelt. Während es sich bei der Entnahme von Blutproben (Serum und Plasma) um invasive Verfahren handelt [DUNN et al., 2005].

Speichelproben wurden bis jetzt im Bereich der Ernährungsforschung noch nicht verbreitet analysiert. Dennoch sehen Gibney et al. durchaus Potential, sei es in der Differenzierung von Metabolitprofilen oder in der Aufzeichnung von Änderungen im Metabolitprofil durch unterschiedliche Ernährungsformen. Speichel enthält größere Mengen an Hormonen (zum Beispiel Testosteron, Cortisol oder Estradiol) und wurde bis jetzt unter anderem auf seine Funktion als Biomarker für Arachidonsäure und auf seine antioxidative Kapazität untersucht.

Bei Blutproben kommen sowohl Serum als auch Plasma bei der Erforschung menschlicher Metabolite zum Einsatz [GIBNEY et al., 2005]. Serum ist die klare Flüssigkeit, welche sich bildet, wenn Blut gerinnt. Um Plasma zu erhalten müssen Antikoagulate zugegeben werden, um die zellulären Komponenten mittels Zentrifuge zu trennen. Für NMR-Analysen von Plasma ist Lithium-Heparinat zu bevorzugen, da es im Gegensatz zu EDTA (Ethylendiamintetraacetat) oder Natriumcitrat die NMR-Signale nicht beeinflusst.

Blut sollten in sauberen Röhrchen gesammelt werden. Wichtige Parameter wie Gerinnungszeit, Temperatur zur Zeit der Entnahme, Art der Ampulle, Antikoagulat, Zeit und Geschwindigkeit bei der Zentrifugation, Transport und Lagerung

müssen bekannt sein. Serum- und Plasmaproben sollten bis zur Analyse bei -80 °C aufbewahrt werden. Zur Aufbewahrung von Teilproben zur späteren Überprüfung der Ergebnisse sollten diese eingefroren werden und zu einem späteren Zeitpunkt auf Eis, in einem kalten Wasserbad oder bei Raumtemperatur aufgetaut werden – keinesfalls jedoch durch Erhitzen [ISSAQ et al., 2009].

Es zeigen sich erhebliche Differenzen zwischen den resultierenden NMR- und MS-Spektren. Der Einsatz von Serum ist jedoch dem von Plasma vorzuziehen, da Plasmaproben Antikoagulantien zur Hemmung der Blutgerinnung beigegeben werden müssen, welche Auswirkungen auf das Stoffwechselprofil haben können [GIBNEY et al., 2005].

Bei einem Vergleich der Kapazitäten von Urin und Plasma im Hinblick auf die enthaltenen Metabolite ist zu erwähnen, dass im Urin neben den Nährstoffen eine höhere Konzentration weiterer Stoffe aus der Nahrung gefunden werden kann, als im Plasma, da diese Stoffe mit dem Urin ausgeschieden werden. Ein weiterer Unterschied besteht darin, dass fettlösliche Verbindungen im Plasma – nicht aber im Urin – enthalten sind.

Trotzdem hat Urin im Bereich der Metabolomik bei Pharmazie und Toxikologie – und gegebenenfalls auch in der Ernährungswissenschaft – einen sehr hohen Stellenwert [GIBNEY et al., 2005]. Urinproben sollten in sterilisierten, sauberen bernsteinfarbenen Röhrchen gesammelt werden und bis zur Analyse auf Eis gelagert werden. Bei langfristiger Lagerung sollte die Temperatur -80 °C betragen. Sollte ein Teil der Probe für spätere Analysen zur Bestätigung der Ergebnisse aufbewahrt werden, gibt es auch die Möglichkeit die Urinproben einzufrieren. Dabei ist darauf zu achten, dass sowohl Einfrieren, als auch Auftauen der Proben zu fehlerhaften Ergebnissen führen können. Eingefrorene Urinproben sollen auf Eis oder bei Raumtemperatur analysiert werden. Erhitzen oder das Auftauen in heißem Wasser ist zu unterlassen, da dies schädlich für die enzymatische Aktivität ist und hitzesensitive Bestandteile schädigen kann. Zur Entfernung etwaiger Verunreinigungen werden die Proben nach dem Auftauen zentrifugiert.

Für NMR-Analysen sollten Konservierungsstoffe (zum Beispiel Citrate oder Ascorbinsäure) gemieden werden, da sie die Signale im NMR-Spektrum beeinflussen können [ISSAQ et al., 2009].

Detaillierte Informationen über Probenentnahme, -bearbeitung und -aufbewahrung bei molekularen epidemiologischen Studien bietet der Review von Holland et al., welcher ebenfalls auf notwendige Schritte hinweist, die unternommen werden müssen um den richtigen Umgang mit biologischen Proben von der Abnahme bis hin zur Analyse zu gewährleisten, um so die Qualität der Proben und die Gültigkeit der Ergebnisse sicherzustellen. Dazu zählen unter anderem die Identifizierung geeigneter biologischer Proben, welche nach Möglichkeit keine invasiven Eingriffe erfordern, die Bestimmung des geeigneten Zeitpunktes zur Probenentnahme, auch hinsichtlich der Stabilität der zu erforschenden Biomarker, das genau Dokumentieren der ausgeführten Arbeitsschritte, die Schulung der Mitarbeiter, auch im Hinblick auf die Einhaltung gesetzlicher Vorschriften zur Gewährleistung der Sicherheit im Umgang mit infektiösem Material [HOLLAND et al., 2003].

## **7 Extraktionsverfahren zum Auftrennen der Stoffmenge und zur Gewinnung der zu untersuchenden Metabolite**

Arbeitsschritte zur Vorbereitung des Probenmaterials, wie zum Beispiel Extraktion, Reinigung, Konzentration usw. sind abhängig von der Art der Probe, der Analysemethode und der Art der Fragestellung, also ob bestimmte oder alle Metabolite in der Probe untersucht werden. Die angewendeten Verfahren sollen einfach und die Ergebnisse nachvollziehbar und reproduzierbar sein. Extraktionsverfahren wie Flüssig-Flüssig-Extraktion, Festphasenextraktion, Festphasenmikroextraktion, Hochdruckextraktion, Filtration usw. sollen so gewählt werden, das möglichst viele Metabolite extrahiert werden. Gefrorene Proben sollen auf Eis aufgetaut werden um die Qualität der Probe zu garantieren. Urinproben werden bei 3000 U/min zentrifugiert um Rückstände zu entfernen. Für die MS-Analyse wird der Überstand entfernt und die Extraktion mit Festphasenextraktion oder Flüssig-Flüssig-Extraktion mit geeigneten Lösungsmitteln, abhängig von den zu untersuchenden Verbindungen, durchgeführt. Bei Analysen im Bereich der Metabolomik sind bei der Untersuchung von Urin, Plasma und Serum mehrere Extraktionsverfahren und verschiedene Analysemethoden, auf Grund der unterschiedlichen chemikalischen und physikalischen Eigenschaften, nötig. Verbindungen wie Aminosäuren, Zucker, Lipide, Steroidhormone, Kohlenwasserstoffe usw. erfordern unterschiedliche Lösungsmittel zur Extraktion der Metabolite. Serum- und Plasmaproben benötigen eine Proteinfällung. Dafür können organische Lösungsmittel, Hitze oder Säuren verwendet werden.

Bei der Analyse von Urin mit MS ist es empfehlenswert, Urin mit Urease zu spalten und den Harnstoff zu entfernen, um diverse Störfaktoren, wie zum Beispiel Peak-Verzerrung, zu verringern. Bei der Verwendung von Urease sollte bedacht werden, dass diverse Metabolite im Harnstoff (zum Beispiel Citrate, Succinate, Ascorbate, Tyrosin, Glycerin,...) entfernt oder reduziert werden können. Bei zielgerichteten Analysen werden nichtradioaktive Isotope zugesetzt.

Die Analyse mit NMR ist von zahlreiche Faktoren (Temperatur und pH der Probe, Ionenstärke, Konzentration, Anwesenheit zweiwertiger Metallkationen,...) abhängig. Zur Standardisierung des pH im Urin kann entweder eine Phosphatpufferlösung in D<sub>2</sub>O bei pH 7,0 oder 7,4 hergestellt werden [ISSAQ et al., 2009].

Slupsky et al. führen eine manuelle pH-Einstellung mit NaOH und HCL durch [SLUPSKY et al., 2007]. Zur Referenzierung und Quantifizierung der NMR-Spektren sind interne und externe Standards (zum Beispiel Natriumtrimethylsilylpropionat oder Formiate) erforderlich. Bei Proben in organischen Lösungsmitteln können Tetramethylsilan oder das Lösungsmittel als interne Referenz herangezogen werden.

Serum- und Plasmaproben können mit NMR unter Zugabe von internem Standard und 10 % D<sub>2</sub>O untersucht werden [ISSAQ et al., 2009]. Eine Verdünnung der Proben wird von Beckonert et al. empfohlen, um die Viskosität der Proben zu verringern und dadurch schärfere Peaks zu erhalten. Dadurch wird auch gleichzeitig die Proteinübertragung minimiert [BECKONERT et al., 2007].

Serum und Plasma enthalten katalytisch aktive Proteine, welche vor allem bei langen NMR-Verfahren aus den Proben, durch Ausfällung mit organischen Lösungsmitteln oder Ultrafiltration, entfernt werden sollen. Tizinai et al. haben verschiedene Fällungsmethoden und Ultrafiltration verglichen und festgestellt, dass es bei den meisten Extraktionsverfahren mit Proteinfällung zu einem Verlust von Metaboliten kam, wohingegen die Ultrafiltration im Hinblick auf die Beibehaltung der Metabolitkonzentrationen und die Reproduzierbarkeit der Ergebnisse den Fällungsmethoden überlegen ist [TIZINAI et al., 2008].

Bruce et al. untersuchten vier verschiedene organische Lösungsmittel (Acetonitril, Aceton, Methanol und Ethanol) hinsichtlich ihrer Verwendung zur Präparation von Blutplasma bei Studien im Bereich metabolic profiling. Zwei Lösungen wurden als optimal bezüglich der Anzahl an extrahierten Biomarkern, Datenqualität, Reproduzierbarkeit und Lebensdauer der Säule angesehen: "Methanol/Ethanol" (1:1) und "Methanol/Acetonitril/Aceton" (1:1:1), welche jeweils zu Plasma im Verhältnis 4:1 zugesetzt werden, wobei das Gesamtvolumen 400 µl beträgt. In Bezug auf die Extraktionsdauer und -temperatur im Vortex wurden die besten Resultate bei 15 min und 4 °C erzielt [BRUCE et al., 2009].

Die Verwendung von Methanol zur Extraktion der Metabolite in Plasma und Serum wird in vielen Studien eingesetzt. A et al. verwendeten fünf verschiedene



organischen Lösungsmitteln (Methanol, Ethanol, Acetonitril, Aceton und Chloroform) und konnten belegen, dass die Extraktion mit Methanol besonders effiziente und reproduzierbare Ergebnisse liefert [A et al., 2005].

Want et al. untersuchten mehrere Methoden zur Extraktion von Metaboliten im Serum und stellten ebenfalls fest, dass eine Extraktion mit Methanol die effektivste und direkteste Methode darstellt und außerdem reproduzierbare Ergebnisse für Studien im Bereich der Metabolomik liefert [Want et al., 2006].

Zhang et al. untersuchten Methoden für die Extraktion intrazellulärer Metaboliten in Erythrozyten. Dabei konnte gezeigt werden, dass eine Lösung aus Methanol/Chloroform/Wasser im Verhältnis 700:200:50 (950 µl Gesamtvolumen) am effizientesten ist [ZHANG et al., 2009].

## 8 Chemisch analytische Methoden zur Analyse der Metabolite

Es existieren verschiedene Analysemethoden zur Identifizierung von Metaboliten. Jedoch ist keine dieser Methoden in der Lage alle Arten von Molekülen auf einmal zu erfassen. Dies liegt vor allem an den unterschiedlichen physikalischen und chemischen Eigenschaften der Metabolite, da ein breites Spektrum von polaren Stoffen mit niedrigem Molekulargewicht (Ethanol) bis hin zu polaren Stoffen mit hohem Molekulargewicht (Glucoside), nichtpolaren Lipiden und anorganischen Stoffen zu untersuchen ist [DUNN et al., 2005].

Die Methoden zur Metabolomanalyse unterscheiden sich vor allem in ihrer Sensitivität, der Menge der Metabolite und Proben, die pro Zeiteinheit analysiert werden können, sowie der Vollständigkeit der Metabolite, die bei der Analyse erfasst werden [WECKWERTH und MORGENTHAL, 2005]. Bei Analysen im Bereich der globalen Metabolomik sind meist mehr als eine chromatographische Technik oder ein Trennverfahren nötig. Für zielgerichtete Analysen ist die Frage nach der zu verwendenden Methoden von den jeweiligen zu untersuchenden Metabolitgruppen abhängig. So wird Gaschromatographie gekoppelt mit Massenspektrometrie eingesetzt um stabile, flüchtige Metabolite und Derivate wie Fettsäuren, Steroide, Flavonoide,... zu analysieren. Hochleistungsflüssigkeitschromatographie wird vorwiegend für die Analyse von Lipiden, Peptiden, Nucleotiden, Ionen,... verwendet [ISSAQ et al., 2009].

Zahlreiche Studien belegen, dass die zwei wichtigsten Techniken zur Analyse des Metaboloms die Massenspektrometrie (MS) und die Kernspinresonanzspektroskopie (NMR) sind. Die Vorteile der NMR bestehen darin, dass es sich um eine nichtdestruktive, nichtselektive Methode handelt, welche kostengünstig und schnell ist und nur minimale Probenvorbereitung benötigt. Dagegen ist MS eine viel empfindlichere Methode, welche allerdings bedeutend weniger Konzentrationen nachweisen kann im Vergleich zur NMR. Einen Überblick über die Vor- und Nachteile der jeweiligen Technologie zeigt die nachfolgende Tabelle [GOLDSMITH et al., 2010].

Variable	NMR	MS
Equipment cost	High	High
Maintenance cost	High	High
Per sample cost	Low	High
Reproducibility (within or across labs)	High	Moderate
Identification of unknown metabolites	Takes time	Can be fast
Sensitivity	Lower than MS but approaches micromolar	Higher than NMR
Quantitation	Routine	Difficult
Resolvable metabolites	Possible	Possible
Potential for sample bias	No	Yes
Data analysis automation	Some	Yes

Abbildung 3: Vor- und Nachteile von Kernspinresonanzspektroskopie (NMR) und Massenspektrometrie (MS) [GOLDSMITH et al., 2010]

Law et al. untersuchten mögliche Auswirkungen von grünem Tee auf das Metabolitprofil im Urin. Die Analyse erfolgte sowohl mit GC/MS und LC/MS, als auch mit  $^1\text{H}$  NMR. Dabei konnte unter anderem gezeigt werden, dass sich die Spektren bei der GC/MS, welche vor und nach der Aufnahme von grünem Tee gemacht wurden unterschieden. Demnach kann GC/MS verwendet werden um molekulare Veränderungen im Urin wiederzugeben. Die Spezifität von LC/MS und  $^1\text{H}$  NMR, sowie ihre Potential zur Identifizierung von Biomarkern konnte ebenfalls gezeigt werden [LAW et al., 2008].

### 8.1 Beschreibung des Einsatzes von Massenspektrometrie (MS) sowie deren Vor- und Nachteile in der Metabolomik

Die Vorteile der Massenspektrometrie liegen sicher in der hohen Selektivität und Empfindlichkeit der Methode. Außerdem ermöglicht die MS, auf Grund der quantitativen Analyse, die Möglichkeit eine große Menge an Metaboliten auf einmal zu bestimmen. Eine Bestimmung der Summenformel durch genaue Messung der Masse ist ebenfalls möglich.

Nachteilig kann sich die Notwendigkeit der Präparation der Probe vor Beginn der Analyse auswirken, welche einen möglichen Verlust von Metaboliten zur

Folge haben kann. Es kann ebenfalls zur Diskriminierung einzelner Metabolitklassen auf Grund der Ionisierung kommen.

Das Koppeln chromatographischer Verfahren mit MS ermöglicht eine bessere Analyse der sehr komplexen Metabolitproben. Flüssigkeitschromatographie gekoppelt mit Massenspektrometrie (LC/MS) wird bei der Analyse von Metaboliten mit niedrigem Molekulargewicht verwendet. Gaschromatographie gekoppelt mit Massenspektrometrie (GC/MS), welche zur Analyse von flüchtigen Stoffen eingesetzt wird hat den Nachteil, dass die meisten Körperflüssigkeiten und Gewebe flüchtig sind [GOLDSMITH et al., 2010]. Aufgrund der hohen Sensitivität und der Möglichkeit sehr viele Substanzen auf einmal zu detektieren ist GC/MS dennoch eine der bedeutendsten Methoden zur Analyse von Metaboliten [WECKWERTH und MORGENTHAL, 2005].

Elektrospray-Ionisation und APCI (atmospheric pressure chemical ionization) sind zwei Ionisierungsverfahren, welche am häufigsten mit LC/MS verwendet werden, wobei es zur Unterdrückung von Elektronen kommen kann. Dies kann bei der Verwendung von Elektrospray-Ionisation in Verbindung mit GC/MS verhindert werden [ISSAQ et al., 2009].

### **8.1.1 Kopplung von Gaschromatographie mit Massenspektrometrie (GC/MS)**

GC/MS ermöglicht die Analyse von flüchtigen, thermisch stabilen polaren und unpolaren Metaboliten. Zur Analyse von flüchtigen Metaboliten ist bei Verwendung der GC/MS keine Derivatisierung nötig. Diese Proben können meist direkt, ohne weitere Probenvorbereitung analysiert werden.

Im Gegensatz dazu ist bei nichtflüchtigen Metaboliten eine Derivatisierung erforderlich, um flüchtige, thermisch stabile Stoffe zu erhalten. Die Analyse von nichtflüchtigen Metaboliten, wie zum Beispiel Aminosäuren, Zucker, phosphorylierte Substanzen, Amine, Alkohole oder Lipide, wird vor allem beim metabolic profiling angewandt. Die Probenvorbereitung ist aufwendiger, vor al-

lem weil die Probe getrocknet werden muss. Dieser Schritt kann mit einem möglichen Verlust von flüchtigen Metaboliten verbunden sein.

Die Derivatisierung umfasst meist eine Oximierung gefolgt von einer Trimethylsilylierung. Dabei werden aktive Wasserstoffe der funktionellen Gruppen durch weniger polare Trimethylsilylgruppen ersetzt. Die Derivatisierung hat aber auch die Vorteile, dass sie effizient, quantitativ und reproduzierbar ist.

Im Anschluss werden die einzelnen Komponenten im Massenspektrometer durch Elektronenstoßionisation oder chemische Ionisation ionisiert. Die im Analysator gemessenen Fragmentgrößen werden als Intensitäten im Computer aufgezeichnet. Identifikation und Nachweis erfolgen anhand der Retentionszeit, als auch durch charakteristische Fragmentationen [DUNN et al., 2005].

Ein wesentlicher Vorteil der GC/MS mit Elektronenstoßionisation ist, dass hier bereits zahlreiche Massenspektren zum Vergleich zur Verfügung stehen [SHULAEV, 2006].

A et al. untersuchten zum Beispiel Metabolite im menschliches Blutplasma mit Hilfe von GC/MS und zeigten so die mögliche Nutzung der Methode zur Identifizierung von Biomarkern auf [A et al., 2005].

### **8.1.2 Kopplung von Flüssigkeitschromatographie beziehungsweise**

#### **Hochleistungsflüssigkeitschromatographie mit Massenspektrometrie (LC/MS beziehungsweise HPLC/MS)**

LC/MS zeichnet sich durch hohe Sensitivität aus. Die zu analysierenden Stoffe werden nach Polarität und Masse getrennt. Flüssigkeitschromatographie und im speziellen Hochleistungsflüssigkeitschromatographie bieten eine hohe chromatographische Auflösung und analytische Flexibilität, da sie für die Analyse von spezifischen Metaboliten oder einer beziehungsweise mehrerer Substanzklassen verwendet werden können. Ein Vorteil gegenüber der GC/MS besteht darin, dass meist keine Derivatisierung erforderlich ist.

Nachteilig ist, dass LC/MS keine Methode zur zielgerichteten Analyse ist und es an bereits vorhandenen Daten mangelt, um die erhaltenen Massenspektren zu vergleichen [SHULAEV, 2006].

Speziell bei der HPLC/MS bestehen die Nachteile sowohl im hohen Druck, als auch in der hohen Temperatur, welche bei der Analyse auf die Stoffe einwirken. Auf Grund der minimalen Probenvorbereitung und der großen Anzahl an analysierten Metaboliten kann die HPLC/MS für die Identifikation von Biomarkern oder zum Vergleich von Stoffwechselprofilen bei Gesunden und Kranken herangezogen werden [DUNN et al., 2005].

Mittlerweile gibt es zahlreiche Weiterentwicklungen des Verfahrens, zum Beispiel die Verwendung von der Elektrospray-Ionisation (ESI) oder Ultra Performance Flüssigkeitschromatographie (UPLC). Ionenfallen-Massenspektrometer können zur Strukturbestimmung unbekannter Verbindungen beitragen [SHULAEV, 2006].

Sana et al. versuchten einen möglichen Standard im Hinblick auf Extraktion und metabolic profiling von Erythrozyten unter Verwendung von LC/MS zu entwickeln [SANA et al., 2008].

Studien, welche sich mit der Untersuchung von menschlichem Blutplasma unter Verwendung von UPLC beschäftigten, wurden von Bruce et al. durchgeführt [BRUCE et al., 2008; BRUCE et al., 2009].

### **8.1.3 Kopplung von Kapillarelektrophorese mit Massenspektrometrie (CE/MS)**

Durch CE/MS erfolgt eine chromatographische Trennung von Ionen in einer flüssigen Phase unter Einfluss eines elektrischen Feldes [DUNN et al., 2005].

CE/MS bietet einige Vorteile gegenüber GC/MS und LC/MS. Zum Beispiel ein sehr hohes chromatographisches Auflösungsvermögen, geringe Probenanforderungen und eine kurze Analysedauer. Des Weiteren ist meist keine Derivatisierung der Proben nötig und es können Kationen, Anionen und ungeladene Moleküle in einem einzigen Analyseschritt getrennt werden.

Kapillarelektrophorese kann sowohl für zielgerichtete Analysen, als auch für globale Analysen (zum Beispiel für anorganische Ionen, organische Säuren,

Aminosäuren, Nukleotide und Nukleoside, Vitamine, Thiole, Kohlenhydrate und Peptide) verwendet werden.

Nachteile der Methoden sind sicherlich die mangelnde Verfügbarkeit kommerzieller Daten zum Abgleich und die Schwierigkeiten der Reproduzierbarkeit von Retentionszeiten [SHULAEV, 2006].

Sogat et al. studierten die quantitative und qualitative Analyse von Aminosäuren im menschlichen Urin mit Hilfe von Kapillarelektrophorese gekoppelt mit Tandem-Massenspektrometrie [SOGAT et al., 2004].

Ramautar et al. analysierten Metabolite in Körperflüssigkeiten (Liquor, Urin, Plasma) unter Verwendung von Kapillarelektrophorese mit nicht kovalent beschichteten Kapillaren [RAMAUTAR et al., 2008].

## **8.2 Beschreibung des Einsatzes von Kernspinresonanzspektroskopie (NMR), sowie deren Vor- und Nachteile in der Metabolomik**

Die Kernspinresonanzspektroskopie bietet detaillierte Informationen über die Molekularstruktur des zu untersuchenden Stoffes. Sie basiert auf dem Eigendrehimpuls („Spin“) von Atomkernen. Wird ein Magnetfeld angelegt, richten sich die Spins danach aus. Die zu untersuchende Probe wird im Magnetfeld einem Radiofrequenzpuls ausgesetzt, wodurch ein höheres Energieniveau erreicht wird. Wird die Radiofrequenz ausgeschaltet, kehrt der Spin wieder auf sein ursprüngliches, niedrigeres Energieniveau zurück. Während dieses Vorganges wird ein NMR-Signal erzeugt, welches mit Hilfe von Computern in Form von Spektren aufgezeichnet und analysiert wird. Jedes Molekül liefert ein individuelles Spektrum, wobei der Bereich unter dem Peak die Konzentration anzeigt. In der Metabolomik dominieren  $^1\text{H}$ -Spektren, da  $^1\text{H}$  ein äußerst sensibles und nicht radioaktives Nuklid ist [GOLDSMITH et al., 2010].

Einer der Vorteile dieser Methode ist, dass es sich um eine nichtdestruktive, spezifische, nichtselektive Technik handelt. So liefert jedes NMR-Spektrum ein spezifisches Spektrum für die jeweilige Verbindung und zahlreiche Informationen über die strukturellen Eigenschaften [DUNN et al., 2005]. Da schon zahlrei-

che Metabolite mit Hilfe von NMR-Spektren analysiert wurden, existieren viele Vorlagen, welche zur Interpretation neuer Spektren herangezogen werden können.

Spektren von Körperflüssigkeiten (zum Beispiel Urin oder Plasma) erzeugen eine große Menge an Signalen, was die Notwendigkeit von Datenreduktion und Mustererkennung nach sich zieht. Schwierigkeiten zeigen sich unter anderem in der Interpretation der Daten. Da Wasser einer der Hauptbestandteile vieler Körperflüssigkeiten ist, ergibt sich ein enormer Peak, welcher Informationen von anderen Molekülen verdecken kann. Techniken zur Unterdrückung dieser Störungen können bei diesem Problem behilflich sein [GOLDSMITH et al., 2010]. Einsatz findet die Kernspinresonanzspektroskopie sowohl beim metabolic fingerprinting als auch beim metabolic profiling [SHULAEV, 2006].

Es gibt bereits eine Vielzahl von Studien, welche sich mit der Analyse von Metaboliten, unter der Verwendung von NMR, im Bereich Ernährung beschäftigt haben [LENZ et al., 2004; WALSH et al., 2006; SOLANKY et al., 2003; BERTRAM et al., 2009].

### **8.3 Beschreibung des Einsatzes von Schwingungsspektroskopie, sowie deren Vor- und Nachteile in der Metabolomik**

Zur Schwingungsspektroskopie zählen die Infrarotspektroskopie (IR), im speziellen die Fourier-Transformations-Infrarotspektroskopie (FTIR), und die Ramanspektroskopie, welche beide im Bereich metabolic fingerprinting eingesetzt werden.

Gemessen werden die Schwingungen und Rotationen von funktionellen Gruppen, welche durch Interaktion von Strahlung mit der Probe entstehen. Es kommt zur Anregung der Energiezustände in den Molekülen, welche von der jeweilig verwendeten Wellenlänge (bei der IR der Infrarotbereich, bei der Ramanspektroskopie monochromatisches Licht mit einer Wellenlänge im sichtbaren oder UV-Bereich) abhängig sind. Gemessen werden die verschiedenen Arten Energiezuständen, welche auf Grund von Änderungen der molekularen Konfiguration oder Änderungen der Elektronenverteilung im Molekül entstehen.



Die Vorteile der beiden Techniken sind eine schnelle – sowohl in Bezug auf die Probenvorbereitung, als auch auf die Analysedauer – (besonders bei FTIR), kostengünstige, reagenzfreie, nichtdestruktive Analyse komplexer biologischer Proben [DUNN et al., 2005].

Nachteile, im Speziellen bei der FTIR, bestehen darin, dass mehr als ein Peak pro Bestandteil auftreten können, die Identifikation von Metaboliten fast unmöglich ist und die Proben getrocknet werden müssen [SHULAEV, 2006].

#### **8.4 Weitere Analysemethoden, welche im Bereich Metabolomik eingesetzt werden**

Neben den oben erwähnten Technologien gibt es eine Reihe weiterer Methoden, welche vielfach erst auf ihre Verwendbarkeit im Bereich Metabolomik getestet werden müssen und hier kurz erwähnt werden sollen. Im Allgemeinen lässt sich ein Trend in die Richtung Kombination verschiedener Techniken oder mehrerer Detektoren erkennen, um eine noch exaktere und quantitativ hochwertigere Identifizierung der Metabolite zu erreichen.

Mögliche Techniken, die zum Einsatz kommen sind zum Beispiel die Verwendung von Dünnschichtchromatographie, die Kombination von HPLC mit UV-, Photodiodenzeile (PDA) oder elektrochemischen Detektoren und enzymatische Nachweisverfahren [SHULAEV, 2006].

## **9 Computergestützte Verfahren zur Roh-Daten-Erfassung der umfangreichen Metabolit-Daten**

Die enormen Datenmengen aus metabolischen Experimenten müssen richtig erfasst und bearbeitet werden, um qualitative und quantitative korrekte Daten zu liefern, welche als Grundlage für die Interpretation der Ergebnisse dienen. Die Bearbeitung der Daten in der Metabolomik lässt sich grob in die Datenverarbeitung und die Datenanalyse unterteilen. Die Datenverarbeitung beinhaltet die grundlegende Bearbeitung der Rohdaten mit Signalverarbeitungsmethoden und der Kombination der Daten mit den Messergebnissen. Dadurch wird die anschließende Analyse der Rohdaten erleichtert. Bei der Datenanalyse erfolgt die Analyse und Interpretation der Daten, meist mit Hilfe multivariater Analysemethoden.

### **9.1 Datenverarbeitung von MS-Daten**

Die Ausgangsdaten bei der massenspektrometrischen Untersuchung von Metaboliten sind Histogramme, also der Aufzeichnung der vom Detektor erfassten ionisierten Moleküle innerhalb eines bestimmten Zeitintervalls. Ein Histogramm zeigt das Verhältnis von Masse zu Ladung und die Intensität an. Ziel bei der Datenverarbeitung ist es diese Rohdaten so darzustellen, dass es möglich ist die charakteristischen Merkmalen (Masse/Ladung, Retentionszeit und Intensität) der Ionen zu erfassen.

Da verschiedene Hersteller von Instrumenten unterschiedliche Datenformate verwenden gibt es mittlerweile Software, welche die Daten in gebräuchliche Datenformate wie netCDF oder mzXML umwandelt.

Die erforderlichen Schritte bei der Datenverarbeitung beinhalten Filterung (Beseitigung von Störeffekten wie Rauschen oder Basislinienmessung), Mustererkennung (Datendarstellung der gemessenen Ionen aus den Rohdatensignalen), Ausrichtung (Gruppierung der Messungen verschiedener Proben) und Normalisierung (Entfernung unerwünschter systematischer Variationen innerhalb der Proben).

Danach können die so verarbeiteten Daten analysiert werden. Eine Übersicht über die notwendigen Arbeitsschritte bei der Datenverarbeitung von Metabolitdaten zeigt die nachfolgende Abbildung [KATAJAMAA und ORESIC, 2007].

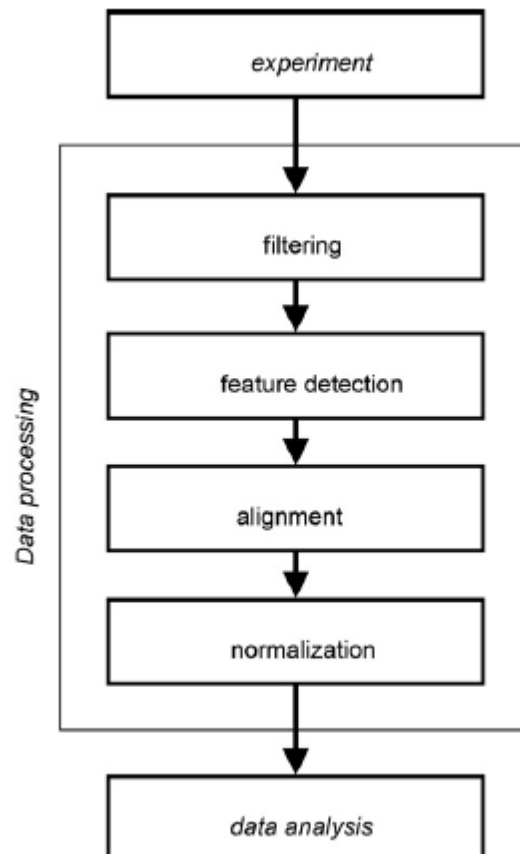


Abbildung 4: Arbeitsschritte bei der Datenverarbeitung von MS-Daten [KATAJAMAA und ORESIC, 2007]

### 9.1.1 Rohdaten Vorverarbeitung und Filterung

Häufig wird zuerst eine Zentrierung der Daten vorgenommen. Das bedeutet, dass jedes Histogramm getrennt bearbeitet wird und mehrere Datenpunkte des gleichen Peaks in einem einzigen Datenpunkt (mit einem Masse/Ladung- und Intensitätswert) zusammengefasst werden. Dadurch werden die Dateien kleiner und deren Management erleichtert.

Andererseits können dadurch auch Störeffekte entstehen, welche die Abschätzung der Daten und die Merkmalerkennung erschweren. Störeffekte auf Grund von chemischen Stoffen (zum Beispiel Lösungsmittelmoleküle) verursachen eine Verschiebung der Basislinie der Massen in den Spektren. Dieser Effekt kann durch die Ermittlung der ursprünglichen Basislinie und deren abziehen aus dem Rohsignal entfernt werden. Störeffekte wie Rauschen werden vor allem durch Detektoren verursacht und müssen aus den Messsignalen entfernt werden. Dies geschieht in der Regel durch Signalverarbeitungstechniken wie Medianfilter oder Wavelet-Transformation [KATAJAMAA und ORESIC, 2007].

### 9.1.2 Mustererkennung

Ziele sind die Erkennung aller Ionensignale und präzise quantitative Informationen über die Ionenkonzentrationen. Zurzeit existieren drei Ansätze zur Lösung, welche allerdings noch keine perfekten Ergebnisse liefern.

Der erste Ansatz beschäftigt sich mit einer Peakdetektion, welche unabhängig von Masse/Ladung und Retentionszeit ist. Der zweite Ansatz schneidet Daten zu extrahierten Ionen Chromatogrammen, wobei jedes eine kleinen Masse/Ladungsbereich abdeckt. Der dritte Ansatz zur Mustererkennung beschäftigt damit, die Datenmodelle mit den Rohdaten abzugleichen.

Mit sanften Ionisationstechniken wie Elektrosprayionisation können in der Regel einzelne Peaks erzeugt werden, generell ist dies jedoch nicht der Fall. Mehrere Ionen können verschiedenen Fragmenten desselben Moleküls entsprechen, was zu einem Problem bei quantitativen Analysen führt. Methoden zur Dekonvolution sind dann erforderlich, um verschiedene Ionen denselben Metaboliten zuzuordnen. Dekonvolution wird zum Beispiel häufig in der Verarbeitung von GC/MS-Daten eingesetzt. Im Bereich der Metabolomik stellt sich hierbei das Problem, dass auf Grund der komplexen Daten es zu einer großen Anzahl an überlappenden Peaks mit ähnlicher Retentionszeit und überlappenden Isotopenmustern kommt. Das führt dazu, dass es schwierig wird zwischen systematische Fehler von biologischer Variabilität zu unterscheiden – was für eine erfolgreiche Dekonvolution nötig ist [KATAJAMAA und ORESIC, 2007].

### 9.1.3 Ausrichtung

Hierbei kommt es zu einer Korrektur unterschiedlicher Retentionszeiten zwischen Chromatographiedurchläufen und einer Kombination unterschiedlicher Probandaten. Meist geschieht dies durch paarweise Angleichung, wobei zu beachten ist, dass die Wahl der Referenzprobe Auswirkungen auf das Resultat hat.

Eine Möglichkeit besteht darin, anhand der Rohdaten eine Reihe von Abbildungen zu erstellen, welche die Retentionszeiten jedes Durchlaufes zu einer gemeinsamen Retentionszeitachse verwandeln [KATAJAMAA und ORESIC, 2007]. Ein bekanntes Beispiel ist die correlation optimized warping-Methode (kurz COW), welche der Retentionszeitachse eines gesamten Ionenchromatogramms einem anderen angleicht. Diese Methode hat den Vorteil, dass keine Datenvorverarbeitung, wie zum Beispiel Peakerkennung, nötig ist. [NIELSEN et al., 1998].

Eine andere Methode ist die Gruppierung detektierter Eigenschaften und das Erzeugen einer Matrix, in der jede Zeile einer Gruppe (zum Beispiel einem Ion) und jede Spalte einigen Messergebnisse (zum Beispiel der Peakfläche) der jeweiligen Probe darstellen. Ausrichtungsmethoden, welche die Merkmale gruppieren, ohne die Retentionszeit zu korrigieren, erfordern eine hohe Reproduzierbarkeit der Chromatographie, da sie nur kleine Variationen in den Retentionszeiten ableiten.

Eine Kombination beider oben erwähnter Methoden ist auch möglich. Die Wahl der jeweiligen Methode beeinflusst die Auswahl der nachfolgenden Datenanalyse. So ist bei der ersten Methode ein Vergleich der Rohdatensignale erforderlich um Unterschiede zwischen den Proben zu identifizieren. Bei der zweiten Methode erfolgt meist eine multivariate Datenanalyse.

Wenn detektierte Merkmale in der Probe sofort identifiziert werden können, kann auf die Ausrichtung verzichtet, beziehungsweise die Ausrichtung erleichtert werden [KATAJAMAA und ORESIC, 2007].

#### 9.1.4 Normalisierung

Ziel ist eine Beseitigung unerwünschter systematischer Verzerrungen von Ionenintensitäten zwischen den Messungen. Dabei muss darauf geachtet werden, Daten über biologische Variationen zu erhalten, da diese von Interesse sind.

Es gibt zwei strategische Ansätze zur Normalisierung von metabolischen Daten. Einerseits die Verwendung von statistischen Modellen um optimale Skalierungsfaktoren für jede Probe, basierend auf den kompletten Datensets, zu erlangen (zum Beispiel durch Normalisierung des Medians oder der Intensitäten). Ein Nachteil beim Verwenden von statistischen Ansätzen ist das Fehlen eines absoluten Referenzwertes verschiedener Metabolitkonzentrationen. Außerdem hat das Einschränken von Daten auf eine spezifische Norm, basierend auf dem Gesamtsignal, Auswirkungen auf die Kovarianzstruktur, was mit Vorsicht in Bezug auf die Verwendung multivariater Datenanalyse zu betrachten ist.

Andererseits die Normalisierung von einem einzigen oder mehreren internen oder externen Standardverbindungen basierend auf empirischen Regeln (zum Beispiel spezifischen Bereichen von Retentionszeiten). Quantitative Analysemethoden verwenden meist standardisierte Isotopen für jede Metabolitmessung. Im Bereich metabolic profiling ist dieser Ansatz jedoch nicht durchführbar, da die Anzahl der Metabolite zu groß ist und die Verfügbarkeit von Referenzisotopen sehr limitiert ist [KATAJAMAA und ORESIC, 2007].

Einen neuen Ansatz bietet NOMIS (Normalization using Optimal selection of Multiple Internal Standards) und zwar eine Normalisierung mit optimaler Auswahl mehrerer interner Standards. Dabei wird eine optimale Zuordnung von internen und externen Standardpeaks für jeden detektierten Peak in der Probe einer reproduzierten Studie erlernt. Durch kontrollieren mehrerer Standardverbindungen über mehrere Probendurchläufe soll festgestellt werden wie die Standards korrelieren, welche Variationen spezifisch für den jeweiligen Standard sind, und welche Variationsmuster von den gemessenen Metaboliten und Standards gemeinsam genutzt werden, so dass diese entfernt werden können [SYSI-AHO et al., 2007].

## 9.2 Datenverarbeitung von NMR-Daten

$^1\text{H}$ -NMR-Daten können durch Binning, Verwendung der Rohdaten oder spektrale Anpassung vorbearbeitet werden.

Bei der zielgerichteten Analyse werden bereits bekannte Metabolite anhand einer Referenzdatenbank, mit Hilfe der dort enthalten Spektren, abgeglichen und absolute Metabolitkonzentrationen bestimmt [ISSAQ et al., 2009]. Nach Vollerendeter Identifizierung und Quantifizierung können statistische Methoden oder Mustererkennungstechniken zur Interpretation der Daten verwendet werden.

Meist ist es erforderlich, dass bei der Arbeit mit NMR die Proben einen bestimmten pH-Wert und/ oder eine bestimmte Temperatur aufweisen. Dies erfordert die Verwendung von Software zur Kurvenanpassung und die Verfügbarkeit von Datenbanken für NMR-Spektren, welche Metabolite mit unterschiedlichen pH-Werten und Frequenzbereichen enthalten [WISHART, 2008]. Ein Nachteil der spektralen Anpassung ist die Limitierung der Suche auf bereits bekannte Spektren [ISSAQ et al., 2009]. Ein Vorteil der NMR im Bereich der quantitativen Metabolomik ist, dass die meisten Metabolite eine einzigartige oder charakteristische chemische Verschiebung aufweisen. Dies reduziert das Problem der spektralen Redundanz, da es unwahrscheinlich ist, dass zwei Verbindungen die gleiche Anzahl an Peaks, gleiche chemische Verschiebung, Peakintensitäten, Spin-Kopplung oder Linienformen haben [WISHART, 2008].

Um Konzentrationsunterschiede zwischen den Proben (vor allem bei Urinproben) auszugleichen werden NMR-Daten in der Regel normalisiert [ISSAQ et al., 2009]. Torgrip et al. präsentieren einen neuen Ansatz zur Normalisierung von NMR-Daten: histogram matching. Mit dieser Methode sollen durch die Normalisierung verursachte Varianzen reduziert werden [TORGRIP et al., 2008].

Normalisierte Daten werden anschließend skaliert, um dynamische Bereiche der Metabolitkonzentrationen auszugleichen und so statistische Abweichungen zu vermeiden. Methoden für die Skalierung umfassen Kombinationen von Mittelwert-Zentrierung, Varianz, Paretoprinzip, logarithmische Transformation, u. s. w. [ISSAQ et al., 2009].

## 10 Verfahren für die statistische Datenanalyse der Metabolitdaten

Die enorm großen Datenmengen der Metabolitdaten können an Hand von statistischen Verfahren oder Algorithmen aus dem Bereich maschinelles Lernen oder evolutionäre Algorithmen analysiert werden.

Eine Schwierigkeit hierbei besteht darin, dass pro Experiment viele tausende Variablen im Verhältnis zu einer relativ kleinen Anzahl an Proben anfallen. Daher ist es wichtig, dass die Anzahl der Variablen reduziert wird [SHULAEV, 2006].

Es gibt verschiedene mehrdimensionale und multivariate statistische Analyseverfahren und Programme zur Mustererkennung, welche alle diverse Vor- und Nachteile zur Folge haben. Multivariate statistische Analyse ermöglicht die Identifizierung von relevanten Metaboliten [ISSAQ et al., 2009]. Dabei geht es um die Feststellung von Varianz und Kovarianz, wobei die später noch genauer erwähnten Hauptkomponentenanalyse und Pfadanalyse dem Prinzip der Kovarianz entsprechen [KEMSLEY et al., 2007].

Bei der Analyse von GC/MS- und HPLC/MS-Daten sind Retentionszeit, Peak-Intensität und das Verhältnis von Masse zu Ladung die benötigten Parameter. Bei NMR-Daten sind entweder chemische Verschiebung und bin Integration oder die Konzentrationen der Metaboliten erforderlich [ISSAQ et al., 2009].

Prinzipiell werden 2 Methoden unterschieden: supervised (überwachte) und unsupervised (unüberwachte) Methoden. Zu den supervised Methoden zählen zum Beispiel Varianzanalyse, Pfadanalyse oder Diskriminanzanalyse. Zu den unsupervised Methoden zählen zum Beispiel Hauptkomponentenanalyse und Clusterverfahren (hierarchische Clusterverfahren, Kohonenkarten) [SHULAEV, 2006].

Supervised Methoden sind leistungsfähiger als unsupervised Methoden, da sie vor allem Hauptkomponentenanalyse und Clusteringverfahren verwenden, welche sich beide mit der Varianz beschäftigen [BROWN et al., 2005].

Softwarepakete, stellen meist mehrere statistische Verfahren zur Verfügung. Beispiele dafür sind R Projekt (<http://www.r-project.org/>), welches Hauptkompo-



nentenanalyse, Singulärwertzerlegung und multidimensionale Skalierung anbietet, LNKnet (<http://www.ll.mit.edu/mission/communications/ist/lnknet/index.html>), welches Clusteringmethoden und supervised Algorithmen anbietet, oder Octave (<http://www.gnu.org/software/octave/>), eine frei verfügbare Version von Matlab [ARITA, 2004].

## **10.1 Supervised Methods– Überwachte Methoden**

Supervised Methoden, wie Regressionsverfahren oder neuronale Netze, benötigen eine bereits identifizierte Teilmenge, an Hand derer eine Kalibrierung vorgenommen werden kann. Diese Methoden haben den Vorteil, dass die Einstufung auf einem gewissen Vorwissen basiert. Ein Nachteil ist zum Beispiel, dass zusätzliche Daten für die Kalibrierung erforderlich sind [MENDES, 2002].

Supervised Methoden werden daher verwendet, wenn Informationen über die zu klassifizierenden Daten vorhanden sind. Dadurch ergibt sich die Möglichkeit dem Algorithmus ein Modell über die Ein- und Ausgaben anzutrainieren. Ein typisches Beispiel hierfür im Bereich der Metabolomik ist die Untersuchung von zwei unterschiedlichen Klassen von Proben – denen von gesunden Patienten im Vergleich zu denen von erkrankten Patienten. Dabei werden Biomarker bestimmt, indem die Proben in gesunde und erkrankte Gruppen unterteilt werden [BROWN et al., 2005].

### **10.1.1 Pfadanalyse**

Das Prinzip der Pfadanalyse ist dem der Hauptkomponentenanalyse ähnlich. Der Unterschied besteht vor allem darin, dass bei der Pfadanalyse die Identitäten der Klassen von vornherein bekannt sind [WISHART, 2007]. Anstatt nur die dominanten Merkmale der Daten darzulegen, wie bei der Hauptkomponentenanalyse, wird bei der Pfadanalyse eine zweite Matrix mit Informationen über die abhängigen Variablen für jede Probe angefertigt. Diese zusätzlichen Informationen reduzieren den Einfluss irrelevanter Daten [KEMSLEY et al., 2007]. Der

Algorithmus maximiert die Kovarianz zwischen den Testvariablen und den abhängigen Variablen [WISHART, 2007].

Ein Problem, welches sich bei der Analyse ergeben kann ist das der Überanpassung. Sie entsteht wenn viel mehr Variablen als Proben vorhanden sind – was eben vor allem bei Metabolitdaten der Fall ist. Problematisch wird es dann, wenn die Überanpassung nicht offensichtlich ist und somit die Ergebnisse verzerrt werden [KEMSLEY et al., 2007].

PLS-DA (Partial least squares – Discriminant Analysis) ist eine Sonderform der Pfadanalyse. PLS-DA wird zur genaueren Separation zwischen den zu beobachteten Klassen verwendet, um zu ermitteln welche Variablen die spezifischen Informationen beinhalten [WISHART, 2007].

### **10.1.2 Varianzanalyse**

Die Varianzanalyse wird verwendet, um die gesamten Daten je nach Varianz zu unterteilen. Ziel ist es, die unterschiedlichen Variationen aufzutrennen und sie spezifischen Faktoren zuzuordnen. Dies geschieht durch Aufteilung der Variationen in orthogonale Daten, welche die bekannten Faktoren repräsentieren, und unabhängige Daten, welche nicht den bekannten Faktoren zuordenbar sind (wie zum Beispiel Messfehler, biologische Variabilität oder Batcheffekte).

Zu erwähnen ist, dass zwischen der univariaten und der multivariaten Varianzanalyse unterschieden werden kann. Bei der univariaten Varianzanalyse wird jeweils eine Variable analysiert. In Kombination mit anderen statistischen Verfahren, etwa der Hauptkomponentenanalyse oder der simultanen Komponentenanalyse, kann eine multivariate Varianzanalyse vorgenommen werden [THISSEN et al., 2009].

### **10.1.3 Künstliche neuronale Netzwerke**

Künstliche neuronale Netzwerke haben ihre Stärke in der Klassifizierung, sind aber schwach in Bezug auf die Darlegung der Grundlagen [BROWN et al., 2005]. Es handelt sich dabei um eine Sammlung identischer, einfacher Algorithmen, welche die jeweilige Zahleneingabe in dazugehörige Ausgabe umwan-

deln. Art und Weise der Umwandlung hängt von der Gewichtigkeit der Verbindung und der Verzerrung ab. Das Ergebnis hängt sowohl von der Stärke der Verbindungen, der Verzerrung und der gewichteten Summe der Eingaben ab. Die Bedeutung von künstlichen neuronalen Netzwerken besteht darin, dass nichtlineare Funktionen über einen Lernalgorithmus durch iterative Vorgehensweise erlernt werden. Stärke und Verzerrung werden so modifiziert, dass das Ergebnis der Netzwerke eine Funktion der Eingaben, entsprechend der erwünschten Ausgaben, ist [MENDES und KELL, 1996].

Support Vector Machines (SVM) zählen zu den supervised Methoden des maschinellen Lernens und werden zur Analyse und Klassifizierung von Metabolitdaten verwendet. Im Gegensatz zur Hauptkomponentenanalyse oder zur PLS-DA können SVM auch nichtlineare Funktionen mit Hilfe eines Kernels bearbeiten. Das Grundprinzip von SVM besteht in einer binären Support Vector Klassifizierung. Anhand von Daten zweier Klassen wird ein optimaler linearer Klassifikator in Form einer Hyperebene gebildet, welcher die meiste Bandbreite beinhaltet. Die Punkte, welche diesen Bereich treffen, werden Support Vectors genannt. Mahadevan et al. haben anhand zweier Fallstudien gezeigt, dass SVM im Hinblick auf die Genauigkeit der Vorhersage bessere Ergebnisse liefert als PLS-DA [MAHADEVAN et al., 2008].

#### **10.1.4 Evolutionäre Algorithmen**

Evolutionäre Algorithmen (dazu zählen genetische Programmierung, genetische Algorithmen, Evolutionsstrategien, evolutionäre Programmierung) werden meist in Kombination mit einer statistischen Analysemethode (zum Beispiel Pfadanalyse oder Diskriminanzanalyse) durchgeführt. Sie beruhen auf den Prinzipien der Evolution und der Selektion der Spezies [SHULAEV, 2006]. Evolutionäre Algorithmen eignen sich vor allem zur Analyse von datenreichen aber gleichzeitig hypothesenarmen Datenmengen. Sie helfen dabei prädiktive Regeln oder Hypothesen aufzustellen und sollen auf diese Weise dabei helfen mögliche Biomarker zu identifizieren [HOLLYWOOD et al., 2006].

Pena-Reyes und Sipper bieten einen umfangreichen Überblick über evolutionäre Algorithmen [PENA-REYES und SIPPER, 2000].

### **10.1.5 Regelinduzierte Algorithmen**

Regelinduzierte Algorithmen zählen zu den kategorischen Algorithmen. Sie produzieren uni- oder multivariate Entscheidungsbäume, welche bei der Entdeckung wichtiger Stoffwechselmetaboliten helfen [HOLLYWOOD et al., 2006].

CART (Classification and Regression Trees) ist ein Algorithmus, welcher einfach zu verstehende Regeln vorgibt, dafür aber nicht so präzise ist [BROWN et al., 2005].

## **10.2 Unsupervised Methods– Unüberwachte Methoden**

Bei unsupervised Methoden, wie Hauptkomponentenanalyse oder Clusterverfahren, sind keine zusätzlichen Kenntnisse über die Proben erforderlich. Die Proben werden in unterschiedliche Klassen unterteilt. Nachteilig ist jedoch, dass es nicht immer einfach ist, eine geeignete Basis für diese Grundeinteilung vorzunehmen [MENDES, 2002].

### **10.2.1 Hauptkomponentenanalyse**

Die Hauptkomponentenanalyse zählt zu den Clusterverfahren und ist eine Methode zur Datenreduktion. Bei der Hauptkomponentenanalyse erfolgt eine lineare Transformation der Datenpunkte, wobei die Eigenschaften der Probe an den Koordinatenachsen aufgelistet werden.

Die Hauptkomponentenanalyse wird vor allem dafür verwendet, um Unterschiede zwischen Proben zu identifizieren. Gleichzeitig wird festgestellt welche Variablen für diesen Unterschied verantwortlich sind und ob diese Variablen korrelieren oder nicht. Des Weiteren wird die Menge an enthaltenen Informationen oder Signalen quantifiziert [WISHART, 2007]. Aus einer großen Menge manifester Variablen werden mit Hilfe der Hauptkomponenten die Informationen in

Gruppen unterteilt, welche an Hand der latenten, zugrundeliegenden Variablen ausgewählt werden. Dies ist vor allem dann nützlich, wenn es sich dabei um sehr ähnliche Daten handelt und die Anzahl der Variablen die Anzahl der Proben übersteigt [KEMSLEY et al., 2007].

Ein Nachteil der Methode besteht darin, dass die Analyse sensitiv auf experimentelle Störfaktoren (Rauschen) ist, da generell alle Daten (eben auch mögliche Störfaktoren) im Endmodell enthalten sind [WISHART, 2007]. Außerdem ist zu beachten, dass bei dieser Methode keine zusätzlichen Informationen, welche eventuell zusätzlich zu den Messungen vorliegen, berücksichtigt werden. Es werden einzig und allein die experimentellen Daten in der Matrix erfasst [KEMSLEY et al., 2007].

Als Methoden zur Datenreduktion erleichtert die Hauptkomponentenanalyse die visuelle und graphische Erkennung von Mustern oder Gruppierungen.

Die Hauptkomponentenanalyse eignet sich für die Analyse von binned NMR-, GC/MS- und HPLC/MS-Daten, da hierbei sowohl Daten über die Höhe, als auch die Lage der Peaks vorhanden sind [WISHART, 2007].

### **10.2.2 Clusterverfahren**

Bei Clusteringalgorithmen werden die Merkmale als Vektoren angezeigt, wobei jeder Vektor einige gemessene Eigenschaften repräsentiert. Ziel ist es, jeden dieser Vektoren einer Gruppe zuzuordnen, so dass die jeweiligen Vektoren in einer Gruppe sich ähnlich sind, wohingegen sich die einzelnen Gruppen voneinander unterscheiden.

Dieses einerseits sehr intuitive Konzept kann in der Praxis Schwierigkeiten mit sich bringen, da nicht immer eine eindeutige Einteilung vorzunehmen ist. Probleme können aber auch bei Experimenten auftreten, bei denen eine eindeutige Aufteilung möglich ist, da bei den meisten Clusterverfahren die Qualität der Einteilung nur von einem Bewertungsmerkmal abhängt. Des Weiteren können Schwierigkeiten hinsichtlich der Bestimmung der Anzahl an Clustern pro Datenmenge auftreten. Wichtig ist daher eine sorgfältige Analyse der Ergebnisse. Ein Problem vieler Clusteringalgorithmen (zum Beispiel k-means-Algorithmen, hierarchische Algorithmen, Kohonenkarten) ist, dass sie eine Aufteilung der

Merkmale vornehmen, ohne eine Abschätzung der Zuverlässigkeit dieser Ergebnisse zu liefern. Ziel ist einzig und allein die Aufteilung aller Merkmale zu einem Cluster – begründet oder unbegründet.

Durch wiederholtes Clustern der Daten, unter Verwendung von randomisierten Algorithmen, verschiedenen Algorithmen oder Resampling der Daten können akzeptable Resultate erzielt werden. Eine andere Möglichkeit besteht darin, die Resultate an Hand interner und externer Größen zu bewerten [BROWN et al., 2005].

## 11 Softwareprogramme zur Datenbearbeitung im Bereich Metabolomik

Die Anforderungen an Softwareprogramme zur Datenverarbeitung im Bereich Metabolomik sind ziemlich einzigartig. Daher bedarf es meist sehr spezieller Programme, welche Wissen aus Chemieinformatik, Bioinformatik und Statistiken kombinieren [XIA et al., 2009].

Auf den Webseiten von MS-Utills (<http://www.ms-utils.org/>) und von Fiehn laboratory ([http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/Peak\\_Alignment/](http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/Peak_Alignment/)) finden sich Auflistung verfügbarer Softwareprogramme. Die zurzeit erhältlichen Softwareprogramme lassen sich in kommerzielle und frei verfügbare Programme unterscheiden. Des Weiteren lassen sich vor allem Unterschiede in Bezug auf die erbrachte Leistung ableiten – entweder sind die Programme auf einen Anwendungsbereich hin spezialisiert, oder sie kombinieren mehrere Prozesse bis hin zur Datenanalyse.

Das Ergebnis der meisten Softwareprogramme ist eine Matrix, welche die Peakintensitäten der detektierten Ionen in den verschiedenen Proben aufzeigt. Diese können anschließend mit verschiedenen statistischen Programmen (vor allem Hauptkomponentenanalyse oder t-Test) bearbeitet werden. Je mehr Funktionen das Programm bietet, desto besser wird ein Überblick über die verarbeiteten Daten ermöglicht, und desto einfach ist es einen Gesamtüberblick von den Rohdaten bis hin zum Ergebnis der Analyse zu erhalten. Vor allem kommerzielle Programme bieten diesen Vorteil.

Im Hinblick auf die Auswahl der Software für die Verarbeitung von Metabolitdaten sind vor allem die Qualität der Verarbeitung, die Benutzerfreundlichkeit, die Leistungsfähigkeit, sowie die Gesamtkosten für die jeweilige Software entscheidend. Im Bezug auch die Benutzerfreundlichkeit sind zum Beispiel Programme mit graphischer Benutzeroberfläche leichter zu handhaben. Bei der Leistungsfähigkeit ist darauf zu achten, dass eine enorme Menge an Rohdaten verarbeitet werden müssen. Das frei verfügbare Softwareprogramm MZmine liefert zum Beispiel die Möglichkeit die Berechnungen auf mehrere Prozessoren oder Computer zu verteilen [KATAJAMAA und ORESIC, 2007].

## **11.1 Beispiele für kommerzielle Software zur Datenbearbeitung im Bereich Metabolomik**

### **11.1.1 BlueFuse for Biomarkers**

BlueFuse for Biomarkers ist eine Softwarelösung, welche umfassende Anwendungen für eine verbesserte Entdeckung und Klassifizierung von möglichen Biomarkern liefert. Sowohl biologische Proben, als auch Rohdaten aus NMR, GC/MS oder LC/MS können als Ausgangsmaterial verwendet werden. Vorteile sind die automatische Qualitätskontrolle, welche während des gesamten Arbeitsablaufes zuverlässige, prüfbare und reproduzierbare Ergebnisse liefert. Anormale Spektren werden sofort automatisch erkannt, geringfügige Mängel korrigiert und unbrauchbare Spektren aus der Analyse ausgeschlossen. Die Daten werden mit Hilfe von speziellen BlueFuse Algorithmen (unter anderem Ausrichtung, Korrektur der Basislinien und Normalisierung) bearbeitet um unerwünschte Effekte (Verzerrungen oder Rauschen) zu entfernen. Die Peakflächen werden unter Verwendung von Algorithmen zur Merkmalerkennung untersucht und anschließend werden die Daten mit univariaten und multivariaten Methoden analysiert. Die Ergebnisse werden von Experten geprüft und in einem Abschlussbericht zusammengefasst [BLUEGENOME, 2008].

### **11.1.2 MarkerLynx**

MarkerLynx ist ein kommerzielles Softwareprogramm, welches bei der Identifizierung von miteinander verknüpften Markern hilft, indem es nach charakteristischen Mustern in LC/MS-Datensätzen sucht. Die Proben werden auf der Grundlage ähnlicher LC/MS-Signale geordnet. Identifizierte Marker können anschließend durch Tandem-Massenspektrometrie oder andere Analysemethoden noch genauer untersucht werden. Die Arbeitsschritte, welche von MarkerLynx durchgeführt werden sind eine automatische Peakerkennung, die Identifizierung der Ionenintensitäten, das Erstellen einer Datenmatrix, in welcher das Masse zu Ladungs-Verhältnis und die Intensität der detektierten Ionen eingetragen werden und die PCA-Analyse. Die Daten, welche nach der PCA-Analyse vorliegen sind mit denen der originalen Chromatogramme und Ionenintensitäten verlinkt.



Ähnliche chromatographische oder massenspektrometrische Daten werden durch die PCA zu Clustern zusammengefasst. Der unmittelbare Übergang von Datenreduktion, -integration und statistischer Analyse der LC/MS-Daten ermöglicht eine schnelle Identifizierung von Biomarkern [WATERS, 2010].

### **11.1.3 MarkerView**

MarkerView™ Software unterstützt, durch spezielle statistische Analysen (zum Beispiel PCA, t-Test,...) und graphische Aufbereitung der MS-Daten, metabolische Studien und die Identifizierung von Biomarker. MarkerView stellt mehrere Funktionen wie chromatographische Erkennung und Peakerkennung oder automatische Ausrichtung von Massen und Retentionszeiten zum genauen Vergleich identischer Verbindungen in verschiedenen Proben zur Verfügung. Eine Verlinkung der Ergebnisse zu den Rohdaten zur Interpretation und zur besseren Identifizierung vermeintlicher Biomarker ist durchführbar. Die Software hilft sowohl bei der Analyse von Experimente mit bereits klassifizierten, als auch mit nicht-klassifizierten Probengruppen oder einer Kombination davon, da sowohl supervised und unsupervised Analysemethoden verwendet werden. Nach der Datenanalyse wird automatisch ein umfangreicher Endbericht erstellt [AB SCIEX, 2010].

### **11.1.4 SIEVE**

Thermo Scientific SIEVE Software ermöglicht eine einfache, automatisierte Analyse von GC/MS- und LC/MS-Daten zur Identifizierung von Biomarkern. Die Software identifiziert unterschiedliche signifikant statistische Häufungen von Proben und ermöglicht so Aussagen über Verbindungen, welche sich in Dosierung, Zeit oder anderen Merkmalen unterscheiden. Sowohl der Vergleich von 2 Proben, als auch die Verarbeitung mehrerer hundert MS-Daten in kurzer Zeit ist möglich. Die zu verarbeiteten, komplexen Daten werden vorgefiltert, worauf hin sich die Anzahl der zu untersuchenden Verbindungen verringert und der Ablauf der Analyse verkürzt wird. Zur Ausrichtung der chromatographischen Daten werden Algorithmen verwendet, welche die Variabilität der Daten minimieren.

Die MS-Intensitäten der Rohdaten werden herangezogen um Unterschiede festzustellen. So wird verhindert, dass Peak-Daten manipuliert werden und wertvolle Daten verloren gehen. Die Identifizierung der Metabolite erfolgt mit Hilfe der öffentlich zugängliche Datenbank ChemSpider. Der gesamte Ablauf kann mit Hilfe von SIEVE Software vollständig automatisiert werden. Dennoch können die einzelnen Schritte jederzeit unterbrochen werden, um zum Beispiel für die Identifizierung Daten einer anderen Datenbank anzeigen zu lassen. Anschließend kann die Analyse von diesem Arbeitsschritt aus fortgesetzt werden, ohne noch einmal von Beginn an Starten zu müssen [VAST SCIENTIFIC, 2009].

## **11.2 Beispiele für frei zugängliche Software zur Datenbearbeitung im Bereich Metabolomik**

### **11.2.1 Automated Mass Spectral Deconvolution and Identification System (AMDIS)**

AMDIS stellt eine zuverlässige und zeitsparende Methode dar, mit deren Hilfe eine automatische Dekonvolution und Identifizierung von Verbindungen aus Chromatogrammen von GC/MS-Daten ermöglicht wird [NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 2010]. Das Programm beinhaltet vier Arbeitsschritte: die Analyse von Störeffekten, der Verbindungsnachweis, die spektrale Dekonvolution und die Identifizierung der Verbindungen. Zuerst werden Störeffekte, wie Rauschen, erkannt und entfernt. Danach werden die einzelnen Verbindungen erfasst und als Peak-Modell aufgezeichnet. Der dritte Schritt extrahiert die "gereinigt" Spektren von den einzelnen Ionenchromatogrammen mit Hilfe der vorher gemachten Peak-Modelle. Der letzte Schritt errechnet Übereinstimmungen für die extrahierten Spektren mit Spektren aus einer Präsenzdatenbank, wobei eine Vielzahl der Informationen, welche bei der Dekonvolution im dritten Schritt erworben wurden mit einbezogen werden [STEIN, 1999].

Die Software ist frei zugänglich und steht unter <http://chemdata.nist.gov/mass-spc/amdis/downloads/> zum Download zur Verfügung. AMDIS wird laufend aktualisiert, wobei Änderungen der letzten Version unter <http://chemdata.nist.gov/mass-spc/amdis/changes.html> abrufbar sind.

### **11.2.2 metaP-Server**

Mit Hilfe dieses frei zugänglichen Programmes (<http://metabolomics.helmholtz-muenchen.de/metap2/>) ist die automatische und standardisierte Analyse von Metabolitdaten möglich. Dabei wird anfangs die Datenqualität überprüft, anschließend Schätzungen über Reproduzierbarkeit und Batcheffekte, Hypothesen- und Korrelationstests für Phänotypen und PCA durchgeführt. Letztendlich werden die erfassten Metaboliten mit den farbigen Stoffwechselwegen der KEGG Datenbank verlinkt. Zur genauen Metabolitidentifizierung sind auch Querverweise zu HMDB, LipidMaps, PubChem und CAS Datenbanken verfügbar. Das Programm beschäftigt sich hauptsächlich mit der Untersuchung und Interpretation von Metabolitdaten (Konzentrationen oder Peaklisten) im Zusammenhang mit multiplen Gruppenklassen und metrischen Daten (zum Beispiel Gewicht oder Alter). Somit kann ein möglicher Einfluss des Phänotyps auf die detektierten Metabolite erfasst werden. metP-Server unterstützt so die Identifizierung von Biomarkern und liefert Daten für epidemiologische und metabolische Studien [KASTENMÜLLER et al., 2011].

### **11.2.3 MetAlign**

MetAlign ist ein frei verfügbares Softwareprogramm ([www.metalalign.nl](http://www.metalalign.nl)), mit dessen Hilfe die großen Datenmengen aus GC/MS- und LC/MS-Daten einfach und automatisch analysiert werden können. Die Konvertierung der Daten in verschiedene Formate (zum Beispiel auch netCDF) ist möglich. Diverse Vorverarbeitungsschritte wie die Korrektur der Basislinien, Rauschunterdrückung oder eine genaue Berechnung der Masse werden, ebenso wie Peakerkennung und Ausrichtung von bis zu 1000 Datensätzen, vorgenommen. Auf Grund der 100 – 1000 fachen Datenreduktion wird die Datenerfassung der umfangreiche

GC/MS- und LC/MS-Daten erleichtert beziehungsweise ermöglicht. Univariate und multivariate statistische Analysemethoden stehen zur Verfügung und die Ergebnisse sind mit den meisten multivariaten Statistikprogrammen kompatibel [LOMMEN, 2009].

#### **11.2.4 MZmine und MZmine 2**

MZmine wurde 2005 als frei verfügbares Softwareprogramm für die einfache Verarbeitung und Visualisierung von LC/MS und GC/MS-Daten eingeführt [PLUSKAL et al., 2010]. Das Programm bietet eine Reihe von Möglichkeiten wie die Bearbeitung der Ausgangsdaten, Spektralfilterung, Peakerkennung, chromatographische Ausrichtung, Normalisierung, Visualisierung und Datenexport. Des Weiteren ist eine automatische Verarbeitung einer großen Anzahl von Spektren möglich. Rohdaten können in Form von NetCDF oder mzXML vorliegen und mit diversen Methoden, zum Beispiel Filterung zur Unterdrückung von Rauschen, bearbeitet werden [KATAJAMAA et al., 2006].

Die neue Version MZmine 2 bietet einige Verbesserungen. Das System wurde flexibler gestaltet und ermöglicht nun eine unkomplizierte Entwicklung neuer Datenverarbeitungsmodule. Die zurzeit bestehenden Datenmodule beschäftigen sich mit dem Import der Rohdaten, der Rohdatenbearbeitung, Peakerkennung, -ausrichtung und -identifizierung, Visualisierung, Normalisierung und statistischer Analyse. Rohdaten können in den Formaten mzML, mzXML, NetCDF und RAW erfasst werden. Zur Unterstützung der Identifizierung ist MZmine 2 mit Onlinedatenbanken wie KEGG, HMDB, METLIN oder PubChem verlinkt. Die graphische Oberfläche vereinfacht und intuitiver gestaltet. Bei jedem Schritt der Datenverarbeitung mit MZmine 2 stehen einem die ursprünglichen Daten auch noch zur Verfügung. Hochauflösende MS-Daten können besser bearbeitet werden und MZmine 2 wurde dahingehend optimiert mit der dadurch erhöhten Datenmenge zurechtzukommen [PLUSKAL et al., 2010].

### 11.2.5 XCMS und XCMS<sup>2</sup>

XCMS und XCMS<sup>2</sup> sind beide frei verfügbar unter <http://metlin.scripps.edu/download/> und sollen vor allem durch Design, Verfügbarkeit und Flexibilität überzeugen. Sie ermöglichen die Datenverarbeitung von LC/MS-Daten in den Formaten NetCDF, mzXML und mzData [<http://metlin.scripps.edu/download/>]. Eigentlich als Programm für die Verarbeitung von Metabolitdaten aus LC/MS gedacht, bietet es auch die Möglichkeit andere Datentypen (zum Beispiel Peptide) oder Instrumenten (zum Beispiel GC/MS) zu bearbeiten. XCMS bietet neuartige nichtlineare Ausrichtungen der Retentionszeiten, Filtration, Peakerkennung und -abstimmung. Mit XCMS können eine Vielzahl von endogenen Metaboliten identifiziert werden. Gleichzeitig wird für jede Probe eine nichtlineare Korrektur der Retentionszeit berechnet, wodurch die relativen Ionenintensitäten der Metabolite direkt miteinander vergleichbar sind. So können Veränderungen in bestimmten endogenen Metaboliten erfasst und potentielle Biomarker identifiziert werden. Außerdem ist es möglich, andere Datenverarbeitungsprogramme zu integrieren, und so die Datenanalyse individuellen Bedürfnissen anzugleichen. XCMS hat des Weiteren den Vorteil, dass es nicht auf die Verwendung von Instrumenten bestimmter Hersteller beschränkt ist. Dadurch wird eine identische, reproduzierbare Datenanalyse unabhängig von Geräteherstellern ermöglicht. Es ermöglicht außerdem für eine Vielzahl komplexer Aufgaben eine automatische Analyse [SMITH et al., 2006]. Eine Erweiterung bietet XCMS<sup>2</sup>, welche die automatische Suche nach Tandem-Massenspektrometrie-Daten unterstützen soll. Dabei werden die Daten der Massen von Fragmentionen der zu analysierten Tandem-Massenspektrometrie-Spektren mit den Referenzspektren aus der Datenbank METLIN verglichen. Es ermöglicht außerdem die Identifizierung von neuartigen Metabolitstrukturen indem struktureller Elemente ermittelt werden, welche möglicherweise charakteristische Fragmentionen darstellen, die in Beziehung mit Referenzverbindungen bei METLIN stehen, auch wenn die Massen der beiden Verbindungen nicht dieselben sind [BENTON et al., 2008].

### 11.2.6 COMparison of SPectrAI Retention Information (COMSPARI)

In der Metabolomik ist einerseits die Identifizierung einzelner Verbindungen, aber auch das Erkennen von Differenzen zwischen Datensätzen verschiedener Proben wichtig, zum Beispiel um die unterschiedliche molekulare Zusammensetzung von Zellen auf Grund genetischer Unterschiede festzustellen. COMSPARI ist eine frei verfügbare Software ([www.biomechanic.org/comspari/](http://www.biomechanic.org/comspari/)), welche einen neuen Ansatz bietet kleinste Unterschiede zwischen Massenspektren oder Chromatogrammen von sehr ähnlichen und dennoch verschiedenen Proben festzustellen. Mögliche Unterschiede bei Intensitäten, welche zeitgleich in Bezug auf Zeit und Masse-Ladungs-Verhältnis auftreten, werden verglichen. Die Daten aus GC/MS oder LC/MS werden jeweils pro einem Datenpaar in einem benutzerfreundlichen und informativen Display in gespiegelter Form wiedergegeben, wobei die Grundlinie sich ungefähr in der Mitte des Displays befindet und jeweils ein Paar ober- beziehungsweise unterhalb dieser Linie angezeigt wird. Diese gespiegelte Anzeige ermöglicht es Unterschiede zwischen den beiden ähnlichen Massenspektren oder Chromatogrammen schnell und intuitiv festzustellen [KATZ et al., 2004].

### 11.2.7 MetaboAnalyst

Hierbei handelt es sich um ein frei zugängliches, web-basiertes Programm um metabolische Daten zu verarbeiten und unter [www.metaboanalyst.ca](http://www.metaboanalyst.ca) aufzurufen. Ein großer Vorteil bei der Verwendung von MetaboAnalyst ist die Möglichkeit, dass sehr viele unterschiedliche Datentypen verarbeitet werden können. So zum Beispiel Daten über Verbindungen oder Konzentrationen, GC/MS und LC/MS Spektren, Peaklisten von NMR- und MS-Daten und binned spektra/ data binning. Die verwendeten Datenformate netCDF, mzXML und mzDATA. Die nötigen Arbeitsschritte bei der Verwendung von MetaboAnalyst sind das Hochladen der Daten, Datenverarbeitung, Normalisierung und die statistische Analyse der Daten. Das Endergebnis stellt eine detaillierte Zusammenfassung dar, welche zum Download zur Verfügung steht. Darin enthalten ist ein ausführlicher Bericht über jede verwendete Methode, ebenso farbige graphische Darstellun-

gen, Informationen über die identifizierten Metabolite und Abbildungen der betreffenden Stoffwechselwege. Das Vorhandensein einer großen Anzahl an Referenzspektren erleichtert die Identifizierung der Metaboliten. Die statistische Analyse erfolgt unter anderem durch fold change analysis, t-Test, PCA, PLS-DA, hierarchische Clusterverfahren und andere maschinellen Lernverfahren. Benutzerfreundliche Bedienbarkeit ist durch eine einfache Schritt-für-Schritt Anweisung gegeben [XIA et al., 2009].

## 12 Ansätze zur Datenspeicherung für den Aufbau von Metabolit-Datenbanken

Metabolit-Datenbanken, insbesondere jene, die sich auf den Bereich der menschlichen Metabolomik beziehen, müssen bestimmte Funktionen erfüllen. Dies liegt vor allem an der schnellen Identifizierung von mehreren Tausenden Metaboliten in kurzer Zeit. Deshalb sollten Metabolomik-Datenbanken nicht nur Informationen über Verbindungen und Reaktionen aufweisen. Wissenschaftler sollten auch die Möglichkeiten haben diese Datenbanken nach Stoffkonzentrationen, Lage (Körperflüssigkeit, Gewebe, subzellulär), physikalische und chemische Eigenschaften, bekannte assoziierte Krankheiten, Nomenklatur, Beschreibung, Daten über Enzyme und Mutationen, NMR- und MS-Spektren zu durchsuchen. Besonders wichtig ist, dass diese Daten schnell verfügbar, validiert, referenziert und leicht auffindbar sind.

Eine Herausforderung bei Spezies-spezifischen Datenbanken ist die Frage, welche Informationen mit einbezogen werden. Sollen nur endogene Metaboliten angeführt werden, oder auch exogene Arzneistoffe und deren Metabolite, pflanzliche Nahrungsbestandteile und chemische Zusatzstoffe? Wo ist die Ober- und Untergrenze in Bezug auf das Molekulargewicht der Metaboliten? Soll die Datenbank nur auf Metabolite beschränkt sein, welche nachweisbar sind [WISHART et al., 2007]?

Theoretisch und teilweise auch praktisch können verschiedene Modelle von Datenbanken unterschieden werden:

- Datenbanken, welche sowohl Metabolitprofile, als auch detaillierte Informationen zu Rohdaten und Metadaten enthalten.
- Datenbanken, welche nur auf eine Spezies spezialisiert sind.
- Datenbanken, welche Daten von experimentellen Ergebnissen unterschiedlichsten Studiendesigns und mehrere Spezies betreffend beinhalten.
- Datenbanken, welche Informationen über alle Metaboliten in jeder biologischen Spezies enthalten.



- Datenbanken, welche biochemische Informationen beinhalten [MENDES, 2002].
- Datenbanken, welche Daten über Metabolome und Genome integrieren, um Modelle des Stoffflusses zu erstellen [GOODACRE et al., 2004].

Die folgende Tabelle zeigt eine Gegenüberstellung der zurzeit wichtigsten Datenbanken, wobei die Datenbanken SMPDB, KEGG, HumanCyc und WikiPathways im Nachhinein noch detaillierter beschrieben werden, da sie vor allem im Bereich der menschlichen Metabolomik eine wichtige Rolle spielen.

Feature	SMPDB	KEGG	Reactome	HumanCyc	BioCarta	EHMN	WikiPathways
Number of metabolic pathways	70	73 (for humans)	64 (for humans)	317 (total), 238 (conf)	55	70	44 (for humans)
Number of disease pathways	113	35	3	0	24	0	4
Number of drug action pathways	168	0	3	3	10	0	4
Provides multiple organism pathways	No	Yes	Yes	No	Yes	No	Yes
Chemical structures shown in diagrams	Yes	No	Some	Yes (when zoomed)	Some	No	No
Protein 4° structures shown in diagrams	Yes	No	No	No	Some	No	No
Cell structures shown in pathway diagrams	Yes	No	No	No	Yes	No	No
Organs shown in pathway diagrams	Some	No	No	No	No	No	No
Descriptions of pathways provided	Yes	No	Yes	Yes	Yes	No	Some
Pathway images are hyperlinked	Yes	Yes	Limited	Yes	Yes	No	No
Pathway images are zoomable	Yes	No	No	Yes	No	Yes	Limited
Information provided on pathway entities	Detailed	Modest	Limited	Moderate	Limited	Limited	Limited
Supports simple text search	Yes	Yes	Yes	Yes	No	Yes	Yes
Supports advanced text search	Yes	No	Yes	Yes	No	No	No
Supports sequence searching	Yes	No	No	Yes	No	No	No
Supports graphical chemical structure search	Yes	No	No	No	No	No	No
Supports chemical expression mapping	Yes	No	Yes	Yes	No	No	No
Supports gene/protein expression mapping	Yes	No	Yes	Yes	No	No	No
Downloadable	Yes	Yes	Yes	Yes	No	No	Yes
Component lists available	Yes	No	Yes	No	Proteins only	Yes	Yes
BioPax, CellML or SBML compatible	No	Partial	Yes	Yes	No	Yes	No

Abbildung 5: Vergleich von Datenbanken [FROLKIS et al., 2010]

Einen Überblick über schon vorhandene Pathway-Datenbanken gibt zum Beispiel die Internetseite [www.pathguide.org](http://www.pathguide.org). [PICO et al., 2008]. Auf [www.pathguide.org](http://www.pathguide.org) finden sich Informationen über etwa 325 verschiedenen Datenbanken [PATHGUIDE, 2010].

## **13 Auflistung bereits vorhandener Datenbanken im Bereich Metabolomik**

### **13.1 Kyoto Encyclopedia of Genes and Genomes (KEGG)**

Kyoto Encyclopedia of Genes and Genomes (KEGG) ist eine umfangreiche, öffentlich zugängliche, täglich aktualisierte Datenbanksammlung, welche unter <http://www.genome.jp/kegg/> verfügbar ist und unter anderem eine PATHWAY Datenbank, eine GENES Datenbank und eine LIGAND Datenbank beinhaltet. Die PATHWAY Datenbank enthält Informationen über Stoffwechselwege, die GENES Datenbank gibt Auskunft über Gene und Proteine und die LIGAND Datenbank beinhaltet Informationen über chemische Verbindungen und Reaktionen.

Die Datenobjekte sind alle graphisch dargestellt und verschiedene Berechnungsmethoden stehen zur Verfügung, um bestimmte Merkmale der Graphiken mit den entsprechenden biologischen Bedeutungen in Verbindung zu bringen. Ziel der KEGG Datenbank ist es ein Grundlagenverständnis für die Beziehung zwischen genomischen und übergeordneten funktionellen Informationen zu erlangen. Daten aus Experimenten werden nicht als Textdateien wiedergegeben, sondern in Form von Graphiken, welche im Nachhinein bearbeitet werden können.

Die KEGG PATHWAY Datenbank enthält Referenzdiagramme molekularer Stoffwechselwege, welche zahlreiche zelluläre Prozesse umfassen, die einfach mit genomischen Informationen integriert werden können. Die Informationen in der KEGG PATHWAY Datenbank sind hierarchisch gegliedert. Die oberste Ebene ist unterteilt in Metabolismus, genetische und umweltbedingte Informationsverarbeitung, zelluläre Prozesse, menschliche Krankheiten und Arzneimittelentwicklung. Jede dieser Ebenen enthält mehrere Unterebenen, welche schließlich mit den Stoffwechseldiagrammen und/ oder orthologische Gruppentabellen (eine Ansammlung von Genen und Proteinen) verlinkt sind [KANEHISA et al., 2002].

Die erworbenen Kenntnisse über zelluläre Funktionen und Verhaltensweisen werden vor allem in Form von molekularen Netzwerken (KEGG pathway maps) und in hierarchischen Listen (BRITE functional hierarchies) dargestellt. Aktuell gilt das Hauptaugenmerk dem Aufbau der beiden Datenbanken KEGG DISEA-

SE und KEGG DRUG, welche Informationen über menschliche Krankheiten und Arzneimitteln beinhalten.

Mittlerweile stehen Benutzern 16 Hauptdatenbanken zur Verfügung, welche in die Kategorien Systeminformationen (KEGG PATHWAY, KEGG BRITE, KEGG MODULE, KEGG DISEASE, KEGG DRUG), genomische Informationen (KEGG ORTHOLOGY, KEGG GENOME, KEGG GENES, KEGG SSDB, KEGG DGENES, KEGG EGENES) und chemische Informationen (KEGG COMPOUND, KEGG GLYCAN, KEGG REACTION, KEGG RPAIR, KEGG ENZYME) unterteilt werden [KANEHISA et al., 2010].

### **13.2 Human Metabolome Database (HMDB)**

Die Human Metabolome Database (HMDB; online verfügbar unter [www.hmdb.ca](http://www.hmdb.ca)) ist eine frei zugängliche Datenbank. Sie umfasst eine umfangreiche Sammlung menschlicher Metabolite und Stoffwechseldaten, welche zahlreichen Wissenschaftlern, zum Beispiel Biochemikern, Ärzten oder auch Ernährungswissenschaftlern, bei ihrer Arbeit zur Verfügung stehen.

Die Informationen werden aus zahlreichen Büchern, Zeitschriften und elektronischen Datenbanken zusammengetragen, sowie durch experimentelle Daten verschiedener Metabolitkonzentrationen aus massenspektroskopischen und kernresonanzspektroskopischen Analysen, welche zum Beispiel aus Urin- oder Blutproben gewonnen werden, ergänzt. Außerdem enthalten sind NMR- und MS-Spektren von Referenz-Metaboliten.

Jeder in der HMDB eingetragene Metabolit wird umfassend beschrieben. So finden sich zum Beispiel Informationen über die Verbindungsbeschreibung, den Namen beziehungsweise etwaige Synonyme, strukturelle Informationen, physikalisch-chemische Daten, Stoffwechselwege sowie umfangreiche Links zu Bildern, Artikeln und anderen öffentlichen Datenbanken (zum Beispiel KEGG oder BioCyc) [WISHART et al., 2007].

Mittlerweile ist eine neuere Version der HMDB verfügbar, welche zahlreiche Erweiterungen im Bereich Inhalt, Website-Interface, Spektraldatenbank, Daten-

abfrage, als auch Datenbankkapazität aufweist, wie in der nachfolgenden Abbildung ersichtlich ist. Auch im Hinblick auf eine benutzerfreundlichere Handhabung wurden Verbesserungen vorgenommen, und mit Hilfe optimierter Suchmöglichkeiten ist eine einfachere Navigation durch die Human Metabolome Database möglich [WISHART et al., 2009].

Database feature or content status	HMDB (v 1.0)	HMDB (v 2.0)
Number of metabolites	2180	6826
Number of unique metabolite synonyms	27 700	43 882
Number of compounds with disease links	862	1002
Number of compounds with biofluid or tissue concentration data	883	4413
Number of compounds with chemical synthesis references	220	1647
Number of compounds with urine concentration data	231	472
Number of compounds with serum concentration data	174	3976
Number of compounds with cerebrospinal fluid concentration data	47	360
Number of compounds with experimental reference <sup>13</sup> C NMR spectra	380	784
Number of compounds with experimental reference <sup>1</sup> H NMR spectra	385	792
Number of compounds with predicted NMR spectra	1900	3044
Number of compounds with reference MS-MS spectra	390	799
Number of compounds with GC-MS reference data	0	279
Number of human-specific pathway maps	26	58
Number of compounds in Human Metabolome Library (HML)	607	920
Number of HMDB data fields	91	102
Pathway search/browse	No	Yes
Disease search/browse	No	Yes
Chemical class search/browse	No	Yes
Chemical substructure search	No	Yes
Biofluid search/sort tools	No	Yes
Advanced (multipeak or multicomponent) NMR search	No	Yes
Advanced (multipeak or multicomponent) MS-MS search	No	Yes
Advanced (retention index or MS peak) GC-MS search	No	Yes

Abbildung 6: Inhaltsvergleich von Version 1.0 und 2.0 der HMDB [WISHART et al., 2009]

### 13.3 The Small Molecule Pathway Database (SMPDB)

Die Small Molecule Pathway Database (SMPDB, verfügbar unter <http://www.smpdb.ca>) ist eine interaktive, visuelle Datenbank, welche mehr als 350 handgezeichnete Beschreibungen von menschlichen Stoffwechselwegen beinhaltet.

Der Grundgedanke, der zur Entwicklung dieser Datenbank geführt hat ist, dass die meisten anderen Stoffwechseldatenbanken entwickelt wurden ohne dabei auf die zunehmende Bedeutung des Themengebietes der Metabolomik zu achten. So bietet SMPDB eine fokussierte Ausrichtung auf menschliche Metabolite und enthält detaillierte Hyperlink-Diagramme des menschlichen Stoffwechsels, Stoffwechselerkrankungen und Signalwegen. Alle Stoffwechselwege enthalten Informationen über die betreffenden Organe, Organellen, subzellulären Kompartimente, Protein-Cofaktoren, Lage der Proteine oder Metaboliten und chemische Strukturen. Jeder Metabolit wird detailliert beschrieben (Name, physikalische und chemische Daten, NMR- und MS-Spektren, sowie Informationen über Konzentration und Lage) und zusätzlich mit Informationen aus der Human Metabolome Database (HMDB), DrugBank und UniProt verlinkt.

Die Datenbank ist sehr benutzerfreundlich aufgebaut und erleichtert so das Durchsuchen, wobei auch eine Volltextsuche, sowie die Suche nach Strukturen oder Sequenzen möglich sind. Alle in der Datenbank enthaltenen Informationen (PowerPoint-Bilder, Beschreibungen und Tabellen) können heruntergeladen und falls erwünscht auch noch nachträglich bearbeitet werden.

SMPDB bezeichnet sich als eine „Nischen-Datenbank“ welche vor allem Informationen über menschliche Metabolite auflisten möchte. Dadurch ergeben sich aber auch einige Einschränkungen, so zum Beispiel die Tatsache, dass keine Protein-Signalwege enthalten sind, wie zum Beispiel in der KEGG-Datenbank [FROLKIS et al., 2010].

### 13.4 WikiPathways

WikiPathways (online verfügbar unter [www.wikipathways.org](http://www.wikipathways.org)) entspricht dem bekannten Grundgedanken und -konzept, welches auch hinter Wikipedia steckt. Es bietet, auf Grund der freien Zugänglichkeit zu der Datenbank einer großen Gruppe von Forschern die Möglichkeit Daten zu sammeln und auszutauschen, was durchaus dem zunehmenden Trend eines weltweiten, möglichst barrierefreien Informationsaustausches entspricht. Aufgrund der einfachen Bedienbarkeit und dem barrierefreien Zugang zu WikiPathways soll eine große Gemeinschaft von Wissenschaftlern – von Studenten bis hin zu einschlägigen Experten – angeregt werden, ihr Wissen mit der Gesellschaft zu teilen und so gemeinsam Probleme zu erarbeiten, Schlüsse zu ziehen und neue Erkenntnisse zu gewinnen. Auf diese Weise soll eine umfassendere und aktuellere Datenbank aufgebaut werden, als es bei den bisherigen, teilweise sehr auf ein spezielles Themengebiet reduzierten, Datenbanken der Fall ist.

Jeder Stoffwechselweg wird bei WikiPathways auf einer eigenen Seite angezeigt, wobei sowohl ein Diagramm, eine Beschreibung, Literaturreferenzen, als auch beteiligte Gene, Proteine oder Metabolite angeführt werden.

Alle Änderungen, die vorgenommen werden, sind ebenfalls ersichtlich, wodurch es jederzeit möglich ist etwaige Fehler auszubessern oder Verbesserungen vorzunehmen [PICO et al., 2008]. Da es besonders wichtig ist Fehler schnell auszubessern gibt es die Möglichkeit diese fehlerhaften Daten automatisch aufzuspüren mit Hilfe sogenannter „bots“. Dabei handelt es sich um selbstständig arbeitende Computerprogramme, die ständig die Homepage überwachen und etwaige Fehler, zum Beispiel fehlende Literaturreferenzen, an einen Administrator melden [KELDER et al., 2009].

Durch Suchhilfen und die Möglichkeit nach Spezies oder Kategorie (zum Beispiel metabolische oder zelluläre Prozesse) zu suchen ist ein einfaches navigieren durch die Datenbank möglich.

Laut Pico et al. sind in Wikipathways 544 artspezifische Stoffwechselwege eingetragen, wobei unter anderem bereits 9 neue Stoffwechselwege beim Menschen erfasst wurden. Von mehr als 250 registrierten Usern sind ungefähr 10% am Aufbau von Wikipathways aktiv beteiligt [PICO et al., 2008].

### 13.5 HumanCyc

HumanCyc ist eine frei zugängliche Pathway/ Genom Datenbank, welche sich mit der Analyse des menschlichen Stoffwechsels beschäftigt und unter <http://humancyc.org/> abrufbar ist. 2709 menschliche Enzyme wurden bereits 896 Bioreaktionen zugeordnet und 622 dieser Enzyme scheinen im Zusammenhang mit 135 Stoffwechselwegen zu stehen. Des Weiteren lassen sich Übereinstimmungen dieser möglichen Stoffwechselwege hinsichtlich der ernährungsphysiologischen Anforderungen des Menschen ableiten.

Durch Analyse des menschlichen Genoms werden Enzyme ihren möglicherweise entsprechenden Stoffwechselwegen zugewiesen und so in biologischen Kontext gestellt, welcher es ermöglichen soll eine quantitative Modellierung des Stoffwechsels auszuarbeiten. HumanCyc bietet demnach eine genom-basierte Sicht auf die menschliche Ernährung, welche essentielle Ernährungsansprüche mit Stoffwechselwegen verknüpft. Die Zusammenführung von Daten des menschlichen Genoms mit klinischen, biochemischen, physiologischen und weiteren Daten soll zu einem tieferen Verständnis des menschlichen Stoffwechsels und den Auswirkungen von Ernährung auf Gesundheit und Krankheit beitragen. Wenn individuelle Genomsequenzen bekannt sind, eröffnet dies die Möglichkeit in Richtung optimale individuelle Ernährung des Menschen.

Mit Hilfe des Softwaretools Omics Viewer werden die Daten der Gen- und Proteinexpression sowie der Metabolomik als farbige Reaktionsschritte im Stoffwechselweg dargestellt [ROMERO et al., 2004].

### 13.6 Reactome

Reactome (<http://www.reactome.org>) ist eine frei zugängliche Datenbank, welche einen Überblick über biologische Prozesse – primär auf den Menschen bezogen – bietet.

Ausgehend von Reaktionen werden Datenmodelle von Stoffwechselwegen erstellt, welche diverse Prozesse, wie zum Beispiel intermediäre Stoffwechselwege, Regulationswege oder Signaltransduktion, darstellen. Die Startseite der Homepage zeigt eine graphische Übersicht über die Reaktionen, welche in der

Datenbank enthalten sind. Jede Reaktion ist als Pfeil dargestellt. Reaktionen, welche in kausalem oder temporärem Zusammenhang stehen, sind in Gruppen zusammengefasst, um den Benutzern eine einfache und schnelle Möglichkeit zu bieten die einzelnen Stoffwechselwege zu unterscheiden. Diese Art der Darstellung soll einerseits eine einfache Navigation durch die Inhalte der Datenbank ermöglichen, als auch als Data Mining Tool Verwendung finden.

Benutzer können die Datenbank entweder mit Hilfe des hierarchischen Inhaltsverzeichnis, diverser Suchfunktionen oder durch klicken auf oben erwähnte Übersicht über die Reaktionswege durchsuchen. Detaillierte Informationen zu den jeweiligen Stoffwechselwegen, wie beteiligte Reaktionen, Komplexe und Makromoleküle, sowie Verbindungen zwischen Stoffwechselwegen, liegen in Form von Textzusammenfassungen, handgezeichneten Diagrammen und computerlesbaren Informationen vor.

Reactome ist ähnlich organisiert wie ein Online-Journal, und gibt qualifizierte Autoren die Möglichkeit zum stetigen Aufbau der Datenbank beizutragen.

Reactome bietet dem Benutzer eine Reihe von Vorteilen, wie benutzerfreundliche Dateneingabe, Annotationen von erfahrener Biologen, Visualisierung und Exploration der Daten in interaktiven Stoffwechselkarten und die gleichzeitig Einblendung mehrere Datensätze.

Um Datenaustausch mit anderen Datenbanken, wie Cycs oder KEGG, zu erleichtern sind die Daten in Reactome in standardisierten Formaten wie BioPAX oder SBML verfügbar [JOSHI-TOPE et al., 2005].



## **14 Zukünftige Anwendungsbereiche für Metabolitdaten**

Die Forschung im Bereich Metabolomik ist in den letzten Jahren massiv vorangetrieben worden. Anwendungsbereiche finden sich zum Beispiel bei der Biomarkeridentifikation für Krankheiten, Arzneimittelinterventionen oder Umwelteinflüsse, Nutrigenomik und personalisierte Medizin, klinische Diagnostik, Studien über Wirkungsweisen, Metabolic Engineering und die Integration vorhandener Daten mit anderen Omiks-Daten [DUNN et al., 2005].

### **14.1 Verknüpfung der Datenbanken mit anderen “Omiks-Daten”**

Um Themenübergreifend zu arbeiten müssen Stoffwechseldatenbanken im Bereich der menschlichen Metabolomik in Zukunft auch mit Phänotyp-Datenbanken verlinkt werden. Dabei muss auf die vielfältigen Ernährungsbedingungen eingegangen werden.

Eine erste Auswahl an Metaboliten, welche für die Ernährungswissenschaft von Interesse sind, muss erstellt werden, welche mit Hilfe von NMR- und MS-Technologien identifiziert werden und in – wenn möglich frei zugängliche – Stoffwechseldatenbanken aufgenommen werden [GIBNEY et al., 2005]. Hierfür ist es jedoch auch nötig, dass die zurzeit vorhandenen Techniken im Bereich Metabolomik noch weiterentwickelt, benutzerfreundlicher und leistungsstärker werden [KOULMAN und VOLMER, 2008].

Des Weiteren ist beim Verlinken der Daten mit anderen (Omiks-)Daten darauf zu achten, dass gewisse Standards erarbeitet werden, um den Austausch zu erleichtern und wissenschaftliches Arbeiten zu ermöglichen.

Standardisierungsvorschläge, wie sie zum Beispiel von der SMRS Arbeitsgruppe erarbeitet werden, sind ein wichtiger Schritt in die richtige Richtung. Dabei muss auf die Standardisierung im Umgang mit Proben, Daten, Analyseverfahren, statistische und bioinformatische Methoden, als auch auf die Verwendung standardisierter Begriffe geachtet werden [LINDON et al., 2005].

## **14.2 Identifizierung von Biomarkern**

Die Verknüpfung von Daten aus der Ernährung und Metabolomik mit phänotypischen Daten wird es Wissenschaftlern ermöglichen es zu erforschen wie Gene mit Nahrungsmitteln interagieren. Die daraus gewonnenen Informationen können genutzt werden, um den individuellen Gesundheitszustand – bezogen auf das spezifische Genprofil – zu verbessern. Auch für die Lebensmittelindustrie werden sich dadurch neue Märkte eröffnen.

Erste Errungenschaften im Bereich der Metabolomik sind die Entdeckung und Anwendung von Biomarkern, welche physiologische Veränderungen oder pathologische Zustände aufzeigen. Aus der Sicht der Ernährungswissenschaft kann dieses Wissen dazu verwendet werden, um einen individuellen, optimalen Stoffwechsel zu erreichen beziehungsweise zu erhalten. Daher wird von der Systembiologie, Nutrigenomik und Metabolomik erhofft, dass sie den Weg ebnen in Bezug auf personalisierte Ernährung, welche anhand der individuellen genetischen Variation und deren Auswirkungen auf den Stoffwechsel Rückschlüsse auf individuelle Ernährung zulassen, um den Gesundheitszustand aufrechtzuerhalten und Krankheiten vorzubeugen [KOULMAN und VOLMER, 2008].

## **14.3 Individuelle Ernährung und individuelle Medizin**

Laut German et al. besteht die Rolle der Ernährungswissenschaft darin, Wissen über das menschliche Genom und Metabolom so zu nutzen, dass Wechselwirkungen zwischen Ernährung, Genotypen und Stoffwechsel nachvollziehbar sind und die Ernährung so variiert wird, dass der allgemeine, individuelle Gesundheitszustand verbessert wird – dies bedeutet einen Schritt in Richtung präventive Ernährung.

War es bisher nur möglich die Wirkungen einzelner Nährstoffe auf einzelne biologische Reaktionen, Enzyme, Gene, etc. zu untersuchen, besteht nun die Möglichkeit auf Grund modernerer molekularer Analysemethoden komplexere Zusammenhänge zu erforschen und so die Auswirkung von Umwelt, Toxinen und

Ernährung auf den gesamten Stoffwechsel zu untersuchen [GERMAN et al., 2002].

Vorweg ist es aber nötig noch mehr Grundlagenforschung zu betreiben. Sei es um akute und chronische Auswirkungen von Ernährung auf Metabolite in Körperflüssigkeiten festzustellen oder die Rolle der Darmflora noch genauer zu betrachten. Des Weiteren muss geklärt werden, welche Bedeutung den Begleitstoffe in der Nahrung zuzuordnen ist und festzustellen inwieweit sich Metabolite in Bezug auf unterschiedliche Ernährungsinterventionen verändern [GIBNEY et al., 2005].

Ein weiterer wichtiger Anwendungsbereich in der Nutrigenomik ist die Nutzung funktioneller Verbindungen – vor allem die Wirkung sekundärer Pflanzenstoffe [GARCIA-CANAS et al., 2010].

Hierzu adaptiveren Rezzi et al. ein graphisches Konzept, um im Speziellen die Zusammenhänge von ernährungsabhängiger Metabolomik und dem Gesundheits- und Risikomanagement darzulegen. Darin wird dargelegt, dass der individuelle Stoffwechsel von Genen, Umwelt, Lebensstil, Ernährung und der Darmflora beeinflusst wird. Die verschiedenen daraus resultierenden Stoffwechseltypen werden als blaue, rote und grüne Linie dargestellt, wobei die grüne Linie Schwankungen innerhalb der gesunden Bandbreite widerspiegelt. Ernährungsabhängige Metabolomik zielt dabei auf die Optimierung der Gesundheit und der Erhaltung der Homöostase ab (blaue Linie) [REZZI et al., 2007].

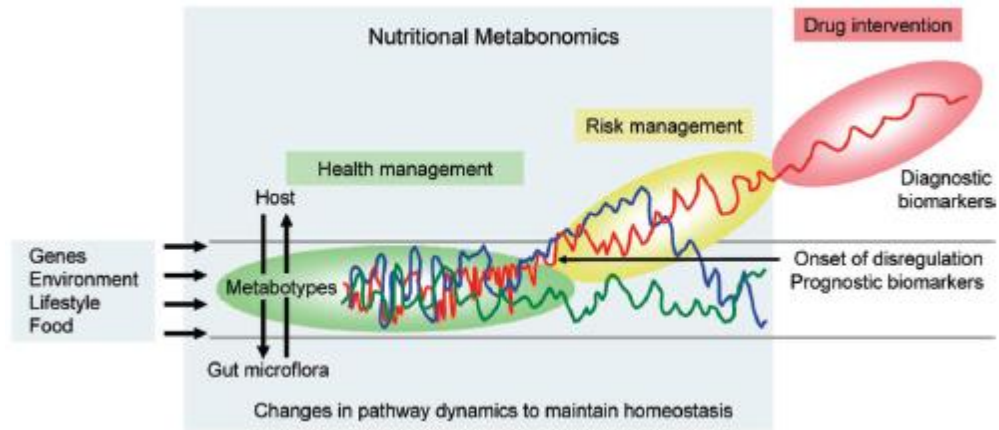


Abbildung 7: Konzeptualisierung von ernährungsabhängiger Metabolomik für Gesundheits- und Risikomanagement [REZZI et al., 2007]

## 15 Schlussbetrachtung

In dieser Arbeit wurde auf die bioinformatischen Methoden zur Erstellung und Auswertung des metabolischen Profils eingegangen. Dabei wurde anfangs auf die Problematik der sich teilweise überschneidenden Begriffsdefinitionen, sowie auf die noch nicht einheitlich festgelegten Standards im Bereich der Metabolomik hingewiesen.

Bei beiden Punkten ist anzumerken, dass hier möglichst bald einheitliche Richtlinien vorgegeben und eingehalten werden müssen. Dies erleichtert nicht nur Wissenschaftler die sich mit der Materie neu befassen den Einstieg in dieses Wissensgebiet, sondern ermöglicht auch eine einfachere Kommunikation zwischen Wissenschaftlern und eine bessere Nutzbarkeit der Daten.

Hinsichtlich der Definition einheitlicher Standards ist darauf zu achten, dass hierbei nicht nur einzelne Standards für einen bestimmten Bereich (zum Beispiel im Bereich chemische Analysen oder im Bereich Berichterstattung) erarbeitet werden. Standardisierungsrichtlinie wie zum Beispiel von Bino et al. oder von CASTLE et al. erarbeitet, welche sich umfassend mit der Berichterstattung von Experimenten im Bereich der Metabolomik befassen, sind zu bevorzugen und weisen den richtigen Weg in Richtung einheitliche Standardisierung [vergleiche BINO et al., 2004 und CASTLE et al., 2006].

Weitere wichtige Aspekte sind die Wahl des geeigneten Probenmaterials, sowie die Ermittlung der endogenen und exogenen Faktoren und deren Auswirkungen auf das metabolische Profil.

Im Bereich der Ernährungswissenschaft sind Blut, Urin und Speichel die drei wichtigsten zu untersuchenden menschlichen Körperflüssigkeiten in der Metabolomik. Wobei hier anzumerken ist, dass nicht invasiv entnommene Proben bevorzugt werden sollten. Grundlegend ist anzumerken, dass vor allem im Bereich der Ernährung noch zu wenige Experimente vorliegen, welche einen direkten Vergleich von Blut, Urin und Speichelproben liefern, und so die jeweiligen Vor- und Nachteile detailliert gegenüberstellen. Walsh et al. untersuchten die Effekte standardisierter Ernährung auf das metabolische Profil in Urin, Plasma und Speichel und konnten so Unterschiede zwischen den drei Körper-

flüssigkeiten aufzeigen und gleichzeitig die Auswirkung von endogenen und exogenen Faktoren auf die Ergebnisse minimieren [vergleiche WALSH et al., 2006].

Die Auswirkung endogener und exogener Faktoren (wie Alter, Geschlecht, Lebensstil, etc.) spielen vor allem im Bereich der Erforschung des menschlichen metabolischen Profils und im Hinblick auf die Vergleichbarkeit dieser Profile eine entscheidende und wichtige Rolle. Da es zu kostenintensiv und zeitaufwendig wäre, hier vor Beginn der jeweiligen Experimente möglichst einheitliche Ausgangssituationen zu schaffen, sollte möglicherweise die Überlegung angestellt werden, Personen mit möglichst ähnlichen körperlichen Voraussetzungen (Verwandte, Zwillinge) oder mit möglichst ähnlichen Umweltbedingungen oder Lebensstil (Lebensgemeinschaften, Internat, Krankenhaus) für Studien auszuwählen.

Hinsichtlich der Wahl der geeigneten Extraktionsverfahren und der Technologie zur chemischen Analyse spielen mehrere Faktoren, wie die Eigenschaften des zu untersuchenden Probenmaterials und die Fragen nach den zu erforschenden Metaboliten, eine Rolle. Zu wählen sind Techniken, bei denen möglichst wenige Metaboliten schon bei der Extraktion oder bei der Analyse zerstört werden. Des Weiteren sollte – je nach Fragestellung des Experiments – darauf geachtet werden, dass die gewonnenen Daten im Anschluss mit schon vorhandenen Daten abgeglichen werden können.

Zur Extraktion der Metabolite in Plasma oder Serum wird in vielen Studien Methanol verwendet [vergleiche BRUCE et al., 2009; A et al., 2005; Want et al., 2006; ZHANG et al., 2009].

Die beiden wichtigsten Analysemethoden im Bereich der Metabolomik sind die Massenspektrometrie und die Kernspinresonanzspektroskopie, deren jeweiligen Vor- und Nachteile in zahlreichen Publikationen aufgezählt werden [vergleiche zum Beispiel GOLDSMITH et al., 2010]. Die am häufigsten in der Metabolomik verwendeten Analysemethoden werden in der vorliegenden Arbeit, jeweils inklusive ihrer Vor- und Nachteile, aufgezählt.

Nach der chemischen Analyse müssen die enorm großen Datenmengen mit Hilfe statistischer und bioinformatischer Methoden analysiert werden. Die Rohdaten aus massenspektrometrischen und kernspinresonanzspektrometrischen Experimenten müssen vorweg vorverarbeitet werden. Anschließend werden die enorm großen Datenmengen der Metabolitdaten mittels statistischer Verfahren oder Algorithmen aus dem Bereich maschinelles Lernen oder evolutionäre Algorithmen analysiert. Dabei unterscheidet man zwischen supervised (überwachte) und unsupervised (unüberwachte) Methoden. Supervised Methoden benötigen eine bereits identifizierte Teilmenge, an Hand derer eine Kalibrierung vorgenommen werden kann, was den Vorteil mit sich bringt, dass die Einstufung auf einem gewissen Vorwissen basiert, wohingegen bei unsupervised Methoden keine zusätzlichen Kenntnisse über die Proben erforderlich sind.

Zahlreiche kommerzielle, sowie frei verfügbare Softwareprogramme zur Datenverarbeitung im Bereich der Metabolomik stehen Wissenschaftlern bereits zur Verfügung. Bei der Auswahl sollte – je nach finanzieller Verfügbarkeit – auch darauf geachtet werden, dass das Programm einen Gesamtüberblick von den Rohdaten bis hin zum Endergebnis liefert, um so eine bessere Nachvollziehbarkeit zu gewährleisten – dies ist vor allem bei kommerziellen Programmen der Fall.

Einen guten ersten Überblick über verfügbare Softwareprogramme liefern die beiden Internetseiten von MS-Utills (<http://www.ms-utils.org/>) und von Fiehn laboratory (<http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/Peak Alignment/>).

Die gewonnenen Daten können nun in Metabolit-Datenbanken eingetragen werden oder mit bereits bestehenden Daten verglichen werden. Dabei ist anzumerken, dass Metabolit-Datenbanken ganz bestimmte Funktionen erfüllen müssen. So sind zum Beispiel Informationen über Stoffkonzentrationen, physikalische und chemische Eigenschaften, detaillierte Beschreibung über Lage, assoziierte Krankheiten oder die Möglichkeit nach NMR- und MS-Spektren zu suchen wichtige Kriterien, welche eine gute, umfassende Metabolit-Datenbank erfüllen sollte.

Dabei gibt es allerdings auch noch einige Fragen – besonders im Bereich Spezies-spezifische Datenbanken – zu klären, zum Beispiel ob und wenn ja, welche weiteren Informationen – außer denen über die Metabolite – in der Datenbank aufscheinen sollten (zum Beispiel Daten über exogene Arzneistoffe, pflanzliche Nahrungsbestandteile, chemische Zusatzstoffe, etc.).

Einige der wichtigsten Datenbanken im Bereich der menschlichen Metabolomik werden in der vorliegenden Arbeit aufgezählt, wobei auf die unterschiedlichen Strukturen der Datenbanken eingegangen wird.

Abschließend erfolgt ein Ausblick auf zukünftige Anwendungsgebiete für die gewonnenen Erkenntnisse aus den Metabolitdaten – speziell auf den Bereich Ernährungswissenschaften bezogen. Hierbei wird auf die Themenbereiche Verknüpfung der Metabolit-Datenbanken mit anderen „Omiks“-Datenbanken, die Identifizierung von Biomarkern und individuelle Ernährung beziehungsweise individuelle Medizin. Vor allem bei dem Verlinken der gewonnenen Metabolit-Daten mit anderen Datenbanken, sowie bei der Identifizierung von Biomarkern spielt, wie schon oben erwähnt, die Frage nach einheitlichen Standards eine wichtige Rolle, denn nur so ist es möglich Daten miteinander zu vergleichen oder zu kombinieren und daraus logische Schlüsse zu ziehen. Besonders zukunftsweisend stellt sich jedoch der Bereich individuelle Ernährung dar, da sich hier die Möglichkeit einer präventiven Ernährungsweise eröffnet – individuell auf die Bedürfnisse jedes einzelnen abgestimmt.

Es ist zu erwähnen, dass bei allen drei angesprochenen Bereichen noch mehr Experimente im Bereich Ernährung erfolgen müssen, um weitere wichtige Erkenntnisse zu erlangen.



## 16 Zusammenfassung

Bioinformatische Methoden werden zur Auswertung von Daten aus Messungen des metabolischen Profils herangezogen. Die jeweiligen Verfahren und Methoden stehen – in diesem Gebiet – vielfach noch in einer sehr frühen Phase der Forschung, und vor allem Studien, den Bereich der Ernährungswissenschaften betreffend, sind eher selten.

Um hier einen Schritt in eine effektive und sinnvolle Richtung zu machen, sind sowohl eine internationale Vereinheitlichung von gebräuchlichem Fachvokabular nötig, als auch weitere Forschung in Bezug auf die unterschiedlichen Vor- und Nachteile des zu verwendeten Probenmaterials erforderlich. Auswirkungen exogener und endogener Effekte auf Probanden und somit auf das Probenmaterial spielen, wie bereits durchgeführte Experimente zeigen, eine nicht zu verachtende Rolle und müssen berücksichtigt werden.

Sowohl die Wahl zweckdienlicher Extraktionsverfahren, als auch der Methoden zur chemischen Analyse unterliegen gewissen Einschränkungen, die in Bezug auf die Auswertung der Daten durch bioinformatische Methoden berücksichtigt werden müssen. Auf die jeweiligen Vor- und Nachteile wird in der vorliegenden Arbeit ausführlich eingegangen. Ein Trend zeigt sich bei den Analysemethoden hin zur Massenspektrometrie und Kernspinresonanzspektroskopie.

Die Auswertung der enormen Datenmengen erfolgt mit Hilfe statistischer und bioinformatischer Methoden. Die Masse und Komplexität der Daten stellen die eingesetzten Softwareprogramme vor beachtliche Anforderungen, die gelöst werden müssen. Einen interessanten Ansatz in Bezug auf Informationsaustausch bietet WikiPathways, das Wissenschaftlern die Möglichkeit bietet, gemeinsam und barrierefrei Wissen auszutauschen und Erkenntnisse zu erlangen.

Handlungsbedarf besteht hinsichtlich der effizienten Weiterverarbeitung der gewonnenen Daten und ihre Einbettung in bereits bestehenden Strukturen (z. B. weitere „Omiks“-Datenbanken).

Interessante Ansätze in Bezug auf zukünftige Anwendungsgebiete – im Bereich der Ernährungswissenschaft – stellt sicherlich die Auswertung des metabolischen Profils im Hinblick auf präventive, individuelle Ernährung dar.

## 17 Abstract

Bioinformatical methods are used for analysis of data that is created from measurements of the metabolic profile. The respective methods and processes of this area are currently still in a very early stage of research. In particular, studies concerning the field of nutrition science are rather rare.

For the next step towards an effective and meaningful direction for the domain, both an international standardization of common terminology is required, as well as further research on the advantages and disadvantages of the use of different sample material. Studies have shown that exogenous and endogenous effects on test persons, and therefore on the sample material, have a significant impact and must be considered.

Both the choice of appropriate extraction methods and the methods of chemical analysis are subject to certain restrictions that must be considered in relation to the evaluation of the data using bioinformatical methods.

This thesis discusses the respective advantages and disadvantages of different methods in detail. A trend towards mass spectrometry and nuclear magnetic resonance spectroscopy is identified as the dominant analytical methods. The analysis of the enormous amounts of generated data is conducted using statistical and bioinformatical methods. The mass and complexity of the data leads to significant requirements towards the software tools that must be solved. An interesting approach to information exchange is taken by the project WikiPathways, which provides scientists with the possibility to exchange knowledge without barriers.

Action is required for the development of methods for efficient processing of the data and its integration into existing structures (for example, other "omics"-databases).

Interesting approaches to future applications – in the field of nutritional science – is certainly the evaluation of metabolic profile in respect to a preventive, individual diet.

## 18 Abbildungsverzeichnis

Abbildung 1: Arbeitsschritte zur Metabolitenerfassung in der Nutrigenomik [GARCIA-CANAS et al., 2010].....	6
Abbildung 2: Phänotyp beeinflussende Faktoren [ZEISEL et al., 2005].....	8
Abbildung 3: Vor- und Nachteile von Kernspinresonanzspektroskopie (NMR) und Massenspektrometrie (MS) [GOLDSMITH et al., 2010] .....	21
Abbildung 4: Arbeitsschritte bei der Datenverarbeitung von MS-Daten [KATAJAMAA und ORESIC, 2007] .....	29
Abbildung 5: Vergleich von Datenbanken [FROLKIS et al., 2010] .....	51
Abbildung 6: Inhaltsvergleich von Version 1.0 und 2.0 der HMDB [WISHART et al., 2009] .....	55
Abbildung 7: Konzeptualisierung von ernährungsabhängiger Metabolomik für Gesundheits- und Risikomanagement [REZZI et al., 2007] .....	62

## 19 Literaturverzeichnis

A J, TRYGG J, GULLBERG J, JOHANSSON A I, JONSSON P, ANTTI H, MARKLUND S L, MORITZ T. Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome. *Anal. Chem.*, 2005; 77: 8086-8094.

AB SCIEX. MarkerView™ Software 1.2.1 for Metabolomic and Biomarker Profiling Analysis. Aktuelle Version von 2010. [http://www.absciex.com/LITERATURE/cms\\_042026.pdf](http://www.absciex.com/LITERATURE/cms_042026.pdf) (bezogen am 26.09.2010).

ARITA M. Computational resources for metabolomics. *BRIEFINGS IN FUNCTIONAL GENOMICS AND PROTEOMICS*, 2004; 3: 84-93.

BECKONERT O, KEUN H C, EBBELS T M, BUNDY J, HOLMES E, LINDON J C, NICHOLSON J K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc.*, 2007; 2: 2692-703.

BENTON H P, WONG D M, TRAUGER S A, SIUZDAK G. XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization. *Anal Chem.*, 2008; 80: 6382-6389.

BERTRAM H C, EGGERS N, ELLER N. Potential of human saliva for nuclear magnetic resonance-based metabolomics and for health-related biomarker identification. *Anal Chem.*, 2009; 81: 9188-93.

BINO R J, HALL R D, FIEHN O, KOPKA J, SAITO K, DRAPER J, NIKOLAU B J, MENDES P, ROESSNER-TUNALI U, BEALE M H, TRETHERWEY R N, LANGE B M, WURTELE E S, SUMNER L W. Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.*, 2004; 9: 418-25.

BLUEGENOME. BlueFuse for biomarkers. Aktuelle Version von 2008. <http://cambridgebluegenome.com/pdfs/biomarker.pdf> (bezogen am 26.09.2010).

BRENNAN L. Personalised nutrition – Metabolomic applications in nutritional research. *Proceedings of the Nutrition Society*, 2008; 67: 404-408.

BROWN M, DUNN W B, ELLIS D I, GOODACRE R, HANDL J, KNOWLES J D, O'HAGAN S, SPASIC I, KELL D B. A metabolome pipeline: from concept to data to knowledge. *Metabolomics*, 2005; 1: 39-51.

BRUCE S J, JONSSON P, ANTTI H, CLOAREC O, TRYGG J, MARKLUND S L, MORITZ T. Evaluation of a protocol for metabolic profiling studies on human blood plasma by combined ultra-performance liquid chromatography/mass spectrometry: From extraction to data analysis. *Analytical Biochemistry*, 2008; 372: 237-249.

BRUCE S J, TAVAZZI I, PARISOD V, REZZI S, KOCHHAR S, GUY P A. Investigation of Human Blood Plasma Sample Preparation for Performing Metabolomics Using Ultrahigh Performance Liquid Chromatography/Mass Spectrometry. *Anal. Chem.*, 2009; 81: 3285-3296.

CASTLE A L, FIEHN O, KADDURAH-DAOUK R, LINDON J C. Metabolomics Standards Workshop and the development of international standards for reporting metabolomics experimental results. *BRIEFINGS IN BIOINFORMATICS*, 2006; VOL 7. NO 2. 159-165.

DETTMER K, ARONOV P A, HAMMOCK B D. MASS SPECTROMETRY-BASED METABOLOMICS. *Mass Spectrometry Reviews*, 2007; 26: 51-78.

DUNN W B, BAILEY N J C, JOHNSON H E. Measuring the metabolome: current analytical technologies. *Analyst*, 2005; 130: 606-625.

FIEHN O, KRISTAL B, VAN OMMEN B, SUMNER L W, SANSONE S, TAYLOR C, HARDY N, KADDURAH-DAOUK R. Establishing Reporting Standards for Metabolomic and Metabonomic Studies: A Call for Participation. *OMICS A Journal of Integrative Biology*, 2006; 10 (2): 158-163.

FROLKIS A, KNOX C, LIM E, JEWISON T, LAW V, HAU D D, LIU P, GAUTAM B, LY S, GOU A C, XIA J, LIANG Y, SHRIVASTAVA S, WISHART D S. SMPDB: The Small Molecule Pathway Database. *Nucleic Acids Research*, 2010; 38: D480-D487.

GARCIA-CANAS V, SIMO C, LEON C, CIFUENTES A. Advances in Nutrigenomics research: Novel and future analytical approaches to investigate the biological activity of natural compounds and food functions. *Journal of Pharmaceutical and Biomedical Analysis*, 2010; 51: 290-304.

GERMAN J B, BAUMAN D E, BURRIN D G, FAILLA M L, FREAKE H C, KING J C, KLEIN S, MILNER J A, PELTO G H, RASMUSSEN K M, ZEISEL S H. Metabolomics in the Opening Decade of the 21st Century: Building the Roads to Individualized Health. *J. Nutr.*, 2004; 134: 2729-2732.

GERMAN J B, ROBERTS M A, FAY L, WATKINS S M. Metabolomics and Individual Metabolic Assessment: The Next Great Challenge for Nutrition. *J. Nutr.*, 2002; 132: 2486-2487.

GIBNEY M J, WALSH M, BRENNAN L, ROCHE H M, GERMAN B, VAN OM-MEN B. Metabolomics in human nutrition: opportunities and challenges. *Am J Clin Nutr*, 2005; 82: 497-503.

GOLDSMITH P, FENTON H, MORRIS-STIFF G, AHMAD N, FISHER J, PRASAD K R. Metabonomics: A Useful Tool for the Future Surgeon. *Journal of Surgical Research*, 2010; 160: 122-132.

GOODACRE R, BROADHURST D, SMILDE A K, KRISTAL B S, BAKER J D, BEGER R, BESSANT C, CONNOR S, CAPUANI G, CRAIG A, EBBELS T, KELL D B, MANETTI C, NEWTON J, PATERNOSTRO G, SOMORJAI R, SJÖSTRÖM M, TRYGG J, WULFERT F. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*, 2007; 3: 231-241.

GOODACRE R, VAIDYANATHAN S, DUNN W B, HARRIGAN G G, KELL D B. Metabolomics by numbers: acquiring and understanding global metabolite data. *TRENDS in Biotechnology*, 2004; 22: 245-52.

HOLLAND N T, SMITH M T, ESKENAZI B, BASTAKI M. Biological sample collection and processing for molecular epidemiological studies. *Mutation Research*, 2003; 543: 217-234.

HOLLYWOOD K, BRISON D R, GOODACRE R. Metabolomics: Current technologies and future trends. *Proteomics*, 2006; 6: 4716-4723.

ISSAQ H J, VAN Q N, WAYBRIGHT T J, MUSCHIK G M, VEENSTRA T D. Analytical and statistical approaches to metabolomics research. *J. Sep. Sci.*, 2009; 32: 2183-2199.

JOSHI-TOPE G, GILLESPIE M, VASTRIK I, D'EUSTACHIO P, SCHMIDT E, DE BONO B, JASSAL B, GOPINATH G R, WU G R, MATTHEWS L, LEWIS S, BIRNEY E, STEIN L. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Research*, 2005; 33: D428-D432.

KANANI H, CHRYSANTHOPOULOS P K, KLAPA M I. Standardizing GC-MS metabolomics. *Journal of Chromatography B*, 2008; 871: 191-201.

KANEHISA M, GOTO S, FURUMICHI M, TANABE M, HIRAKAWA M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research*, 2010; 38: D355-D360.

KANEHISA M, GOTO S, KAWASHIMA S, NAKAYA A. The KEGG databases at GenomeNet. *Nucleic Acids Research*, 2002; 30: 42-46.

KASTENMÜLLER G, RÖMISCH-MARGL W, WÄGELE B, ALTMAIER E, SUHRE K. metaP-Server : A Web-Based Metabolomics Data Analysis Tool. *Journal of Biomedicine and Biotechnology*, 2011; Article ID 839862, 7 pages.

KATAJAMAA M, MIETTINEN J, ORESIC M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics*, 2006; 22: 634-636.

KATAJAMAA M, ORESIC M. Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography A*, 2007; 1158: 318-328.

KATZ J E, DUMLAO D S, CLARKE S HAU J. A new technique (COMSPARI) to facilitate the identification of minor compounds in complex mixtures by GC/MS and LC/MS: tools for the visualization of matched datasets. *Journal of the American Society for Mass Spectrometry*, 2004; 15: 580-584.

KEGG: KYOTO ENCYCLOPEDIA OF GENES AND GENOMES. KEGG Overview. Aktuelle Version vom 01.07.2010. <http://www.genome.jp/kegg/kegg1a.html> (bezogen am 10.09.2010).

KELDER T, PICO A R, HANSPERS K, VAN IERSEL M P, EVELO C, CONKLIN B R. Mining Biological Pathways Using WikiPathways Web Services. *PLoS One*, 2009; 4 (7): e6447: 1-4.

KEMSLEY E K, GALL G L, DAINTY J R, WATSON A D, HARVEY L J, TAPP H S, COLQUHOUN I J. Multivariate techniques and their application in nutrition: a metabolomics case study. *British Journal of Nutrition*, 2007; 98: 1-14.

KOCHHAR S, JACOBS D M, RAMADAN Z, BERRUEX F, FUERHOLZ A, FAY L B. Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabonomics. *Analytical Biochemistry*, 2006; 352: 274-281.

KOULMAN A, VOLMER D A. Perspectives for metabolomics in human nutrition: an overview. *British Nutrition Foundation Nutrition Bulletin*, 2008; 33: 324-330.

KUSSMANN M, REZZI S, DANIEL H. Profiling techniques in nutrition and health research. *Current Opinion in Biotechnology*, 2008; 19: 83-99.



LAW W S, HUANG P Y, ONG E S, ONG C N, LI S F, PASIKANTI K K, CHAN E C. Metabonomics investigation of human urine after ingestion of green tea with gas chromatography/mass spectrometry, liquid chromatography/mass spectrometry and <sup>1</sup>H NMR spectroscopy. *Rapid Commun Mass Spectrom.*, 2008; 22: 2436-46.

LENZ E M, BRIGHT J, WILSON I D, HUGHES A, MORRISSON J, LINDBERG H, LOCKTON A. Metabonomics, dietary influences and cultural differences: a <sup>1</sup>H NMR-based study of urine samples obtained from healthy British and Swedish subjects. *Journal of Pharmaceutical and Biomedical Analysis*, 2004; 36: 841-9.

LINDON J C, NICHOLSON J K, HOLMES E, KEUN H C, CRAIG A, PEARCE J T, BRUCE S J, HARDY N, SANSONE S A, ANTTI H, JONSSON P, DAYKIN C, NAVARANGE M, BEGER R D, VERHEIJ E R, AMBERG A, BAUNSGAARD D, CANTOR G H, LEHMAN-MCKEEMAN L, EARLL M, WOLD S, JOHANSSON E, HASELDEN J N, KRAMER K, THOMAS C, LINDBERG J, SCHUPPE-KOISTINEN I, WILSON I D, REILY M D, ROBERTSON D G, SENN H, KROTZKY A, KOCHHAR S, POWELL J, VAN DER OUDERAA F, PLUMB R, SCHAEFER H, SPRAUL M. Standard Metabolic Reporting Structures working group. *Nat Biotechnol.*, 2005; 23: 833-8.

LOMMEN A. MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing. *Anal. Chem.*, 2009; 81: 3079-3086.

MAHADEVAN S, SHAH S L, MARRIE T J, SLUPSKY C M. Analysis of Metabolomic Data Using Support Vector Machines. *Anal. Chem.*, 2008; 80: 7562-7570.

MENDES P. Emerging bioinformatics for the metabolome. *BRIEFINGS IN BIOINFORMATICS*, 2002; 3: 134-145.

MENDES P, KELL D B. On the analysis of the inverse problem of metabolic pathways using artificial neural networks. *BioSystems*, 1996; 38: 15-28.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. AMDIS — USER GUIDE. Aktuelle Version vom 15.09.2010. <http://chemdata.nist.gov/mass-spc/amdis/docs/amdis.pdf> (bezogen am 16.09.2010).

NIELSEN N-P V, CARSTENSEN J M, SMEDSGAARD J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *J. Chromatogr. A*, 1998; 805: 17-35.

ORESIC M. Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutrition, Metabolism & Cardiovascular Diseases*, 2009; 19: 816-824.

PATHGUIDE – THE PATHWAY RESOURCE LIST. Complete Listing of All Pathguide Resources. Aktuelle Version vom Mai 2010. <http://www.pathguide.org> (bezogen am 09.09.2010).

PENA-REYES C A, SIPPER M. Evolutionary computation in medicine: an overview. *Artif Intell Med.*, 2000; 19: 1-23.

PICO A R, KELDER T, VAN IERSEL M P, HANSPERS K, CONKLIN B R, EVELO C. WikiPathways: Pathway Editing for the People. *PLoS Biol.*, 2008; 6 (7): e148: 1403-1407.

PLUSKAL T, CASTILLO S, VILLAR-BRIONES A, ORESIC M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, 2010; 11: 395.

RAMAUTAR R, MAYBORODA O A, DEELDER A M, SOMSEN G W, DE JONG G J. Metabolic analysis of body fluids by capillary electrophoresis using noncovalently coated capillaries. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, 2008; 871: 370-4.

REZZI S, RAMADAN Z, FAY L B, KOCHHAR S. Nutritional Metabonomics: Applications and Perspectives. *Journal of Proteome Research*, 2007; 6: 513-525.

ROMERO P, WAGG J, GREEN M L, KAISER D, KRUMMENACKER M, KARP P D. Computational prediction of human metabolic pathways from the complete human genome. *Genome Biology*, 2004; 6: R2.1-R2.17.

SANA T R, WADDELL K, FISCHER S M. A sample extraction and chromatographic strategy for increasing LC/MS detection coverage of the erythrocyte metabolome. *Journal of Chromatography B*, 2008; 871: 314-321.

SCRIPPS CENTER FOR METABOLOMICS AND MASS SPECTROMETRY. XCMS Download. Aktuelle Version vom 26.09.2010. <http://metlin.scripps.edu/download/> (bezogen am 25.09.2010).

SHULAEV V. Metabolomics technology and bioinformatics. *Brief Bioinform.*, 2006; 7: 128-39.

SLUPSKY C M, RANKIN K N, WAGNER J, FU H, CHANG D, WELJIE A M, SAUDE E J, LIX B, ADAMKO D J, SHAH S, GREINER R, SYKES B D, MARRIE T J. Investigations of the Effects of Gender, Diurnal Variation, and Age in Human Urinary Metabolomic Profiles. *Anal Chem.*, 2007; 79: 6995-7004.

SMITH C A, WANT E J, O'MAILLE G, ABAGYAN R, SIUZDAK G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.*, 2006; 78: 779-787.

SOGA T, KAKAZU Y, ROBERT M, TOMITA M, NISHIOKA T. Qualitative and quantitative analysis of amino acids by capillary electrophoresis-electrospray ionization-tandem mass spectrometry. *Electrophoresis*, 2004; 25: 1964-1972.

SOLANKY K S, BAILEY N J, BECKWITH-HALL B M, DAVIS A, BINGHAM S, HOLMES E, NICHOLSON J K, CASSIDY A. Application of biofluid <sup>1</sup>H nuclear magnetic resonance-based metabonomic techniques for the analysis of the bio-

chemical effects of dietary isoflavones on human plasma profile. *Anal Biochem.*, 2003; 323: 197-204.

SOLANKY K S, BAILEY N J, BECKWITH-HALL B M, BINGHAM S, DAVIS A, HOLMES E, NICHOLSON J K, CASSIDY A. Biofluid  $^1\text{H}$  NMR-based metabolomic techniques in nutrition research – metabolic effects of dietary isoflavones in humans. *J Nutr Biochem.*, 2005; 16: 236-244.

STEIN S E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry*, 1999; 10: 770-781.

STELLA C, BECKWITH-HALL B, CLOAREC O, HOLMES E, LINDON J C, POWELL J, VAN DER OUDERAA F, BINGHAM S, CROSS A J, NICHOLSON J K. Susceptibility of human metabolic phenotypes to dietary modulation. *Journal of Proteome Research*, 2006; 5: 2780-2788.

STRELKOV S, VON ELSTERMANN M, SCHOMBURG D. Comprehensive analysis of metabolites in *Corynebacterium glutamicum* by gas chromatography/mass spectrometry. *Biol Chem*, 2004; 385 (9): 853-861.

SUMNER L W, AMBERG A, BARRETT D, BEGER R, BEALE M H, DAYKIN C, FAN T W M, FIEHN O, GOODACRE R, GRIFFIN J L, HANKEMEIER T, Hardy N, HIGASHI R, KOPKA J, LINDON J C, LANE A N, MARRIOTT P, NICHOLLS A W, REILLY M D, VIANT M. Proposed minimum reporting standards for chemical analysis. *Metabolomics*, 2007; 3: 211-221.

SYSI-AHO M, KATAJAMAA M, YETUKURI L, ORESIC M. Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinformatics*, 2007; 8: 93.

THISSEN U, WOPEREIS S, VAN DEN BERG S A, BOBELDIJK I, KLEEMANN R, KOOISTRA T, VAN DIJK K W, VAN OMMEN B, SMILDE A K. Improving the analysis of designed studies by combining statistical modelling with study design information. *BMC Bioinformatics*, 2009; 10: 52.

TIZINAI S, EMWAS A H, LODI A, LUDWIG C, BUNCE C M, VIANT M R, GÜNTHER U L. Optimized metabolite extraction from blood serum for  $^1\text{H}$  nuclear magnetic resonance spectroscopy. *Anal Biochem.*, 2008; 377: 16-23.

TORGRIP R J O, ABERG K M, ALM E, SCHUPPE\_KOISTINEN I, LINDBERG J. A note on normalization of biofluid  $^1\text{H}$ -NMR data. *Metabolomics*, 2008; 4: 114-121.

VAN DORSTEN F A, DAYKIN C A, MULDER T P J, VAN DUYNHOVEN J P M. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. *J Agr Food Chem*, 2006; 54: 6929-6938.

VAST SCIENTIFIC. Automated, label-free, semi-quantitative analysis of proteins, peptides, and metabolites based on comparisons of LC/MS and GC/MS data. Aktuelle Version vom Dezember 2009. [http://www.vastscientific.com/sieve/63141\\_SIEVE\\_BR\\_121610.pdf](http://www.vastscientific.com/sieve/63141_SIEVE_BR_121610.pdf) (bezogen am 26.09.2010).

WALSH M C, BRENNAN L, MALTHOUSE J P G, ROCHE H M, GIBNEY M J. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *Am J Clin Nutr*, 2006; 84: 531-9.

WANG Y, TANG H, NICHOLSON J K, HYLANDS P J, SAMPSON J, HOLMES E. A metabonomic strategy for the detection of the metabolic effects of chamomile (*Matricaria recutita* L.) ingestion. *J Agric Food Chem*, 2005; 53: 191-196.

WANT E J, O'MAILLE G, SMITH C A, BRANDON T R, URITBOONTHAI W, QIN C, TRAUGER S A, SIUZDAK G. Solvent-Dependent Metabolite Distribution, Clustering, and Protein Extraction for Serum Profiling with Mass Spectrometry. *Anal. Chem.*, 2006; 78: 743-752.

WATERS. THE MARKERLYNX APPLICATION MANAGER: INFORMATICS FOR MASS SPECTROMETRIC METABONOMIC DISCOVERY. Aktuelle Version von 2010.

<http://www.waters.com/webassets/cms/library/docs/720001056en.pdf> (bezogen am 26.09.2010).

WECKWERTH W, MORGENTHAL K. Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today*, 2005; 10 (22): 1551-1558.

WHITFIELD P D, GERMAN A J, NOBLE P-J M. Metabolomics: an emerging post-genomic tool for nutrition. *British Journal of Nutrition*, 2004; 92: 549-555.

WISHART D S. Current Progress in computational metabolomics. *Brief Bioinform.*, 2007; 8: 279-93.

WISHART D S. Quantitative metabolomics using NMR. *Trends in Analytical Chemistry*, 2008; 27: 228-237.

WISHART D S, KNOX C, GUO A C, EISNER R, YOUNG N, GAUTAM B, HAU D D, PSYCHOGIOS N, DONG E, BOUATRA S, MANDAL R, SINELNIKOV I, XIA J, JIA L, CRUZ J A, LIM E, SOBSEY C A, SHRIVASTAVA S, HUANG P, LIU P, FANG L, PENG J, FRADETTE R, CHENG D, TZUR D, CLEMENTS M, LEWIS A, SOUZA A D, ZUNIGA A, DAWE M, XIONG Y, CLIVE D, GREINER R, NAZYROVA A, SHAYKHUTDINOV R, LI L, VOGEL H J, FORSYTHE I. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Research*, 2009; 37: D603-D610.

WISHART D S, TZUR D, KNOX C, EISNER R, GUO A C, YOUNG N, CHENG D, JEWELL K, ARNDT D, SAWHNEY S, FUNG C, NIKOLAI L, LEWIS M, COUTOULY M A, FORSYTHE I, TANG P, SHRIVASTAVA S, JERONCIC K, STOTHARD P, AMEGBEY G, BLOCK D, HAU D D, WAGNER J, MINIACI J, CLEMENTS M, GEBREMEDHIN M, GUO N, ZHANG Y, DUGGAN G E, MACINNIS G D, WELJIE A M, DOWLATABADI R, BAMFORTH F, CLIVE D, GREINER R, LI L, MARRIE T, SYKES B D, VOGEL H J, QUERENGESSER L. HMDB: the Human Metabolome Database. *Nucleic Acids Research*, 2007; 35: D521-D526.

XIA J, PSYCHOGIOS N, YOUNG N, WISHART D S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research*, 2009; 37: W652-W660.

ZHANG Y, A J, WANG G, HUANG Q, YAN B, ZHA W, GU S, LIU L, REN H, REN M, SHENG L. Organic solvent extraction and metabonomic profiling of the metabolites in erythrocytes. *Journal of Chromatography B*, 2009; 877: 1751-1757.

ZEISEL S H, FREAKE H C, BAUMAN D E, BIER D M, BURRIN D G, GERMAN J B, KLEIN S, MARQUIS G S, MILNER J A, PELTO G H, RASMUSSEN K M. The nutritional phenotype in the age of metabolomics. *J Nutr.*, 2005; 135: 1613-1616.

**20 Lebenslauf****Angaben zur Person**

Name	Klug Barbara
Staatsangehörigkeit	Österreich
Geburtsdatum	20. Februar 1980

**Berufserfahrung**

Datum	Seit 01.12.2006 – laufend
Beruf oder Funktion	Freie Dienstnehmerin
Wichtigste Tätigkeiten und Zuständigkeiten	Marketing und Thea Kochbuchverkauf, Kundenkontakt, organisatorische Abwicklung, fachspezifische Beantwortung von Kundenanfragen
Name und Adresse des Arbeitgebers	Unilever Austria GmbH, Stella-Klein-Löw-Weg, 131023 Wien
Branche	Lebensmittelherstellung und -vertrieb

Datum	02.02.2009 – 04.03.2009
Beruf oder Funktion	Praktikum
Wichtigste Tätigkeiten und Zuständigkeiten	Dateneigabe im Rahmen des Projektes “nutrition-Day in Europe”
Name und Adresse des Arbeitgebers	Arbeitsgemeinschaft für klinische Ernährung, Höfergasse 13/1, 1090 Wien
Branche	Forschungseinrichtung

Datum	01.01.2007 – 31.01.2007
Beruf oder Funktion	Praktikum
Wichtigste Tätigkeiten und Zuständigkeiten	Administrative Tätigkeiten, schriftliche Beantwortung von Kundenfragen, Mitarbeit beim Relaunch der ÖGE-Homepage und der ÖGE online-Redaktion, Mitarbeit der Bibliotheksverwaltung, Literaturrecherchen



Name und Adresse des Arbeitgebers	Österreichische Gesellschaft für Ernährung, Spar- gelfeldstraße 191, 1220 Wien
Branche	Öffentlichkeitsarbeit
Datum	04.07.2005 – 12.08.2005
Beruf oder Funktion	Ferialpraktikum
Wichtigste Tätigkeiten und Zuständigkeiten	Urlaubsvertretung im Direktionssekretariat. Bearbei- tung und Aktualisierung von Kundendaten. Termin- vereinbarungen. Dokumentenarchivierung.
Name und Adresse des Arbeitgebers	BBRZ Linz, Grillparzerstraße 50, 4021 Linz
Branche	Berufsbildung
Datum	05.08.2002 – 30.08.2002
Beruf oder Funktion	Ferialpraktikum
Wichtigste Tätigkeiten und Zuständigkeiten	Urlaubsvertretung im Direktionssekretariat. Termin- vereinbarungen. Unterstützung bei der Vorbereitung und Durchführung einer Konferenz.
Name und Adresse des Arbeitgebers	BBRZ Linz, Grillparzerstraße 50, 4021 Linz
Branche	Berufsbildung
Datum	06.08.2001 – 31.08.2001
Beruf oder Funktion	Ferialpraktikum
Wichtigste Tätigkeiten und Zuständigkeiten	Aushilfe im Direktionssekretariat. Korrespondenz. Zusammenstellung von Unterlagen. Administrative Unterstützung beim Abschluss eines EU- geförderten Bildungsprojektes.
Name und Adresse des Arbeitgebers	BBRZ Linz, Grillparzerstraße 50, 4021 Linz
Branche	Berufsbildung

**Hochschulbildung**

Datum	Seit Oktober 1999
Studienrichtung	Ernährungswissenschaften
	Schwerpunkt: Ernährungsökonomie
Bildungseinrichtung	Universität Wien

Datum	Oktober 1998 – Juni 1999
Studienrichtung	Technische Physik
Bildungseinrichtung	Johannes Kepler Universität Linz

**Schulbildung**

1990 - 1998	Bundesrealgymnasium Peuerbach, Linz
1986 - 1990	Volksschule Harbach, Linz